Advanced Computational MRI of the Breast and Abdomen

by

Timothy Jameson Allen

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Medical Physics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 05/01/2023

The dissertation is approved by the following members of the Final Oral Committee:

Walter Block, Professor, Medical Physics
Diego Hernando, Associate Professor, Medical Physics
James Holmes, Research Assistant Professor, Radiology
Kevin Johnson, Assistant Professor, Medical Physics
Roberta Strigel, Associate Professor, Radiology

Abstract

Breast MRI is a critically important tool in screening for cancer in high-risk women, assessing the extent of known cancers, and in evaluating the integrity of breast implants. It also shows promise for evaluating therapy response. Emerging advanced techniques in breast MRI include abbreviated screening protocols and simultaneous PET/MR. However, there are several technical challenges which limit breast MRI. First, breast MRI generally compromises on image quality to reduce imaging time. Recent developments in deep learning MRI reconstructions may reduce the need for compromise. However, there is no existing research using these reconstruction techniques in a breast setting. Second, abbreviated breast protocols require highly optimized workflows to achieve their stated goal of shortened exam times. Elements of current breast MRI setup are performed manually by MR technologists, are prone to errors, and subject to intra- and inter-operator variability. Automated methods could allow for more rapid and consistent image setup, but such techniques do not currently exist for breast MRI. Third, simultaneous PET/MR of the breast requires administration of both a PET radiotracer and a gadolinium-based contrast agent. Gadolinium is a heavy metal and there is concern that it may attenuate PET annihilation photons and bias quantification of radiotracer uptake.

Three chapters in this work are dedicated to addressing each of the challenges just introduced. First, a deep learning reconstruction was applied to existing and novel breast MR sequences, and the specific benefits to image quality were assessed. Second, an automated method for breast MR imaging setup was developed, and its performance compared to that of human users. Third, the impact of gadolinium-based contrast agents on PET radiotracer uptake measurements was quantified.

MRI is also a sensitive tool for detection of lesions in the abdomen. Specifically, diffusion weighted imaging provides a measure of tissue microstructure which can help distinguish cancerous and normal tissues. Current clinical diffusion weighted imaging allows for a mostly qualitative assessment of diffusion. Quantification through the apparent diffusion coefficient or other diffusion metrics can provide additional insight into tissue microstructure. However, such quantification in the abdomen is particularly difficult due to motion-induced artifacts. While methods to help suppress these artifacts exists, they have been limited to specific scanner hardware configurations. The final chapter in this work describes the implementation of motion-robust diffusion methods on scanners with multiple hardware configurations. The reproducibility of quantitative diffusion metrics across different scanners with different hardware was also assessed.

Overall, the work presented in this thesis made substantial progress in overcoming the challenges of advanced MR imaging techniques in the breast and abdomen.

Acknowledgments

This Ph.D. journey has been an extraordinary one. It has been filled with moments of immense joy and instances of unceasing tears; sometimes with nary a few moments between them. A particular challenge of this Ph.D. work was the unique circumstances brought about by the COVID-19 pandemic. The majority of this work was performed during lockdown being buttressed by two seemingly brief periods of relative normalcy. Even without the pandemic, the journey required to complete this Ph.D. has been one of immense growth and self-reflection. There were many times in which I wanted to give up, to forgo finishing my Ph.D, abandoning my projects and my hard work. It was the people around me who kept me motivated, who pushed me to do the things I didn't believe I could do, who were my biggest advocates and cheerleaders when I didn't have the heart to cheer for myself. While my name is on the cover, the writing of this dissertation is a culmination of all the support and guidance that I have received throughout the years. Therefore, I would like to take this time to acknowledge, by name, those who made this dissertation a reality.

First off, I would like to thank my main Ph.D. advisor during my time at UW-Madison, Dr. James Holmes. My biggest struggle during my Ph.D. was the writing. I never believed that I would be a good writer, and I thought that I would always shy away from it. Jim has convinced me otherwise. During our time together, he patiently worked with me to improve my writing. He was willing, at great time expense to him, to go through a manuscript line-by-line mentoring me through the scientific writing process and providing tips on how to improve my manuscript. Most importantly, he provided the edification necessary for me to develop into a scientific author. He took the time to carefully explain his own thought process and why he certain changes were necessary to the construction of the manuscript. However, he was always open and nurturing,

listening carefully to my own reasoning and process, so we could write something we both were proud of. My writing skill has grown tremendously, which I hope this dissertation demonstrates, and Jim played a huge role in its development. Thank you, Jim.

Another huge thank you must go to Dr. Diego Hernando. After Jim accepted a role at the University of Iowa, Diego graciously accepted me into his lab. He never hesitated to take me as a student though I came with the baggage of many projects unrelated to his own work. Diego has a way of making you reason deeply about your projects. This is due to his mastery in asking truly salient and meaningful questions about your work. Far from feeling accusatory, his questions are genuine and quite often are the beginning to a new and rich path of inquiry. His unending interest and curiosity regarding his students' work is what makes him such a great advisor. Thank you, Diego, for being there exactly when I needed you.

There are many others who also played an advisory role in both official and unofficial capacities, and I am deeply grateful to each of them. Thank you Dr. Frank Korosec, Dr. Roberta Strigel, and Dr. Amy Fowler who were instrumental in guiding my journey through medical physics, radiology, and science. An additional special thank you to the remaining members of my Ph.D.: Dr. Kevin Johnson and Dr. Wally Block.

I had the privilege to work in three incredible MRI research labs here at UW-Madison. The individuals making up these groups were core to my success and each has impressed on me priceless life lessons that I will continue to cherish throughout my life. Thank you to all the members of the breast MRI research group: Dr. Leah Henze Bancroft, Dr. Pingni Wang, Dr. Fred Kelcz, and Dr. Erin MacDonald. Thank you to all the members of the LIRP and QIML research groups: Dr. Scott Reader, Dr. David Harris, Dr. Nate Roberts, Dr. Rianne A Van der Heijden, Dr. Ali Pirasteh, Dr. Jitka Starekova, Dr. Thekla Oechtering, Dr. Nikolaos Panagiotopoulos, Dr.

Devashish Joshi, Dr. Ruiyang Zhao, Dr. Daiki Tamada, Dr. Gregory Simchick, Dr. Raphael Tadeu Do Vale Souza, Aidan Tollefson, Yavuz Muslu, Katrina Ruedinger, Sheena Chu, Andrea Houck, Garrett Fullerton, Jiayi Tang, Srijostna Volety, Collin Buelo, and Ruiqi Geng.

I would also like to thank some additional members of the extended MR and Radiology family: Dr. Samuel A. Hurley, Dr. Karl Vigen, Dr. Cari Glide Hurst, Bruce Collick, and Dr. Tyler Bradshaw. Special thanks to my fellow MR group reps: Daniel Seiter, Ruvini Navaratna, and Srijostna Volety, Brock Jolicoeur. Thank you to everyone from GE Healthcare who helped make this work a reality: Dr. Kang Wang, Dr. Ty Cashen, Lloyd Estkowski, Dr. Arnuad Guidon, Dr. Patricia Lan, and Dr. Ersin Bayram.

Finally, I would like to reserve a very special thanks for my family and friends who kept me sane during my Ph.D. and the pandemic. Thank you, Mom and Dad, for your never-ending support. Your weekly phone calls reminded me of how blessed and thankful I am to have you as my parents. Thanks to my siblings: Jessica, Tyler, Maddi, and Tori for keeping my life fun. Thank you Grandpa Pete, Grandma Carla, and Grandma Sharon who helped raise me into who I am today. Thank you to my friends who kept me from spending all my time locked in my room. Many have been mentioned already, but additional thanks to Tan Phan, Nadeem Shaheen, Robin Wells, Dani Mischke, and Josh Yeend.

Finally, the biggest thank you to my loving wife, Hannah Samowtiz, who was and is, undoubtedly, my greatest support and my best friend. This achievement is as much hers as it is mine.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2D	Two-dimensional
3D	Three dimensional
μ-map	Attenuation Map
ADC	Apparent Diffusion Coefficient
AP	Anterior-Posterior
BI-RADS	Breast Imaging Reporting and Data System
cIoU	Combined Intersection Over Union
CPU	Central Processing Unit
CT	Computed Tomography
DCE	Dynamic Contrast Enhanced
DL	Deep Learning
DM1/DM2	Dynamic Reconstruction Method 1/2
DWI	Diffusion Weighted Imaging
FDG	^{18}F-fluorodeoxglucose
FOV	Field of View
FSE	Fast Spin Echo
GBCA	Gadolinium-based Contrast Agent
gIoU	Generalized Intersection Over Union
GPU	Graphics Processing Unit
HR	High Spatial Resolution
IoU	Intersection over Union
LR	Left-Right
MIP	Maximum Intensity Projection
MODI	M1-Optimized Diffusion Imaging
MONO	Monopolar or Stejskal-Tanner Imaging
MRI/MR	Magnetic Resonance Imaging/Magnetic Resonance
MRAC	MR-based Attenuation Correction
OSEM	Ordered Subsets Expectation Maximum
PET	Positron Emission Tomography

PET/MR	Simultaneous PET/MR Exam
RDC	Reproducibility Coefficient
RMSE	Root Mean Squared Error
SI	Superior-Inferior
SNR	Signal-To-Noise Ratio
SOC	Standard-of-Care
SPGR	Spoiled Gradient Recalled
SUV	Standard Uptake Value
T1w	T1-weighted
T2w	T2-weighted
tMODI	Tetrahedral MODI
VOI	Volume of Interest

Thesis Outline

This dissertation will cover work that falls into one of two main categories: 1) advances in breast MRI or 2) advances in abdominal diffusion weighted imaging (DWI). Traditionally, a dissertation would present all the introductory information at the beginning, followed by all the novel scientific contributions, and end with closing remarks. However, due to the two distinct categories of this work, this dissertation will be split into two parts. It is my intent that this decision will aid in the comprehension of this work.

In Part I, an overview of breast MRI and advanced breast MRI techniques relevant to this dissertation's content is discussed. Then, the novelty and significance of this dissertation in the setting of breast MRI is presented. The main body of Part I consists of three chapters detailing the major scientific contributions of this work to breast MRI.

Part II is structured similarly to Part I but focuses on the advancements made in DWI of the abdomen. An introduction is presented followed by a brief outline of the novelty and significance in the setting of abdominal DWI. The main body of Part II will introduce in detail the major scientific contributions of this work in regard to abdominal DWI. Part II will conclude with a summary of the entire dissertation and a discussion potential future work that logically follows from this dissertation.

Part I

Advances in Breast MRI

Chapter One

Introduction

1.1 Breast MRI and Its Challenges

1.1.1 Breast MRI

The American Cancer Society estimates that 1 out of 8 women will develop breast cancer in their lifetime[1] and MRI plays an important role in breast cancer care. MRI is recommended for breast cancer screening in high-risk populations, for evaluating extent of known disease, and as an accurate method to assess response to neoadjuvant therapy[2–4]. Typical MRI protocols for breast imaging include the following acquisitions: T1-weighted (T1w) dynamic contrast enhanced (DCE), T2-weighted (T2w), and T1w without fat saturation. Diffusion weighted imaging may also be included. Of these acquisitions, DCE MRI provides the greatest sensitivity for detecting breast cancer[5–7]. The DCE consists of a single pre-contrast phase and multiple post-contrast phases of T1w imaging, which allows for assessment of tissue enhancement over time.

1.1.2 T2w Imaging and Protocol Compromises

T2w imaging is used in tandem with DCE MRI to detect T2 hyperintense findings, such as cysts, lymph nodes, and edema. T2 hyperintensity assists radiologists in differentiating benign and malignant findings; benign findings, such as fibroadenomas, are more likely to appear hyperintense on T2w images whereas malignant findings are most often isointense or hypointense. T2w imaging requires time-consuming acquisitions with long repetition times, which can lead to

patient discomfort and motion-related artifacts. Breast radiologists generally prefer to compromise on signal-to-noise ratio (SNR) and spatial resolution to achieve shorter acquisition times. While T1w DCE images frequently acquire sub-millimeter in-plane spatial resolution, T2w imaging is often limited to resolutions well above one millimeter. As an example, at the University of Wisconsin-Madison, the DCE acquisition has a matrix of 448 x 448 while the T2w acquisition has a 288 x 288 matrix. Both have a typical field of view around 32 cm. This results in a mismatch in image characteristics when trying to compare T1w and T2w images. T2w images are comparatively blurry with decreased spatial detail associated with volume averaging of small findings. Attempting to increase the spatial resolution while maintaining a reasonable acquisition time results in prohibitively low SNR.

1.1.3 Abbreviated Breast Protocols

Cancer screening with abbreviated breast MR has seen rapid growth in the past few years[8,9]. Abbreviated breast exams use fewer sequences than conventional breast MRI, often including only a T2w sequence and a truncated DCE sequence with a single post-contrast phase. The goal of abbreviated breast MRI is to allow more individuals to benefit from cancer screening by reducing the time and monetary cost of screening. While typical breast exams require 30 to 45 minutes to complete, abbreviated protocols target much shorter imaging times. Depending on the exact abbreviated protocol, this may be around 10 minutes or as little as 3 minutes. Achieving these target exam times requires highly optimized imaging protocols and workflow. In contrast, workflows that include highly-tailored, patient specific imaging can help mitigate image quality issues such as unsuppressed fat caused by field inhomogeneities[10–12]. Although, this complex workflow increases the technologist burden and takes substantial time to perform. Clinics are forced to choose between quick workflows and workflows that provide robust image quality.

1.1.4 MR Pre-scan in Breast

Image quality in breast MRI is known to be highly variable. Clinical MRI vendors try to maximize image quality for each acquisition using scan-specific system settings that can include the x, y, and z linear shims, center frequency, transmit gain, and receive gain. Optimizing these settings on a scan-by-scan basis helps to maintain image quality by preventing several sources of potential artifacts. Consequences of poor system settings can include incomplete water excitation, flip angle inaccuracy, image distortion, or fat suppression failure. For example, recent work by Zhou et al. found a statistically significant relationship between the x, y, and z linear shims and the quality of fat suppression in breast MRI[12]. While their work focused on the linear shims, fat suppression quality relies on the totality of scan-specific system settings. Breast MRI involves complex anatomical geometry, a high number of air-tissue interfaces, and variable fat/water tissue composition that makes determination of scan-specific settings challenging[13].

Typically, scan-specific settings are determined using a short system calibration, also known as the pre-scan, at the start of each image series. Two general approaches to pre-scan exist. Scan settings can be optimized globally across the entire scan volume or locally for a subregion within that volume. Since the axial breast imaging volume includes the breast, heart, lungs, arms, chest wall, and liver, the global approach may result in non-optimal settings for the breast tissue itself. Local pre-scan has been shown to be beneficial in breast MR by helping to achieve, for instance, a more homogenous B0 field using bilateral (two independent) pre-scan volumes, one for each breast[10,11].

Selection of the breast tissue for local pre-scan can be done using a single or multiple pre-scan volumes. Manually aligning these volumes to the individual patient anatomy in all three dimensions is prone to error and intra- and inter-reader variability. The aforementioned work by

Zhou et al. highlighted the variability in pre-scan placement by clinical technologists and how that variability could impact scan-specific system settings and image quality[12].

Scan volume selection is similarly prone to error and variability. The American College of Radiology (ACR) stipulates that the scan volume should be adequately positioned so that all of the breast tissue including the axillary tail is covered[4,14]. Variability in scan volume selection can contribute to variable image quality by affecting signal-to-noise ratio, image resolution, and artifacts such as phase wrap. Thus, accuracy and consistency of scan volume placement are necessary for high-quality breast MRI. However, as previously mentioned, this workflow often conflicts with the strict scan time goals of abbreviated MR protocols.

1.2 Approaches to Improve Breast MRI

1.2.1 Deep Learning Reconstructions

Machine learning (ML) and the sub-field of deep learning (DL) have led to many innovations and opportunities in the field of medical imaging[15,16]. ML and DL tools may help minimize image quality compromises in T2w breast MRI. DL-based reconstruction algorithms are particularly impactful and help to improve medical images via noise reduction, resolution enhancement, and/or artifact suppression[17]. DL reconstructions have been investigated for neurological[18], musculoskeletal[19], abdominal[20], and cardiac MRI imaging[21]. Despite their promising results throughout the body, research into DL reconstructions for breast imaging is lacking. Existing ML and DL research in breast MRI has largely focused on breast segmentation, lesion segmentation, lesion classification, and prognosis as discussed in multiple review articles on the topic[22,23].

The paucity of research on DL reconstruction for breast MRI may be due, in part, to the need for a large, diverse set of high-quality MR exams to serve as the ground truth. Compiling

such a dataset is especially challenging in breast MRI due to the extensive anatomical variability. Differences in breast size, density of fibroglandular tissue, presence of implants, implant type, and tissue distribution must be accounted for. Many patients also have post-surgical changes (e.g., mastectomy) adding further complexity. Even with sufficient anatomical variability, the training data would ideally need to encompass the range of expected acquisition parameters (TE, TR, parallel imaging, matrix size, etc.) and hardware configurations (magnet strength, receive coil, scanner model, etc.). Without this diversity, a change in acquisition protocol might require a complete retraining of the DL algorithm which is unfeasible in a typical clinical setting.

A general-purpose DL algorithm such as the one introduced by Lebel[24] may enable DL reconstructions for breast MRI while simultaneously avoiding curation of extensive training data and the time-consuming training process. Rather than being developed for a specific use (e.g., T2w breast imaging), it is designed to offer general image quality improvements such as reduced noise, decreased truncation artifact, and enhanced edge sharpness. The algorithm was trained on a dataset of 4 million image pairs encompassing a broad variety of image content and contrasts. In theory, this makes it compatible with many anatomical sites and acquisitions. The performance of the DL reconstruction has been evaluated in anatomic settings including pituitary[25], cardiac[26], prostate[27], and orthopedic hip and shoulder imaging[28]. Notably, the model did not require re-training to achieve increased image quality in these scenarios.

If the performance of DL in the breast setting mirrors its performance in other anatomical sites, it may reduce protocol compromises (e.g., low spatial resolution) regularly required for T2w breast MRI. However, there are unique challenges to overcome in breast imaging. As mentioned, breast anatomy is highly variable between individuals with implants and post-operative findings being common. In particular, T2w breast images have unique contrast with potentially very sparse

areas of fibroglandular tissue over a dark, fat-suppressed background. This is distinct from other anatomical regions previously investigated with this DL algorithm. Due to these idiosyncrasies, performance of the DL algorithm in the setting of breast MR imaging is currently unknown.

1.2.2 Automated MR Volume Prescription

An automated method for scan and pre-scan volume placement in breast MR could limit variability and reduce workload for MR technologists. Existing methods for scan volume placement have been investigated in a limited number of anatomies outside of the breast. Atlas- and template-based techniques have been shown in the brain[29,30], knee[31], and liver[32]. Alternative approaches use segmentation algorithms or feature localization[33,34]. Ozhinsky et al. developed placement of pre-scan volumes in addition to scan volumes to help with fat suppression during spectroscopic imaging[34]. These existing techniques are encouraging but may be challenging to implement in breast. Atlas-based approaches would require extensive anatomical atlases to capture the vast variation in breast anatomy. Implementation of other methods would require robust segmentation or landmark identification which is difficult in breast due to the intermingling of fibroglandular and fat tissue and lack of suitably consistent landmarks to guide volume placement.

More recently, machine learning approaches for automated scan volume placement using deep, convolutional neural networks, a form of DL, have been demonstrated[35–38]. However, these highly anatomically specific networks do not address the placement of either single or multiple pre-scan volumes. Another limitation of existing approaches is the lack of uncertainty given with their predictions. There is a growing need in uncertainty estimates which would assist interpretation of DL predictions[39]. Knowing when a DL prediction is certain of its output and when it is merely guessing would provide users an idea of the trustworthiness of a given outcome.

Techniques such as Monte Carlo dropout can provide uncertainty estimates but are not incorporated into the existing deep learning placement models[40–44].

1.3 Simultaneous PET/MR of the Breast

Simultaneous PET/MR is a hybrid imaging modality with applications in oncology, cardiology, and neurology[45]. For breast cancer, combined PET and MR data have been studied as a potential source for novel biomarkers useful in tumor staging, prognostication, and therapy response prediction[46–50]. Although, the development and validation of quantitative PET/MR imaging biomarkers is contingent on accurate methods for quantification of radiotracer uptake.

Loss of coincident 511 keV photons due to absorption from different depths within the body contributes to quantitative errors in PET imaging if attenuation correction is not performed during PET data reconstruction. While attenuation correction for PET/CT imaging is based on the CT data, attenuation correction in a combined PET/MR system must be accomplished using dedicated MR imaging (MRAC) to derive attenuation maps (μ-maps). Most MRAC methods include automatic segmentation of MR images followed by assignment of attenuation coefficients (segmentation methods) or registration of MR images to an MR/CT atlas which can be used for pseudo-CT generation (atlas methods). The challenges of MRAC and the various methods for MRAC generation are discussed in various review articles[51,52]. Recently, machine-learning based solutions have also been demonstrated[53]. It is important to evaluate potential intrinsic and extrinsic factors influencing attenuation of PET annihilation photons for accurate quantification of radiotracer uptake using PET/MR imaging.

1.3.1 Potential Attenuation via Gadolinium Contrast Agents

Gadolinium-based contrast agents (GBCAs) are used in MR imaging to enhance T1-weighted contrast and identify tissue with altered perfusion such as invasive cancers or cardiac perfusion defects[54,55]. For DCE breast MRI, use of GBCAs is required to achieve high sensitivity for cancer detection (90-99%)[56–58]. Gadolinium (Gd) is a rare-earth heavy metal element with a high atomic number (Z=64) and the primary functional component of any GBCA. Thus, GBCAs have the potential to attenuate annihilation photons causing loss in the PET signal. For hybrid PET/MR using DCE imaging, questions have arisen regarding possible effects of GBCAs on PET quantitative accuracy[47,59–62].

GBCAs dilute following initial injection but could be at a high enough concentration immediately after injection to introduce an apparent decrease in radiotracer uptake. As a "worst-case scenario", one study estimated a maximum GBCA concentration in the left ventricle of 70 mM based on a 20 mL bolus of 0.5 M GBCA diluted within the average end diastolic left ventricle volume of 150 mL and that the concentration would diminish over 30 to 50 s[59]. Attenuation by GBCAs may be particularly relevant to 1) dynamic PET acquisition protocols due to the initial influx of contrast occurring in a single short time frame and 2) for breast and pulmonary PET/MR imaging since the field-of-view includes the heart where GBCA will be near its highest concentration. Ultimately, an apparent decrease in radiotracer uptake would lead to underestimation of PET metrics such as standardized uptake values, SUV_{mean} and SUV_{max}, and prevent accurate pharmacokinetic modeling.

None of the standard MRAC techniques incorporate attenuation due to GBCAs. The mechanism by which GBCAs are responsible for enhanced contrast in T1-weighted MRI (local magnetic field increase with shortening of the T1 relaxation time) is independent of the mechanism

of attenuation (photon-matter interactions). Therefore, changes in MR image contrast are not necessarily reflective of attenuation changes. Two potential mechanisms have been proposed by which GBCAs could affect PET quantification: 1) directly through increased attenuation of annihilation photons resulting in false reduction of radiotracer uptake and 2) indirectly through errors in tissue segmentation in attenuation map generation. The second mechanism has been investigated by others by comparing MRAC sequences before and after GBCA administration in studies of oncologic patients undergoing PET/MR[60–63]. Previous investigators concluded that intravenous GBCAs have no effect on segmentation or atlas-based MRAC methods[60]. Additionally, other investigators found an effect of GBCA on tissue misclassification in the brain, but not for other body regions or for various cancerous lesions[61]. A more recent study demonstrated an influence of GBCA on MRAC for certain tissues such as lung and fat[63]. A practical solution that resolves this potential issue is to acquire MRAC images prior to GBCA administration, which is generally performed in most clinical PET/MR protocols.

Previous research on the first potential mechanism of direct attenuation by GBCAs has been limited to phantom-based studies using static PET acquisitions and found that GBCAs did not affect PET quantification[59,60]. Previous investigators acknowledged that their work does not investigate dynamic imaging with time-dependent PET and GBCA concentrations and is not a true patient representation[59]. They proposed that more complicated, anthropomorphic phantoms could be used as a more realistic comparison to actual clinical scenarios. A direct in vivo experiment with human subjects would provide results which are more readily applicable to clinical imaging. Other previous investigators were also aware of the limitations of their study concerning static versus dynamic reconstructions and indicated that additional dedicated studies are required to rule

out any effect[60]. Thus, previous investigators have encouraged further study using human subjects to confirm their prior results using simplified phantom models.

1.4 Innovations in Breast MRI

This section will briefly discuss the innovations of Part I with a paragraph devoted to each of the challenges of breast imaging introduced above.

Deep Learning Reconstruction for T2-weighted Imaging. Chapter 2 investigated a deep learning reconstruction to improve the quality of standard-of-care T2w breast MR images. It is the first work to apply such a deep learning reconstruction to breast imaging. Also introduced in that chapter is a novel high-resolution protocol tailored to optimally benefit from a deep learning reconstruction. Changes in the deep learning reconstruction output as a function of acquisition resolution were also evaluated for the first time. Overall, Chapter 2 discusses advances to the field of breast MRI by demonstrating the feasibility of using a deep learning reconstruction to improve the quality of breast MRI.

Automated Prescription of Scan and Pre-scan Volumes. Chapter 3 introduces the first algorithm for automated placement of scan volumes in breast MRI. While an algorithm for automated placement of pre-scan volumes using a neural network was developed by our group previously, it was limited to placing cubic volumes[64]. The work of chapter 3 extended the capabilities of the existing automated method by allowing for placement of pre-scan volumes with an independent size in three dimensions. Additionally, our automated method provided a way of estimating the uncertainty in the volume placement predictions, which is novel for automated MRI prescription techniques based on a neural network.

GBCA Attenuation Effect on PET Quantification in Breast MRI. Chapter 4 is the next logical step in investigating the GBCA attenuation effect in PET/MR. Previous investigations into

the effect of GBCAs on PET quantification have been purely phantom-based[59,60]. This is the first work which investigated the effect in vivo using human subjects. Additionally, the effect was evaluated using a dynamic PET reconstruction for the first time.

Chapter Two

Deep Learning Reconstruction for T2-weighted Imaging[1]

2.1 Introduction

As discussed in **Section 1.1.2**, T2w imaging of the breast suffers from lower resolution and worse overall image quality compared to DCE sequence due to the long acquisition times. The purpose of this chapter was to assess the ability of a general-purpose DL algorithm to reduce image quality compromises in T2w imaging. In particular, the use of DL is investigated with two T2w imaging protocols: 1) a standard-of-care (SOC) T2w protocol and 2) a novel high spatial resolution (HR) T2w protocol developed for this work.

2.2 Methods

To assess the performance of the DL algorithm, one phantom study and two in vivo studies were performed. Effects of the DL reconstruction on SNR and image sharpness were measured in a phantom before and after application of deep learning. Then, in-vivo imaging of patient volunteers with a SOC protocol and a HR protocol was performed with image quality being evaluated by breast-trained radiologists.

[1] This chapter has been submitted to Magnetic Resonance Imaging for publication consideration as a research article.

2.2.1 MRI Protocols

A summary of T2w protocols used in this study is provided in

Table 1. Imaging was performed on one of two 3.0T MR scanners: 1) a SIGNA Premier (GE Healthcare, Waukesha, WI) or 2) a Discovery MR750w (GE Healthcare). Breast specific 16-channel receive coils were used (Sentinelle, Dunlee, The Netherlands). The SOC T2w acquisition was performed with a 2D fast spin echo (FSE) sequence with echo train length of 16. Imaging parameters included a 288 x 288 acquired in-plane matrix with 2 mm slices and a variable field of view (32 cm to 36 cm) as required to achieve full coverage of the breasts and axilla while maximizing spatial resolution in smaller patients. Parallel imaging was performed with an acceleration factor of 3 (ASSET on scanner 1 or ARC on scanner 2, GE Healthcare). The difference in parallel imaging technique between the two MRI scanners was necessary to avoid changes to our institution's clinical imaging protocols on these two MRI systems. A HR acquisition was created from the SOC T2w sequence by increasing the acquisition matrix size to 448 x 448 and increasing the parallel imaging acceleration factor from 3 to 4. The 448 x 448 matrix size results in an in-plane resolution that matches the T1w DCE acquisition performed routinely at our institution. The parallel imaging increase was implemented to reduce the time penalty resulting from the increased number of phase-encoding lines. The mean acquisition times for the SOC acquisition was 247 or 237 seconds for scanners 1 and 2, respectively, while the HR acquisition had mean times of 308 and 284 seconds.

Table 1: T2w Acquisition Sequences

	Scanner 1		Scanner 2	
	SIGNA Premier		Discovery MR750w	
Parameter	SOC	HR	SOC	HR
Plane	Axial	Axial	Axial	Axial
Type	2D FSE	2D FSE	2D FSE	2D FSE
TR/TE (ms)	3500/85	3500/85	4000/85	4000/85
Acquisition Matrix	288x288	448x448	228x228	448x448
FOV (cm)	32-36	32-36	32-36	32-36
Slice Thickness (mm)	2	2	2	2
PI Factor	3/ASSET	4/ASSET	3/ARC	4/ARC
ETL	16	16	16	16
Mean Acquisition Time (s)	247	308	237	284

T2w: T2-weighted, SOC: Standard-of-care, HR: High-resolution, FSE: Fast Spin Echo, FOV: Field-of-view, PI: Parallel Imaging, ETL: Echo Train Length

2.2.2 Deep Learning Algorithm

An early prototype DL reconstruction was developed and provided by collaborators at GE Healthcare for early feasibility testing in the setting of breast imaging. The work described in this chapter used a pre-product prototype, however a product version has since been released (AIR Recon DL, GE Healthcare). The reconstruction algorithm has been described by Lebel[24] and is based on a convolutional neural network that is trained to reduce image noise, increase image sharpness, and reduce the prevalence of simple artifacts such as truncation ringing. The algorithm performs its DL inference in k-space and is compatible with the conventional MR processing pipeline including parallel imaging, image intensity correction, and gradient warping correction. The algorithm includes a user-adjustable, tunable noise parameter which controls the amount of estimated noise variance to be removed from the acquired images. This parameter is generally set below 100% noise removal to allow a residual level of noise which, as demonstrated by others, helps preserve the expected visual appearance of the images[27]. For this work, a noise reduction parameter of 75% was used.

A key feature of the algorithm is its generic nature; it is not trained on a specific image type, but rather with a broad, general dataset consistent of a wide variety of image contrasts and anatomical sites. This, in theory, makes it agnostic to the specifics of the imaging protocol and useful in many applications. However, validation of its performance for specific applications is still necessary. Due to the DL algorithm's generalizability, no additional training was performed as part of this study. DL images were generated from the raw MR k-space data through an offline prototype implementation provided by the vendor. Non-DL images were generated using the conventional on-scanner product reconstruction pipeline.

2.2.3 Technical Analysis in Phantom

Phantom experiments were conducted to assess the potential denoising and sharpness improvements using the DL algorithm. An ACR accreditation phantom (J.M. Specialty Parts, San Diego, CA) was imaged on scanner 1 using an 11-channel head and neck receive RF coil (GE Healthcare). The ACR phantom was used for its large volume and homogeneity which provides excellent opportunity for SNR measurements. It also has numerous internal sharp edges oriented perpendicularly to each other allowing for multiple independent measures of edge sharpness. The coil was selected to accommodate the geometry of the phantom. Imaging of the phantom was first performed using the SOC T2w acquisition. Then, parallel imaging was increased from 3 to 4 and the acquisition matrix was systematically increased from 288 x 288 to 448 x 448, corresponding with acquisition pixel sizes of 1.0 mm^2, 0.9 mm^2, and 0.8 mm^2 at a 32 cm field of view (320 x 320, 356 x 356, and 400 x 400 matrices).

SNR and image sharpness of the resulting non-DL and DL images were measured within the phantom. SNR was expected to increase due to a reduction in noise achieved with DL. SNR measurements were performed using the subtraction method described by NEMA[65] over multiple

slices near the center of the phantom. Sharpness was measured using the edge response at two vertical and two horizontal edges within a single central slice of the phantom. The sharpness was defined as the width of the edge measured from 20% to 80% of the signal intensity. The final size of DL reconstructed images depends on the acquisition matrix (see discussion). Thus, a discrepancy between the DL images size and the non-DL image size could exist. In such cases, a second non-DL reconstruction with an increased reconstruction matrix size was obtained through an offline version of the conventional, non-DL vendor reconstruction.

2.2.4 SOC Imaging

All human subject data were HIPAA compliant and collected under IRB approval. T2w series from twenty consecutive clinical breast MR exams on scanner 2 were retrospectively collected. Any breast MRI (with or without intravenous contrast) including a T2w acquisition was eligible. The T2w images were always collected prior to contrast administration. Both non-DL and DL T2w images series were generated as described in subsection 2.2.2.

Image quality of the T2w series was assessed via a reader-study. Four radiologists participated in the reads: One board-certified, fellowship-trained breast radiologist with 15 years-experience, a board-certified breast radiologist with 25 years-experience, and two radiologists who were undergoing breast fellowship training at the time of the study. Each radiologist independently viewed the non-DL and DL SOC series side-by-side. Readers were blinded to image reconstruction type and allowed to scroll through the 3D image volume. Image order was not randomized. Image scoring was based on four image quality criteria: 1) presence of artifacts, 2) perceived SNR, 3) perceived sharpness, and 4) overall image quality. The word perceived is used here to distinguish the radiologists' perception of SNR and sharpness from the technical measurements that were performed in phantom experiments. Each image quality category was scored using a 5-point Likert

scale with 1 indicating poor performance and 5 indicating excellent performance. The rubric for overall image quality is shown in

Table 1 with the remaining categories included in the appendix (Table A1). Notably, the scoring criteria included language asking readers to assess if poor category performance impacted the diagnostic utility of the image. If the readers felt diagnostic capability was impacted, they were instructed to score the performance no higher than 2 for that category. Wilcoxon signed-rank non-parametric tests were used to test for significant differences in average image quality scores between non-DL and DL images using a significance level of $p = 0.05$.

Table 2: T2w Likert Scoring Criteria

Score	Overall Image Quality
5	Excellent: no artifacts and anatomical detail well visualized
4	Good: minor artifacts, some blurriness, no impact on diagnostic capability
3	Fair: major or multiple minor artifacts, blurriness, no impact on diagnostic capability
2	Poor: multiple major or minor artifacts, loss of detail, impact on diagnostic capability
1	Non-diagnostic: severe artifacts, and complete loss of anatomical detail

Readers were also asked to state an overall preference for either non-DL or DL-enhanced images on a case-by-case basis. Responding “no preference” was allowed. Since signal intensity in T2w breast is used to assist interpretation of DCE findings, one task-based question was designed to mimic clinical interpretation of the T2w series. Readers were asked to identify hyperintense features in the T2w series (e.g., cysts, lymph nodes, edema, etc.) and respond if the signal intensity was different between the non-DL and DL reconstruction. Reponses could be brighter in non-DL, brighter in DL, or equally as bright.

2.2.5 HR Imaging

Fifty-six patient volunteers were prospectively imaged using the HR T2w sequence added to their clinical exam alongside the SOC T2w series. Prospective data was collected only after subjects provided informed consent. Patients were imaged on either scanner 1 or scanner 2 based on where their clinical exam was scheduled. Individuals with implants were excluded because, at our institution, these patients already have additional implant-specific sequences performed as part of their SOC, lengthening the exam. The HR acquisition was performed immediately after the SOC T2w acquisition and prior to contrast agent injection. The HR k-space data were used to reconstruct non-DL and DL T2w image series. The SOC T2w (non-DL) from the patients' clinical exam was also obtained.

A second reader study was performed to assess DL in the setting of HR acquisitions. Three image series were read: 1) the SOC T2w without DL, 2) the HR T2w without DL, and 3) the HR T2w with DL. Two board-certified, fellowship-trained breast radiologists with 15 and 13 years of experience, respectively, scored all cases independently. Readers were asked to identify a preferred series in each case. Stating no preference was allowed. Again, Wilcoxon signed-rank non-parametric tests weree used to test for significance difference between the three series.

2.3 Results

2.3.1 Technical Analysis in Phantom

Application of the DL reconstruction to phantom data resulted in a 2.1 to 2.8 times increase in the SNR (**Figure 1)**. This increase was present for both SOC and HR images. Edge sharpness was increased by 1.2 to 1.3 times for acquisition matrices below 320 x 320 and 1.4 to 1.7 times for matrices of 356 x 356 and above **(Figure 2)**. An example edge response demonstrates a reduction in truncation artifact **(Figure 3)**.

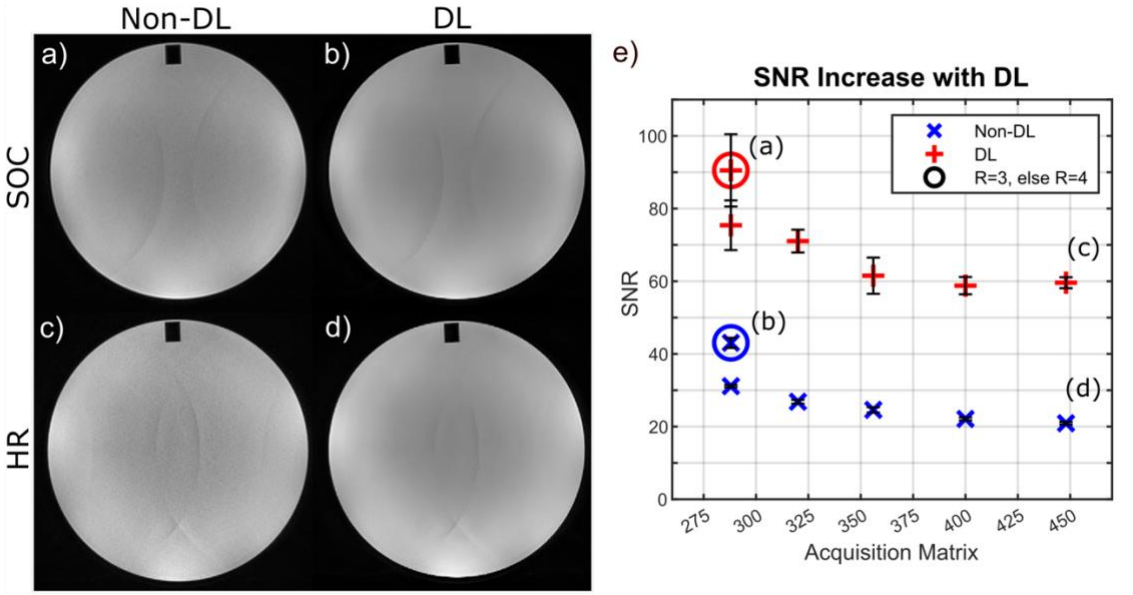

Figure 1: Effect of HR Imaging and a DL Reconstruction on SNR in a Phantom. Application of a deep learning (DL) reconstruction provided a 2.1 to 2.8 times increase in the measured SNR. **a-d)** Phantom images collected with either standard-of-care (SOC) or high spatial resolution (HR) acquisitions and reconstructed with either non-DL or DL pipelines. **e)** SNR measurements averaged over multiple slices within the phantom. Results are shown for non-DL and DL images. Points corresponding to images a-d are labelled. Error bars show one standard deviation. R stands for acceleration factor which is either 3 or 4.

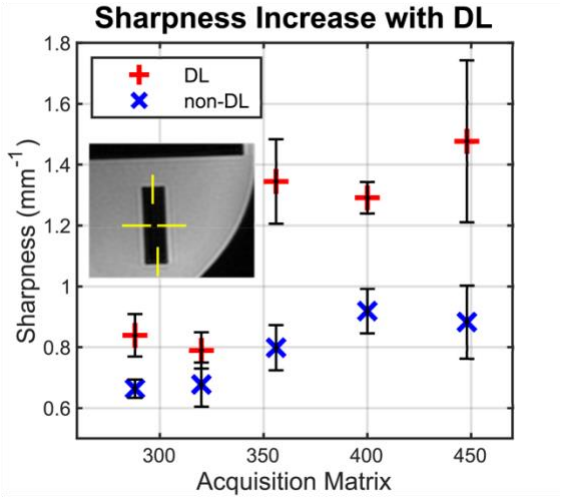

Figure 2: Effect of HR Imaging and DL Reconstruction on Edge Sharpness in a Phantom. Thedeep learning reconstruction (DL) provided a 1.2 to 1.7 times increase in the image sharpness compared to a conventional reconstruction. Shown here are sharpness measurements averaged over four internal edges of the phantom imaged with the high-resolution protocol (inset). Non-DL and DL results are shown. The increase in sharpness is more pronounced at higher acquisition

matrices. Error bars represent standard deviation of measured sharpness.

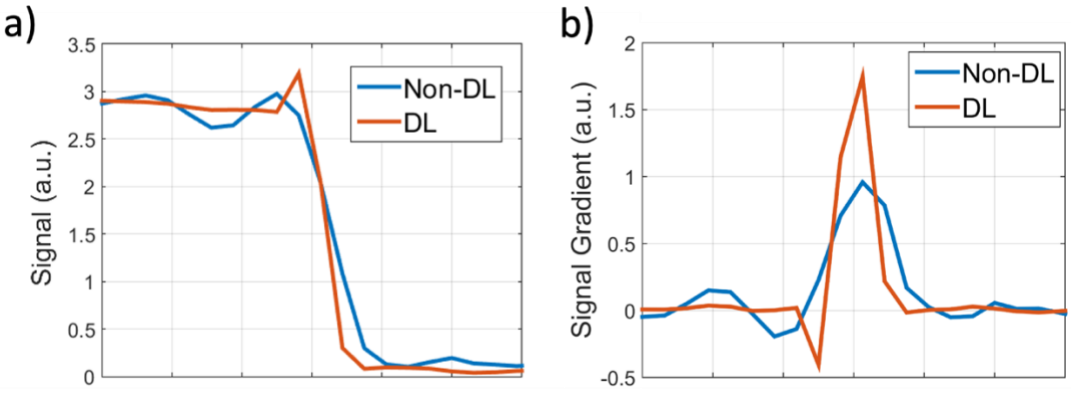

Figure 3: Edge Response With and Without the DL Reconstruction. a) Edge response in a phantom both with (DL) and without (non-DL) a DL reconstruction applied. With the DL reconstruction there was a reduction in truncation artifact and a steeper edge response consistent with enhanced image sharpness. **b)** The derivative of the edge response further demonstrated a reduction in truncation artifact and enhanced sharpness.

2.3.2 SOC Imaging

Included in the collected SOC cases were two subjects with bilateral silicone implants, two with unilateral silicone implants, and one subject imaged during lactation. One subject was excluded because the k-space data were unavailable due to an error in data transfer, leaving 19 for the analysis. Example cases with both non-DL and DL images are shown in **Figure 4** and **Figure 5**. Images reconstructed using DL had a statistically significant increase in each of the four tested categories (presence of artifacts, perceived sharpness, perceived SNR, and overall quality) when compared to the non-DL reference **(Figure 6)**.

Of the 76 reader/subject combinations with the SOC protocol (19 patient MR cases with 4 readers each), the DL images were preferred in 71 instances (93%). Radiologists stated that the preference was due to the decreased image noise and increased sharpness. The non-DL images were preferred in 4 instances and no preference was found in 1 instance. Given reasons for preference of the non-DL images mentioned increased artifact intensity in DL images. Details of

the preferences for each reader can be found in **Table 3**. T2 bright features in DL images had equal to or greater intensity than the same features in non-DL images for 65 out of 76 cases (86%).

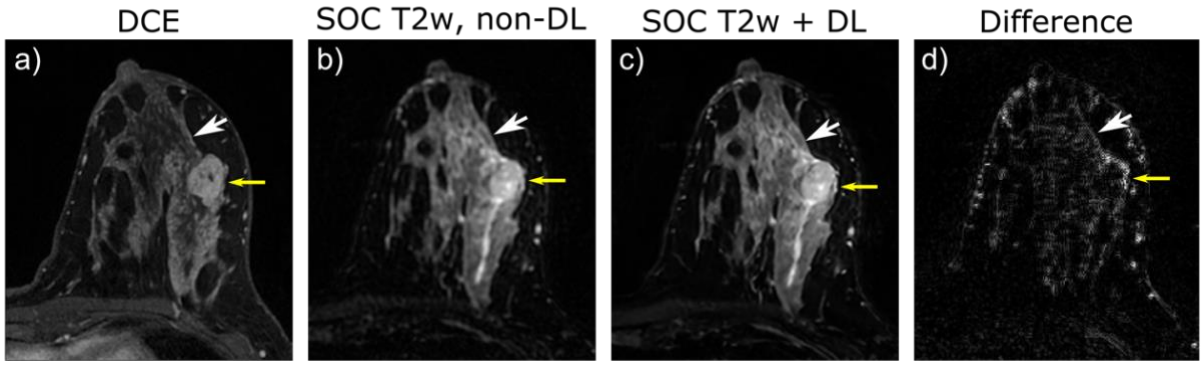

Figure 4: SOC T2w Imaging with DL Reconstruction – Fibroadenoma. 37-year-old female who underwent breast MRI for high-risk screening. Left breast contained a biopsy-proven benign fibroadenoma with intermediate signal on the T2w images (yellow arrow, small arrowhead). **a)** Contrast enhanced T1w image showed fibroadenoma enhancement (yellow arrow, small arrowhead). **b)** Standard-of-care (SOC) T2w acquisition without deep learning reconstruction (DL). **c)** Same raw data as b) but with DL. The DL reduced the image noise and made the fine fibroglandular tissue details (white arrow, large arrowhead) more visible compared with the non-DL image. **d)** The absolute difference between b and c demonstrated the reduction of noise and edge sharpening.

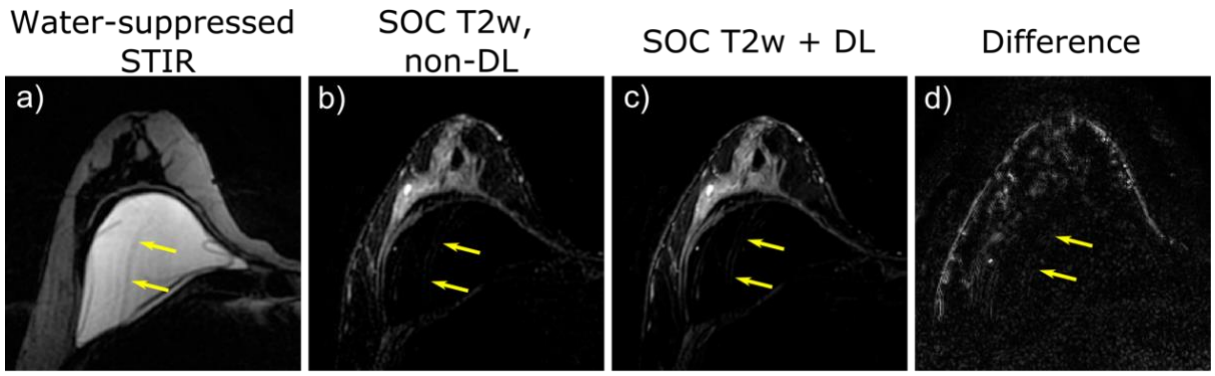

Figure 5: SOC T2w Imaging with DL Reconstruction – Silicone Implant. A 65-year-old with bilateral silicone implants. The deep learning reconstruction (DL) provided an apparent decrease in image noise and increased image sharpness. In this example, a motion artifact is intensified within the implant (arrows). **a)** Water-suppressed STIR image showed the silicone implant and the motion artifact. **b)** Standard-of-care (SOC) T2w acquisition without deep learning. **c)** Same raw data as b but with DL which showed enhancement of the motion artifact. **d)** Absolute value of difference between b and c further showed the artifact enhancement.

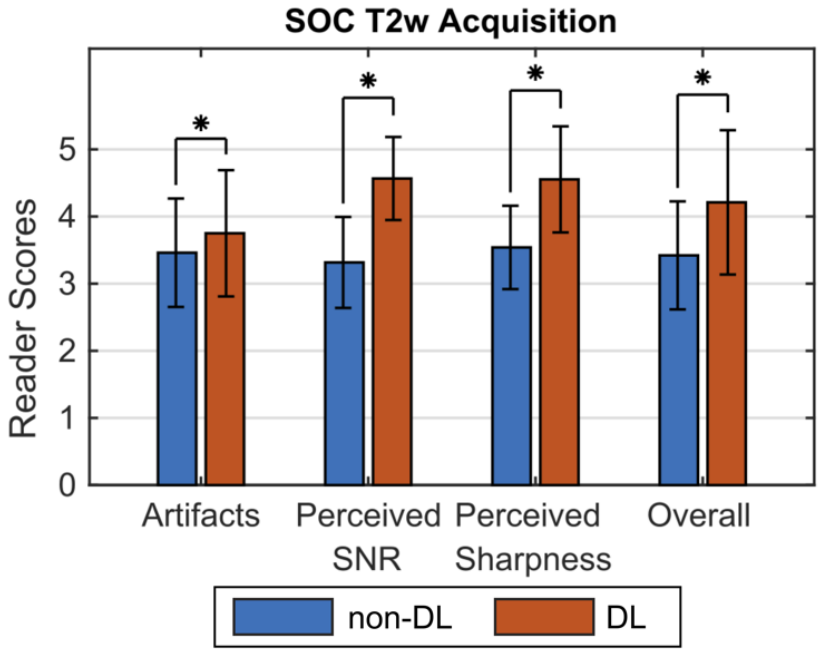

Figure 6: Image Quality Assessment of Standard-of-Care T2w with DL Reconstruction. Semi-quantitative comparison of mean image quality showed that images with the deep learning reconstruction (DL) scored significantly higher than non-DL images in each measured category. Stars indicate statistically significant differences with $p < 0.05$. Error bars demonstrate one standard deviation.

Table 3: Reader Preference for DL or non-DL Reconstruction.

SOC Reader Preference			
	SOC, non-DL	SOC, DL	No Preference
Reader 1	1	18	-
Reader 2	-	19	-
Reader 3	3	15	1
Reader 4	-	19	-

HR Reader Preference			
	SOC, non-DL	HR, non-DL	HR, DL
Reader 1	2	2	50
Reader 2	23	31	-

SOC: Standard of Care, HR: High Resolution, DL: Deep Learning.

2.3.3 HR Imaging

HR protocol data from two subjects were excluded due to a technical and operator error during the pre-scan calibration procedure leaving 54 subjects in the analysis. The SOC sequences

required an average acquisition length of 241 seconds while the HR required 293 seconds (20% increase) due to an increase in the number of phase encodings. Without the increase of parallel imaging acceleration factor from 3 to 4, the HR acquisition would have necessitated an estimated 374 second acquisition (55% increase).

Example HR images can be found in **Figure 7** and **Figure 8**. Results of the image quality scoring are summarized in **Figure 9**. The non-DL HR images scored significantly higher in perceived image sharpness, but significantly worse in image artifacts and perceived SNR when compared to SOC images. No significant change was detected in overall quality ($p = 1.0$). Application of the DL reconstruction to HR data led to a significant increase in perceived sharpness, perceived SNR, and overall quality compared to non-DL HR images. However, DL HR images scored significantly worse in the artifact category. Comparing DL HR images to SOC images demonstrated a significant increase in SNR, sharpness, and overall quality but significantly worse performance in the artifact category.

Image preference was split between the two radiologists (**Table 3**). One reader preferred the HR T2w images with DL in 50 of 54 cases. Reasons given were the decreased image noise and enhanced sharpness. The second reader was split between SOC T2w images (23 out of 54) and the HR images without DL (31 out of 54). Comments from this radiologist indicated that they felt the HR images with DL “looked too sharp” or had a “fake look to anatomy outside [the] breast”.

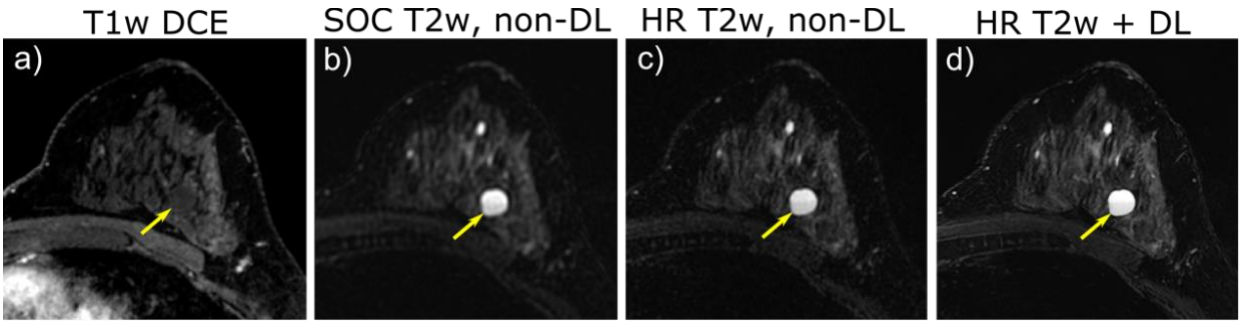

Figure 7: HR T2w Imaging with DL Reconstruction – Cysts. A 48-year-old female undergoing high risk screening MRI with multiple benign T2 hyperintense cysts including a complicated cyst with fluid-fluid level (yellow arrows). **a)** Contrast-enhanced T1w image. **b)** Standard-of-care (SOC) T2w image without deep learning (DL). **c)** T2w imaging with a higher spatial resolution HR showed image sharpening with the boundary between the two cyst fluids clearly visible. However, the poor SNR made visualization of the fibroglandular tissue difficult. **d)** The HR T2w image with DL maintained superior image sharpness while improving overall SNR.

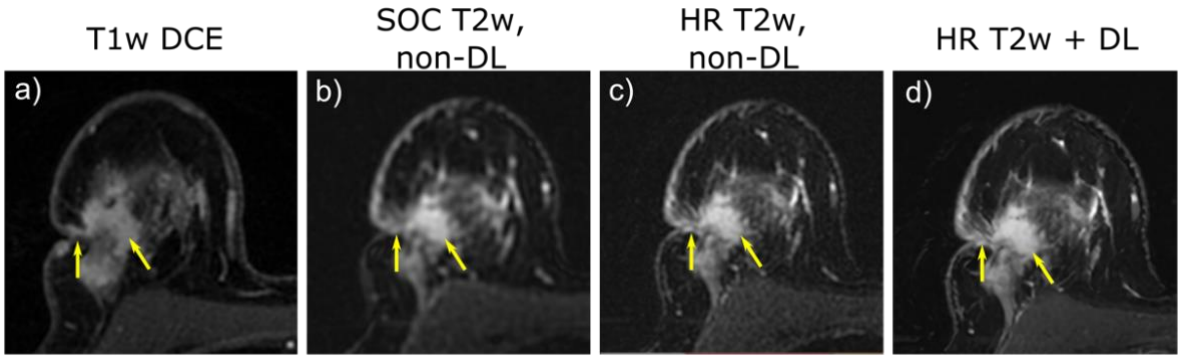

Figure 8: HR T2w Imaging with DL Reconstruction – Skin Invasion. Images from a 52-year-old female with recurrent cancer and suspected skin invasion (arrows). Skin invasion was later confirmed by pathology. **a)** Contrast-enhanced T1w image (DCE) showed the lesion enhancement. **b)** Standard-of-care (SOC) T2w imaging, without deep learning (DL) reconstruction was relatively blurry and lacked spatial detail. **c)** A high-spatial resolution (HR) T2w acquisition provided more detail but at the cost of SNR. **d)** Application of a DL reconstruction allowed for detailed images with good SNR.

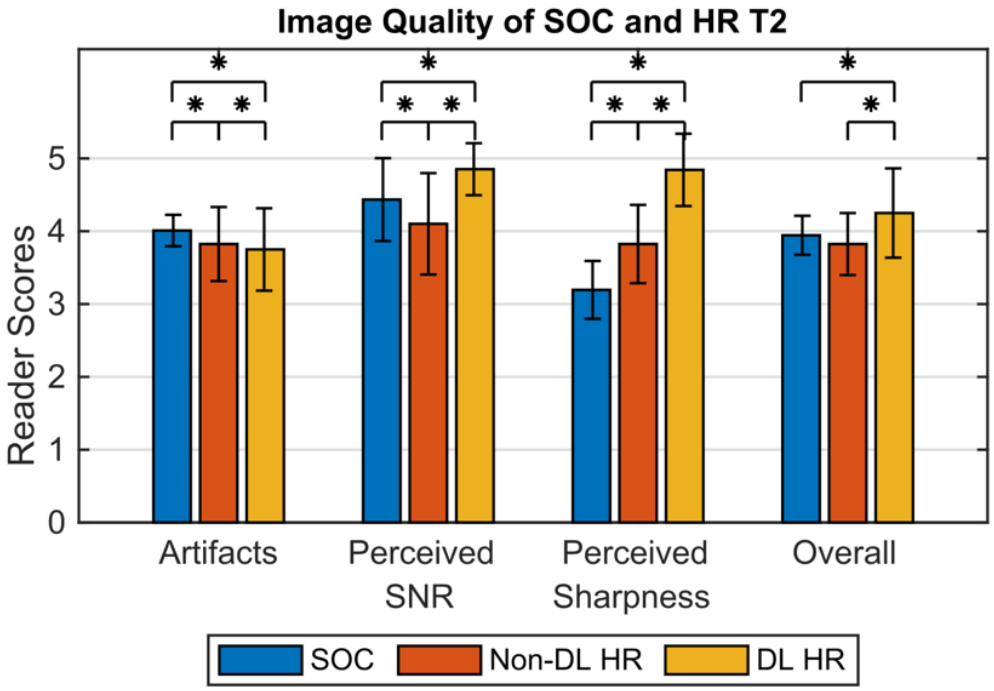

Figure 9: Image Quality Assessment of High Resolution T2w with DL Reconstruction. The high spatial resolution (HR) protocol without a deep learning (DL) reconstruction led to a significant increase in perceived image sharpness but significantly worsened scoring for both artifacts and perceived SNR. Subsequent application of the DL reconstruction led to significantly higher perceived SNR with scoring exceeding that of the standard-of-care (SOC) images. Further, there was an additional increase in perceived sharpness and overall quality for HR images with the DL reconstruction. Data was from 54 subjects evaluated by 2 radiologists. Stars indicate statistically significant differences in the scores with $p<0.05$.

2.4 Discussion

This work evaluated the performance of a DL reconstruction in T2w breast MRI for both a SOC and a HR protocol. Overall, the DL-based reconstruction performed well in the setting of T2w breast MRI. Phantom imaging showed increased SNR and image sharpness when DL was used. In human subjects, application of the DL reconstruction to the SOC protocol improved all measured image quality metrics. HR imaging without the DL reconstruction caused a drop in perceived SNR scoring compared with SOC imaging but application of the DL reconstruction increased perceived SNR above that of SOC images. This was despite the increase in parallel imaging acceleration. The higher acceleration limited the acquisition time increase to 20%

allowing for HR imaging in clinically feasible scan time. In general, it appears that the DL algorithm provides a reduction in protocol compromises allowing for T2w imaging in aa reasonable scan time, with higher spatial resolution, and a higher SNR.

The relative signal intensity remained consistent after application of DL with T2 bright features retaining their hyperintensity in 86% of cases. This suggests that the contrast and, subsequently, the clinical utility of the images were preserved. Additionally, the breast radiologists in this study largely preferred the SOC DL images over their non-DL counterparts. Together with the higher perceived quality, these findings indicate that the DL could potentially find acceptance in a clinical imaging scenario. DL may provide sharper T2w images with increased SNR that better assist in interpretation of T1w DCE images. This may lead to superior diagnostic performance of T2w imaging. Future work assessing the diagnostic performance of DL images compared to the non-DL images is needed.

During phantom experiments, a substantial increase in sharpening occurred when the matrix was increased from 320 x 320 to 356 x 356. At the same time, the reconstructed DL images increased in size from 512 x 512 to 1024 x 1024 pixels. This interesting result is explained by looking at the DL reconstruction in more detail. The DL reconstruction pipeline involves an interpolation of the acquired data based on the truncation artifact[24]. The increase in image size was necessary to allow for sufficient interpolation of the acquired data and was performed automatically in the prototype reconstruction as the matrix size exceeded 320 x 320. The additional interpolation explains the observed increase in sharpening at higher acquisition matrix sizes.

While DL improved artifact scores for SOC imaging, scores worsened when used with HR imaging. In both scenarios, the radiologists noted that artifacts were sometimes “intensified” after application of DL. Previously published work reported similar worsening of pulsation and motion

ghost artifacts with the same DL reconstruction[66]. The DL algorithm is not trained to detect any artifacts other than those due to data truncation[24]. Therefore, it is expected that existing artifacts may become more pronounced due to sharpening of artifact edges and a reduction in surrounding noise. In this work, HR imaging may have worsened already existing artifacts (such as ghosting) or introduced new artifacts which are then "intensified" by the DL. This idea is supported by our data in which HR imaging showed an initial decrease in artifact scores and a further decrease after DL was applied. The radiologist reported that the "intensified" artifacts were not necessarily detrimental to their image interpretation. Although more prominent with DL than without, artifacts retain a familiar appearance. Furthermore, readers reported that the "intensified" artifacts were sometimes easier to identify as being artifactual and consequently ignored. However, this may not be the case for all specific clinical scenarios. Because it is impossible to fully eliminate artifacts from raw data, it is prudent to consider what types of artifacts are to be expected with a particular sequence and how they may be impacted by the DL reconstruction.

A differential image preference was observed between reader 1 and reader 2 when viewing HR images. Reader 1 generally preferred the HR DL images and reader 2 preferred non-DL images citing over-smoothing of features and a "fake look" to the DL images. However, reader 2 still gave DL images higher image quality scores. The reader's proclivity for non-DL images appears to be a preference and not indicative of any image quality issues. A difference between reader 1 and reader 2 was that reader 1 participated in both reader studies while reader 2 only participated in the HR study. Reader 1 was therefore more familiar with the appearance of the DL reconstructions at the time of review. According to the global-focal search model of image perception, radiologists have an expert's schemata, that is, a mental picture of what the typical MR image should look like[67]. The DL reconstructed images may have been too different from reader 2's expert schemata

to be accepted as “real” images. Concerns about “over smoothing” may be addressed by reducing the noise reduction parameter of the DL reconstruction. In this work, we used a value of 0.75 which indicates a high amount of denoising. Other investigators have compared radiologists’ preference for images with 0.75 and 0.50 noise reduction in hip and shoulder MRI[28]. They found radiologist preference was split between the 0.50 and 0.75 levels of noise reduction. Decreasing the noise reduction parameter in our case may help to alleviate some of the concerns raised by reader 2.

This work has several limitations. For clinical interpretation, T2w breast images are typically generally viewed in context with the T1w DCE sequence. However, in this study the radiologists were asked to perform their assessment of the T2w images independently of T1w DCE images focusing on radiologists’ perception of image quality and not on clinical performance. As such, an assessment of diagnostic performance was not performed. Instead, this work provides the foundation of quality improvements to justify further study of the DL algorithm’s impact on clinical interpretation and decision making. Further evaluation of the algorithm is needed to assess diagnostic performance. Another limitation is the receive coil array used for phantom studies was not identical to the breast coil used for human subjects imaging. Therefore, the exact SNR increase measured in phantom may not be directly replicable during in vivo imaging. Finally, this study was performed exclusively at 3 T and did not include 1.5 T MR imaging. The DL algorithm may provide additional benefit to imaging at 1.5 T due to the overall increase in image noise expected at lower field strengths. Future work should include imaging at 1.5 T.

2.5 Conclusion

The DL reconstruction performed well in the setting of T2w breast MRI using multiple imaging protocols. It allowed for superior image quality in SOC and HR protocols while maintaining hyperintense signal in T2 bright features, and most radiologists preferred the DL

images. While image artifacts may be intensified, it remains to be determined if this is detrimental to image viewing or interpretation. Overall, the DL reconstruction is a promising technique for reducing image quality compromises in T2w imaging of the breast.

Chapter Three

Automated Prescription of Scan and Pre-scan Volumes[2]

3.1 Introduction

As outlined in **Section 1.1.4**, setup of breast MRI exams include placement of both a single scan volume and bilateral pre-scan volumes. This process is time-consuming, tedious, prone to errors, and subject to both intra- and inter-operator variability. Thus, there is a need for an automated, fast, and reliable technique for placement of scan and pre-scan volumes in breast MRI. The purpose of this chapter is to demonstrate the feasibility of a breast-specific, deep-learning approach for automated placement of these volumes in breast MRI.

3.2 Methods

A deep convolution neural network was trained to place both scan and pre-scan volumes. The training image data was collected from clinical exams performed on scanners across the UW Health system. Scan volumes were collected from the clinical exams while pre-scan volumes were generated offline. The neural network was trained separately to place both scan and pre-scan volumes generating two separate volume placement models. Model performance was assessed

[2]This chapter has been published as a research article in the journal *Tomography.* Allen, T. J., Henze Bancroft, L. C., Wang, K., Wang, P. N., Unal, O., Estkowski, L. D., Cashen, T. A., Bayram, E., Strigel, R. M., & Holmes, J. H. (2023). Automated Placement of Scan and Pre-Scan Volumes for Breast MRI Using a Convolutional Neural Network. *Tomography*, *9*(3), 967–980. https://doi.org/10.3390/tomography9030079

using multiple metrics assessing the agreement of model-predicted volumes with the volume used for network training. The methods are further detailed in the following sections.

3.2.1 Data Collection and Curation

Clinical breast MR exams (N = 413) from 12/07/2016 to 3/14/2018 were de-identified and retrospectively collected from different 1.5 T and 3 T MRI scanners (N = 11) from a single MRI vendor (GE Healthcare, Waukesha, WI) within the UW Health system. All data collection was performed under IRB approval. For this feasibility work, patients with implants (N = 42) and substantial post-operative deformation due to surgery (e.g. mastectomy or lumpectomy, N = 21) were excluded. Exams with incomplete imaging (N = 13) or insufficient scan volume placement (N=4) were also excluded leaving a final set of 333 cases. Patient age, indication for their clinical exam, and final BI-RADS assessment category were recorded. Each breast exam began with 3-plane anatomical scout imaging based on a 2D acquisition which collected images in axial, sagittal, and coronal orientations. Select scan parameters from the 3-plane anatomical scout are summarized in Table 4. If multiple scout image series existed in an exam, the latest series in the exam was used.

Table 4: Summary of Breast MR Exam Data for Auto Prescription Training

Cases Collected					413	
Exclusions						
Implants					42	
Surgical Changes*					21	
Incomplete Data					13	
Poor Scan Volume Placement					4	
Inclusions						
#	Model	Field Strength (T)	Slice Quantity	FOV (cm)	Scan Volume	Pre-scan Volume
1	SIGNA HDxt	1.5	30-44	40	110	74
2	SIGNA Artist	1.5	45	38-46	21	7
3	Optima MR450w	1.5	35-60	40	22	22
4	Optima MR450w	1.5	45	38-44	24	12
5	SIGNA HDxt	1.5	45	40	2	1
6	Discovery MR750	3	37	44-46	0	0
7	SIGNA Premier	3	45	44	1	1
8	SIGNA Architect	3	45	44	41	17
9	SIGNA PET/MR	3	81	44	25	21
10	Discovery MR750w	3	40-44	40	72	42
11	SIGNA Premier	3	45	44	15	5
Total					333	202

*Surgical changes refer to exams excluded from training due to morphological changes brought about by surgery (such as mastectomy or lumpectomy). FOV: Axial field of view. All exams from scanner 6 were excluded due to incomplete anatomic coverage of the scout images.

The scan volumes and bilateral pre-scan volumes placed by technologist during the clinical exams were collected for each exam. A large amount of variation was observed in the clinically placed pre-scan volumes (**Figure 10**), thus, they were determined to be unsuitable to serve as training volumes for a machine learning network. An alternative set of bilateral pre-scan volumes were generated by a team of expert users in an offline configuration. The offline pre-scan volumes were placed by one medical physics graduate student and then reviewed in consensus by two experienced clinical physicists. Volumes were iteratively adjusted until all three reviewers agreed that they conformed to guidelines for good pre-scan volume placement provided by the MR

scanner vendor. Notice that the scan and pre-scan volumes come from two different sources; scan volumes were placed by technologist during the clinical breast MR while pre-scan volumes were placed offline by expert users.

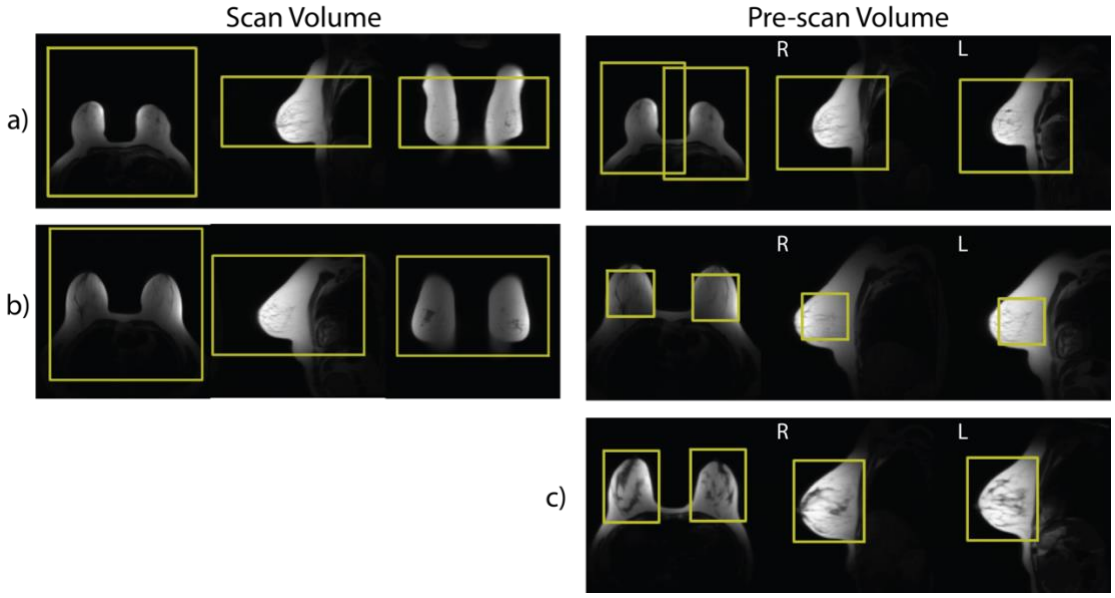

Figure 10: Example of Clinical Scan and Pre-scan Volumes in Breast MRI. Examples **a)** and **b)** show scan volumes and pre-scan volumes as prescribed during clinical breast MRI. Note that the placement of scan volumes was consistent across cases. However, the pre-scan volume placement was found to vary dramatically between cases even when the underlying anatomy is similar as is the case in a and b. Example **c)** shows an example of offline volumes generated by expert users used to train pre-scan volume placement providing more consistent technique.

Each volume (scan or pre-scan) was cuboid in shape as demonstrated in **Figure 10** and **Figure 11**. The scan volume was square in the axial plane and was described by five placement parameters: the positions in the left-right (LR), anteroposterior (AP), superior/inferior (SI) directions, axial size (i.e. field-of-view, FOV), and extent of SI coverage. The bilateral pre-scan volumes were positioned independently, one over each breast, but the two volumes shared a common size. Therefore, the pre-scan volumes were parameterized by 9 values: LR, AP, SI position for each of the two volumes, and the size in each dimension.

3.2.2 Network Training

A convolutional neural network based on Alexnet[68] was trained for the placement of scan volumes and pre-scan volumes using the scout images as input. The network contained 5 convolutional layers and 3 fully connected layers. Dropout layers with dropout rate of 0.5 were included between the fully connected layers. For the purpose of this work, the network in combination with a set of trained weights will be referred to as a placement model. The single network was trained under two conditions resulting in two placement models: 1) scan volume placement trained on clinically placed volumes and 2) pre-scan placement trained on the offline, consensus-based volumes. Although, the network structure was identical for both trainings, the size of the final output layer was adjusted to output scan or pre-scan placement parameters (5 vs 9, respectively).

Before the scout images were used to train the neural network, several pre-processing steps were performed **(Figure 11)**. The axial, sagittal, and coronal images were separated, and maximum intensity projections (MIPs) were calculated for each view. Each MIP was then binarized via Otsu's method of automated thresholding to create three binary masks[69]. The three binary masks were then used as input into the network.

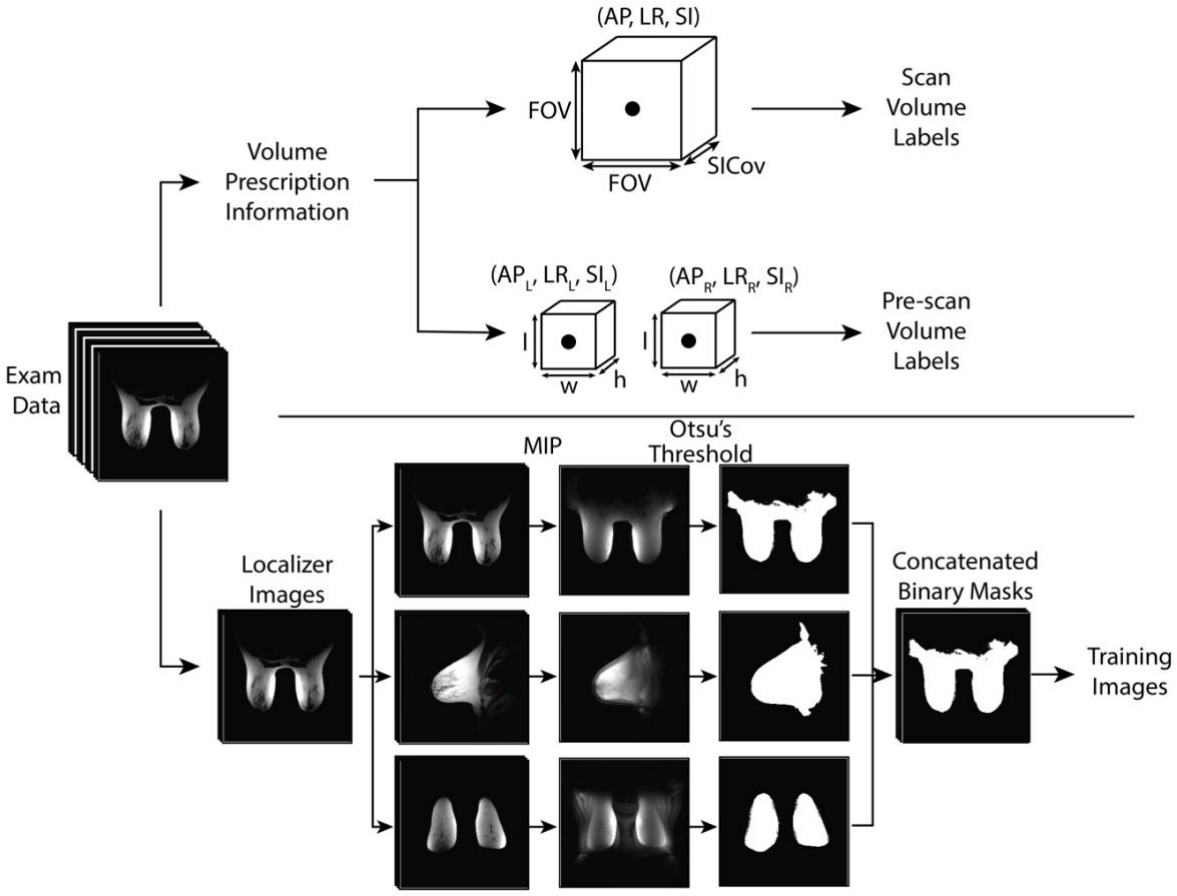

Figure 11: Volume Parameterization and Image Preprocessing. Parameterization of volume information (top) and image pre-processing of anatomical scout images (bottom) prior to neural network training. Note that the two shim volumes (represented by two cuboids) have a shared size. FOV: Axial field of view. SICov: Coverage in the SI direction. MIP: Maximum intensity projection.

GPU-accelerated training was performed on a NVIDIA DGX A100 system with a batch size of 32. Using conventional approaches for training a deep learning network, weights are iteratively updated according to the loss function used. In this work, two loss functions were tested, and the best performing approach was used for the final model. The two loss functions were: 1) a 3D version of the generalized intersection over union (GIoU[70]) and 2) the root-mean-squared error (RMSE). The optimizer used was RMSprop[71]. Five times data augmentation was performed in the

form of shifts in all 3-dimensions. Five-fold cross validation of each placement model was performed. Using this technique, the training data (images and their associated volumes) are split into five groups with each group serving as the testing data in a round-robin fashion. This is thought to provide a more comprehensive assessment of the model performance across all available data compared to a simple training/validation/testing split.

Estimates of the model uncertainty were generated using Monte Carlo dropout. This technique generates multiple predictions for a given input and selects a random subset of the model weights that is used for each prediction[43]. Monte Carlo dropout can also increase model accuracy since distribution averages may be more accurate than any individual prediction. In this work, the mean of 100 predictions was used as the final placement prediction and one standard deviation represented the prediction uncertainty.

3.2.3 Model Performance

Model performance was characterized by several metrics including the 3D intersection over union (IoU), 2D IoU, distance between volume positions, and percent difference in volume size. 3D IoU quantifies the agreement of the known and predicted 3D volumes, while the 2D IoU quantifies agreement of the 2D boxes obtained by projecting the 3D volumes into axial, sagittal, and coronal planes. For the bilateral pre-scan volumes, metrics were calculated for each volume independently. The dependence of model performance on the number of Monte Carlo predictions was evaluated as was the performance across different scanners. A comprehensive agreement score was developed by summing the IoU scores for the scan volume and the two pre-scan volumes. This metric is referred to as the combined IoU or cIoU and can range from 0 to 3 with higher scores indicating higher levels of agreement.

The time to prediction is important to characterize for any future clinical implementation. Therefore, the time to prediction starting from the raw, unprocessed images was measured. Timing measurements were obtained from 10 cases and averaged to find the mean time to prediction. The timing study was performed on a 2018 MacBook Pro with 2.6 GHz 6-Core Intel Core i7 processor. No GPU acceleration was used during the timing estimation.

3.3 Results

3.3.1 Data Collection

A total of 333 exams were used to train the scan volume placement model. The majority (72%) of these exams were performed to screen for breast cancer, 16% were to evaluate the extent of known disease, 6% were for neoadjuvant chemotherapy response assessment, and the final 6% were in response to other imaging or clinical indications. Final BI-RADS assessment categories ranged from 0 to 6 with at least 5 cases from each category. BI-RADS assessment categories of 1 and 2 comprised 71% of the exams. Patient age ranged from 20 to 77 years old with a mean age of 47. A total of 208 sets of pre-scan volumes were generated offline by the expert users and used to train the pre-scan model. Exam indication percentages, distribution of BI-RADS categories, and patient ages for the exams used for pre-scan volumes training were similar to those used for scan volume training.

3.3.2 Model Performance – Scan Volume

Training with the RMSE loss function provided the highest average 3D IoU for scan volume placement. The histogram showing the 3D IoU distribution for all cases (**Figure 12**) demonstrates good overall performance of the scan volume placement model with a mean 3D IoU of 0.68. **Figure 12** also shows examples of excellent and average scan volume placements using

the RMSE-trained network and highlights the spatial relationship between the predicted volume and the anatomy. Distribution statistics for the various placement metrics can be found in **Table 5**. The average error in model predicted volume size was 4%, and the average distance between the center of the model predicted volume and clinically placed volume was 3.1 cm. The average amount of overlap between the model predicted and technologist placed volumes was 82%. Results obtained using the GIoU loss function for training can be found in appendix B (**Table B1**).

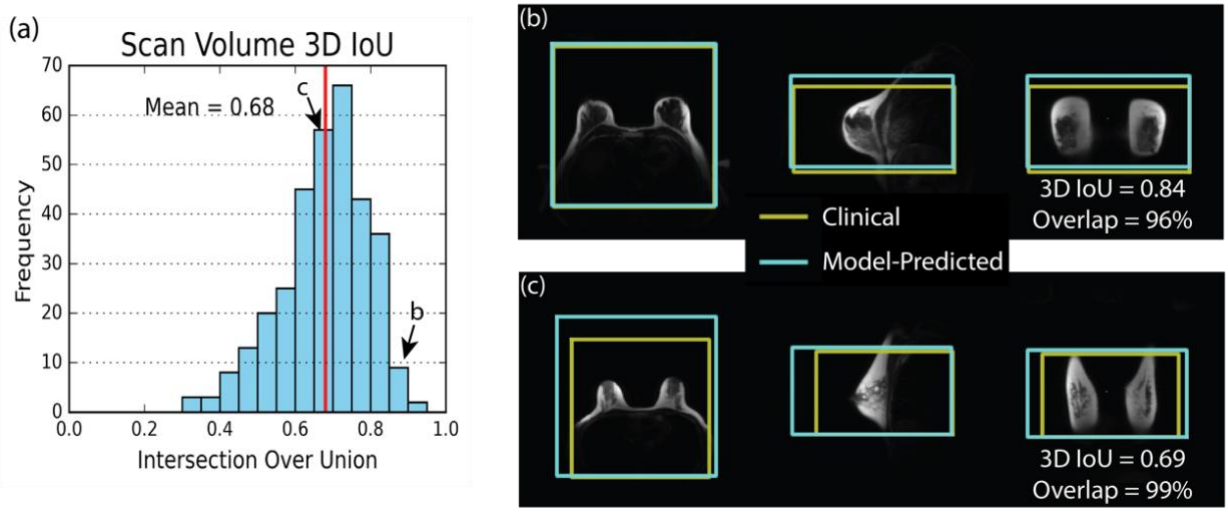

Figure 12: Histogram and Examples of Scan Volume Placement Performance. The scan volume 3D intersection over union (IoU) histogram, **(a)**, and two placement examples, **(b,c)**, demonstrate that the model placed a scan volume that is similar to scan volumes from the clinical exam. The red line indicates the mean value of 3D IoU. In **(b)** and **(c)**, the yellow box shows the technologist-placed volume generated during the clinical exam. The blue box shows the model-predicted volume. **(b)** shows an excellent case with a high 3D IoU of 0.84 and 96% overlap of the clinically placed volume. **(c)** shows an average performing example with 3D IoU of 0.69 and overlap of 99%.

Table 5: Scan Volume Placement Metrics

Metric	5th %	Median	95th %	Average
3D IoU	0.46	0.69	0.85	0.68
Axial IoU	0.61	0.81	0.95	0.79
Sagittal IoU	0.53	0.73	0.89	0.71
Coronal IoU	0.6	0.78	0.92	0.77
Distance (cm)	0.9	2.7	6.6	3.1
Volume Error (%)	-30	2	45	4
Overlap (%)	57	84	99	82
RMSE (cm)	0.9	1.9	3.6	2.0

IoU: intersection over union, Distance: distance between the model predicted volume centers and the technologist prescribed volume center, RMSE: Root mean squared error between all 5 scan volume placement parameters. 5th and 95th % stand for the 5th and 95th percentiles of the distributions, respectively.

3.3.3 Model Performance – Pre-scan

Training for pre-scan placement using the GIoU loss function resulted in higher average 3D IoU than with the RMSE loss function. The average 3D IoU was 0.66 and the histogram showing the distribution of 3D IoU for all cases is shown in **Figure 13**. Examples of excellent and average placement are also shown with anatomical context. There was no significant difference in mean 3D IoU between the right and left volumes (p=0.68). The mean distance between expert-placed and model-predicted volume centers was 1.4 cm. No significant difference between left and right mean distance was observed (p=0.80). The average volume size error was 3%. Distribution summaries of these metrics and given in **Table 6**. Results from the RMSE-trained model can be found in Appendix B (**Table B2**).

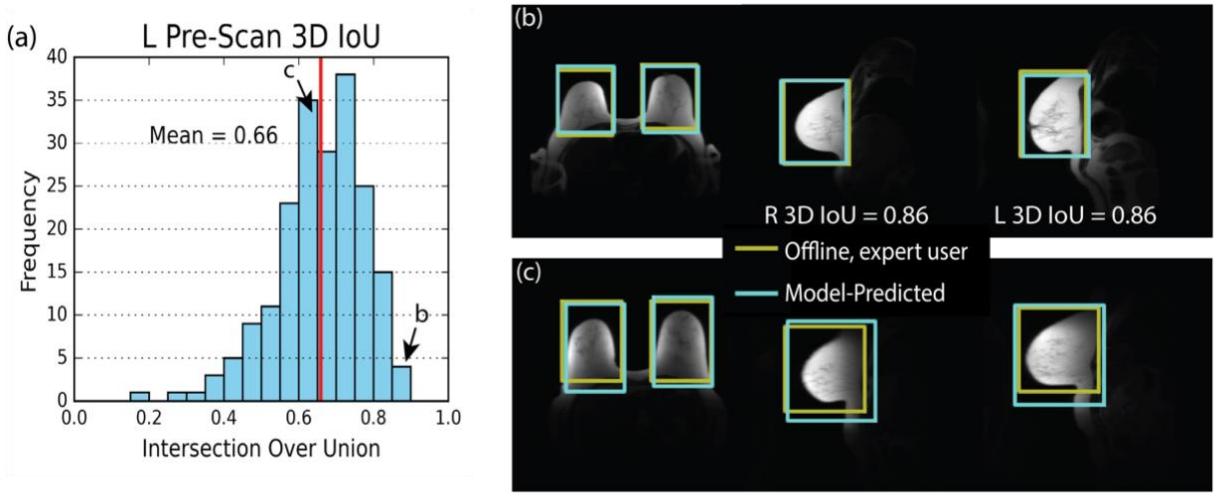

Figure 13: Histogram and Examples of Pre-Scan Volume Placement Performance. The pre-scan volume 3D intersection over union (IoU) histogram, **(a)**, and the two examples, **(b,c)**, demonstrate that the model predicted a pair of pre-scan volumes that closely match the offline, expert-placed prescan volumes. The histogram shows results of the left (L) volume only. However, the right (R) volume's distribution is similar. The red line indicates the mean value of 3D IoU. In **(b)** and **(c)**, the yellow box shows the expert user placed volume generated in an offline setting. The blue box shows the model-predicted volume. **(b)** shows an excellent case with a high 3D IoU for both pre-scan volumes. **(c)** shows an average performing case.

Table 6: Pre-scan Volume Placement Metrics

Parameter	Side	5th %	Median	95th %	Average
3D IoU	R	0.45	0.65	0.83	0.65
	L	0.43	0.68	0.83	0.66
Axial IoU	R	0.52	0.75	0.90	0.73
	L	0.60	0.78	0.90	0.76
Sagittal IoU	R	0.51	0.73	0.87	0.71
	L	0.53	0.73	0.87	0.72
Coronal IoU	R	0.49	0.73	0.89	0.72
	L	0.55	0.75	0.90	0.74
Distance (cm)	R	0.5	1.3	2.9	1.5
	L	0.5	1.2	3.0	1.4
Volume Error (%)	N/A	-35	-2	56	3
RMSE (cm)	N/A	0.6	1.2	2.2	1.2

IoU: intersection over union, Distance: distance between the model predicted volume centers and the expert prescribed pre-scan volume centers, Side: Pre-scan volume placed on right (R) or left (L) breast, RMSE: Root mean squared error between all 9 pre-scan placement parameters. 5th and 95th % stand for the 5th and 95th percentiles of the distributions, respectively.

3.3.4 Uncertainty Estimate

Table 7 demonstrates uncertainty estimates for the predicted placement parameters derived using Monte Carlo dropout. For scan volume placement, the mean positioning uncertainty was 2.2 cm for AP position, 0.8 cm for LR position, and 1.4 cm for SI position. The mean uncertainties in the size parameters were 1.3 cm for axial size and 1.7 cm for SI coverage. 95% of predictions for all 5 placement parameters had an estimated uncertainty of less than 2.7 cm. For context, the average scan volume size is 33 x 33 x 19 cm^3.

The mean uncertainty estimates for the AP, LR, and SI position of both left and right side volumes were 0.5 cm, 0.3 cm, and 0.6 cm, respectively. The shared size parameters of the pre-scan volumes had mean uncertainties of 0.7 cm, 0.4 cm, and 0.8 cm in the AP, LR, and SI directions. Overall, 95% of predictions for the 9 pre-scan volume placement parameters had uncertainties under 1.0 cm. For context, the average pre-scan volume size is 13 x 11 x 16 cm^3.

Table 7: Uncertainty Estimates for Automated Prescription of Breast MRI

Scan Volume				Pre-scan Volume			
Parameter	5th %	Mean	95th %	Parameter	5th %	Mean	95th %
AP Position	1.3	2.2	3.4	AP Position L	0.3	0.5	0.9
LR Position	0.5	0.8	1.3	LR Position L	0.2	0.3	0.6
SI Position	0.9	1.4	2.2	SI Position L	0.3	0.6	1.1
Axial Size (FOV)	0.8	1.3	2.1	AP Position R	0.3	0.5	0.8
SI Coverage	1.0	1.7	2.6	LR Position R	0.2	0.3	0.6
				SI Position R	0.3	0.6	1.0
				AP Size	0.5	0.7	1.1
				LR Size	0.3	0.4	0.6
				SI Size	0.5	0.8	1.1

All values are in cm. AP: Anterior-Posterior, LR: Left-Right, SI: Superior-Inferior, L: Left, R: Right, 5th %: 5th percentile, 95th %: 95th percentile.

3.3.5 Overall Model Performance

Model performance increased with the number of Monte Carlo dropout predictions. The average 3D IoU of scan volume model was 0.62 when a single prediction was used, 0.67 when 10 predictions were used, and 0.68 when 20 predictions were used. Further increase in the 3D IoU after 20 predictions was less than 0.01 (**Figure 14**). The average 3D IoU of pre-scan volumes exhibited similar behavior.

Figure 15 shows the scanner-by-scanner results for 3D IoU performance of both scan and pre-scan models. For scan volume placement on scanners with at least 5 cases, the average 3D IoU ranged from 0.73 to 0.46. The average 3D IoU for left pre-scan volumes on scanners with at least 5 cases ranged from 0.70 to 0.65. The behavior of the right pre-scan volumes was similar (results not shown). Scanners with a higher number of cases tended to exhibit a higher 3D IoU.

The average and median cIoU across exams which had both scan and pre-scan volumes was 2.0. The 5th percentile of 3D cIoU values was 1.5 and the 95th percentile was 2.5. The

standard deviation between the IoU of the three volumes was less than 0.1 for 98% of the cases. **Figure 16** shows an example exam that scored average in the cIoU metric.

Once trained, successive prediction of the scan and pre-scan volumes was achieved in an average of 4 seconds when only a single Monte Carlo dropout prediction was used. With 20 predictions, an average time of 16 seconds was required.

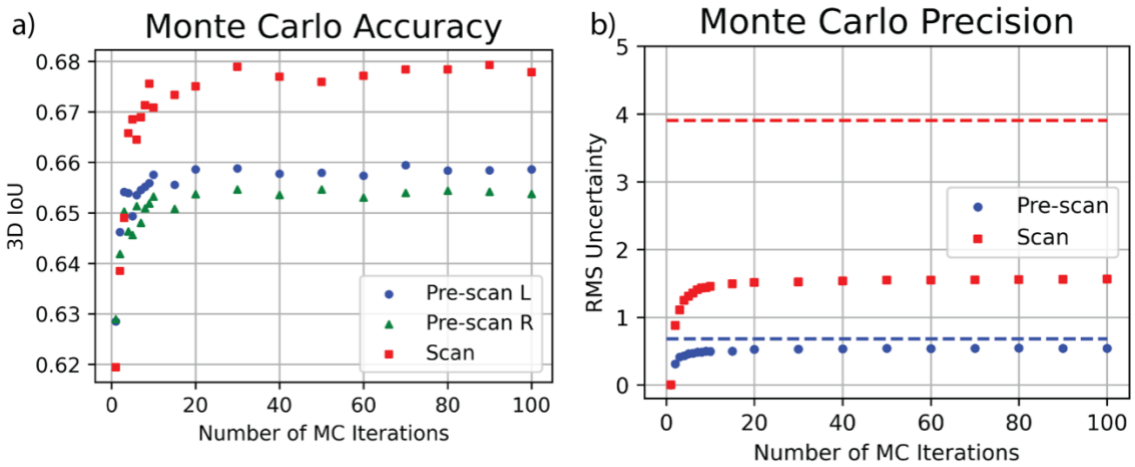

Figure 14: Effect of Monte Carlo Dropout on Prescription Accuracy and Precision. a) The 3D intersection over union (IoU) of the models increased with the number of Monte Carlo (MC) predictions/iterations. Performing at least 20 predictions achieved similar 3D IoU as 100 predictions. **b)** The average root-mean-squared (RMS) uncertainty of the models was accurately estimated using 20 predictions. The dotted lines show an upper bound on the uncertainty obtained by feeding in pure noise images to the trained models.

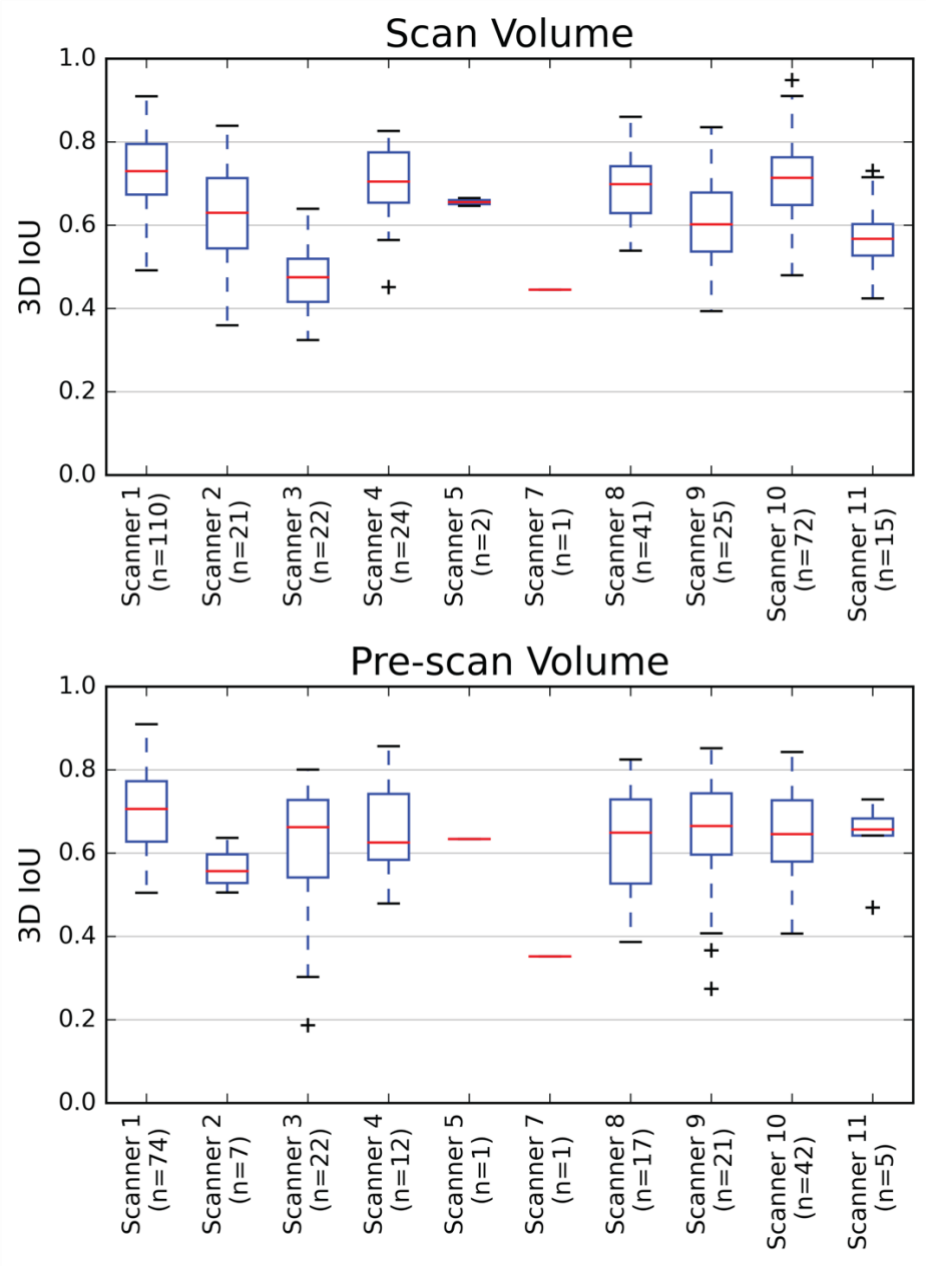

Figure 15: Variation in Volume Prescription by Scanner. Average 3D intersection over unions (IoU) for both scan volume and pre-scan volumes were good across multiple scanners from which the data were obtained. 3D IoU for the left pre-scan volume only is shown. The right pre-scan volume showed similar behavior. Generally, scanners with a larger number of cases tended to have higher 3D IoUs.

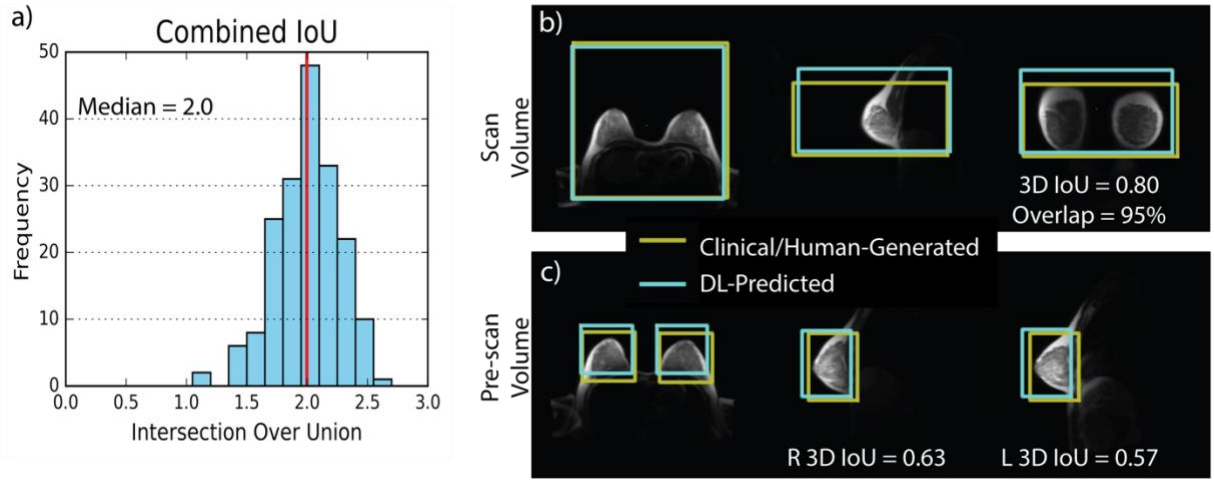

Figure 16: Histogram and Examples of Combined Scan and Pre-scan Volume Placement Performance. a) A mean value of 2.0 was measured for the histogram of the combined intersection over union (IoU). **b,c)** An example with an average combined IoU score of 2.0, consistent with the mean value for all cases studied, shows good agreement between DL-predicted and clinical or human generated volumes for both scan volume or pre-scan volumes.

3.4 Discussion

In this work, we demonstrated the feasibility of using a convolutional neural network for automated placement of scan and pre-scan volumes in breast MRI. Overall, we found that the trained models generated volumes which closely matched the human-placed volumes as evidence by the good values for the 3D IoU, the small difference in volume centers, and small difference in volume size. Model performance was consistent on data from multiple MR systems and with a variety of scout acquisition protocols. The uncertainty estimation feature of the models provided a method to gauge model confidence with results here on the order of a few cm. Prediction and uncertainty estimates required a reasonable number of Monto Carlo predictions and could be completed in a short amount of time. These results are promising and were obtained using a relatively small dataset. Additional refinement of these models may allow quick, reproducible, and automated placement of volumes needed for clinical breast MRI.

A novel feature of this work is the placement of bilateral pre-scan volumes in addition to the scan volume. Values for the 3D IoU of the pre-scan volumes were slightly lower than for the scan volumes. However, IoU measurements are sensitive to the size of the volumes being assessed. Smaller volumes will produce smaller IoU values for a given error in position or size. The other placement metrics indicate that the pre-scan and scan volumes are similar in performance and the difference in IoU is likely due to the smaller size of the pre-scan volumes. Placement of bilateral pre-scan volumes such as those used in this work is expected to be more difficult than scan volumes due to the additional placement parameters (nine vs five) necessary to predict pre-scan size and location. Therefore, the similar performance to scan placement seen in this work is encouraging.

Scan volume cases with the worst performance were observed to have anatomic variations that fell on the more extreme limits of our dataset, such as in cases with very large or small breast size. For the pre-scan volume, the worst cases were identified as subjects with small breast size and who were imaged with their arms above their head. In this setup, the breast tissue was stretched in the SI direction, leading to pre-scan volumes that were relatively small in the axial plane when compared to their length in the SI direction. With this shape, any location error in the axial plane led to a severe decrease in the 3D IoU metric. In general, for both the scan and pre-scan volumes, the error in location influenced the 3D IoU more than the error in size. While the volume size errors seemed to be quite high in some cases (from −40% up to 50%), it is useful to note that the volume scales quickly with small size errors. For example, a 1 cm overestimation of each side of a $10 \times 10 \times 10$ cube leads to a size error of 33% and a 3D IoU of 0.75. Conversely, a 1 cm positional error in each direction for the same cube leads to an absolute distance error of only 1.7 cm, and a 3D IoU of 0.57. This example demonstrates the general fact that 3D IoU is more sensitive to

distance error than size error. In accordance with this, the location errors in this work were likely also the dominant driver of low 3D IoU.

Another innovation of this work is the inclusion of uncertainty estimates in the placement predictions. Often, machine learning results do not include any measure of uncertainty making it difficult to interpret the output. The models developed in this work were able togive an uncertainty estimate with every prediction. The estimates are academically useful in that they provide a measure of the overall model uncertainty. However, they could also be useful in a clinical imaging environment to alert the MR technologist to high levels of uncertainty and the need for further review or human intervention.

The models introduced here directly predict a full 3D volume. Previously described models required further processing to obtain the final scanning volume. For example, the approach described by Blansit et al. first generated landmark heatmaps then predicted slice planes based on these heatmaps[38]. Alternatively, Geng et al. predicted 2D boxes on individual scout image slices and the final volumes was that which contains all the predicted 2D boxes[36]. These additional processing steps introduce additional sources of error. For example, generating a 3D volume from 2D boxes will tend to overestimate the size of the 3D volume since it can only be as small as the largest 2D box. Because our approach predicts the 3D volume directly, additional errors due to conversion from 2D box to 3D volumes are avoided. This is particularly salient for pre-scan volume placement since overestimating the size of pre-scan volumes may result in sub-optimal scan-specific system settings and poor image quality.

Further development of neural network-based volume placement could benefit clinical breast imaging. Observations of clinical technologists at our institution indicate that the entire volume placement process typically takes around 1-2 minutes but can take up to 3 minutes. This

amount of time is substantial, especially when considering abbreviated breast MRI protocols that attempt to screen for breast cancer using shortened protocols, reduced table time, and ultimately decreased cost[8,9]. Conversely, the models introduced here can place a complete set of volumes in less than 20 seconds. Additionally, the models are expected to provide more reproducible volume placements than human technologists since the same input images will always yield the same result. Specifically, there should be no inter- or intra- operator variability with a neural network-based model. This could help achieve more consistent quality in breast MRI as well as more repeatable and reproducible quantitative MRI. However, the models are still sensitive to factors such as noise and patient position and, thus, the true placement consistency would need to be evaluated. Another potential use for the networks could be to assist in the training of MR technologists by comparing their own scan and pre-scan volumes with those generated by the placement models.

In this work, training the placement model using the RMSE loss function provided superior performance for the placement of scan volumes in terms of 3D IoU. This is inconsistent with the expected dependency where the use of a loss function similar to the performance metric would be maximize model performance. In the setting of 2D object detection, other researchers have commented on the instability of GIoU loss and the difficulty in obtaining accurate regression using GIoU loss with some groups proposing more sophisticated loss functions to improve the performance[72,73].

This study has limitations. First, the training dataset was relatively small for neural network training and cases with saline or silicone implants as well as major surgical changes such as mastectomy were excluded. While these are important clinical cases, they were not included for this initial feasibility work. Further research is necessary to assess the performance of the model

when using a larger training dataset and in the setting of implants or mastectomy cases. This study only utilized single shot fast spin echo-based scout acquisitions on a single vendor platform and future work is needed to expand the training data to include a broader range of imaging protocols and vendor implementations. Ultimately, the goal of the pre-scan volumes is to improve image quality by local optimization of pre-scan settings. This retrospective work focused on replicating human-placed volumes and did not evaluate how the model's placements impacted image quality. Further study of these models should include acquisition of prospective data with an evaluation of the image quality both using technologist and network derived pre-scan volumes. Finally, only one neural network, AlexNet, was evaluated. Study of additional models may lead to further improvement in volume prediction.

3.5 Conclusion

In summary, this chapter demonstrates the feasibility of a deep neural network to accurately replicate human placement of scan and pre-scan volumes in breast MRI. The resultant network-based placements closely agreed with human users despite the use of a relatively small training dataset. Further improvements are anticipated with inclusion of additional training data. These models show promise for quick and consistent placement of pre-scan volumes which may help to improve consistency in breast MR image quality.

Chapter Four

GBCA Attenuation Effect on PET Quantification in Breast MRI[3]

4.1 Introduction

The purpose of this chapter was to evaluate whether gadolinium-based contrast agents (GBCAs) significantly attenuate annihilation photons and bias quantitation of PET as discussed in **Section 1.3.1**. Significant attenuation of photons could result in biased PET quantification in the setting of contrast enhanced PET/MR. Based on results from phantom studies, it was hypothesized that SUV quantification of ^{18}F-fluorodeoxglucose (FDG) uptake using pre-contrast MR attenuation correction would not significantly change after injection of GBCA for patients undergoing simultaneous PET/MR.

4.2 Methods

To evaluate potential bias caused by GBCA attenuation in-vivo, clinical breast PET/MR exams were collected. The PET data from the exams were used to reconstruct images which simulated dynamic imaging scenarios. For each scenario, ROIs were drawn in multiple organs throughout the body and SUV_{mean} and SUV_{max} were calculated for each. Statistical testing was performed to determine any significant SUV changes following GBCA administration. Additional

[3]This chapter has been published in the journal Medical Physics: Allen TJ, Henze Bancroft LC, Kumar M, Bradshaw TJ, Strigel RM, McMillan AB, Fowler AM. Gadolinium-based contrast agent attenuation does not impact PET quantification in simultaneous dynamic contrast enhanced breast PET/MR. Med Phys. 2022 Aug;49(8):5206-5215. doi: 10.1002/mp.15781. Epub 2022 Jun 8. PMID: 35621727.

details are provided in the following sections. A calculation of the attenuation coefficients for a representative GBCA is presented in **Appendix C**.

4.2.1 Participant Information

This study was IRB-approved and HIPAA compliant. Women 18 years or older with biopsy-proven, invasive breast cancer measuring at least 1 cm by conventional imaging and receiving pre-operative breast MRI were eligible. Patients who were pregnant, lactating, had breast implants, had surgical excision or underwent neoadjuvant therapy for the biopsy-proven breast malignancy, or had a contraindication to MRI or PET were not eligible. The PET/MR data used in this analysis was collected as part of a previously published prospective study performed at our institution[74].

4.2.2 Imaging Acquisition and Reconstruction

Participants fasted for at least 6 hours before receiving a 370 MBq (10 mCi) injection of ^{18}F-fluorodeoxyglucose (FDG) intravenously approximately 85 minutes (range: 79 to 117 minutes) prior to imaging. Participants underwent a simultaneous PET/MR breast exam on a Signa 3.0T PET/MR system (GE Healthcare, Waukesha WI) using an 8-channel dedicated breast coil. The following software versions were used during the study: MP24.0_R01_1524, MP24.0_R02_1629, and MP24.0_R03_1712. PET acquisition duration was 30 minutes in the prone position with one bed position at the level of the breasts. Data was saved in list-mode format so that multiple PET reconstructions could be performed. Concurrent with the PET acquisition, participants received a standard-of-care breast MRI exam including a 2D fat-saturated T2-weighted fast spin-echo (FSE) sequence, a 3D non-fat saturated T1-weighted spoiled gradient recalled (SPGR) sequence, a dynamic contrast enhanced (DCE) series consisting of four

consecutive phases of a 3D fat-saturated T1w SPGR sequence, and diffusion weighted imaging. Pre-contrast, two-point Dixon-based in-phase and out-of-phase images were acquired and used for the vendor-supplied PET attenuation correction via the segmentation method (air, lung, fat, and soft tissue).[22] The fat image and water image from the Dixon sequence was used to generate the attenuation map (μ-map). Scan parameters are included in **Table 8**. A diagram showing PET and MR scan timing for a single subject is shown in **Figure 17**.

Between the first and second time-phase of the DCE series, 0.1 mmol/kg of a representative gadolinium-based MR contrast agent (gadobenate dimeglumine, MultiHance, Bracco Diagnostics, Monroe Township, NJ) was injected intravenously by a power injector at 2 mL/sec followed by a 10 mL saline flush. The concentration of the GBCA provided by the manufacturer was 0.5 M. The GBCA injection time was measured from the beginning of PET acquisition.

Table 8: MR and PET Sequence Parameters

MR Sequence	T2W	T1W Pre-Contrast	DCE	DWI	Dixon-Based MRAC
Plane	Axial	Axial	Axial	Axial	Axial
TR/TE (ms)	4000/85	5.4/2.6	5.4/2.6	4000/MinTE	4.0/1.1+2.2
FOV (mm)	320	320	320	360	500
Matrix Size	288/288	448/448	448/448	128/192	256/128
Slice Thickness (mm)	2	1.4	1.4	4.0	5.2
Bandwidth (kHz)	31.25	83.33	83.33	250	167

PET Sequence	Static	Dynamic
FOV (mm)	600	600
Matrix Size	256/256	256/256
Acquisition Time (min)	30	30
Frame Length (min)	30	0.5
Recon Method	OSEM	OSEM
Iterations	3	3
Subsets	28	28
Filter	4 mm Gaussian	4 mm Gaussian

MinTE: minimum allowed echo time, OSEM: Ordered subsets expectation maximum, MRAC: MR-based attenuation correction.

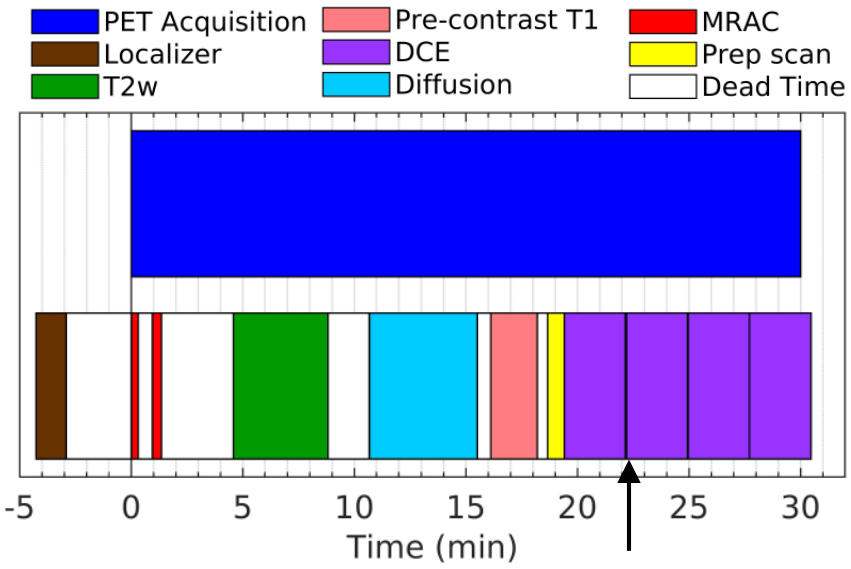

Figure 17: PET/MR Timing Diagram. Diagram showing the timing of PET and MR acquisitions for a single subject. PET data collection and MRAC began simultaneously before the administration of GBCA. GBCA administration occurred between the first and second phase of the DCE series (arrow). Because of this timing, the MRAC was not affected by the GBCA.

PET images were reconstructed from raw data using an ordered subsets expectation maximization (OSEM) algorithm provided by the vendor for offline use written in MATLAB (The Mathworks Inc., Natick, MA). PET acquisition and reconstruction parameters are included in **Table 8**. Corrections to account for random coincidences, dead time, and scatter were applied along with attenuation correction. The PET data were reconstructed in three distinct ways. 1) PET data were dynamically reconstructed into 60 time frames of 30 seconds, independent of the GBCA administration as would be the case in a likely clinical image scenario. This reconstruction method is referred to as dynamic method 1 (DM1). A 30 second frame duration was chosen as a compromise between short frames required to capture the estimated injection duration during which GBCA attenuation might be seen (seconds to tens of seconds) and the image noise. 2) A second reconstruction explicitly selected the 30 second frames so that the start of GBCA

administration occurred between two frames (DM2). 3) Finally, a static PET reconstruction incorporating the full 30 minutes of PET data was performed.

For DM1 and DM2, frames of the dynamic PET reconstruction were chosen to represent pre-contrast activity and post-contrast activity. Frames that started prior to GBCA injection are defined as pre-contrast frames. Frames that started following GBCA injection are defined as post-contrast frames. Three timing windows were chosen consisting of a single frame (30 seconds), two frames (60 seconds), or four frames (120 seconds) immediately before and after contrast administration (**Figure 18**). In the DM1 reconstruction, pre-contrast frames could include data acquired after GBCA administration since the pre-contrast frames were defined by the frame start time. The distribution of contrast times within the pre-contrast frame nearest GBCA administration is shown in **Figure 19**. Despite the inclusion of some contrast data in pre-contrast frames, the DM1 reconstruction was performed to replicate clinical timing scenarios where PET frames are sequential from the beginning of the acquisition period. The 30, 60, and 120 second timing windows are used to investigate the tradeoff between image noise and potential GBCA attenuation. At short frame duration, noise is high but the initial administration of concentrated GBCA is more likely to be captured. With longer frames, noise is reduced but the potential effect of GBCA is lost due to averaging over times when GBCA is much more dilute. Thus, multiple timing windows were used in an attempt to identify a frame duration where the potential GBCA effect could be detected through the image noise.

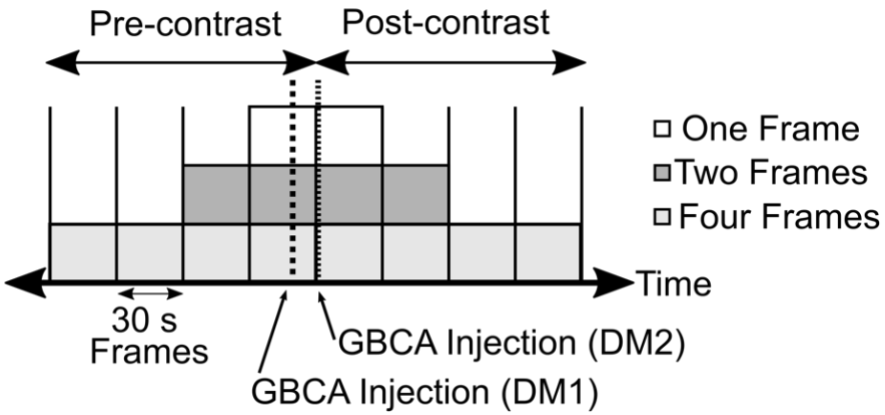

Figure 18: Diagram of 2 Distinct PET Reconstruction Paradigms. SUV_{mean} and SUV_{max} were calculated for both pre-contrast and post-contrast timing intervals. These intervals consisted of either one, two, or four 30 second frames. In this figure, solid vertical lines represent the segmentation of PET data into 30 second frames. Pre-contrast frames are those which began prior to the injection of GBCA. Post-contrast frames are those which began after GBCA injection. For the DM1 reconstruction, the administration of GBCA did not always occur at the same time within the nearest 30 s dynamic frame, e.g., the beginning of a frame (see **Figure 19**). For the DM2 reconstruction, frames were selected such that GBCA administration always occurred at the beginning of a frame.

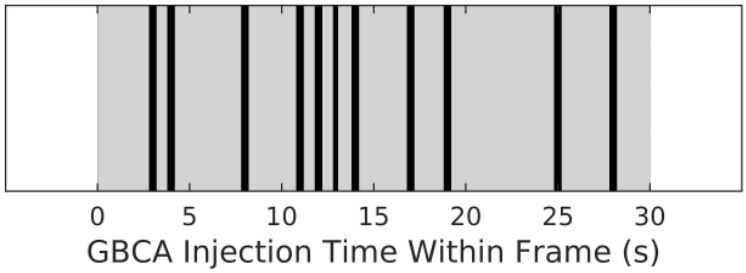

Figure 19: Distribution of Gadolinium Administration in DM1 Frames. For the DM1 reconstruction, GBCA administration timing within the last pre-contrast frame is shown. Each black bar indicates when the GBCA was administered in the pre-contrast frame. The injection times vary uniformly throughout the frame as would be the case in a clinical image scenario where administration timing is independent of the beginning of PET frames.

4.2.3 Imaging Analysis and Quantification

A board-certified, fellowship-trained radiologist with 8 years of experience identified four volumes of interest (VOIs) within each subject using the full 30-minute static PET reconstruction. The VOIs included 1) a contour of the known biopsy-proven breast malignancy using 30% thresholding of the maximum pixel intensity; 2) a 1 cm diameter sphere within the fibroglandular

tissue of the contralateral normal breast; 3) a 1 cm diameter sphere within the descending aorta at the lower thoracic level; and 4) a 3 cm diameter sphere within the right lobe of the liver (**Figure 20**). PET images were viewed alongside DCE MR images for anatomical reference. Drawing of VOIs was performed using MiradaXD3 v3.6 (Mirada Medical, Denver, CO).

Two metrics were used for quantification of FDG uptake in each of the four VOIs: SUV_{mean} and SUV_{max}. SUV_{mean} is the average standardized uptake value among all pixels in the VOI while the SUV_{max} is the uptake for the highest intensity pixel. For both dynamic reconstructions, the SUV_{mean} and SUV_{max} for pre-contrast and post-contrast periods were calculated using the three timing windows discussed earlier.

4.2.4 Statistical Analysis

Percent change in the SUV metrics was calculated to test for a statistically significant, non-zero percent change. Wilcoxon signed-rank tests were performed for each volume of interest and each timing interval used. Utilizing Bonferroni correction for multiple comparisons, p-values of 0.0125 or less were considered significant at the $\alpha<0.05$ level. Statistical analysis was performed within MATLAB (The Mathworks Inc., Natick, MA).

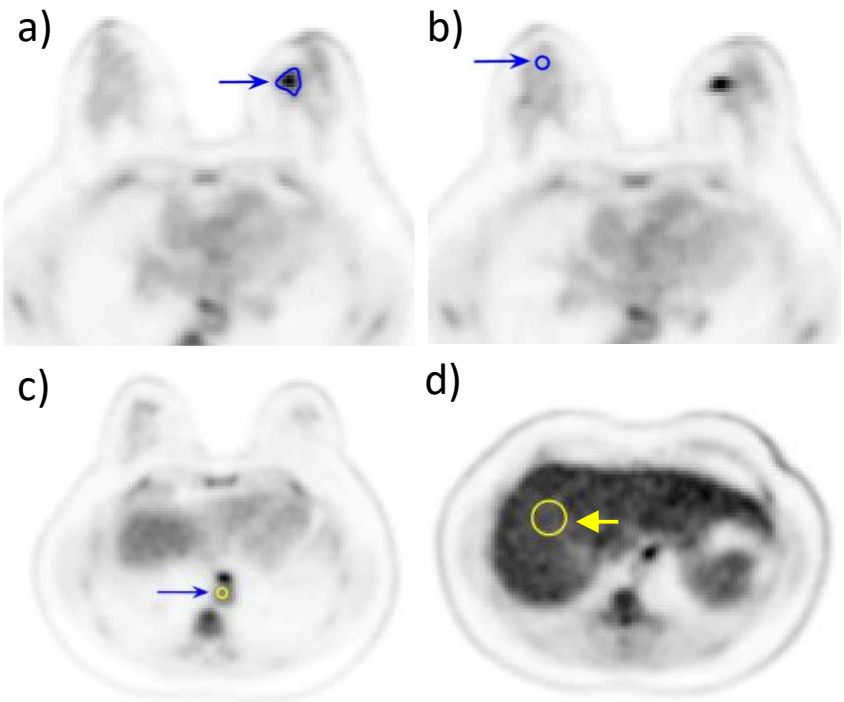

Figure 20: Example of Volumes of Interests for Simultaneous PET/MR. SUV_{mean} and SUV_{max} were calculated for four volumes of interest (VOIs) in each subject. 30-minute reconstructed axial PET images demonstrating: **a)** biopsy-proven left breast malignancy with VOI using 30% of the maximum pixel intensity threshold (arrow), **b)** one cm diameter VOI sphere within the fibroglandular tissue of the contralateral normal breast (arrow), **c)** one cm diameter VOI sphere within the descending aorta (arrow), **d)** 3 cm diameter VOI sphere in the right lobe of the liver (arrow). VOIs were drawn using DCE MRI as anatomical reference (not shown).

4.3 Results

Twenty-three women with biopsy-confirmed, invasive breast cancer were imaged in the initial prospective study[74]. For the current analysis, three subjects were excluded due to contrast injection occurring after the PET acquisition had ended. Seven subjects were excluded because contrast injection occurred within 120 seconds of the end of PET acquisition. The remaining 13 subjects included in this analysis were aged 33 to 70 years (mean, 49 years). Malignant lesion sizes varied from 1.7 cm to 8.8 cm (mean, 4.5 cm) as measured by the longest dimension using MRI. One subject had two separate, biopsy-confirmed sites of invasive carcinoma within the same breast. Only the larger of the two lesions was included in this analysis. The injected volumes of

0.5 M GBCA ranged from 10 mL to 18 mL for a standardized 0.1 mmol/kg patient weight-based dosing.

4.3.1 DM1 Reconstruction

Images from a representative subject along with the corresponding SUV_{mean} and SUV_{max} time activity curves for the DM1 reconstruction are provided in **Figure 21**. Using the 30 second timing interval, no significant change in SUV_{mean} was measured for the biopsy-proven breast malignancies (p=0.07), contralateral normal fibroglandular tissue (p=0.41), descending aorta (p=1.0), or liver (p=0.64) (**Figure 22**). In the 60 second analysis, no significant change was detected in any tissue type tested (breast malignancies, p=0.03; fibroglandular tissue, p=0.95; aorta, p=0.74; liver, p=0.19). Finally, in the 120 second time interval, no significant change was detected in the tissues analyzed (breast malignancies, p=0.31; fibroglandular tissue, p=0.50; aorta, p=0.95; liver, p=0.17).

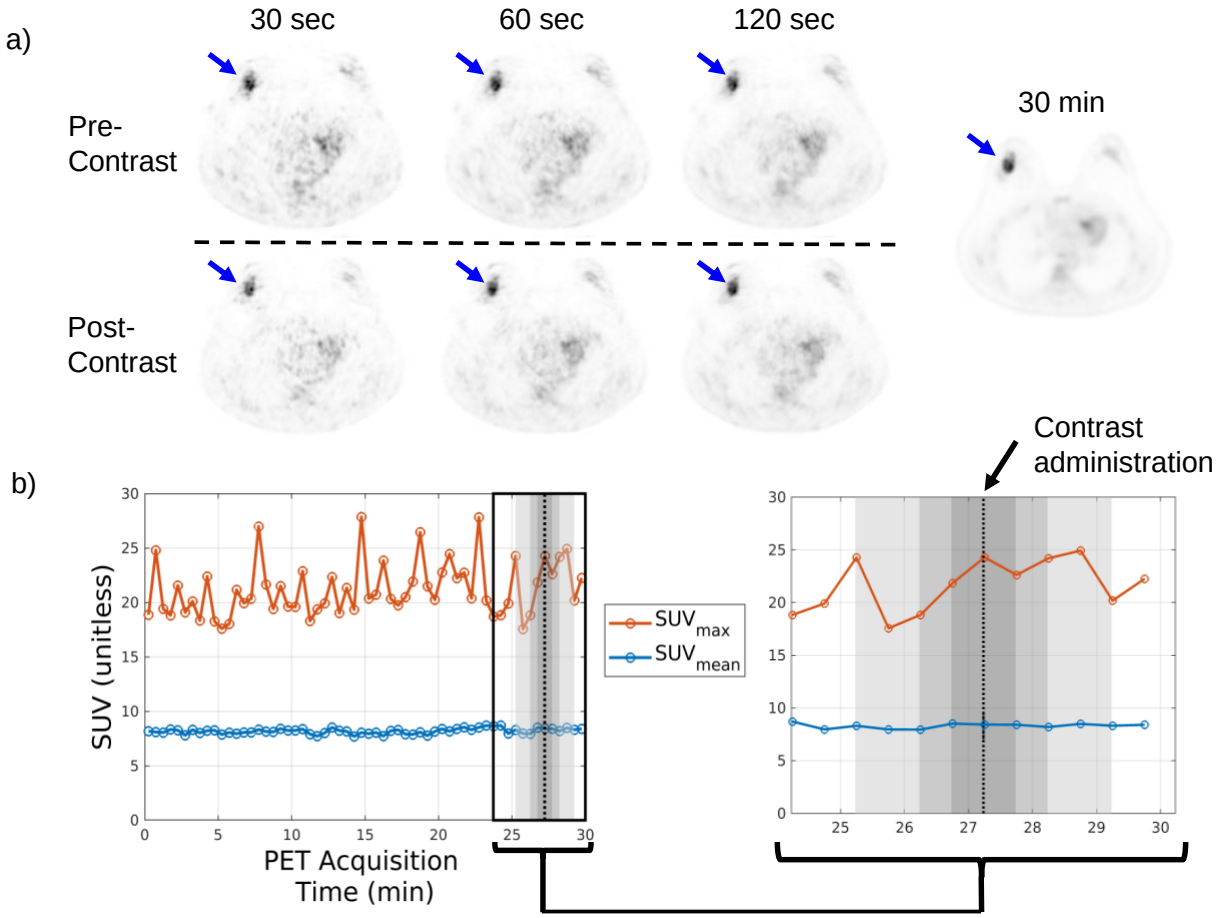

Figure 21: Example PET Images Using Multiple Timing Intervals and Time Activity Curves. **a)** Axial breast PET images in a representative patient with known right breast cancer (increased FDG uptake indicated by arrow). Pre- and post-contrast SUV PET images are shown for each of the tested timing interval frames (30 s, 60 s, and 120 s). The static PET reconstruction using the entire 30 minutes of PET data is included for reference. **b)** Corresponding SUV_{mean} and SUV_{max} time activity curves for this subject are included. The dashed vertical line indicates the time of gadolinium-based contrast agent administration. The shaded portion of the plot corresponds to the tested timing interval frames (dark gray: 30 s, medium gray: 60 s, light gray: 120 s). The inset plot provides a more detailed view of the data immediately before and after contrast administration.

Similarly, no change in SUV_{max} was observed for any of the investigated tissues in the 30 second interval (breast malignancies, p=0.64; fibroglandular tissue, p=0.34; aorta, p=0.34; liver, p=0.03) (**Figure 23**) in the 60 second interval (breast malignancies, p=0.59; fibroglandular tissue, p=0.84; aorta, p=0.27; liver, p=0.03), or the 120 second interval (breast cancer lesions, p=0.74; fibroglandular tissue, p=0.59; aorta, p=0.59; liver, p=0.17).

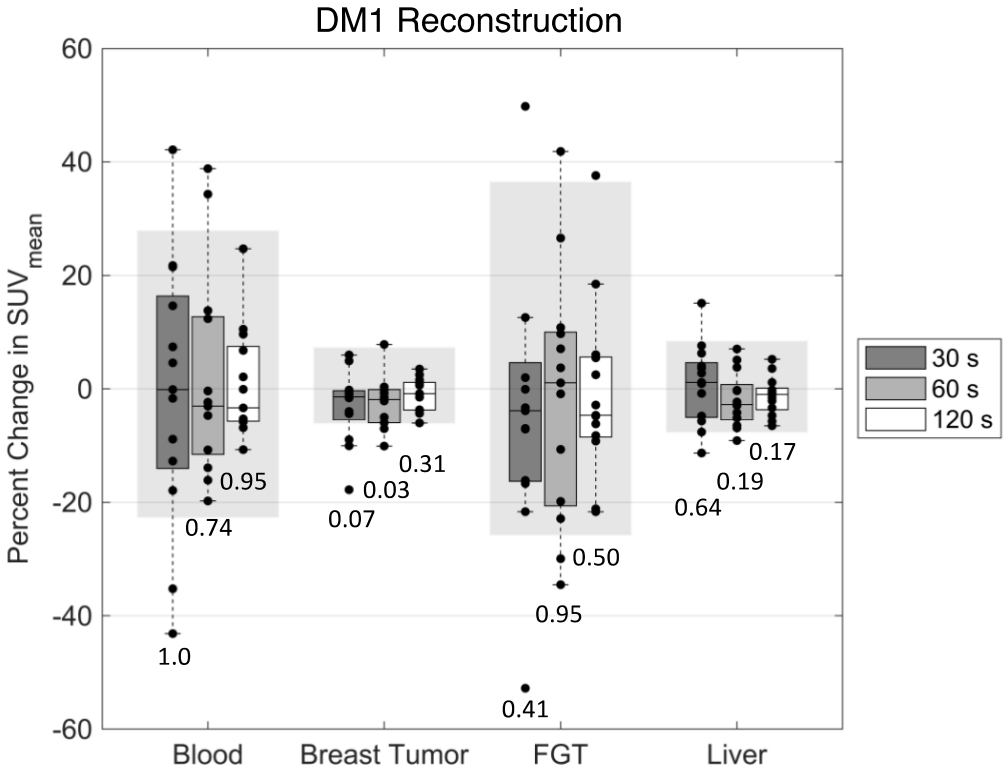

Figure 22: SUV_{mean} Before and After GBCA Administration Using DM1 Reconstruction. Percent change in SUV_{mean} for each of the measured VOIs: aorta/blood, breast tumor, normal fibroglandular tissue (FGT), and liver for each timing interval using DM1 reconstruction. The shaded regions show the frame-to-frame variation of SUV_{mean} for each tissue as the 95% confidence interval. Our data mostly falls in the region expected for frame-to-frame variation indicating no effect of GBCAs. The value below each box and whisker is the p-value corresponding to a test of difference from zero change. One extreme outlier is not shown for the sake of readability (Blood, 30 s, Change in SUV_{mean} = 89%).

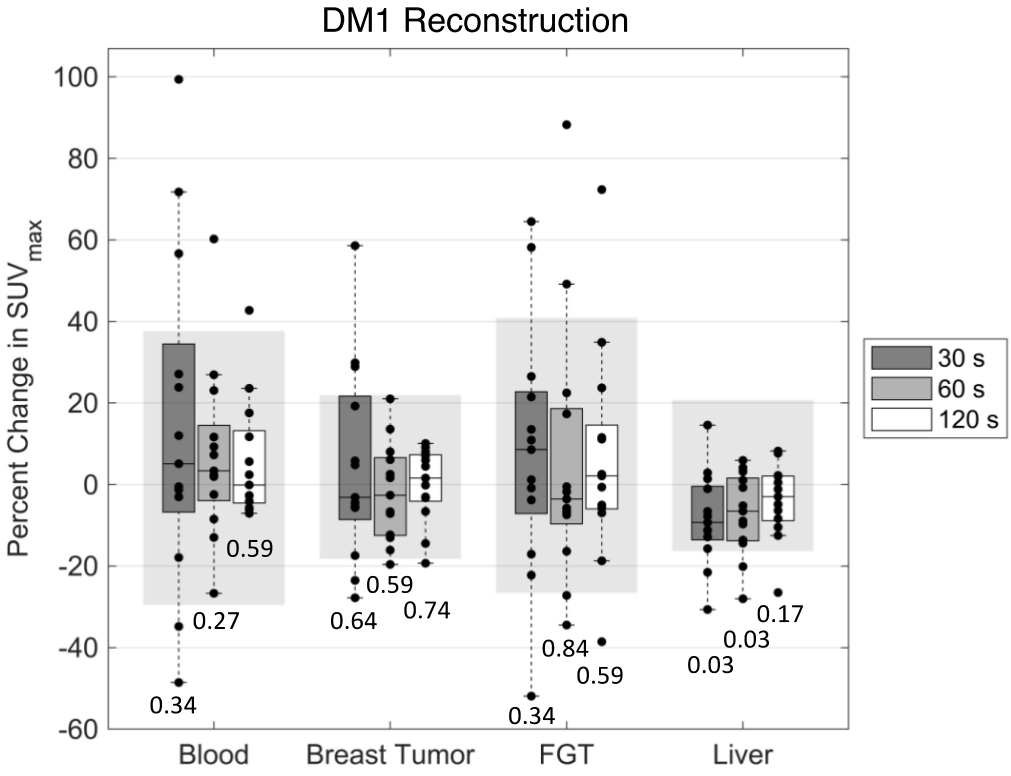

Figure 23: SUV_{max} Before and After GBCA Administration Using DM1 Reconstruction. Percent change in SUV_{max} for each of the measured VOIs: aorta/blood, breast tumor, normal fibroglandular tissue (FGT), and liver for each timing interval using DM1 reconstruction. The shaded regions show the frame-to-frame variation of SUV_{max} for each tissue as the 95% confidence interval. Our data mostly falls in the region expected for frame-to-frame variation indicating no effect of GBCAs. The value below each box and whisker is the p-value corresponding to a test of difference from zero change.

Since PET annihilation photons from left breast malignancies may be more likely to exhibit attenuation from GBCA due to the proximity to the heart, a subset analysis was performed comparing pre-contrast and post-contrast SUVs for the 8 subjects with left breast cancer. Analysis of this subset of subjects showed no significant difference in either SUV_{mean} (30 seconds, p=0.74; 60 seconds, p=0.38; 120 seconds, p=0.95) or SUV_{max} (30 seconds, p=0.25; 60 seconds, p=0.38; 120 seconds, p=0.05) in the tumor VOIs for left breast malignancies.

4.3.2 DM2 Reconstruction

Similar to the results obtaining using the DM1 reconstruction, no significant change in SUV_{mean} was found for known cancers (30 seconds, p=0.17; 60 seconds, p=0.013; 120 seconds, p=0.17), normal fibroglandular tissue (30 seconds, p=0.59; 60 seconds, p=0.84; 120 seconds, p=0.34), aorta (30 seconds, p=0.59; 60 seconds, p=0.84; 120 seconds, p=0.95), or liver (30 seconds, p=0.74; 60 seconds, p=0.54; 120 seconds, p=0.34) (**Figure D1**, **Appendix D**). No significant change in SUV_{max} was observed for known cancers (30 seconds, p=0.03; 60 seconds, p=0.17; 120 seconds, p=1.0), normal fibroglandular tissue (30 seconds, p=0.84; 60 seconds, p=0.74; 120 seconds, p=0.54), aorta (30 seconds, p=0.19; 60 seconds, p=0.45; 120 seconds, p=0.24), or liver (30 seconds, p=0.79; 60 seconds, p=0.50; 120 seconds, p=0.84) (**Figure D2**, **Appendix D**). Sub-analysis of cancers of the left breast found no significant change in either SUV_{mean} (30 seconds, p=0.46; 60 seconds, p=0.15; 120 seconds, p=0.55), or SUV_{max} (30 seconds, p=0.15; 60 seconds, p=0.95; 120 seconds, p=0.11).

4.4 Discussion

This work aimed to determine if the use of GBCAs during simultaneous PET/MR imaging could directly impact PET attenuation and thus bias quantification of radiotracer uptake. Using simultaneous breast PET/MR with FDG for patients with newly diagnosed invasive breast cancer, SUV metrics were compared in dynamic PET reconstructed data from timing window frames before and after GBCA administration using pre-contrast MR attenuation correction. Two dynamic reconstruction methods were investigated: DM1 with frames chosen independent of GBCA administration and DM2 with frames selected such that GBCA administration fell exactly between two frames. We observed no detectable change in SUV_{mean} or SUV_{max} in the biopsy-proven breast malignancies or in normal tissues (aorta/blood, fibroglandular tissue, and liver) using either

reconstruction method. The lack of effect of GBCA administration on PET quantification is likely due to the small total mass of gadolinium contained in a standard dose of contrast and the rapid dilution of the agent after intravenous injection. This study addresses gaps in the existing literature by providing a direct comparison of quantitative PET values immediately before and after the injection of GBCA in a cohort of 13 patient subjects using a different scanner vendor than previous work with phantom models[59,60]. Since the data acquisition for the MR attention correction was performed prior to GBCA administration, the potential confounding effect of gadolinium on the μ-map accuracy was removed.

This study focused on patients undergoing simultaneous breast PET/MR imaging since it offers a unique opportunity to study the attenuation effects of GBCAs on PET data quantification and is expected to be one of the clinical scenarios potentially most impacted. First, GBCA injection is required for the dynamic contrast enhanced portion of the MRI exam. Second, the field of view used in breast imaging includes the heart. As GBCA first passes through the heart it will be near its highest concentration and provide the greatest attenuation. Third, accurate quantification of SUV in breast is clinically important. FDG PET can be used to assist in initial staging of cancer and for detection of distant metastases[74,75]. SUV_{max} in particular has been of interest for early prediction of neoadjuvant chemotherapy response[76–79]. Early identification of non-responders would allow treatment to pivot away from ineffectual chemotherapy and toward more effective treatment options. Use of SUV_{max} for clinical decision making relies on accurate quantification of the tracer uptake and attenuation due to gadolinium presents a possible source of inaccuracy, thus resulting in impaired clinical decision making.

The results of this multi-patient study expand the existing knowledge regarding the potential impact of GBCAs on radiotracer uptake for simultaneous PET/MR imaging. Using

phantom models, Lois et al. demonstrated that the GBCA, gadobutrol (Gadovist, Bayer HealthCare, Whippany, NJ) had similar PET attenuation values as water under clinically relevant concentrations (0.2%) and had no effect on the attenuation of PET emission data[60]. Furthermore, data from one patient with bronchial carcinoma imaged with simultaneous FDG PET/MR before and after GBCA administration demonstrated that the MR-based attenuation maps were not significantly affected by the presence of GBCA (<0.5% and <3% for the two correction algorithms tested). However, a direct comparison of attenuation corrected PET values before and after GBCA administration could not be performed due to different study conditions. By using dynamic PET reconstructed data from timing windows before and after GBCA administration, our study provides this direct comparison. In another phantom-based study using simultaneous PET/MR imaging with 18F-sodium fluoride and the GBCA gadoteric acid (Dotarem, Guerbet, Villepinte, France), O'Doherty and Schleyer demonstrated that the presence of high GBCA concentrations produced minimal effect on image-based PET activity values[59].

In agreement with the prior phantom-based studies, our results show that no significant PET quantification bias was present in patients undergoing simultaneous breast PET/MR imaging despite the proximity to the heart. We therefore expect no significant SUV quantification errors in other anatomy, such as the brain, where no large pool of gadolinium exists within the imaging volume at any time during PET acquisition. Although certain classes of GBCAs have been associated with gadolinium deposits within brain structures such as the dentate nucleus[80,81], the higher end of deposited gadolinium concentrations is orders of magnitude less than the concentration of a typical GBCA (2.1 µg/g of tissue versus 340 mg/g of GBCA)[81]. Since our protocol was performed at a single thoracic bed position and did not include whole-body imaging, the effect of GBCAs in other anatomic regions outside of the field-of-view could not be evaluated.

We used timing windows to represent dynamic frames as short as 30 seconds. Using smaller windows, we might expect a larger effect due to a higher average gadolinium concentration over the course of the dynamic frame. However, such short frames would suffer from low signal-to-noise, and any prompted change in PET signal would likely be undetectable due to high noise. Future PET systems with higher sensitivity might facilitate the detection of gadolinium effects at shorter time frames.

This study focused on SUV (both mean and max) as quantitative PET metrics, though the results should be generally applicable to any PET metric. SUV is directly proportional to the detected signal in a voxel. Since SUV is unaffected, the signal itself should be unaffected and other quantification metrics, such as radiotracer pharmacokinetic parameters, will be similarly unbiased. However, effect of GBCAs on PET pharmacokinetic parameters was not directly tested in this study.

Only one GBCA (gadobenate dimeglumine) was evaluated here. Other FDA-approved GBCAs in clinical use have a similar elemental composition, density, and concentration as gadobenate dimeglumine. Gadobutrol (Gadovist) is the only FDA-approved GBCA that is supplied at a higher concentration (1.0 M) allowing for lower volume injections to achieve the same recommended dosing (0.1 mmol/kg body weight). All of the FDA-approved GBCAs contain the same amount of gadolinium per recommended dose. Thus, we assert that the tested GBCA (gadobenate dimeglumine) is representative of all GBCAs and would anticipate similar results for other GBCAs.

This study investigated one possible way that GBCA can affect PET quantification, namely additional attenuation caused by gadolinium atoms not accounted for by current MRAC methods. However, the GBCA also affects T1-weighted tissue contrast which could impact the accuracy of

image-based MRAC approaches and thus indirectly affect SUV quantification. Though not addressed in this work, it has been investigated by others and can be readily addressed by routinely acquiring MRAC images prior to GBCA administration[60,61]. The results of these studies varied from no effect to potential tissue-specific effects of GBCA on MRAC methods for brain, lung, and fat.

Historically, a similar question was posed with contrast-enhanced PET/CT exams if iodinated contrast agents used in CT affect SUV values by either attenuating annihilation photons or by impacting the image data used for attenuation correction. There does not appear to be significant attenuation of annihilation photons by iodine contrast agents at 511 keV in general use cases[82]. However, changes in Hounsfield units due to iodinated contrast in PET/CT leads to an overestimation of attenuation by the CT attenuation correction leading to elevated SUV[82–84].

4.5 Conclusion

The results of this study demonstrated that SUV_{mean} and SUV_{max} were not affected by the injection of a GBCA during simultaneous breast PET/MR imaging. The attenuation of annihilation photons caused by the presence of the GBCA is likely not substantial enough to cause a significant difference in detected signal. Thus, GBCAs may be used during a simultaneous PET/MR acquisition without imparting bias onto PET quantification metrics and an additional correction to the PET signal to account for GBCA attenuation is not necessary when pre-contrast MR-based attenuation correction sequences are obtained.

Part II

Advances in Abdominal Diffusion Weighted Imaging and Final Thoughts

Chapter Five

ADC Reproducibility Using M1-optimized DWI

5.1 Introduction

5.1.1 Diffusion Weighted Imaging

Diffusion weighted imaging (DWI) enables noninvasive probing of tissue microstructure with multiple clinical and research applications[85,86]. In the liver, DWI is a clinically established tool for detection of focal liver lesions and particularly for liver metastases[87,88]. Further, quantitative mapping of liver diffusion metrics may enable improved diagnosis, staging, and treatment monitoring with applications in focal and diffuse disease. The apparent diffusion coefficient (ADC) parameter has shown promise for lesion characterization[89], assessment of response to treatment[90], and in evaluation of diffuse liver disease such as fibrosis[91]. Quantification using more advanced signal representations such as intra-voxel incoherent motion[92], kurtosis[93], or recent tissue models[94,95] may provide enhanced sensitivity and specificity.

For quantitative diffusion measurements to be broadly useful, they must be reproducible. However, multiple studies have demonstrated the lack of reproducibility in liver ADC and IVIM modeling[96–98]. A contributing factor to the poor reproducibility is the sensitivity of DWI signals to multiple sources of undesirable variability. The diffusion gradient waveforms necessary to apply the diffusion weighting extend TEs and causes low SNR, particularly in short-T2 organs such as the liver[99]. Stejskal-Tanner ("MONO") diffusion waveforms are commonly used but are prone to motion-induced artifacts[100]. Cardiac motion compresses the left lobe of the liver and causes signal

dropout[101]. Imaging with high b-values is more sensitive to cardiac-induced signal dropout causing ADC values to be artifactually increased in the left liver lobe.

5.1.2 Motion-Robust Liver DWI

Attempts to desensitize ADC measurements to motion have included both prospective[97,102–106] and retrospective methods[107–111]. Retrospective mitigation of motion-induced signal dropout can be performed through averaging across repetitions. In body DWI, multiple repetitions for each b-value and diffusion direction are generally acquired to increase SNR. The averaging of these repetitions helps reduce the effect of signal dropout from any individual repetition. While a simple averaging technique is generally used, weighted averaging methods may further reduce signal dropout by giving less weight to repetitions with severe dropouts[107,110,111]. Generally, retrospective correction may fail in the absence of sufficient repetitions without signal dropouts.

Prospective motion correction methods include respiratory gating, cardiac triggering, and/or the use of motion compensated diffusion weighting gradients[102–105]. Cardiac gating is not often used for liver imaging since it severely lengthens overall scan times. Therefore, suppression of cardiac motion is more commonly performed using motion compensated gradients. These diffusion gradient waveforms often use first moment (M1) nulling, which limits DWI's sensitivity to compressive tissue motion at the cost of longer TEs and lower SNR. An additional downside of M1-nulled waveforms is the potential for hyperintense vasculature due to compensation of blood signal. Vessels appear bright in M1-nulled DWI, and thus may mimic liver lesions and confound the interpretation of DWI. To suppress this hyperintense vessel signal, an alternative technique known as M1-optimized DWI (MODI) has been proposed[106]. This technique uses waveforms with a small, but non-zero M1-value. MODI shows promise in achieving motion-robust DWI while

suppressing hyperintense vessel signals, but still has the downside of longer TEs and reduced SNR compared to MONO diffusion encoding.

The minimum TE achievable with MODI depends on the gradient coil performance including the maximum gradient strength, G_{max}, and maximum slew rates, S_{max}. Currently, MODI has only been demonstrated on an MR system with high performance gradients (G_{max} = 70 mT/m, S_{max} = 150 mT/m/msec). Direct implementation of MODI in MR systems with conventional gradient performance results in extended TEs and prohibitively low SNR especially in short-T2 organs such as the liver.

Therefore, there is an unmet need to overcome the limitations of existing motion-robust methods for liver DWI. Previous implementations of MODI used only a single gradient coil to apply the diffusion weighting. Use of multiple gradient coils simultaneously, such as in tetrahedral diffusion encoding[102,112] may allow for shortened TEs and enable motion-robust DWI using MODI on MR systems with conventional gradient performance. Further, a combination of prospective (MODI) and retrospective (weighted averaging) methods for motion-robust liver DWI may lead to additional improvements in motion robustness and reproducibility[111,113].

Therefore, the purpose of this work is to introduce technical advances to enable motion-robust DWI and diffusion quantification on MR systems with conventional gradient performance, and to evaluate the reproducibility of liver ADC mapping across multiple different MR systems.

5.2 Innovations in Abdominal DWI

This is the first work to demonstrate the MODI technique on a 1.5T scanner and scanners with conventional gradient performance. This is achieved by implementing, for the first time, tetrahedral waveform design with MODI. Additionally, this work provides the first assessment of MODI-based ADC measurement reproducibility across various MR scanners.

5.3 Theory

The original MODI algorithm for waveform generation is based on optimized diffusion gradient design (ODGD)[105,106]. As a brief overview, MODI generates a diffusion-weighting gradient waveform, G(t), that achieves a desired b-value with minimum TE subject to timing, hardware, and design constraints. MODI with tetrahedral diffusion encoding (tMODI) allows for an increase in the effective gradient strength, G_{eff}, by √3. Since the b-value and M1-values are proportional to G_{eff}^2 and G_{eff}, respectively, shorter TEs can be achieved with a tetrahedral design for a given b-value and M1-value enabling higher SNR images.

A potential challenge with tMODI is the increase in concomitant gradient (CG) effects, especially at lower field strengths[114,115]. Existing methods to suppress CG effects in diffusion waveform design have been previously described[105,116] and one of these techniques has been used with MODI previously[105]. This CG-suppression technique equates CG-induced phase differences before and after the 180-degree RF pulse in a spin echo DWI sequence by enforcing,

$$\int_{A1} G(t)^2 dt = \int_{A2} G(t)^2 dt, \quad \textbf{[1]}$$

where A1 is the time interval prior to the 180 RF pulse and A2 is the time interval following the 180 RF pulse. It can be shown that this technique also suppresses CG effects when using tetrahedral gradient design. The net amplitude due to CG for x- and y- gradient coils at 90° to each other can be modeled using a 2nd order Taylor expansion[115,117]:

$$B_C = \frac{\left(G_x^2 + G_y^2\right)^2 z^2}{2B_0} + \frac{G_z^2(x^2 + y^2)}{2B_0} - \frac{G_x G_z xz}{2B_0} - \frac{G_y G_z yz}{2B_0}, \quad \textbf{[2]}$$

and the spatially varying phase difference due to this magnetic field is:

$$\phi(x,y,z) = \gamma \int_{A1} B_C(x,y,z,t)dt \; - \int_{A2} B_C(x,y,z,t)dt. \quad \textbf{[3]}$$

If the gradient coils play waveforms with the same shape (i.e., scaled versions or identical versions of the same waveform), the cross terms in equation

$$B_C = \frac{\left(G_x^2 + G_y^2\right)^2 z^2}{2B_0} + \frac{G_z^2(x^2+y^2)}{2B_0} - \frac{G_x G_z xz}{2B_0} - \frac{G_y G_z yz}{2B_0}, \quad \textbf{[2]}$$

reduce to self-squared terms. For tetrahedral encoding, each coil plays an identical waveform so satisfying equation [1] ensures that the tMODI design is CG-nulled.

An additional limitation of the prior MODI algorithm was the occasional convergence to suboptimal local minima, leading to longer than necessary TEs. In this work, the MODI optimization algorithm was updated to include multiple initializations for each optimization step. Using these multiple initializations helps avoid suboptimal local solutions. The complete updated waveform design pipeline used in this chapter is shown in **Figure 24** with unique contributions of this work highlighted.

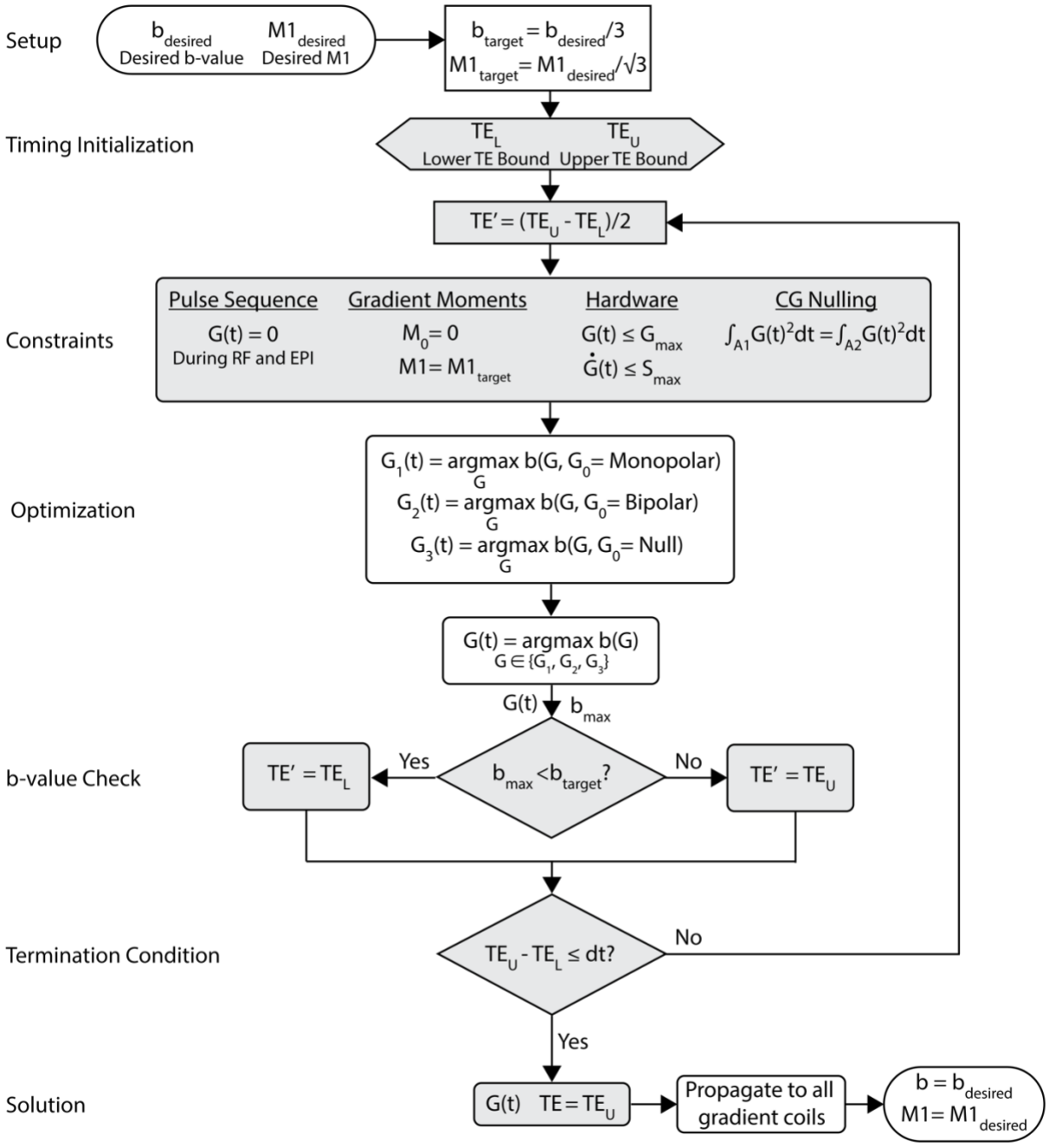

Figure 24: Overview of MODI Optimization and Novel Contributions. Modification to the existing MODI optimization algorithm allows for waveforms with a tetrahedral design and helps to prevent the optimization algorithm from converging to local minima. Shown here is a flowchart illustrating the steps of the optimization. White blocks are steps that are novel to this work of this chapter. The new blocks in the setup and solution categories allow for tetrahedral design while the blocks in optimization help to avoid local minima. The parameter dt is the temporal resolution of the waveform.

5.4 Methods

A brief overview of the methods is presented here, and further details follow in subsequent paragraphs. This chapter implemented updates to the MODI optimization to accommodate tetrahedral gradients and help avoid local minima. Simulations of TEs achievable for given b-values and M1 values were used to demonstrate the TE savings achievable with the MODI updates. Phantom imaging was performed to validate CG-nulling. A small reproducibility study with 8 subjects was performed using 3 different MR scanners. Reproducibility of liver ADC measurements was assessed for both MONO and MODI imaging which were used in combination with one of three weighted averaging methods.

5.4.1 MR Systems

Throughout this chapter, three different clinical MR scanners from the same manufacturer (GE Healthcare, Waukesha, WI, USA) were used **(Table 9)**. Scanner 1 was a 3T SIGNA Premier system with high performance gradient coils (G_{max}= 70 mT/m, S_{max} = 150 mT/m/msec) which had been used in a previous study to demonstrate the feasibility of MODI. Scanners 2 and 3 were a 3T Signa PET/MR system and a 1.5T SIGNA Artist system, respectively. Both have “conventional” performance gradient coils (G_{max}= 33 mT/m and S_{max} = 120 mT/m/msec). The feasibility of MODI on these systems has not been previously demonstrated.

Table 9: Scanner and Sequence Details

Scanner	Model	Field Strength	Max Slew Rate (mT/m/msec)	Max Gradient Strength (mT/m)
Scanner 1	Signa Premier	3.0 T	150	70
Scanner 2	Signa PET/MR	3.0 T	120	33
Scanner 3	Signa Artist	1.5 T	120	33

Sequence Parameters

Sequence	Diffusion Directions	Scanner	|M1| (s/mm) low b, high b	TE (ms)
MONO	3	Scanner 1	0.22, 0.68	47
		Scanner 2	0.26, 0.85	64
		Scanner 3	0.26, 0.85	62
tMODI	4	Scanner 1*	0.10	66
		Scanner 2		72
		Scanner 3		70

Common Parameters

b-values (s/mm^2)	Repetitions per b-value	Acceleration	Slice Thickness	FOV (cm)
50,500	2,4	ASSET (R=2)	6 mm	38 x 38

tMODI waveforms include concomitant gradient nulling. *The gradient hardware on Scanner 1 prevents the use of tetrahedral gradients. On that scanner, tMODI experiments were performed with conventional MODI.

5.4.2 Echo Time Simulations

Simulations were performed to evaluate the differences in TE between the existing single coil MODI design and the proposed tMODI design. Diffusion gradient waveforms were generated for Scanner 3 with b-values between 25 and 800 s/mm^2 and M1-values with magnitude between 0 and 1 s/mm. The minimum echo time achievable for each waveform was determined. MODI/tMODI waveforms with controlled M1 values were generated as described above. Note that MODI is generally not used to generate waveforms with M1 values larger than those of MONO. When generating MODI/tMODI waveforms with M1-value near to or greater than the

corresponding MONO M1-value, a MONO solution with the desired b and M1-values was directly constructed. This was achieved by iteratively stretching and scaling the MONO solution for that b-value until the desired M1-value was achieved. The maximum allowed TE was 100 ms at which point the iteration was abandoned.

5.4.3 Experimental Validation of CG-nulling

A phantom study was conducted to assess CG-induced signal bias using tetrahedral waveforms both with and without CG-nulling. A silicone oil phantom was imaged on Scanner 3 using a 2D spin-echo diffusion sequence. CG effects are inversely proportional to the static field strength and so the 1.5T scanner provided a worst-case assessment of CG effects. The Gaussian diffusion of silicone oil is very slow so any substantial signal loss in the phantom was due to CG effects[118]. The phantom was imaged using tMODI waveforms both with and without CG nulling. Details of the acquisition parameters can be found in **Table 9**. The percent difference in high b-value signal compared to low b-value signal was calculated within a large ROI inside of the phantom.

5.4.4 Averaging Techniques

As discussed in section 5.1.2, individual repetitions are combined into a single image using simple arithmetic averaging in conventional DWI. In this chapter, three different repetition averaging techniques were employed and tested with liver imaging:

1. “Simple Averaging” - Arithmetic averaging of repetitions,
2. “Weighted Averaging” - Weighted averaging using the normalized Gaussian filtered image as weighting coefficients,
3. “Squared Weighted Averaging” - Weighted averaging using the square of the Gaussian filtered image as the weighting coefficients.

This can be described mathematically if we let S_i be the signal intensity for repetition i and $S_{i,G}$ the signal intensity after Gaussian blurring:

$$\bar{S} = \frac{\sum_i S_i \, S_{i,G}^n}{\sum_i S_{i,G}}, \qquad [4]$$

where simple averaging is n = 0, weighted averaging is n = 1, and squared weighted averaging is n = 2. The goal of this evaluation was to compare the robustness to motion-induced localized signal dropouts achievable with each of these averaging methods. Further details of the weighted and squared weighted averaging techniques has been provided by Führes et al[111]. For weighted averaging, this work uses a Gaussian kernel size of 32 mm.

5.4.5 Healthy Subject ADC Reproducibility

With IRB approval and informed consent, healthy volunteers (N=8) were imaged on the recruited for this study. Each volunteer was imaged on the three MR scanners consecutively on the same day **(Table 9).** On each scanner, multiple DWI sequences were performed using a 2D spin-echo diffusion sequence with both MONO and tMODI waveforms. tMODI waveforms were CG-nulled. Scanner 1 was not able to use tMODI since that system is designed for isotropic gradient coils (i.e., combinations of gradient coils are limited to the strength of a single gradient coil). Therefore, conventional MODI was used in place of tMODI on Scanner 1. For the in-vivo study, MODI/tMODI implementations on the various scanners will be referred to collectively as tMODI for simplicity. B-values of 50 and 500 s/mm^2 were used and a M1 value of 0.1 s/mm. Acquisitions were performed using respiratory triggering controlled by respiratory bellows. The number of respiratory cycles was fixed at 2 for consistency across scanners, which limited the maximum number of acquired slices. To achieve whole liver coverage within 2 respiratory cycles,

a slice thickness of 6 mm was used and spacing in-between slices was increased as necessary. **Table 9** provides detailed acquisition parameters.

The three different repetition averaging methods were applied to each acquisition resulting in three sets of reconstructed DW images and ADC maps. Subsequently, 9 ROIs were drawn corresponding to the 9 Couinaud liver segments by a radiologist using the b = 50 s/mm^2 images with simple averaging as an anatomic guide. ROIs were placed on single slices and constructed to be as large as possible within the segment of interest while avoiding large vessels and unsuppressed fat artifacts. This was done for each sequence so no registration between sequences was performed. Mean ADC values were calculated for each ROI, acquisition sequence, and averaging method across all healthy volunteers. Because only healthy normal volunteers were used, ADC measurements are expected to be consistent across all liver lobes. Similar to previous works, higher ADC measurements in the left lobe were assumed to be due to cardiac motion and not underlying subject disease. For this reason, in this work, we refer to these higher ADC values as overestimated despite the lack of an ADC reference value.

5.4.6 In-Vivo Data Analysis

A linear mixed-effects model was used to estimate systematic differences in ADC measurements between segments in the left and right liver lobes, between high and low gradient performance, and between 1.5T and 3.0T field strengths, while controlling for random subject effects. Statistics were performed in MATLAB R2022b (MathWorks, Natick, MA, USA) using the fitlme function. This modeling was repeated for each combination of acquisition sequence and averaging technique. ADC measurements from segment 1 were excluded from mixed-effects analysis since it is not considered a part of the right or left liver lobe.

The reproducibility of ROI-based ADC measurements was assessed between each pair of scanners (3 pairs) and across all 3 scanners collectively. Between pairs of scanners, i and j, the difference between each individual ROI measurement was determined, and the standard deviation of the differences, $\sigma_{ROI,i,j}$, was calculated. The ROI-based reproducibility coefficient, $RDC_{i,j}$, was defined as $1.96\sigma_{ROI,i,j}$. It follows that 95% of all potential measurements between scanner i and j are estimated to have an absolute difference of less than $RDC_{i,j}$. The average difference in ADC measurements between scanners i and j, $\Delta_{i,j}$, was also measured. An RDC value comparing measurements among all scanners was also calculated. The standard deviation of all measured differences between the 3 scanners, σ_C, was used to determine the collective RDC, $RDC_C = 1.96\sigma_C$. The RDC_C is the estimated absolute difference between 95% of all potential measurements between any two scanners. RDC calculations were performed for each acquisition sequence and averaging technique.

5.5 Results

5.5.1 Echo Time Simulations

Figure 25 shows the achievable TEs from MODI and tMODI using scanner 3 (conventional gradient performance) over a wide range of b-values and M1 values. For relevant b-values and M1 values, tMODI leads to shorter TEs. For example, motion-robust imaging of the liver has been previously demonstrated (on a system with high-performance gradients) using a high b-value of 500 s/mm2 and M1-value of 0.1 s/mm. With these parameters, the achievable TE using the previously proposed MODI on a system with conventional gradients was 102 ms. With tetrahedral gradients, TE was reduced to 75.5 ms, and with the updated optimization TE was further reduced to 69 ms. Therefore, the total reduction in TE for this parameter set was 33 ms or 32%.

5.5.2 Experimental Validation of CG-nulling

Figure 26 shows the effect of CGs in the silicone oil phantom. Using tMODI without CG-nulling, negative signal bias is observed which increases with the distance from imaging isocenter. This signal bias was up to -8% at a distance of 10 cm from the isocenter. When CG-nulling was used, the signal bias was reduced (e.g., from -8% to -3% at 10 cm from isocenter).

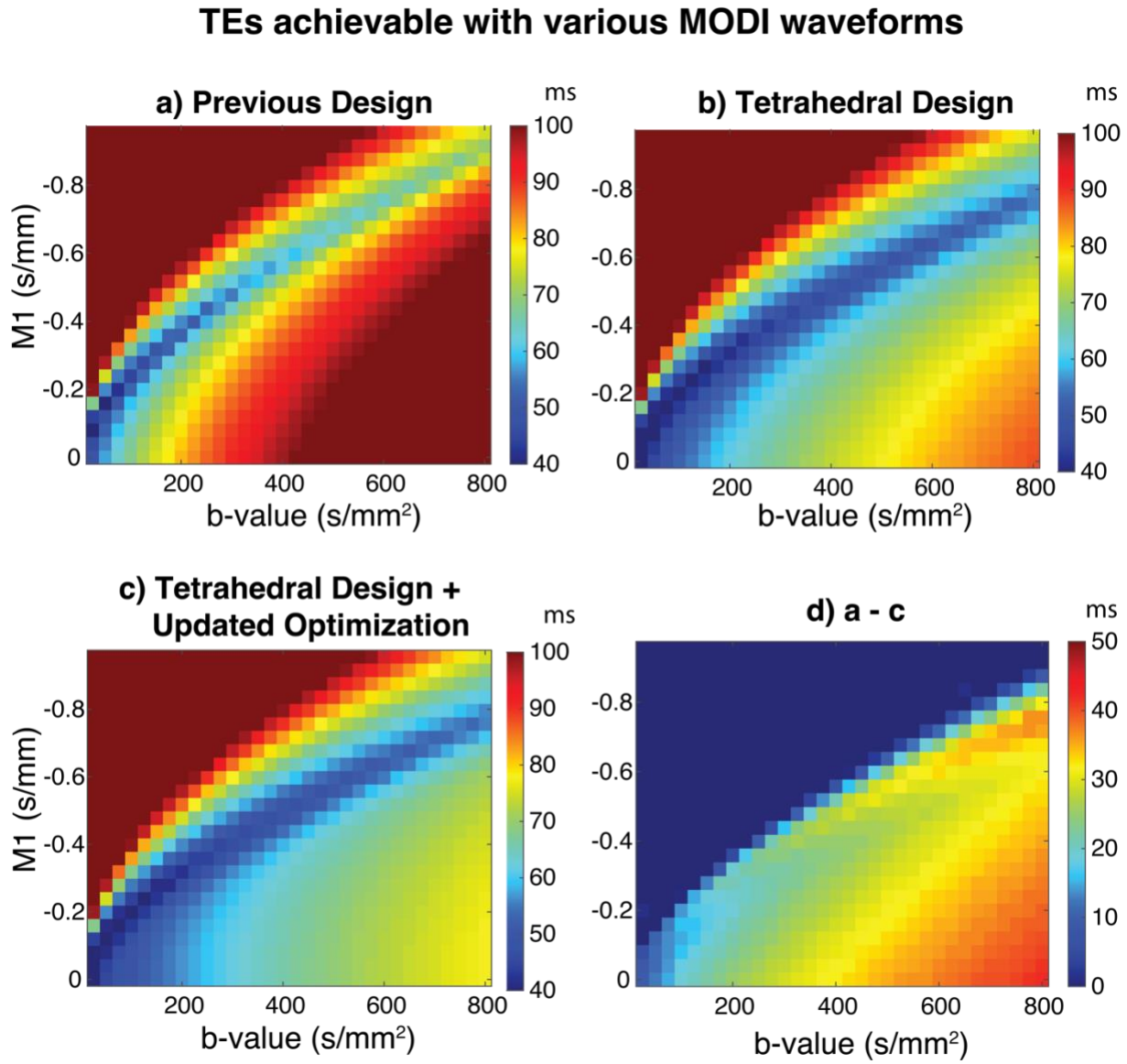

Figure 25: Echo Times Achievable with MODI. The tetrahedral waveform design with the updated optimization allows for substantial TE savings particularly for higher b-values and lower M1 values. **a)** Previously achievable echo times (TEs) for given b-value and M1 value with MODI optimization. **b)** TEs with tetrahedral design. **c)** TEs with tetrahedral design and updated optimization. **d)** Difference between a and c.

5.5.3 ADC Reproducibility

Example ADC maps from one subject on a single scanner using the different acquisitions and the different post-processing techniques are shown in **Figure 27**. The distributions of ADC measurements from each individual liver segment across all subjects and scanners are shown in

Figure 28. Both the example ADC maps and liver segment measurements showed overestimated ADC in segment I and the left liver lobe (segments II-IVb) using MONO with simple averaging. Use of weighted or squared weighted averaging reduced the ADC overestimation with squared weighted averaging having a greater effect. tMODI showed less overestimation than MONO, but still had residual ADC overestimation in segment I and the left liver lobe. Application of weighted and squared weighted averaging with tMODI acquisitions also demonstrated reduced overestimation.

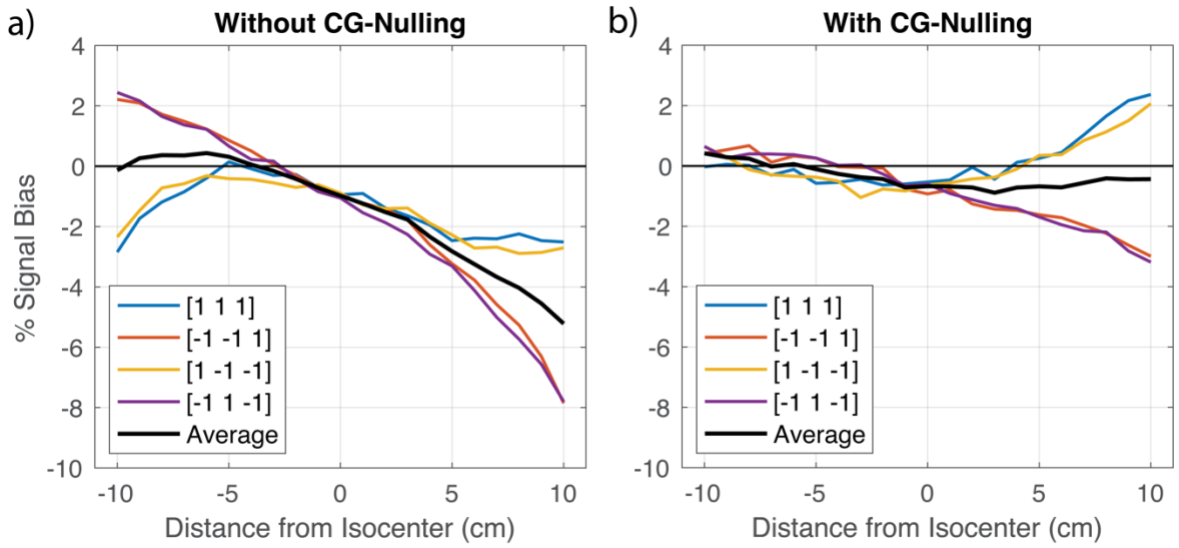

Figure 26: Experimental Validation of Concomitant Gradient Correction. a) DWI of a silicone oil phantom using tMODI without concomitant gradient (CG) nulling showed negative signal bias. **b)** tMODI with CG-nulling reduced the bias. Distance from the isocenter is along the direction of the B0 field (slice-direction). Thinner colorful lines represent individual diffusion directions while the thicker black line represents the average among all diffusion directions.

Results of mixed-effects analysis are presented in **Table 10**. Results of the mixed-effects analysis are presented in Table 2. Compared to the right lobe, ADC values in the left lobe were significantly higher using MONO with any averaging technique and tMODI with simple averaging (all $p<0.001$) with an effect size ranging from 0.14 to 0.43 $\times 10^{-3}$ mm^2/s. Application of weighted averaging to tMODI data still had significant left-right lobe difference ($p=0.01$) although the effect

size was small (0.06×10^{-3} mm^2/s). No significant difference was detected when tMODI was used with squared weighted averaging (p=0.59). Additionally, there was no significant difference observed between gradient performance or field strength for any acquisition/averaging technique (p > 0.06 across all comparisons).

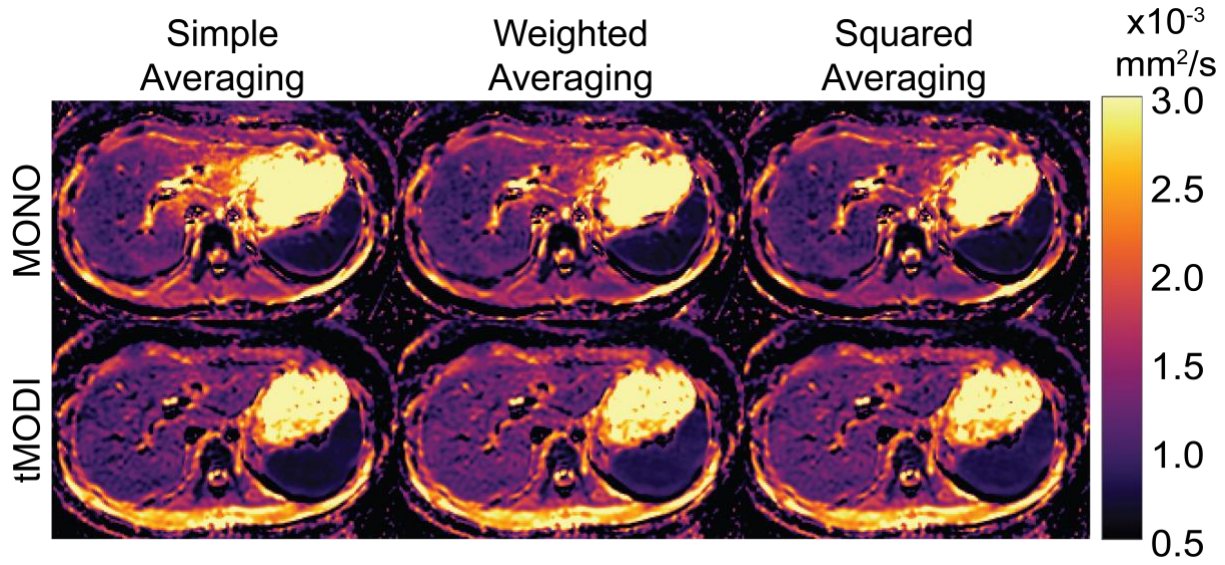

Figure 27: Example ADC Maps Using Various DWI Acquisitions and Averaging Techniques. Example ADC maps from one healthy volunteer on a single scanner demonstrate the overestimated ADC in the left lobe of the liver using MONO and reduction of this overestimation using tMODI. Applying weighted or squared weighted averaging helped to suppress the left lobe overestimation in MONO but left-lobe ADC was still overestimated compared to tMODI with the same averaging technique. In this example, the tMODI acquisition largely avoided residual motion artifacts and the different averaging techniques show little ADC difference visually.

Figure 29 contains ADC maps from a single subject across the three scanners. Visually, MONO ADC maps showed overestimated ADC values in the left lobe. The intensity and spatial distribution of this overestimation varies from scanner to scanner. tMODI acquisitions had less overestimated ADC values that appeared more consistent across the three scanners. Quantitative analysis of the reproducibility among scanners can be found in the scatterplots of **Figure 30.** The RDC_C using MONO with simple averaging was 0.84×10^{-3} mm^2/s. In contrast, tMODI ADC measurements with the same averaging had an RDC_C of 0.63×10^{-3} mm^2/s. Use of weighted or

squared weighted averaging reduced the RDC_C for both MONO and tMODI. The lowest RDC_C was observed using tMODI with squared weighted averaging ($RDC_C = 0.47 \times 10^{-3}$ mm^2/s).

Individual scanner-to-scanner comparisons can be found in

Table 11. All mean differences between each scanner pair had a magnitude below 0.1×10^{-3} mm^2/s. tMODI resulted in smaller RDC values compared to MONO between each scanner pair and for all averaging techniques. RDC values were similar between each scanner pair and ranged from 0.38×10^{-3} mm^2/s to 0.70×10^{-3} mm^2/s.

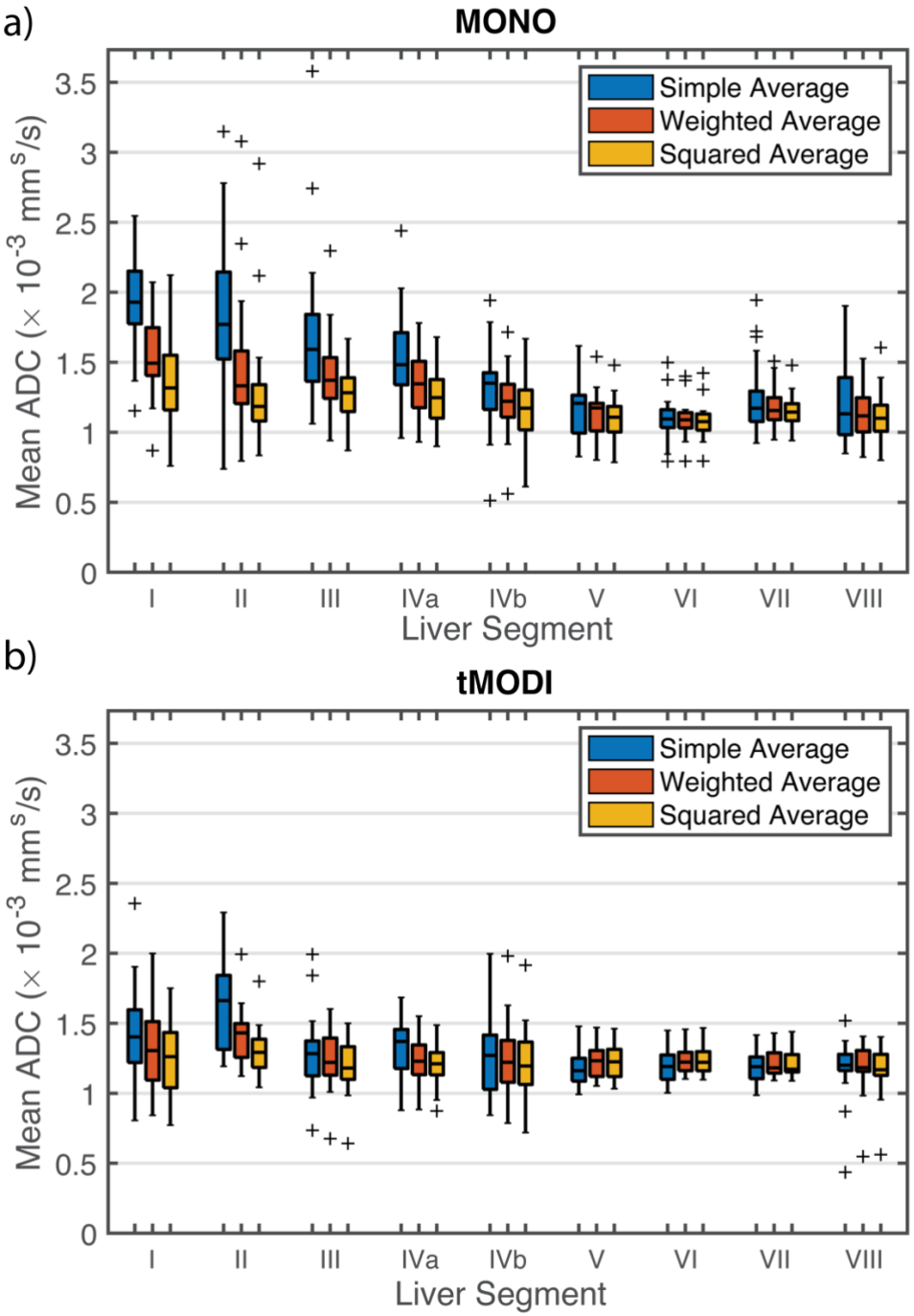

Figure 28: ROI Measurements Across All Liver Segments in Healthy Volunteers. a) MONO with simple averaging led to overestimated ADC measurements in healthy volunteers' liver segments I-IVb. **b)** MODI reduced this overestimation making ADC measurements more consistent across the different segments. Both MONO and MODI showed reduced intra-segment and inter-segment variation after application of weighted averaging. Individual box plots contain ADC measurements across all scanners and subjects for a given segment, acquisition type, and averaging technique.

Table 10: Results of Mixed Effect Modeling Predicting ADC Differences.

Component	Averaging Method	MONO ADC Change ($\times 10^{-3}$ mm^2/s)	MONO p-value	MODI ADC Change ($\times 10^{-3}$ mm^2/s)	MODI p-value
	Simple	1.13 (0.99, 1.26)	<0.001	1.16 (1.09, 1.22)	<0.001
Intercept	Weighted	1.11 (1.04, 1.18)	<0.001	1.19 (1.13, 1.24)	<0.001
	Squared	1.10 (1.02, 1.17)	<0.001	1.18 (1.13, 1.23)	<0.001
Lobe					
Right		Reference		Reference	
	Simple	0.43 (0.33, 0.52)	<0.001	0.19 (0.12, 0.26)	<0.001
Left	Weighted	0.23 (0.16, 0.30)	<0.001	0.06 (0.01, 0.11)	0.01
	Squared	0.14 (0.08, 0.19)	<0.001	-0.01 (0.03, 0.06)	0.59
Gradients					
High		Reference		Reference	
	Simple	0.11 (-0.01, 0.23)	0.06	0.03 (-0.5, 0.11)	0.46
Conventional	Weighted	0.02 (-0.07, 0.11)	0.65	-0.01 (-0.07, 0.05)	0.74
	Squared	0.01 (-0.06, 0.09)	0.72	-0.01 (-0.07, 0.04)	0.64
Field Strength					
3.0T		Reference		Reference	
	Simple	0.01 (-0.11, 0.13)	0.92	0.03 (-0.05, 0.11)	0.44
1.5T	Weighted	0.03 (-0.06, 0.12)	0.54	0.05 (-0.01, 0.11)	0.10
	Squared	0.01 (-0.06, 0.09)	0.69	0.05 (-0.01, 0.11)	0.07

Data in parentheses are 95% confidence intervals. ADC measurements from segment I were excluded from mixed-effects analysis since it is not considered a part of the right or left liver lobe.

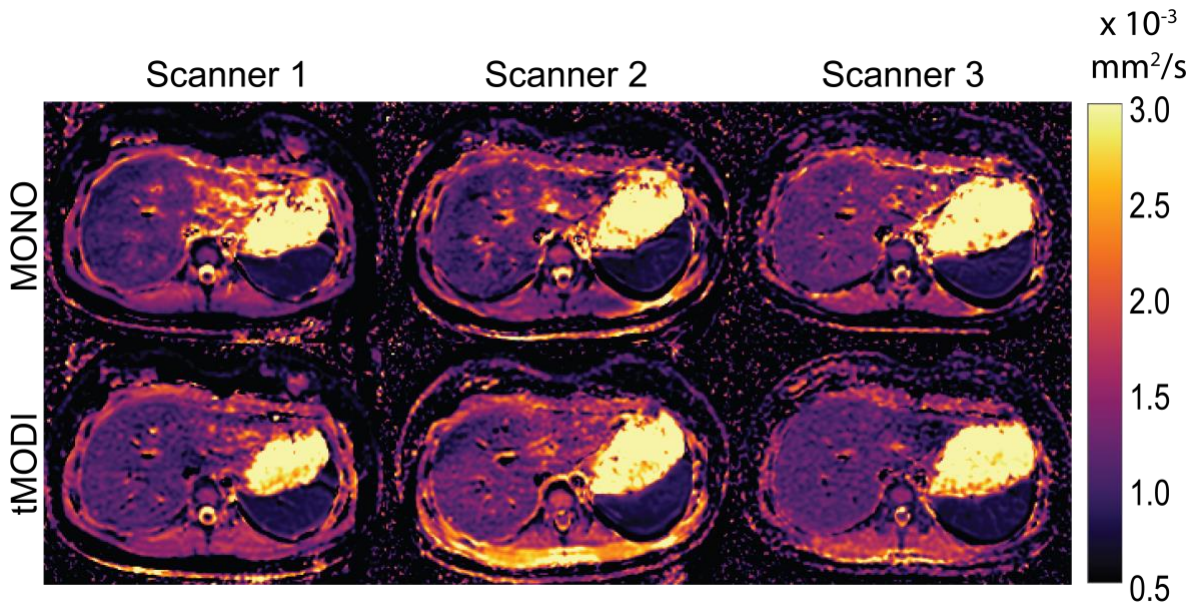

Figure 29: Reproducibility of ADC Maps Across Multiple MR Scanners. Example ADC maps from a single healthy volunteer on three different scanners show the visually overestimated and variable ADC measurements in the left lobe of the liver due to motion artifacts using MONO. tMODI demonstrated visually more consistent left lobe ADC measurements across the three scanners. Examples shown here used the squared weighted averaging technique demonstrating the presence of motion artifacts with MONO even when advanced averaging is used.

Figure 30: Scatter Plots of Within ROI Variability of ADC Measurements with Different DWI Acquisitions and Averaging Techniques. (See following page.) Scatterplots illustrate that the range of ADC measurements and the value of RDC_C measurements were reduced using tMODI and/or advanced averaging techniques. Each point in the above scatter plots represents a single ROI from a single subject compared across two scanners (e.g., blue circles show the difference between scanner 1 and scanner 2). The x-position of each dot is the mean ADC value for that ROI in that subject across all 3 scanners. RDC_C is the absolute difference expected between 95% of measurements between any two scanners. Dotted lines represent +/- RDC_C**.**

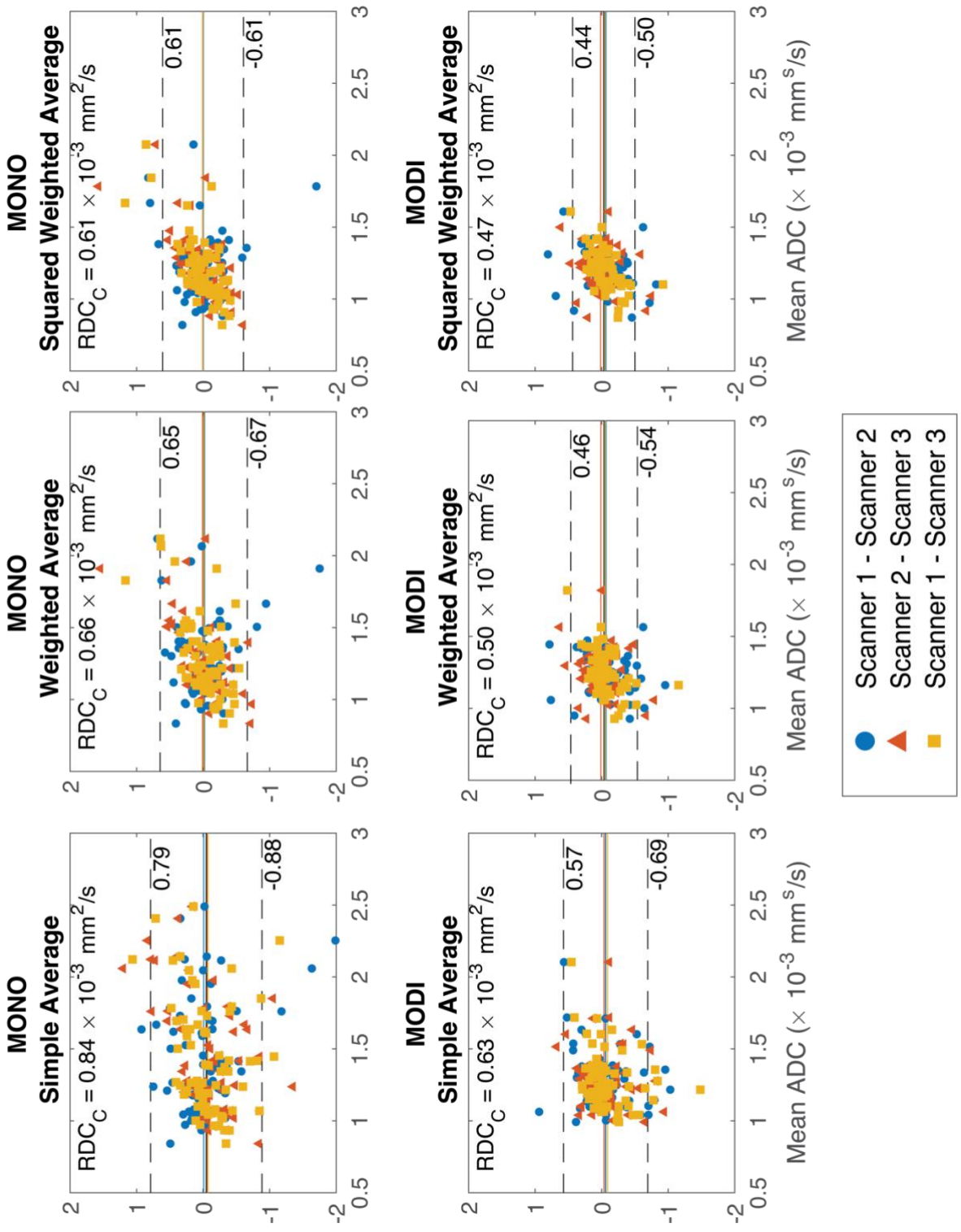
MONO
Simple Average
RDC_C = 0.84 × 10^-3 mm^2/s
0.79
-0.88
MONO
Weighted Average
RDC_C = 0.66 × 10^-3 mm^2/s
0.65
-0.67
MONO
Squared Weighted Average
RDC_C = 0.61 × 10^-3 mm^2/s
0.61
-0.61
MODI
Simple Average
RDC_C = 0.63 × 10^-3 mm^2/s
0.57
-0.69
MODI
Weighted Average
RDC_C = 0.50 × 10^-3 mm^2/s
0.46
-0.54
MODI
Squared Weighted Average
RDC_C = 0.47 × 10^-3 mm^2/s
0.44
-0.50
Within ROI Variability (× 10^-3 mm^s/s)
Mean ADC (× 10^-3 mm^s/s)
Scanner 1 - Scanner 2
Scanner 2 - Scanner 3
Scanner 1 - Scanner 3

Table 11: Summary of Reproducibility Coefficients and Differences in Scanner ADC Measurements.

Acquisition	Averaging Method	RDC_C	$\overline{\Delta}_{1,2}$	$RDC_{1,2}$	$\overline{\Delta}_{2,3}$	$RDC_{2,3}$	$\overline{\Delta}_{1,3}$	$RDC_{1,3}$
MONO	Simple	0.84	0.01	0.88	0.06	0.87	0.07	0.76
	Weighted	0.66	0.02	0.70	-0.01	0.70	0.01	0.57
	Squared	0.61	0.01	0.56	-0.01	0.63	0.00	0.54
tMODI	Simple	0.63	0.06	0.70	0.03	0.58	0.09	0.61
	Weighted	0.50	0.07	0.59	-0.01	0.46	0.05	0.43
	Squared	0.47	0.06	0.56	-0.02	0.44	0.05	0.38

All values are in $\times 10^{-3}$mm^2/s/ $\overline{\Delta}_{i,j}$ is the average difference between ADC measurements from scanner i and scanner j (scanner j – scanner i). $RDC_{i,j}$ is the reproducibility coefficient between scanner i and scanner j.

5.6 Discussion

This chapter demonstrated the feasibility of M1-optimized DWI of the liver on clinical MR systems with conventional gradient performance. The proposed updates to the existing MODI optimization including implementation of tetrahedral gradients allowed for shortened TEs on these systems. Further, the work assessed the combination of tMODI waveforms and advanced averaging techniques to mitigate motion artifacts and the associated overestimated ADC. Acquisitions based on MONO and tMODI benefited from advanced averaging techniques by reducing the overestimation of ADC in the left lobe of the liver. The impact of tMODI and advanced averaging on ADC reproducibility was also assessed. Compared to conventional MONO diffusion waveforms, the proposed tMODI approach led to improved ADC reproducibility across MR systems. The reproducibility was improved further when advanced averaging techniques were applied. Overall, tMODI appears as a promising technique for motion-robust DWI on multiple scanner platforms with various field strengths and gradient performance characteristics.

In agreement with previous studies, conventional monopolar gradient waveforms produced ADC overestimation in the left liver lobe compared to the right lobe[101,119,120]. In contrast, when

using tMODI, the interlobar difference was substantially reduced[106]. This work additionally confirmed that application of advanced averaging techniques mitigates residual motion artifacts[107,111]. Importantly, this work applies this advanced averaging to M1-optimized diffusion imaging to enable blood signal suppression with improved motion-robustness. tMODI with conventional arithmetic averaging produced a modest, but significant difference in ADC measurements between the left and right liver lobes. Using tMODI in conjunction with square weighted averaging resulted in no significant difference between ADC values in both liver lobes. These results demonstrate the potential of combining both prospective motion correction in the form of tMODI and retrospective motion correction with advanced averaging techniques.

This prospective-retrospective combination of motion-robust DWI also improved the reproducibility of ADC measurements across scanners. A prior study found that MONO waveforms with a respiratory triggered acquisition found ADC repeatability (i.e., repeated ADC measurements on the same scanner) to have a 95% limit of agreement of 0.60×10^{-3} mm^2/s[97]. In this work, ADC reproducibility (i.e., repeated ADC measurement across different scanners) using MONO waveforms with simple averaging was found to be 0.84×10^{-3} mm^2/s. tMODI with simple averaging led to an RDC_C of 0.63×10^{-3} mm^2/s. Given that reproducibility measures are generally more variable than test-retest repeatability, our reproducibility results appear consistent with the literature. Use of advanced averaging with either MONO or tMODI resulted in RDCs lower than previously reported repeatability for MONO measurements.

This work has several potential impacts on clinical practice and body DWI research. First, the reduction in left lobe signal dropout using tMODI acquisitions may enable improved lesion detection throughout the liver, particularly in regions affected by compressive motion. Importantly, this work expands previous MODI prototypes at 3T on systems with high

performance gradients, demonstrating that DWI using tMODI is feasible on a wider range of clinical MR systems. This work also suggests that tMODI may provide more reproducible ADC measurements in the liver, and, therefore, may help increase the utility of ADC measurements clinically. While this work focuses on quantification of liver DWI using the ADC, tMODI and advanced averaging techniques improve the consistency of the underlying DWI signal. Therefore, these developments may also help improve the precision of more advanced quantitative diffusion MRI methods such as kurtosis[93] or recent biophysical diffusion models[94,95]. Finally, while this work focuses on the liver, the techniques shown here may benefit DWI and quantification in other organs such as pancreas or kidneys which are also sensitive to motion[121,122].

This study is limited by the small number of subjects included in the reproducibility analysis. Additionally, these subjects were healthy volunteers with no known pathologies. Further study of these techniques with a more extensive patient cohort is necessary. This work also used a single MR vendor at a single site. Extension to additional vendors is important future work but was infeasible in this project due to the challenges of multi-vendor implementation. The use of vendor-neutral pulse sequences constitutes an intriguing approach to overcome this limitation. Further, this work uses limited diffusion parameters (e.g., a single b-value pair and single M1-value, along with the resulting ADC map), due to scan time constraints. Even though the TEs and therefore diffusion times were slightly different for each scanner, the dependence of ADC on TE was not explicitly modeled for this study. Future work may include additional testing with multiple vendors at multiple sites, and/or testing with a wider range of diffusion parameters.

Concomitant gradient corrections were included in this work, but tetrahedral gradients may also result in additional distortions or signal bias due to induced eddy currents. While not included in this work, the potential of including eddy current nulling in the optimization was investigated.

A preliminary evaluation similar to existing work by others was performed[103]. Including eddy current nulling in this case increased the TE by approximately 10 ms. This TE increase along with the lack of evidence suggesting that eddy current nulling was required led to the decision to exclude eddy current correction from this work.

This work evaluated two previously proposed weighted averaging methods to address residual motion artifacts in the acquired signal. These methods were shown to improve reproducibility. Other groups have found that using even more aggressive averaging techniques such as weighted averaging with weights raised to the third or fourth power may provide additional reduction in motion artifact[111]. However, a comprehensive characterization of the bias-variance tradeoff[111] and an analysis of additional averaging methods[107,111,123] remain as future work.

5.7 Conclusions

In conclusion, this chapter demonstrates the feasibility of motion-robust DWI and ADC mapping on scanners with conventional gradient strength and demonstrates improved reproducibility of ADC mapping compared to conventional DWI of the liver.

Chapter Six

Summary and Future Work

6.1 Summary

The chapters presented in this dissertation represent substantial advances in the field of MRI with particular focus on breast MRI and abdominal DWI. In this final summary chapter, a brief review of the novel scientific contributions of this work is presented along with the significance of each contribution. The next section will briefly discuss potential future work which immediately follows from the results presented in this dissertation.

In Chapter 2, the effects of a DL-based reconstruction on the quality of T2w breast MRI was investigated. The DL reconstruction was shown to improve image quality when applied to a SOC T2w acquisition. Further, a HR acquisition was developed that was designed to capitalize on the image quality benefits offered by the DL reconstruction. Improvements in SNR and image sharpness were demonstrated for both acquisition techniques in a phantom study and in human subject data. This work demonstrates the feasibility of using a DL reconstruction to improve image quality in T2w breast MRI and suggests that a relaxation on previous compromises between acquisition time and image quality may be possible.

In Chapter 3, a novel automated technique for placement of scan and pre-scan volumes in breast MRI was introduced. This technique was based on a neural network and predicted both pre-scan and scan volumes which closely resembled volumes placed by human users. The

demonstrated technique was also able to estimate the uncertainty in volume placements which is a unique feature for neural network-based placement techniques in MRI. This work is significant as it introduces an automated technique for placement that is less prone to intra- and inter-operator variability and is faster than human technologists. Development of automated placement techniques is critical in applications such as abbreviated breast MRI where optimized, streamlined workflows are essential to comply with strict exam time constraints.

In Chapter 4, the effect of GBCA attenuation on PET quantification during simultaneous PET/MR was measured. Overall, no significant attenuation due to GBCA was observed for any timing intervals used in the study or for any of the reconstructions tested. This work greatly expands upon previous phantom work in this space by being the first to validate the lack of effect in vivo. This work confirms that GBCA may be used in simultaneous PET/MR examinations without biasing quantitative measures of radiotracer uptake.

In Chapter 5, motion-robust DWI using MODI and advanced averaging techniques were shown to be feasible on a variety of scanners with various gradient performance characteristics and field strengths. This was enabled by the introduction of tMODI which greatly reduced the achievable TEs on systems with conventional gradient performance. tMODI was observed to provide more reproducible ADC measurements compared to a conventional MONO diffusion acquisition. This work is significant as it allows for use of MODI-based techniques across many MR scanners and demonstrates that specialized, high-performance gradient systems are not necessary for motion-robust DWI. It provides evidence that a combination of MODI techniques and weighted averaging may allow for consistent and unbiased measurements of ADC measurements in the abdomen. Overall, it helps address a major concern in abdominal DWI,

namely, motion-induced artifacts in DWI and associated poor reproducibility in ADC measurements.

6.2 Extensions and Future Work

6.2.1 Deep Learning Reconstruction for T2-weighted Imaging

This work demonstrated the feasibility of using a deep learning reconstruction to increase the quality of T2w imaging in breast MRI. Further work in this space could include an assessment of the diagnosis capability of images generated using the deep learning reconstruction. Metrics such as the diagnostic accuracy, sensitivity, and specificity could be measured and reported. Additional future work could assess the image quality using various levels of the DL reconstruction. In this work 75% was used, but some radiologists found that this level of denoising produced ‘fake’ looking images. A lower level such as 50% would provide less denoising that may result in images more appealing to certain radiologists.

Finally, further development of the DL reconstruction could allow for application to 3D sequences. This work could then be further expanded upon to include the DCE image series. As the most sensitive sequence for diagnosis, application of the DL reconstruction to the DCE may have profound impacts on diagnostic accuracy, sensitivity, and specificity. With either 2D or future 3D sequences, clinical protocols could and should be developed with potential deep learning reconstructions in mind in order to take full advantage of the possible time savings and/or increased image quality.

6.2.2 Automated Prescription of Scan and Pre-scan Volumes

The automated prescription technology introduced in this work is a promising way to provide quick and consistent prescription of breast exams. Future work using the automated prescription technology could expand the training and testing data to include implants and cases

with unique breast morphology due to mastectomy or lumpectomy. These are critically important cases since many patients who receive breast MRI have a previous history of cancer and are highly likely to have implants or surgical changes. Because of the projection and binarization performed as part of image pre-processing, inclusion of implants is expected to have little effect on the overall model performance. However, this hypothesis requires further testing. Mastectomy or lumpectomy cases may pose more of a challenge due to the presence of unusual morphology resulting from surgery. For example, an exam with unilateral mastectomy will have extreme asymmetry due to the removal of breast tissue from one side of the patient. It is unknown how the automated placement would respond to such a case. It may not even be sensical to place bilateral pre-scan volumes in such a scenario. Additional consideration should be given to exams with mastectomy or lumpectomy in the future.

A limitation of this work is that it does not evaluate the actual clinical performance of the volumes, i.e. image quality. Scan volumes clinical performance can likely be inferred based on the overlap metric which was high and showed that relevant anatomy was being covered. The pre-scan volumes' clinical performance is harder to evaluate. Ideally, better placed shim volumes would improve image quality by enabling more optimal scan-specific system settings. This was not evaluated in this work and is left as a future research topic. In addition, the reproducibility of model-based volume placement should be compared to the inter- and intra-operator variability of the clinical technologists to verify if the automated placements are indeed more consistent.

Another potential area of future research for this work would be to try using a different neural network architecture or expanding the training data with exams from different institutions. This would assist in further dissemination of this tool and validate or call into question the reproducibility of its overall performance. Finally, implementation of the automated technique on

a scanner is necessary to assess volume placement prospectively and should be included in future work.

6.2.3 GBCA Attenuation Effect on PET Quantification in Breast MRI

The work presented in Chapter 4 demonstrated that GBCA attenuation did not impact PET quantification using SUV measurements. Overall, this information should be used to inform clinical protocols. Protocol designers generally do not have to consider the effects of GBCA attenuation when creating their simultaneous PET/MR exams. However, it should be noted that the effects of GBCA on MR-based attenuation correction should still be considered and studied further.

Dynamic reconstructions using PET frames as short as 30 s were used which is near the shortest time duration that is feasible for dynamic PET reconstructions. Advances in radiotracers, PET detectors, and imaging technique may allow dynamic imaging with even shorter frame durations. As frame duration shrinks, latent attenuation effects due to GBCA may appear due to higher average concentrations of GBCA in frames immediately following contrast administration. Thus, future work may consider using shorter dynamic frames to probe for any potential effect.

In addition, this work focused on quantification using SUV_{mean} and SUV_{max}. It is hypothesized that other quantification metrics such as those obtained from radiotracer pharmacokinetic modeling would be similarly unaffected, however, this hypothesis requires further investigation.

6.2.4 ADC Reproducibility Using M1-optimized DWI

Chapter 5 showed the feasibility of MODI on a variety of MR systems including 1.5 T and 3.0 T and systems with variable gradient performance. ADC reproducibility across these systems was assessed. Future study in this area should expand the study of reproducibility by including

systems from multiple institutions and multiple manufacturers. This dissertation's updates to MODI assist in this by enabling scanning across a broader range of systems and sites.

Additional future work could investigate motion-robust diffusion quantification in other organs. For example, the pancreas is also susceptible to cardiac-induced motion artifacts and has been studied with MODI previously[121]. However, previous pancreas-focused work did not utilize the MODI updates presented in this work and was limited to a single scanner with high performance gradients. It also did not utilize advanced repetition averaging techniques. Future study of motion-robust DWI in the pancreas, kidney, heart and other organs would likely benefit from the MODI updates and averaging techniques presented in this dissertation.

Finally, future work should also test MODI using a broad range of b-values and M1-values. Confirmation that MODI is feasible using a large range of b- and M1-values is necessary for more advanced diffusion modelling techniques such as kurtosis. Liver and other organs may benefit from ADC estimation using a different combination of b-values which require further validation beyond this work.

Thank you for reading this dissertation! I hope you found it interesting and potentially useful to your own work. Cheers!

References

1. American Cancer Society. American Cancer Society. Cancer Facts & Figures 2021. Atlanta: American Cancer Society; 2021. Published online 2021. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2021.html. Accessed: 06/08/2023.

2. Mann RM, Kuhl CK, Kinkel K, Boetes C. Breast MRI: Guidelines from the European Society of Breast Imaging. *Eur Radiol*. 2008;18(7):1307-1318. doi:10.1007/s00330-008-0863-7.

3. Sardanelli F, Boetes C, Borisch B, et al. Magnetic resonance imaging of the breast: Recommendations from the EUSOMA working group. *Eur J Cancer*. 2010;46(8):1296-1316. doi:10.1016/j.ejca.2010.02.015

4. American College of Radiology. Practice Parameter for the Performance of Contrast-Enhanced Magnetic Resonance Imaging (CE-MRI) of the Breast. 2018;1076:1-11. https://www.acr.org/-/media/ACR/Files/Practice-Parameters/MR-Contrast-Breast.pdf?la=en. Accessed: 06/08/2023.

5. Kuhl C. The Current Status of Breast MR Imaging Part I. Choice of Technique, Image Interpretation, Diagnostic Accuracy, and Transfer to Clinical Practice. *Radiology*. 2007;244(2):356-378. doi:10.1148/radiol.2442051620

6. DeMartini W, Lehman C. A Review of Current Evidence-Based Clinical Applications for Breast Magnetic Resonance Imaging. *Topics in Magnetic Resonance Imaging*. 2008;19(3):143-150. doi:10.1097/RMR.0b013e31818a40a5

7. Lehman CD, Isaacs C, Schnall MD, et al. Cancer Yield of Mammography, MR, and US in High-Risk Women: Prospective Multi-Institution Breast Cancer Screening Study. *Radiology*. 2007;244(2):381-388. doi:10.1148/radiol.2442060461

8. Kuhl CK. Abbreviated breast MRI for screening women with dense breast: The EA1141 trial. British Journal of Radiology. 2018;91(1090). doi:10.1259/bjr.20170441

9. Kuhl CK, Schrading S, Strobel K, Schild HH, Hilgers RD, Bieling HB. Abbreviated breast Magnetic Resonance Imaging (MRI): First postcontrast subtracted images and maximum-intensity projection - A novel approach to breast cancer screening with MRI. *Journal of Clinical Oncology*. 2014;32(22):2304-2310. doi:10.1200/JCO.2013.52.5386

10. Maril N, Collins CM, Greenman RL, Lenkinski RE. Strategies for shimming the breast. *Magn Reson Med*. 2005;54(5):1139-1145. doi:10.1002/mrm.20679

11. Hancu I, Govenkar A, Lenkinski R, Lee SK. On shimming approaches in 3T breast MRI. *Magn Reson Med*. 2013;69(3):862-867. doi:10.1002/mrm.24307

12. Zhou W, Favazza CP, Axmacher JA, Trzasko JD, Geske JR, Lee CU. Evaluation of Shimming Techniques on MRI Breast Image Quality at 1.5T. *J Breast Imaging*. 2019;1(3):199-204. doi:10.1093/jbi/wbz045

13. Lee SK, Hancu I. Patient-to-patient variation of susceptibility-induced B0 field in bilateral breast MRI. *Journal of Magnetic Resonance Imaging*. 2012;36(4):873-880. doi:10.1002/jmri.23715

14. American College of Radiology. MRI Exam-Specific Parameters: Breast. Revised: 24 February 2023. Available online: https://accreditationsupport.acr.org/support/solutions/articles/11000114407-mri-exam-specific-parameters-breast-revised-2- 24-2023-#:~:text=MRI%20Exam-Specific%20Parameters%3A%20Breast%20%28Revised%202-24-2023%29%201%20Do,for%20the%20pre-contrast%20and%20the%20post-contrast%20sequences.%20 (accessed on 2 May 2023).

15. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500-510. doi:10.1038/s41568-018-0016-5

16. Saba L, Biswas M, Kuppili V, et al. The present and future of deep learning in radiology. *Eur J Radiol*. 2019;114(September 2018):14-24. doi:10.1016/j.ejrad.2019.02.038

17. Lin DJ, Johnson PM, Knoll F, Lui YW. Artificial Intelligence for MR Image Reconstruction: An Overview for Clinicians. *Journal of Magnetic Resonance Imaging*. 2021;53(4):1015-1028. doi:10.1002/jmri.27078

18. Aggarwal HK, Mani MP, Jacob M. MoDL: Model Based Deep Learning Architecture for Inverse Problems. *IEEE Trans Med Imaging*. 2017;38(2). doi:10.1109/TMI.2018.2865356

19. Herrmann J, Koerzdoerfer G, Nickel D, et al. Feasibility and Implementation of a Deep Learning MR Reconstruction for TSE Sequences in Musculoskeletal Imaging. *Diagnostics*. 2021;11(8):1484. doi:10.3390/diagnostics11081484

20. Chen F, Taviani V, Malkiel I, et al. Variable-density single-shot fast spin-echo MRI with deep learning reconstruction by using variational networks. *Radiology*. 2018;289(2):366-373. doi:10.1148/radiol.2018180445

21. Fuin N, Bustin A, Küstner T, et al. A multi-scale variational neural network for accelerating motion-compensated whole-heart 3D coronary MR angiography. *Magn Reson Imaging*. 2020;70:155-167. doi:10.1016/j.mri.2020.04.007

22. Reig B, Heacock L, Geras KJ, Moy L. Machine learning in breast MRI. *Journal of Magnetic Resonance Imaging*. 2020;52(4):998-1018. doi:10.1002/jmri.26852

23. Balkenende L, Teuwen J, Mann RM. Application of Deep Learning in Breast Cancer Imaging. *Semin Nucl Med*. Published online 2022. doi:10.1053/j.semnuclmed.2022.02.003

24. Lebel RM. Performance characterization of a novel deep learning-based MR image reconstruction pipeline. *ArXiv*. Published online 2020.

25. Kim M, Kim HS, Kim HJ, et al. Thin-slice pituitary mri with deep learning–based reconstruction: diagnostic performance in a postoperative setting. *Radiology*. 2021;298(1):114-122. doi:10.1148/RADIOL.2020200723

26. van der Velde N, Hassing HC, Bakker BJ, et al. Improvement of late gadolinium enhancement image quality using a deep learning–based reconstruction algorithm and its influence on myocardial scar quantification. *Eur Radiol*. 2021;31(6):3846-3855. doi:10.1007/s00330-020-07461-w

27. Wang X, Ma J, Bhosale P, et al. Novel deep learning-based noise reduction technique for prostate magnetic resonance imaging. *Abdominal Radiology*. 2021;46(7):3378-3386. doi:10.1007/s00261-021-02964-6

28. Koch KM, Sherafati M, Emre Arpinar V, et al. Analysis and evaluation of a deep learning reconstruction approach with denoising for orthopedic mri. *Radiol Artif Intell*. 2021;3(6). doi:10.1148/ryai.2021200278

29. Itti L, Chang L, Ernst T. Automatic Scan Prescription for Brain MRI. *Magn Reson Med*. 2001;45(3):486-494. doi:10.1002/1522-2594(200103)45:3<486::aid-mrm1064>3.0.co;2-#

30. Park YW, Deelchand DK, Joers JM, et al. AutoVOI: real-time automatic prescription of volume-of-interest for single voxel spectroscopy. *Magn Reson Med*. 2018;80(5):1787-1798. doi:10.1002/mrm.27203

31. Bystrov D, Pekar V, Young S, Dries SPM, Heese HS, van Muiswinkel AM. Automated planning of MRI scans of knee joints. In: *Medical Imaging 2007: Visualization and Image-Guided Procedures*. Vol 6509. SPIE; 2007:65092Z. doi:10.1117/12.709255

32. Goto T, Kabasawa H. Automated scan prescription for MR imaging of deformed and normal livers. *Magnetic Resonance in Medical Sciences*. 2013;12(1):11-20. doi:10.2463/mrms.2012-0006

33. Ozhinsky E, Vigneron DB, Chang SM, Nelson SJ. Automated prescription of oblique brain 3D magnetic resonance spectroscopic imaging. *Magn Reson Med*. 2013;69(4):920-930. doi:10.1002/mrm.24339

34. Ozhinsky E, Vigneron DB, Nelson SJ. Improved spatial coverage for brain 3D PRESS MRSI by automatic placement of outer-volume suppression saturation bands. *Journal of Magnetic Resonance Imaging*. 2011;33(4):792-802. doi:10.1002/jmri.22507

35. Geng R, Sundaresan M, Starekova J, et al. Automated Image Prescription for Liver MRI using Deep Learning. In: *Proc. Intl. Soc. Mag. Reson. Med.* ; 2021.

36. Geng R, Buelo CJ, Sundaresan M, et al. Automated MR Image Prescription of the Liver Using Deep Learning: Development, Evaluation, and Prospective Implementation. *Journal of Magnetic Resonance Imaging*. Published online December 30, 2022. doi:10.1002/jmri.28564

37. Kojima S, Hirata M, Shinohara H, Ueno E. Reproducibility of scan prescription in follow-up brain MRI: manual versus automatic determination. *Radiol Phys Technol*. 2013;6(2):375-384. doi:10.1007/s12194-013-0211-8

38. Blansit K, Retson T, Masutani E, Bahrami N, Hsiao A. Deep Learning–based Prescription of Cardiac MRI Planes. *Radiol Artif Intell*. 2019;1(6):e180069. doi:10.1148/ryai.2019180069

39. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. *NPJ Digit Med*. 2021;4(1). doi:10.1038/s41746-020-00367-3

40. Combalia M, Vilaplana V. Monte-Carlo Sampling applied to Multiple Instance Learning for Histological Image Classification. Published online December 30, 2018. doi:10.1007/978-3-030-00889-5

41. Lemay A, Hoebel K, Bridge CP, et al. Improving the repeatability of deep learning models with Monte Carlo dropout. *NPJ Digit Med*. 2022;5(1). doi:10.1038/s41746-022-00709-3

42. Srivastava N, Hinton G, Krizhevsky A, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*. 2014;15:1929-1958.

43. Gal Y, Ghahramani Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In: *International Conference on Machine Learning*. ; 2016:1050-1059. http://arxiv.org/abs/1506.02142

44. Leibig C, Allken V, Ayhan MS, Berens P, Wahl S. Leveraging uncertainty information from deep neural networks for disease detection. *Sci Rep*. 2017;7(1). doi:10.1038/s41598-017-17876-z

45. Fendler WP, Czernin J, Herrmann K, Beyer T. Variations in PET/MRI operations: Results from an international survey among 39 active sites. *Journal of Nuclear Medicine*. 2016;57(12):2016-2021. doi:10.2967/jnumed.116.174169

46. Huang S ying, Franc BL, Harnish RJ, et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *NPJ Breast Cancer*. 2018;4(1). doi:10.1038/s41523-018-0078-2

47. Umutlu L, Kirchner J, Bruckmann NM, et al. Multiparametric integrated18f-fdg pet/mri-based radiomics for breast cancer phenotyping and tumor decoding. *Cancers (Basel)*. 2021;13(12):1-13. doi:10.3390/cancers13122928

48. Lim I, Noh WC, Park J, et al. The combination of FDG PET and dynamic contrast-enhanced MRI improves the prediction of disease-free survival in patients with advanced breast cancer after the first cycle of neoadjuvant chemotherapy. *Eur J Nucl Med Mol Imaging*. 2014;41(10):1852-1860. doi:10.1007/s00259-014-2797-4

49. Choi JH, Kim HA, Kim W, et al. Early prediction of neoadjuvant chemotherapy response for advanced breast cancer using PET/MRI image deep learning. *Sci Rep*. 2020;10(1):1-11. doi:10.1038/s41598-020-77875-5

50. Cho N, Im SA, Cheon GJ, et al. Integrated 18F-FDG PET/MRI in breast cancer: early prediction of response to neoadjuvant chemotherapy. *Eur J Nucl Med Mol Imaging*. 2018;45(3):328-339. doi:10.1007/s00259-017-3849-3

51. Mehranian A, Arabi H, Zaidi H. Vision 20/20: Magnetic resonance imaging-guided attenuation correction in PET/MRI: Challenges, solutions, and opportunities. *Med Phys*. 2016;43(3):1130-1155. doi:10.1118/1.4941014

52. Chen Y, An H. Attenuation Correction of PET/MR Imaging. *Magn Reson Imaging Clin N Am*. 2017;25(2):245-255. doi:10.1016/j.mric.2016.12.001

53. Liu F, Jang H, Kijowski R, Bradshaw T, Mcmillan AB. Deep Learning MR Imaging-based Attenuation Correction for PET/MR imaging. *Radiology*. 2018;286(2):676-684. doi:10.1148/radiol.2017170700

54. Moon M, Cornfeld D, Weinreb J. Dynamic Contrast-Enhanced Breast MR Imaging. *Magn Reson Imaging Clin N Am*. 2009;17(2):351-362. doi:10.1016/j.mric.2009.01.010

55. Hamirani YS, Kramer CM. Cardiac MRI assessment of myocardial perfusion. *Future Cardiol*. 2014;10(3):349-358. doi:10.2217/fca.14.18

56. Peters NHGM, Borel Rinkes IHM, Zuithoff NPA, Mali WPTM, Moons KGM, Peeters PHM. Meta-analysis of MR imaging in the diagnosis of breast lesions. *Radiology*. 2008;246(1):116-124. doi:10.1148/radiol.2461061298

57. Medeiros LR, Duarte CS, Rosa DD, et al. Accuracy of magnetic resonance in suspicious breast lesions: a systematic quantitative review and meta-analysis. *Breast Cancer Res Treat*. 2011;126(2):273-285. doi:10.1007/s10549-010-1326-9

58. Bennani-Baiti B, Bennani-Baiti N, Baltzer PA. Diagnostic Performance of Breast Magnetic Resonance Imaging in Non-Calcified Equivocal Breast Findings: Results from a Systematic Review and Meta-Analysis. *PLoS One*. 2016;11(8):e0160346. doi:10.1371/journal.pone.0160346

59. O' Doherty J, Schleyer P. An experimental phantom study of the effect of gadolinium-based MR contrast agents on PET attenuation coefficients and PET quantification in PET-MR imaging: application to cardiac studies. *EJNMMI Phys*. 2017;4(1). doi:10.1186/s40658-017-0173-8

60. Lois C, Bezrukov I, Schmidt H, et al. Effect of MR contrast agents on quantitative accuracy of PET in combined whole-body PET/MR imaging. *Eur J Nucl Med Mol Imaging*. 2012;39(11):1756-1766. doi:10.1007/s00259-012-2190-0

61. Ruhlmann V, Heusch P, Kühl H, et al. Potential influence of Gadolinium contrast on image segmentation in MR-based attenuation correction with Dixon sequences in whole-body 18F-FDG PET/MR. *Magnetic Resonance Materials in Physics, Biology and Medicine*. 2016;29(2):301-308. doi:10.1007/s10334-015-0516-1

62. Lee W, Park JA, Kim KM, et al. Effect of Gd-based MR contrast agents on CT attenuation of PET/CT for quantitative PET-MRI study. *Journal of Nuclear Medicine*. 2011;52(supplement 1 53):130-136.

63. Bruckmann NM, Lindemann ME, Grueneisen J, et al. Comparison of pre- and post-contrast-enhanced attenuation correction using a CAIPI-accelerated T1-weighted Dixon 3D-VIBE sequence in 68Ga-DOTATOC PET/MRI. *Eur J Radiol*. 2021;139:109691. doi:https://doi.org/10.1016/j.ejrad.2021.109691

64. Wang K, Gui D, Holmes J, et al. Toward Fully Automated Breast MR Exams Using Deep Learning. In: *Proceedings of the 2018 ISMRM Annual Meeting*.

65. National Electrical Manufacturers Association. NEMA Standards Publication MS 1-2008, Determination of Signal-to-Noise Ratio (SNR) in Diagnostic Magnetic Resonance Imaging. Published online 2008:1-19. http://www.nema.org/standards/pages/determination-of-signal-to-noise-ratio-in-diagnostic-magnetic-resonance-imaging.aspx?%5Cnpapers2://publication/uuid/CDE4FB6F-AC32-4A9E-B2F8-C0ADB3715C24

66. Zochowski KC, Tan ET, Argentieri EC, et al. Improvement of peripheral nerve visualization using a deep learning-based MR reconstruction algorithm. *Magn Reson Imaging*. 2022;85(June 2021):186-192. doi:10.1016/j.mri.2021.10.038

67. Nodine CF, Kundel HL. Using eye movements to study visual search and to improve tumor detection. *RadioGraphics*. 1987;7(6):1241-1250. doi:10.1148/radiographics.7.6.3423330

68. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: F. Pereira, C.J. Burges, L. Bottou, K.Q. Weinberger, eds. *Advances in Neural Information Processing Systems*. ; 2012. http://code.google.com/p/cuda-convnet/

69. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst Man Cybern*. 1979;9(1):62-66. doi:10.1109/TSMC.1979.4310076

70. Rezatofighi H, Tsoi N, Gwak J, Sadeghian A, Reid I, Savarese S. Generalized Intersection over Union. In: *The IEEE Conference on Computer Vision and Pattern REcognition (CVPR)*. ; 2019.

71. Hinton G. Coursera Neural Networks for Machine Learning Lecture 6. Accessed June 2, 2023. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf

72. Tong C, Yang X, Huang Q, Qian F. NGIoU Loss: Generalized Intersection over Union Loss Based on a New Bounding Box Regression. *Applied Sciences (Switzerland)*. 2022;12(24). doi:10.3390/app122412785

73. Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020;34(07):12993-13000. doi:10.1609/aaai.v34i07.6999

74. Fowler AM, Kumar M, Bancroft LH, et al. Measuring Glucose Uptake in Primary Invasive Breast Cancer Using Simultaneous Time-of-Flight Breast PET/MRI: A Method Comparison Study with Prone PET/CT. *Radiol Imaging Cancer*. 2021;3(1):e200091. doi:10.1148/rycan.2021200091

75. Chakraborty D, Basu S, Ulaner GA, Alavi A, Kumar R. Diagnostic Role of Fluorodeoxyglucose PET in Breast Cancer: A History to Current Application. *PET Clin*. 2018;13(3):355-361. doi:10.1016/j.cpet.2018.02.011

76. Pahk K, Kim S, Choe JG. Early prediction of pathological complete response in luminal B type neoadjuvant chemotherapy-treated breast cancer patients: comparison between interim 18F-FDG PET/CT and MRI. *Nucl Med Commun*. 2015;36(9):887-891.

77. Fletcher JW, Kinahan PE. PET/CT Standardized Uptake Values (SUVs) in Clinical Practice and Assessing Response to Therapy. *National Institute of Health*. 2010;31(6):496-505. doi:10.1053/j.sult.2010.10.001.PET/CT

78. Groheux D. Role of Fludeoxyglucose in Breast Cancer: Treatment Response. *PET Clin*. 2018;13(3):395-414. doi:10.1016/j.cpet.2018.02.003

79. Fowler AM, Strigel RM. Clinical advances in PET–MRI for breast cancer. *Lancet Oncol*. 2022;23(1):e32-e43. doi:https://doi.org/10.1016/S1470-2045(21)00577-5

80. Kanda T, Nakai Y, Oba H, Toyoda K, Kitajima K, Furui S. Gadolinium deposition in the brain. *Magn Reson Imaging*. 2016;34(10):1346-1350. doi:10.1016/j.mri.2016.08.024

81. Kanda T, Fukusato T, Matsuda M, et al. Gadolinium-based contrast agent accumulates in the brain even in subjects without severe renal dysfunction: Evaluation of autopsy brain specimens with inductively coupled plasma mass spectroscopy. *Radiology*. 2015;276(1):228-232. doi:10.1148/radiol.2015142690

82. Yau YY, Chan WS, Tam YM, et al. Application of intravenous contrast in PET/CT: Does it really introduce significant attenuation correction error? *Journal of Nuclear Medicine*. 2005;46(2):283-291.

83. Mawlawi O, Erasmus JJ, Munden RF, et al. Quantifying the effect of IV contrast media on integrated PET/CT: Clinical evaluation. *American Journal of Roentgenology*. 2006;186(2):308-319. doi:10.2214/AJR.04.1740

84. Shapira N, Scheuermann J, Perkins AE, et al. Quantitative positron emission tomography imaging in the presence of iodinated contrast media using electron density quantifications from dual-energy computed tomography. *Med Phys*. 2021;48(1):273-286. doi:10.1002/mp.14589

85. Jones DK. *Diffusion MRI*. Oxford University Press; 2010.

86. Haacke EM, Brown RW, Thompson MR, Venkatesan R. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Vol 82. Wiley-liss New York; 1999.

87. Palmucci S, Mauro LA, Messina M, et al. Diffusion-weighted MRI in a liver protocol: its role in focal lesion detection. *World J Radiol*. 2012;4 7:302-310.

88. Parikh TN, Drew SJ, Lee VS, et al. Focal liver lesion detection and characterization with diffusion-weighted MR imaging: comparison with standard breath-hold T2-weighted imaging. *Radiology*. 2008;246 3:812-822.

89. Bruegel M, Holzapfel K, Gaa J, et al. Characterization of focal liver lesions by ADC measurements using a respiratory triggered diffusion-weighted single-shot echo-planar MR imaging technique. *Eur Radiol*. 2008;18(3):477-485. doi:10.1007/s00330-007-0785-9

90. Cui Y, Zhang XP, Sun YS, Tang L, Shen L. Apparent Diffusion Coefficient: Potential Imaging Biomarker for Prediction and Early Detection of Response to Chemotherapy in Hepatic Metastases. *Radiology*. 2008;248(3):894-900. doi:10.1148/radiol.2483071407

91. Sandrasegaran K, Territo P, Elkady RM, et al. Does intravoxel incoherent motion reliably stage hepatic fibrosis, steatosis, and inflammation? *Abdominal Radiology*. 2018;43:600-606.

92. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*. 1988;168(2):497-505.

93. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med*. 2005;53.

94. Grussu F, Barba I, Bernatowicz K, et al. DR-HIGADOS: a new diffusion-relaxation framework for clinically feasible microstructural imaging of the liver. In: *Proc. Intl. Soc. Mag. Reson. Med.* ; 2021.

95. Grussu F, Bernatowicz K, Barba I, et al. Histological correlates of DR-HIGADOS microstructural metrics in the mouse and human liver. In: *Pod. Intl. Soc. Mag. Reson. Med.*; 2022.

96. Cieszanowski A, Pasicz K, Podgórska J, et al. Reproducibility of intravoxel incoherent motion of liver on a 3.0T scanner: free-breathing and respiratory-triggered sequences acquired with different numbers of excitations. *Pol J Radiol*. 2018;83:437-445. doi:10.5114/pjr.2018.79651

97. Kwee TC, Takahara T, Koh DM, Nievelstein RAJ, Luijten PR. Comparison and reproducibility of ADC measurements in breathhold, respiratory triggered, and free-breathing diffusion-weighted MR imaging of the liver. *Journal of Magnetic Resonance Imaging*. 2008;28(5):1141-1148. doi:https://doi.org/10.1002/jmri.21569

98. Donati OF, Chong D, Nanz D, et al. Diffusion-weighted MR Imaging of Upper Abdominal Organs: Field Strength and Intervendor Variability of Apparent Diffusion Coefficients. *Radiology*. 2013;270(2):454-463. doi:10.1148/radiol.13130819

99. De Bazelaire CMJ, Duhamel GD, Rofsky NM, Alsop DC. MR Imaging Relaxation Times of Abdominal and Pelvic Tissues Measured in Vivo at 3.0 T: Preliminary Results. *Radiology*. 2004;230(3):652-659. doi:10.1148/radiol.2303021331

100. Liau J, Lee J, Schroeder ME, Sirlin CB, Bydder M. Cardiac motion in diffusion-weighted MRI of the liver: Artifact and a method of correction. *Journal of Magnetic Resonance Imaging*. 2012;35(2):318-327. doi:10.1002/jmri.22816

101. Kwee TC, Takahara T, Niwa T, et al. Influence of cardiac motion on diffusion-weighted magnetic resonance imaging of the liver. *Magnetic Resonance Materials in Physics, Biology and Medicine*. 2009;22(5):319-325. doi:10.1007/s10334-009-0183-1

102. Ozaki M, Inoue Y, Miyati T, et al. Motion artifact reduction of diffusion-weighted MRI of the liver: Use of velocity-compensated diffusion gradients combined with

tetrahedral gradients. *Journal of Magnetic Resonance Imaging*. 2013;37(1):172-178. doi:10.1002/jmri.23796

103. Aliotta E, Moulin K, Ennis DB. Eddy current–nulled convex optimized diffusion encoding (EN-CODE) for distortion-free diffusion tensor imaging with short echo times. *Magn Reson Med*. 2018;79(2):663-672. doi:10.1002/mrm.26709

104. Aliotta E, Wu HH, Ennis DB. Convex optimized diffusion encoding (CODE) gradient waveforms for minimum echo time and bulk motion–compensated diffusion-weighted MRI. *Magn Reson Med*. 2017;77(2):717-729. doi:10.1002/mrm.26166

105. Peña-Nogales Ó, Zhang Y, Wang X, et al. Optimized Diffusion-Weighting Gradient Waveform Design (ODGD) formulation for motion compensation and concomitant gradient nulling. *Magn Reson Med*. 2019;81(2):989-1003. doi:10.1002/mrm.27462

106. Zhang Y, Peña-Nogales Ó, Holmes JH, Hernando D. Motion-robust and blood-suppressed M1-optimized diffusion MR imaging of the liver. *Magn Reson Med*. 2019;82(1):302-311. doi:10.1002/mrm.27735

107. Ichikawa S, Motosugi U, Tamada D, et al. Improving the Quality of Diffusion-weighted Imaging of the Left Hepatic Lobe Using Weighted Averaging of Signals from Multiple Excitations. *Magnetic Resonance in Medical Sciences*. 2019;18(3):225-232. doi:10.2463/mrms.mp.2018-0085

108. Murphy P, Wolfson T, Gamst A, Sirlin C, Bydder M. Error model for reduction of cardiac and respiratory motion effects in quantitative liver DW-MRI. *Magn Reson Med*. 2013;70(5):1460-1469. doi:https://doi.org/10.1002/mrm.24563

109. Raspe JKJ, Van AT, Harder F, et al. Spatial Scaling of Respiratory-Triggered Liver Diffusion Weighted Imaging. In: *Proc. Intl. Soc. Mag. Reson. Med.* ; 2022.

110. Gadjimuradov F, Benkert T, Nickel MD, Führes T, Saake M, Maier A. Deep learning–guided weighted averaging for signal dropout compensation in DWI of the liver. *Magn Reson Med*. 2022;88(6):2679-2693. doi:https://doi.org/10.1002/mrm.29380

111. Führes T, Saake M, Lorenz J, et al. Reduction of the cardiac pulsation artifact and improvement of lesion conspicuity in flow-compensated diffusion images in the liver—A quantitative evaluation of postprocessing algorithms. *Magn Reson Med*. 2023;89(1):423-439. doi:10.1002/mrm.29427

112. Rauh SS, Riexinger AJ, Ohlmeyer S, et al. A mixed waveform protocol for reduction of the cardiac motion artifact in black-blood diffusion-weighted imaging of the liver. *Magn Reson Imaging*. 2020;67:59-68. doi:10.1016/j.mri.2019.12.011

113. Laun FB, Führes T, Seuss H, et al. Flow-compensated diffusion encoding in MRI for improved liver metastasis detection. *PLoS One*. 2022;17(5):e0268843-. https://doi.org/10.1371/journal.pone.0268843

114. Joe Zhou X, Tan SG, Bernstein MA. Artifacts Induced by Concomitant Magnetic Field in Fast Spin-Echo Imaging.; 1998.

115. Bernstein MA, Zhou J, Polzin JA, et al. Concomitant Gradient Terms in Phase Contrast MR: Analysis and Correction.; 1998.

116. Szczepankiewicz F, Westin CF, Nilsson M. Maxwell-compensated design of asymmetric gradient waveforms for tensor-valued diffusion encoding. *Magn Reson Med*. 2019;82(4):1424-1437. doi:10.1002/mrm.27828

117. Meier C, Zwanger M, Feiweier T, Porter D. Concomitant field terms for asymmetric gradient coils: Consequences for diffusion, flow, and echo-planar imaging. *Magn Reson Med*. 2008;60(1):128-134. doi:10.1002/mrm.21615

118. Baron CA, Lebel RM, Wilman AH, Beaulieu C. The effect of concomitant gradient fields on diffusion tensor imaging. *Magn Reson Med*. 2012;68(4):1190-1201. doi:https://doi.org/10.1002/mrm.24120

119. Norris DG. Implications of bulk motion for diffusion-weighted imaging experiments: Effects, mechanisms, and solutions. *Journal of Magnetic Resonance Imaging*. 2001;13(4):486-495. doi:10.1002/jmri.1072

120. Chen X, Qin L, Pan D, et al. Liver Diffusion-weighted MR Imaging: Reproducibility Comparison of ADC Measurements Obtained with Multiple Breath-hold, Free-breathing, Respiratory-triggered, and Navigator-triggered Techniques. *Radiology*. 2014;271(1):113-125. doi:10.1148/radiol.13131572

121. Geng R, Zhang Y, Starekova J, et al. Characterization and correction of cardiovascular motion artifacts in diffusion-weighted imaging of the pancreas. *Magn Reson Med*. 2021;86(4):1956-1969. doi:10.1002/mrm.28846

122. McTavish S, Van AT, Peeters JM, et al. Motion compensated renal diffusion weighted imaging. *Magn Reson Med*. 2023;89(1):144-160. doi:https://doi.org/10.1002/mrm.29433

123. Gadjimuradov F, Benkert T, Nickel MD, Führes T, Saake M, Maier A. Deep learning–guided weighted averaging for signal dropout compensation in DWI of the liver. *Magn Reson Med*. 2022;88(6):2679-2693. doi:10.1002/mrm.29380

124. Hubbell J, Seltzer S. Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients 1 keV to 20 MeV for Elements Z = 1 to 92 and 48 Additional Substances of Dosimetric Interest. Published 1995. http://physics.nist.gov/PhysRefData/XrayMassCoef/cover.html [online].

125. Multihance [package insert]. Monroe Township, NJ; Bracco Diagnostics Inc.; 2018.

Appendix A

T2w Reader Study Rubric

Table A1: T2w Reader Study Rubric

Score	SNR	Artifacts	Perceived Sharpness	Overall
5	Excellent	Excellent: no artifacts	Excellent: edges are sharp and distinct	Excellent: no artifacts and anatomical detail well visualized
4	Good	Good: minor artifacts, no impact on diagnostic capability	Good: edges seen clearly, slight blurriness, no impact on diagnostic capability	Good: minor artifacts, some blurriness, no impact on diagnostic capability
3	Fair	Fair: major or multiple minor artifacts, no impact on diagnostic capability	Fair: some blurring, loss of anatomical detail, no impact on diagnostic capability	Fair: major or multiple minor artifacts, blurriness, no impact on diagnostic capability
2	Poor	Poor: multiple major or minor artifacts, impact on diagnostic capability	Poor: blurring, loss of anatomical detail, impact on diagnostic capability	Poor: multiple major or minor artifacts, loss of detail, impact on diagnostic capability
1	Non-diagnostic	Non-diagnostic: image is unreadable	Non-diagnostic: features are blurred beyond recognition	Non-diagnostic: severe artifacts, and complete loss of anatomical detail

Minor artifact-> the artifact does not affect areas of interest, the artifact is easily read through, the artifact has only limited effect.

Major artifact-> the artifact obscures the area of interest; the artifact greatly impacts the visualization of tissues.

Impact on Diagnostic Capability-> The reader feels that the content of a section (SNR, artifact, etc.) is impacting their ability to confidently view or identify areas of interest in the image.

Appendix B

Additional results from automated breast volume prescription

Table B1: Scan Volume Placement Statistics for the GIoU-trained Model

Metric	5th %	Median	95th %	Average
3D IoU	0.43	0.65	0.82	0.64
Axial IoU	0.55	0.8	0.93	0.77
Sagittal IoU	0.46	0.65	0.85	0.65
Coronal IoU	0.52	0.73	0.9	0.72
Distance (cm)	1.1	3.2	7.2	3.7
Volume Error(%)	-28	13	76	17
Overlap (%)	60	84	100	83
RMSE (cm)	1.1	2.3	4.2	2.5

GIoU: Generalized Intersection Over Union, IoU: intersection over union, Distance: distance between the model predicted volume centers and the technologist prescribed volume center. RMSE: Root mean squared error between all 5 scan volume placement parameters. 5th and 95th % stand for the 5th and 95th percentiles of the distributions, respectively.

Table B2: Pre-Scan Volume Placement Statistics for the RMSE-trained Model

Parameter	Side	5th %	Median	95th %	Average
3D IoU	R	0.42	0.65	0.82	0.65
	L	0.39	0.65	0.82	0.65
Axial IoU	R	0.47	0.75	0.90	0.73
	L	0.57	0.78	0.92	0.76
Sagittal IoU	R	0.49	0.73	0.87	0.71
	L	0.52	0.73	0.89	0.71
Coronal IoU	R	0.51	0.73	0.88	0.72
	L	0.56	0.76	0.91	0.74
Distance (cm)	R	0.5	1.3	3.5	1.5
	L	0.5	1.3	3.3	1.5
Volume Error (%)	N/A	-29	6	69	13
RMSE (cm)	N/A	0.6	1.2	2.3	1.3

RMSE: Root Mean Squared Error, IoU: intersection over union, Distance: distance between the model predicted volume centers and the expert prescribed pre-scan volume centers, Side: Pre-scan volume placed on right (R) or left (L) breast, RMSE: Root mean squared error between all 9 pre-scan placement parameters. 5th and 95th % stand for the 5th and 95th percentiles of the distributions, respectively.

Appendix C

Calculation of Gadobenate Dimeglumine Attenuation

The attenuation values for the representative GBCA of Chapter Three, gadobenate dimeglumine, were estimated using the mixtures and compounds rule from Hubbell and Seltzer[124] and data from the United States Food and Drug Administration initial approval label[125]. The chemical formula of gadobenate dimeglumine is $C_{22}H_{28}GdN_3O_{11}\cdot 2C_7H_{17}NO_5$. It is supplied in an aqueous solution with a concentration of 529 mg/mL. The total density of the solution is 1.220 g/mL, meaning that 56.6% of the solution is water. From the chemical formula and the atomic masses of the constituent elements, the percent weight of each element for the gadobenate dimeglumine molecule and then the aqueous solution can be calculated. These percent weights are shown in **Table C1** along with the mass attenuation coefficient (μ/ρ) for each element obtained from Hubbell and Setzler[124]. To determine the mass attenuation for the solution, the individual mass attenuation coefficients are weighted by percent weight (w) and then added as described by Hubbell and Seltzer[124]:

$$\frac{\mu}{\rho} = \sum_i w_i \left(\frac{\mu}{\rho}\right)_i . \qquad [4]$$

To determine the linear attenuation coefficient (μ), the mass attenuation coefficient for the solution is multiplied by the density of the solution (ρ). **Table C2** shows the result of the calculation. Gadobenate dimeglumine at its concentration from the manufacturer has a linear attenuation coefficient between soft tissue and bone. However, immediately after administration

the contrast agent begins to dilute. The actual attenuation inside the body is expected to be much less.

Table C1: Mass Attenuation Coefficient of Gadobenate Dimeglumine

Material	% by weight	MAC @ 511 keV (cm^2/g)
Gd	6.44	1.12×10^{-1}
C	17.7	8.64×10^{-2}
H	2.56	1.71×10^{-1}
N	2.87	8.65×10^{-2}
O	13.8	8.06×10^{-2}
H_2O	56.6	9.61×10^{-2}
Gadobenate Dimeglumine	100	9.49×10^{-2}

Mass attenuation coefficient (MAC) of gadobenate dimeglumine is calculated through a linear combination of the MAC of its constituent elements. Each constituent MAC is multiplied by its percent weight of the total gadobenate dimeglumine solution and added to estimate the solution MAC.

Table C2: Attenuation Coefficients of Various Materials at 511 keV

Material	Mass Attenuation Coefficient (cm^2/g)	Density (g/cm^3)	Linear Attenuation Coefficient (cm^{-1})	Linear Attenuation Coefficient Relative to Water
Water	0.9687×10^{-1}	1.00	0.969×10^{-1}	1.00
Muscle	0.9518×10^{-1}	1.04	0.991×10^{-1}	1.03
Adipose Tissue	0.9616×10^{-1}	0.916	0.881×10^{-1}	0.84
Bone	0.8946×10^{-1}	1.65	1.47×10^{-1}	1.52
Elemental Gd	1.1679×10^{-1}	7.90	9.23×10^{-1}	9.53
Gadobenate dimeglumine	9.49×10^{-2}	1.22	1.16×10^{-1}	1.20

Attenuation coefficients and density for water, muscle, adipose tissue, bone, and elemental gadolinium (Gd) were obtained from Johns and Cunninghan[124]. Information for Gadobenate dimeglumine was from its United States Food and Drug Administration label[125].

Appendix D

GBCA effect with DM2 reconstruction

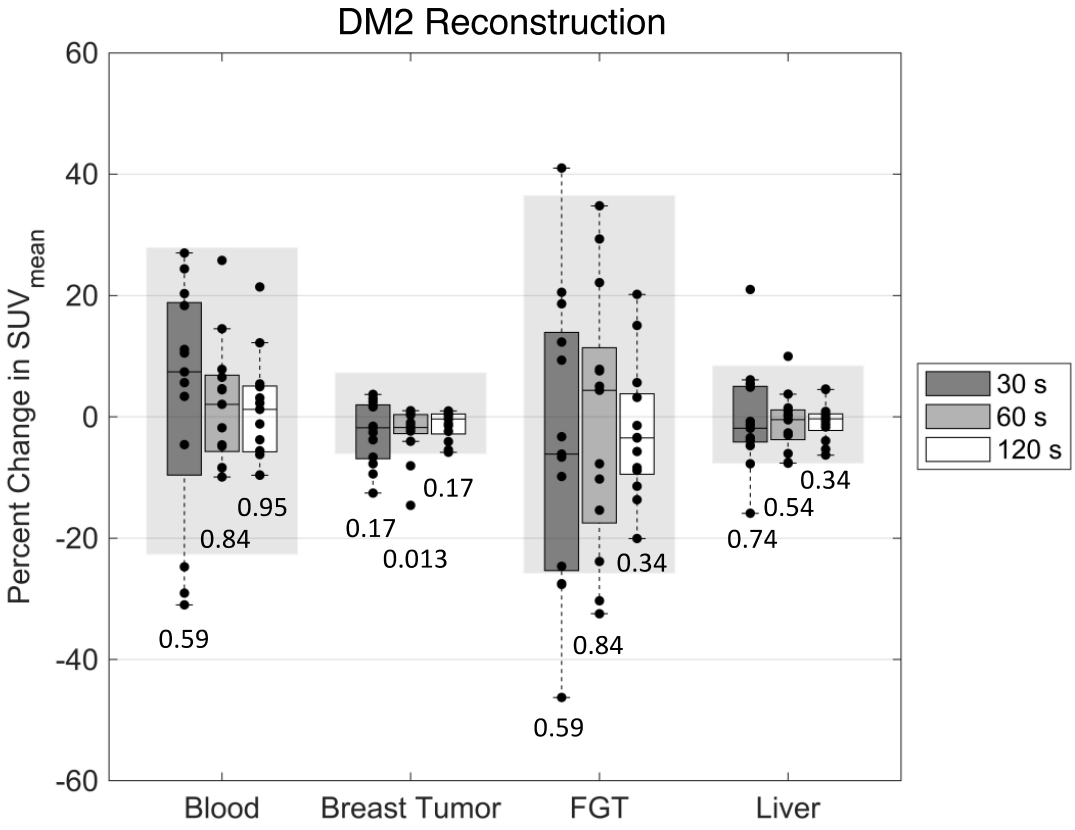

Figure D1: SUV_{mean} Before and After GBCA Administration Using DM2 Reconstruction. Percent change in SUV_{mean} for each of the measured VOIs, aorta/blood, breast tumor, normal fibroglandular tissue (FGT), and liver, for each timing interval using DM2 reconstruction. Format is the same as described in **Error! Reference source not found.**.

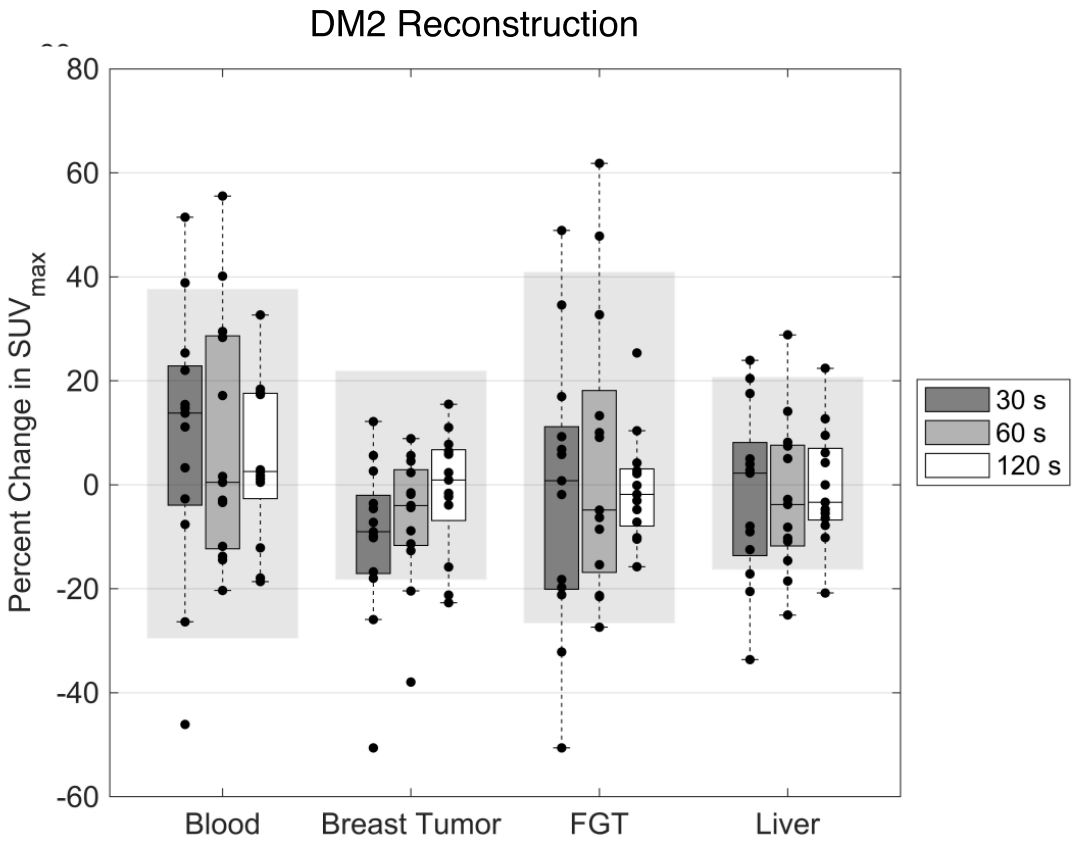

Figure D2: SUV_{max} Before and After GBCA Administration Using DM2 Reconstruction. Percent change in SUV_{max} for each of the measured VOIs, aorta/blood, breast tumor, normal fibroglandular tissue (FGT), and liver, for each timing interval using DM2 reconstruction. Format is the same as described in **Figure 23**.