Model Observers for Mass Lesion and Microcalcification Detection in Breast CT

By

SU HYUN LYU

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

John M. Boone, Chair

Elizabeth A. Morris

Guobao Wang

Committee in Charge

2023

This dissertation is dedicated to my parents

Acknowledgments

First and foremost, I would like to thank my Ph.D. advisor, Dr. John Boone, for his support and guidance over these past five years. It has been a privilege to learn from you, and I am a better scientist, writer, speaker, and person because of all that you have taught me. Thank you for helping me navigate the tumultuous years of COVID-19, for enabling me to do research through those challenges, for availing yourself to help me at any time, and for supporting my personal and career endeavors.

Thank you to Dr. Andrew Hernandez and Dr. Craig Abbey for your expertise and guidance. I am very grateful for all the time you spent with me in the details of the science. Thank you to my dissertation committee, Dr. Guobao Wang and Dr. Elizabeth Morris, and to my qualifying exam committee, Dr. Guobao Wang, Dr. J. Anthony Seibert, and Dr. Audrey Fan for your valuable feedback and support in my research studies. Thank you to Dr. Shadi Aminololama-Shakeri and Dr. Elizabeth Morris for their clinical expertise. Thank you to Dr. Amy Becker, Janani Aiyer, and George Burkett for helping me get acquainted in the lab and for extending kindness and friendship. Thank you to the members of the Physics and Engineering in Medical Imaging (PEMI) research group at UC Davis for the scientific discussions which helped me to learn a lot. Thank you to everyone in the UC Davis Department of Biomedical Engineering which has been my home for the past nine years. Thank you to the UC Davis Radiology clinical and administrative teams for their collaboration and support in these research projects.

Lastly, I am very grateful to my family. Thank you for your love and care that has enabled me to finish graduate school. I am deeply grateful to my husband, John, for his daily love and patience.

Abstract

Breast computed tomography (CT) is a relatively new breast imaging modality based upon cone-beam CT geometry, and there are ongoing efforts to optimize breast CT for the detection of breast cancer lesions. Clinical trials are considered the ideal approach when conducting protocol optimization studies. However, the time, expense, and extensive database required to conduct these studies pose significant challenges and costs. As an alternative to clinical studies, simulation studies have been proposed, where synthetic images are generated using phantom imaging or computer simulations. In addition, model observers have been proposed in lieu of human observers to evaluate the simulated images. By using simulated images and model observers, thousands of images can be generated and evaluated in a short period of time. In this work, we developed methods to simulate “hybrid” breast CT images, where mathematically generated mass lesions and microcalcifications are inserted into actual patient breast CT volume data sets. Then, we used model observers, namely the pre-whitened matched filter (PWMF) and convolutional neural networks (CNNs), to detect the simulated lesions.

First, we simulated contrast-enhanced mass lesions and explored the improvement in mass lesion detectability due to contrast enhancement across lesion diameter, section thickness, breast density, and view plane. An average 20% improvement was observed, and a larger improvement was observed for patients with dense breasts. Small lesions are generally harder to detect in dense breasts, but these results demonstrated that injected contrast can substantially improve detection performance in dense breasts.

Next, we compared the PWMF model observer with the CNN model observer for detecting mathematically generated unenhanced mass lesions inserted into 1) breast CT background and 2) Gaussian background, where the PWMF is known to be an ideal observer.

In Gaussian background, the CNN performed essentially identically to the PWMF across lesion sizes and section thicknesses. In breast CT background, the CNN outperformed the PWMF across lesion size, breast density, and most section thicknesses. These findings suggest that in breast CT images, CNNs capture more diagnostic information than PWMFs and may be a more pertinent observer when conducting optimal performance studies.

Lastly, we simulated microcalcifications and microcalcification clusters. The loss of intensity owing to partial volume effects was modeled and used to mathematically insert microcalcifications into acquired patient breast CT images. 2D and 3D CNNs were used to evaluate the detectability of simulated calcifications across clinical parameters. Our results demonstrated the utility of the maximum intensity projection (MIP) for displaying image volumes containing microcalcification clusters. We found that there was no statistically significant difference in detection performance when using the MIP compared to all slices in the native section thickness, but that thicker sections led to reduced detection performance. The MIP procedure essentially compresses 3D images to 2D images, resulting in efficient and better detection for microcalcifications.

Collectively, these studies elucidate the key factors affecting mass lesion and microcalcification detectability in unenhanced and contrast-enhanced breast CT and demonstrate the utility of model observers for examining breast CT images when human observers are unavailable. As breast CT advances towards translation to the clinic, these studies will be useful for optimizing breast CT protocols.

Table of Contents

Chapter 1: Introduction

1.1. Breast cancer and breast imaging

The American Cancer Society predicts that 1 in 8 women will develop breast cancer during her lifetime[1]. The five-year survival rate for an individual receiving diagnosis of localized breast cancer is 99%, whereas the five-year survival rate for an individual receiving diagnosis of metastatic breast cancer is 27%[1]. Clearly, early detection of breast cancer substantially decreases mortality.

Mammography was developed in the early 1980's as a screening tool for breast cancer and remains the standard-of-care screening technique for the normal risk population at many institutions[2]. In mammography, the breast is compressed between the compression paddle and detector, and projection images of the compressed breast are acquired at the mediolateral oblique angle and the craniocaudal angle. Other institutions such as UC Davis are transitioning to using digital breast tomosynthesis (DBT)[3] for screening and diagnostic imaging. DBT is another X-ray based imaging technique that produces a pseudo-3D image from projection images acquired over a limited angle range (15-60^{o}) around the breast. Synthetic mammograms (SM) can be reconstructed from DBT data and are often read by breast radiologists in conjunction with DBT images[4]. Multiple studies have demonstrated lower recall rates and higher cancer detection rates using DBT with SM[4,5,6]. The normal risk population is recommended to undergo screening mammography annually starting at age 40, and individuals who are at higher risk of breast cancer due to family history or genetic disposition are recommended to undergo biannual screening starting at age 20[7] via mammography[2], DBT[3], ultrasound[8], breast magnetic resonance imaging (bMRI)[9], or a combination of these modalities.

Breast cancer varies morphologically but is generally categorized into two categories: masses and microcalcifications. Mass lesions are developments of abnormal cell growth which tend to burgeon in milk ducts and glands[10]. They are typically oval, round, or irregular in shape, and can appear smooth and obscured at their margins or spiculated and more conspicuous at their margins[11]. Microcalcifications are small calcium deposits (<1 mm) that are frequently associated with premalignant or malignant lesions[12–14]. Because mass lesions and microcalcifications are commonly seen on mammograms, there are several well-known morphological patterns that help in identifying them and assessing their potential for malignancy. A summary of these patterns, reported by the American College of Radiology in the Breast Imaging Reporting and Data System (BI-RADS)[11], is shown in **Table 1.1**:

Table 1.1: Morphological features of mass lesions and microcalcifications from BI-RADS.

Mass lesions		
Shape	**Margin**	**Density**
Oval	Circumscribed	High density
Round	Obscured	Equal density
Irregular	Microlobulated	Low density
	Indistinct	Fat-containing
	Spiculated	

Microcalcifications		
Typically benign	**Suspicious morphology**	**Distribution**
Skin	Amorphous	Diffuse
Vascular	Coarse heterogenous	Regional
Coarse or “popcorn-like”	Fine pleomorphic	Grouped
Large rod-like	Fine linear or fine-linear branching	Linear
Round		Segmental
Rim		
Dystrophic		
Milk of calcium		

The breast is composed primarily of two tissues: adipose tissue, which appears dark on X-ray images, and fibroglandular tissue, which appears light on X-ray images. Breast lesions tend to form within fibroglandular tissue structures, specifically ductal or lobular cells, and also appear light[16]. The proportion of fibroglandular tissue compared to adipose tissue in the breast constitutes a patient's breast density, where a denser breast consists of more fibroglandular tissue. When dense breasts are imaged via mammography, fibroglandular tissue can obscure the visualization of tumors when they are superimposed onto a planar projection image. Malignant lesions can be obscured by overlapping normal tissue which can cause a false negative reading. Superimposition of normal tissue can cause a summation artifact resembling a mass, and this can cause a false positive reading. In 2023, the FDA updated U.S. mammography regulations to require that patients be notified of their breast density due to the influence of breast density on mammography accuracy[15].

1.2. Breast CT

To address the issue of superimposition of glandular tissue in projection imaging, our lab has focused on the development of breast computed tomography (CT), a true 3D X-ray imaging modality[16]. In breast CT, the patient lies prone with her breast inserted in a hole on the tabletop such that it is pendant and not compressed. Cone-beam projections are collected 360° around the breast and reconstructed using the Feldkamp algorithm[17] to produce a volume data set containing the entire breast. One notable advantage of breast CT is its ability to separate overlapping anatomical structures in the breast into individual tomographic images, enabling clearer visualization of anatomical structures inside the breast.

Over the past two decades our lab has developed four generations of breast CT scanners[18] and has performed several clinical trials enrolling hundreds of patients. The

prototype scanners are shown in **Figure 1.1**. Since 2005, our breast CT protocol has involved the intravenous injection of iodine-based contrast agent to enhance the conspicuity of lesions[4]. Contrast injection is particularly useful for the visualization of malignant tumors because malignant tumors uptake contrast agent more than benign tumors and thus appear brighter when imaged. Contrast injection is also used in breast MRI, but breast MRI utilizes a gadolinium-based contrast agent whereas breast CT utilizes an iodine-based contrast agent[4,14]. Contrast-enhanced breast CT and contrast-enhanced breast MRI are minimally invasive procedures, therefore, they may not be accepted as screening methods for the normal-risk population. However, these tools could be utilized for screening in high-risk populations and for diagnostic imaging, and they are currently being studied in those settings.

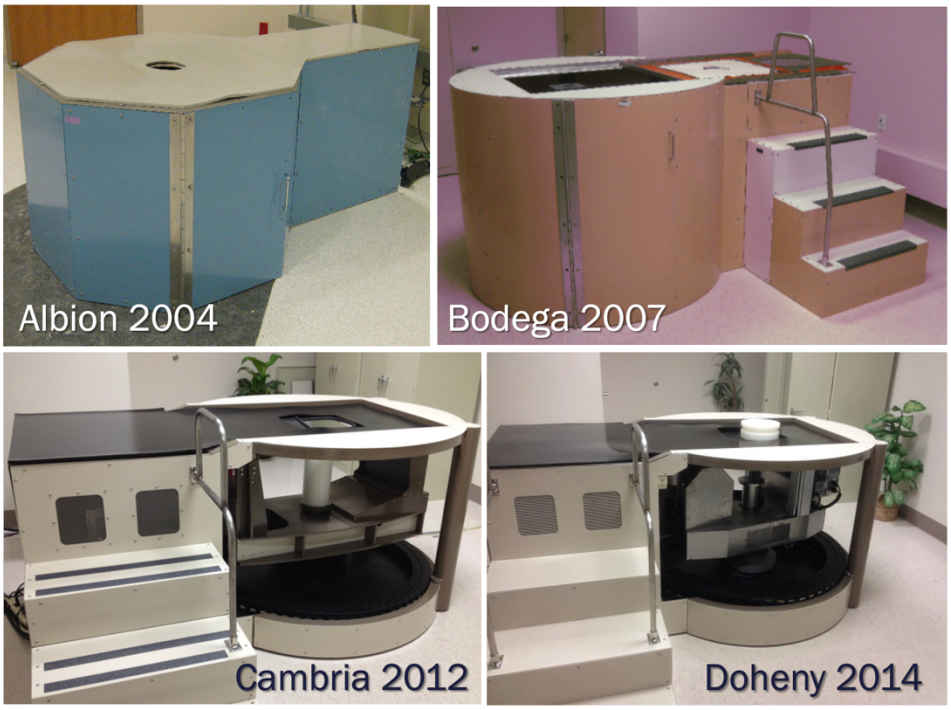

Figure 1.1: Four generations of breast CT scanner prototypes developed at UC Davis.

1.3. Simulation studies and model observers

There are ongoing efforts[19–21] to optimize breast CT for the detection of breast cancer lesions. Lesion detectability in breast CT is dependent on a host of factors. Lesion size, shape, and contrast contribute to detectability. Patient-specific factors such as breast density can affect detectability. Imaging protocols, including contrast injection protocol, image acquisition methods, image reconstruction methods, and image display methods can affect detectability. It is likely that the optimal parameters for detecting mass lesions are suboptimal for detecting microcalcifications since mass lesions are larger, low-contrast objects, while microcalcifications are smaller, high-contrast objects. Protocol optimization can be complex, seeing that it is crucial to detect both mass lesions and microcalcifications.

Clinical trials are considered the ideal approach when conducting protocol optimization studies. However, the time, expense, and extensive database required to adequately represent the screening population pose significant challenges and costs. In addition, radiation risks are introduced when patients are repeatedly imaged under varying protocols. Simulation studies have been proposed as an alternative to clinical studies, where synthetic images are generated using phantom imaging or computer simulations. These studies are also called "virtual clinical trials"[22–24]. This dissertation will report on simulations of hybrid images, where mathematically generated mass lesions and microcalcifications are inserted into actual patient breast CT volume data sets.

Furthermore, model observers are used in lieu of human observers to evaluate lesion detectability in the hybrid images. Model observers refer to mathematical or computational techniques used to assess the performance of imaging systems[25,26]. They provide quantitative and standardized ways to evaluate imaging systems. Generally, model observers are used for two main purposes: 1) to evaluate *ideal observer* performance, where all

statistical information available in the image is used to determine maximum task performance[27], and 2) to predict *human observer* performance, often through channelized observers, which mimic the human visual system's processing of image information through channels or frequencies[28,29]. Model observers have shown to reasonably predict human observer performance in breast CT[30,31] and other breast imaging modalities[32–34]. By using simulated images and model observers, thousands of images can be generated and evaluated in a short period of time.

1.4. Dissertation outline

This dissertation will report on key factors that contribute to mass lesion and microcalcification detectability in breast CT using simulated images and model observers. Chapter 2 will describe simulation methods for contrast-enhanced mass lesions, which will be evaluated using a pre-whitened matched filter (PWMF) mathematical model observer. Chapter 3 will compare the PWMF with a convolutional neural network (CNN) model observer for the detection of simulated lesions in breast CT and Gaussian background. Chapter 4 will describe simulation methods for microcalcifications and microcalcification clusters, which will be evaluated using a CNN model observer. Finally, Chapter 5 will conclude by providing a summary of the results and a discussion of future directions.

Chapter 2: Model observer performance in contrast-enhanced lesions in breast CT: The influence of contrast concentration on detectability

2.1. Introduction

The use of iodine-based contrast agent for better delineation of tumors in breast computed tomography (CT) has been shown to be compelling[35], similar to the tumor enhancement in contrast-enhanced breast MRI (CE-bMRI)[9] and contrast-enhanced dual-energy digital mammography (CE-DM)[36]. Indeed, the use of intravenous contrast injection in whole body CT applications also adds important diagnostic information to the CT examination. While different organs have different vascular perfusion characteristics which impact the timing of and peak enhancement levels for that organ[37], in the case of solid tumors the "leaky vessel" phenomena resulting from angiogenesis leads to the "wash in" and "wash out" of iodinated contrast agent in the interstitial space surrounding the tumor[38]. This has been called the enhanced permeability and retention (EPR)[39] effect. CE-bMRI and CE-DM also capitalize on the EPR effect, although CE-DM is a 2D imaging modality.

Contrast-enhanced breast CT is a relatively new tool, and a structured evaluation of different imaging parameters at play has yet to be conducted. This study aims to examine the independent and co-dependent effects of contrast, breast density, lesion size, view plane, and section thickness on contrast-enhanced lesion detectability, building from previous work on unenhanced lesion detectability[40]. While a clinical trial is the ideal approach for this kind of study, the time, expense, and large database needed to adequately represent the screening population makes this approach challenging and costly. A simulation study is used here in lieu of a clinical study, and we have chosen a hybrid simulation method, in which mathematically generated lesions are inserted into actual patient CT images. This method

has been implemented with mammographic backgrounds[32,41,42] and breast CT background[40]. Moreover, a model observer is proposed in lieu of a human observer to evaluate lesion detectability in the simulated images. Mathematical model observers have shown to reasonably predict human observer performance in breast CT[30,31] as well as in other breast imaging modalities[32–34].

In this investigation, data sets of acquired breast CT images from a cohort of patients imaged at our institution were used in concert with simulated mathematically inserted spherical lesions to study the role of contrast injection. To compare performance with previous work, unenhanced simulated lesions were inserted into the acquired breast CT image data using a previously-reported algorithm, and results of a pre-whitened matched filter model observer were compared to previously reported work[40].

The lesion enhancement arising from contrast-enhanced breast CT images in 22 patients was evaluated, and the contrast enhancement (average increase in Hounsfield Unit, ΔHU) was characterized. These clinical results were used to compute contrast levels for the simulated lesions, over a range from 0% to 100% contrast enhancement levels. More complex methods have been proposed to simulate clinically-relevant iodinated lesions involving heterogenous enhancement patterns and rim enhancement[43]. Our model does not attempt to simulate intra-lesion heterogeneity, but this simplified approach may be useful in evaluating more general effects of contrast across a range of clinical parameters. The pre-whitened matched filter was used to quantify detection performance using the area under the receiver operating characteristic curve (AUC) as a performance metric. The improvement in detection due to contrast injection was quantified across breast density, lesion size, view plane, and section thickness.

2.2. Methods

2.2.1. Breast CT system and imaging technique

Prototype breast CT systems built in our laboratory at UC Davis were used to acquire breast CT data sets from patients in an IRB-approved study. The laboratory has developed four generations of breast CT scanners as described in Ghazi *et al.*[18]. Image data sets from the first- and second-generation breast CT scanners, which are very similar in design, were used in this study and are briefly described here.

A tungsten anode x-ray tube, collimator, and flat-panel x-ray detector were integrated into a cone-beam breast CT scanner and powered by an integrated bearing-motor-encoder system. The gantry rotates in the horizontal plane during acquisition, using an 80 kV x-ray beam with 0.2 mm copper filtration. The flat-panel detector (Varian 4030CB; Varian, Palo Alto, CA) was operated in 2 × 2 binning mode, with a native dexel size of 0.194 mm. The x-ray tube and detector rotated ≥ 360 degrees around the breast to capture 500 cone-beam projection images in about 16 seconds. Image acquisition protocols were not altered for contrast imaging.

A variation of the Feldkamp algorithm[17] was used to reconstruct the projection images. The projection data were reconstructed into a 512 × 512 matrix with an isotropic voxel size of 0.4 mm. Low-frequency cupping artifacts were corrected using a previously-reported polynomial-fitting algorithm[44]. A previously-reported segmentation algorithm[45] was then used to label each voxel in each breast CT image as either air, skin, adipose tissue, dense fibroglandular tissue, or sparse fibroglandular tissue, producing segmentation volumes which are perfectly aligned with the gray-scale images.

2.2.2. Patient study

A total of 322 women were imaged on our first- and second-generation breast CT scanners as part of several IRB-approved clinical trials. Patients receiving a score of BIRADS 4 or 5 based on their diagnostic breast examination were eligible to enroll in our studies. These patients eventually underwent breast biopsy, and the resulting pathology report was considered ground truth (benign or malignant). In our research imaging protocol, the breast CT scan took place just prior to the breast biopsy. For research scans involving contrast, pre- and post-contrast scans were acquired on both breasts. 100 mL of contrast agent (Omnipaque-350; GE Healthcare, Waukesha, WI)[46] was intravenously injected at a rate of 4 mL/sec using a power-injector. Post-contrast scanning in the affected breast began on average 90 seconds after the start of injection.

Most patients received four breast CT scans according to contrast imaging protocol, but patients electing not to receive contrast and/or contralateral imaging received one to three scans. Of 913 total scans, 253 pre-contrast scans that were artifact-free, contained the breast in the field of view, and were void of breast implants were selected for this study.

2.2.3. Relationship between iodine concentration and Hounsfield units

Our study depends upon the assumption that HU and iodine concentration are linearly related[47,48]. This relationship was measured using spectral modeling. The breast CT system used to acquire patient data used an 80 kV tube potential with a 0.2 mm copper filter as described in **Section 2.2.1**. This spectrum was modeled using the tungsten anode spectral model using interpolating cubic splines (TASMICS) model[49], and a thickness of 140 mm of soft tissue was added to the beam to simulate the attenuation (and beam hardening) of the median breast. The attenuation of 10 mm added thickness of water with various

concentrations of iodine (0, 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 mg/mL) was simulated, allowing the estimation of the HU for each iodine concentration. Linear regression was used to compute the relationship between iodine concentration [I] and HU. An R^2 value of 1.000 was found. The spectral model demonstrated that [I] (mg/mL) is linearly related to HU. The linear relationship is given in **Equation 2.1**:

$$HU = 5.246 \times [I] \tag{2.1}$$

2.2.4. Simulation of lesions

2.2.4.1. Overview: Generation of hybrid lesion images

The methodology for generating hybrid lesion images was described by Packard *et al.*[40] and is summarized here. Mathematically generated lesions were inserted at random locations in patient breast CT images with the lesion centers (x, y, z) selected using a random number generator. A check was performed on the surrounding $64 \times 64 \times 64$ voxels to ensure that the inserted lesion was fully contained within the breast parenchyma (adipose and fibroglandular tissue). Synthetic unenhanced lesions were inserted into these subvolumes. The resulting volume with the synthetic lesion, $f_{sim}[i,j,k]$, is shown in **Equation 2.2**:

$$f_{sim}[i,j,k] = f[i,j,k] + \left(\Delta I \times M(d_{TB}[i,j,k]) \times M\left(\frac{1}{2}D - d_{LC}[i,j,k]\right)\right) \tag{2.2}$$

where $f[i,j,k]$ is the breast CT volume and ΔI is the average difference in HU between adipose and fibroglandular tissue in the breast CT image. $M(d_{TB}[i,j,k])$ is the tissue boundary modulation term, ranging from 0-1, which modulates based on $d_{TB}[i,j,k]$, the distance to the nearest tissue boundary from voxel $[i,j,k]$. D is the lesion diameter, $d_{LC}[i,j,k]$ is the distance to the lesion center from voxel $[i,j,k]$, and $M(\frac{1}{2}D - d_{LC}[i,j,k])$ is the lesion

boundary modulation term, also ranging from 0 - 1, which blurs the lesion edges based on the distance to the nearest lesion boundary from voxel $[i, j, k]$. Both modulation functions are determined on a patient data set basis by mathematically modeling the edge roll-off at adipose/fibroglandular tissue boundaries to ensure that the resolution of the added lesion matches the resolution of the background breast CT image.

After a lesion is inserted into the 3D volume, a $64 \times 64 \times 1$ image of the lesion is generated by either extracting the center slice of the 3D volume for 0.4 mm slice thickness or by slice averaging across multiple adjacent slices.

2.2.4.2. Synthetic contrast enhancement

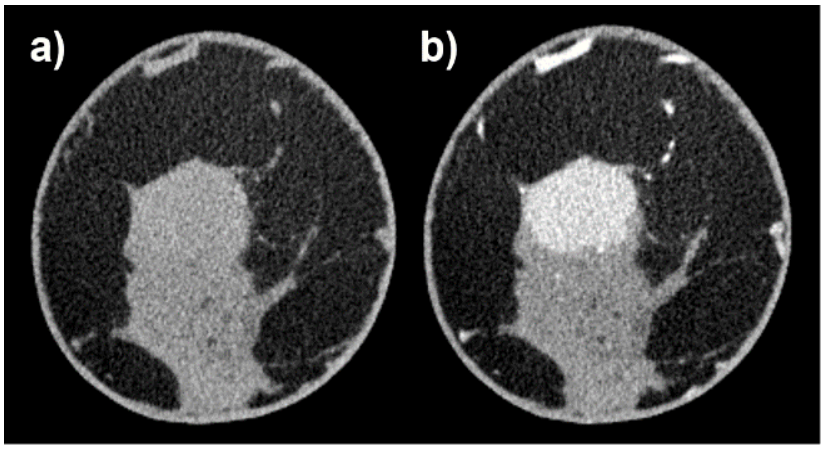

Figure 2.1: Pre-contrast (a) and post-contrast (b) coronal breast CT images of invasive mammary carcinoma. Local enhancement is visible in the post-contrast image.

It has been observed in unenhanced breast CT images that when a lesion infiltrates a region of fibroglandular tissue, it does not change the attenuation coefficient of the native fibroglandular region[40,50]. However, when the unenhanced lesion extends into a region of adipose tissue, it increases the attenuation coefficient of that region to essentially match the

intensity of the native fibroglandular tissue. In iodinated contrast imaging, the same process is observed, but the attenuation coefficients of both fibroglandular and adipose regions are further increased to the same value such that the entire lesion is enhanced. **Figure 2.1** displays pre- and post-contrast images for one patient from our study with an enhancing invasive mammary carcinoma, where these phenomena can be observed. These observations inform our methodology for simulating contrast enhancement: after unenhanced lesions are mathematically inserted into breast subvolumes, each voxel within the spherical lesion boundary is enhanced to simulate contrast uptake. While there are BI-RADS lexicons describing enhancing lesions in terms of shape, spiculation, and enhancement pattern, our simple addition of a spherical, uniformly enhancing lesion does not attempt to simulate these clinical features.

The extent to which each lesion is enhanced was investigated. Leaky vasculature resulting from angiogenesis is known to be the primary cause of rapid local enhancement in malignant tumors. We hypothesized that enhancement could also be patient-dependent, and examined potential relationships between contrast enhancement and patient-specific factors. 22 patients from our data set who had biopsy-proven malignant mass lesions and received pre-contrast and post-contrast imaging were studied. For each lesion, contrast enhancement was quantified by computing the increase in HU (ΔHU) in the lesion between the post-contrast and pre-contrast images. The effects of tumor volume, differential glandular intensity (ΔI), and contrast delay, i.e., the time between the start of contrast injection until post-injection scanning, on ΔHU were investigated. In addition, the effect of patient size in terms of body mass index (BMI) was studied to determine potential dependencies on these parameters.

No notable relationships were found between ΔHU and tumor volume (R^2 = 0.0253), ΔI (R^2 = 0.163), contrast delay, or BMI (R^2 = 0.0406). It was concluded that ΔHU is primarily

related to the extent of angiogenesis and the permeability of angiogenetic microvessels in any given tumor. With this observation, the method for enhancing lesions was simplified such that each voxel within the simulated lesion would increase by $\overline{\Delta HU}$, the mean enhancement observed in the 22-patient lesion cohort. The mean enhancement $\overline{\Delta HU}$ was found to be 76.0, and the standard deviation σ was 33.5. The equivalent iodine concentrations based on **Equation 2.1** are 14.5 mg/cm^3 and 6.39 mg/cm^3, respectively.

The resulting equation for simulating contrast-enhanced lesions is shown in **Equation 2.3**:

$$f_{sim}[i,j,k] = f[i,j,k] + \left((\Delta I \times M(d_{TB}[i,j,k])) + (\alpha \times \overline{\Delta HU})\right) \times M\left({}^{D}/_{2} - d_{LC}[i,j,k]\right) \quad (2.3)$$

where $\overline{\Delta HU} = 76.0$, the mean enhancement observed in the clinical data set of malignant lesions and α is a scaling term to study the effect of different amounts of contrast on lesion detectability. Five different values of α were studied: 0, 0.25, 0.50, 0.75, 1. The corresponding HUs are 0, 19.0, 38.0, 57.0, and 76.0, respectively, and the equivalent iodine concentrations in mg/cm^3 are: 0.0, 3.6, 7.2, 10.9, and 14.5, respectively. In addition to contrast level, the effects of lesion diameter, section thickness, breast density, and view plane on lesion detectability were also studied. Six different lesion diameters (1, 3, 5, 9, 11, 15 mm) and six different section thicknesses (0.4, 1.2, 1.9, 3.5, 5.8, 19.8 mm) were used. The selected lesion diameters represent the size range of mass lesions found over 322 clinical breast CT images. Microcalcifications were not simulated for this study. Data sets spanning a broad range of known breast densities were used to address the role of breast density.

Breast density was quantified in terms of the volumetric glandular fraction (VGF), which is defined as:

$$VGF = \frac{n_g}{n_g + n_a} \tag{2.4}$$

where n_g represents the number of fibroglandular voxels and n_a represents the number of adipose voxels in the segmentation volume of the breast. VGF was computed for each patient and then categorized into one of six bins (0-10%, 11-20%, 21-30%, 31-40%, 41-50%, >50%). 2D images were generated from the coronal and axial view planes to study the role of view plane. The range of parameters studied herein is summarized in **Table 2.1**.

Table 2.1: Range of each parameter studied: alpha (equivalent iodine concentration listed in italics), lesion diameter, section thickness, and view plane.

Alpha	*Equivalent [I] (mg/cm³)*	Lesion Diameter (mm)	Section Thickness (mm)	Plane
0.00	*0.0*	1	0.4	Coronal
0.25	*3.6*	3	1.2	Axial
0.50	*7.2*	5	1.9	
0.75	*10.9*	9	3.5	
1.00	*14.5*	11	5.8	
		15	19.8	

Examples of inserted non-contrast (α = 0) and contrast-enhanced (α = 1) lesions are shown in **Figure 2.2.** In the added lesion columns, a black voxel signifies 0 HU was added to that region, a gray voxel signifies ΔI HU was added to that region, and a white voxel signifies $\Delta I + (\alpha \times \overline{\Delta HU})$ was added to that region. As described in **Equation 2.3**, each computed voxel was additionally modulated by M depending on its distance to the nearest adipose/fibroglandular tissue boundary and to the nearest lesion boundary to ensure smoothness of edges.

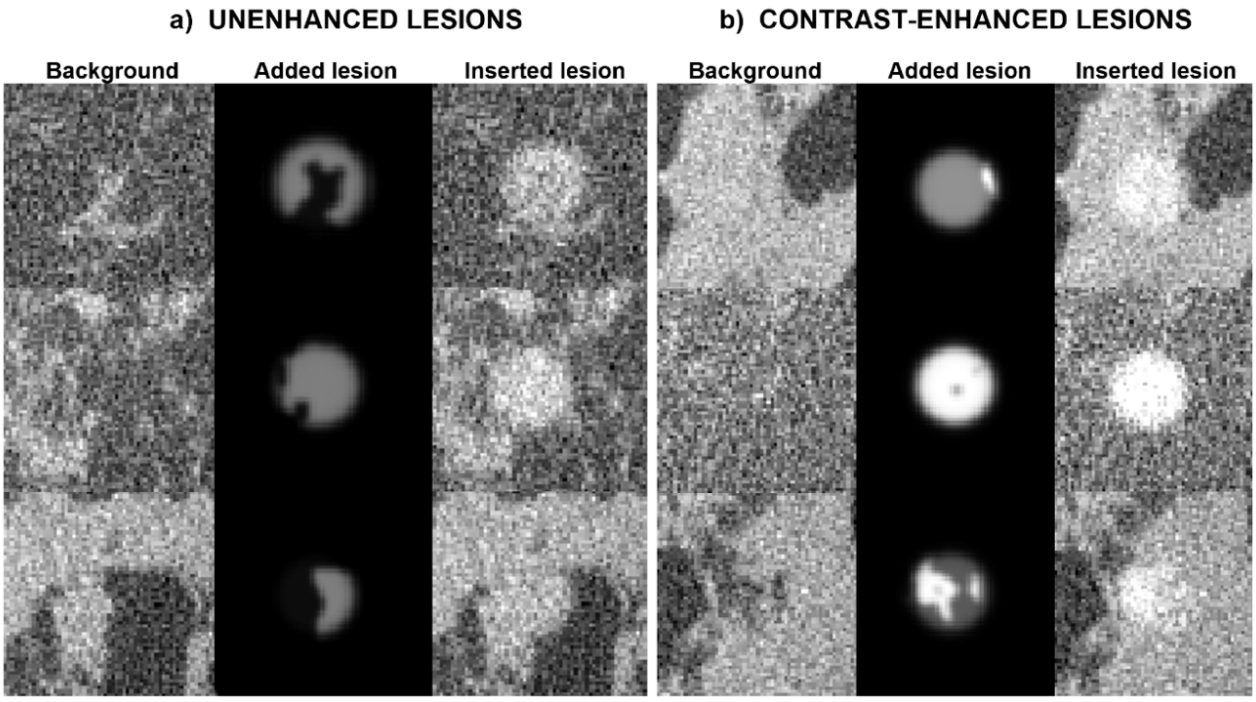

Figure 2.2: Example image patches of (a) simulated unenhanced lesions in breast background and (b) simulated contrast-enhanced lesions in breast background. In unenhanced lesions, intensity is only added to adipose regions. In contrast-enhanced lesions, intensity is added to both adipose and fibroglandular regions, with more intensity being added to adipose regions.

2.2.5. Model observer for detectability evaluation

In the ideal scenario, human observers (e.g., expert breast imaging radiologists) would be used evaluate the detectability of simulated lesions. Given that we study the independent and codependent effects of the parameters listed in **Table 2.1** across 253 patient data sets, at least 91,080 hybrid images would need to be evaluated, making a human observer study infeasible. Accordingly, a pre-whitened matched filter (PWMF)[32,51], a mathematical observer, was used to evaluate the detectability of simulated lesions in a signal known exactly (SKE), location known exactly (LKE) detection study. The PWMF is considered the ideal observer under the assumption of a stationary Gaussian image background and has been shown to outperform human observers in breast CT images[52]. In an SKE/LKE task, the shape of the signal is known, and the location of the lesion is known exactly, and the PWMF makes use of the signal profile and image power spectrum to compute a decision variable which is used to evaluate detectability. Packard *et al*. also utilized a PWMF in lieu of human observers to evaluate detectability of simulated *unenhanced* lesions across clinically relevant parameters in previous work[40]. While the PWMF may not be perfectly accurate with respect to human observers, its *relative* performance is considered reliable in this comparative analysis across clinically relevant parameters.

To implement the PWMF, the mean signal profile and the power spectrum of the image background must first be computed. For each breast CT volume data set, for a given combination of contrast level, lesion diameter, section thickness, and view plane, *N* viable lesion centers are first found, which are used to compute *N* $64 \times 64 \times 1$ lesion-present patches and *N* $64 \times 64 \times 1$ lesion-absent patches. A patch generated with an added lesion is denoted as $I_n^+(x,y)$ and a patch generated at the same lesion center without an added lesion is denoted as $I_n^-(x,y)$, where *n* represents the n^{th} lesion. The mean signal across the

N lesion-present and *N* lesion-absent patches is denoted as $\overline{I^+}(x,y)$ and $\overline{I^-}(x,y)$, respectively, and the mean signal profile between them is defined as:

$$\bar{S}(x,y) = \overline{I^+}(x,y) - \overline{I^-}(x,y) \tag{2.5}$$

$\Delta I_n^-(x,y)$ is defined by:

$$\Delta I_n^-(x,y) = H[x,y] \times (I_n^-[x,y] - \overline{I^-}[x,y]) \tag{2.6}$$

where $H[x,y]$ is defined as the inner product of two 1-D Hamming filters[53] and serves to mitigate artifacts arising from the cyclic nature of the discrete Fourier transform by attenuating the difference as it approaches the edge of the image[51,54]. $I_n^-[x,y]$ represents the lesion-absent patch at the *n*th lesion, and $\overline{I^-}[x,y]$ represents the average of all lesion-absent patches. The mean noise power spectrum in Fourier space is computed from the lesion-absent patches as follows:

$$\overline{PS}(f_x,f_y) = \frac{1}{N-1}\sum_{n=1}^{N}\left|\widehat{\Delta I_n^-}(x,y)\right|^2 \tag{2.7}$$

where $N-1$ represents the loss of one degree of freedom since $\Delta I_n^-(x,y)$ involves a subtraction, and the caret represents the 2D Fourier transform. By characterizing the frequency dependence of the anatomical background in **Equation 2.7**, dividing by this frequency dependence flattens the resulting frequency dependence, which is the "whitening" part of the PWMF observer. The PWMF in the spatial domain is then defined as:

$$w[x,y] = FT^{-1}\left\{\frac{FT\{\bar{S}(x,y)\}}{\overline{PS}(f_x,f_y)+R}\right\} \tag{2.8}$$

where FT represents the 2D Fourier transform, FT^{-1} represents the 2D inverse Fourier transform, and R is a regularization constant used to mitigate high frequency noise in the filter. R was empirically determined to be 10^6. **Figure 2.3** shows the mean signal $\bar{S}(x,y)$, the

2D PWMF, and a surface-contoured plot of the PWMF for a 15 mm lesion. We observe that the PWMF weights intensities at the lesion boundary.

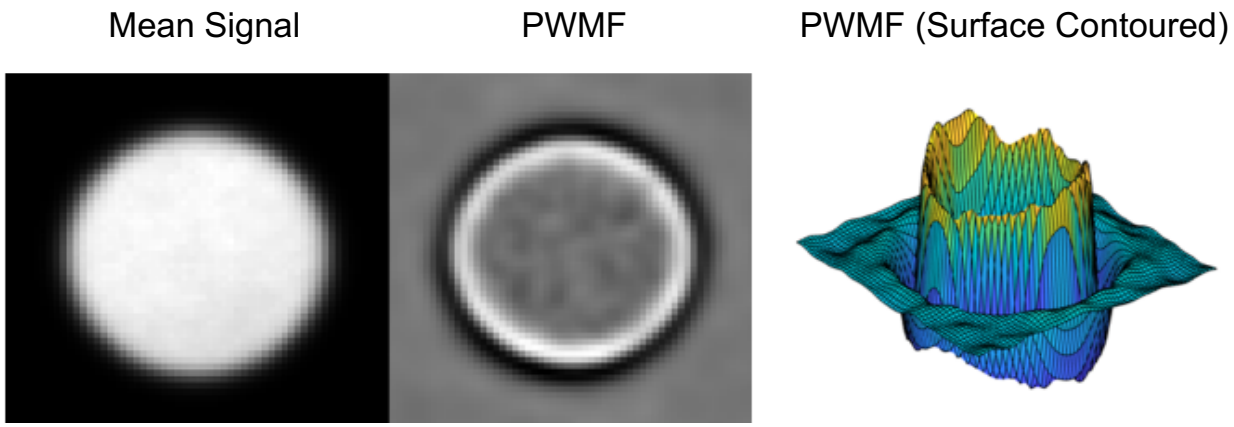

Figure 2.3: Computation of PWMF. The leftmost column shows the mean signal of N inserted 15 mm unenhanced lesions. The center column shows the 2D PWMF for the same lesion size. A surface contoured plot of the PWMF is shown in the rightmost column. The PWMF weights intensities at the lesion boundary.

2.2.6. PWMF performance evaluation

The PWMF was then used to detect lesions in a separate set of hybrid breast CT image patches as a signal known exactly, location known exactly detection task. The PWMF was applied to each patch to compute a scalar valued decision variable λ_n. For each breast CT volume data set, *N* unique viable lesion centers were first generated using a random number generator. For each lesion center and combination of parameters, a lesion-present patch $I_n^+[x,y]$ was generated by the lesion insertion process. For that lesion center, the decision variable λ_n is the sum of the pixel intensity of the image patch $I_n[x,y]$ weighted by the PWMF, $w[x,y]$:

$$\lambda_n = \sum_{x,y} I_n[x,y] \times w[x,y] \quad (2.9)$$

N additional unique lesion centers were then found and used to compute N lesion-absent decision variables using **Equation 2.9**. Independent lesion centers were generated between the sets of lesion-present and lesion-absent patches so as not to correlate the decision variables. The decision variables were plotted on a histogram as depicted in **Figure 2.4** and converted to an empirical receiver operating characteristic (ROC) curve[55] by plotting sensitivity and false positive rate over a range of detection thresholds. For each ROC curve, the area under the ROC curve (AUC) was computed, where the AUC characterizes the detectability performance of that combination of parameters on that specific patient breast CT image. To get an estimate of the overall detectability related to a combination of parameters, the individual AUCs were averaged over the K image volumes (K = 253). The standard deviation of AUCs was also computed.

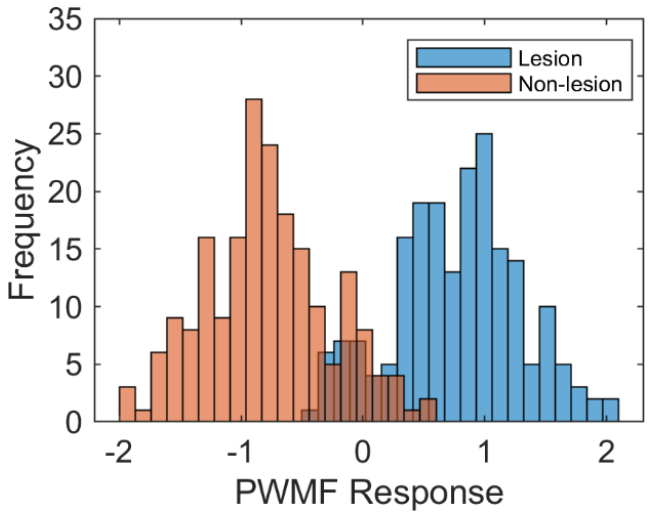

Figure 2.4: Example PWMF response histogram for one breast. PWMF responses (i.e., decision variables) are computed and distinguished between input patches with lesions ("Lesion") and without lesions ("Non-lesion"). Responses are used to generate the ROC curve, which can be used to compute the AUC.

2.2.7. Statistical analysis

Performance estimates were plotted with 95% confidence error bars in **Section 2.3**, which are given by the following equation:

$$CI_{95} = \overline{AUC} \pm 1.96 \times \frac{\sigma}{\sqrt{K}} \tag{2.10}$$

where CI_{95} is the 95% confidence interval, $\overline{AUC}$ is the mean AUC across all breast CT images, σ is the standard deviation of AUCs across all breast CT images, and K is the number of image volumes. Confidence intervals are plotted to represent uncertainty in the plots.

In **Section 2.3**, the improvement in average detectability performance, $\Delta\overline{AUC}$, is computed between detectability estimates in contrast-enhanced ($\alpha = 1$) and unenhanced plots ($\alpha = 0$) using **Equation 2.11:**

$$\Delta\overline{AUC} = \overline{AUC}_{contrast-enhanced} - \overline{AUC}_{unenhanced} \quad (2.11)$$

The percent increase is then calculated using **Equation 2.12**, recognizing that detection performance is 0% when $AUC = 0.5$, and detection performance is 100% when $AUC = 1.0$.

$$\%\ increase = 100 \times \frac{\Delta\overline{AUC}}{0.5} \quad (2.12)$$

The percent increase serves as a metric of improvement using contrast enhancement and is used to observe trends in the study across clinical parameters. For all comparative tests, a non-parametric approach is taken using the Mann-Whitney U-test. Statistical significance is defined as a difference with $p < .05$. When multiple comparisons are made, a Bonferroni correction was applied by dividing the desired significance (0.05) by the number of comparisons. All tests were two-sided. Statistical analyses were performed in MATLAB (MATLAB; TheMathWorks Inc., Natick, MA).

2.2.8. Number of inserted lesions, N

N is the number of lesions used to compute the mean signal profile in **Equation 2.5**, the mean noise power spectrum in **Equation 2.7**, and the PWMF response histogram for each image. Packard *et al.* used $N = 500$. Due to the computationally expensive process of lesion simulation, the feasibility of reducing *N* was investigated. Detectability estimates for 1- and 5-mm unenhanced lesions in the thinnest section were generated for *N* = 10, 25, 50, 100, 200, 300, 400, and 500. Each detectability estimate (AUC) was repeated ten times to infer error. Two-sided Student's t-tests were used to determine whether differences in detectability

estimates using different *N*'s were statistically significant. Mean AUCs were plotted as a function of the number of inserted lesions, *N,* in **Figure 2.5**.

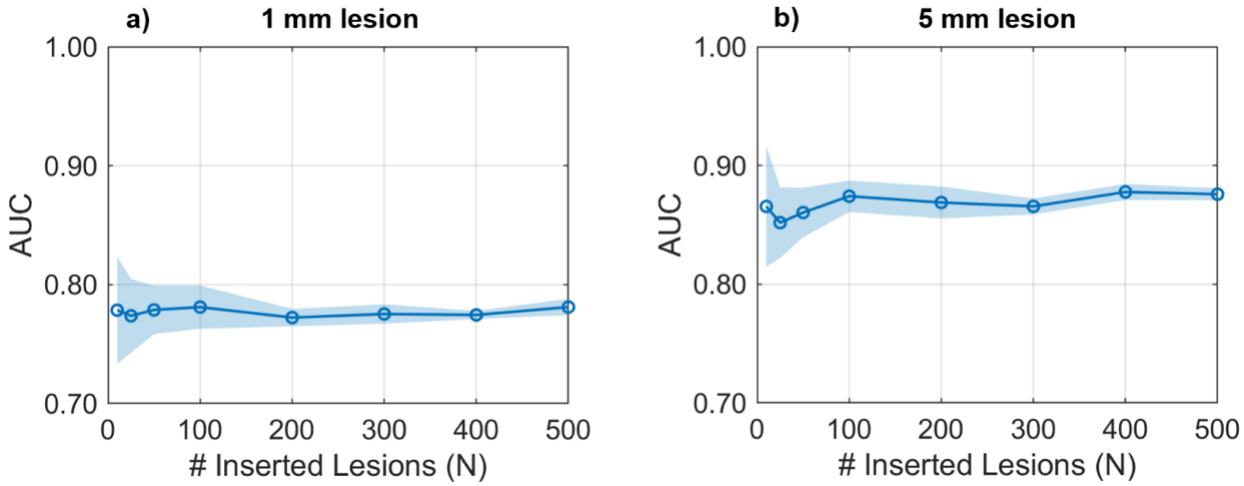

Figure 2.5: Effect of number of inserted lesions, N, on AUC estimation for (a) 1 mm lesion and (b) 5 mm lesion. Detectability estimates were repeated ten times for N = 25, 50, 100, 200, 300, 400, and 500. The mean AUC was plotted for each N. Shaded regions correspond to 95% confidence interval for each estimate.

For both 1- and 5-mm lesions, it was found that when *N* is less than 100, there is notable fluctuation and larger error when estimating AUC. As *N* increases, error steadily decreases, and AUC estimates begin to converge. Based on **Figure 2.5**, we determined that reducing *N* to 200 could be used to reduce computational time but still provide an accurate estimation of detectability with high precision. The differences in performance estimates using *N* = 500 and *N* = 200 were not statistically significant ($p = .117, p = .351$ for 1- and 5-mm lesions, respectively).

2.3. Results

2.3.1. Model validation

Detection performance for unenhanced lesions was plotted against similar results from Packard *et al.* in **Figure 2.6**. Performance estimates for 1-, 5-, and 15-mm diameter lesions averaged over all breast densities are plotted as a function of section thickness. Notably, our results are of the axial view plane only, while Packard *et al*.'s results are averaged over three views planes: coronal, axial, and sagittal. There were differences between the projection planes that likely explain the differences between the two data sets. Moreover, our data set contained 253 image volumes, and Packard utilized 151 image volumes. Considering the differences between the studies being compared, there is reasonable agreement between trends observed in this study and Packard *et al*.'s.

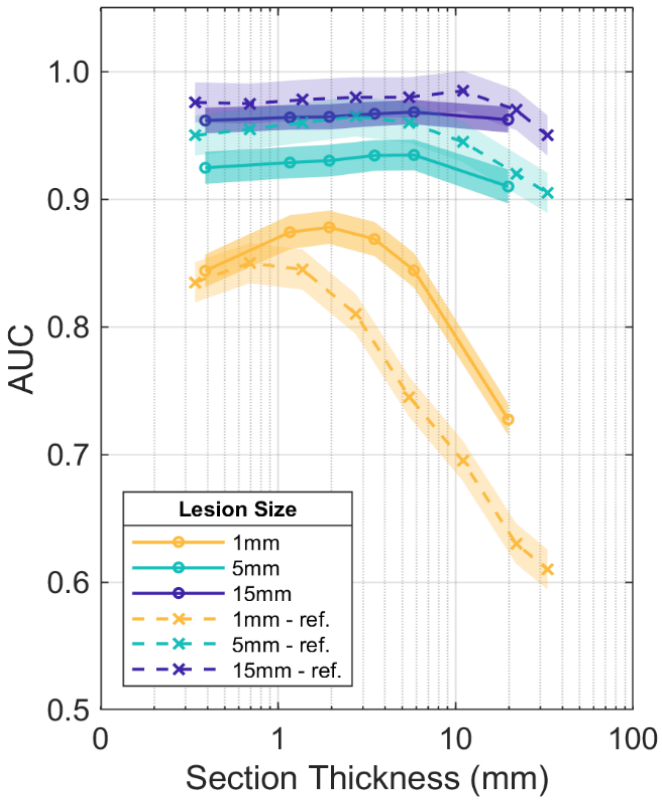

Figure 2.6: Model validation. AUC is plotted as a function of section thickness for 1-, 5-, and 15-mm lesions in the axial view plane. Results from this study are compared with results from the reference study (Packard *et al*). Shaded regions correspond to 95% confidence interval for each estimate.

2.3.2. Effects of contrast enhancement

Detection performance was averaged across all breast densities (i.e., VGFs) and both view planes and plotted as a function of contrast level (α) for three lesion sizes in **Figure 2.7**. As expected, contrast injection improved detection performance for all lesion sizes (1 mm: $p <$ .01; 5 mm: $p < .01$; 15 mm: $p < .01$). For the 1 mm lesion, detection performance steadily improved by 20% from $AUC = 0.85$ with no contrast enhancement ($\alpha = 0$) to $AUC = 0.95$ at full contrast enhancement ($\alpha = 1$).

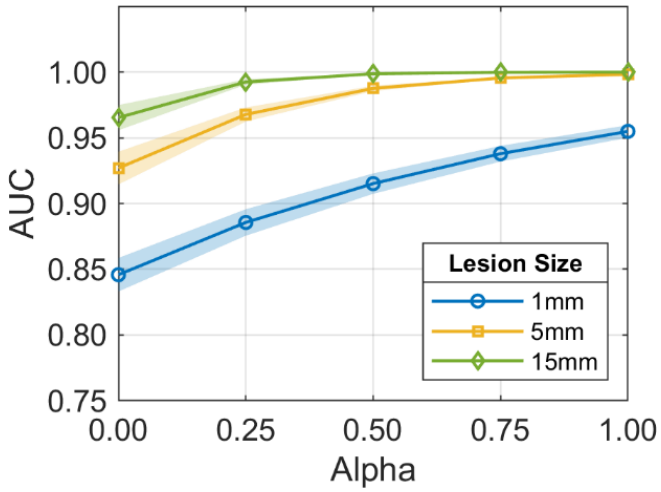

Figure 2.7: Effect of contrast levels. AUC is plotted as a function of contrast level α for 1-, 5-, and 15-mm lesion sizes. Shaded regions correspond to 95% confidence interval for each estimate.

2.3.3. Effects of section thickness

Detection performance was averaged across all breast densities (i.e., VGFs) and both view planes and plotted as a function of section thickness for six lesion sizes in **Figure 2.8**. **Figure 2.8a** displays performance estimates for unenhanced lesions ($\alpha = 0$). The thinnest section (0.4 mm) may not be the ideal section thickness for unenhanced lesion detection. For the 1 mm lesion, detection performance rises to a mild peak around 1.2 mm thickness as section thickness increases before falling off more rapidly. For larger lesions, detection performance is generally unaffected as section thickness increases up until thickness reaches 5.8 mm, where performance begins to decrease.

Figure 2.8b displays performance estimates for contrast-enhanced lesions ($\alpha = 1$). Detection performance improves for contrast-enhanced lesions across all section thicknesses and lesion sizes (family-wise $p < .01$). With contrast, detection improves approximately 5-

20% until maximum detection performance (i.e., AUC = 1) is reached. When section thickness is less than 2 mm, the AUC estimate for contrast-enhanced lesions greater than 1 mm in diameter is approximately 1.0.

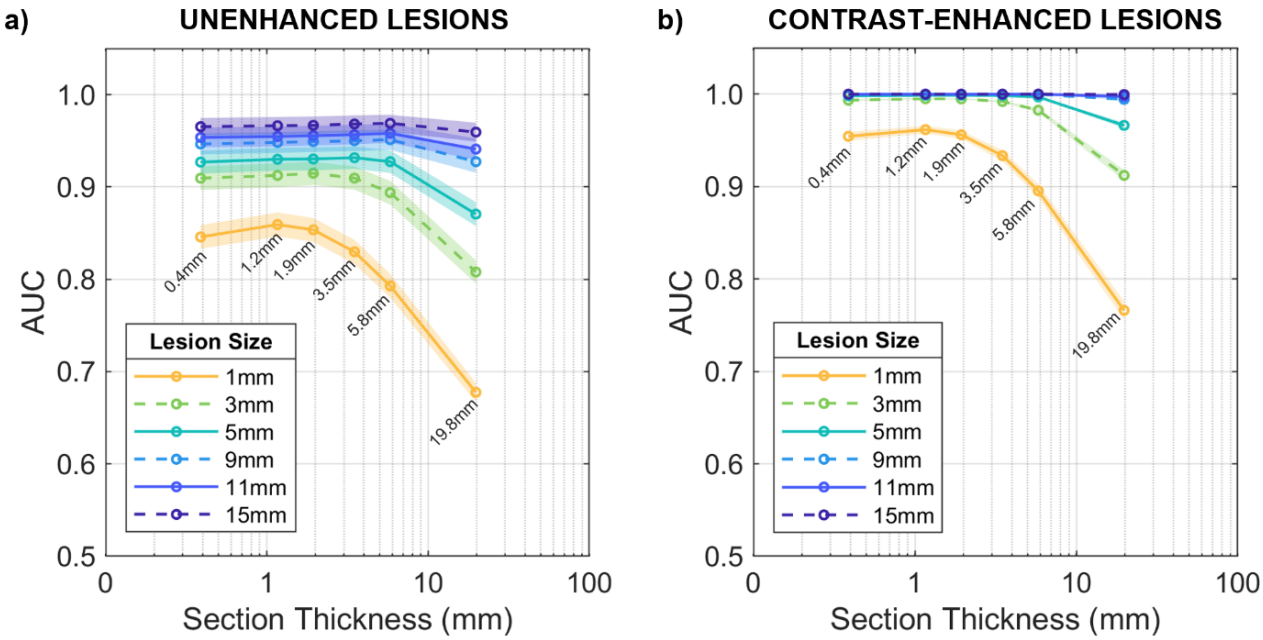

Figure 2.8: Effects of section thickness in (a) unenhanced lesion detection and (b) contrast-enhanced lesion detection for six lesion sizes. Shaded regions correspond to 95% confidence interval for each estimate.

2.3.4. Effects of lesion size

Detection performance was averaged across all breast densities (i.e., VGFs) and both view planes and plotted as a function of lesion diameter for a thin section (0.4 mm) and a thick section (19.8 mm) in **Figure 2.9**. As might be expected, larger lesions are uniformly easier to detect than smaller lesions in both thin and thick sections. Detection performance improves for contrast-enhanced lesions across lesion size and section thickness (family-wise $p < .01$).

When contrast is used, detection performance improves up to 20% (until AUC → 1) across lesion diameter for both thin and thick sections.

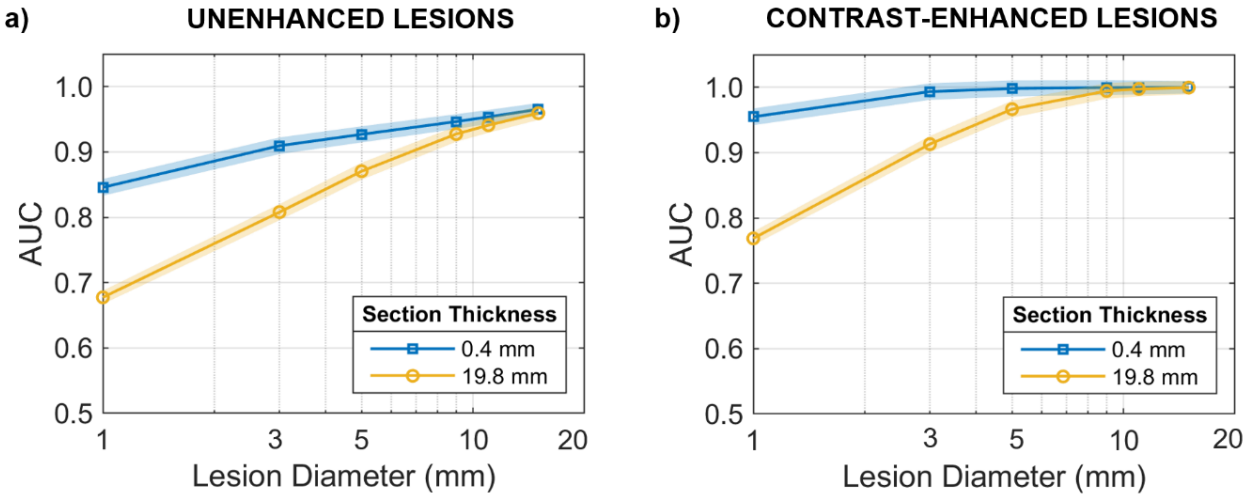

Figure 2.9: Effects of lesion size. (a) Performance estimates for unenhanced lesions ($\alpha = 0$) in the thinnest (0.4 mm) and thickest sections (19.8 mm). (b) Performance estimates for contrast-enhanced lesions ($\alpha = 1$) for the thinnest (0.4 mm) and thickest (19.8 mm) sections. Shaded regions correspond to 95% confidence interval for each estimate

2.3.5. Effects of breast density

The distribution of patient breast densities from our patient data set is shown in **Table 2.2**:

Table 2.2: Distribution of breast densities from patient data set, represented by volumetric glandular fraction (VGF).

VGF	0-10%	11-20%	21-30%	31-40%	41-50%	>50%
N	3	70	104	48	18	10

A small fraction of patients (3/253 patients, 1.2%) had primarily fatty breasts (VGF ≤ 10%), and a small fraction of patients (10/253 patients, 3.9%) had very dense breasts (VGF > 50%). The remaining patients (240/253 patients, 94.9%) had low to moderately dense breasts ranging from 11-50% glandularity. Detection performance was averaged across all section thicknesses and both view planes and plotted as a function of VGF for three lesion sizes.

Effects of breast density on detection performance can be observed in **Figure 2.10**. **Figure 2.11** displays improvement in detectability with contrast injection in terms of ΔAUC and percent increase. For patients with VGF ≤ 40%, detection performance improves up to 20% (until AUC → 1), and for patients with denser breasts (VGF > 40%), detection performance improves more drastically, ranging from 20-80% for 1- and 5-mm lesions. There is no statistically significant difference in detection performance for a 1 mm lesion in a fatty breast with and without contrast (VGF $\leq 10\%; p = 0.4$). For every other breast density category, the improvement in detection performance for the 1 mm lesion with and without contrast is statistically significant (11-20%: $p < .01$; 21-30%: $p < .01$; 31-40%: $p < .01$; 41-50%: $p < .01$; >50%: $p < .01$).

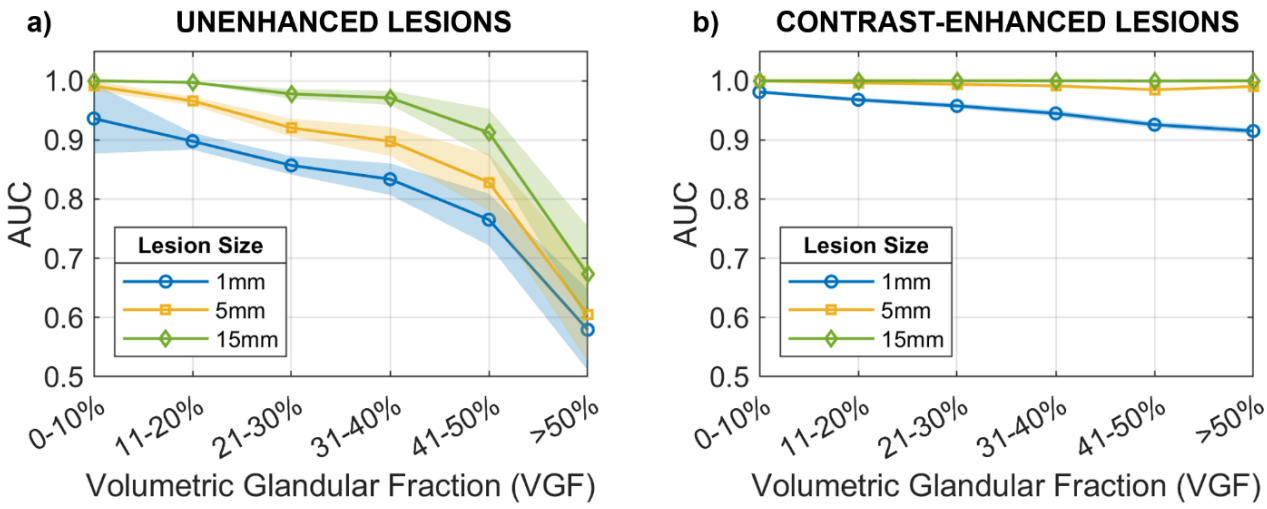

Figure 2.10: Effects of breast density. (a) Performance estimates for unenhanced lesions ($\alpha = 0$) for three lesion sizes. (b) Performance estimates for contrast-enhanced lesions ($\alpha = 1$) for three lesion sizes. Shaded regions correspond to 95% confidence interval for each estimate.

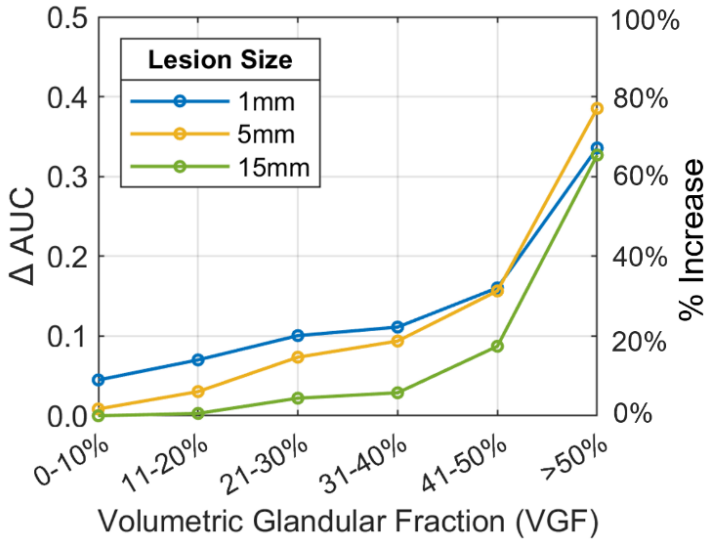

Figure 2.11: Improvement in detectability (ΔAUC) between contrast-enhanced lesions and unenhanced lesions for three lesion sizes as a function of VGF. Equivalent percent increase is denoted on the right vertical axis.

2.3.6. Effects of view plane

Effects of view plane on lesion detectability can be observed in **Figure 2.12**. Detection performance was averaged across all breast densities (i.e., VGFs) and plotted as a function of section thickness for 1- and 5-mm lesions viewed in the coronal and axial planes.

Equivalent or greater detection performance is achieved using the axial view compared to the coronal view for all section thicknesses, both lesion sizes, and with and without contrast. For the 1 mm unenhanced lesion, there is no statistically significant difference in detection performance between axial and coronal views using 0.4 mm section thickness ($p = 0.61$). For thicker sections, the axial view outperforms the coronal view (z = 1.2 mm: $p < .01$; z = 1.9 mm: $p < .01$; z = 3.5 mm: $p < .01$; z = 5.8 mm: $p < .01$; z = 19.8 mm: $p < .01$). For the 5 mm unenhanced lesion, there is no statistically significant difference in detection performance between axial and coronal views using section thicknesses of 0.4 mm ($p = 0.45$), 1.2 mm ($p = 0.54$), 1.9 mm ($p = 0.65$), and 3.5 mm ($p = 0.15$). The axial view outperforms the coronal view at thicker sections (z = 5.8 mm: $p < .01$; z = 19.8 mm: $p < .01$). Contrast injection improves detection performance approximately 20% ($\Delta AUC = 0.1$) for both lesion sizes across all section thicknesses.

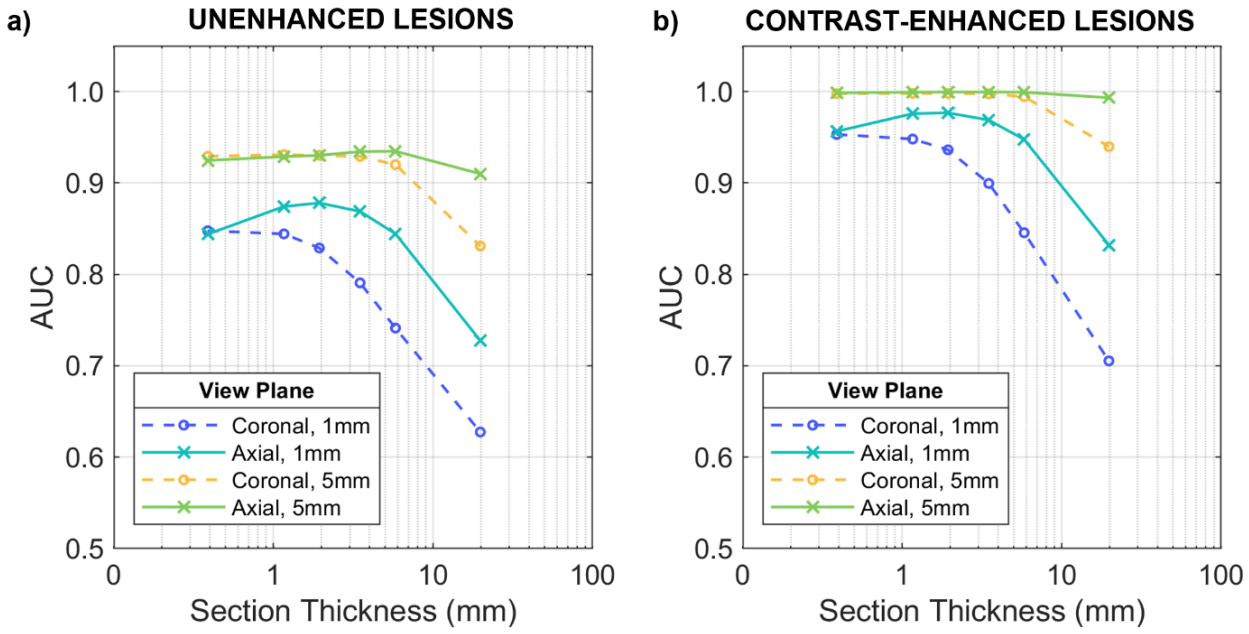

Figure 2.10: Effects of view plane on lesion detectability. (a) Detection performance for unenhanced 1 mm and 5 mm lesions ($\alpha = 0$) viewed in coronal and axial planes. (b) Detection performance for 1 mm and 5 mm contrast-enhanced lesions ($\alpha = 1$) viewed in coronal and axial planes.

2.4. Discussion

We sought to quantify improvement in lesion detectability due to contrast enhancement across lesion diameter, section thickness, breast density, and view plane. **Figure 2.7** shows the effect of contrast level on lesion detectability. At full contrast ($\alpha = 1$)., detection performance improves up to 20% (until AUC →1) from unenhanced detection performance. Similar improvement is also observed when visually comparing unenhanced and contrast-enhanced plots in **Figures 2.8, 2.9, and 2.12.** A larger improvement in detection performance is observed in **Figure 2.10** when stratifying patients based on breast density. In unenhanced breast CT, we observe that lesion detectability has dependencies on patient breast density. The AUC for unenhanced lesions in patients with mostly fatty breasts (VGF < 20%) exceeds

0.9, while the AUC for unenhanced lesions in patients with very dense breasts (VGF > 50%) ranges from 0.55 – 0.70. With contrast, AUC exceeds 0.9 for all lesion sizes across all breast densities. In effect, when using injected contrast, lesion detectability becomes only slightly dependent on breast density.

Section thickness is an important factor when displaying tomographic images, but not the only one. Image acquisition and reconstruction methods, kernel size in filtered backprojection, and reconstructed voxel size also matter, and the tradeoff between image resolution and image noise is well known. Therefore, it is hard to determine a universal optimal section thickness for displaying breast CT. Aside from very thin sections, detection performance generally decreases as section thickness increases, as might be expected. Thicker sections often suffer from superposition of fibroglandular anatomy which can obscure lesions and reduce detectability. This is the primary limitation of 2D projection imaging modalities such as mammography or tomosynthesis. Reduced detectability is more pronounced in denser breasts compared to fatty breasts since more fibroglandular tissue is superimposed[40], therefore, optimal section thickness is also dependent on breast density. For this reason, we have built our in-house breast CT viewer to enable the reader to adjust the display slice thickness in all three projections simultaneously in real-time. For the general breast CT system, a middle ground between selecting an overly thin slice at the cost of quantum noise and an overly thick slice at the cost of anatomical noise should be found. Intravenous contrast appears to improve detection performance around 20% on average and may compensate for losses in detection performance due to selecting an overly thin or overly thick section.

In **Section 2.3.6** it was determined that equivalent or greater detection performance is achieved using the axial plane compared to the coronal plane across clinical parameters.

In a previous study, Packard *et al.* demonstrated that the sagittal view performed very similarly to the axial view, which is expected as the breast is approximately radially symmetrical along the anterior-posterior line. For this reason, we only evaluated the two view planes and averaged other parameters across these two view planes. In our validation study in **Section 2.3.1**, there are differences between Packard's performance estimates and ours, although both curves take on similar trends. The improved detection performance for the 1 mm lesion in our study may be due to advanced image reconstruction techniques[18] that were used in this study but not in Packard *et al*.'s study. These techniques enabled sharper image resolution which in turn improved the added lesion resolution, since the added lesion boundaries were blurred to match the resolution of the patient image. The improved resolution appears to be more impactful when detecting smaller lesions, i.e., 1 mm lesions. Another likely reason for these differences is the differences in patient data sets since Packard used a 151-patient data set, and a 253-patient data set was used in this study. Despite the axial view performing better for lesion detection, breast CT images are traditionally viewed in the coronal plane, or with all three orthogonal views on the same display. In clinical practice, radiologists tend to view breast CT images in thin sections, therefore, differences in detectability between the two views are likely to be minor.

This study had limitations. The patient images used to generate hybrid images contained pre-contrast images of the affected breast. It is possible that some lesion-absent ROIs contained real mass lesions, but these instances were likely few given the small relative size of a mass lesion compared to the volume of a breast in pendant position. Furthermore, AUC was averaged over 400 ROIs across 253 images, therefore, any bias introduced by individual ROIs containing mass lesions was likely mitigated by averaging procedures. Breast cancer presents as both masses and microcalcifications, but only mass lesions were

simulated in this study. Partial volume effects appear to influence lesion detectability as mass lesions become very small. In this study, partial volume was modeled using a single modulation function, M, independent of lesion size. It is possible that this function may have been insufficient to fully model partial volume effects for the 1 mm diameter lesion. Only malignant masses from our clinical contrast-enhanced breast CT data set were studied for enhancement calibration, but benign masses are known to minimally enhance as well[35]. The methodology for adding contrast was simplified in this study such that each voxel within the inserted lesion was uniformly enhanced. In clinical images, enhancing lesions can appear more visibly heterogenous, particularly for lesions containing microinvasions of ductal carcinoma in situ (DCIS), cysts, or distortions. Furthermore, a simple SKE/LKE detection task was implemented such that the PWMF fixed the signal profile and did not take lesion shape into consideration. Accordingly, a spherical lesion was inserted in 3D, although in reality, lesions vary in shape and size. Lesion phenotypes that are known to indicate malignancy, e.g., spiculations, were not simulated in this study. Furthermore, despite the simplifications of fixing an SKE signal, it has been shown that the PWMF models optimal detection performance and is similar to human detection performance[52]. For the purpose of modeling human observers, it appears that the PWMF model is useful.

2.5. Conclusion

Contrast enhancement was simulated for synthetic spherical lesions, and enhancement levels were calibrated based on 22 malignant masses from our clinical contrast-enhanced breast CT data set. In this model observer study, the improvement in detection due to contrast injection was quantified across breast density, lesion size, view plane, and section thickness. Small lesions are generally harder to detect in dense breasts, but these results suggest that injected contrast can substantially improve detection performance in dense breasts. Optimal

section thickness for detectability has dependencies on breast density and lesion size, therefore, display thickness should be adjusted in real-time using display software.

For emerging diagnostic tools like contrast-enhanced breast CT, it is important to optimize imaging protocols for lesion detection. These findings are not only useful for the optimization and development of contrast-enhanced breast CT but may also be relevant in other x-ray-based breast imaging modalities such as contrast-enhanced breast tomosynthesis[56] and dual energy contrast-enhanced mammography[57].

Chapter 3: Pre-whitened matched filter and convolutional neural network-based model observer performance for mass lesion detection in non-contrast breast CT

3.1. Introduction

Breast computed tomography (CT) is a relatively new breast imaging modality based upon cone-beam CT geometry[16,58], or helical CT geometry[59]. While a number of human observer studies have been published on the performance of breast CT[16,30,60], such studies can be limited when assessing a large number of images, which is often the case when fine-tuning an imaging system or identifying optimal parameter settings for lesion detectability. Mathematical model observers have been shown to reasonably predict human observer performance[32,33], and are useful when human observer studies are infeasible[20]. Recently, convolutional neural networks (CNNs) have also been used as substitutes for human observers[61,62], where they are referred to as anthropomorphic models. Studies have shown that appropriately trained CNNs have utility as an optimal observer, the so-called ideal observer[27,63,64]. Ideal observers are useful for assessing how much diagnostic information is contained in an image in advance of processing or display effects that make this information accessible to human observers[65–67].

In this study, a CNN model observer is compared to a more conventional ideal-observer model, the pre-whitened matched filter (PWMF). The PWMF has an appealing definition that involves the signal to be detected as well as the texture of the image background, as specified by the power spectrum. The PWMF is known to be an optimal detection filter for images with variability that is described by a stationary Gaussian distribution[65,68,69] and it is related to image-quality measures like noise-equivalent quanta and

detective quantum efficiency. However, it has been demonstrated that breast CT images are not Gaussian distributed[70], and so it is not clear that the PWMF represents an ideal observer in this case. This means that it may be possible that the PWMF is systematically missing diagnostic information in breast CT images and may therefore underestimate optimal performance across patient and imaging factors.

Neural networks[71] have been suggested as a way to implement the ideal observer when the analytical approach of defining a likelihood ratio is not feasible, and more recently CNNs specifically have been evaluated for this purpose[72,73]. This general approach is based on the flexibility of the network architecture and the ability to train the model on samples of data rather than derive accurate probabilistic models of images with complex non-Gaussian statistical properties. Flexible network architectures allow the network models to extract image information that may not be accessible to a model like the PWMF that is constrained to be linear.

The field has generally found that network models are capable of higher performance than an optimal linear filter like the PWMF, and this includes recent studies by Baek and colleagues[74–76] using synthetic breast CT images with simulated anatomical backgrounds. In this report, we seek to extend these results to breast CT images acquired from patients for the task of detecting a simulated lesion that is embedded in real anatomical background. This is, to our knowledge, the first comparison between CNNs and mathematical model observers in breast CT images with real anatomical backgrounds, and may provide a more accurate assessment of the model observers when applied to breast CT. This detection paradigm has been studied previously using the PWMF[21,40], and was found useful for understanding how detection performance is dependent on the interaction between lesion diameter and section thickness.

This study involves a data set of 322 patient breast CT images acquired at the UC Davis Medical Center under an IRB approved protocol[16,60]. We also investigate Gaussian images matched to the mean and power-spectrum of the breast CT images. This allows us to implement our particular CNN in an imaging condition where the PWMF is known to be an ideal observer. If the CNN is able to closely approximate the ideal observer, then we have some confidence that the architecture and training process is adequate for comparing performance more generally. In the breast CT images, we compare the CNN to the PWMF across lesion diameter, slice thickness, and breast-density categories.

3.2. Methods

3.2.1. Image generation

3.2.1.1. Insertion of lesions into breast background

Spherical lesions were mathematically inserted into reconstructed patient breast CT image volumes to mimic 3D mass lesions in breast parenchyma. Human breast CT images were acquired at the UC Davis Medical Center under IRB-approved clinical trials which recruited patients receiving BIRADS 4 or 5 on their breast screening exams[16,60]. Enrolled patients were scanned on prototype breast CT scanners[18] developed in our laboratory. All patients subsequently underwent biopsy to yield the ground truth diagnosis for suspicious lesions as benign or malignant. Of the four existing iterations of prototype breast CT scanners, the first two scanners, which are very similar in design, were used to scan a total of 322 women. From this cohort we selected 253 image volume data sets for this study on the basis of not containing artifacts and not involving contrast-imaging. Each volume data set contained 300-500 reconstructed slices (512 × 512 matrix size) with isotropic voxel sizes of 0.4 mm.

A previously published method developed by Packard *et al.*[40] was used to insert spherical lesions into breast CT images and is briefly detailed here. Let $f[i,j,k]$ be a reconstructed breast CT image volume and $s[i,j,k]$ be a segmented version of $f[i,j,k]$, where each voxel is segmented as adipose tissue, fibroglandular tissue, skin, or air. $s[i,j,k]$ is used to compute $TB[i,j,k]$, a binary volume identifying tissue boundaries (TB) in the breast, and having a value of 1 for every voxel segmented as fibroglandular tissue but having one of its six adjacent voxels (in 3D) segmented as adipose tissue, and 0 for all other voxels. $TB[i,j,k]$ is used to compute $d_{TB}[i,j,k]$, the distance from every voxel to the nearest tissue boundary identified in $TB[i,j,k]$. $d_{TB}[i,j,k]$ is signed such that it is positive for voxels segmented as adipose tissue and negative for voxels segmented as fibroglandular tissue.

Let the index location $[i_{LC}, j_{LC}, k_{LC}]$ be a randomly generated lesion center "LC" where the lesion is to be inserted. The lesion location $[i_{LC}, j_{LC}, k_{LC}]$ is kept if the surrounding $64 \times 64 \times 64$ volume is fully contained within the patient breast and does not contain skin; otherwise, the lesion center coordinates are re-generated. Let D be the diameter of the lesion to be inserted and $d_{LC}[i,j,k]$ be the distance from each voxel to lesion center $[i_{LC}, j_{LC}, k_{LC}]$. The distance to the nearest lesion boundary $d_{LB}[i,j,k]$ is then defined as:

$$d_{LB}[i,j,k] = \frac{1}{2}D - d_{LC}[i,j,k] \tag{3.1}$$

and is positive for voxels within the spherical lesion and negative for voxels outside the spherical lesion. Let ΔI be the mean differential intensity between all fibroglandular and adipose voxels in the image. Then, the resulting image volume with the inserted spherical lesion $f_{sim}[i,j,k]$ is:

$$f_{sim}[i,j,k] = f[i,j,k] + \left(\Delta I \times M(d_{TB}[i,j,k]) \times M(d_{LB}[i,j,k])\right) \tag{3.2}$$

Intensity is added on a voxel-by-voxel basis in order to preserve the native image noise. Outside of the spherical lesion, the added term becomes zero. The added intensity at each voxel is modulated by $M(d_{TB}[i,j,k])$, the tissue-boundary modulation term, which ranges from 0-1 and approaches 0 when $d_{TB}[i,j,k]$ is negative. In effect, this term allows intensity only to be added to adipose regions and smooths the regions within the inserted lesion where adipose and fibroglandular tissue coincide. The added intensity at each voxel is further modulated by $M\left(\frac{1}{2}D - d_{LC}[i,j,k]\right)$, the lesion-boundary modulation term, which also ranges from 0-1 and serves to smooth the edge of the spherical inserted lesion. The modulation function M is derived by mathematically modeling the edge-blurring at boundaries between adipose and fibroglandular tissue in the native patient image. These modulation terms serve to retain the native image resolution (~ modulation transfer function).

A 2D $64 \times 64 \times 1$ image was then generated from the 3D image volume by extracting the center slice along the axial dimension or by slice averaging across adjacent slices to model thicker sections. Previous model observer studies demonstrated that higher detection performance was found in the axial and sagittal views compared to the coronal view in breast CT[40]. The resulting patch with the added lesion is denoted as $I_n^+(x,y)$, where *n* represents the *n*th lesion, and the same patch without the added signal is denoted as $I_n^-(x,y)$. Four lesion diameters (1, 3, 5, 9 mm) and six section thicknesses (0.4, 1.2, 2.0, 6.0, 12.4, 20.4 mm) were studied. Sample lesion-present patches with varying lesion diameters and section thicknesses are shown in **Figure 3.1**.

The role of breast density was retrospectively studied using the range of breast densities spanning the patient breast CT data sets. For every patient, breast density was quantified by the volumetric glandular fraction (VGF). Let n_g represent the number of voxels

segmented as fibroglandular tissue and n_a represent the number of voxels segmented as adipose tissue in the segmentation volume $s[i,j,k]$. The VGF is then defined as:

$$VGF = \frac{n_g}{n_g + n_a} \tag{3.3}$$

3.2.1.2. Insertion of lesions into Gaussian background

Mass lesions were mathematically inserted into Gaussian images matched to the mean and noise power spectrum of the breast CT images. For breast CT images a 3D lesion was inserted into a 3D background volume. In comparison, for Gaussian images a 2D lesion (derived from the breast CT patches) was inserted into a 2D background patch. Let $\overline{I^+}(x,y)$ and $\overline{I^-}(x,y)$ denote the mean lesion-present and mean lesion-absent patches, respectively. The added lesion is the mean signal $\bar{S}(x,y)$ across image patches from all breast CT training images:

$$\bar{S}(x,y) = \overline{I^+}(x,y) - \overline{I^-}(x,y) \tag{3.4}$$

This mean signal inherently smooths the lesion boundary and dampens the signal in thicker sections.

Generating the 2D Gaussian background patch requires knowledge of the breast CT power spectrum. The power spectrum was estimated from the training breast CT background images. Let $I_n^-[x,y]$ represent the n^{th} lesion-absent patch and $H[x,y]$ represent a 2D Hamming filter. The windowed deviation function $\Delta I_n^-(x,y)$ is then defined as:

$$\Delta I_n^-(x,y) = H[x,y] \times (I_n^-[x,y] - \overline{I^-}[x,y]) \tag{3.5}$$

where $H[x,y]$ is a windowing function used to attenuate artifacts arising from the cyclic nature of the discrete Fourier transform as it approaches the edge of an image[32]. Let N represent

the total number of lesion-absent patches. The mean power spectrum $\overline{PS}(f_x, f_y)$ of the image backgrounds is then estimated as:

$$\overline{PS}(f_x, f_y) = \frac{1}{N-1}\sum_{n=1}^{N}\left|\Delta\widehat{I_n^-}(x,y)\right|^2 \quad (3.6)$$

where the caret is used to represent the Fourier transform. The mean power spectrum was then normalized to compensate for the windowing that occurs in **Equation 3.5**. To normalize the power spectrum, the mean pixel variance at each pixel across all the background images was first computed:

$$V = \frac{1}{N-1}\sum_{n=1}^{N}(I_n^-[x,y] - \overline{I^-}[x,y])^2 \quad (3.7)$$

Then, the normalized mean noise power spectrum $\overline{PS}(f_x, f_y)_{norm}$ is defined as:

$$\overline{PS}(f_x, f_y)_{norm} = \frac{\sum_{x,y} V}{\sum_{x,y}\overline{PS}(f_x, f_y)} \times \overline{PS}(f_x, f_y) \quad (3.8)$$

$\overline{PS}(f_x, f_y)_{norm}$ was then converted to the spatial domain using a 2D inverse Fourier transform. Gaussian background patches were then generated by convolving the square root of $\overline{PS}(x,y)_{norm}$ with a $64 \times 64 \times 1$ patch of random white noise $E(x,y)$. In using the normalized mean noise power spectrum of breast CT images for Gaussian simulation, pixel variance between the two image types is maintained. The final Gaussian patch $G^+(x,y)$ with the inserted lesion is defined as:

$$G^+(x,y) = \bar{S}(x,y) + \left(E(x,y) * \sqrt{\overline{PS}(x,y)_{norm}}\right) \quad (3.9)$$

and the same patch without the inserted lesion is defined as $G^-(x,y)$. Sample Gaussian lesion-present patches with varying lesion diameters and section thicknesses are shown in **Figure 3.1**.

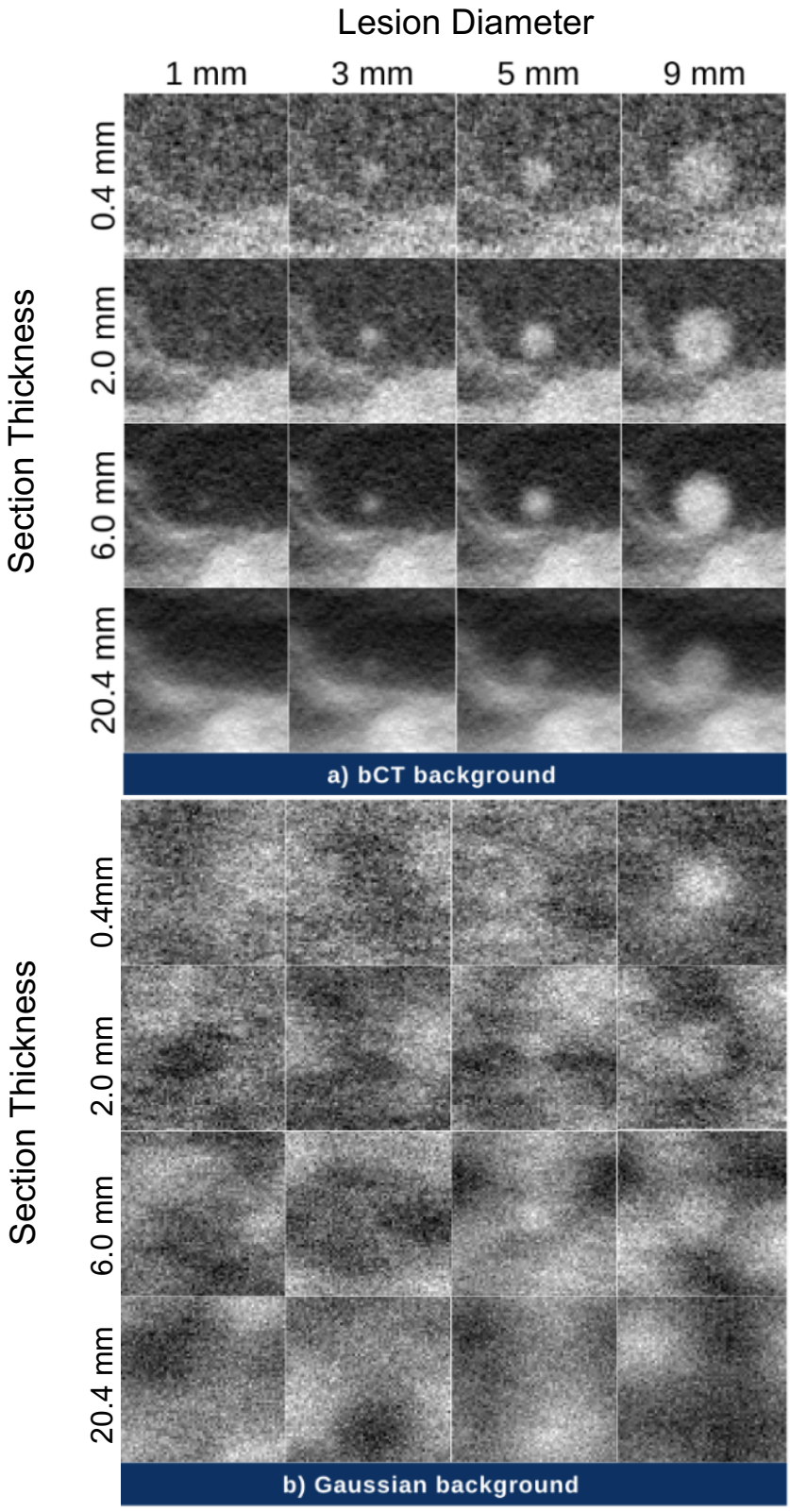

Figure 3.1: Example lesion-present patches in a) breast CT background and b) simulated Gaussian background for varying lesion diameters and section thicknesses.

3.2.2. Model observers: Pre-whitened matched filter

3.2.2.1. PWMF computation

The PWMF is a mathematical model observer that makes use of the mean signal and background power spectrum of a set of images to compute a decision variable[69,70]. A unique filter was computed for each combination of lesion diameter, section thickness, and background condition (breast CT or Gaussian) to tune the filter to the environment. Let FT denote the 2D Fourier transform, and FT^{-1} denote the 2D inverse Fourier transform. Let $\bar{S}(x,y)$ be the mean signal across all *training* images, and $\overline{PS}(f_x,f_y)_{norm}$ be the normalized mean noise power spectrum of *training* image backgrounds. The PWMF $w[x,y]$ is then defined as:

$$w[x,y] = FT^{-1}\left\{\frac{FT\{\bar{S}(x,y)\}}{\overline{PS}(f_x,f_y)_{norm}}\right\} \tag{3.10}$$

Once the PWMF was computed from a set of *training* images, it was then applied to an independent set of *testing* image patches in order to evaluate lesion detection performance. For breast CT conditions, lesion-present and lesion-absent testing patches were generated from completely separate breast CT volume data sets using the lesion insertion process described in **Section 3.2.1**. Let $I_n[x,y]$ represent a testing image patch and $w[x,y]$ be the PWMF tuned to that specific lesion diameter, section thickness, and background condition. A scalar-valued decision variable λ_n was then computed for each testing patch as follows:

$$\lambda_n = \sum_{x,y} I_n[x,y] \times w[x,y] \tag{3.11}$$

3.2.2.2. Training and testing the PWMF

Out of the 253 total breast CT volume data sets, 229 data sets (N = 229, ~90%) were used for training the PWMF. For each training data set, 200 unique lesion centers were identified and used to generate 200 lesion-present patches and 200 lesion-absent patches. The remaining 24 data sets (K = 24, ~10%) were used for testing the PWMF. For testing, 200 unique lesion centers were first identified to generate 200 lesion-present patches, and 200 additional different lesion centers were identified to generate 200 lesion-absent patches so as not to correlate the decision variables. In total, for a given lesion diameter and section thickness, 91,600 breast CT training patches (i.e., 229 × 400) were generated and 9,600 breast CT testing patches (i.e., 229 × 400) were generated. Though simulated Gaussian image patches were not dependent on patient data sets or lesion centers, the same number of training and testing data sets were generated to match the hybrid breast CT image data set.

3.2.3. Model observers: Convolutional neural network

3.2.3.1. CNN architecture

A convolutional neural network (CNN) was implemented to perform a simple binary classification task and compute a decision variable. The input to the CNN was a single 2D image patch, and the output was a scalar-valued decision variable between 0 and 1, scaled by the sigmoid function. The network consisted of two convolutional layers followed by one fully-connected layer. The two convolutional layers served to extract feature maps from the preceding layer, and the fully-connected layer condensed the feature maps into a scalar-valued decision variable. The first convolutional layer contained 3 × 3 filters with a stride of 1, and the second convolutional layer contained 3 × 3 filters with a stride of 1. Batch

normalization was implemented after the first convolutional layer. Max pooling layers were implemented after each convolutional layer with a pool size of 2 × 2. Dropout was implemented after the first max pooling layer with a rate of 0.2, after the second max pooling layer with a rate of 0.2, and after the fully-connected layer with a rate of 0.5. The rectified linear unit (ReLU) activation function was used in all layers, including the fully-connected layer.

The choice of a three-layered architecture was due to the relatively simple task of binary classification in a signal-known-exactly (SKE) setting. Similar studies have also employed this architecture[74,75]. Increasing the depth of the network could have resulted in a more complex model but may have led to overfitting[77]. In this model, the total number of parameters was 821,889. A diagram of the CNN architecture is shown in **Figure 3.2**.

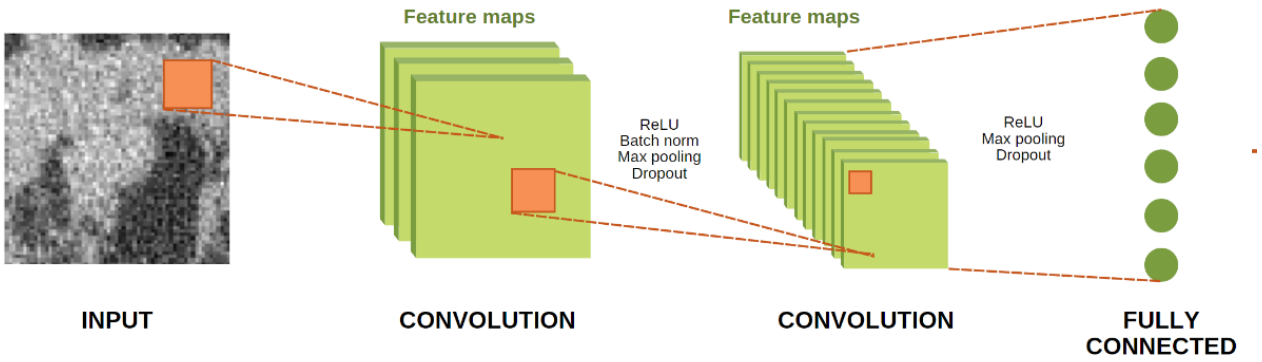

Figure 3.2: Convolutional neural network architecture.

3.2.3.2. Training and testing the CNN

The same training and testing simulated data sets described in **Section 2.2.2** for breast CT and Gaussian background images were used to train and test the CNN model observer. For

each combination of lesion diameter, section thickness, and background condition, 91,600 images were used for training and validation, and 9,600 images were used for testing.

The CNN was trained to minimize the binary-cross entropy (BCE) loss. Let y be the ground truth label (0 or 1), $\tilde{y}$ be the predicted value, and N be the number of samples. BCE loss is then defined as:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\tilde{y}) + (1 - y_i)\log(1 - \tilde{y_i})] \tag{3.12}$$

The training metric was accuracy:

$$Accuracy = \frac{\#\ correct\ predictions}{\#\ total\ predictions} \tag{3.13}$$

The Adam optimizer[78] was used with a learning rate of 1e-5 and a batch size of 64. The maximum number of training epochs was set to 150 but early stopping was implemented such that if the validation loss did not decrease after 4 epochs, training was stopped. The CNN model was implemented in Python using the Keras library[79]. An NVIDIA GeForce GTX 1080 GPU was used.

3.2.4. Performance evaluation and statistical analysis

For detection performance evaluation, receiver operating characteristic (ROC) curve analysis was used on the scalar-valued decision variables produced by the model observers. For a range of thresholds that discriminated each variable as true positive, false positive, true negative, or false negative, the true positive rate was plotted against the false positive rate to produce an empirical ROC curve. The area under the ROC curve (AUC) was computed individually for each testing data set (K = 24) instead of from one pool of all decision variables in order to study the effect of breast density on individual breast CT images. In addition, the

mean and standard deviation of AUCs across all 24 data sets were also computed to characterize the average detectability for that combination of parameters. Mean AUCs were plotted with 95% confidence error bars in **Section 3.3**. Let $\overline{AUC}$ and σ be the mean and standard deviation, respectively, of AUCs across the K testing data sets. The 95% confidence interval CI_{95} is then defined as:

$$CI_{95} = \overline{AUC} \pm 1.96 \times \frac{\sigma}{\sqrt{K}} \tag{3.14}$$

In **Section 3.3.1**, PWMF and CNN detection performance was compared on Gaussian background images. To quantify the similarity between the model observers, the maximum absolute difference between individual and mean AUCs were computed across clinical parameters. The maximum absolute difference between individual PWMF and CNN AUCs from 24 testing data sets is defined as:

$$|\Delta AUC|_{max} = \max\left(|AUC_{PWMF} - AUC_{CNN}|\right) \tag{3.15}$$

In **Section 3.3.2**, PWMF and CNN detection performance was compared across clinical parameters using paired t-tests. To address the multiple comparisons problem, we employed the Bonferroni correction to adjust the family-wise error rate. One asterisk (*) is used to indicate $p < .05$ and two asterisks (**) are used to indicate $p < .01$.

3.3. Results

3.3.1. Comparison of PWMF and CNN model observers in Gaussian background

The CNN observer closely matched the PWMF observer in detection performance across all section thicknesses and lesion diameters. Across these parameters, $|\Delta AUC|_{max}$ was 0.0096. PWMF and CNN detection performance on Gaussian images for the native section thickness (Z = 0.4 mm) as a function of lesion diameter are displayed in **Figure 3.3a**. Detection performance is nearly identical. **Figures 3.3b-c** show PWMF and CNN detection performance on Gaussian images as a function of section thickness for 1- and 5-mm lesions. In these settings, the CNN observer also closely aligned with the PWMF observer.

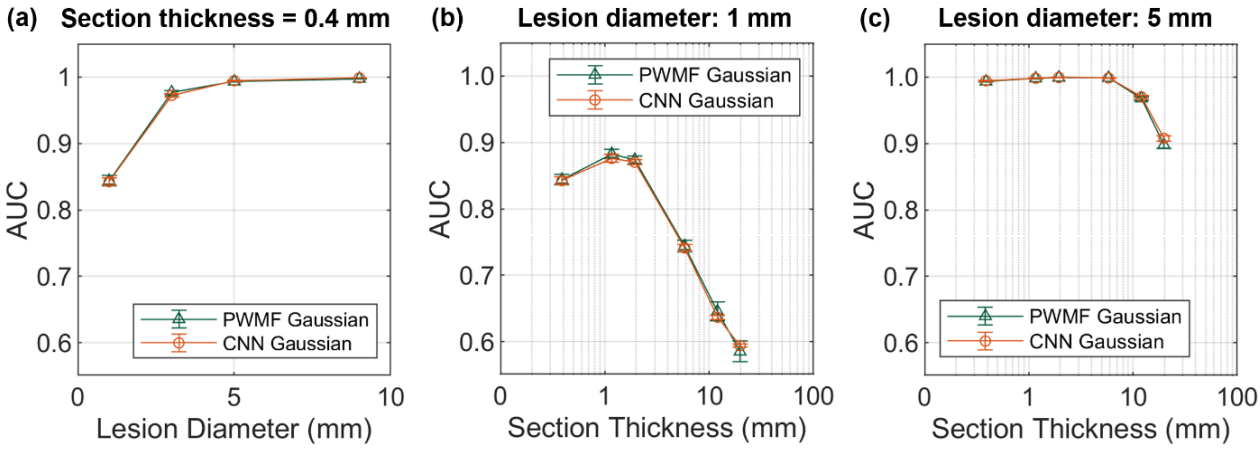

Figure 3.3: Comparison of PWMF and CNN model observers on Gaussian background images as a function of (a) lesion diameter, displayed in the native section thickness (0.4 mm), (b) section thickness for a 1-mm lesion, and (c) section thickness for a 5-mm lesion. Detection performance is nearly identical across all parameters. Error bars correspond to 95% confidence intervals for each performance estimate.

3.3.2. Comparison of PWMF and CNN model observers in breast CT background

3.3.2.1. Model observer comparison across lesion diameter

Model observer performance on breast CT images displayed in the native section thickness (Z = 0.4 mm) were averaged across breast densities and plotted as a function of lesion diameter in **Figure 3.4a**. Across all lesion diameters, the CNN outperformed the PWMF ($p < .01$). These findings were further analyzed in the context of breast density in **Figures 3.4b-c**. Of the 24 testing data sets, breast CT patches extracted from patients with lower VGF breasts (N = 12) were evaluated and the mean AUC was plotted in **Figure 3.4b**, and breast CT patches extracted from patients with higher VGF breasts (N = 12) were evaluated and the mean AUC was plotted in **Figure 3.4c**. The range of breast densities in the testing data set was [0.118, 0.597]. The CNN consistently outperformed the PWMF in higher- and lower-density breasts ($p < .05$).

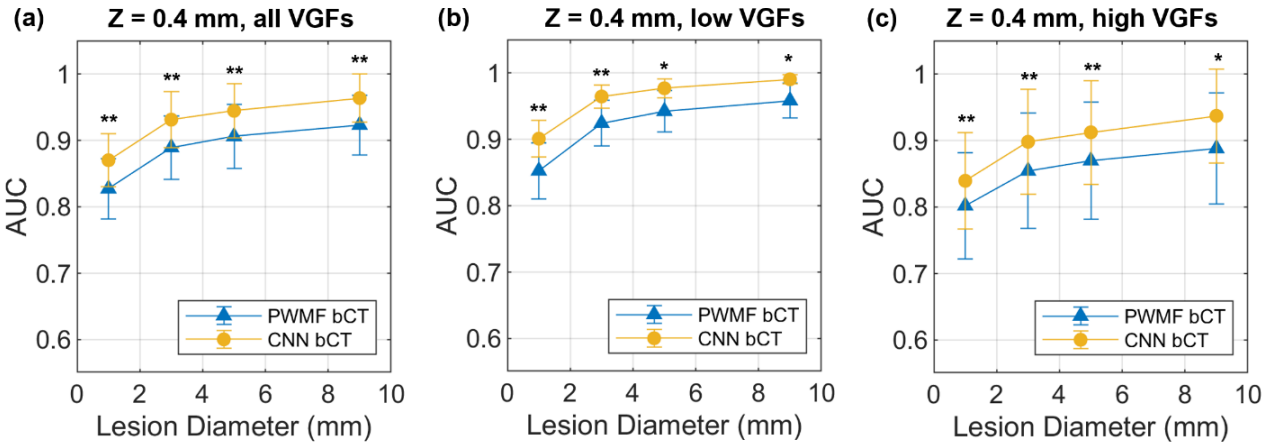

Figure 3.4: Comparison of PWMF and CNN model observers on breast CT images displayed in the native section thickness (Z = 0.4 mm) as a function of lesion diameter across (a) all VGFs (N=24), (b) low VGFs (N=12), and (c) high VGFs (N=12). Paired t-tests were used with Bonferroni correction to adjust for multiple comparisons. One asterisk (*) is used to indicate p < .05, and two asterisks (**) are used to indicate p < .01. Error bars correspond to 95% confidence intervals for each performance estimate.

3.3.2.2. Model observer comparison across section thickness

Model observer performance on breast CT images containing 1- and 5- mm lesions were averaged across all breast densities and plotted as a function of section thickness in **Figures 3.5a** and **3.6a,** respectively. Across all section thicknesses for both lesion sizes, the CNN outperformed the PWMF ($p < .01$). These findings were further analyzed in the context of breast density in **Figures 3.5b-c** and **3.6b-c**. Of the 24 testing data sets, breast CT patches extracted from patients with lower VGF breasts (N = 12) were evaluated and the mean AUC was plotted in **Figures 3.5b** and **3.6b**, and breast CT patches extracted from patients with higher VGF breasts (N = 12) were evaluated and the mean AUC was plotted in **Figure 3.5c** and **3.6c**. The CNN outperformed the PWMF in higher- and lower-density breasts for the detection of both lesion sizes ($p < .05$) except when detecting the 1 mm in high VGF breasts

in section thicknesses of 1.2 and 2 mm. As might be expected, both observers detected larger lesions better than smaller lesions.

Figure 3.5 shows that the thinnest section (0.4 mm) is not the ideal display thickness for detecting small (≤ 1 mm) lesions. Rather, the 1.2- and 2-mm section thicknesses enabled peak detection for both model observers. For the 5 mm lesion (**Figure 3.6**), detectability was minimally affected by section thickness, and detection performance decreased only slightly when section thickness exceeded 6 mm. In **Figure 3.6**, we observe that CNN detection performance of 5 mm lesions is minimally dependent on breast density. In comparison, the PWMF detection performance decreases for higher VGF.

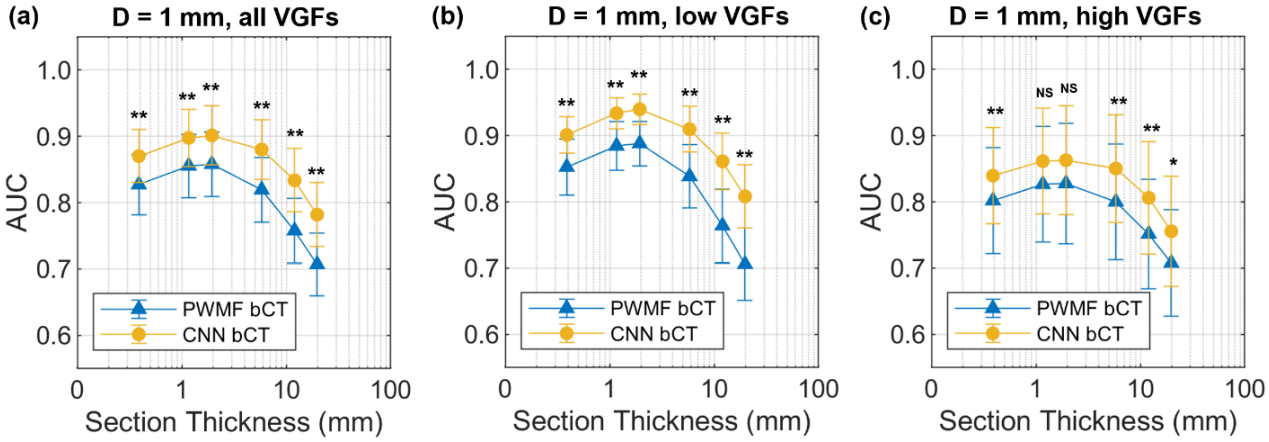

Figure 3.6: Comparison of PWMF and CNN model observers on breast CT images with 1 mm lesions as a function of section thickness across (a) all VGFs (N=24), (b) low VGFs (N=12), and (c) high VGFs (N=12). Paired t-tests were used with Bonferroni correction to adjust for multiple comparisons. One asterisk (*) is used to indicate $p < .05$, and two asterisks (**) are used to indicate $p < .01$. Error bars correspond to 95% confidence intervals for each performance estimate.

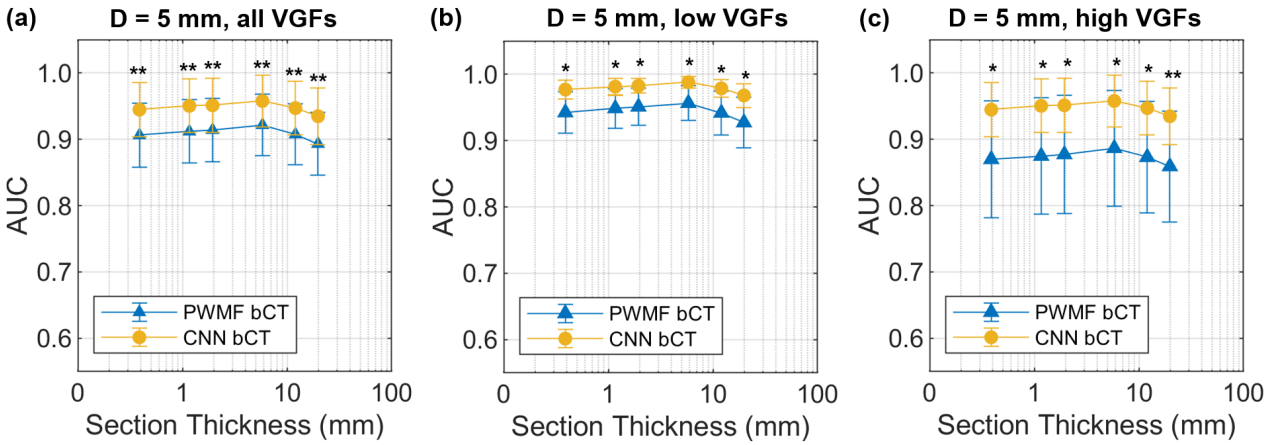

Figure 3.6: Comparison of PWMF and CNN model observers on breast CT images with 5 mm lesions as a function of section thickness across **(a)** all VGFs (N=24), **(b)** low VGFs (N=12), and **(c)** high VGFs (N=12). Paired t-tests were applied with Bonferroni correction to adjust for multiple comparisons. One asterisk (*) is used to indicate $p < .05$, and two asterisks (**) are used to indicate $p < .01$. Error bars correspond to 95% confidence intervals for each performance estimate.

3.4. Discussion

In this study, a CNN model observer was compared to a more conventional model observer, the PWMF, for the binary detection task of detecting a SKE mass lesion. The PWMF is a well-established linear detection filter that has proven to be the ideal observer in background conditions with stationary Gaussian noise. CNNs have likewise been implemented as potential ideal observers given their ability to learn and extract relevant features from images even in the presence of complex backgrounds.

The model observers were used to detect mass lesions in Gaussian background to better understand the performance of the CNN in the context of a known ideal observer (the PWMF). The added signal and simulated Gaussian noise were matched to the mean signal and mean noise power spectrum of breast CT images. We observed that the CNN performed nearly identically to the PWMF across lesion sizes and section thicknesses, with the largest absolute difference between AUCs being 0.0096. This result suggests that the CNN is effectively an ideal observer in Gaussian textures and that it extracts at least as much diagnostic information from an image as the PWMF does. It is likely that the CNN model can be applied to non-Gaussian imaging conditions and at minimum it would perform at least as well as the PWMF.

The model observers were used to detect mass lesions mathematically inserted into breast CT images with real anatomical background. The CNN outperformed the PWMF in detecting mass lesions across lesion size, most section thicknesses, and breast density. As expected, both observers detected larger lesions better than smaller lesions. For both observers, the optimal section thickness of display was between 1.2 - 2.0 mm, the equivalent of 3-5 reconstructed slices, and reduced detection performance was observed in thinner and thicker sections. We suspect that quantum noise interference contributes to reduced

detection performance in thinner sections, and that anatomical noise contributes to reduced detection performance in thicker sections due to the superposition of fibroglandular anatomy. Packard et al. observed similar results in a previous study[40].

It is notable that the CNN outperformed the PWMF in breast CT background while in Gaussian conditions the two observers performed equally. These findings suggest that there is a substantial amount of diagnostic information in breast CT images that the CNN captures that is not accessible to the PWMF. The PWMF employs only the mean signal and mean noise power spectrum of an image to formulate a decision variable. The PWMF and linear model observers in general are evidently limited. In contrast, neural networks, which are non-linear, can perform (perhaps ideally) in an SKE setting. Our results should serve as motivation for future studies that identify the specific informative features that allow the CNN to outperform the PWMF (e.g., using reverse-correlation methods[61]).

Previous studies have demonstrated that the PWMF is the ideal observer in Gaussian image backgrounds[65,68,69]. This study suggests that in non-Gaussian backgrounds (such as breast CT), the PWMF fails to recognize higher order statistical information and image features, whereas the CNN clearly yields superior performance. Furthermore, previous studies in our laboratory suggested that the PWMF outperforms human observers (i.e. expert breast radiologists) in detecting simulated lesions embedded in real breast CT background across the parameters of lesion size and section thickness[30]. Therefore, these studies collectively suggest that the CNN observer may be most appropriate for estimating peak performance in breast CT images.

This study had limitations. Model observers were compared for the detection of relatively simple signals: SKE-LKE mass lesions. Microcalcifications were not simulated in this study, and an appropriate simulation would require more complex modeling of partial

volume effects. Previously, human observer studies have indicated that microcalcifications are more difficult to detect than mass lesions in breast CT[16]. Therefore, to fully understand the utility of PWMF and CNN model observers in breast CT, an evaluation of their abilities to detect microcalcifications is necessary. Future studies will investigate this. A three-layered CNN was used in this study. We recognize that for more complex detection tasks, a deeper architecture may be required and that additional training methods such as transfer learning may be useful. Furthermore, we recognize that the CNN performed exceptionally well in this study primarily due to two reasons: 1) the signal was known exactly, and 2) there was a large amount of labeled training data. These conditions were only possible since we mathematically simulated the training signals, albeit on actual breast anatomy. Finally, we chose the PWMF among other linear observers for this study. Future studies comparing CNNs with other linear observers such as the Hotelling observers may be useful to underscore the utility of the CNN-based observer in breast CT.

3.5. Conclusion

In this study, we used PWMF and CNN model observers to detect SKE mass lesions in patient breast CT images. The CNN outperformed the PWMF across lesion size, most section thicknesses, and breast density. We conclude that the CNN captures more diagnostic information from breast CT images than the PWMF and may be a more suitable observer when conducting optimal performance studies.

While model observer studies are important, they do not fundamentally replace the need for human observer studies. However, there is an increasing emphasis on virtual clinical trials in the literature. The power in these model observer studies is of course that many more lesions and lesion placements can be studied than with human observers, and this provides

the ability to generate statistically meaningful results which can aid in optimizing breast CT parameters prior to human observer studies.

Chapter 4: Microcalcification detectability in hybrid breast CT images using CNN observers

4.1. Introduction

Microcalcifications are key indicators of potential early-stage breast cancer which can manifest in various morphologies and distributions[12–14]. The American College of Radiology has released a BI-RADS report[11] on morphological patterns of microcalcifications commonly seen in mammography along with their potential for malignancy. Large rod-like calcifications, for instance, are typically benign and do not necessitate further diagnostic imaging or biopsy. Clustered microcalcifications – dense groupings of small (< 1 mm) calcifications within a small region – are likely to be malignant[80,81], and generally necessitate biopsy. During breast cancer screening exams, it is crucial to accurately detect and characterize microcalcifications.

Digital mammography has historically been used for breast cancer screening, and its capabilities for imaging with high resolution have been valuable for microcalcification detection[82,83]. Mammography's sensitivity substantially decreases, however, when used to image patients with dense breasts due to the superposition of fibroglandular tissue which can obscure lesions. In recent years, breast computed tomography (CT) has emerged as a promising alternative imaging tool[16,84], with the key advantage of accessing fibroglandular anatomy inside the breast without the superposition of neighboring tissue. Initial clinical studies[16] suggested that breast CT is better than mammography at detecting mass lesions, but that mammography is better than breast CT at detecting microcalcifications. Since this study was published, our laboratory at UC Davis has developed a higher-resolution breast CT scanner which achieves nearly four times the spatial resolution than that of the earlier generation scanner based on their system modulation transfer functions[18]. This scanner's

ability to detect microcalcifications is under evaluation, and protocols for optimizing the scanner for microcalcification detection are being developed. Ideally, such optimization would involve extensive clinical trials with human observers to evaluate image quality, diagnostic performance, and the impact of various imaging parameters. However, the challenges of conducting large-scale clinical trials, such as time and cost limitations, necessitate the exploration of alternative approaches.

Simulation studies have been proposed as an alternative to clinical studies, where synthetic images are generated using phantom imaging[20,85,86] or computer simulations[40,41,87–89]. These studies are also called "virtual clinical trials" [22–24]. In this work, hybrid images are computer simulated, where mathematically generated microcalcifications are embedded into clinical patient breast CT images acquired at our institution using the high-resolution breast CT scanner. This approach enables the investigation of many lesion-related parameters (e.g., size) while preserving crucial patient- (e.g., breast density) and imaging-related parameters (e.g., resolution). To detect the simulated microcalcifications, a convolutional neural network (CNN) is used in lieu of a human observer. CNN-based model observers are thought to approximate human visual perception[62,74,90], providing an efficient and reproducible means of quantifying microcalcification detectability across several important parameters. CNNs have shown to be useful and versatile across imaging contexts including breast CT[62,74,75]. While CNNs do not necessarily accurately predict human observer performance, they can be useful in assessing trends across individual parameters as well as the interplay between multiple parameters that affect detection performance[76,90]. In this study, 2D and 3D CNN models are used to detect individual microcalcifications and microcalcification clusters embedded in 109 patient breast CT data sets, and parameters related to calcification size, calcification contrast,

cluster diameter, cluster density, and image display method (i.e., single slices, slice averaging, and maximum-intensity projections) are studied.

4.2. Methods

4.2.1. Breast CT system

Four generations of pendant geometry, cone-beam breast CT scanners have been designed and fabricated in our laboratory over the past two decades. The fourth-generation prototype scanner follows a similar design as previously reported scanners[16] but achieves nearly four times the spatial resolution of the early scanners based upon modulation transfer function (MTF) analysis[18]. This is due to the combination of a pulsed X-ray source, a smaller focal spot, and a higher resolution flat-panel detector. A comparison of system MTFs for the second (Bodega), third (Cambria), and fourth (Doheny) generation breast CT systems is shown in **Figure 4.1**[18].

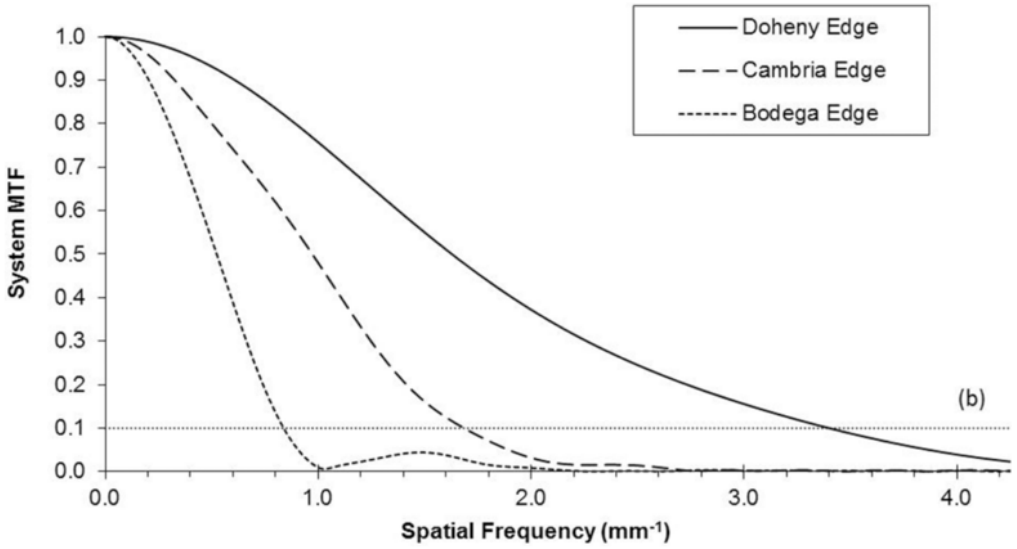

Figure 4.1: From Gazi *et al.*: Comparison of system MTFs for second (Bodega), third (Cambria), and fourth (Doheny) generation breast CT systems at most challenging periphery of the FOV.

4.2.2. Patient images

An IRB-approved clinical trial was conducted at UC Davis Medical Center evaluating the fourth-generation breast CT system as a tool for breast cancer screening and for the diagnostic breast exam. Patients with suspicious lesions (BI-RADS 4 or 5) based on screening breast imaging (mammography or tomosynthesis) were eligible to participate in the study. The scanning protocol involved four sequential scans of the 1) unaffected breast prior to contrast-injection, 2) affected breast prior to contrast injection, 3) affected breast after contrast injection, and 4) unaffected breast after contrast injection. After scans were acquired, patients underwent breast biopsy on areas of suspicion based upon standard of care which provided the ground truth diagnosis. To date, 58 women have been scanned on the Doheny breast CT system resulting in 222 breast CT volume data sets. Contrast injection is

advantageous for visualizing malignant solid tumors due to local perfusion of contrast agent through "leaky" angiogenetic vessels,[38,39] and this contrast greatly improves lesion conspicuity[60]. Because this study focuses on microcalcification detectability, only pre-contrast volume data sets were used to avoid contrast enhancement in and around microcalcifications. In total, 109 pre-contrast volume data sets were selected for this study to be used as anatomical background for mathematical microcalcification insertion. Each volume data set contained 800-900 reconstructed slices (1024 × 1024 matrix size) with isotropic voxel dimensions of 0.2 mm.

4.2.3. Microcalcification simulation

Spherical microcalcification profiles were generated using methodology developed by Hernandez et al.[20] First, a single high-resolution microcalcification profile was generated by inserting a sphere of intensity H in the center of a $330 \times 330 \times 330$ matrix with an isotropic voxel dimension of 0.01 mm. Then, the volume was blurred in the xy plane by (i) converting each coronal slice to the frequency domain using the 2D Fourier transform, (ii) multiplying the slice by the 2D coronal plane MTF measured on the breast CT system, and (iii) converting the blurred slice to the spatial domain using the 2D inverse Fourier transform. The volume was then blurred in the z dimension by (i) converting each vector in the z-direction to the frequency domain using the 1D Fourier transform, (ii) multiplying the vector by the 1D z-direction MTF measured on the breast CT system, and (iii) converting the blurred vector to the spatial domain using the 1D inverse Fourier transform. The MTFs were not assumed to be equal in the xy and z directions due to the role of the Shepp-Logan reconstruction kernel in the xy dimensions but not the z dimension. The high-resolution profile was then down sampled to match the voxel dimensions of the breast CT system (0.2 mm), resulting in a

$17 \times 17 \times 17$ matrix, $m(x, y, z)$. The microcalcification profile generation process is illustrated in **Figure 4.2** for a 0.2 mm diameter calcification.

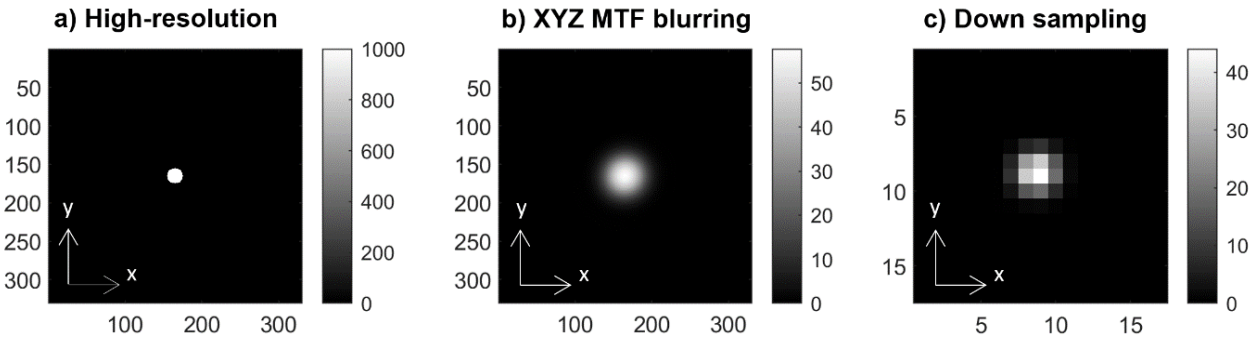

Figure 4.2: Simulation of a 0.2 mm microcalcification profile. As a (a) mathematically generated high-resolution microcalcification undergoes (b) XYZ MTF blurring and (c) down sampling, the microcalcification's edges are blurred and signal is attenuated.

In addition to edge blurring, the blurring procedure resulted in grayscale reduction due to partial volume effects. Let the intensity of a simulated high-resolution calcification, H, be 1000 HU, and the intensity of the background volume, B, be -200 HU, as shown in **Figure 4.2**. Let y be the peak intensity in HU of the calcification after the blurring procedure described above. y was computed for a range of computer-generated calcification diameters between 0 and 1 mm. A plot of peak intensity as a function of calcification diameter in mm, D_{calc}, is shown in **Figure 4.3**.

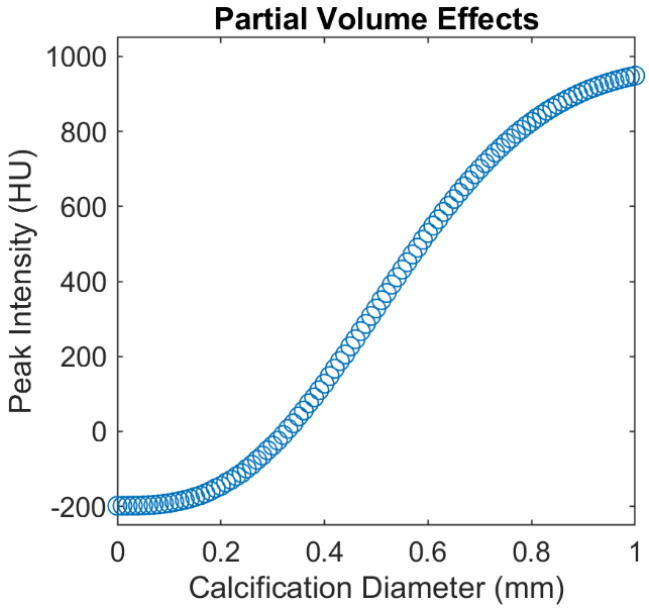

Figure 4.3: Peak intensity in HU of calcifications after blurring as a function of calcification diameter, D_{calc}showing the partial volume effects in the system. The grayscale values of the initial high-resolution calcification and background were 1000 HU and -200 HU, respectively.

Based on **Figure 4.3**, it was found that intensity reduction resulting from the blurring procedure can be modeled using a logistic function for any input intensity in HU, H, calcification diameter in mm, D_{calc}, and background intensity in HU, B, using **Equation 4.1**:

$$y(D_{calc}, B, H) = B + \frac{H-B}{1+e^{-\left(\frac{D_{calc}-.52}{.15}\right)}} \quad (4.1)$$

Equation 4.1 is unique to our imaging system but illustrates the key principle of partial volume that occurs in any imaging system: objects smaller than the spatial resolution of the imaging system experience blurring and signal attenuation when imaged. In **Figure 4.**, as the calcification diameter approaches zero, the peak intensity of the blurred calcification approaches the intensity of the background, B, and as calcification diameter exceeds 1 mm, the peak intensity of the blurred calcification approaches the native intensity, H. In this study,

B was computed for each simulated microcalcification as the median intensity of a $7 \times 7 \times 7$ subvolume of anatomical background surrounding the location of microcalcification insertion.

H was estimated empirically by applying **Equation 4.1** to select calcifications found in our clinical patient data sets. Two large calcifications in two separate patients were identified using the patients' screening exam radiology reports and verified using the corresponding pathology reports. Only large, pathology-confirmed calcifications were used for determining D_{calc} in order to maximize accuracy in delineating calcification edges. Custom-built breast CT viewing software was used to measure the diameter and peak intensity for each calcification. The median intensity of the tissue surrounding the calcifications was also measured. These variables were applied to **Equation 4.1** and resulted in native intensity values, H, of 361 HU and 768 HU. The range of native intensity values corresponds to varying densities and compositions of individual calcifications in our clinical data sets[12,91].

4.2.4. Hybrid image generation

For every breast CT volume, lesion-present and lesion-absent region of interest (ROI) patches were generated at random locations within the breast. For each ROI, a lesion center was computer generated using a random number generator and kept if the surrounding volume $V[i,j,k]$ was fully contained within the patient's breast and did not contain skin. Otherwise, a new lesion center was computer generated. For viable lesion centers, the surrounding volume $V[i,j,k]$ served as the anatomical background for mathematical microcalcification insertion. The lesion centers were saved and used for every iteration of ROI generation in this study to reduce variability, but the selection of training ROIs were different than testing ROIs because different patient data sets were held-out for testing as explained in the following section.

4.2.4.1. 2D image generation

First, 2D hybrid image ROIs were generated such that each ROI contained a single microcalcification of known diameter at the center of the ROI. This setting can be described as signal-known-exactly (SKE), location-known-exactly (LKE). SKE, LKE tasks can be used to establish baselines for evaluating an observer's detection performance prior to experimenting with unknown or uncertain signals and locations.[92]

Let $V[i,j,k]$ be a 120 × 120 × 120 anatomical background volume, $M[i,j,k]$ be a volume of the same size with a blurred microcalcification profile $m[x,y,z]$ in the center, and zeros elsewhere, and $V_{sim}[i,j,k]$ be the simulated volume with the inserted calcification. $M[i,j,k]$ is scaled such that when it is added to $V[i,j,k]$, the peak intensity of $V_{sim}[i,j,k]$ equals $y(D_{calc},B,H)$ from **Equation 4.1**. This process simulates the partial volume-related blurring of a calcification with surrounding voxels. Let $[i_{max},j_{max},k_{max}]$ be the indices of the peak intensity voxel in $M[i,j,k]$. The simulated volume $V_{sim}[i,j,k]$ is then defined as:

$$V_{sim}[i,j,k] = V[i,j,k] + \left(\frac{y(D_{calc},B,H) - V[i_{max},j_{max},k_{max}]}{M[i_{max},j_{max},k_{max}]} \times M[i,j,k]\right) \quad (4.2)$$

$V_{sim}[i,j,k]$ was generated at each lesion center and displayed in 2D by either extracting the center slice of $V_{sim}[i,j,k]$ in the axial view plane or by slice averaging across adjacent slices to model thicker sections. Similarly, lesion-absent ROIs were generated and displayed in 2D by either extracting the center slice of $V[i,j,k]$ in the axial view plane or by slice averaging across adjacent slices. The resulting 2D ROIs had dimensions of 120 × 120 × 1. Five microcalcification diameters D_{calc} (0.04, 0.10, 0.15, 0.20, 0.40 mm), three input intensities H (361, 565, 768 HU), and five section thicknesses (0.2, 0.6, 1.0, 3.0, 11.0 mm) were studied. The equivalent number of slices n_{slices} for each section thickness was 1, 3, 5, 15, and 55 slices, respectively.

Out of the 109 total breast CT volume data sets, 99 data sets (N = 99, ~90%) were used to generate training ROIs for training the model observer. For each volume data set, 200 lesion-present ROIs and 200 lesion-absent ROIs were generated from different lesion centers resulting in 39,600 total training ROIs. 10% of the training ROIs (3960 ROIs) were reserved for validation during the training process. The remaining 10 breast CT volume data sets (N = 10, ~10%) were used to generate a testing data set. Again, 200 lesion-present ROIs and 200 lesion-absent ROIs were generated from each volume data set resulting in 4000 total testing ROIs. Sample lesion-present ROIs of simulated SKE LKE microcalcifications are shown in **Figures 4.4 & 4.5.**

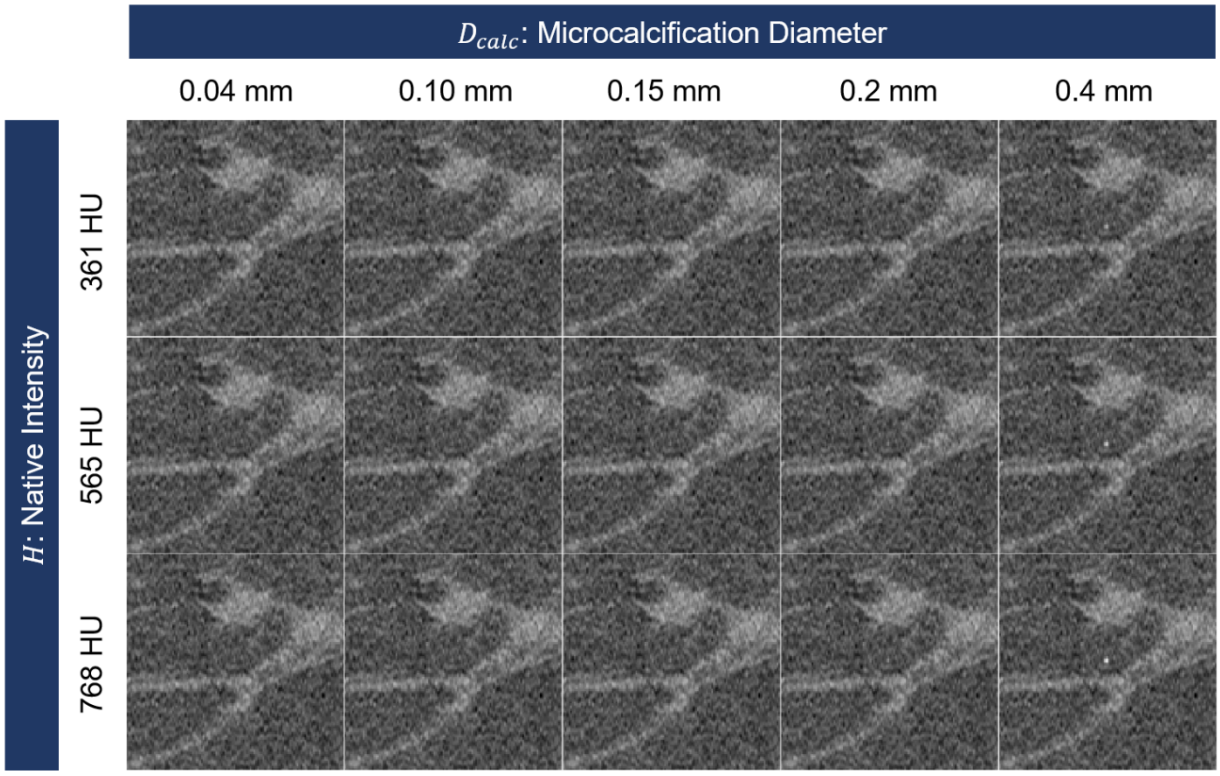

Figure 4.4: Example ROIs of a single simulated microcalcification centered in the field of view. Microcalcification diameter is varied across columns. Native intensity of microcalcifications prior to the blurring procedure (H) is varied across rows. Partial volume effects reduce microcalcification conspicuity as microcalcification diameter becomes smaller than the resolution of the imaging system (0.2 mm). Field of view within each ROI: 24 mm × 24 mm.

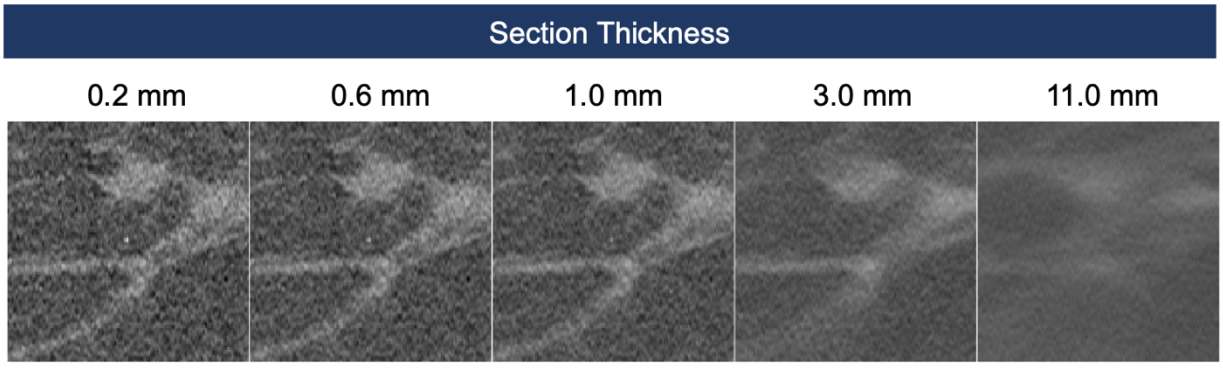

Figure 4.5: Example ROIs of a simulated 0.4 mm microcalcification displayed using different section thicknesses. Field of view within each ROI: 24 mm × 24 mm.

4.2.4.2. 3D image generation

3D hybrid image volumes were generated with simulated microcalcification clusters. A single cluster was inserted at the center of every volume of interest (VOI), and microcalcifications of uniform diameters were randomly placed within the cluster. This task had similarities to location-known-exactly tasks because the cluster was centered at a known location, but similarities as well to location-known-statistically tasks because individual calcifications were randomly placed within the cluster across different lesions.

Let $V[i,j,k]$ be a 50 × 50 × 50 anatomical background volume at a random lesion center in the breast. Let $M_n[i,j,k]$ be a volume of zeros of the same size. A spherical boundary is defined within $M_n[i,j,k]$ based on the cluster diameter, $D_{cluster}$, representing the boundary of the cluster. A random number generator was used to define the location, $[i_n, j_n, k_n]$, within the cluster where a blurred microcalcification profile, $m[x,y,z]$, was inserted. $m[x,y,z]$ is inserted at $M_n[i_n, j_n, k_n]$. $M_n[i,j,k]$ is then scaled such that when it is added to $V[i,j,k]$, the peak intensity of $V_{sim}[i,j,k]$ equals $y_n(D_{calc}, B, H)$ from **Equation 4.1**. Microcalcifications are repeatedly inserted at random locations within the cluster until the desired number of microcalcifications, N_{calcs}, are inserted. Because $y_n(D_{calc}, B, H)$ is computed at every location $[i_n, j_n, k_n]$ within the cluster, microcalcifications within the same cluster varied in intensity, as is common in breast CT images. Let $[i_{max}, j_{max}, k_{max}]$ be the indices of the peak intensity voxel in $M_n[i,j,k]$. The simulated VOI with inserted clusters is defined as:

$$V_{sim}[i,j,k] = V[i,j,k] + \sum_{n=1}^{N_{calcs}} \left(\frac{y_n(D_{calc},B,H) - V[i_{max},j_{max},k_{max}]}{M_n[i_{max},j_{max},k_{max}]} \times M_n[i,j,k] \right) \quad (4.3)$$

$V_{sim}[i,j,k]$ volumes (50 × 50 × 50) were used as lesion-present VOIs, and $V[i,j,k]$ volumes (50 × 50 × 50) generated from separate lesion centers were used as lesion-absent VOIs.

Thicker sections were modeled by averaging every n_{slices} slices, resulting in 50 × 50 × N VOIs, where N is defined as:

$$N = \left\lceil \frac{50}{n_{slices}} \right\rceil \tag{4.4}$$

The maximum intensity projection (MIP) was also generated from each VOI in the axial view plane, resulting in 50 × 50 × 1 ROIs. Five microcalcification diameters D_{calc} (0.20, 0.25, 0.30, 0.35, 0.40 mm), six N_{calcs} (1, 3, 5, 7, 10, 15), and six cluster diameters $D_{cluster}$ (1, 3, 5, 6, 8, 10 mm) were studied. Six section thicknesses (0.2, 0.6, 1.0, 2.2, 3.0, 9.8 mm) in addition to the MIP were studied to understand the role of image display method on microcalcification detectability. The equivalent number of slices n_{slices} for each section thickness was 1, 3, 5, 11, 15, and 49 slices, respectively. Native intensity H was fixed at 565 HU.

Out of the 109 total breast CT volume data sets, 99 data sets (N = 99, ~90%) were used to generate the training data set for training the model observer. For each volume data set, 150 lesion-present VOIs and 150 lesion-absent VOIs were generated from different lesion centers resulting in 29,700 total training VOIs. 10% of the training VOIs (2970 VOIs) were reserved for validation during the training process. The remaining 10 breast CT volume data sets (N = 10, ~10%) were used to generate the testing data set. Again, 150 lesion-present VOIs and 150 lesion-absent VOIs were generated from each volume data set resulting in 3000 total testing VOIs. Sample VOIs of simulated microcalcification clusters displayed using MIP are shown in **Figure 4.6.**

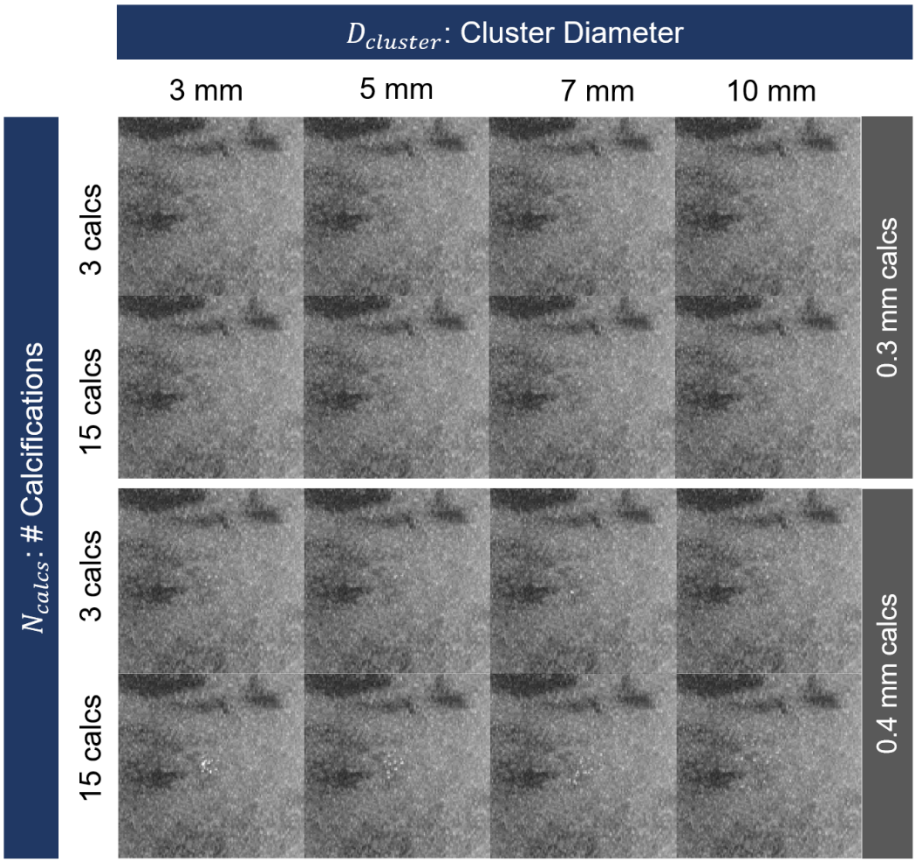

Figure 4.6: Example images of simulated microcalcification clusters displayed using maximum intensity projection (MIP). Calcification diameter, cluster diameter, and number of calcifications are varied. Field of view within each ROI: 24 mm × 24 mm.

4.2.5. Detectability estimation using CNN

4.2.5.1. 2D CNN

A 2D convolutional neural network (CNN) was used to detect simulated microcalcifications in the SKE, LKE image data sets. The input to the CNN was a 2D ROI (120 × 120 × 1), and the output was a decision variable between 0 and 1 scaled by the sigmoid function. The decision variables were used for ROC curve analysis to estimate overall detectability of a data set.

The CNN architecture is shown in **Figure 4.7.** The CNN consisted of three layers: two convolutional layers and one fully-connected layer. The choice of the three-layered architecture was due to the relative simplicity of the SKE, LKE detection task, similar to previous studies[93]. The convolutional layers used 3 × 3 convolutional filters with strides of 1. Batch normalization was implemented after the first convolutional layer. Max pooling was then implemented after both convolutional layers with a pool size of 2 × 2 and a stride of 1. Dropout was implemented after both max pooling layers with a rate of 0.2, and after the fully-connected layer with a rate of 0.5. The rectified linear unit (ReLU) activation function was applied in all three layers. The binary-cross entropy (BCE) loss function was used:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\tilde{y}) + (1 - y_i)\log(1 - \tilde{y_i})] \quad (4.5)$$

where y is the ground truth label (0 or 1), $\tilde{y}$ is the predicted value, and N is the number of samples. The training metric was accuracy:

$$Accuracy = \frac{\#\ correct\ predictions}{\#\ total\ predictions} \quad (4.6)$$

The CNN model was implemented in Python using the Keras library[79]. The Adam optimizer[78] was used with a learning rate of 1e-5 and a batch size of 64. Training ran between 50 – 150 epochs. An NVIDIA GeForce GTX 1080 GPU was used.

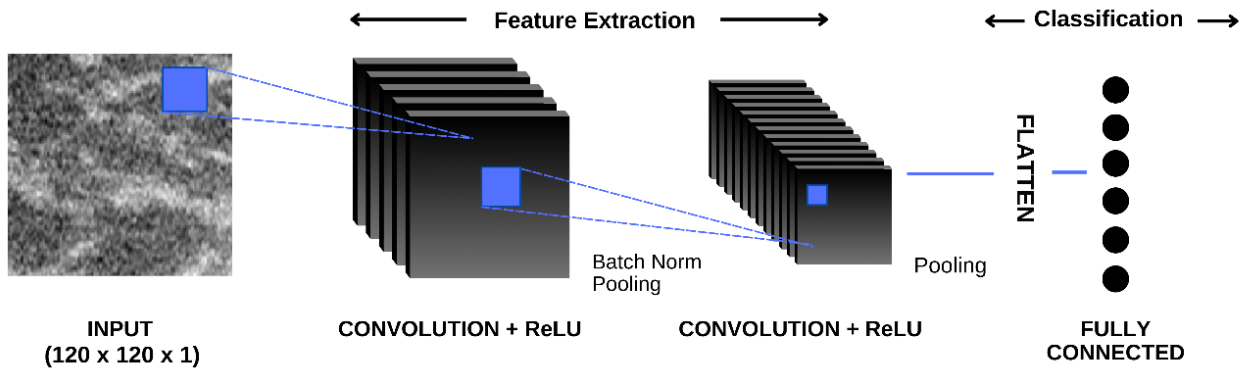

Figure 4.7: 2D CNN architecture: two convolutional layers followed by a fully-connected layer.

4.2.5.2. 3D CNN

A 3D CNN was used to detect simulated microcalcification clusters in the hybrid 3D image VOIs. The input to the CNN was a 50 × 50 × N VOI, where N varied with section thickness, and the output was a scalar-valued decision variable between 0 and 1 scaled by the sigmoid function. The 3D CNN model was designed to mimic the sequential 2D analysis performed by human observers when examining 3D breast CT volumes. Human observers review 3D image volumes slice-by-slice without immediate access to true 3D depth information. Findings from each slice are synthesized into one classification decision for each image volume. To mimic this process, a 3 × 3 × 1 convolutional kernel was used instead of a 3 × 3 × 3 convolutional kernel, which is commonly employed for 3D image data. The 3 × 3 × 1 kernel capitalizes on local spatial information within each slice while disregarding inter-slice information. The CNN synthesizes information from each slice into one decision variable in the fully-connected layer. This kernel choice also suited the range of slices comprising different data sets owing to section thickness: slice averaging across the entire VOI resulted

in a single 50 × 50 × 1 slice which required 2D convolutional kernels, while VOIs displayed in the native section thickness contained 50 slices.

The CNN consisted of four layers: three convolutional layers and one fully-connected layer. While the 2D and 3D CNN architectures were similar, an extra convolutional layer was added to the 3D CNN due to the difficulty of detecting relatively small targets (1-15 calcifications) compared to the background space (up to 50^3 voxels). Moreover, the microcalcifications in the 3D generated data sets were signal-known-statistically (SKS), whereas the microcalcifications in the 2D generated data sets were SKE, LKE, increasing the complexity of the detection task. All three convolutional layers used 3 × 3 × 1 convolutional kernels with strides of 1. Batch normalization was implemented after the first convolutional layer. Max pooling was implemented after all three convolutional layers with a pool size of 2 × 2 × 1 and a stride of 1. The pooling layers further downscaled the spatial dimensions of the feature maps, reducing computational complexity and providing translational invariance. Dropout was implemented after the three max pooling layers with a rate of 0.2, and after the fully-connected layer with a rate of 0.5 to enhance generalization by randomly deactivating neurons during training. The rectified linear unit (ReLU) activation function was applied in all four layers.

The 3D CNN model was trained using the Adam optimizer with a learning rate of 1e-5, a batch size of 64, and the BCE loss function. Training occurred over 100-300 epochs. More epochs were required for training the 3D CNN compared to the 2D CNN due to the increased difficulty of the task. The model's performance was evaluated using accuracy metrics. The CNN was implemented in Python using an NVIDIA GeForce GTX 1080 GPU. The 3D CNN model with the 3 × 3 × 1 convolutional kernel is shown in **Figure 4.8**.

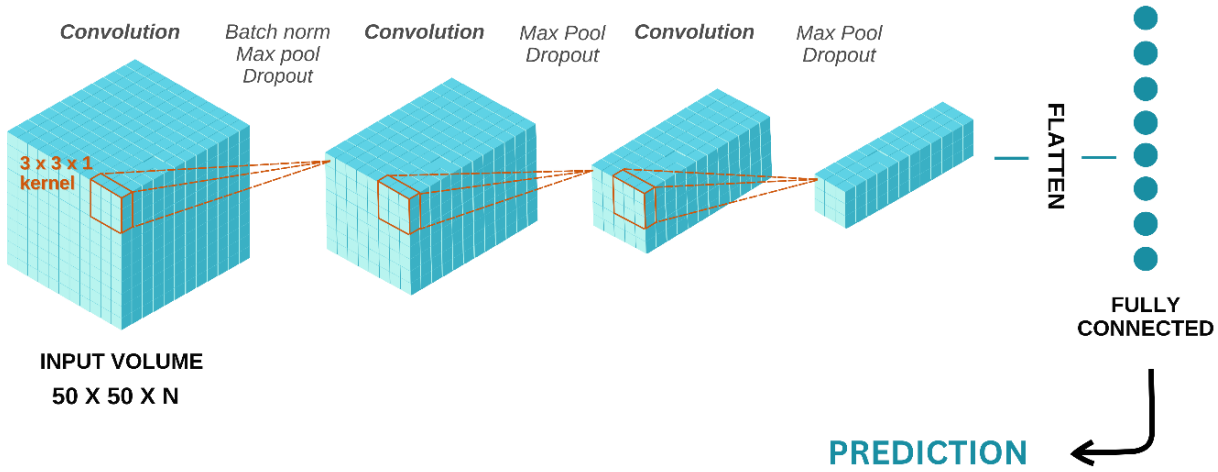

Figure 4.8: 3D CNN architecture: three convolutional layers followed by a fully-connected layer.

4.2.6. Performance evaluation and statistical analysis

CNN-generated decision variables were used with receiver operating characteristic (ROC) curve analysis to estimate overall detection performance. For all decision variables related to one testing breast CT volume, an empirical ROC curve was constructed by plotting the true positive rate against the false positive rate at various threshold values. The area under the ROC curve (AUC) was computed, signifying detection performance in that breast CT volume. AUCs were computed on a breast-volume basis to elucidate variability related to breast density, anatomical and quantum noise, and motion artifacts. To estimate overall detection performance across all the testing breast CT volumes, the mean and standard deviation of testing-volume AUCs were also computed. In **Section 4.3**, $\overline{AUC}$ was plotted with 95% confidence error bars to represent uncertainty in the detectability estimations. 95% confidence error bars were computed using **Equation 4.7:**

$$CI_{95} = \overline{AUC} \pm 1.96 \times \frac{\sigma}{\sqrt{K}} \quad (4.7)$$

where CI_{95} is the 95% confidence interval, $\overline{AUC}$ is the mean AUC across K testing breast CT volumes, σ is the standard deviation of AUCs across K testing breast CT volumes, and K is the number of testing volumes (K = 10). $\overline{AUC}$ was the primary performance metric in this study.

For all comparative tests, the Mann-Whitney U-test was employed. All tests were two-sided, and statistical significance was defined as a difference with a p-value less than 0.05. In cases where multiple comparisons were conducted, Bonferroni correction was applied by dividing the desired significance level by the number of comparisons. Statistical analysis was performed using Matlab (Matlab; TheMathWorks Inc., Natick, MA).

4.3. Results

4.3.1. SKE, LKE detection task

4.3.1.1. Effect of microcalcification size and native intensity

SKE LKE detection performance is plotted as a function of microcalcification diameter for three native intensities in **Figure 4.9**. When the signal and location are known, microcalcifications that are notably smaller than the resolution of the imaging system (0.20 mm) can be detected despite partial volume effects. Detection performance increases monotonically with calcification diameter and native intensity, and maximum detection performance is achieved when calcification diameter equals or exceeds 0.40 mm. As expected, larger and "brighter" calcifications are more conspicuous than smaller and "dimmer" calcifications. For the remainder of the study, H = 565 HU is used.

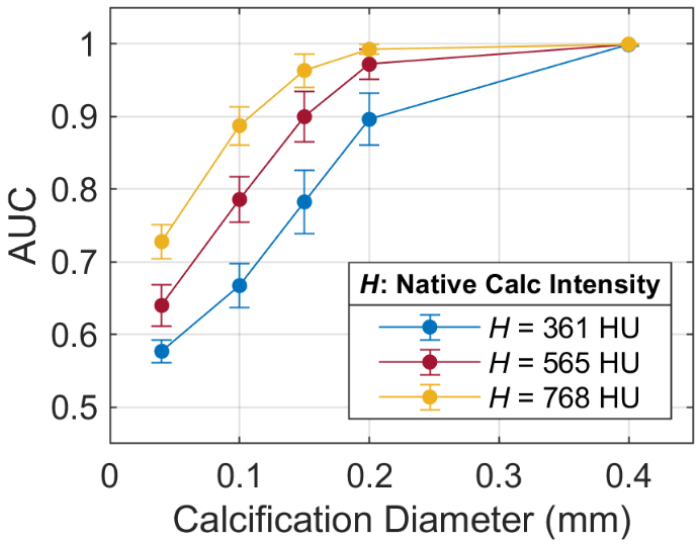

Figure 4.9: Effect of size and native intensity on SKE, LKE microcalcification detectability. A single microcalcification was placed at the center of the ROI for this SKE LKE task. AUC is plotted as a function of calcification diameter D_{calc} for three native intensities H. Error bars correspond to 95% confidence intervals for each estimate.

4.3.1.2. Effect of section thickness

SKE LKE detection performance is plotted as a function of section thickness for five calcification diameters in **Figure 4.10**. For the largest calcification (0.40 mm), detection performance is generally unaffected as section thickness increases until the thickness reaches 1.0 mm (5 slices), where performance begins to decrease. For calcifications smaller than 0.40 mm, thicker sections reduce detection performance. The native section thickness (0.20 mm) of breast CT enables peak detection performance across all calcification diameters, and detection performance degrades precipitously as section thickness increases.

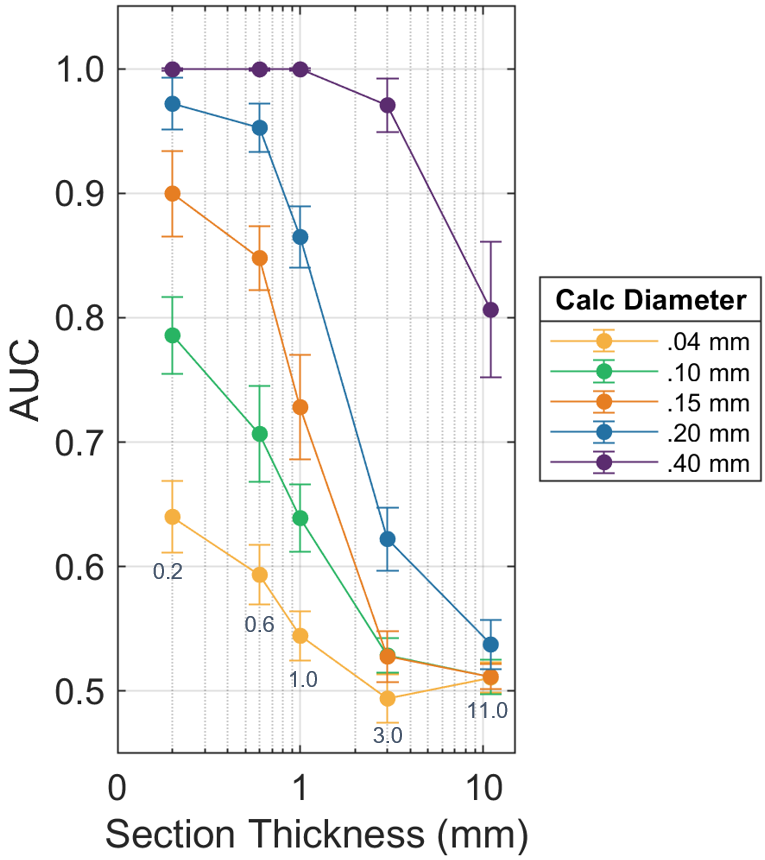

Figure 4.10: Effect of section thickness on SKE LKE microcalcification detectability for five calcification diameters and five section thicknesses (0.2, 0.6, 1.0, 3.0, 11.0 mm). A single microcalcification was placed at the center of the ROI for this SKE LKE task. Error bars correspond to 95% confidence intervals for each estimate.

4.3.2. Cluster detection task

4.3.2.1. Effect of image display method

Figure 4.11 shows the effect of image display method on CNN detection performance of a 5-mm diameter cluster containing 3 microcalcifications. Mean AUC is plotted as a function of section thickness, and detection performance using maximum intensity projection (MIP) is also plotted for five calcification diameters. Data points representing 0.20- and 0.25-mm

calcifications across section thicknesses are omitted from the plot for ease in visualization of other differences.

Detection performance decreases with increasing section thickness, and peak detection performance occurs using the native section thickness (0.2 mm) and the MIP. When slices are averaged across the entire volume (50 slices or 9.8 mm thickness), microcalcifications of all three sizes are indistinguishable from anatomical background (AUC ~ 0.5). For the three microcalcification sizes, there was no statistically significant difference in detectability between using the MIP and using the native section thickness ($p = .08, p = .52, p = .76$ for 0.30, 0.35, and 0.40-mm microcalcifications, respectively). The MIP images contained a single slice while the native section thickness images contained 50 slices, thus computing time for MIP images was nearly 50 times shorter than the computing time of the native section thickness images. For the remainder of the study, MIP was used to display the simulated images.

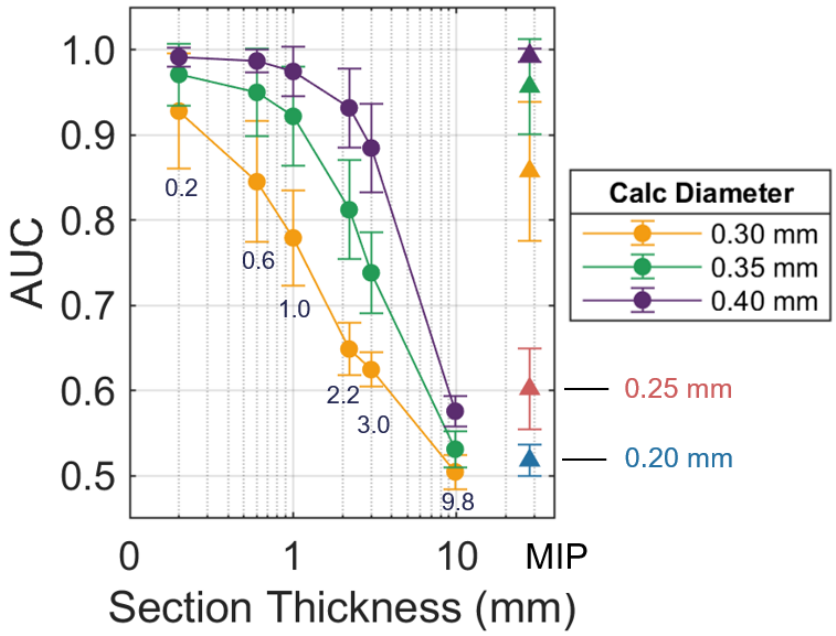

Figure 4.11: Effect of image display method. Detection performance using slice averaging (i.e., section thickness) for a 5-mm cluster of 3 calcifications is plotted for three calcification diameters (0.30, 0.35, 0.40 mm) and denoted using solid circles. Detection performance using maximum intensity projection (MIP) is plotted for five calcification diameters (0.20, 0.25, 0.30, 0.35, 0.40 mm) and denoted using solid triangles. Error bars correspond to 95% confidence intervals for each estimate.

4.3.2.2. Effect of cluster diameter

Figure 4.12 shows the effect of cluster diameter on CNN detection performance for clusters containing 3 microcalcifications displayed using MIP. Mean AUC is plotted as a function of cluster diameter for four calcification diameters.

The 0.40 mm microcalcification diameter data points frame the diagnostic performance of breast CT when the detection target is 3 calcifications. For the 0.40 mm microcalcification, near optimal detection performance is achieved (AUC ~ 1.0), and cluster diameter has minimal effect on cluster detectability. For microcalcifications on the threshold of detectability (i.e., 0.20 mm, 0.25 mm, and 0.30 mm microcalcifications), cluster diameter

becomes important, and larger cluster diameters result in lowered overall detectability, while smaller cluster diameters result in increased overall detectability.

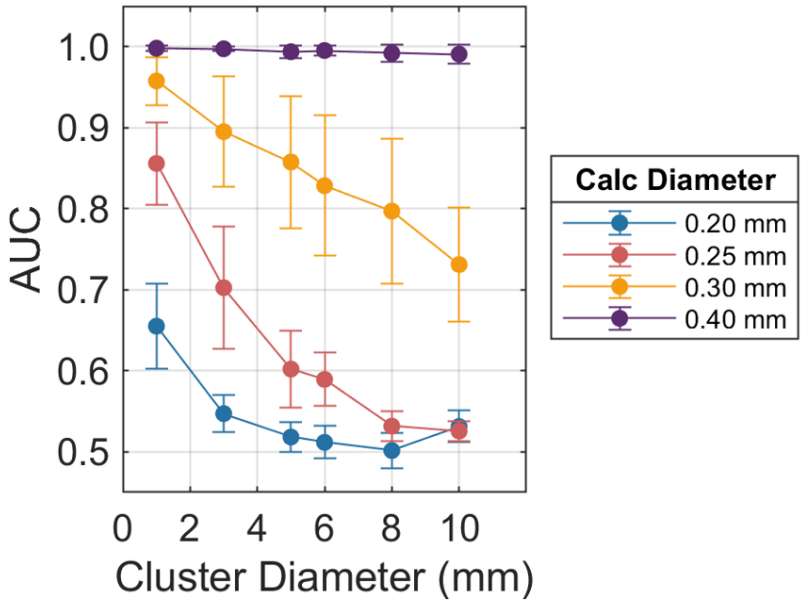

Figure 4.12: Effect of cluster diameter on detection performance for clusters of 3 microcalcifications displayed using MIP. Error bars correspond to 95% confidence intervals for each estimate.

4.3.2.3. Effect of number of calcifications

Figure 4.13 displays the effect of the number of calcifications on CNN detection performance for clusters measuring 5-mm in diameter displayed using MIP. Mean AUC is plotted as a function of number of calcifications for four calcification diameters.

Similar to **Figure 4.12**, **Figure 4.13** shows that clusters of relatively large (0.40 mm) microcalcifications are minimally affected by the number of microcalcifications. For microcalcifications at the threshold of detectability (i.e., 0.20 mm, 0.25 mm, and 0.30 mm microcalcifications), the presence of more microcalcifications improves the overall cluster

detectability, while fewer calcifications within a cluster decreases the overall cluster detectability.

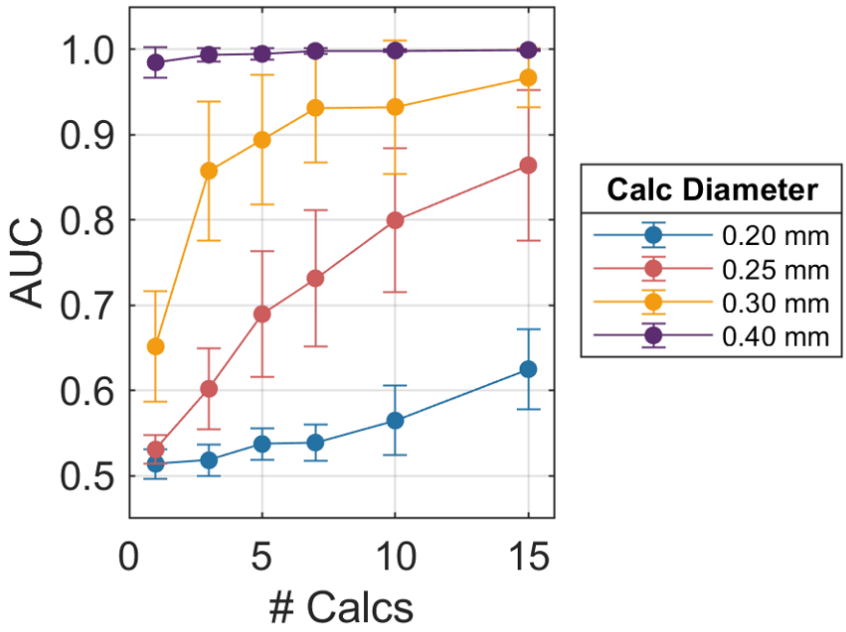

Figure 4.13: Effect of number of calcifications on detection performance for clusters measuring 5-mm in diameter displayed using MIP. Error bars correspond to 95% confidence intervals for each estimate.

4.4. Discussion

Our results demonstrated the utility of the maximum intensity projection (MIP) for displaying image volumes containing microcalcification clusters. We found that there was no statistically significant difference in detection performance when using the MIP compared to the native section thickness, but that thicker sections led to reduced detection performance. This result suggests that the CNN is primarily accessing maximum intensities within breast CT image volumes to determine microcalcification presence. The MIP display method, despite using a single slice, yielded comparable performance to the native section thickness, which employed 50 slices. Reduction in slices did not sacrifice detection accuracy, and this is useful in the

context of CNNs, where there are significant computational advantages to using MIPs over multi-slice image volumes. The MIP procedure essentially compresses the 3D image to a 2D image, resulting in efficient and better detection for microcalcifications. Future studies will capitalize on the MIP for investigating additional parameters with improved efficiency. While CNNs do not necessarily predict human observer performance, it is likely that human observers may also benefit from viewing MIPs compared to individual slices.

It is noteworthy that the native section thickness was the optimal section thickness for detecting individual microcalcifications and clusters of microcalcifications. These findings differ from those of previous studies conducted in our laboratory for the detection of mass lesions, where the optimal section thickness for detecting small (1 mm), unenhanced mass lesions was the equivalent of 3-5 reconstructed slices[40]. We suspect that the lowered detection performance of mass lesions in the thinnest section was due to the interference of quantum noise with the signal. Microcalcifications are higher-contrast objects compared to mass lesions and may be more immune to quantum noise. One advantage of breast CT over mammography is the ability to adjust the display (section thickness or MIP) of the image volume in real-time using viewing software, and this ability will be important during screening exams when lesion positions are not known *a priori*, or when lesions of interest vary in morphology.

The effects of cluster diameter and number of calcifications on overall cluster detection performance was also investigated. Our results indicated that cluster diameter affects the detectability of microcalcification clusters, particularly for smaller microcalcifications near the threshold of detectability. Larger cluster diameters resulted in reduced overall detectability, while smaller cluster diameters led to increased detectability. Additionally, we observed that the presence of more calcifications within a cluster improved the overall detectability, while

fewer calcifications decreased it. These results begin to elucidate how model observers interact with microcalcification clusters in a 3D volume.

These results underscore the well-known reality that the challenge of detecting microcalcifications is the challenge of resolving small objects. The voxel size of our current breast CT scanner is 0.2 mm. Assuming two calcifications of the same composition but differing only in that one is larger than the voxel size resolution, and the other is smaller than the voxel size resolution, the larger calcification will appear notably brighter when imaged on our breast CT system due to partial volume effects. In this study, we mathematically model the loss of intensity owing to partial volume effects specific to our system, and then evaluate the detectability of partial-volumed calcifications across clinical parameters. These models may be useful to quantitatively estimate the improvement in detectability that may arise from adjusting components of the breast CT system or protocols such as reconstructed voxel size.

This study had a number of limitations. The field of view of the 2D 120 × 120 × 1 ROIs was 24 mm × 24 mm. Due to GPU memory constraints, when generating 3D VOIs, the field of view was reduced to 10 mm × 10 mm × 10 mm (50 × 50 × 50 voxels). The reduced volumetric field of view limited our ability to simulate larger cluster diameters and limited the CNN's learning of anatomical background. Nevertheless, the extensive training dataset consisting of 29,700 VOIs derived from 109 distinct breast CT datasets likely enabled the CNN to effectively grasp the nuances of breast anatomical background during its training process. Moreover, the methodology for simulating microcalcifications and microcalcification clusters was simplified such that each microcalcification was spherical, and that each cluster contained microcalcifications of homogeneous size. While this simplification does not capture the full complexity of real-world microcalcifications and clusters, it enabled a controlled investigation into specific parameters of interest: microcalcification size, native intensity,

cluster diameter, number of calcifications, and image display methods. Contrast imaging was not explored in this study, but initial studies have demonstrated the utility of contrast-enhanced breast CT in improving the conspicuity of small, malignant calcifications such as ductal carcinoma in situ (DCIS)[94], and will likely bring added benefit. Future studies should investigate the influence of these additional factors to obtain a more comprehensive understanding of breast CT optimization for microcalcification detection.

4.5. Conclusion

This study investigated individual effects and the interplay of parameters affecting microcalcification detectability in breast CT. We inserted mathematically generated microcalcifications into acquired patient breast CT images and used CNN-based model observers to evaluate microcalcification detectability across clinical and imaging parameters. As breast CT is still a relatively new breast imaging modality, there is an ongoing need to identify optimal imaging protocols. The results of this investigation will be useful for future studies investigating a broader set of parameters related to breast CT and microcalcification detection.

Chapter 5: Conclusions

This dissertation focused on key factors that contribute to mass lesion and microcalcification detectability in breast CT using hybrid images and model observers.

In Chapter 1, a new method was developed for simulating contrast-enhanced mass lesions based on enhancement levels arising from contrast-enhanced mass lesions from our clinical data set. Acquired patient breast CT images were used in concert with simulated contrast-enhanced mass lesions to study the role of contrast enhancement on detectability. A mathematical model observer, the pre-whitened matched filter (PWMF), was used to evaluate detectability across clinical parameters, and the improvement in detection due to contrast injection was quantified. An average 20% improvement in lesion detectability due to contrast enhancement was observed across lesion diameter, section thickness, breast density, and view plane, and a larger improvement was observed for patients with dense breasts. Small lesions are generally harder to detect in dense breasts, but these results demonstrated that injected contrast can substantially improve detection performance in dense breasts.

In Chapter 2, a CNN model observer was compared to the PWMF model observer for detecting simulated unenhanced mass lesions mathematically inserted into real patient breast CT images. The model observers were used to detect mass lesions in Gaussian background to better understand the performance of the CNN in the context of a known ideal observer (the PWMF). In the Gaussian background, the CNN performed essentially identically to the PWMF across lesion sizes and section thicknesses. In the breast CT background, the CNN outperformed the PWMF across lesion size, breast density, and most section thicknesses. These findings suggest that in breast CT images, CNNs capture more

diagnostic information than PWMFs and may be a more pertinent observer when conducting optimal performance studies in breast CT images.

In Chapter 3, a new method was developed for mathematically simulating microcalcifications and microcalcification clusters. The loss of intensity owing to partial volume effects was modeled and used to mathematically insert microcalcifications into acquired patient breast CT images. 2D and 3D CNN models were used to evaluate the detectability of simulated calcifications across clinical parameters. We found that there was no statistically significant difference in detection performance when using the MIP compared to the native section thickness, but that thicker sections led to reduced detection performance.

Together, these studies elucidate the key factors affecting mass lesion and microcalcification detectability in unenhanced and contrast-enhanced breast CT and demonstrate the utility of model observers for examining breast CT images when human observers are unavailable. Future studies will investigate a broader set of parameters such as the role of morphology in lesion detection, the influence of contrast injection in microcalcification detection, and the role of reconstruction kernel in mass lesion and microcalcification detection. As breast CT advances towards translation to the clinic, these studies will be useful for optimizing breast CT protocols.

References

1. Cancer Facts & Figures 2021. Published online 1930:72.

2. Lewin JM, D'Orsi CJ, Hendrick RE. Digital mammography. *Radiol Clin North Am*. 2004;42(5):871-884, vi. doi:10.1016/j.rcl.2004.06.004

3. Niklason LT, Christian BT, Niklason LE, et al. Digital tomosynthesis in breast imaging. *Radiology*. 1997;205(2):399-406. doi:10.1148/radiology.205.2.9356620

4. Bernardi D, Macaskill P, Pellegrini M, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. *Lancet Oncol*. 2016;17(8):1105-1113. doi:10.1016/S1470-2045(16)30101-2

5. Liberatore M, Cucchi JM, Fighiera M, et al. Interest of systematic tomosynthesis (3D mammography) with synthetic 2D mammography in breast cancer screening. *Horm Mol Biol Clin Investig*. 2017;32(2):/j/hmbci.2017.32.issue-2/hmbci-2017-0024/hmbci-2017-0024.xml. doi:10.1515/hmbci-2017-0024

6. Zuckerman SP, Sprague BL, Weaver DL, Herschorn SD, Conant EF. Multicenter Evaluation of Breast Cancer Screening with Digital Breast Tomosynthesis in Combination with Synthetic versus Digital Mammography. *Radiology*. 2020;297(3):545-553. doi:10.1148/radiol.2020200240

7. Cancer Screening Overview (PDQ®)–Patient Version - National Cancer Institute. Published January 13, 2010. Accessed July 6, 2021. https://www.cancer.gov/about-cancer/screening/patient-screening-overview-pdq

8. Brem RF, Lenihan MJ, Lieberman J, Torrente J. Screening breast ultrasound: past, present, and future. *AJR Am J Roentgenol*. 2015;204(2):234-240. doi:10.2214/AJR.13.12072

9. Morris EA. Breast cancer imaging with MRI. *Radiol Clin North Am*. 2002;40(3):443-466. doi:10.1016/s0033-8389(01)00005-7

10. Arpino G, Laucirica R, Elledge RM. Premalignant and In Situ Breast Disease: Biology and Clinical Implications. *Ann Intern Med*. 2005;143(6):446. doi:10.7326/0003-4819-143-6-200509200-00009

11. Breast Imaging Reporting & Data System. Accessed May 23, 2023. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads

12. Fandos-Morera A, Prats-Esteve M, Tura-Soteras JM, Traveria-Cros A. Breast tumors: composition of microcalcifications. *Radiology*. 1988;169(2):325-327. doi:10.1148/radiology.169.2.2845471

13. Henrot P, Leroux A, Barlier C, Génin P. Breast microcalcifications: The lesions in anatomical pathology. *Diagn Interv Imaging*. 2014;95(2):141-152. doi:10.1016/j.diii.2013.12.011

14. Le Gal M, Chavanne G, Pellier D. Diagnostic value of clustered microcalcifications discovered by mammography (apropos of 227 cases with histological verification and without a palpable breast tumor). *Bull Cancer*. 1984;71(1):57-64.

15. Mammography Quality Standards Act. Federal Register. Published March 10, 2023. Accessed May 23, 2023. https://www.federalregister.gov/documents/2023/03/10/2023-04550/mammography-quality-standards-act

16. Lindfors KK, Boone JM, Nelson TR, Yang K, Kwan ALC, Miller DF. Dedicated Breast CT: Initial Clinical Experience. *Radiology*. 2008;246(3):725-733. doi:10.1148/radiol.2463070410

17. Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. *J Opt Soc Am A*. 1984;1(6):612. doi:10.1364/JOSAA.1.000612

18. Gazi PM, Yang K, Burkett GW, Aminololama-Shakeri S, Anthony Seibert J, Boone JM. Evolution of spatial resolution in breast CT at UC Davis. *Med Phys*. 2015;42(4):1973-1981. doi:10.1118/1.4915079

19. Packard NJ, Abbey CK, Yang K, Boone JM. Effect of slice thickness on detectability in breast CT using a prewhitened matched filter and simulated mass lesions: Effect of slice thickness on breast CT detectability. *Med Phys*. 2012;39(4):1818-1830. doi:10.1118/1.3692176

20. Hernandez AM, Becker AE, Hyun Lyu S, Abbey CK, Boone JM. High-resolution μCT imaging for characterizing microcalcification detection performance in breast CT. *J Med Imaging*. 2021;8(05). doi:10.1117/1.JMI.8.5.052107

21. Lyu SH, Hernandez AM, Shakeri SA, Abbey CK, Boone JM. Model observer performance in contrast-enhanced lesions in breast CT: The influence of contrast concentration on detectability. *Med Phys*. Published online August 28, 2023:mp.16667. doi:10.1002/mp.16667

22. Bakic PR, Barufaldi B, Higginbotham D, et al. Virtual clinical trial of lesion detection in digital mammography and digital breast tomosynthesis. In: Chen GH, Lo JY, Gilat Schmidt T, eds. *Medical Imaging 2018: Physics of Medical Imaging*. SPIE; 2018:5. doi:10.1117/12.2294934

23. Barufaldi B, Vent TL, Bakic PR, Maidment ADA. Computer simulations of case difficulty in digital breast tomosynthesis using virtual clinical trials. *Med Phys*. 2022;49(4):2220-2232. doi:10.1002/mp.15553

24. Badano A, Graff CG, Badal A, et al. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. *JAMA Netw Open*. 2018;1(7):e185474. doi:10.1001/jamanetworkopen.2018.5474

25. Barrett HH, Yao J, ROLLANDt JP, MYERSt KJ. Model observers for assessment of image quality. *Proc Natl Acad Sci USA*. Published online 1993.

26. Eckstein MP, Abbey CK. CHAPTER 10 A Practical Guide to Model Observers for Visual Detection in Synthetic and Natural.

27. Burgess A. Image quality, the ideal observer, and human performance of radiologic decision tasks. *Acad Radiol*. 1995;2(6):522-526. doi:10.1016/S1076-6332(05)80411-8

28. Brankov JG, Yongyi Yang, Liyang Wei, El Naqa I, Wernick MN. Learning a Channelized Observer for Image Quality Assessment. *IEEE Trans Med Imaging*. 2009;28(7):991-999. doi:10.1109/TMI.2008.2008956

29. He X, Park S. Model Observers in Medical Imaging Research. *Theranostics*. 2013;3(10):774-786. doi:10.7150/thno.5138

30. Chen L, Boone JM, Abbey CK, et al. Simulated lesion, human observer performance comparison between thin-section dedicated breast CT images versus computed thick-section simulated projection images of the breast. *Phys Med Biol*. 2015;60(8):3347-3358. doi:10.1088/0031-9155/60/8/3347

31. Han M, Kim B, Baek J. Human and model observer performance for lesion detection in breast cone beam CT images with the FDK reconstruction. Sidky E, ed. *PLOS ONE*. 2018;13(3):e0194408. doi:10.1371/journal.pone.0194408

32. Burgess AE, Jacobson FL, Judy PF. Human observer detection experiments with mammograms and power-law noise. *Med Phys*. 2001;28(4):419-437. doi:10.1118/1.1355308

33. Eckstein MP, Abbey CK, Whiting JS. Human vs model observers in anatomic backgrounds. In: Kundel HL, ed. ; 1998:16-26. doi:10.1117/12.306180

34. Carton AK, de Carvalho PM, Li Z, Dromain C, Muller S. Assessment of mass detection performance in contrast enhanced digital mammography. In: Mello-Thoms CR, Kupinski MA, eds. ; 2015:941608. doi:10.1117/12.2081772

35. Prionas ND, Lindfors KK, Ray S, et al. Contrast-enhanced Dedicated Breast CT: Initial Clinical Experience1. *Radiology*. 2010;256(3):714-723. doi:10.1148/radiol.10092311

36. Jochelson MS, Dershaw DD, Sung JS, et al. Bilateral Contrast-enhanced Dual-Energy Digital Mammography: Feasibility and Comparison with Conventional Digital Mammography and MR Imaging in Women with Known Breast Carcinoma. *Radiology*. 2013;266(3):743-751. doi:10.1148/radiol.12121084

37. Curry FRE, Adamson RH. Vascular permeability modulation at the cell, microvessel, or whole organ level: towards closing gaps in our knowledge. *Cardiovasc Res*. 2010;87(2):218-229. doi:10.1093/cvr/cvq115

38. Miles KA. Tumour angiogenesis and its relation to contrast enhancement on computed tomography: a review. *Eur J Radiol*. 1999;30(3):198-205. doi:10.1016/S0720-048X(99)00012-1

39. Maeda H. The enhanced permeability and retention (EPR) effect in tumor vasculature: the key role of tumor-selective macromolecular drug targeting. *Adv Enzyme Regul*. 2001;41(1):189-207. doi:10.1016/S0065-2571(00)00013-3

40. Packard NJ, Abbey CK, Yang K, Boone JM. Effect of slice thickness on detectability in breast CT using a prewhitened matched filter and simulated mass lesions. *Med Phys*. 2012;39(4):1818-1830. doi:10.1118/1.3692176

41. Timberg P, Båth M, Andersson I, Mattsson S, Tingberg A, Ruschin M. In-plane visibility of lesions using breast tomosynthesis and digital mammography: In-plane visibility of lesions using BT and DM. *Med Phys*. 2010;37(11):5618-5626. doi:10.1118/1.3488899

42. Shaheen E, De Keyzer F, Bosmans H, Dance DR, Young KC, Ongeval CV. The simulation of 3D mass models in 2D digital mammography and breast tomosynthesis: Simulation of 3D mass models. *Med Phys*. 2014;41(8Part1):081913. doi:10.1118/1.4890590

43. de la Rosa RS, Carton AK, de Carvalho PM, Li Z, Muller S, Bloch I. Preliminary study of CEDBT and CESM performances using simulated analytical contrast uptakes. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE; 2018:792-795. doi:10.1109/ISBI.2018.8363691

44. Ghazi P, Hernandez AM, Abbey C, Yang K, Boone JM. Shading artifact correction in breast CT using an interleaved deep learning segmentation and maximum-likelihood polynomial fitting approach. *Med Phys*. 2019;46(8):3414-3430. doi:10.1002/mp.13599

45. Packard N, Boone JM. Glandular segmentation of cone beam breast CT volume images. In: *Medical Imaging 2007: Physics of Medical Imaging*. Vol 6510. SPIE; 2007:1091-1098. doi:10.1117/12.713911

46. Omnipaque. Accessed September 25, 2022. https://www.gehealthcare.com/products/contrast-media/omnipaque

47. Wang L, Liu B, Wu XW, et al. Correlation between CT attenuation value and iodine concentration in vitro: discrepancy between gemstone spectral imaging on single-source dual-energy CT and traditional polychromatic X-ray imaging. *J Med Imaging Radiat Oncol*. 2012;56(4):379-383. doi:10.1111/j.1754-9485.2012.02379.x

48. Ehn S, Sellerer T, Muenzel D, et al. Assessment of quantification accuracy and image quality of a full-body dual-layer spectral CT system. *J Appl Clin Med Phys*. 2017;19(1):204-217. doi:10.1002/acm2.12243

49. Hernandez AM, Boone JM. Tungsten anode spectral model using interpolating cubic splines: Unfiltered x-ray spectra from 20 kV to 640 kV. *Med Phys*. 2014;41(4):042101. doi:10.1118/1.4866216

50. Johns PC, Yaffe MJ. X-ray characterisation of normal and neoplastic breast tissues. *Phys Med Biol*. 1987;32(6):675-695. doi:10.1088/0031-9155/32/6/002

51. Burgess AE, Judy PF. Signal detection in power-law noise: effect of spectrum exponents. *JOSA A*. 2007;24(12):B52-B60. doi:10.1364/JOSAA.24.000B52

52. Chen L, Abbey CK, Boone JM. Association between power law coefficients of the anatomical noise power spectrum and lesion detectability in breast imaging modalities. *Phys Med Biol*. 2013;58(6):1663-1681. doi:10.1088/0031-9155/58/6/1663

53. Kaiser J, Hamming R. Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter. *IEEE Trans Acoust Speech Signal Process*. 1977;25(5):415-422. doi:10.1109/TASSP.1977.1162980

54. Bendat JS, Piersol AG. *Random Data: Analysis and Measurement Procedures*. John Wiley & Sons; 2011.

55. Burgess AE. Comparison of receiver operating characteristic and forced choice observer performance measurement methods. *Med Phys*. 1995;22(5):643-655. doi:10.1118/1.597576

56. Chong A, Weinstein SP, McDonald ES, Conant EF. Digital Breast Tomosynthesis: Concepts and Clinical Practice. *Radiology*. 2019;292(1):1-14. doi:10.1148/radiol.2019180760

57. Lewin JM, Isaacs PK, Vance V, Larke FJ. Dual-energy contrast-enhanced digital subtraction mammography: feasibility. *Radiology*. 2003;229(1):261-268. doi:10.1148/radiol.2291021276

58. O'Connell A, Conover DL, Zhang Y, et al. Cone-Beam CT for Breast Imaging: Radiation Dose, Breast Coverage, and Image Quality. *Am J Roentgenol*. 2010;195(2):496-509. doi:10.2214/AJR.08.1017

59. Kalender WA, Beister M, Boone JM, Kolditz D, Vollmar SV, Weigel MCC. High-resolution spiral CT of the breast at very low dose: concept and feasibility considerations. *Eur Radiol*. 2012;22(1):1-8. doi:10.1007/s00330-011-2169-4

60. Prionas ND, Lindfors KK, Ray S, et al. Contrast-enhanced Dedicated Breast CT: Initial Clinical Experience. *Radiology*. 2010;256(3):714-723. doi:10.1148/radiol.10092311

61. Abbey CK, Sengupta S, Zhou W, et al. Analyzing neural networks applied to an anatomical simulation of the breast. In: Mello-Thoms CR, Taylor-Phillips S, eds. *Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment*. SPIE; 2022:14. doi:10.1117/12.2612614

62. Fan F, Ahn S, Man BD, et al. Deep learning-based model observers that replicate human observers for PET imaging. In: *Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment*. Vol 11316. SPIE; 2020:53-58. doi:10.1117/12.2547505

63. Barlow HB. Increment thresholds at low intensities considered as signal/noise discriminations. *J Physiol*. 1957;136(3):469-488. doi:10.1113/jphysiol.1957.sp005774

64. Burgess AE, Wagner RF, Jennings RJ, Barlow HB. Efficiency of Human Visual Signal Discrimination. *Science*. 1981;214(4516):93-94. doi:10.1126/science.7280685

65. Wagner RF, Brown DG. Unified SNR analysis of medical imaging systems. *Phys Med Biol*. 1985;30(6):489-518. doi:10.1088/0031-9155/30/6/001

66. Vennart W. ICRU Report 54: Medical imaging-the assessment of image quality - ISBN 0-913394-53-X. April 1996, Maryland, U.S.A. *Radiography*. 1997;3(3):243-244. doi:10.1016/S1078-8174(97)90038-9

67. Barrett HH, Abbey CK, Clarkson E. Objective assessment of image quality III ROC metrics, ideal observers, and likelihood-generating functions. *J Opt Soc Am A*. 1998;15(6):1520. doi:10.1364/JOSAA.15.001520

68. Sharp P, Barber DC, Brown DG, et al. Appendix E: The Hotelling Observer. *Rep Int Comm Radiat Units Meas*. 1996;os-28(1):63-65. doi:10.1093/jicru_os28.1.63

69. Barrett HH, Yao J, ROLLANDt JP, MYERSt KJ. Model observers for assessment of image quality. *Proc Natl Acad Sci USA*. Published online 1993.

70. Abbey CK, Nosratieh A, Sohl-Dickstein J, Yang K, Boone JM. Non-Gaussian statistical properties of breast images: Non-Gaussian statistical properties of breast images. *Med Phys*. 2012;39(11):7121-7130. doi:10.1118/1.4761869

71. Boone JM, Gross GW, Greco-Hunt V. Neural Networks in Radiologic Diagnosis. *Invest Radiol*. 1990;25(9):1012-1016.

72. Kupinski MA, Edwards DC, Giger ML, Metz CE. Ideal observer approximation using Bayesian classification neural networks. *IEEE Trans Med Imaging*. 2001;20(9):886-899. doi:10.1109/42.952727

73. Zhou W, Li H, Anastasio MA. Approximating the Ideal Observer and Hotelling Observer for binary signal detection tasks by use of supervised learning methods. *IEEE Trans Med Imaging*. 2019;38(10):2456-2468. doi:10.1109/TMI.2019.2911211

74. Han M, Baek J. A convolutional neural network-based anthropomorphic model observer for signal-known-statistically and background-known-statistically detection tasks. *Phys Med Biol*. 2020;65(22):225025. doi:10.1088/1361-6560/abbf9d

75. Kim B, Han M, Baek J. A Convolutional Neural Network-Based Anthropomorphic Model Observer for Signal Detection in Breast CT Images Without Human-Labeled Data. *IEEE Access*. 2020;8:162122-162131. doi:10.1109/ACCESS.2020.3021125

76. Kim G, Han M, Shim H, Baek J. A convolutional neural network-based model observer for breast CT images. *Med Phys*. 2020;47(4):1619-1632. doi:10.1002/mp.14072

77. Jimmy L, Caruana R. Do Deep Nets Really Need to be Deep?

78. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. Published online January 29, 2017. Accessed January 26, 2023. http://arxiv.org/abs/1412.6980

79. Chollet F. Keras. Published online 2015. https://github.com/fchollet/keras

80. Kallergi M. Computer-aided diagnosis of mammographic microcalcification clusters. *Med Phys*. 2004;31(2):314-326. doi:10.1118/1.1637972

81. Bent CK, Bassett LW, D'Orsi CJ, Sayre JW. The Positive Predictive Value of BI-RADS Microcalcification Descriptors and Final Assessment Categories. *Am J Roentgenol*. 2010;194(5):1378-1383. doi:10.2214/AJR.09.3423

82. Fischer U, Baum F, Obenauer S, et al. Comparative study in patients with microcalcifications: full-field digital mammography vs screen-film mammography. *Eur Radiol*. 2002;12(11):2679-2683. doi:10.1007/s00330-002-1354-x

83. Arodź T, Kurdziel M, Popiela TJ, Sevre EOD, Yuen DA. Detection of clustered microcalcifications in small field digital mammography. *Comput Methods Programs Biomed*. 2006;81(1):56-65. doi:10.1016/j.cmpb.2005.10.002

84. Boone JM, Nelson TR, Lindfors KK, Seibert JA. Dedicated Breast CT: Radiation Dose and Image Quality Evaluation. *Radiology*. 2001;221(3):657-667. doi:10.1148/radiol.2213010334

85. Suryanarayanan S, Karellas A, Vedantham S, Sechopoulos I, D'Orsi CJ. Detection of Simulated Microcalcifications in a Phantom with Digital Mammography: Effect of Pixel Size. *Radiology*. 2007;244(1):130-137. doi:10.1148/radiol.2441060977

86. Ghammraoui B, Zidan A, Alayoubi A, Zidan A, Glick SJ. Fabrication of microcalcifications for insertion into phantoms used to evaluate x-ray breast imaging systems. *Biomed Phys Eng Express*. 2021;7(5):055021. doi:10.1088/2057-1976/ac1c64

87. Shaheen E, Van Ongeval C, Zanca F, et al. The simulation of 3D microcalcification clusters in 2D digital mammography and breast tomosynthesis: Simulation of 3D microcalcification clusters. *Med Phys*. 2011;38(12):6659-6671. doi:10.1118/1.3662868

88. Lefebvre F, Benali H, Gilles R, Di Paola R. A simulation model of clustered breast microcalcifications. *Med Phys*. 1994;21(12):1865-1874. doi:10.1118/1.597186

89. Gong X, Glick SJ, Liu B, Vedula AA, Thacker S. A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT breast imaging: Comparison of lesion detectability with 3 breast imaging modalities. *Med Phys*. 2006;33(4):1041-1052. doi:10.1118/1.2174127

90. Massanes F, Brankov JG. Evaluation of CNN as anthropomorphic model observer. In: *Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*. Vol 10136. SPIE; 2017:188-194. doi:10.1117/12.2254603

91. Frappart L, Remy I, Lin HC, et al. Different types of microcalcifications observed in breast pathology: Correlations with histopathological diagnosis and radiological

examination of operative specimens. *Virchows Arch A Pathol Anat Histopathol*. 1987;410(3):179-187. doi:10.1007/BF00710823

92. Eckstein MP, Abbey CK. Model observers for signal-known-statistically tasks (SKS). In: Krupinski EA, Chakraborty DP, eds. ; 2001:91-102. doi:10.1117/12.431177

93. Lyu SH, Abbey CK, Hernandez AM, Boone JM. Pre-whitened matched filter and convolutional neural network based model observer performance for mass lesion detection in non-contrast breast CT.

94. Aminololama-Shakeri S, Abbey CK, Gazi P, et al. Differentiation of ductal carcinoma in-situ from benign micro-calcifications by dedicated breast computed tomography. *Eur J Radiol*. 2016;85(1):297-303. doi:10.1016/j.ejrad.2015.09.020