Title
Automatic Reconstruction, Synthesis, and Processing of Musculoskeletal Magnetic Resonance Images Using Deep Learning

Author
Tolpadi, Aniket Anil

Publication Date
2023

Peer reviewed|Thesis/dissertation

Automatic Reconstruction, Synthesis, and Processing of Musculoskeletal Magnetic Resonance Images Using Deep Learning

by

Aniket Tolpadi

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
AND
UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:
Sharmila Majumdar
ED3F278527E84B0...

Sharmila Majumdar

Chair

DocuSigned by:
[signature]
DocuSigned by:

Miki Lustig

DocuSigned by:
Valentina Pedoia
DocuSigned by:

Valentina Pedoia

DocuSigned by:
[signature]
C02E7B87A580493...

Cynthia Chin

Committee Members

Acknowledgements

I am beyond grateful for all the amazing people that have been around me throughout my Ph.D. First and foremost, I am thankful for my advisor, Sharmila Majumdar, who welcomed a student with very limited background in machine learning and medical imaging, and was patient with me while I learned and developed the skillset and confidence to take on projects in image reconstruction and image synthesis. Her support, energy, and ideas left me leaving many meetings excited to take projects in slightly different directions, and her encouragement at many junctures such as writing my first manuscript, practicing my first few conference presentations, and putting together my qualifying exam proposal was very validating. I am thankful that she valued my input and scientific curiosity, and grateful that she supported me in traveling to conferences across the world to present my work.

To my other mentor, Valentina Pedoia: thank you for absolutely everything. Valentina's been an incredible sounding board for my ideas over the years—from things like architecture designs, image preprocessing strategies, and general perspective on medical imaging and machine learning literature. I owe a lot of my development as a scientist to Valentina's hands-on mentorship, open-door and open-message policies, and general love and advocacy for her students. Thank you for giving me so many opportunities to mentor students and taking on projects—I truly learned something from every such experience. I was blessed to have two amazing mentors: Sharmila and Valentina, I am grateful for your energy and that you made it fun to go to work every day. I could not imagine a Ph.D. experience without you two.

To my qualifying exam and dissertation committees, thank you for the valuable feedback and encouragement. My qualifying exam was the most difficult academic experience in my life,

and my committee was amazing through it all: thank you to Peder Larson for your wisdom and knowledge in answering my MR physics questions, and general feedback on my experimental design; to Grace O'Connell, whose spine mechanics questions were critical in helping me better understand and fine-tune my proposal; to Cynthia Chin, whose feedback from a radiologist's perspective as part of my qualifying and dissertation committees ensured the questions I tried answering through both works had clinical relevance. From my dissertation committee, special thanks to Miki Lustig, whose questions ensured I was not committing "implicit data crimes" with my MR reconstruction work, and that I better understood what patterns my models were drawing on to make predictions.

I'd also like to thank my lab-mates and other members at the UCSF Center for Intelligent Imaging for making this experience as enjoyable as it was. Thanks in particular to Rupsa Bhattacharjee, Francesco Calivà, Felix Gassert, Kenneth Gao, Madeline Hess, Claudia Iriondo, Johanna Luitjens, Alejandro Morales Martinez, Radhika Tibrewala and Michelle Tong. I enjoyed our conversations about science, and more generally, becoming friends with you all. Special thanks also to Misung Han, who patiently answered my many questions as I tried to better understand MR pulse sequences and the standard post-processing of MR images; I learned so much from her. Thanks to Jenny Lee, who was my rotation mentor when I joined the lab and was patient with me in those early days, even when I was writing Python scripts she described as "engineer's code." Thanks as well to other lab and CI^2 members that I have gotten to know over the years: Erin Argentieri, Bruno Astuto, Alex Beltran, Upasana Bharadwaj, Pablo Damasceno, Io Flament, Gabrielle Hoyer, Natalia Konovalova, Eugene Ozhinsky, Koren Roach, Rutwik Shah, and Rafeek Thahakoya. I am excited to see the amazing things you accomplish.

To the friends I have made in the UC-Berkeley/UCSF Graduate Program in Bioengineering, thank you for forming a community that made it a simple decision for me to do my Ph.D. here, and for providing a support structure that made it easy to forge ahead with my work, even when my science was not working. I'm especially thankful for the best friends I have made in this program—Justin Zhong and Vivien Tran. The ski trips, hiking trips, camping trips, and board game nights were fun and welcome reprieves from work. Thanks also to my roommate for most of my Ph.D., Ameesh Shah—it's amazing how much easier work is when you have a good living situation, and I'm grateful to have found that. I was blessed to find such a great group of people in the Bay Area with whom to go through my Ph.D.

To my college friends, I am happy I have been able to keep in touch with so many of you through ski trips and road trips. Those are some of my favorite weeks of the year, and I love getting updates about your lives during it. I hope we keep doing them as long as we can.

To my friends from high school, you guys are amazing—I am reminded every time I see you all that you are my best friends. Though they have become more sporadic with time, I enjoy our video chats and virtual game nights. The week that I see you all is always the best week of my year. Special thanks here to the chess squad—Conor McDonough, Kumar Muthukumar, and Arth Naik. Chess nights are my favorite nights of the week, and particularly during Covid-19, I owe a lot to those nights and other virtual game nights for keeping me sane and giving me things to look forward to in otherwise depressing times.

To my partner Preethi, I feel very fortunate to have met you. Thank you for being a sounding board when I need it and generally for making days more enjoyable. I am excited to see where we go from here.

And last, but certainly not least, thank you to my family for being so supportive and for keeping me grounded. My parents have always pushed me academically but have encouraged me to study things that I am most passionate about; I am grateful for their support and love. To my sister Anagha, a fun consequence of me doing my Ph.D. in the Bay Area is that we got to spend a lot more time together (even living together for 3 months during the pandemic) and became much closer. She is amazing. I love you all.

I was surrounded by so many amazing people that helped make my Ph.D. possible. Thank you all for everything.

Aniket Anil Tolpadi

Contributions

I must thank the coauthors of the works in this dissertation. Chapter 5 is adapted from an ISMRM abstract titled “A Cartilage-Specific Loss Function Improves Image Reconstruction Performance in Multiple Tissues of Clinical Interest,” which features work from Francesco Calivà, Misung Han, Emma Bahroos, Peder Larson, Sharmila Majumdar, and Valentina Pedoia. Chapter 6 is adapted from a manuscript titled “Region of Interest-Specific Loss Functions Improve T_2 Quantification with Ultrafast T_2 Mapping MRI Sequences in Knee, Hip and Lumbar Spine” featuring work from coauthors Misung Han, Francesco Calivà, Valentina Pedoia, and Sharmila Majumdar. Chapter 7 is adapted from a manuscript titled “K2S Challenge: From Undersampled K-Space to Automatic Segmentation,” which features work from coauthors Upasana Bharadwaj, Kenneth T. Gao, Rupsa Bhattacharjee, Felix G. Gassert, Johanna Luitjens, Paula Giesler, Jan Nikolas Morshuis, Paul Fischer, Matthias Hein, Christian F. Baumgartner, Artem Razumov, Dmitry Dylov, Quintin van Lohuizen, Stefan J. Fransen, Xiaoxia Zhang, Radhika Tibrewala, Hector Lise de Moura, Kangning Liu, Marcelo V.W. Zibetti, Ravinder Regatte, Sharmila Majumdar and Valentina Pedoia. Chapter 8 is adapted from a manuscript titled “Synthetic Inflammation Imaging with PatchGAN Deep Learning Networks,” with work from the following coauthors: Johanna Luitjens, Felix G. Gassert, Xiaojuan Li, Thomas Link, Sharmila Majumdar, and Valentina Pedoia. Lastly, chapter 9 is adapted from the manuscript “Deep Learning Predicts Total Knee Replacement from Magnetic Resonance Images,” with work from coauthors Jinhee Lee, Valentina Pedoia, and Sharmila Majumdar.

Automatic Reconstruction, Synthesis, and Processing of Musculoskeletal Magnetic Resonance Images Using Deep Learning

Aniket Tolpadi

Abstract

Musculoskeletal (MSK) diseases are widespread, with the World Health Organization estimating in 2019 that 1.71 billion people worldwide are afflicted with the condition [1]. MSK conditions include low back pain, knee osteoarthritis, and rheumatoid arthritis, among others, all of which induce debilitating pain and require early diagnosis to improve prognosis of treatment outcomes. Imaging is a crucial tool for diagnosis, and among available options, Magnetic Resonance Imaging (MRI) is a preferred modality for its sharp soft-tissue contrast, high-resolution images, and lack of ionizing radiation. However, acquisition and processing of MR images has numerous challenges: (1) acquisitions are time-consuming, and therefore expensive and susceptible to motion artifacts; (2) special sequences require toxic contrast agent administration, which have safety concerns; and (3) analysis of acquired images to identify patients most requiring clinical intervention is laborious. This work proposes using deep learning to address various aspects of these challenges. I will be presenting 5 applications and uses of deep learning algorithms:

1. To accelerate a 3D fat-suppressed knee MR sequence, showing that optimizing reconstruction algorithms for one tissue of clinical interest can improve its performance in other tissues of clinical interest.

2. For image reconstruction of accelerated compositional MR acquisitions in the knee, hip and lumbar spine, optimizing reconstructed images for tissues of heightened clinical interest (cartilage and intervertebral discs).
3. To automatically segment bone and cartilage from 8X accelerated knee MR acquisitions.
4. To synthesize post-contrast wrist MR images from pre-contrast scans in rheumatoid arthritis patients.
5. To predict if patients would require a total knee replacement within 5 years, using MR imaging and demographic variables.

Table of Contents

Chapter 1 - Overview 1

Chapter 2 - Relevant Musculoskeletal Anatomy 2

2.1 Knee 2

2.1.1 Knee Anatomy 2

2.1.2 Knee Pathophysiology 3

2.2 Hip 4

2.2.1 Hip Anatomy 4

2.2.2 Hip Pathophysiology 4

2.3 Lumbar Spine 5

2.3.1 Lumbar Spine Anatomy 5

2.3.2 Lumbar Spine Pathophysiology 6

2.4 Wrist 8

2.4.1 Wrist Anatomy 8

2.4.2 Wrist Pathophysiology 8

Chapter 3 - Clinical and Quantitative Imaging of Musculoskeletal Tissues 10

3.1 Basic MR Physics 10

3.2 Components of a MR Pulse Sequence 12

3.3 Standard Postprocessing of MR Acquisitions....14

3.4 Conventional MR Protocols....16

3.5 Compositional MR Imaging....18

Chapter 4 - Deep Learning in Medical Imaging....21

4.1 Basics of Deep Learning....21

4.2 Anomaly Detection and Prognosis Prediction....25

4.3 Image Segmentation....27

4.4 Image Synthesis....28

4.5 Image Reconstruction....29

Chapter 5 - A Cartilage-Specific Loss Function Improves Image Reconstruction Performance in Multiple Tissues of Clinical Interest....33

5.1 Introduction....33

5.2 Methods....34

5.2.1 Image Acquisition....34

5.2.2 Undersampling and Pre-Processing....34

5.2.3 Training....35

5.3 Results....36

5.4 Discussion and Conclusions....38

Chapter 6 - Region of Interest-Specific Loss Functions Improve T_2 Quantification with Ultrafast T_2 Mapping MRI Sequences in Knee, Hip and Lumbar Spine..... 40

6.1 Abstract 40

6.2 Introduction 41

6.3 Methods 46

6.3.1 MAPSS Acquisitions 46

6.3.2 T_2 Fitting and Spatial Undersampling 48

6.3.3 DL Pipeline Training 49

6.3.3.1 DL Architecture 49

6.3.3.2 Loss Function 51

6.3.3.3 Training and Segmentation Details 52

6.3.4 Experiments 54

6.3.4.1 Loss Function Ablation Study 54

6.3.4.2 Evaluation of Accelerated Acquisition Scheme Performance 55

6.3.4.2.1 Comparison of Global and ROI-Specific Performance 56

6.3.4.2.2 Standard Reconstruction Metrics 56

6.3.4.2.3 T_2 Value Retention 56

6.3.4.2.4 Texture Retention 57

6.3.4.3 Repeatability Study 57

6.3.4.4 Raw Multicoil Data Assessment 58

6.4 Results 58

6.4.1 Ablation Study Results 58

6.4.2 Visuals of Network Performance and Comparison with State-of-the-Art Models 59

6.4.3 Evaluation of T_2 Quantification Performance and Comparison with State-of-the-Art Models 65

6.4.3.1 Voxel-wise T_2 Evaluation Fidelity 65

6.4.3.2 T_2 Value Retention on Region of Interest Averages 68

6.4.3.3 Texture Retention 69

6.4.4 Repeatability Study 71

6.4.5 Raw Multicoil Data Assessment 71

6.5 Discussion and Conclusions 72

Chapter 7 - K2S Challenge: From Undersampled K-Space to Automatic Segmentation 78

7.1 Abstract 78

7.2 Introduction 79

7.3 Methods 83

7.3.1 Challenge 83

7.3.2 Dataset 85

7.3.2.1 Subject Eligibility and Sequence Information 85

7.3.2.2 Extraction of ARC-Reconstructed Multicoil Raw k-Space Data 86

7.3.2.2.1 k-Space Post-Processing 87

7.3.2.2.2 Image Space Post-Processing 88

7.3.2.3 Ground Truth Segmentation Generation 89

7.3.2.3.1 Cartilage Segmentation Pipeline....89
7.3.2.3.2 Bone Segmentation Pipeline....90
7.3.2.4 Selection of Cases for K2S Dataset....91
7.3.2.5 Final K2S Dataset Characteristics....93
7.3.3 Evaluation Process....94
7.3.4 Timeline....94
7.3.5 Overview of Top Submission Methodologies....95
7.3.5.1 K-nirsh (University of Tübingen, Tübingen, Germany)....95
7.3.5.2 UglyBarnacle (Skolovo Institute of Science and Technology, Moscow, Russia)....96
7.3.5.3 FastMRI-AI (University Medical Center Groningen, Groningen, Netherlands)....97
7.3.5.4 NYU-Knee AI (NYU Grossman School of Medicine, New York, USA)....97
7.3.6 Further Analysis of Submissions....98
7.3.6.1 Intermediate Pipeline Reconstruction Performance....98
7.3.6.2 Comparison of Reconstruction and Segmentation Performance....99
7.3.6.3 Biomarker Analysis: Cartilage Thickness....99
7.3.6.4 Biomarker Analysis: Bone Shape....100
7.4 Results....100
7.4.1 Segmentation Metrics....101
7.4.2 Reconstruction Metrics....101
7.4.3 Comparison of Reconstruction and Segmentation Performance....103
7.4.4 Biomarker Analysis: Cartilage Thickness....105
7.4.5 Biomarker Analysis: Bone Shape....108

7.5 Discussion and Conclusions 110

Chapter 8 - Synthetic Inflammation Imaging with PatchGAN Deep Learning Networks..... 116

8.1 Abstract 116

8.2 Introduction 117

8.3 Methods 123

8.3.1 Study Group 123

8.3.2 MR Acquisition 124

8.3.3 Anomaly Segmentations and Evaluations 124

8.3.4 Image Preprocessing 125

8.3.5 Data Partitioning 126

8.3.6 Network Architecture 127

8.3.7 Training Details 128

8.3.8 Hyperparameter Search and Model Selection 130

8.3.9 Model Performance Evaluation 131

8.3.10 Enhancement Maps 132

8.3.11 Occlusion Maps 133

8.3.12 Uncertainty Maps 133

8.3.13 Statistical Analysis 134

8.4 Results 134

8.4.1 Model Parameter Selection 134

8.4.2 Utility of Deconvolution Operators in UNet Decoders 135

8.4.3 Standard Reconstruction Metrics Performance 138

8.4.4 Comparison of Reconstruction Performance Across Synovitis Severity 139

8.4.5 Enhancement Maps Analysis 141

8.4.6 Occlusion and Uncertainty Maps Analysis 141

8.5 Discussion 144

8.6 Conclusions 149

Chapter 9 - Deep Learning Predicts Total Knee Replacement from MR Images 151

9.1 Abstract 151

9.2 Introduction 152

9.3 Novelty 156

9.4 Methods 157

9.4.1 Data 157

9.4.2 Pipeline Architecture 160

9.4.3 Training 161

9.4.4 Integration of Imaging and Non-Imaging Data 162

9.4.5 Statistical Analysis 163

9.4.6 Imaging Biomarker Identification 164

9.5 Results 165

9.5.1 OA Pretrain Utility in TKR Prediction 165

9.5.2 X-Ray Pipeline Optimization and Performance 165

9.5.3 MRI Pipeline Optimization and Performance 168

9.5.4 Comparison of MRI and Radiograph Pipeline Performances.....169

9.5.5 Biomarker Identification and Analysis 172

9.6 Discussion and Conclusions 175

References 180

Appendix A - Supplementary Information to Chapter 6 229

Appendix B - Supplementary Information to Chapter 8 255

Appendix C - Supplementary Information to Chapter 9..... 261

List of Figures

Figure 5.1 KIKI-Net inspired architecture for reconstruction 36

Figure 5.2 R=4 pipeline performances.....37

Figure 5.3 R=8 pipeline performances.....38

Figure 6.1 Proposed MAPSS Acceleration Pipeline.....50

Figure 6.2 Comparison of predicted T_2 maps with ROI-specific methodologies to past approaches 62

Figure 6.3 T_2 quantification performance in knee cartilage 62

Figure 6.4 T_2 quantification performance in hip cartilage 63

Figure 6.5 T_2 quantification performance in lumbar spine intervertebral discs 64

Figure 7.1 K2S challenge schematic.....84

Figure 7.2 k-Space and image space post-processing steps 86

Figure 7.3 Intermediate outputs within post-processing pipeline 87

Figure 7.4 1–5 LIKERT cartilage segmentation scores overlaid on ground truth knee scans.....92

Figure 7.5 1–5 LIKERT bone segmentation scores overlaid on ground truth knee scans 93

Figure 7.6 Intermediate pipeline reconstruction outputs for top 4 submissions 102

Figure 7.7 Sagittal slice segmentations overlaid on intermediate pipeline reconstructions 104

Figure 7.8 Reconstruction metrics plotted against weighted DSC for top submissions.....105

Figure 7.9 Femoral cartilage thickness maps projected onto femur 106

Figure 7.10 Bland-Altman and correlation plots between predicted and ground truth cartilage thicknesses.....107

Figure 7.11 Femoral bone shape feature visualizations 109

Figure 7.12 Correlations along femoral, tibial, and patellar bone shape features between submissions and ground truth.....110

Figure 8.1 Network Architectures 128

Figure 8.2 Network Performance with and without Deconvolutions for UNets.....136

Figure 8.3 Network Performance with and without Deconvolutions for PatchGANs.....137

Figure 8.4 Visual Comparison of Reconstructed Post-Gadolinium Images with and without Imaging Findings of RA.....140

Figure 8.5 Predicted Enhancement Maps with PatchGAN, UNet, and Ground Truth.....142

Figure 8.6 Occlusion Maps for PatchGAN and UNet Pipelines 143

Figure 8.7 Uncertainty Maps for PatchGAN and UNet Pipelines.....144

Figure 9.1 Pipeline predicting if patient will undergo TKR within 5 years 156

Figure 9.2 ROC curves for X-ray and MRI architectures on test data.....167

Figure 9.3 ROC curves for MRI and X-ray pipelines at select OA classifications 171

Figure 9.4 MRI pipeline true positive occlusion map slices.....173

Supp. Figure A.1 Accelerated MAPSS reconstruction network architecture.....246

Supp. Figure A.2 Modified sigmoid function for knee, hip and lumbar spine pipelines 247

Supp. Figure A.3 Predicted T_2 maps for proposed models and equivalent pipelines trained without ROI-specific loss 249

Supp. Figure A.4 Global T_2 retention for proposed pipelines and state-of-the-art models 251

Supp. Figure A.5 Comparison of biases in predicted T_2 maps in knee cartilage, hip cartilage, and intervertebral discs 252

Supp. Figure A.6 Multicoil data assessment of pipeline performance.....254

Supp. Figure B.1 Example Registrations 258

Supp. Figure B.2 Example Anomaly Distance Map.....259

Supp. Figure B.3 Sample Hyperparameter Search Slices.....260

Supp. Figure C.1 Sample slices of DESS MRI and compressed slice versions 264

List of Tables

Table 5.1 Model performance metrics in test set....36

Table 6.1 Knee, hip and lumbar spine datasets and splits....46

Table 6.2 ROI-specific model performance in standard metrics from R=2 through R=12 66

Table 6.3 Texture retention analysis in No RNN and Full Model pipelines 69

Table 7.1 Acquisition parameters for 3D fat-suppressed CUBE sequence used in K2S 85

Table 7.2 Segmentation performance across test set (n = 50) for top 4 pipelines 101

Table 7.3 Standard reconstruction metrics for intermediate pipeline outputs 103

Table 8.1 Full Dataset and Splits Information....126

Table 8.2 Coronal IDEAL Post-Gd T_1 Image Synthesis Performance for Select Pipelines 138

Table 9.1 List of non-imaging variables fed into logistic regression models to predict TKR 159

Table 9.2 Data used to train 3D DESS MRI and 2D radiograph architectures 160

Table 9.3 Performance in TKR prediction of OA pretrained models....165

Table 9.4 Performance of X-ray and MRI architectures on test data....167

Table 9.5 Summary of occlusion map analysis comparing frequencies with which selected knee joint tissues were indicated as hotspots....174

Supp. Table A.1 Acquisition times for MAPSS at tested R....229

Supp. Table A.2 Information for cross-validation splits in knee, hip and lumbar spine 229

Supp. Table A.3 Ablation study T_2 quantification error rates....230

Supp. Table A.4 T_2 value equivalents of ablation study quantification error rates 231

Supp. Table A.5 ROI and global correlations between predicted and ground truth maps....232

Supp. Table A.6 T_2 value equivalents of quantification errors in cartilage, IVDs....233

Supp. Table A.7 T_2 quantification errors in knee cartilage compartments.....234

Supp. Table A.8 Correlation between predicted and ground truth T_2 in knee cartilage compartments.....235

Supp. Table A.9 T_2 quantification errors in hip cartilage compartments.....236

Supp. Table A.10 Correlation between predicted and ground truth T_2 in hip cartilage compartments.....237

Supp. Table A.11 T_2 quantification errors in lumbar spine IVD levels.....238

Supp. Table A.12 Correlation between predicted and ground truth T_2 in lumbar spine IVDs....239

Supp. Table A.13 Optimized loss function weightings for best pipelines.....241

Supp. Table A.14 T_2 quantification error rates across 3 splits in tissues of interest.....242

Supp. Table A.15 Correlations between predicted and ground truth T_2 maps across 3 splits in tissues of interest.....243

Supp. Table A.16 Texture retention performance of knee, hip and lumbar spine pipelines across 3 splits in tissues of interest.....244

Supp. Table B.1 MR Acquisition Parameters.....255

Supp. Table B.2 SSIM for full volumes and wrist tissue from hyperparameter search.....255

Supp. Table B.3 nRMSEs for full volumes and wrist tissue from hyperparameter search.....256

Supp. Table B.4 Reconstruction metrics with and without Imaging Findings of RA.....256

Supp. Table C.1 Non-imaging variables identified from literature as correlated with OA progression or eventual TKR.....261

Supp. Table C.2 Percentages of selected tissues identified as hotspots in true positive.....262

Supp. Table C.3 Percentages of selected tissues identified as hotspots in true negatives.....263

Chapter 1 - Overview

Chapter 2 will serve as a basic introduction to the musculoskeletal (MSK) joints and tissues that will be further examined in the thesis. Next, chapter 3 will introduce Magnetic Resonance Imaging (MRI), an essential imaging modality commonly used to better understand these tissues, while chapter 4 will provide background on deep learning (DL) and how it can be applied to various aspects of the clinical imaging workflow. Finally, chapters 5-9 are self-contained studies, complete with relevant background, methods, results, and conclusions, each of which details some application of deep learning to streamline, accelerate, or automate some portion of the imaging workflow.

Chapter 2 - Relevant Musculoskeletal Anatomy

The MSK system is a complex framework consisting of various tissue types, responsible for crucial tasks that include load bearing, posture maintenance, and facilitating locomotion. Hierarchically, the organ system is stratified into numerous joints and sub-anatomies such the ankle and shoulder, each of which can be broken down into components such as cartilage, bones, muscles, and ligaments. This thesis will center around four such anatomies: knee, hip, lumbar spine and wrist.

2.1 Knee

2.1.1 Knee Anatomy

The knee is a modified hinge joint allowing for sagittal flexion and extension, and varus and valgus rotation in the frontal plane [2]. As one of the largest joints in the human body, it has substantial weight-bearing responsibilities, carried out through a complex configuration of bones, ligaments, cartilage, and muscles. Four bones are observed within the knee: the femur, tibia, fibula, and patella. Among these, the tibiofemoral joint is particularly important in weight-bearing, whereas the patellofemoral joint is responsible for frictionless transfer of flexion and extension forces about the knee, facilitating motility. These muscles are stabilized primarily by a series of ligaments, and secondarily by surrounding muscles. Ligaments connect bone to bone, and within the knee, the anterior cruciate ligament (ACL) is the most important stabilizer, responsible for up to 85% of the joint's stability [3]. In conjunction with the posterior cruciate ligament (PCL), the ACL prevents anteroposterior motion of the femur with respect to the tibia, whereas the medical collateral ligament (MCL) and lateral collateral ligament (LCL) prevent

mediolateral relative motion between the bones. Other relevant stabilizing ligaments include the popliteal ligament. To withstand compressive and shear forces observed in weight-bearing, flexion, extension, varus rotation, and valgus rotation, the articulating surfaces of the tibiofemoral and patellofemoral joints are lined with hyaline cartilage and encased in a fibrous, synovial fluid filled capsule. Cartilage is a well-hydrated, collagen-rich tissue that reduces friction and acts as a shock absorber for the entire knee. Moreover, the tibiofemoral joint is also equipped with medical and lateral fibrocartilaginous structures known as menisci. Like cartilage, menisci also function as a shock absorber, while also designed to prevent excessive varus or valgus rotation [4].

2.1.2 Knee Pathophysiology

Among the most common pathophysiological knee anomalies is osteoarthritis (OA), for which incidence rate estimates range from 14 to 30 million in the United States alone [5,6]. In OA, the knee undergoes structural changes that may include cartilage loss, alterations in subchondral bone properties, and narrowing of the tibiofemoral joint space, among others [7]. Ultimately, these changes can cause inflammation, debilitating pain, and generally reduced quality of life. Unfortunately, however, knee OA is irreversible: while treatments exist for early-stage OA that can mitigate symptoms, late-stage OA has no noninvasive treatments, underscoring the importance of regular monitoring of joint health for early OA identification and treatment initiation [8,9]. Other anomalies that can occur as a precursor, concurrently, or due to OA include cartilage and meniscal lesions, which can compromise the ability of both tissues to withstand compressive forces and generally induce pain [10]. Aside from these lesions, bone

marrow edema is also common, in which fluid collects in extracellular marrow spaces, possibly due to trauma, infection, or cancer [11,12]. Diagnosis of these anomalies is typically done with a combination of medical imaging and monitoring symptoms, necessitating automatic tools to track patient health and identify at-risk populations for these conditions.

2.2 Hip

2.2.1 Hip Anatomy

Like the knee, the hip also carries substantial weight-bearing responsibilities, but differs in that it is a ball-and-socket joint rather than a hinge joint. Three bones—the ilium, ischium, and pubis—intersect to form the "socket" of the hip joint, known as the acetabulum [13]. The proximal head of the femur forms a ball that inserts into this socket, allowing for flexion, extension, abduction, adduction, and internal and external rotation. The joint is stabilized by a series of ligaments, the most significant of which are the iliofemoral, pubofemoral, and ischiofemoral, all of which thicken the hip joint capsule and limit internal rotation, abduction and extension, and extension, respectively [13]. The femoral head and acetabulum are both lined with cartilage, which has similar roles in the hip as in the knee: resist compressive forces and reduce friction associated with motion. Also as in the knee, the hip is a synovial joint, contained in a fibrous sac and lubricated with synovial fluid [14].

2.2.2 Hip Pathophysiology

OA is a similarly common pathology in the hip as in the knee. In the hip, chronic usage and general wear-and-tear of the hip joint can see damage to femoral and acetabular articular

cartilage, the formation of bony osteophytes that induce pain, and additional damage to stabilizing ligaments and muscles [15]. Hip OA is widespread: in the United Kingdom, 10-25% of those older than 55 suffer from the condition, and its effects are not solely limited to the elderly population [16]. Also very commonly observed are hip fractures, observed in one (or more) of the bones constituting the hip joint. Due to poor blood supply, healing from hip fractures is often slow and affliction with the condition very painful, with potentially dire consequences: 5-10% of those with hip fracture die within one month of the fracture date [17]. While OA and fractures are both ongoing areas of research, particularly related to their detection from medical images, this thesis will focus more so on tissues implicated in hip OA.

2.3 Lumbar Spine

2.3.1 Lumbar Spine Anatomy

The lumbar spine is a crucial structure in the lower back with numerous responsibilities, including protecting the spinal cord and the numerous nerves that emerge from it, and providing structural support for the spinal column and torso. The vertebral column features alternating intervertebral discs (IVDs) and vertebral body bones. IVDs are avascular tissues lying between adjacent vertebrae, and are responsible for resisting spinal compression while absorbing axial and torsional stresses that may be imparted on the spine. These functions are accomplished due to the IVD structure, which can be stratified into three regions: the nucleus pulposus (NP), annulus fibrosis (AF), and cartilaginous end plates (CEP). The NP is centrally located within the IVD, rich with proteoglycans that have negatively charged side chains, allowing the region to be well hydrated and deform reversibly, thereby resisting axial loads. The

AF is radially located, consisting of 10-20 concentric rings of “lamellae” composed primarily of collagen I, with each concentric ring having fibers oriented roughly 60-65 degrees from the vertical in an alternating fashion [18]. This structure allows the AF to resist tensile stresses exerted by the radial pressure of the NP. CEPs lie at the superior and inferior ends of the IVD, composed of hyaline cartilage at early stages of development and fibrocartilage in adults, and is crucial in delivering nutrients to the disc [19]. Vertebral bodies consist of anteriorly located bodies that sit on top of IVDs, and several processes located posterior to the discs: the transverse processes protrude laterally, limiting left/right rotation of the spine; the spinous process protrudes posteriorly, limiting anteroposterior motion of the spine. Each vertebra also consists of two superior and inferior articular processes, allowing adjacent vertebra to sit on top of one another. The spinal cord itself passes through the central canals of vertebrae, whereas adjacent vertebrae form foramen on the left and right sides of the spinal column at each vertebral level, through which nerves pass that innervate various muscles (primarily in the lower body for the lumbar spine). The spinal column is stabilized by numerous paraspinal muscles [20–22].

2.3.2 Lumbar Spine Pathophysiology

The overarching anomaly observed in the lumbar spine is low back pain (LBP), which is the fifth most common reason why Americans seek medical care and is the nation’s second leading cause of disability [23]. A precise cause for LBP can be difficult to identify, as 90 percent of all LBP cases are non-specific [24]: causes can be anatomic or psychosocial. In anatomic cases, degeneration of IVDs is one possibility: NP proteoglycan side chains can degrade or shift from

chondroitin sulfate to keratin sulfate [25], decreasing the relative proteoglycan content within the NPs, reducing their ability to stay hydrated, thereby reducing its ability to resist axial loads [26]. AF changes with aging and trauma can include cracking or tears in the lamellae, allowing the NP to protrude through the AF and breaking down the barrier between the NP and AF in the tissue. This further reduces the disc's ability to resist axial compression and torsion and may also lead to a herniated disc that impinges on the spinal cord or other nerves, causing pain [27,28]. The lumbar spine work in this thesis will focus primarily on the IVDs, but various other anomalies can arise within the lumbar spine that cause pain. For instance, lesions can emerge between vertebral and cartilaginous endplates at the intersection of IVDs and vertebral bodies, possibly causing the bony marrow to inflame and convert into fat, while the end plates can undergo some degree of ossification; these are known as Modic changes [29,30]. Elsewhere, the central canal of the spinal cord or the foramen can constrict for various reasons, impinging on the spinal cord and/or foraminal nerves, causing pain that can radiculate to regions innervated by the nerves [31,32]. Fractures can also be observed within the vertebral bodies, altering the load bearing mechanics of the spine while also causing pain [33]. Other sources of pain can include misalignment of the spine, which is observed in scoliosis, lordosis and kyphosis [34–36]. Beyond anatomic factors, psychosocial factors such as depression, stress, and reliving pain episodes can also induce pain [37].

2.4 Wrist

2.4.1 Wrist Anatomy

The wrist joint is one of the most complex in the human body, encompassing a wide assortment of relatively small bones. Proximally, the wrist joint begins at the forearms, where two bones are found: the medially located ulna, and the laterally located radius. Moving distally into the wrist is the distal carpal row (DCR) of bones: the trapezium, trapezoid, capitate and hamate [38]. Proximal to the DCR is the proximal carpal row (PCR) bones: the scaphoid, lunate, triquetrum and pisiform [39]. The PCR then attaches to the metacarpal bones of the hand by strong ligaments, causing the DCR to function fundamentally as a single unit. Relative motion of these carpal bones is restricted by a complex system of 33 intraarticular and intracapsular ligaments [40]. Six total forearm extensor and flexor muscles attach distally within the wrist into the DCR through a network of tendons, allowing for wrist motion [41].

2.4.2 Wrist Pathophysiology

Numerous anomalies can occur in the wrist, including fractures of the many bones and ligament tears, all of which can cause pain and joint instability [42]. While not a weight-bearing joint, the wrist can also become afflicted with OA, the causes of which are unknown but likely involve biomechanical factors, as well as biochemical factors such as proteinases and proinflammatory cytokines [43]. The wrist-related work in this thesis, however, will focus on Rheumatoid Arthritis (RA), a widespread autoimmune disorder observed in 0.5-1.0% of Americans, with an incidence rate in women that is 2-3 times higher than that of men [44]. RA is systemic, mainly affecting joints (particularly the feet and hands), and is characterized by synovial joint inflammation,

bone tissue erosion and soft tissue breakdown [45]. In addition to synovial inflammation, bone marrow edema (BME) can also result from RA, all of which can cause pain and severely degrade quality of life [46]. RA is typically treated using Disease-Modifying Anti-Rheumatic Drugs (DMARDs), which see 75-80% of patients attain intended treatment outcomes within a year of treatment initiation. However, this spikes to 90% when treatments are initiated in early RA stages, underscoring the importance of early diagnosis [47].

Chapter 3 - Clinical and Quantitative Imaging of Musculoskeletal Tissues

Diagnosing anomalies such as knee OA, hip OA, and wrist RA require holistic assessment of symptoms such as pain and quality of life, and in the case of wrist RA, laboratory tests for C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR). However, central to diagnostic schemes for these conditions and lumbar spine anomalies, is medical imaging. Imaging provides a noninvasive means of depicting anomalies and can help identify potential pain sources and guide treatment courses. Widely used clinical imaging modalities include X-ray, Computed Tomography (CT), Ultrasound (US), and Magnetic Resonance Imaging (MRI). Compared to its counterparts, MR has advantages with its sharp contrast, exquisite depiction of soft tissue, ability to image three-dimensional volumes, and lack of ionizing radiation [48]. Its primary drawbacks, however, are long acquisition times, high costs, and limited accessibility in low-resource regions. Much of the work in this thesis will center around addressing MR drawbacks with an eye towards its faster, cheaper, and more widespread clinical use. This chapter will begin by discussing the fundamentals of MR imaging, after which standard clinical MR sequences and contrast mechanisms will be described. To conclude, compositional MR will be introduced, and the advantages it offers over conventional approaches.

3.1 Basic MR Physics

MR entails imaging one of several biologically active atomic nuclei with nonzero nuclear magnetic moments. Nuclear magnetic moments form in an atomic nucleus: for a nucleus with an equal number of protons and neutrons, nuclear charges can be distributed roughly evenly, causing the nucleus to have a nuclear magnetic moment of essentially zero [49]. Alternately,

nuclei with unequal numbers of protons and neutrons necessarily will have at least one "unpaired" proton or neutron, impairing the atom's ability to evenly distribute nuclear charges and causing it to have a nuclear magnetic moment [50]. Such atoms, such as ^{1}H, ^{23}Na, and ^{31}P, therefore are susceptible to extrinsic magnetic fields [51]. While not detectable for a single nucleus, these nuclei behave as small magnetic dipoles that, when gathered in sufficiently large numbers (~10^{15}), yield a nuclei conglomerate that generates an observable magnetic signal, forming the basis of MR signal [52]. Due to its abundance in the human body, ^{1}H MR is by far the most widely used nucleus for MR, and is the basis of both clinical MR imaging and the imaging analyzed in this thesis.

In the absence of an external magnetic field, these dipoles are randomly oriented, essentially cancelling one another out and not generating net magnetization. In the presence of an external magnetic field, however, some alignment of the dipoles is observed: most along the direction of the magnetic field, but some in a high-energy state, antiparallel to the external field, generating net magnetization [53]. In MR imaging, this alignment is induced with the B_0 field, a 1.5 or 3 Tesla (T) field for most current clinical imaging applications, although higher strength B_0 fields are being investigated in research settings. Crucially, when placed in an external B_0 field, these nuclei begin precessing about the B_0 field at a specific frequency known as the Larmor frequency ($\omega_0 = B_0 \times \gamma$, where ω_0 is the Larmor frequency, B_0 is the strength of the external B_0 field, and γ is the gyromagnetic ratio, or a constant for every nucleus at a given field strength) [54]. Control of this frequency, and the location of spins within their precession cycles ("phase") are crucial to localizing MR signals.

3.2 Components of a MR Pulse Sequence

In scanners, patients are subjected to strong B_0 fields, inducing net magnetization along the longitudinal axis (along the B_0 field direction) [55]. To initiate imaging, a radiofrequency (RF) pulse is applied at frequency ω_0 perpendicular to the B_0 field axis; this applies a torque that "tips" the net magnetization away from the B_0 axis by an angle determined by pulse strength and duration ("flip angle") [56]. This separates the net magnetization into longitudinal (along the B_0 axis) and transverse (perpendicular to the B_0 axis) components. Once the RF pulse concludes, net magnetization begins recovering along the longitudinal axis ("longitudinal recovery") and decaying along the transverse axis ("transverse decay") [53]. Due to the heterogeneity in intrinsic macromolecular properties such as free and bound water content across tissues, longitudinal recovery and transverse decay occur at different rates across a tissue sample, which are exploited to produce MR images [57]. Also due to this heterogeneity, however, the phase coherence observed in nuclei immediately after an RF pulse is quickly lost and must be reestablished during signal acquisition. Typically, phase coherence is induced after excitation RF pulses by using 180° refocusing RF pulses or additional gradients [58]. Receiving coils then measure signal in the transverse plane. In modern MR, several coils are usually used and located around the periphery of the anatomy being imaged; the use of multiple coils allows for considerably better spatial sensitivity to regions throughout the imaged volume than a single coil setup. Receiving coils are tuned to the Larmor frequency and acquire the coherent signal. The time between the initial RF pulse and signal acquisition is known as echo time (TE), whereas the time between successive non-refocusing RF pulses is known as repetition time (TR).

Signal is generated and recorded using these mechanisms but must also be localized in the tissue. Receiving coils can be tuned a particular frequency and phase of spins within their precession. As such, if signal from only a particular voxel is desired, its frequency and phase must be modulated such that receiving coils can be tuned to isolate the voxel. Gradient coils accomplish this frequency and phase modulation: most MR scanners are equipped with two gradient coils along the transverse axes and one along the longitudinal axis. For 2D MR imaging, a slice-selecting gradient coil is activated simultaneously with the excitatory RF pulse ("slice-selecting pulse"); when the slice-selecting gradient coil is activated, it causes slight local variations in magnetic field strength (from the original B_0), thereby locally altering the Larmor frequency. When applied, the RF pulse is tuned to the desired frequency, thereby only "tipping" longitudinal magnetization into the transverse plane for a slice of tissue. After excitation, phase is encoded along an orthogonal axis by the phase encoding gradient coil: when applied, the local alterations in Larmor frequency cause nuclei along the phase-encoding direction to reach slightly differing starting points in their precession cycles. When the phase-encoding gradient is turned off, the Larmor frequency returns to its original frequency, but due to the phase encoding, nuclei along the phase-encoding axis now lie at different phases within their precession cycles despite subsequently precessing at the same frequency. This accounts for localization in two dimensions, and the third is accomplished with the frequency encoding gradient coil. The frequency encoding gradient is applied while signal is read by the receiving coils: just as the previous gradient coils, it induces local variations in the frequency of nuclei precession by modulating the Larmor frequency, thereby allowing for localization of signal along

the remaining axis [53]. In 3D imaging, an entire volume of tissue is instead excited rather than a slice, with the slice-selecting gradient coil instead used as a second phase-encoding gradient coil [59].

A standard MR sequence has several components: an RF pulse to excite tissue, refocusing pulses or gradients to induce spatial coherence while signal is recorded, and a series of gradient coils to localize signal to given voxels. Numerous parameters can affect resulting image appearance, such as gradient coil strength, RF pulse strength and duration, and the bandwidth of frequencies accepted by receiving coils during acquisition. Additional techniques can be added, for example, to suppress or saturate signal from tissues such as fat or fluid [60–62]. The most important parameters, however, are TR and TE: by controlling RF pulse and/or refocusing gradient timing, the degree of longitudinal recovery and transverse decay undergone by a tissue prior to signal acquisition can be adjusted. Strategic selection of TR and TE therefore controls which properties are accentuated, and thus, the structures a resulting image will illuminate.

3.3 Standard Postprocessing of MR Acquisitions

An underreported aspect of the MR acquisition pipeline are the many steps involved in postprocessing MR scanner data to arrive at images ultimately used in the clinic. First off, MR images are acquired in the frequency domain, which is also described as k-space. A series of postprocessing steps must be performed in k-space and image space to attain desired images.

In k-space, standard clinical sequences may be accelerated using parallel imaging (PI)— briefly, it exploits the redundancy in acquiring signal in multiple receiving coils to acquire fewer less data in k-space, thereby reducing acquisition time. Some PI methods will impute unacquired k-space points, and if so, they must be imputed in an initial step [63]. Subsequently, many MR acquisitions will acquire data using a given size of k-space (the "acquisition matrix"), but the intended dimensions of the output resolution could be larger to increase the sequence's apparent spatial resolution. To accomplish this, images are zero-padded in k-space to the desired dimensions. When transformed into Cartesian space (also called "image space," or the domain in which medical images are viewed by clinicians), the sharp boundaries between nonzero acquired k-space points and the peripheral zeros can induce undesired Gibbs artifacts. This is mitigated by filtering the k-space, often with a Fermi or Hamming filter, that softens the boundary between nonzero and zero points within k-space, mitigating the Gibbs artifacts at the expense of some sharpness in the resultant image [64,65]. Processed k-space is subsequently inverse Fourier transformed (IFFT) into image space.

Upon IFFT, one image remains for each of the receiving coils; these must be integrated into one coil-combined image, thereby improving signal-to-noise (SNR) ratio of the resulting image. Coil combination can be accomplished with various approaches, the simplest of which summing the squares of corresponding pixels across coils, while more complicated approaches will rely on scanner calibration data to optimize the coil combination approach [66,67]. After coil-combination, multiple corrections are applied: surface coil intensity correction (SCIC) and gradient coil inhomogeneity correction. SCIC entails pixel intensity adjustment across the image

to allow for more uniform contrast and SNR throughout the image [68], whereas gradient coil inhomogeneity corrections warp images to adjust for nonlinearities that may be observed in the localizing gradient coils [69]. Pixel magnitudes are typically stored and images scaled to a desired range of pixel values, yielding standard images used in clinical applications. More sophisticated MR sequences may require further processing, such as potential separation of phase and magnitude signals, but these baseline steps will be required for nearly every sequence.

3.4 Conventional MR Protocols

T_1, T_2 and proton density (PD) weighted images dominate current clinical MR protocols. T_1 relaxation time is also referred to as longitudinal relaxation time, and is a time constant describing the rate at which magnetization will recover along the longitudinal axis after an excitatory RF pulse is applied [70]. T_1-weighted images then select parameters to accentuate differences in these relaxation times across a given slice of tissue; this is accomplished by a sequence with a short TR and TE. In doing so, T_1 images will see fat appear as very bright, soft tissues such as muscles and ligaments appear moderately bright, while fluid will be dark [71]. Contrarily, T_2 relaxation time is also referred to as transverse relaxation time; in this case, T_2 is a time constant describing the rate of magnetization decay along the transverse axis after RF pulse application [72]. T_2-weighted images are attained by using a long TR and TE, instead attaining images in which fluid is very bright, fat appears moderately bright, and soft tissues generally appear less bright [71]. The final of these major sequences, PD, does not refer to a given type of relaxation, but instead selects parameters to maximize signal imaged in the

transverse plane, doing so using a long TR and short TE. As such, tissues with the highest concentrations of protons will appear brightest in PD imaging: fat and fluid will be very bright, while soft tissue will appear gray and bones dark.

Clinical imaging protocols will administer multiple sequences to acquire multiple weightings of tissue for radiologist assessment. Furthermore, current clinical sequences are overwhelmingly acquired in 2D; as such, clinical imaging protocols will acquire the aforementioned sequences in multiple planes (axial, sagittal and coronal) [73]. Furthermore, signal from fat, particularly from tissues such as bone, can obscure the identification of more nuanced and clinically relevant findings such as bone marrow edema and inflammation; as such, fat suppression or short tau inversion recovery (STIR) can be integrated into these sequences to null fat signal. Lastly, in extreme cases, a contrast agent such as Gadolinium (Gd) can be administered to improve the diagnostic quality of MR images. Here, Gd is injected intravenously into a patient and, due to its paramagnetic properties, will shorten both T_1 and T_2 relaxation times. This means that, with otherwise identical acquisition parameters, all tissues would appear brighter in a T_1-weighted scan and darker in a T_2-weighted scan [74]. Specific tissues will see Gd uptake at differing rates, meaning that resulting images will exhibit enhancement (alteration in pixel intensities with respect to non-contrast baselines) in accordance with underlying anatomy, making anomalies such as tumors and active sites of inflammation easier to identify.

These conventional sequences provide rich structural information that has formed the basis of modern radiology. It is important to note, however, that all the sequences discussed to this

point are fundamentally qualitative sequences: that is, while the intrinsic T_2 relaxation times of tissues in a sample affect how that sample appears in T_2-weighted MR scans, those exact T_2 relaxation times are not identified from a conventional sequence. In other words, conventional MR sequences are weighted such that the relative intensities of pixels give rise to anatomy (i.e. fluid is known to be brighter than soft tissue in T_2 imaging), but actual pixel intensities themselves are meaningless. This is the fundamental difference between conventional and compositional MR scans, the latter of which is described in the next section.

3.5 Compositional MR Imaging

Compositional MR scans yield maps of MR parameters such as T_1 and T_2, allowing the intrinsic parameters to be visualized rather than qualitative images weighted by these parameters. Longitudinal magnetization can be represented by the following equation:

$M_{z,TE} = M_z\left(1 - e^{-\frac{TE}{T_1}}\right)$; where $M_{z,TE}$ is the longitudinal magnetization observed at echo time TE, M_z is the baseline longitudinal magnetization before the initial RF excitation pulse is applied, and T_1 is the intrinsic T_1 of the given tissue [75]. Contrarily, transverse magnetization is given by the following: $M_{xy,TE} = M_{xy,0}\left(e^{-\frac{TE}{T_2}}\right)$, where $M_{xy,TE}$ is the transverse magnetization at echo time TE, $M_{xy,0}$ is the transverse magnetization time at echo time 0 (maximal transverse magnetization), and T_2 is the intrinsic tissue T_2 [75]. Conventional MR uses predefined TR and TE to enable the desired weighting of an acquired image. In a single compositional MR sequence, however, a given volume is essentially acquired multiple times using multiple TEs (or in the case of a more complex intrinsic MR parameter such as $T_{1\rho}$, multiple spin-lock times), making for a

substantially lengthier acquisition than a conventional MR sequence with similar parameters [76,77]. After acquisition, an image obtains intensities and corresponding TEs for each pixel, which are then fit pixelwise to the appropriate magnetization equation, possibly using a technique like Levenberg-Marquardt fitting [78]. This allows solving voxel-wise for an MR parameter such as T_1 or T_2, yielding maps of these parameters.

This approach affords compositional MR sequences (including T_1 and T_2 mapping) many advantages over conventional MR. First off, pixel values resulting from compositional MR acquisitions carry physiological meaning: elevated T_2 values, for example, can be an indication cartilage degeneration in the knee and hip, while low T_2 values can indicate degeneration in IVDs [79–82]. Additionally, conventional MR images are sensitive to morphological changes in tissues, but compositional MR can be sensitive to biochemical changes that precede morphological changes. These changes may include alterations in collagen, proteoglycan, and water content, and have most thoroughly been characterized in the knee [83,84]. Despite these advantages, however, acquisition times for compositional MR are necessarily longer than conventional sequences, and at least to this point are too long to reasonably be implemented in clinical imaging protocols. Furthermore, while substantial progress has been made in compositional MR scan-rescan reproducibility, additional improvements are needed before widespread clinical adoption [85–87]. Nonetheless, compositional sequences offer a promising alternative to deliver useful and quantitative information to clinical imaging protocols with further development, complementing the information obtained from conventional sequences.

Between these conventional and quantitative sequences, the work in this thesis will use the following: 3D fat-saturated PD knee scans, coronal 2D fat-suppressed T_1-weighted scans, and coronal 2D fast-suppressed Gd T_1-weighted scans. An additional project will use the magnetization-prepared angle-modulated partitioned k-space spoiled gradient-echo snapshots (MAPSS) compositional MR sequence that simultaneously acquires images to calculate $T_{1\rho}$ and T_2 maps, focusing on the T_2 mapping acquisitions [88]. The final work will use 3D double-echo steady-state (DESS) knee MR images, a more complex sequence that acquires two signals per slice that can be sum-of-squares combined into a single morphological MR sequence [89]. The acquisition of two echo images per slice has the added advantage of allowing for compositional imaging along with the more conventional structural MR, but that facet of DESS is not explored in this thesis.

Chapter 4 - Deep Learning in Medical Imaging

Deep Learning (DL) has brought overwhelming changes to medical imaging, and more broadly, radiology. Its applications have spanned the entire imaging lifecycle, with ongoing avenues of research applying DL for assigning MR protocols to patients (the full list of MR sequences to be acquired), accelerating image acquisition, and automating the processing and interpretation of medical images. This chapter will introduce the basics of DL, with particular focus on the computer vision architectures. It will then overview medical imaging applications in which computer vision has been utilized and that constitute the core works of this thesis.

4.1 Basics of Deep Learning

DL entails efficient usage of large amounts of data to train multilayered, deep architectures and solve tasks such as classification and regression. Conceptually, one of the simplest DL architectures is an artificial neural network (ANN), which accepts input variables such as age, BMI and OA severity, and can predict binary variables such as whether a patient experiences pain or continuous variables such as pain severity. Each ANN layer consists of a series of nodes, with the number of nodes in each layer being predetermined and intrinsic to precise network design [90]. Each node performs two operations: a weighted sum of node outputs from the previous layer (for the first layer, a weighted sum of the input variables), and applies a nonlinearity such as a sigmoid operator, hyperbolic tangent, or rectified linear unit (ReLU) [91]. The nonlinearity applications are crucial in allowing network generalization to a wider variety of data than a strictly linear approach would permit. Furthermore, weights used in weighted summations are learned during network training: this is typically accomplished using an

optimization approach and a loss function. A loss function is a penalty term that describes the degree to which network predictions are incorrect: widely used examples include categorical cross entropy for classification tasks and mean-squared error for regression tasks such as age prediction [92,93]. Given network weights, a model can be inferred on a batch of data, allowing loss function calculation; an optimization approach determines how to update network weights in accordance with the loss. Conceptually, the simplest of these approaches is gradient descent: gradients are calculated with respect to each network weight, essentially determining the direction in which individual weights will be updated at the next network step. Furthermore, a learning rate is set prior to model training, and is multiplied by the network gradients (all multiplied by -1 to ensure correct direction of weight updates), determining the size of steps used to update model weights [94]. In stochastic gradient descent, one data point is randomly selected from a training batch to evaluate gradients, whereas in batch gradient descent, all data points of a training batch are used [95]. More sophisticated approaches exist such as Adam Optimizer that use higher-level gradient information and automatically adjust learning rate to update network weights, whereas other approaches such as AdaGrad, AdaDelta, and RMSProp are also available for this purpose [96–98].

These ANN principles are crucial and form the foundation of more sophisticated DL approaches. Particularly in medical image analysis, however, architectures often must be adapted to handle visual inputs. Referred to as “computer vision” algorithms in the literature, the most used family of algorithms in this space are convolutional neural networks (CNNs), which are used extensively in this thesis. While overarching ANN principles apply to CNNs, some components

are modified for visual input. Appropriately, the basis of CNNs are convolutions, which take the place of ANN nodes in architectures. Here, fixed-size filters do elementwise multiplication of filter values with a region of corresponding size from an input image, summing outputs to obtain an output pixel value. The filter is then shifted to a new region within the input image, also of corresponding size, in a process known as "striding." Here, the process repeats, yielding another output pixel value. By striding the filter throughout an entire image, filters yield a feature map representation of the input image, which is fundamentally another image sensitive to the filter's properties [99]. In classical approaches, filters were handcrafted to maximize sensitivity to desired attributes such as horizontal lines, vertical lines, or higher-level features [100]. In DL, however, the filter weights are learned, generally yielding far superior performance to handcrafted features for similar tasks. Just as nonlinearities needed to be applied to nodes in ANNs to more accurately model intricate patterns, nonlinearities are applied to convolution outputs, or feature maps. Also as was the case in ANNs, CNN feature maps of a given layer serve as inputs to the convolutions and resulting feature maps of the next layer. By convolving images in this manner, CNNs learn increasingly complex image representations: for instance, while filters in the first layer may be sensitive to lines at different orientations, filters in the second layer may be sensitive to corners, and so on, until a final layer may be sensitive to something as specific as animal species [101]. Many ANN loss functions and optimization techniques that held are applicable to CNNs, but working with images allows for usage of image-processing metrics such as structural similarity index (SSIM) as a loss function during training [102].

Properly training CNNs, or any DL network, has a series of required steps and challenges. First, a dataset must be understood and split into multiple datasets: (1) a training set, seen by the model during training and used to update model weights; (2) a validation set, seen by the model during training but only used to evaluate model performance and not update weights; (3) a test set, not seen by the network until final parameters have been selected and only inferred on once to evaluate model performance. Particularly with images, pre-processing of images is required to ensure consistent scaling and distribution of pixel values. After data splitting and preprocessing, a network architecture must be designed, a loss function identified and an optimization approach selected to carry out training. Considerations to improve training include learning rate optimization: a small learning rate will cause training to be slow, while a large one can cause difficulties optimizing network parameters. Data augmentation techniques such as random rotation, translation, and addition of noise to training set images are common to improve robustness of learned features, likely improving network performance on the test set and external data. Also worth of considering are a stopping criterion and the number of epochs to train the model: an “epoch” refers to a complete cycle of the network seeing all training data, and while some approaches train for a fixed number of epochs, others may stop early in accordance with criteria such as limited improvement in validation set performance. Some medical imaging-specific challenges include robustness of trained algorithms to different MR sequences, to MR images acquired from multiple scanners, and to MR images acquired from different vendors.

4.2 Anomaly Detection and Prognosis Prediction

Considerable effort from radiologists goes into mundane tasks: staging IVD degeneration severity, assessing knee OA severity, and staging lumbar spine stenosis are among the most common for MSK radiologists. These tasks are extremely repetitive, placing immense loads on radiologists and possibly contributing to burnout [103]. Furthermore, radiologist expertise is most needed in a limited number of cases with truly unique and subtle findings, such as small tissue lesions, as opposed to other cases in which anomalies are common and more easily spotted. As such, the medical imaging community has spent years training algorithms to automate anomaly detection, imagining a future in which, given a lumbar spine MR scan, as an example, a radiologist is provided automated assessments of IVD health, Modic changes in vertebral endplates, detection of vertebral fractures, classification of stenosis, and so on. In this manner, radiologists could edit model predictions as needed, but having them as a baseline rather than grading from scratch can ease burden and redirect their attention to patients where their expertise is most needed.

Examples of DL anomaly detection algorithms in MSK work are widespread. A Siamese-network style architecture was used to automatically diagnose knee OA from radiographic images, achieving strong test set performance, with an area under the receiver operating characteristic curve of 0.93 [9]. For more detailed analysis, another approach used a 3D V-Net architecture to segment tissues such as cartilage and menisci compartments, which were then fed into 3D CNN architectures with residual connections, showing strong performance in diagnosing anomalies in cartilage, bone marrow, menisci and the ACL [104]. Another approach used an MRNet-style

architecture to automatically diagnose ACL and meniscal tears [105]. Outside the knee, hip fractures can be difficult to identify in medical imaging, but a DenseNet architecture saw strong sensitivity and specificity in doing so from radiographic images [106]. The lumbar spine has similarly seen substantial ongoing work: Modic changes have classically been used to track conversions between different states within vertebral endplates, but a recent work used a V-Net to develop a voxel-wise, more nuanced assessment of these changes throughout entire vertebra [107]. Similarly, Faster R-CNN and ResNet architectures have been used to effectively classify degree of lumbar spinal stenosis from MR images [108]. One of the first works to thoroughly assess numerous spinal anomalies was SpineNet, which automatically classifies IVD degeneration, vertebral endplate defects and changes, and stenosis, among others [109].

Numerous aspects of MSK imaging have thus seen DL algorithms applied for anomaly detection. From a data science perspective, prognosis prediction can be framed as a very similar problem: rather than training algorithms to automatically diagnose current anomalies, they can instead predict if the anomaly will be present in the future, requiring longitudinal datasets. Practically, prognosis prediction algorithms have obvious applications: algorithms predicting otherwise healthy patients have some length of time until they develop an anomaly such as OA can be the impetus for initiating treatments that extend the time to disease or invasive treatment. In this vein, prognosis prediction is of clear clinical interest, and ongoing works doing so in MSK include prediction of OA progression from MR images using an EfficientNet-B0 architecture [110] and investigating knee phenotypes associated with risk of radiographic knee OA [111]. Additional

work is required to develop similar predictive models to inform preventative measure initiation in clinical decision making, which this thesis addresses in one project.

4.3 Image Segmentation

Tissue segmentation is a booming area of MSK and DL research, with widespread clinical applications. For instance, segmentations are crucial in surgical planning, identifying incision paths that minimally damage tissue peripheral to an anomaly [112]. Beyond this, emerging avenues of MSK research include identifying imaging biomarkers for OA, while compositional MR research continues to characterize soft tissue mass composition, among numerous others [113–115]. A necessary step in identifying imaging biomarkers is tissue segmentation: before evaluating, for instance, if T_2 values within knee cartilage or specific femoral bone shape phenotypes are imaging biomarkers for OA risk, cartilage and bones must be segmented. Probably more than any other medical image processing task, tissue segmentations are laborious and slow when done manually. As such, substantial research has gone into designing and training DL segmentation algorithms to automate this process.

Any discussion of medical imaging DL segmentation approaches must start with the UNet, whose development in 2015 revolutionized segmentation and image synthesis algorithms alike [116]. The key innovation with the UNet was the introduction of skip connections: the network has an encoding path that creates low-resolution, high-dimensional representations of input images, and a decoding path that decodes the high-dimensional representation into a predicted segmentation. Importantly, at corresponding levels on the encoding and decoding paths, skip

connections concatenate information from the encoding path to the decoding path, providing crucial contextual information to the decoding path that helps localize predicted segmentations. UNets and closely related architectures have seen substantial applications in segmenting knee cartilage and lumbar spine IVDs from MR scans and have been repurposed into flexible frameworks that streamline much of the data preprocessing and hyperparameter tuning involved in training segmentation models [117–120]. More recently, advanced DL approaches such as transformers have also seen application in tissue segmentation [121].

DL has seen considerable development for medical imaging tissue segmentation, and these approaches show extremely promising performance. Many current approaches, however, show human-like performance for healthy patients but can fail in rare cases such as cartilage lesion anomalies, making this fine-tuning an area of ongoing research.

4.4 Image Synthesis

Image synthesis entails predicting the appearance of one image from another. In the context of MR imaging, it will usually involve using at least one MR sequence to predict the appearance of another (i.e. using axial T_1 images to predict axial T_2 images). This is of clinical interest for several reasons, the most basic of which is saving time and costs. Properly trained image-to-image translation algorithms can synthesize, for example, a T_2 image from a T_1, eliminating the need to acquire a T_2 weighted image in the scanner, shortening the MR imaging protocol length. A more exciting reason, however, is its potential to eliminate the need to administer toxic contrast agents such as Gd. Often required for imaging inflammation or when a tumor is

suspected, Gd is toxic and shows evidence of deposition within the bone and brain, meaning its administration should be avoided whenever possible [122]. In this vein, image-to-image translation algorithms can be applied to predict post-contrast MR image appearance from a pre-contrast image. As such, a properly trained image synthesis algorithm here could dramatically improve patient safety and possibly eliminate the need for toxic contrast agents.

Most applications of image synthesis algorithms for post-contrast images have come in brain MR, where Gd is usually administered to rule out tumors. 2D and 3D UNets, and UNet-style architectures have been applied to eliminate or reduce Gd dosage required in brain post-contrast imaging applications, synthesizing post-Gd images with minimal loss in image quality and similar utility of images for downstream tasks such as tumor segmentation compared to full-dose post-contrast images [123–125]. Outside of brain applications, cardiac imaging can also entail Gd administration to diagnose myocardial infarction, where DL has seen applications to eliminate required Gd dosage [126]. Due to the temporal component inherent in cardiac imaging, however, the proposed pipeline was more complex than conventional approaches. MSK, however, has seen less widespread application of image synthesis algorithms than brain imaging, with essentially no work done in synthesizing post-contrast imaging of inflammation. This stands as a limitation of previous literature that one of the works in this thesis will address.

4.5 Image Reconstruction

Image reconstruction refers to converting MR imaging from k-space, the frequency-based domain in which MR signals are acquired, into the image-space DICOM images that are

interpreted by radiologists. A major drawback in MR imaging is long acquisition time; consequently, much of recent image reconstruction research has entailed devising k-space undersampling schemes and reconstruction algorithms. More specifically, undersampling schemes sample fewer points in k-space by sampling fewer phase-encoding lines, resulting in a faster MR acquisition that produces blurry and aliased images. Reconstruction algorithms, on the other hand, accept the aliased image and/or the undersampled k-space and predict the appearance of a full-length acquisition image. In combining the two, image reconstruction schemes reduce MR acquisition time while attempting to minimally sacrifice image quality.

A wide array of reconstruction approaches has gained traction, not all of which use DL. Some such approaches include parallel imaging (PI), in which the redundance of acquiring k-space using multiple coils is exploited to unalias undersampled images, reducing acquisition time at the expense of SNR [63]. In compressed sensing (CS), image reconstruction is done in an iterative manner, minimizing an objective function that ensures a predicted image maintains fidelity to acquired k-space points while preserving sparsity of the predicted image in an alternate domain, such as wavelets [127,128]. Low rank and spare modeling approaches share some similarities to CS in requiring minimization of an objective function but have some other methodological differences and are best studied for dynamic MR acceleration [129]. MR Fingerprinting (MRF) has also shown considerable promise, and marks a fundamentally different approach: here, MR images are acquired by pseudorandomizing acquisition parameters such as TR, TE and flip angle, after which acquired images are compared against a lookup dictionary to simultaneous predict multiple MR image weightings for the patient [130]. MRF acquisitions can

be combined with classical reconstruction approaches to further accelerate the acquisition scheme.

These non-DL approaches all show substantial promise and will be described in further detail in Chapter 6. That said, DL applications in image reconstruction have exploded in recent years, with the subject matter gathering considerable interest annually at major medical imaging conferences. Here, image-to-image DL architectures are trained in one of several manners: they can predict fully-sampled image space from an aliased image space, fully-sampled k-space from undersampled k-space, or fully-sampled image space directly from undersampled k-space [131–133]. Loss function selection is crucial: reconstruction algorithms often use a data consistency loss, ensuring k-space of predicted images corresponds with acquired k-space points, just as CS does in its objective function [134]. Data consistency aside, other common terms include a pixel-based loss function, structural similarity index, and feature based loss functions. In this manner, many of the same image-to-image architectures used in image segmentation and image synthesis can be adapted for reconstruction, which fundamentally is a special image-to-image translation problem. UNet style networks and skip connections formed the basis of early DL works for reconstruction, but recent advances have popularized variational and unrolled architectures, in which input aliased images/undersampled k-space/coil sensitivity maps are fed through multiple versions of the same architecture, slightly improving predictions each iteration and ultimately predicting final reconstructed images [135].

There is no question that reconstruction has made substantial progress in recent years. Some challenges remain, however: as it pertains to DL, it is important to note that CNNs are typically translationally invariant. While this is useful when in image space, it is cumbersome in k-space, where points at the center of k-space are orders of magnitude higher than peripheral k-space points; properly training a network that operates directly in k-space and having peripheral k-space points contribute meaningfully to final network parameters is therefore challenging, making some DL approaches difficult to train. Beyond this, small, sharp features are easily lost in the undersampling process, and are usually among the most important clinically yet the most difficult for a reconstruction algorithm to recapture [136]. The stability of architectures in reconstructing these features have also been challenged [137,138]. GANs have often been added to training schemes to better preserve these features, but their retention remains a challenge [139]. As with other DL applications in imaging, robustness of trained algorithms to multiple vendors and MR sequences remains challenging, as does robustness to different undersampling patterns [140]. Lastly, conventional metrics used to assess algorithm performance such as SSIM, peak signal-to-noise ratio (PSNR), and normalized root mean squared error (nRMSE) correlate poorly with gold-standard radiologist annotations, making optimization of algorithms and assessment of their efficacy difficult [141,142]. The image reconstruction works in this thesis will address some of these challenges.

Chapter 5 - A Cartilage-Specific Loss Function Improves Image Reconstruction Performance in Multiple Tissues of Clinical Interest

The following has been reformatted and reproduced with full permission from the publisher. It was a conference presentation that appeared as:

Tolpadi A. A., Calivà F., Han M., Bahroos E., Larson P. Majumdar S & Pedoia V. A Cartilage-Specific Loss Function Improves Image Reconstruction Performance in Multiple Tissues of Clinical Interest. *In Proceedings of the 30th Annual Meeting of ISMRM, London, England, United Kingdom*. 948 (2022).

5.1 Introduction

For musculoskeletal imaging, MRI is a premier option, offering high-resolution images with exquisite soft tissue contrast [143]. A major drawback, however, is long acquisition time. Consequently, image reconstruction has been a major recent research focus, with deep learning, compressed sensing, and model-based approaches among those under development to accelerate acquisition [144–148]. However, most published approaches are optimized for entire imaging volumes rather than specific tissues of interest. While not necessarily a drawback, musculoskeletal imaging often necessitates strong image quality most so in a specific tissue such as cartilage [149]. Moreover, most reconstruction algorithms are assessed with metrics like structural similarity index (SSIM) [102] that are agnostic to tissues of interest and may not accurately capture a model's clinical utility [141]. As such, optimization of reconstruction algorithms for specific tissues remains an open question, and the limitations of standard image reconstruction metrics is worth investigating.

5.2 Methods

5.2.1 Image Acquisition

For 3D-Fast-Spin-Echo fat suppressed CUBE images acquired at a UCSF GE Signa 3T MRI scanner, an in-house pipeline was developed that leveraged GE Orchestra 1.10 and other postprocessing tools to reconstruct images from raw scanner data, allowing multicoil k-space, pre-processed coil-combined images, post-processed coil-combined images, and other intermediate files from the image processing pipeline to be stored. Acquisition parameters were as follows: FOV=15cm^2; acquisition matrix=256×256×200; ±62.5kHz readout bandwidth; TR=1002ms; TE=29ms; ARC acceleration by a factor of 4 [150]. Images are zero-filled in k-space to obtain final 512×512×200 resolution. Scans from 62 patients were split 38/12/12 into training, validation, and test.

5.2.2 Undersampling and Pre-Processing

3D multicoil k-space was undersampled in k_y-k_z with a center-weighted Poisson pattern while fully sampling the 5% central square in k-space, projecting the pattern along k_x. Both undersampled and fully-sampled k-space were 1D inverse Fourier transformed along the slice direction, yielding undersampled and corresponding ground truth k_x-k_y-z 2D k-space for each coil, along with associated multicoil images. Root sum of squares coil combination of fully sampled coil images was used to calculate ground truth coil-combined images. A 5-class 3D V-Net pipeline obtained segmentations for 4 knee cartilage compartments and the menisci [151].

5.2.3 Training

A KIKI-Net [132] inspired architecture was designed to take 2D undersampled multicoil images as input, predict fully sampled multicoil images, and do root sum of squares coil combination, yielding a coil-combined prediction (Figure 5.1). A 4-component loss function was used to train baseline networks at R=4 and R=8: (1) multicoil image space L_1 loss; (2) multicoil k-space L_1 data consistency loss; (3) coil-combined image space L_1 loss; (4) 1 - coil-combined SSIM. In addition, separate networks were trained with an additional loss component: coil-combined L_1 loss in cartilage. Training was done for 20 epochs, with loss function weightings and learning rate being optimized in a hyperparameter search. Standard reconstruction metrics such as SSIM, as well as tissue-specific metrics like normalized root mean square error (nRMSE) and peak signal-to-noise ratio (PSNR) were used to assess performance [152].

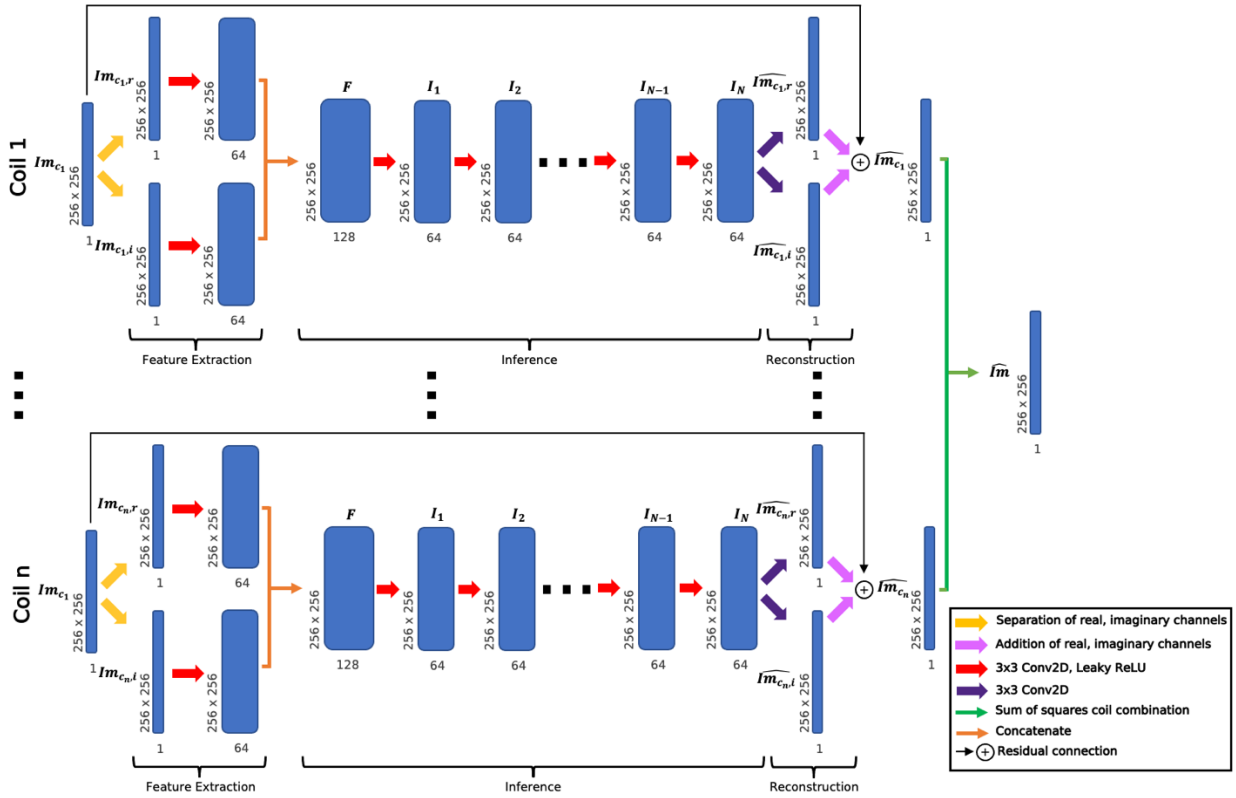


Figure 5.1 KIKI-Net inspired architecture for reconstruction. KIKI-Net inspired architecture predicts coil-combined images from undersampled multicoil image-space inputs. Undersampled coil images were fed through feature extractors for real and imaginary channels, inference convolutions, and a reconstruction convolution, and subsequently sum of squares combined to yield single-coil predictions. Weights were shared across coils, and N=10 inference layers used. Baseline networks were trained with a multi-component loss: multicoil image space L_1, multicoil k-space L_1, coil-combined L_1, coil-combined SSIM.

5.3 Results

Standard reconstruction metrics show that at R=4 and R=8, baseline models perform well in recovering ground truth images, with reasonably high SSIM and low nRMSEs (Table 5.1). At a tissue level, these metrics clearly show addition of the cartilage L_1 loss reduced nRMSE and increased PSNR not only within cartilage, but also within menisci.

Table 5.1 Model performance metrics in test set. Reconstructions within cartilage and menisci show slight to substantially lower errors and higher PSNR at expense of slight drops in full-slice performance with cartilage-specific loss usage. Tissue-specific losses can thus improve reconstruction performance in multiple clinically relevant tissues. Lower SSIM compared to other published models is likely in part due to the challenges associated with reconstructing such a high-resolution, fat suppressed sequence.

		R=4		R=8	
		No Cartilage Loss	**With Cartilage Loss**	**No Cartilage Loss**	**With Cartilage Loss**
SSIM (± 1 s.d.)		0.712 ± 0.09	0.703 ± 0.09	0.612 ± 0.07	0.599 ± 0.07
nRMSE (% ± 1 s.d.)	**Full Volume**	1.24 ± 0.62	1.28 ± 0.71	2.0 ± 0.84	2.08 ± 0.9
	Cartilage	1.03 ± 0.31	0.91 ± 0.32	1.68 ± 0.53	1.55 ± 0.59
	Meniscus	2.93 ± 1.98	2.81 ± 2.05	4.18 ± 2.24	4.04 ± 2.39
PSNR (± 1 s.d.)	**Full Volume**	27.5 ± 1.54	27.4 ± 1.53	25.4 ± 1.53	25.3 ± 1.52
	Cartilage	22.1 ± 1.36	22.7 ± 1.54	20.0 ± 1.45	20.5 ± 1.67
	Meniscus	21.8 ± 1.61	22.0 ± 1.7	20.1 ± 1.52	20.3 ± 1.5

Visualizations of reconstructions at R=4 show both pipelines recover fine details and yield dramatic improvements in image quality over zero-filling (Figure 5.2). Moreover, reconstructions show the cartilage L_1 loss mitigates spurious signal elevations in menisci present in baseline

models while more accurately reconstructing sharp, local signal elevations. While trends were similar at R=8, the cartilage L_1 loss also mitigated aliasing artifacts in menisci slightly better than the baseline model (Figure 5.3).

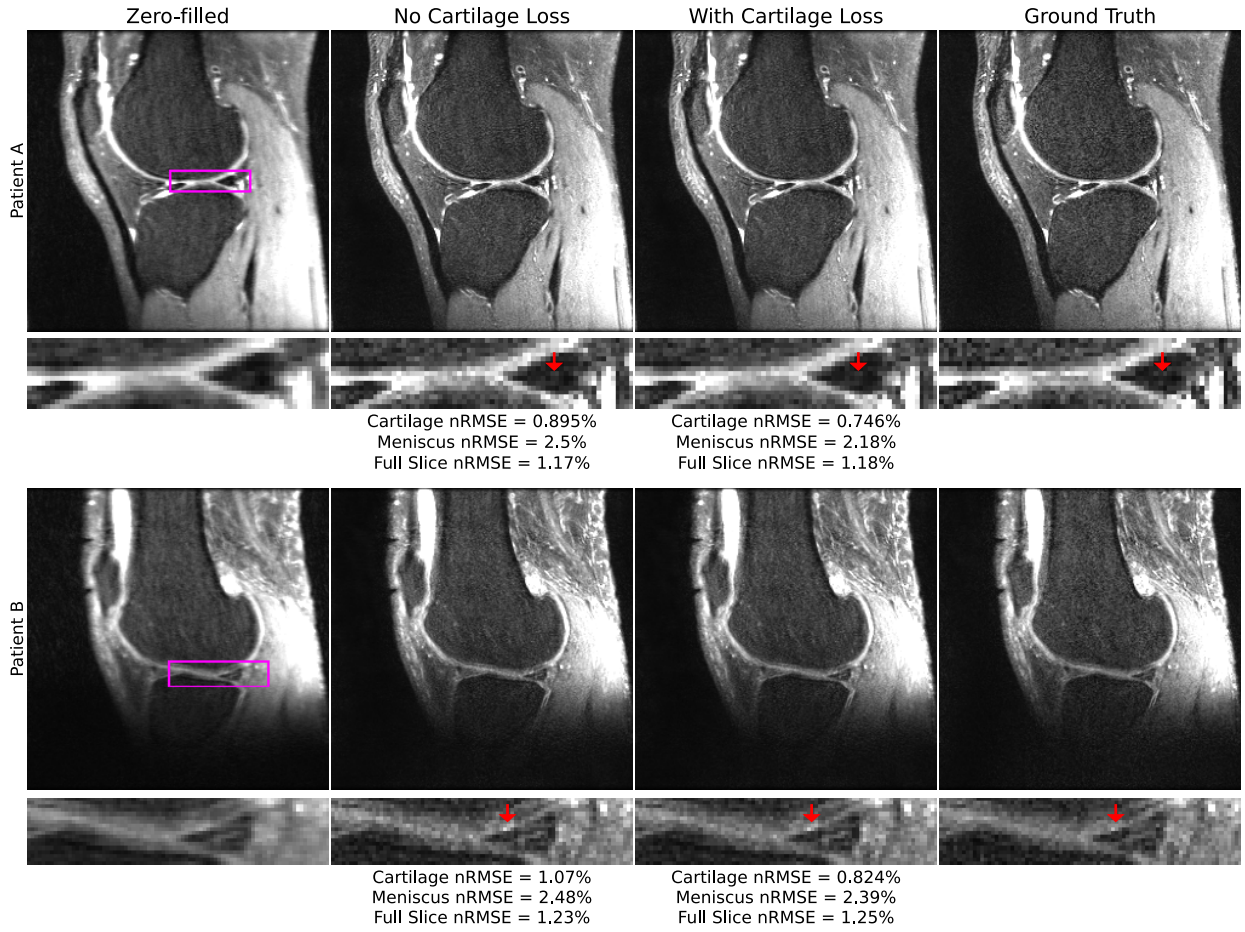


Figure 5.2 R=4 pipeline performances. In both patients, nRMSEs show improvement in cartilage and menisci reconstructions with use of a cartilage-specific loss. In patient A, standard reconstruction reveals a slight, spurious elevation in posterior meniscal horn signal that is less apparent with cartilage-specific loss. Similarly in patient B, a sharp signal elevation in lateral femoral cartilage is better reconstructed with cartilage-specific loss.

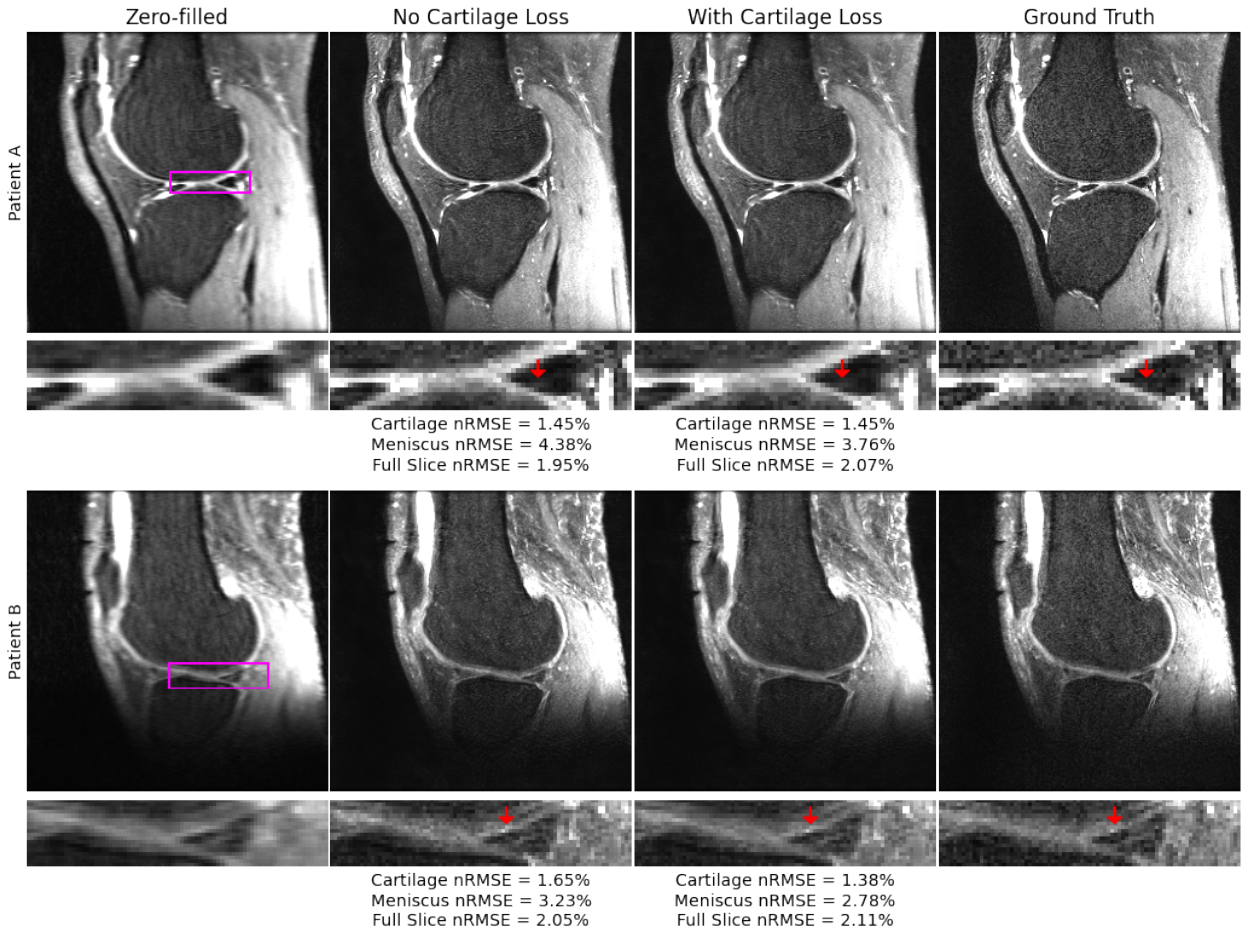


Figure 5.3 R=8 pipeline performances. Reconstructions at R=8 show similar or improved nRMSE in cartilage and menisci with cartilage-specific loss use. In patient A, an aliasing artifact in both reconstructions is better managed, although not eliminated, in the posterior meniscal horn with use of cartilage specific loss. In patient B, similar to the R=4 pipeline, the sharp local elevation in lateral femoral cartilage signal is better reproduced with a cartilage-specific loss, indicating improvements in both cartilage and menisci reconstructions with the tissue-specific loss.

5.4 Discussion and Conclusions

In tandem with the proposed architecture, a cartilage L_1 loss term improved reconstruction performance not only within cartilage, but also within menisci. While the former is interesting and useful, it is unsurprising. However, a cartilage loss term that improves meniscal reconstruction in addition to cartilage indicates adding tissue-specific loss terms can improve reconstruction performance in the multiple clinically crucial tissues, offering a simple means of

improving reconstruction performance for clinical settings worthy of exploration for all reconstruction algorithms. Also noteworthy is that at R=4 and R=8, full-slice SSIM, nRMSE, and PSNR worsened despite improvements in cartilage and menisci reconstructions. This indicates that, just as pulse sequence techniques such as fat suppression improve image quality in one tissue at the expense of another, reconstruction pipelines too can and perhaps should be optimized for tissue-specific performance. Another option may be training multiple tissue-specific pipelines and aggregating predictions to obtain improved full-volume reconstructions. Beyond these possibilities, that full-volume metrics worsened as cartilage and meniscal reconstructions improved raises the question of whether these standard metrics are optimal to evaluate clinical utility of reconstruction pipelines, seeing that the most clinically useful pipeline is likely one with very strong cartilage and menisci reconstructions.

This work elucidates the potential of tissue-specific losses to improve clinical utility of reconstruction models. To further investigate this, future work will include extension of this approach to other anatomies. Lastly, particularly for 3D sequences, finding innovative ways to exploit all dimensions of the acquisition in training while managing computational constraints is another avenue of future exploration.

Chapter 6 - Region of Interest-Specific Loss Functions Improve T_2 Quantification with Ultrafast T_2 Mapping MRI Sequences in Knee, Hip and Lumbar Spine

The following has been reformatted and reproduced with full permission from the publisher. It appeared in *Nature Scientific Reports* as:

Tolpadi, A.A., Han, M., Calivà, F. *et al.* Region of interest-specific loss functions improve T_2 quantification with ultrafast T_2 mapping MRI sequences in knee, hip and lumbar spine. *Sci Rep* **12**, 22208 (2022). https://doi.org/10.1038/s41598-022-26266-z

6.1 Abstract

MRI T_2 mapping sequences quantitatively assess tissue health and depict early degenerative changes in musculoskeletal (MSK) tissues like cartilage and intervertebral discs (IVDs) but require long acquisition times. In MSK imaging, small features in cartilage and IVDs are crucial for diagnoses and must be preserved when reconstructing accelerated data. To these ends, we propose region of interest-specific postprocessing of accelerated acquisitions: a recurrent UNet deep learning architecture that provides T_2 maps in knee cartilage, hip cartilage, and lumbar spine IVDs from accelerated T_2-prepared snapshot gradient-echo acquisitions, optimizing for cartilage and IVD performance with a multi-component loss function that most heavily penalizes errors in those regions. Quantification errors in knee and hip cartilage were under 10% and 9% from acceleration factors R=2 through 10, respectively, with bias for both under 3 ms for most of R=2 through 12. In IVDs, mean quantification errors were under 12% from R=2 through 6. A Gray Level Co-Occurrence Matrix-based scheme showed knee and hip pipelines outperformed state-of-the-art models, retaining smooth textures for most R and sharper ones through

moderate R. Our methodology yields robust T_2 maps while offering new approaches for optimizing and evaluating reconstruction algorithms to facilitate better preservation of small, clinically relevant features.

6.2 Introduction

Magnetic Resonance Imaging (MRI) has emerged as a crucial part of diagnosing pathologies such as osteoarthritis, ligament damage, tumors, and others [153–155]. Within MRI, several sequences can be deployed that exploit intrinsic tissue properties, providing images of varying weightings that effectively visualize tissues such as muscle, ligaments, bone marrow, and others [156]. In musculoskeletal (MSK) applications, clinical imaging protocols consist mostly of 2D fast spin echo (FSE) acquisitions with T_1 or T_2 weighting in various acquisition planes, which do well in depicting the structure and morphology of the underlying anatomy [157]. However, compositional MRI (cMRI) techniques to assess actual tissue parameters are gaining more attention as a complement of qualitative imaging.

cMRI techniques like T_2 relaxometry can provide maps of T_2 values (or another intrinsic MR parameter) across an imaging volume rather than a morphological image. For MSK applications, T_2 relaxometry offers sensitivity to water content, collagen content, and collagen fiber orientation in cartilage [149], making it sensitive to biochemical changes that can precede morphological changes across several tissues and anatomies [158,159]. Pre-morphological change sensitivity has been best characterized in the knee, where T_2 values are significantly higher across most cartilage compartments in healthy patients that later develop osteoarthritis

(OA) compared to controls [160,161]. Additionally, T_2 relaxometry offers quantitative MSK tissue health assessments, correlating with measures of hip cartilage and intervertebral disc (IVD) health [81,162–164], whereas in conventional clinical imaging, only semiquantitative tissue health assessments are obtainable with expert annotation [27,165]. All of this makes cMRI a promising potential addition to clinical imaging protocols.

A major challenge facing clinical adoption of cMRI, however, is acquisition time: while mapping sequences like the magnetization-prepared angle-modulated partitioned k-space spoiled gradient echo snapshots (MAPSS) can provide robust MR parameter maps, their acquisition times can exceed 5-6 minutes, making their addition to a clinical scan protocol difficult [88]. Acquisitions can be accelerated by sampling fewer points in k-space, inducing aliasing artifacts in resulting images that must be removed through subsequent postprocessing. Some proposed approaches to these ends are reconstruction strategies such as parallel imaging (PI), compressed sensing (CS), model-based reconstructions, deep learning (DL), low-rank and sparse modeling methods, and MR Fingerprinting (MRF). Most of these approaches design an algorithm or exploit the redundancy of k-space acquisition across multiple coils to predict the appearance of the fully-sampled reconstructed image.

PI was one of the earliest techniques to accelerate MRI acquisition and has seen clinical adoption. Here, the redundancy of a multiple coil acquisition is leveraged to mitigate aliasing artifacts [63,166,167], reducing clinical scan time up to acceleration factor R=3 for MSK applications [168,169]. CS [128] has also shown promise, where aliased images are iteratively

reconstructed by minimizing an objective function, retaining fidelity to acquired k-space and imposing sparsity on the reconstructed image in another domain. CS has attained clinically acceptable MSK image quality through roughly R=4 [168–171], and up to R=8 in research settings for knee cartilage $T_{1\rho}$ mapping [172]. Similarly, PI and CS have also been applied sequentially (and simultaneously) for further acceleration [173].

For cMRI acceleration, model-based reconstructions have gained traction, integrating the physics of T_2/T_2^* decay and T_1 recovery into an objective function iteratively optimized to reconstruct maps, showing promise in brain and lumbar spine T_2 mapping [174–176]. More generally, incorporation of the physics of MRI parameter recovery/decay has seen applications not just in model-based approaches, but in various aspects of other methodologies as well [177]. DL approaches have gained prominence in solving inverse problems such as reconstruction, allowing for cMRI reconstructions at higher R than other methods. Standalone DL approaches have seen promising results in knee MAPSS acceleration, T_1 mapping, and T_2 mapping sequences [116,178–181]. In other methodologies, DL has been integrated with model-based approaches while introducing loss functions to maintain fidelity to acquired k-space, seeing promise up to R=8 in knee and brain T_1 and T_2 mapping [134,139,140]. DL has been applied to accelerate T_2 mapping in MR Fingerprinting, where DL can remove aliasing artifacts from undersampled acquisitions and/or replacing time-consuming dictionary lookup steps to predict MR parameter maps, and exploiting spatial correlations within maps to improve reconstructions [182,183]. Lastly, aside from DL, low-rank and sparse modeling methods have emerged as a means of accelerating acquisitions, where several MRI images acquired at

different echo times are decomposed into temporal basis functions and spatial coefficients to model an MRI parameter, showing promise through R=8 [184].

These works represent great progress, although avenues for improvement remain. Above all, these methods have optimized reconstructed images for full-volume performance; however, in MSK applications, clinical assessment relies on the inspection of precise anatomic features in specific anatomic regions, and consequently, the reconstruction quality cannot be compromised within these regions. Put differently, given clinical context, strong image quality may be most important in specific regions of an image, leaving room for algorithm optimization. Furthermore, most recent published approaches leverage k-space data in formal reconstruction approaches, but for niche applications such as region of interest (ROI)-focused optimization, such approaches may be outperformed by DL-based post-processing algorithms that denoise and fit undersampled T_2-weighted images without using raw k-space. Moreover, performance of standard reconstruction algorithms is typically evaluated using metrics such as structural similarity index (SSIM), normalized root mean square error (NRMSE), and peak signal-to-noise ratio (PSNR), but recent works show these metrics may not provide the best correspondence with radiologist annotations [141,142], leading other groups to propose alternate metrics to fill this niche [185].

To these ends, this study proposes a recurrent UNet pipeline to postprocess undersampled coil-combined T_2-weighted echo images, fitting and predicting T_2 maps from accelerated MAPSS acquisitions in the knee, hip and lumbar spine [186,187]. These algorithms are trained with

multi-component, ROI-specific losses that optimize predicted maps for T_2 value and textural retention in cartilage and IVDs. In doing so, our approach allows for ROI-specific optimization, facilitating retention of small, crucial clinical features in tissues of interest while building on past applications of weighted loss functions for image processing tasks [188].

To summarize, the contributions of this work are as follows:

- By using a 4-component loss function in network training, we introduce the concept of "ROI-specific optimization" of cMRI accelerated acquisition pipelines.
- We conduct a thorough ablation study of these 4 loss function components, proving the value of all in retaining textures in predicted maps while retaining high fidelity to ground truth T_2 values.
- Acknowledging that standard evaluation metrics such as SSIM and NRMSE provide suboptimal sensitivity to clinically relevant metrics, we conduct a thorough Gray Level Co-Occurrence Matrix (GLCM)-metric-based analysis of smooth and sharp textural retention in predicted maps, with an eye towards better evaluation of retention of small features crucial to clinical diagnoses [189,190].
- We build on limited literature in hip and lumbar spine cMRI accelerated acquisition schemes by developing and evaluating our pipeline not only in knee cartilage, as several other works have done, but also for hip cartilage and lumbar spine IVD in ultrafast acquisitions.

6.3 Methods

6.3.1 MAPSS Acquisitions

Retrospective datasets including MAPSS in the knee (n=244 patients, 446 scans), hip (n=67 patients, 89 scans), and lumbar spine (n=21 patients, 24 scans) acquired from clinical 3T MRI scanners was used. Patients were scanned in accordance with all pertinent guidelines, including approval from the University of California, San Francisco Institutional Review Board (Human Research Protection Program), and informed consent was obtained from all study participants. MAPSS simultaneously acquired multiple $T_{1\rho}$ and T_2 weighted images, using $T_{1\rho}$ or T_2 preparation followed by 3D RF-spoiled gradient-echo Cartesian acquisition in a segmented radial centric view ordering during a transient state. A fat-selective inversion pulse was applied before either $T_{1\rho}$ [191,192] or T_2 preparation [193]. Each acquisition included $T_{1\rho}$-prepared images at four spin-lock times (TSLs) for $T_{1\rho}$ quantification, and three additional T_2-prepared images for T_2 quantification (TSL=0 ms images were shared for TE=0 ms images). In this study, only T_2-prepared images at four different TEs and corresponding T_2 maps from the MAPSS sequence were used. k_y-k_z space was acquired within an elliptical coverage (area=0.7 compared to rectangular k_y-k_z, not acquiring corner space). Knee images were acquired from patients having ACL injuries, with scans taken at baseline and 3 years post-reconstruction. Hip images were acquired from patients having hip OA. Lumbar spine images were acquired from healthy subjects or patients with low back pain. Table 6.1 shows acquisition parameters.

Table 6.1 Knee, hip and lumbar spine datasets and splits. MAPSS acquisition parameters for all datasets, with corresponding training, validation and test splits. ARC refers to Auto-calibrating Reconstruction for Cartesian Imaging [150]. For hip acquisitions, no phase wrap was applied: k_y was oversampled by a factor of 2X, with space outside the prescribed y-FOV eliminated after reconstruction. In some cases, multiple acquisitions were taken per patient due to having

multiple knees/hips scanned, or due to having follow-up scans for the same patient. Age and weight are reported mean ±1 standard deviation (s.d.). Datasets were split into training, validation and test, ensuring all scans of a particular patient were only placed into one of the three datasets. Unless otherwise noted, all results are reported on the test set are described by this table; to ensure robustness of trained pipeline to data splits, additional versions were trained on 2 more splits detailed in Supp. Tables A.2 and A.13, with results on those splits described in Supp. Tables A.14-A.16.

		Knee	**Hip**	**Lumbar Spine**
Acquisition Parameters	Scanner(s)	GE Discovery MR750w (GE Healthcare, Waukesha, WI), GE Discovery MR750 (GE Healthcare, Waukesha, WI)	GE Discovery MR750w (GE Healthcare, Waukesha, WI), GE Discovery MR750 (GE Healthcare, Waukesha, WI)	GE Discovery MR750w (GE Healthcare, Waukesha, WI), GE Signa PET/MR (GE Healthcare, Waukesha, WI)
	Coil(s)	8-channel T/R knee array (Invivo, Gainesville, FL)	32-channel cardiac array (Invivo, Gainesville, FL)	Geometry embracing method (GEM) posterior array (GE Healthcare, Aurora, OH)
	FOV	14×14 cm^2	14×14 cm^2	20×20 cm^2
	Acq. Matrix	256×128	256×128	256×128
	Slice Thickness	4.0mm	4.0mm	8.0mm
	Slices	22	28	12
	TEs	0 ms, 12.9 ms, 25.7 ms, 51.4 ms	0 ms, 10.4 ms, 20.8 ms, 41.7 ms	0 ms, 12.9 ms, 25.7 ms, 51.4 ms
	Readout BW	±62.5kHz	±62.5kHz	±62.5kHz
	Magnetization Recovery Time	1.3 s	1.2 s	1.5 s
	ARC	2X	2X	None
	No Phase Wrap	None	2X k_y oversampling	None
	Other	64-view acquisition/T_2 preparation	64-view acquisition/T_2 preparation	64-view acquisition/T_2 preparation
Demographics Information	Sex (M/F)	140/104	35/32	10/11
	Age	29.7±12.9	48.9±13.2	45.3±14.7
	Weight	74.3±12.7 kg	69.8±12.4 kg	69.6±11.0 kg
Training Information Details	Learning Rate	0.001	0.001	0.001
	Batch Size	1	1	1
Training	Patients	144	39	13
	Scans	265	59	14
	Slices	5,591	1,533	112

		Knee	Hip	Lumbar Spine
Validation	Patients	50	15	4
	Scans	91	15	5
	Slices	1,952	390	42
Test	Patients	50	13	4
	Scans	90	15	5
	Slices	1,928	390	40
Total	Patients	244	67	21
	Scans	446	89	24
	Slices	9,471	2,313	194

6.3.2 T_2 Fitting and Spatial Undersampling

Later T_2 weighted echo time images for each slice were registered to corresponding TE=0 ms images using a 3D rigid registration algorithm with a normalized mutual information criterion [194]. Levenberg-Marquardt fitting of registered T_2 weighted images yielded ground truth T_2 maps [195].

To simulate accelerated acquisition, coil-combined T_2 weighted magnitude images after reconstruction (ARC for knee and hip) were Fourier transformed and retrospectively undersampled using a center-weighted Poisson disc pattern, fully sampling a central 5% square in k_y-k_z (R=2, 3, 4, 6, 8, 10, 12). Acquisition times associated with ground truth and accelerated MAPSS acquisitions in each body part can be found in Supp Table A.1. As MAPSS acquires phase-encode lines with elliptical coverage in k_y-k_z (relative area of 0.7 compared to rectangular coverage), phase encoding lines solely within the sampling ellipse were undersampled. Although working with synthesized k-space data generated from coil-combined magnitude images, retrospective undersampling was done and R reported with respect to elliptical coverage in k_y-k_z to accurately simulate an actual undersampling pattern and not overstate model performance [196]. However, for hip acquisitions, reconstructed space outside the y-FOV

had already been discarded; thus, simulating acquisitions with application of 'no phase wrap' was not possible and undersampling patterns would differ from those implemented on a scanner. T_2 weighted images from each echo time were undersampled with a unique pattern. For k_y-k_z lines not sampled at a given echo time, those k_y-k_z lines were initialized with the corresponding k_y-k_z from the image with the temporally closest echo time for which that k_y-k_z was sampled. Only k_y-k_z lines not sampled in images acquired at all echo times were zero-filled. k-Space was subsequently inverse Fourier transformed, yielding undersampled, aliased images.

6.3.3 DL Pipeline Training

6.3.3.1 DL Architecture

An overview of the data processing and training schemes is shown in Figure 6.1, while a detailed diagram depicting our proposed network architecture is in Supp. Fig. A.1 ("Full Model"; 39,808,710 trainable parameters). Magnitude images from data undersampled as specified were fed into a recurrent UNet network. The network contains an initial recurrent portion: aliased images from each T_2 echo time have a 5-layer processing stream of 2D 3×3 convolutions with stride 1, yielding layers of depth 64, 128, 256, 512, and 1. Residual connections connect input aliased images with processing stream outputs. 2D 3×3 convolutions with stride 1 and residual connections transfer information between temporally adjacent corresponding hidden echo time processing layers with weighting parameter λ_w=0.2 [197]. This soft-weighted view-sharing of neighboring T_2 weighted echo time images facilitated sharing of feature map information between temporally adjacent echo time images, which can augment sharing of k_y-k_z initializations to improve network image predictions. Outputs of all 4 echo time image

processing streams were concatenated and fed to a UNet that predicted T_2 maps. 2D 3×3 convolutions with stride 2 were used for the encoder, and 2D 4×4 transpose convolutions with stride 2 for the decoder. Two additional architecture versions were also trained: one UNet with no recurrent portion ("No RNN"; 35,116,037 trainable parameters) and a second in which all layers apart from inputs to the recurrent portion and UNet had half the depth listed in Supp. Fig. A.1 ("Reduced Parameters"; 9,958,246 trainable parameters).

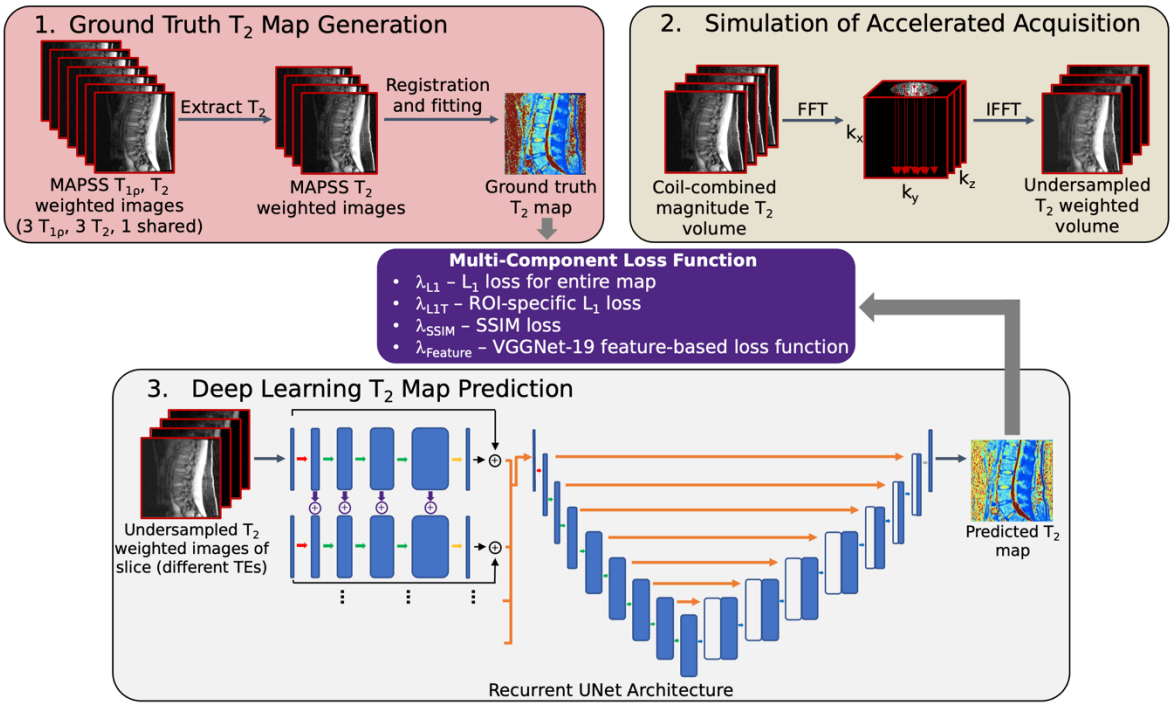


Figure 6.1 Proposed MAPSS Acceleration Pipeline. Proposed MAPSS Acceleration Pipeline. Experiments in proposed study entail generating ground truth T_2 maps from MAPSS, simulating accelerated acquisition of T_2-weighted MAPSS images, and training a network to predict T_2 maps from undersampled images. (1) MAPSS contains 7 images, 3 that are T_2 weighted, 3 $T_{1\rho}$ weighted, and 1 shared; the T_2 and shared image weightings are extracted, registered, and fitted slice-wise to yield ground truth T_2 maps. To simulate accelerated acquisition, each volume of coil-combined magnitude T_2 weighted images acquired at a given echo time are Fourier transformed, undersampled along the k_y-k_z plane with a center-weighted Poisson disc pattern, and inverse Fourier transformed to yield a simulated accelerated acquisition of a volume. Finally, undersampled T_2 weighted images acquired at all echo times for the same anatomic slice are concatenated and fed to the proposed recurrent UNet architecture, which predicts the T_2 map appearance for the slice. Training is done slice-wise with a multi-component loss function that includes a novel ROI-specific L_1 loss that optimizes predicted T_2 maps in cartilage and IVD

ROIs, with other components that improve training stability and encourage retention of textures.

6.3.3.2 Loss Function

Networks were trained with the multi-part loss function shown in Equation 6.1:

$$L_{network} = \lambda_{L_1} L_{L_1} + \lambda_{L_{1,\phi}} L_{L_{1,\phi}} + \lambda_{SSIM} L_{SSIM} + \lambda_{Feature} L_{Feature}$$

Equation 6.1 L_{L_1} is a scaled global loss function detailed in equation 6.2, $L_{L_{1,\phi}}$ is the ROI-specific L_1 loss described in equation 6.4, L_{SSIM} is an SSIM loss described in equation 6.5, and $L_{Feature}$ is a feature-based loss function designed to retain sharper textures and described in equation 6.6. λ_{L_1}, $\lambda_{L_{1,\phi}}$, λ_{SSIM}, $\lambda_{Feature}$ are loss component weightings optimized through a hyperparameter search.

L_{L_1} is a scaled global L_1 loss:

$$L_{L_1} = \left| S(T_2) - S(\hat{T}_2) \right|$$

Equation 6.2 T_2 represents ground truth T_2, $\hat{T}_2$ represents predicted T_2, and $S(x)$ is a translated and scaled sigmoid operator that assigns more weight to higher T_2 values. Sharp contrasts and high T_2 values can easily be lost in accelerated acquisition schemes, so $S(x)$ proved useful through empirical testing in focusing networks to preserve these details. $S(x)$ is defined below in Equation 6.3:

The translated and scaled sigmoid operator $S(x)$ is calculated as follows:

$$S(x) = y_l + (y_h - y_l)\left(1 + exp\left(-(10/(x_h - x_l))(x - (x_l + x_h)/2\,)\right)\right)^{-1}$$

Equation 6.3 x_l and x_h are the low and high T_2 value limits where the sigmoid operator weighting will transition from y_l to y_h. Parameters selected for the knee were as follows: x_l=0 ms, x_h=100 ms, y_l=0.1, y_h=1.0. In the hip: x_l=0 ms, x_h=60 ms, y_l=0.5, y_h=1.0. In the lumbar spine: x_l=0 ms, x_h=150 ms, y_l=0.25, y_h=1.0. A schematic of the operator that results from parameters of all three anatomies can be found as Supp. Fig. A.2.

$L_{L_{1,\phi}}$ is the ROI-specific L_1 loss:

$$L_{L_{1,\phi}} = \left| S\left(T_{2,\phi}\right) - S\left(\hat{T}_{2,\phi}\right)) \right|$$

Equation 6.4 $T_{2,\phi}$ represents ground truth T_2 values in the tissue of interest ϕ (IVD or cartilage), scaled by $S(x)$ (Equation 6.3), and $\hat{T}_{2,\phi}$ is the same for predicted T_2. Pixels corresponding to ϕ

are obtained from segmentation masks, the generation of which is described in section 6.3.3.3. For both L_{L_1} and $L_{L_{1,\phi}}$, L_1 norms were used instead of L_2 due to reduced sensitivity to outliers, leading to more stable trainings.

L_{SSIM} is an SSIM loss:

$$L_{SSIM} = 1 - SSIM$$

Equation 6.5 SSIM is the structural similarity index between predicted and target maps. $L_{Feature}$ is a feature-based loss function designed to retain sharper textures:

$$L_{Feature} = \left| VGG_{T_2} - VGG_{\hat{T}_2} \right|$$

Equation 6.6 VGG_{T_2} and $VGG_{\hat{T}_2}$ were the outputs of the 21st layer of a VGG-19 [198] network pretrained on ImageNet when fed resized and normalized target and predicted T_2 maps, respectively. Maps were resized to 224×224×1, concatenated with themselves along the channel axis to yield 224×224×3 inputs, and normalized such that the channels had mean pixel values of 0.485, 0.456 and 0.406, with standard deviations of 0.229, 0.224, and 0.225, respectively.

Loss component weightings were optimized through constrained random hyperparameter searches with the following ranges:

- Knee: λ_{L_1}=1, $\lambda_{L_{1,\phi}} = 50 - 150$, $\lambda_{SSIM} = 0 - 2$, $\lambda_{Feature} = 0 - 0.5$
- Hip: λ_{L_1}=1, $\lambda_{L_{1,\phi}} = 0 - 3$, $\lambda_{SSIM} = 0 - 2$, $\lambda_{Feature} = 0 - 1$
- Spine: λ_{L_1}=1, $\lambda_{L_{1,\phi}} = 1 - 10$, $\lambda_{SSIM} = 10 - 100$, $\lambda_{Feature} = 5 - 55$

6.3.3.3 Training and Segmentation Details

Scans of all three anatomies were split into training, validation and test sets as shown in Table 6.1. In the knee, cartilage was segmented manually. In the hip, cartilage was segmented manually for 4 central slices per volume. Segmentation in both was performed by research assistants trained by radiologists with over 20 years of experience. Since the hip dataset had substantially fewer segmented than unsegmented slices, the hip training set was bootstrapped

to equalize the number of slices with and without segmentations (1,068 bootstrapped slices). Finally, in the lumbar spine, IVDs were segmented with an ensemble of coarse-to-fine context memory (CFCM) networks [117]. To calculate performance metrics and implement ROI-specific training losses, these segmentation masks were leveraged to identify pixels in tissues of interest (cartilage or IVD).

Signal values were scaled per slice for the middle 95% of pixel values to fall between 0 and 500 for the knee and lumbar spine, and 0 and 100 for the hip; these ranges were optimized empirically. During training, imaging volumes were augmented with random translation (±10 pixels across phase and frequency directions) and random rotation (±5 degrees about slice direction). All models were trained with learning rate 0.001 and Adam optimizer on an NVIDIA Titan Xp 12 GB GPU with batch size of 1 so the model would fit on a single GPU. Separate pipelines were trained for all 3 anatomies at R=2, 3, 4, 6, 8, 10, and 12. For each pipeline, and at each trained R, a constrained random hyperparameter search was done for 15 iterations at 10 epochs per iteration to optimize λ_{L_1}, $\lambda_{L_{1,\phi}}$, λ_{SSIM}, and $\lambda_{Feature}$ for visual fidelity of predicted maps to ground truth. Visual fidelity was assessed in the search using NRMSE and Pearson's r in the tissue of interest [199].

$$(7)\ NRMSE = \left\| T_2 - \widehat{T_2} \right\|_{2,\phi} \left(\| T_2 \|_{2,\phi} \right)^{-1}$$

Equation 6.7 Calculation of NRMSE in tissues of interest, where T_2 are ground truth T_2 values and $\widehat{T_2}$ are predicted T_2 values. L_2 norms are calculated in both only for pixels in tissues of interest ϕ.

Final pipelines across all anatomies and R were trained using optimized parameter sets until validation loss did not decrease for 10 epochs. Key training details are summarized as part of Table 6.1.

6.3.4 Experiments

6.3.4.1 Loss Function Ablation Study

An ablation study is key to understand contributions of loss components. Given optimized loss function weights, every combination of loss components was ablated and corresponding models were retrained until validation loss no longer decreased. “No RNN” and “Reduced Parameters” networks were also trained while maintaining loss function components at optimized values to assess the utility of simpler architectures. NRMSE and Pearson’s correlation coefficient (r) were calculated in tissues of interest across the test set for original and ablated models to determine loss component contributions to performance. Pearson’s r was deemed an appropriate statistical test for this and subsequent experiments, as it is useful in assessing the linear relationship between related pairs of interval data. While no formal NRMSE test was done, it nonetheless allows for quantitative assessment of T_2 quantification quality and easy comparison with results from other approaches. NRMSE is reported ±1 standard deviation (s.d.); Pearson’s r was deemed significant in accordance with corresponding *P* values, α=0.001, 0.01, and 0.05. NRMSEs within tissues of interest of a given scan were also multiplied by mean T_2 values within the tissue of interest of that patient, generating T_2 value equivalents of error rates.

To more specifically evaluate the utility of the ROI-specific loss component, two loss function configurations from the ablation study were further analyzed at all R: no ROI-specific loss component ($\lambda_{L_{1,\phi}} = 0;\ \ \lambda_{L_1}, \lambda_{SSIM}, \lambda_{Feature} \neq 0$) and no ROI-specific or feature-based components ($\lambda_{L_{1,\phi}}, \lambda_{Feature} = 0;\ \ \lambda_{L_1}, \lambda_{SSIM} \neq 0$). These models were intended to represent baselines in which all loss functions were preserved except the ROI-specific component, and a standard reconstruction loss function of pixel and SSIM-based loss components, respectively. Pearson's r—evaluated in tissues of interest and globally—was calculated to determine the degree and significance of correlation between predicted maps and ground truth, both globally and within tissues of interest, α=0.001, 0.01, and 0.05.

6.3.4.2 Evaluation of Accelerated Acquisition Scheme Performance

Three versions of our pipeline (full pipeline, "No RNN," and "Reduced Parameters") were compared to state-of-the-art CS, DL, and DL/model-based solutions. At each R, MANTIS (54,413,056 trainable parameters) and MANTIS-GAN (54,413,056 [Generator] and 2,763,648 [Discriminator] trainable parameters) pipelines were trained using published network architectures, loss functions and undersampling strategies [134,139]. Loss function weightings for both were optimized through grid hyperparameter searches yielding the following: (MANTIS) λ_{data}=0.1, λ_{cnn}=1; (MANTIS-GAN) λ_{data}=0.1, λ_{cnn}=1, λ_{GAN}=0.01. To apply CS reconstruction, original MAPSS T_2-prepared images were Fourier transformed into coil-combined k-space, 1D-inverse Fourier transformed along the readout direction, and individual slices in $k_y - k_z$ reconstructed using an L_1 wavelet-based algorithm with regularization coefficient 0.001 [148]. CS reconstructed images were registered to the TE=0 ms echo time image using a 3D rigid

registration algorithm with a normalized mutual information criterion and fitted using Levenberg-Marquardt fitting to yield T_2 maps. Performance of these approaches and our proposed methods was evaluated through the following:

6.3.4.2.1 Comparison of Global and ROI-Specific Performance

To test for completeness of training, performance of our proposed pipelines was compared against state-of-the-art models that did not use ROI-specific components in predicting T_2 maps. Pearson's r (α=0.001, 0.01, and 0.05) was used to compare model performances and assess strength of correlations to ground truth T_2.

6.3.4.2.2 Standard Reconstruction Metrics

Performance was reported in tissues of interest with standard reconstruction metrics: NRMSE (mean ±1 s.d.) and Pearson's r (α=0.001, 0.01, and 0.05). NRMSEs were also converted into T_2 value equivalents by tissue compartment as in the ablation study.

6.3.4.2.3 T_2 Value Retention

Fidelity of predicted maps to ground truth T_2 was also assessed. First, predicted and ground truth T_2 values were compared across tissues of interest within the test set (mean ±1 s.d.), generating violin plots for all three anatomies with overlaid boxplots for T_2 value distribution comparison. T_2 agreement was also assessed through Bland-Altman analysis.

6.3.4.2.4 Texture Retention

Gray Level Co-Occurrence Matrix (GLCM) [200] metrics were used to assess texture retention within tissues of interest. GLCM contrast and dissimilarity are maximized by large local pixel value changes and thus by sharper textures. GLCM homogeneity is maximized by small local pixel value changes, while GLCM energy and angular second moment (ASM) are maximized by few total pixel values within an image; hence, all three are maximized by smoothness. For each anatomy and R, we calculated these texture metrics at 4 orientations (θ=0°, 45°, 90° and 135°; d=1 pixel) and averaged across all orientations. Finally, we calculated intraclass correlation coefficients (ICCs) for all metrics with respect to ground truth (two-way mixed effects, single rater [201]) and reported 95% ICC confidence intervals (α=0.001, 0.01, and 0.05). These tests were chosen as appropriate, as they assess both reliability and agreement of associated metrics, and in this use case, individual GLCM metric values themselves are considered the only rater, justifying the ICC test type selected.

6.3.4.3 Repeatability Study

To assess the robustness of pipelines to different datasets, two additional splits of the knee, hip and spine datasets were made, ensuring no patient was part of multiple validation and/or test datasets and that all scans from a given patient were only in one of training, validation and test for each split (folds 2 and 3 in Supp. Table A.2, where fold 1 is the original split). Additional hyperparameters searches optimized loss function weights on the two new splits. Optimized loss weights and corresponding T_2 quantification and texture retention performance for each splits is presented at all tested R in the same manner as for the primary split.

6.3.4.4 Raw Multicoil Data Assessment

An in-house pipeline was developed that leveraged GE Orchestra 1.10 and other postprocessing tools to reconstruct coil-combined images from raw k-space data. As a proof of concept, knee MAPSS scans were performed on 3 volunteers, hip scans for 2, and lumbar spine for 2, all using the acquisition parameters listed for the retrospective datasets used for algorithm training, with raw k-space data saved for all. Multicoil k-space data (after ARC for knee and hip) was undersampled with the same center-weighted Poisson disc pattern described earlier, with each coil seeing the same undersampling pattern and k_y-k_z lines being shared across different T_2 weighted echo time k-spaces as previously described. Coil-combined images resulting from undersampled multi-coil data at all tested R were fed through corresponding post-processing pipelines to predict T_2 map appearance. A radiologist with 2 years of experience segmented knee cartilage, hip cartilage, and intervertebral discs from these acquisitions, allowing for visualizations of predicted T_2 maps and NRMSE calculations in ROIs.

6.4 Results

6.4.1 Ablation Study Results

Voxel-wise performance metrics for ablation study models at R=8 are shown in Supp. Table A.3, with T_2 value NRMSE equivalents in Supp. Table A.4. Within the knee and hip, all loss components were necessary to obtain the optimal combination of high Pearson's r and low NRMSE in cartilage. For the lumbar spine, while all loss components proved vital in maximizing Pearson's r and minimizing NRMSE in IVDs, performance improved when the initial recurrent

network was omitted. Though quantitative analysis is shown for all three pipeline versions in subsequent experiments, the full model is designated as best for knee and hip, and the no RNN for the spine.

ROI-specific and global assessments of best models and corresponding models trained without an ROI-specific loss ($\lambda_{1,\phi}$=0) and models trained with a generic loss ($\lambda_{1,\phi}$=0, λ_{Feat}=0) are shown in Supp. Table A.5. In the knee and hip, across nearly all R, ROI-specific loss addition leads to improved correlations between predicted and ground truth cartilage T_2, with diminished performance globally. In the lumbar spine, which was trained with a substantially fewer batches than the knee and hip pipelines, these trends were inconsistent across tested R. Example predictions and ground truth for one slice of a patient in each pipeline are shown in Supp. Fig. A.3, showing that patterns of local T_2 value elevations in cartilage and IVDs are better preserved with an ROI-specific loss as opposed to pipelines trained without the loss component.

6.4.2 Visuals of Network Performance and Comparison with State-of-the-Art Models

Predicted T_2 maps are displayed at select R for knee, hip and lumbar spine models in Figure 6.2 for our three pipelines and three methods from the literature. In knee, hip, and lumbar spine, T_2 quantification performance is strongest with our proposed methods, maintaining low error rates, showing promising results compared with state-of-the-art methods through R=10. Optimal architecture performances are further explored in Figures 6.3-6.5**.** As shown in Figure 6.3, predicted T_2 knee maps retained strong fidelity to ground truth within tibiofemoral joint cartilage. Patterns within predicted maps became slightly more diffuse as R increased to 10, as

indicated by a slight rise in NRMSE for cartilage in the slice, but visually, T_2 values and map patterns are preserved. As seen in Figure 6.4, hip predicted maps preserve T_2 values well in femoral and acetabular cartilage through R=10, although T_2 patterns become more diffuse by R=10. Figure 6.5 shows T_2 map predictions in the lumbar spine. The L4-L5 IVD is shown in more detail, where T_2 quantification performance was acceptable at R=3, moderate at R=6, and worse at R=10, as indicated by rising IVD NRMSEs.

ROI and global performance comparisons of our selected pipelines against state-of-the-art approaches are in Supp. Table A.5. Across piplines trained with relatively large dataset (knee and hip), DL and model-based approaches (MANTIS and MANTIS-GAN) outperformed our proposed pipeline globally, but within cartilage ROIs, our pipeline exhibited stronger Pearson's r at each tested R. These trends were not as strong in the lumbar spine pipelines, possibly owing to the randomness of training with a smaller dataset. Global and ROI-specific T_2 predictions are further visualized in Supp. Fig. A.4, showing predicted T_2 values exhibit substantially more visual fidelity to ground truth and lower NRMSE in state-of-the-art models compared to our pipeline, but a reversal of that trend in cartilage. In the lumbar spine, at some but not all R, those trends held, yielding similar conclusions to the Pearson's r analysis.

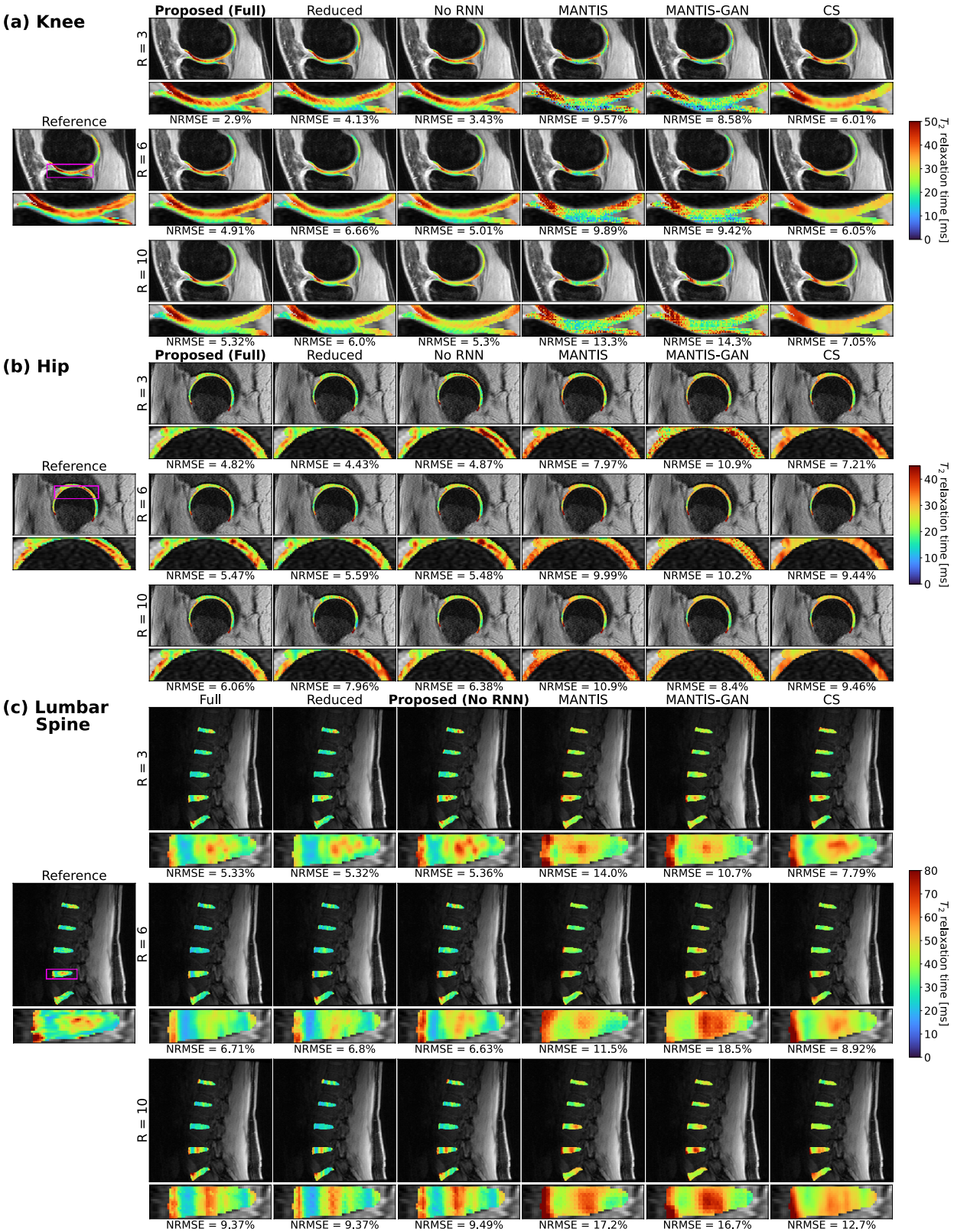

(a) Knee
Proposed (Full)
Reduced
No RNN
MANTIS
MANTIS-GAN
CS
R = 3
Reference
R = 6
R = 10
NRMSE = 2.9%
NRMSE = 4.13%
NRMSE = 3.43%
NRMSE = 9.57%
NRMSE = 8.58%
NRMSE = 6.01%
NRMSE = 4.91%
NRMSE = 6.66%
NRMSE = 5.01%
NRMSE = 9.89%
NRMSE = 9.42%
NRMSE = 6.05%
NRMSE = 5.32%
NRMSE = 6.0%
NRMSE = 5.3%
NRMSE = 13.3%
NRMSE = 14.3%
NRMSE = 7.05%
T_2 relaxation time [ms]
(b) Hip
Proposed (Full)
Reduced
No RNN
MANTIS
MANTIS-GAN
CS
R = 3
Reference
R = 6
R = 10
NRMSE = 4.82%
NRMSE = 4.43%
NRMSE = 4.87%
NRMSE = 7.97%
NRMSE = 10.9%
NRMSE = 7.21%
NRMSE = 5.47%
NRMSE = 5.59%
NRMSE = 5.48%
NRMSE = 9.99%
NRMSE = 10.2%
NRMSE = 9.44%
NRMSE = 6.06%
NRMSE = 7.96%
NRMSE = 6.38%
NRMSE = 10.9%
NRMSE = 8.4%
NRMSE = 9.46%
T_2 relaxation time [ms]
(c) Lumbar Spine
Full
Reduced
Proposed (No RNN)
MANTIS
MANTIS-GAN
CS
R = 3
Reference
R = 6
R = 10
NRMSE = 5.33%
NRMSE = 5.32%
NRMSE = 5.36%
NRMSE = 14.0%
NRMSE = 10.7%
NRMSE = 7.79%
NRMSE = 6.71%
NRMSE = 6.8%
NRMSE = 6.63%
NRMSE = 11.5%
NRMSE = 18.5%
NRMSE = 8.92%
NRMSE = 9.37%
NRMSE = 9.37%
NRMSE = 9.49%
NRMSE = 17.2%
NRMSE = 16.7%
NRMSE = 12.7%
T_2 relaxation time [ms]

Figure 6.2 Comparison of predicted T_2 maps with ROI-specific methodologies to past approaches. (a) Predicted T_2 maps in knee cartilage for a representative patient within test set. T_2 quantification performance was best in pipelines trained with ROI-specific losses (Full Model, Reduced Parameters, and No RNN), where strong fidelity to T_2 values and patterns of local elevations within cartilage were maintained through R=10, while other tested approaches did a poorer job in predicting T_2 values in these maps. (b) Predicted hip cartilage T_2 maps showed similar trends, where performance of the full model was especially strong, showing low T_2 quantification error and better retaining local T_2 elevations through R=10 than other approaches. (c) Predicted T_2 maps in lumbar spine IVDs show higher T_2 quantification errors than in hip and knee cartilage, but ROI-specific loss pipelines best preserved map textures and values.

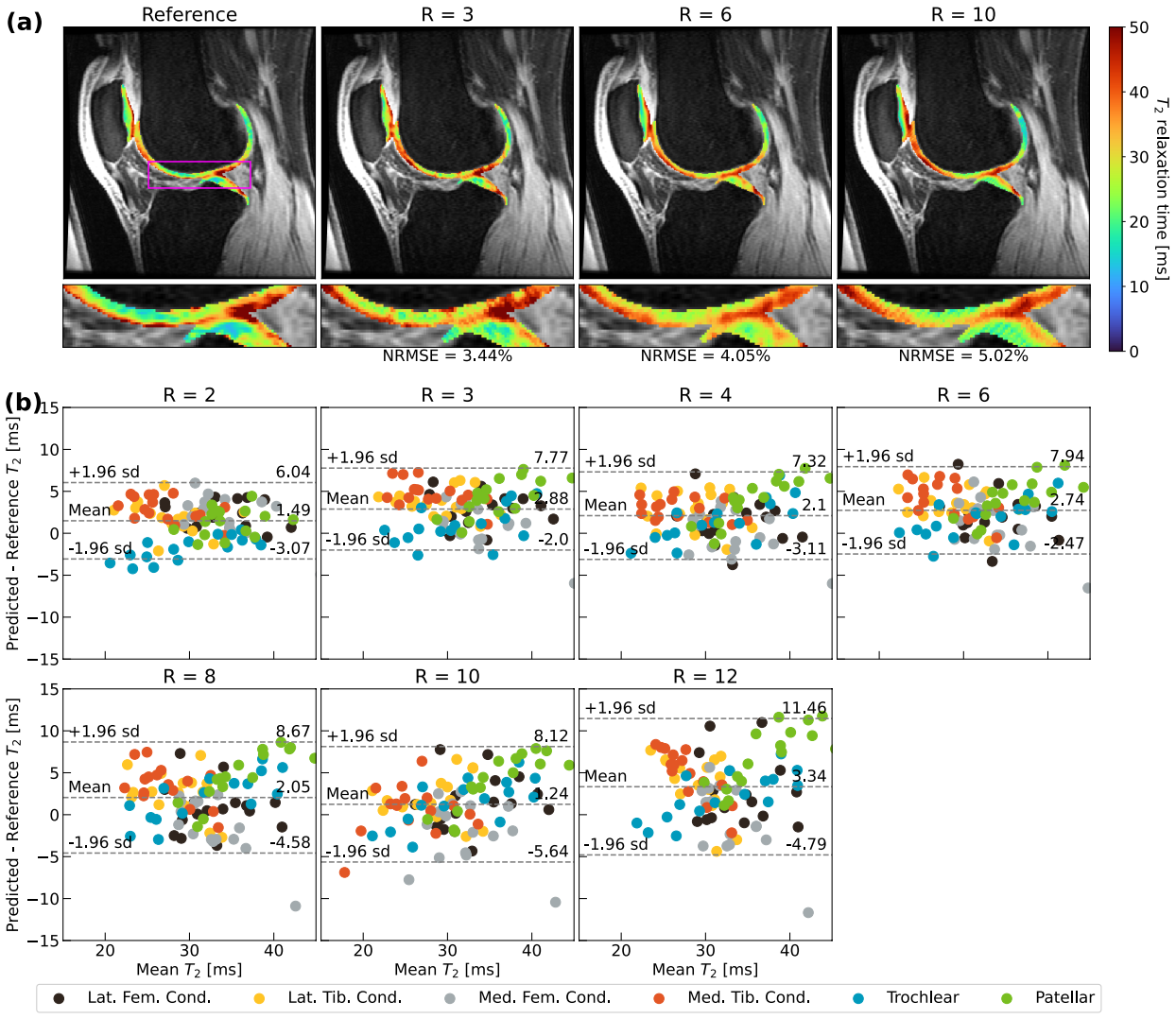


Figure 6.3 T_2 quantification performance of optimal ROI-specific pipeline in knee cartilage. T_2 quantification performance of optimal ROI-specific pipeline in knee cartilage. (a) Visual pipeline

performance within the knee for a representative patient, with corresponding NRMSEs for cartilage in the predicted T_2 map slice. Performance remains strong through R=10, maintaining T_2 patterns in the medial tibiofemoral cartilage, indicating pipeline utility. Predicted maps generated by the network are masked using a cartilage segmentation mask and superimposed on the ground truth, fully sampled TE=0 ms MAPSS echo time image. (b) Bland-Altman plots for all scans within test set for which multiclass cartilage compartment segmentations were available (n=16, 6 cartilage compartments for each). Predicted T_2 values demonstrate minimal bias and tight limits of agreement across most tested R, with best performance coming from patellofemoral cartilage.

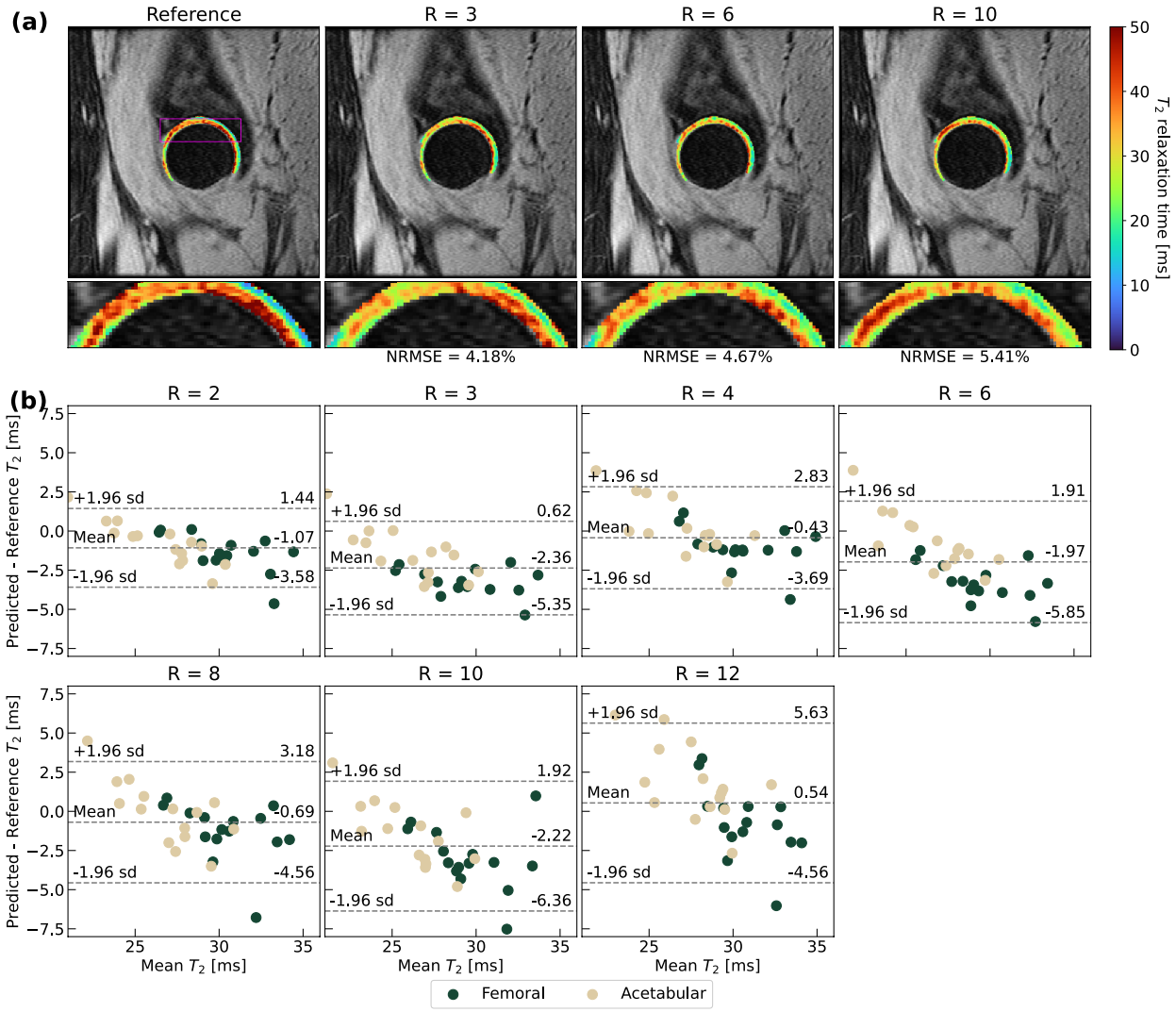


Figure 6.4 T_2 quantification performance of optimal ROI-specific pipeline in hip cartilage. T_2 quantification performance of optimal ROI-specific pipeline in hip cartilage. (a) Visual pipeline performance within the hip for a representative patient, with corresponding NRMSEs for cartilage in the predicted T_2 map slice. Predicted maps are masked using a cartilage

segmentation mask and superimposed on the ground truth, fully sampled TE=0 ms MAPSS echo time image. For this patient, T_2 patterns maintain through R=10, although local T_2 elevations are more diffusely predicted at higher R. (b) Bland-Altman plots for all scans within test set (n=15, 2 cartilage compartments for each). Plots demonstrate very limited bias and even tighter limits of agreement from R=2 through R=12 than knee pipeline, showing hip pipeline effectiveness in reproducing T_2 values from accelerated MAPSS acquisitions.

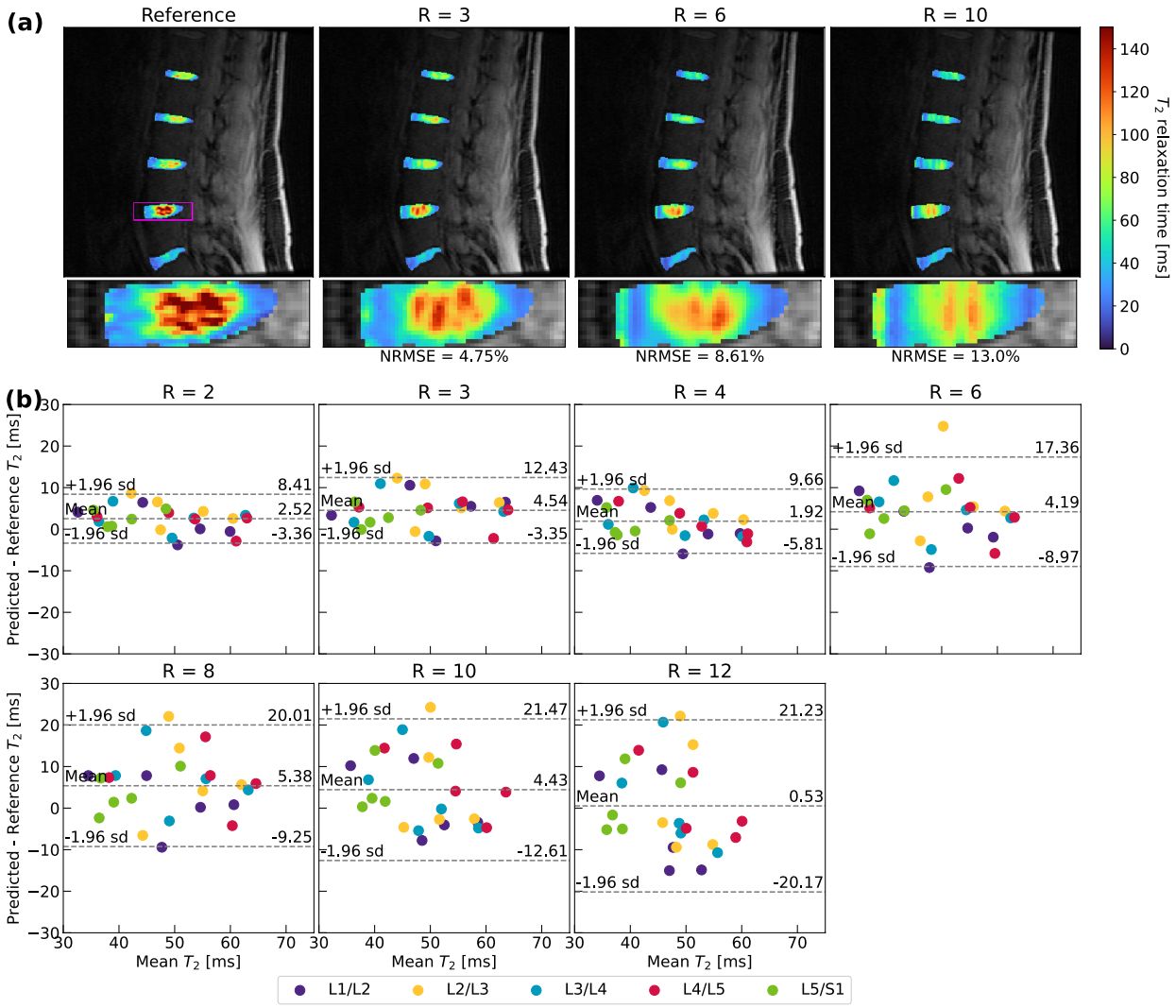


Figure 6.5 T_2 quantification performance of optimal ROI-specific pipeline in lumbar spine intervertebral discs. T_2 quantification performance of optimal ROI-specific pipeline in lumbar spine intervertebral discs. (a) Visual pipeline performance within the lumbar spine IVDs for a representative patient, with corresponding NRMSEs for IVDs in the predicted T_2 map slice. Predicted maps are masked using an IVD segmentation mask and superimposed on the ground truth, fully sampled TE=0 ms MAPSS echo time image. Network performance is best through R=6, after which local T_2 elevations are diffuse and underestimated. (b) Bland-Altman plots for all scans within test set (n=5, 5 IVDs plotted for each if segmentation of disc available). T_2 value

predictions reflect some bias and fairly wide limits of agreement, particularly above R=4. These results indicate progress but the need for improvement. Smaller lumbar spine dataset and test set size are likely responsible for poorer model when compared to hip and knee performance, as well as the relatively smaller number of slices in k_z, which exacerbates undersampling effects.

6.4.3 Evaluation of T_2 Quantification Performance and Comparison with State-of-the-Art Models

6.4.3.1 Voxel-wise T_2 Evaluation Fidelity

Pearson's r and NRMSE across all anatomies and R for our approaches and state-of-the-art methods are in Table 6.2. T_2 value NRMSE equivalents are in Supp. Table A.6. For all anatomies and across nearly all R, T_2 quantification performance is strongest in our methods, particularly in the No RNN and full model pipelines, compared to state-of-the-art models.

An exhaustive examination of knee T_2 quantification performance, stratified by cartilage compartments, is in Supp. Tables A.7 and A.8. For the full model, across all cartilage compartments, T_2 estimation errors remained under 10% through R=10 across all cartilage compartments while Pearson's r ranged from 0.748 at R=2 to 0.491 at R=12, indicating strong correlations between predictions and ground truth at R=2 and moderate correlations through R=12 [202]. For some cartilage compartments and R, performance was stronger in the No RNN pipeline. Interestingly, quantification performance was strongest in patellofemoral joint cartilage, generally exhibiting lower NRMSE and stronger correlations. Our ROI-specific loss pipelines outperformed state-of-the-art models in each cartilage compartment.

Supp. Tables A.9 and A.10 show hip T_2 quantification performance across cartilage compartments. As in the knee, quantification performance was strong, with error rates across all cartilage under 9% through R=12 for the no RNN and full model pipelines. While the no RNN pipeline had stronger quantification errors, the full model had higher Pearson's r, which ranged from 0.794 at R=2 to 0.517 at R=12, showing strong correlations between predictions and ground truth through R=3 and moderate correlations through R=12. T_2 quantification performance was slightly stronger in femoral than acetabular cartilage. Our pipelines again outperformed state-of-the-art models in each cartilage compartment.

Supp. Tables A.11 and A.12 show lumbar spine T_2 quantification performance, which was mixed. Pearson's r across all discs was very high, ranging from 0.884 at R=2 to 0.643 at R=12 for the no RNN model, indicating strong correlations through R=8 and moderate correlations through R=12 to ground truth. That said, IVD error rates were markedly higher across all R than in hip and knee cartilage, ranging from 4.86% to 18.8%. Though there was some volatility, error rates and Pearson's r generally showed poorest T_2 quantification in L1/L2 and L2/L3 discs. Through R=8, ROI-specific loss pipelines outperformed state-of-the-art models at nearly all disc levels, with stronger Pearson's r in most IVD levels through R=12.

Table 6.2 ROI-specific model performance in standard metrics from R=2 through R=12. Performances of pipelines trained with ROI-specific losses and other state-of-the-art methods in T_2 quantification error rates in knee cartilage, hip cartilage, and lumbar spine IVDs. NRMSEs are reported ±1 s.d., and Pearson's r is reported with significances as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (knee: n=90; hip: n=15; lumbar spine: n=5). Across all anatomies, performances were strongest in ROI-specific loss pipelines (Full Model, Reduced Parameters, and No RNN): in the knee, the No RNN and Full Model pipelines particularly excelled across all tested R; in the hip, the No RNN pipeline was strong in maintaining minimal T_2 quantification errors, while the Full Model and Reduced Parameters models had strongest correlations between predicted maps and ground truth; in the lumbar spine, the No RNN pipeline especially had strong T_2

quantification performance. Performance in the knee and hip pipelines is strong and below clinically significant T_2 changes across nearly all tested R, while Pearson's r indicates strong T_2 value preservation in the lumbar spine through R=6. T_2 quantification performance is thus promising in all three pipelines, but particularly for the knee and hip.

	R	Metric	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
Knee Cartilage	2	NRMSE	5.52 ± 1.25	6.07 ± 3.21	**4.76 ± 1.78**	14.4 ± 2.85	13.5 ± 3.3	8.92 ± 3.2
		Pearson's r	0.748***	0.736***	**0.807*** **	0.587***	0.611***	0.620***
	3	NRMSE	6.52 ± 2.17	7.18 ± 3.08	**6.39 ± 2.59**	16.5 ± 3.43	15.1 ± 2.89	9.92 ± 3.23
		Pearson's r	0.695***	0.668***	**0.722*** **	0.467***	0.502***	0.559***
	4	NRMSE	**7.54 ± 2.96**	9.56 ± 5.47	7.56 ± 3.19	16.6 ± 3.73	15.7 ± 4.5	11.8 ± 3.73
		Pearson's r	0.651***	0.637***	**0.677*** **	0.451***	0.467***	0.486***
	6	NRMSE	**8.09 ± 2.65**	10.7 ± 6.67	8.44 ± 3.49	15.2 ± 2.33	16.3 ± 3.91	12.4 ± 4.1
		Pearson's r	0.612***	0.610***	**0.629*** **	0.397***	0.378*	0.445***
	8	NRMSE	**8.94 ± 2.66**	9.59 ± 3.83	10.1 ± 4.42	16.7 ± 2.73	17.3 ± 2.39	12.9 ± 3.93
		Pearson's r	0.585***	0.574***	**0.609*** **	0.352***	0.364**	0.410***
	10	NRMSE	9.77 ± 3.44	10.2 ± 3.61	**9.35 ± 3.5**	17.6 ± 2.48	16.7 ± 3.44	13.4 ± 3.76
		Pearson's r	0.555***	0.514***	**0.565*** **	0.327***	0.333***	0.386***
	12	NRMSE	10.7 ± 2.32	**9.93 ± 3.76**	10.5 ± 3.37	18.2 ± 4.5	20.5 ± 5.58	13.4 ± 3.96
		Pearson's r	0.491***	**0.545*** **	0.511***	0.290***	0.287***	0.381***
Hip Cartilage	2	NRMSE	3.97 ± 1.03	4.1 ± 1.1	**3.79 ± 0.807**	4.58 ± 0.993	8.21 ± 1.42	14.8 ± 2.78
		Pearson's r	**0.794*** **	0.782***	0.770***	0.716***	0.514***	0.310***
	3	NRMSE	6.53 ± 1.63	5.63 ± 1.68	**5.25 ± 1.13**	6.41 ± 1.31	10.0 ± 1.57	12.9 ± 3.15
		Pearson's r	0.705***	**0.726*** **	0.703***	0.596***	0.372***	0.332***
	4	NRMSE	6.15 ± 1.01	6.17 ± 1.47	**5.84 ± 0.891**	7.33 ± 1.67	9.97 ± 1.74	11.8 ± 2.03
		Pearson's r	0.646***	**0.665*** **	0.648***	0.510***	0.333***	0.339***
	6	NRMSE	8.1 ± 1.85	8.22 ± 2.06	**7.48 ± 1.52**	8.63 ± 2.32	9.68 ± 1.92	11.8 ± 2.14
		Pearson's r	0.587***	**0.597*** **	0.570***	0.382***	0.321***	0.334***
	8	NRMSE	6.97 ± 1.93	**6.33 ± 1.33**	6.98 ± 1.45	10.2 ± 2.72	12.0 ± 2.64	10.5 ± 2.3
		Pearson's r	**0.598*** **	0.588***	0.558***	0.334***	0.237***	0.347***
	10	NRMSE	8.99 ± 2.65	**8.12 ± 1.28**	8.7 ± 3.46	9.74 ± 2.24	10.5 ± 1.91	10.2 ± 2.4
		Pearson's r	**0.558*** **	0.534***	0.522***	0.279***	0.268***	0.335***
	12	NRMSE	7.75 ± 1.5	8.27 ± 2.19	**7.34 ± 1.38**	9.74 ± 2.23	11.5 ± 2.36	10.3 ± 2.52
		Pearson's r	**0.517*** **	0.566***	0.512***	0.280***	0.228***	0.349***
Lumbar Spine IVD	2	NRMSE	6.71 ± 1.7	6.86 ± 1.57	**4.86 ± 1.16**	8.78 ± 2.08	8.95 ± 1.91	10.1 ± 3.06
		Pearson's r	0.865***	0.866***	**0.884*** **	0.784***	0.785***	0.802***
	3	NRMSE	9.92 ± 2.39	8.76 ± 2.16	**7.13 ± 1.69**	11.0 ± 1.17	11.3 ± 1.74	9.48 ± 1.4
		Pearson's r	**0.836*** **	0.823***	0.832***	0.717***	0.712***	0.777***
	4	NRMSE	10.3 ± 3.02	9.73 ± 3.07	**7.42 ± 1.1**	12.1 ± 1.24	12.6 ± 1.35	11.3 ± 2.31
		Pearson's r	0.799***	0.813***	**0.819*** **	0.680***	0.671***	0.723***
	6	NRMSE	12.1 ± 3.58	12.2 ± 4.11	**10.3 ± 3.31**	15.6 ± 2.65	12.1 ± 1.9	12.0 ± 2.76
		Pearson's r	**0.776*** **	0.764***	0.771***	0.660***	0.658***	0.728***
	8	NRMSE	13.4 ± 3.89	13.0 ± 2.63	**12.0 ± 3.07**	13.2 ± 1.42	12.7 ± 1.7	12.8 ± 2.53
		Pearson's r	**0.742*** **	0.723***	**0.742*** **	0.631***	0.645***	0.695***

	R	Metric	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
IVDs	10	NRMSE	15.3 ± 3.22	14.8 ± 2.78	14.8 ± 2.26	13.8 ± 1.57	**13.2 ± 1.81**	15.0 ± 3.77
		Pearson's r	0.695***	**0.700*** **	0.672***	0.647***	0.636***	0.648***
	12	NRMSE	18.1 ± 1.95	23.1 ± 2.71	18.8 ± 2.76	14.8 ± 3.03	**14.1 ± 1.88**	24.8 ± 11.2
		Pearson's r	**0.664*** **	0.320***	0.643***	0.651***	0.614***	0.586***

6.4.3.2 T_2 Value Retention on Region of Interest Averages

Bland-Altman plots are provided for the knee, hip and lumbar spine in Figures 6.3-6.5. In knee and hip, T_2 values are predicted with minimal bias with respect to ground truth. The ±1.96 s.d. limits of agreement were less than approximately ±6 ms with mean biases under ±3 ms through R=8 for knee cartilage (Figure 6.3). Among cartilage compartments, predictions in trochlear and patellar cartilage showed the least bias, while tibiofemoral cartilage T_2 was generally slightly overestimated. In the hip (Figure 6.4), ±1.96 s.d. limits of agreement were less than approximately ±5 ms with mean biases under ±3 ms through R=12, although T_2 quantification performance was similar across femoral and acetabular cartilage. In the lumbar spine (Figure 6.5), limits of agreement were considerably wider than the hip and knee pipelines, particularly above R=4. While the line of equality was contained in these limits at all R, spine pipelines generally overestimated T_2 values. While at some particular R, a disc level saw poorer T_2 quantification than others (i.e. L2/L3 at R=6), on balance, predicted maps yielded similar bias and error across all discs.

Supp. Figure A.5 shows T_2 value distributions in violin and boxplots. Plots reveal minimal bias in hip cartilage predicted T_2 maps and slight but limited bias towards overestimating T_2 in knee cartilage. In the lumbar spine, more volatility was observed in predicted T_2 distributions, likely

due to small test set size (n=5), but at least through R=6, these deviations had limited magnitude.

6.4.3.3 Texture Retention

ICCs ±1 s.d. for GLCM metrics are in Table 6.3 for our best performing pipelines: no RNN and full model. In knee cartilage, ICCs showed significant correlations between predicted and ground truth GLCM metrics at all R for smooth textures and many R for sharp textures, indicating good to excellent reliability in preserving smooth textures (ASM and energy) at all R and moderate reliability in preserving sharper textures at low R (dissimilarity). In hip cartilage, ICCs showed significant correlations across all R in preserving smooth textures, and at low to moderate R for sharper textures. Reliability in smooth texture preservation ranged from good to excellent for all R and moderate for sharper textures at low to medium R. In both knee and hip cartilage, the full pipeline saw substantially higher GLCM ICCs for smooth and sharper texture across nearly all R. Within the lumbar spine, ICCs were significant across nearly all R for smoother textures. While ICCs were reasonable high for some R in contrast metrics, confidence intervals were wide, limiting findings of significant correlations. ICCs showed moderate to excellent reliability in preserving smoother textures, and poor to moderate reliability for sharper textures. For the spine, the No RNN model yielded optimal texture retention.

Table 6.3 Texture retention analysis in No RNN and Full Model pipelines. Intraclass correlation coefficients (ICCs) of Gray Level Co-Occurrence Matrix (GLCM)-based metrics. Contrast and dissimilarity are most sensitive to sharper image textures, while homogeneity, ASM, and energy are most sensitive to smoother image textures. Significance in correlations is noted as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (knee: n=16; hip: n=15; lumbar spine: n=5). In the knee and hip, Full Model pipelines outperformed No RNN versions in retention of smooth and sharp textures. In the lumbar spine, the No RNN pipeline outperformed the Full Model version, possibly because the smaller lumbar spine dataset size made training a larger network with a

multi-component loss more difficult. In conjunction with standard reconstruction metrics, the Full Model pipeline was selected as the best knee and hip model, whereas the No RNN pipeline was selected as the best lumbar spine model. Top models in all anatomies preserved smoother textures at nearly all tested R, while dissimilarity texture metrics showed sharper textures were significantly correlated with ground truth and preserved in the knee and hip at low to medium R. In the lumbar spine, mean ICCs for sharper textures at many tested R also were high, but small dataset size likely led to wide standard deviations, preventing significant conclusions from being reached. Many textures are preserved in predicted T_2 maps, particularly knee and hip.

		R	GLCM Texture Metric				
			Contrast	**Dissimilarity**	**Homogeneity**	**ASM**	**Energy**
Knee	Full Model	2	0.307 ± 0.18**	0.638 ± 0.12***	0.734 ± 0.09***	0.966 ± 0.015***	0.954 ± 0.02***
		3	0.153 ± 0.2	0.521 ± 0.15***	0.735 ± 0.09***	0.962 ± 0.015***	0.95 ± 0.02***
		4	0.11 ± 0.2	0.387 ± 0.17***	0.61 ± 0.12***	0.973 ± 0.01***	0.95 ± 0.02***
		6	0.0667 ± 0.2	0.22 ± 0.19*	0.382 ± 0.17***	0.97 ± 0.015***	0.94 ± 0.025***
		8	0.061 ± 0.2	0.111 ± 0.2	0.0615 ± 0.2	0.952 ± 0.02***	0.9 ± 0.04***
		10	0.0594 ± 0.2	0.218 ± 0.19*	0.307 ± 0.18**	0.961 ± 0.015***	0.928 ± 0.03***
		12	0.0032 ± 0.2	-0.066 ± 0.2	-0.178 ± 0.19	0.927 ± 0.03***	0.861 ± 0.055***
	No RNN	2	0.455±0.16***	0.599 ± 0.13***	0.32 ± 0.18***	0.898 ± 0.04***	0.904 ± 0.04***
		3	0.394±0.17***	0.523 ± 0.15***	0.383 ± 0.17***	0.709 ± 0.11***	0.802 ± 0.07***
		4	0.262 ± 0.18**	0.305 ± 0.18**	0.244 ± 0.18**	0.646 ± 0.12***	0.754 ± 0.09***
		6	0.103 ± 0.2	0.0574 ± 0.2	0.061 ± 0.2	0.874 ± 0.045***	0.869 ± 0.05***
		8	0.0645 ± 0.2	0.0411 ± 0.2	0.0435 ± 0.2	0.922 ± 0.03***	0.911 ± 0.035***
		10	0.0474 ± 0.2	0.0382 ± 0.2	0.093 ± 0.2	0.92 ± 0.035***	0.913 ± 0.035***
		12	0.0568 ± 0.2	0.0315 ± 0.2	0.0885 ± 0.2	0.818 ± 0.065***	0.862 ± 0.055***
Hip	Full Model	2	0.312 ± 0.34*	0.633 ± 0.23***	0.837 ± 0.12***	0.945 ± 0.04***	0.957 ± 0.035***
		3	0.369 ± 0.32*	0.671 ± 0.21***	0.816 ± 0.14***	0.976 ± 0.02***	0.98 ± 0.015***
		4	0.328 ± 0.33*	0.597 ± 0.25***	0.801 ± 0.15***	0.957 ± 0.035***	0.954 ± 0.04***
		6	0.235 ± 0.35	0.475 ± 0.3**	0.645 ± 0.23***	0.939 ± 0.05***	0.941 ± 0.045***
		8	0.199 ± 0.36	0.487 ± 0.28**	0.823 ± 0.13***	0.923 ± 0.06***	0.933 ± 0.055***
		10	0.127 ± 0.36	0.308 ± 0.34	0.48 ± 0.29**	0.862 ± 0.11***	0.855 ± 0.11***
		12	0.198 ± 0.36	0.38 ± 0.32*	0.523 ± 0.28**	0.927 ± 0.06***	0.914 ± 0.07***
	No RNN	2	0.285 ± 0.34	0.399 ± 0.32*	0.406 ± 0.31*	0.855 ± 0.11***	0.841 ± 0.12***
		3	0.15 ± 0.36	0.241 ± 0.35	0.292 ± 0.34	0.867 ± 0.1***	0.85 ± 0.12***
		4	0.113 ± 0.36	0.202 ± 0.36	0.282 ± 0.34	0.836 ± 0.12***	0.813 ± 0.14***
		6	0.0394 ± 0.36	0.0504 ± 0.36	0.0785 ± 0.36	0.793 ± 0.15***	0.767 ± 0.16***
		8	0.0229 ± 0.37	0.000593 ± 0.37	-0.0583 ± 0.37	0.682 ± 0.21***	0.653 ± 0.22***
		10	-0.00292 ± 0.36	-0.0328 ± 0.37	-0.196 ± 0.36	0.644 ± 0.23***	0.621 ± 0.24***
		12	-0.00208 ± 0.37	-0.0312 ± 0.36	-0.0646 ± 0.36	0.712 ± 0.2***	0.687 ± 0.2***
Lumbar Spine	Full Model	2	0.557 ± 0.7	0.695 ± 0.62	0.744 ± 0.57*	0.892 ± 0.35**	0.923 ± 0.27**
		3	0.499 ± 0.73	0.615 ± 0.67	0.644 ± 0.66	0.819 ± 0.48*	0.872 ± 0.39*
		4	0.236 ± 0.8	0.421 ± 0.76	0.497 ± 0.73	0.67 ± 0.64	0.775 ± 0.54*
		6	0.341 ± 0.78	0.428 ± 0.76	0.262 ± 0.8	0.566 ± 0.7	0.67 ± 0.64*
		8	0.0633 ± 0.81	0.152 ± 0.8	0.276 ± 0.79	0.685 ± 0.62	0.728 ± 0.58*
		10	-0.0393 ± 0.81	-0.0631 ± 0.81	-0.0699 ± 0.81	0.403 ± 0.76	0.479 ± 0.74*
		12	-0.0697 ± 0.81	-0.156 ± 0.8	-0.424 ± 0.76	0.16 ± 0.8	0.198 ± 0.8*

		R	GLCM Texture Metric				
			Contrast	Dissimilarity	Homogeneity	ASM	Energy
Lumbar Spine	No RNN	2	0.496 ± 0.73	0.731 ± 0.58*	0.883 ± 0.37**	0.967 ± 0.14***	0.975 ± 0.11***
		3	0.357 ± 0.78	0.615 ± 0.67	0.807 ± 0.5*	0.909 ± 0.31**	0.934 ± 0.24**
		4	0.336 ± 0.78	0.607 ± 0.68	0.771 ± 0.54*	0.874 ± 0.38*	0.91 ± 0.31**
		6	0.307 ± 0.78	0.53 ± 0.72	0.604 ± 0.68	0.903 ± 0.32**	0.916 ± 0.29**
		8	0.2 ± 0.8	0.4 ± 0.76	0.59 ± 0.68	0.847 ± 0.44*	0.871 ± 0.39*
		10	0.0696 ± 0.81	0.184 ± 0.8	0.386 ± 0.76	0.692 ± 0.62	0.726 ± 0.59*
		12	0.0157 ± 0.82	0.0858 ± 0.81	0.325 ± 0.78	0.561 ± 0.7	0.591 ± 0.68*

6.4.4 Repeatability Study

Optimal loss weightings from hyperparameter searches on the two additional splits are in Supp. Table A.13. Results of trainings on additional splits in T_2 quantification error, Pearson's r, and texture metrics are in Supp. Tables A.14-A.16. In the knee and hip pipelines, experiments show comparable results across all folds for these metrics. In the lumbar spine, Pearson's r exhibited similar values across all folds, but in some cases, mean texture metric ICCs and NRMSEs exhibited substantial differences. However, confidence intervals were very wide for ICCs and NRMSEs in the lumbar spine, likely due to limited test set size (n=5).

6.4.5 Raw Multicoil Data Assessment

Supp. Fig. A.6 shows T_2 maps predicted from our proposed pipelines on retrospectively undersampled raw k-space data. In the knee, T_2 quantification errors were low through R=12, with local T_2 elevations preserved and little dip in performance compared to corresponding retrospectively undersampled coil-combined knee data. In the hip, T_2 quantification errors were low, with local T_2 elevations reproduced at most R; while performance at higher R matched expected performance from coil-combined experiments, lower R quantification errors were slightly higher. Performance was more volatile in the lumbar spine, where through R=4, T_2

quantification errors matched expected results and local T_2 patterns were generally preserved, but performance degraded substantially above R=4.

6.5 Discussion and Conclusions

In this work, we present data-driven pipelines that leverage recurrent UNet architectures and multi-component losses to accelerate MAPSS T_2 mapping for anatomies where a subset of tissues is of particular clinical interest. By image processing and standard reconstruction metrics, through R=10, our knee pipelines retained fidelity to T_2 values with tight limits of agreement, preserving smooth textures with good to excellent reliability and sharper ones with moderate reliability for most tested R. While the no RNN pipeline delivered lower NRMSEs and higher Pearson's r across many cartilage compartments and R than full model, its texture retention was poorer, making the full model better suited to preserve small, key diagnostic features. In hip cartilage, predicted maps retained T_2 fidelity through R=12 with tight limits of agreement, preserved smooth textures with good to excellent agreement across tested R, and maintained sharper textures at low to moderate R. As with the knee, texture retention was strongest in the full pipeline despite lower no RNN NRMSEs. In IVDs, the no RNN pipeline delivered best standard reconstruction metric and texture retention performance. Despite maintaining smoother textures with moderate to excellent agreement across tested R and preserving sharper textures at lower R, the IVD pipeline revealed biases and fairly wide limits of agreement in T_2 preservation, particularly at R=6 and higher. When assessed on retrospectively undersampled multicoil raw k-space data, the knee and hip pipelines saw minimal degradation in performance as compared to results from images undersampled via synthetic k-space,

whereas the lumbar spine pipeline exhibited similar performance through R=4. Furthermore, repeatability studies indicated that, particularly for the hip and knee, performance was stable with respect to datasets. All told, these metrics indicate promise for the knee and hip pipelines in MAPSS T_2 mapping acceleration, and progress but room for improvement in IVDs.

Assessments of ROI-specific loss component utility showed its potential for improving predictions in accelerated acquisition schemes. When trained with sufficiently large datasets, as our knee and hip pipelines were, its inclusion saw stronger fidelity to local T_2 patterns in cartilage ROIs and reduced T_2 quantification errors compared to analogous pipelines trained without the ROI-specific loss component. Compared to state-of-the-art DL pipelines, knee and hip pipelines saw improved Pearson's r in cartilage ROIs but poorer global Pearson's r, as expected from the focused training approach. Interestingly, CS approaches exhibit relatively strong NRMSEs while generating relatively smooth predicted T_2 maps; this is possibly because in training, DL-based approaches simultaneously removed aliasing artifacts and performed T_2 fitting, and could attempt to preserve finer details than a CS approach performing those steps sequentially. While our approaches outperformed state-of-the-art methods at many R and tissue compartments in the lumbar spine, global Pearson's r indicated this may have been partially due to some models being more completely trained than others. These results may have been different with a larger lumbar spine training set. Nonetheless, the value of ROI-specific loss functions in accelerated acquisition pipelines is clear: with sufficiently large datasets, they can optimize for ROIs and outperform state-of-the-art approaches at high R, as existing approaches are optimized for global and not ROI-specific performance.

We can contextualize performance by comparing quantification errors to clinically significant T_2 changes. In the knee, T_2 increases 13.4% in lateral femoral condyle (LFC) cartilage, 12.3% in medical femoral condyle (MFC) cartilage, and 8.1% in medial tibial condyle (MTC) cartilage among patients with mild OA compared to controls [203]. Our top-performing knee pipeline saw errors below this benchmark through R=12 in the LFC and at R=2 in the MTC. In IVDs, T_2 decreases 36.3% in the nucleus pulposus and 24.2% in the annulus fibrosus from healthy to degenerative discs [204]. Our top-performing pipeline saw quantification errors for each disc below the more stringent 24.2% through R=12. In the hip, T_2 values among healthy patients that progress to OA within 18 months are 7.3% higher in femoral and 5.2% higher in acetabular cartilage compared to controls [205]. Our top-performing hip pipeline had errors below these benchmarks at all R in femoral cartilage and at R=2 in acetabular cartilage. Clinical metrics thus depict promise for pipelines in all three anatomies in maintaining sub-clinical-significance quantification errors.

Clinical and standard metrics show knee and hip pipeline performances to be particularly promising—the T_2 values, map texture preservation, and error rates relative to clinical benchmarks all mark meaningful progress towards reducing cMRI acquisition time for eventual clinical use. That said, while lumbar spine performance was strong by clinical metrics, it lagged the knee and hip by standard reconstruction metrics. One explanation is dataset size: the lumbar spine dataset had substantially fewer scans and imaging slices than the knee and hip. This has twofold impact: (1) the strength of a model trained from a smaller dataset is inherently

limited, and (2) having only 5 test set scans limits statistical power and induces wide standard deviations of metrics, preventing significant conclusions from being reached. The effects of this small dataset size particularly surface in repeatability studies. Furthermore, lumbar spine acquisitions were more susceptible to breathing artifacts and had fewer slices than the hip and knee; undersampling therefore left fewer lumbar spine k_y-k_z lines sampled compared to the hip and knee, inducing worse initializations and possibly poorer performance. Nonetheless, to our knowledge, this is the first DL application to accelerate lumbar spine cMRI, marking progress that must be furthered with additional data procurement and algorithm development for clinical utility.

The GLCM-based textural retention evaluation demonstrated a framework through which reconstruction performance can be better evaluated than through standard metrics like SSIM, NRMSE, and PSNR. ICCs of GLCM metrics between predicted and ground truth T_2 maps allow for intuitive, scaled measurements that can reflect how well a particular texture was preserved: for example, visual inspection of predicted T_2 maps in knee and hip cartilage in Figures 3-4 indicate that sharp textures are preserved better by the hip pipeline. This qualitative observation is confirmed by the GLCM Dissimilarity ICCs observed for the full model in the hip and knee pipelines in Table 3 at several tested R. This work could be furthered by extending this analysis to additional GLCM metrics for an even more thorough assessment of textural feature retention. Additional future improvements could also include pre-processing cartilage and IVD tissues prior to GLCM metric calculation to improve stability of these metrics, as other groups have started to do [206].

Moreover, by showing results at 7 acceleration factors instead of the 2-3 typical in the literature, we found performance did not always degrade steadily as R increased. Networks therefore may be sensitive not just to general undersampling patterns, but also the specific nature of the pattern. Thus, when future DL reconstruction pipelines are trained, a library of undersampling patterns may be advisable to encourage robustness to sampling patterns [140].

This study has limitations. First, we used retrospectively undersampled coil-combined magnitude echo time images that, in the knee and hip, had undergone ARC processing in their reconstruction, with 4 edge slices discarded for all data. Due to coil combination and post-processing, the k-space being undersampled would not match the acquisition's multi-coil k-space. Additionally, while we undersampled the MAPSS acquisition ellipse for each anatomy, the hip acquisitions had 'no phase wrap' applied, meaning that tested undersampling patterns would differ from those implemented on the scanner. While our raw k-space experiments show performance degradation was limited compared to coil-combined magnitude image experiments, models would be stronger if trained with a similarly sized multicoil k-space dataset. Second, this network is specific to our sampling patterns and acquisition parameters, and new pipelines would need to be trained should parameters like MAPSS T_2 echo times be substantially changed. Finally, the lumbar spine dataset size is rather small, limiting the power of conclusions.

To conclude, this study shows a novel means of training DL pipelines to accelerate cMRI in anatomies where specific tissues are of heightened clinical importance. In knee and hip, pipelines were effective at high R in maintaining textures, keeping fidelity to T_2 values, and minimizing T_2 quantification errors, whereas in the lumbar spine, the pipeline performed reasonably by those same criteria, but poorer in T_2 value fidelity and quantification errors. This reflects progress towards clinically useful pipelines that specialize in MSK T_2 mapping. The GLCM-based textural retention analysis elucidates an alternate to standard reconstruction metrics, allowing for intuitive measures of the types of features best preserved by a accelerated acquisition schemes, potentially allowing for better quantitative assessment of model performance. Future directions include multicoil k-space training, simultaneous MAPSS $T_{1\rho}$ and T_2 acceleration, and temporal undersampling of T_2 weighted echo time images.

Chapter 7 - K2S Challenge: From Undersampled K-Space to Automatic Segmentation

The following has been reformatted and reproduced with full permission from the publisher. It appeared in *Bioengineering* as:

Tolpadi, A.A., Bharadwaj, U., Gao, K.T., *et al*. K2S Challenge: From Undersampled K-Space to Automatic Segmentation. *Bioengineering* **10**:267 (2023). https://doi.org/10.3390/bioengineering10020267

7.1 Abstract

Magnetic Resonance Imaging (MRI) offers strong soft tissue contrast but suffers from long acquisition times and requires tedious annotation from radiologists. Traditionally, these challenges have been addressed separately with reconstruction and image analysis algorithms. To see if performance could be improved by treating both as end-to-end, we hosted the K2S challenge, in which challenge participants segmented knee bones and cartilage from 8× undersampled k-space. We curated the 300-patient K2S dataset of multicoil raw k-space and radiologist quality-checked segmentations. 87 teams registered for the challenge and there were 12 submissions, varying in methodologies from serial reconstruction and segmentation to end-to-end networks to another that eschewed a reconstruction algorithm altogether. Four teams produced strong submissions, with the winner having a weighted Dice Similarity Coefficient of 0.910 ± 0.021 across knee bones and cartilage. Interestingly, there was no correlation between reconstruction and segmentation metrics. Further analysis showed the top four submissions were suitable for downstream biomarker analysis, largely preserving cartilage thicknesses and key bone shape features with respect to ground truth. K2S thus showed the

value in considering reconstruction and image analysis as end-to-end tasks, as this leaves room for optimization while more realistically reflecting the long-term use case of tools being developed by the MR community.

7.2 Introduction

Magnetic Resonance Imaging (MRI) has emerged as one of the strongest medical imaging modalities for clinical use, offering exquisite soft tissue contrast for visualizing tissues such as ligaments, cartilage, and muscle [207,208]. Conventional MR sequences see the weighting of images in accordance with intrinsic MR parameters such as T1 and T2 and allow for suppression or saturation of signal from tissue types such as fat or fluid [57,61]. As such, MR images can be tailored for a given clinical context. Furthering this are recent developments of advanced sequences such as zero echo time (ZTE) and ultrashort echo time (UTE), which allow for high-resolution imaging of additional tissues such as tendons in musculoskeletal imaging [209–212]. MR has the added advantage of not exposing subjects to ionizing radiation compared to alternatives such as radiographs and computed tomography (CT). Despite these advantages, however, MR faces several challenges, including (1) long acquisition times and (2) the requirement of time-consuming and laborious radiologist annotation and interpretation of images to extract clinical meaning [213,214].

Fortunately, several tools have been developed to address these concerns. In the case of long MR scan times, acquisitions can be accelerated by sampling fewer points in k-space, the raw frequency-based domain in which MRI signals are obtained. This undersampling induces aliasing

artifacts in resulting images that can be removed by image reconstruction algorithms. In recent years, considerable effort has been put into developing several families of reconstruction approaches: (1) compressed sensing (CS) algorithms iteratively reconstruct images by ensuring consistency with acquired k-space and imposing sparsity on the reconstructed image in an alternate domain [128,170,171,215]; (2) parallel imaging (PI) algorithms exploit the redundancy of using multiple coils to acquire the same imaging volume to reduce acquisition times at the expense of signal-to-noise ratio (SNR) [63,166,167]; (3) deep learning (DL) approaches use complex, nonlinear models to impute full-length acquisition images from aliased images and/or undersampled k-space [134,216]. Other approaches growing in popularity include magnetic resonance fingerprinting (MRF) and low-rank and sparse modeling approaches [172,217,218]. On the other hand, a host of DL tools have emerged to automate mundane MR image-processing tasks. For instance, the introduction of the U-Net in 2015 seeded major advances in medical image segmentation from limited data, paving the way for more complicated architectures that have been applied for accurate lumbar spine and knee cartilage segmentation, among others [116,219–221]. Yet other DL applications include automatic assessments of cartilage thickness, staging of anterior cruciate ligament injury severity, diagnosis of lumbar spine anomalies, and analysis of bone shape [109,222–224].

This body of work unquestionably reflects substantial advances made by the MR research community. But it is noteworthy that with extremely few exceptions, the challenges of long acquisition times and image analysis have been treated as separate entities [225]. The long-term vision, however, would be a software package that addresses these challenges

simultaneously, raising a niche for optimization. Namely, image analysis algorithms are designed for full-length acquisition of MR image inputs, but there is no guarantee and little investigation that they would perform similarly well on reconstructed images from the accelerated acquisition. On the other hand, image reconstruction algorithms are overwhelmingly optimized for metrics such as normalized root mean square error (nRMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM), which correlate to the perceptual quality of a reconstructed image [226–228]. In other words, reconstruction algorithm image outputs are optimized for visually appealing images and radiologist interpretation, but what if their outputs were instead intended as features for subsequent image analysis pipeline input? Could the features required for accurate radiologist readings with respect to ground truth differ from those required for DL image analysis pipeline input to yield strong performance? More generally, if image reconstruction and annotation are viewed as end-to-end rather than serial tasks, is it possible to attain stronger image analysis performance?

To answer these questions, we hosted the K2S challenge at the 25th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) in Singapore. Previously, challenges and/or the releases of large datasets have spurred major advances in the MR research community. The release of the Osteoarthritis Initiative (OAI) and Multicenter Osteoarthritis Study (MOST) precipitated substantial advances in understanding osteoarthritis, total knee replacement, and knee pain, among others [110,229–232]. On the other hand, the fastMRI challenge was crucial in (1) making image reconstruction more accessible to the MR research community by releasing large datasets including raw k-space, and (2) seeding major

advances in reconstruction research such as popularizing unrolled DL architectures [131,233,234]. Our objective was to fill a similar niche in the end-to-end reconstruction and image analysis space. As such, we curated the K2S dataset, which consists of 300 patients that underwent 3D fat-suppressed knee scans, for each of whom k-space data and radiologist-approved 6-compartment tissue segmentations were released. The use of Fourier-transformed DICOM images as k-space would be problematic, not maintaining consistency with the multicoil nature of most MR acquisitions and the numerous post-processing steps to convert raw multicoil k-space into DICOM images, while also likely overstating performance [196]; importantly, our released dataset thus was of raw multicoil k-space data. Challenge participants were to train algorithms that segmented knee bones and cartilage from 8× undersampled acquisitions. Winners were selected using a weighted dice similarity coefficient (DSC) that assessed the accuracy of resulting segmentations, but additional analyses were conducted to assess segmentation quality, determine if strong image reconstruction was a prerequisite for strong segmentation, and gage the suitability of submitted segmentations for biomarker analysis [235].

In short, the contributions of the K2S challenge and this paper are as follows:
Reframing image reconstruction and annotation as end-to-end tasks for an eventual clinical workflow rather than sequential steps.

- Curating a large dataset (n = 300) with 3D raw k-space data and tissue segmentations to allow training of segmentation algorithms directly from undersampled k-space, and

whatever additional research objectives may emerge from having raw k-space and segmentations in the same dataset.

- Investigating whether strong image reconstructions are a prerequisite for strong tissue segmentations.
- Assessing if segmentation algorithms trained from 8× undersampled data are suitable for biomarker analysis.

7.3 Methods

7.3.1 Challenge

K2S challenge participants were responsible for predicting 6-class knee tissue segmentations (femur, tibia, patella, femoral cartilage, tibial cartilage, and patellar cartilage) from 8× undersampled k-space data. An overview of the steps involved in dataset curation, the challenge objective, and the timeline can be viewed in Figure 7.1, with details on all steps and evaluation criteria described below.

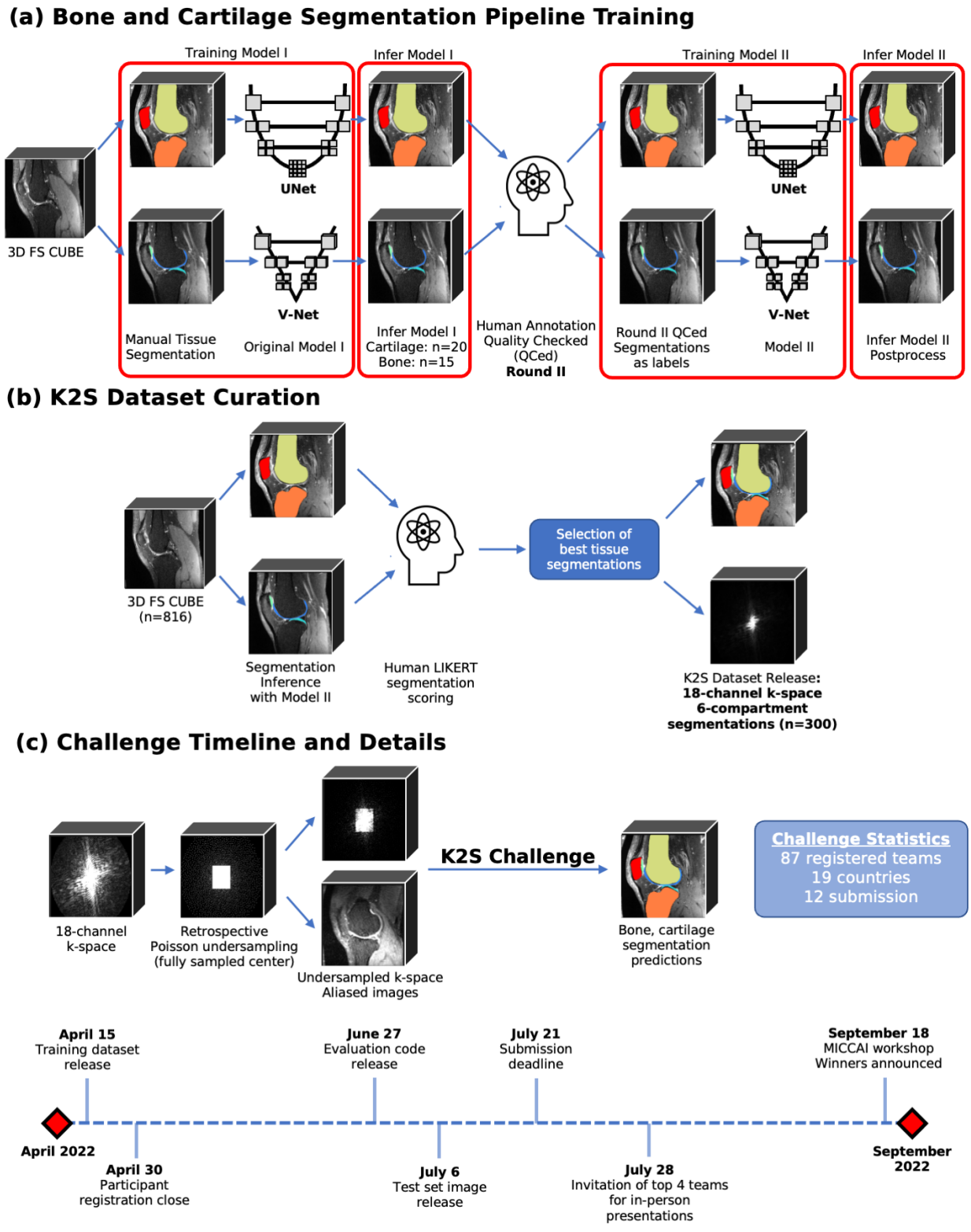


Figure 7.1 K2S challenge schematic. Overview of steps involved in human-in-the-loop training of models to generate ground truth bone and cartilage segmentations, and the process for

radiologist approval of final 300 segmentations to be included in K2S dataset. The K2S challenge was for participants to segment knee bones and cartilage from 8× undersampled k-space, with the training set released on 15 April, the test set released on 6 July, and the submission deadline on 21 July.

7.3.2 Dataset

7.3.2.1 Subject Eligibility and Sequence Information

Subjects at the UCSF Orthopedic Institute between 14 June 2021 and 21 June 2022 were scanned with an imaging protocol that included 3D fat-suppressed CUBE acquisitions (n = 816). There were no exclusion criteria placed on patients for inclusion in the eventual K2S dataset, and patients were scanned in accordance with all pertinent guidelines, including approval from the UCSF Institutional Review Board (Human Research Protection Program), obtaining informed consent from all study participants. The 3D fat-suppressed CUBE sequence was selected for K2S, as 3D sequences have higher SNR compared to 2D imaging, allowing for higher resolution acquisitions that can be reformatted into multiple planes for subsequent research objectives. Scans were performed on a GE Discovery MR750 3T Scanner using an 18-channel knee transmit/receive coil. The full-length acquisition time of the sequence was 4 min and 58 s. Complete acquisition parameters are listed in Table 7.1.

Table 7.1 Acquisition parameters for 3D fat-suppressed CUBE sequence used in K2S dataset, and for this challenge.

MR Acquisition Information		
Scanner: GE Discovery MR750 3T Scanner (GE Healthcare, Milwaukee, WI) **Gradient System Max Strength:** 50 mT/m **Max Slew Rate:** 200 mT/m/ms		
Coil: 18-channel knee transmit/receive coil (Quality Electrodynamics (QED), Mayfield Village, OH)		
TR/TE: 1002/29 ms	**FOV**: 150 mm	**Slice Thickness**: 0.6 mm (0.6 mm spacing between slices)
Flip Angle: 90	**SAR**: 0.0939	**Echo Train Length**: 36
Frequency: 128	**Bandwidth**: 244	**ARC** [150]: 4 (R = 2 in k_y, k_z)

MR Acquisition Information	
Acquisition Matrix: 256 × 256 × 200	**Image Dimensions**: 512 × 512 × 200 **Resolution**: 0.586 mm × 0.586 mm × 0.6 mm **Voxel Size**: 0.293 mm × 0.293 mm × 0.6 mm

7.3.2.2 Extraction of ARC-Reconstructed Multicoil Raw k-Space Data

An in-house pipeline was developed replicating all post-processing steps done on an MR scanner to go from raw k-space data to DICOM images viewed by clinicians for diagnostic decisions. To the best of the authors' knowledge, no centralized resource is available describing all these steps, which can make it difficult for those interested in reconstruction to familiarize themselves with the process before model development. The authors thus saw value in describing these steps, shown schematically in Figure 7.2, with examples of pipeline intermediates at several steps in Figure 7.3. Unless otherwise specified, all post-processing steps were implemented using functions in GE Orchestra 1.10.

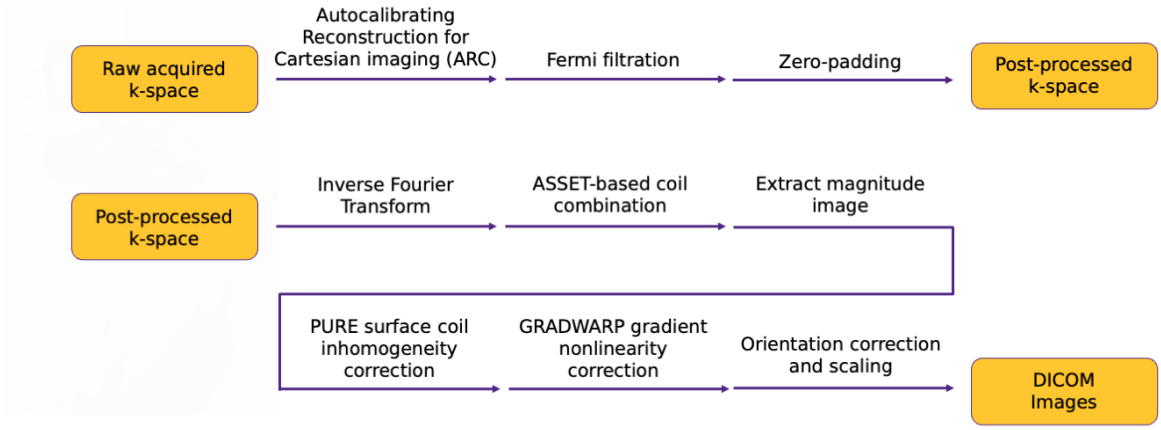


Figure 7.2 k-Space and image space post-processing steps for the in-house pipeline to reconstruct DICOM images from raw scanner data. Briefly, the steps in k-space are as follows: ARC reconstruction (parallel imaging), Fermi filtration to remove Gibbs artifacts, and zero-padding to bring the image to the intended output resolution. Image-space processing included coil combination, surface coil intensity correction, and gradient coil inhomogeneity correction.

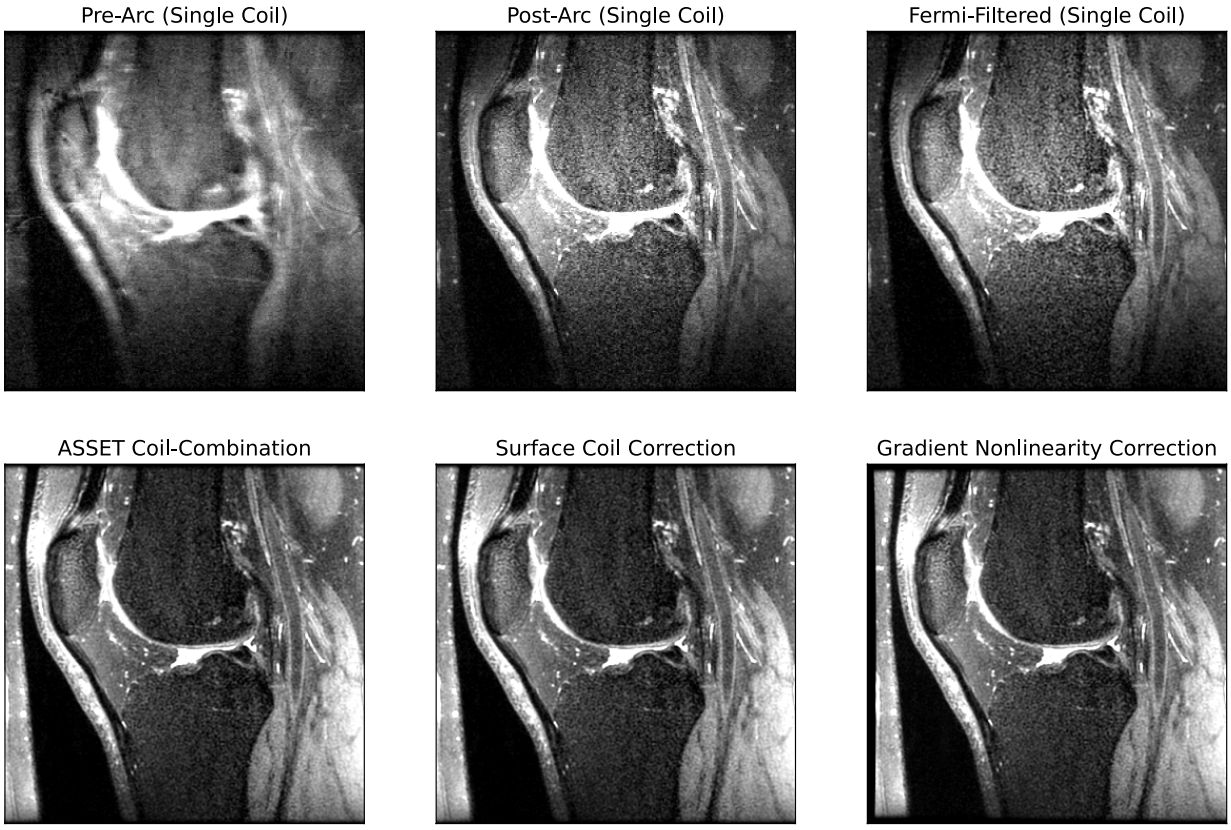


Figure 7.3 Intermediate outputs within the post-processing pipeline going from raw k-space to DICOM images. Each pane of the image reflects the output of the image after the step described by the pane title.

7.3.2.2.1 k-Space Post-Processing

Some sequences may leverage PI techniques (such as ARC or GeneRalized Autocalibrating Partial Parallel Acquisition (GRAPPA)) to acquire fewer lines within k-space, instead exploiting already acquired data across multiple coils to mitigate aliasing artifacts at the expense of SNR [63,150]. This was the case for our sequence; consequently, the first step in post-processing raw multicoil k-space data was applying ARC to impute unacquired k-space lines. Subsequently, Fermi filtration was applied: given MR images are often zero-padded in k-space, ringing artifacts can emerge from the sharp boundary in k-space between nonzero and zero points [236]. A

Fermi filter smooths this boundary, reducing ringing artifacts at the expense of sharpness in the reconstructed image. A custom Fermi filtration function was used, using the Fermi filtration radius and width parameters extracted from raw sequence metadata. After Fermi filtration, k-space was zero-padded to the intended image dimensions (in our case, from 256 × 256 × 200 to 512 × 512 × 200), completing k-space post-processing. All k-space post-processing was on multicoil data.

7.3.2.2.2 Image Space Post-Processing

Post-processed k-space was 3D inverse Fourier transformed to image space for each of the 18 coils and coil-combined to yield a single-coil image. The most basic means of coil combination is root sum-of-squares, but GE provides another method based on Array coil Spatial Sensitivity Encoding (ASSET), which leverages sensitivity maps in a PI-inspired technique to do coil combination [237]. Magnitude images were then calculated, after which GE's Phased array Uniformity Enhancement (PURE) was used to perform surface coil intensity correction [238,239]. This was followed by GRADWARP, which warps images to correct for inhomogeneities in gradient coils [240]. A final step in post-processing was correcting image orientation and scaling pixel values, yielding DICOM images used by clinicians for diagnostic purposes.

In the context of segmenting undersampled images, one complication emerges: in GRADWARP, the MR image is warped such that it no longer corresponds to k-space. As such, the post-processing pipeline intermediate prior to GRADWARP must be segmented, or the GRADWARP function must be integrated into model training itself while segmenting DICOM images. Due to

the difficulties of implementing the latter (backpropagating through GRADWARP would not be trivial), our solution was the former.

7.3.2.3 Ground Truth Segmentation Generation

Ground truth knee cartilage and bone segmentations were generated by separate DL pipelines and post-processing techniques, each trained with a radiologist in the loop.

7.3.2.3.1 Cartilage Segmentation Pipeline

480 3D fat-suppressed CUBE sequences were acquired across three sites (UCSF, San Francisco, CA, USA; Hospital for Special Surgery, New York, NY, USA; Mayo Clinic, Rochester, MN, USA) with similar acquisition parameters to the 3D fat-suppressed CUBE sequences ultimately used in K2S. These volumes were manually segmented by readers trained by a senior radiologist with over 25 years of experience, split 400/80 into training and validation, and used to train a 3D V-Net for multiclass cartilage segmentation [104,241]. This initial pipeline was inferred on 20 3D fat-suppressed CUBE sequences from the UCSF Orthopedic Institute with K2S acquisition parameters, but on volumes acquired prior to the eligibility window for K2S inclusion. The 20 inferred segmentations were manually corrected and quality checked (QC) by an intern under radiologist supervision. 15 of the 20 cases were used to fine-tune the pipeline in a second training, seeing convergence reached after 5 epochs, and the remaining 5 cases were used to select final model parameters.

After the second training, the V-Net was inferred on all 816 cases eligible for K2S. The following post-processing steps were selected and applied under radiologist supervision: 3D morphological opening, 3D connected components analysis (preserving the largest femoral and patellar and the 2 largest tibial cartilage components), and 2D sagittal connected components analysis (preserving all connecting components larger than 150 pixels).

7.3.2.3.2 Bone Segmentation Pipeline

40 3D fat-suppressed CUBE sequences acquired at the UCSF Orthopedic Institute prior to the eligibility window for K2S inclusion were manually segmented by a trained reader for bone, tibia, and patella. These cases were used to train a baseline 3D U-Net for a binary bone segmentation model. An additional 15 cases acquired using the K2S acquisition parameters were also manually segmented by three radiologists with three (J.L.), three (P.G.), and four (F.G.) years of experience. The trained baseline model was inferred from these cases, which were used for model fine-tuning.

The fine-tuned U-Net was inferred on the 816 cases with the following post-processing steps, applied under radiologist supervision: filling holes, morphological opening, and connected components analysis (preserving all connecting components larger than 1000 voxels and with centroids in central 50% of slices). Finally, the sizes of connected components were used to extract bone labels (femur, tibia, and patella).

7.3.2.4 Selection of Cases for K2S Dataset

Of the 816 potential cases, the target was selecting 300, with the intent of maintaining sufficient cases for training reconstruction and segmentation models, maintaining some variety of anomalies in included cases, and ensuring a reasonable memory footprint given computational constraints. Radiologists with three (J.L.) and four (F.G.) years of experience developed 5-point LIKERT scales to assess segmentation quality: (1) unusable; (2) poor, with some mislabeling of bones or cartilage; (3) useable, with some major issues, but correct labeling of bone or cartilage; (4) good, with some minor but acceptable issues; (5) (near) human-like. Examples of cartilage segmentation LIKERT scores for the 5 classes are seen in Figure 7.4, and for bone in Figure 7.5. Segmentation LIKERT scores were calculated for bone and cartilage from videos of the segmentations that cycled through all sagittal slices. Cases with acceptable segmentation quality for both cartilage and bone were selected as the K2S dataset. Cartilage LIKERT scores for K2S were as follows: 5:14; 4:175; 3:110; 1:1. Bone LIKERT scores were as follows: 5:112; 4:179; 3:9.

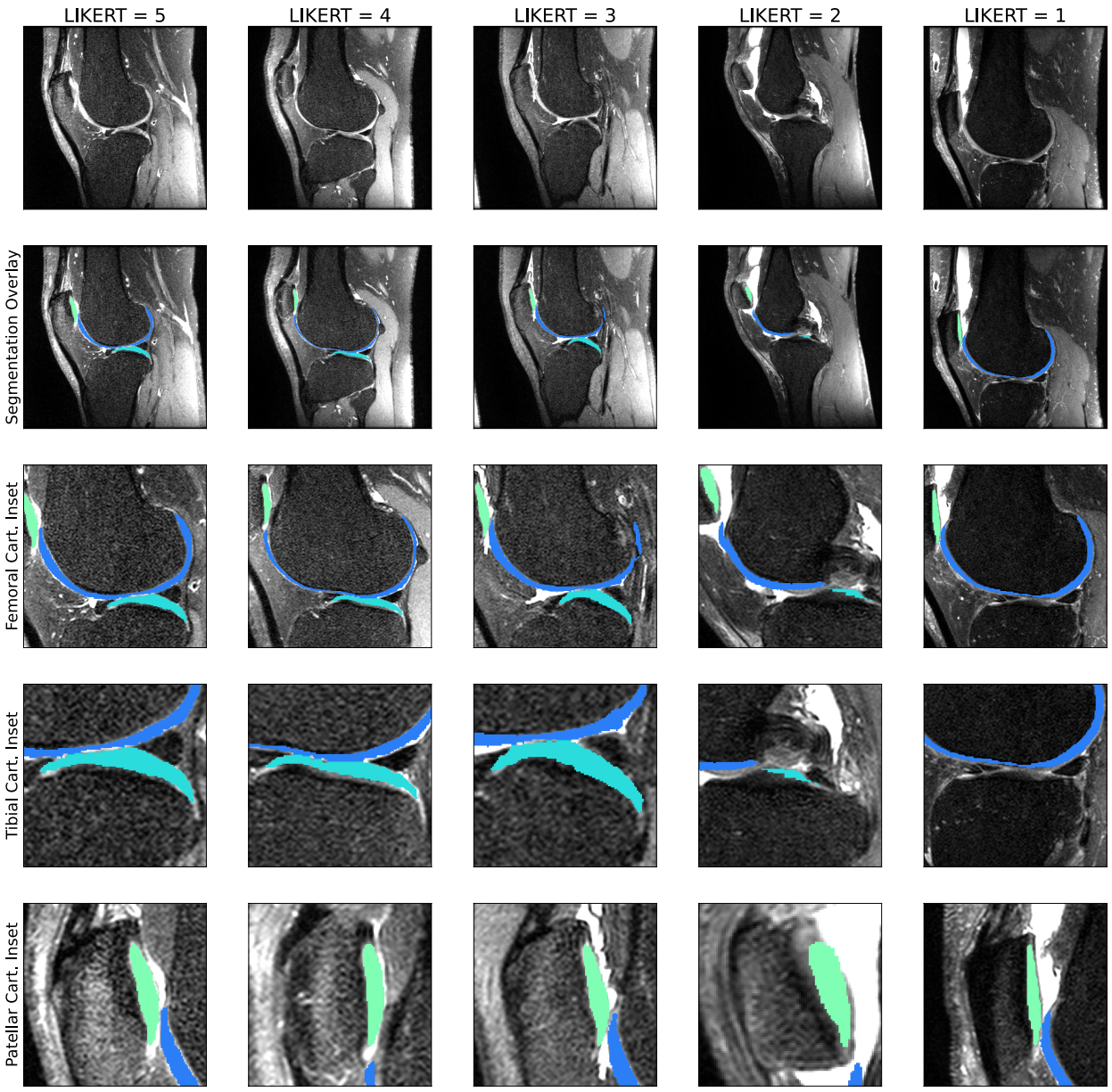


Figure 7.4 1–5 LIKERT cartilage segmentation scores overlaid on ground truth knee scans. In this example, the LIKERT of 5 indicates human-like segmentation; the LIKERT of 4 shows a slight underestimation of patellar and tibial cartilage; the LIKERT of 3 is assigned due to minor underestimation of patellar and tibial cartilage, with soft tissue detected as femoral cartilage; the LIKERT of 2 is assigned due to missing mask areas for patellar and tibial cartilage, with femoral cartilage overestimation; the LIKERT of 1 is missing a tibial cartilage mask.

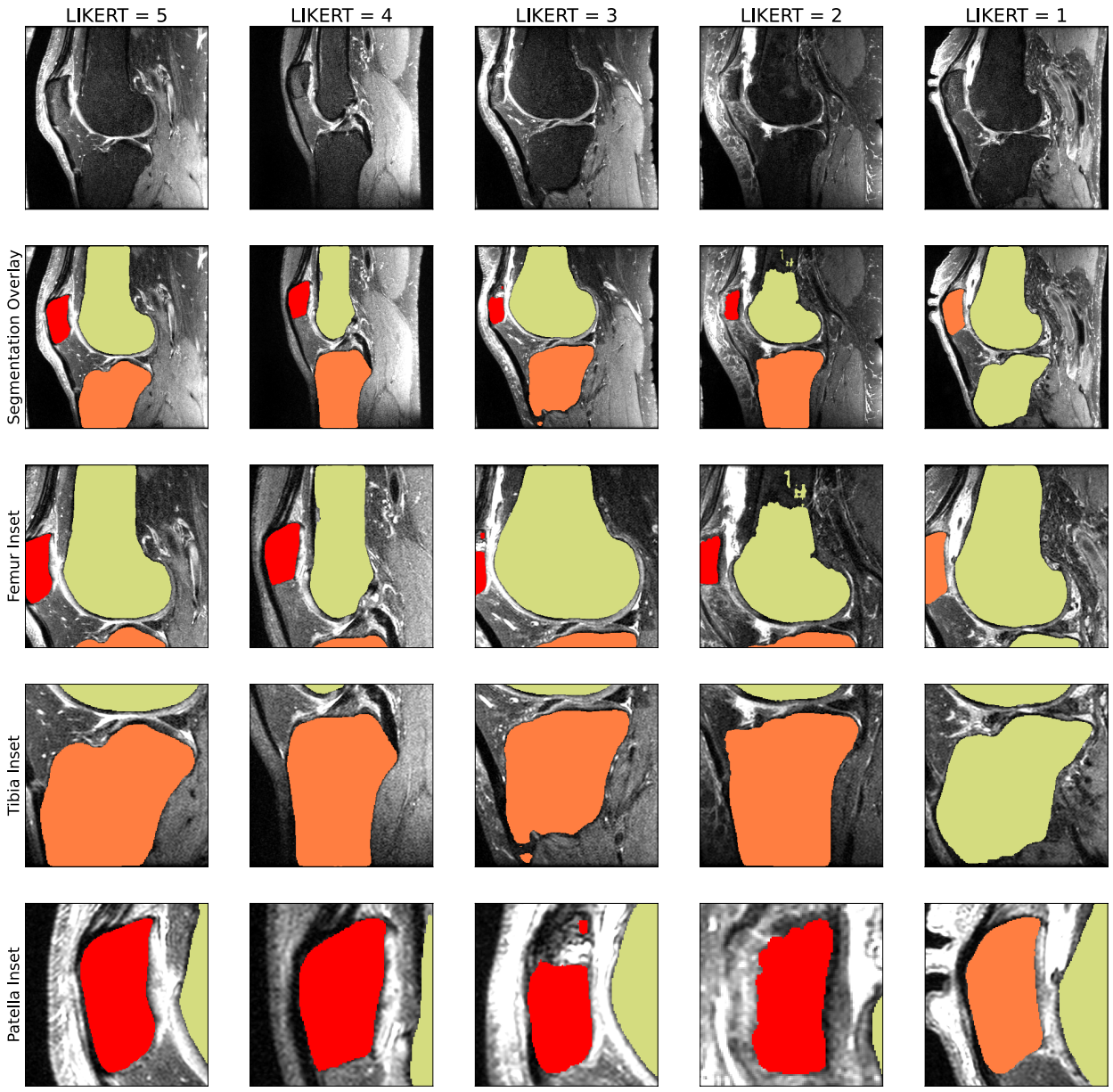


Figure 7.5 1–5 LIKERT bone segmentation scores overlaid on ground truth knee scans. In this example, the LIKERT of 5 indicates human-like segmentation; the LIKERT of 4 shows minor missing components in the femoral bone; the LIKERT of 3 shows missing components of the patellar bone mask; the LIKERT of 2 shows major missed regions within the tibial and patellar bone; the LIKERT of 1 has patella and tibia masks misassigned.

7.3.2.5 Final K2S Dataset Characteristics

The K2S training dataset (n = 300) had the following demographic characteristics: age of 44.3 ± 13.9 years, weight of 75.6 ± 14.9 kg, 160/140 male to female. The test dataset followed the

same described steps (n = 50): age of 44.5 ± 14.4 years, weight of 70.5 ± 16.6 kg, 26/24 male to female (all mean ± standard deviation).

The training dataset included the following: multicoil ARC-reconstructed k-space and multiclass segmentation for each patient (n = 300), 8× center-weighted Poisson undersampling mask with a fully sampled central 5% square of k-space in k_y-k_z, and a file detailing the quality of the segmentations and any radiologist notes associated with each patient. The released test dataset was solely the 8× undersampled multicoil ARC-reconstructed k-space.

7.3.3 Evaluation Process

Submissions were evaluated using a weighted sum of DSC. Namely, DSC was calculated in each of the 6 tissue compartments, and combined as follows into a weighted DSC that assigned each compartment a weight inversely proportional to the size of the tissue compartment, as shown in Equation 7.1:

$$Weighted\ DSC = \frac{\sum_t \frac{DSC_t}{n_t}}{\sum_t \frac{1}{n_t}}$$

Equation 7.1 Weighted DSC calculation, where *t* refers to the tissue compartment, DSC_t refers to the DSC within that tissue compartment, and n_t is the number of pixels in the ground truth segmentation for tissue *t* [235].

7.3.4 Timeline

- 15 April 2022: Training dataset release
- 30 April 2022: Participant registration close

- 27 June 2022: Release of code used to evaluate submissions
- 6 July 2022: Test dataset release
- 21 July 2022: Submission deadline
- 28 July 2022: Invitation of top 4 teams for in-person presentations
- 18 September 2022: In-person workshop at MICCAI 2022, winners announced

All told, 87 teams registered for the K2S challenge from 19 countries, and 12 teams made submissions for the challenge.

7.3.5 Overview of Top Submission Methodologies

7.3.5.1 K-nirsh (University of Tübingen, Tübingen, Germany)

K-nirsh's submission involved two cascaded nnUNet architectures, a first for reconstruction and a second for segmentation [242]. Multicoil k-space was inverse Fourier transformed and coil-combined using root sum-of-squares coil combination, yielding coil-combined 8× undersampled images. An initial nnUNet was pretrained to predict fully-sampled coil-combined images from 8× undersampled coil-combined inputs using a mean square error (MSE) loss. A second nnUNet was pretrained to predict multiclass cartilage and bone segmentations from a 2-channel input (8× undersampled coil-combined image and fully sampled coil-combined image), using DSC segmentation loss. After pretraining, these models were trained end-to-end, with the initial nnUNet regression output replacing the fully sampled coil-combined image as input for the second segmentation nnUNet. The model was fine-tuned for over 1000 epochs on NVIDIA V100 GPUs, using only the segmentation loss and implementing a weight scheduler that linearly increased small class weighting (cartilage). The weighted DSC loss used to evaluate the

challenge submission was used as a validation loss and the best model according to this metric was chosen for the challenge submission. The output of the first nnUNet was considered the reconstruction output of this pipeline, whereas the output of the second nnUNet was the segmentation submission.

7.3.5.2 UglyBarnacle (Skolovo Institute of Science and Technology, Moscow, Russia)

UglyBarnacle's submission differed from other top submission methodologies by leveraging CS as opposed to DL for reconstruction. An initial reconstruction pipeline accepted as input the 18-channel, 256 × 256 × 200 8× undersampled k-space array, performing a CS reconstruction with a combined L_1-wavelet and total variation (TV) regularization function, imposing 3 times the weight on TV as opposed to L_1-wavelet. The CS reconstruction was solved as an optimization problem: the goal was to find the undersampled part of the k-space that minimized the target value function (weighted sum of L_1-wavelet and TV of volumetric image). The optimization problem was solved using the Adam optimization algorithm over 50 iterations for each scan. Reconstructed images were fed to an architecture similar to V-Net for tissue segmentation. The segmentation network was implemented in 3D, with the following feature map depths at V-Net stages: 16, 32, 64, 128, 256. Max-pooling was used to compress the representation of feature maps in the encoder, and upsampling to increase resolution in the decoder, with skip connections transferring information from the encoder to corresponding parts of the decoder. The network output was fed through two final convolutions (one with a feature map depth of 7 and the last with a depth of 1) to yield predicted segmentations.

7.3.5.3 FastMRI-AI (University Medical Center Groningen, Groningen, The Netherlands)

As with K-nirsh, k-space was zero-padded to 512 × 512 along k_x and k_y, inverse Fourier transformed, and root sum-of-squares combined, yielding coil-combined 8× undersampled image space. Unlike other top submissions, FastMRI-AI did not implement a reconstruction framework, choosing instead to directly segment the undersampled image; the root sum-of-squares coil combined images were thus considered the reconstruction outputs for this approach in subsequent analysis. A 3D U-Net featuring a squeeze and excite attention layer was trained on 160 × 160 × 48 patches, selected with stride 51 × 51 × 16, yielding around 27 predictions per voxel [243]. Networks were trained with weighted DSC loss, giving twice the weight to cartilage afforded to bone and background. Predictions were post-processed with simulated extended image boundaries by mirror padding, self-ensembling for overlapping sliding window prediction, and connected component analysis for each class, removing objects that were less than 60% the size of the largest object in the given class.

7.3.5.4 NYU-Knee AI (New York University Grossman School of Medicine, New York, USA)

NYU-Knee AI trained multiple components individually: a Variational Network (VN) for image reconstruction, followed by an ensemble of 2D U-Nets to predict tissue segmentations [119,135,244–246]. For reconstruction, eSPIRIT was used to calculate coil sensitivity maps for undersampled and ground truth data using the central 24 × 24 region in k-space [67]. Zero-filled k-space was then fed through a VN for K = 10 iterations, at each iteration using calculated coil sensitivity maps and acquired k-space to ensure data consistency with intermediate reconstructed images, while also feeding iteration outputs through a convolutional, ReLU, and

transpose convolutional layer to encourage recovery of details lost from undersampling. The VN was trained with an MSE loss function between the 256 × 256 ground truth and the reconstructed coil-combined images for 200 epochs. VN outputs were fed to 2D U-Nets, predicting 256 × 256 segmentations that were upsampled and convolved to the intended 512 × 512 output resolution. Multiple networks were trained with either focal loss, cross-entropy loss, or a hybrid of both for 300 epochs; an internal validation set was used to choose the best-performing network for each of the 6 tissue classes, ultimately using 3 focal loss networks and 1 weighted cross entropy loss network in the final submission.

7.3.6 Further Analysis of Submissions

7.3.6.1 Intermediate Pipeline Reconstruction Performance

The objective of the challenge was segmenting bones and cartilage, and no part of the evaluation criteria nor any communication between organizers and challenge participants prior to submissions discussed a requirement for reconstruction submissions. However, at some level, each of the top-performing pipelines fed some image (either directly undersampled for FastMRI-AI, or after reconstruction for the other 3 top submissions) through a segmentation pipeline. As such, it was instructive to see how reconstruction metrics of images fed to segmentation pipelines compared to segmentation metrics. Challenge organizers thus requested the top four teams provide intermediate reconstruction outputs for the test set. Using these images, standard reconstruction metrics were calculated: nRMSE, PSNR, and SSIM.

7.3.6.2 Comparison of Reconstruction and Segmentation Performance

In addition to the visual comparison of reconstructions and segmentations, Pearson's r was calculated between weighted DSC and nRMSE, PSNR, and SSIM for each of the top 4 teams [199]. Given the wide variety of approaches used by the teams, these experiments investigated a correlation between reconstruction and segmentation performance.

7.3.6.3 Biomarker Analysis: Cartilage Thickness

Previously developed tools were used to calculate cartilage thicknesses for ground truth and submissions [222]. Briefly, Euclidean distance transforms on each cartilage compartment of each patient were used to generate skeletonizations. The skeletonizations were sampled and distances from skeletonized points to cartilage surfaces were calculated for each compartment and each patient. Skeleton-to-surface distances were averaged across a cartilage compartment for a given patient to obtain mean cartilage thickness measurements, which were then compared between ground truth and each of the submissions in Bland-Altman and correlation plots. Pearson correlation coefficients were calculated for each submission to assess the correlation of submitted cartilage thicknesses to ground truth, as a proxy for assessing the suitability of submissions for biomarker analysis.

To visualize cartilage thickness maps, voxel-based segmentations were converted into triangulated meshes using a Marching Cubes algorithm, and cartilage thickness maps were projected onto bones for select cases [232]. Maps were then compared for a qualitative assessment of regions best and most poorly preserved by sample submissions.

7.3.6.4 Biomarker Analysis: Bone Shape

To analyze the bone shape, previously developed tools again were applied [224]. Triangulated meshes of each bone of the ground truth segmentations were generated using a Marching Cubes algorithm, after which Euclidean coordinates of each point in the mesh were flattened into a 1D vector for each test set case. Principal component analysis (PCA) was used to reduce the dimensionality of these vectors, preserving the top 5 PCs, which constituted bone shape features. Statistical parameterization was used to extract the mean and standard deviation of each PC. For visualization purposes, mean +3 standard deviations (s.d.) and mean −3 s.d. bone shapes were generated for each PC, with qualitative interpretations of the features varying most with each PC being described (i.e., volume). Segmentations of each submission were similarly transformed into 1D Euclidean coordinate vectors and projected into the PC space generated from the ground truth. Correlations between submissions and ground truth along these shape features were calculated for each.

7.4 Results

12 submissions were received for the K2S, for which weighted DSC was calculated across the test set as described in Equation 7.1. The top four submissions by weighted DSC were analyzed further, with results discussed below.

7.4.1 Segmentation Metrics

Segmentation results are shown in Table 7.2, stratified by tissue compartment but also showing the weighted DSC that determined challenge winners. K-nirsh delivered strong segmentation performance in each tissue compartment, closely rivaling ground truth, and interestingly did so from intermediate reconstruction outputs exhibiting poor reconstruction metrics. FastMRI-AI also yielded high-quality segmentations despite not implementing any reconstruction framework. Overarchingly, segmentation performance for all four pipelines was strong, given that severely aliased images served as model input. To differentiate between the top two submissions, which showed similar weighted DSC, a paired t-test was run to assess for significant difference in performance: K-nirsh performance indeed was significantly better than UglyBarnacle, even after adjusting for Bonferroni correction ($n = 50$, $\alpha = 0.05$).

Table 7.2 Segmentation performance across test set (n = 50) for each of the top 4 pipelines, stratified by tissue compartment. Results are presented mean ± 1 s.d. K-nirsh showed the strongest results in each tissue compartment and overall, and is shown in bold.

	Cartilage			Bone			Full
Team	**Femoral**	**Tibial**	**Patellar**	**Femur**	**Tibia**	**Patella**	**Weighted DSC**
K-nirsh	**0.904 ± 0.014**	**0.899 ± 0.015**	**0.910 ± 0.034**	**0.989 ± 0.002**	**0.985 ± 0.004**	**0.966 ± 0.012**	**0.910 ± 0.021**
UglyBarnacle	0.895 ± 0.016	0.890 ± 0.017	0.903 ± 0.032	0.984 ± 0.004	0.980 ± 0.004	0.961 ± 0.015	0.903 ± 0.021
FastMRI-AI	0.845 ± 0.124	0.862 ± 0.126	0.843 ± 0.124	0.964 ± 0.078	0.952 ± 0.138	0.834 ± 0.306	0.849 ± 0.123
NYU-Knee AI	0.798 ± 0.029	0.756 ± 0.04	0.796 ± 0.043	0.980 ± 0.004	0.975 ± 0.005	0.939 ± 0.014	0.795 ± 0.030

7.4.2 Reconstruction Metrics

Example sagittal slices of intermediate pipeline reconstruction outputs are shown in Figure 7.6, with corresponding reconstruction metrics. NYU-Knee AI and particularly UglyBarnacle

produced intermediate reconstruction outputs with strong fidelity to ground truth, recovering fine details lost to aliasing. On the other hand, K-nirsh yielded an image with more distinct tissue boundaries, but with noise and pixel intensity distributions that clearly differed from the ground truth. Complete metrics of reconstruction performance are shown in Table 7.3.

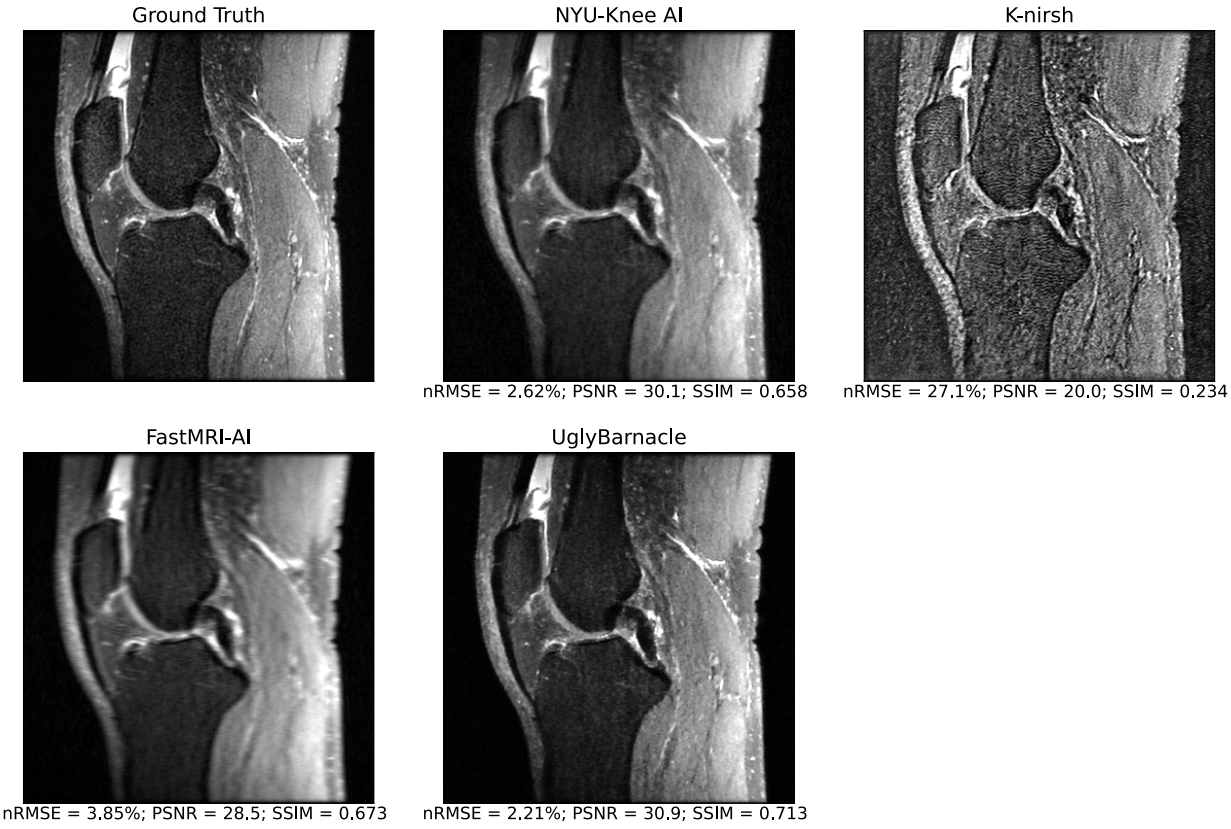


Figure 7.6 Intermediate pipeline reconstruction outputs for each of the top 4 submissions in an example sagittal slice, as well as ground truth, with reconstruction metrics displayed for the volume including the visualized slice. For this volume, UglyBarnacle delivers the highest quality reconstruction, followed closely by NYU-Knee AI, recovering sharpness and many fine details lost to aliasing during 8× Poisson undersampling. K-nirsh delivers an intermediate reconstruction that was poor by standard reconstruction metrics, but perceptually, made boundaries between tissues much more distinct and perhaps easier to segment. This is likely due to K-nirsh fine-tuning the reconstruction and segmentation networks in an end-to-end manner, unlike other top submissions.

Table 7.3 Standard reconstruction metrics for intermediate pipeline outputs from all top submissions across the released test set (n = 50). Results are presented mean ± 1 s.d. The top pipeline by each of these metrics was UglyBarnacle, shown in bold.

Team	nRMSE	PSNR	SSIM
K-nirsh	31.2 ± 4.26	19.7 ± 0.68	0.217 ± 0.059
UglyBarnacle	**2.07 ± 0.25**	**31.5 ± 0.87**	**0.693 ± 0.043**
FastMRI-AI	3.05 ± 0.68	29.8 ± 0.99	0.681 ± 0.061
NYU-Knee AI	2.18 ± 0.33	31.3 ± 0.87	0.672 ± 0.029

7.4.3 Comparison of Reconstruction and Segmentation Performance

Example slices of predicted segmentations, overlaid on intermediate reconstruction outputs, are shown for all four teams alongside ground truth in Figure 7.7. For each reconstruction metric, and for each of the top 4 performing pipelines, weighted DSC was plotted against the reconstruction metric in Figure 7.8, with Pearson's correlation coefficients being calculated for each pair. The highest correlation coefficient in this study was between nRMSE and weighted DSC for the NYU-Knee AI submission, at 0.284, with all other correlation coefficients being substantially lower. This indicates that, at the absolute best, there was a weak correlation between segmentation and reconstruction metrics, and in most cases, there was a negligible or even slightly negative correlation between the two.

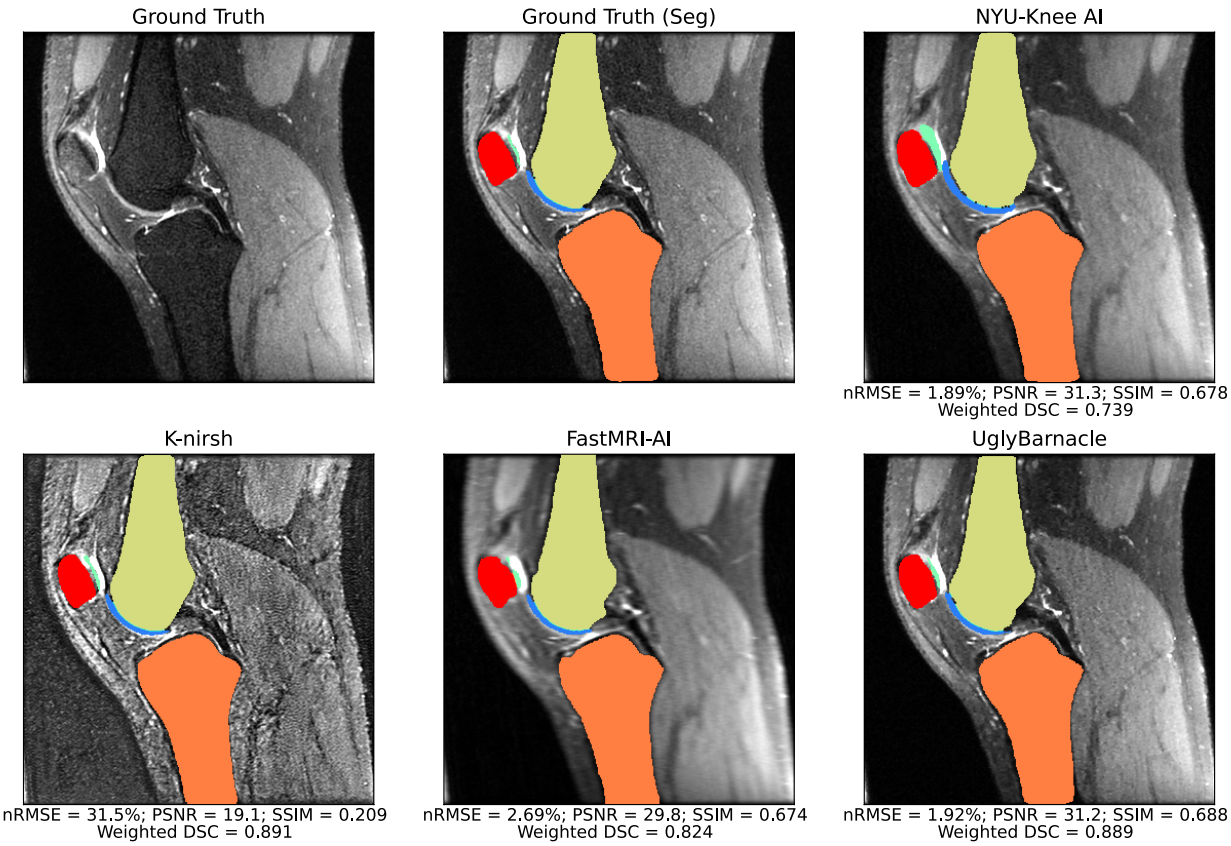


Figure 7.7 Sagittal slice segmentations overlaid on intermediate pipeline reconstructions, with reconstruction and segmentation metrics for the volume including the slice displayed. Background anatomy slices were thus blurrier for some teams than for others, as different teams had different quality intermediate pipeline reconstruction outputs. In this example, segmentation quality was strong for all top submissions, with only some overestimation of cartilage thickness from the NYU-Knee AI pipeline being apparent. K-nirsh maintains a slight edge over UglyBarnacle in reconstruction metrics for this volume.

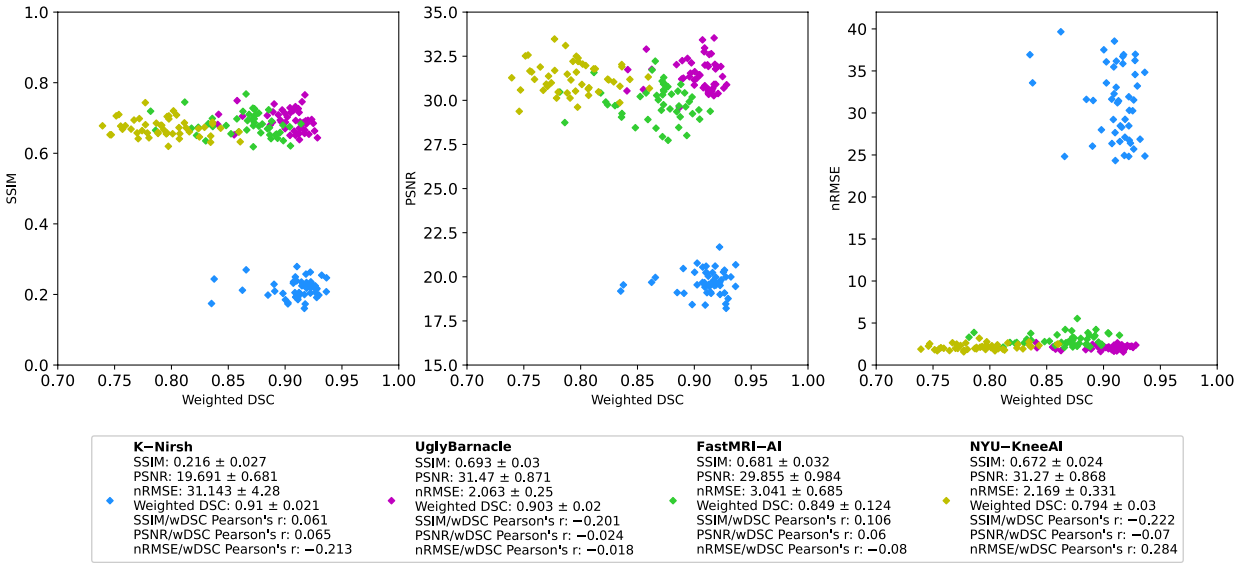


Figure 7.8 Reconstruction metrics (nRMSE, PSNR, SSIM) plotted against weighted DSC for each of the top four submissions, with each point denoting a subject in the test set (n = 50). Pearson's correlation coefficient was calculated for each pair and is displayed on the chart, indicating that at absolute best, there was a weak correlation between segmentation and reconstruction metrics, and that in most cases, there was no or even negative correlation.

7.4.4 Biomarker Analysis: Cartilage Thickness

Example femoral cartilage thickness maps projected onto the femur are shown in Figure 7.9, with corresponding femoral cartilage segmentation DSCs. These results elucidate added complexity: while FastMRI-AI and NYU-Knee AI lagged K-nirsh and UglyBarnacle in weighted DSC, they did a better job preserving certain thick and thin cartilage regions. Qualitatively, however, these maps show K-nirsh, UglyBarnacle, and FastMRI-AI perform especially well in reconstructing cartilage thicknesses; Bland-Altman plots in Figure 7.10 confirm these results, showing cartilage thicknesses across all three compartments were predicted with minimal bias and strong fidelity to ground truth by these three teams. Interestingly, bias in retaining femoral cartilage thicknesses decreased with larger ground truth cartilage thicknesses, regardless of submission. More granularly, while fastMRI-AI slightly overestimated patellar cartilage

thicknesses, they also reflected the least bias in maintaining femoral cartilage thickness, showing some discordance between weighted DSC and downstream biomarker analysis. Contrarily, thicknesses were overestimated by NYU-Knee AI, particularly in tibiofemoral regions. In comparing the top two challenge finishers, K-nirsh and UglyBarnacle, biases in predicted cartilage thicknesses were slightly lower for UglyBarnacle in femoral and tibial cartilage, and slightly higher in patellar cartilage (UglyBarnacle: femoral: 0.088 ± 0.07, tibial: 0.036 ± 0.09, patellar: 0.114 ± 0.13; K-nirsh: femoral: 0.096 ± 0.08, tibial: 0.049 ± 0.09, patellar: 0.097 ± 0.11; all in units of mm, mean ± 1 s.d.). However, paired t-tests showed none of these differences were significant even after Bonferroni correction (n = 50, α = 0.05).

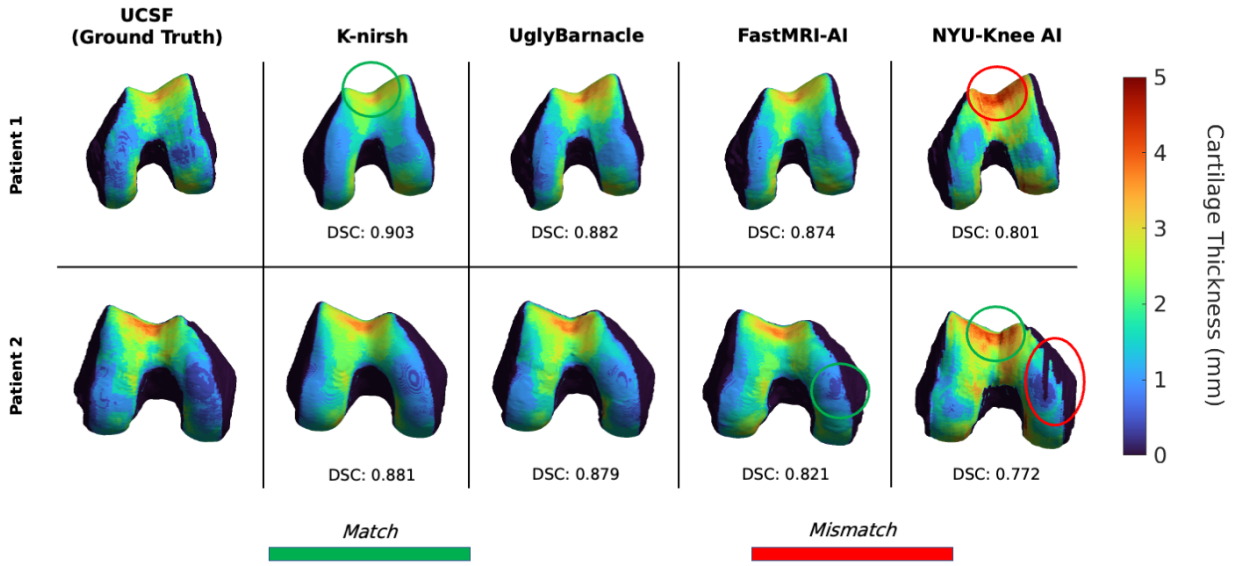


Figure 7.9 Femoral cartilage thickness maps projected onto voxel-based femoral bone shapes for each of the top 4 teams, as well as ground truth. While all submissions exhibit a degree of smoothness that is not reflected in the ground truth, the top three especially were strong in preserving cartilage thicknesses (K-nirsh, UglyBarnacle, FastMRI-AI), with NYU-Knee AI slightly overestimating cartilage thicknesses but still preserving key features in some regions.

Correlation plots in Figure 7.10 showed K-nirsh, UglyBarnacle, and NYU-Knee AI yielded high Pearson correlation coefficients with respect to ground truth, indicating high-quality

segmentations. Interestingly, UglyBarnacle showed a slightly higher correlation to ground truth cartilage thickness in tibiofemoral cartilage than K-nirsh, despite lower DSCs in both tissues. Visually, FastMRI-AI also appeared to show a strong correlation between predicted and ground truth cartilage thickness, although poor prediction in one case appeared to severely degrade the correlation coefficient.

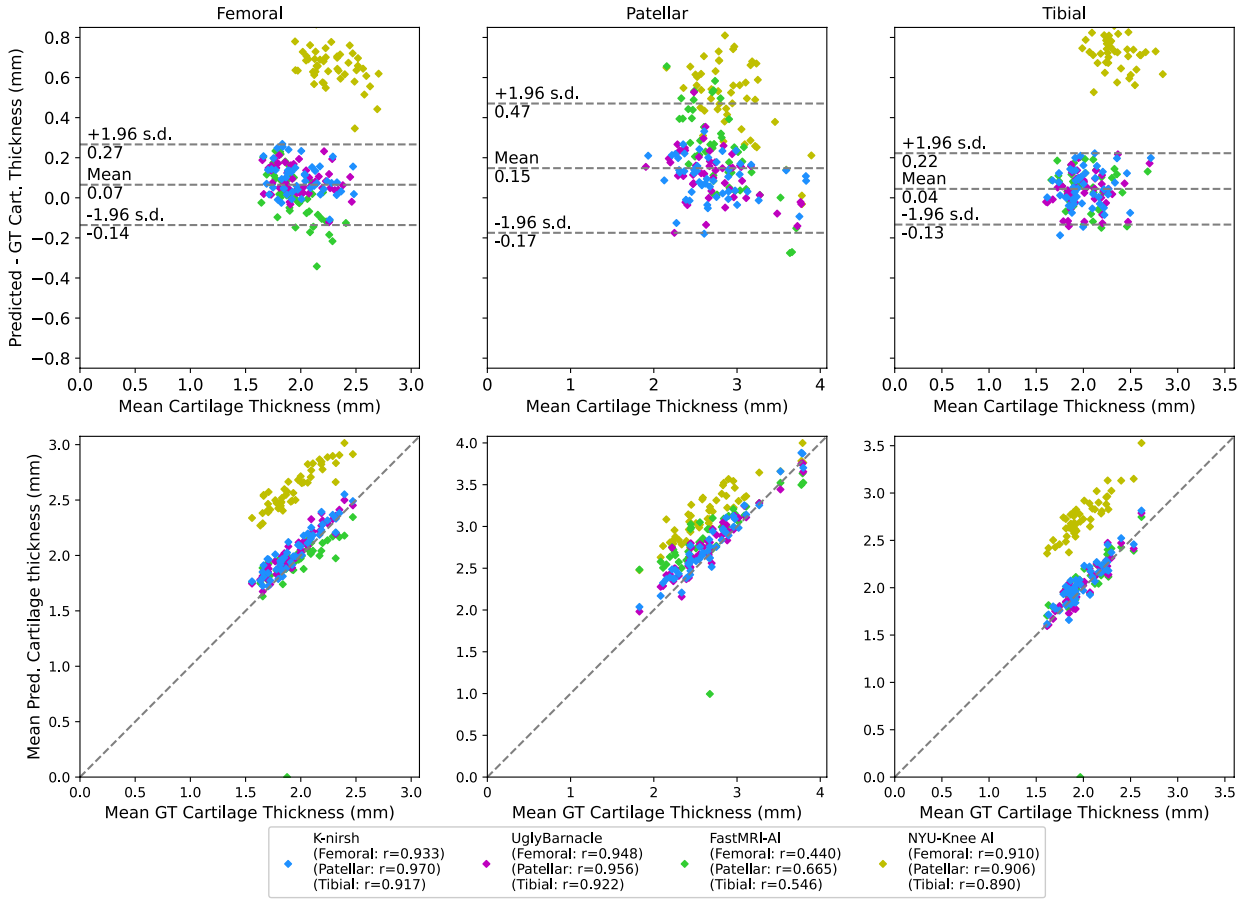


Figure 7.10 Bland-Altman and correlation plots between predicted and ground truth cartilage thicknesses for each of the top 4 submissions, across each of the 3 cartilage compartments. The mean and standard deviations for these plots were calculated using the data points from K-nirsh, UglyBarnacle, and FastMRI-AI, given the thickness overestimations seen from NYU-Knee AI. The top three submissions saw minimal bias and strong fidelity to ground truth, while NYU-Knee AI appeared to slightly overestimate particularly tibial and femoral cartilage thicknesses. That said, correlation plots showed strong correlations between predicted and ground truth thicknesses for K-nirsh, UglyBarnacle, and NYU-Knee AI. FastMRI-AI visually appeared to have strong correlation as well, but an outlier case appears to have severely degraded the correlation

coefficient. All told, these results collectively are quite promising that submissions are suitable for some downstream biomarker analysis.

7.4.5 Biomarker Analysis: Bone Shape

Statistical shape modeling identified 5 femoral shape features most contributing to variation within the test set, as illustrated in Figure 7.11: femoral volume, medial wall incline slope, condylar posterior protrusion, intercondylar notch width, and width-to-height ratio. Similar features were identified for the patella and tibia, and the correlation between submitted bone shapes and ground truth was calculated for the top PCs (and thus, top shape features) for each submission. Those correlation coefficients are shown in Figure 7.12: while each of the top four submissions performed best in at least one of the 15 shape features across the 3 bones, generally K-nirsh had the strongest performance among the teams in the femur, while NYU-Knee AI did best within the tibia and patella. Correlations for all teams were moderate to strong for many of the shape features.

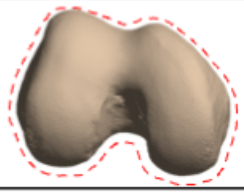
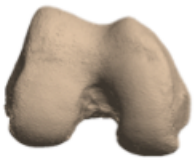
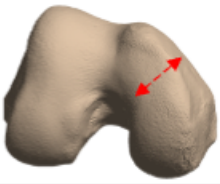
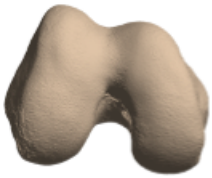
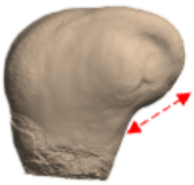
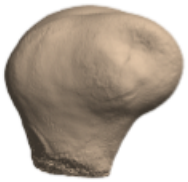
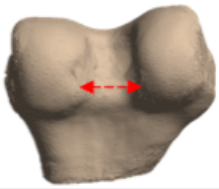
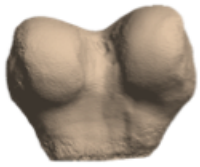
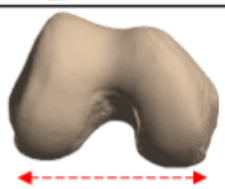
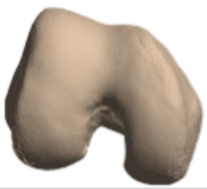
	+3SD	-3SD	Shape Feature Description
1			Femoral volume
2			Medial wall incline slope
3			Condylar posterior protrusion
4			Intercondylar notch width
5			Width-to-height ratio

Figure 7.11 Femoral bone shape features, visualized after statistical parametrization, with qualitative descriptions of shape features. Similar features were also generated for the tibia and patella by the same procedure: extracting Euclidean points of bone surfaces, converting them into 1D vectors, using PCA to compress the resulting matrix into a 5-dimensional one, and visualizing each of the PCs.

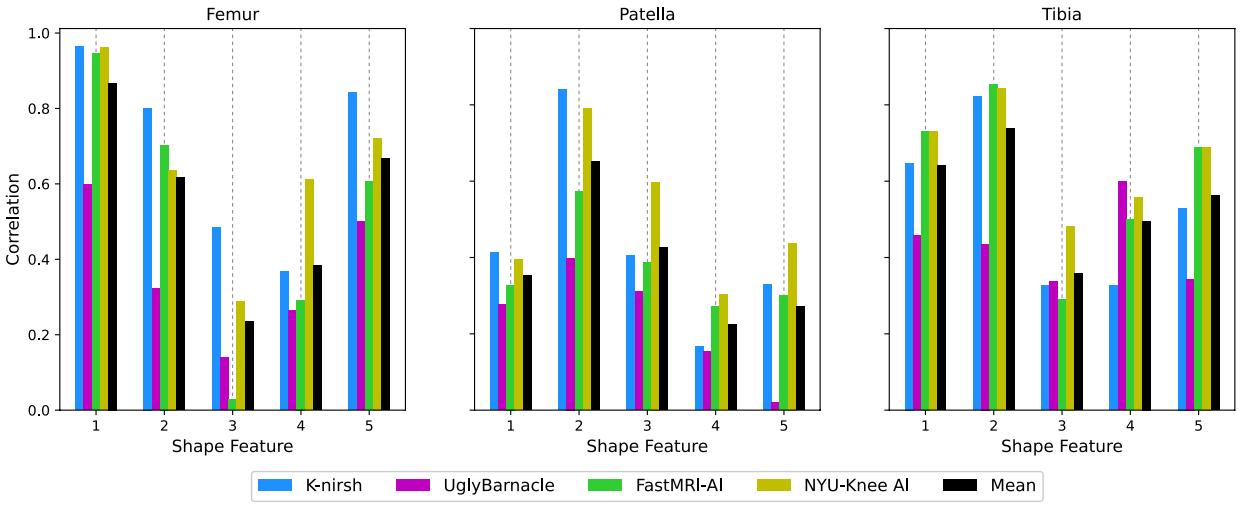


Figure 7.12 Correlations along femoral, tibial, and patellar bone shape features between submissions and ground truth. For many of the bone shape features, correlations were moderate to strong, indicating another means in which submitted segmentations from 8× undersampled images at times were suitable for downstream biomarker analysis. K-nirsh and NYU-Knee AI appeared to have strong correlations most consistently between predicted and ground truth bone shapes among the top 4 submissions.

7.5 Discussion and Conclusions

In this work, we describe the K2S challenge, which aims to reframe image reconstruction and image analysis as end-to-end rather than serial tasks, opening room for optimization. We curated the K2S dataset of 300 patients that had undergone 3D fat-suppressed knee MRI acquisitions, each with 3D raw k-space and bone and cartilage segmentations, challenging participants to segment the tissues directly from 8× undersampled k-space. A variety of solutions were submitted for the challenge. Some, like NYU-Knee AI and UglyBarnacle, spent considerable time optimizing reconstruction networks, leveraging VN and CS frameworks to attain high-quality reconstructions that served as inputs for standard segmentation networks. Interestingly, FastMRI-AI did not pursue a reconstruction network at all, choosing exclusively to optimize the segmentation network and develop unique postprocessing techniques, attaining

very competitive results. K-nirsh, on the other hand, pretrained separate reconstruction and segmentation networks, performing end-to-end optimization of both for weighted DSC. The end-to-end optimization made this the only approach that implicitly optimized reconstruction outputs for segmentation inputs, possibly playing a role in their top finish within the challenge. All told, however, all top submissions produced high-quality segmentations in knee cartilage and bone, maintaining accuracy with respect to ground truth despite working originally from 8× undersampled multicoil k-space.

Beyond strong DSC metrics, predicted segmentations from all top submissions produced cartilage thickness maps that either maintained minimal bias or strong correlation to ground truth cartilage thicknesses. Statistical shape modeling generated five features that captured the most variance in bone shape for each of the patella, tibia, and femur. Each of the top submissions was most correlated to ground truth along at least one of the features, with moderate to good correlations seen in many, while K-nirsh and NYU-Knee AI generally showed the best performance in retaining bone shape. As such, for both bone and cartilage, all top submissions yielded cartilage and bone segmentations that to varying degrees were suitable for subsequent biomarker analysis. An added observation was that downstream biomarker performance did not always correspond with segmentation metrics: for instance, NYU-Knee AI delivered among the best correlations between predicted and ground truth bone shape features despite obtaining the poorest weighted DSC among the top 4 submissions, with segmentations that often appeared slightly dilated compared to ground truth but preserved shape. Likewise, UglyBarnacle slightly outperformed K-nirsh in correlations between

tibiofemoral cartilage thicknesses and ground truth despite slightly poorer weighted DSCs, but its slightly reduced bias was not statistically significant. This accentuates the complexity of segmentation as an image analysis task: there is no all-encompassing, perfect metric to quantify segmentation quality.

A noteworthy finding from this challenge was that strong reconstruction performance was not a prerequisite for strong segmentation performance. K-nirsh had by far the poorest metrics of the submitted pipeline reconstruction intermediates, poorer even than the root sum-of-squares coil-combined 8× undersampled images that FastMRI-AI used as pipeline inputs. Despite this, K-nirsh yielded the strongest segmentation performance; visual inspection of K-nirsh reconstructions reveals sharp images that enhance contrast at boundaries between different tissues such as cartilage/bone boundaries, yielding an image that is perhaps easier to segment than ground truth. This demonstrates that ideal features for radiologist interpretation of an MR image can differ from those optimal for processing by an image analysis algorithm. That FastMRI-AI showed competitive segmentation performance despite directly segmenting undersampled images is a testament to this. Furthering this, was there essentially no correlation between reconstruction and segmentation metrics for any of the top submissions on a per-patient basis. There is therefore room for optimizing image analysis algorithms when trained end-to-end with reconstruction algorithms instead of training separate algorithms and inferring serially. It is important to note that segmentation performance from undersampled k-space depends not only on the segmentation algorithm but also on the undersampling pattern, which was fixed in this challenge. More complicated joint optimization of segmentation and

undersampling can further improve end-to-end MRI reconstruction and image analysis outcomes [247,248].

Apart from the specific challenge, the curation and release of the K2S dataset marks an important initiative that can seed advances in both reconstruction and image analysis algorithm development. To our knowledge, this is the largest released dataset that pairs raw k-space data with tissue segmentations (n = 300 patients, each with an 18-coil, 200-slice k-space). While a dataset of this size is more than sufficient for training most reconstruction algorithms, image annotation algorithms generally require considerably larger datasets to sufficiently represent rare anomaly classes. Our hope is that the release of this dataset can allow research groups to investigate objectives such as ROI-specific image reconstruction, end-to-end reconstruction and segmentation, and more generally end-to-end reconstruction and image analysis tasks.

This challenge had some limitations. First off, the k-space provided to challenge participants had undergone R = 4 ARC, and thus does not reflect the full-length acquisition k-space that would ordinarily be undersampled. Given that the full 3D fat-suppressed CUBE sequence without ARC would require nearly 20 min for acquisition, this compromise was made to make curate a larger dataset suitable for algorithm development. Additionally, while substantial work was done by challenge organizers and radiologists (J.L. and F.G.) in inspecting segmentation quality, bone and cartilage segmentations ultimately were model generated, and were not the gold-standard manual annotations that are desired for training models. It is therefore more accurate to describe the challenge objective as achieving on 8× undersampled data the same segmentation

performance seen on fully sampled data, albeit the latter was carefully monitored and quality checked by radiologists. This tradeoff was taken to obtain a substantially larger dataset than would have been possible if exclusively using manual segmentations. We would expect these findings to hold on a dataset with purely manual segmentations but confirming so would require inferring trained models on such a dataset. Furthermore, this challenge provided a fixed undersampling pattern: a center-weighted Poisson pattern with a fully-sampled center. This undersampling pattern was selected such that potential challenge solutions would not be biased towards or against a given reconstruction backbone (i.e., compressed sensing, deep learning), but there conceivably would be room for further optimization of segmentations with respect to the undersampling pattern. Additionally, since all submissions were trained and tested on a fixed undersampling pattern, the robustness of solutions to other R = 8 undersampling patterns was not assessed and is an important research objective for the reconstruction community to pursue. Lastly, there is no perfect solution to the gradient inhomogeneity correction step (GRADWARP) in the standard processing pipeline of raw scanner data. Once applied, correspondence between k-space and image space is lost, meaning ordinary DICOM image segmentations would not match k-space. In the K2S dataset, segmentations were provided on images prior to GRADWARP application, meaning that gradient coil inhomogeneities manifested themselves into segmentations. Due to the difficulty in backpropagating through GRADWARP, this was viewed as the easier choice for pipeline development, with the understanding that resulting segmentations could be processed by GRADWARP to perform necessary corrections. Nonetheless, this is an unavoidable limitation

that must be discussed at greater length for this and other datasets that may be released pairing k-space and tissue segmentations.

In conclusion, the K2S challenge curated a landmark dataset, tasking participants with segmenting bone and cartilage from 8× undersampled knee MRI images. Through it, the top four teams produced submissions that yielded high-quality segmentations, showing highly varied methodologies and very strong performance that was suitable for downstream biomarker analysis in cartilage thickness and bone shape assessments. Through the submissions of two teams with unconventional approaches—K-nirsh and FastMRI-AI—we clearly see that features required for radiologist annotation differ from those required for DL model input, there is room for image analysis pipeline optimization when trained end-to-end with reconstruction, and strong reconstruction is not a prerequisite for strong segmentation. These findings can motivate similar efforts for end-to-end optimization of image analysis and reconstruction tasks, not only for segmentation, but for anomaly detection, prognosis prediction, bone shape assessment, and others.

Chapter 8 - Synthetic Inflammation Imaging with PatchGAN Deep Learning Networks

The following has been reformatted and reproduced with full permission from the publisher. It appeared in *Bioengineering* as:

Tolpadi, A.A., Luitjens, J., Gassert, F.G., *et al*. Synthetic Inflammation Imaging with PatchGAN Deep Learning Networks. *Bioengineering* **10**:516 (2023).

https://doi.org/10.3390/bioengineering10050516

8.1 Abstract

Background: Gadolinium (Gd)-enhanced Magnetic Resonance Imaging (MRI) is crucial in several applications, including oncology, cardiac imaging, and musculoskeletal inflammatory imaging. One use case is rheumatoid arthritis (RA), a widespread autoimmune condition for which Gd MRI is crucial in imaging synovial joint inflammation, but Gd administration has well-documented safety concerns. As such, algorithms that could synthetically generate post-contrast peripheral joint MR images from non-contrast MR sequences would have immense clinical utility. Moreover, while such algorithms have been investigated for other anatomies, they are largely unexplored for musculoskeletal applications such as RA, and efforts to understand trained models and improve trust in their predictions have been limited in medical imaging.

Methods: A dataset of 27 RA patients was used to train algorithms that synthetically generated post-Gd IDEAL wrist coronal T_1-weighted scans from pre-contrast scans. UNets and PatchGANs were trained, leveraging an anomaly-weighted L_1 loss and global generative adversarial network

(GAN) loss for the PatchGAN. Occlusion and uncertainty maps were also generated to understand model performance.

Results: UNet synthetic post-contrast images exhibited stronger normalized root mean square error (nRMSE) than PatchGAN in full volumes and the wrist, but PatchGAN outperformed UNet in synovial joints (UNet nRMSEs: volume = 6.29 ± 0.88, wrist = 4.36 ± 0.60, synovial = 26.18 ± 7.45; PatchGAN nRMSEs: volume = 6.72 ± 0.81, wrist = 6.07 ± 1.22, synovial = 23.14 ± 7.37; n = 7). Occlusion maps showed that synovial joints made substantial contributions to PatchGAN and UNet predictions, while uncertainty maps showed that PatchGAN predictions were more confident within those joints.

Conclusions: Both pipelines showed promising performance in synthesizing post-contrast images, but PatchGAN performance was stronger and more confident within synovial joints, where an algorithm like this would have maximal clinical utility. Image synthesis approaches are therefore promising for RA and synthetic inflammatory imaging.

8.2 Introduction

Rheumatoid arthritis (RA) is a widespread autoimmune disorder observed in 0.5–1.0% of the American population, with incidence rates being two to three times higher in women than in men [44]. RA mainly affects the joints, typically the hands and feet, and is characterized by synovial joint inflammation. In the joints it can lead to bone tissue erosions and soft tissue breakdown, often inducing stiffness and debilitating pain, but may also show systemic effects in

the skin, heart or lungs if left untreated [45]. It is typically diagnosed through a holistic assessment that begins with a medical history examination, paying particular attention to pain, swelling, peripheral joint pain, and swelling/tenderness, all of which can be indicative of RA. Furthermore, laboratory tests for rheumatoid factor (RF), C-reactive protein (CRP), and erythrocyte sedimentation rate (ESR) are often performed to confirm other RA indications. Lastly, medical imaging plays a crucial role in distinguishing inflammatory phenotypes, providing additional evidence to confirm RA [249]. Once diagnosed, RA is usually treated with Disease-Modifying Anti-Rheumatic Drugs (DMARDs), which see 75–80% of patients attain intended treatment outcomes, but 90% when initiated in the early stages of RA [47]. Robust tools such as imaging are thus necessary for screening and diagnosing RA at early stages, maximizing the odds of successful treatment.

Radiographs have traditionally been the clinical standard imaging modality for RA diagnosis, as their acquisition is quick, inexpensive, and widely accessible, yielding two-dimensional images that are effective in visualizing late-stage bone erosions [250]. In recent years, however, Magnetic Resonance Imaging (MRI) has gained prominence despite its higher costs and longer acquisition time, producing three-dimensional anatomic images with excellent depiction of soft tissues and sharp details [251]. As a result, it has emerged as a superior option for visualizing early-stage bone erosions and bone marrow edema (BME) that can result from RA [46]. An added advantage of MR is the ability to administer contrast agents such as Gadolinium (Gd) prior to scans, altering the magnetic properties of underlying tissue to improve the visualization of numerous pathologies [252]. In RA imaging, a post-contrast Gd MRI can better distinguish

active soft tissue RA sites in joints, such as synovitis, from general effusion [253], conveying critical information that conventional MRI cannot provide [254]. However, Gd administration has long-term concerns such as deposition in brain and bone [255,256], is contra-indicated in patient subgroups such as those with renal diseases and pregnant women [257], and, more generally, adds scan time, cost, and patient discomfort to the imaging protocol. As such, if post-contrast MR images could be synthetically generated without Gd administration, the implications for RA diagnosis and other musculoskeletal (MSK) inflammatory conditions or even sarcomas would be significant.

The problem posed by this clinical context is one of "image synthesis," or the designing of algorithms to generate images from some input. While these inputs can be multimodal, including text or patches of images, the focus here will be on synthesis algorithms that accept full image inputs [258,259]. For image synthesis tasks, deep learning (DL), and particularly convolutional neural networks (CNNs) [260], have taken on an outsized role in recent years. When trained with sufficiently large datasets, CNN filters can be optimized for a given task, with filters in early network layers typically being sensitive to generic features such as edges, while those in later layers are typically sensitive to far more complex, task-specific features [261]. The UNet is a commonly used image synthesis algorithm in which inputted images are encoded by convolutional filters into a low-resolution, high-dimensional representation that is decoded using deconvolutional filters, yielding an output image. Originally designed for segmentation, the UNet has seen substantial application in image synthesis for its ease of training and relatively low dataset size requirements compared to other DL approaches [116]. Another

prominent approach is generative adversarial networks (GANs), where an image-to-image translation network such as a UNet ("generator") is paired with a discriminator network that is trained to distinguish between synthetic and real images [262]. By setting up training as a min-max game in which generator and discriminator networks continually try to fool one another, substantially sharper images can be obtained, although GANs are more difficult to train and are prone to hallucinating artifacts compared to conventional approaches [263]. Other approaches such as variational autoencoders (VAEs) and transformer networks have been investigated in this space [264,265].

These methods have seen considerable application for medical imaging tasks. In brain MRI, image synthesis has been studied for the reduction or elimination of the Gd dosage required for post-contrast tumor imaging. In several studies, standard UNet or encoder-decoder style architectures accepted reduced-dose Gd post-contrast images and/or other MR sequences as inputs, were trained to predict full-dose post-contrast Gd images, and quantified model efficacy through radiologist assessment or the suitability of synthetic images for downstream tasks [123–125]. Another approach in eliminating Gd dosage for brain MRI used an innovative training scheme, training a network for tumor detection and passing convolutional feature maps from that network as inputs to a conventional image synthesis architecture. This allowed the image synthesis architecture to focus on pathologic regions when optimizing parameters to produce synthetic post-contrast images [266]. Some approaches beyond image synthesis have also been investigated to eliminate the need for Gd administration. For instance, Gd is administered in cardiac MRI to identify regions of myocardial infarction. Here, DL pipelines have been developed

to accept exclusively non-contrast MR images as inputs, localize the left ventricle, extract motion-based features inherent to cardiac MRI, and integrate both to predict if a patient suffered from infarction [126,267]. On the other hand, features from non-contrast MR sequences such as synthetic MRI and diffusion weighted imaging (DWI) have proven effective in differentiating benign and metastatic retropharyngeal lymph nodes, a task that usually requires a post-contrast MRI [268]. Also worthy of mention are recent image synthesis applications in biomedical imaging outside of MRI: in histopathology, standard image synthesis generator networks have been paired with multiple discriminators to generate synthetic stained images, while in microscopy, GAN image synthesis pipelines have been applied for synthetic cell painting, identifying cellular components from brightfield microscopy images [269,270].

These works mark substantial progress, with well-validated frameworks yielding promising results on a wide variety of biomedical image synthesis tasks, including post-contrast MR image synthesis. That said, there are some clear gaps in the literature. For RA imaging, the authors are not aware of any previous work developing post-contrast MR image synthesis algorithms. Such algorithms would have immense clinical utility, synthesizing post-Gd images that could be used to identify synovitis and active inflammation sites in RA patients, while eliminating the risks associated with administering Gd. More generally, Gd is used in brain imaging to identify tumors and distinguish tumor types, while in cardiac imaging it helps identify myocardial infarction sites, among others; in MSK, however, it is administered to image inflammation. Synthetic inflammatory MSK imaging has seen little to no investigation in previous works. Particularly in comparison with brain applications, synthetic Gd dosage reduction in MSK applications, such as

wrist imaging, brings about additional challenges such as severe motion artifacts, reduced signal-to-noise ratio (SNR), and considerably smaller datasets [271]. Lastly, despite all these image synthesis works in biomedical applications, efforts to understand the basis of model predictions have been limited; this work would be critical for radiologists to gain confidence in model predictions, a prerequisite for eventual clinical deployment. As such, post-contrast MSK MR image synthesis confers numerous unique challenges that must be managed methodologically, and has been largely unexplored, making it ripe for an initial proof-of-concept study.

This is precisely the niche this work seeks to fill: the purpose of this study was to develop DL pipelines that generate synthetic post-contrast wrist MR images from their pre-contrast counterparts [272], thereby marking the first known effort for synthetic MSK inflammatory imaging. We use image quality metrics to assess the diagnostic and perceptual quality of model-generated synthetic post-contrast images relative to true post-contrast images. We also generate occlusion and uncertainty maps to better understand model performance, making its predictions more trustworthy. More specifically, the contributions and novelty of our work are as follows:

1. To our knowledge, this proof-of-concept study is the first application of DL techniques for generating synthetic post-contrast images for MSK inflammatory imaging.

2. We show that our trained pipelines perform strongly with regards to predicting post-contrast image appearance, particularly in regions afflicted with synovitis, where these models would see the most clinical utility.
3. We investigate the deconvolution operator, checkerboarding artifacts that can be intrinsic to architectures that use it, and how they surface in conventional and adversarial network training schemes.
4. We conduct a rigorous analysis of model predictions, identifying regions in pre-contrast image inputs that were most important to predicted post-contrast images, and regions in which predictions were most uncertain. This provides a straightforward framework that can be used to understand predictions made by image synthesis architectures in biomedical imaging applications.

8.3 Methods

8.3.1 Study Group

All studies performed in this retrospective study were Health Insurance Portability and Accountability Act (HIPAA) compliant, approved by the UCSF Institutional Review Board (Human Research Protection Program, IRB# 12-10418) and registered under Clinical Trial NCT01773681. Informed consent was obtained from all study participants. Twenty-seven UCSF patients with RA were recruited that met the following criteria: at least 18 years old and fulfilled the 2010 ACR/EULAR criteria for the classification of RA. Patients were treated with either methotrexate or a combination of methotrexate and tumor necrosis factor alpha inhibitors (anti-TNFα) based on RA disease activity; intended sample sizes were thus as large as feasible given the exclusion

criteria and the requirements of informed consent from study participants. Data was collected from patients as part of this cohort from 20 March 2014 to 8 February 2018. Patients were imaged at baseline, 3-months, and 1-year follow-up time points, conducting MR imaging, sampling serum to measure ESR, and recording clinical notes at each time point. As the dataset used in this study was from a UCSF clinical trial, data privacy and patient confidentiality concerns prevent its public release, but codes used in generating results can be obtained from the authors upon reasonable request.

8.3.2 MR Acquisition

All patients underwent a standardized protocol that included coronal T_1 IDEAL scans pre- and post-Gd administration on a 3.0-T wide bore scanner (MR Discovery 750w, GE Healthcare, Waukesha, WI, USA) using 8-channel HD wrist array coils (GE Healthcare, Waukesha, WI, USA). Scans were done with acquisition matrices of 384 × 256 (n = 58) or 256 × 224 (n = 6), a slice thickness of 2 mm, a TR of 457 to 793 ms, and a TE of 10.06–12.48 ms. Complete acquisition parameters for both sequences can be found in Supp. Table B.1.

8.3.3 Anomaly Segmentations and Evaluations

In post-contrast images, synovitis was segmented in the following synovial joints: intercarpal joints, carpometacarpal joints, the radioulnar joint, and radiolunar joints. Regions with bone marrow edema (BME) were segmented in the following bones: the first to fifth metacarpals, capitate, hamate, lunate, pisiform, scaphoid, trapezium, trapezoid, triquetrum, ulna, and radius. Anomaly segmentations were performed by a radiologist with over 30 years of experience (T.L.)

using the Image Processing Package (version 6.43.01) developed by the University of California, San Francisco Musculoskeletal Quantitative Imaging Research Group.

T.L. also quantified synovitis severity for each patient at each time point with the Rheumatoid Arthritis Magnetic Resonance Imaging Score (RAMRIS) for synovitis [273], a 0–9 scale in which a higher score is associated with more severe imaging findings of RA.

Lastly, bounding boxes delineating wrist tissue and background were drawn using the software MD.ai by a radiologist with two years of experience (J.L.), such that reconstruction metrics for synthetic post-Gd images could be evaluated solely in wrist tissue and not be sensitive to textures and noise in background pixels.

8.3.4 Image Preprocessing

Six of 64 acquired imaging volumes had slices that were 256 × 256 pixels, with the remainder being 512 × 512; the slices of these six volumes were upsampled to 512 × 512 using third-order b-spline interpolation. Pre-Gd volumes were then registered to post-Gd volumes with a three-step process: (1) translation, (2) affine, and (3) third order b-spline registration (maximum iterations = 256, 256, 512, respectively; Advanced Mattes Mutual Information [274] criterion for all). B-spline registration was only done for scans where the structural similarity index (SSIM) [228] between pre and post-Gd acquisitions was above 0.5; other scans had motion artifacts so severe that non-rigid registration was not possible. All registrations were performed using SimpleITK 2.0.0 in Python (version 3.7.11) [275–277]. Example slices before and after

registration can be found in Supp. Fig. B.1. Pixel values in the slices of pre-Gd scans were scaled such that the middle 95% of pixel values were between 0 and 1. The unscaled pixel values in pre-Gd slices that corresponded to 0 and 1 in the scaled slices were also mapped to 0 and 1 in the post-Gd slices, thereby scaling post-Gd slices while preserving the relative enhancement across the volume.

8.3.5 Data Partitioning

The data were partitioned into training, validation, and test datasets, splitting such that all scans from a given patient were in only one of the three datasets. Furthermore, four patients without imaging findings of synovitis were in the dataset (RAMRIS synovitis of 0); splits ensured at least 1 of these patients were in each of training, validation and test. Splits were intended to maintain similar age, BMI, and ESR across the three datasets, but the relatively small overall dataset required some compromise. The full characteristics of the data splits can be found in Table 8.1.

Table 8.1 Full Dataset and Splits Information. Demographics and patient information for the entire dataset and splits into training, validation and test. All data are presented as mean ± 1 s.d. The dataset consisted of 27 patients diagnosed with RA, each of whom were scanned up to three times (baseline, 3-month, and 1-year follow-up after one of two treatments). Data splitting was done at a patient level while ensuring each of the training, validation and test datasets included at least one patient with a RAMRIS synovitis of 0. The small dataset size and splitting conditions caused slight imbalances in demographic and health variables across the splits.

	Train	Validation	Test	Full
Age	53.38 ± 13.50	45.94 ± 16.16	52.12 ± 18.60	52.41 ± 14.65
BMI	29.35 ± 8.90	25.32 ± 3.06	28.33 ± 1.26	28.79 ± 8.03
ESR [mm/hr]	29.06 ± 26.07	32.00 ± 24.00	27.00 ± 20.12	29.05 ± 25.32
RAMRIS Synovitis	4.57 ± 2.13	2.33 ± 2.62	1.67 ± 1.25	4.00 ± 2.37
Slices	783	87	105	975
Volumes	51	6	7	64

8.3.6 Network Architecture

All network architectures were implemented in PyTorch (version 1.10.2). Two-dimensional UNet [116] architectures were used as image-to-image synthesizers in our approaches, accepting as input a pre-processed pre-Gd coronal T_1 IDEAL slice and outputting the corresponding synthetic post-Gd slice. A baseline UNet model was trained, and in a separate pipeline version, an identical UNet was treated as a PatchGAN generator and paired with a PatchGAN discriminator [278]. The PatchGAN discriminator accepted concatenated inputs of the pre-processed pre-Gd slice and either the corresponding synthetic post-Gd slice or the ground truth post-Gd slice, yielding a 16 × 16 output in which each output pixel had a corresponding receptive field "patch" in the concatenated inputs. The 16 × 16 outputs were trained to predict whether synthetic post-Gd generator outputs were real or synthetic. Multiple baseline UNet and PatchGAN generator versions were trained: one set in which all steps of the UNet/generator decoding path used a deconvolution operator, and another in which the deconvolutions were replaced by either a 2 × 2 bilinear upsampling interpolation operator followed by a convolution [279], or just the 2 × 2 bilinear interpolation. The exact network architecture and layers can be seen in Figure 8.1. Weights for the UNets, UNet generators, and PatchGAN discriminators were initialized randomly to have a mean of 0 and a standard deviation of 0.02.

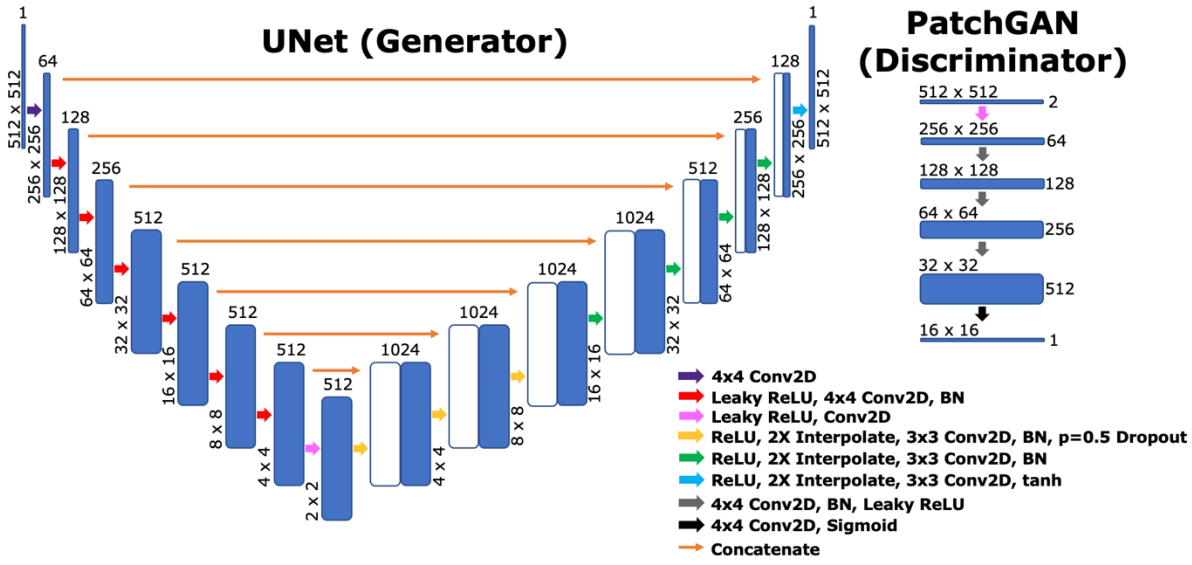


Figure 8.1 Network Architectures. The baseline UNet and PatchGAN generators used identical architectures, while the PatchGAN pipeline also trained a discriminator whose architecture is pictured. All generator encoding path convolutions had a stride of 2 and a padding 1, while all decoding path convolutions had a stride and padding of 1. The first three discriminator convolutions had a stride of 2 and a padding of 1, while the final two had a stride and a padding of 1. For PatchGAN and UNet pipelines with deconvolutions, all "2X interpolate, 4 × 4 Conv2D" steps would be replaced by 4 × 4 2D transposed convolutions with a stride of 2 and a padding of 1. All leaky ReLU layers had a negative slope of 0.2.

8.3.7 Training Details

The baseline UNets were trained with a weighted L_1 loss, as shown below in Equation 8.1, with loss function variables as follows: n = number of samples; S_i = anomaly segmentation mask for slice i; $\hat{y}_i$ = synthetic post-Gd image slice; y_i = ground truth post-Gd slice. The anomaly segmentation mask S_i used to weight the L_1 loss was calculated as follows: anomaly segmentations were turned into binary masks, any pixel more than 20 pixels from the nearest anomaly was set to a background value λ_B, pixels within anomalies were set to 1, and intermediate pixels were set to a range from λ_B to 1 based on their Euclidean distance from an anomaly segmentation. A sample distance map can be found in Supp. Fig. B.2.

$$L_{UNet} = \frac{1}{n}\sum_{i=0}^{n} S_i(\widehat{y_i} - y_i)$$

Equation 8.1 Baseline UNet, pixel-based loss function. Variables: n = number of samples; S_i = anomaly segmentation mask for slice i; $\widehat{y_i}$ = synthetic post-Gd image slice; y_i = ground truth post-Gd slice.

On the other hand, PatchGAN generators were trained with the same weighted L_1 loss and a GAN loss, as shown in Equation 8.2, while PatchGAN discriminators were trained with the loss function shown in Equation 8.3. Additional variables for these loss functions are as follows: x_i = pre-Gd image slice; $D(a, b)$ = PatchGAN discriminator output for concatenated inputs a and b; λ_{L_1} = anomaly-weighted L_1 loss weighting for generator; λ_{GAN} = discriminator loss weighting for generator. With this loss function setup, the discriminator was trained to predict values of 1 when fed ground truth data and 0 when fed generator predictions, while the generator was trained to do the opposite. For any training batch, the following scheme was followed: (1) synthetic post-Gd generator predictions were calculated; (2) pre-Gd, synthetic post-Gd, and ground truth post-Gd images were used to calculate L_{Dis} and update discriminator parameters; (3) synthetic post-Gd generator predictions and corresponding discriminator outputs were recalculated with new model parameters, L_{Gen} was calculated, and generator parameters were updated; (4) steps (1) and (2) were repeated again to update the discriminator parameters. This approach of two discriminator steps and one generator step per training batch was empirically useful in yielding similar generator and discriminator strength during training.

$$L_{Gen} = \frac{1}{n}\sum_{i=0}^{n} \lambda_{L_1} S_i(\widehat{y_i} - y_i) - \lambda_{GAN} \log D(x_i, \widehat{y_i})$$

Equation 8.2 PatchGAN generator loss function. Variables: n = number of samples; S_i = anomaly segmentation mask for slice i; x_i = pre-Gd image slice; $\widehat{y_i}$ = synthetic post-Gd image slice; y_i =

ground truth post-Gd slice; $D(a, b)$ = PatchGAN discriminator output for concatenated inputs a and b; λ_{L_1} = anomaly-weighted L_1 loss weighting for generator; λ_{GAN} = discriminator loss weighting for generator.

$$L_{Dis} = \frac{1}{2n}\sum_{i=0}^{n} \log D(x_i, \hat{y_i}) - logD(x_i, y_i)$$

Equation 8.3 PatchGAN discriminator loss function. Variables: n = number of samples; x_i = pre-Gd image slice; $\hat{y_i}$ = synthetic post-Gd image slice; y_i = ground truth post-Gd slice; $D(a, b)$ = PatchGAN discriminator output for concatenated inputs a and b.

Baseline UNets, PatchGAN generators, and PatchGAN discriminators were all trained with a learning rate of 0.001, an Adam optimizer ($\beta_1 = 0.5$, $\beta_2 = 0.999$), and batch size of 1 to ensure that full batches fit on a single GPU [97]. All pipelines were trained on an NVIDIA Titan Xp 12 GB GPU. For baseline UNet and PatchGAN generator inputs, the following augmentations were done on the training set, each with a probability 0.5: [−2,2] degree random rotation, [−10,10] pixel random translation along both directions in a slice, [−5,5] percent random zoom, and Gaussian noise addition with a mean of 0 and standard deviation of 0.02. Training was done in two stages: initially for 10 epochs in a hyperparameter search to optimize λ_{GAN} and λ_B (more thoroughly described in the following subsection), and finally for 35 epochs with optimized parameters. With 783 pairs of pre and post-Gd slices seen in the training set, this means that 27,405 total slices were seen by all selected models during training (3,045 additional slices for validation).

8.3.8 Hyperparameter Search and Model Selection

For each of the four pipelines trained (UNet and PatchGAN, both with and without deconvolutions), grid hyperparameter searches were carried out to optimize the background pixel weighting in segmentation distance maps (0, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2) and λ_{GAN}

(0.001–0.01, spaced by 0.001). λ_{L_1} was held constant at 1 for all searches. In hyperparameter searches, models were trained for 10 epochs and model performances were evaluated on the validation set. The most promising parameter set for each of the four pipelines was then trained from scratch for 35 epochs to yield the final models.

The selection of optimal parameter sets was done through a combination of standard reconstruction metrics and visual inspection. For each of the four pipelines, SSIM and normalized root mean square error (nRMSE) were used to screen for top candidate models, whose performance on the validation set was then assessed by visual inspection. The primary criteria for evaluating model performance were (1) the synthesis of new information not obvious from pre-Gd scans, (2) the preservation of sharp textures in synthetic post-Gd scans compared to ground truth post-Gd scans, and (3) the absence of obvious algorithm-generated artifacts that may cause a radiologist to lose confidence in the reconstructed image quality.

8.3.9 Model Performance Evaluation

The assessment of whether to use or omit the deconvolutions in the UNet decoding path was done visually for the UNet and PatchGAN approaches; the best performing models for both methods were then used for a more rigorous analysis. The quantitative assessment of synthetic post-Gd image quality was performed using three standard reconstruction metrics: SSIM, nRMSE, and peak signal-to-noise ratio (PSNR) [227]. Due to the slight misregistration of corresponding slices that may have been present even after previous preprocessing, metrics were presented both with and without slice-wise registration: ((1) 256-iteration translation, (2)

256-iteration affine, and then (3) 512-iteration third order b-spline with a transformation bending penalty of 500, all with the Advanced Mattes Mutual Information criterion). The slice-wise registration was solely for the calculation of model performance metrics; only unregistered model outputs are presented in figures. The reconstruction metrics were evaluated per-volume in the following regions: full imaging volumes, wrist anatomy bounding boxes, and synovial joints. While these metrics do not correlate well with gold-standard radiologist annotations when evaluated on full image volumes or slices, they are widely used in the image reconstruction and image synthesis literature, and thus facilitate easy comparison of model performance with those performing similar tasks [141,142]. Furthermore, our dataset affords us wrist and anomaly bounding boxes; the calculation of these metrics specifically in these regions—one discarding background, and another focusing specifically on tissues of highest clinical interest when administering Gadolinium—can overcome the limitations of these metrics when used conventionally, affording them more clinical significance.

8.3.10 Enhancement Maps

For UNet, PatchGAN, and ground truth post-Gd images, pixels among the top 10% in predicted signal enhancement were identified. Enhancement maps were shown as follows: pre-Gd slice, post-Gd slice, and post-Gd slice with the degree of enhancement overlaid for the most enhancing pixels (top 10%), colored by the predicted extent of the enhancement. For visual consistency, colormap ranges for the enhancement map were calculated with respect to the enhancement observed in ground truth, with the same ranges being used for the maps regardless of algorithmic approach.

8.3.11 Occlusion Maps

For each slice, pre-contrast IDEAL T_1 images were pre-processed using previously described techniques, which were used as inputs for UNet and PatchGAN generator architectures, generating network outputs. The pixel values were then set to zero in a 32 × 32 occlusion, and the occluded image was fed through the same architecture, recording the absolute difference in predicted pixel magnitude as compared to the unoccluded image. This procedure was repeated for all 32 × 32 occlusions throughout the slice (with a stride length of 8), summing up the predicted changes in pixel magnitudes in an aggregate array and dividing each pixel by the number of occlusions in which it was contained. The aggregate array values were then min-max normalized, divided by pre-contrast IDEAL T_1 pixel values (to incorporate into resulting maps information for regions other than areas of high pixel intensity), and again min-max normalized, yielding occlusion maps. For display purposes, the maps are thresholded such that only the top 5% of the occlusion map magnitudes were visualized.

8.3.12 Uncertainty Maps

The uncertainty maps of the model predictions were generated by corrupting the latent representations of a given slice [280]. Namely, for 100 iterations, Gaussian noise with a mean of 0 and a standard deviation of 0.5 was added to the encoding path outputs at each of the eight levels (seven layers that were concatenated to the corresponding decoding path levels and the bottom of the encoder). The variance of the predicted pixel intensities from these 100 perturbed latent spaces was then calculated, min-max normalized, and thresholded for display

purposes such that only the 15% most variant pixels would display, thereby generating uncertainty maps for each slice.

8.3.13 Statistical Analysis

To assess if synthetic post-Gd scans provided significant improvements over baseline pre-Gd images, 2-sample t-tests [281] were conducted. On a per-scanned-volume-basis, these tests compared the metrics of model outputs (nRMSE, SSIM, PSNR) to those of the pre-Gd scanned volumes; a Bonferroni correction [282] was applied when necessary to adjust for multiple comparisons.

8.4 Results

The hyperparameter search results are presented on the validation set, which was used to select optimal values for λ_{GAN} and λ_B in training loss functions. The results from finalized models are presented on the test set, on which finalized models were run just one time. Key demographic information on the test set is available in Table 8.1.

8.4.1 Model Parameter Selection

The reconstruction performance metrics evaluating the similarity of the synthetic post-Gd model outputs to ground truth were calculated for all 70 tested hyperparameter combinations for each of the four model type configurations (PatchGAN and baseline UNet, with and without decoding path deconvolutions). Sample results are shown for PatchGAN without generator deconvolutions for SSIMs in Supp. Table B.2, and for nRMSEs in Supp. Table B.3.

Hyperparameter combinations with strong performances in either approach were carried onto a visual inspection of post-Gd synthesis performance, an example of which is shown for several hyperparameter combinations in Supp. Fig. B.3, also for the PatchGAN without deconvolutions. Hyperparameters associated with the selected best models through this process are listed below:

- PatchGAN, no deconvolutions: λ_B=0.05, λ_{GAN}=0.01;
- PatchGAN, with deconvolutions: λ_B=0.15, λ_{GAN}=0.001;
- UNet, no deconvolutions: λ_B=0.05;
- UNet, with deconvolutions: λ_B=0.15.

8.4.2 Utility of Deconvolution Operators in UNet Decoders

A comparison of sample synthetic post-Gd slices with and without deconvolutions in the UNet decoding path can be found in Figure 8.2, while a comparison of synthetic post-Gd slices with and without deconvolutions in the PatchGAN generator decoding path can be found in Figure 8.3. In baseline UNet pipelines, checkerboarding artifacts were apparent when deconvolutions were used, particularly in regions of relatively homogenous pixel values, such as the muscles around the radius and ulna. When those deconvolutions were replaced by 2 × 2 upsampling and standard convolutions, the checkerboarding artifacts were largely absent. These checkerboarding artifacts were less apparent in both PatchGAN pipelines, but in the version that used deconvolutions, they were evident at the extended boundaries of sharp changes in pixel intensities. Checkerboarding was thus best avoided by PatchGAN and UNet pipelines without

deconvolutions, and these pipeline versions were selected as top-performing pipelines for both approaches in the remaining experiments.

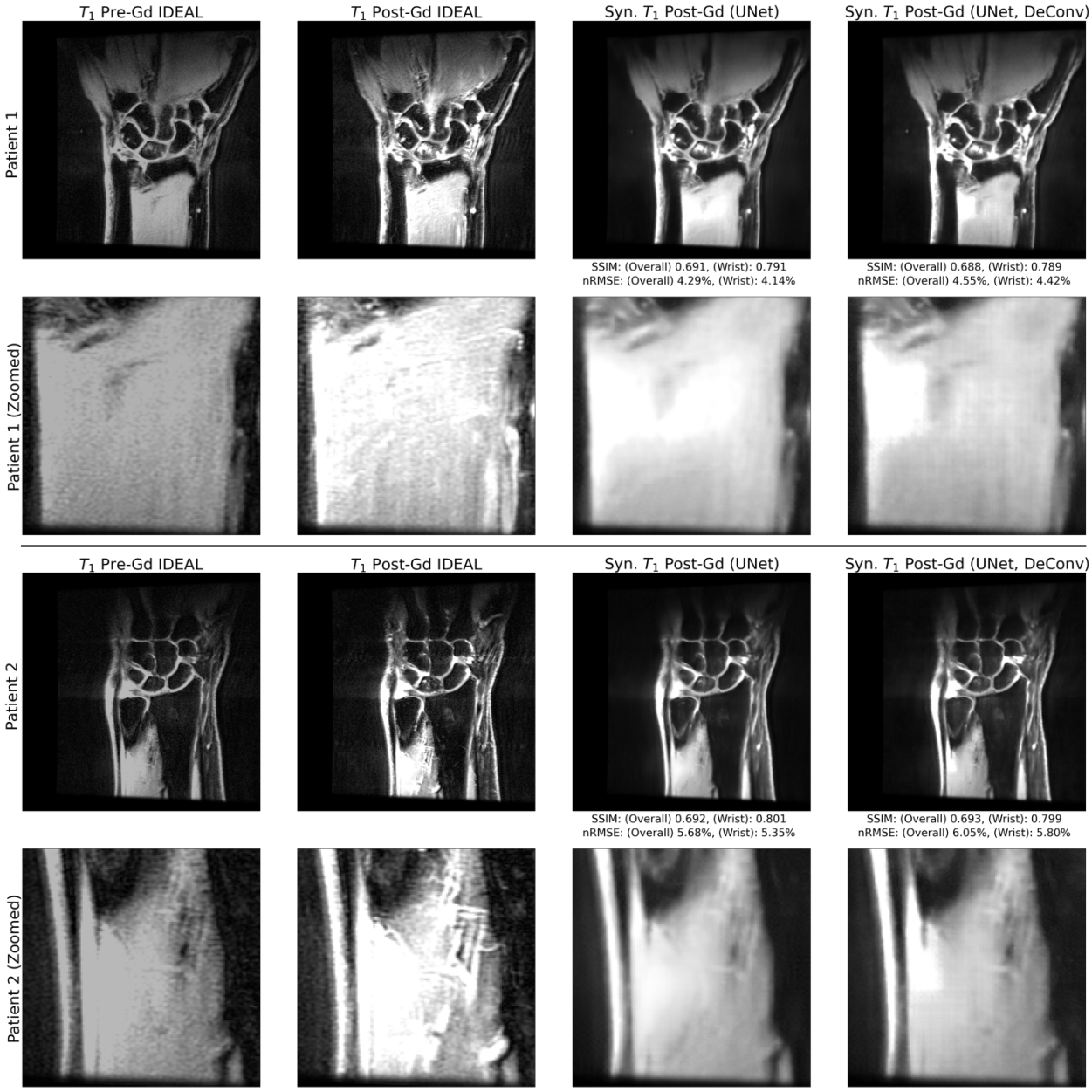


Figure 8.2 Network Performance with and without Deconvolutions in Decoding Path of Baseline UNet. The performance on example test set slices for baseline UNet, with and without decoding path deconvolutions, with zoomed insets. The use of decoding path deconvolutions in baseline UNets induces checkerboarding artifacts in larger regions of relatively homogenous pixel values,

such as the forearm muscle insets (particularly evident in patient 1). When replaced with convolution and interpolation operators, these artifacts were substantially mitigated, making this the preferred architecture when training baseline UNets.

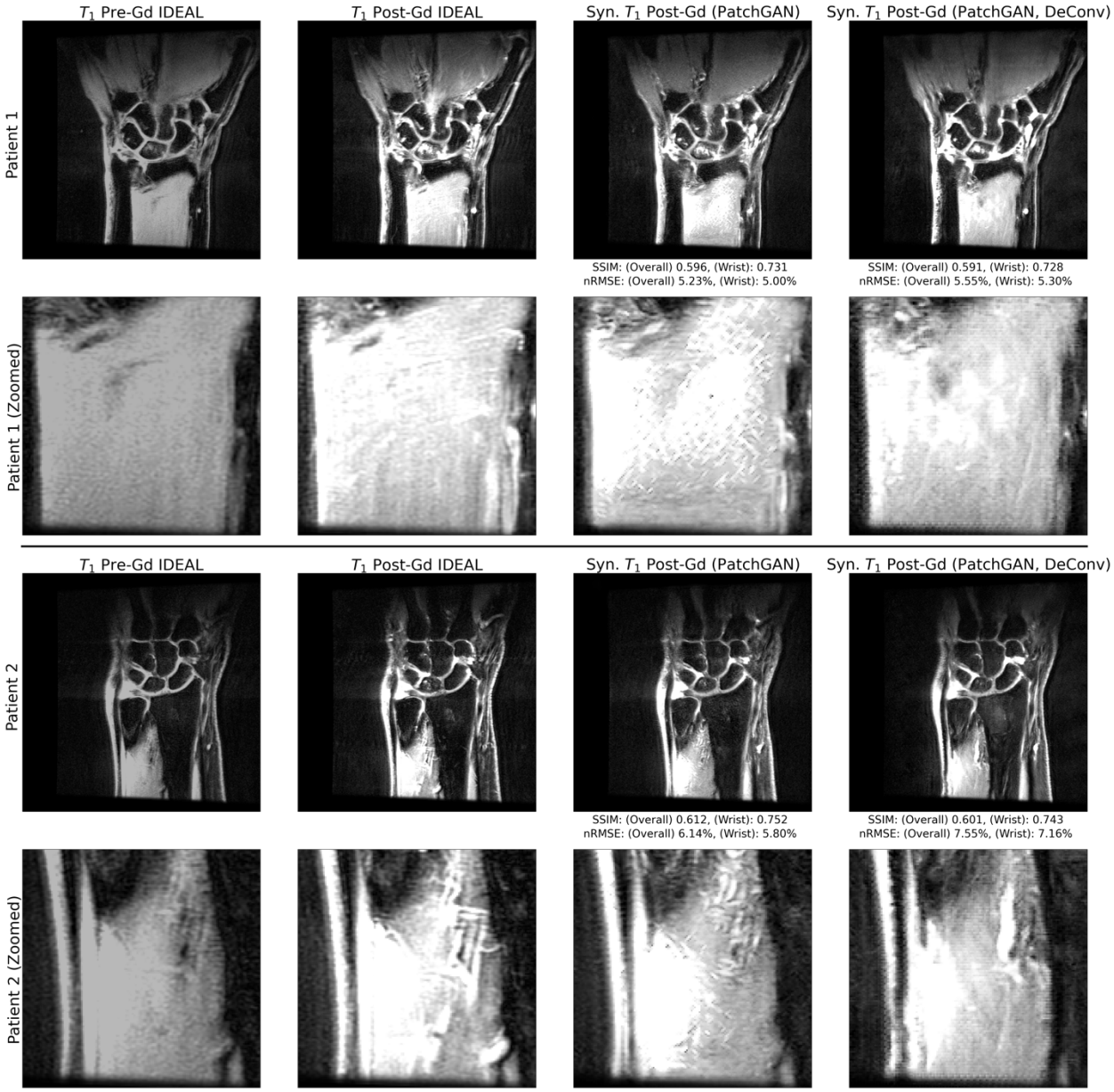


Figure 8.3 Network Performance with and without Deconvolutions in Decoding Path of PatchGAN Generator. The performance on example test set slices for PatchGAN pipelines, with and without generator decoding deconvolutions, with zoomed insets. At sharp transitions in pixel intensities, such as intersections of the radius and ulna with muscles displayed in insets, clear checkerboarding is observed when deconvolutions are used. This was substantially reduced when deconvolutions were replaced with convolutions and interpolation; PatchGAN

generators with these decoding path operations were thus used when training PatchGAN pipelines in the remainder of this paper.

8.4.3 Standard Reconstruction Metrics Performance

Standard reconstruction metrics across the test set are shown in Table 8.2 for full imaging volumes, wrist volumes, and synovial joints. Both synthetic post-Gd volumes had showed significant improvements over pre-Gd volumes in PSNR and nRMSE, with the baseline UNet pipeline also showing significantly higher SSIM. While the UNet baseline model showed stronger performance in all metrics within full volumes and the wrist, the PatchGAN showed stronger reconstruction performance in synovial joints when measured by nRMSE and PSNR.

Table 8.2 Coronal IDEAL Post-Gd T_1 Image Synthesis Performance for Select Pipelines. Standard reconstruction metrics of the PatchGAN and baseline UNet pipelines were evaluated on a per-patient basis within the test set (n = 7) for entire imaging volumes ("full"), wrist tissue in each volume ("wrist"), and synovial joints. Metrics were calculated with and without three-stage nonlinear registration of synthetic post-Gd volumes to ground truth. UNet pipelines reflect the stronger bulk reconstruction metrics in full volumes and within wrist tissue, but the PatchGAN pipeline shows stronger performance in synovial joints in which an algorithm like this would see most clinical utility. Bonferroni-corrected 2-sample t-tests showed nearly all pipelines offered significantly better metrics than Pre-Gd baselines (n = 7; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

		Full	Wrist Only	Synovial Joints
Pre-Gd	**nRMSE**	26.30 ± 9.16	17.82 ± 6.31	260.24 ± 158.56
	PSNR	17.77 ± 0.95	22.99 ± 0.91	8.94 ± 1.64
	SSIM	0.60 ± 0.03	0.94 ± 0.00	
PatchGAN Registered	**nRMSE**	6.72 ± 0.81***	6.07 ± 1.22***	23.14 ± 7.37**
	PSNR	20.77 ± 0.65***	25.40 ± 1.24**	12.10 ± 1.34**
	SSIM	0.58 ± 0.02	0.94 ± 0.01	
PatchGAN Unregistered	**nRMSE**	8.46 ± 1.03***	7.68 ± 1.41**	28.96 ± 10.57**
	PSNR	19.85 ± 0.69***	24.38 ± 1.21*	11.23 ± 1.52*
	SSIM	0.56 ± 0.02*	0.94 ± 0.01	
UNet Registered	**nRMSE**	6.29 ± 0.88***	4.36 ± 0.60***	26.18 ± 7.45**
	PSNR	22.03 ± 0.60***	27.13 ± 0.69***	11.58 ± 0.93**
	SSIM	0.69 ± 0.02***	0.95 ± 0.00**	
UNet Unregistered	**nRMSE**	7.73 ± 1.03***	5.38 ± 0.73***	29.69 ± 7.60**
	PSNR	21.20 ± 0.62***	26.20 ± 0.77***	10.98 ± 0.87*
	SSIM	0.68 ± 0.02***	0.95 ± 0.01*	

8.4.4 Comparison of Reconstruction Performance Across Synovitis Severity

The image quality metrics for synthetic post-Gd volumes are shown in Supp. Table A.4 for test set patients without imaging findings of RA synovitis (RAMRIS synovitis = 0, n = 2) and those with imaging findings of RA synovitis (RAMRIS synovitis > 0, n = 5). Though the sample size limits the power of these conclusions, the metrics were slightly stronger for RAMRIS > 0 than for RAMRIS = 0. Visual examples of the reconstructed post-Gd volumes for a RAMRIS = 0 and RAMRIS > 0 patient are shown in Figure 8.4. In the RAMRIS = 0 patient with no imaging findings of synovitis, the absence of synovial enhancement was captured by both pipelines, whereas in the RAMRIS > 0 patient, UNet and PatchGAN pipelines illuminated similar enhancement patterns in intercarpal regions, with the PatchGAN pipeline depicting sharper enhancement pattern contours, particularly in the muscles and bones.

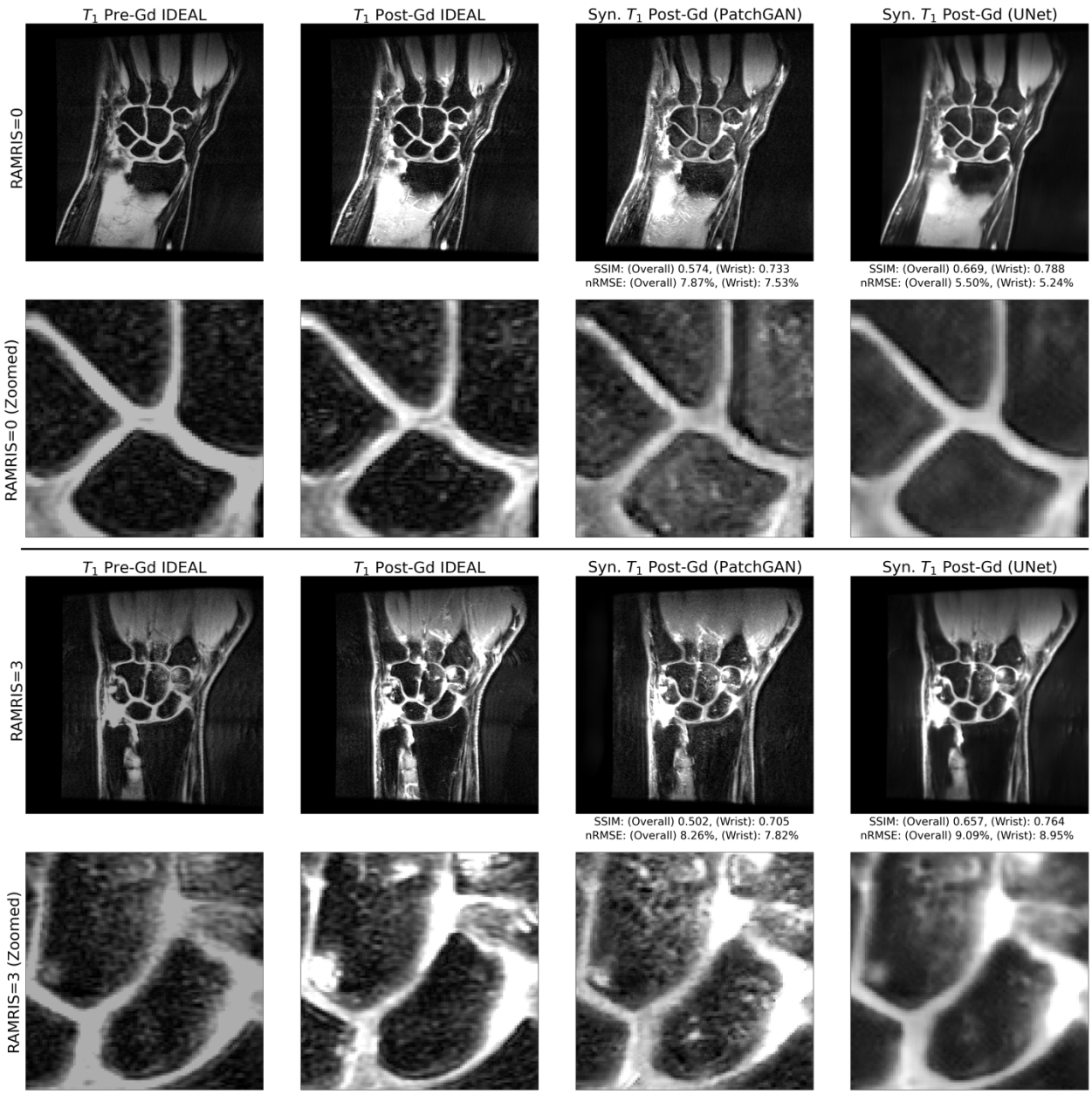


Figure 8.4 Visual Comparison of Reconstructed Post-Gadolinium Images with and without Imaging Findings of RA. Two example test set slices reconstructed by baseline UNet and PatchGAN pipelines for patients with and without imaging findings of RA (RAMRIS = 3, RAMRIS = 0, respectively). There was little to no enhancement in the synovial joints of the RAMRIS = 0 patient, which is captured by both pipelines, as seen in the zoomed insets. In the RAMRIS = 3 patient, the contours of enhancement in the zoomed inset were captured well within the intercarpal joint for both pipelines, with noise distribution patterns better reconstructed by the PatchGAN. The reconstruction performance thus shows promise for patients with and without imaging findings of RA.

8.4.5 Enhancement Maps Analysis

Enhancement maps are shown for an example slice for the PatchGAN and UNet models, as well as ground truth, in Figure 8.5. The enhancement maps show that for the PatchGAN model, general magnitudes of uptake were much more accurately preserved than for the UNet, most notably across intercarpal joints. The predicted enhancement locations were visually very similar for both pipelines.

8.4.6 Occlusion and Uncertainty Maps Analysis

Occlusion maps for the UNet and PatchGAN pipelines in sample test set slices are shown in Figure 8.6. Encouragingly, occlusion maps for both pipelines show a substantial focus on intercarpal joint regions in terms of their relative importance to the predicted pixel values. Peripherally to the intercarpal joint, the occlusion maps show some focus on muscles as well, perhaps slightly more so for the UNet than for the PatchGAN. On the other hand, the uncertainty maps are shown in an example test set slice for UNet and PatchGAN pipelines in Figure 8.7. The UNet shows considerable uncertainty in intercarpal joint region predicted pixel values, whereas for the PatchGAN, uncertainty was highest in the background and within the muscles. PatchGAN also showed some uncertainty in predictions within bones such as the radius and ulna, as well as within bone marrow edema regions; notably, however, uncertainty was limited in the synovial joints.

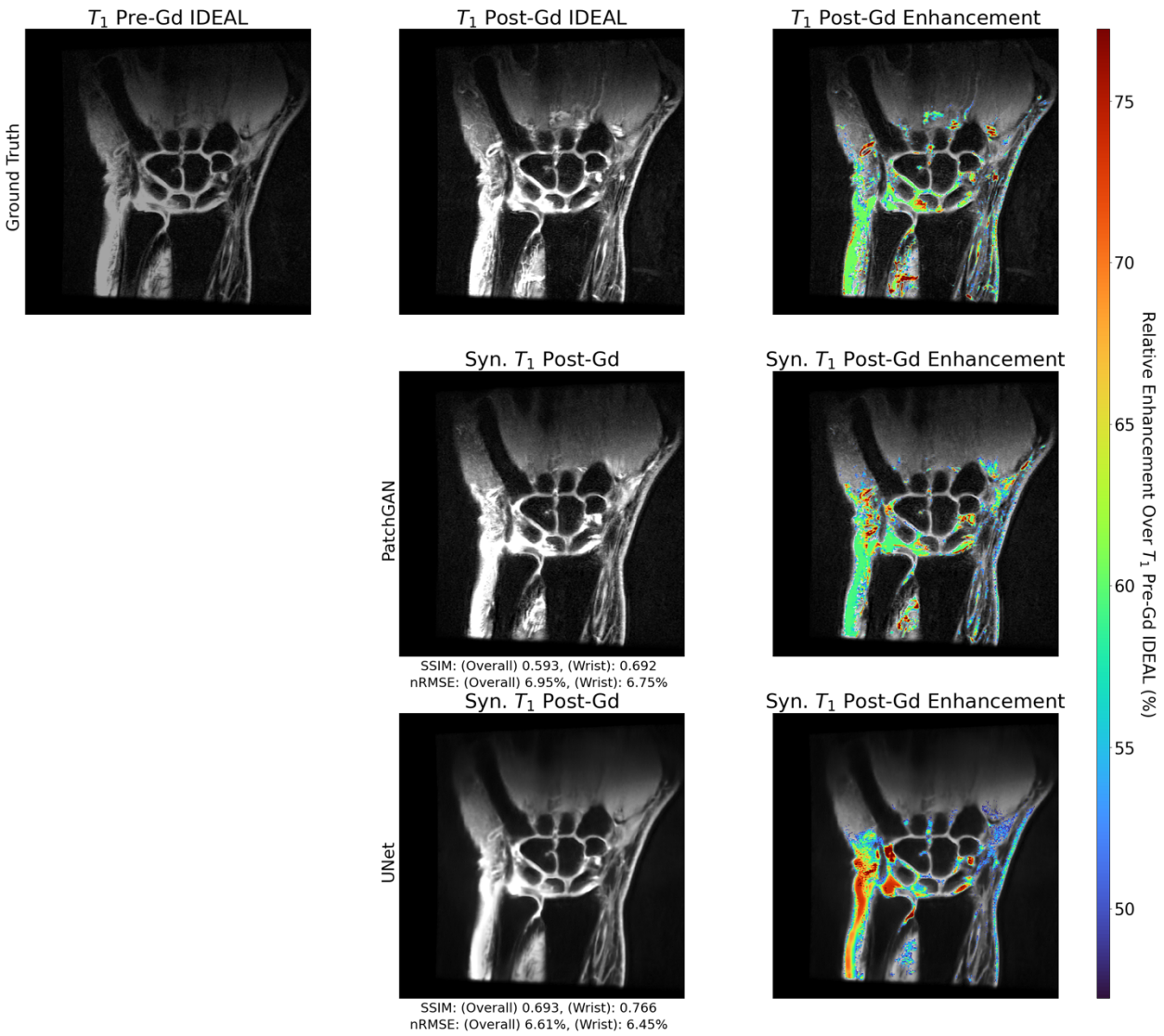


Figure 8.5 Predicted Gadolinium Enhancement Maps with PatchGAN, UNet, and Ground Truth Models. Enhancement maps were generated by identifying the magnitude of pixel intensity increase from synthetic or ground truth Post-Gd slices compared to corresponding pre-Gd slices, and by highlighting the top 10%. While the performance in preserving the location of these top 10% of enhancing pixels was similar for the baseline UNet and PatchGAN, the enhancement magnitudes were far better preserved globally by the PatchGAN, including intercarpal regions susceptible to synovitis. These maps reflect the long-term vision of a pipeline like this: given a pre-Gd scan, the algorithm can identify locations susceptible to synovitis and distinguish active inflammatory sites from general effusion with additional model development.

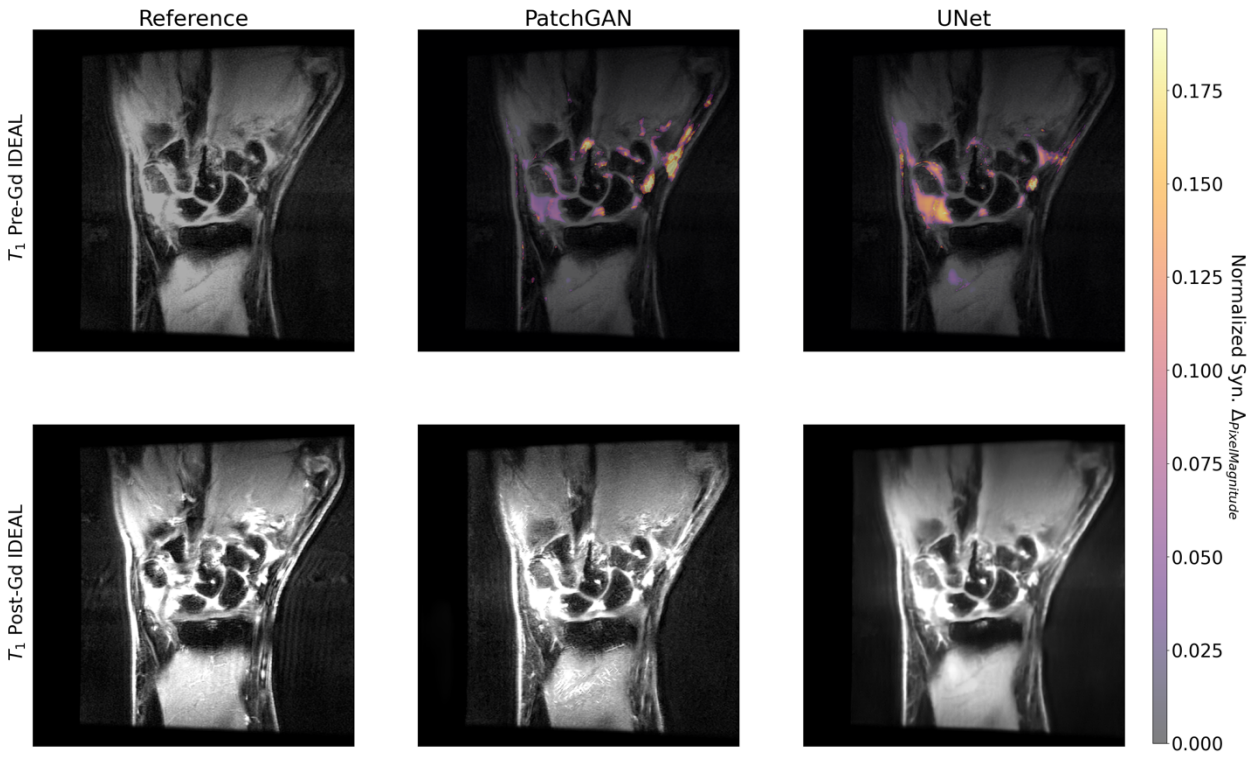


Figure 8.6 Occlusion Maps for PatchGAN and UNet Pipelines. Occlusion maps were generated for PatchGAN and UNet by occluding 32 × 32 patches of the input slices and assessing changes in predicted pixel values compared to unoccluded slices. Occlusion maps were then normalized by pre-Gd pixel intensities and thresholded to identify hotspots most impactful in model predictions. For UNet and PatchGAN, hotspots primarily included intercarpal joint regions. Particularly for the UNet, the maps also showed some emphasis on the forearm muscles. Given that the synovial joints are where an inflammatory imaging algorithm would see the most utility, the fact that both algorithms placed heavy emphasis on the intercarpal regions was promising, indicating that both focused on synovitis-relevant regions to make predictions.

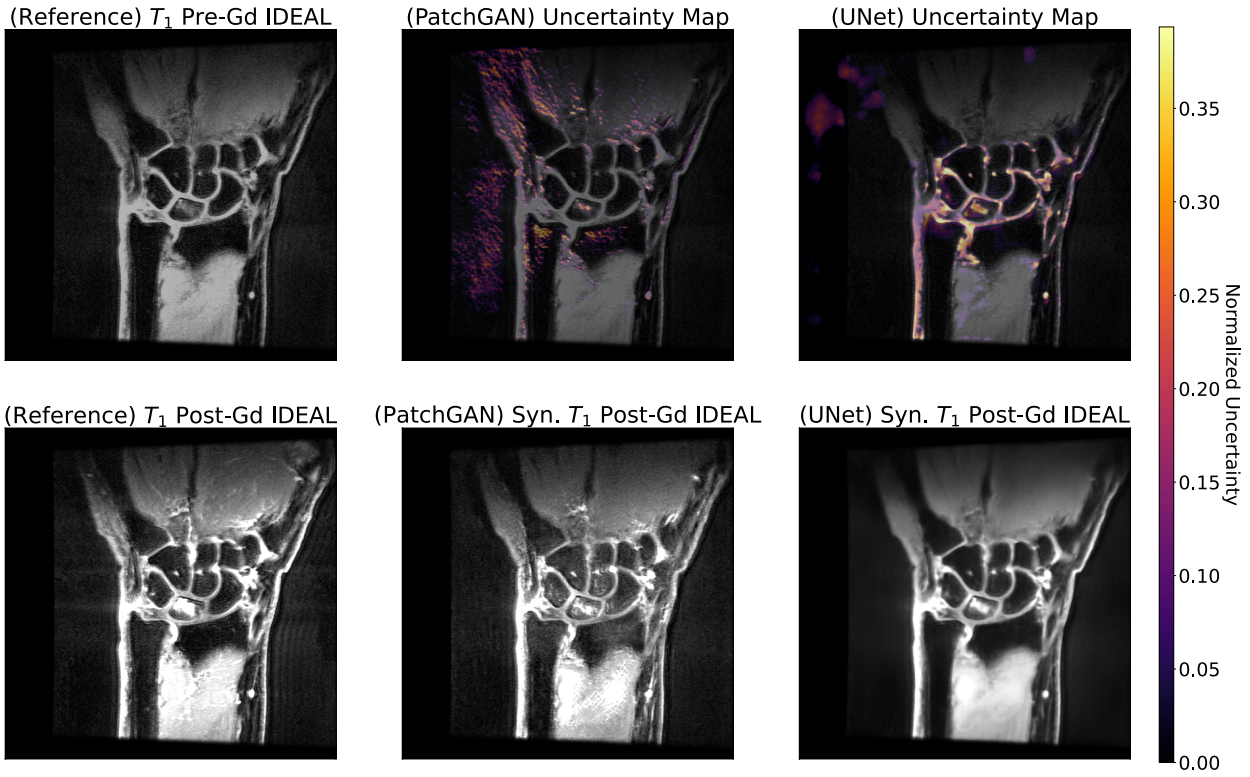


Figure 8.7 Uncertainty Maps for PatchGAN and UNet Pipelines. Uncertainty maps for the PatchGAN generator and baseline UNet were generated by corrupting the latent space of all encoding path outputs, adding Gaussian noise with a mean of 0 and a standard deviation of 0.5 for 100 iterations, and calculating the variance in the predicted pixel magnitudes for each output pixel across these iterations. The most variant pixels were designated as the most uncertain ones. For the PatchGAN, the uncertain regions were mainly in the background, muscles, and within bones. For baseline UNet, the uncertainty maps placed a heavy emphasis on the intercarpal joint, with some residual highlighting of background. In conjunction with occlusion maps, PatchGAN generator predictions were more confident and less uncertain within intercarpal joint regions compared to the baseline UNet. Considering that the intercarpal joint is crucial for synovitis diagnosis and is where both algorithms would be the most useful, the PatchGAN's confident predictions within it were promising.

8.5 Discussion

In this work, we developed multiple strong-performing DL pipelines that synthetically generate post-contrast coronal IDEAL T_1 wrist MR images from pre-contrast coronal IDEAL T_1 wrist images, marking steps toward synthetic inflammatory imaging of MSK tissues for conditions such as RA. Reconstruction metrics show reasonably strong performances for UNet and

PatchGAN pipelines without generator decoding path deconvolutions—PatchGAN nRMSEs in the wrist were 7.68 ± 1.41 (6.07 ± 1.22 after registration, mean ± standard deviation (s.d.)) and for the UNet they were 5.38 ± 0.73 (4.36 ± 0.60 after registration, mean ± s.d.). Standard reconstruction metrics—nRMSE, PSNR, and SSIM—showed the UNet to have superior performance across full volumes and within the wrist, but purely in the synovial joints, where a pipeline like this would see the most utility, the PatchGAN outperformed the UNet. These findings provide yet additional evidence to a growing body of literature which suggests that standard reconstruction metrics do not provide great correlation with clinically useful metrics when evaluated in a classical fashion (across an entire tissue) [141,142,185]. This, in addition to a perceptually stronger performance replicating sharper textures (particularly within muscles and bones, but at times in the synovial joints as well), shows the PatchGAN pipeline without deconvolutions to be the strongest tested version and with the most potential for eventual clinical use with further development. Additionally, enhancement maps showed that while both pipelines exhibited similar performance in identifying the location of the top 10% of enhancing pixels, the PatchGAN did a substantially better job in preserving the enhancement magnitudes. These trends particularly held in the muscles and vessels, but also in many synovial joints.

To build clinicians' trust in medical image processing algorithms, experiments such as the proposed occlusion map and uncertainty analyses are vital to address the criticism of deep learning algorithms being "black boxes." These techniques yielded notable insights in the PatchGAN and UNet pipelines: occlusion maps showed that both pipelines focused heavily on intercarpal regions and synovial joints as a basis for generating model predictions. At the same

time, uncertainty maps yielded diverging conclusions: whereas the PatchGAN was most uncertain in background, muscles, and within bones, the UNet pipeline was the most uncertain within the intercarpal joints themselves. Given that intercarpal joints—and more generally synovial joints—are where a synthetic inflammatory imaging algorithm would see maximal utility in RA imaging, it is extremely encouraging that the PatchGAN based much of its predictions on the intercarpal joints and was relatively confident in its predictions. This, combined with the superior reconstruction metrics obtained in synovial joints by the PatchGAN as compared to the UNet, confirms it to be the pipeline with the most potential for clinical utility, and indicates that the combination of a GAN and a focused, ROI-based loss can yield promising results for optimizing image synthesis algorithms. Uncertainty and occlusion map approaches such as those applied in this work are straightforward to implement and can be extended to other deep learning applications such as image synthesis, image segmentation, and image reconstruction. In doing so, they can make the findings of such algorithms easier to interpret while providing valuable insights into how they work. From a clinical perspective, they can not only build trust in algorithm outputs, but also direct a radiologist's attention to uncertain regions in an image that require closer examination.

The exploration of architectural designs also yielded interesting insights. Checkerboarding artifacts have long been reported as a shortcoming of CNNs, and more specifically UNets, with many strategies being proposed to mitigate them [283–285]. Our investigation of UNet pipelines with and without one such mitigating strategy—replacing deconvolutions with interpolation and standard convolutions—showed checkerboarding artifacts to be widespread in larger areas of

relatively homogenous pixel intensity with the standard deconvolutions, but absent with the mitigating strategy implemented. When paired with a PatchGAN discriminator, even a UNet generator with deconvolutions resolved the checkerboarding artifacts in larger homogenous pixel intensity areas, but saw minor checkerboarding emerge at the boundaries between pixel intensities. Checkerboarding artifacts are thus intrinsic to the standard UNet architecture, and among the tasks a discriminator must learn in adversarial training is their removal. When deconvolutions are replaced with interpolation and standard convolutions, the artifact removal responsibility is simplified for a GAN discriminator, in theory allowing the discriminator to focus on more minute differences between real and synthetic images and, thus, possibly producing stronger synthetic images. These lessons can be translated to GAN training strategies in other settings—training schemes may yield stronger results after the thorough inspection of generator architectures to ensure that obvious artifacts are not intrinsic to the network design.

It is clear from our work that larger sample sizes are needed to derive statistical conclusions with more power and to assess algorithm efficacy stratifying by race, RA status, and others. However, this study nonetheless serves as a strong proof-of-concept indicating the potential for DL algorithms to synthesize post-contrast images for inflammatory imaging in MSK applications. Importantly, these algorithms can synthesize images in a negligible amount of time, essentially providing free information for radiologists examining inflammation, even for the many patients for whom contrast MR sequences would otherwise not be prescribed. With additional validation, and through building clinicians' trust in these algorithms, they can allow for safer, more comfortable, and less time-consuming RA diagnosis and treatment through synthetic

imaging. Beyond the proof-of-concept wrist RA post-contrast synthesis, this work can seed new efforts in other MSK applications such as synthetic RA imaging in other joints [286], synthetic screening for sarcoma [287], more thorough investigations associating contrast and non-contrast MRI of Hoffa's fat pad with pain [288], larger cohort studies assessing bone perfusion [289], and safer imaging techniques to diagnose spondylodiscitis [290]. In all these applications, Gd is administered in standard imaging protocols, so similar datasets can be curated and used to train synthetic post-contrast imaging algorithms to reduce and hopefully eliminate the need for Gd administration. Furthermore, validated algorithms could synthesize post-contrast images from existing large datasets such as the Osteoarthritis Initiative (OAI), K2S, and fastMRI+ to allow for large cohort studies to facilitate a better understanding of inflammation [229,291,292].

This study had several limitations. Ideally, there would be a true comparison of algorithm performance in patients with and without RA to ensure strong performance in both, but ethical considerations prevented us from administering Gd to healthy controls. In the absence of this, we used RAMRIS scores to stratify RA patients into subgroups of those with and without imaging findings of RA for a pseudo-control study, but this is not a true control study. Furthermore, the desire to compare algorithm performance in patients with and without imaging findings of RA in a pseudo-control study, combined with the small dataset size, led to some imbalance in demographic characteristics across training, validation, and test datasets. Namely, test set patients had the least severe RA. Additionally, pre-Gd coronal IDEAL images were registered to corresponding post-Gd images in data preprocessing. Radiologist anomaly segmentations were performed only on post-Gd images, so doing so allowed segmentations to

be used in weighting loss functions and assessing model performance in anomalous regions, but this registration step would not be possible at the inference time. There was thus a tradeoff between optimizing trained algorithms for strong performance in synovial joints and using a realistic workflow for eventual clinical utility; the authors viewed the former as more important in a proof-of-concept approach. Lastly, standard imaging protocols would typically use T_1 pre-contrast scans and fat-saturated post-contrast T_1 scans for RA imaging. Our approach used IDEAL scans before and after contrast administration, as these sequences were available in our dataset, but for true clinical translation an algorithm should be trained on these other sequences. The structure of our dataset thus conferred many limitations on our work, but nonetheless, it represents a meaningful first step towards making synthetic inflammatory imaging a larger research focus for the MSK community.

8.6 Conclusions

To the best of the authors' knowledge, our work marks the first concerted effort at leveraging DL for synthetic inflammation imaging for an MSK application. We developed PatchGAN and baseline UNet pipelines that showed strong performance synthesizing post-contrast IDEAL T_1 images from corresponding pre-contrast IDEAL T_1 images, with the PatchGAN pipeline outperforming the UNet in synovial joints, generating more accurate and confident predictions where a model would have the most utility. The PatchGAN also showed magnitudes of signal enhancement that more closely match that of ground truth images and retained sharp textures in synthetic images. As such, the PatchGAN model was particularly promising in synthesizing post-contrast inflammatory images, and with further development, it could reduce or eliminate

the need for Gadolinium administration in treating patients with RA. There are numerous future directions for research: (1) more sophisticated GANs such as CycleGAN can be implemented to improve the sharpness in reconstructed images; (2) generator architectures that learn registration transforms and predict images can also be investigated, eliminating the need for the registering of pre-Gd images to post-Gd images, which would not be possible at inference time in the clinic; (3) investigating other loss functions, such as other types of GAN distances; and (4) assessing model robustness by inferring from conventional wrist coronal T_1 scans to evaluate predicted post-contrast scans on conventionally used clinical sequences in inflammatory imaging. For substantial progress, however, the MSK field will require concerted efforts to curate larger datasets for inflammatory RA conditions that will allow for more statistically powerful conclusions, more complicated models, and comparisons across population subgroups. Our hope is that the promise of our results can motivate efforts to do so.

Chapter 9 - Deep Learning Predicts Total Knee Replacement from MR Images

The following has been reformatted and reproduced with full permission from the publisher. It appeared in *Nature Scientific Reports* as:

Tolpadi, A.A., Lee, J.J., Pedoia, V. *et al.* Deep Learning Predicts Total Knee Replacement from Magnetic Resonance Images. *Sci Rep* **10**, 6371 (2020). https://doi.org/10.1038/s41598-020-63395-9

9.1 Abstract

Knee Osteoarthritis (OA) is a common musculoskeletal disorder in the United States. When diagnosed at early stages, lifestyle interventions such as exercise and weight loss can slow OA progression, but at later stages, only an invasive option is available: total knee replacement (TKR). Though a generally successful procedure, only 2/3 of patients who undergo the procedure report their knees feeling “normal” post-operation, and complications can arise that require revision. This necessitates a model to identify a population at higher risk of TKR, particularly at less advanced stages of OA, such that appropriate treatments can be implemented that slow OA progression and delay TKR. Here, we present a deep learning pipeline that leverages MRI images and clinical and demographic information to predict TKR with AUC 0.834 ± 0.036 ($p < 0.05$). Most notably, the pipeline predicts TKR with AUC 0.943 ± 0.057 ($p < 0.05$) for patients without OA. Furthermore, we develop occlusion maps for case-control pairs in test data and compare regions used by the model in both, thereby identifying TKR imaging biomarkers. As such, this work takes strides towards a pipeline with clinical utility,

and the biomarkers identified further our understanding of OA progression and eventual TKR onset.

9.2 Introduction

Knee Osteoarthritis (OA) is one of the most common musculoskeletal disorders in the United States, with estimates of its incidence rate ranging from 14 to 30 million [5,6]. Annual arthritis-related medical expenditures are nearly $140 million, and hip and knee OA together are the 11th highest contributor to global disability [293,294]. The propensity of knee OA to induce eventual disability can be attributed to structural changes in the joint that characterize the disease, as well as symptoms that can include inflammation, debilitating pain, and functional limitations [295,296]. Progression of the full-joint disease is typically assessed using the Kellgren-Lawrence (KL) scale, a 0-4 scale in which a higher score is associated with narrowing of the tibiofemoral joint (TFJ) space and other radiographic changes, and thus, a more advanced stage of knee OA [165]. When diagnosed at early stages (KL = 0, 1), knee OA can be managed through nonsurgical treatment options, including exercise and/or weight loss, oral medications such as acetaminophen or NSAIDs, or intra-articular injections such as corticosteroids and hyaluronic acid, all of which have varying degrees of success in reducing pain [8]. At late stages (KL = 4), however, no noninvasive option exists [9]; here, the only option is total knee replacement (TKR).

TKR is an elective procedure in which the knee joint is resurfaced with a metal or plastic implant intended to restore function, provide pain relief, and improve quality of life [297]. In the United States, estimates of TKR incidence lie at 400,000 each year, a figure expected to grow 143% by

2050 even through conservative projections [298]. While TKR is considered one of the most effective procedures in orthopedic surgery, electing for it is far from straightforward: noninvasive alternatives such as weight loss, physical therapy, and NSAIDs are first exhausted. If unsuccessful, a patient will undergo a thorough examination of clinical history and comprehensive imaging of the joint to determine if a TKR is feasible, and if so, the desired implant design and size [299,300]. The procedure is also imperfect: only 66% of patients report their knees feeling "normal," and 33% of patients report some degree of pain post-implant [301]. Furthermore, the implant can fail under some circumstances: periprosthetic joint infection and wound complications can be observed, and implant instability can occur due to aseptic loosening, malpositioning of the implant, and wear of joint components [302,303]. It is thus much preferable to prolong the good health of the knee, particularly in patients where OA has not advanced to the most severe stages, thereby delaying TKR as long as possible. This necessitates a model to identify patients at higher risk of TKR such that appropriate treatment options can be pursued.

Given the multitude of factors on which a decision to pursue TKR is made, devising a model to predict if the invasive intervention will be necessary is a difficult task, but with obvious utility. For a patient in earlier stages of OA, a model predicting the patient to be at risk of TKR can be the impetus for a more aggressive nonsurgical treatment. Meanwhile, for a late-stage OA patient, a model predicting them to undergo TKR may facilitate a doctor and patient opting for the treatment earlier than they otherwise would, thereby reducing time spent pursuing nonsurgical alternatives with minimal probability of success while dealing with serious pain.

Beyond this, if the model were to draw from medical images of the knee, it could identify anatomic regions most correlated with a TKR prediction. To this point, few studies have been conducted in this space, and those that have primarily investigate the importance of cartilage volume loss, subchondral bone defects, and bone marrow lesions [304–306]. An identification of more such biomarkers for TKR, however, could greatly improve understanding of both OA and TKR, and ultimately guide treatment strategies.

Predictive modeling of TKR, however, has a limited history, particularly with models that use medical images. A few studies have leveraged random forest regression, Cochran-Armitage tests for trend, and t-tests to identify demographic, general health, and physical examination measurements that most strongly correlate with TKR or total joint arthroplasty (TJA) [307,308]. Others have taken these efforts further, using techniques such as multiple regression and multivariate risk prediction models to predict TKR outright [309,310]. To our knowledge, only one group has developed a predictive model of TKR that accepts image inputs, attaining performance that surpasses that of models using only clinical and demographic information [311]. Notably, past TKR predictive models largely measure performance by evaluating the area under the receiver operating characteristic (ROC) curve, which plots true positive rate against false positive rate [312]. However, in most datasets used in this space, the number of patients who eventually undergo TKR is dramatically higher among those who have advanced OA as opposed to those with no or moderate OA. Consequently, this performance metric (AUC), while effectively capturing a model's combination of sensitivity and specificity, can be inflated for TKR prediction by indiscriminately predicting patients without OA not to undergo TKR, while more

accurately predicting patients with severe OA to undergo TKR, the latter of which is easier. As a result, while past works have made clear progress in predicting TKR, none have overcome datasets imbalanced with respect to OA severity to report sensitive and specific prediction at these early stages, where a model would have the most utility.

One technique that has shown promise in delivering such performance is deep learning (DL). DL, especially convolutional neural networks (CNNs), has made strides in image classification tasks, attaining performances on the popular ImageNet classification challenge that approach or surpass human performance [313–315]. DL shines when afforded large datasets, as its automated feature extraction allows one to solve problems too complex for conventional approaches[29]. Given the complex prognostic features in TKR recommendation, CNNs become more promising for TKR prediction. In the past, DL had seen limited utility in OA and TKR prediction due to the large dataset requirement for efficacy; that limitation has been somewhat mitigated by the curation of large-sized cohort studies such as the Osteoarthritis Initiative (OAI) [229]. Consequently, DL has recently been applied for knee OA classification and progression prediction [9,316–318]. The success of these works further suggests the feasibility of leveraging DL to predict TKR.

In this study, we formulate a DL-based pipeline that incorporates knee joint images in addition to clinical and demographic information to predict the onset of TKR (Figure 9.1). We demonstrate that the pipeline's predictions using solely Magnetic Resonance Imaging (MRI) images matches that of past work, while the integration of MRI image-based predictions with

non-imaging variables facilitates TKR prediction with especially high sensitivity and specificity for patients without radiographic OA. Furthermore, we show the increase in pipeline performance when using 3D MRI images as opposed to 2D radiographs, suggesting MRI may have a role in TKR risk screening despite higher costs and more limited availability. And finally, we leverage occlusion maps to conduct a thorough analysis of tissues that most significantly affect the output model metric associated with TKR prediction confidence, thereby identifying a set of imaging biomarkers for eventual TKR onset.

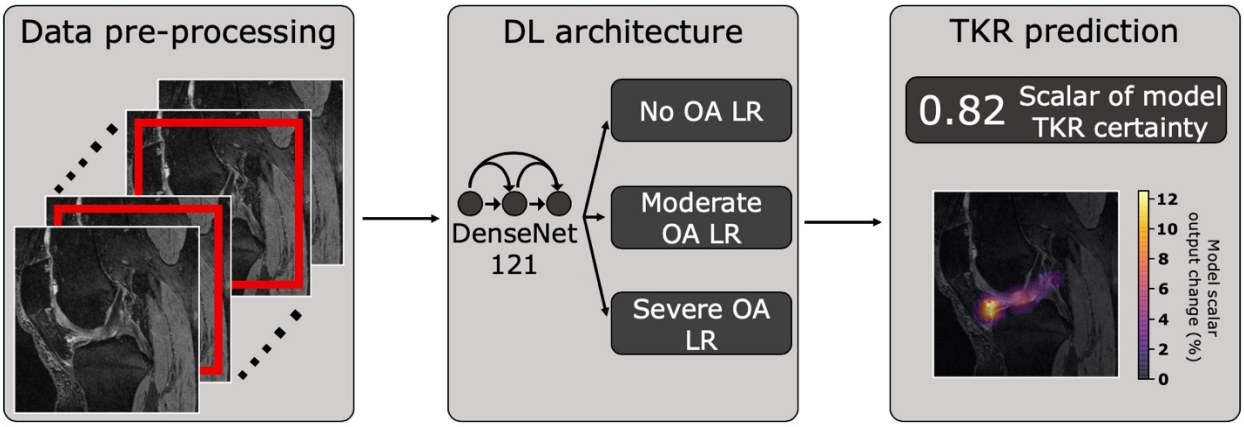


Figure 9.1 Pipeline predicting if patient will undergo TKR within 5 years from MRI/X-ray images and non-imaging variables. MRI and X-ray images are center-cropped and cropped to a region centered around the joint, respectively, and normalized. DenseNet-121 is pretrained to predict OA and fine-tuned to predict TKR. Image-based predictions and clinical information are fed to a logistic regression (LR) ensemble based on OA severity. Each ensemble, whose hyperparameters were optimized for Youden's index in a hyperparameter search, averages predictions of LR models in its OA severity for final TKR prediction. Pipeline is subsequently analyzed through occlusion map analysis to identify imaging biomarkers of TKR.

9.3 Novelty

This work reports a methodology and results that are novel in the following manners:

1. This model is the first to apply a 3-dimensional DenseNet CNN for prediction of TKR from MRI.

2. The TKR prediction model is evaluated for patients stratified by OA severity, which has not been reported in previous studies.
3. With the aim of improving model interpretability and clinical utility, we report the first comprehensive, case-control study to identify imaging biomarkers for TKR.

9.4 Methods

9.4.1 Data

Data was acquired from a prospective observational study conducted by OAI. The dataset followed 4,796 patients and acquired images including 2D posteroanterior radiographs and 3D Sagittal Double Echo Steady-State (DESS) MRI images over the course of 10 years. Details of data collection and study design have been previously reported [229]. The OAI study protocol was approved by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and is registered on ClinicalTrials.gov as “Osteoarthritis Initiative (OAI): A Knee Health Study”, NCT#00080171. The study was carried out in accordance with all pertinent guidelines and regulations, and written and informed consent was obtained from participants prior to each clinical visit in the study.

Both posteroanterior radiographs and DESS MRI images were evaluated as data sources for TKR prediction models. Patients for whom KL grade was not recorded at any point in the longitudinal study were excluded. To homogenize datasets, radiograph and MRI images were only taken from patients and time points at which both were available (n = 35,482). We labeled entries as cases if the patient underwent a first TKR within 5 years of the given time point (n = 1,043). We

labeled entries as controls if patients did not undergo a TKR or eventually underwent one but the time to it was longer than 5 years (n = 34,439). Contralateral TKRs were not considered.

The radiographs and MRI images were preprocessed for training and model evaluation. Radiographs were cropped to a 500 × 500 region centered around the knee joint. Briefly, 2D cross-correlation template matching was used to identify a 500 × 500 bounding box centered around the knee joint in 450 joints, and these cases were used to train a U-Net architecture that identified this region for all posteroanterior radiographs from the OAI study [318]. DESS MRIs were center-cropped to a 120 × 320 × 320 region, after which both sets of cropped images were normalized. Normalized MRI pixel values were then rounded to nearest integers, compressing the MRI image to 14 possible pixel values. This rounding approach was initially tested as a strategy to accelerate training of a 3D CNN, given the large imaging volumes and large dataset on which it was being trained, believing the approach could suppress information extraneous to eventual TKR. Empirically, this approach yielded superior validation performance to leaving pixel values unrounded, so it was utilized. Examples of the results of this compression strategy are in Supp. Fig. C.1.

Non-imaging variables were screened for among studies and reviews detailing risk factors for knee OA progression and TKR onset [308–310,319–322]. Variables such as KL grade known to be deducible directly from MRI images and radiographs were not considered. From these studies, 40 non-imaging variables of interest were identified (Supp. Table C.1). The OAI database was then parsed for corresponding variables, and these corresponding variables were added as

potential non-imaging variables for our study, yielding 44 potential non-imaging variables. In some cases, multiple OAI metrics corresponded to non-imaging variables of interest, causing the number of OAI non-imaging variables to exceed what was identified from literature. Missing data points were imputed with k-nearest neighbors. These potential variables were used to train a random forest with 100 trees to predict onset of TKR within 5 years, and the minimum depth at which each feature was used across all trees in the forest was identified. Features whose minimum depth was below the average minimum depth of all features were preserved as non-imaging variables [323]. This yielded 27 non-imaging variables that are displayed in Table 9.1.

Table 9.1 List of non-imaging variables fed into logistic regression models to make predictions of whether a patient would undergo TKR within 5 years. Abbreviations used: Body Mass Index (BMI), Nonsteroidal Anti-inflammatory drugs (NSAIDS), Blood Pressure (BP), Physical Activity Scale for the Elderly (PASE), Knee Injury and Osteoarthritis Outcome Score (KOOS), Quality of Life (QOL), Western Ontario and McMaster Universities Arthritis Index (WOMAC), Short Form 12 (SF-12).

Non-imaging variables used to augment image-based predictions	
Age	Comorbidity score
BMI	Injections to treat arthritis in previous 6 months
Education	Seen physician for arthritis in previous year
Ethnicity	Knee valgus negative alignment (degrees)
Income	Isometric leg strength
NSAID usage	Back pain in previous 30 days
Analgesics usage	Difficulty squatting in previous 7 days
Systolic BP	Difficulty kneeling in previous 7 days
Considering TKR	Baseline frequent knee pain status
PASE	Previous knee injury that limited walking
KOOS QOL	0-10 global rating assessing effect of knee pain
KOOS pain	SF-12 physical component score
WOMAC pain	SF-12 mental component score
WOMAC disability	

The data were then split into training, validation, and test with a 65%/20%/15% split, ensuring entries of any patient were only in one of the three datasets to prevent data leakage. Within the

training set, imbalance between TKR and non-TKR cases was addressed with data augmentation, drawing bootstrap samples from the rare class with replacement [324]. A summary of the data prior to augmentation is provided in Table 9.2, detailing the number of cases and controls while showing descriptive statistics regarding demographics in each of the three datasets.

Table 9.2 Data used to train 3D DESS MRI and 2D radiograph architectures. After exclusion criteria were applied, 35,482 qualifying entries were found in the OAI dataset across 4,790 unique patients, all of which were split into training, validation, and test sets as displayed in table. To prevent data leakage, all entries from any given patient were only allowed to be in one of the three sets. S.d. is reported for age, BMI, and KOOS pain score within the table.

		Training		Validation		Test	
		Control	**Case**	**Control**	**Case**	**Control**	**Case**
	Age	62.5 ± 9.15	66.3 ± 8.38	62.4 ± 9.21	66.1 ± 8.76	62.8 ± 9.55	66.4 ± 7.78
	BMI	28.3 ± 4.75	29.6 ± 4.79	28.4 ± 4.64	29.8 ± 4.61	28.4 ± 4.81	29.9 ± 3.96
	KOOS Pain	87.2 ± 16.2	67.2 ± 19.6	87.6 ± 15.9	66.2 ± 19.1	87.4 ± 16.5	68.7 ± 20.6
	Male	9,708	291	2,876	70	2,126	59
	Female	12,731	396	4,035	134	2,963	93
OA	**None**	12,721	41	4,118	13	2,892	12
	Moderate	8,950	357	2,611	93	2,056	83
	Severe	768	289	182	98	141	57
	Total Entries	23,126	0	7,115	0	5,241	0
	Unique Patients	3,114	0	957	0	719	0

9.4.2 Pipeline Architecture

The DL-based pipeline is based on a DenseNet-121 with the following parameters: 16 filters in initial layer, growth rate of 32, pooling block configuration of [6, 12, 24, 16], 4 bottleneck layers, 2 classes. The same architecture was used for the radiograph and MRI pipelines, but for the MRI pipeline, we modified the convolutional layers, batch normalization layers, pooling layers, and

leaky rectified linear unit (ReLU) layers to allow for 3D image input [325]. The network yielded a scalar reflecting certainty of TKR within 5 years, which was added to the non-imaging variables. The 28 resulting variables were fed into one of three sets of Logistic Regression (LR) ensembles, with each ensemble optimized to maximize sensitivity and specificity in cases of no (KL = 0, 1), moderate (KL = 2, 3), and severe OA (KL = 4). Based on the KL grade of a sample, it was fed into an LR ensemble, yielding a prediction as to whether the patient will undergo a TKR within 5 years.

9.4.3 Training

A DenseNet-121 was initially pretrained to predict knee OA using the entire training set, assessing cross-entropy loss and accuracy on the validation set after completion of each epoch. The pre-train was stopped when validation loss began to increase. The pretrained model was subsequently fine-tuned to predict TKR. We utilized a random search to determine optimal learning rate, dropout rate, weights of the cross-entropy loss function, and number of layers to freeze during fine-tuning. The search was carried out for 25 iterations, after which a set of parameters were selected that yielded the best combination of accuracy, sensitivity, and specificity on the validation set. Due to computational intensity, the hyperparameter search was not conducted on the entire dataset: for the 2D DenseNet-121, 10% of training and validation sets were used, whereas for the 3D DenseNet-121, 2.5% of both were used. After the search, the model fine-tuned using the subset of the training set was further fine-tuned on the entire training set using optimal parameters until validation loss began to increase. The test set was

held out during training and predictions for it evaluated just once after fine-tuning, which marked the end of model optimization.

9.4.4 Integration of Imaging and Non-Imaging Data

Random forest regression, support vector machine, neural network, and LR architectures were assessed for efficacy of integrating imaging and non-imaging predictions, with LR providing best results on validation data. The LR architecture was thus used: all 28 imaging and non-imaging models were fed into an LR model, the optimal parameters of which were also identified through a random search. The search was conducted for 100 iterations, seeking to optimize the cross-entropy loss function weights afforded to both classes. For the cases of no, moderate, and severe OA, ideal parameters were identified by selecting those that maximized Youden's index within each OA classification in the search [326]. Predictions of the best few models in each classification were averaged to yield final TKR predictions. The number of predictions averaged in each classification was selected by finding a value that optimized validation accuracy, AUC, and Youden's index. The resulting LR models were ensembled and run on test data just once. Confidence intervals of accuracy, sensitivity, and specificity for each OA severity were obtained by bootstrapping, sampling 100% of test data with replacement (B = 100). Confidence intervals for AUC were calculated in the same manner. Results are reported on 3 versions of each model: the sole DenseNet-121 output (image only), output of a single LR model trained to predict TKR using solely the 27 non-imaging variables while not weighting the loss function class weights (non-imaging info. only), and output of the LR ensemble with image predictions (integrated model).

9.4.5 Statistical Analysis

The accuracies of X-ray and MRI pipeline performances within each OA classification and overall were compared using McNemar's test [327,328]. This test was appropriate because it specifically tests for differences in a dichotomous variable in matched groups. In our case, the variable was correct TKR prediction and the groups were the X-ray and MRI pipelines. Initially, the McNemar test statistic was modeled with a chi-squared distribution to test for significant differences between the pipelines, and if one existed, a binomial distribution was used to interrogate which pipeline yielded the significantly higher performance. All tests were carried out at $\alpha = 0.05$.

Relative sensitivity and specificity of the X-ray and MRI pipelines were assessed by comparing their AUCs within each OA classification and overall. This test is appropriate because the ROC curve plots true positive rate (sensitivity) against false positive rate (1 – specificity); consequently, the closer the AUC is to 1, the better the combination of sensitivity and specificity. 100% of test data was sampled with replacement (B = 100), and for each corresponding pair of X-ray and MRI pipelines (matched by OA classification and use of images only or both image and non-image information), AUCs were calculated. To test if one outperformed the other, differences in AUCs were calculated at each iteration, and the mean and standard deviation of the differences used to conduct a student's t-test with 99 degrees of freedom. This test is applicable on each matched pair of X-ray and MRI pipelines due to the number of iterations for which test data was sampled, allowing the central limit theorem to

apply. For confidence intervals, mean and standard deviation of AUCs of individual models were calculated and used to report 95% intervals.

9.4.6 Imaging Biomarker Identification

For all 124 true positives in the test data for the integrated MRI pipeline, corresponding controls were identified by randomly sampling from test data true negatives, keeping OA status distributions identical and using a student's t-test with 123 degrees of freedom to ensure no significant difference in KOOS pain scores across cases and corresponding controls at $\alpha = 0.05$. Occlusion maps were generated for all cases and controls using voxel size of $12 \times 32 \times 32$ and stride of 12. For each pixel, the value displayed represented the magnitude of change in the scalar pipeline output resulting when that pixel was occluded, averaged across all occlusions in which that pixel existed. Pixels for which scalar pipeline output change lied in the top 5% were designated as "hotspots." Anatomic regions of these hotspots were identified and odds ratios (OR) calculated to interrogate possible imaging biomarkers of TKR. 95% OR confidence intervals were calculated for each anatomic region investigated in this analysis using Cornfield's method, as this method performs well with relatively small sample sizes [329]. P values of ORs were calculated using a two-tailed Fisher's exact test [330]. Tissues where p values fell below the significance level of $\alpha = 0.05$ and in which 95% OR confidence intervals did not include 1 were deemed significant. These test selections were appropriate, as they allowed for direct comparison of the frequencies at which several tissues were hotspots across cases and controls, and as such, identified significant tissues with regards to TKR onset.

9.5 Results

9.5.1 OA Pretrain Utility in TKR Prediction

To test information learned from the OA pretrain, pretrained models themselves were used to predict TKR, with results depicted in Table 9.3. Predictably, the radiograph OA pretrain model had poor sensitivity for patients without OA, and poor specificity in moderate and severe cases of OA. While the MRI OA pretrain model expectedly yielded more balanced sensitivity and specificity across all OA stages, it too left room for improvement, particularly in sensitivity at no OA and specificity at severe OA. This confirmed the pretrain provided useful information to both architectures but fine-tuning and integration of non-imaging variables were necessary to attain desired TKR prediction performance.

Table 9.3 Performance in TKR prediction of OA pretrained models for radiographs and MRI, stratified by severity of OA. Pretraining strategy yields useful information to both models, but performance at no OA in particular leaves room for improvement, justifying subsequent model fine-tuning. Standard errors used to calculate confidence intervals.

OA status	Model type	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Non-TKR cases	TKR cases
None	Radiograph	92.1 ± 0.083	25.2 ± 2.16	92.4 ± 0.081	2,892	12
	MRI	94.3 ± 0.070	48.7 ± 2.48	94.4 ± 0.070		
Moderate	Radiograph	29.3 ± 0.151	93.8 ± 0.439	26.7 ± 0.156	2,056	83
	MRI	65.4 ± 0.154	65.5 ± 0.848	65.4 ± 0.158		
Severe	Radiograph	29.7 ± 0.488	100.0 ± 0.000	1.4 ± 0.180	141	57
	MRI	33.4 ± 0.523	82.2 ± 0.824	14.0 ± 0.441		
All	Radiograph	64.2 ± 0.124	90.7 ± 0.378	63.4 ± 0.126	5,089	152
	MRI	80.2 ± 0.079	70.4 ± 0.595	80.5 ± 0.082		

9.5.2 X-Ray Pipeline Optimization and Performance

For the X-Ray model, hyperparameter tuning steps found the following to yield the best combination of validation accuracy, sensitivity, and specificity: learning rate of 3.981×10^{-6}, TKR

class weight in cross-entropy loss function of 0.927 and non-TKR class weight of 0.073, dropout rate of 0.375, and only the last 2 layers fine-tuned after OA pretrain.

A radiograph model was fine-tuned to predict TKR with these parameters, and its predictions fed into an LR ensemble. Averaging predictions of the best 5 LR models found through random search in the 3 OA categories yielded best validation performance, so this ensemble was used on the test set. Test accuracy, sensitivity, and specificity are provided in Table 9.4, and ROC curves of all three versions of this pipeline are found in Figure 9.2. AUCs are as follows: 0.848 **±** 0.039 (image only), 0.868 **±** 0.028 (non-imaging info. only), 0.890 **±** 0.021 (integrated model). Furthermore, AUCs for the image-only and combined versions of the pipeline at no OA are as follows: 0.514 **±** 0.087 (image only); 0.799 **±** 0.055 (integrated model). At moderate OA: 0.788 **±** 0.025 (image only); 0.865 **±** 0.016 (integrated model). At severe OA: 0.552 **±** 0.040 (image only); 0.641 **±** 0.044 (integrated model). All AUC intervals are calculated using standard deviation (s.d.), $p < 0.05$.

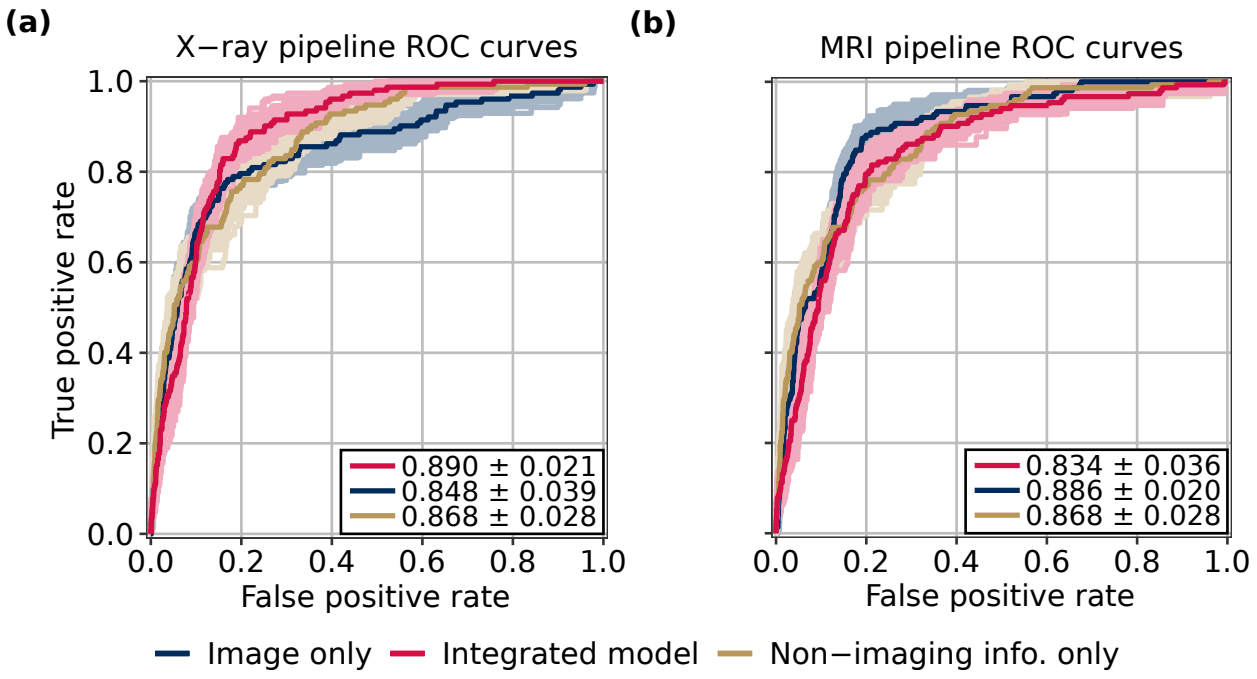


Figure 9.2 ROC curves for X-ray and MRI architectures on test data. X-ray pipeline ROC curves are shown in (a), with AUCs as follows, $p < 0.05$: 0.848 ± 0.039 (image only), 0.868 ± 0.028 (non-imaging info. only), 0.890 ± 0.021 (integrated model). MRI pipeline ROC curves are shown in (b), with AUCs as follows, $p < 0.05$: 0.886 ± 0.020 (image only), 0.868 ± 0.028 (non-imaging info. only), 0.834 ± 0.036 (integrated model). Standard deviations used to calculate confidence intervals. ROC curves with AUCs within 1 standard deviation of the mean for each model type during bootstrapping are also shown on plots.

Table 9.4 Performance of X-ray and MRI architectures on test data. While integrated X-ray pipeline delivers higher accuracy than integrated MRI pipeline, integrated MRI pipeline yields improved sensitivity over integrated X-ray pipeline across all stages of OA, markedly so at no OA. Standard errors used to calculate confidence intervals.

OA status	Image source	Model type	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Non-TKR cases	TKR cases
None	X-ray	Non-imaging info. only	89.1 ± 0.139	49.7 ± 2.80	89.3 ± 0.140	2,892	12
		Image only	95.0 ± 0.089	7.8 ± 1.64	95.4 ± 0.089		
		Integrated model	95.4 ± 0.081	8.6 ± 1.95	95.8 ± 0.077		
	MRI	Non-imaging info. only	89.1 ± 0.139	49.7 ± 2.80	89.3 ± 0.140		
		Image only	95.2 ± 0.088	66.9 ± 3.23	95.3 ± 0.089		
		Integrated model	82.4 ± 0.171	92.2 ± 1.68	82.4 ± 0.173		

OA status	Image source	Model type	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Non-TKR cases	TKR cases
Moderate	X-ray	Non-imaging info. only	72.9 ± 0.208	70.0 ± 1.16	73.0 ± 0.212	2,056	83
		Image only	79.9 ± 0.196	66.7 ± 1.23	80.4 ± 0.195		
		Integrated model	81.4 ± 0.178	76.0 ± 1.12	81.6 ± 0.179		
	MRI	Non-imaging info. only	72.9 ± 0.208	70.0 ± 1.16	73.0 ± 0.212		
		Image only	68.8 ± 0.225	78.3 ± 0.952	68.4 ± 0.227		
		Integrated model	74.9 ± 0.216	78.9 ± 0.974	74.7 ± 0.228		
Severe	X-ray	Non-imaging info. only	51.3 ± 0.744	89.4 ± 0.864	35.8 ± 0.925	141	57
		Image only	32.1 ± 0.714	94.5 ± 0.735	7.2 ± 0.467		
		Integrated model	60.5 ± 0.775	64.0 ± 1.57	59.0 ± 0.959		
	MRI	Non-imaging info. only	51.3 ± 0.744	89.4 ± 0.864	35.8 ± 0.925		
		Image only	34.6 ± 0.775	98.3 ± 0.390	9.2 ± 0.632		
		Integrated model	59.6 ± 0.770	84.0 ± 1.03	49.6 ± 1.04		
All	X-ray	Non-imaging info. only	81.1 ± 0.118	75.6 ± 0.776	81.2 ± 0.122	5,089	152
		Image only	86.4 ± 0.095	72.5 ± 0.864	86.9 ± 0.095		
		Integrated model	88.4 ± 0.094	66.3 ± 0.924	89.1 ± 0.090		
	MRI	Non-imaging info. only	81.1 ± 0.118	75.6 ± 0.776	81.2 ± 0.122		
		Image only	82.1 ± 0.118	84.9 ± 0.636	82.1 ± 0.119		
		Integrated model	78.5 ± 0.134	81.8 ± 0.643	78.4 ± 0.138		

9.5.3 MRI Pipeline Optimization and Performance

Similarly, a hyperparameter search was carried out for the MRI pipeline to optimize parameters for eventual fine-tuning. The following hyperparameters were found optimal: learning rate of 1.906×10^{-2}, TKR class cross-entropy weight of 0.902 and non-TKR class weight of 0.098, dropout rate of 0.329, only last layer of model fine-tuned after OA pretrain.

An MRI-based model was fine-tuned from these parameters. The resulting predictions were fed into an LR ensemble, where averaging predictions of the best 4 models in each OA category optimized validation performance. Performance of the resulting architecture on test data is reported in the same manner as the radiograph pipeline, in Table 9.4 and Figure 9.2. AUCs are as follows: 0.886 **±** 0.020 (image only), 0.868 **±** 0.028 (non-imaging info. only), 0.834 **±** 0.036 (integrated model). AUCs for the image-only and combined pipeline versions at no OA are as follows: 0.897 **±** 0.039 (image only); 0.943 **±** 0.029 (integrated model). At moderate OA: 0.764 **±** 0.020 (image only); 0.830 **±** 0.024 (integrated model). At severe OA: 0.560 **±** 0.042 (image only); 0.726 **±** 0.038 (integrated model). Again, all AUC intervals are calculated using s.d., $p < 0.05$.

9.5.4 Comparison of MRI and Radiograph Pipeline Performances

A comparison of overall AUCs attained by the integrated MRI and X-ray pipelines across OA grades and overall shows that at no OA and severe OA, the MRI pipeline outperformed the X-ray pipeline (No OA, B = 100: $p = 3.04 \times 10^{-2}$; Moderate OA, B = 100: $p = 9.55 \times 10^{-1}$; Severe OA, B = 100: $p = 4.57 \times 10^{-2}$; Overall, B = 100: $p = 9.94 \times 10^{-1}$). The MRI pipeline thus has a superior combination of sensitivity and specificity than does the X-ray pipeline for patients without OA and those with severe OA. The AUCs obtained by the image-only pipelines also were compared, and showed the MRI pipeline to outperform the X-ray pipeline for patients without OA and overall (No OA, B = 100: $p = 6.10 \times 10^{-5}$; Moderate OA, B = 100: $p = 7.58 \times 10^{-1}$; Severe OA, B = 100: $p = 4.37 \times 10^{-1}$; Overall, B = 100: $p = 1.16 \times 10^{-2}$). These results follow intuition: while radiographic imaging is primarily capable of illuminating bones in the joint, MRI can visualize soft tissues such as cartilage, muscle, and meniscus [143,331]. It follows that an MRI model will

exhibit a better combination of sensitivity and specificity, especially in early OA stages at which few radiographic changes in the knee have occurred. ROC curves for pipeline versions and OA classifications in which the MRI architecture yielded a significantly better AUC than its X-ray counterpart are shown in Figure 9.3.

McNemar's test assessed relative accuracies of these pipelines. There was a statistically significant difference between the accuracies of the integrated X-ray and MRI pipelines for patients at no OA, moderate OA, and overall (No OA, n = 537: $p = 1.65 \times 10^{-59}$; Moderate OA, n = 521: $p = 1.13 \times 10^{-9}$; Severe OA, n = 47: $p = 8.84 \times 10^{-1}$; Overall, n = 1,105: $p = 1.52 \times 10^{-54}$), and in each of those 3 statistically significant cases, the X-ray pipeline outperformed the MRI pipeline (No OA, n= 537: $p = 1.11 \times 10^{-16}$; Moderate OA, n = 521: $p = 5.97 \times 10^{-10}$; Overall, n = 1,105: $p = 1.11 \times 10^{-16}$). In interpreting these tests and the AUC tests holistically, it is evident that the X-ray pipeline is able to attain superior accuracy in several OA classifications by compromising on its combination of sensitivity and specificity. This is further supported by the accuracies and sensitivities reported for the respective pipelines in Table 9.4, which show that while the X-ray pipeline is more accurate than its MRI counterpart at every OA classification, the opposite is true for sensitivity—drastically so for patients without OA. In the clinic, where sensitivity as to whether a patient is at risk of eventual TKR is paramount, these results would show the MRI pipeline to be the more useful model.

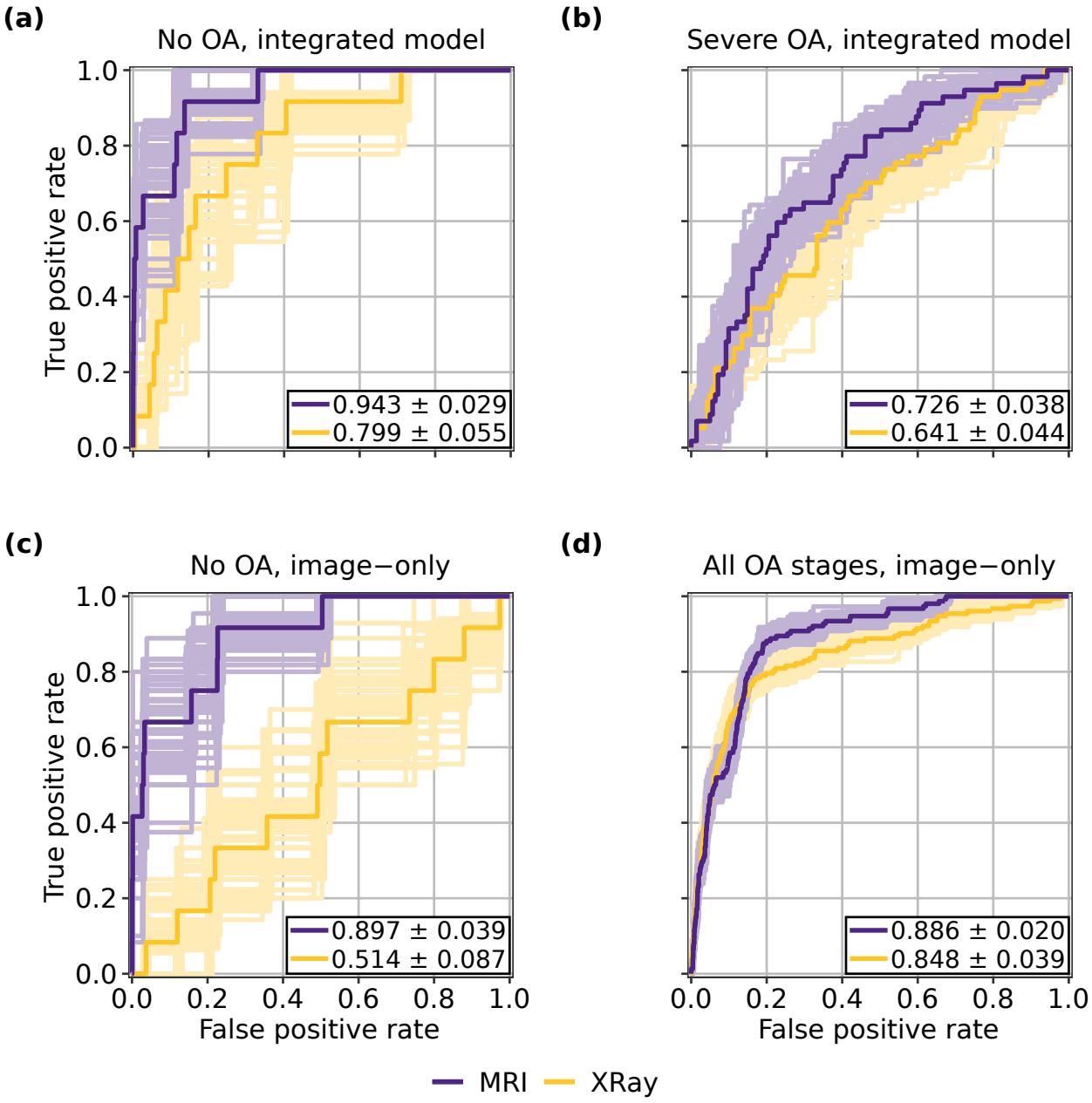


Figure 9.3 ROC curves for MRI and X-ray pipelines at selected OA classifications and pipeline versions in which MRI performance was significantly better than that of X-ray. MRI pipeline outperforms X-ray pipeline at no OA for both image-only and integrated models, as seen in (a) and (c). As shown in (b), integrated MRI pipeline also outperformed integrated X-ray pipeline for patients with severe OA, while (d) shows image-only MRI pipeline outperformed image-only X-ray pipeline across all OA stages. AUCs are displayed in the figure with $p < 0.05$. Standard deviations used to calculate confidence intervals. ROC curves with AUCs within 1 standard deviation of the mean for each pipeline version during bootstrapping are also shown on plots.

It is also worthy to note the improvement in performance that occurs for patients without OA when imaging predictions are added to non-imaging variables in both pipelines. In the X-ray pipeline, the model's AUC increased from 0.514 ± 0.087 to 0.799 ± 0.055 when non-imaging variables were added to the radiographs, a sizeable increase when compared to the MRI pipeline performance, which saw AUC increase from 0.897 ± 0.039 to 0.943 ± 0.029 ($p < 0.05$ for all). This demonstrates that non-imaging variables such as various pain scales seem to add critical information to the X-ray pipeline, while the same information is less important in the MRI pipeline.

9.5.5 Biomarker Identification and Analysis

Of the 152 patients in test data who underwent a TKR, 124 were detected by the MRI pipeline. Occlusion maps were generated for these cases and their corresponding true negative controls, an example of which is shown in Figure 9.4. Tissues and their hotspot percentages across these true positives and corresponding true negative controls can be found in Supp. Table C.2 and Supp. Table C.3 online, respectively. ORs, 95% confidence intervals, and associated p values for each tissue can be found in Table 9.5.

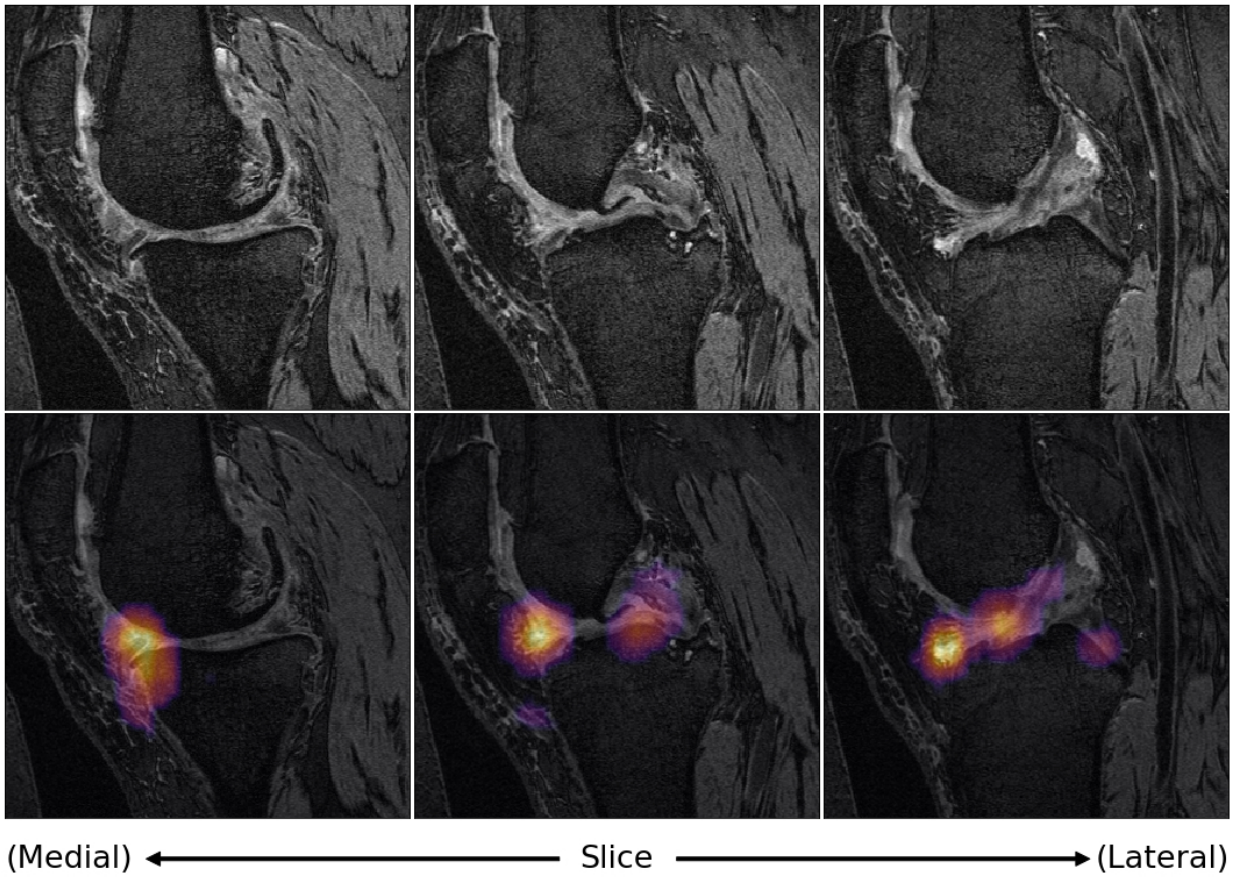


Figure 9.4 Slices of occlusion map of true positive detected by MRI pipeline, overlaid on corresponding slices of DESS MRI. Such maps were generated and analyzed for all 124 true positives and corresponding true negative controls of the integrated MRI pipeline.

Three tissues saw ORs and 95% confidence intervals that lied above 1 and p values below α = 0.05: the medial patellar retinaculum, gastrocnemius tendon, and plantaris muscle. Thus, we conclude there is a substantial and statistically significant difference in the risk of TKR within 5 years when these tissues are identified as hotspots by the pipeline. From the ORs, we see that the risk of TKR increases when any of the three are identified as hotspots: for the medial patellar retinaculum, the risk is 1.98 times higher with a 95% confidence interval from 1.02 to 3.99; for the gastrocnemius tendon, it is 2.97 times higher with a 95% confidence interval from 1.12 to 10.0; and for the plantaris muscle, it is 2.84 times higher with a 95% confidence interval

from 1.47 to 5.82. As such, these results provide evidence that all are imaging biomarkers of TKR.

Table 9.5 Summary of occlusion map analysis comparing frequencies with which selected knee joint tissues were indicated as hotspots in analysis. Hotspots were defined as pixels that, when occluded, were among the top 5% of all pixels in change of pipeline TKR prediction output metric when occluded. Odds ratios, 95% confidence intervals calculated using Cornfield's method, and p values calculated using Fisher's exact test are displayed. Tissues that were significant at $\alpha = 0.05$ are designated with a *. N value for all tests was n = 124.

Tissue type	Tissue	OR (95% CI, n = 124)	P value (n = 124)
Cartilage	TFJ medial*	0.05 (0.00 - 0.48)	3.36×10^{-3}
	TFJ lateral*	0.03 (0.00 - 0.25)	3.89×10^{-5}
	PFJ	1.03 (0.60 - 1.77)	1.00×10^{0}
Meniscus	Medial anterior*	0.33 (0.12 - 0.79)	1.04×10^{-2}
	Medial posterior*	0.40 (0.16 - 0.89)	2.37×10^{-2}
	Lateral anterior*	0.23 (0.05 - 0.67)	5.05×10^{-3}
	Lateral posterior*	0.26 (0.06 - 0.80)	1.49×10^{-2}
Bone	TFJ medial*	0.17 (0.00 - 0.91)	3.57×10^{-2}
	TFJ lateral*	0.02 (0.00 - 0.22)	8.48×10^{-6}
	PFJ	1.11 (0.64 - 1.92)	7.93×10^{-1}
Ligament	ACL*	0.49 (0.23 - 0.99)	4.72×10^{-2}
	PCL	1.58 (0.89 - 2.87)	1.27×10^{-1}
	Popliteal	1.62 (0.96 - 2.77)	7.51×10^{-2}
Tendon	Medial patellar retinaculum*	1.98 (1.02 - 3.99)	4.19×10^{-2}
	Lateral patellar retinaculum	1.08 (0.60 - 1.96)	8.88×10^{-1}
	Popliteal	1.49 (0.87 - 2.57)	1.56×10^{-1}
	Patellar	1.76 (0.92 - 3.48)	9.00×10^{-2}
	Gastrocnemius*	2.97 (1.12 - 10.0)	2.67×10^{-2}
	Semimembranosus	0.50 (0.23 - 1.03)	6.17×10^{-2}
	Quadriceps	3.18 (0.88 - 20.4)	8.38×10^{-2}
	Gracilis	4.52 (0.74 - 290)	1.20×10^{-1}
Fat pad	Hoffa	2.38 (0.92 - 7.38)	7.80×10^{-2}
Muscle	Popliteus	1.98 (1.00 - 4.14)	5.11×10^{-2}
	Vastus medialis	1.26 (0.54 - 3.00)	6.93×10^{-1}
	Gastrocnemius	1.35 (0.73 - 2.54)	3.76×10^{-1}
	Plantaris*	2.84 (1.47 - 5.82)	1.29×10^{-3}
	Biceps femoris	4.52 (0.74 - 290)	1.20×10^{-1}
	Tibialis anterior	2.37 (0.24 - 161)	6.22×10^{-1}
	Semimembranosus	0.35 (0.05 - 1.32)	1.36×10^{-1}
Synovium	General	1.17 (0.50 - 2.82)	8.41×10^{-1}

On the other hand, several tissues located within or near the tibiofemoral joint—namely, cartilage and bone in both medial and lateral locations of the joint, menisci in all tested regions, and the ACL—saw ORs and 95% confidence intervals entirely below 1 and p values below α = 0.05. Consequently, for all of these tissues, we find a statistically significant difference in the risk of TKR within 5 years when these tissues are identified as hotspots. In the case of each, the risk of TKR appears to decrease when these tissues are identified as hotspots. Interestingly, each of these tissues have either been implicated as imaging biomarkers of OA progression, or damage within them is associated with OA onset [332–334]. These results, in conjunction with the three tissues in which risk of TKR increased when identified as hotspots, suggest that compared to OA progression, TKR onset relies less on tissues in and around the tibiofemoral joint and more on tissues in other locations of the joint to make predictions. TKR has been considered an outcome of OA progression, but these results demonstrate in part how it is a more nuanced problem.

9.6 Discussion and Conclusions

In this work, we present a pipeline that integrates MR imaging and non-imaging features to attain strong TKR prediction performance, reporting accuracy of 78.5 **±** 0.134%, sensitivity of 81.8 **±** 0.643%, and specificity of 78.4 **±** 0.138% (intervals calculated with standard error of measurement (s.e.m.), $p < 0.05$). Comparisons of AUCs showed the MRI pipeline to outperform the X-ray pipeline for patients without OA and with severe OA, thereby showing the MRI model to have a better combination of sensitivity and specificity in these OA classifications. That it did so particularly for patients without OA shows the utility of the MRI pipeline in screening for patients at risk of TKR despite higher costs. It was also interesting that, particularly among

patients with no OA, the X-ray model improved drastically more than the MRI model when non-imaging information was added, judging by disparities in AUCs. This suggests the MRI-trained DenseNet-121 may have learned to predict some of the non-imaging features from the images themselves, indicating that MRI images may intrinsically contain information regarding pain, quality of life, and physical performance, among other non-imaging variables used in this study. The utility of MRI in predicting these variables through DL is certainly worth further investigation.

A comparison of the MRI pipeline performance to past work is insightful. The closest analog to our work was conducted by Wang, T. *et al*. [311], who trained independent residual networks to predict TKR from both DESS and Turbo Spin Echo (TSE) MRI images, integrating both predictions with non-imaging variables in an LR model to yield a final TKR prediction. This yielded a model with AUC of 0.86 ± 0.01 ($p < 0.01$) when solely DESS or TSE images were used, and 0.88 ± 0.02 ($p < 0.01$) when both images and non-imaging features were integrated. Our MRI image-only model saw AUC of 0.886 ± 0.020 (image only, $p < 0.05$) and an integrated AUC of 0.834 ± 0.036 (combined, $p < 0.05$). Our image-only model thus yields performance superior to its image-only counterpart, with a 95% confidence interval lying entirely above the mean AUC of the image-only model by Wang, T. *et al*. [311]. Our integrated model, as discussed previously, was optimized to maximize Youden's index within each OA classification rather than overall AUC, explaining why our integrated model has a lower overall AUC than our image-only model. However, due to this decision, we obtained strong performance at early and moderate OA stages, with sensitivity and specificity of 92.2 ± 1.68% and 82.4 ± 0.173% at no OA, respectively,

and 78.9 ± 0.974% and 74.7 ± 0.228% at moderate OA (intervals calculated using s.e.m., $p < 0.05$). In particular, the AUC of 0.943 ± 0.029 (interval calculated with s.d., $p < 0.05$) obtained by the MRI pipeline for patients without OA, the most difficult OA classification from which to predict TKR, by far surpasses that of past TKR predictive models that include patients across all stages of OA. This performance marks progress towards a model that identifies patients at risk for TKR such that nonsurgical treatment strategies can be implemented to delay TKR.

The biomarker analysis conducted also has implications, as it identified several tissues located within or near the tibiofemoral joint as reducing risk of TKR when identified as hotspots by the full MRI pipeline—namely, these were medially and laterally located cartilage and bone, all examined meniscal regions, and the ACL. These tissues or damage within them all have been associated with progression or onset of OA, and that our model shows TKR onset to be less reliant on these imaging features in cases compared to controls demonstrates TKR onset to be a more complicated problem than OA progression, despite the relationship between the two. On the other hand, the model identifies three tissues as increasing risk of TKR when identified as hotspots in the pipeline: the medial patellar retinaculum, gastrocnemius tendon, and plantaris muscle. The medial patellar retinaculum is crucial for lateral stabilization of the knee joint, and as such, damage to it results in a patella that more easily dislocates [335]. Past work has shown patellar dislocation increases risk for OA, and TKR can be an effective procedure to treat inveterate patellar dislocation, showing a previous link between this tissue's functionality and eventual OA and TKR [336,337]. The gastrocnemius tendon and plantaris muscle, on the other hand, are both posteriorly located tissues within the knee that play a key role in knee flexion

[338]. While literature regarding the plantaris muscle is rather sparse, injuries to the muscle can be implicated in knee and calf pain felt by a patient [339]. Given their related functionality and location, the gastrocnemius tendon and plantaris muscle can jointly be implicated in conditions such as “tennis leg,” which refers to mid-calf pain felt during extension of the leg, usually due to damage to one of these tissues or their associated muscles or tendons [340]. The significance of the plantaris muscle and gastrocnemius tendon to OA progression and TKR, however, have not been well characterized, and these results justify future studies to these ends.

This study had some limitations. The first is specific to the OAI dataset, which tends towards older, female patients, all from the United States: across 4,796 patients, the mean age is 61 years and 58% of patients are female. This is not emblematic of the general population, so the robustness of the pipeline could be strengthened by testing on a dataset such as the Multicenter Osteoarthritis Study (MOST). A further limitation of the dataset is that, despite the fairly large size, there are a very limited number of patients with the classification of most interest: those without radiographic OA that still undergo TKR within 5 years. Only 66 such cases existed in the entire OAI dataset, and 12 were in the test set. As such, the OAI dataset and the number of comparison experiments we ran within and across OA classifications limits the statistical power of our conclusions. Furthermore, in this study, pixels in MRI images were compressed to 14 possible values to optimize performance—a version of the pipeline was also constructed and evaluated without the compression, but its TKR prediction performance was not as strong. Ideally, a model that uses all available information would be used in occlusion map analysis to draw more precise conclusions regarding anatomic regions that associate with

TKR, but this compromise was necessary to improve performance. A final limitation was computational intensity in occlusion map generation: the voxel size and stride used were $12 \times 32 \times 32$ and 12, respectively. These ideally would be smaller so maps could yield more precise insights but doing so was infeasible in a reasonable amount of time.

To conclude, this work presents a predictive model that delivers performance not previously seen in predicting TKR, especially for patients without OA. By delivering such performance, this pipeline can identify patients at risk of TKR with high sensitivity and specificity, and for patients with no or moderate OA, this can allow a non-invasive treatment to be implemented that prolongs good health of the knee and delays TKR. The biomarker analysis identifies the medial patellar retinaculum, gastrocnemius tendon, and plantaris muscle as increasing risk of TKR when identified as a hotspot by the model, while its assessment that several tissues within and near the tibiofemoral joint appear to reduce risk of TKR helps demonstrate the added complexity of predicting TKR onset as opposed to OA progression. Beyond this, additional directions include investigating a more effective means of integrating non-image information with image predictions to improve TKR prediction performance, assessing the efficacy of alternate network architectures, and reducing computational time to make predictions.

References

1. Cieza, A.; Causey, K.; Kamenov, K.; Hanson, S.W.; Chatterji, S.; Vos, T. Global Estimates of the Need for Rehabilitation Based on the Global Burden of Disease Study 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *The Lancet* **2020**, *396*, 2006–2017, doi:10.1016/S0140-6736(20)32340-0.
2. Abulhasan, J.F.; Grey, M.J. Anatomy and Physiology of Knee Stability. *Journal of Functional Morphology and Kinesiology 2017, Vol. 2, Page 34* **2017**, *2*, 34, doi:10.3390/JFMK2040034.
3. Rodriguez, K.; Soni, M.; Joshi, P.K.; Patel, S.C.; Shreya, D.; Zamora, D.I.; Patel, G.S.; Grossmann, I.; Sange, I. Anterior Cruciate Ligament Injury: Conservative Versus Surgical Treatment. **2021**, doi:10.7759/cureus.20206.
4. Hirschmann, M.T.; Müller, W. Complex Function of the Knee Joint: The Current Understanding of the Knee. *Knee Surgery, Sports Traumatology, Arthroscopy* **2015**, *23*, 2780–2788, doi:10.1007/S00167-015-3619-3/FIGURES/8.
5. Deshpande, B.R.; Katz, J.N.; Solomon, D.H.; Yelin, E.H.; Hunter, D.J.; Messier, S.P.; Suter, L.G.; Losina, E. Number of Persons With Symptomatic Knee Osteoarthritis in the US: Impact of Race and Ethnicity, Age, Sex, and Obesity. *Arthritis Care Res (Hoboken)* **2016**, *68*, 1743–1750, doi:10.1002/ACR.22897.
6. Kremers, H.M.; Larson, D.R.; Crowson, C.S.; Kremers, W.K.; Washington, R.E.; Steiner, C.A.; Jiranek, W.A.; Berry, D.J. Prevalence of Total Hip and Knee Replacement in the United States. *J Bone Joint Surg Am* **2015**, *97*, 1386, doi:10.2106/JBJS.N.01141.

7. Cox, L.G.E.; van Rietbergen, B.; van Donkelaar, C.C.; Ito, K. Bone Structural Changes in Osteoarthritis as a Result of Mechanoregulated Bone Adaptation: A Modeling Approach. *Osteoarthritis Cartilage* **2011**, *19*, 676–682, doi:10.1016/J.JOCA.2011.02.007.

8. RINGDAHL, E.; PANDIT, S. Treatment of Knee Osteoarthritis. *Am Fam Physician* **2011**, *83*, 1287–1292.

9. Tiulpin, A.; Thevenot, J.; Rahtu, E.; Lehenkari, P.; Saarakkala, S. Automatic Knee Osteoarthritis Diagnosis from Plain Radiographs: A Deep Learning-Based Approach. *Scientific Reports 2018 8:1* **2018**, *8*, 1–10, doi:10.1038/s41598-018-20132-7.

10. Falah, M.; Nierenberg, G.; Soudry, M.; Hayden, M.; Volpin, G. Treatment of Articular Cartilage Lesions of the Knee. *Int Orthop* **2010**, *34*, 621, doi:10.1007/S00264-010-0959-Y.

11. Hofmann, S.; Kramer, J.; Breitenseher, M.; Pietsch, M.; Aigner, N. Bone Marrow Edema in the Knee. Differential Diagnosis and Therapeutic Possibilities. *Orthopade* **2006**, *35*, 463–477, doi:10.1007/S00132-006-0952-8/FIGURES/12.

12. Fowkes, L.A.; Toms, A.P. Bone Marrow Oedema of the Knee. *Knee* **2010**, *17*, 1–6, doi:10.1016/J.KNEE.2009.06.002.

13. Glenister, R.; Sharma, S. Anatomy, Bony Pelvis and Lower Limb, Hip. *StatPearls* **2022**.

14. Parvaresh, K.C.; Pennock, A.T.; Bomar, J.D.; Wenger, D.R.; Upasani, V. v. Analysis of Acetabular Ossification From the Triradiate Cartilage and Secondary Centers. *J Pediatr Orthop* **2018**, *38*, e145–e150, doi:10.1097/BPO.0000000000001120.

15. Altman, R.; Asch, E.; Bloch, D.; Bole, G.; Borenstein, D.; Brandt, K.; Christy, W.; Cooke, T.D.; Greenwald, R.; Hochberg, M.; et al. Development of Criteria for the Classification and Reporting of Osteoarthritis. Classification of Osteoarthritis of the Knee. Diagnostic and

Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* **1986**, *29*, 1039–1049, doi:10.1002/ART.1780290816.

16. Sandiford, N.; Kendoff, D.; Muirhead-Allwood, S. Osteoarthritis of the Hip: Aetiology, Pathophysiology and Current Aspects of Management. *Ann Jt* **2020**, *5*, 8–8, doi:10.21037/AOJ.2019.10.06.

17. Parker, M.; Johansen, A. Hip Fracture. *BMJ* **2006**, *333*, 27–30, doi:10.1136/BMJ.333.7557.27.

18. Oichi, T.; Taniguchi, Y.; Oshima, Y.; Sakae Tanaka, |; Saito, T. Pathomechanism of Intervertebral Disc Degeneration. **2020**, doi:10.1002/jsp2.1076.

19. Gruber, H.E.; Ashraf, N.; Kilburn, J.; Williams, C.; Norton, H.J.; Gordon, B.E.; Hanley, E.N. Vertebral Endplate Architecture and Vascularization: Application of Micro-Computerized Tomography, a Vascular Tracer, and Immunocytochemistry in Analyses of Disc Degeneration in the Aging Sand Rat. *Spine (Phila Pa 1976)* **2005**, *30*, 2593–2600, doi:10.1097/01.BRS.0000187877.30149.83.

20. Mäki, T.; Oura, P.; Paananen, M.; Niinimäki, J.; Karppinen, J.; Junno, J.-A. Longitudinal Analysis of Paraspinal Muscle Cross-Sectional Area During Early Adulthood-A 10-Year Follow-Up MRi Study., doi:10.1038/s41598-019-56186-4.

21. Jelvéus, A. Applied Stretches to Common Muscle Groups. *Integrated Sports Massage Therapy* **2011**, 97–117, doi:10.1016/B978-0-443-10126-7.00007-1.

22. Lee, L.J.; Coppieters, M.W.; Hodges, P.W. Differential Activation of the Thoracic Multifidus and Longissimus Thoracis during Trunk Rotation. *Spine (Phila Pa 1976)* **2005**, *30*, 870–876, doi:10.1097/01.BRS.0000158956.77897.EC.

23. Prevalence and Most Common Causes of Disability Among Adults --- United States, 2005 Available online: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5816a2.htm (accessed on 1 March 2023).

24. Maher, C.; Underwood, M.; Buchbinder, R. Non-Specific Low Back Pain. *The Lancet* **2017**, *389*, 736–747, doi:10.1016/S0140-6736(16)30970-9.

25. Cs-Szabo, G.; Ragasa-San Juan, D.; Turumella, V.; Masuda, K.; Thonar, E.J.M.A.; An, H.S. Changes in MRNA and Protein Levels of Proteoglycans of the Anulus Fibrosus and Nucleus Pulposus during Intervertebral Disc Degeneration. *Spine (Phila Pa 1976)* **2002**, *27*, 2212–2219, doi:10.1097/00007632-200210150-00006.

26. Kos, N.; Gradisnik, L.; Velnar, T. A Brief Review of the Degenerative Intervertebral Disc Disease. **2019**, doi:10.5455/medarh.2019.73.421-424.

27. Pfirrmann, C.W.A.; Metzdorf, A.; Zanetti, M.; Hodler, J.; Boos, N. Magnetic Resonance Classification of Lumbar Intervertebral Disc Degeneration. *Spine (Phila Pa 1976)* **2001**, *26*, 1873–1878, doi:10.1097/00007632-200109010-00011.

28. Marc, J.; Derek, V.; Cawley, T. Modern Concepts Spinal Anatomy.

29. Toyone, T.; Takahashi, K.; Kitahara, H.; Yamagata, M.; Murakami, M.; Moriya, H. Vertebral Bone-Marrow Changes in Degenerative Lumbar Disc Disease: An MRI Study of 74 Patients with Low Back Pain. *Journal of Bone and Joint Surgery - Series B* **1994**, *76*, 757–764, doi:10.1302/0301-620X.76B5.8083266.

30. Modic, M.T.; Steinberg, P.M.; Ross, J.S.; Masaryk, T.J.; Carter, J.R. Degenerative Disk Disease: Assessment of Changes in Vertebral Body Marrow with MR Imaging. *Radiology* **1988**, *166*, 193–199, doi:10.1148/RADIOLOGY.166.1.3336678.

31. Ishimoto, Y.; Yoshimura, N.; Muraki, S.; Yamada, H.; Nagata, K.; Hashizume, H.; Takiguchi, N.; Minamide, A.; Oka, H.; Kawaguchi, H.; et al. Associations between Radiographic Lumbar Spinal Stenosis and Clinical Symptoms in the General Population: The Wakayama Spine Study. *Osteoarthritis Cartilage* **2013**, *21*, 783–788, doi:10.1016/J.JOCA.2013.02.656.

32. Cochrane Library Cochrane Database of Systematic Reviews Surgical versus Non-Surgical Treatment for Lumbar Spinal Stenosis (Review). **2016**, doi:10.1002/14651858.CD010264.pub2.

33. Gagnet, P.; Kern, K.; Andrews, K.; Elgafy, H.; Ebraheim, N. Spondylolysis and Spondylolisthesis: A Review of the Literature. **2018**, doi:10.1016/j.jor.2018.03.008.

34. Wong, E.; Altaf, F.; Lawrence, J.; Gray, R.J. Adult Degenerative Lumbar Scoliosis. *Orthopedics* **2017**, *40*, e930–e939, doi:10.3928/01477447-20170606-02.

35. Rosenberg, J.J. Scoliosis. *Pediatr Rev* **2011**, *32*, 397–398, doi:10.1542/PIR.32-9-397.

36. Burnei, G. Congenital Scoliosis: An up-to-Date. *J Med Life 8*, 388–397.

37. Alhowimel, A.; AlOtaibi, M.; Radford, K.; Coulson, N. Psychosocial Factors Associated with Change in Pain and Disability Outcomes in Chronic Low Back Pain Patients Treated by Physiotherapist: A Systematic Review. *SAGE Open Med* **2018**, *6*, 205031211875738, doi:10.1177/2050312118757387.

38. Kijima, Y.; Viegas, S.F. Wrist Anatomy and Biomechanics. *J Hand Surg Am* **2009**, *34*, 1555–1563, doi:10.1016/J.JHSA.2009.07.019.

39. Kuo, C.E.; Wolfe, S.W. Scapholunate Instability: Current Concepts in Diagnosis and Management. *J Hand Surg Am* **2008**, *33*, 998–1013, doi:10.1016/J.JHSA.2008.04.027.

40. Ayhan, Ç.; Ayhan, E. Kinesiology of the Wrist and the Hand. *Comparative Kinesiology of the Human Body: Normal and Pathological Conditions* **2020**, 211–282, doi:10.1016/B978-0-12-812162-7.00013-8.

41. Kamal, R.N.; Starr, A.; Akelman, E. Carpal Kinematics and Kinetics. *J Hand Surg Am* **2016**, *41*, 1011–1018, doi:10.1016/J.JHSA.2016.07.105.

42. FORMAN, T.A.; FORMAN, S.K.; ROSE, N.E. A Clinical Approach to Diagnosing Wrist Pain. *Am Fam Physician* **2005**, *72*, 1753–1758.

43. Akhondi, H.; Panginikkod, S. Wrist Arthritis. *Hand Clin* **2022**, *21*, 11, doi:10.1016/j.hcl.2005.08.013.

44. Silman, A.J.; Pearson, J.E. Epidemiology and Genetics of Rheumatoid Arthritis. *Arthritis Res* **2002**, *4 Suppl 3*, S265–S272, doi:10.1186/AR578.

45. Aletaha, D.; Smolen, J.S. Diagnosis and Management of Rheumatoid Arthritis: A Review. *JAMA* **2018**, *320*, 1360–1372, doi:10.1001/JAMA.2018.13103.

46. Rubin, D.A. MRI and Ultrasound of the Hands and Wrists in Rheumatoid Arthritis. I. Imaging Findings. *Skeletal Radiol* **2019**, *48*, 677–695, doi:10.1007/S00256-019-03179-Z.

47. Goekoop-Ruiterman, Y.P.M.; De Vries-Bouwstra, J.K.; Allaart, C.F.; Van Zeben, D.; Kerstens, P.J.S.M.; Hazes, J.M.W.; Zwinderman, A.H.; Ronday, H.K.; Han, K.H.; Westedt, M.L.; et al. Clinical and Radiographic Outcomes of Four Different Treatment Strategies in Patients with Early Rheumatoid Arthritis (the BeSt Study): A Randomized, Controlled Trial. *Arthritis Rheum* **2005**, *52*, 3381–3390, doi:10.1002/ART.21405.

48. Smith, K.A.; Carrino, J. MRI-Guided Interventions of the Musculoskeletal System. *Journal of Magnetic Resonance Imaging* **2008**, *27*, 339–346, doi:10.1002/JMRI.21274.

49. Bushberg, J.T.; Seibert, J.A.; Leidholdt, E.M.; Boone, J.M.; Abbey, C.K. The Essential Physics of Medical Imaging. 1177.

50. Atta-ur-Rahman; Choudhary, M.I. *Applications of NMR Spectroscopy: Volume 1*; Elsevier, 2015; ISBN 9781608059638.

51. Chen, Q.; Shah, N.J.; Worthoff, W.A. Compressed Sensing in Sodium Magnetic Resonance Imaging: Techniques, Applications, and Future Prospects. *Journal of Magnetic Resonance Imaging* **2022**, *55*, 1340–1356, doi:10.1002/JMRI.28029.

52. Levitt, M.H. Spin Dynamics: Basics of Nuclear Magnetic Resonance. *Wiley* **2015**, 709.

53. Bitar, R.; Leung, G.; Perng, R.; Tadros, S.; Moody, A.R.; Sarrazin, J.; McGregor, C.; Christakis, M.; Symons, S.; Nelson, A.; et al. MR Pulse Sequences: What Every Radiologist Wants to Know but Is Afraid to Ask1. *https://doi.org/10.1148/rg.262055063* **2006**, *26*, 513–537, doi:10.1148/RG.262055063.

54. Plewes, D.B.; Kucharczyk, W. Physics of MRI: A Primer. *Journal of Magnetic Resonance Imaging* **2012**, *35*, 1038–1054, doi:10.1002/JMRI.23642.

55. Currie, S.; Hoggard, N.; Craven, I.J.; Hadjivassiliou, M.; Wilkinson, I.D. Understanding MRI: Basic MR Physics for Physicians. *Postgrad Med J* **2013**, *89*, 209–223, doi:10.1136/POSTGRADMEDJ-2012-131342.

56. Shahzad, K.; Mati, W. Advances in Magnetic Resonance Imaging (MRI). *Advances in Medical and Surgical Engineering* **2020**, 121–142, doi:10.1016/B978-0-12-819712-7.00009-7.

57. Bley, T.A.; Wieben, O.; François, C.J.; Brittain, J.H.; Reeder, S.B. Fat and Water Magnetic Resonance Imaging. *Journal of Magnetic Resonance Imaging* **2010**, *31*, 4–18, doi:10.1002/JMRI.21895.

58. Introduction to MR Imaging. *Specialty Imaging: Temporomandibular Joint* **2016**, 346–353, doi:10.1016/B978-0-323-37704-1.50057-7.

59. Johnson, G.; Zain Wadghiri, Y.; Turnbull, D.H. 2D Multislice and 3D MRI Sequences Are Often Equally Sensitive. *Reson Med* **1999**, *41*, 824–828.

60. Golfieri, R.; Baddeley, H.; Pringle, J.S.; Souhami, R. The Role of the STIR Sequence in Magnetic Resonance Imaging Examination of Bone Tumours. *Br J Radiol* **1990**, *63*, 251–256, doi:10.1259/0007-1285-63-748-251.

61. Delfaut, E.M.; Beltran, J.; Johnson, G.; Rousseau, J.; Marchandise, X.; Cotten, A. Fat Suppression in MR Imaging: Techniques and Pitfalls. *Radiographics* **1999**, *19*, 373–382, doi:10.1148/RADIOGRAPHICS.19.2.G99MR03373.

62. Bakshi, R.; Ariyaratana, S.; Benedict, R.H.B.; Jacobs, L. Fluid-Attenuated Inversion Recovery Magnetic Resonance Imaging Detects Cortical and Juxtacortical Multiple Sclerosis Lesions. *Arch Neurol* **2001**, *58*, 742–748, doi:10.1001/ARCHNEUR.58.5.742.

63. Griswold, M.A.; Jakob, P.M.; Heidemann, R.M.; Nittka, M.; Jellus, V.; Wang, J.; Kiefer, B.; Haase, A. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). *Magn Reson Med* **2002**, *47*, 1202–1210, doi:10.1002/mrm.10171.

64. Lowe, M.J.; Sorenson, J.A. Spatially Filtering Functional Magnetic Resonance Imaging Data. *Magn Reson Med* **1997**, *37*, 723–729, doi:10.1002/MRM.1910370514.

65. Friedman, L.; Glover, G.H.; Krenz, D.; Magnotta, V. Reducing Inter-Scanner Variability of Activation in a Multicenter FMRI Study: Role of Smoothness Equalization. *Neuroimage* **2006**, *32*, 1656–1668, doi:10.1016/J.NEUROIMAGE.2006.03.062.

66. Walsh, D.O.; Gmitro, A.F.; Marcellin, M.W. Adaptive Reconstruction of Phased Array MR Imagery. **2000**, doi:10.1002/(SICI)1522-2594(200005)43:5.

67. Uecker, M.; Lai, P.; Murphy, M.J.; Virtue, P.; Elad, M.; Pauly, J.M.; Vasanawala, S.S.; Lustig, M. ESPIRiT — An Eigenvalue Approach to Autocalibrating Parallel MRI: Where SENSE Meets GRAPPA. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* **2014**, *71*, 990, doi:10.1002/MRM.24751.

68. Ren, H.; Lin, W.; Ding, X. Surface Coil Intensity Correction in Magnetic Resonance Imaging in Spinal Metastases. *Open Medicine* **2017**, *12*, 138, doi:10.1515/MED-2017-0021.

69. Hui, C.; Zhou, Y.X.; Narayana, P. A Fast Algorithm for Calculation of Inhomogeneity Gradient in MRI Data. *J Magn Reson Imaging* **2010**, *32*, 1197, doi:10.1002/JMRI.22344.

70. Dietrich, O.; Gaass, T.; Reiser, M.F. T1 Relaxation Time Constants, Influence of Oxygen, and the Oxygen Transfer Function of the Human Lung at 1.5T-A Meta-Analysis. *Eur J Radiol* **2017**, *86*, 252–260, doi:10.1016/J.EJRAD.2016.11.027.

71. Gold, G.E.; Han, E.; Stainsby, J.; Wright, G.; Brittain, J.; Beaulieu, C. Musculoskeletal MRI at 3.0 T: Relaxation Times and Image Contrast. *AJR* **2004**, 183.

72. Weiger, M.; Pruessmann, K.P. Short-T2 MRI: Principles and Recent Advances. *Prog Nucl Magn Reson Spectrosc* **2019**, *114–115*, 237–270, doi:10.1016/J.PNMRS.2019.07.001.

73. Chien, A.; Weaver, J.S.; Kinne, E.; Omar, I. Magnetic Resonance Imaging of the Knee. *Pol J Radiol* **2020**, *85*, e509, doi:10.5114/PJR.2020.99415.

74. Ibrahim, M.A.; Hazhirkarzar, B.; Dublin, A.B. Gadolinium Magnetic Resonance Imaging. *StatPearls* **2022**.

75. Lugauer, F.; Wetzl, J. Magnetic Resonance Imaging. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2018**, *11111 LNCS*, 91–118, doi:10.1007/978-3-319-96520-8_6.

76. Alsayyad, M.A.I.; Ali Shehata, K.A.; Khattab, R.T. Role of Adding T2 Mapping Sequence to the Routine MR Imaging Protocol in the Assessment of Articular Knee Cartilage in Osteoarthritis. *Egyptian Journal of Radiology and Nuclear Medicine* **2021**, *52*, 1–9, doi:10.1186/S43055-021-00453-W/FIGURES/7.

77. Wáng, Y.-X.J.; Zhang, Q.; Li, X.; Chen, W.; Ahuja, A.; Yuan, J. T1ρ Magnetic Resonance: Basic Physics Principles and Applications in Knee and Intervertebral Disc Imaging. *Quant Imaging Med Surg* **2015**, *5*, 858, doi:10.3978/J.ISSN.2223-4292.2015.12.06.

78. Marquardt, D.W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics* **1963**, *11*, 431–441, doi:10.1137/0111030.

79. Yang, L.; Sun, C.; Gong, T.; Li, Q.; Chen, X.; Zhang, X. T1ρ, T2 and T2* Mapping of Lumbar Intervertebral Disc Degeneration: A Comparison Study. *BMC Musculoskelet Disord* **2022**, *23*, 1–10, doi:10.1186/S12891-022-06040-Y/TABLES/5.

80. Zhao, H.; Li, H.; Liang, S.; Wang, X.; Yang, F. T2 Mapping for Knee Cartilage Degeneration in Young Patients with Mild Symptoms. *BMC Med Imaging* **2022**, *22*, 1–9, doi:10.1186/S12880-022-00799-1/FIGURES/4.

81. Friedrich, K.M.; Shepard, T.; De Oliveira, V.S.; Wang, L.; Babb, J.S.; Schweitzer, M.; Regatte, R. T2 Measurements of Cartilage in Osteoarthritis Patients With Meniscal Tears. *AJR Am J Roentgenol* **2009**, *193*, W411, doi:10.2214/AJR.08.2256.

82. Wyatt, C.; Kumar, D.; Subburaj, K.; Lee, S.; Nardo, L.; Narayanan, D.; Lansdown, D.; Vail, T.; Link, T.M.; Souza, R.B.; et al. Cartilage T1r and T2 Relaxation Times in Patients With Mild-to-Moderate Radiographic Hip Osteoarthritis. *ARTHRITIS & RHEUMATOLOGY* **2015**, *67*, 1548–1556, doi:10.1002/art.39074.

83. Liess, C.; Lüsse, S.; Karger, N.; Heller, M.; Glüer, C.G. Detection of Changes in Cartilage Water Content Using MRI T2-Mapping in Vivo. *Osteoarthritis Cartilage* **2002**, *10*, 907–913, doi:10.1053/joca.2002.0847.

84. Nishioka, H.; Hirose, J.; Nakamura, E.; Oniki, Y.; Takada, K.; Yamashita, Y.; Mizuta, H. T1ρ and T2 Mapping Reveal the in Vivo Extracellular Matrix of Articular Cartilage. *Journal of Magnetic Resonance Imaging* **2012**, *35*, 147–155, doi:10.1002/JMRI.22811.

85. Wongratwanich, P.; Nagasaki, T.; Shimabukuro, K.; Konishi, M.; Ohtsuka, M.; Suei, Y.; Nakamoto, T.; Akiyama, Y.; Awai, K.; Kakimoto, N. Intra- and Inter-Examination Reproducibility of T2 Mapping for Temporomandibular Joint Assessment at 3.0 T. *Scientific Reports 2022 12:1* **2022**, *12*, 1–9, doi:10.1038/s41598-022-15184-9.

86. Gracien, R.M.; Maiworm, M.; Brüche, N.; Shrestha, M.; Nöth, U.; Hattingen, E.; Wagner, M.; Deichmann, R. How Stable Is Quantitative MRI? – Assessment of Intra- and Inter-

Scanner-Model Reproducibility Using Identical Acquisition Sequences and Data Analysis Programs. *Neuroimage* **2020**, *207*, 116364, doi:10.1016/J.NEUROIMAGE.2019.116364.

87. Li, X.; Pedoia, V.; Kumar, D.; Rivoire, J.; Wyatt, C.; Lansdown, D.; Amano, K.; Okazaki, N.; Savic, D.; Koff, M.F.; et al. Cartilage T1ρ and T2 Relaxation Times: Longitudinal Reproducibility and Variations Using Different Coils, MR Systems and Sites. *Osteoarthritis Cartilage* **2015**, *23*, 2214–2223, doi:10.1016/J.JOCA.2015.07.006.

88. Li, X.; Wyatt, C.; Rivoire, J.; Han, E.; Chen, W.; Schooler, J.; Liang, F.; Shet, K.; Souza, R.; Majumdar, S. Simultaneous Acquisition of T1ρ and T2 Quantification in Knee Cartilage – Reproducibility and Diurnal Variation. *J Magn Reson Imaging* **2014**, *39*, 1287, doi:10.1002/JMRI.24253.

89. Hardy, P.A.; Recht, M.P.; Piraino, D.; Thomasson, D. Optimization of a Dual Echo in the Steady State (DESS) Free-Precession Sequence for Imaging Cartilage. *Journal of Magnetic Resonance Imaging* **1996**, *6*, 329–335, doi:10.1002/JMRI.1880060212.

90. Sharma, V.; Rai, S. A Comprehensive Study of Artificial Neural Networks. *International Journal of Advanced Research in Computer Science and Software Engineering* **2012**, *2*, 2277.

91. Zhang, X.; Chang, D.; Qi, W.; Zhan, Z. A Study on Different Functionalities and Performances among Different Activation Functions across Different ANNs for Image Classification. *J Phys Conf Ser* **2021**, *1732*, 12026, doi:10.1088/1742-6596/1732/1/012026.

92. Zhang, Z.; Sabuncu, M.R. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. *Adv Neural Inf Process Syst* **2018**, *31*.

93. Köksoy, O. Multiresponse Robust Design: Mean Square Error (MSE) Criterion. *Appl Math Comput* **2006**, *175*, 1716–1729, doi:10.1016/J.AMC.2005.09.016.

94. Smith, S.L.; Kindermans, P.J.; Ying, C.; Le, Q. V. Don't Decay the Learning Rate, Increase the Batch Size. *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings* **2017**, doi:10.48550/arxiv.1711.00489.

95. Ruder, S. An Overview of Gradient Descent Optimization Algorithms. **2016**, doi:10.48550/arxiv.1609.04747.

96. Duchi, J.; Singer, Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization * Elad Hazan. *Journal of Machine Learning Research* **2011**, *12*, 2121–2159.

97. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* **2014**, doi:10.48550/arxiv.1412.6980.

98. Zeiler, M.D. ADADELTA: An Adaptive Learning Rate Method. **2012**, doi:10.48550/arxiv.1212.5701.

99. Wang, Z.J.; Turko, R.; Shaikh, O.; Park, H.; Das, N.; Hohman, F.; Kahng, M.; Polo Chau, D.H. CNN Explainer: Learning Convolutional Neural Networks with Interactive Visualization. *IEEE Trans Vis Comput Graph* **2021**, *27*, 1396–1406, doi:10.1109/TVCG.2020.3030418.

100. Mehrotra, R.; Namuduri, K.R.; Ranganathan, N. Gabor Filter-Based Edge Detection. *Pattern Recognit* **1992**, *25*, 1479–1494, doi:10.1016/0031-3203(92)90121-X.

101. Shang, W.; Sohn, K.; Almeida, D.; Lee, H. Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units. *33rd International Conference on Machine Learning, ICML 2016* **2016**, *5*, 3276–3284, doi:10.48550/arxiv.1603.05201.

102. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing* **2004**, *13*, 600–612, doi:10.1109/TIP.2003.819861.

103. Bailey, C.R.; Bailey, A.M.; McKenney, A.S.; Weiss, C.R. Understanding and Appreciating Burnout in Radiologists. *Radiographics* **2022**, *42*, E137–E139, doi:10.1148/RG.220037/ASSET/IMAGES/LARGE/RG.220037.TBL1.JPEG.

104. Astuto, B.; Flament, I.; Namiri, N.K.; Shah, R.; Bharadwaj, U.; Link, T.M.; Bucknor, M.D.; Pedoia, V.; Majumdar, S. Automatic Deep Learning–Assisted Detection and Grading of Abnormalities in Knee MRI Studies. *Radiol Artif Intell* **2021**, *3*, doi:10.1148/RYAI.2021200165/ASSET/IMAGES/LARGE/RYAI.2021200165.FIG6.JPEG.

105. Bien, N.; Rajpurkar, P.; Ball, R.L.; Irvin, J.; Park, A.; Jones, E.; Bereket, M.; Patel, B.N.; Yeom, K.W.; Shpanskaya, K.; et al. Deep-Learning-Assisted Diagnosis for Knee Magnetic Resonance Imaging: Development and Retrospective Validation of MRNet. *PLoS Med* **2018**, *15*, e1002699, doi:10.1371/JOURNAL.PMED.1002699.

106. Krogue, J.D.; Cheng, K. V.; Hwang, K.M.; Toogood, P.; Meinberg, E.G.; Geiger, E.J.; Zaid, M.; McGill, K.C.; Patel, R.; Sohn, J.H.; et al. Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning. *Radiol Artif Intell* **2020**, *2*, doi:10.1148/RYAI.2020190023/ASSET/IMAGES/LARGE/RYAI.2020190023.FIG8.JPEG.

107. Gao, K.T.; Tibrewala, R.; Hess, M.; Bharadwaj, U.U.; Inamdar, G.; Link, T.M.; Chin, C.T.; Pedoia, V.; Majumdar, S. Automatic Detection and Voxel-Wise Mapping of Lumbar Spine Modic Changes with Deep Learning. *JOR Spine* **2022**, *5*, e1204, doi:10.1002/JSP2.1204.

108. Hallinan, J.T.P.D.; Zhu, L.; Yang, K.; Makmur, A.; Algazwi, D.A.R.; Thian, Y.L.; Lau, S.; Choo, Y.S.; Eide, S.E.; Yap, Q.V.; et al. Deep Learning Model for Automated Detection and Classification of Central Canal, Lateral Recess, and Neural Foraminal Stenosis at Lumbar Spine MRI. *Radiology* **2021**, *300*, 130–138, doi:10.1148/RADIOL.2021204289/ASSET/IMAGES/LARGE/RADIOL.2021204289.TBL4.JPEG .

109. Jamaludin, A.; Kadir, T.; Zisserman, A. SpineNet: Automated Classification and Evidence Visualization in Spinal MRIs. *Med Image Anal* **2017**, *41*, 63–73, doi:10.1016/J.MEDIA.2017.07.002.

110. Schiratti, J.B.; Dubois, R.; Herent, P.; Cahané, D.; Dachary, J.; Clozel, T.; Wainrib, G.; Keime-Guibert, F.; Lalande, A.; Pueyo, M.; et al. A Deep Learning Method for Predicting Knee Osteoarthritis Radiographic Progression from MRI. *Arthritis Res Ther* **2021**, *23*, 1–10, doi:10.1186/S13075-021-02634-4/FIGURES/6.

111. Lee, J.J.; Namiri, N.K.; Astuto, B.; Link, T.M.; Majumdar, S.; Pedoia, V. A Personalized Risk Model Leverages MRI-Based Structural Phenotypes and Clinical Factors to Predict Incidence of Radiographic Osteoarthritis. *Arthritis Care Res (Hoboken)* **2022**, *75*, 501–508, doi:10.1002/ACR.24877/ABSTRACT.

112. Ozdemir, F.; Karani, N.; Fürnstahl, P.; Goksel, O. Interactive Segmentation in MRI for Orthopedic Surgery Planning: Bone Tissue. *Int J Comput Assist Radiol Surg* **2017**, *12*, 1031–1039, doi:10.1007/S11548-017-1570-0/FIGURES/8.

113. Morales Martinez, A.; Caliva, F.; Flament, I.; Liu, F.; Lee, J.; Cao, P.; Shah, R.; Majumdar, S.; Pedoia, V. Learning Osteoarthritis Imaging Biomarkers from Bone Surface Spherical Encoding. *Magn Reson Med* **2020**, *84*, 2190–2203, doi:10.1002/MRM.28251.

114. Chalian, M.; Li, X.; Guermazi, A.; Obuchowski, N.A.; Carrino, J.A.; Oei, E.H.; Link, T.M. The QIBA Profile for MRI-Based Compositional Imaging of Knee Cartilage. *Radiology* **2021**, *301*, 423–432, doi:10.1148/RADIOL.2021204587/ASSET/IMAGES/LARGE/RADIOL.2021204587TBL6.JPEG.

115. Ahlawat, S.; Fritz, J.; Morris, C.D.; Fayad, L.M. Magnetic Resonance Imaging Biomarkers in Musculoskeletal Soft Tissue Tumors: Review of Conventional Features and Focus on Nonmorphologic Imaging. *Journal of Magnetic Resonance Imaging* **2019**, *50*, 11–27, doi:10.1002/JMRI.26659.

116. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2015**, *9351*, 234–241, doi:10.1007/978-3-319-24574-4_28/COVER.

117. Iriondo, C.; Pedoia, V.; Majumdar, S. Lumbar Intervertebral Disc Characterization through Quantitative MRI Analysis: An Automatic Voxel-Based Relaxometry Approach. *Magn Reson Med* **2020**, *84*, 1376–1390, doi:10.1002/MRM.28210.

118. Wang, C.; Guo, Y.; Chen, W.; Yu, Z. Fully Automatic Intervertebral Disc Segmentation Using Multimodal 3D U-Net. *Proceedings - International Computer Software and Applications Conference* **2019**, *1*, 730–739, doi:10.1109/COMPSAC.2019.00109.

119. Norman, B.; Pedoia, V.; Majumdar, S. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. *Radiology* **2018**, *288*, 177–185, doi:10.1148/RADIOL.2018172322.

120. Isensee, F.; Petersen, J.; Klein, A.; Zimmerer, D.; Jaeger, P.F.; Kohl, S.; Wasserthal, J.; Koehler, G.; Norajitra, T.; Wirkert, S.; et al. NnU-Net: Self-Adapting Framework for U-Net-Based Medical Image Segmentation. *Informatik aktuell* **2018**, 22, doi:10.48550/arxiv.1809.10486.

121. Liu, Q.; Xu, Z.; Jiao, Y.; Niethammer, M. ISegFormer: Interactive Segmentation via Transformers with Application to 3D Knee MR Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2022**, *13435 LNCS*, 464–474, doi:10.1007/978-3-031-16443-9_45/TABLES/4.

122. Davies, J.; Siebenhandl-Wolff, P.; Tranquart, F.; Jones, P.; Evans, P. Gadolinium: Pharmacokinetics and Toxicity in Humans and Laboratory Animals Following Contrast Agent Administration. *Arch Toxicol* **2022**, *96*, 403, doi:10.1007/S00204-021-03189-8.

123. Calabrese, E.; Rudie, J.D.; Rauschecker, A.M.; Villanueva-Meyer, J.E.; Cha, S. Feasibility of Simulated Postcontrast Mri of Glioblastomas and Lower-Grade Gliomas by Using Three-Dimensional Fully Convolutional Neural Networks. *Radiol Artif Intell* **2021**, *3*, doi:10.1148/RYAI.2021200276/ASSET/IMAGES/LARGE/RYAI.2021200276.FIG7.JPEG.

124. Gong, E.; Pauly, J.M.; Wintermark, M.; Zaharchuk, G. Deep Learning Enables Reduced Gadolinium Dose for Contrast-Enhanced Brain MRI. *Journal of Magnetic Resonance Imaging* **2018**, *48*, 330–340, doi:10.1002/JMRI.25970.

125. Pasumarthi, S.; Tamir, J.I.; Christensen, S.; Zaharchuk, G.; Zhang, T.; Gong, E. A Generic Deep Learning Model for Reduced Gadolinium Dose in Contrast-Enhanced Brain MRI. *Magn Reson Med* **2021**, *86*, 1687–1700, doi:10.1002/MRM.28808.

126. Zhang, N.; Yang, G.; Gao, Z.; Xu, C.; Zhang, Y.; Shi, R.; Keegan, J.; Xu, L.; Zhang, H.; Fan, Z.; et al. Deep Learning for Diagnosis of Chronic Myocardial Infarction on Nonenhanced Cardiac Cine MRI. *Radiology* **2019**, *291*, 606–607, doi:10.1148/RADIOL.2019182304/ASSET/IMAGES/LARGE/RADIOL.2019182304.FIG6.JPEG .

127. Lustig, M.; Donoho, D.L.; Santos, J.M.; Pauly, J.M. Compressed Sensing MRI: A Look at How CS Can Improve on Current Imaging Techniques. *IEEE Signal Process Mag* **2008**, *25*, 72–82, doi:10.1109/MSP.2007.914728.

128. Lustig, M.; Donoho, D.; Pauly, J.M. Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging. *Magn Reson Med* **2007**, *58*, 1182–1195, doi:10.1002/MRM.21391.

129. Chen, J.; Liu, S.; Huang, M. Low-Rank and Sparse Decomposition Model for Accelerating Dynamic MRI Reconstruction. *J Healthc Eng* **2017**, *2017*, doi:10.1155/2017/9856058.

130. Ma, D.; Gulani, V.; Seiberlich, N.; Liu, K.; Sunshine, J.L.; Duerk, J.L.; Griswold, M.A. Magnetic Resonance Fingerprinting. *Nature 2013 495:7440* **2013**, *495*, 187–192, doi:10.1038/nature11971.

131. Muckley, M.J.; Riemenschneider, B.; Radmanesh, A.; Kim, S.; Jeong, G.; Ko, J.; Jun, Y.; Shin, H.; Hwang, D.; Mostapha, M.; et al. Results of the 2020 FastMRI Challenge for Machine Learning MR Image Reconstruction. *IEEE Trans Med Imaging* **2021**, *40*, 2306, doi:10.1109/TMI.2021.3075856.

132. Eo, T.; Jun, Y.; Kim, T.; Jang, J.; Lee, H.J.; Hwang, D. KIKI-Net: Cross-Domain Convolutional Neural Networks for Reconstructing Undersampled Magnetic Resonance Images. *Magn Reson Med* **2018**, *80*, 2188–2201, doi:10.1002/MRM.27201.

133. Zhu, B.; Liu, J.Z.; Cauley, S.F.; Rosen, B.R.; Rosen, M.S. Image Reconstruction by Domain-Transform Manifold Learning. *Nature 2018 555:7697* **2018**, *555*, 487–492, doi:10.1038/nature25988.

134. Liu, F.; Feng, L.; Kijowski, R. MANTIS: Model-Augmented Neural NeTwork with Incoherent k-Space Sampling for Efficient MR Parameter Mapping. *Magn Reson Med* **2019**, *82*, 174–188, doi:10.1002/MRM.27707.

135. Hammernik, K.; Klatzer, T.; Kobler, E.; Recht, M.P.; Sodickson, D.K.; Pock, T.; Knoll, F. Learning a Variational Network for Reconstruction of Accelerated MRI Data. *Magn Reson Med* **2018**, *79*, 3055–3071, doi:10.1002/MRM.26977/ASSET/SUPINFO/MRM26977-SUP-0005-SUPPINFO05.MP4.

136. Calivá, F.; Cheng, K.; Shah, R.; Pedoia, V. Adversarial Robust Training of Deep Learning MRI Reconstruction Models. *Journal of Machine Learning for Biomedical Imaging* **2020**, 7–8, doi:10.48550/arxiv.2011.00070.

137. Darestani, M.Z.; Chaudhari, A.S.; Heckel, R. Measuring Robustness in Deep Learning Based Compressive Sensing. **2021**.

138. Antun, V.; Renna, F.; Poon, C.; Adcock, B.; Hansen, A.C. On Instabilities of Deep Learning in Image Reconstruction and the Potential Costs of AI. *Proc Natl Acad Sci U S A* **2020**, *117*, 30088–30095, doi:10.1073/PNAS.1907377117/SUPPL_FILE/PNAS.1907377117.SAPP.PDF.

139. Liu, F.; Kijowski, R.; Feng, L.; El Fakhri, G. High-Performance Rapid MR Parameter Mapping Using Model-Based Deep Adversarial Learning. *Magn Reson Imaging* **2020**, *74*, 152–160, doi:10.1016/J.MRI.2020.09.021.

140. Liu, F.; Samsonov, A.; Chen, L.; Kijowski, R.; Feng, L. SANTIS: Sampling-Augmented Neural NeTwork with Incoherent Structure for MR Image Reconstruction. *Magn Reson Med* **2019**, *82*, 1890–1904, doi:10.1002/MRM.27827.

141. Knoll, F.; Murrell, T.; Sriram, A.; Yakubova, N.; Zbontar, J.; Rabbat, M.; Defazio, A.; Muckley, M.J.; Sodickson, D.K.; Zitnick, C.L.; et al. Advancing Machine Learning for MR Image Reconstruction with an Open Competition: Overview of the 2019 FastMRI Challenge. *Magn Reson Med* **2020**, *84*, 3054–3070, doi:10.1002/MRM.28338.

142. Mason, A.; Rioux, J.; Clarke, S.E.; Costa, A.; Schmidt, M.; Keough, V.; Huynh, T.; Beyea, S. Comparison of Objective Image Quality Metrics to Expert Radiologists' Scoring of Diagnostic Quality of MR Images. *IEEE Trans Med Imaging* **2020**, *39*, 1064–1072, doi:10.1109/TMI.2019.2930338.

143. Roos, E.M.; Arden, N.K. Strategies for the Prevention of Knee Osteoarthritis. *Nat Rev Rheumatol* **2016**, *12*, 92–101, doi:10.1038/NRRHEUM.2015.135.

144. Sriram, A.; Zbontar, J.; Murrell, T.; Zitnick, C.L.; Defazio, A.; Sodickson, D.K. GrappaNet: Combining Parallel Imaging with Deep Learning for Multi-Coil MRI Reconstruction.

145. Schlemper, J.; Oksuz, I.; Clough, J.R.; Duan, J.; King, A.P.; Schnabel, J.A.; Hajnal, J. V.; Rueckert, D. DAUTOMAP: Decomposing AUTOMAP to Achieve Scalability and Enhance Performance. **2019**, doi:10.48550/arxiv.1909.10995.

146. Yaman, B.; Amir, S.; Hosseini, H.; Akçakaya, M. Zero-Shot Self-Supervised Learning for MRI Reconstruction. **2021**, doi:10.48550/arxiv.2102.07737.

147. Aggarwal, H.K.; Jacob, M. J-MoDL: Joint Model-Based Deep Learning for Optimized Sampling and Reconstruction. *IEEE Journal on Selected Topics in Signal Processing* **2019**, *14*, 1151–1162, doi:10.1109/JSTSP.2020.3004094.

148. Dwork, N.; O'Connor, D.; Baron, C.A.; Johnson, E.M.I.; Kerr, A.B.; Pauly, J.M.; Larson, P.E.Z. Utilizing the Wavelet Transform's Structure in Compressed Sensing. *Signal Image Video Process* **2021**, *15*, 1407–1414, doi:10.1007/S11760-021-01872-Y.

149. Crema, M.D.; Roemer, F.W.; Marra, M.D.; Burstein, D.; Gold, G.E.; Eckstein, F.; Baum, T.; Mosher, T.J.; Carrino, J.A.; Guermazi, A. Articular Cartilage in the Knee: Current MR Imaging Techniques and Applications in Clinical Practice and Research. *Radiographics* **2011**, *31*, 37–61, doi:10.1148/RG.311105084.

150. Brau, A.C.S.; Beatty, P.J.; Skare, S.; Bammer, R. Comparison of Reconstruction Accuracy and Efficiency among Autocalibrating Data-Driven Parallel Imaging Methods. *Magn. Res. Med.* **2008**, *59*, 382–395, doi:10.1002/MRM.21481.

151. Arouche Nunes, B.A.; Flament, I.; Shah, R.; Bucknor, M.; Link, T.; Pedoia, V.; Majumdar, S. MRI-Based Multi-Task Deep Learning for Cartilage Lesion Severity Staging in Knee Osteoarthritis. *Osteoarthritis Cartilage* **2019**, *27*, S398–S399, doi:10.1016/j.joca.2019.02.399.

152. Pal, A.; Rathi, Y. A Review and Experimental Evaluation of Deep Learning Methods for MRI Reconstruction. *The journal of machine learning for biomedical imaging* **2022**, *1*.

153. Braun, H.J.; Gold, G.E. Diagnosis of Osteoarthritis: Imaging. *Bone* **2012**, *51*, 278–288, doi:10.1016/J.BONE.2011.11.019.

154. Wang, X.; Oo, W.M.; Linklater, J.M. What Is the Role of Imaging in the Clinical Diagnosis of Osteoarthritis and Disease Management? *Rheumatology (Oxford)* **2018**, *57*, iv51–iv60, doi:10.1093/RHEUMATOLOGY/KEX501.

155. Abd-Ellah, M.K.; Awad, A.I.; Khalaf, A.A.M.; Hamed, H.F.A. A Review on Brain Tumor Diagnosis from MRI Images: Practical Implications, Key Achievements, and Lessons Learned. *Magn Reson Imaging* **2019**, *61*, 300–318, doi:10.1016/J.MRI.2019.05.028.

156. Widmann, G.; Henninger, B.; Kremser, C.; Jaschke, W. MRI Sequences in Head & Neck Radiology - State of the Art. *Rofo* **2017**, *189*, 413–422, doi:10.1055/S-0043-103280.

157. Chaudhari, A.S.; Kogan, F.; Pedoia, V.; Majumdar, S.; Gold, G.E.; Hargreaves, B.A. Rapid Knee MRI Acquisition and Analysis Techniques for Imaging Osteoarthritis. *J Magn Reson Imaging* **2020**, *52*, 1321–1339, doi:10.1002/JMRI.26991.

158. Soellner, S.T.; Goldmann, A.; Muelheims, D.; Welsch, G.H.; Pachowsky, M.L. Intraoperative Validation of Quantitative T2 Mapping in Patients with Articular Cartilage Lesions of the Knee. *Osteoarthritis Cartilage* **2017**, *25*, 1841–1849, doi:10.1016/J.JOCA.2017.07.021.

159. Roemer, F.W.; Kijowski, R.; Guermazi, A. Editorial: From Theory to Practice - the Challenges of Compositional MRI in Osteoarthritis Research. *Osteoarthritis Cartilage* **2017**, *25*, 1923–1925, doi:10.1016/J.JOCA.2017.08.007.

160. Atkinson, H.F.; Birmingham, T.B.; Moyer, R.F.; Yacoub, D.; Kanko, L.E.; Bryant, D.M.; Thiessen, J.D.; Thompson, R.T. MRI T2 and T1ρ Relaxation in Patients at Risk for Knee Osteoarthritis: A Systematic Review and Meta-Analysis. *BMC Musculoskeletal Disorders 2019 20:1* **2019**, *20*, 1–18, doi:10.1186/S12891-019-2547-7.

161. Liebl, H.; Joseph, G.; Nevitt, M.C.; Singh, N.; Heilmeier, U.; Subburaj, K.; Jungmann, P.M.; McCulloch, C.E.; Lynch, J.A.; Lane, N.E.; et al. Early T2 Changes Predict Onset of Radiographic Knee Osteoarthritis: Data from the Osteoarthritis Initiative. *Ann Rheum Dis* **2015**, *74*, 1353–1359, doi:10.1136/ANNRHEUMDIS-2013-204157.

162. Wang, Y.X.J.; Zhao, F.; Griffith, J.F.; Mok, G.S.P.; Leung, J.C.S.; Ahuja, A.T.; Yuan, J. T1rho and T2 Relaxation Times for Lumbar Disc Degeneration: An in Vivo Comparative Study at 3.0-Tesla MRI. *Eur Radiol* **2013**, *23*, 228–234, doi:10.1007/S00330-012-2591-2.

163. Bashir, A.; Gray, M.L.; Burstein, D. Gd-DTPA2- as a Measure of Cartilage Degradation. *Magn Reson Med* **1996**, *36*, 665–673, doi:10.1002/MRM.1910360504.

164. Fernquest, S.; Palmer, A.; Gammer, B.; Hirons, E.; Kendrick, B.; Taylor, A.; De Berker, H.; Bangerter, N.; Carr, A.; Glyn-Jones, S. Compositional MRI of the Hip: Reproducibility, Effect of Joint Unloading, and Comparison of T2 Relaxometry with Delayed Gadolinium-Enhanced Magnetic Resonance Imaging of Cartilage. *Cartilage* **2021**, *12*, 418–430, doi:10.1177/1947603519841670.

165. KELLGREN, J.H.; LAWRENCE, J.S. Radiological Assessment of Osteo-Arthrosis. *Ann Rheum Dis* **1957**, *16*, 494–502, doi:10.1136/ARD.16.4.494.

166. Glockner, J.F.; Hu, H.H.; Stanley, D.W.; Angelos, L.; King, K. Parallel MR Imaging: A User's Guide. *Radiographics* **2005**, *25*, 1279–1297, doi:10.1148/RG.255045202.

167. Pruessmann, K.P.; Weiger, M.; Scheidegger, M.B.; Boesiger, P. SENSE: Sensitivity Encoding for Fast MRI., doi:10.1002/(SICI)1522-2594(199911)42:5.

168. Altahawi, F.F.; Blount, K.J.; Morley, N.P.; Raithel, E.; Omar, I.M. Comparing an Accelerated 3D Fast Spin-Echo Sequence (CS-SPACE) for Knee 3-T Magnetic Resonance Imaging with Traditional 3D Fast Spin-Echo (SPACE) and Routine 2D Sequences. *Skeletal Radiol* **2017**, *46*, 7–15, doi:10.1007/S00256-016-2490-8.

169. Niitsu, M.; Ikeda, K. Routine MR Examination of the Knee Using Parallel Imaging. *Clin Radiol* **2003**, *58*, 801–807, doi:10.1016/S0009-9260(03)00260-5.

170. Ye, J.C. Compressed Sensing MRI: A Review from Signal Processing Perspective. *BMC Biomedical Engineering 2019 1:1* **2019**, *1*, 1–17, doi:10.1186/S42490-019-0006-Z.

171. Iuga, A.I.; Abdullayev, N.; Weiss, K.; Haneder, S.; Brüggemann-Bratke, L.; Maintz, D.; Rau, R.; Bratke, G. Accelerated MRI of the Knee. Quality and Efficiency of Compressed Sensing. *Eur J Radiol* **2020**, *132*, doi:10.1016/J.EJRAD.2020.109273.

172. Zibetti, M.V.W.; Sharafi, A.; Otazo, R.; Regatte, R.R. Accelerating 3D-T1ρ Mapping of Cartilage Using Compressed Sensing with Different Sparse and Low Rank Models. *Magn Reson Med* **2018**, *80*, 1475–1491, doi:10.1002/MRM.27138.

173. Pandit, P.; Rivoire, J.; King, K.; Li, X. Accelerated T1ρ Acquisition for Knee Cartilage Quantification Using Compressed Sensing and Data-Driven Parallel Imaging: A Feasibility Study. *Magn Reson Med* **2016**, *75*, 1256–1261, doi:10.1002/MRM.25702.

174. Sumpf, T.J.; Uecker, M.; Boretius, S.; Frahm, J. Model-Based Nonlinear Inverse Reconstruction for T2 Mapping Using Highly Undersampled Spin-Echo MRI. *J Magn Reson Imaging* **2011**, *34*, 420–428, doi:10.1002/JMRI.22634.

175. Hilbert, T.; Sumpf, T.J.; Weiland, E.; Frahm, J.; Thiran, J.P.; Meuli, R.; Kober, T.; Krueger, G. Accelerated T2 Mapping Combining Parallel MRI and Model-Based Reconstruction: GRAPPATINI. *J Magn Reson Imaging* **2018**, *48*, 359–368, doi:10.1002/JMRI.25972.

176. Raudner, M.; Schreiner, M.M.; Hilbert, T.; Kober, T.; Weber, M.; Szelényi, A.; Windhager, R.; Juras, V.; Trattnig, S. Clinical Implementation of Accelerated T2 Mapping: Quantitative Magnetic Resonance Imaging as a Biomarker for Annular Tear and Lumbar Disc Herniation. *Eur Radiol* **2021**, *31*, 3590–3599, doi:10.1007/S00330-020-07538-6.

177. Desai, A.D.; Gunel, B.; Ozturkler, B.M.; Beg, H.; Vasanawala, S.; Hargreaves, B.A.; Ré, C.; Pauly, J.M.; Chaudhari, A.S. VORTEX: Physics-Driven Data Augmentations Using Consistency Training for Robust Accelerated MRI Reconstruction. **2021**, doi:10.48550/arxiv.2111.02549.

178. Chaudhari, A.; Fang, Z.; Hyung Lee, J.; Gold, G.; Hargreaves, B. Deep Learning Super-Resolution Enables Rapid Simultaneous Morphological and Quantitative Magnetic Resonance Imaging. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2018**, *11074 LNCS*, 3–11, doi:10.1007/978-3-030-00129-2_1/TABLES/2.

179. Cai, C.; Wang, C.; Zeng, Y.; Cai, S.; Liang, D.; Wu, Y.; Chen, Z.; Ding, X.; Zhong, J. Single-Shot T2 Mapping Using Overlapping-Echo Detachment Planar Imaging and a Deep Convolutional Neural Network. *Magn Reson Med* **2018**, *80*, 2202–2214, doi:10.1002/MRM.27205.

180. Jeelani, H.; Yang, Y.; Zhou, R.; Kramer, C.M.; Salerno, M.; Weller, D.S. A Myocardial T1-Mapping Framework with Recurrent and U-Net Convolutional Neural Networks.

Proceedings - International Symposium on Biomedical Imaging **2020**, *2020-April*, 1941–1944, doi:10.1109/ISBI45749.2020.9098459.

181. Li, H.; Yang, M.; Kim, J.; Liu, R.; Zhang, C.; Huang, P.; Gaire, S.K.; Liang, D.; Li, X.; Ying, L. Ultra-Fast Simultaneous T1rho and T2 Mapping Using Deep Learning. In Proceedings of the Proceedings of the 28th Annual Meeting of ISMRM, Virtual Conference and Exhibition; 2020; p. 2669.
182. Fang, Z.; Chen, Y.; Liu, M.; Xiang, L.; Zhang, Q.; Wang, Q.; Lin, W.; Shen, D. Deep Learning for Fast and Spatially Constrained Tissue Quantification From Highly Accelerated Data in Magnetic Resonance Fingerprinting. *IEEE Trans Med Imaging* **2019**, *38*, 2364–2374, doi:10.1109/TMI.2019.2899328.
183. Cohen, O.; Zhu, B.; Rosen, M.S. MR Fingerprinting Deep RecOnstruction NEtwork (DRONE). *Magn Reson Med* **2018**, *80*, 885–894, doi:10.1002/MRM.27198.
184. Meng, Z.; Guo, R.; Li, Y.; Guan, Y.; Wang, T.; Zhao, Y.; Sutton, B.; Li, Y.; Liang, Z.P. Accelerating T2 Mapping of the Brain by Integrating Deep Learning Priors with Low-Rank and Sparse Modeling. *Magn Reson Med* **2021**, *85*, 1455–1467, doi:10.1002/MRM.28526.
185. Adamson, P.M.; Gunel, B.; Dominic, J.; Desai, A.D.; Spielman, D.; Vasanawala, S.; Pauly, J.M.; Chaudhari, A. SSFD: Self-Supervised Feature Distance as an MR Image Reconstruction Quality Metric. In Proceedings of the NeurIPS; December 2 2021.
186. Tolpadi, A.; Han, M.; Caliva, F.; Pedoia, V.; Majumdar, S. A Recurrent Encoder-Decoder Network Accelerates T2 Mapping in the Knee, Hip and Lumbar Spine. In Proceedings of the Proceedings of the 30th Annual Meeting of ISMRM; London, United Kingdom, 2022; Vol. 14, p. 1339.

187. Tolpadi, A.; Caliva, F.; Han, M.; Pedoia, V.; Majumdar, S. 8X Accelerated Intervertebral Disc Compositional Evaluation with Recurrent Encoder-Decoder Deep Learning Network. In Proceedings of the In Proceedings of the 29th Annual Meeting of ISMRM, Virtual Conference and Exhibition; 2021; p. 2179.

188. Caliva, F.; Iriondo, C.; Martinez, A.M.; Majumdar, S.; Pedoia, V. Distance Map Loss Penalty Term for Semantic Segmentation. **2019**, doi:10.48550/arxiv.1908.03679.

189. Joseph, G.B.; Baum, T.; Carballido-Gamio, J.; Nardo, L.; Virayavanich, W.; Alizai, H.; Lynch, J.A.; McCulloch, C.E.; Majumdar, S.; Link, T.M. Texture Analysis of Cartilage T2 Maps: Individuals with Risk Factors for OA Have Higher and More Heterogeneous Knee Cartilage MR T2 Compared to Normal Controls--Data from the Osteoarthritis Initiative. *Arthritis Res Ther* **2011**, *13*, doi:10.1186/AR3469.

190. Urish, K.L.; Keffalas, M.G.; Durkin, J.R.; Miller, D.J.; Chu, C.R.; Mosher, T.J. T2 Texture Index of Cartilage Can Predict Early Symptomatic OA Progression: Data from the Osteoarthritis Initiative. *Osteoarthritis Cartilage* **2013**, *21*, 1550–1557, doi:10.1016/J.JOCA.2013.06.007.

191. Chen, W.; Takahashi, A.; Han, E. Quantitative T(1)(ρ) Imaging Using Phase Cycling for B0 and B1 Field Inhomogeneity Compensation. *Magn Reson Imaging* **2011**, *29*, 608–619, doi:10.1016/J.MRI.2011.02.002.

192. Dixon, W.T.; Oshinski, J.N.; Trudeau, J.D.; Arnold, B.C.; Pettigrew, R.I. Myocardial Suppression in Vivo by Spin Locking with Composite Pulses. *Magn Reson Med* **1996**, *36*, 90–94, doi:10.1002/MRM.1910360116.

193. Foltz, W.D.; Stainsby, J.A.; Wright, G.A. T2 Accuracy on a Whole-Body Imager. *Magn Reson Med* **1997**, *38*, 759–768, doi:10.1002/MRM.1910380512.

194. Weingärtner, S. Manual for RView. **2010**.

195. Moré, J.J. The Levenberg-Marquardt Algorithm: Implementation and Theory. **1978**, 105–116, doi:10.1007/BFB0067700.

196. Shimron, E.; Tamir, J.I.; Wang, K.; Lustig, M. Implicit Data Crimes: Machine Learning Bias Arising from Misuse of Public Data. *Proc Natl Acad Sci U S A* **2022**, *119*, e2117203119, doi:10.1073/PNAS.2117203119/SUPPL_FILE/PNAS.2117203119.SAPP.PDF.

197. Srivastava, R.K.; Greff, K.; Ch, K.; Urgen Schmidhuber, J.¨ Highway Networks. **2015**, doi:10.48550/arxiv.1505.00387.

198. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* **2014**, doi:10.48550/arxiv.1409.1556.

199. Schober, P.; Schwarte, L.A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg* **2018**, *126*, 1763–1768, doi:10.1213/ANE.0000000000002864.

200. Haralick, R.M.; Dinstein, I.; Shanmugam, K. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern* **1973**, *SMC-3*, 610–621, doi:10.1109/TSMC.1973.4309314.

201. Koo, T.K.; Li, M.Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* **2016**, *15*, 155–163, doi:10.1016/J.JCM.2016.02.012.

202. Ratner, B. The Correlation Coefficient: Its Values Range between +1/-1, or Do They. *Journal of Targeting, Measurement and Analysis for Marketing* **2009**, *17*, 139–142, doi:10.1057/JT.2009.5/TABLES/2.

203. Dunn, T.C.; Lu, Y.; Jin, H.; Ries, M.D.; Majumdar, S. T2 Relaxation Time of Cartilage at MR Imaging: Comparison with Severity of Knee Osteoarthritis. *Radiology* **2004**, *232*, 592, doi:10.1148/RADIOL.2322030976.

204. Cui, Y.-Z.; Yang, X.-H.; Liu, P.-F.; Chen, W.-J. Preliminary Study on Diagnosis of Lumbar Disc Degeneration with Magnetic Resonance T1p, T2 Mapping and DWI Quantitative Detection Technologies. *Eur Rev Med Pharmacol Sci* **2016**, *20*, 3344–3350.

205. Gallo, M.C.; Wyatt, C.; Pedoia, V.; Kumar, D.; Lee, S.; Nardo, L.; Link, T.M.; Souza, R.B.; Majumdar, S. T1ρ and T2 Relaxation Times Are Associated with Progression of Hip Osteoarthritis. *Osteoarthritis Cartilage* **2016**, *24*, 1399–1407, doi:10.1016/J.JOCA.2016.03.005.

206. Janacova, V.; Szomolanyi, P.; Vilimek, D.; Trattnig, S.; Juras, V. GLCM Texture Analysis of Knee Cartilage T2 Maps: Machine Learning Based Selection of Important Features. In Proceedings of the Proceedings of the 30th Annual Meeting of ISMRM; London, United Kingdom, 2022; p. 1498.

207. Dean Deyle, G. The Role of MRI in Musculoskeletal Practice: A Clinical Perspective. *J. Man. Manip. Ther.* **2011**, *19*, 152–161, doi:10.1179/2042618611Y.0000000009.

208. del Grande, F.; Guggenberger, R.; Fritz, J. Rapid Musculoskeletal MRI in 2021: Value and Optimized Use of Widely Accessible Techniques. *AJR* **2021**, *216*, 704–717, doi:10.2214/AJR.20.22901.

209. Aydıngöz, Ü.; Yıldız, A.E.; Ergen, F.B. Zero Echo Time Musculoskeletal MRI: Technique, Optimization, Applications, and Pitfalls. *RadioGraphics* **2022**, *42*, 1398–1414, doi:10.1148/RG.220029/ASSET/IMAGES/LARGE/RG.220029.FIG18.JPEG.

210. Larson, P.E.Z.; Han, M.; Krug, R.; Jakary, A.; Nelson, S.J.; Vigneron, D.B.; Henry, R.G.; McKinnon, G.; Kelley, D.A.C. Ultrashort Echo Time and Zero Echo Time MRI at 7T. *Magma* **2016**, *29*, 359–370, doi:10.1007/S10334-015-0509-0.

211. Afsahi, A.M.; Ma, Y.; Jang, H.; Jerban, S.; Chung, C.B.; Chang, E.Y.; Du, J. Ultrashort Echo Time Magnetic Resonance Imaging Techniques: Met and Unmet Needs in Musculoskeletal Imaging. *J. Magn. Reson. Imaging* **2022**, *55*, 1597–1612, doi:10.1002/JMRI.28032.

212. Chang, E.Y.; Du, J.; Chung, C.B. UTE Imaging in the Musculoskeletal System. *J. Magn. Reson. Imaging* **2015**, *41*, 870, doi:10.1002/JMRI.24713.

213. Yoon, J.H.; Nickel, M.D.; Peeters, J.M.; Lee, J.M. Rapid Imaging: Recent Advances in Abdominal MRI for Reducing Acquisition Time and Its Clinical Applications. *KJR* **2019**, *20*, 1597–1615, doi:10.3348/KJR.2018.0931.

214. Esteban, O.; Blair, R.W.; Nielson, D.M.; Varada, J.C.; Marrett, S.; Thomas, A.G.; Poldrack, R.A.; Gorgolewski, K.J. Crowdsourced MRI Quality Metrics and Expert Quality Annotations for Training of Humans and Machines. *Sci Data* **2019**, *6*, 1–7, doi:10.1038/s41597-019-0035-4.

215. Su, Y.; Anderson, M.; Tamir, J.I.; Lustig, M.; Li, K. Compressed Sensing MRI Reconstruction on Intel HARPv2. *Proceedings - 27th IEEE International Symposium on Field-Programmable Custom Computing Machines* **2019**, 254–257, doi:10.1109/FCCM.2019.00041.

216. Shimron, E.; de Goyeneche, A.; Wang, K.; Halgren, A.; Syed, A.B.; Vasanawala, S.; Lustig, M. BladeNet: Rapid PROPELLER Acquisition and Reconstruction for High Spatio-Temporal

Resolution Abdominal MRI. In Proceedings of the Proceedings of the 31st Annual International Society for Magnetic Resonance in Medicine; London, 2022; p. 684.

217. Tippareddy, C.; Zhao, W.; Sunshine, J.L.; Griswold, M.; Ma, D.; Badve, C. Magnetic Resonance Fingerprinting: An Overview. *EJNMMI* **2021**, *48*, 4189–4200, doi:10.1007/S00259-021-05384-2/FIGURES/8.

218. Boyacioglu, R.; Wang, C.; Ma, D.; McGivney, D.F.; Yu, X.; Griswold, M.A. 3D Magnetic Resonance Fingerprinting with Quadratic RF Phase. *Magn. Reson. Med.* **2021**, *85*, 2084–2094, doi:10.1002/MRM.28581.

219. Baccouche, A.; Garcia-Zapirain, B.; Castillo Olea, C.; Elmaghraby, A.S. Connected-UNets: A Deep Learning Architecture for Breast Mass Segmentation. *NPJ Breast Cancer* **2021**, *7*, 1–12, doi:10.1038/s41523-021-00358-x.

220. Hess, M.; Allaire, B.; Gao, K.T.; Tibrewala, R.; Inamdar, G.; Bharadwaj, U.; Chin, C.; Pedoia, V.; Bouxsein, M.; Anderson, D.; et al. Deep Learning for Multi-Tissue Segmentation and Fully Automatic Personalized Biomechanical Models from BACPAC Clinical Lumbar Spine MRI. *Pain Medicine* **2022**, doi:10.1093/PM/PNAC142.

221. Zhang, X.; Li, Z.; Shi, H.; Deng, Y.; Zhou, G.; Tang, S. A Deep Learning-Based Method for Knee Articular Cartilage Segmentation in MRI Images. *10th International Conference on Control, Automation and Information Sciences, ICCAIS 2021 - Proceedings* **2021**, 690–694, doi:10.1109/ICCAIS52680.2021.9624655.

222. Iriondo, C.; Liu, F.; Calivà, F.; Kamat, S.; Majumdar, S.; Pedoia, V. Towards Understanding Mechanistic Subgroups of Osteoarthritis: 8-Year Cartilage Thickness Trajectory Analysis. *J. Orthop. Res.* **2021**, *39*, 1305–1317, doi:10.1002/JOR.24849.

223. Namiri, N.K.; Flament, I.; Astuto, B.; Shah, R.; Tibrewala, R.; Caliva, F.; Link, T.M.; Pedoia, V.; Majumdar, S. Deep Learning for Hierarchical Severity Staging of Anterior Cruciate Ligament Injuries from Mri. *Radiol. Artif. Intell.* **2020**, *2*, 1–8, doi:10.1148/RYAI.2020190207/ASSET/IMAGES/LARGE/RYAI.2020190207.FIG6.JPEG.

224. Gao, K.T.; Pedoia, V.; Young, K.A.; Kogan, F.; Koff, M.F.; Gold, G.E.; Potter, H.G.; Majumdar, S. Multiparametric MRI Characterization of Knee Articular Cartilage and Subchondral Bone Shape in Collegiate Basketball Players. *J. Orthop. Res.* **2021**, *39*, 1512–1522, doi:10.1002/JOR.24851.

225. Calivá, F.; Leynes, A.P.; Shah, R.; Bharadwaj, U.U.; Majumdar, S.; Larson, P.E.Z.; Pedoia, V. Breaking Speed Limits with Simultaneous Ultra-Fast MRI Reconstruction and Tissue Segmentation. *Proc Mach Learn Res* 2020, *121*, 94–110.

226. Fienup, J.R. Invariant Error Metrics for Image Reconstruction. *Appl. Opt.* **1997**, *36*, 8352–8357, doi:10.1364/AO.36.008352.

227. Horé, A.; Ziou, D. Is There a Relationship between Peak-Signal-to-Noise Ratio and Structural Similarity Index Measure? *IET Image Process.* **2013**, *7*, 12–24, doi:10.1049/IET-IPR.2012.0489.

228. Dosselmann, R.; Yang, X.D. A Comprehensive Assessment of the Structural Similarity Index. *Signal, Image and Video Process.* **2009**, *5*, 81–91, doi:10.1007/S11760-009-0144-1.

229. Peterfy, C.G.; Schneider, E.; Nevitt, M. The Osteoarthritis Initiative: Report on the Design Rationale for the Magnetic Resonance Imaging Protocol for the Knee. *Osteoarthr. Cartil.* **2008**, *16*, 1433–1441, doi:10.1016/J.JOCA.2008.06.016.

230. Segal, N.A.; Nevitt, M.C.; Gross, K.D.; Hietpas, J.; Glass, N.A.; Lewis, C.E.; Torner, J.C. The Multicenter Osteoarthritis Study: Opportunities for Rehabilitation Research. *PM&R* **2013**, *5*, 647–654, doi:10.1016/J.PMRJ.2013.04.014.

231. Tolpadi, A.A.; Lee, J.J.; Pedoia, V.; Majumdar, S. Deep Learning Predicts Total Knee Replacement from Magnetic Resonance Images. *Sci. Rep.* **2020**, *10*, 1–12, doi:10.1038/s41598-020-63395-9.

232. Morales, A.G.; Lee, J.J.; Caliva, F.; Iriondo, C.; Liu, F.; Majumdar, S.; Pedoia, V. Uncovering Associations between Data-Driven Learned QMRI Biomarkers and Chronic Pain. *Sci. Rep.* **2021**, *11*, 1–14, doi:10.1038/s41598-021-01111-x.

233. Ramzi, Z.; Starck, J.L.; Ciuciu, P. Density Compensated Unrolled Networks for Non-Cartesian MRI Reconstruction. *Proceedings - International Symposium on Biomedical Imaging* **2021**, 1443–1447, doi:10.1109/ISBI48211.2021.9433912.

234. Fabian, Z.; Soltanolkotabi, M. HUMUS-Net: Hybrid Unrolled Multi-Scale Network Architecture for Accelerated MRI Reconstruction. *ArXiv* **2022**, doi:10.48550/arxiv.2203.08213.

235. Eelbode, T.; Bertels, J.; Berman, M.; Vandermeulen, D.; Maes, F.; Bisschops, R.; Blaschko, M.B. Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index. *IEEE Trans. Med. Imaging* **2020**, *39*, 3679–3690, doi:10.1109/TMI.2020.3002417.

236. Chen, B.; Hsu, E.W. Noise Removal in Magnetic Resonance Diffusion Tensor Imaging. *Magn. Res. Med.* **2005**, *54*, 393–401, doi:10.1002/MRM.20582.

237. Noël, P.; Bammer, R.; Reinhold, C.; Haider, M.A. Parallel Imaging Artifacts in Body Magnetic Resonance Imaging. *Can. Assoc. Radiol. J.* **2009**, *60*, 91, doi:10.1016/J.CARJ.2009.02.036.

238. Liney, G.P.; Owen, S.C.; Beaumont, A.K.E.; Lazar, V.R.; Manton, D.J.; Beavis, A.W. Commissioning of a New Wide-Bore MRI Scanner for Radiotherapy Planning of Head and Neck Cancer. *Brit. J. Radiol.* **2013**, *86*, doi:10.1259/BJR.20130150/ASSET/IMAGES/LARGE/BJR-86-1027-D13150-G006.JPEG.

239. Thomas, M.S.; Greenwood, R.; Nolan, C.; Malcolm, P.N.; Toms, A.P. Optimizing MRI of Small Joints and Extremities. *Clin. Radiol.* **2014**, *69*, e414–e421, doi:10.1016/J.CRAD.2014.06.002.

240. Sekihara, K.; Kuroda, M.; Kohno, H. Image Restoration from Non-Uniform Magnetic Field Influence for Direct Fourier NMR Imaging. *Phys. Med. Biol* **1984**, *29*, 15–24.

241. Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *Proceedings - 2016 4th International Conference on 3D Vision* **2016**, 565–571, doi:10.48550/arxiv.1606.04797.

242. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. NnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nat. Methods* **2020**, *18*, 203–211, doi:10.1038/s41592-020-01008-z.

243. Saha, A.; Hosseinzadeh, M.; Huisman, H. End-to-End Prostate Cancer Detection in BpMRI via 3D CNNs: Effects of Attention Mechanisms, Clinical Priori and Decoupled False Positive Reduction. *Med Image Anal* **2021**, *73*, doi:10.1016/J.MEDIA.2021.102155.

244. Zibetti, M.V.W.; Johnson, P.M.; Sharafi, A.; Hammernik, K.; Knoll, F.; Regatte, R.R. Rapid Mono and Biexponential 3D-T1p Mapping of Knee Cartilage Using Variational Networks. *Sci. Rep.* **2020**, *10*, 1–14, doi:10.1038/s41598-020-76126-x.

245. Zibetti, M.V.W.; Knoll, F.; Regatte, R.R. Alternating Learning Approach for Variational Networks and Undersampling Pattern in Parallel MRI Applications. *IEEE Trans. Comput. Imaging* **2022**, *8*, 449–461, doi:10.1109/TCI.2022.3176129.

246. Ebrahimkhani, S.; Jaward, M.H.; Cicuttini, F.M.; Dharmaratne, A.; Wang, Y.; de Herrera, A.G.S. A Review on Segmentation of Knee Articular Cartilage: From Conventional Methods towards Deep Learning. *Artif. Intell. Med.* **2020**, *106*, doi:10.1016/J.ARTMED.2020.101851.

247. Bahadir, C.D.; Wang, A.Q.; Dalca, A. v.; Sabuncu, M.R. Deep-Learning-Based Optimization of the Under-Sampling Pattern in MRI. *IEEE Trans Comput Imaging* **2020**, *6*, 1139–1152, doi:10.1109/TCI.2020.3006727.

248. Razumov, A.; Rogov, O.Y.; Dylov, D. v. Optimal MRI Undersampling Patterns for Pathology Localization. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2022**, *13436 LNCS*, 768–779, doi:10.1007/978-3-031-16446-0_73/FIGURES/4.

249. Taylor, P.C. Update on the Diagnosis and Management of Early Rheumatoid Arthritis. *Clinical Medicine* **2020**, *20*, 561, doi:10.7861/CLINMED.2020-0727.

250. Kgoebane, K.; Ally, M.M.T.M.; Duim-Beytell, M.C.; Suleman, F.E. The Role of Imaging in Rheumatoid Arthritis. *SA J Radiol* **2018**, *22*, doi:10.4102/SAJR.V22I1.1316.

251. De Schepper, A.M.; De Beuckeleer, L.; Vandevenne, J.; Somville, J. Magnetic Resonance Imaging of Soft Tissue Tumors. *Eur Radiol* **2000**, *10*, 213–222, doi:10.1007/S003300050037/METRICS.

252. Zhou, Z.; Lu, Z.R. Gadolinium-Based Contrast Agents for MR Cancer Imaging. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2013**, *5*, 1, doi:10.1002/WNAN.1198.

253. Eshed, I.; Feist, E.; Althoff, C.E.; Hamm, B.; Konen, E.; Burmester, G.R.; Backhaus, M.; Hermann, K.G.A. Tenosynovitis of the Flexor Tendons of the Hand Detected by MRI: An Early Indicator of Rheumatoid Arthritis. *Rheumatology (Oxford)* **2009**, *48*, 887–891, doi:10.1093/RHEUMATOLOGY/KEP136.

254. Tamai, M.; Kawakami, A.; Uetani, M.; Fukushima, A.; Arima, K.; Fujikawa, K.; Iwamoto, N.; Aramaki, T.; Kamachi, M.; Nakamura, H.; et al. Magnetic Resonance Imaging (MRI) Detection of Synovitis and Bone Lesions of the Wrists and Finger Joints in Early-Stage Rheumatoid Arthritis: Comparison of the Accuracy of Plain MRI-Based Findings and Gadolinium-Diethylenetriamine Pentaacetic Acid-Enhanced MRI-Based Findings. *Mod Rheumatol* **2012**, *22*, 654–658, doi:10.1007/S10165-011-0575-8.

255. Boyd, A.S.; Zic, J.A.; Abraham, J.L. Gadolinium Deposition in Nephrogenic Fibrosing Dermopathy. *J Am Acad Dermatol* **2007**, *56*, 27–30, doi:10.1016/J.JAAD.2006.10.048.

256. Murata, N.; Gonzalez-Cuyar, L.F.; Murata, K.; Fligner, C.; Dills, R.; Hippe, D.; Maravilla, K.R. Macrocyclic and Other Non-Group 1 Gadolinium Contrast Agents Deposit Low Levels of Gadolinium in Brain and Bone Tissue: Preliminary Results from 9 Patients with Normal Renal Function. *Invest Radiol* **2016**, *51*, 447–453, doi:10.1097/RLI.0000000000000252.

257. Gulani, V.; Calamante, F.; Shellock, F.G.; Kanal, E.; Reeder, S.B. Gadolinium Deposition in the Brain: Summary of Evidence and Recommendations. *Lancet Neurol* **2017**, *16*, 564–570, doi:10.1016/S1474-4422(17)30158-8.

258. Tseng, H.Y.; Lee, H.Y.; Jiang, L.; Yang, M.H.; Yang, W. RetrieveGAN: Image Synthesis via Differentiable Patch Retrieval. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2020**, *12353 LNCS*, 242–257, doi:10.1007/978-3-030-58598-3_15.

259. Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; Sutskever, I. Zero-Shot Text-to-Image Generation. **2021**.

260. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst* **2012**, *25*.

261. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2013**, *8689 LNCS*, 818–833, doi:10.1007/978-3-319-10590-1_53.

262. Huang, H.; Yu, P.S.; Wang, C. An Introduction to Image Synthesis with Generative Adversarial Nets. **2018**.

263. Fulgeri, F.; Fabbri, M.; Alletto, S.; Calderara, S.; Cucchiara, R. Can Adversarial Networks Hallucinate Occluded People with a Plausible Aspect? *Computer Vision and Image Understanding* **2019**, *182*, 71–80, doi:10.1016/J.CVIU.2019.03.007.

264. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings* **2013**.

265. Esser, P.; Rombach, R.; Ommer, B. Taming Transformers for High-Resolution Image Synthesis. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2020**, 12868–12878, doi:10.1109/CVPR46437.2021.01268.

266. Xie, H.; Lei, Y.; Wang, T.; Roper, J.; Axente, M.; Bradley, J.D.; Liu, T.; Yang, X. Magnetic Resonance Imaging Contrast Enhancement Synthesis Using Cascade Networks with Local Supervision. *Med Phys* **2022**, *49*, 3278–3287, doi:10.1002/MP.15578.

267. Xu, C.; Xu, L.; Gao, Z.; Zhao, S.; Zhang, H.; Zhang, Y.; Du, X.; Zhao, S.; Ghista, D.; Liu, H.; et al. Direct Delineation of Myocardial Infarction without Contrast Agents Using a Joint Motion Feature Learning Architecture. *Med Image Anal* **2018**, *50*, 82–94, doi:10.1016/J.MEDIA.2018.09.001.

268. Wang, P.; Hu, S.; Wang, X.; Ge, Y.; Zhao, J.; Qiao, H.; Chang, J.; Dou, W.; Zhang, H. Synthetic MRI in Differentiating Benign from Metastatic Retropharyngeal Lymph Node: Combination with Diffusion-Weighted Imaging. *Eur Radiol* **2023**, *33*, 152–161, doi:10.1007/S00330-022-09027-4/FIGURES/3.

269. Kausar, T.; Kausar, A.; Ashraf, M.A.; Siddique, M.F.; Wang, M.; Sajid, M.; Siddique, M.Z.; Haq, A.U.; Riaz, I. SA-GAN: Stain Acclimation Generative Adversarial Network for Histopathology Image Analysis. *Applied Sciences 2022, Vol. 12, Page 288* **2021**, *12*, 288, doi:10.3390/APP12010288.

270. Cross-Zamirski, J.O.; Mouchet, E.; Williams, G.; Schönlieb, C.B.; Turkki, R.; Wang, Y. Label-Free Prediction of Cell Painting from Brightfield Images. *Scientific Reports 2022 12:1* **2022**, *12*, 1–13, doi:10.1038/s41598-022-12914-x.

271. Vassa, R.; Garg, A.; Omar, I.M. Magnetic Resonance Imaging of the Wrist and Hand. *Pol J Radiol* **2020**, *85*, e461, doi:10.5114/PJR.2020.99034.

272. Reeder, S.B.; McKenzie, C.A.; Pineda, A.R.; Yu, H.; Shimakawa, A.; Brau, A.C.; Hargreaves, B.A.; Gold, G.E.; Brittain, J.H. Water–Fat Separation with IDEAL Gradient-Echo Imaging. *Journal of Magnetic Resonance Imaging* **2007**, *25*, 644–652, doi:10.1002/JMRI.20831.

273. Østergaard, M.; Peterfy, C.G.; Bird, P.; Gandjbakhch, F.; Glinatsi, D.; Eshed, I.; Haavardsholm, E.A.; Lillegraven, S.; Bøyesen, P.; Ejbjerg, B.; et al. The OMERACT Rheumatoid Arthritis Magnetic Resonance Imaging (MRI) Scoring System: Updated Recommendations by the OMERACT MRI in Arthritis Working Group. *J Rheumatol* **2017**, *44*, 1706–1712, doi:10.3899/JRHEUM.161433.

274. Mattes, D.; Haynor, D.; Vesselle, H.; Lewellyn, T.; Eubank David Mattes, W.; Haynor, D.R.; Lewellyn, T.K.; Eubank, W.; multimodality, N.; Lewellen, T.K. Nonrigid Multimodality Image Registration. *https://doi.org/10.1117/12.431046* **2001**, *4322*, 1609–1620, doi:10.1117/12.431046.

275. Yaniv, Z.; Lowekamp, B.C.; Johnson, H.J.; Beare, R. SimpleITK Image-Analysis Notebooks: A Collaborative Environment for Education and Reproducible Research. *J Digit Imaging* **2018**, *31*, 290, doi:10.1007/S10278-017-0037-8.

276. Lowekamp, B.C.; Chen, D.T.; Ibáñez, L.; Blezek, D. The Design of SimpleITK. *Front Neuroinform* **2013**, *7*, 45, doi:10.3389/FNINF.2013.00045/BIBTEX.

277. Beare, R.; Lowekamp, B.; Yaniv, Z. Image Segmentation, Registration and Characterization in R with SimpleITK. *J Stat Softw* **2018**, *86*, doi:10.18637/JSS.V086.I08.

278. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2016**, *2017-January*, 5967–5976, doi:10.48550/arxiv.1611.07004.

279. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks.

280. Tomczak, A.; Gupta, A.; Ilic, S.; Navab, N.; Albarqouni, S. What Can We Learn About a Generated Image Corrupting Its Latent Representation? *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **2022**, *13436 LNCS*, 505–515, doi:10.1007/978-3-031-16446-0_48/TABLES/4.

281. De Winter, J.C.F. Using the Student's t-Test with Extremely Small Sample Sizes. *Practical Assessment, Research, and Evaluation Practical Assessment* **2013**, *18*, 10, doi:10.7275/e4r6-dj05.

282. Nahler, G. Bonferroni Correction. *Dictionary of Pharmaceutical Medicine* **2009**, 18–18, doi:10.1007/978-3-211-89836-9_140.

283. Shi, W.; Caballero, J.; Huszar, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2016**, *2016-December*, 1874–1883, doi:10.48550/arxiv.1609.05158.

284. Sugawara, Y.; Shiota, S.; Kiya, H. Checkerboard Artifacts Free Convolutional Neural Networks. *APSIPA Trans Signal Inf Process* **2019**, *8*, e9, doi:10.1017/ATSIP.2019.2.

285. Kamrul Hasan, S.M.; Linte, C.A. U-NetPlus: A Modified Encoder-Decoder U-Net Architecture for Semantic and Instance Segmentation of Surgical Instruments from Laparoscopic Images. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* **2019**, 7205–7211, doi:10.1109/EMBC.2019.8856791.

286. Guermazi, A.; Roemer, F.W.; Hayashi, D.; Crema, M.D.; Niu, J.; Zhang, Y.; Marra, M.D.; Katur, A.; Lynch, J.A.; El-Khoury, G.Y.; et al. Assessment of Synovitis with Contrast-Enhanced MRI Using a Whole-Joint Semiquantitative Scoring System in People with, or at High Risk of, Knee Osteoarthritis: The MOST Study. *Ann Rheum Dis* **2011**, *70*, 805–811, doi:10.1136/ARD.2010.139618.

287. Amini, B.; Murphy, W.A.; Haygood, T.M.; Kumar, R.; McEnery, K.W.; Madewell, J.E.; Mujtaba, B.M.; Wei, W.; Costelloe, C.M. Gadolinium-Based Contrast Agents Improve Detection of Recurrent Soft-Tissue Sarcoma at Mri. *Radiol Imaging Cancer* **2020**, *2*, doi:10.1148/RYCAN.2020190046/ASSET/IMAGES/LARGE/RYCAN.2020190046.FIG4.JPEG.

288. Crema, M.D.; Felson, D.T.; Roemer, F.W.; Niu, J.; Marra, M.D.; Zhang, Y.; Lynch, J.A.; El-Khoury, G.Y.; Lewis, C.E.; Guermazi, A. Peripatellar Synovitis: Comparison between Non-Contrast-Enhanced and Contrast-Enhanced MRI and Association with Pain. The MOST Study. *Osteoarthritis Cartilage* **2013**, *21*, 413–418, doi:10.1016/J.JOCA.2012.12.006.

289. Lee, J.H.; Dyke, J.P.; Ballon, D.; Ciombor, D.M.K.; Tung, G.; Aaron, R.K. Assessment of Bone Perfusion with Contrast-Enhanced Magnetic Resonance Imaging: Imaging of Bone

Marrow Edema Associated with Osteoarthritis and Avascular Necrosis. *Orthop Clin North Am* **2009**, *40*, 249, doi:10.1016/J.OCL.2008.12.003.

290. Salaffi, F.; Ceccarelli, L.; Carotti, M.; Di Carlo, M.; Polonara, G.; Facchini, G.; Golfieri, R.; Giovagnoni, A. Differentiation between Infectious Spondylodiscitis versus Inflammatory or Degenerative Spinal Changes: How Can Magnetic Resonance Imaging Help the Clinician? *Radiol Med* **2021**, *126*, 843, doi:10.1007/S11547-021-01347-7.

291. Tolpadi, A.A.; Bharadwaj, U.; Gao, K.T.; Bhattacharjee, R.; Gassert, F.G.; Luitjens, J.; Giesler, P.; Morshuis, J.N.; Fischer, P.; Hein, M.; et al. K2S Challenge: From Undersampled K-Space to Automatic Segmentation. *Bioengineering 2023, Vol. 10, Page 267* **2023**, *10*, 267, doi:10.3390/BIOENGINEERING10020267.

292. Zhao, R.; Yaman, B.; Zhang, Y.; Stewart, R.; Dixon, A.; Knoll, F.; Huang, Z.; Lui, Y.W.; Hansen, M.S.; Lungren, M.P. FastMRI+: Clinical Pathology Annotations for Knee and Brain Fully Sampled Multi-Coil MRI Data. **2021**.

293. Cross, M.; Smith, E.; Hoy, D.; Nolte, S.; Ackerman, I.; Fransen, M.; Bridgett, L.; Williams, S.; Guillemin, F.; Hill, C.L.; et al. The Global Burden of Hip and Knee Osteoarthritis: Estimates from the Global Burden of Disease 2010 Study. *Ann Rheum Dis* **2014**, *73*, 1323–1330, doi:10.1136/ANNRHEUMDIS-2013-204763.

294. Murphy, L.B.; Cisternas, M.G.; Pasta, D.J.; Helmick, C.G.; Yelin, E.H. Medical Expenditures and Earnings Losses Among US Adults With Arthritis in 2013. *Arthritis Care Res (Hoboken)* **2018**, *70*, 869–876, doi:10.1002/ACR.23425.

295. Lawrence, R.C.; Felson, D.T.; Helmick, C.G.; Arnold, L.M.; Choi, H.; Deyo, R.A.; Gabriel, S.; Hirsch, R.; Hochberg, M.C.; Hunder, G.G.; et al. Estimates of the Prevalence of Arthritis

and Other Rheumatic Conditions in the United States. Part II. *Arthritis Rheum* **2008**, *58*, 26–35, doi:10.1002/ART.23176.

296. Ostrander, R. V; Leddon, C.E.; Hackel, J.G.; O'grady, C.P.; Roth, C.A. Efficacy of Unloader Bracing in Reducing Symptoms of Knee Osteoarthritis. *Am J Orthop*.

297. Nguyen, L.C.L.; Lehil, M.S.; Bozic, K.J. Trends in Total Knee Arthroplasty Implant Utilization. *J Arthroplasty* **2015**, *30*, 739–742, doi:10.1016/J.ARTH.2014.12.009.

298. Inacio, M.C.S.; Paxton, E.W.; Graves, S.E.; Namba, R.S.; Nemes, S. Projected Increase in Total Knee Arthroplasty in the United States – an Alternative Projection Model. *Osteoarthritis Cartilage* **2017**, *25*, 1797–1803, doi:10.1016/J.JOCA.2017.07.022.

299. Tanzer, M.; Makhdom, A.M. Preoperative Planning in Primary Total Knee Arthroplasty. *Journal of the American Academy of Orthopaedic Surgeons* **2016**, *24*, 220–230, doi:10.5435/JAAOS-D-14-00332.

300. Sassoon, A.; Nam, D.; Nunley, R.; Barrack, R. Systematic Review of Patient-Specific Instrumentation in Total Knee Arthroplasty: New but Not Improved. *Clin Orthop Relat Res* **2015**, *473*, 151–158, doi:10.1007/S11999-014-3804-6.

301. Parvizi, J.; Nunley, R.M.; Berend, K.R.; Lombardi, A. V.; Ruh, E.L.; Clohisy, J.C.; Hamilton, W.G.; Della Valle, C.J.; Barrack, R.L. High Level of Residual Symptoms in Young Patients After Total Knee Arthroplasty. *Clin Orthop Relat Res* **2014**, *472*, 133, doi:10.1007/S11999-013-3229-7.

302. Chang, M.J.; Lim, H.; Lee, N.R.; Moon, Y.W. Diagnosis, Causes and Treatments of Instability Following Total Knee Arthroplasty. *Knee Surg Relat Res* **2014**, *26*, 61, doi:10.5792/KSRR.2014.26.2.61.

303. Kim, K.T.; Lee, S.; Ko, D.O.; Seo, B.S.; Jung, W.S.; Chang, B.K. Causes of Failure after Total Knee Arthroplasty in Osteoarthritis Patients 55 Years of Age or Younger. *Knee Surg Relat Res* **2014**, *26*, 13, doi:10.5792/KSRR.2014.26.1.13.

304. Raynauld, J.P.; Martel-Pelletier, J.; Berthiaume, M.J.; Abram, F.; Choquette, D.; Haraoui, B.; Beary, J.F.; Cline, G.A.; Meyer, J.M.; Pelletier, J.P. Correlation between Bone Lesion Changes and Cartilage Volume Loss in Patients with Osteoarthritis of the Knee as Assessed by Quantitative Magnetic Resonance Imaging over a 24-Month Period. *Ann Rheum Dis* **2008**, *67*, 683–688, doi:10.1136/ARD.2007.073023.

305. Raynauld, J.P.; Martel-Pelletier, J.; Dorais, M.; Haraoui, B.; Choquette, D.; Abram, F.; Beaulieu, A.; Bessette, L.; Morin, F.; Wildi, L.M.; et al. Total Knee Replacement as a Knee Osteoarthritis Outcome: Predictors Derived from a 4-Year Long-Term Observation Following a Randomized Clinical Trial Using Chondroitin Sulfate. *Cartilage* **2013**, *4*, 219, doi:10.1177/1947603513483547.

306. Everhart, J.S.; Abouljoud, M.M.; Kirven, J.C.; Flanigan, D.C. Full-Thickness Cartilage Defects Are Important Independent Predictive Factors for Progression to Total Knee Arthroplasty in Older Adults with Minimal to Moderate Osteoarthritis: Data from the Osteoarthritis Initiative. *J Bone Joint Surg Am* **2019**, *101*, 56–63, doi:10.2106/JBJS.17.01657.

307. Riddle, D.L.; Kong, X.; Jiranek, W.A. Two-Year Incidence and Predictors of Future Knee Arthroplasty in Persons with Symptomatic Knee Osteoarthritis: Preliminary Analysis of Longitudinal Data from the Osteoarthritis Initiative. *Knee* **2009**, *16*, 494, doi:10.1016/J.KNEE.2009.04.002.

308. Hawker, G.A.; Guan, J.; Croxford, R.; Coyte, P.C.; Glazier, R.H.; Harvey, B.J.; Wright, J.G.; Williams, J.I.; Badley, E.M. A Prospective Population-Based Study of the Predictors of Undergoing Total Joint Arthroplasty. *Arthritis Rheum* **2006**, *54*, 3212–3220, doi:10.1002/ART.22146.

309. Lewis, J.R.; Dhaliwal, S.S.; Zhu, K.; Prince, R.L. A Predictive Model for Knee Joint Replacement in Older Women. *PLoS One* **2013**, *8*, e83665, doi:10.1371/JOURNAL.PONE.0083665.

310. Yu, D.; Jordan, K.P.; Snell, K.I.E.; Riley, R.D.; Bedson, J.; Edwards, J.J.; Mallen, C.D.; Tan, V.; Ukachukwu, V.; Prieto-Alhambra, D.; et al. Clinical Science: Development and Validation of Prediction Models to Estimate Risk of Primary Total Hip and Knee Replacements Using Data from the UK: Two Prospective Open Cohorts Using the UK Clinical Practice Research Datalink. *Ann Rheum Dis* **2019**, *78*, 91, doi:10.1136/ANNRHEUMDIS-2018-213894.

311. Wang, T.; Leung, K.; Cho, K.; Chang, G.; Deniz, C.M. Total Knee Replacement Prediction Using Structural MRIs and 3D Convolutional Neural Networks. In Proceedings of the International Conference on Medical Imaging with Deep Learning - Extended Abstract Track; 2019; p. 79.

312. Akobeng, A.K.; Akobeng, A.K. Understanding Diagnostic Tests 3: Receiver Operating Characteristic Curves. *Acta Paediatr* **2007**, *96*, 644–647, doi:10.1111/J.1651-2227.2006.00178.X.

313. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2015**, *2016-December*, 770–778, doi:10.48550/arxiv.1512.03385.

314. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2016**, *2017-January*, 2261–2269, doi:10.48550/arxiv.1608.06993.

315. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vis* **2014**, *115*, 211–252, doi:10.48550/arxiv.1409.0575.

316. Tiulpin, A.; Klein, S.; Bierma-Zeinstra, S.M.A.; Thevenot, J.; Rahtu, E.; Meurs, J. van; Oei, E.H.G.; Saarakkala, S. Multimodal Machine Learning-Based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data. *Scientific Reports 2019 9:1* **2019**, *9*, 1–11, doi:10.1038/s41598-019-56527-3.

317. Antony, J.; McGuinness, K.; O'Connor, N.E.; Moran, K. Quantifying Radiographic Knee Osteoarthritis Severity Using Deep Convolutional Neural Networks. *Proceedings - International Conference on Pattern Recognition* **2016**, *0*, 1195–1200, doi:10.48550/arxiv.1609.02469.

318. Norman, B.; Pedoia, V.; Noworolski, A.; Link, T.M.; Majumdar, S. Applying Densely Connected Convolutional Neural Networks for Staging Osteoarthritis Severity from Plain Radiographs. *J Digit Imaging* **2019**, *32*, 471–477, doi:10.1007/S10278-018-0098-3.

319. Heidari, B. Knee Osteoarthritis Prevalence, Risk Factors, Pathogenesis and Features: Part I. *Caspian J Intern Med* **2011**, *2*, 205.

320. Cooper, C.; Snow, S.; Mcalindon, T.E.; Kellingray, S.; Stuart, B.; Coggon, D.; Dieppe, P.A.; McAlindon, T.E. RISK FACTORS FOR THE INCIDENCE AND PROGRESSION OF

RADIOGRAPHIC KNEE OSTEOARTHRITIS. *Arthritis Rheum* **2000**, *43*, 995–1000, doi:10.1002/1529-0131.

321. Pisters, M.F.; Veenhof, C.; van Dijk, G.M.; Heymans, M.W.; Twisk, J.W.R.; Dekker, J. The Course of Limitations in Activities over 5 Years in Patients with Knee and Hip Osteoarthritis with Moderate Functional Limitations: Risk Factors for Future Functional Decline. *Osteoarthritis Cartilage* **2012**, *20*, 503–510, doi:10.1016/J.JOCA.2012.02.002.
322. Sharma, L.; Cahue, S.; Song, J.; Hayes, K.; Pai, Y.C.; Dunlop, D. Physical Functioning over Three Years in Knee Osteoarthritis: Role of Psychosocial, Local Mechanical, and Neuromuscular Factors. *Arthritis Rheum* **2003**, *48*, 3359–3370, doi:10.1002/ART.11420.
323. Dietrich, S.; Floegel, A.; Troll, M.; Kühn, T.; Rathmann, W.; Peters, A.; Sookthai, D.; Von Bergen, M.; Kaaks, R.; Adamski, J.; et al. Random Survival Forest in Practice: A Method for Modelling Complex Metabolomics Data in Time to Event Analysis. *Int J Epidemiol* **2016**, *45*, 1406–1420, doi:10.1093/IJE/DYW145.
324. Chen, C.; Liaw, A. Using Random Forest to Learn Imbalanced Data. *University of California, Berkeley* **2004**.
325. Hara, K.; Kataoka, H.; Satoh, Y. Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet? *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2017**, 6546–6555, doi:10.48550/arxiv.1711.09577.
326. Youden, W.J. INDEX FOR RATING DIAGNOSTIC TESTS. *Cancer* **1950**, *3*, 32–35, doi:10.1002/1097-0142.
327. Salzberg, S.L. On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach. *Data Min Knowl Discov* **1997**, *1*, 317–328, doi:10.1023/A:1009752403260/METRICS.

328. Dietterich, T.G. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Comput* **1998**, *10*, 1895–1923, doi:10.1162/089976698300017197.

329. Lawson, R. Small Sample Confidence Intervals for the Odds Ratio. *Commun Stat Simul Comput* **2007**, *33*, 1095–1113, doi:10.1081/SAC-200040691.

330. Upton, G.J.G. Fisher's Exact Test. *J R Stat Soc Ser A Stat Soc* **2016**, *155*, 395–402, doi:10.2307/2982890.

331. Menashe, L.; Hirko, K.; Losina, E.; Kloppenburg, M.; Zhang, W.; Li, L.; Hunter, D.J. The Diagnostic Performance of MRI in Osteoarthritis: A Systematic and Meta-Analysis. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* **2012**, *20*, 13, doi:10.1016/J.JOCA.2011.10.003.

332. Collins, J.E.; Losina, E.; Nevitt, M.C.; Roemer, F.W.; Guermazi, A.; Lynch, J.A.; Katz, J.N.; Kent Kwoh, C.; Kraus, V.B.; Hunter, D.J. Semi-Quantitative Imaging Biomarkers of Knee Osteoarthritis Progression: Data from the FNIH OA Biomarkers Consortium. *Arthritis Rheumatol* **2016**, *68*, 2422, doi:10.1002/ART.39731.

333. Khan, T.; Alvand, A.; Prieto-Alhambra, D.; Culliford, D.J.; Judge, A.; Jackson, W.F.; Scammell, B.E.; Arden, N.K.; Price, A.J. ACL and Meniscal Injuries Increase the Risk of Primary Total Knee Replacement for Osteoarthritis: A Matched Case-Control Study Using the Clinical Practice Research Datalink (CPRD). *Br J Sports Med* **2019**, *53*, 965–968, doi:10.1136/BJSPORTS-2017-097762.

334. Simon, D.; Mascarenhas, R.; Saltzman, B.M.; Rollins, M.; Bach, B.R.; MacDonald, P. The Relationship between Anterior Cruciate Ligament Injury and Osteoarthritis of the Knee. *Adv Orthop* **2015**, *2015*, doi:10.1155/2015/928301.

335. Diederichs, G.; Issever, A.S.; Scheffler, S. MR Imaging of Patellar Instability: Injury Patterns and Assessment of Risk Factors1. *RadioGraphics* **2010**, *30*, 961–981, doi:10.1148/RG.304095755.

336. Moatshe, G.; Dornan, G.J.; Ludvigsen, T.; Løken, S.; LaPrade, R.F.; Engebretsen, L. High Prevalence of Knee Osteoarthritis at a Minimum 10-Year Follow-up after Knee Dislocation Surgery. *Knee Surgery, Sports Traumatology, Arthroscopy* **2017**, *25*, 3914–3922, doi:10.1007/S00167-017-4443-8/TABLES/6.

337. Figueroa, D.; Garín, A.; Figueroa, F. Total Knee Replacement in Patients with Osteoarthritis and Concomitant Inveterate Patellar Dislocation. *Arthroplast Today* **2019**, *5*, 68, doi:10.1016/J.ARTD.2018.04.003.

338. Souza, R. *Advances in MRI of the Knee for Osteoarthritis*; Majumdar, S., Ed.; World Scientific, 2010; ISBN 978-9-81-427170-7.

339. Spina, A.A. The Plantaris Muscle: Anatomy, Injury, Imaging, and Treatment. *J Can Chiropr Assoc* **2007**, *51*, 158.

340. Zetaruk, M.; Hyman, J. *Leg Injuries*; Frontera, W.R., Herring, S.A., Micheli, L.J., Silver, J.K., Young, T.P., Eds.; W.B. Saunders, 2007; ISBN 9781437710632.

Appendix A - Supplementary Information to Chapter 6

Supp. Table A.1 Acquisition times for MAPSS at tested R. Acquisition times for the full MAPSS sequence if the proposed undersampling patterns were implemented for $T_{1\rho}$ and T_2 preparation and image acquisition. If acquisition times were desired solely for T_2 weighted images acquired from MAPSS, all acquisition times in this table would need to be multiplied by 0.571.

Acquisition Type	Knee	Hip	Lumbar Spine
Ground Truth	5 minutes, 53 seconds	8 minutes, 15 seconds	5 minutes, 17 seconds
R=2	2 minutes, 56 seconds	4 minutes, 8 seconds	2 minutes, 39 seconds
R=3	1 minute, 58 seconds	2 minutes, 45 seconds	1 minute, 46 seconds
R=4	1 minute, 28 seconds	2 minutes, 4 seconds	1 minute, 19 seconds
R=6	59 seconds	1 minute, 23 seconds	53 seconds
R=8	44 seconds	1 minute, 2 seconds	40 seconds
R=10	35 seconds	50 seconds	32 seconds
R=12	29 seconds	41 seconds	26 seconds

Supp. Table A.2 Information for cross-validation splits in knee, hip and lumbar spine datasets. Training, validation, and test data splits of MAPSS acquisitions for all three folds, by patient and by total number of scans. In lumbar spine, total scans exceeded number of patients because some patients had spines scanned multiple times, whereas for knee and hip, both knees or both hips of some patients may have been scanned. To prevent data leakage, all scans of a particular patient was only placed into one of the three datasets. Unless otherwise specified, all results in paper are reported from fold 1 split; additional splits 2 and 3 were made to assess robustness of T_2 quantification performance and texture retention to different datasets used, as described in subsequent Supplementary information tables.

Anatomy	Fold	Training			Validation			Test		
		Patients	Scans	Slices	Patients	Scans	Slices	Patients	Scans	Slices
Knee	1	144	265	5,591	50	91	1,952	50	90	1,928
	2	144	262	5,619	50	93	1,960	50	90	1,892
	3	144	259	5,480	50	81	1,739	50	106	2,252
Hip	1	39	59	1,533	15	15	390	13	15	390
	2	39	59	1,533	12	15	390	13	15	390
	3	40	59	1,533	13	15	390	15	15	390
Lumbar spine	1	13	14	112	4	5	42	4	5	40
	2	15	16	130	4	4	32	4	4	32
	3	14	16	130	4	4	32	4	4	32

Supp. Table A.3 Network was trained with 4 loss function components: L_1 loss across slice, L_1 loss in tissue of interest, SSIM loss, and VGG-19 feature-based loss function. Weightings of those components were λ_{L_1}, $\lambda_{L_{1,\phi}}$, λ_{SSIM}, and $\lambda_{Feat.}$, respectively. Ablation study results are shown at R=8, with "X's" in table noting that the given loss function component was not ablated in that trained pipeline version. If loss component was ablated, its λ was set to 0; if not, it remained at its value optimized through the constrained hyperparameter search. In addition, the RNN portion of the network was ablated while keeping all loss components constant, while a separate pipeline version was trained with all layer depths being half of what is depicted in Fig. 1. "X's" mark ablations in which the RNN was part of the network design and/or when the reduced overall number of network parameters was used in the architecture. All metrics are calculated in the tissues of interest (cartilage or IVD) and calculated at a cartilage compartment/IVD level. NRMSEs of ablated and original pipelines are presented ±1 s.d., and significance of Pearson's r is denoted as follows: * P < 0.05, ** P < 0.01, *** P < 0.001 (knee: n=90; hip: n=15; lumbar spine: n=5). The top 4 ablations in NRMSE and Pearson's r are bolded. In knee and hip pipelines, all loss function components are necessary to obtain among the strongest performing models in both metrics. In the lumbar spine, substantial performance improvements are seen with ablated RNN.

λ_{L_1}	$\lambda_{L_{1,\phi}}$	λ_{SSIM}	$\lambda_{Feat.}$	RNN	Reduced Params	Knee		Hip		Lumbar Spine	
						NRMSE (%)	Pearson's r	NRMSE (%)	Pearson's r	NRMSE (%)	Pearson's r
			X	X		10900 ± 1740	0.006	12100 ± 3440	-0.006	43700 ± 14800	-0.008
		X		X		17.1 ± 4.62	0.355***	24.1 ± 7.27	0.509***	26.3 ± 18.1	0.696***
		X	X	X		13.9 ± 2.05	0.432***	12.0 ± 2.38	0.468***	31.4 ± 19.8	0.676***
	X			X		12.5 ± 7.75	**0.598*** **	10.8 ± 3.62	0.557***	19.0 ± 7.79	0.671***
	X		X	X		10.9 ± 6.09	0.587***	13.5 ± 8.41	0.380***	67700 ± 26700	0.002
	X	X		X		11.9 ± 7.35	**0.613*** **	2800 ± 2570	0.196***	30.4 ± 24.2	0.703***
	X	X	X	X		**9.63 ± 4.92**	**0.617*** **	**9.45 ± 3.0**	**0.599*** **	**12.7 ± 3.75**	**0.746*** **
X				X		22.9 ± 3.36	0.315**	21.9 ± 6.49	0.253***	37.2 ± 21.2	0.552***
X			X	X		14.0 ± 2.41	0.295***	16.7 ± 5.14	0.198***	37600 ± 13100	-0.003
X		X		X		17.4 ± 5.3	0.419***	12.5 ± 2.47	0.494***	13.1 ± 3.38	0.747***
X		X	X	X		14.9 ± 3.61	0.317***	12.4 ± 2.61	0.379***	59.5 ± 32.3	0.245***
X	X			X		10.1 ± 4.18	0.552***	22.6 ± 5.51	0.529***	**12.1 ± 3.86**	0.720***
X	X		X	X		11.6 ± 4.78	0.563***	19.0 ± 4.67	0.427***	24200 ± 8850	-0.008
X	X	X		X		**8.97 ± 3.6**	0.578***	11.9 ± 3.66	0.569***	13.6 ± 6.88	**0.747*** **
X	X	X	X	X		**9.15 ± 2.56**	0.586***	**6.94 ± 1.94**	**0.596*** **	13.3 ± 4.48	**0.742*** **
X	X	X	X	X	X	**9.59 ± 3.83**	0.574***	**6.33 ± 1.33**	**0.588*** **	**13.0 ± 2.63**	0.723***
X	X	X	X			10.1 ± 4.42	**0.609*** **	**6.98 ± 1.45**	**0.558*** **	**12.0 ± 3.07**	**0.742*** **

Supp. Table A.4 Loss function ablation study T2 value mean errors. Additional details on loss function ablation study. NRMSE was calculated across all cartilage or IVD on a patient level for all ablated models, and these NRMSEs were multiplied by corresponding mean reference T2 values to convert T2 quantification error rates into T2 estimation errors. Estimation errors are reported mean ±1 s.d. [ms] (knee: n=90; hip: n=15; lumbar spine: n=5).

λ_{L_1}	$\lambda_{L_{1,\phi}}$	λ_{SSIM}	$\lambda_{Feat.}$	**RNN**	**Reduced Params**	**Knee**	**Hip**	**Lumbar Spine**
			X	X		3440.0 ± 864.0	3540.0 ± 1060.0	20500.0 ± 10100.0
		X		X		5.4 ± 1.19	7.03 ± 2.17	12.4 ± 9.51
		X	X	X		4.39 ± 0.6	3.52 ± 0.902	14.8 ± 10.6
	X			X		3.94 ± 2.6	3.17 ± 1.18	8.92 ± 3.7
	X		X	X		3.43 ± 2.05	3.93 ± 2.92	31800.0 ± 17000.0
	X	X		X		3.77 ± 2.44	817.0 ± 823.0	14.3 ± 12.6
	X	X	X	X		3.04 ± 1.67	2.76 ± 1.07	5.94 ± 2.1
X				X		7.24 ± 1.18	6.41 ± 2.32	17.5 ± 11.9
X			X	X		4.43 ± 0.601	4.87 ± 1.79	17700.0 ± 8830.0
X		X		X		5.49 ± 1.3	3.66 ± 0.782	6.17 ± 0.821
X		X	X	X		4.7 ± 0.811	3.63 ± 0.676	27.9 ± 18.4
X	X			X		3.2 ± 1.39	6.59 ± 2.02	5.68 ± 1.43
X	X		X	X		3.67 ± 1.54	5.55 ± 1.65	11400.0 ± 5910.0
X	X	X		X		2.84 ± 1.21	3.46 ± 1.31	6.37 ± 3.59
X	X	X	X	X		2.89 ± 0.906	2.03 ± 0.668	6.24 ± 2.46
X	X	X	X	X	X	3.03 ± 1.29	1.85 ± 0.418	6.11 ± 1.15
X	X	X	X			3.21 ± 1.52	2.04 ± 0.488	5.62 ± 1.26

Supp. Table A.5 ROI and global correlations between predicted and ground truth T_2 maps. Pearson's r between predicted and ground truth T_2 maps for proposed model trained with full 4-component loss function, proposed model trained with just the ROI-specific loss component ablated ($\lambda_{1,\phi}$=0), and proposed model trained with an ordinary loss function of L_1 and SSIM ($\lambda_{1,\phi}$=0, λ_{Feat}=0) as part of ablation study. Correlations are also provided between predictions and ground truth for 3 state-of-the-art models. Significance of Pearson's r is denoted as follows: * P < 0.05, ** P < 0.01, *** P < 0.001 (knee: n=90; hip: n=15; lumbar spine: n=5). For any given R, the strongest correlation within the ROI is highlighted in red, whereas the strongest correlation globally is highlighted in blue. Across the hip and knee pipelines, each of which had large datasets available for training, correlations are strongest within cartilage ROIs for the proposed pipelines across all R, while for all R, state-of-the-art DL pipelines (MANTIS, MANTIS-GAN) exhibited stronger correlations globally to ground truth. Similarly, when the ROI-specific loss function was ablated, for nearly all tested R in hip and knee, correlations became stronger globally than for the proposed pipelines. This is indicative of successful training and the role of the ROI-specific loss function: with a sufficiently large training set, it improves results within cartilage ROIs at the expense of global performance, allowing for ROI-specific model optimization. These trends were inconsistent inn the lumbar spine, likely owing to the very small dataset size that added randomness to the training process; some results thus may be a result of more complete training rather than the specific utility of the ROI-specific loss. If trained with a larger dataset, the lumbar spine results would be expected to mirror the knee and hip.

Tissue	R	Tissue	Proposed Model	$\lambda_{1,\phi}$=0	$\lambda_{1,\phi}$=0, λ_{Feat}=0	MANTIS	MANTIS-GAN	CS
Knee	2	Cartilage	0.748***	0.658***	0.685***	0.587***	0.611***	0.620***
		Global	0.667***	0.675***	0.677***	0.700***	0.703***	0.581***
	3	Cartilage	0.695***	0.491***	0.573***	0.467***	0.502***	0.559***
		Global	0.383***	0.618***	0.683***	0.681***	0.693***	0.566***
	4	Cartilage	0.651***	0.376***	0.558***	0.451***	0.467***	0.486***
		Global	0.149***	0.575***	0.693***	0.672***	0.670***	0.554***
	6	Cartilage	0.612***	0.465***	0.442***	0.397***	0.378*	0.445***
		Global	0.487***	0.663***	0.687***	0.667***	0.659***	0.547***
	8	Cartilage	0.585***	0.383***	0.450***	0.352***	0.364**	0.410***
		Global	0.440***	0.661***	0.672***	0.659***	0.654***	0.548***
	10	Cartilage	0.555***	0.124**	0.346***	0.327***	0.333***	0.386***
		Global	0.165***	0.540***	0.623***	0.651***	0.653***	0.552***
	12	Cartilage	0.491***	0.339***	0.396***	0.290***	0.287***	0.381***
		Global	0.559***	0.613***	0.630***	0.659***	0.656***	0.557***
Hip	2	Cartilage	0.794***	0.562***	0.655***	0.716***	0.514***	0.310***
		Global	0.683***	0.713***	0.727***	0.755***	0.699***	0.534***
	3	Cartilage	0.705***	0.541***	0.612***	0.596***	0.372***	0.332***
		Global	0.607***	0.669***	0.660***	0.723***	0.659***	0.549***
	4	Cartilage	0.646***	0.459***	0.540***	0.510***	0.333***	0.339***
		Global	0.609***	0.594***	0.610***	0.707***	0.641***	0.562***
	6	Cartilage	0.587***	0.458***	0.480***	0.382***	0.321***	0.334***
		Global	0.555***	0.595***	0.643***	0.690***	0.634***	0.573***

Tissue	R	Tissue	Proposed Model	$\lambda_{1,\phi}=0$	$\lambda_{1,\phi}=0$, $\lambda_{Feat}=0$	MANTIS	MANTIS-GAN	CS
Hip	8	Cartilage	0.598***	0.344***	0.437***	0.334***	0.237***	0.347***
		Global	0.630***	0.577***	0.627***	0.684***	0.660***	0.574***
	10	Cartilage	0.558***	0.355***	0.434***	0.279***	0.268***	0.335***
		Global	0.412***	0.429***	0.551***	0.686***	0.615***	0.569***
	12	Cartilage	0.517***	0.427***	0.423***	0.280***	0.228***	0.349***
		Global	0.619***	0.625***	0.622***	0.682***	0.622***	0.578***
Lumbar Spine	2	IVDs	0.884***	0.850***	0.879***	0.784***	0.785***	0.802***
		Global	0.836***	0.673***	0.772***	0.816***	0.821***	0.812***
	3	IVDs	0.832***	0.823***	0.846***	0.717***	0.712***	0.777***
		Global	0.797***	0.707***	0.711***	0.784***	0.786***	0.788***
	4	IVDs	0.819***	0.804***	0.827***	0.680***	0.671***	0.723***
		Global	0.783***	0.696***	0.743***	0.771***	0.774***	0.772***
	6	IVDs	0.771***	0.764***	0.761***	0.660***	0.658***	0.728***
		Global	0.766***	0.737***	0.712***	0.764***	0.770***	0.760***
	8	IVDs	0.742***	0.245***	0.747***	0.631***	0.645***	0.695***
		Global	0.749***	0.664***	0.720***	0.756***	0.757***	0.752***
	10	IVDs	0.672***	0.651***	0.698***	0.647***	0.636***	0.648***
		Global	0.728***	0.707***	0.669***	0.762***	0.761***	0.747***
	12	IVDs	0.643***	0.581***	0.654***	0.651***	0.614***	0.586***
		Global	0.707***	0.661***	0.686***	0.760***	0.762***	0.746***

Supp. Table A.6 T_2 value equivalents of quantification errors in cartilage, IVDs across all models. Additional details on model performance from R=2 through R=12 within cartilage compartments and at disc levels. As in Supp. Table A.4**,** NRMSE was calculated across cartilage or IVD compartment on a patient level for all ablated models, and these NRMSEs were multiplied by corresponding mean reference T2 values to convert quantification error rates into T2 estimation errors. Estimation errors are reported mean ±1 s.d. [ms] (knee: n=90; hip: n=15; lumbar spine: n=5).

Tissue	R	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
Knee Cartilage	2	1.75 ± 0.449	1.92 ± 1.1	1.5 ± 0.62	4.56 ± 0.631	4.28 ± 0.726	2.84 ± 0.951
	3	2.06 ± 0.773	2.27 ± 1.07	2.02 ± 0.884	5.23 ± 0.694	4.79 ± 0.652	3.15 ± 1.08
	4	2.38 ± 1.01	3.02 ± 1.81	2.39 ± 1.08	5.24 ± 0.869	4.95 ± 1.03	3.74 ± 1.26
	6	2.56 ± 0.933	3.38 ± 2.18	2.67 ± 1.21	4.82 ± 0.709	5.16 ± 0.903	3.93 ± 1.44
	8	2.82 ± 0.93	3.03 ± 1.29	3.21 ± 1.52	5.28 ± 0.673	5.46 ± 0.689	4.1 ± 1.41
	10	3.09 ± 1.14	3.21 ± 1.26	2.95 ± 1.24	5.55 ± 0.654	5.29 ± 0.822	4.26 ± 1.38
	12	3.37 ± 0.822	3.14 ± 1.3	3.31 ± 1.21	5.77 ± 1.0	6.48 ± 1.22	4.27 ± 1.46
Hip Cartilage	2	1.16 ± 0.34	1.2 ± 0.349	1.11 ± 0.229	1.34 ± 0.287	2.4 ± 0.327	4.31 ± 0.745
	3	1.91 ± 0.516	1.64 ± 0.519	1.54 ± 0.332	1.87 ± 0.363	2.93 ± 0.382	3.74 ± 0.814
	4	1.8 ± 0.348	1.8 ± 0.525	1.71 ± 0.273	2.14 ± 0.465	2.91 ± 0.467	3.42 ± 0.593
	6	2.37 ± 0.598	2.4 ± 0.676	2.19 ± 0.484	2.52 ± 0.643	2.83 ± 0.497	3.43 ± 0.559

Tissue	R	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
Hip Cart.	8	2.04 ± 0.676	1.85 ± 0.418	2.04 ± 0.488	2.98 ± 0.789	3.52 ± 0.767	3.04 ± 0.681
	10	2.63 ± 0.873	2.37 ± 0.4	2.54 ± 1.22	2.85 ± 0.637	3.05 ± 0.563	2.96 ± 0.685
	12	2.26 ± 0.499	2.42 ± 0.737	2.14 ± 0.433	2.85 ± 0.614	3.35 ± 0.685	3.0 ± 0.748
Lumbar Spine IVDs	2	3.15 ± 0.602	3.22 ± 0.571	2.28 ± 0.4	4.12 ± 1.06	4.2 ± 1.06	4.74 ± 1.93
	3	4.66 ± 0.854	4.11 ± 0.434	3.35 ± 0.478	5.17 ± 0.931	5.29 ± 1.14	4.45 ± 1.14
	4	4.85 ± 0.736	4.57 ± 0.859	3.48 ± 0.473	5.7 ± 1.17	5.91 ± 1.29	5.3 ± 1.65
	6	5.68 ± 1.74	5.75 ± 1.24	4.84 ± 1.3	7.32 ± 2.28	5.7 ± 1.07	5.62 ± 1.63
	8	6.28 ± 0.967	6.11 ± 1.15	5.62 ± 1.26	6.22 ± 1.27	5.96 ± 0.983	6.02 ± 1.53
	10	7.18 ± 0.725	6.95 ± 0.705	6.93 ± 1.1	6.48 ± 1.56	6.19 ± 1.29	7.04 ± 2.35
	12	8.48 ± 1.18	10.8 ± 2.82	8.84 ± 2.24	6.94 ± 2.18	6.62 ± 1.53	11.7 ± 7.04

Supp. Table A.7 T_2 quantification errors in knee cartilage compartments for 6 tested models. Performances of all methods – our ROI-specific loss approaches, other DL and DL/model-based approaches, and a CS approach in predicting T_2 maps in knee cartilage. NRMSEs are reported ±1 s.d., with the top performing model in each cartilage compartment at a given R shown in bold (n=16). Top performing pipelines were all pipelines with ROI-specific loss functions used in training, particularly with our full pipeline and its no RNN version being strongest. In the lateral and medial femoral condyles, T_2 quantification performance was below clinically significant thresholds for all tested R of the full pipeline, and for nearly all tested R for the no RNN pipeline.

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
2	Lateral Femoral Condyle	5.76 ± 2.63	8.95 ± 10.7	**5.67 ± 4.72**	13.9 ± 4.21	13.1 ± 4.29	11.5 ± 9.38
	Lateral Tibial Condyle	6.02 ± 2.35	6.9 ± 4.56	**5.49 ± 3.72**	17.6 ± 7.43	17.9 ± 6.71	9.56 ± 6.7
	Medial Femoral Condyle	5.31 ± 2.06	4.54 ± 1.75	**4.18 ± 1.69**	15.2 ± 6.93	14.2 ± 6.71	12.2 ± 8.76
	Medial Tibial Condyle	8.1 ± 4.94	10.1 ± 12.4	**7.73 ± 6.48**	22.5 ± 7.98	20.6 ± 6.68	14.9 ± 7.87
	Trochlear	7.19 ± 4.45	6.56 ± 5.32	**5.14 ± 2.7**	16.6 ± 5.3	17.1 ± 8.73	5.23 ± 3.26
	Patellar	4.08 ± 1.24	4.69 ± 1.55	**3.52 ± 1.38**	12.3 ± 2.89	11.2 ± 3.95	4.02 ± 2.15
	All Cartilage	5.52 ± 1.25	6.07 ± 3.21	**4.76 ± 1.78**	14.4 ± 2.85	13.5 ± 3.3	8.92 ± 3.2
3	Lateral Femoral Condyle	**7.37 ± 5.35**	9.74 ± 8.88	8.25 ± 7.43	16.3 ± 4.71	14.9 ± 5.24	14.4 ± 10.4
	Lateral Tibial Condyle	8.01 ± 3.83	7.99 ± 5.36	**7.41 ± 4.7**	18.8 ± 5.91	17.1 ± 6.03	10.1 ± 4.98
	Medial Femoral Condyle	**5.77 ± 2.19**	6.46 ± 2.79	6.33 ± 2.74	17.5 ± 3.18	15.6 ± 3.79	13.6 ± 7.82
	Medial Tibial Condyle	10.6 ± 8.22	10.9 ± 11.4	**10.2 ± 9.52**	22.9 ± 7.21	19.9 ± 7.21	16.1 ± 8.27
	Trochlear	7.13 ± 4.43	7.55 ± 5.3	**6.71 ± 4.26**	20.4 ± 8.83	18.9 ± 8.3	6.02 ± 2.4
	Patellar	5.03 ± 1.39	5.29 ± 1.49	**4.32 ± 1.79**	14.1 ± 4.94	12.9 ± 3.81	4.65 ± 2.18
	All Cartilage	6.52 ± 2.17	7.18 ± 3.08	**6.39 ± 2.59**	16.5 ± 3.43	15.1 ± 2.89	9.92 ± 3.23
4	Lateral Femoral Condyle	**10.7 ± 10.1**	17.4 ± 20.8	**10.7 ± 9.41**	17.2 ± 5.44	15.6 ± 6.26	15.1 ± 10.5
	Lateral Tibial Condyle	9.2 ± 4.64	12.3 ± 7.56	**8.98 ± 5.28**	20.5 ± 6.87	18.9 ± 5.65	12.7 ± 7.0
	Medial Femoral Condyle	**7.16 ± 2.85**	11.8 ± 5.28	8.44 ± 4.67	17.7 ± 5.0	17.9 ± 4.71	16.1 ± 8.94
	Medial Tibial Condyle	**11.3 ± 10.2**	16.6 ± 17.7	12.0 ± 11.2	23.6 ± 7.58	21.9 ± 6.81	19.0 ± 9.19
	Trochlear	7.4 ± 4.97	8.88 ± 8.47	**7.03 ± 5.1**	19.3 ± 10.3	21.7 ± 17.6	8.15 ± 2.97
	Patellar	5.11 ± 1.26	**4.61 ± 1.73**	4.76 ± 1.99	13.3 ± 3.5	12.3 ± 4.45	4.86 ± 1.79
	All Cartilage	**7.54 ± 2.96**	9.56 ± 5.47	7.56 ± 3.19	16.6 ± 3.73	15.7 ± 4.5	11.8 ± 3.73

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
6	Lateral Femoral Condyle	**10.9 ± 8.25**	20.1 ± 25.0	11.9 ± 11.0	14.9 ± 4.08	16.8 ± 5.3	16.0 ± 11.2
	Lateral Tibial Condyle	**10.3 ± 4.75**	15.7 ± 9.91	**10.3 ± 6.11**	21.6 ± 7.29	19.6 ± 4.92	14.2 ± 7.92
	Medial Femoral Condyle	**7.93 ± 3.36**	11.2 ± 6.09	9.76 ± 5.98	16.6 ± 5.09	20.1 ± 8.09	17.2 ± 8.43
	Medial Tibial Condyle	**12.4 ± 8.32**	17.3 ± 20.1	13.0 ± 10.9	23.3 ± 5.51	22.5 ± 5.66	19.3 ± 9.17
	Trochlear	**7.0 ± 4.39**	10.4 ± 11.5	7.43 ± 4.81	17.6 ± 9.21	19.5 ± 10.5	9.7 ± 3.28
	Patellar	5.69 ± 1.29	5.45 ± 2.11	**5.45 ± 1.83**	12.0 ± 3.1	12.4 ± 3.5	5.76 ± 2.33
	All Cartilage	**8.09 ± 2.65**	10.7 ± 6.67	8.44 ± 3.49	15.2 ± 2.33	16.3 ± 3.91	12.4 ± 4.1
8	Lateral Femoral Condyle	**11.6 ± 9.1**	14.2 ± 14.6	14.7 ± 13.6	16.2 ± 4.83	16.4 ± 4.89	17.2 ± 10.3
	Lateral Tibial Condyle	**10.4 ± 4.13**	11.0 ± 5.98	11.9 ± 6.75	20.4 ± 5.84	19.3 ± 5.03	14.8 ± 7.17
	Medial Femoral Condyle	**9.5 ± 5.11**	12.8 ± 4.97	14.2 ± 8.74	18.3 ± 5.42	17.2 ± 4.03	17.5 ± 8.66
	Medial Tibial Condyle	**12.9 ± 8.83**	14.5 ± 10.8	15.8 ± 14.0	23.2 ± 5.11	23.1 ± 6.36	19.3 ± 8.85
	Trochlear	8.49 ± 4.77	8.64 ± 6.49	**8.38 ± 5.54**	17.3 ± 4.21	18.7 ± 4.49	10.7 ± 3.58
	Patellar	**6.33 ± 1.24**	6.34 ± 1.6	5.68 ± 2.18	13.6 ± 3.46	14.8 ± 3.29	6.05 ± 2.35
	All Cartilage	**8.94 ± 2.66**	9.59 ± 3.83	10.1 ± 4.42	16.7 ± 2.73	17.3 ± 2.39	12.9 ± 3.93
10	Lateral Femoral Condyle	13.2 ± 10.9	13.4 ± 10.9	**12.5 ± 8.94**	16.7 ± 4.32	18.1 ± 6.8	16.6 ± 9.32
	Lateral Tibial Condyle	**11.5 ± 5.28**	13.7 ± 5.03	11.7 ± 6.06	21.0 ± 4.54	20.0 ± 4.26	14.8 ± 6.67
	Medial Femoral Condyle	12.2 ± 5.9	12.0 ± 6.17	**11.3 ± 6.13**	18.7 ± 5.24	21.4 ± 8.17	17.0 ± 7.71
	Medial Tibial Condyle	17.4 ± 14.3	**14.4 ± 9.27**	15.0 ± 11.8	23.9 ± 6.07	22.7 ± 6.22	20.0 ± 8.92
	Trochlear	8.88 ± 6.27	8.88 ± 5.54	**8.25 ± 5.51**	18.9 ± 6.22	19.0 ± 9.7	12.2 ± 4.15
	Patellar	6.31 ± 1.67	6.47 ± 1.52	**5.54 ± 1.96**	14.9 ± 3.48	11.6 ± 3.25	6.8 ± 2.72
	All Cartilage	9.77 ± 3.44	10.2 ± 3.61	**9.35 ± 3.5**	17.6 ± 2.48	16.7 ± 3.44	13.4 ± 3.76
12	Lateral Femoral Condyle	**13.0 ± 7.39**	13.8 ± 12.4	13.6 ± 7.85	17.7 ± 5.29	21.2 ± 7.46	16.7 ± 9.75
	Lateral Tibial Condyle	13.3 ± 4.03	**12.3 ± 5.86**	13.7 ± 5.92	22.7 ± 6.26	23.9 ± 6.26	15.0 ± 6.26
	Medial Femoral Condyle	**12.1 ± 5.68**	12.6 ± 6.34	14.0 ± 9.33	22.4 ± 8.65	28.3 ± 10.6	17.5 ± 8.03
	Medial Tibial Condyle	15.0 ± 7.12	**14.0 ± 9.78**	15.3 ± 8.63	24.2 ± 5.75	26.2 ± 5.02	19.9 ± 8.33
	Trochlear	8.6 ± 4.45	8.2 ± 5.31	**8.18 ± 4.79**	19.8 ± 9.12	23.3 ± 11.6	12.3 ± 4.22
	Patellar	8.18 ± 1.95	6.7 ± 1.66	**6.29 ± 2.19**	14.5 ± 6.52	15.5 ± 6.5	7.09 ± 2.73
	All Cartilage	10.7 ± 2.32	**9.93 ± 3.76**	10.5 ± 3.37	18.2 ± 4.5	20.5 ± 5.58	13.4 ± 3.96

Supp. Table A.8 Correlation between predicted and ground truth T_2 in knee cartilage compartments for all tested pipelines. Pearson's r between predicted and ground truth T_2 maps in knee cartilage, the significance of which is noted as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (n=16). The top performing model in each cartilage compartment at each R is shown in bold. The no RNN pipeline is the best across most cartilage compartments and R, but the pipelines with ROI-specific losses substantially outperform their counterparts in most cases, exhibiting strong map performance at high R.

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
2	Lateral Femoral Condyle	0.712***	0.678***	**0.779*** **	0.522***	0.558***	0.555**
	Lateral Tibial Condyle	0.801***	0.775***	**0.829*** **	0.573***	0.586***	0.689***
	Medial Femoral Condyle	0.759***	0.775***	**0.826*** **	0.548***	0.605***	0.565*
	Medial Tibial Condyle	0.721***	0.697***	**0.760*** **	0.460**	0.510***	0.503***
	Trochlear	0.780***	0.762***	**0.826*** **	0.692***	0.700***	0.811***
	Patellar	0.671***	0.686***	**0.764*** **	0.527***	0.557***	0.694***
	All Cartilage	0.748***	0.736***	**0.807*** **	0.587***	0.611***	0.620***

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
3	Lateral Femoral Condyle	0.655***	0.562***	**0.660*** **	0.404***	0.469***	0.447*
	Lateral Tibial Condyle	0.715***	0.720***	**0.738*** **	0.416*	0.463*	0.627***
	Medial Femoral Condyle	0.723***	0.701***	**0.748*** **	0.435***	0.509***	0.505***
	Medial Tibial Condyle	0.655*	0.643**	**0.675** **	0.267	0.330	0.408*
	Trochlear	0.711***	0.715***	0.759***	0.638***	0.664***	**0.773*** **
	Patellar	0.618***	0.584***	**0.679*** **	0.363***	0.391***	0.633***
	All Cartilage	0.695***	0.668***	**0.722*** **	0.467***	0.502***	0.559***
4	Lateral Femoral Condyle	0.528***	0.512*	**0.572*** **	0.353***	0.407***	0.421**
	Lateral Tibial Condyle	0.682***	0.670***	**0.705*** **	0.368*	0.373*	0.538*
	Medial Femoral Condyle	0.666***	0.638***	**0.674*** **	0.402*	0.443***	0.463**
	Medial Tibial Condyle	0.600*	0.586***	**0.616** **	0.199*	0.218	0.319**
	Trochlear	0.710***	0.708***	**0.750*** **	0.636***	0.638***	0.700***
	Patellar	0.627***	0.629***	**0.645*** **	0.362*	0.377*	0.597***
	All Cartilage	0.651***	0.637***	**0.677*** **	0.451***	0.467***	0.486***
6	Lateral Femoral Condyle	0.475***	0.481***	**0.527*** **	0.306***	0.309***	0.349
	Lateral Tibial Condyle	0.600**	0.652***	**0.646*** **	0.245*	0.246*	0.462
	Medial Femoral Condyle	0.606***	0.629***	**0.627*** **	0.339*	0.366***	0.394
	Medial Tibial Condyle	0.530**	0.558***	**0.562** **	0.130	0.082	0.219*
	Trochlear	**0.728*** **	0.672***	0.715***	0.619***	0.599***	0.656***
	Patellar	0.557***	**0.568*** **	0.561***	0.308***	0.247***	0.527***
	All Cartilage	0.612***	0.610***	**0.629*** **	0.397***	0.378*	0.445***
8	Lateral Femoral Condyle	0.473***	0.445**	**0.486*** **	0.271*	0.278	0.285
	Lateral Tibial Condyle	0.601***	0.614**	**0.648*** **	0.216***	0.237*	0.423*
	Medial Femoral Condyle	**0.590*** **	0.558***	0.586***	0.308	0.330***	0.377*
	Medial Tibial Condyle	0.502*	0.520**	**0.536** **	0.058*	0.034	0.177*
	Trochlear	0.691***	0.673***	**0.695*** **	0.585***	0.586***	0.625***
	Patellar	0.554***	**0.564*** **	0.549***	0.270***	0.255***	0.482***
	All Cartilage	0.585***	0.574***	**0.609*** **	0.352***	0.364**	0.410***
10	Lateral Femoral Condyle	0.399	0.366***	**0.405*** **	0.242	0.275*	0.287
	Lateral Tibial Condyle	0.548*	0.504***	**0.568*** **	0.185*	0.174*	0.406*
	Medial Femoral Condyle	0.509***	0.480***	**0.511*** **	0.280*	0.347***	0.362*
	Medial Tibial Condyle	0.427***	0.433**	**0.465** **	0.016	-0.008	0.132
	Trochlear	**0.650*** **	0.628***	**0.650*** **	0.564***	0.561***	0.590***
	Patellar	**0.537*** **	0.443***	0.527***	0.203	0.285***	0.414***
	All Cartilage	0.555***	0.514***	**0.565*** **	0.327***	0.333***	0.386***
12	Lateral Femoral Condyle	0.359*	0.409*	**0.377** **	0.214	0.198	0.266
	Lateral Tibial Condyle	0.470***	**0.575*** **	0.494***	0.174*	0.184**	0.375*
	Medial Femoral Condyle	0.444***	**0.493*** **	0.435***	0.249*	0.244*	0.344
	Medial Tibial Condyle	0.339*	**0.467*** **	0.360*	-0.001	-0.022	0.090
	Trochlear	0.656***	**0.679*** **	0.672***	0.520***	0.501***	0.582***
	Patellar	0.434***	**0.512*** **	0.465***	0.157**	0.153***	0.392**
	All Cartilage	0.491***	**0.545*** **	0.511***	0.290***	0.287***	0.381***

Supp. Table A.9 T_2 quantification errors in hip cartilage compartments for 6 tested models. Performances of all methods in predicting T_2 maps in hip cartilage. NRMSEs are reported ±1 s.d., and the top performing model in each cartilage compartment at a given R is bolded (n=15). The no RNN pipeline version performs strongest at most R and cartilage compartments, and ROI-specific losses see stronger performance at most R and cartilage compartments than alternatively. In femoral cartilage, T_2 quantification errors are below clinically significant

thresholds at nearly all tested R for the full model and no RNN pipelines. In acetabular cartilage, error rates were below clinically significant thresholds at R=2.

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
2	Femoral	3.69 ± 1.0	3.94 ± 1.05	**3.33 ± 0.65**	3.85 ± 0.746	7.05 ± 1.09	9.56 ± 2.82
	Acetab.	4.4 ± 1.46	**4.37 ± 1.72**	4.54 ± 1.52	5.8 ± 1.92	9.89 ± 2.99	18.9 ± 5.8
	All Cart.	3.97 ± 1.03	4.1 ± 1.1	**3.79 ± 0.807**	4.58 ± 0.993	8.21 ± 1.42	14.8 ± 2.78
3	Femoral	6.14 ± 1.61	5.54 ± 1.71	**4.72 ± 1.07**	5.55 ± 0.94	9.06 ± 1.21	8.56 ± 2.21
	Acetab.	7.16 ± 2.31	5.78 ± 2.44	**6.15 ± 2.07**	7.8 ± 2.37	11.3 ± 2.95	16.5 ± 5.75
	All Cart.	6.53 ± 1.63	5.63 ± 1.68	**5.25 ± 1.13**	6.41 ± 1.31	10.0 ± 1.57	12.9 ± 3.15
4	Femoral	5.66 ± 1.01	5.81 ± 1.53	**5.23 ± 0.8**	6.27 ± 1.23	9.09 ± 1.68	8.78 ± 2.94
	Acetab.	7.0 ± 1.76	6.84 ± 2.06	**6.79 ± 1.55**	8.88 ± 3.0	11.1 ± 3.23	14.5 ± 3.76
	All Cart.	6.15 ± 1.01	6.17 ± 1.47	**5.84 ± 0.891**	7.33 ± 1.67	9.97 ± 1.74	11.8 ± 2.03
6	Femoral	7.99 ± 1.9	8.26 ± 2.47	**7.1 ± 1.57**	7.42 ± 1.79	8.31 ± 1.64	8.32 ± 2.03
	Acetab.	8.23 ± 2.52	8.15 ± 2.58	**8.01 ± 2.4**	10.2 ± 3.53	11.3 ± 3.69	15.2 ± 3.91
	All Cart.	8.1 ± 1.85	8.22 ± 2.06	**7.48 ± 1.52**	8.63 ± 2.32	9.68 ± 1.92	11.8 ± 2.14
8	Femoral	6.54 ± 2.22	**5.75 ± 1.55**	6.19 ± 1.61	9.25 ± 2.67	12.2 ± 3.45	8.17 ± 2.32
	Acetab.	7.67 ± 2.48	**7.22 ± 2.37**	8.2 ± 2.4	11.4 ± 3.26	11.8 ± 2.98	12.9 ± 3.52
	All Cart.	6.97 ± 1.93	**6.33 ± 1.33**	6.98 ± 1.45	10.2 ± 2.72	12.0 ± 2.64	10.5 ± 2.3
10	Femoral	8.77 ± 3.12	**7.6 ± 1.46**	8.63 ± 4.69	8.38 ± 1.96	9.88 ± 2.35	8.1 ± 2.83
	Acetab.	9.26 ± 3.04	9.03 ± 2.49	**8.8 ± 2.92**	11.4 ± 3.5	11.1 ± 3.47	12.4 ± 3.37
	All Cart.	8.99 ± 2.65	**8.12 ± 1.28**	8.7 ± 3.46	9.74 ± 2.24	10.5 ± 1.91	10.2 ± 2.4
12	Femoral	7.13 ± 1.78	7.92 ± 2.44	**6.49 ± 1.34**	8.34 ± 1.71	10.9 ± 2.92	8.69 ± 3.09
	Acetab.	8.64 ± 2.24	8.77 ± 2.5	**8.59 ± 2.29**	11.5 ± 3.72	12.1 ± 3.6	11.9 ± 3.37
	All Cart.	7.75 ± 1.5	8.27 ± 2.19	**7.34 ± 1.38**	9.74 ± 2.23	11.5 ± 2.36	10.3 ± 2.52

Supp. Table A.10 Correlation between predicted and ground truth T_2 in hip cartilage compartments for 6 tested models. Pearson's r between predicted and ground truth T_2 maps in hip cartilage with significances reported as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (n=15). The top performing model in each cartilage compartment at each R is shown in bold. The full model and reduced parameters pipelines generally show highest correlations between predicted and ground truth maps, but similar to the knee, networks with ROI-specific losses all show strong performance at high R.

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
2	Femoral	**0.773*** **	0.765***	0.760***	0.717***	0.540***	0.399***
	Acetab.	**0.788*** **	0.780***	0.753***	0.676***	0.481***	0.309***
	All Cart.	**0.794*** **	0.782***	0.770***	0.716***	0.514***	0.310***

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
3	Femoral	0.711***	**0.723*** **	0.712***	0.587***	0.420***	0.414***
	Acetab.	0.660***	**0.709*** **	0.663***	0.545***	0.315***	0.331***
	All Cart.	0.705***	**0.726*** **	0.703***	0.596***	0.372***	0.332***
4	Femoral	0.628***	0.635***	0**.641*** **	0.528***	0.368***	0.408***
	Acetab.	0.620***	**0.656*** **	0.616***	0.440***	0.294***	0.328***
	All Cart.	0.646***	**0.665*** **	0.648***	0.510***	0.333***	0.339***
6	Femoral	0.589***	**0.608*** **	0.593***	0.422***	0.371***	0.428***
	Acetab.	0.551***	**0.558*** **	0.521***	0.316	0.296***	0.292***
	All Cart.	0.587***	**0.597*** **	0.570***	0.382***	0.321***	0.334***
8	Femoral	**0.579*** **	0.564***	0.555***	0.402***	0.331***	0.423***
	Acetab.	**0.576*** **	0.579***	0.517***	0.229*	0.177**	0.323***
	All Cart.	**0.598*** **	0.588***	0.558***	0.334***	0.237***	0.347***
10	Femoral	0.523***	**0.528*** **	0.511***	0.369***	0.333***	0.416***
	Acetab.	**0.542*** **	0.482***	0.490***	0.177*	0.242***	0.308*
	All Cart.	**0.558*** **	0.534***	0.522***	0.279***	0.268***	0.335***
12	Femoral	0.521***	**0.563*** **	0.508***	0.336***	0.299***	0.416***
	Acetab.	0.471***	**0.521*** **	0.455***	0.192*	0.187**	0.333**
	All Cart.	0.517***	**0.566*** **	0.512***	0.280***	0.228***	0.349***

Supp. Table A.11 T_2 quantification errors in lumbar spine IVD levels for 6 tested models. Performances of all methods in predicting T_2 maps in lumbar spine IVDs. NRMSEs are reported ±1 s.d., and the top performing model in each cartilage compartment at a given R is bolded (n=5). Through R=8, the no RNN pipeline performs best in predicting T_2 maps, while at R=10, the MANTIS-GAN pipeline performs best. It's possible that, given the substantially smaller lumbar spine dataset from which DL models were trained, the more complicated loss functions of the ROI-specific loss approaches make it difficult to train at ultrafast R, and collection of a larger dataset is required to see higher quality predictions. Regardless, all pipeline versions saw predictions with error rates below clinically significant thresholds aggregated across all discs for all tested R, except CS at R=12.

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
2	L1/L2	6.11 ± 1.3	6.17 ± 1.64	**5.36 ± 1.16**	6.91 ± 1.63	7.54 ± 2.19	10.0 ± 5.72
	L2/L3	9.08 ± 4.57	9.55 ± 5.41	**7.71 ± 5.98**	12.6 ± 11.1	13.0 ± 10.6	13.5 ± 6.28
	L3/L4	5.93 ± 1.1	5.82 ± 0.731	**3.99 ± 1.27**	7.93 ± 2.43	7.77 ± 2.56	10.1 ± 4.69
	L4/L5	5.86 ± 2.29	6.02 ± 2.22	**3.66 ± 0.817**	9.51 ± 3.23	9.71 ± 3.2	9.64 ± 3.82
	L5/S1	6.37 ± 2.1	7.77 ± 2.57	**4.18 ± 1.46**	7.48 ± 1.57	7.28 ± 2.04	6.31 ± 3.5
	All Discs	6.71 ± 1.7	6.86 ± 1.57	**4.86 ± 1.16**	8.78 ± 2.08	8.95 ± 1.91	10.1 ± 3.06
3	L1/L2	9.4 ± 1.83	8.16 ± 1.25	**7.37 ± 1.86**	9.59 ± 1.06	10.4 ± 2.51	10.9 ± 3.45
	L2/L3	12.6 ± 5.13	11.3 ± 6.64	**10.4 ± 7.42**	14.0 ± 7.2	15.4 ± 10.2	12.5 ± 5.32

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
3	L3/L4	9.0 ± 2.76	7.96 ± 2.67	**6.14 ± 2.07**	10.8 ± 2.62	9.34 ± 2.1	9.75 ± 3.15
	L4/L5	8.57 ± 2.29	7.31 ± 1.84	**5.41 ± 0.985**	11.1 ± 1.86	12.1 ± 3.33	8.2 ± 1.33
	L5/S1	10.1 ± 2.22	10.7 ± 2.63	**5.9 ± 2.02**	8.9 ± 2.56	8.93 ± 2.32	5.82 ± 2.37
	All Discs	9.92 ± 2.39	8.76 ± 2.16	**7.13 ± 1.69**	11.0 ± 1.17	11.3 ± 1.74	9.48 ± 1.4
4	L1/L2	12.4 ± 3.58	10.3 ± 3.95	**9.21 ± 1.43**	11.3 ± 2.9	11.6 ± 4.09	13.2 ± 5.38
	L2/L3	12.1 ± 6.67	11.9 ± 7.16	**10.4 ± 7.1**	16.0 ± 8.03	18.9 ± 12.0	15.3 ± 5.64
	L3/L4	8.68 ± 3.86	8.42 ± 2.69	**6.14 ± 1.37**	11.1 ± 1.56	10.7 ± 3.07	9.66 ± 3.77
	L4/L5	9.64 ± 2.48	9.44 ± 2.74	**5.85 ± 1.11**	12.2 ± 2.82	13.2 ± 3.08	10.1 ± 2.94
	L5/S1	10.3 ± 2.44	9.86 ± 2.73	**6.49 ± 2.15**	9.55 ± 2.29	9.38 ± 1.96	6.78 ± 2.28
	All Discs	10.3 ± 3.02	9.73 ± 3.07	**7.42 ± 1.1**	12.1 ± 1.24	12.6 ± 1.35	11.3 ± 2.31
6	L1/L2	16.7 ± 14.0	13.2 ± 4.88	**11.9 ± 5.19**	14.0 ± 4.9	10.9 ± 2.36	14.4 ± 5.43
	L2/L3	15.7 ± 9.23	14.5 ± 8.95	**13.6 ± 9.98**	22.6 ± 12.1	17.7 ± 12.3	15.3 ± 7.14
	L3/L4	11.0 ± 3.69	11.5 ± 3.46	**9.0 ± 2.42**	13.6 ± 4.32	10.2 ± 1.65	11.3 ± 4.24
	L4/L5	10.5 ± 1.14	11.5 ± 4.11	**8.4 ± 2.1**	15.7 ± 5.61	12.2 ± 2.86	11.3 ± 2.61
	L5/S1	11.5 ± 1.73	12.5 ± 2.76	**8.82 ± 2.77**	12.5 ± 3.21	10.8 ± 3.11	8.19 ± 2.29
	All Discs	12.1 ± 3.58	12.2 ± 4.11	**10.3 ± 3.31**	15.6 ± 2.65	12.1 ± 1.9	12.0 ± 2.76
8	L1/L2	15.3 ± 3.94	15.9 ± 5.87	**13.6 ± 4.45**	12.8 ± 4.49	12.3 ± 3.96	13.3 ± 3.8
	L2/L3	17.2 ± 8.8	17.1 ± 8.64	**16.3 ± 8.99**	19.5 ± 9.27	20.0 ± 14.2	16.8 ± 6.38
	L3/L4	12.3 ± 4.85	11.7 ± 3.84	10.9 ± 3.22	11.6 ± 2.54	**10.4 ± 1.87**	13.5 ± 6.05
	L4/L5	12.3 ± 4.25	11.6 ± 2.18	**10.3 ± 2.79**	11.7 ± 2.63	12.4 ± 3.83	11.6 ± 2.44
	L5/S1	12.6 ± 3.94	10.5 ± 2.74	**9.01 ± 2.43**	10.8 ± 1.87	9.97 ± 3.12	9.18 ± 3.05
	All Discs	13.4 ± 3.89	13.0 ± 2.63	**12.0 ± 3.07**	13.2 ± 1.42	12.7 ± 1.7	12.8 ± 2.53
10	L1/L2	17.2 ± 3.6	18.2 ± 3.37	18.0 ± 2.72	15.2 ± 5.16	**12.9 ± 3.58**	16.4 ± 6.34
	L2/L3	19.7 ± 6.8	**17.9 ± 5.99**	18.1 ± 7.09	20.5 ± 10.9	19.5 ± 12.8	18.9 ± 8.12
	L3/L4	14.9 ± 4.38	14.0 ± 2.94	14.7 ± 3.62	11.8 ± 3.72	**12.3 ± 3.79**	15.5 ± 6.71
	L4/L5	14.1 ± 5.13	13.9 ± 4.83	12.8 ± 3.39	12.6 ± 1.79	**12.2 ± 1.84**	14.0 ± 3.9
	L5/S1	12.1 ± 5.09	12.3 ± 3.77	10.5 ± 4.38	11.6 ± 3.35	10.5 ± 2.69	**9.97 ± 2.36**
	All Discs	15.3 ± 3.22	14.8 ± 2.78	14.8 ± 2.26	13.8 ± 1.57	**13.2 ± 1.81**	15.0 ± 3.77
12	L1/L2	26.9 ± 10.1	25.5 ± 4.78	26.8 ± 12.0	**14.3 ± 4.15**	**14.3 ± 3.96**	33.0 ± 19.8
	L2/L3	23.2 ± 5.89	25.2 ± 3.31	23.3 ± 7.19	24.7 ± 18.7	**19.6 ± 12.2**	30.1 ± 11.3
	L3/L4	16.8 ± 4.11	24.5 ± 3.2	18.1 ± 3.88	14.1 ± 4.45	**13.2 ± 2.61**	26.1 ± 14.2
	L4/L5	15.3 ± 4.13	21.9 ± 5.59	15.3 ± 3.43	14.0 ± 5.81	**13.5 ± 5.25**	22.7 ± 11.0
	L5/S1	13.6 ± 4.66	17.9 ± 3.72	13.1 ± 4.29	9.84 ± 1.98	**10.5 ± 2.49**	11.3 ± 2.7
	All Discs	18.1 ± 1.95	23.1 ± 2.71	18.8 ± 2.76	14.8 ± 3.03	**14.1 ± 1.88**	24.8 ± 11.2

Supp. Table A.12 Correlation between predicted and ground truth T_2 in lumbar spine IVDs for 6 tested models. Pearson's r between predicted and ground truth T_2 maps lumbar spine IVDs with significances reported as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (n=5). The top performing model at each IVD level at each R is shown in bold. The full model and no RNN pipelines show highest correlations between predicted and ground truth maps, and all pipelines with ROI-

specific losses performed well, apart from the reduced parameters pipeline at R=12. T_2 quantification performances of our methods were therefore strong across in IVDs across all R.

R	Tissue	Full Model	Reduced Parameters	No RNN	MANTIS	MANTIS-GAN	CS
2	L1/L2	0.849***	0.855***	**0.853*** **	0.782***	0.782***	0.749***
	L2/L3	0.826***	0.830***	**0.832*** **	0.750***	0.745***	0.751***
	L3/L4	0.861	0.869***	**0.886*** **	0.764***	0.755***	0.830***
	L4/L5	0.859***	0.856***	**0.888*** **	0.738***	0.739***	0.823***
	L5/S1	0.793***	0.771***	**0.832*** **	0.696***	0.671***	0.712***
	All Discs	0.865***	0.866***	**0.884*** **	0.784***	0.785***	0.802***
3	L1/L2	0.796***	**0.802*** **	0.778***	0.668***	0.644***	0.703***
	L2/L3	**0.791*** **	0.782***	0.779***	0.654***	0.658***	0.715***
	L3/L4	0.824***	0.809***	**0.830*** **	0.673***	0.695***	0.759***
	L4/L5	0.841***	0.826***	**0.853*** **	0.696***	0.676***	0.795***
	L5/S1	0.710***	0.682***	**0.763*** **	0.584***	0.611***	0.741***
	All Discs	**0.836*** **	0.823***	0.832***	0.717***	0.712***	0.777***
4	L1/L2	0.732***	**0.766*** **	0.745***	0.623***	0.641***	0.579***
	L2/L3	0.749***	**0.756*** **	0.751***	0.610***	0.596***	0.622***
	L3/L4	0.810***	0.814***	**0.826*** **	0.652***	0.666***	0.747***
	L4/L5	0.818***	0.828***	**0.853*** **	0.654***	0.618***	0.751***
	L5/S1	0.675***	0.694***	**0.730*** **	0.557***	0.542***	0.672***
	All Discs	0.799***	0.813***	**0.819*** **	0.680***	0.671***	0.723***
6	L1/L2	**0.722*** **	0.717***	0.689***	0.626***	0.612***	0.653***
	L2/L3	**0.735*** **	0.718***	0.724***	0.586***	0.584***	0.677***
	L3/L4	**0.762*** **	0.745***	0.762***	0.608***	0.641***	0.708***
	L4/L5	0.794***	0.772***	**0.797*** **	0.656***	0.626***	0.727***
	L5/S1	0.642***	0.618***	**0.661*** **	0.462***	0.472***	0.620***
	All Discs	**0.776*** **	0.764***	0.771***	0.660***	0.658***	0.728***
8	L1/L2	**0.659*** **	0.613***	0.637***	0.583***	0.574***	0.599***
	L2/L3	**0.699*** **	0.686***	0.696***	0.513***	0.544	0.652***
	L3/L4	**0.737*** **	0.719***	0.734***	0.591***	0.644***	0.674***
	L4/L5	0.745***	0.731***	**0.752*** **	0.613***	0.624***	0.725***
	L5/S1	0.606***	0.578***	**0.627*** **	0.464***	0.501***	0.557***
	All Discs	**0.742*** **	0.723***	**0.742*** **	0.631***	0.645***	0.695***
10	L1/L2	0.586***	**0.610*** **	0.514***	0.587***	0.557***	0.563***
	L2/L3	0.656***	**0.660*** **	0.632***	0.557	0.518*	0.590***
	L3/L4	**0.681*** **	**0.681*** **	0.652***	0.636***	0.614***	0.644***
	L4/L5	**0.715*** **	0.706***	0.701***	0.649***	0.614***	0.686***
	L5/S1	0.564***	**0.568*** **	0.557***	0.426***	0.498***	0.502***
	All Discs	0.695***	**0.700*** **	0.672***	0.647***	0.636***	0.648***
12	L1/L2	0.545***	0.202**	0.513***	**0.566*** **	0.509***	0.453***
	L2/L3	**0.607*** **	0.237***	0.572***	0.561***	0.513***	0.488***
	L3/L4	**0.652*** **	0.273***	0.616***	0.586***	0.527***	0.542***
	L4/L5	**0.684*** **	0.435***	0.668***	0.654***	0.585**	0.669***
	L5/S1	0.508***	0.233***	0.525***	**0.569*** **	0.507***	0.436***
	All Discs	**0.664*** **	0.320***	0.643***	0.651***	0.614***	0.586***

Supp. Table A.13 Optimized loss function weightings for best pipelines in each anatomy. Loss function weightings found to yield optimal pipeline performance through constrained hyperparameter searches in each anatomy, across each of the three folds. Across different folds at a given R in the same anatomy, optimized loss function weights generally, although not always, exhibited consistency with one another, indicating stability of the entire training procedure.

			R						
	Fold	**Loss Component**	**2**	**3**	**4**	**6**	**8**	**10**	**12**
Knee	**1**	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	123.5	144.5	145.6	120.8	144	110.8	78.9
		λ_{SSIM}	1.151	0.803	0.507	0.561	0.542	0.029	0.297
		$\lambda_{Feat.}$	0.101	0.499	0.447	0.132	0.128	0.138	0.131
	2	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	117.5	100.2	124.3	120.8	61.8	64.7	76
		λ_{SSIM}	1.574	0.635	0.62	0.561	0.447	0.45	0.655
		$\lambda_{Feat.}$	0.433	0.423	0.395	0.132	0.451	0.367	0.134
	3	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	144.4	100.2	145.6	137.9	118.4	134.9	106.9
		λ_{SSIM}	0.469	0.635	0.507	1.363	0.793	0.371	0.403
		$\lambda_{Feat.}$	0.119	0.423	0.447	0.429	0.389	0.115	0.035
Hip	**1**	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	1.275	0.778	0.789	1.221	1.376	2.82	1.227
		λ_{SSIM}	0.27	1.7	1.609	1.727	1.253	0.948	0.961
		$\lambda_{Feat.}$	0.313	0.487	0.632	0.995	0.296	0.996	0.101
	2	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	1.275	1.294	1.322	2.575	0.784	2.82	1.445
		λ_{SSIM}	0.27	0.728	1.215	1.749	1.27	0.948	0.917
		$\lambda_{Feat.}$	0.313	0.43	0.873	0.798	0.01	0.996	0.265
	3	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	1.275	1.294	0.789	2.575	0.782	2.82	1.445
		λ_{SSIM}	0.27	0.728	1.609	1.749	1.576	0.948	0.917
		$\lambda_{Feat.}$	0.313	0.43	0.632	0.798	0.348	0.996	0.265
Lumbar Spine	**1**	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	2.107	7.607	6.787	8.269	6.952	3.284	9.145
		λ_{SSIM}	40.336	69.261	95.355	70.068	19.915	72.269	88.492
		$\lambda_{Feat.}$	9.16	20.149	19.651	6.368	6.347	21.222	33.647
	2	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	1.961	3.057	3.33	1.737	3.284	9.234	9.964
		λ_{SSIM}	73.233	87.205	67.15	94.58	72.269	85.043	62.233
		$\lambda_{Feat.}$	37.602	48.727	5.506	28.705	21.222	44.598	22.239

			R						
	Fold	**Loss Component**	**2**	**3**	**4**	**6**	**8**	**10**	**12**
L. Spine	**3**	λ_{L_1}	1	1	1	1	1	1	1
		$\lambda_{L_{1,\phi}}$	9.914	3.057	8.09	8.942	7.613	4.565	3.551
		λ_{SSIM}	49.478	87.205	41.3	67.49	61.799	69.246	72.122
		$\lambda_{Feat.}$	35.377	48.727	26.156	20.46	15.093	19.953	13.167

Supp. Table A.14 T_2 quantification error rates across 3 splits in tissues of interest. NRMSEs reported reported ±1 s.d. between ground truth and predicted T_2 maps in across cartilage compartments and IVD levels in 3 data splits (knee: n=16, n=9, n=16 for folds 1-3, respectively; hip: n=15 for each of folds 1-3; lumbar spine: n=5, n=4, n=4 for folds 1-3, respectively). Particularly for knee and hip pipelines, performance is consistent across data splits in cartilage compartments and overall at all tested R. In lumbar spine, performance showed increased variability compared to knee and hip pipelines, but mean T_2 quantification errors were all within a standard deviation of one another. Relatively small lumbar spine dataset size relative to knee and hip dataset sizes are likely responsible for considerably wider confidence intervals and increased variance in performance for lumbar spine.

			R						
Tissue	**Tissue Type**	**Fold**	**2**	**3**	**4**	**6**	**8**	**10**	**12**
Knee	Lateral Femoral Condyle	1	5.76 ± 2.63	7.37 ± 5.35	10.7 ± 10.1	10.9 ± 8.25	11.6 ± 9.1	13.2 ± 10.9	13.0 ± 7.39
		2	3.73 ± 1.27	5.99 ± 2.24	7.51 ± 4.79	8.93 ± 4.96	11.0 ± 4.8	9.67 ± 5.43	11.9 ± 5.55
		3	5.22 ± 1.73	6.76 ± 1.99	7.33 ± 2.8	7.4 ± 2.34	8.66 ± 3.54	9.52 ± 3.64	11.6 ± 3.59
	Lateral Tibial Condyle	1	6.02 ± 2.35	8.01 ± 3.83	9.2 ± 4.64	10.3 ± 4.75	10.4 ± 4.13	11.5 ± 5.28	13.3 ± 4.03
		2	5.98 ± 2.51	8.83 ± 3.17	12.1 ± 3.45	14.9 ± 3.69	13.7 ± 3.97	14.9 ± 3.77	15.7 ± 5.28
		3	8.42 ± 2.53	8.33 ± 2.73	12.1 ± 7.25	10.9 ± 4.61	10.9 ± 4.5	14.1 ± 4.5	14.8 ± 4.24
	Medial Femoral Condyle	1	5.31 ± 2.06	5.77 ± 2.19	7.16 ± 2.85	7.93 ± 3.36	9.5 ± 5.11	12.2 ± 5.9	12.1 ± 5.68
		2	3.34 ± 1.19	4.84 ± 2.22	6.12 ± 3.41	6.89 ± 3.98	7.78 ± 3.78	8.2 ± 4.84	11.0 ± 6.39
		3	4.18 ± 1.05	5.34 ± 1.32	6.09 ± 2.41	7.51 ± 2.57	7.64 ± 2.21	8.13 ± 2.18	10.0 ± 2.15
	Medial Tibial Condyle	1	8.1 ± 4.94	10.6 ± 8.22	11.3 ± 10.2	12.4 ± 8.32	12.9 ± 8.83	17.4 ± 14.3	15.0 ± 7.12
		2	4.5 ± 2.41	8.29 ± 4.0	10.3 ± 5.05	12.9 ± 5.7	10.1 ± 4.49	11.5 ± 4.53	11.7 ± 4.71
		3	8.51 ± 3.2	8.1 ± 3.32	13.5 ± 12.0	11.6 ± 4.74	12.3 ± 5.5	14.5 ± 4.17	15.2 ± 5.2
	Trochlear	1	7.19 ± 4.45	7.13 ± 4.43	7.4 ± 4.97	7.0 ± 4.39	8.49 ± 4.77	8.88 ± 6.27	8.6 ± 4.45
		2	4.67 ± 4.1	6.66 ± 3.86	8.47 ± 4.49	9.24 ± 4.35	13.9 ± 6.01	8.82 ± 4.28	12.8 ± 6.16
		3	5.11 ± 3.31	6.58 ± 3.59	5.67 ± 3.07	6.13 ± 2.6	7.83 ± 4.55	8.69 ± 4.59	7.72 ± 4.23
	Patellar	1	4.08 ± 1.24	5.03 ± 1.39	5.11 ± 1.26	5.69 ± 1.29	6.33 ± 1.24	6.31 ± 1.67	8.18 ± 1.95
		2	3.43 ± 2.44	7.4 ± 3.0	8.51 ± 2.82	11.3 ± 3.55	13.4 ± 4.01	8.98 ± 3.88	12.5 ± 5.65
		3	5.51 ± 2.9	6.65 ± 2.87	6.17 ± 2.68	7.32 ± 2.65	7.34 ± 3.2	7.69 ± 3.4	9.99 ± 3.84
	All Cartilage	1	5.52 ± 1.25	6.52 ± 2.17	7.54 ± 2.96	8.09 ± 2.65	8.94 ± 2.66	9.77 ± 3.44	10.7 ± 2.32
		2	4.11 ± 2.12	6.8 ± 2.38	8.45 ± 2.86	10.4 ± 3.12	12.6 ± 3.47	9.76 ± 3.47	12.7 ± 4.52
		3	5.82 ± 1.98	6.88 ± 2.17	7.33 ± 2.35	7.99 ± 2.2	8.7 ± 2.71	9.8 ± 2.61	11.1 ± 2.5
Hip	Femoral	1	3.69 ± 1.0	6.14 ± 1.61	5.66 ± 1.01	7.99 ± 1.9	6.54 ± 2.22	8.77 ± 3.12	7.13 ± 1.78
		2	4.0 ± 1.69	4.93 ± 1.68	7.86 ± 2.86	6.66 ± 2.51	8.38 ± 3.04	6.86 ± 2.66	6.79 ± 3.02
		3	3.36 ± 0.86	4.17 ± 1.05	5.14 ± 1.02	5.89 ± 1.72	7.36 ± 1.56	6.6 ± 2.24	6.42 ± 1.63
	Acetabular	1	4.4 ± 1.46	7.16 ± 2.31	7.0 ± 1.76	8.23 ± 2.52	7.67 ± 2.48	9.26 ± 3.04	8.64 ± 2.24
		2	5.01 ± 3.84	5.9 ± 4.09	8.65 ± 5.94	8.0 ± 5.15	9.42 ± 5.42	8.54 ± 5.72	9.38 ± 5.54
		3	3.18 ± 0.629	4.41 ± 1.11	5.52 ± 1.18	6.73 ± 1.99	8.3 ± 2.34	6.97 ± 2.29	7.06 ± 2.1
	All Cartilage	1	3.97 ± 1.03	6.53 ± 1.63	6.15 ± 1.01	8.1 ± 1.85	6.97 ± 1.93	8.99 ± 2.65	7.75 ± 1.5
		2	4.42 ± 2.44	5.33 ± 2.5	8.23 ± 3.81	7.21 ± 3.46	8.84 ± 3.71	7.52 ± 3.72	7.84 ± 3.94
		3	3.31 ± 0.661	4.29 ± 0.865	5.29 ± 0.889	6.28 ± 1.42	7.79 ± 1.32	6.78 ± 1.98	6.73 ± 1.35

			R						
Tissue	Tissue Type	Fold	2	3	4	6	8	10	12
Lumbar Spine	L1/L2	1	6.11 ± 1.3	9.4 ± 1.83	12.4 ± 3.58	16.7 ± 14.0	15.3 ± 3.94	17.2 ± 3.6	26.9 ± 10.1
		2	8.93 ± 4.5	16.0 ± 7.29	16.7 ± 8.83	25.4 ± 13.0	25.5 ± 14.3	20.1 ± 9.02	20.7 ± 9.87
		3	8.88 ± 5.43	8.84 ± 2.08	16.4 ± 3.98	19.1 ± 9.7	20.2 ± 12.7	20.0 ± 8.87	27.3 ± 9.65
	L2/L3	1	9.08 ± 4.57	12.6 ± 5.13	12.1 ± 6.67	15.7 ± 9.23	17.2 ± 8.8	19.7 ± 6.8	23.2 ± 5.89
		2	9.09 ± 4.82	15.6 ± 10.0	17.6 ± 11.7	33.4 ± 24.7	34.7 ± 23.8	27.8 ± 14.0	24.4 ± 15.8
		3	8.82 ± 4.73	7.74 ± 1.56	16.6 ± 10.2	16.8 ± 10.8	20.6 ± 12.2	20.1 ± 13.3	26.6 ± 14.4
	L3/L4	1	5.93 ± 1.1	9.0 ± 2.76	8.68 ± 3.86	11.0 ± 3.69	12.3 ± 4.85	14.9 ± 4.38	16.8 ± 4.11
		2	6.45 ± 3.06	15.1 ± 9.9	10.3 ± 5.24	27.6 ± 19.1	25.7 ± 16.7	26.1 ± 11.5	23.3 ± 14.0
		3	6.52 ± 2.02	7.85 ± 1.13	13.6 ± 4.0	13.3 ± 5.22	17.1 ± 8.4	23.4 ± 13.4	35.3 ± 22.0
	L4/L5	1	5.86 ± 2.29	8.57 ± 2.29	9.64 ± 2.48	10.5 ± 1.14	12.3 ± 4.25	14.1 ± 5.13	15.3 ± 4.13
		2	9.08 ± 4.79	16.5 ± 8.71	16.4 ± 8.58	27.2 ± 15.4	24.1 ± 12.6	27.4 ± 12.8	24.7 ± 12.0
		3	7.65 ± 3.82	8.68 ± 1.01	14.8 ± 7.36	16.0 ± 8.91	17.5 ± 8.36	20.4 ± 9.49	29.7 ± 13.9
	L5/S1	1	6.37 ± 2.1	10.1 ± 2.22	10.3 ± 2.44	11.5 ± 1.73	12.6 ± 3.94	12.1 ± 5.09	13.6 ± 4.66
		2	10.2 ± 7.62	17.8 ± 14.2	23.4 ± 17.0	20.8 ± 12.7	21.8 ± 12.6	30.8 ± 13.7	23.0 ± 13.5
		3	5.17 ± 1.3	9.42 ± 2.17	7.96 ± 1.16	9.01 ± 1.46	8.99 ± 0.531	12.4 ± 1.59	13.5 ± 1.23
	All Discs	1	6.71 ± 1.7	9.92 ± 2.39	10.3 ± 3.02	12.1 ± 3.58	13.4 ± 3.89	15.3 ± 3.22	18.1 ± 1.95
		2	8.45 ± 3.74	16.3 ± 8.33	14.8 ± 7.18	28.3 ± 16.5	26.5 ± 14.5	26.5 ± 11.5	23.5 ± 11.4
		3	7.65 ± 3.58	8.66 ± 0.601	13.8 ± 5.74	14.4 ± 7.25	16.2 ± 8.06	19.2 ± 10.1	26.3 ± 13.4

Supp. Table A.15 Correlations between predicted and ground truth T_2 maps across 3 splits in tissues of interest. Pearson's r between predicted and ground truth T_2 maps in tissues of interest for knee, hip and lumbar spine pipelines, with significances reported as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (knee: n=16, n=9, n=16 for folds 1-3, respectively; hip: n=15 for each of folds 1-3; lumbar spine: n=5, n=4, n=4 for folds 1-3, respectively). Performance is reported across each of the 3 data splits. With few exceptions across some IVD levels for some R, deviations in Pearson's r were relatively small across splits for the same tissue of interest at a given R, indicating stability of pipelines to datasets used.

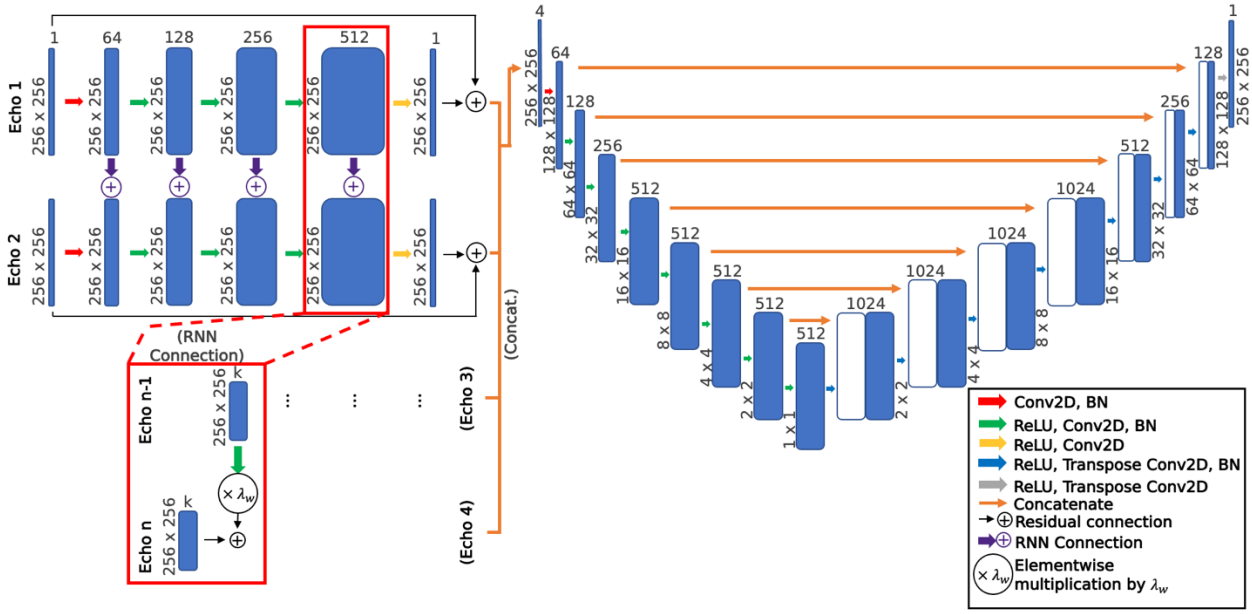
			R						
Tissue	Tissue Type	Fold	2	3	4	6	8	10	12
Knee	Lateral Femoral Condyle	1	0.712***	0.655***	0.528***	0.475***	0.473***	0.399	0.359*
		2	0.769***	0.666***	0.604***	0.535***	0.454***	0.458***	0.380***
		3	0.770***	0.695***	0.633***	0.588***	0.528**	0.516***	0.334***
	Lateral Tibial Condyle	1	0.801***	0.715***	0.682***	0.600**	0.601***	0.548*	0.470***
		2	0.821***	0.715***	0.591***	0.537***	0.536***	0.434***	0.451***
		3	0.760***	0.728***	0.658***	0.622***	0.591***	0.478	0.428***
	Medial Femoral Condyle	1	0.759***	0.723***	0.666***	0.606***	0.590***	0.509***	0.444***
		2	0.757***	0.658***	0.580***	0.528***	0.484***	0.443***	0.438***
		3	0.771***	0.683***	0.628***	0.476***	0.499***	0.462***	0.305***
	Medial Tibial Condyle	1	0.721***	0.655*	0.600*	0.530**	0.502*	0.427***	0.339*
		2	0.784***	0.620***	0.495***	0.360*	0.461***	0.308***	0.321
		3	0.656***	0.637***	0.553*	0.484*	0.407*	0.247	0.153
	Trochlear	1	0.780***	0.711***	0.710***	0.728***	0.691***	0.650***	0.656***
		2	0.824***	0.773***	0.728***	0.717***	0.664***	0.660***	0.657***
		3	0.817***	0.762***	0.715***	0.717***	0.701***	0.661***	0.678***
	Patellar	1	0.671***	0.618***	0.627***	0.557***	0.554***	0.537***	0.434***
		2	0.839***	0.738***	0.683***	0.609***	0.557***	0.578***	0.504***
		3	0.805***	0.771***	0.733***	0.613***	0.659***	0.590***	0.477***
	All Cartilage	1	0.748***	0.695***	0.651***	0.612***	0.585***	0.555***	0.491***
		2	0.812***	0.709***	0.633***	0.569***	0.532***	0.519	0.485***
		3	0.775***	0.728***	0.666***	0.597***	0.583***	0.521***	0.450***

			R						
Tissue	Tissue Type	Fold	2	3	4	6	8	10	12
Hip	Femoral	1	0.773***	0.711***	0.628***	0.589***	0.579***	0.523***	0.521***
		2	0.755***	0.686***	0.611***	0.585***	0.564***	0.518***	0.554***
		3	0.802***	0.738***	0.667***	0.623***	0.537***	0.581***	0.593***
	Acetabular	1	0.788***	0.660***	0.620***	0.551***	0.576***	0.542***	0.471***
		2	0.792***	0.722***	0.613***	0.586***	0.541***	0.553***	0.481***
		3	0.821***	0.744***	0.677***	0.566***	0.468***	0.552***	0.543***
	All Cartilage	1	0.794***	0.705***	0.646***	0.587***	0.598***	0.558***	0.517***
		2	0.782***	0.714***	0.624***	0.594***	0.564***	0.554***	0.539***
		3	0.818***	0.753***	0.687***	0.616***	0.519***	0.586***	0.589***
Lumbar Spine	L1/L2	1	0.849***	0.796***	0.732***	0.722***	0.659***	0.586***	0.545***
		2	0.793***	0.704***	0.715***	0.572***	0.492***	0.323***	0.462***
		3	0.812***	0.741***	0.722***	0.707***	0.696***	0.579***	0.521***
	L2/L3	1	0.826***	0.791***	0.749***	0.735***	0.699***	0.656***	0.607***
		2	0.869***	0.824***	0.824***	0.701***	0.641***	0.375***	0.584***
		3	0.852***	0.791***	0.776***	0.779***	0.728***	0.651***	0.569***
	L3/L4	1	0.861***	0.824***	0.810***	0.762***	0.737***	0.681***	0.652***
		2	0.896***	0.850***	0.855***	0.740***	0.687***	0.291***	0.604***
		3	0.861***	0.783***	0.755***	0.756***	0.722***	0.628***	0.514***
	L4/L5	1	0.859***	0.841***	0.818***	0.794***	0.745***	0.715***	0.684***
		2	0.835***	0.757***	0.709***	0.642***	0.607***	0.357***	0.539***
		3	0.823***	0.738***	0.709***	0.717***	0.671***	0.578***	0.540***
	L5/S1	1	0.793***	0.710***	0.675***	0.642***	0.606***	0.564***	0.508***
		2	0.769***	0.657***	0.649***	0.546***	0.490***	0.150	0.469***
		3	0.878***	0.792***	0.737***	0.747***	0.707***	0.602***	0.563***
	All Discs	1	0.865***	0.836***	0.799***	0.776***	0.742***	0.695***	0.664***
		2	0.855***	0.798***	0.788***	0.682***	0.622***	0.281***	0.565***
		3	0.859***	0.781***	0.767***	0.774	0.724***	0.647***	0.576***

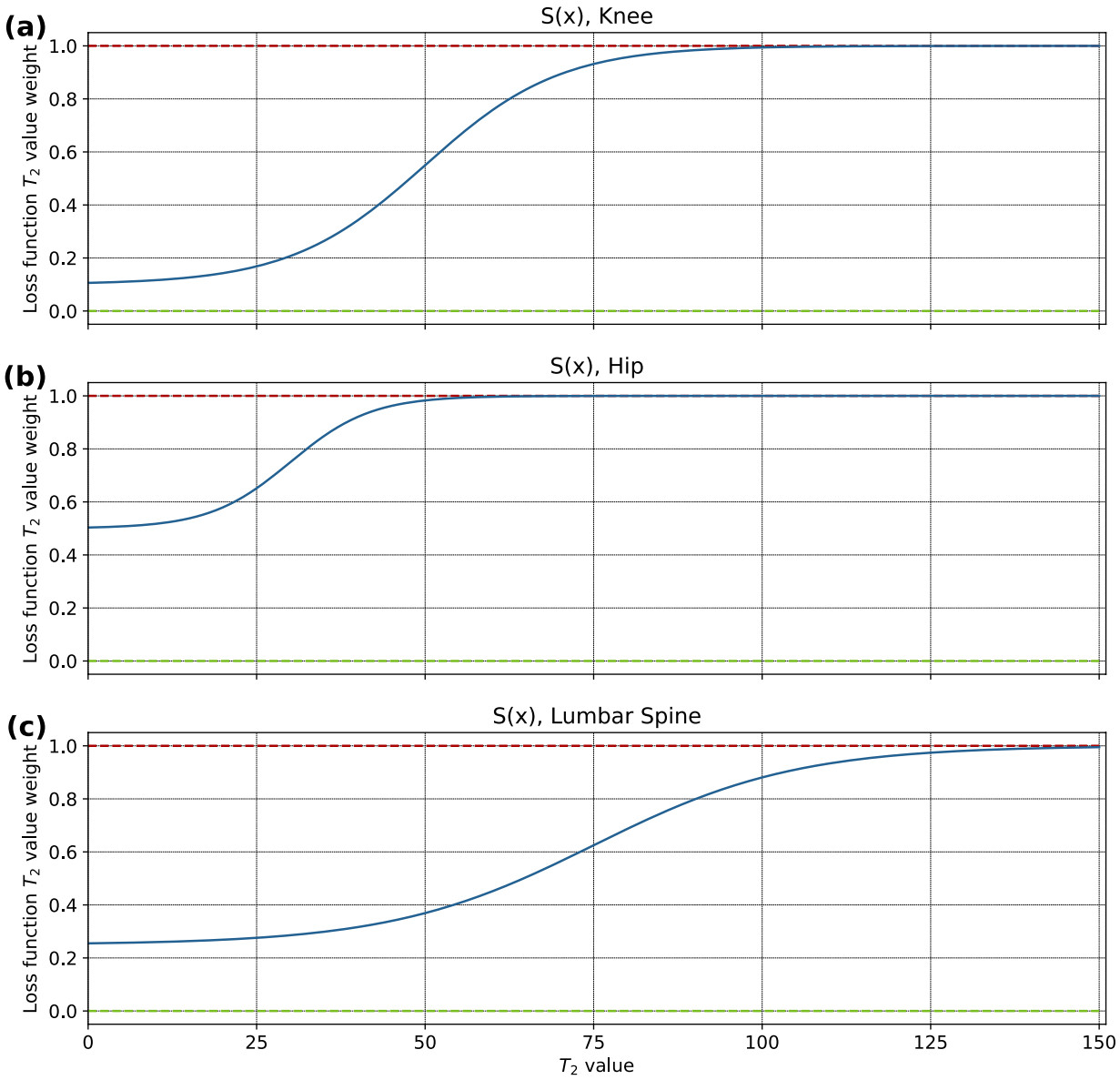
Supp. Table A.16 Texture retention performance of knee, hip and lumbar spine pipelines across 3 splits in tissues of interest. Intraclass correlation coefficients (ICCs) of Gray Level Co-Occurrence Matrix (GLCM)-based metrics for knee, hip and lumbar spine pipelines across 3 data splits. Significance of ICCs is reported as follows: * P < 0.05, ** P < 0.01, *** P < 0.001 (knee: n=16, n=9, n=16 for folds 1-3, respectively; hip: n=15 for each of folds 1-3; lumbar spine: n=5, n=4, n=4 for folds 1-3, respectively). Deviation in ICCs for texture metrics is minimal in the hip and knee pipelines for all cartilage compartments at all tested R. In the lumbar spine pipeline, more deviation existed in texture metrics, although due to small test set sizes (n=4), confidence intervals for ICCs are very wide, so at least some of the differences in texture retention performance can be attributed to this.

	R	Fold	GLCM Texture Metric				
			Contrast	Dissimilarity	Homogeneity	ASM	Energy
Knee	2	1	0.307 ± 0.18**	0.638 ± 0.12***	0.734 ± 0.09***	0.966 ± 0.015***	0.954 ± 0.02***
		2	0.344 ± 0.24**	0.493 ± 0.2***	0.673 ± 0.15***	0.908 ± 0.05***	0.902 ± 0.05***
		3	0.261 ± 0.18**	0.444 ± 0.16***	0.579 ± 0.13***	0.906 ± 0.04***	0.909 ± 0.035***
	3	1	0.153 ± 0.2	0.521 ± 0.15***	0.735 ± 0.09***	0.962 ± 0.015***	0.95 ± 0.02***
		2	0.0972 ± 0.26	0.157 ± 0.26	0.299 ± 0.24*	0.874 ± 0.07***	0.875 ± 0.07***
		3	0.256 ± 0.18**	0.41 ± 0.17***	0.474 ± 0.16***	0.885 ± 0.045***	0.895 ± 0.04***
	4	1	0.11 ± 0.2	0.387 ± 0.17***	0.61 ± 0.12***	0.973 ± 0.01***	0.95 ± 0.02***
		2	0.0554 ± 0.26	-5.23e-06 ± 0.27	0.0101 ± 0.26	0.838 ± 0.08***	0.819 ± 0.09***
		3	0.179 ± 0.2*	0.448 ± 0.16***	0.625 ± 0.12***	0.908 ± 0.04***	0.908 ± 0.035***

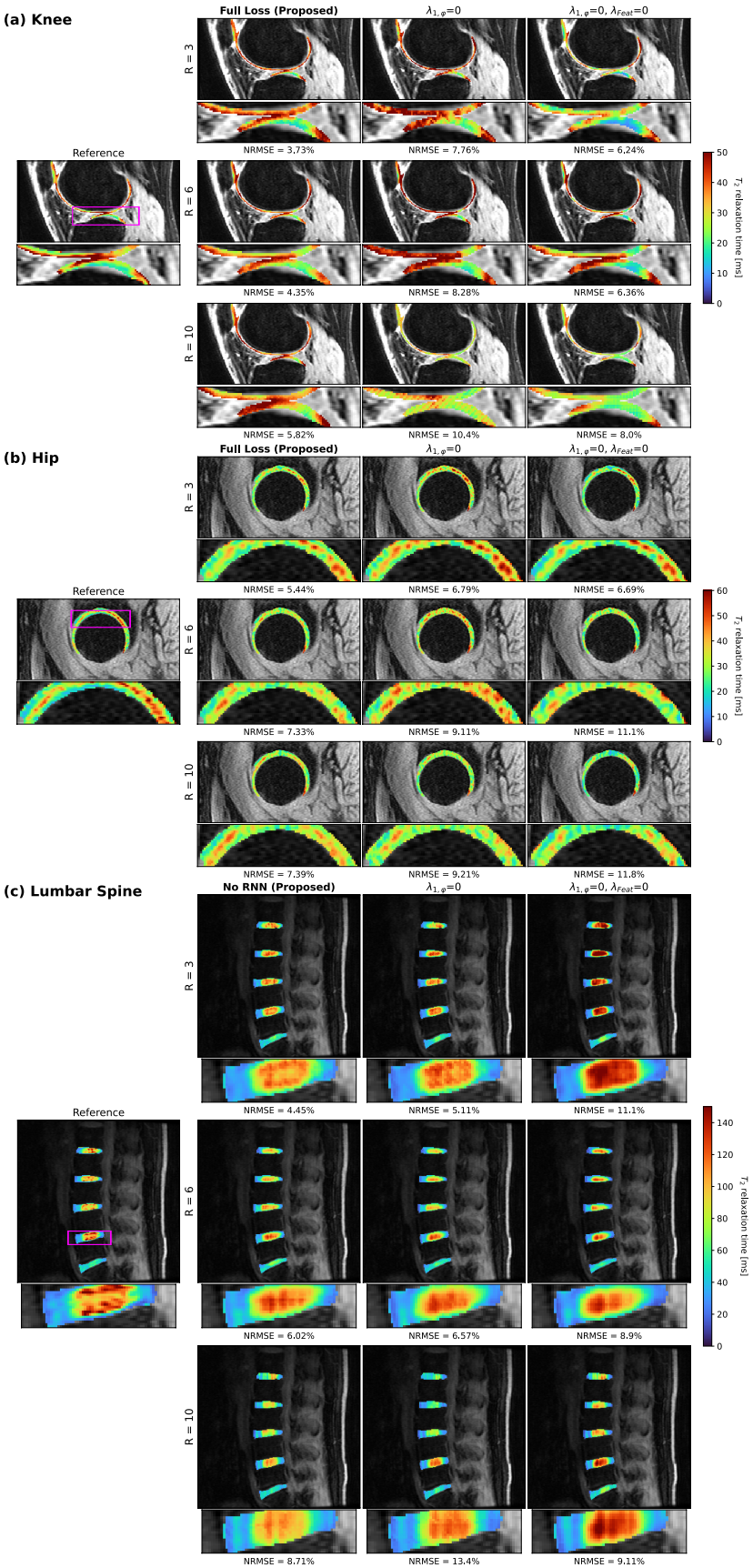
	R	Fold	GLCM Texture Metric				
			Contrast	Dissimilarity	Homogeneity	ASM	Energy
Knee	6	1	0.0667 ± 0.2	0.22 ± 0.19*	0.382 ± 0.17***	0.97 ± 0.015***	0.94 ± 0.025***
		2	0.0219 ± 0.27	-0.0819 ± 0.26	-0.125 ± 0.26	0.819 ± 0.09***	0.808 ± 0.095***
		3	0.304 ± 0.18**	0.458 ± 0.16***	0.525 ± 0.15***	0.879 ± 0.05***	0.885 ± 0.045***
	8	1	0.061 ± 0.2	0.111 ± 0.2	0.0615 ± 0.2	0.952 ± 0.02***	0.9 ± 0.04***
		2	-0.00262 ± 0.26	-0.133 ± 0.26	-0.187 ± 0.26	0.825 ± 0.085***	0.799 ± 0.1***
		3	0.0692 ± 0.2	0.0851 ± 0.2	0.15 ± 0.2	0.835 ± 0.065***	0.829 ± 0.065***
	10	1	0.0594 ± 0.2	0.218 ± 0.19*	0.307 ± 0.18**	0.961 ± 0.015***	0.928 ± 0.03***
		2	0.0279 ± 0.26	-0.0231 ± 0.26	-0.0865 ± 0.26	0.831 ± 0.085***	0.808 ± 0.095***
		3	-0.0143 ± 0.2	-0.0469 ± 0.2	0.0504 ± 0.2	0.769 ± 0.085***	0.764 ± 0.085***
	12	1	0.0032 ± 0.2	-0.066 ± 0.2	-0.178 ± 0.19	0.927 ± 0.03***	0.861 ± 0.055***
		2	-0.0137 ± 0.26	-0.186 ± 0.26	-0.327 ± 0.24	0.807 ± 0.095***	0.772 ± 0.11***
		3	-0.00712 ± 0.2	-0.133 ± 0.2	-0.166 ± 0.19	0.756 ± 0.09***	0.742 ± 0.09***
Hip	2	1	0.312 ± 0.34*	0.633 ± 0.23***	0.837 ± 0.12***	0.945 ± 0.04***	0.957 ± 0.035***
		2	0.116 ± 0.35	0.345 ± 0.32*	0.72 ± 0.18***	0.902 ± 0.07***	0.915 ± 0.065***
		3	0.274 ± 0.33	0.476 ± 0.28**	0.698 ± 0.19***	0.884 ± 0.085***	0.889 ± 0.085***
	3	1	0.369 ± 0.32*	0.671 ± 0.21***	0.816 ± 0.14***	0.976 ± 0.02***	0.98 ± 0.015***
		2	0.146 ± 0.35	0.415 ± 0.3*	0.836 ± 0.12***	0.923 ± 0.06***	0.917 ± 0.065***
		3	0.285 ± 0.33	0.504 ± 0.28**	0.721 ± 0.18***	0.942 ± 0.045***	0.937 ± 0.05***
	4	1	0.328 ± 0.33*	0.597 ± 0.25***	0.801 ± 0.15***	0.957 ± 0.035***	0.954 ± 0.04***
		2	0.0992 ± 0.36	0.294 ± 0.33	0.677 ± 0.2***	0.913 ± 0.065***	0.918 ± 0.065***
		3	0.423 ± 0.3**	0.649 ± 0.22***	0.824 ± 0.12***	0.914 ± 0.065***	0.913 ± 0.065***
	6	1	0.235 ± 0.35	0.475 ± 0.3**	0.645 ± 0.23***	0.939 ± 0.05***	0.941 ± 0.045***
		2	0.124 ± 0.35	0.386 ± 0.3*	0.791 ± 0.14***	0.907 ± 0.07***	0.902 ± 0.075***
		3	0.238 ± 0.34	0.464 ± 0.28**	0.731 ± 0.18***	0.892 ± 0.08***	0.893 ± 0.08***
	8	1	0.199 ± 0.36	0.487 ± 0.28**	0.823 ± 0.13***	0.923 ± 0.06***	0.933 ± 0.055***
		2	0.128 ± 0.35	0.375 ± 0.31*	0.764 ± 0.16***	0.848 ± 0.11***	0.82 ± 0.12***
		3	0.286 ± 0.33	0.521 ± 0.27**	0.772 ± 0.16***	0.839 ± 0.12***	0.844 ± 0.11***
	10	1	0.127 ± 0.36	0.308 ± 0.34	0.48 ± 0.29**	0.862 ± 0.11***	0.855 ± 0.11***
		2	0.0664 ± 0.36	0.209 ± 0.34	0.438 ± 0.3**	0.809 ± 0.13***	0.799 ± 0.14***
		3	0.135 ± 0.35	0.27 ± 0.33	0.473 ± 0.28**	0.695 ± 0.19***	0.73 ± 0.18***
	12	1	0.198 ± 0.36	0.38 ± 0.32*	0.523 ± 0.28**	0.927 ± 0.06***	0.914 ± 0.07***
		2	0.12 ± 0.36	0.376 ± 0.31*	0.708 ± 0.19***	0.818 ± 0.13***	0.8 ± 0.14***
		3	0.142 ± 0.34	0.317 ± 0.32*	0.57 ± 0.25***	0.795 ± 0.15***	0.813 ± 0.14***
Lumbar Spine	2	1	0.557 ± 0.7	0.695 ± 0.62	0.744 ± 0.57*	0.892 ± 0.35**	0.923 ± 0.27**
		2	0.449 ± 0.83	0.589 ± 0.78	0.788 ± 0.64	0.859 ± 0.54*	0.903 ± 0.44*
		3	-0.233 ± 0.86	-0.38 ± 0.84	0.376 ± 0.84	0.969 ± 0.2**	0.996 ± 0.03***
	3	1	0.499 ± 0.73	0.615 ± 0.67	0.644 ± 0.66	0.819 ± 0.48*	0.872 ± 0.39*
		2	0.419 ± 0.84	0.659 ± 0.74	0.616 ± 0.77	0.414 ± 0.84	0.532 ± 0.8
		3	-0.163 ± 0.87	-0.195 ± 0.87	0.391 ± 0.84	0.991 ± 0.065***	0.987 ± 0.09***
	4	1	0.236 ± 0.8	0.421 ± 0.76	0.497 ± 0.73	0.67 ± 0.64	0.775 ± 0.54*
		2	0.291 ± 0.86	0.515 ± 0.81	0.67 ± 0.74	0.551 ± 0.8	0.653 ± 0.75
		3	-0.175 ± 0.87	-0.186 ± 0.87	0.346 ± 0.85	0.942 ± 0.32**	0.954 ± 0.26**
	6	1	0.341 ± 0.78	0.428 ± 0.76	0.262 ± 0.8	0.566 ± 0.7	0.67 ± 0.64
		2	0.109 ± 0.88	0.356 ± 0.85	0.417 ± 0.84	0.347 ± 0.85	0.438 ± 0.84
		3	-0.189 ± 0.88	-0.219 ± 0.87	0.159 ± 0.88	0.936 ± 0.34**	0.958 ± 0.25**
	8	1	0.0633 ± 0.81	0.152 ± 0.8	0.276 ± 0.79	0.685 ± 0.62	0.728 ± 0.58
		2	-0.169 ± 0.87	0.043 ± 0.88	0.319 ± 0.86	0.67 ± 0.74	0.681 ± 0.74
		3	-0.061 ± 0.88	0.00747 ± 0.88	0.376 ± 0.84	0.938 ± 0.33**	0.958 ± 0.24**
	10	1	-0.0393 ± 0.81	-0.0631 ± 0.81	-0.0699 ± 0.81	0.403 ± 0.76	0.479 ± 0.74
		2	-0.393 ± 0.84	-0.361 ± 0.85	-0.194 ± 0.88	0.473 ± 0.82	0.496 ± 0.82
		3	-0.0191 ± 0.88	0.13 ± 0.88	0.434 ± 0.84	0.925 ± 0.37*	0.95 ± 0.29**
	12	1	-0.0697 ± 0.81	-0.156 ± 0.8	-0.424 ± 0.76	0.16 ± 0.8	0.198 ± 0.8
		2	0.104 ± 0.88	0.377 ± 0.84	0.435 ± 0.84	0.505 ± 0.82	0.548 ± 0.8
		3	-0.0992 ± 0.88	-0.000956 ± 0.88	0.249 ± 0.86	0.968 ± 0.2**	0.958 ± 0.25**



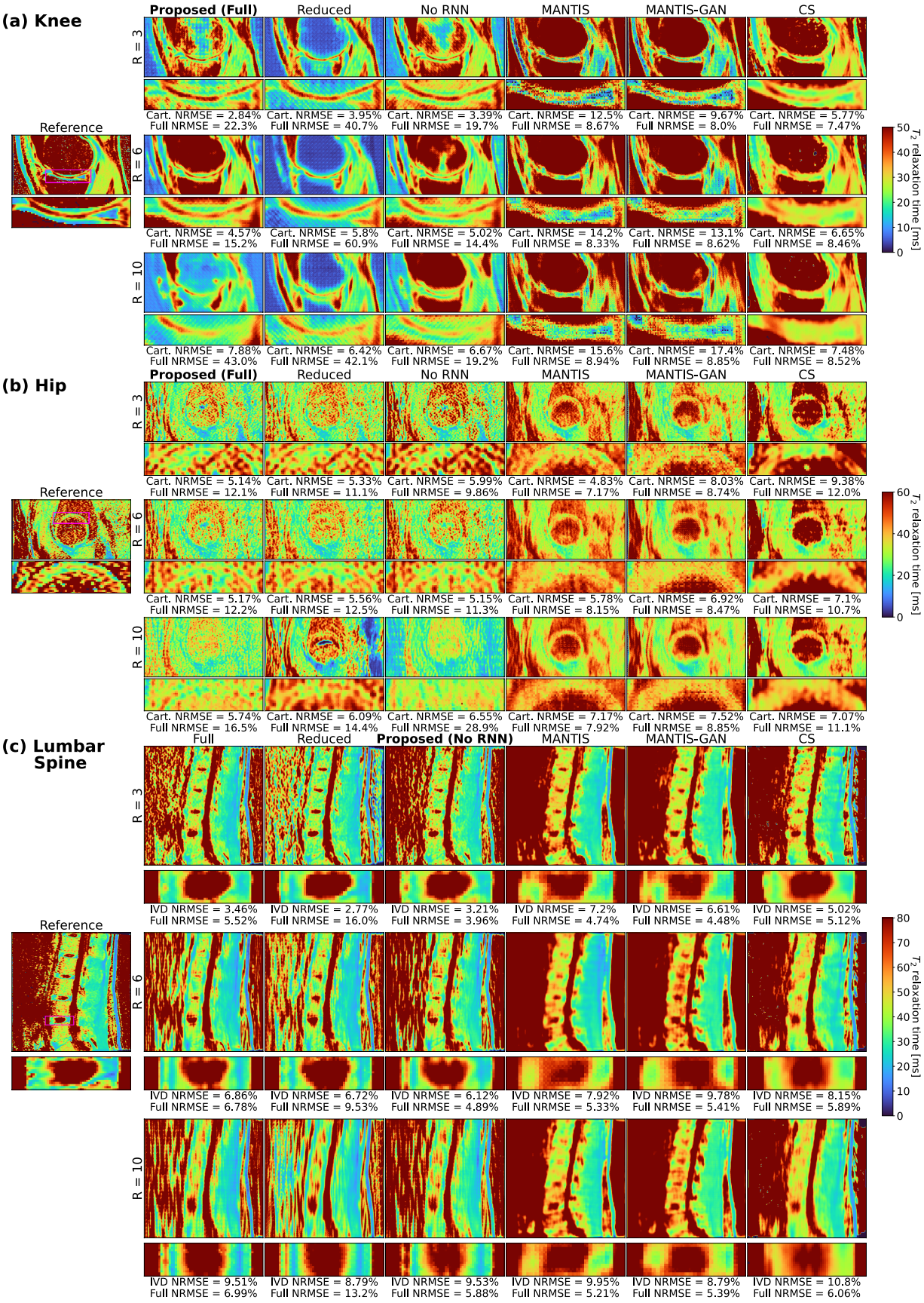
Supp. Figure A.1 Network architecture. Recurrent UNet network to predict T_2 map appearance from spatially undersampled T_2-specific MAPSS acquisition echo time images. T_2 weighted images at each echo time have a unique, 5-layer processing stream, with information passed between corresponding layers in adjacent temporal processing streams through RNN connections: ReLU, 2D convolutional layer, batch normalization, and elementwise multiplication by weighting parameter λ_w=0.2 before being added to corresponding layers of the next stream. Processing stream outputs are concatenated and fed to the UNet, which predicts T_2 maps. Depths and dimensions are provided at each layer. This schematic reflects the "full model"; additional versions were trained without the initial RNN and solely with the UNet network (No RNN) and a streamlined version in which the depth of each layer was half what is depicted in this full model schematic (Reduced Parameters).



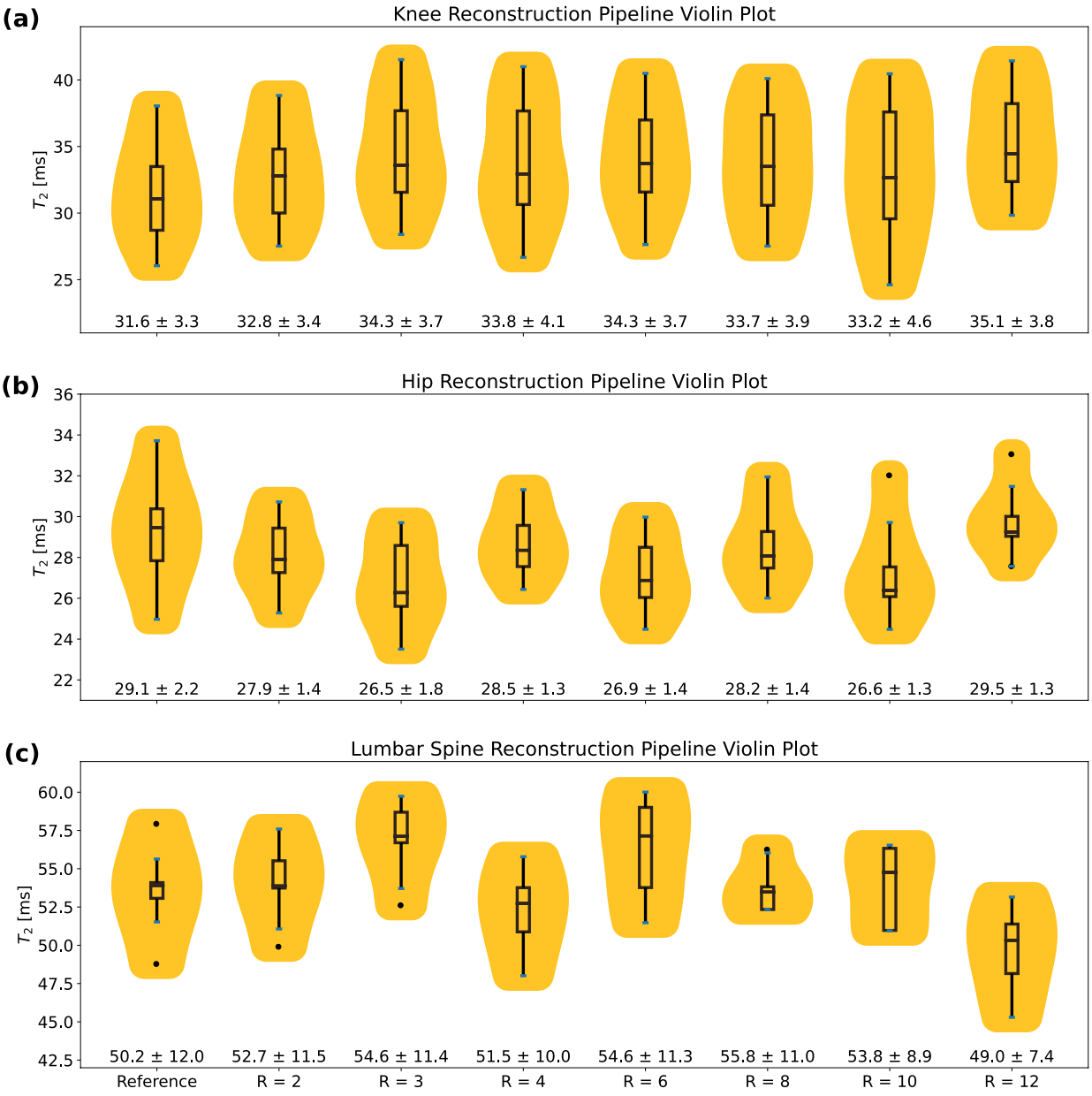
Supp. Figure A.2 Modified sigmoid function for knee, hip and lumbar spine pipelines. T_2 values in each architecture were fed through these sigmoid functions to determine an equivalent S(x) for the given pixel to be used for ROI-specific L_1 losses in network training. Sigmoid functions thus assign higher weight to correct prediction of higher T_2 values, which can be lost due to aliasing when undersampling images, particularly in local T_2 value elevations. Additionally, S(x) saturates signal above some threshold, allowing network training to focus on correct predictions in T_2 value ranges that are more physiologically realistic for cartilage and IVDs.


(a) Knee
Full Loss (Proposed)
$\lambda_{1,\varphi}$=0
$\lambda_{1,\varphi}$=0, λ_{Feat}=0
R = 3
Reference
R = 6
R = 10
NRMSE = 3.73%
NRMSE = 7.76%
NRMSE = 6.24%
NRMSE = 4.35%
NRMSE = 8.28%
NRMSE = 6.36%
NRMSE = 5.82%
NRMSE = 10.4%
NRMSE = 8.0%
T_2 relaxation time [ms]
(b) Hip
Full Loss (Proposed)
$\lambda_{1,\varphi}$=0
$\lambda_{1,\varphi}$=0, λ_{Feat}=0
NRMSE = 5.44%
NRMSE = 6.79%
NRMSE = 6.69%
NRMSE = 7.33%
NRMSE = 9.11%
NRMSE = 11.1%
NRMSE = 7.39%
NRMSE = 9.21%
NRMSE = 11.8%
(c) Lumbar Spine
No RNN (Proposed)
$\lambda_{1,\varphi}$=0
$\lambda_{1,\varphi}$=0, λ_{Feat}=0
NRMSE = 4.45%
NRMSE = 5.11%
NRMSE = 11.1%
NRMSE = 6.02%
NRMSE = 6.57%
NRMSE = 8.9%
NRMSE = 8.71%
NRMSE = 13.4%
NRMSE = 9.11%

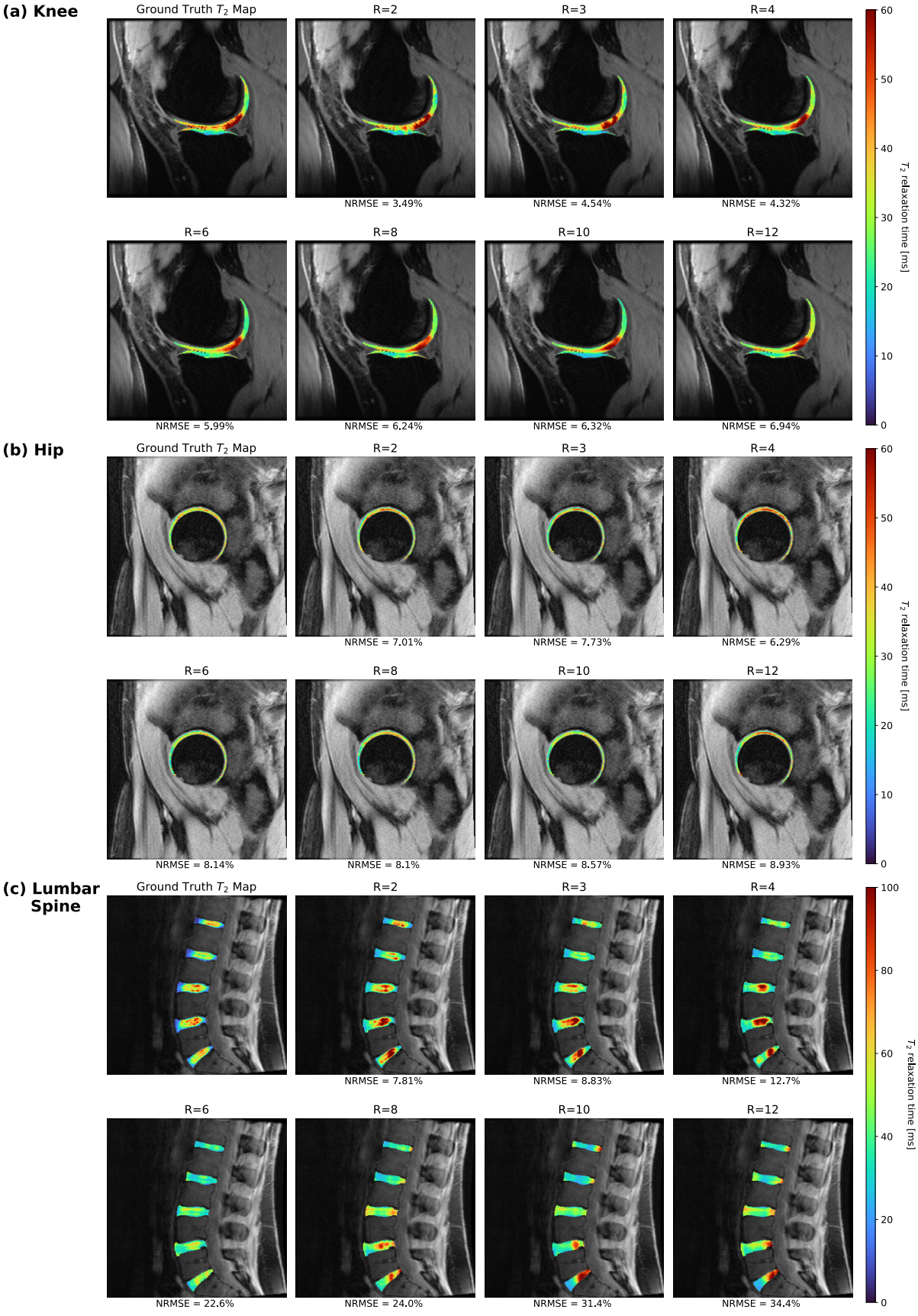
Supp. Figure A.3 Predicted T_2 maps for proposed models and equivalent pipelines trained without ROI-specific loss. Top-performing pipelines for knee (recurrent UNet), hip (recurrent UNet), and lumbar spine (UNet, or "No RNN"), with corresponding versions trained with ablated loss functions. Middle column ($\lambda_{1,\phi}=0$) was trained with proposed loss function with ROI-specific component ablated (global L_1, SSIM, feature-based losses remained). Right column ($\lambda_{1,\phi}=0$, $\lambda_{Feat}=0$) with an ordinary loss function (global L_1 and SSIM). Results show that full loss pipelines have lower T_2 quantification error rates across all anatomies at R=3, 6, and 10 for visualized slices than do pipelines trained without ROI-specific loss, demonstrating its value in maintaining low errors and maintaining visual fidelity to ground truth.


(a) Knee
Proposed (Full)
Reduced
No RNN
MANTIS
MANTIS-GAN
CS
R = 3
Cart. NRMSE = 2.84%
Full NRMSE = 22.3%
Cart. NRMSE = 3.95%
Full NRMSE = 40.7%
Cart. NRMSE = 3.39%
Full NRMSE = 19.7%
Cart. NRMSE = 12.5%
Full NRMSE = 8.67%
Cart. NRMSE = 9.67%
Full NRMSE = 8.0%
Cart. NRMSE = 5.77%
Full NRMSE = 7.47%
Reference
R = 6
Cart. NRMSE = 4.57%
Full NRMSE = 15.2%
Cart. NRMSE = 5.8%
Full NRMSE = 60.9%
Cart. NRMSE = 5.02%
Full NRMSE = 14.4%
Cart. NRMSE = 14.2%
Full NRMSE = 8.33%
Cart. NRMSE = 13.1%
Full NRMSE = 8.62%
Cart. NRMSE = 6.65%
Full NRMSE = 8.46%
R = 10
Cart. NRMSE = 7.88%
Full NRMSE = 43.0%
Cart. NRMSE = 6.42%
Full NRMSE = 42.1%
Cart. NRMSE = 6.67%
Full NRMSE = 19.2%
Cart. NRMSE = 15.6%
Full NRMSE = 8.94%
Cart. NRMSE = 17.4%
Full NRMSE = 8.85%
Cart. NRMSE = 7.48%
Full NRMSE = 8.52%
T_2 relaxation time [ms]
0
10
20
30
40
50
(b) Hip
Proposed (Full)
Reduced
No RNN
MANTIS
MANTIS-GAN
CS
R = 3
Cart. NRMSE = 5.14%
Full NRMSE = 12.1%
Cart. NRMSE = 5.33%
Full NRMSE = 11.1%
Cart. NRMSE = 5.99%
Full NRMSE = 9.86%
Cart. NRMSE = 4.83%
Full NRMSE = 7.17%
Cart. NRMSE = 8.03%
Full NRMSE = 8.74%
Cart. NRMSE = 9.38%
Full NRMSE = 12.0%
Reference
R = 6
Cart. NRMSE = 5.17%
Full NRMSE = 12.2%
Cart. NRMSE = 5.56%
Full NRMSE = 12.5%
Cart. NRMSE = 5.15%
Full NRMSE = 11.3%
Cart. NRMSE = 5.78%
Full NRMSE = 8.15%
Cart. NRMSE = 6.92%
Full NRMSE = 8.47%
Cart. NRMSE = 7.1%
Full NRMSE = 10.7%
R = 10
Cart. NRMSE = 5.74%
Full NRMSE = 16.5%
Cart. NRMSE = 6.09%
Full NRMSE = 14.4%
Cart. NRMSE = 6.55%
Full NRMSE = 28.9%
Cart. NRMSE = 7.17%
Full NRMSE = 7.92%
Cart. NRMSE = 7.52%
Full NRMSE = 8.85%
Cart. NRMSE = 7.07%
Full NRMSE = 11.1%
T_2 relaxation time [ms]
0
20
40
60
(c) Lumbar Spine
Full
Reduced
Proposed (No RNN)
MANTIS
MANTIS-GAN
CS
R = 3
IVD NRMSE = 3.46%
Full NRMSE = 5.52%
IVD NRMSE = 2.77%
Full NRMSE = 16.0%
IVD NRMSE = 3.21%
Full NRMSE = 3.96%
IVD NRMSE = 7.2%
Full NRMSE = 4.74%
IVD NRMSE = 6.61%
Full NRMSE = 4.48%
IVD NRMSE = 5.02%
Full NRMSE = 5.12%
Reference
R = 6
IVD NRMSE = 6.86%
Full NRMSE = 6.78%
IVD NRMSE = 6.72%
Full NRMSE = 9.53%
IVD NRMSE = 6.12%
Full NRMSE = 4.89%
IVD NRMSE = 7.92%
Full NRMSE = 5.33%
IVD NRMSE = 9.78%
Full NRMSE = 5.41%
IVD NRMSE = 8.15%
Full NRMSE = 5.89%
R = 10
IVD NRMSE = 9.51%
Full NRMSE = 6.99%
IVD NRMSE = 8.79%
Full NRMSE = 13.2%
IVD NRMSE = 9.53%
Full NRMSE = 5.88%
IVD NRMSE = 9.95%
Full NRMSE = 5.21%
IVD NRMSE = 8.79%
Full NRMSE = 5.39%
IVD NRMSE = 10.8%
Full NRMSE = 6.06%
T_2 relaxation time [ms]
0
10
20
30
40
50
60
70
80

Supp. Figure A.4 Global T_2 value retention performance for proposed pipelines and state-of-the-art models. ROI and global T_2 quantification errors are shown for a slice within the test set for each of the knee, hip and lumbar spine pipelines. In the knee and hip, both visually and quantitatively, T_2 maps predicted by global approaches show substantially lower global errors than do our proposed pipelines, but within cartilage ROIs, our pipelines exhibit stronger performance. These results are as expected—the ROI-specific loss function improves predictions in cartilage ROIs and degrades them globally, indicating successful training of these pipelines. In the lumbar spine, these trends are more inconsistent, possibly due to the substantially larger datasets and number of batches seen in knee and hip pipeline training as compared to the lumbar spine; some of the lumbar spine findings thus may be attributed to the randomness of training with a small dataset. Nonetheless, when afforded a sufficiently large dataset for training, the ROI-specific loss performs as expected.



Supp. Figure A.5 Comparison of biases in predicted T_2 maps in knee cartilage, hip cartilage, and intervertebral discs. Violin plots of T_2 values for reference and each tested pipeline, R=2 through R=12, for (a) knee (n=90), (b) hip (n=15), and (c) lumbar spine (n=4). Boxplots are overlaid on violin plots, and the T_2 values are also displayed mean ±1 s.d. In conjunction with Bland-Altman plots in Figures 3-4, violin plots show that for knee and hip pipelines, T_2 values with are preserved with minimal bias. In the lumbar spine, while violin plots indicate some volatility in T_2 value preservation in predicted performance, bias in predicted maps was minimal at most tested R. Knee and hip pipelines thus generally maintain strong fidelity to T_2 values, whereas lumbar spine pipeline retains reasonable fidelity to T_2 values.


(a) Knee
Ground Truth T_2 Map
R=2
R=3
R=4
NRMSE = 3.49%
NRMSE = 4.54%
NRMSE = 4.32%
R=6
R=8
R=10
R=12
NRMSE = 5.99%
NRMSE = 6.24%
NRMSE = 6.32%
NRMSE = 6.94%
(b) Hip
Ground Truth T_2 Map
R=2
R=3
R=4
NRMSE = 7.01%
NRMSE = 7.73%
NRMSE = 6.29%
R=6
R=8
R=10
R=12
NRMSE = 8.14%
NRMSE = 8.1%
NRMSE = 8.57%
NRMSE = 8.93%
(c) Lumbar Spine
Ground Truth T_2 Map
R=2
R=3
R=4
NRMSE = 7.81%
NRMSE = 8.83%
NRMSE = 12.7%
R=6
R=8
R=10
R=12
NRMSE = 22.6%
NRMSE = 24.0%
NRMSE = 31.4%
NRMSE = 34.4%
T_2 relaxation time [ms]
0
10
20
30
40
50
60
T_2 relaxation time [ms]
0
20
40
60
80
100

Supp. Figure A.6 Assessment of proposed pipeline performance on multicoil raw k-space data. MAPSS sequences were acquired for 3 knee, 2 hip, and 2 lumbar spine volunteers. Multicoil raw k-space data (after ARC reconstruction for knee and hip) was undersampled with the same patterns applied on retrospectively undersampled coil-combined images used during training, and k-space lines were also shared as with the coil-combined approach. Resulting k-space was filtered, inverse Fourier transformed, and processed with an in-house pipeline developed to replicate all image post-processing steps normally used in generating DICOM images, thereby generated coil-combined magnitude images of multi-coil undersampled images. These coil-combined equivalents of undersampled data were fed through corresponding pipelines, yielding predicted T_2 maps. Visually, and by T_2 quantification errors, knee and hip pipelines exhibit strong performance through R=12, preserving local T_2 value elevations at most R. Slight degradation in performance was observed at lower R for hip pipelines relative to benchmarks attained on coil-combined magnitude undersampled images, but at other R for knee and hip, performance matched expectations. In the lumbar spine, performance matched expectations through R=4, declining sharply at R=6 and higher. This may be due to the smaller dataset used for lumbar spine pipeline training fewer k_z slices in lumbar spine MAPSS acquisitions that exacerbated effects of undersampling as compared to knee and hip, yielding much lower SNR in aliased model input images.

Appendix B - Supplementary Information to Chapter 8

Supp. Table B.1 MR Acquisition Parameters. Acquisition parameters for the 64 scanned volume pairs in the 27-patient dataset of RA patients. Parameters for any given patient at any given time point were the same for both pre and post-Gadolinium coronal T_1 IDEAL wrist scans.

Scanner	GE Signa Discovery MR750w
Coil	8-channel HD Wrist Array
Field Strength	3T
Slice Thickness	2 mm
Spacing between Slices	2 mm
TR	457-793 ms
TE	10.06-12.48 ms
Frequency	127.8 Hz
Bandwidth	195.3 Hz (384x256, n=58), 390.6 Hz (256x224, n=6)
Acquisition Matrix	384x256 (n=58), 256x224 (n=6)
Flip Angle	90 (n=2) or 111 (n=62)
SAR	1.578-3.259
Pixel Spacing	0.234x0.234 mm (n=58), 0.469x0.469 mm (n=6)

Supp. Table B.2 SSIM for full volumes and wrist tissue from hyperparameter search. SSIMs obtained for 10-epoch trains of all 70 hyperparameter combinations for the PatchGAN pipeline without generator decoding deconvolutions. The top-12 performing parameter sets by SSIMs are highlighted in bold and were examined further visually to find the best-performing parameter set.

		λ_B						
λ_{GAN}		**0**	**0.025**	**0.050**	**0.075**	**0.100**	**0.150**	**0.200**
0.001	**Full**	0.596	0.590	0.593	0.560	0.584	0.573	0.572
	Wrist	0.713	0.713	**0.718**	0.708	0.711	0.712	0.705
0.002	**Full**	0.555	0.553	0.548	0.537	0.567	0.557	0.556
	Wrist	0.697	0.709	0.673	0.662	0.715	0.694	0.690
0.003	**Full**	0.538	0.577	0.541	0.590	0.538	0.634	0.592
	Wrist	0.681	0.703	0.689	0.712	0.665	**0.747**	**0.725**
0.004	**Full**	0.496	0.550	0.582	0.594	0.548	0.555	0.581
	Wrist	0.713	0.680	0.705	0.711	0.709	0.707	0.709
0.005	**Full**	0.591	0.554	0.585	0.575	0.584	0.622	0.578
	Wrist	**0.721**	0.698	0.708	0.692	**0.721**	**0.738**	0.716
0.006	**Full**	0.519	0.527	0.596	0.583	0.565	0.505	0.577
	Wrist	0.695	0.687	0.705	0.703	0.698	0.687	0.703
0.007	**Full**	0.557	0.587	0.607	0.491	0.579	0.555	0.56
	Wrist	0.712	0.714	**0.722**	0.682	0.703	0.695	0.685

λ_{GAN}		λ_B 0	0.025	0.050	0.075	0.100	0.150	0.200
0.008	**Full**	0.561	0.567	0.585	0.553	0.565	0.526	0.567
	Wrist	0.694	**0.732**	0.715	0.688	0.717	0.675	0.681
0.009	**Full**	0.560	0.510	0.624	0.550	0.571	0.580	0.507
	Wrist	0.684	0.674	**0.748**	0.705	0.693	0.704	0.703
0.010	**Full**	0.518	0.496	0.574	0.582	0.502	0.567	0.632
	Wrist	0.689	**0.720**	**0.733**	0.701	0.646	0.694	**0.737**

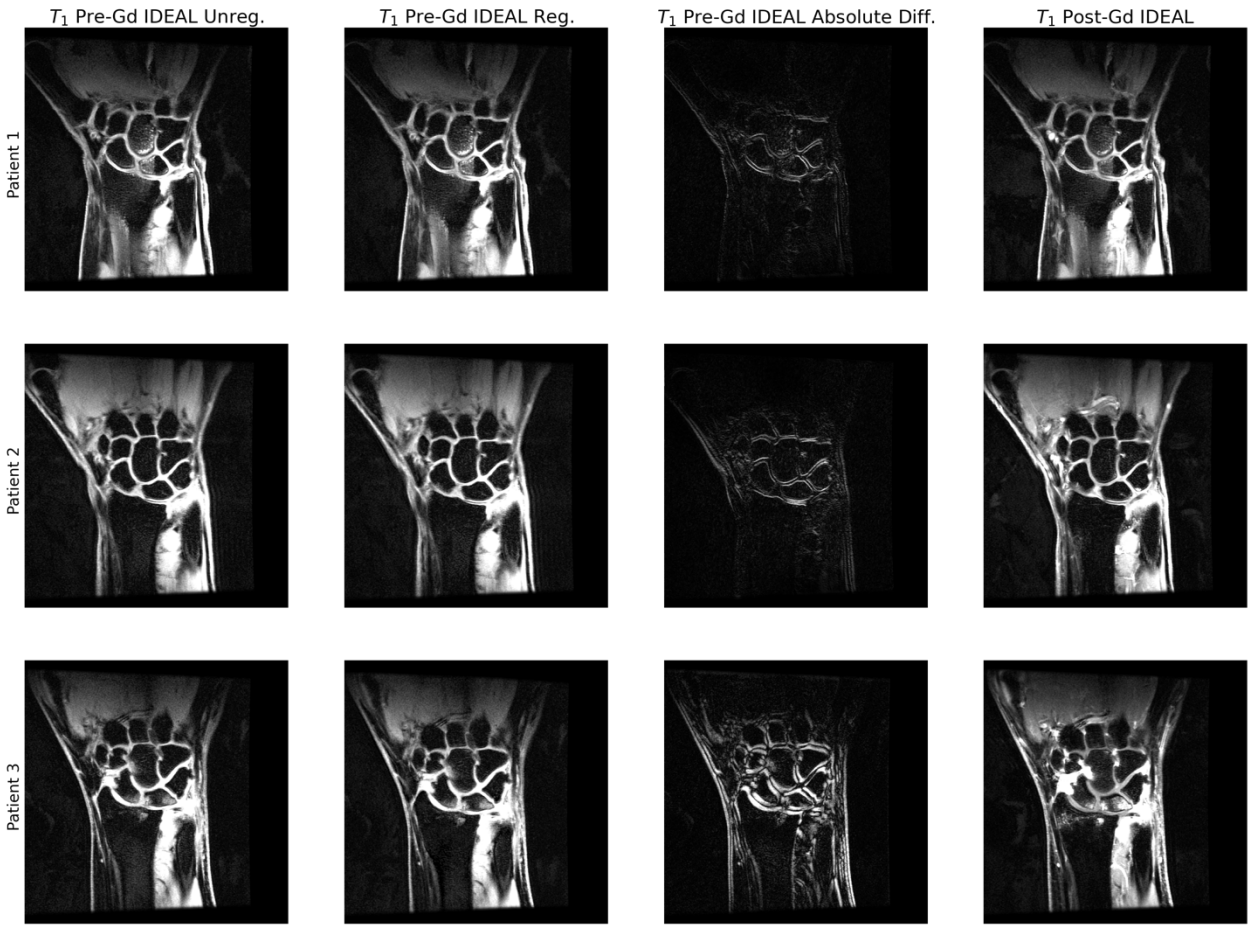
Supp. Table B.3 nRMSEs for full volumes and wrist tissue from hyperparameter search. nRMSEs obtained for 10-epoch trains of all 70 hyperparameter combinations for the PatchGAN pipeline without generator decoding deconvolutions. The top-12 performing parameter sets by nRMSEs are highlighted in bold and were examined further visually to find the best-performing parameter set.

λ_{GAN}		λ_B 0	0.025	0.050	0.075	0.100	0.150	0.200
0.001	**Full**	11.5	6.1	11.4	7.8	14.8	8.7	24.1
	Wrist	11.1	**5.8**	10.8	**7.4**	14.5	**8.2**	23.1
0.002	**Full**	12.6	9.7	6.2	10	23.9	20.3	20.3
	Wrist	12	9.1	**5.9**	9.5	23.7	20	20.1
0.003	**Full**	10.6	12	9.9	10.6	38.4	10.1	12.6
	Wrist	10.2	11.2	9.5	10	38.1	9.6	11.7
0.004	**Full**	9.2	18.5	22.5	11.6	10	33.8	10
	Wrist	8.8	18	21.8	11.1	9.7	33.4	9.5
0.005	**Full**	12.8	16.7	11.8	11.3	15.3	24.5	18.9
	Wrist	12.4	15.9	11.3	10.6	14.9	24.2	18.3
0.006	**Full**	21.2	6.2	17.9	13.1	5.8	14.5	12.3
	Wrist	20.9	**5.9**	17	12.6	**5.5**	13.9	11.1
0.007	**Full**	10.1	6.4	22	11.4	10.4	8.9	19.1
	Wrist	9.5	**6.1**	21.4	11	9.9	**8.4**	18.1
0.008	**Full**	9.9	11.4	9.4	8.9	9.6	9.8	7.9
	Wrist	9.3	10.9	9	8.4	9.2	8.9	**7.5**
0.009	**Full**	15.3	47	10.5	12.7	9.9	10.7	13.2
	Wrist	14.7	**7.4**	10	12.4	9.5	10.3	12.6
0.010	**Full**	10.9	12.9	11	10.6	15.4	6.8	8.7
	Wrist	10.5	12.6	10.4	10.1	14.7	**6.4**	**8.4**

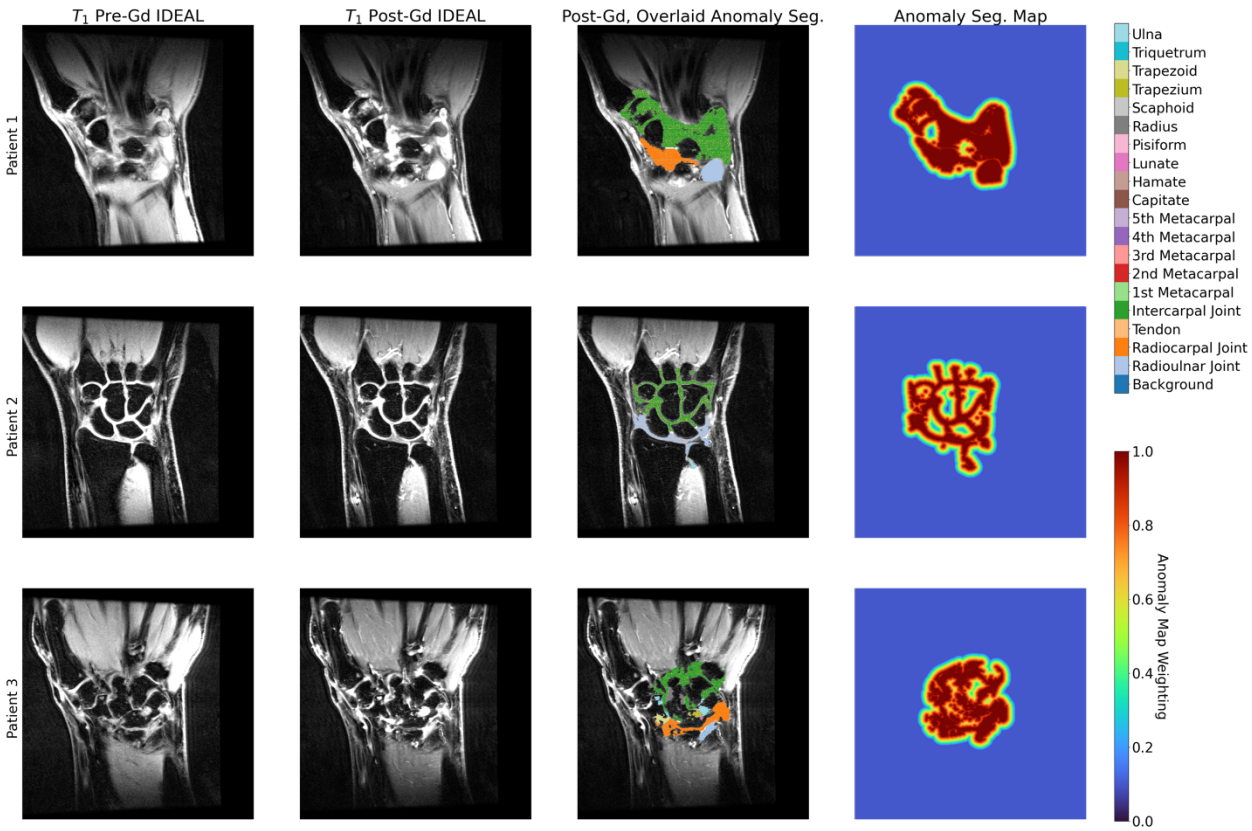
Supp. Table B.4 Reconstruction Metrics for Patients with and without Imaging Findings of RA. Bulk reconstruction metrics in full imaging volumes, wrist tissue, and synovial joints in patients without imaging findings of synovitis (RAMRIS=0, n=2) and patients with imaging findings of synovitis (RAMRIS>0, n=5) within the test set. All metrics are evaluated on a per-patient basis.

Sample sizes are quite small, making proper statistical comparisons difficult, but reconstruction metrics are generally stronger for the patients with RAMRIS>0, as is expected given most patients within this dataset prior to data splitting (24 of 28 patients) had RAMRIS synovitis scores greater than 0. Performance thus seems slightly better in patients with imaging findings of synovitis, but that may be more a function of the dataset used for training than the methodology.

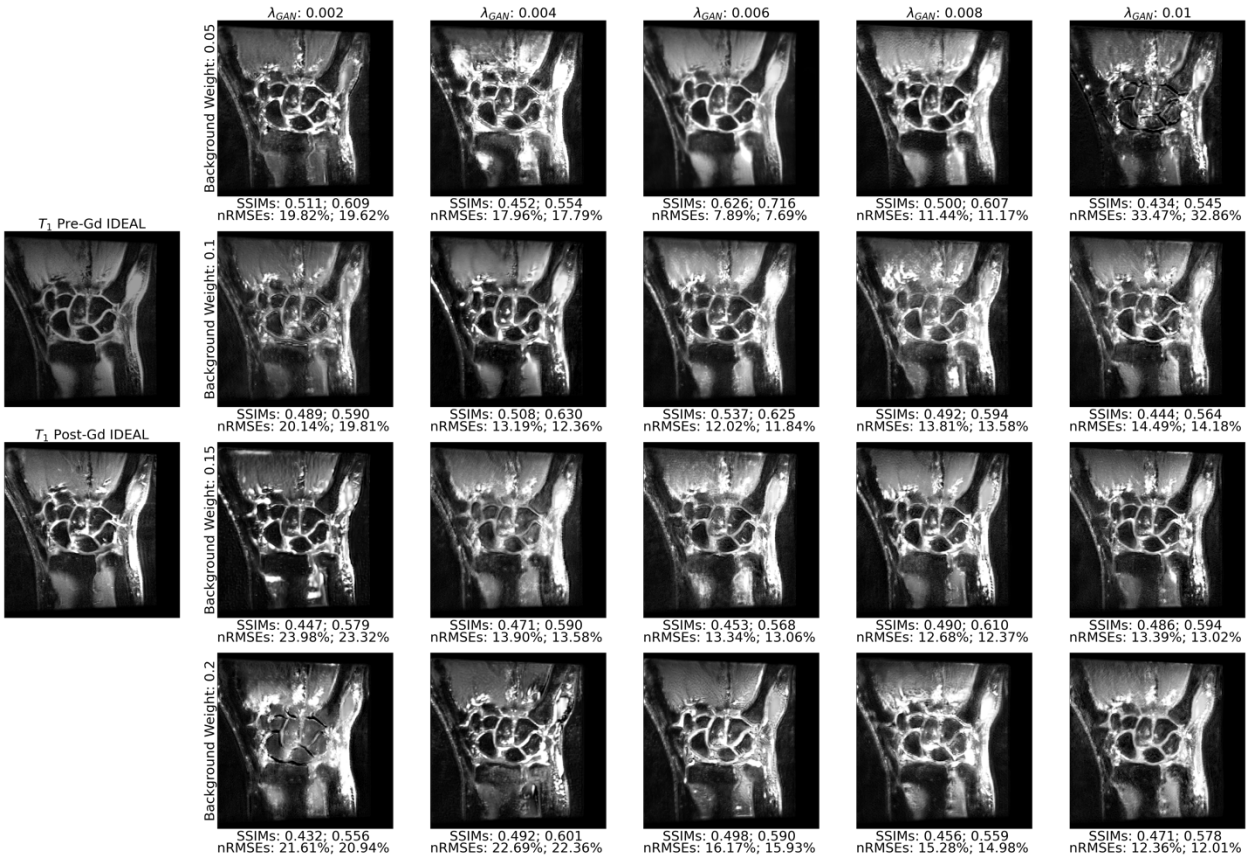
		RAMRIS = 0			RAMRIS > 0		
		Full	**Wrist Only**	**Synovial Joints**	**Full**	**Wrist Only**	**Synovial Joints**
Pre-Gd	**nRMSE**	23.95 ± 4.71	23.72 ± 4.83	133.52 ± 43.31	27.24 ± 10.27	26.94 ± 10.43	310.93 ± 159.54
	PSNR	17.76 ± 0.38	17.89 ± 0.38	9.38 ± 0.40	17.77 ± 1.10	17.95 ± 1.23	8.77 ± 1.90
	SSIM	0.62 ± 0.03	0.75 ± 0.01		0.59 ± 0.03	0.73 ± 0.01	
PatchGAN Reg.	**nRMSE**	7.13 ± 0.80	6.85 ± 0.77	28.20 ± 11.85	6.55 ± 0.75	6.26 ± 0.73	21.12 ± 2.37
	PSNR	20.44 ± 0.52	20.68 ± 0.51	11.53 ± 2.19	20.90 ± 0.64	21.16 ± 0.65	12.33 ± 0.64
	SSIM	0.58 ± 0.02	0.72 ± 0.00		0.58 ± 0.02	0.73 ± 0.01	
PatchGAN Unreg.	**nRMSE**	8.93 ± 0.69	8.65 ± 0.68	36.44 ± 17.23	8.28 ± 1.09	7.96 ± 1.06	25.98 ± 2.54
	PSNR	19.54 ± 0.34	19.73 ± 0.33	10.60 ± 2.45	19.98 ± 0.75	20.21 ± 0.75	11.49 ± 0.76
	SSIM	0.56 ± 0.02	0.70 ± 0.00		0.56 ± 0.02	0.71 ± 0.01	
UNet Reg.	**nRMSE**	7.26 ± 0.20	7.00 ± 0.17	28.74 ± 10.61	5.90 ± 0.73	5.67 ± 0.76	25.15 ± 5.39
	PSNR	21.36 ± 0.07	21.57 ± 0.04	11.26 ± 1.58	22.30 ± 0.49	22.54 ± 0.51	11.71 ± 0.40
	SSIM	0.68 ± 0.02	0.78 ± 0.01		0.69 ± 0.02	0.79 ± 0.01	
UNet Unreg.	**nRMSE**	8.35 ± 0.45	8.11 ± 0.45	31.51 ± 9.96	7.48 ± 1.09	7.24 ± 1.09	28.96 ± 6.28
	PSNR	20.82 ± 0.15	21.00 ± 0.14	10.67 ± 1.36	21.36 ± 0.67	21.57 ± 0.67	11.10 ± 0.51
	SSIM	0.67 ± 0.02	0.77 ± 0.01		0.68 ± 0.02	0.78 ± 0.01	



Supp. Figure B.1 Example Registrations. Coronal IDEAL Pre-Gd slices being registered to coronal IDEAL Post-Gd slices, with absolute difference maps of registration shown for 3 patients. Registration was done to account for motion and slight alterations to patient position that may have occurred between the sequences. Registration was done on a per-volume basis in a 3-stage algorithm: (1) translation, (2) affine, and (3) 3rd order b-spline registration (maximum iterations = 256, 256, 512, respectively; Advanced Mattes Mutual Information criterion for all). B-spline registration was done only for patients where SSIM between unregistered Pre-Gd and Post-Gd scans was above 0.5, which was used as a proxy for detecting motion artifacts so severe that any non-linear registration would lead to overfitting.



Supp. Figure B.2 Example Anomaly Distance Map. Example anomaly segmentations and corresponding anomaly segmentation maps that would result and be used to weight pixel-based L_1 loss functions during training.



Supp. Figure B.3 Sample Hyperparameter Search Slices. Examples of 10-epoch training results across one slice for 20 hyperparameter combinations; results are shown for the PatchGAN pipeline without generator decoding path deconvolutions. After initial screening using SSIM and nRMSE, optimal hyperparameter combinations were selected based visual inspection, with primary criteria being synthesis of new information, fidelity of reconstructed volumes to ground truth, and absence of obvious algorithm-generated artifacts that could cause a radiologist to lose confidence in the quality of the synthetic post-Gd images. Models with optimal hyperparameter sets were then trained from scratch for 35 epochs.

Appendix C - Supplementary Information to Chapter 9

Supp. Table C.1 Non-imaging variables identified from literature as correlated with OA progression or eventual TKR. These non-imaging variables were taken to the OAI database, and, if present, added as potential non-imaging variables to supplement image-based predictions.

Variable grouping	Variable	Source
Demographics	Age	(Lewis, 2013) [309]
	Obesity/BMI	(Lewis, 2013) [309]
	Gender	(Heidari 2011) [319]
	Ethnicity	(Yu, 2019) [310]
	Income	(Hawker, 2006) [308]
	Education level	(Pisters, 2012) [321]
Previous knee trauma and pain	Knee pain	(Lewis, 2013) [309]
	Previous knee trauma	(Heidari 2011) [319]
	Repetitive knee trauma	(Heidari 2011) [319]
	Previous meniscal injuries	(Heidari 2011) [319]
	Previous knee injury	(Cooper, 2000) [320]
Knee physical activity and functionality	Mechanical forces exerted on knee	(Heidari 2011) [319]
	Frequent kneeling	(Heidari 2011) [319]
	Frequent squatting	(Heidari 2011) [319]
	Physical activity level	(Pisters, 2012) [321]
	Muscular weakness	(Heidari 2011) [319]
	Joint range of motion	(Pisters, 2012) [321]
	Lower knee extension muscle strength	(Pisters, 2012) [321]
Previous actions to treat knee pain	Previous joint injections	(Yu, 2019) [310]
	Previous knee arthroscopy	(Yu, 2019) [310]
	Previous analgesics or opioid usage	(Lewis, 2013) [309]
	Previous NSAID usage	(Yu, 2019) [310]
	Number of previous knee referrals	(Yu, 2019) [310]
	Number of previous consultations	(Yu, 2019) [310]
	Willingness to consider TJA as treatment	(Hawker, 2006) [308]
	Seen physician for arthritis in previous year	(Hawker, 2006) [308]
Preexisting health conditions	Heberden's nodes	(Cooper, 2000) [320]
	Recorded diagnosis of joint-specific OA	(Yu, 2019) [310]
	Low back pain	(Yu, 2019) [310]
	Hypertension	(Yu, 2019) [310]
	Smoking status	(Yu, 2019) [310]
	Drinking status	(Yu, 2019) [310]
	Asthma	(Yu, 2019) [310]
	COPD	(Yu, 2019) [310]
	Diabetes mellitus	(Yu, 2019) [310]
	Comorbidities	(Pisters, 2012) [321]
Miscellaneous	Knee joint laxity	(Heidari 2011) [319]
	Genetic susceptibility to knee OA	(Heidari 2011) [319]

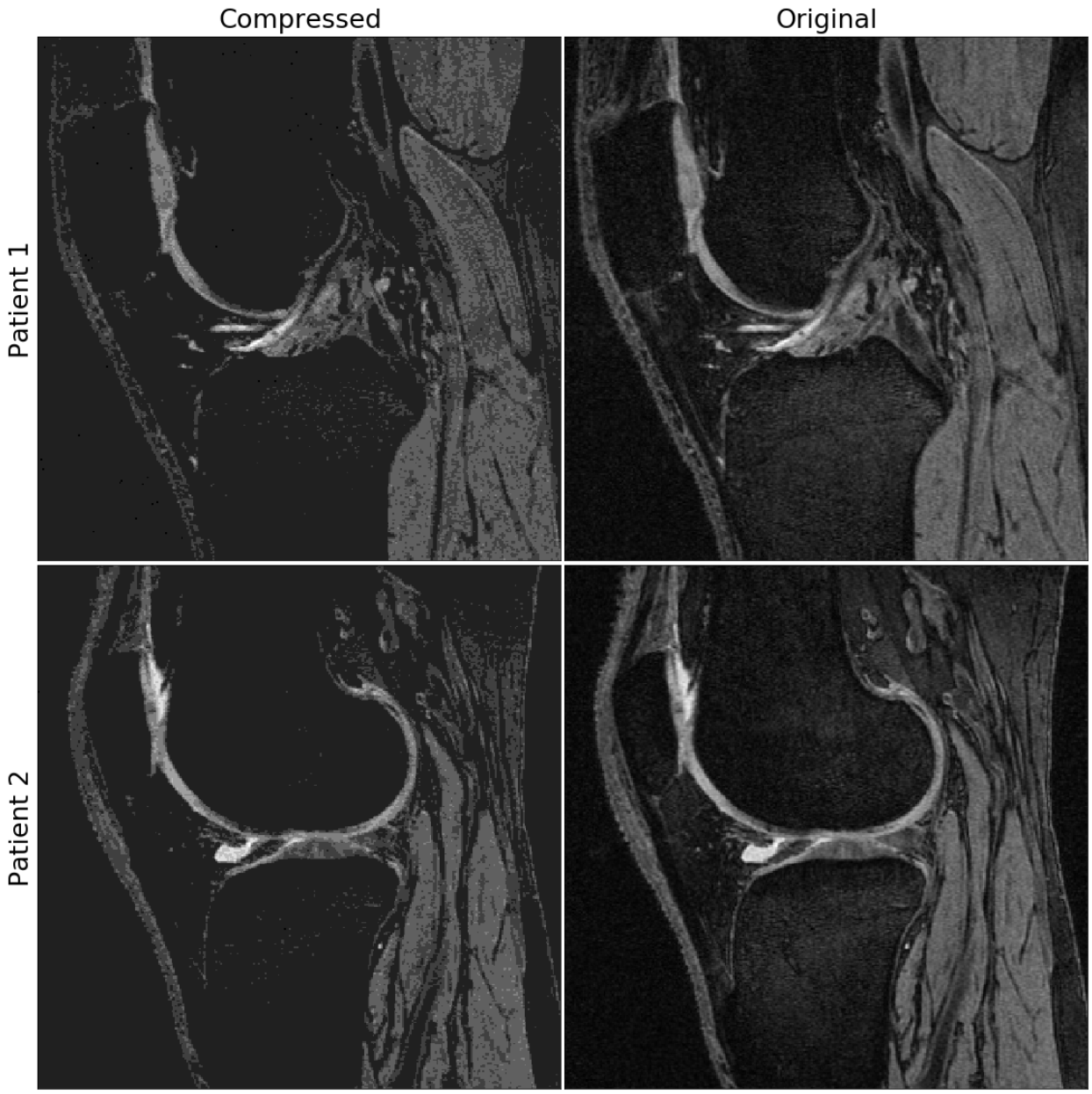
Variable grouping	Variable	Source
Miscellaneous	Mental health measures	(Sharma, 2003) [322]
	SF36 score	(Hawker, 2006) [308]

Supp. Table C.2 Percentages of selected tissues identified as hotspots among 124 true positives detected by integrated MRI pipeline, stratified by OA severity.

Tissue type	Tissue	No OA (n = 11)	Moderate OA (n = 65)	Severe OA (n = 48)	Total (n = 124)
Cartilage	TFJ medial	100.0	95.4	87.5	92.7
	TFJ lateral	100.0	87.7	85.4	87.9
	PFJ	27.3	43.1	41.7	41.1
Meniscus	Medial anterior	100.0	84.6	75.0	82.3
	Medial posterior	90.9	87.7	70.8	81.5
	Lateral anterior	100.0	87.7	81.3	86.3
	Lateral posterior	100.0	90.8	81.3	87.9
Bone	TFJ medial	100.0	95.4	89.6	93.5
	TFJ lateral	90.9	87.7	83.3	86.3
	PFJ	27.3	35.4	45.8	38.7
Ligament	ACL	100.0	81.5	64.6	76.6
	PCL	72.7	73.8	77.1	75.0
	Popliteal	54.5	56.9	58.3	57.3
Tendon	Medial patellar retinaculum	90.9	78.5	91.7	84.7
	Lateral patellar retinaculum	54.5	21.5	33.3	29.0
	Popliteal	36.4	49.2	43.8	46.0
	Patellar	27.3	27.7	25.0	26.6
	Gastrocnemius	36.4	9.2	14.6	13.7
	Semimembranosus	27.3	13.8	6.3	12.1
	Quadriceps	0.0	4.6	14.6	8.1
	Gracilis	0.0	4.6	6.3	4.8
Fat pad	Hoffa	100.0	90.8	97.9	94.4
Muscle	Popliteus	18.2	35.4	10.4	24.2
	Vastus medialis	18.2	7.7	18.8	12.9
	Gastrocnemius	36.4	26.2	27.1	27.4
	Plantaris	27.3	32.3	31.3	31.5
	Biceps femoris	0.0	4.6	6.3	4.8
	Tibialis anterior	0.0	4.6	0.0	2.4
	Semimembranosus	0.0	3.1	2.1	2.4
Synovium	General	81.8	87.7	93.8	89.5

Supp. Table C.3 Percentages of selected tissues identified as hotspots among 124 true negative controls detected by integrated MRI pipeline, stratified by OA severity.

Tissue type	Tissue	No OA (n = 11)	Moderate OA (n = 65)	Severe OA (n = 48)	Total (n = 124)
Cartilage	TFJ medial	100.0	100.0	100.0	100.0
	TFJ lateral	100.0	100.0	100.0	100.0
	PFJ	18.2	27.7	62.5	40.3
Meniscus	Medial anterior	100.0	92.3	93.8	93.5
	Medial posterior	100.0	96.9	83.3	91.9
	Lateral anterior	90.9	96.9	97.9	96.8
	Lateral posterior	100.0	100.0	91.7	96.8
Bone	TFJ medial	100.0	100.0	97.9	99.2
	TFJ lateral	100.0	100.0	100.0	100.0
	PFJ	9.1	26.2	56.3	36.3
Ligament	ACL	100.0	90.8	79.2	87.1
	PCL	45.5	63.1	72.9	65.3
	Popliteal	54.5	46.2	41.7	45.2
Tendon	Medial patellar retinaculum	54.5	66.2	87.5	73.4
	Lateral patellar retinaculum	18.2	23.1	35.4	27.4
	Popliteal	45.5	33.8	37.5	36.3
	Patellar	9.1	16.9	18.8	16.9
	Gastrocnemius	0.0	9.2	0.0	4.8
	Semimembranosus	36.4	26.2	12.5	21.8
	Quadriceps	0.0	0.0	6.3	2.4
	Gracilis	0.0	1.5	0.0	0.8
Fat pad	Hoffa	81.8	81.5	95.8	87.1
Muscle	Popliteus	18.2	15.4	10.4	13.7
	Vastus medialis	0.0	7.7	16.7	10.5
	Gastrocnemius	45.5	29.2	6.3	21.8
	Plantaris	18.2	15.4	10.4	13.7
	Biceps femoris	0.0	1.5	0.0	0.8
	Tibialis anterior	9.1	0.0	0.0	0.8
	Semimembranosus	27.3	7.7	2.1	7.3
Synovium	General	90.9	87.7	87.5	87.9



Supp. Figure C.1 Sample slices of DESS MRI and their corresponding compressed versions when rounding pixel values after normalization.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Aniket Tolpadi
1FE1E65796FB4EB...
Author Signature

5/1/2023
Date