DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

ADVANCED MRI RECONSTRUCTION AND DETECTION TECHNIQUES FOR MENISCAL TEAR DIAGNOSIS AT HIGH ACCELERATION FACTORS

by

Fatma Harman

February, 2024

İZMİR

ADVANCED MRI RECONSTRUCTION AND DETECTION TECHNIQUES FOR MENISCAL TEAR DIAGNOSIS AT HIGH ACCELERATION FACTORS

A Thesis Submitted to the
Graduate School of Natural And Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Department of Electrical and Electronics Engineering

by

Fatma Harman

February, 2024
İZMİR

Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**ADVANCED MRI RECONSTRUCTION AND DETECTION TECHNIQUES FOR MENISCAL TEAR DIAGNOSIS AT HIGH ACCELERATION FACTORS**" completed by **FATMA HARMAN** under supervision of **PROF. DR. MUSTAFA ALPER SELVER** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

.....

Prof. Dr. Mustafa Alper SELVER

Supervisor

.....	.....
Prof. Dr. Oğuz DİCLE	Assoc. Prof. Dr. Ahmet ÖZKURT
Thesis Committee Member	Thesis Committee Member
.....	.....
Prof. Dr. Adil ALPKOÇAK	Prof. Dr. Devrim ÜNAY
Examining Committee Member	Examining Committee Member

Prof. Dr. Okan FISTIKOĞLU
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Prof. Dr. M. Alper Selver, for his unwavering support, invaluable guidance, and encouragement throughout my doctoral journey. His expertise and commitment played a pivotal role in shaping this thesis, and I consider myself fortunate to have had such a dedicated mentor. I extend sincere appreciation to Prof. Dr. M. Alper Selver for his invaluable contributions to both my academic and personal development. His mentorship has been a continuous source of inspiration, and I am sincerely thankful for the opportunities for learning and growth that it provided. I couldn't have asked for a better PhD advisor.

I also want to express my gratitude to Assoc. Prof. Dr. İlkay Öksüz for his time and continuous guidance, even with his busy schedule.

Special thanks are extended to my PhD committee members, Prof. Dr. Oguz Dicle and Assoc. Prof. Dr. Ahmet Ozkurt, for their selflessness and unwavering support during the monitoring of my PhD thesis.

I am also deeply grateful to Quan Zhou and Chinmay Singhal from MD.ai for their invaluable assistance and support. With sincere gratitude, I extend my thanks to the MD.ai platform for allowing us to conduct collaborative model validation and rapid annotation projects using modern web browsers.

Furthermore, I would like to express my appreciation to Dr. Ali Canturk and Assoc. Prof. Dr. M. Mahmut Baris for their significant contributions.

Next, I extend my gratitude to my colleague Bilal Karaman and the MCBU staff in the Department of Information Technologies. Special thanks to Prof. Dr. Muhammed G. Erdem from the Computer Engineering department and all the staff in the MCBU EEE department for their support.

Heartfelt thanks to my parents for their boundless support and sacrifices. I am acutely aware that I can never repay even a fraction of what they have done for me.

Special acknowledgment to my mother, whose lessons on hard work and dedication have shaped my character, and to my father, whose unwavering belief in me has been a constant source of strength. Thanks also to my brother and my nephew for bringing joy and smiles into my life.

Fatma HARMAN

ADVANCED MRI RECONSTRUCTION AND DETECTION TECHNIQUES FOR MENISCAL TEAR DIAGNOSIS AT HIGH ACCELERATION FACTORS

ABSTRACT

Magnetic Resonance Imaging (MRI) is a non-invasive diagnostic method that allows contrast-enhanced imaging of soft tissues. However, the slow data acquisition process in MRI leads to long waiting times for patients. To address this issue and accelerate MRI scans, extensive research has been conducted in the field of image reconstruction using various algorithms after data reduction (undersampling) in k-space.

This study explores various applications of deep learning (DL) methods in the field of MRI. Initially, ensemble methods in deep neural networks for MRI image reconstruction were investigated using the FastMRI dataset. In this comprehensive evaluation employing different models and non-trainable ensemble combiners, ensemble models demonstrated a significant advantage over baseline models in evaluation metrics such as PSNR, SSIM, MSE, and NMSE.

Although deep learning methods provide notable reconstruction quality at high acceleration factors, clinically important details tend to be partially visible or lost during the reconstruction process. Addressing this issue, our study examines meniscus tears detected using DL in accelerated knee MRI images. Our analysis highlights challenges related to deep learning reconstruction quality, pointing out issues such as hallucinations and false positive artifacts. We propose an innovative approach combining deep learning reconstruction and detection for meniscus tear detection in the FastMRI + data set, highlighting the clinical effectiveness of this method.

Furthermore, our study presents a comprehensive analysis involving evaluations by blind and non-blind experts. In the blind study, an expert radiologist labels meniscus tears without prior knowledge of acceleration levels or applied reconstruction

methods. Conversely, the non-blind study evaluates reconstructed images considering specific acceleration factors and reconstruction models. This evaluation enhances understanding of the clinical applicability and reliability of the proposed deep learning reconstruction and detection system.

In conclusion, this study reveals the yet-to-be-fully-explored potential of DL methods in the MRI domain. Particularly, it provides a comprehensive evaluation of accurate classification of meniscus tears using k-space data and DL-generated reconstruction results. While demonstrating superior performance across various metrics, this evaluation also challenges traditional assumptions regarding image quality and classifier performance at different DL reconstruction models and acceleration rates, offering valuable insights.

Keywords: Magnetic resonance imaging, reconstruction, task analysis, deep learning, meniscus tear detection, pathology

YÜKSEK HIZLANDIRMA FAKTÖRLERİNDE MENİSKÜS YIRTIĞI TEŞHİSİ İÇİN GELİŞMİŞ MANYETİK REZONANS GÖRÜNTÜLEME (MRG) GERİÇATIM VE TESPİT TEKNİKLERİ

ÖZ

Manyetik Rezonans Görüntüleme (MRG), yumuşak dokuların kontrastlı bir şekilde görüntülenmesini sağlayan ve invazif olmayan önemli bir tanı aracıdır. Ancak MRG'de veri toplama sürecinin yavaşlığı, hastalar için uzun bekleme sürelerine neden olmakta, görüntü kalitesini etkilemekte ve iş yükünü artırmaktadır. Bu sorunu çözerek MRG taramalarının hızlandırılması amacıyla k-uzayında veri eksiltme (seyrek örnekleme) sonrası çeşitli algoritmalar kullanılarak görüntü geriçatma alanında kapsamlı araştırmalar yapılmıştır.

Bu tez, hızlandırılmış (seyrek / eksik örneklenmiş) MRG alanında derin öğrenme (DÖ) yöntemlerinin geriçatma ve saptama alanlarında kullanımını araştırmaktadır. İlk olarak, MRG görüntü geriçatımı için derin sinir ağlarında topluluk yöntemleri, FastMRI veri seti üzerinde incelenmiştir. Farklı modeller ve eğitilmeyen topluluk birleştiricilerin kullanıldığı bu kapsamlı değerlendirmede, topluluk modellerinin PSNR, SSIM, MSE ve NMSE gibi değerlendirme metrikleri açısından temel modellere göre belirgin bir üstünlük gösterdiği görülmüştür.

Derin öğrenme yöntemleri yüksek hızlanma faktörlerinde dikkate değer bir geriçatım kalitesi sağlarken, klinik açıdan önemli detaylar, geriçatma sürecinde kısmen görünme veya kaybolma eğilimindedir. Bu sorun ele alınarak, hızlandırılmış diz MRG görüntülerinde DÖ ile tespit edilen menisküs yırtıklarına odaklanılmıştır. Analizler, özellikle derin öğrenme kullanılarak geriçatım kalitesiyle ilgili zorlukları vurgulamakta, halüsinasyonlar ve yanlış pozitif artefaktlar gibi sorunlara işaret etmektedir.

Tez kapsamında, FastMRI+ veri setinde, menisküs yırtıklarını tespit etmek için derin öğrenme tabanlı geriçatma ve saptamayı birleştiren yenilikçi bir yaklaşım

önerilmiş ve bu yöntemin klinik etkinliği incelenmiştir.

Bu çalışmaya ek olarak, kör ve kör olmayan uzmanlar tarafından yapılan değerlendirmeleri içeren kapsamlı bir analiz sunulmuştur. Kör çalışmada, uzman radyolog, hızlandırma seviyeleri veya uygulanan geriçatım yöntemleri hakkında önceden bilgi sahibi olmadan menisküs yırtıklarını etiketlemiştir. Diğer taraftan, kör olmayan çalışma, belirli hızlandırma faktörleri ve geriçatım modelleri dikkate alınarak yeniden oluşturulan görüntüleri değerlendirmiştir. Bu değerlendirme, önerilen derin öğrenme geriçatım ve tespit sisteminin klinik uygulanabilirliği ve güvenilirliği hakkındaki geri bildirimi artıran sonuçlar üretmiştir.

Sonuç olarak, bu çalışma, MRG alanında DÖ yöntemlerinin henüz tam olarak analiz edilmemiş özelliklerini ortaya koymaktadır. Özellikle k-uzayı verileri ve DÖ ile elde edilen geriçatım sonuçlarını kullanarak, menisküs yırtıklarının doğru sınıflandırılmasında kapsamlı bir değerlendirme sağlamaktadır. Bu değerlendirme, çeşitli metriklerde üstün performans gösterirken, aynı zamanda farklı derin öğrenme geriçatım modelleri ve hızlanma oranlarındaki görüntü kalitesi ile sınıflandırıcı performansı arasındaki geleneksel varsayımları sorgulayarak yeni bir bakış açısı sunmaktadır.

Anahtar kelimeler: Manyetik rezonans görüntüleme, geriçatım, görev analiz, derin öğrenme, menisküs yırtığı tespiti, patoloji

CONTENTS

Page

Ph.D. THESIS EXAMINATION RESULT FORM ii
ACKNOWLEDGEMENTS iii
ABSTRACT v
ÖZ vii
LIST OF FIGURES xxiii
LIST OF TABLES xxv
LIST OF SYMBOLS xxvi
ABBREVIATIONS xxviii

CHAPTER ONE – INTRODUCTION 1

CHAPTER TWO – MRI ESSENTIALS 5

2.0.1 Magnetic Resonance Imaging 5
2.0.1.1 Physics of Magnetic Resonance 6
2.0.2 Accelerated MRI 9
2.0.2.1 Reconstruction 10
2.0.2.2 Analysis 12

CHAPTER THREE – ENSEMBLE DEEP LEARNING MODELS FOR ENHANCED MRI RECONSTRUCTION: A FastMRI STUDY 15

3.1 Introduction 15
3.2 FastMRI Dataset 19
3.3 Trajectories 21
3.4 Methods 22
3.4.1 Deep Learning Approaches for MRI Reconstruction in Single-Coil Knee Dataset 22
3.4.2 Deep Learning Approaches for MRI Reconstruction in Multi-Coil Knee Dataset 24

3.4.3 Basic Ensemble Methods 25
3.5 Evaluation Metrics 26
3.6 Evaluation Results 28
3.6.1 Quantitative Results 28
3.6.1.1 Quantitative Results in Single-Coil FastMRI Dataset 28
3.6.1.2 Quantitative Results in Multi-Coil FastMRI Dataset 30
3.6.2 Qualitative Results 31
3.6.2.1 Qualitative Results in Single-Coil FastMRI Dataset 31
3.6.2.2 Qualitative Results in Multi-Coil FastMRI Dataset 33
3.7 Conclusion and Discussion 34

CHAPTER FOUR – DEEP LEARNING-BASED MENISCUS TEAR DETECTION FROM ACCELERATED MRI 38

4.1 Introduction 38
4.2 Background and Related Works 43
4.3 Materials and Methods 45
4.3.1 Dataset 46
4.3.2 Reconstruction Models 47
4.3.3 Detection 51
4.4 Evaluation Metrics 54
4.4.1 Reconstruction Metrics 54
4.4.2 Detection Metrics 54
4.5 Results and Analysis 56
4.5.1 Quantitative Reconstruction Results 56
4.5.2 Quantitative Detection Results 58
4.5.3 Expert Analysis of Reconstruction and Detection Results 60
4.6 Conclusion and Future Work 66

CHAPTER FIVE – ENHANCING DEEP LEARNING PERFORMANCE IN ACCELERATED KNEE MRI RECONSTRUCTION: A FOCUS ON BLIND AND NON-BLIND EVALUATIONS 71

5.1 Introduction 71
5.2 Background and Related Research 75
5.3 Materials and Methods: Experimental Design and Implementation 77
5.3.1 Dataset 78
5.3.2 Reconstruction Models 79
5.3.3 Detection Model 81
5.4 Evaluation Metrics 83
5.4.1 Reconstruction Metrics 83
5.4.2 Detection Metrics 84
5.5 Results and Interpretations 84
5.5.1 Quantitative Reconstruction Performances 85
5.5.2 Quantitative Detection Performances 86
5.5.3 Expert Study 88
5.5.3.1 Non-blind Expert Study 89
5.5.3.2 Blind Expert Study 93
5.6 Conclusion and Discussion 97

CHAPTER SIX – INTEGRATED APPROACH TO MENISCAL TEAR RECONSTRUCTION AND CLASSIFICATION IN ACCELERATED MRI . 101

6.1 Introduction 101
6.2 Material and Methods 104
6.2.1 FastMRI Knee Dataset and FastMRI+ Annotations 104
6.2.2 Preprocessing of the Images 105
6.2.3 MRI Reconstruction 108
6.2.4 Transfer Learning Methodologies for Meniscus Tear Detection 110
6.2.5 Evaluation Metrics 112
6.3 Results 115
6.4 Discussion and Conclusion 119

CHAPTER SEVEN – CONCLUSION AND DISCUSSION 125

7.1 Conclusion 125

7.1.1 Basic Ensembles For MRI Reconstruction Task 130
7.1.2 Meniscus Tear Detection in Accelerated MRI 130
7.1.3 Improving DL Performance in Accelerated Knee MRI: Blind and Non-Blind Assessments 130
7.1.4 Meniscus Tear Classification From Undersampled MRI 131
7.2 Discussion and Future Endeavors 131

REFERENCES 134

LIST OF FIGURES

Page

Figure 2.1 A knee MRI slice captured using Echo 1 (a) and Echo 2 (b) sequences…6

Figure 2.2 This overview outlines the MR image reconstruction process, which involves collecting data with undersampling in k-space. The acquired data are then processed using a specialized reconstruction algorithm, highlighting the key steps in achieving high-quality MR images..... 13

Figure 2.3 An overview of the integrated multi-coil MRI pipeline and its associated machine learning tasks. The process initiates with the transformation of undersampled data collected from multiple coils into high-quality images (depicted by the blue arrow during the reconstruction stage). Following this, image analysis tasks, including the localization of pathology regions of interest (e.g., detection and classification), are carried out..... 14

Figure 3.1 A proton-density weighted slice (a) with fat suppression (PDFS) and (b) without fat suppression (PD). 20

Figure 3.2 K-space trajectories at a) 4-fold acceleration b) 8-fold acceleration 21

Figure 3.3 Reconstruction results for a specific slice in the single-coil validation set at 4-fold and 8-fold acceleration (27th slice of file1000041, part of the PD validation set). The first two rows display the reference and DL reconstruction results for 4-fold acceleration, along with the absolute error (residual error) compared to the reference. The following two rows show the reference and DL reconstruction results for 8-fold acceleration, along with the corresponding residual error. 34

Figure 3.4 Reconstruction results for a specific slice in the single-coil validation set with 4-fold and 8-fold acceleration (27th slice of file1000041, part of the PD validation set). The first two rows present the reference and ensemble-based reconstruction results for 4-fold acceleration, accompanied by the absolute error (residual error) in comparison to the reference. The subsequent two rows depict the reference and ensemble-based reconstruction results for 8-fold acceleration, along with the corresponding residual error. 35

Figure 3.5 Reconstruction results for a specific slice in the multi-coil validation set at 4-fold and 8-fold acceleration (27th slice of file1000041, part of the PD validation set). The first two rows display the reference and DL reconstruction results for 4-fold acceleration, along with the absolute error (residual error) compared to the reference. The following two rows show the reference and DL reconstruction results for 8-fold acceleration, along with the corresponding residual error. 36

Figure 3.6 Reconstruction results for a specific slice in the multi-coil validation set with 4-fold and 8-fold acceleration (27th slice of file1000041, part of the PD validation set). The first two rows present the reference and ensemble-based reconstruction results for 4-fold acceleration, accompanied by the absolute error (residual error) in comparison to the reference. The subsequent two rows depict the reference and ensemble-based reconstruction results for 8-fold acceleration, along with the corresponding residual error. 37

Figure 4.1 Examples of Meniscus Tear bounding boxes and their corresponding cropped images extracted from the FastMRI+ knee dataset. The arrows highlight the tears. The first row shows instances from 1.5 Tesla examinations, and their corresponding cropped images are displayed just below. The second row presents cases from 3 Tesla examinations, along with their respective cropped images below. The images in the first column represent Proton Density (PD) contrast, while the ones in the second column represent Proton Density Fat Suppression (PDFS). 48

Figure 4.2 The study design through the model training and testing phases. Initially, we undersampled k-space and generated zero-filled aliased images. Subsequently, three distinct reconstruction models were employed to obtain high-quality images. For the evaluation of meniscus tear, expert study and analysis using DL processes were performed separately. Data annotated as "Meniscus Tear" were collected and prepared for our nnDetection model. The model was then trained with this data and tested with the three reconstruction outcomes. Finally, we compared the quantitative DL detection results with the expert study findings. 49

Figure 4.3 An overview of the comprehensive multi-coil MRI reconstruction, detection, and analysis pipeline. Initially, the acquired data are undersampled using cartesian trajectories, resulting in an aliased image. Subsequently, the aliased image is processed through three different reconstruction models (termed upstream, represented by the blue arrow) to transform it into a high-quality image. These reconstructed outputs are then assessed using both Image Quality Assessment (IQA) and expert studies. Following the evaluation, the region of interest (in this case, the meniscus tear) is localized (e.g., detected) and trained using a 3D nnDetection model with fully sampled scans. The trained model is then tested with reconstructed scans for further analysis (referred to as downstream, indicated by the red arrow) 53

Figure 4.4 A demonstration of the nnDetection outcomes represented by bounding boxes for different reconstructions. At 4-fold and 8-fold acceleration (top row and bottom row), the detector successfully identified the meniscus tear across all reconstruction methods except for ZeroFilled. These bounding boxes shown in the illustration are 2D representations of the 3D prediction boxes. The arrow in the first image of the second row points the meniscus tear. 60

Figure 4.5 Illustrations of hallucinations observed in the DL-based reconstruction outcomes, specifically on the femur, medial, or distal femur regions. Among these examples, (a), (c), and (d) represent linear, (b) exhibits fogging-style artifacts. The hallucinations are not confined to a single location, but are scattered across the images, highlighted by arrows. ... 64

Figure 4.6 Illustrations of hallucinations resulting from DL-based reconstruction, primarily found in the tibia and scattered throughout the images. The examples are identified in various regions including (a) tibia, (c) fibula, (d) both tibia and fibula, (e) femur and tibia, and (f) proximal tibia and fibula. All appear as lineal artifacts. Additionally, (b) shows minimal hypo-intense lineal artifact in the proximal tibia. The instances are marked with an arrow. 65

Figure 4.7 The hallucinations pointed by arrows exhibit realistic structures artificially generated by the DL-based methods and are referred as the "FP Artifact"s. Examples (a), (d), and (b) demonstrate increased pseudo and linear signals in the meniscus, potentially leading to misdiagnosis. Example (c) presents artificially generated "Bone Marrow Edema" that is not present in the fully sampled images. 66

Figure 4.8 The hallucinations observed in the femur, tibia, and FP artifacts were carefully diagnosed and counted slice by slice by an expert. The counts were then scaled logarithmically using the formula (1+log10(count)). The resulting counts are presented as bar plots for both acceleration rates. The bar plots in (a) display the hallucination counts from DL-reconstruction results in the femur and tibia, while (b) shows the count of FP artifacts. 67

Figure 4.9 The plot illustrates the average diagnostic scores from radiological assessments for the reconstructed images. The scoring system is summarized as follows: 1 for far worse, 2 for worse, 3 for alike, 4 for better, and 5 for far better compared to the reference image. The signs on the plot represent the calculated average of expert scores. 67

Figure 4.10 The plot displays the average scores obtained from radiological evaluations of pathological features in the reconstructed images. The scoring system used is as follows: 0 indicates that the pathology is not observable, 1 for far worse, 2 for worse, 3 for similar, and 4 for better compared to the reference image. 68

Figure 5.1 An overview of the comprehensive multi-coil MRI reconstruction, detection, and analysis pipeline. The acquired data undergo initial undersampling using cartesian trajectories, resulting in an aliased image. The aliased image is then processed through three distinct reconstruction models(UNET, iRIM, and E2E VarNET) (referred to as upstream, depicted by the blue arrow), transforming it into a high-quality image. The reconstructed outputs undergo assessment through both Image Quality Assessment (IQA) and non-blind expert evaluations, covering both diagnosis and the identification of pathological features. Following the evaluation, the region of interest, in this case, the meniscus tear, is localized (detected) and trained using a 3D nnDetection model, utilizing fully sampled scans. This trained model is subsequently tested with reconstructed scans for further analysis (referred to as downstream, indicated by the red arrow). For the comparison of meniscus tears, reconstruction outcomes are also evaluated by a blind study, conducted without any information about the models or accelerations used. The blind expert annotates meniscus tears, and the annotation results are finally compared with both FastMRI+ annotations and nnDetection results for a comprehensive assessment. 78

Figure 5.2 Meniscus Tear Bounding Boxes in FastMRI+ Knee Dataset. The Tear Area is highlighted in zoomed-in views with indicators using white arrows. The first and second rows showcase examples from 1.5 and 3 Tesla examinations, while the first and second columns display PD and PDFS Contrasts, respectively. 80

Figure 5.3 The study design encompasses the model training and testing phases, incorporating evaluation through blind/non-blind expert studies. Initially, k-space was undersampled, and zerofilled aliased images were generated. Subsequently, three distinct reconstruction models were employed to obtain high-quality images. For the assessment of meniscus tear, non-blind expert studies and analyses, involving diagnosis and pathological evaluation using deep learning processes, were conducted independently. Data annotated as "Meniscus Tear" were collected and prepared for our nnDetection model. The model underwent training with this data using 5-fold cross-validation and was tested using the three reconstruction outcomes. The reconstruction outcomes were also assessed with blind expert studies and annotated based on the existence of meniscus tear. Finally, we compared the quantitative deep learning detection results with blind expert study findings and reconstruction outcomes, taking into account the influence of non-blind expert assessments on the detection process. 82

Figure 5.4 The plot depicts the average diagnostic scores assigned by radiologists during assessments of the reconstructed images. The scoring system is condensed as follows: 1 for significantly worse, 2 for worse, 3 for similar, 4 for better, and 5 for significantly better compared to the reference image. The symbols on the plot indicate the computed average of expert scores. 90

Figure 5.5 Expert evaluations of 4x accelerated image reconstruction by DL methods. The radiologist assigned scores: 2 for UNET (b), 3 for iRIM (c), and 4 for E2E VarNET (d). Notably, in the case of E2E VarNET, the radiologist emphasized higher contrast in the femur medial condyle (yellow arrowed) compared to the reference image. Furthermore, background noise was effectively suppressed, indicating a significant enhancement in image quality beyond that of the reference image. These findings suggest that the DL model not only preserved but also improved certain aspects of the image, particularly in terms of contrast and noise reduction. 91

Figure 5.6 The plot showcases the average scores derived from radiological assessments of pathological features in the reconstructed images. The scoring system employed is as follows: 0 indicates that the pathology is not observable, 1 for far worse, 2 for worse, 3 for similar, and 4 for better compared to the reference image. 92

Figure 5.7 Hallucinations in DL-Based Reconstruction Outcomes: (a) Reference fully sampled image; (b) Arrows indicate DL-generated "Bone Marrow Edema" not present in (a). (c) Reference fully sampled image; (d) DL-generated increased pseudo and linear signals in the meniscus, posing misdiagnosis risks. (e) Fully sampled image; (f) DL-generated fogging-style artifacts. (g) Fully sampled image; (h) DL-generated lineal artifacts surrounding femur distal and tibia. 93

Figure 6.1 Unified multi-coil MRI reconstruction and analysis pipeline: (1) Acquired data is initially undersampled, resulting in an aliased image. (2) The aliased image is then processed through three different reconstruction models to transform it into a high-quality image. (3) These reconstructed outputs are further preprocessed through scaling, rotation, cropping, and resizing. (4) Following the preprocessing, the region of interest (in this case, the meniscus tear) is localized and trained using transfer learning models with fully sampled scans. (5) Finally, the trained models are tested with different reconstructed scans for comprehensive analysis.. 104

Figure 6.2 Illustrations of MRI scans depicting both healthy and meniscus tear cases. The top row showcases scans acquired with a 1.5T scanner, while the bottom row displays scans obtained using a 3T scanner. The first two columns represent healthy scans, and the subsequent two columns depict meniscus tear cases. In the first and third columns, Proton-Density (PD) contrast is utilized, while the second and fourth columns employ Proton-Density with Fat Suppression (PDFS) contrast. The presence of a meniscus tear is indicated by a white arrow. 106

Figure 6.3 Different types and orientations of meniscus tears within the menisci. The top row displays fully sampled scans, and the rows just below illustrate specific tear instances with corresponding cropped views: (a) Bucket-handle meniscus tear shown in a cropped view. (b) Double tears on the left side observed in the cropped view. (c) Multiple tears showcased in the cropped view. The last two rows represent cropped scans of the corresponding slice reconstructed at 4-fold and 8-fold accelerations. Each tear is marked with a white arrow. 107

Figure 6.4 Visual representation of cropped scan coordinates (highlighted in red) utilized as a frame of reference. Both the training and validation datasets are cropped according to the designated red box coordinates, while yellow boxes indicate FastMRI+ annotations. (a) Heatmap presenting the training dataset with the selected red box coordinate, encompassing all bounding boxes. (b) An example scan extracted from the training set. (c) Heatmap showcasing the validation dataset, also including all bounding boxes in the validation set. (d) An illustrative scan from the validation set. 108

Figure 6.5 Study design encompassing model training and testing phases. (1) Undersampling of k-space and generation of zerofilled aliased images. (2) Utilization of three reconstruction models for high-quality image generation. (3) Collection and preprocessing of "Meniscus Tear" and "Healthy" data for transfer learning. (4) Training of models and testing with different reconstruction outcomes. (5) Classification results obtained. 109

Figure 6.6 An illustration of a validation dataset image used for testing (a), where a previously unseen meniscus tear was accurately classified by AlexNET. Additionally, (b) presents the same image with a superimposed saliency mask, highlighting the network's focus on non-relevant areas, often overlooking torn menisci. Moreover, (c) displays the identical image cropped based on designated annotated coordinates, showcasing the targeted region of interest, while (d) shows the corresponding cropped image with an overlay of the saliency map, indicating areas of significance detected by the model. 110

Figure 6.7 An example slice displaying a meniscus tear is showcased, along with the corresponding ground truth bounding box (highlighted in yellow). The zoomed-in region, indicated by a white arrow, specifically highlights the presence of the meniscus tear. Notably, the classification networks exhibited less favorable performance in cases of ZeroFilled FFT reconstruction (RSS), attributable to its low Signal-to-Noise Ratio (SNR) and pronounced artifacts at an acceleration rate of 8. A comparative analysis reveals that, in terms of reconstruction, UNET, iRIM, E2E VarNET, and their respective ensembles consistently yield intricate textures, with the exception of UNET at acceleration rates of 4 and 8. 116

Figure 6.8 Meniscus Tear Detection Using Class Activation Map (CAM) with AI Classifiers Achieving High Probability Scores. Class activation maps (CAM) reveal the specific regions in images contributing to accurate meniscus tear classification. Images (a) and (b) depict 4-fold and 8-fold accelerated and reconstructed scans. Probability scores for each classifier, including ResNET50, AlexNET, VGG16, VGG19, ResNET50, ResNET101, GoogleNET, and NasNET Mobile, are presented from left to right. The CAM analysis effectively identifies meniscus tears, with high-importance areas highlighted in red. 119

Figure 6.9 Meniscus Tear Localization with Class Activation Map (CAM) in Reconstructed Scans: AI-classifiers exhibit high probability scores for identifying meniscus tears but may mislocate them at 4-8 fold acceleration rates. Images in (a) and (b) depict 4-fold and 8-fold accelerated and reconstructed scans, respectively. Probability scores for each class are provided by various classifiers, including ResNET50, AlexNET, VGG16, VGG19, ResNET50, ResNET101, GoogleNET, and NasNET Mobile, ordered from left to right. The CAM analysis effectively highlights critical areas in red. 120

LIST OF TABLES

Page

Table 3.1 Categorization of Volumes and Slices in Each Set..... 20

Table 3.2 MSE, NMSE, PSNR, and SSIM for single-coil fastMRI dataset reconstruction baselines for All, PD and PDFS accelerated (Acc.) at 4x and 8x. 29

Table 3.3 MSE, NMSE, PSNR, and SSIM metrics for the reconstruction of the single-coil fastMRI dataset using basic ensemble methods. Results are presented for acceleration rates of 4x and 8x in the All, PD, and PDFS contrasts..... 30

Table 3.4 Performance of Base Learner Results on Multi-Coil Validation Data ... 32

Table 3.5 Performance of Ensembles of Base Learners on Multi-Coil Validation Data 33

Table 4.1 Summary of Meniscus Tear and Artifact Labels with Subset Divisions. 47

Table 4.2 Reconstruction performances: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) along with their standard deviations for fastMRI dataset reconstruction baselines for All, PD and PDFS acquisitions accelerated at 4x and 8x. 57

Table 4.3 Reconstruction performances of the ensemble methods: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) along with their standard deviations for fastMRI dataset reconstruction baselines for All, PD, and PDFS acquisitions accelerated at 4x and 8x. 59

Table 4.4 Detection performances: FROC score IOU 0.1 and AP@0.1 Max. Det. 100 results for All, PD, and PDFS Accelerated at 4x and 8x. 61

Table 4.5 The proposed framework for meniscus tear detection compared with existing methods available for this task..... 62

Table 5.1 Categorization of Meniscus Tear and Artifact Labels with Subset Breakdown 79

Table 5.2 Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) scores, along with their corresponding standard deviations, are presented for the baseline reconstructions of the fastMRI dataset. The evaluations cover All, PD, and PDFS acquisitions accelerated at 4x and 8x. 87

Table 5.3 Detection Performances: AP@0.1, Max. Det. 100, Precision, and Recall results for All (ALL), Proton Density (PD), and Proton Density with Fat Suppression (PDFS) accelerated at 4x and 8x..... 88

Table 5.4 Performance of Blind Expert Annotations with FastMRI+ Reference: Mean Average Precision (mAP), Precision, and Recall Results for Accelerated All (ALL), Proton Density (PD), and Proton Density with Fat Suppression (PDFS) at 4x and 8x. 95

Table 5.5 Performance of nnDetection with Blind Expert Study Annotations as Reference: Mean Average Precision (mAP), Precision, and Recall Results for Accelerated All (ALL), Proton Density (PD), and Proton Density with Fat Suppression (PDFS) at 4x and 8x. 96

Table 6.1 Summary of Meniscus Tear Labels and Healthy Subjects with Subset Divisions 105

Table 6.2 Comparison of Meniscus Tear Classification Results on Different Reconstruction Methods..... 123

Table 6.3 Comparison of average SSIM between True Positives (TP) and False Negatives (FN) for reconstructed slices at 4x and 8x acceleration rates, across different classifiers. Scores in bold represent the highest SSIM values 124

LIST OF SYMBOLS

B_0	: External static magnetic field
M	: Bulk magnetization vector
$M^0(x, y, z)$	: Equilibrium magnetization distribution
N_s	: Number of individual nuclear magnetic moments
μ	: Magnetic moment vector
^{1}H	: Protons in hydrogen nuclei
γ	: Gyromagnetic ratio
T_1	: Time constant for the recovery of longitudinal magnetization
T_2	: Time constant for the decay of transverse magnetization
G_x, G_y, G_z	: Gradient magnetic fields along the x, y, and z axes
TR	: Repetition time of the pulse sequence
TE	: Echo time
$s(t)$	: Recorded signal as a function of time
k_x, k_y, k_z	: Spatial frequencies in k-space
$\rho(r)$	: Proton density
k	: Constant
x_{rec}	: Reconstructed image
λ	: Regularization parameter
x_u^k	: Undersampled k-space data
$\Psi(x)$	: Sparsity transform
$N(x_u; \theta^*)$	: Output of a neural network with parameters θ^*
N_c	: Number of data consistency layers

$N_{i,d}$: Number of convolutional blocks between consecutive data consistency layers

l_2-norm : Euclidean norm or 2-norm

l_1-norm : Absolute norm or 1-norm

F_u^{-1} : Inverse Fourier transform applied to the undersampled k-space

Σ : Sigma - Symbol for summation

ζ : Zeta - Represents a constant

α : IOU threshold

TP : True Positive

FP : False Positive

FN : False Negative

mAP : Mean Average Precision

N : Total number of classes

AP_k : Average precision of class k

μ_x : Average pixel intensity in the image x

μ_y : Average pixel intensity in the image y

μ_x^2 : Variance of pixel intensities in the image x

μ_y^2 : Variance of pixel intensities in the image y

σ_{xy} : Covariance between pixel intensities of images x and y

C_1 : Constant used in the SSIM calculation

C_2 : Constant used in the SSIM calculation.

ABBREVIATIONS

MRI	: Magnetic Resonance Imaging
MR	: Magnetic Resonance
CT	: Computed Tomography
RF	: Radio-frequency
PD	: Proton Density
PDFS	: Proton Density Fat Suppression
CS	: Compressed Sensing
HGGDP	: Homotopic Gradients of Generative Density Priors
SFL	: Spatial and Fourier Layer
Co-VeGAN	: Complex-Valued Generative Adversarial Network
DL	: Deep Learning
FastMRI	: Fast Magnetic Resonance Imaging
qMRI	: Quantitative Magnetic Resonance Imaging
iRIM	: Invertible Recurrent Inference Machine
E2E	: End-to-End
PSNR	: Peak Signal-to-Noise Ratio
SSIM	: Structural Similarity Index
MSE	: Mean Squared Error
NMSE	: Normalized Mean Squared Error
SNR	: Signal-to-Noise Ratio
CNR	: Contrast-to-Noise Ratio
PI-RADS	: Prostate Imaging Reporting and Data System

T2WI : T2-weighted Imaging

DWI : Diffusion-weighted Imaging

DWIBS : Diffusion-weighted Imaging with Background Suppression

ML : Machine Learning

E2E VarNET : End-to-End Variational Network

PD-Net : Primal-Dual Net

DONet : Dual-Octave Network

SDEs : Stochastic Differential Equations

HP-ALF : Hierarchical Perception Adversarial Learning Framework

CAMP-Net : Consistency-Aware Multi-Prior Network

jCAN : Joint Cross-Attention Network

PnP-HQS : Plug and Play Half-Quadratic-Splitting

RNLFNet : Residual Non-Local Fourier Network

GCESS : Graph Convolutional Enhanced Self-Similarity Network

k-UNN : k-Space Interpolation With Untrained Neural Network

PDHG : Primal-Dual Hybrid Gradient

LUNA16 : Lung Nodule Analysis 2016

ADAM : Anomaly Detection in Abdominal Multimodalities

CADA : Cerebral Aneurysm Detection and Annotation

RibFrac : Rib Fracture Detection

LIDC-IDRI : Lung Image Database Consortium and Image Database Resource Initiative

Kits19 : Kidney Tumor Segmentation Challenge 2019

SKM-TEA : Stanford Knee MRI with Multi-Task Evaluation

OAI : Osteoarthritis Initiative

MRNet	: Magnetic Resonance Network
RBF	: Radial Basis Function
Av.	: Averaging
4x	: 4-fold acceleration
8x	: 8-fold acceleration
RMS Prop	: Root Mean Square Propagation
IOU	: Intersection over Union
NMS	: Non-Maximum Suppression
FP	: False Positive
FN	: False Negative
TP	: True Positive
TN	: True Negative
FROC	: Free-Response Receiver Operating Characteristic
AP	: Average Precision
FPPI	: False Positives Per Image
mAP	: Mean Average Precision
IQA	: Image Quality Assessment
TSE	: Turbo Spin Echo
NYU	: New York University
2D / 3D	: Two-Dimensional / Three-Dimensional
PI	: Parallel Imaging
Adam optimizer	: Adaptive Moment Estimation optimizer
PReLU	: Parametric Rectified Linear Unit
ReLU	: Rectified Linear Unit

CNN	: Convolutional Neural Network
RNN	: Recurrent Neural Network
DNN	: Deep Neural Network
VGG-16	: Visual Geometry Group 16
VGG-19	: Visual Geometry Group 19
ResNet50	: Residual Network 50
ResNet101	: Residual Network 101
RSS	: Root Sum Square
AI	: Artificial Intelligence
CAM	: Class Activation Map
FFT	: Fast Fourier Transform
Acc.	: Acceleration
Ac.	: Accuracy
Prec.	: Precision
Rec.	: Recall
Spec.	: Specificity
F_1 Sc.	: F_1 Score
GRAPPA	: GeneRalized Autocalibrating Partial Parallel Acquisition
SENSE	: SENSitivity Encoding
1.5T / 3T	: 1.5 Tesla / 3 Tesla
Max. / Min.	: Maximum / Minimum

CHAPTER ONE
INTRODUCTION

Magnetic Resonance Imaging (MRI) stands as a paramount diagnostic tool, renowned for its exceptional soft tissue contrast that facilitates non-invasive diagnosis. However, the sluggish data acquisition process inherent to MRI presents challenges, leading to prolonged waiting times for patients. In response to this issue, extensive research has been dedicated to accelerating MR acquisitions. A prevalent approach involves undersampling in k-space, followed by the reconstruction of missing samples using specific algorithms.

Compressed Sensing (CS), due to the intrinsic compressible nature of MR images, has emerged as a potent technique for recovering undersampled acquisitions (Lustig et al., 2008, 2007; Haldar et al., 2010; Quan et al., 2018). CS addresses the ill-posed inverse image recovery problem by incorporating regularization in the form of sparsity in known transform domains. Despite its success in enhancing scan efficiency, challenges arise from pre-defined sparsifying transform domains and suboptimal hyperparameter selection, which can hinder recovery performance.

To overcome these limitations, Deep Learning (DL) techniques have surfaced as promising alternatives, exhibiting superior image quality even at high acceleration rates (Yang et al., 2017; Lee et al., 2018; Sun et al., 2016; Schlemper et al., 2017; Hammernik et al., 2018; Akçakaya et al., 2019 ; Zhu et al., 2018; Hyun et al., 2018; Han et al., 2018; Cheng et al., 2018; Dar et al., 2020; Van Nguyen et al., 2015; Han et al., 2019; Wang et al., 2017; Sriram et al., 2020; Putzky & Welling, 2019; Montalt-Tordera et al., 2021; Aggarwal et al., 2018; Eo et al., 2018; Adler & Öktem, 2018; Schlemper et al., 2017). Unlike CS, DL employs neural networks to learn the transform domain from extensive datasets without the need for online hyperparameter tuning. While DL exhibits remarkable recovery performance, challenges arise from the dependence on large datasets, which may not be readily available in some cases, hindering reproducibility and comparison with existing literature.

The FastMRI dataset has alleviated these challenges by providing MRI knee raw

volumes, including both single-coil and multi-coil data. Despite the success of DL techniques in MRI reconstruction, there is a gap in the analysis of identifying and quantifying pathology in reconstruction outcomes. Some studies have addressed this limitation by performing reconstruction and classification/detection tasks, both with and without an unified pipeline (Patsanis et al., 2023; Ebner et al., 2020; Baalbaki et al., 2023; Litjens et al., 2017; Zhao et al., 2021).

Moreover, the degraded performance of reconstruction approaches at very high acceleration rates necessitates exploration beyond traditional DL models. Ensemble models, combining the strengths of deep learning and ensemble learning, offer a potential solution, enhancing generalization performance (Ganaie et al., 2022). Scarcity of literature on this topic underscores the necessity for a comprehensive investigation into the potential benefits of ensembles for MR reconstruction tasks. One such approach focuses on Sensitivity Networks, utilizing diverse networks with varying losses to compare image enhancement and reconstruction approaches within the same model capacity (Kerstin et al., 2019). This study aims to bridge existing gaps in the field by exploring the effectiveness of ensemble models for MR reconstruction tasks and conducting a thorough analysis of pathology identification and quantification in the reconstructed images.

In this thesis, we contribute to addressing various challenges in the field of medical imaging. Firstly, we investigate ensemble methods in deep neural networks for MRI reconstruction, specifically applying them to the FastMRI dataset. Our exploration involves evaluating off-the-shelf models like U-Net (UNET), i-RIM (iRIM), End-to-End (E2E) Variational Network for the multi-coil dataset, and established networks like KIKI-NET and UNET for the single-coil dataset. The utilization of three non-trainable ensemble combiners —Max/Min, Averaging, and Weighted Averaging Combiner— provides comprehensive insights into their impact on reconstruction performance. Through rigorous analysis using metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Mean Squared Error (MSE), and Normalized Mean Squared Error (NMSE), we highlight the marked superiority of ensemble models over individual ones across both single-coil and

multi-coil datasets.

Next, our focus shifts to the intricacies of DL-detected meniscal tears in retrospectively undersampled knee MRI. A detailed analysis uncovers challenges presented by hallucinations and false positive artifacts in DL reconstructions, shedding light on the influence of reconstruction quality on the identification of clinically relevant details. Furthermore, we explore the alignment between DL detectors and traditional image quality metrics, along with expert evaluations, offering a comprehensive understanding of the reliability of DL-based detection outcomes compared to conventional imaging methods. We then extend our investigation by proposing an innovative integrated pipeline that combines DL reconstruction and detection for meniscal tears on fastMRI+ data. The results showcase state-of-the-art performance, emphasizing the effectiveness of this integrated approach in clinical applications.

Furthermore, our study contributes significantly to the field through the inclusion of both blind and non-blind evaluations. In the blind study, expert radiologists meticulously annotate meniscal tears without prior knowledge of the acceleration levels or reconstruction methods applied. This unbiased evaluation provides valuable insights into the effectiveness of the DL reconstruction and detection pipeline, offering a rigorous assessment of its performance without any preconceived information. In contrast, the non-blind study evaluates reconstructed images, taking into account specific acceleration factors and reconstruction models used. This assessment, covering diagnostic and pathological features, adds depth to our understanding of the clinical applicability and robustness of the proposed DL reconstruction and detection system. By incorporating both blind and non-blind evaluations, our research not only addresses the challenges of meniscal tear detection but also establishes a comprehensive framework that enhances the reliability and clinical relevance of DL-based approaches in accelerated knee MRI.

Finally, we broaden our study to explore the untapped potential of DL methods in MRI, with a specific focus on accurate classification of meniscal tears using

undersampled k-space data and DL-produced reconstruction outcomes. Our innovative sequential process, encompassing both reconstruction and classification stages, achieves outstanding performance in accuracy, precision, recall, specificity, and F_1 score. The study also highlights consistent accuracy across models, even when focusing on varying image zones, as revealed by activation maps. Importantly, our results challenge the conventional assumption that elevated SSIM scores necessarily correlate with improved classifier performance across all DL reconstruction models at different acceleration rates.

The structure of the thesis is outlined as follows. Chapter 2 provides a concise overview of fundamental concepts in MRI data acquisition and accelerated MRI. The subsequent chapters, namely 3, 4, and 5, are dedicated to the contributions mentioned earlier. In Chapter 3, ensemble methods in deep neural networks for MR image reconstruction are elaborated, with a specific exploration of their application using the FastMRI dataset. Chapter 4 delves into the complexities of DL-detected meniscal tears in retrospectively undersampled knee MRI. Additionally, Chapter 5 provides distinct discussions on blind and non-blind expert studies, contributing to a comprehensive understanding. Moving on to Chapter 6, the focus is on the precise classification of meniscal tears within the framework of undersampled k-space data and the outcomes of DL-produced reconstruction. The thesis concludes with Chapter 7, encompassing final remarks and a discussion on potential avenues for future research.

CHAPTER TWO
MRI ESSENTIALS

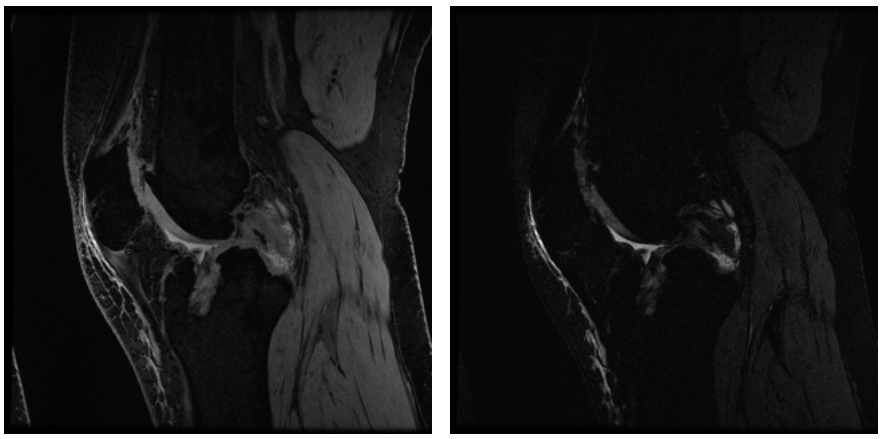
2.0.1 Magnetic Resonance Imaging

In MRI, we use signals from tissues' nuclear magnetic resonance properties, activated by fixed magnetic and changing radio frequency fields. What makes MRI versatile is that operators can adjust pulse sequences to control how these fields are applied over time. Similar to nuclear medicine, MRI can create different images, each showing unique signals. By combining pulse sequences cleverly, we can tell the difference between fat and water, check water diffusion, see bone structures, measure motion, and estimate blood flow.

A regular MRI machine looks like a CT scanner, with a big structure through which patients on a bed pass. Inside this structure, there are tools for applying magnetic and radio frequency fields, and receiver coils that capture the raw signal data. Like CT, MRI gives us cross-sectional images, and the way we rebuild these images, though similar to CT, comes directly from the characteristics of the raw signals in frequency space.

In practice, we mainly use MRI to see anatomy, creating high-contrast cross-sectional images of the entire body. But MRI goes a step further by offering different ways to get images—by tweaking the pulse sequences—to show different contrasts and reveal pathology more effectively (Prince & Links, 2023).

Figure 2.1 illustrates various image pairs, highlighting the differential diagnostic capabilities of different pulse sequences. In Figure 2.1 (a), the Echo 1 image of the knee reveals a horizontal, complex, and extrusion meniscus tear, along with a ligament tear. Conversely, Figure 2.1 (b) displays an Echo 2 image capturing the same tear while also effectively revealing changes in cartilage lesions.



(a) (b)

Figure 2.1 A knee MRI slice captured using Echo 1 (a) and Echo 2 (b) sequences.

2.0.1.1 Physics of Magnetic Resonance

In whole-body MRI, the primary focus is on the abundant ^{1}H nuclei within our bodies, primarily due to water content. These ^{1}H nuclei, similar to spinning protons, play a crucial role in generating a robust MRI signal. While the common expression in MRI refers to imaging protons, it specifically pertains to ^{1}H nuclei, distinguishing them from protons in other atoms. Each nucleus, with a positive charge and spin, contributes to the creation of a microscopic magnetic field represented by a magnetic moment vector μ.

Considering a specific spin system, such as the nuclei of ^{1}H within a material volume, the application of an external static magnetic field (B_0) results in magnetization. This magnetization process is accurately characterized by a bulk magnetization vector (M), representing the cumulative effect of a large number (N_s) of individual nuclear magnetic moments.

$$M = \sum_{n=1}^{N_s} \mu_s \tag{2.1}$$

If the sample remains undisturbed in the static magnetic field (B_0) for an extended period, M will eventually stabilize at an equilibrium value.

In the context of magnetic resonance imaging (MRI), applying a radio-frequency (RF) field at the Larmor frequency of protons plays a pivotal role. This application steers the net magnetization into the transverse direction, influencing the subsequent imaging process. The duration of the RF pulse is crucial, determining the magnetization along both the longitudinal and transverse directions. Following the application of the RF pulse, the net transverse magnetization experiences decay with a time constant T_2, signifying the rate at which the signal diminishes. Simultaneously, the net longitudinal magnetization undergoes recovery with a time constant T_1, illustrating the pace of return to the equilibrium state.

The dynamic interplay between the magnetic field and net magnetization is mathematically described by the Bloch equation:

$$\frac{d\mathbf{M}}{dt} = \mathbf{M} \times \gamma\mathbf{B} - \frac{M_x\mathbf{i} + M_y\mathbf{j}}{T_2} - \frac{M_z\mathbf{i} + M_0\mathbf{k}}{T_1} \tag{2.2}$$

Here, $\mathbf{M}$ is the net magnetization vector, γ is the gyromagnetic ratio specific to the nucleus, $\mathbf{B}$ is the applied magnetic field vector, and M_x, M_y, and M_z represent the respective components of the net magnetization along the x, y, and z axes. The terms T_1 and T_2 are the time constants associated with the recovery of longitudinal magnetization and the decay of transverse magnetization, respectively. This equation provides a quantitative framework for understanding the dynamic behavior of magnetization in the MRI process.

MRI employs receiver coils to detect changes in flux along the transverse direction. However, the received signal lacks spatial information, reflecting alterations in overall flux throughout the volume. To introduce spatial encoding, additional gradient magnetic fields (G_x, G_y, G_z) are applied along the (x, y, z) axes, complementing the main static magnetic field (B_0). This introduces a spatial variation in the magnetic field across the imaging volume. By manipulating these gradient fields, we can discern the spatial distribution of magnetization. The signal recorded by the receiver coil is mathematically expressed as:

$$s(t) = \int_x \int_y \int_z M^0(x, y, z) \exp^{-t/T_2(r)} \exp^{-i\gamma \int_0^t G(\tau) r . d\tau} \, dx \, dy \, dz \tag{2.3}$$

Here, $s(t)$ represents the recorded signal as a function of time t, $M^0(x, y, z)$ is the equilibrium magnetization distribution, $T_2(r)$ is the relaxation time constant, $G(\tau)$ is the gradient magnetic field as a function of time τ, and γ is the gyromagnetic ratio. The integral encompasses the spatial dimensions (x, y, z) of the imaged volume. This equation encapsulates the temporal and spatial dynamics in the recorded MRI signal.

Traditionally, the integral involving gradient terms is substituted with k_x, k_y, and k_z, collectively known as k-space. By defining $k_{x(t)} = \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau$, $k_{y(t)} = \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau$, and $k_{z(t)} = \frac{\gamma}{2\pi} \int_0^t G_z(\tau) d\tau$, the signal received at each time can be formulated as:

$$s(t) = \int_x \int_y \int_z M^0(x, y, z) \exp^{-t/T_2(r)} \exp^{-i2\pi[k_{x(t)} + k_{y(t)} + k_{z(t)}]} \, dx \, dy \, dz \tag{2.4}$$

This expression represents the 3D Fourier transform of the imaged volume, where the signal at each time corresponds to a Fourier transform at a spatial frequency $k = [k_x, k_y, k_z]$. Adjusting G determines the sampling trajectory in k-space.

The manipulation of transverse magnetization using RF pulses and gradient fields aids in imaging specific tissue properties, enabling the acquisition of multiple contrasts. For example, the widely used spin-echo pulse sequence yields an acquired signal expressed as:

$$S_0 = k\rho(r)[1 - \exp^{-TR/T_1(r)}] \exp^{-TE/T_2(r)}$$

Within this equation, $\rho(r)$ represents proton density, k stands as a constant, TR signifies the repetition time of the pulse sequence, and TE denotes the echo time. The adjustment of TR and TE is instrumental in accentuating either T_1-weighted or T_2-weighted contrasts within the images, providing flexibility in catering to diverse imaging preferences. However, Proton Density (PD)-weighted contrasts exhibit a

slight divergence from T_1 and T_2 contrasts as they are linked to the quantity of nuclei in the imaged area, specifically the number of hydrogen protons, rather than the magnetic characteristics of these nuclei. These PD-weighted images, originating from the first echo, emerge when the influence of both T_1 and T_2 contrast is minimized.

2.0.2 Accelerated MRI

Despite the crucial soft tissue contrast it provides for non-invasive diagnosis, the inherently slow data acquisition process of MRI poses challenges, leading to prolonged wait times for patients. In response to this issue, the pursuit of accelerated MR acquisitions without compromising image quality emerges as a practical remedy. This involves employing advanced techniques and technologies aimed at expediting the imaging process and, significantly, implementing strategic data reduction in k-space.

In k-space, where raw data is initially acquired, reducing the amount of information collected can be strategically applied without sacrificing the diagnostic value of the images. This streamlined approach not only accelerates the imaging process but also optimizes data storage and processing resources. Subsequently, employing sophisticated reconstruction algorithms becomes paramount to recreate a high-fidelity image from the reduced k-space data.

By strategically reducing and reconstructing data in k-space, not only can patient waiting times be significantly reduced, but the overall efficiency and accessibility of MRI procedures can also be enhanced. This approach not only addresses the challenge of prolonged wait times but also contributes to the continuous advancement of MRI technology, ultimately leading to improved healthcare outcomes.

2.0.2.1 Reconstruction

A widely used method to speed up MRI acquisition is undersampling in k-space, followed by a customized reconstruction approach. This technique selectively skips some k-space samples, exploiting redundancies in the data while preserving diagnostic precision. The recovery problem within this framework can be expressed as:

$$x = F_u^{-1} x_u^k \tag{2.5}$$

In the given equation, F_u^{-1} represents the inverse Fourier transform operator at the undersampled k-space locations, where x is the image targeted for reconstruction, and x_u^k corresponds to the acquired k-space data. It is crucial to note that Equation 2.5 characterizes an ill-posed inverse problem, necessitating prior information about x. This essential information is introduced in the form of a regularization term $\Psi(x)$:

$$x_{rec} = \underset{x}{\operatorname{argmin}} \|x - F_u^{-1} x_u^k\|_2^2 + \lambda \|\Psi(x)\|_1 \tag{2.6}$$

In the image recovery formulation, F_u^{-1} stands for the inverse Fourier transform operator applied to the undersampled k-space locations. Here, x represents the image undergoing reconstruction, x_u^k is the acquired k-space data, Ψ denotes the sparsity transform, and λ serves as a regularization parameter. The objective is to minimize the l_2-norm between the reconstructed image and the inverse Fourier transform of the acquired k-space data. Simultaneously, sparsity in the transformed image domain is encouraged through the l_1-norm term with the regularization parameter λ.

Due to the inherent compressible nature of MR images, compressed sensing (CS) has emerged as a prominent technique for recovering undersampled MR acquisitions (Lustig et al., 2008, 2007; Haldar et al., 2010; Quan et al., 2018). In CS, the inverse image recovery problem in Equation 2.6 is addressed by associating $\Psi(x)$ with sparsity in a predefined transform domain. Despite its advancements in scanning efficiency,

CS relies on predetermined sparsifying transform domains, which may not always be optimal. Additionally, achieving optimal performance in CS often requires careful tuning of hyperparameters.

Recently, there has been a notable shift towards utilizing Deep Neural Networks (DNNs) in MRI reconstruction, marking a departure from reliance on sparsity priors in predefined transform domains. This paradigm shift directly tackles the limitations encountered in traditional methods. Deep Learning (DL) techniques, notably, have emerged as promising alternatives, showcasing superior image quality even under high acceleration rates (Yang et al., 2017; Lee et al., 2018; Sun et al., 2016; Schlemper et al., 2017; Hammernik et al., 2018; Akçakaya et al., 2019; Zhu et al., 2018; Hyun et al., 2018; Han et al., 2018, 2019; Cheng et al., 2018; Dar et al., 2020; Van Nguyen et al., 2015; Wang et al., 2017; Sriram et al., 2020; Putzky & Welling, 2019; Montalt-Tordera et al., 2021; Aggarwal et al., 2018; Eo et al., 2018; Adler & Öktem, 2018; Schlemper et al., 2017). Unlike Compressed Sensing (CS), Deep Learning leverages neural networks to autonomously learn the transform domain from extensive datasets, eliminating the need for real-time hyperparameter tuning. This innovative approach not only overcomes the limitations associated with CS but also showcases superior performance across diverse scenarios. DNNs achieve this by learning these priors offline in a data-driven manner using large datasets consisting of pairs of fully sampled and undersampled images. The mathematical representation of this approach is captured in Equation 2.7:

$$x_{rec} = \underset{x}{\operatorname{argmin}} \|x - F_u^{-1} x_u^k\|_2^2 + \lambda \|N(x_u; \theta^*) - x\|_2 \tag{2.7}$$

In the equation presented, x_{rec} is determined as the solution to a minimization problem. This problem aims to minimize the l_2-norm, representing the difference between the reconstructed image x and the inverse Fourier transform of the acquired k-space data x_u^k. The operator F_u^{-1} symbolizes the inverse Fourier transform specifically applied to the undersampled k-space locations. Moreover, the regularization term introduces the l_2-norm, measuring the difference between the

output of a neural network $N(x_u; \theta^*)$ with parameters θ^* and the true image x. This regularization process is governed by the parameter λ. The formulation seamlessly integrates traditional image reconstruction principles with the capabilities of Deep Neural Networks (DNNs), enhancing the quality of MRI reconstruction.

This entire process is visually summarized in Figure 2.2, providing a comprehensive illustration of the key steps involved in MR image reconstruction. The figure outlines the data collection through undersampling in k-space and the subsequent application of a specialized reconstruction algorithm. The interplay between traditional image reconstruction principles and the advanced capabilities of Deep Neural Networks (DNNs) is visually represented, offering a clear overview of how this integrated approach contributes to the enhancement of MRI reconstruction quality.

2.0.2.2 Analysis

Despite the successful validation of deep learning techniques through standard image quality metrics, their ability to effectively evaluate clinically relevant details may be constrained. Even though these approaches showcase high reconstruction quality at accelerated rates, concerns arise regarding the visibility or preservation of clinically important details during the reconstruction phase. Although radiologists have previously compared the diagnostic equivalence of accelerated DL-reconstructed images to conventional ones for assessing knee internal derangement, uncertainties persist regarding the accuracy of DL detectors/classifiers in identifying crucial clinical details when utilizing reconstruction models. These disparities in clinically relevant measures pose challenges to clinical deployment and have the potential to undermine clinician trust in the technology. Figure 2.3 provides a visual representation of the sequential pipeline for further context.

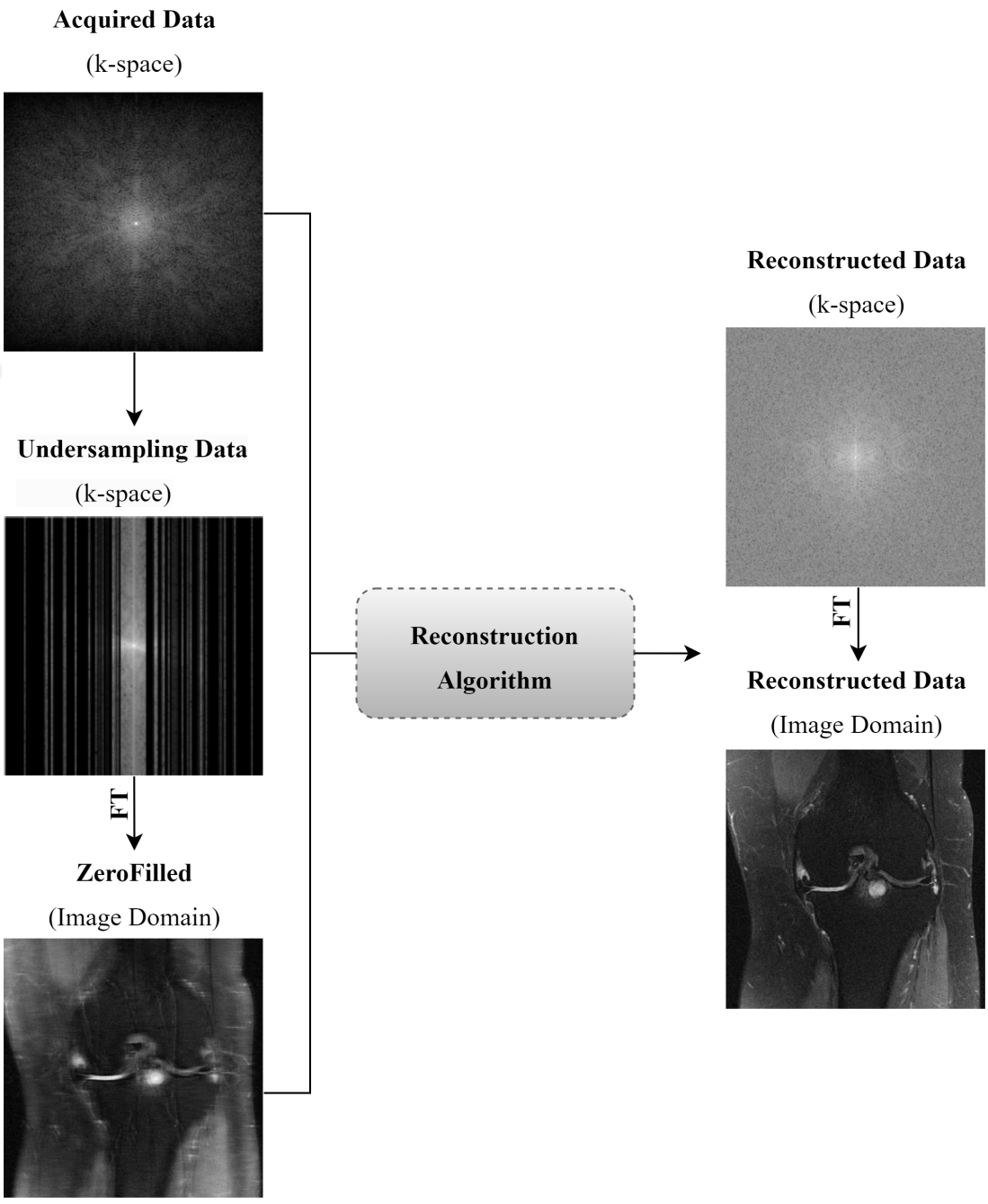


Figure 2.2 This overview outlines the MR image reconstruction process, which involves collecting data with undersampling in k-space. The acquired data are then processed using a specialized reconstruction algorithm, highlighting the key steps in achieving high-quality MR images.

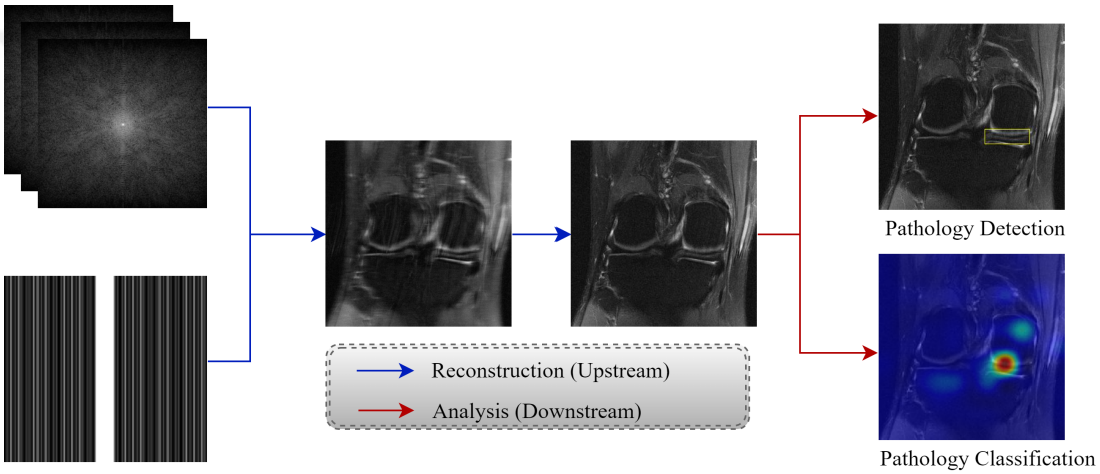


Figure 2.3 An overview of the integrated multi-coil MRI pipeline and its associated machine learning tasks. The process initiates with the transformation of undersampled data collected from multiple coils into high-quality images (depicted by the blue arrow during the reconstruction stage). Following this, image analysis tasks, including the localization of pathology regions of interest (e.g., detection and classification), are carried out.

CHAPTER THREE
ENSEMBLE DEEP LEARNING MODELS FOR ENHANCED MRI RECONSTRUCTION: A FastMRI STUDY

3.1 Introduction

In recent years, the field of MRI reconstruction has witnessed a significant surge in research, particularly leveraging machine learning methodologies, driven by notable advancements in algorithms and software platforms (Abadi et al., 2016; Paszke et al., 2019). Until recently, MRI reconstruction studies were constrained by small datasets, often requiring agreements with scanner vendors for data acquisition. Consequently, numerous research groups faced challenges due to inadequate infrastructure for assembling the requisite datasets, leading to difficulties in achieving reproducible results and comparing methodologies with existing literature (Knoll et al., 2020).

The availability of the fastMRI dataset has addressed these limitations by providing MRI knee raw volumes, encompassing both single-coil and multi-coil data. This dataset has played a pivotal role in advancing MRI reconstruction methodologies, facilitating result replication and fair comparisons. Traditional techniques such as Parallel Imaging (PI) and Compressed Sensing (CS) have long been employed in musculoskeletal MRI as conventional methods (Pal & Rathi, 2022; Gu et al., 2021; Hou et al., 2023; Zhang et al., 2021; Yan et al., 2020; Knoll et al., 2020). However, their application is often restricted by artifacts and noise, limiting the achievable acceleration rate. Deep Learning (DL) techniques present a promising alternative, demonstrating enhanced image quality even at high acceleration rates.

For the single-coil dataset, significant advancements have been achieved in the literature, particularly concerning the fastMRI knee dataset. These approaches can be broadly categorized into supervised and unsupervised schemes.

In the realm of supervised methods, Liu et al. (Li et al., 2021) introduced an edge-enhanced dual discriminator generative adversarial network, effectively striking

a balance between abstract global high-level features and edge features. In a different vein, Zabihi et al. (Zabihi et al., 2021) proposed a separated UNET, reducing the number of trainable parameters while maintaining high-quality scans. The utilization of deformable convolutions by Xiao et al. (Xiao et al., 2021) provided a novel feature extractor, facilitating the extraction of valuable information in MRI reconstruction tasks.

Other researchers have explored unique methodologies such as homotopic gradients of generative density priors (HGGDP) by Quan et al. (Quan et al., 2021) to address challenges in low-dimensional manifolds and low data density regions. The introduction of Spatial and Fourier Layer (SFL) by Yi et al. (Yi et al., 2021) allowed simultaneous learning of local and global information in spatial and Fourier domains. Ziller et al. (Ziller et al., 2021) presented a complex-valued UNET model architecture, utilizing stochastic gradient descent training with ζ-DP to achieve effective results.

Recent developments in 2022 include Liu et al.'s (Liu et al., 2022) multiscale convolutional dictionary structure and Peng et al.'s (Peng et al., 2022) diffusion model-based MR reconstruction, which guides the generation process based on observed signals and a pre-trained diffusion model. Zhao et al. (Zhao et al., 2022) proposed a transformer-based framework for direct processing of signals in k-space, addressing limitations associated with regular grids. Vasudeva et al. (Vasudeva et al., 2022) introduced a complex-valued generative adversarial network (Co-VeGAN) framework, showcasing the ongoing exploration of diverse methodologies.

The benchmark provided by Ramzi et al. (Ramzi et al., 2020) has been instrumental in evaluating and comparing these advancements. This benchmark, including architectures like UNET (Ronneberger et al., 2015), Cascade-NET (Schlemper et al., 2017), KIKI-NET (Eo et al., 2018), and PD-NET (Adler & Öktem, 2018), has paved the way for standardized evaluations and facilitated meaningful comparisons within the community.

In the realm of sampling trajectory exploration, Huang et al. (Huang &

Ravishankar, 2022) introduced MNet, a data-driven sampler utilizing a convolutional neural network for adaptive, object-specific sampling patterns. Additionally, strategies for optimizing both sampling patterns and network parameters jointly have been presented by Aggarwal et al. (Aggarwal & Jacob, 2020).

Unsupervised approaches, where paired training data is absent, have allowed for the learning of flexible regularizers adaptable to changes in the forward problem. Variational autoencoders proposed by Duff et al. (Duff et al., 2022) and unsupervised denoising networks introduced by Moreno et al. (Moreno López et al., 2021) offer solutions to denoise MRI images in the complex image space using sophisticated algorithms.

In the domain of multi-coil MRI reconstruction, significant strides have been made, offering a spectrum of methodologies to address the complexities associated with capturing data from multiple receiver coils.

In 2021, Pawar et al. (Pawar et al., 2021) introduced a novel loss function, combining mean absolute error, structural similarity, and Sobel edge loss, enhancing multi-coil MRI reconstruction. Darestani et al. (Darestani & Heckel, 2021) optimized untrained CNNs, demonstrating efficiency in handling multi-coil data. Yaman et al. (Yaman et al., 2021) pioneered a zero-shot self-supervised approach for accelerated MRI reconstruction.

Recent breakthroughs have ushered in a new era of architectural innovations and methodologies in 2022. Feng et al. (Feng et al., 2022) introduced the Dual-Octave Network (DONet), focusing on learning multiscale spatial-frequency features from both real and imaginary components of MR data. Fabian et al. (Fabian et al., 2022) proposed a hybrid architecture that seamlessly combines convolutional and Transformer blocks, showcasing proficiency in unrolled and multi-scale reconstruction. Addressing diffusion models with continuous stochastic differential equations (SDEs), Cao et al. (Cao et al., 2022) leveraged them as a generative prior for MR reconstruction. In 2022, Yan et al. (Yan et al., 2022) presented a framework for anatomy-shared and anatomy-specific parameterized learners.

Liu et al. (Liu et al., 2023) advanced the field with a multi-level densely connected framework designed for high-resolution MRI. Xin et al. (Xin et al., 2023) introduced a two-stage MRI reconstruction pipeline, while Cui et al. (Cui et al., 2023) proposed k-space Interpolation With Untrained Neural Network (k-UNN) for optimal random sampling trajectories. Gao et al. (Gao et al., 2023) brought forth a hierarchical perception adversarial learning framework (HP-ALF), and Zhang et al. (Zhang et al., 2023) showcased the Consistency-Aware Multi-Prior Network (CAMP-Net). Sun et al. (Sun et al., 2023) presented an unrolling-based joint Cross-Attention Network (jCAN), and Pal et al. (Pal et al., 2023) introduced the Weight Agnostic Randomly Weighted Network method. Rasti et al. (Rasti-Meymandi et al., 2023) demonstrated the effectiveness of sparse recovery with Plug and Play Half-Quadratic-Splitting (PnP-HQS), and Zhou et al. (Zhou et al., 2023) introduced the Deep Residual Non-Local Fourier Network (RNLFNet). Ma et al. (Ma et al., 2023) incorporated non-local information with Enhanced Self-Similarity Using Graph Convolutional Network (GCESS), while Kusakunniran et al. (Kusakunniran et al., 2023) introduced StrainNet, designed for encoding under-sampled data into high-level features in 2023.

Ensemble methodologies have exhibited notable success in classification tasks (Kuncheva, 2014) and segmentation competitions (Kamnitsas et al., 2018; Isensee et al., 2018), showcasing improved outcomes compared to individual classifiers. However, the application of ensembles in the realm of Magnetic Resonance (MR) reconstruction remains underexplored, introducing a significant gap in the literature. The limited literature on this topic underscores the need for a comprehensive investigation into the potential benefits of ensembles for MR reconstruction tasks.

Σ-Net stands as a singular ensemble study, focusing on Sensitivity Networks and employing diverse networks with varying losses to compare image enhancement and reconstruction approaches within the same model capacity (Kerstin et al., 2019). Motivated by this gap and the success of ensembles in other domains, our study aims to unravel the untapped potential of ensembles in the context of MR reconstruction.

In this research endeavor, we propose to assemble well-established deep learning

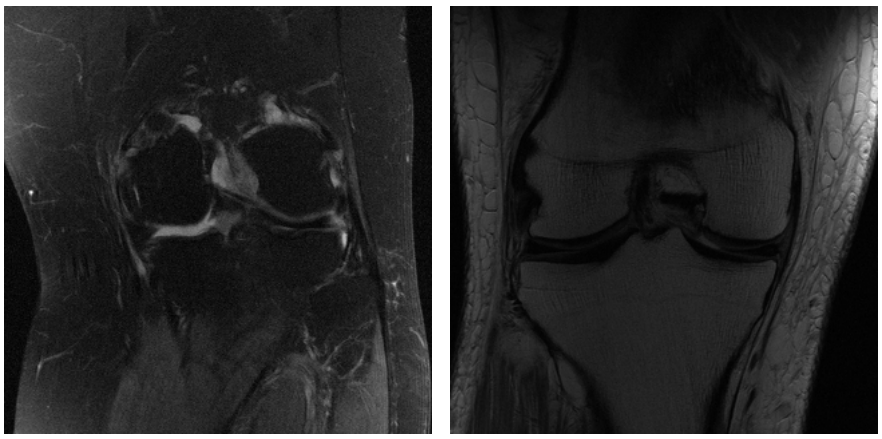
models that consistently yield impressive results with their default parameters, eliminating the need for extensive tuning. This strategic choice ensures that practitioners can readily implement our proposed system without the necessity for fine-tuning or structural modifications. The ensemble is constructed using three fundamental combination methods: Max/Min, Averaging, and Weighted Averaging Combiner.

The subsequent sections of this study are organized as follows: Section 3.2 provides an overview of the fastMRI dataset and its characteristics. Section 3.3 details the trajectories implemented for k-space undersampling. Section 3.4 offers an overview of the deep learning models and ensemble methods, along with combination rules tailored for both single-coil and multi-coil datasets. Section 3.5 introduces commonly used evaluation metrics, as well as perceptual and generative metrics relevant to the reconstruction task. In Section 3.6, we present both qualitative and quantitative results, culminating in our conclusion and a comprehensive discussion in Section 3.7.

3.2 FastMRI Dataset

The fastMRI dataset encompasses an extensive collection of raw MRI data, comprising 8,344 volumes and totaling 167,375 slices, derived from in vivo examinations across various MRI systems. Each fully sampled volume originates from either one of three clinical 3T systems (Siemens Magnetom Skyra, Prisma, and Biograph MR) or a clinical 1.5T system (Siemens Magnetom Aera). The data acquisition involves a conventional Cartesian 2D Turbo Spin Echo (TSE) protocol and a knee coil array with 15 channels, commonly used by the NYU School of Medicine. The dataset incorporates two pulse sequences, generating coronal proton-density with/without fat suppression (PDFS, 798 scans) and (PD, 796 scans) (Zbontar et al., 2018), as depicted in Figure 3.1.

Table 3.1 illustrates the distribution of volumes in each dataset. The training and



(a) (b)

Figure 3.1 A proton-density weighted slice (a) with fat suppression (PDFS) and (b) without fat suppression (PD).

validation datasets are employed for determining model parameters and hyperparameter values, while the test dataset serves for evaluating various approaches. It's noteworthy that the ground truths of the test dataset are not publicly disclosed to prevent model overfitting. For the assessment of model results on the test dataset, participants are required to submit their outcomes to the public leaderboard at http://fastmri.org/.

Table 3.1 Categorization of Volumes and Slices in Each Set

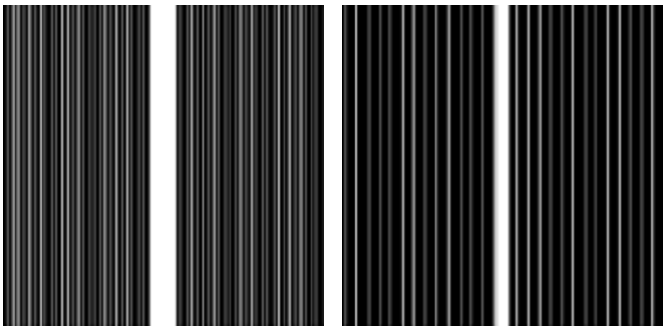
	Volumes		**Slices**	
	Multi-Coil	**Single-Coil**	**Multi-Coil**	**Single-Coil**
Training	973	973	34,742	34,742
Validation	199	199	7135	7135
Test	118	108	4092	3903
Challenge	104	92	3810	3305

In the course of this investigation, assessments were conducted using knee volumes from both multi-coil and single-coil configurations, facilitating a comprehensive evaluation of networks and subsequent analyses.

3.3 Trajectories

The undersampling segment involves retrospectively masking k-space lines from a fully sampled acquisition, with undersampled k-spaces incorporated in both test and challenge volumes. To emulate physically plausible accelerations in 2D data acquisitions, k-space lines are excluded in the phase encoding direction. A consistent undersampling mask is applied across all slices within a volume. The acceleration factors, representing four-fold or eight-fold acceleration, are arbitrarily set with equal probability for each.

To generate undersampling masks, a certain number of adjoint lowest frequency k-space lines are initially incorporated to form a fully k-space region. When the acceleration factor is four, the central region of the fully sampled includes 8% of all k-space lines, and for an acceleration factor of eight, 4% of all k-space lines are involved. For the knee dataset, the remaining k-space lines are uniformly and randomly included with the specified probability, ensuring that, on average, the undersampling mask achieves the desired acceleration factor. Figure 3.2 illustrates the k-space trajectories at four and eight-fold acceleration factors.



(a) (b)

Figure 3.2 K-space trajectories at a) 4-fold acceleration b) 8-fold acceleration

3.4 Methods

In this study, our primary objective is to capitalize on the advancements achieved in deep learning models, particularly in handling undersampled k-space data, which corresponds to aliased images. Moreover, we aim to enhance and extend these advancements through the utilization of ensemble methods.

3.4.1 Deep Learning Approaches for MRI Reconstruction in Single-Coil Knee Dataset

This study delves into the realm of magnetic resonance imaging (MRI) reconstruction, a domain significantly impacted by the advancements in deep learning models. Various state-of-the-art architectures, including UNET, Cascade-NET, Invertible Recurrent Inference Machine (iRIM), KIKI-NET, and Primal-Dual (PD) NET (PD-NET), have been employed to tackle the challenges posed by undersampled k-space data. The selected deep learning models in this study are described below:

- **UNET:** Several deep learning techniques, particularly those based on Convolutional Neural Networks (CNNs), have been explored for addressing MRI reconstruction challenges arising from undersampled k-space data. One prevalent architecture in this domain is the UNET, which has been widely applied in various image-to-image prediction tasks, including segmentation (Long et al., 2015) and reconstruction (Ramzi et al., 2020). The UNET model used in this study follows a standard design, featuring a down-sampling path followed by an upsampling path without an additional data consistency step. Specifically, it incorporates a cascade of 16, 32, 64, and 128 filters. Two final convolutional blocks were employed, with one comprising 4 filters and the last one featuring 2 filters (representing real and imaginary parts) without any non-linearity and having a kernel size of 1.

- **Cascade-NET:** Cascade-NET, as mentioned in (Schlemper et al., 2017), is

rooted in dictionary learning optimization (Caballero et al., 2014). It replaces the dictionary learning step with convolutional neural networks while retaining the data consistency step in k-space. The optimization algorithm is unrolled, allowing back-propagation through the linear data consistency step to obtain the corresponding Jacobian. Each convolutional block consists of 48 filters, and there are $N_{i,d}$=5 convolutional blocks between each consecutive pair of $N_c = 5$ data consistency layers. The final data consistency layer has only 2 filters, representing real and imaginary parts, without any non-linearity.

- **KIKI-NET:** The KIKI-NET (Eo et al., 2018) is an extension of the Cascade-NET, introducing a non-residual convolutional block to address k-space completion. Following the replacement of k-space values in sampled locations with the original measurements, a Convolutional Neural Network (CNN) is applied in this architecture. The KIKI-NET utilizes only two data consistency layers ($N_c = 2$), with $N_{k,d} = N_{i,d} = 16$ convolutional blocks in both image and k-space. Each block contains 48 filters, except for the last one, which has only 2 filters (representing real and imaginary parts) without any non-linearity and a kernel size of 1. The chosen non-linearity for this network is leaky ReLU. The training process involves sequential training of K-net, KI-net with fixed K-net weights, KIK-net, and finally the complete KIKI-NET.

- **Invertible Recurrent Inference Machine (iRIM):** To address accelerated MRI challenges in the fastMRI challenge (Knoll et al., 2020), an Invertible Recurrent Inference Machine (iRIM) was trained for both multi-coil and single-coil datasets (Putzky & Welling, 2019). The iterative learning-to-infer approach gains popularity as a solver in inverse problems, despite its memory requirements during training. This implementation employs an iterative inverse model with constant memory, relying on invertible networks to avoid storing intermediate activations. This approach allows training models with 400 layers on 3D MRI volumes for the reconstruction problem (Putzky & Welling, 2019). System checkpoints were directly downloaded and applied to the multi-coil, single-coil fastMRI data (Putzky et al., 2019).

- **Primal-Dual (PD) NET:** The Primal-Dual (PD) NET, proposed by Adler et al. (Adler & Öktem, 2018), is based on wavelet-based denoising and solving the corresponding optimization problem with Primal-Dual Hybrid Gradient (PDHG) algorithms (Chambolle & Pock, 2011). Cheng et al. (Cheng et al., 2019) implemented PD Net in the MRI dataset. The PDHG algorithm is unrolled, and convolutional neural networks replace the proximity operators present in the general case of PDHG. For a fair comparison with Cascade-NET and UNET, ReLU is applied instead of Parametric Rectified Linear Unit (PReLU) as a nonlinearity function (He et al., 2015). The parameters utilized in this context are consistent with the original implementation (Adler & Öktem, 2018), which includes $N_c = 10$, $N_{k,d} = N_{i,d} = 3$, and 32 filters in the convolutional blocks.

3.4.2 Deep Learning Approaches for MRI Reconstruction in Multi-Coil Knee Dataset

To provide a comprehensive overview of the deep learning models employed in this study, let's begin by introducing each model:

- **UNET:** The network employed in this study follows a standard UNET model, featuring a down-sampling path followed by an upsampling path without an additional data consistency step. The architecture comprises cascades of 16, 32, 64, 128 and 256 filters, with two 3×3 convolution blocks in each path, followed by normalization and ReLU activation functions. Additionally, the upsampling path incorporates interleaved bilinear upsampling layers, doubling the resolution between blocks (Zbontar et al., 2018).

- **Invertible Recurrent Inference Machine (iRIM):** The iRIM, initially employed for the single-coil dataset, has also been extended for application on the multi-coil dataset. The recommended methodology involves training models with 400 layers on 3D MRI volumes, specifically addressing the challenges posed in the reconstruction problem (Putzky & Welling, 2019).

- **End-to-End Variational Network:** The E2E Variational Network represents an extended version of variational methods, introducing an end-to-end learning approach (Sriram et al., 2020). Generalized compressed sensing methods, integrated into an unrolled gradient descent scheme, form the foundation of variational networks. In the generalized compressed sensing method, all parameters, including the prior model defined by kernel filters and activation functions, undergo offline training to learn the weights (Hammernik et al., 2018). In E2E VarNET, the input consists of undersampled k-space, which undergoes multiple refinement steps of the same nature. Unlike variational networks that operate in image space quantities, E2E VarNET utilizes k-space quantities. Notably, VarNET models typically employ shallow convolutional networks with radial basis function (RBF) kernels. However, E2E VarNET replaces these shallow CNNs with UNETs, resembling a Variational UNET with k-space data consistency. It's worth noting that E2E VarNET differs in learning sensitivity maps, as opposed to using fixed sensitivity maps computed with parallel imaging (PI) in VarNET.

3.4.3 Basic Ensemble Methods

Ensemble learning combines diverse individual models to achieve enhanced generalization performance. In recent times, deep learning models with multilayer architectures have exhibited superior performance compared to shallow or traditional models, even in instances of potential overtraining. Deep ensemble models leverage the strengths of both deep learning models and ensemble learning, resulting in improved generalization performance for the final model (Ganaie et al., 2022). Additionally, ensemble learning mitigates prediction variance and generalization errors by combining predictions from multiple neural network models. The nonlinearity and high variance inherent in deep neural models can pose limitations on making predictions in the final model. Ensemble learning techniques encompass various aspects such as training data, model selection, and the methodology used to combine predictions. By employing these techniques, the hope is that predictions

from different networks, each with unique strengths and perspectives, converge in a complementary manner. This collaborative approach contributes diverse viewpoints, fostering a more nuanced and comprehensive understanding of the subject matter (Bazeley, 2017). In this research, we introduce basic ensemble combination methods, representing the simplest form of ensemble combiners that do not necessitate further training or parameter tuning (Kavur et al., 2022). The three fundamental combination methods explored in this study are outlined as follows:

- **Min / Max Combiner:** The combination rules for minimum and maximum are consistent for each pixel value. In this combiner, the minimum and maximum values of each pixel, derived from the outcomes of all networks, are respectively assigned to the reconstruction results for each volume.
- **Averaging Combiner:** For each volume, each pixel is computed by straightforward averaging across all network results.
- **Weighted Averaging Combiner:** For each network, coefficients used in weighted averaging are computed by obtaining the pseudo-inverse between the average of each reconstructed slice from all networks and its corresponding ground truth in the validation dataset. These coefficients are then applied for the reconstruction of volumes in both the validation and test datasets.

3.5 Evaluation Metrics

Three commonly used quantitative evaluation metrics, are calculated in this study to assess the performance of the reconstructed images. The first metric (one of the three commonly used metrics) is the normalized mean square error (NMSE), defined in Equation 3.1 as:

$$\text{NMSE} = \frac{||\tilde{u} - u||_2^2}{||u||_2^2} \tag{3.1}$$

Here, $\tilde{u}$ represents the reconstructed image volume, and u is the fully sampled ground truth image volume. It is important to note that NMSE normalizes entire image volumes rather than individual slices due to significant variations across a volume (Knoll et al., 2020).

The second quality metric is the Peak Signal-to-Noise Ratio (PSNR) (another of the three commonly used metrics), a measure that compares local differences. The higher the PSNR, the better the quality of the reconstructed image. It is defined in Equation 3.2 as:

$$PSNR(u, \tilde{u}) = 10 \log_{10}[\frac{(Max.(u))^2}{MSE(u, \tilde{u})}] \tag{3.2}$$

Here, Mean Square Error (MSE) is the mean square error between u and $\tilde{u}$, defined as $\frac{\|\tilde{u}-u\|_2^2}{n}$, where n is the number of entries in the volume u.

The third quality metric is the Structural Similarity Index (SSIM) (the last of the three commonly used metrics), which measures the similarity between two images by considering the interdependencies among nearby pixels. The resulting similarity is calculated as shown in Equation 3.3.

$$\text{SSIM} = \frac{(2\mu_{m_1}\mu_{m_2} + c_1)(2\sigma_{m_1 m_2} + c_2)}{(\mu_{m_1}^2 + \mu_{m_2}^2 + c_1)(\sigma_{m_1}^2 + \sigma_{m_2}^2 + c_2)} \tag{3.3}$$

Here, μ_{m_1} and μ_{m_2} are the average pixel intensities in m_1 and m_2, $\sigma_{m_1}^2$ and $\sigma_{m_2}^2$ are their variances, $\sigma_{m_1 m_2}$ is the covariance between m_1 and m_2, and c_1 and c_2 are two variables used to balance the division.

For SSIM values, a window size of 7×7 is chosen, k_1=0.01, k_2=0.03 are set, and L is defined as the maximum value of the target volume, $L = \max(v)$ (Zbontar et al., 2018). SSIM values range between 0 and 1, where 1 indicates a perfect match with the ground truth. SSIM values of 0.97, 0.98, and 0.99 are generally considered good for quality reconstruction.

3.6 Evaluation Results

In this section, we present a comprehensive analysis of the performance of both single-coil and multi-coil models on the FastMRI dataset. Our evaluation encompasses both qualitative and quantitative assessments to provide a thorough understanding of the models' effectiveness in reconstructing magnetic resonance imaging (MRI) data.

3.6.1 Quantitative Results

We begin by delving into the quantitative aspects of our evaluation, where we employ established metrics to measure the structural fidelity of the reconstructed images and overall efficacy of the models. Through rigorous numerical evaluations, we aim to quantify the extent to which each model excels in reconstructing images from single-coil and multi-coil MRI data.

3.6.1.1 Quantitative Results in Single-Coil FastMRI Dataset

The evaluation results for networks employing individual base learners and their ensembles are presented in Table 3.2 and Table 3.3. When considering a single-coil validation dataset and examining the base learners individually, iRIM achieves the highest score at both acceleration rates, as indicated in Table 3.2. In terms of basic ensemble configurations, averaging demonstrates superior performance across all aspects, specifically achieving higher results in PD contrast at 4x acceleration and PDFS contrast at 8x acceleration. Conversely, weighted averaging outperforms in PDFS contrast at 4x acceleration and attains higher results in all and PD contrast at 8x acceleration, as highlighted in Table 3.3. It is noteworthy that basic ensemble results consistently exhibit slightly superior performance across both acceleration rates.

Table 3.2 MSE, NMSE, PSNR, and SSIM for single-coil fastMRI dataset reconstruction baselines for All, PD and PDFS accelerated (Acc.) at 4x and 8x.

Modality	Acc.	Acq.	MSE	NMSE	PSNR	SSIM
ZeroFilled	4x	ALL	1.692e-10	0.05	29.57	0.6558
		PD	2.785e-10	0.028	30.63	0.728
		PDFS	5.89e-11	0.066	28.47	0.58
	8x	ALL	3.893e-10	0.082	26.96	0.5505
		PD	6.844e-10	0.065	27.18	0.629
		PDFS	9.127e-11	0.099	26.81	0.474
KIKI-NET	4x	ALL	1.175e-10	0.038	31.26	0.7143
		PD	1.833e-10	0.019	32.91	0.7983
		PDFS	5.099e-11	0.057	29.62	0.628
	8x	ALL	3.478e-10	0.073	27.76	0.6102
		PD	6.106e-10	0.058	27.96	0.6886
		PDFS	8.23e-11	0.088	27.54	0.528
PD-NET	4x	ALL	8.844e-11	0.033	32.16	0.7285
		PD	1.306e-10	0.015	34.18	0.819
		PDFS	4.59e-11	0.052	**30.04**	0.638
	8x	ALL	2.365e-10	0.056	29.15	0.6403
		PD	4.03e-10	0.039	29.84	0.7311
		PDFS	6.825e-11	0.073	28.31	0.547
UNET	4x	ALL	9.686e-11	0.034	31.79	0.7215
		PD	1.465e-10	0.016	33.78	0.8113
		PDFS	4.678e-11	0.053	29.96	0.638
	8x	ALL	3.375e-10	0.071	27.61	0.6049
		PD	5.924e-10	0.057	28.03	0.6911
		PDFS	8.002e-11	0.086	27.66	0.528
iRIM	4x	ALL	**7.688e-11**	**0.034**	**32.45**	**0.7501**
		PD	**1.058e-10**	**0.013**	**35.03**	**0.8382**
		PDFS	**4.77e-11**	**0.055**	29.85	**0.661**
	8x	ALL	**1.144e-10**	**0.044**	**30.95**	**0.6916**
		PD	**1.668e-10**	**0.019**	**33.14**	**0.7938**
		PDFS	**6.151e-11**	**0.070**	**28.75**	**0.5883**
Cascade-NET	4x	ALL	9.129e-11	0.034	31.98	0.7187
		PD	1.35e-10	0.015	34.17	0.8143
		PDFS	4.714e-11	0.054	29.87	0.625
	8x	ALL	2.639e-10	0.061	28.74	0.624
		PD	4.519e-10	0.043	29.34	0.7103
		PDFS	7.397e-11	0.079	28.07	0.534

Table 3.3 MSE, NMSE, PSNR, and SSIM metrics for the reconstruction of the single-coil fastMRI dataset using basic ensemble methods. Results are presented for acceleration rates of 4x and 8x in the All, PD, and PDFS contrasts

Modality	**Acc.**	**Acq.**	**MSE**	**NMSE**	**PSNR**	**SSIM**
Averaging	4x	ALL	**6.623e-11**	**0.027**	**33.39**	**0.7704**
		PD	**9.517e-11**	**0.011**	**35.83**	**0.8525**
		PDFS	3.7e-11	0.042	30.96	0.6811
	8x	ALL	1.903e-10	0.048	30.88	0.6925
		PD	3.198e-10	0.032	32.39	0.7874
		PDFS	**5.948e-11**	**0.065**	**29.36**	**0.5967**
Weighted Averaging	4x	ALL	6.596e-11	0.027	33.24	0.7644
		PD	9.393e-11	0.011	35.58	0.8482
		PDFS	**3.77e-11**	**0.043**	**31**	**0.6853**
	8x	ALL	**1.298e-10**	**0.043**	**31.32**	**0.6969**
		PD	**2.021e-10**	**0.022**	**33.38**	**0.7965**
		PDFS	5.673e-11	0.064	29.23	0.5964
Max. Combiner	4x	ALL	1.26e-10	0.043	30.8	0.752
		PD	1.928e-10	0.021	32.61	0.8314
		PDFS	5.851e-11	0.066	28.96	0.6718
	8x	ALL	2.995e-10	0.066	28.21	0.6605
		PD	5.192e-10	0.049	28.63	0.7401
		PDFS	7.757e-11	0.083	27.79	0.5802
Min. Combiner	4x	ALL	1.203e-10	0.042	30.96	0.7324
		PD	1.818e-10	0.01987	32.88	0.8212
		PDFS	5.808e-11	0.065	29.03	0.6427
	8x	ALL	3.068e-10	0.074	27.85	0.6326
		PD	5.222e-10	0.052	28.5	0.7253
		PDFS	8.919e-11	0.097	27.18	0.539

3.6.1.2 Quantitative Results in Multi-Coil FastMRI Dataset

The detailed outcomes for both individual networks and their ensembles are presented comprehensively in Table 3.4 and Table 3.5. In the multi-coil validation

dataset, the E2E Variational Network stands out for its performance in PD acquisition, except for a slight deviation in NMSE, as shown in Table 3.4.

The effectiveness of ensemble methods becomes evident in their adeptness at mitigating prediction variance, thereby manifesting in notably improved results within the ambit of noisy predictions. Table 3.5 demonstrates that ensemble approaches consistently achieve statistically superior outcomes, especially in the PDFS contrast.

Delving into the nuances of performance at different acceleration rates and contrasts, it is discerned that averaging presents a marginal advantage in PSNR and SSIM scores at the 4x acceleration rate. Meanwhile, Weighted Averaging demonstrates a nuanced superiority, yielding slightly higher scores at the 8x acceleration rate across all contrasts. These nuanced distinctions underline the robustness and versatility of ensemble techniques in enhancing the overall quantitative reconstruction outcomes.

3.6.2 Qualitative Results

Beyond numerical metrics, we conduct a nuanced qualitative assessment of the reconstructed images. This involves a visual inspection of the output, focusing on factors such as image clarity, structural fidelity, and the ability to capture fine details. Our qualitative analysis provides valuable insights into the visual prowess of the models and their potential applicability in clinical settings.

3.6.2.1 Qualitative Results in Single-Coil FastMRI Dataset

To assess the concordance between quantitative and qualitative results, an evaluation from a qualitative perspective becomes essential. Figures 3.3 and 3.4 present the qualitative outcomes of DL and ensemble-based reconstructions at both 4-8 fold acceleration rates, including the absolute error (residual error) when compared to the reference.

Table 3.4 Performance of Base Learner Results on Multi-Coil Validation Data

Model	**Acc.**	**Acq.**	**MSE**	**NMSE**	**PSNR**	**SSIM**
iRIM	4x	ALL	1.882e-11	0.006153	38.94	0.9157
		PD	2.897e-11	0.002995	40.44	0.9513
		PDFS	8.569e-12	0.009343	37.42	0.8797
	8x	ALL	4.005e-11	0.01024	36.3	0.8858
		PD	6.619e-11	0.006344	37.06	0.9231
		PDFS	1.365e-11	0.01417	35.53	0.8481
UNET	4x	ALL	2.898e-11	0.007459	37.56	0.9061
		PD	4.805e-11	0.004624	38.36	0.9363
		PDFS	9.719e-12	0.01032	36.75	0.8757
	8x	ALL	5.77e-11	0.01212	35.15	0.8736
		PD	1.001e-10	0.008954	35.36	0.9055
		PDFS	1.484e-11	0.01531	34.94	0.8413
E2E Variational Network	4x	ALL	1.293e-11	0.00531	39.99	0.9283
		PD	**1.848e-11**	0.002103	**42.2**	**0.9631**
		PDFS	7.322e-12	0.00855	37.76	0.8931
	8x	ALL	2.472e-11	0.007479	37.82	0.9037
		PD	3.928e-11	0.003969	39.13	0.9423
		PDFS	1e-11	0.01102	36.5	0.8647

The qualitative results not only complement but also affirm the quantitative findings reflected in Table 3.2 and Table 3.3. In these tables, iRIM, serving as a base learner, demonstrates superior performance, and the ensemble results from Weighted Average (Weighted Av.) and Averaging show better evaluation metrics, as corroborated by the corresponding residual errors.

While the figures may appear quite similar at first glance, a closer inspection reveals subtle differences in the distributed residual errors, providing a more nuanced understanding of the image disparities.

Table 3.5 Performance of Ensembles of Base Learners on Multi-Coil Validation Data

Model	Acc.	Acq.	MSE	NMSE	PSNR	SSIM
Averaging	4x	ALL	**1.301e-11**	**0.00471**	**40.26**	**0.9296**
		PD	1.944e-11	**0.002074**	42.1	0.9624
		PDFS	**6.516e-12**	**0.007372**	**38.4**	**0.8964**
	8x	ALL	2.676e-11	0.007399	37.73	0.9027
		PD	4.357e-11	0.00422	38.77	0.9397
		PDFS	9.792e-12	0.01061	36.68	0.8653
Weighted Averaging	4x	ALL	1.476e-11	0.004933	39.88	0.9266
		PD	2.266e-11	0.002344	41.48	0.9589
		PDFS	6.773e-12	0.007549	38.27	0.8939
	8x	ALL	**2.247e-11**	**0.006829**	**38.25**	**0.9069**
		PD	**3.562e-11**	**0.003577**	**39.58**	**0.9451**
		PDFS	**9.182e-12**	**0.01011**	**36.91**	**0.8683**
Max. Combiner	4x	ALL	2.464e-11	0.007492	37.83	0.9226
		PD	3.903e-11	0.003946	39.09	0.9545
		PDFS	1.009e-11	0.01107	36.56	0.8904
	8x	ALL	4.813e-11	0.01135	35.64	0.8944
		PD	8.132e-11	0.007542	36.15	0.9293
		PDFS	1.461e-11	0.01519	35.13	0.859
Min. Combiner	4x	ALL	2.281e-11	0.006542	38.34	0.9212
		PD	3.673e-11	0.003651	39.46	0.9553
		PDFS	8.758e-12	0.009461	37.21	0.8867
	8x	ALL	4.707e-11	0.01091	35.76	0.8914
		PD	7.994e-11	0.007395	36.27	0.9286
		PDFS	1.387e-11	0.01447	35.24	0.8539

3.6.2.2 Qualitative Results in Multi-Coil FastMRI Dataset

To assess the alignment between quantitative and qualitative results in the context of multi-coil reconstructions, Figures 3.5 and 3.6 provide a visual representation of the DL and ensemble-based reconstructions at both 4-8 fold acceleration rates, including the absolute error (residual error) compared to the reference.

The qualitative findings substantiate the quantitative results regarding image quality, as outlined in Table 3.4 and Table 3.5. In these tables, E2E VarNET, serving

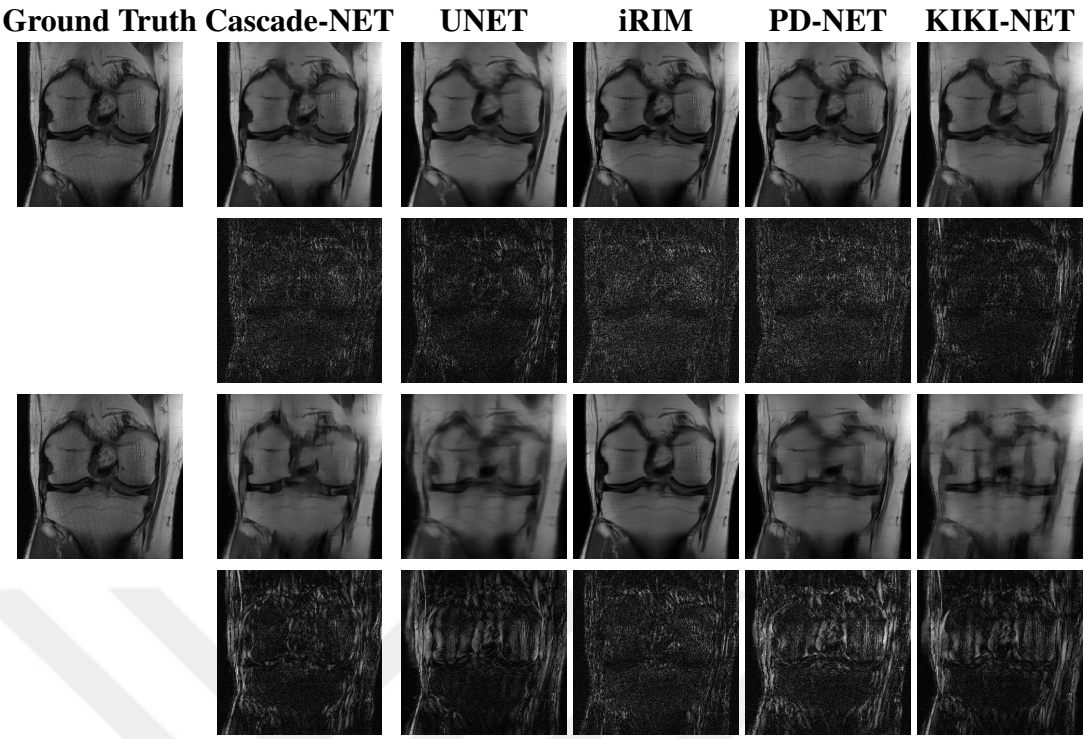


Figure 3.3 Reconstruction results for a specific slice in the single-coil validation set at 4-fold and 8-fold acceleration (27th slice of file1000041, part of the PD validation set). The first two rows display the reference and DL reconstruction results for 4-fold acceleration, along with the absolute error (residual error) compared to the reference. The following two rows show the reference and DL reconstruction results for 8-fold acceleration, along with the corresponding residual error.

as a base learner, exhibits superior performance, while the ensemble results from Weighted Average and Averaging demonstrate better evaluation metrics, as evident in the corresponding residual errors.

While the figures may seem quite similar at first glance, a more detailed examination reveals subtle differences in the distributed residual errors, providing valuable insights into the variations among the images as presented in the single-coil qualitative results.

3.7 Conclusion and Discussion

In this chapter, we implemented various neural network architectures, including KIKI-NET, Cascade-NET, UNET, iRIM, PD-NET for the single-coil knee dataset, and E2E VarNET, UNET, and iRIM for the multi-coil dataset.

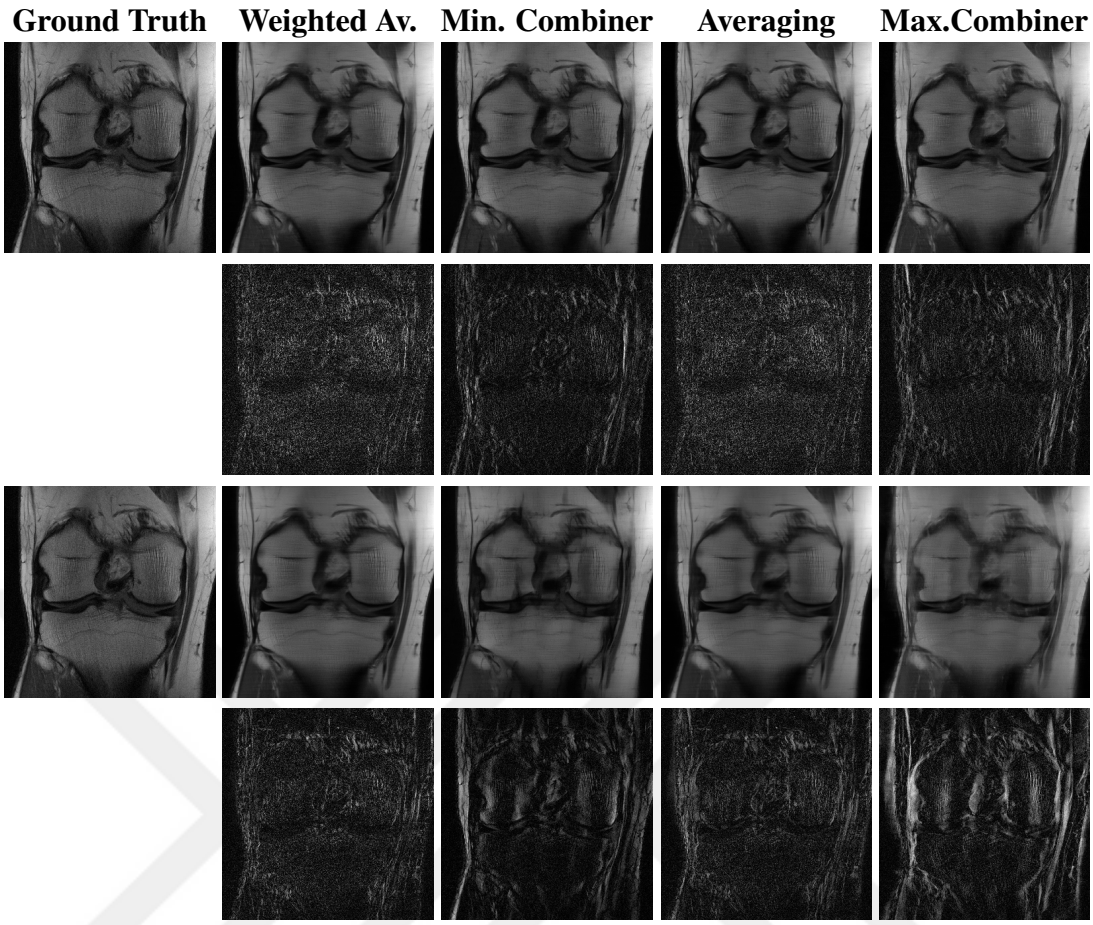


Figure 3.4 Reconstruction results for a specific slice in the single-coil validation set with 4-fold and 8-fold acceleration (27th slice of file1000041, part of the PD validation set). The first two rows present the reference and ensemble-based reconstruction results for 4-fold acceleration, accompanied by the absolute error (residual error) in comparison to the reference. The subsequent two rows depict the reference and ensemble-based reconstruction results for 8-fold acceleration, along with the corresponding residual error.

Our results demonstrate that basic ensembles consistently outperform individual base learners in both single-coil and multi-coil knee validation datasets. This observation is particularly encouraging as it implies generalizability across different datasets without the need for dataset-specific design modifications or extensive parameter tuning, thereby addressing concerns related to long training times and computational costs.

Notably, ensembles exhibit a more pronounced positive impact on fat-suppressed scans, even at highly accelerated rates. It would be interesting to explore their performance further with other types of undersampling methods, such as after

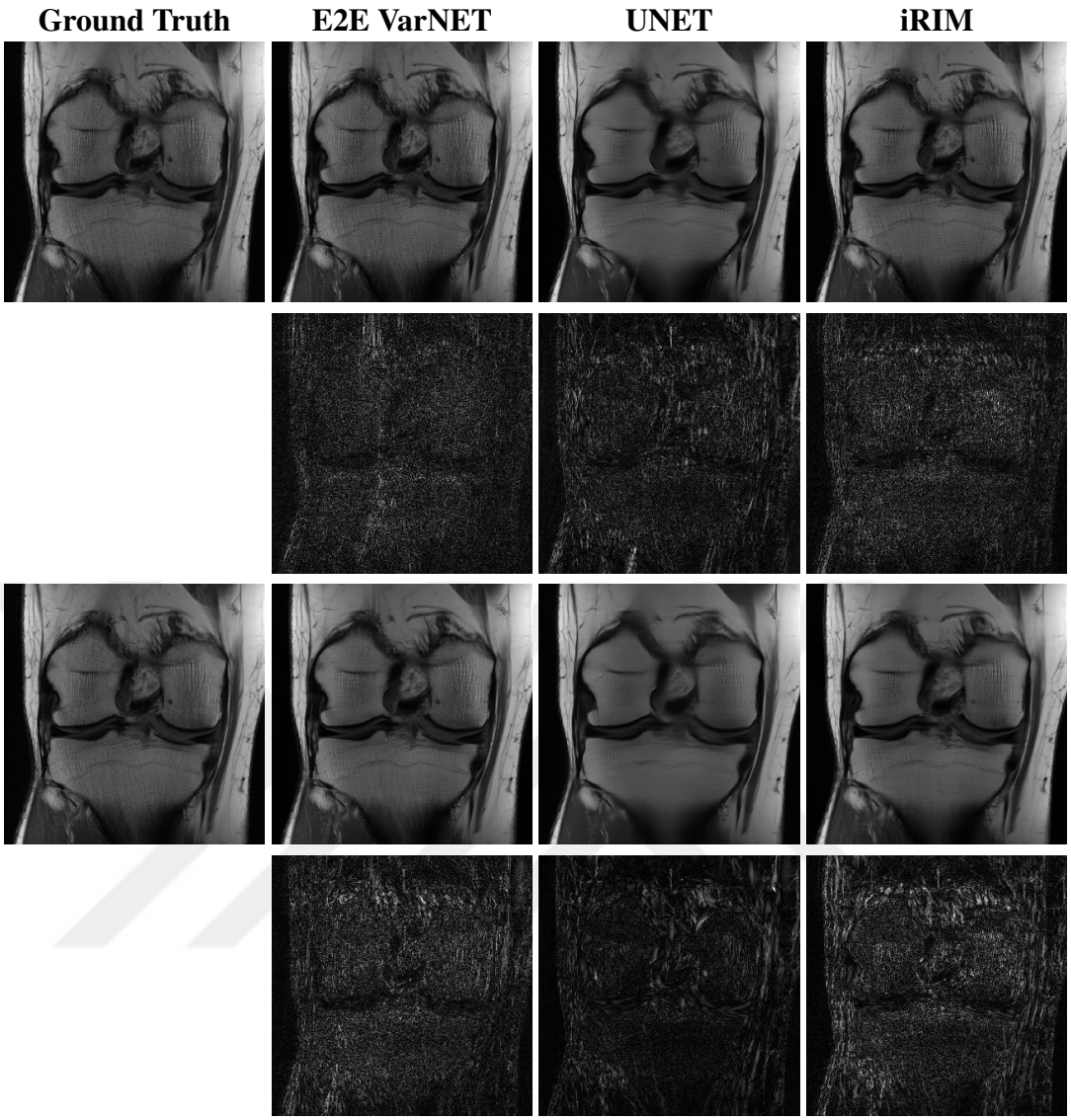


Figure 3.5 Reconstruction results for a specific slice in the multi-coil validation set at 4-fold and 8-fold acceleration (27th slice of file1000041, part of the PD validation set). The first two rows display the reference and DL reconstruction results for 4-fold acceleration, along with the absolute error (residual error) compared to the reference. The following two rows show the reference and DL reconstruction results for 8-fold acceleration, along with the corresponding residual error.

regridding for non-cartesian undersampling, and additional schemes. However, generalizing networks in non-cartesian sampling poses challenges due to the unconventional implementation of data fidelity steps and unfamiliar positioning over the image.

Despite the high reconstruction quality achieved by deep learning approaches, especially at high acceleration rates, our study highlights a potential limitation.

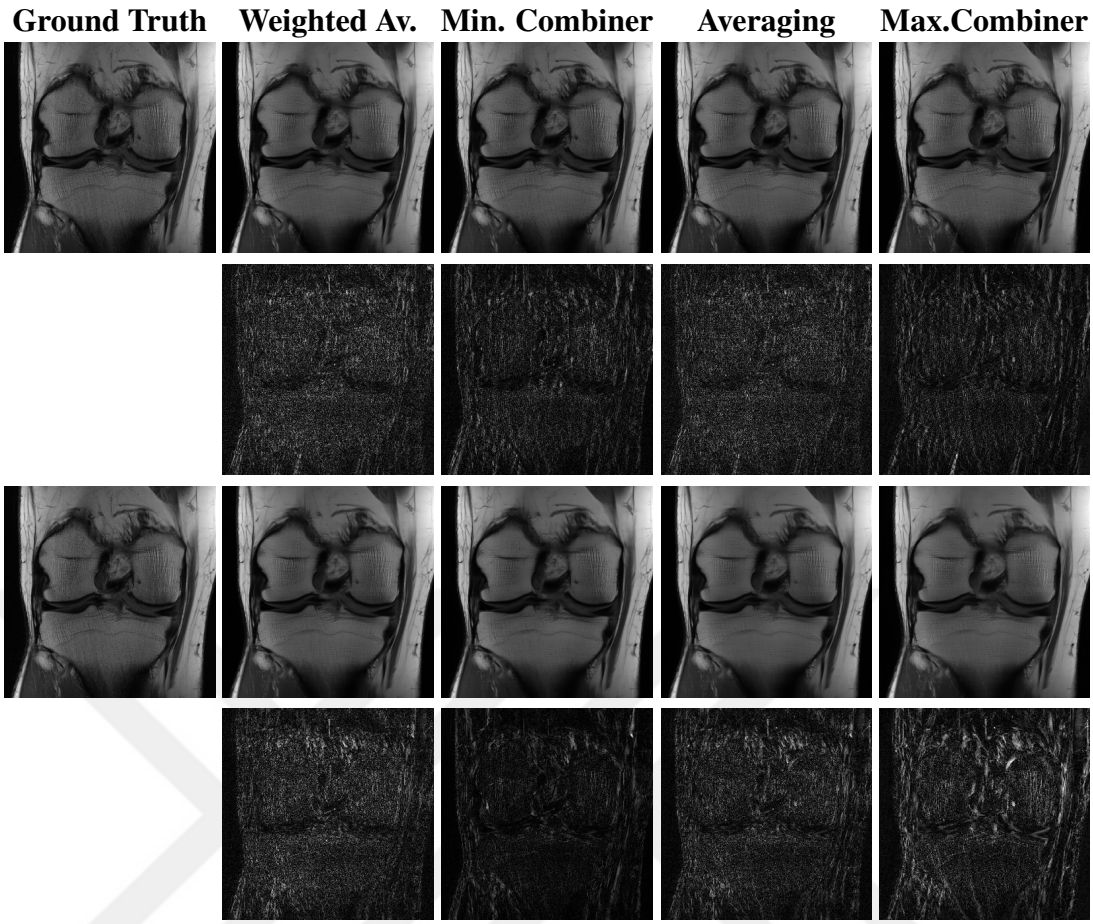


Figure 3.6 Reconstruction results for a specific slice in the multi-coil validation set with 4-fold and 8-fold acceleration (27th slice of file1000041, part of the PD validation set). The first two rows present the reference and ensemble-based reconstruction results for 4-fold acceleration, accompanied by the absolute error (residual error) in comparison to the reference. The subsequent two rows depict the reference and ensemble-based reconstruction results for 8-fold acceleration, along with the corresponding residual error.

Clinically important details, such as Meniscus Tears, may be partially visible or lost during the reconstruction phase. This emphasizes the need for the development of comprehensive metrics capable of detecting and preserving clinically relevant details during the reconstruction process. Further research and refinement in this area could enhance the clinical applicability of deep learning-based reconstruction methods.

CHAPTER FOUR
DEEP LEARNING-BASED MENISCUS TEAR DETECTION FROM ACCELERATED MRI

4.1 Introduction

Magnetic Resonance Imaging (MRI) is a powerful diagnostic tool known for its exceptional soft tissue contrast, making it invaluable in musculoskeletal diagnosis (van Beek et al., 2019). While the slow data acquisition process of MRI has historically posed challenges, the advancements in technology and imaging protocols have significantly improved the efficiency of MRI procedures. Contemporary approaches, such as accelerated imaging techniques and novel reconstruction algorithms, contribute to reducing acquisition times. In this context, our study focuses on MRI reconstruction from undersampled k-space and its effects on representation and detection of clinically important information. Recently, Deep Learning (DL) techniques have emerged as a promising solution, surpassing traditional methods and enabling accelerated acquisition with fewer deviations from fully-sampled k-space reconstructions (Johnson et al., 2022). Despite their reported success based on the commonly used image quality metrics, such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), the outcomes lack comprehensive clinical feedback. Thus, the clinical translation of these performance improvements are still being investigated case by case.

To bridge this gap and achieve reliable and repeatable results, there is a strong need to design and conduct comprehensive studies, which focus on comparing expert assessments with evaluation metrics. For instance, Lee et al. (Lee et al., 2023) presented complementary perspectives, emphasizing the effectiveness of DL-based reconstruction in reducing acquisition time, improving Signal-to-Noise Ratio (SNR) and Contrast-to-Noise Ratio (CNR), and maintaining image quality and Prostate Imaging Reporting and Data System (PI-RADS) classification in both prostate T2-weighted (T2WI) and diffusion-weighted (DWI) imaging, supported by expert

assessments. Building on these findings, Desai et al. (Desai et al., 2022) highlighted that current evaluation metrics for image and segmentation quality don't align well with qMRI biomarkers, especially in relevant tissue areas. This suggests a need for a new evaluation framework to directly estimate these biomarkers. However, in the domain of musculoskeletal applications, the utilization of DL-reconstructed MRI images for classification and detection tasks remains limited (Gassenmaier et al., 2021). Although Machine Learning (ML), especially DL models, shows promise, their outcomes still fall short when compared to the inter-expert variations (Chen et al., 2022). Furthermore, the prospective implementation of these models in clinical settings is very rare (Chaudhari et al., 2021).

In this study, we focus on the evaluation of DL reconstructed MRI images for musculoskeletal diagnostic tasks, particularly the detection of meniscal tears. we address the limitations inherent in traditional MRI reconstruction and detection methods in detail. Our primary goal is to systematically compare the performance of DL detectors against conventional imaging and expert evaluations. Utilizing a comprehensive evaluation pipeline, we aim to uncover the potential of DL methods in capturing clinically relevant details and facilitating clinical utility in both MRI reconstruction and detection. Ultimately, our objective is to contribute to diagnostic decision-making improvement in routine radiology practice.

Our central focus revolves around evaluating ML techniques, specifically in the context of meniscal tear detection and MRI reconstruction, with a keen eye on the perspective of clinicians. This study is geared towards providing a comprehensive assessment of the proposed models' effectiveness and reliability in these specific medical imaging tasks. We aspire to offer valuable insights into the potential clinical applicability and diagnostic enhancement that these ML methods bring to knee MRI examinations.

To address the limitations associated with both image reconstruction and classification/detection tasks, a long-term benchmarking study was initiated in 2019 (Knoll et al., 2020). The study began with the release of the fastMRI dataset, which

provided an opportunity to evaluate the performance of state-of-the-art ML methods for MRI reconstruction on a large-scale standardized dataset (Zbontar et al., 2018). This dataset includes undersampled k-space data for knee MRI, along with fully sampled ground truth data. The challenge is ongoing, and its outcomes are regularly disseminated through individual studies and reviews.

In the subsequent phase, the FastMRI+ dataset was introduced to address the lack of clinical annotations in the fastMRI images (Zhao et al., 2021). This extension includes expert-labeled bounding boxes, enabling the identification of twenty-one different pathologies from the multi-coil knee dataset. In addition to standard medical image classification and detection tasks, FastMRI+ allows for the analysis of reconstruction frameworks to retain crucial clinical information by observing classification and detection outcomes (Zhao et al., 2021).

In the realm of fastMRI reconstruction, various DL models with diverse network topologies have been proposed and evaluated (Sriram et al., 2020, Sriram et al., 2020; Kerstin et al., 2019, Hammernik et al., 2018; Pezzotti et al., 2020; Ramzi et al., 2020; Liu et al., 2022; Tong et al., 2022; Pal & Rathi, 2022). These models primarily aim for high-quality reconstruction from accelerated MRI acquisition, often incorporating thorough evaluations involving expert readings (Muckley et al., 2021; Knoll et al., 2020). Recent advancements showcase the potential of DL reconstruction in significantly reducing scan time while enhancing image quality for knee MRI (Johnson et al., 2023). However, these methods predominantly focus on image reconstruction and lack integration for automatic disease detection.

Among these architectures, the UNET model (Zbontar et al., 2018) stands out for its widespread adoption in image reconstruction tasks. Renowned for its superior performance in feature extraction and utilization of skip connections, the UNET architecture is often used in fastMRI research. The Invertible Recurrent Inference Machine (iRIM) (Putzky & Welling, 2019) introduces a solution to memory challenges in learning. It employs invertible networks to ensure constant memory usage during training, making it particularly effective for deep models. Conversely,

the End-to-End Variational Network (E2E VarNET) (Sriram et al., 2020; Hammernik et al., 2018), excels in achieving learning by incorporating sensitivity maps.

In tackling uncertainties related to automated disease detection within deep learning-based reconstructions, particularly in the FastMRI+ context, we utilize the nnDetection tool—a self-configuring model with automatic adaptability (Baumgartner et al., 2021). This tool has demonstrated outstanding performance across various medical object detection tasks on well-known public datasets, including LUNA16 (Setio et al., 2017), ADAM (Timmins et al., 2020), CADA (Hennemuth et al., 2021), RibFrac (Jin et al., 2020), LIDC-IDRI (Armato III et al., 2011), Kits19 (Heller et al., 2019), and others. Notably, it has excelled in Decathlon tasks like Pancreas, Prostate, and Hepatic Vessel (Simpson et al., 2019). The established versatility and success of nnDetection underscore its significance in automating meniscal tear detection, offering automation, specialization, and design flexibility. The integration of nnDetection strengthens the reliability of our approach and addresses methodological gaps of current studies.

Accordingly, the main contributions of this study can be summarized under three headings as follows:

- We apply MRI reconstruction and analysis using different DL models and regenerating the results for fastMRI. Despite variations in model architecture and training, all methods exhibit closely aligned evaluation results (Putzky & Welling, 2019; Zbontar et al., 2018; Sriram et al., 2020). Our key contribution lies in conducting a comprehensive analysis that reveals diverse outcomes from these DL models, indicating their potential for effectively preserving meniscus tears when combined via ensembles.

- The second contribution of our study is a detailed comparative analysis of both the original and DL-reconstructed images by two expert radiologists. During this assessment, the radiologists scored various factors, including hallucinations, artifacts, and other elements that could potentially influence the diagnostic process. The results indicate that the UNET model produces more

reconstruction artifacts (such as hallucinations) when compared to the E2E VarNET and iRIM. Nevertheless, the UNET exhibits fewer false positives in detecting meniscal tears than the other models. In other words, E2E VarNET and iRIM are more suitable for reconstruction by generating fewer artifacts and are more suitable for expert visual analyses. However, UNET generates artifacts that do not have an impact on the performance of DL detectors for meniscal tear detection.

- In addition to our contributions outlined above, our study introduces a comprehensive evaluation pipeline for meniscal tear diagnosis, combining both reconstruction and detection processes (Figure 4.3). Building on the analyses conducted, we assess the performance of a state-of-the-art DL detector, nnDetection (Baumgartner et al., 2021), on retrospectively undersampled DL-reconstructed knee MRI. Through a quantitative comparison with conventional imaging for meniscus tears, we rigorously evaluate the detector's performance. Our findings indicate that when combined with the E2E VarNET, nnDetection demonstrates the ability to effectively preserve and detect clinically significant outcomes at both 4x and 8x acceleration rates. Our results surpass the current literature, achieving average precision scores of 0.69 and 0.67 at 4- and 8-fold accelerations, respectively. Notably, the UNET model achieves average precision values of 0.66 and 0.64, while the iRIM model achieves values of 0.67 and 0.65 at 4-8 fold accelerated scans, further highlighting the superior performance of our proposed pipeline.

The rest of this study is structured as follows. Section 4.2 presents the background of MRI reconstruction and reviews related studies. The materials and methods employed for reconstruction and analysis are described and we discuss the fastMRI dataset and its constraints in Section 4.3. In Section 4.4, we present the image quality metrics and expert evaluations. The results and analysis are presented in Section 4.5. Finally, in Section 4.6, we draw conclusions and highlight avenues for future research.

4.2 Background and Related Works

The slow data acquisition speed of MRI, attributed to the point-by-point transformation (Fourier) of an image known as k-space, presents a significant challenge. To avoid artifacts, adhering to the Nyquist criteria while sampling k-space is essential. However, accelerated MRI involves sub-sampling k-space below the Nyquist rate to expedite the process, resulting in aliasing artifacts, which represent a classical, ill-posed inverse problem. The primary goal of accelerated scans is to address this ill-posed problem and obtain a high-quality original image.

In recent years, DL models have been widely adopted for various critical tasks such as detection, classification, and reconstruction, showcasing significant success in different applications (Litjens et al., 2017). Despite their success, these models often overlook important factors like acquisition variability, orientation, pathology, and artifacts. For instance, recent research by Patsanis et al. (Patsanis et al., 2023) showcased the sparse presence of pathology annotations and how it affects prostate cancer detection. They highlighted the significance of carefully selecting appropriate training data in this context. Ebner et al. (Ebner et al., 2020) demonstrated superior performance in fetal brain MRI reconstruction, detection of pathological brains, and expert reader quality assessments. Baalbaki et al. (Baalbaki et al., 2023) proposed the identification of atherosclerosis in reconstructed images.

In the realm of musculoskeletal imaging, several studies have addressed specific pathologies. Dung et al. (Dung et al., 2023) and Ni et al. (Ni et al., 2023) proposed approaches for classifying anterior cruciate ligament tears and diagnosing anterior talofibular ligament tears, respectively, in 2023. In a different study, Zhao et al. (Zhao et al., 2021) attempted to use a method based on natural image processing to capture clinically important details with limited annotations, but it did not achieve the desired outcome. Translating architectures or improvements from natural images to medical images is challenging and seldom results in improved clinical utility. Additionally, evaluating pathological features like meniscal tears using only 2D models lacks clinical reliability. In the clinical process, the detection of meniscal tears involves

examining the possible tear area and following the tear across adjacent slices, requiring radiologists' expertise and careful consideration.

In the context of musculoskeletal MRI, combining reconstruction and MRI tasks, such as classification and detection, has proven to be particularly effective, especially in dealing with undersampled k-space acquisitions. However, the availability of limited benchmarking datasets and sparse annotations poses challenges for further advancements. While several datasets have been released to enable benchmarking for ML-based MRI research, they may have limitations in terms of size, application area, or evaluation metrics. Overcoming these challenges and leveraging the potential of DL models in preserving clinically relevant information in musculoskeletal MRI is crucial for enhancing their clinical applicability.

For instance, the 2018 release of Mridata.org provided fully sampled raw k-spaces for 19 healthy patients. In this context, Du et al. (Du et al., 2021) proposed an adaptive Convolutional Neural Network (CNN) model for k-space data interpolation by integrating complementary information of contiguous slices, while Huang et al. (Huang et al., 2022) focused on quantitatively evaluating the impact of image contrast, acceleration, and sampling patterns over deep cascade CNNs. Despite these efforts, the small size and homogeneity of these datasets limit their clinical usage. Consequently, assessing the effect of various reconstruction methods on subtle but vital pathologies through clinical outcomes poses another challenge (Ong et al., 2018; Darestani et al., 2021; FSE'XL, PD & FAT, FSE'XL et al.).

In 2019, the fastMRI dataset offered large MRI raw k-spaces, facilitating progress in MRI reconstruction methods and promoting replication of results and fair evaluation (Knoll et al., 2020, Knoll et al., 2020; Zbontar et al., 2018). This dataset played a pivotal role in advancing ML-based MRI reconstruction tasks (Muckley et al., 2021). Subsequently, the Stanford Knee MRI with Multi-Task Evaluation (SKM-TEA) dataset was introduced in 2021, providing an end-to-end evaluation of clinically important details with 155 anonymized MRI scans, including qMRI parameter maps. Peng et al. (Peng et al., 2022) proposed the use of

diffusion-model-based reconstruction with a Monte Carlo sampling scheme in SKM-TEA. However, the limited dataset size acquired from only 3T scanners and a single orientation makes it challenging to analyze the effects of different reconstructions on subtle clinically important details (Desai et al., 2022).

In parallel with these developments, the Osteoarthritis Initiative (OAI) (Nevitt et al., 2006) and MRNet (Bien et al., 2018) have been instrumental in classifying knee abnormalities in 2006 and 2018, respectively. More recently, Nasser et al. (Nasser et al., 2020) introduced a discriminative model to enhance classification performance, while Hu et al. (Hu et al., 2022) explored the severity grade of longitudinal osteoarthritis using an adversarial evolved neural network in the OAI dataset. Dai et al. (Dai et al., 2021) combined transformer and CNN for multi-modal classification, and Kara et al. (Kara & Hardalaç, 2021) used the ResNET50 model to classify and detect knee abnormalities in the MRNet dataset. However, these datasets provide only patient-level labels and lack bounding boxes for localizing abnormalities in MRI volumes. Additionally, these datasets do not offer raw k-space data, which hinders the end-to-end evaluation of the impact of MRI reconstruction in a clinical context.

FastMRI+ extends the capabilities of fastMRI by incorporating detection labels for clinical pathologies, enabling the localization of pathology in reconstructed scans (Zhao et al., 2021). The availability of these bounding boxes enhances the evaluation of clinically important pathologies in both upstream and downstream approaches. Despite the potential, the utilization of this dataset in studies remains limited, with one notable investigation by Razumov et al. (Razumov et al., 2022) focusing on determining the optimal k-space undersampling pattern for effective localization or segmentation of pathology in a downstream manner.

4.3 Materials and Methods

In this work, we aim to leverage the progress made in DL models concerning undersampled k-space (aliased images). Additionally, we intend to extend the

pipeline to encompass the downstream task of disease detection across different reconstruction outcomes. The study pipeline is represented in Figure 4.2. Undersampled k-space is fed to three distinct reconstruction models. For the evaluation of meniscus tear, expert study and analysis using DL processes were performed separately. nnDetection model is with the three reconstruction outcomes. Finally, we compared the quantitative DL detection results with the expert study findings.

4.3.1 Dataset

The fastMRI dataset, as extensively described in reference (Zbontar et al., 2018) (https://fastmri.med.nyu.edu/), constitutes a foundational element in our study, contributing various crucial attributes. It is strategically partitioned into training (973 volumes), validation (199 volumes), and test sets (118 volumes), providing a robust framework for model development and comprehensive evaluation. To prevent overfitting, ground truths for the test set are deliberately withheld. Our study leverages the retrospective k-space undersampling technique, emulating 2D physically achievable accelerations, applied to fully sampled acquisitions for accelerated MRI reconstruction. This process selectively excludes k-space lines in the phase-encoding direction and uniformly applies an undersampling mask to all slices. Accelerations of four-fold and eight-fold are implemented, retaining only 8% and 4% of the central k-space, respectively. The study benefits significantly from the predefined acquisition trajectories provided by fastMRI, enhancing the realism and applicability of our findings.

FastMRI+ dataset (Zhao et al., 2021), an extension of fastMRI, is designed for comprehensive evaluation in MRI reconstruction. It features meticulously annotated clinically significant pathologies and introduces clinically limiting artifacts. With 1172 reconstructed k-space volumes in the training and validation sets, including 6154 bounding boxes and 13 study-level labels, FastMRI+ ensures a diverse dataset.

Incorporating fully sampled images from the fastMRI multicoil knee dataset

ensures the availability of critical clinical information. This aspect is pivotal for assessing how diverse reconstruction algorithms preserve clinical details. Through the analysis of outcomes across various tasks, we showcase the efficacy of different methods, contributing to a deeper comprehension of the role of FastMRI+ in MRI reconstruction.

FastMRI+ (Zhao et al., 2021), available at https://github.com/microsoft/fastmri-plus, is pivotal for advancing MRI reconstruction research. With 815 pathology-labeled volumes and 12 volumes marked as artifacts in the training set, and 154 pathology-labeled volumes with one artifact in the validation set, the dataset provides a comprehensive evaluation. Table 4.1 details the categorization of meniscal tears and artifacts, with a visual representation in Figure 4.1.

Table 4.1 Summary of Meniscus Tear and Artifact Labels with Subset Divisions

Label	**Annotation Count**	**Subject Count**	
		Training Set	**Validation Set**
Meniscus Tear	**5658**	**560**	**103**
Artifact	–	12	1

4.3.2 Reconstruction Models

We utilized three advanced deep learning models that have exhibited outstanding performance. The following section provides detailed information about these models.

- U-NET (UNET): Numerous techniques built on the UNET architecture have been presented in the literature (Ronneberger et al., 2015; Long et al., 2015, Zbontar et al., 2018). The network utilized in this study is a basic model with two paths: down-sampling and up-sampling, excluding an additional data consistency step. It consists of filter cascades with sizes 32, 64, 128, and 256. In both the down-sampling and up-sampling paths, two 3x3 convolution blocks are followed by normalization and rectifier linear units (RELU) activation functions, maintaining a similar structure. These blocks are separated by a max-pooling



Figure 4.1 Examples of Meniscus Tear bounding boxes and their corresponding cropped images extracted from the FastMRI+ knee dataset. The arrows highlight the tears. The first row shows instances from 1.5 Tesla examinations, and their corresponding cropped images are displayed just below. The second row presents cases from 3 Tesla examinations, along with their respective cropped images below. The images in the first column represent Proton Density (PD) contrast, while the ones in the second column represent Proton Density Fat Suppression (PDFS).

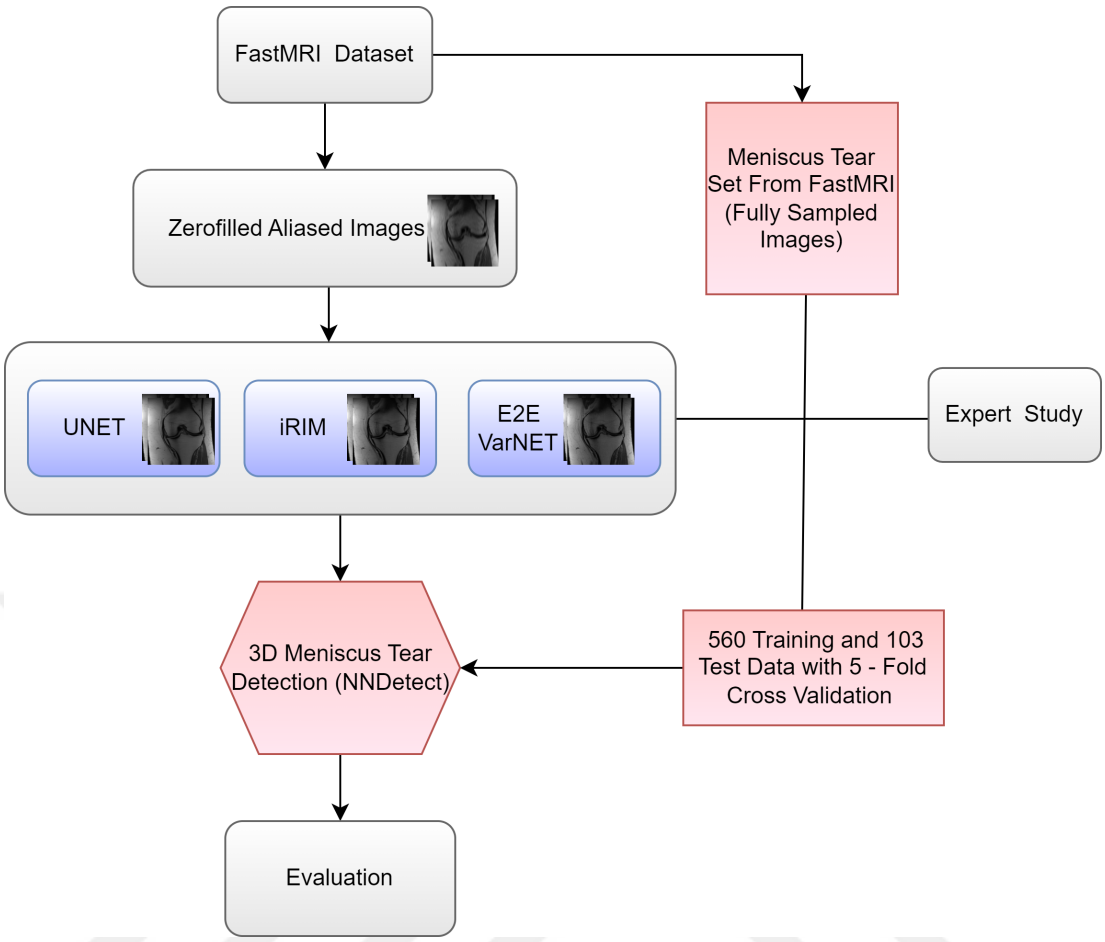


Figure 4.2 The study design through the model training and testing phases. Initially, we undersampled k-space and generated zero-filled aliased images. Subsequently, three distinct reconstruction models were employed to obtain high-quality images. For the evaluation of meniscus tear, expert study and analysis using DL processes were performed separately. Data annotated as "Meniscus Tear" were collected and prepared for our nnDetection model. The model was then trained with this data and tested with the three reconstruction outcomes. Finally, we compared the quantitative DL detection results with the expert study findings.

layer with two strides in the down-sampling path and a bilinear upsampling layer in the up-sampling path, doubling the resolution between blocks (Zbontar et al., 2018).

The UNET model underwent a maximum of 50 epochs with RMS Prop as the optimizer, using a learning rate of 0.001. The architecture comprised 4 UNET pooling layers, and a learning step size of 40 epochs for decreasing the learning rate, governed by a learning gamma of 0.1. Additionally, the UNET model employed 1 channel for both input and output to UNET.

- Invertible Recurrent Inference Machine (iRIM): Iterative learning, while popular for addressing inverse problems, poses challenges with memory requirements during training, growing linearly with the model's depth, making it difficult to use extensive models. In this study, an alternative approach is utilized—an iterative inverse model based on invertible networks with constant memory, eliminating the need to store intermediate activations. This enables the training of the model with 400 layers on 3D MRI volumes for reconstruction tasks (Putzky & Welling, 2019, Putzky et al., 2019). The iRIM model also underwent a maximum of 50 epochs, employing the Adam optimizer with an initial learning rate of 0.0001, reduced by a factor of 10 every 30 epochs. The learning gamma for iRIM was set to 0.1.

- E2E Variational Network (E2E VarNET): The E2E Variational Network is an extended version of variational methods allowing for fully end-to-end learning (Sriram et al., 2020). Variational networks were originally designed using generalized compressed sensing methods within an unrolled gradient descent scheme (Hammernik et al., 2018). In these methods, all parameters, including prior models specified by kernel filters and activation functions, were learned offline during training (Hammernik et al., 2018). E2E Variational Network underwent a maximum of 50 epochs, utilizing the Adam optimizer with a learning rate of 0.001. This model featured 4 UNET pooling layers, 18 top-level channels for UNET, and a learning step size of 40 epochs for decreasing the learning rate, with a learning gamma of 0.1. However, there were some distinctions. In E2E VarNET, undersampled k-space is performed on various refinement steps of the same form, using k-space quantities instead of image-space quantities. Additionally, total shallow convolutional networks with radial basis function (RBF) kernels, present in variational networks, are replaced by UNETs in E2E Variational Networks. This modification effectively turns them into variational UNETs. However, E2E VarNET also involves an intermediate data consistency step with learned sensitivity maps, unlike fixed sensitivity maps computed using Parallel Imaging (PI) in a Variational Network.

4.3.3 Detection

FastMRI+ (Zhao et al., 2021), an extension of fastMRI, provides detection labels for clinical pathologies, aiding in the identification of pathology in reconstructed scans. The presence of bounding boxes facilitates the comprehensive assessment of clinically significant pathologies in both upstream and downstream processes. In this study, nnDetection (Baumgartner et al., 2021), a systematic and automated self-configuration method, was utilized for downstream analysis. nnDetection introduces a novel rule-based, empirical design with fixed choices and adopts the simple Retina UNET architecture (Jaeger et al., 2020). This method has the capability to fully automate the detection of meniscal tears without requiring additional computed sources beyond standard training. nnDetection systematically configures the entire detection pipeline and allows for comprehensive design possibilities. It follows a similar approach to nnUNet (Isensee et al., 2021), where domain knowledge is divided into fixed, rule-based, and empirical parameter forms.

Fixed parameters: nnDetection offers design opportunities that do not necessitate dataset-specific adaptation for robust generalization. The Retina UNET, a variation of RetinaNet, was selected as the architecture template due to its ability to leverage pixel-level annotations (Jaeger et al., 2020). The key elements of the architecture, such as pool strides, pooling, and kernel sizes, were adapted through rule-based parameters.

Rule-based parameters: An interdependent heuristic approach was employed to formulate distinct dependencies between the data and design options. This methodology also incorporates information about object sizes compared to nnUNet. The optimization process for determining network parameters, such as pooling strides, numbers, and kernel size, remains consistent with nnUNet. However, to enhance training stability, a fixed batch size of 4 was utilized.

Empirical parameters: The post-processing aspect of this task, which typically deals with clustering overlaps of bounding box predictions, was addressed in nnDetection. The approach involved using sliding window inference with overlapping

patches. To handle overlap between neighboring patches, non-maximum suppression (NMS) was applied, with predictions near the patch center given higher importance than those at the border. Weighted box clustering (Jaeger et al., 2020) was employed to cluster predictions from test time augmentation or multiple models. The empirical parameters used exclusively during the test were optimized and designed based on the validation set. nnDetection utilized predefined initialization parameters due to their interdependencies and optimized them through a rigorous process.

In the initial phase, nnDetection, an automatic self-configuring model, extracts a data map and applies heuristic rules to determine rule-based parameters. The training process involves the application of a combination of cross-entropy (classification) and generalized intersection-over union (regression) loss functions over 60 epochs, with a batch size of 4. Notably, half of each batch must include at least one ground-truth box. Optimization is performed using stochastic gradient descent with a Nesterov momentum of 0.9. The learning rate undergoes a linear ramp-up from 1e-6 to 1e-2 over the first 4000 iterations, followed by a poly learning rate schedule until epoch 50. The final 10 epochs employ training with a cyclic learning rate oscillating between 1e-3 and 1e-6 each epoch. Empirical parameters are determined post-training, and the final predictions result from ensembling the predictions from five models obtained through cross-validation with empirically selected settings. Throughout this process, standard data augmentation techniques, such as Gaussian noise, Gaussian blur, adjustments in brightness and contrast, simulation of low resolution, gamma augmentation, elastic deformation, scaling, flipping, mirroring, and rotation, are dynamically implemented to ensure a robust and diverse training dataset.

Additionally, it is worth noting that nnDetection has demonstrated outstanding performance in various medical object detection tasks on public datasets, including LUNA16 (Setio et al., 2017), ADAM (Timmins et al., 2020), CADA (Hennemuth et al., 2021), RibFrac (Jin et al., 2020), LIDC-IDRI (Armato III et al., 2011), Kits19 (Heller et al., 2019), and 11 other tasks such as Decathlon (Simpson et al., 2019)-Pancreas, Prostate, and Hepatic Vessel. This track record highlights its versatility and potential for broader medical imaging applications, underlining its

significance in our study. In conclusion, nnDetection is a powerful method for meniscal tear detection, offering automation, specialization, and flexibility in design. These advantages make it a valuable tool in improving the accuracy and reliability of meniscal tear diagnosis in knee MRI images.

In this study, nnDetection was employed for meniscus tear detection. It was trained using 560 fully sampled training volumes with 5-fold cross-validation and then tested over 103 different reconstructed outcomes. An illustration of the overview of the proposed sequential process is depicted in Figure 4.3. The corresponding figure represents each step of the comprehensive multi-coil MRI reconstruction and detection pipeline together with expert analysis.

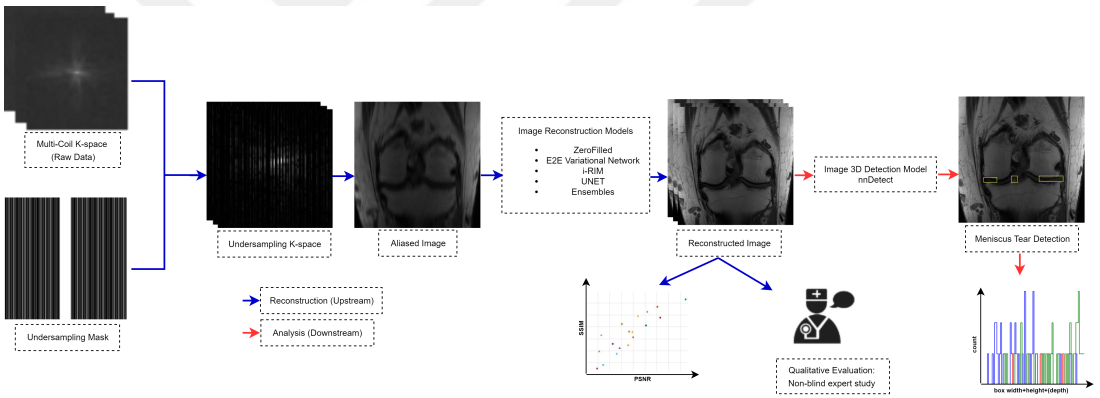


Figure 4.3 An overview of the comprehensive multi-coil MRI reconstruction, detection, and analysis pipeline. Initially, the acquired data are undersampled using cartesian trajectories, resulting in an aliased image. Subsequently, the aliased image is processed through three different reconstruction models (termed upstream, represented by the blue arrow) to transform it into a high-quality image. These reconstructed outputs are then assessed using both Image Quality Assessment (IQA) and expert studies. Following the evaluation, the region of interest (in this case, the meniscus tear) is localized (e.g., detected) and trained using a 3D nnDetection model with fully sampled scans. The trained model is then tested with reconstructed scans for further analysis (referred to as downstream, indicated by the red arrow)

4.4 Evaluation Metrics

Within this section, we delve into a detailed examination of the performance metrics employed for both image reconstruction and lesion detection tasks on the FastMRI dataset. By utilizing a dual focus on reconstruction and detection, we aim to provide a thorough understanding of the models' capabilities in addressing these critical aspects of medical imaging.

4.4.1 Reconstruction Metrics

Two established metrics were employed to evaluate the effectiveness of our reconstruction process. The Peak Signal-to-Noise Ratio (PSNR), a key metric, focuses on assessing local differences between the reconstructed and reference images. Additionally, the Structural Similarity Index (SSIM), another crucial quality metric, quantifies image similarity by considering the interdependencies among neighboring pixels, as highlighted in the work by Zbontar et al., 2018. These metrics provide valuable insights into the reconstruction performance, with PSNR emphasizing local fidelity and SSIM capturing broader image similarities. A detailed explanation of these metrics is provided in Section 3.5, and their mathematical representations are denoted in Equations 3.2 and 3.3, respectively.

4.4.2 Detection Metrics

To evaluate the object detector's performance accurately, it's essential to assess the accuracy of the detections, which involves various terms like True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN). These terms play a vital role in gauging the effectiveness of DL detectors in identifying meniscal tears, ultimately contributing to the assessment of their clinical relevance.

It is important to quantify these terms using a metric, and Intersection over Union (IOU) is one such metric. IOU calculates the extent of overlap between the ground

truth (gt) and the prediction (pd) by dividing the area of their intersection by the area of their union. In mathematical terms, $\text{IOU}(\text{gt}, \text{pd}) = \frac{\text{area}(\text{gt} \cap \text{pd})}{\text{area}(\text{gt} \cup \text{pd})}$. This metric provides a measure of how well the prediction aligns with the actual ground truth, indicating the accuracy of the detection.

The IOU metric varies from 0 to 1, with 1 indicating a perfect overlap between the prediction and the ground truth, and 0 indicating no overlap at all. By setting a specific IOU threshold value, denoted as α, we can determine the following:

- TP refers to detections where the IOU is greater than or equal to α, meaning the prediction aligns well with the ground truth.
- FP refers to detections where the IOU is less than α, indicating that the prediction is incorrect and does not match the ground truth.
- FN refers to missed ground truth instances where the IOU is less than α, showing that the detector failed to detect the actual object.

In summary, the IOU threshold α is used to classify detections as true positive, false positive, or false negative based on the level of overlap between the predicted and actual objects.

The performance metrics *Precision*, *Recall*, and *mean Average Precision (mAP)* were derived from TP, FP, and FN. These well-established metrics provide valuable insights into the accuracy of the model. The summary of these performance metrics is presented as follows:

- *Precision:* This metric quantifies the probability that positive class predictions are correctly classified as members of the positive class and is computed as TP/(TP + FP)
- *Recall:* This metric measures the probability of positive class predictions being correctly retrieved from all positive data and is calculated as TP/(TP + FN).

- *Mean Average Precision (mAP)*: The mean Average Precision (mAP) is computed as the sum of the average precision (AP) values for all classes divided by the total number of classes (N). This metric represents the area under the precision-recall curve over all classes and can be expressed as $\text{mAP} = \frac{\sum_{k=1}^{N} \text{AP}_k}{N}$, where AP_k corresponds to the average precision of class k, and N is the number of classes. The mean Average Precision (mAP) at an intersection over union (IOU) threshold of 0.1 and maximum 100 predictions per image for each class is determined.

Another important detection metric used in this study is the Free-Response Receiver Operating Characteristic (FROC) curve, which is widely employed for evaluating detection and localization tasks. The FROC curve assesses per-lesion sensitivity in relation to the average number of false positives per image (1/8, 1/4, 1/2, 1, 2, 4, 8) for varying detection thresholds. It provides valuable insights into the model's ability to detect and precisely localize abnormalities, making it a critical tool in performance evaluation for detection (Bandos et al., 2009).

4.5 Results and Analysis

In order to assess the performance of MRI reconstruction and detection, two sets of experiments were conducted using a standardized evaluation pipeline. The first set focused on comparing various reconstruction models to enhance analysis performance. Meanwhile, the second set examined how well these models could detect clinically significant outcomes using a detector. In this section, we present the results obtained from these two sets of experiments.

4.5.1 Quantitative Reconstruction Results

The primary focus of the quantitative evaluation process is on addressing background noise, which significantly influences the SSIM and PSNR scores. DL

models have demonstrated notable achievements in noise reduction, leading to improved PSNR and SSIM scores. Among the reconstruction models we implemented, E2E VarNET outperformed the UNET and iRIM models, achieving SSIM scores of 0.94 and 0.91, as well as PSNR scores of 40.29 and 37.94, for all contrasts (PD+PDFS) at 4x and 8x acceleration rates, respectively. The detailed results of the quantitative reconstruction evaluation can be found in Table 4.2.

Table 4.2 Reconstruction performances: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) along with their standard deviations for fastMRI dataset reconstruction baselines for All, PD and PDFS acquisitions accelerated at 4x and 8x.

Acceleration	Modality	Acquisition	PSNR (db)	SSIM
4x	**ZeroFilled**	ALL	31.73(±3.75)	0.83(±0.08)
		PD	32.08(±3.30)	0.85(± 0.06)
		PDFS	31.35(± 4.06)	0.81(± 0.09)
	UNET	ALL	37.65(±3.72)	0.91(±0.09)
		PD	38.4(±3.64)	0.94 (±0.04)
		PDFS	36.85(±3.13)	0.88(±0.10)
	iRIM	ALL	39.08(±4.90)	0.92(±0.10)
		PD	40.05(±4.15)	0.95(±0.04)
		PDFS	37.6(±3.68)	0.89(±0.10)
	E2E Variational Network	ALL	**40.29(±6.42)**	**0.94(±0.10)**
		PD	**42.45(±4.88)**	**0.97(±0.04)**
		PDFS	**38.0 (±4.37)**	**0.90(±0.10)**
8x	**ZeroFilled**	ALL	28.01(±3.49)	0.73(±0.10)
		PD	27.54(±3.13)	0.75(± 0.07)
		PDFS	28.51(± 3.60)	0.72(± 0.11)
	UNET	ALL	35.11(± 2.81)	0.88(± 0.10)
		PD	35.28(± 3.02)	0.91(± 0.15)
		PDFS	34.93(± 2.54)	0.85(± 0.11)
	iRIM	ALL	36.26(± 3.84)	0.89(± 0.11)
		PD	37.03(± 3.84)	0.92(± 0.05)
		PDFS	35.45(± 3.12)	0.86(± 0.11)
	E2E Variational Network	ALL	**37.94(± 4.40)**	**0.91(± 0.11)**
		PD	**39.19(± 3.75)**	**0.95(± 0.04)**
		PDFS	**36.61(± 3.39)**	**0.88(± 0.11)**

Drawing inspiration from the success of ensemble methods in classification tasks (Kuncheva, 2014) and segmentation competitions (Kamnitsas et al., 2018; Isensee et al., 2018), we investigated their potential in the context of MRI reconstruction. Employing three fundamental combination methods (Kavur et al., 2022) (Min/Max Combiner, Averaging Combiner, and Weighted Averaging), we evaluated their efficacy in improving the reconstruction process. The results, as presented in Table 4.3, reveal that, surprisingly, basic ensembles yielded slightly higher performance in the reconstruction task. Notably, Averaging and Weighted Averaging exhibited slightly superior PSNR and SSIM values compared to individual base learners (i.e. UNET, E2E VarNET, and iRIM).

4.5.2 Quantitative Detection Results

To assess the detection performance, we implemented the nnDetection self-configuring framework, which we fine tuned for 3D meniscus tear detection. The reconstruction results is presented in Table 4.4. Remarkably, the E2E VarNET model outperformed others, achieving impressive mAP and FROC scores of 0.69 and 0.72 at 4x, and 0.67 and 0.69 at 8x accelerations, respectively. These illustrate the superior capability of E2E VarNET in accurately identifying meniscus tears. Additionally, Figure 4.4 showcases the qualitative results of this detector at both acceleration rates. It becomes evident that at 4-fold and 8-fold acceleration, the detector successfully identifies the meniscus tear across all reconstruction outputs except ZeroFilled. This firmly establishes the effectiveness of the integrated framework when applied to raw k-space data. These findings indicate the potential of our approach to achieve significant and promising outcomes in prospective applications.

Moreover, our study encompassed a comprehensive comparison between our meniscus tear detection framework and existing methods reported in the literature, focusing specifically on precision and recall metrics. Our main goal was to conduct a rigorous evaluation of our approach's performance in comparison to established techniques for detecting meniscus tears. This detailed analysis aimed to highlight the

Table 4.3 Reconstruction performances of the ensemble methods: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) along with their standard deviations for fastMRI dataset reconstruction baselines for All, PD, and PDFS acquisitions accelerated at 4x and 8x.

Acceleration	**Modality**	**Acquisition**	**PSNR (db)**	**SSIM**
4x	**Averaging**	ALL	**40.76**(±**6.01**)	**0.94**(±**0.09**)
		PD	**42.76**(±**4.59**)	**0.97**(± **0.04**)
		PDFS	**38.64**(± **4.17**)	**0.91**(± **0.09**)
	Weighted Averaging	ALL	40.05(±4.97)	0.93(±0.09)
		PD	41.56(±3.97)	0.96 (±0.04)
		PDFS	38.44(±3.78)	0.90(±0.09)
	Max. Combiner	ALL	37.97(±4.42)	0.93(±0.09)
		PD	39.17(±3.79)	0.95(±0.04)
		PDFS	36.69(±3.54)	0.89(±0.09)
	Min. Combiner	ALL	38.47 ±4.13)	0.92(±0.10)
		PD	39.52(±3.71)	0.95(±0.04)
		PDFS	37.36 (±3.31)	0.90(±0.10)
8x	**Averaging**	ALL	38.13(±4.18)	0.91(±0.10)
		PD	39.27(±3.57)	0.95(± 0.04)
		PDFS	36.92(± 3.34)	0.88(± 0.11)
	Weighted Averaging	ALL	**38.31**(± **4.38**)	**0.91**(± **0.11**)
		PD	**39.56**(± **3.69**)	**0.95**(± **0.04**)
		PDFS	**36.99**(± **3.40**)	**0.88**(± **0.11**)
	Max. Combiner	ALL	35.62(± 3.15)	0.90(± 0.10)
		PD	36.10(± 3.2)	0.93(± 0.04)
		PDFS	35.11(± 2.79)	0.87(± 0.10)
	Min. Combiner	ALL	35.73(± 3.13)	0.89(± 0.11)
		PD	36.21(± 3.19)	0.93(± 0.04)
		PDFS	35.22(± 2.73)	0.86(± 0.11)

advantages and potential enhancements provided by nnDetection over existing methods, especially those that relied on limited annotations. The conclusive results of this extensive assessment, as depicted in Table 4.5, incontrovertibly establish our proposed method's superior performance compared to Zhao et al.'s approach in meniscus tear detection.

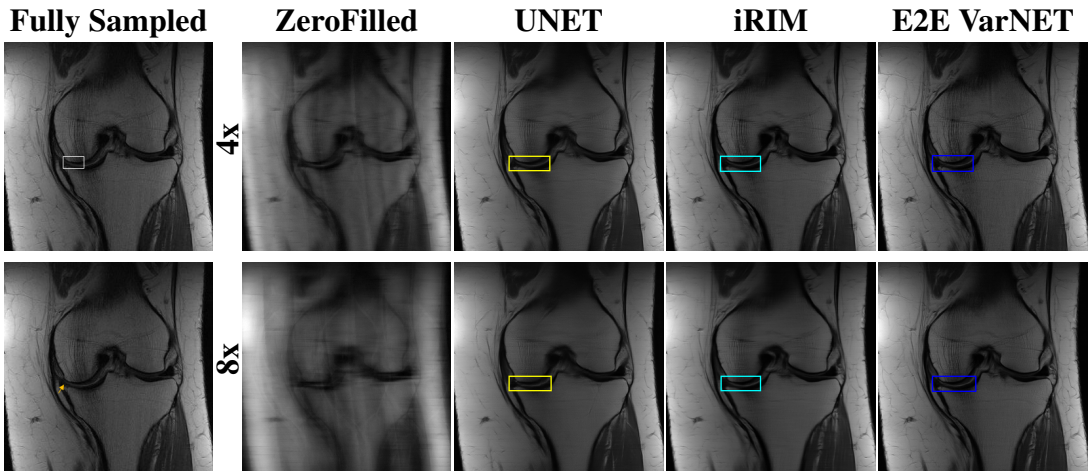


Figure 4.4 A demonstration of the nnDetection outcomes represented by bounding boxes for different reconstructions. At 4-fold and 8-fold acceleration (top row and bottom row), the detector successfully identified the meniscus tear across all reconstruction methods except for ZeroFilled. These bounding boxes shown in the illustration are 2D representations of the 3D prediction boxes. The arrow in the first image of the second row points the meniscus tear.

While the quantitative reconstruction results indicated a slightly higher performance of basic ensembles in the reconstruction task, these results did not translate into superior performance in the detection task. Therefore, we decided not to include ensemble results, as they demonstrated similar performance to the base learners (i.e. UNET, E2E VarNET, and iRIM).

The conclusive results of this extensive assessment, presented in Table 4.5, undeniably demonstrate the superior performance achieved by the proposed method in meniscus tear detection.

4.5.3 Expert Analysis of Reconstruction and Detection Results

In this section, we aim to analyze the image and diagnostic quality of image reconstructions using two expert radiologists. In previous research, Muckley et al. (Muckley et al., 2021) introduced a radiologists' ranking system based on the quality of pathology depiction and identified hallucinations in the fastMRI brain dataset challenge. Knoll et al. (Knoll et al., 2020), on the other hand, ranked radiologists' submissions in the fastMRI knee challenge by considering both artifacts and Image

Table 4.4 Detection performances: FROC score IOU 0.1 and AP@0.1 Max. Det. 100 results for All, PD, and PDFS Accelerated at 4x and 8x.

Acceleration	Modality	Acquisition	FROC score IOU 0.1	AP@0.1 Max.Det. 100
4x	UNET	ALL	0.68	0.66
		PD	0.69	0.68
		PDFS	0.67	0.65
	iRIM	ALL	0.7	0.67
		PD	0.69	0.66
		PDFS	0.7	0.68
	E2E Variational Network	ALL	**0.72**	**0.69**
		PD	**0.73**	**0.7**
		PDFS	**0.7**	**0.68**
8x	UNET	ALL	0.66	0.64
		PD	0.7	0.68
		PDFS	0.63	0.61
	iRIM	ALL	0.68	0.65
		PD	0.69	0.64
		PDFS	0.66	0.66
	E2E Variational Network	ALL	**0.69**	**0.67**
		PD	**0.71**	**0.69**
		PDFS	**0.67**	**0.66**

Quality Assessment (IQA) metrics.

Despite these efforts, certain aspects remained unaddressed, including the specific types and locations of artifacts/hallucinations present in the knee dataset, as well as the explicit count of these occurrences. To provide answers to these questions, we conducted an expert analysis.

The expert study involves qualitative evaluation of the original and DL-reconstructed images. Two radiologists, one with with 20-years of experience (referred as M.M.B.) and the other with 5-years of experience (referred as A.C.), participated to our study. They meticulously observed and documented the hallucinatory effects and artifacts introduced by the reconstruction models. The final

Table 4.5 The proposed framework for meniscus tear detection compared with existing methods available for this task

Author	Acceleration	Modality	Precision	Recall
Zhao (Zhao et al., 2021) (For a limited dataset of 205 samples)	4x	**UNET**	0.58	0.34
		iRIM	-	-
		E2E Variational Network	0.52	0.43
Proposed		**UNET**	**0.82**	**0.48**
		iRIM	**0.87**	**0.52**
		E2E Variational Network	**0.87**	**0.56**
Zhao (Zhao et al., 2021) (For a limited data of 205 samples from FastMRI+)	8x	**UNET**	0.4	0.2
		iRIM	-	-
		E2E Variational Network	0.52	0.37
Proposed (Complete FastMRI+ dataset is used)		**UNET**	**0.8**	**0.49**
		iRIM	**0.80**	**0.5**
		E2E Variational Network	**0.87**	**0.57**

conclusions are established through their mutual agreement. Examples of these hallucinations are presented in Figure 4.5, 4.6, and 4.7. Some of the models transformed certain structures into abnormalities, which could potentially impact clinical evaluations. These abnormalities were identified as FP artifacts, as they appeared in the reconstructed images, but were not present in the original ones. Additionally, artifacts were observed to extend across anatomical areas such as the femur, tibia, fibula, and other regions. For instance, Figure 4.5 illustrates scattered linear artifacts surrounding the femur, with some appearing in a fogging-style Figure 4.5 (b). Moreover, linear artifacts were also evident in the tibia, proximal tibia, and fibula, as shown in Figure 4.6.

The presence of hallucinations has a negative impact on expert sentiments and raises concerns for clinical applications. These hallucinations often mimic normal structures that either do not exist or indicate abnormalities, posing significant

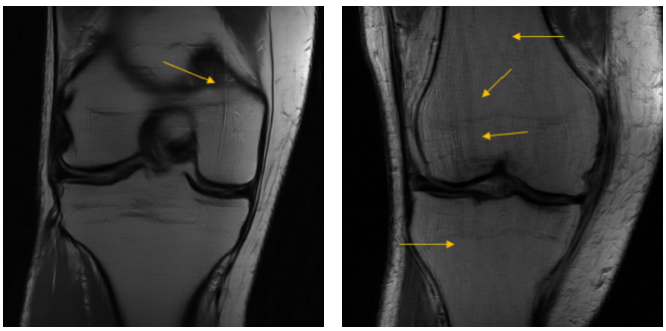
challenges. Moreover, some of these hallucinations lack realism, complicating the issue even further. As shown in Figure 4.7, increased pseudo and linear signal artifacts in the meniscus can lead to confusion in tear assessment, and bone marrow edema can be misdiagnosed, potentially resulting in incorrect treatment.

These indicate that, DL may not be fully optimized to handle hallucinations despite their high performance in terms quantitative metrics. Moreover, adversarial perturbations have demonstrated the instability of DL models (Antun et al., 2020). These hallucinations and artifacts persist as significant challenges across all acceleration rates. In our research, experts thoroughly diagnosed and counted these hallucinations in the femur, tibia, and FP artifacts on a slice-by-slice basis. To ensure a clear illustration, all artifact counts were subjected to logarithmic scaling (using the formula 1+log10(count)). This scaling method enhances the visibility of differences in artifact countsand facilitates both quantitative analysis and visual interpretation. The bar plots for both acceleration rates are depicted in Figure 4.8, showing the total number of artifacts.

Figure 4.8 (a) indicates that UNET exhibits more artifacts in the tibia and femur at both acceleration rates compared to the other models. On the other hand, Figure 4.8 (b) demonstrates that E2E VarNET and iRIM have more FP artifacts at both acceleration rates.

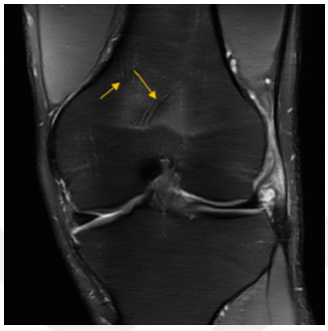
The observed hallucinations and artifacts present significant challenges for future research and development. Addressing and resolving these issues will be a critical focus to improve the reliability and accuracy of deep learning models in medical image reconstruction and analysis. Tackling these challenges is essential to ensure that DL models produce more realistic and clinically valid results, reducing the impact of hallucinations and artifacts on diagnostic accuracy and treatment decisions. Future efforts should concentrate on finding solutions and advancements to enhance the overall performance and robustness of DL-based approaches in medical imaging, thereby advancing patient care and outcomes.

Furthermore, as part of this study's framework, a comprehensive analysis on image

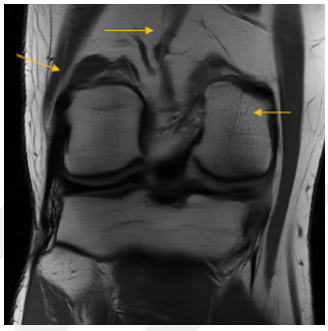


(a) Artifacts in the medial femur

(b) Fogging-style artifact in the distal femur



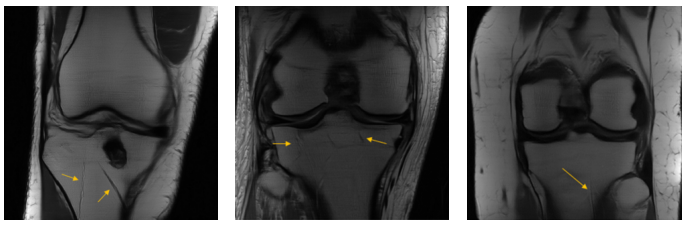
(c) Artifacts in the distal femur



(d) Lineal artifacts in the femur

Figure 4.5 Illustrations of hallucinations observed in the DL-based reconstruction outcomes, specifically on the femur, medial, or distal femur regions. Among these examples, (a), (c), and (d) represent linear, (b) exhibits fogging-style artifacts. The hallucinations are not confined to a single location, but are scattered across the images, highlighted by arrows.

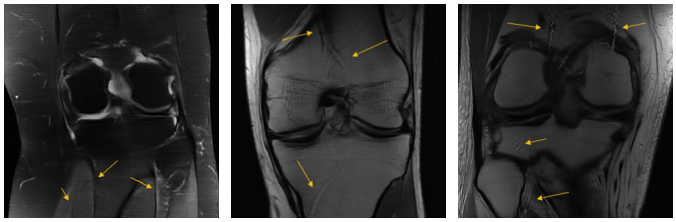
quality was conducted. This thorough assessment encompassed a meticulous examination of each dataset, involving a detailed analysis of individual slices. The radiologists maintained an unbiased approach throughout their assessment, focusing solely on the reconstruction techniques while remaining unaware of patient-specific details or external factors. The assessment of the reconstructed images was carried out through a structured 5-point scoring system, ranging from "1," indicating a notable deterioration compared to the reference image, to "5," denoting a significant enhancement surpassing the reference image's quality. Both the original reference images and the reconstructed counterparts underwent this evaluation, with the evaluator assigning standardized quality scores using the 5-point ordinal scale. This rigorous evaluation process allowed for a comprehensive understanding of the impact and efficacy of the employed reconstruction methods. The averaged results, as depicted in Figure 4.9, indicate that E2E VarNET showed the best performance at



(a) Lineal artifacts in the tibia

(b) Minimal hypointense lineal artifacts in the proximal tibia

(c) Artifacts in the fibula.



(d) Artifacts in the tibia and fibula

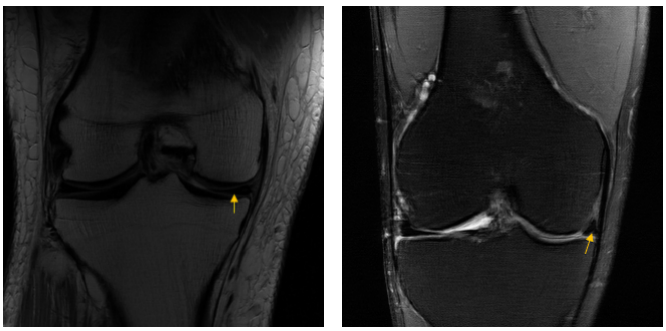
(e) Lineal artifacts observed in the femur and tibia

(f) Artifacts observed in the proximal tibia and fibula

Figure 4.6 Illustrations of hallucinations resulting from DL-based reconstruction, primarily found in the tibia and scattered throughout the images. The examples are identified in various regions including (a) tibia, (c) fibula, (d) both tibia and fibula, (e) femur and tibia, and (f) proximal tibia and fibula. All appear as lineal artifacts. Additionally, (b) shows minimal hypo-intense lineal artifact in the proximal tibia. The instances are marked with an arrow.

both acceleration rates, followed by iRIM and UNET. In addition to this, no visual received a perfect ’5’ rating in the evaluation. These quality assessment findings align with the image quality metrics (SSIM and PSNR) across all images and acceleration rates. This consistency further supports the superior performance of E2E VarNET in the reconstruction task compared to the other models.

To evaluate the pathological features related to meniscal tear in all volumes, a 0-4 ordinal scoring system was used for all slices. In this scoring system, a score of "0" indicated that the meniscus tear was not observable, "1" and "2" indicated far worse and worse compared to the reference, "3" indicated similarity to the reference, and "4" indicated better than the reference. Figure 4.10 displays the mean overall pathological assessment scores for meniscal tear. The results demonstrate that all DL models achieved similar scoring outcomes, with E2E VarNET showing a slight advantage over iRIM and UNET at both acceleration rates. These averaged expert pathological scores align well with the performance of the nnDetection detector, as

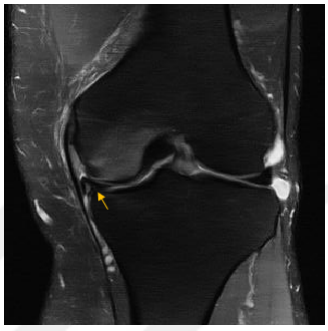


(a) Increased pseudo signal in the meniscus

(b) Linear increased signal in the meniscus



(c) False-positive artifacts (Bone Marrow Edema)



(d) Increased pseudo signal observed in the lateral meniscus

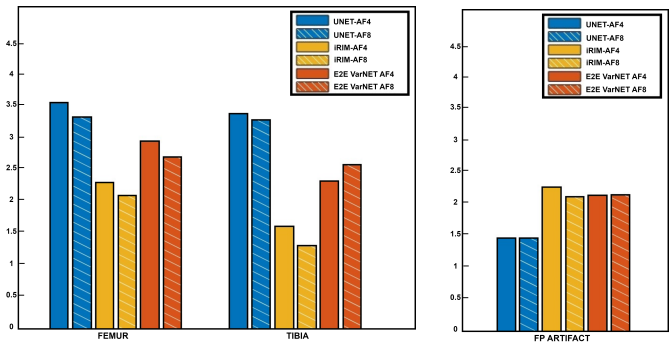
Figure 4.7 The hallucinations pointed by arrows exhibit realistic structures artificially generated by the DL-based methods and are referred as the "FP Artifact"s. Examples (a), (d), and (b) demonstrate increased pseudo and linear signals in the meniscus, potentially leading to misdiagnosis. Example (c) presents artificially generated "Bone Marrow Edema" that is not present in the fully sampled images.

evidenced by both the FROC and mAP scores. The expert evaluations of meniscal tear pathology are consistent with the results obtained from the DL models, validating the reliability of the detection outcomes.

4.6 Conclusion and Future Work

In this research, we conducted a comprehensive evaluation of MRI reconstruction and its analysis, specifically focusing on clinically relevant details related to meniscus tears. Our study aimed to assess the performance of deep learning (DL) detection for different DL reconstructions, which had been analyzed separately in the existing literature so far.

Our findings demonstrate that DL reconstruction performance has a substantial



(a) (b)

Figure 4.8 The hallucinations observed in the femur, tibia, and FP artifacts were carefully diagnosed and counted slice by slice by an expert. The counts were then scaled logarithmically using the formula (1+log10(count)). The resulting counts are presented as bar plots for both acceleration rates. The bar plots in (a) display the hallucination counts from DL-reconstruction results in the femur and tibia, while (b) shows the count of FP artifacts.

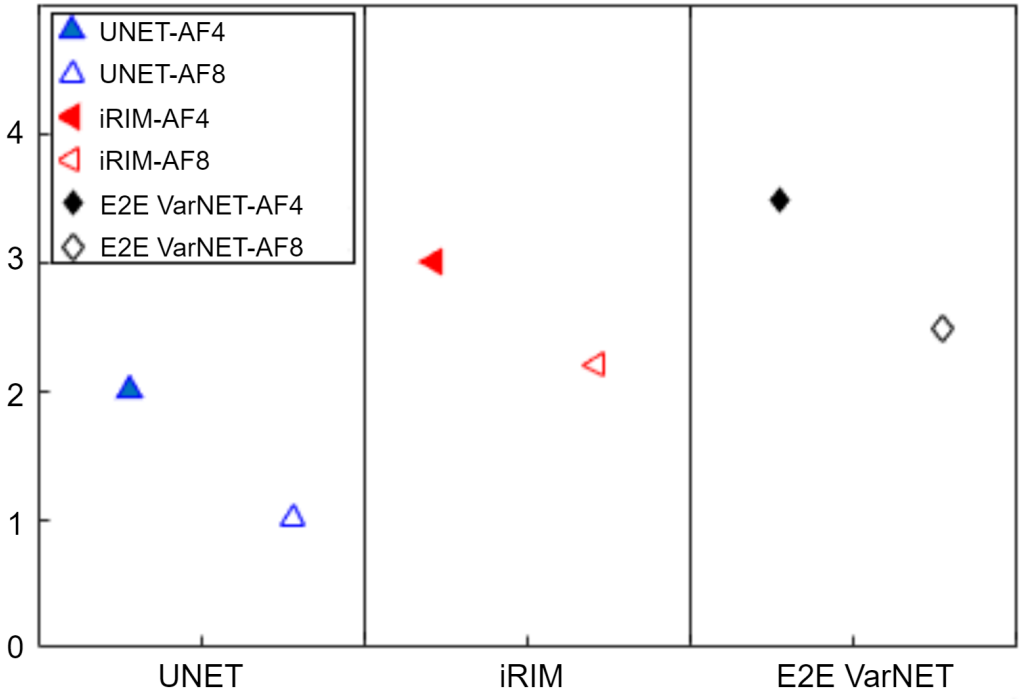


Figure 4.9 The plot illustrates the average diagnostic scores from radiological assessments for the reconstructed images. The scoring system is summarized as follows: 1 for far worse, 2 for worse, 3 for alike, 4 for better, and 5 for far better compared to the reference image. The signs on the plot represent the calculated average of expert scores.

impact on detection performance, and these effects vary significantly between DL detectors and expert radiologists. Through our extensive analysis, we propose a comprehensive approach that integrates reconstruction and detection strategies. As a

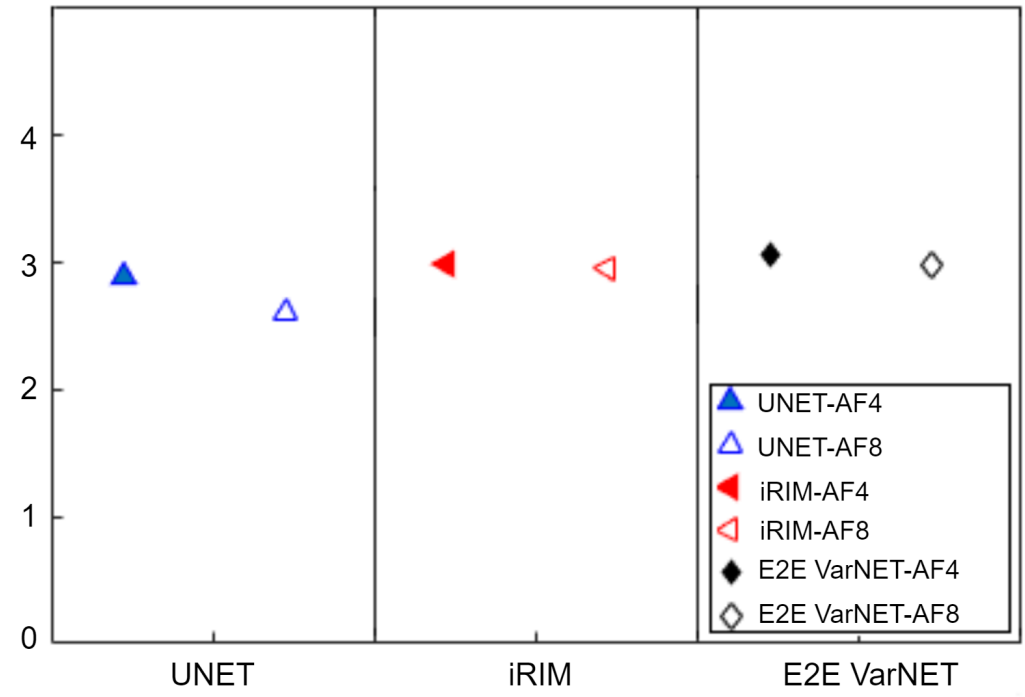


Figure 4.10 The plot displays the average scores obtained from radiological evaluations of pathological features in the reconstructed images. The scoring system used is as follows: 0 indicates that the pathology is not observable, 1 for far worse, 2 for worse, 3 for similar, and 4 for better compared to the reference image.

result, we achieved state-of-the-art (SOTA) performance in detecting meniscal tears in the FastMRI+ dataset. Overall, our study enriches the field by providing valuable insights into the dynamic synergy between MRI reconstruction and DL-based detection, revealing the effectiveness of a comprehensive strategy for enhancing meniscal tear detection.

In our study, we employed three distinct DL models for the MRI reconstruction task and evaluated their results using a well-established DL detector, nnDetection, on the FastMRI+ dataset to detect Meniscus Tear. The reconstruction outcomes obtained from these models, which displayed high and comparable Image Quality Assessment (IQA) scores, were utilized as input for the DL detector. Furthermore, the clinical relevance and applicability of our approach were underscored by the meticulous evaluations conducted by expert radiologists. By incorporating these elements into our study design, we ensured both the technical effectiveness and clinical significance of our DL methods for diagnosing meniscal tears in knee MRI. Additionally, expert

radiologists evaluated these outcomes for their suitability in clinical diagnosis.

In terms of clinical diagnostic quality assessment, the UNET model did not perform as well as iRIM and E2E VarNET. However, when specifically evaluating the detection of meniscus tears, UNET showed similar performance to iRIM and E2E VarNET. Interestingly, when analyzing hallucination and False Positive (FP) artifact counts in relation to tibia and femur artifacts, the results indicated a different trend, with UNET showing more of these artifacts compared to iRIM and E2E VarNET.

Although the presence of the most counted and scattered artifacts in the femur and tibia has negatively affected experts' diagnosis assessments, it has not significantly influenced pathological evaluation. Surprisingly, E2E VarNET, which performed best in both diagnosis and pathological assessment, had the highest number of FP artifacts, along with iRIM. This discrepancy is not consistent with the IQA, detection metric scores, and the evaluations made by radiologists regarding pathological and diagnostic aspects. Additionally, the issue of hallucinations presents a significant concern in the assessment process, particularly in clinical applications. These observations emphasize the complexity of the relationship between image artifacts, diagnostic accuracy, and the effectiveness of deep learning models in medical imaging. Future research endeavors should prioritize the mitigation of these challenges to further enhance the clinical applicability of DL-based approaches in medical imaging.

Nevertheless, we observed that the existing evaluation results for both reconstruction and detection are in agreement with our radiologists' pathological and diagnosis assessments of clinically important pathologies, considering both qualitative and quantitative aspects. By introducing this comprehensive evaluation pipeline for clinically important pathologies, we anticipate that the proposed framework will facilitate clinical utility in MRI reconstruction and detection.

In future research, we aim to assess the efficacy of our proposed framework using a prospective sampling strategy to enhance its relevance in clinical applications. It is crucial to recognize that retrospective undersampling may not fully account for signal

relaxation differences during echo trains and lacks equivalence in a prospective setting. Furthermore, our vision includes a comprehensive evaluation of the framework's performance under various acquisitions, orientations, and across different vendors. This extended scope offers opportunities to delve into and comprehend the relative strengths and weaknesses of parallel imaging approaches such as GRAPPA, SENSE, and PI. Analyzing these methods in future research promises valuable insights for augmenting the overall quality and applicability of MRI reconstruction in clinical practice.

Additionally, we plan to conduct a thorough performance evaluation of our framework with diverse acquisitions, orientations, and different vendors by incorporating more prospective cases to simulate real clinical setups. We also aspire to explore and implement genuine end-to-end models where a singular neural network is responsible for both image reconstruction and subsequent pathology detection. This investigative step can leverage the benefits of end-to-end training, potentially enhancing the overall performance and efficiency of our diagnostic pipeline.

In conclusion, this study represents significant outcomes in clinical translation of MRI reconstruction and detection. By leveraging both reconstruction and detection from k-space, we integrated both radiologist and patient perspectives into clinical settings. Our proposed framework demonstrates the ability to deliver high-quality results that align well with radiologist assessment. This approach has the potential to meet the growing demand for advanced imaging techniques in clinical practice and cater to a broader patient population. The combination of reconstruction and detection holds promise for improving diagnostic accuracy and enhancing the overall efficacy of MRI in healthcare.

CHAPTER FIVE
ENHANCING DEEP LEARNING PERFORMANCE IN ACCELERATED KNEE MRI RECONSTRUCTION: A FOCUS ON BLIND AND NON-BLIND EVALUATIONS

5.1 Introduction

In the rapidly evolving landscape of medical imaging, recent studies have showcased the transformative potential of employing deep learning methods for knee MRI reconstruction from undersampled k-space. This approach not only holds promise in significantly improving the detection of meniscus tears but also demonstrates the ability to enhance diagnostic accuracy, addressing inherent limitations in traditional MRI techniques (Harman et al., 2023). Alongside these advancements, brain MRI super-resolution has emerged as a powerful tool, showcasing its capability to elevate the visualization of cortical parcellation patterns (Tian et al., 2021). Moreover, prospective super-resolution techniques have proven effective in simultaneously elevating image quality and preserving quantitative accuracy for parametric T_2 mapping, as well as enhancing the diagnostic accuracy of rapid MRI scans (Chaudhari et al., 2018, 2021).

Nevertheless, given that the establishment of innovative evaluation criteria to enhance image quality is an area of ongoing research, the utilization of reconstruction methods free from predefined models and the assessment of their clinical utility downstream offer valuable perspectives on the general effectiveness and applicability of the proposed techniques.

In this context, this study addresses the critical gap associated with both image reconstruction and detection tasks using the benchmarking fastMRI dataset (Zbontar et al., 2018) and its extended clinical annotations in the FastMRI+ dataset (Zhao et al., 2021). These datasets offer a standardized large-scale evaluation platform for state-of-the-art machine learning (ML) methods in magnetic resonance (MR) reconstruction and detection. The fastMRI dataset includes undersampled k-space

data for knee MR images, along with fully sampled ground truth data, while the FastMRI+ dataset provides expert-labeled bounding boxes for identifying twenty-one different pathologies from the multi-coil knee dataset. In addition to standard medical image classification and detection tasks, FastMRI+ enables the analysis of reconstruction frameworks to retain crucial clinical information by observing classification and detection outcomes (Zhao et al., 2021).

While many DL methods primarily focus on achieving high-quality reconstruction from accelerated MRI acquisition (Fabian et al., 2022; Xin et al., 2023; Liu et al., 2023; Sriram et al., 2020, Sriram et al., 2020; Kusakunniran et al., 2023; Yan et al., 2022), the thorough evaluations involving expert readings are still evolving (Muckley et al., 2021; Knoll et al., 2020). Recent prospective results demonstrated that DL reconstruction of accelerated knee MRI allowed for nearly a two-fold reduction in scan time, improved image quality, and provided equivalent diagnostic utility (Johnson et al., 2023). However, these methods have predominantly focused on reconstructing the image and have not addressed automatic deep learning-based lesion detection and its expert evaluation in both the reconstruction and detection phases.

Moreover, to the best of our knowledge, no study on automated pathology detection, as well as blind and non-blind evaluation on deep learning-based reconstructions, has been conducted for the fastMRI dataset. Therefore, it remains uncertain whether DL detectors can accurately identify clinically significant details and their concordance with radiologists, with or without blind assessment, when reconstruction models are applied. Additionally, existing studies have limitations, including the inadequate alignment between radiologists' evaluations and the utilization of evaluation metrics for image quality determination.

To address these concerns, we employed the nnDetection tool to detect meniscal tear pathology in images generated from various DL reconstruction models. For the reconstruction phase, three radiologists (two non-blind, who see the reconstruction models and fully sampled scans, and one blind, who has never seen any information about the models and only assesses the reconstruction outcomes) evaluated the

reconstruction for both diagnostic and pathological quality. Additionally, the radiologist conducted blind assessments annotated meniscus tears using a commercial browser-based annotation platform (MD.ai, New York, NY), allowing adjustments of brightness, contrast, and magnification. This study aims to explore the concordance between blind and non-blind assessments, the existence of meniscus tears in reconstruction outcomes, and the performance of DL detectors, shedding light on their potential clinical utility.

In light of the study's objectives and findings, the primary contributions can be outlined as follows:

- **Evaluation Framework Development:** We introduce a comprehensive evaluation framework that seamlessly integrates both blind and non-blind assessments for image reconstruction and lesion detection tasks. This framework is purposefully designed to unveil the limitations of traditional image quality metrics. It accomplishes this by aligning reconstruction results with quantitative standard evaluations (SSIM, PSNR) and non-blind pathological assessments. In addition to its focus on reconstruction, the framework extends its scope to uncover the limitations of traditional detection metrics. This expansion is achieved by aligning quantitative detection evaluations (mAP, Precision, Recall scores) with blind annotation assessments. The synergy between blind annotations and quantitative detection results (mAP, Precision, Recall scores) enhances the framework's capacity to thoroughly assess the clinical utility of DL-accelerated MRI reconstructions. The non-blind diagnostic assessment establishes a strong correlation between reconstruction outcomes, diagnostic evaluations, and established quantitative standards. Simultaneously, the blind annotation results align closely with quantitative detection metrics, providing a nuanced understanding of the overall performance. This integrated approach ensures a comprehensive evaluation of both the reconstruction process and the detection of pathological features, contributing to a thorough assessment of DL-accelerated MRI reconstructions.

- **Concordance Analysis:** We perform an exhaustive examination of the concordance between blind and non-blind assessments and image quality metrics (reconstruction + detection). This investigation delves into how the inclusion of blind evaluations shapes the interpretation of DL-generated reconstructions, alongside non-blind diagnoses and pathological assessments. Despite the presence of apparent hallucinations impacting all aspects, the analysis reveals alignment with quality metrics. This comprehensive assessment significantly enriches our understanding of the reliability and consistency of expert evaluations in the context of DL-accelerated MRI images.

- **Performance Assessment of DL Detectors:** Our study assesses the effectiveness of DL detectors in identifying clinically relevant features, specifically meniscal tears, within accelerated knee MRI reconstructions. Through a comparison of outcomes with both blind and non-blind expert evaluations, we offer valuable insights into the reliability and accuracy of DL detectors in a clinical context. Our findings reveal a stable alignment between DL detector results and image quality assessment metrics, as well as non-blind expert scores. An integrated approach, combining reconstruction and detection processes for meniscal tears on FastMRI+ data, demonstrates superior performance. The E2E VarNET achieves higher average precision scores of 0.69 and 0.67 at 4- and 8-fold accelerations, respectively, compared to blind expert study scores of 0.60 and 0.44 with the E2E VarNET and iRIM models. Notably, the UNET model attains average precision values of 0.66 and 0.64, surpassing scores of 0.29 and 0.31, while the iRIM model achieves values of 0.67 and 0.65 at 4-8 fold accelerated scans, outperforming scores of 0.40 and 0.44. These results underscore the superior performance of our proposed pipeline. In a blind expert study with E2E VarNET, an mAP of 0.32 at 8-fold acceleration is observed, which is comparatively lower than the detector's performance. This discrepancy emphasizes the potential advantages of DL detectors in enhancing sensitivity and detection accuracy, even under challenging conditions, further solidifying their clinical relevance in the context of accelerated MRI scans.

- **Dataset Utilization:** We leverage the benchmarking fastMRI dataset and its extended FastMRI+ clinical annotations, offering a standardized and large-scale platform for evaluating state-of-the-art ML methods in both MRI reconstruction and lesion detection. The use of these datasets enhances the generalizability of our findings to real-world clinical scenarios.

- **Clinical Applicability Insights:** Our study offers valuable insights into the broader clinical applicability of DL methods in musculoskeletal diagnosis, illuminating both the potential benefits and challenges associated with DL-accelerated MRI reconstructions. This contribution advances the ongoing discussion surrounding the integration of DL techniques into routine clinical practice. However, it is important to note the inherent challenges posed by artifacts in this context. These artifacts, though challenging, are evident and can potentially mask or obscure primary pathologies or mimic different pathologies, leading to the possibility of misdiagnosis. This emphasizes the need for a nuanced understanding of the limitations and considerations associated with DL-accelerated MRI reconstructions in the clinical setting.

The remainder of this chapter is organized as follows. Section 5.2 provides a background on MRI reconstruction and reviews relevant studies. Section 5.3 details the materials, methods, and datasets utilized for reconstruction and analysis. In Section 5.4, we shortly present the image quality and detection metrics. The results, along with the analysis and blind/non-blind expert evaluations, are discussed in Section 5.5. Finally, Section 5.6 concludes the study, summarizing key findings and outlining potential avenues for future research.

5.2 Background and Related Research

Traditional MRI encounters a significant challenge in slow data acquisition, attributed to the point-by-point Fourier transformation in k-space. Adhering to Nyquist criteria during k-space sampling is crucial for artifact avoidance. However,

accelerated MRI introduces sub-sampling below the Nyquist rate, expediting the process but leading to aliasing artifacts and a classical ill-posed inverse problem. The overarching goal of accelerated scans is to surmount these challenges and achieve high-quality original images.

In recent years, DL models have become integral to critical radiological tasks, including detection, classification, and reconstruction, yielding notable success across diverse applications (Litjens et al., 2017). Despite these advancements, studies addressing key aspects such as expert evaluation, concordance with traditional metrics, pathology, and artifacts remain limited.

Several investigations delve into DL-based reconstruction's radiological efficacy. Lee et al. (Lee et al., 2023) highlight DL's impact on prostate imaging, demonstrating enhancements in acquisition time, Signal-to-Noise Ratio (SNR), Contrast-to-Noise Ratio (CNR), and overall image quality through expert assessments. Zhou et al. (Zhou et al., 2022) present a self-supervised learning approach for non-cartesian MRI, showcasing successful training without fully sampled data. Other studies (van Amerom et al., 2018; Ghodrati et al., 2019; Han et al., 2018; Tolpadi et al., 2023) explore applications in fetal cardiac anatomy, cardiac MR imaging, k-space trajectory enhancement, and end-to-end MRI reconstruction, showcasing DL's adaptability in addressing diverse challenges.

Furthermore, investigations by Johnson et al. (Johnson et al., 2023, 2022) and Ni et al. (Ni et al., 2023) evaluate DL-reconstructed knee MRI and AI-assisted compressed sensing in knee MRI. Bae et al. (Bae et al., 2022) and Afat et al. (Afat et al., 2023) demonstrate the clinical feasibility of accelerated deep learning–reconstructed diffusion-weighted imaging (DWI) in abdominal and liver MRI. Studies by Lee et al. (Lee et al., 2022), Sauer et al. (Sauer et al., 2023), and Tajima et al. (Tajima et al., 2022) explore DL-accelerated imaging in breast cancer, breast DWI, and prostate cancer patients undergoing Diffusion-weighted Imaging with Background Suppression (DWIBS).

Additionally, Kim et al. (Kim et al., 2024), Fujima et al. (Fujima et al., 2023), Kim

et al. (Kim et al., 2023), and Wilpert et al. (Wilpert et al., 2023) delve into DL's applications in whole-spine DWI, head and neck DWI, liver evaluation, and breast MRI. These studies collectively contribute to the evolving landscape of DL applications in MRI, illustrating its potential to elevate image quality, shorten acquisition times, and enhance diagnostic confidence across diverse anatomical regions.

5.3 Materials and Methods: Experimental Design and Implementation

In this study, our primary objective is to capitalize on advancements in Deep Learning (DL) models, particularly in the context of undersampled k-space (aliased images). Our focus extends beyond traditional applications, aiming to integrate this progress into a comprehensive pipeline that addresses both upstream and downstream challenges. Specifically, we aim to tackle the downstream task of pathology detection, encompassing various reconstruction outcomes, while also incorporating blind and non-blind radiologist assessments in both directions. The study pipeline is visually represented in Figure 5.1.

To achieve this, undersampled k-space serves as the input to three distinct reconstruction models. Our evaluation strategy involves a meticulous examination of meniscus tears through separate blind and non-blind expert studies, alongside DL processes. For the non-blind expert study, we conducted a detailed analysis, considering diagnosis and pathological assessments by radiologists. This involved a thorough examination of imaging data and clinical records to ensure a comprehensive evaluation, leading to accurate and reliable conclusions. Furthermore, the Detection model is integrated with the three distinct reconstruction outcomes. In the final step of our approach, we compare quantitative DL detection results with blind expert study findings, ensuring an unbiased and rigorous assessment.

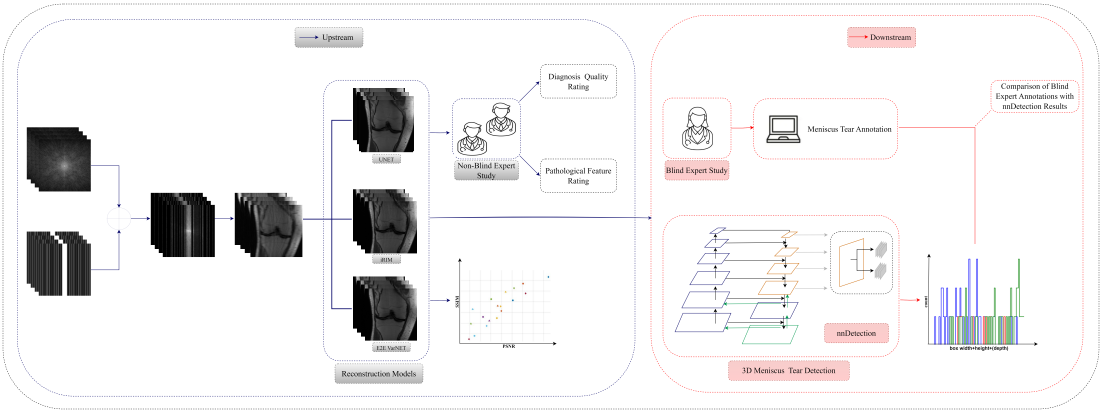


Figure 5.1 An overview of the comprehensive multi-coil MRI reconstruction, detection, and analysis pipeline. The acquired data undergo initial undersampling using cartesian trajectories, resulting in an aliased image. The aliased image is then processed through three distinct reconstruction models(UNET, iRIM, and E2E VarNET) (referred to as upstream, depicted by the blue arrow), transforming it into a high-quality image. The reconstructed outputs undergo assessment through both Image Quality Assessment (IQA) and non-blind expert evaluations, covering both diagnosis and the identification of pathological features. Following the evaluation, the region of interest, in this case, the meniscus tear, is localized (detected) and trained using a 3D nnDetection model, utilizing fully sampled scans. This trained model is subsequently tested with reconstructed scans for further analysis (referred to as downstream, indicated by the red arrow). For the comparison of meniscus tears, reconstruction outcomes are also evaluated by a blind study, conducted without any information about the models or accelerations used. The blind expert annotates meniscus tears, and the annotation results are finally compared with both FastMRI+ annotations and nnDetection results for a comprehensive assessment.

5.3.1 Dataset

Our study relies on the fastMRI dataset, thoroughly explained in (Zbontar et al., 2018) and Section 4.3.1, accessible at https://fastmri.med.nyu.edu/. This dataset is crucial for our research, providing diverse data neatly divided into training (973 volumes), validation (199 volumes), and test sets (118 volumes). It forms the foundation for developing our model and conducting a comprehensive evaluation. The undersampling trajectory we used is based on predefined acquisition paths provided by the fastMRI dataset. For additional details about the fastMRI dataset and undersampling trajectories, please see Section 4.3.1.

The FastMRI+ dataset, an extension introduced by Zhao et al. (Zhao et al., 2021), can be found at https://github.com/microsoft/fastmri-plus and is detailed in Section 4.3.1. Specifically designed for comprehensive MRI reconstruction

evaluations, FastMRI+ includes fully sampled images from the fastMRI multicoil knee dataset. It goes beyond by annotating clinically significant pathologies and intentionally introducing clinically limiting artifacts, ensuring the inclusion of critical clinical information. This is crucial for assessing how different reconstruction algorithms preserve these clinical details. The dataset includes 1172 reconstructed k-space volumes in the training and validation sets, featuring 6154 bounding boxes and 13 study-level labels, providing a diverse dataset for comprehensive evaluation.

In the training set, 815 volumes are labeled with pathology, and 12 volumes are marked as artifacts. The validation set consists of 154 pathology-labeled volumes, including one volume marked with an artifact, contributing to the dataset's comprehensiveness. Among these annotated volumes, the training set includes 560 volumes mostly labeled with 4844 annotations as meniscus tear, while the validation dataset has 103 volumes with 814 annotations for meniscus tears.

For a thorough breakdown of meniscal tears and artifacts, refer to the specifics outlined in Table 5.1, accompanied by a visual representation in Figure 5.2.

Table 5.1 Categorization of Meniscus Tear and Artifact Labels with Subset Breakdown

Label	**Annotation Count**		**Subject Count**	
	Training Set	**Validation Set**	**Training Set**	**Validation Set**
Meniscus Tear	**4844**	**814**	**560**	**103**
Artifact	–	–	12	1

5.3.2 Reconstruction Models

In the context of accelerated MRI reconstruction endeavors, we utilized the capabilities of three distinct deep learning models, each meticulously selected for its proven proficiency in addressing the challenges associated with undersampled k-space. The strategic use of multiple models enables us to explore diverse approaches and maximize the potential for high-fidelity image reconstruction.

As we delve into the intricate details of each model, unraveling their respective

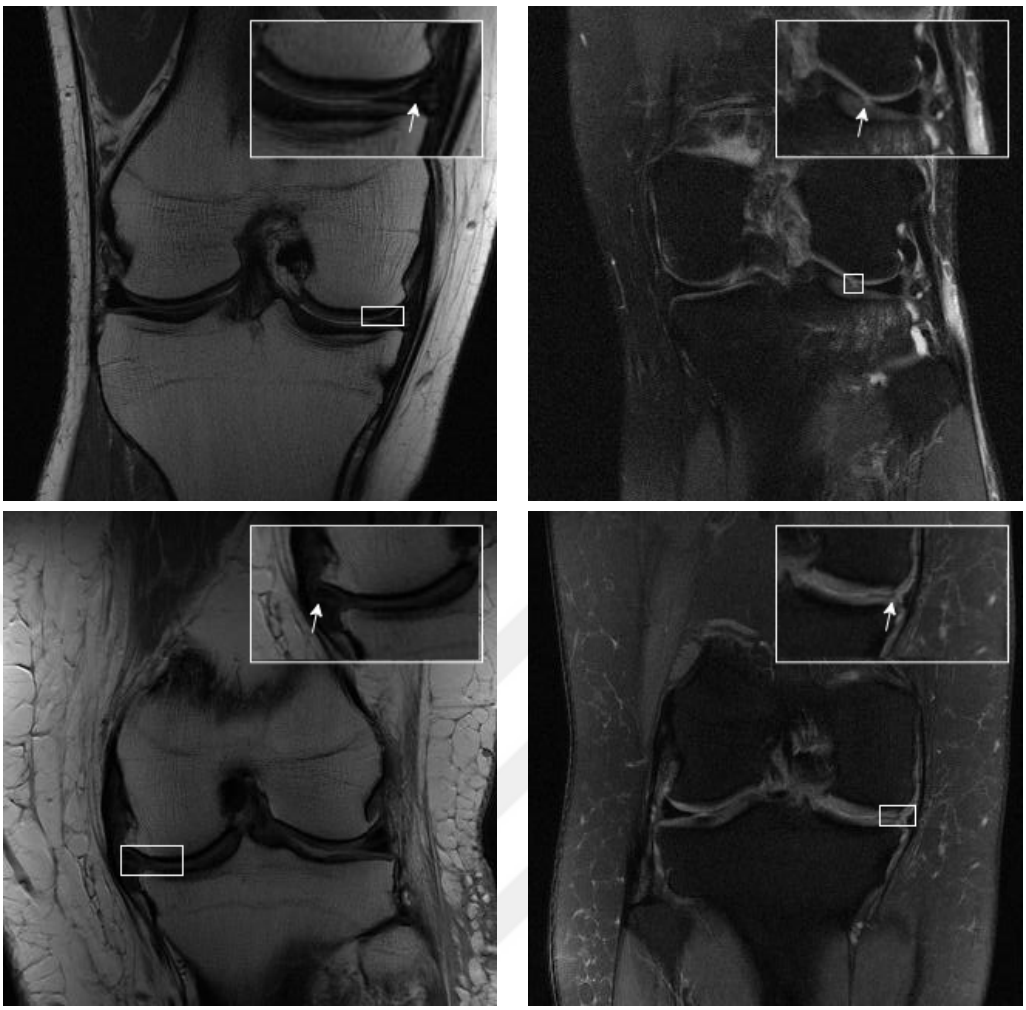


Figure 5.2 Meniscus Tear Bounding Boxes in FastMRI+ Knee Dataset. The Tear Area is highlighted in zoomed-in views with indicators using white arrows. The first and second rows showcase examples from 1.5 and 3 Tesla examinations, while the first and second columns display PD and PDFS Contrasts, respectively.

architectures, fundamental principles, and specific attributes contributing to their efficacy, a comprehensive explanation of these models (UNET, iRIM, and E2E VarNET) is provided in Section 4.3.2 of Chapter 4. Through the implementation of this multi-model strategy, our objective extends beyond merely understanding how each model uniquely contributes to overall success. We also aim to investigate how these models perform under radiological evaluation, shedding light on their effectiveness in capturing clinically relevant features in accelerated MRI reconstructions. The overarching goal of this thorough analysis is to bridge the gap between technical intricacies and practical clinical applicability, ensuring a

comprehensive assessment of the proposed multi-model reconstruction framework.

5.3.3 Detection Model

In our study on detecting meniscal tears, we use nnDetection (Baumgartner et al., 2021), an advanced and automated self-configuration method. This method, which relies on the established Retina UNET architecture (Jaeger et al., 2020), stands out for its ability to autonomously identify meniscal tears without needing additional computed sources beyond standard training data.

In Section 4.3.3, we thoroughly explain how nnDetection systematically sets up the entire detection process with a structured and adaptive approach. This method, similar to nnUNet (Isensee et al., 2021), organizes domain knowledge into fixed, rule-based, and empirical parameters, showcasing its effectiveness in handling the specific challenges of meniscal tear detection. We also highlight nnDetection's outstanding performance in various medical object detection tasks on public datasets like LUNA16 (Setio et al., 2017), ADAM (Timmins et al., 2020), and 11 other tasks, such as Decathlon (Simpson et al., 2019)-Pancreas, Prostate, and Hepatic Vessel. This impressive track record underscores its versatility and potential for broader applications in medical imaging, emphasizing its crucial role in our study. For detailed information, please refer to Section 4.3.3.

In the initial phase, nnDetection, an automated self-configuring model, initiates by extracting a data map and applying heuristic rules to determine rule-based parameters. After the training process, empirical parameters are determined, and the final predictions are generated by ensembling the predictions from five models obtained through cross-validation with empirically selected settings.

In conclusion, nnDetection emerges as a robust method for meniscal tear detection, offering automation, specialization, and flexibility in design. These advantages render it a valuable tool in improving the accuracy and reliability of meniscal tear diagnosis in knee MRI images. In this research, nnDetection was employed for meniscus tear

detection, trained using 560 fully sampled training volumes with 5-fold cross-validation, and tested over 103 different reconstructed outcomes.

The training process involves the application of a combination of cross-entropy (classification) and generalized intersection-over union (regression) loss functions over 60 epochs, with a batch size of 4. Notably, half of each batch must include at least one ground-truth box. Optimization is performed using stochastic gradient descent with a Nesterov momentum of 0.9. The learning rate undergoes a linear ramp-up from 1e-6 to 1e-2 over the first 4000 iterations, followed by a poly learning rate schedule until epoch 50. The final 10 epochs employ training with a cyclic learning rate oscillating between 1e-3 and 1e-6 each epoch. An overview of the sequential framework is depicted in Figure 5.3.

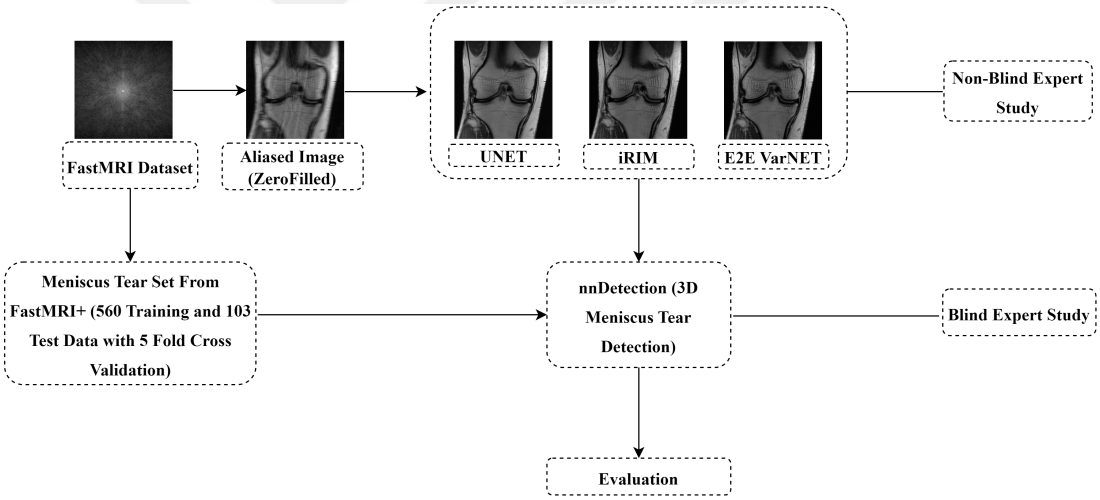


Figure 5.3 The study design encompasses the model training and testing phases, incorporating evaluation through blind/non-blind expert studies. Initially, k-space was undersampled, and zerofilled aliased images were generated. Subsequently, three distinct reconstruction models were employed to obtain high-quality images. For the assessment of meniscus tear, non-blind expert studies and analyses, involving diagnosis and pathological evaluation using deep learning processes, were conducted independently. Data annotated as "Meniscus Tear" were collected and prepared for our nnDetection model. The model underwent training with this data using 5-fold cross-validation and was tested using the three reconstruction outcomes. The reconstruction outcomes were also assessed with blind expert studies and annotated based on the existence of meniscus tear. Finally, we compared the quantitative deep learning detection results with blind expert study findings and reconstruction outcomes, taking into account the influence of non-blind expert assessments on the detection process.

5.4 Evaluation Metrics

Exploring the metrics employed to assess the performance of our proposed multi-model accelerated MRI reconstruction framework and meniscal tear detection method, this section is divided into two main subsections: Reconstruction Metrics and Detection Metrics. Each subsection presents unique evaluation criteria and objectives for a comprehensive assessment.

Reconstruction Metrics focus on evaluating the quality of accelerated MRI reconstruction, using specific metrics such as PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) to provide insights into the fidelity and visual quality of the reconstructed images.

Detection Metrics play a crucial role in evaluating the effectiveness of our meniscal tear detection method. These metrics, including Precision, Recall, and mAP (mean Average Precision), quantify the precision, recall, and overall performance of the detection algorithm.

5.4.1 Reconstruction Metrics

In the assessment of our accelerated MRI reconstruction framework, two widely recognized metrics were utilized for performance evaluation. The first, Peak Signal-to-Noise Ratio (PSNR), quantifies local differences between reconstructed and ground truth images. This is elaborated in Section 3.5, where its mathematical expression is denoted by Equation 3.2. The second metric employed is the Structural Similarity Index (SSIM), which gauges image similarity by considering interdependencies between neighboring pixels (Zbontar et al., 2018). Section 3.5 provides insights into SSIM, and its mathematical representation is presented in Equation 3.3.

5.4.2 Detection Metrics

To accurately evaluate the object detector, key metrics such as True Positive (TP), False Positive (FP), False Negative (FN) are essential. These metrics are determined using the Intersection over Union (IOU) threshold (α), which classifies detections based on the overlap between predicted and actual objects.

The IOU metric calculates this overlap by dividing the area of intersection by the area of union between ground truth (gt) and prediction (pd). According to this threshold, performance metrics, including Precision, Recall, and mean Average Precision (mAP), are derived from TP, FP, and FN, as detailed in Section 4.4.2. These metrics play a critical role in assessing the effectiveness of deep learning detectors in identifying meniscal tears, significantly contributing to their clinical relevance evaluation.

5.5 Results and Interpretations

In our pursuit to advance accelerated MRI reconstruction and bolster meniscal tear detection, our study unfolds through two extensive experiments. The initial phase focuses on a comparative analysis of diverse reconstruction models, aiming to discern their nuanced performance under standardized conditions. Subsequently, attention shifts to the second experiment, where these models undergo scrutiny by a dedicated detector to evaluate their efficacy in detecting clinically significant outcomes.

This section encompasses the outcomes of these dual investigations, providing a detailed exposition of the findings derived from our systematic evaluation pipeline. It is crucial to note that the reconstruction results underwent a non-blind assessment by expert radiologists, offering insights into their performance in a controlled setting. To ensure a comprehensive evaluation, the reconstruction results were also subjected to a blind assessment by radiologists, facilitating a comparison with the detector outcomes.

The non-blind evaluation by expert radiologists offers valuable insights into the

intrinsic qualities of various reconstruction models. Quantitative benchmarks, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), provide a detailed understanding of local differences and overall structural similarity between the reconstructed and ground truth images. Moreover, this evaluation incorporates diagnostic and pathological assessments, introducing a clinical perspective to the analysis. The examination also provides crucial information about artifact identification and the potential presence of hallucinations in the reconstructed images, enhancing the integrated evaluation of the models.

In a blind evaluation, radiologists assessed the reconstructed images without knowledge of the specific models employed. This approach ensures an unbiased judgment of the models' performance, reflecting their real-world applicability. The blind assessment serves as a crucial benchmark to gauge how well the reconstruction models perform when subjected to rigorous clinical scrutiny.

Following the reconstruction assessments, the study delves into the meniscal tear detection results. Intersection over Union (IOU), Precision, Recall, and mean Average Precision (mAP) metrics serve as key indicators, allowing for a comprehensive comparison of the detection models' performance. The outcomes are not only measured quantitatively but are also interpreted through the lens of radiological expertise.

In the subsequent sections, we unravel the intricacies of each experiment, offering a nuanced interpretation of the results obtained from both the non-blind and blind assessments. Through this comprehensive analysis, we aim to provide a thorough understanding of the strengths and potential limitations of our proposed multi-model accelerated MRI reconstruction framework and meniscal tear detection methodology.

5.5.1 Quantitative Reconstruction Performances

The quantitative evaluation process primarily targets the mitigation of background noise, a crucial factor influencing SSIM and PSNR scores. Deep learning models

have showcased significant advancements in noise reduction, leading to notable improvements in PSNR and SSIM scores. Among the implemented reconstruction models, E2E VarNET demonstrated superior performance compared to the UNET and iRIM models. Specifically, E2E VarNET achieved SSIM scores of 0.94 and 0.91, along with PSNR scores of 40.29 and 37.94 for all contrasts (PD+PDFS) at 4x and 8x acceleration rates, respectively.

For the same acceleration rates and contrasts, iRIM and UNET models yielded SSIM scores of 0.92, 0.89, and 0.91, 0.88, as well as PSNR scores of 39.08, 36.26, and 37.65, 35.11 dB, respectively. A comprehensive overview of the quantitative reconstruction evaluation results is provided in Table 5.2.

5.5.2 Quantitative Detection Performances

To evaluate the detection performance, we employed the nnDetection self-configuring framework, fine-tuned specifically for the task of 3D meniscus tear detection. The reconstruction results are presented in Table 5.3.

Remarkably, the E2E VarNET model exhibited superior performance compared to other models, achieving impressive mean average precision (mAP), Precision, and Recall scores. At 4x acceleration, the E2E VarNET model demonstrated mAP of 0.69, Precision of 0.88, and Recall of 0.58. Similarly, at 8x acceleration, the model showcased excellent performance with mAP of 0.67, Precision of 0.87, and Recall of 0.57.

Comparatively, the UNET model achieved competitive average precision values of 0.66, 0.82, 0.48 at 4x, and 0.64, 0.80, 0.49 at 8x accelerations, respectively. On the other hand, the iRIM model exhibited precision values of 0.67, 0.87, 0.52 at 4x, and 0.65, 0.80, 0.50 at 8x fold accelerated scans. These findings underscore the stability and effectiveness of our proposed pipeline.

The notable superiority of the E2E VarNET model in accurately identifying meniscus tears is evident from the achieved metrics. These results reinforce the

Table 5.2 Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) scores, along with their corresponding standard deviations, are presented for the baseline reconstructions of the fastMRI dataset. The evaluations cover All, PD, and PDFS acquisitions accelerated at 4x and 8x.

Modality	Acceleration	Acquisition	PSNR (dB)	SSIM
ZeroFilled	**4x**	ALL	31.73(±3.75)	0.83(±0.08)
		PD	32.08(±3.30)	0.85(±0.06)
		PDFS	31.35(±4.06)	0.81(±0.09)
	8x	ALL	28.01(±3.49)	0.73(±0.10)
		PD	27.54(±3.13)	0.75(±0.07)
		PDFS	28.51(±3.60)	0.72(±0.11)
UNET	**4x**	ALL	37.65(±3.72)	0.91(±0.09)
		PD	38.4(±3.64)	0.94(±0.04)
		PDFS	36.85(±3.13)	0.88(±0.10)
	8x	ALL	35.11(±2.81)	0.88(±0.10)
		PD	35.28(±3.02)	0.91(±0.15)
		PDFS	34.93(±2.54)	0.85(±0.11)
iRIM	**4x**	ALL	39.08(±4.90)	0.92(±0.10)
		PD	40.05(±4.15)	0.95(±0.04)
		PDFS	37.6(±3.68)	0.89(±0.10)
	8x	ALL	36.26(±3.84)	0.89(±0.11)
		PD	37.03(±3.84)	0.92(±0.05)
		PDFS	35.45(±3.12)	0.86(±0.11)
E2E Variational Network	**4x**	ALL	**40.29(±6.42)**	**0.94(±0.10)**
		PD	**42.45(±4.88)**	**0.97(±0.04)**
		PDFS	**38.0 (±4.37)**	**0.90(±0.10)**
	8x	ALL	**37.94(±4.40)**	**0.91(±0.11)**
		PD	**39.19(±3.75)**	**0.95(±0.04)**
		PDFS	**36.61(±3.39)**	**0.88(±0.11)**

exceptional capability of our proposed framework for 3D meniscus tear detection.

Table 5.3 Detection Performances: AP@0.1, Max. Det. 100, Precision, and Recall results for All (ALL), Proton Density (PD), and Proton Density with Fat Suppression (PDFS) accelerated at 4x and 8x.

Modality	Acceleration	Acquisition	mAP	Precision	Recall
UNET	**4x**	ALL	0.66	0.82	0.48
		PD	0.68	0.79	0.48
		PDFS	0.65	0.84	0.48
	8x	ALL	0.64	0.80	0.49
		PD	0.68	0.83	0.51
		PDFS	0.61	0.77	0.48
iRIM	**4x**	ALL	0.67	0.87	0.52
		PD	0.66	0.88	0.53
		PDFS	0.68	0.87	0.52
	8x	ALL	0.65	0.80	0.50
		PD	0.64	0.80	0.52
		PDFS	0.66	0.82	0.48
E2E Variational Network	**4x**	ALL	**0.69**	**0.87**	**0.56**
		PD	**0.7**	**0.88**	**0.58**
		PDFS	**0.68**	**0.85**	**0.53**
	8x	ALL	**0.67**	**0.87**	**0.57**
		PD	**0.69**	**0.90**	**0.58**
		PDFS	**0.66**	**0.83**	**0.56**

5.5.3 Expert Study

In addition to quantitative assessments, the detection methodology underwent evaluations by experienced experts in the field. Radiologists were invited to participate in two distinct expert studies: a Blind Expert Study and a Non-Blind Expert Study.

5.5.3.1 Non-blind Expert Study

In the non-blind expert study, participating radiologists were provided with contextual information regarding the acceleration levels and reconstruction techniques, aiming to assess the impact of such information on the experts' diagnostic evaluations. During both expert studies, participants critically examined the reconstructed images, emphasizing the clarity of meniscus structures, the identification of abnormalities, and overall diagnostic utility. Their feedback and observations offer valuable qualitative insights into the reliability and practical applicability of the proposed methodology in clinical settings.

As an integral part of our study's framework, we conducted an in-depth examination of image quality in the non-blind expert study section. This thorough assessment involved a meticulous scrutiny of each dataset, emphasizing a detailed analysis of individual slices. Operating under a non-blinded condition, radiologists focused exclusively on evaluating the reconstruction techniques, having access to contextual information such as acceleration levels and reconstruction methods.

For the evaluation of reconstructed images, a structured 5-point scoring system was employed, ranging from "1" to "5." A score of "1" denoted a significant deterioration compared to the reference image, while a score of "5" indicated a substantial enhancement surpassing the reference image's quality. Both the original reference images and their reconstructed counterparts underwent this comprehensive evaluation, with evaluators assigning standardized quality scores using the 5-point ordinal scale.

This rigorous evaluation process facilitated a comprehensive understanding of the impact and efficacy of the employed reconstruction methods. The averaged results, as depicted in Figure 5.4, highlighted E2E VarNET's superior performance at both acceleration rates, followed by iRIM and UNET. Notably, no visual received a perfect '5' rating in the evaluation, underscoring the strictness of the quality assessment. These findings from the quality assessment align with the image quality metrics

(SSIM and PSNR) across all images and acceleration rates, providing further support for the superior performance of E2E VarNET in the reconstruction task compared to the other models.

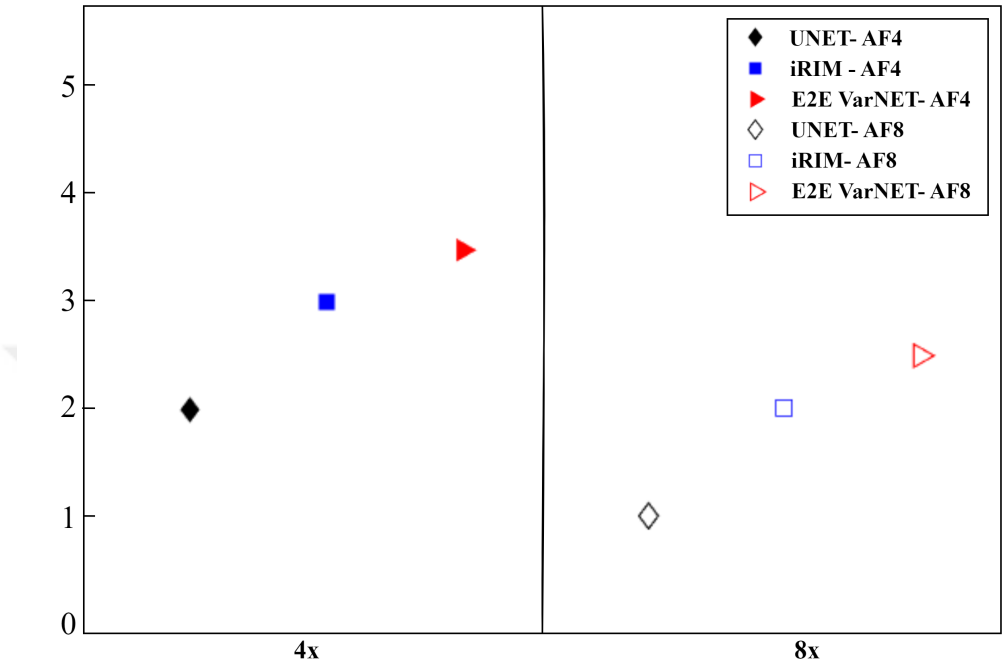
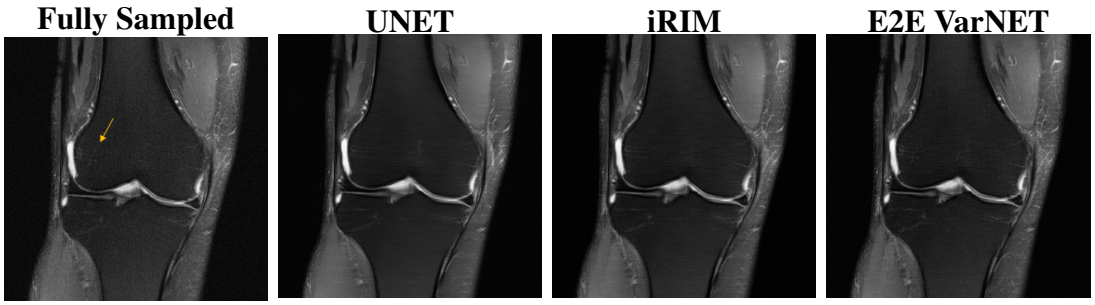


Figure 5.4 The plot depicts the average diagnostic scores assigned by radiologists during assessments of the reconstructed images. The scoring system is condensed as follows: 1 for significantly worse, 2 for worse, 3 for similar, 4 for better, and 5 for significantly better compared to the reference image. The symbols on the plot indicate the computed average of expert scores.

In some instances, experts pointed out a noteworthy improvement in certain aspects of the images reconstructed by the Deep Learning (DL) models compared to the reference images. This concept is visually exemplified in Figure 5.5, which was specifically highlighted by the radiologists during their assessments.

To assess the pathological features associated with meniscal tears across all volumes, we employed a 0-4 ordinal scoring system for all slices. Within this scoring system, a score of "0" signified that the meniscus tear was not observable, while "1" and "2" indicated conditions that were far worse and worse compared to the reference, respectively. A score of "3" suggested similarity to the reference, and "4" indicated a better condition than the reference.



(a) (b) (c) (d)

Figure 5.5 Expert evaluations of 4x accelerated image reconstruction by DL methods. The radiologist assigned scores: 2 for UNET (b), 3 for iRIM (c), and 4 for E2E VarNET (d). Notably, in the case of E2E VarNET, the radiologist emphasized higher contrast in the femur medial condyle (yellow arrowed) compared to the reference image. Furthermore, background noise was effectively suppressed, indicating a significant enhancement in image quality beyond that of the reference image. These findings suggest that the DL model not only preserved but also improved certain aspects of the image, particularly in terms of contrast and noise reduction.

Figure 5.6 presents the mean overall pathological assessment scores for meniscal tears. The results indicate that all Deep Learning (DL) models produced similar scoring outcomes, with E2E VarNET exhibiting a slight advantage over iRIM and UNET at both acceleration rates. These averaged expert pathological scores align well with the performance metrics of the nnDetection detector, as evident from both the mean Average Precision (mAP) and Precision and Recall scores. The expert evaluations of meniscal tear pathology consistently correlate with the results obtained from the DL models, affirming the reliability of the detection outcomes.

Radiologists carefully examined the artifacts and hallucinatory effects introduced by the reconstruction models in the non-blind expert study. Some models transformed specific structures into abnormalities, raising concerns about their potential impact on clinical evaluations. The presence of artifacts and hallucinations negatively affected expert opinions, especially regarding their suitability for clinical applications. Identified artifacts extended across anatomical areas like the femur, tibia, fibula, and other regions. These hallucinations often mimicked normal structures that either do not exist or falsely indicated abnormalities, presenting significant challenges for accurate diagnosis. Additionally, some hallucinations lacked realism, adding complexity to the problem.

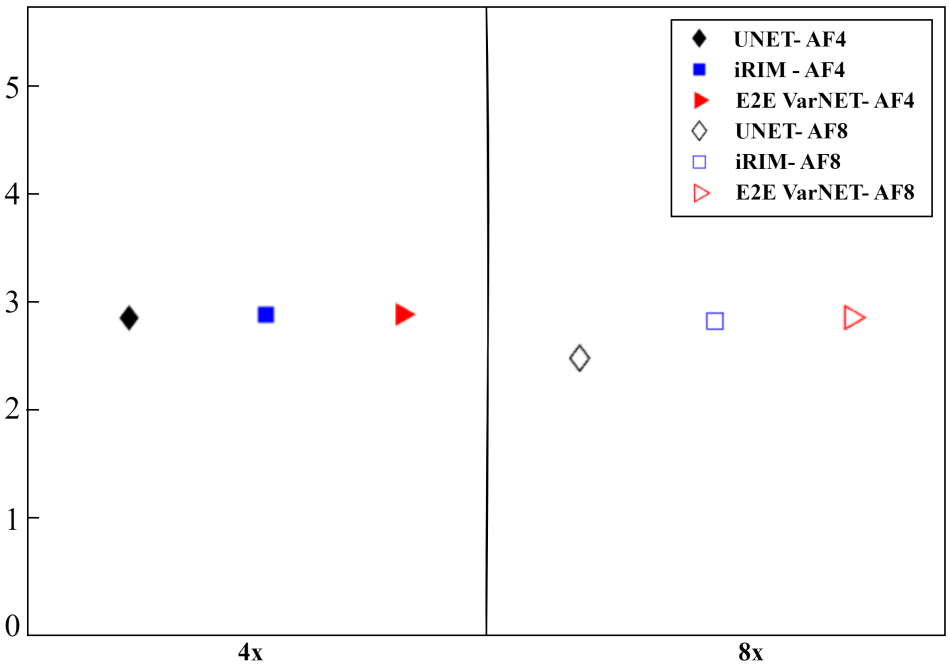


Figure 5.6 The plot showcases the average scores derived from radiological assessments of pathological features in the reconstructed images. The scoring system employed is as follows: 0 indicates that the pathology is not observable, 1 for far worse, 2 for worse, 3 for similar, and 4 for better compared to the reference image.

Figure 5.7 visually illustrates scattered linear artifacts surrounding the femur and tibia (h), with some appearing in a fogging-style (f), resulting in increased pseudo (d) and linear signal artifacts in the meniscus. The scattered linear artifacts around the femur and tibia could potentially mask or obscure background anatomy, including primary pathology. Additionally, these artifacts might mimic an old non-displaced fracture line, posing challenges for less-experienced radiologists. The fogging artifact, although not inherently clinically problematic, has the potential to obscure normal anatomy, and if it occurs in the exact area with the primary pathology, it could complicate the diagnostic process.

In some cases, a high-intensity artifact with unclear borders (Figure 5.7b) may resemble bone marrow edema. This poses challenges, especially in patients evaluated for trauma-induced meniscal tears or ligament injuries. A misleading pseudo-increased meniscal tear image could lead to a misdiagnosis, suggesting a longer or more extensive tear than actually present.

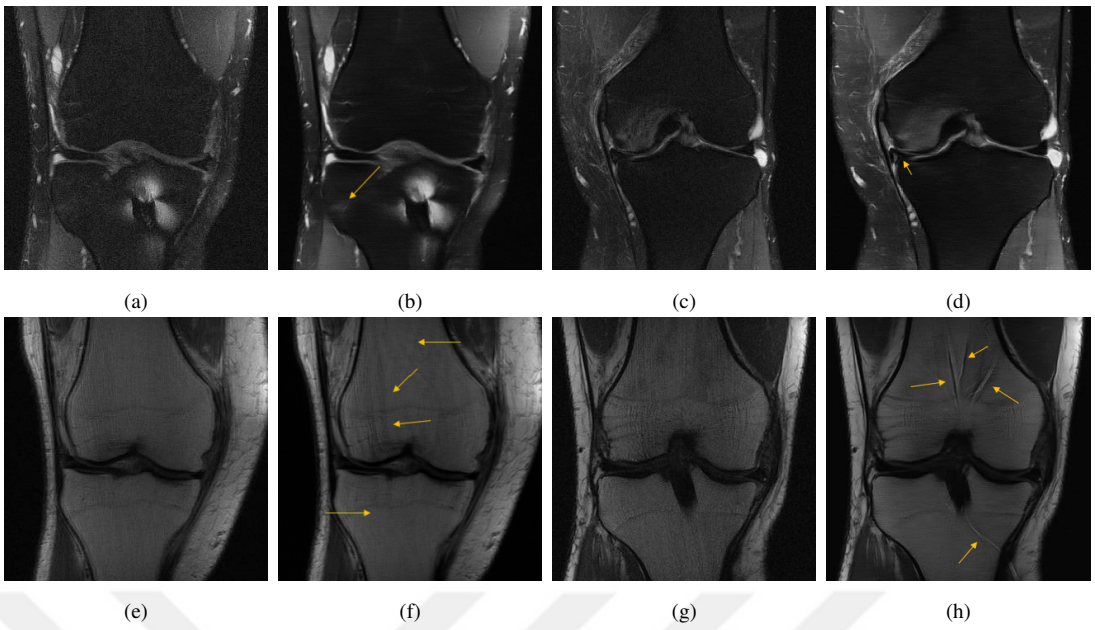


Figure 5.7 Hallucinations in DL-Based Reconstruction Outcomes: (a) Reference fully sampled image; (b) Arrows indicate DL-generated "Bone Marrow Edema" not present in (a). (c) Reference fully sampled image; (d) DL-generated increased pseudo and linear signals in the meniscus, posing misdiagnosis risks. (e) Fully sampled image; (f) DL-generated fogging-style artifacts. (g) Fully sampled image; (h) DL-generated lineal artifacts surrounding femur distal and tibia.

5.5.3.2 Blind Expert Study

During the blind expert study, radiologist conducted assessments in a blinded manner without prior knowledge of the acceleration levels or reconstruction methods, exclusively annotating meniscus tears using the MD.ai annotation platform. The annotation process included all meniscus tear and other annotated volumes in FastMRI+, categorized according to contrasts equally. The expert was blinded to the reconstruction outcomes of 20 PD + 20 PDFS for UNET, iRIM, and E2E VarNET, which were divided equally based on 4x and 8x acceleration rates.

The annotated meniscus tears from this study were then compared with annotations available in the FastMRI+ dataset and the results obtained through the nnDetection framework.

The comparison with FastMRI+ annotations aimed to evaluate the concordance

between expert annotations and those present in the publicly available dataset. Any discrepancies, areas of agreement, or unique observations made by the expert radiologist were carefully examined.

Furthermore, the results of the blind expert study were contrasted with the outcomes of the nnDetection framework. This analysis focused on assessing the performance of automated detection in comparison to the expert annotations. Sensitivity, precision, and overall agreement were scrutinized to gain insights into the reliability of the proposed methodology.

In the comprehensive examination of blind expert study results, this study provides valuable insights into the alignment of expert annotations with both FastMRI+ annotations and the automated detection outcomes generated by the nnDetection framework.

The blind expert annotations are meticulously evaluated against the FastMRI+ Reference, as depicted in Table 5.4. Noteworthy results emerge, especially at 4x acceleration, where E2E VarNET showcases superior performance compared to other models. It achieves remarkable metrics, including a mean Average Precision (mAP) of 0.60, Precision of 0.98, and Recall of 0.31. Similarly, at 8x acceleration, iRIM demonstrates noteworthy performance with an average precision of 0.44, Precision of 0.92, and Recall of 0.23. These findings underscore the efficacy of the models in clinically significant meniscal tear detection.

Adding nuance to the evaluation, in the PDFS contrast at 8x acceleration, E2E VarNET stands out with higher performance when compared to iRIM. This nuanced observation contributes to a comprehensive understanding of the models' effectiveness across different acceleration rates and contrasts in meniscal tear detection.

The selection of reconstruction results for blind expert annotation is not specific; it is only considered for the equal distribution of contrasts. Additionally, FastMRI+ annotations are significantly smaller in comparison with expert annotations. These

differences can potentially contribute to variations, especially at an 8-fold acceleration rate, when compared with nnDetection results.

Table 5.4 Performance of Blind Expert Annotations with FastMRI+ Reference: Mean Average Precision (mAP), Precision, and Recall Results for Accelerated All (ALL), Proton Density (PD), and Proton Density with Fat Suppression (PDFS) at 4x and 8x.

Modality	Acceleration	Acquisition	mAP	Precision	Recall
UNET	**4x**	ALL	0.29	0.85	0.16
		PD	0.32	0.81	0.19
		PDFS	0.29	1	0.14
	8x	ALL	0.31	0.98	0.16
		PD	0.13	1	0.07
		PDFS	0.51	0.97	0.26
iRIM	**4x**	ALL	0.40	0.79	0.23
		PD	0.40	0.91	0.21
		PDFS	0.44	0.7	0.26
	8x	ALL	**0.44**	**0.92**	**0.23**
		PD	**0.49**	**0.90**	**0.25**
		PDFS	0.41	0.90	0.22
E2E Variational Network	**4x**	ALL	**0.60**	**0.98**	**0.31**
		PD	**0.62**	**0.98**	**0.31**
		PDFS	**0.60**	**0.99**	**0.30**
	8x	ALL	0.32	0.75	0.20
		PD	0.15	0.68	0.09
		PDFS	**0.46**	**0.81**	**0.27**

At 8-fold acceleration, both iRIM and E2E VarNET have the same volume counts annotated, but the annotated slice numbers make the difference, despite the higher label count in E2E VarNET. It is important to note that such variations may arise due to differences in dataset sizes and the distribution of annotations.

However, in all quantitative detection metrics, nnDetection results consistently show higher scores compared to blind expert results at both acceleration rates. This

observation suggests that the automated detection method outperforms human annotations in terms of precision, recall, and mean average precision, providing stable and reliable results even in challenging scenarios, such as higher acceleration rates and variations in dataset sizes.

Additionally, when using blind expert annotations as a reference, variations in the alignment with reconstruction outcomes are observed in nnDetection results. UNET exhibits a higher concordance with blind expert annotations at both acceleration rate , as shown in Table 5.5. This may be attributed to lower annotated volumes and a higher degree of annotation box size overlapping.

Table 5.5 Performance of nnDetection with Blind Expert Study Annotations as Reference: Mean Average Precision (mAP), Precision, and Recall Results for Accelerated All (ALL), Proton Density (PD), and Proton Density with Fat Suppression (PDFS) at 4x and 8x.

Modality	Acceleration	Acquisition	mAP	Precision	Recall
UNET	**4x**	ALL	**0.91**	**1**	**0.45**
		PD	**0.95**	**1**	**0.48**
		PDFS	**0.83**	**1**	**0.42**
	8x	ALL	**0.93**	**1**	**0.47**
		PD	**0.8**	**1**	**0.4**
		PDFS	**0.97**	**1**	**0.49**
iRIM	**4x**	ALL	0.59	0.91	0.31
		PD	0.37	0.71	0.23
		PDFS	0.74	1	0.37
	8x	ALL	0.64	0.88	0.35
		PD	0.57	1	0.28
		PDFS	0.74	0.83	0.41
E2E Variational Network	**4x**	ALL	0.77	0.97	0.41
		PD	0.9	0.98	0.46
		PDFS	0.68	0.96	0.36
	8x	ALL	0.48	0.79	0.29
		PD	0.49	0.87	0.32
		PDFS	0.52	0.83	0.29

5.6 Conclusion and Discussion

In this chapter, we conducted a thorough examination of MRI reconstruction and its analysis, specifically emphasizing clinically relevant details related to meniscus tears. Our research aimed to evaluate the effectiveness of deep learning (DL) detection across various DL reconstructions, which had been individually scrutinized in prior literature. Furthermore, we explored the diagnostic and pathological performance of these multi-model DL reconstructions, taking into account assessments from non-blind expert studies. These evaluations were centered exclusively on the assessment of pathology and diagnosis in the reconstructed images, conducted with awareness of the acceleration factor and reconstruction models used. Additionally, we incorporated blind expert study annotations, where experts, devoid of any prior knowledge about the specific models or acceleration factors, independently examined the images and annotated meniscus tears.

The results of our study highlight the significant impact of DL reconstruction performance on detection outcomes, revealing noteworthy variations in assessments conducted by DL detectors and blind/non-blind expert radiologists. Our in-depth analysis advocates for a comprehensive approach that aligns reconstruction and detection strategies with the evaluations of both blind and non-blind expert radiologists. Achieving state-of-the-art (SOTA) detection performance in identifying meniscal tears within the FastMRI+ dataset, our research makes a substantial contribution to the field. It provides valuable insights into the intricate relationship between MRI reconstruction and DL-based detection, emphasizing the effectiveness of a comprehensive strategy to enhance meniscal tear detection. These findings are in concordance with blind/non-blind expert evaluations, emphasizing the study's significance in improving clinical reliability.

In our study, we employed three distinct DL models for the MRI reconstruction task and evaluated their results using a well-established DL detector, nnDetection, on the FastMRI+ dataset to identify Meniscus Tear. The outcomes of the reconstruction from these models, showcasing high and comparable Image Quality Assessment

(IQA) scores, were utilized as input for the DL detector and assessed by non-blind radiologists. These IQA scores align with non-blind expert diagnosis and pathological assessment scores. Regarding clinical diagnostic quality assessment, the UNET model demonstrated lower performance compared to iRIM and E2E VarNET. However, when specifically assessing the presence of meniscus tears, UNET exhibited similar performance to iRIM and E2E VarNET at both acceleration rate. Furthermore, the clinical relevance and practicality of our approach were highlighted through thorough evaluations conducted by these radiologists. They also emphasized the hallucinatory effects introduced by the reconstruction models, negatively impacting expert sentiments, particularly concerning their suitability for clinical applications. The identified hallucinations and artifacts present significant challenges for future research and development. Addressing and resolving these issues will be a crucial focus to enhance the reliability and accuracy of deep learning models in medical image reconstruction and analysis.

In addition to this evaluation, the outcomes of the reconstruction from these three distinct models were utilized for the DL detector and blind expert study. The DL detector performance scores align with IQA metrics at both acceleration rates, with E2E VarNET exhibiting the highest detection metric scores, followed by iRIM and UNET at both acceleration rates, respectively. However, in the blind expert study, there have been some changes in ranking concerning the acceleration rate. In the blind assessment against the FastMRI+ annotations as reference, E2E VarNET showcases superior performance compared to other models, especially at 4x acceleration. However, at 8x acceleration, iRIM demonstrates noteworthy performance with detection metrics. This change in ranking might be attributed to the selection of reconstruction results for blind expert annotation, which is not specific and is only considered for the equal distribution of contrasts. Such variations may arise due to differences in dataset sizes and the distribution of annotations. Additionally, FastMRI+ annotations are significantly smaller compared to expert annotations, and these differences can potentially contribute to variations, especially at an 8-fold acceleration rate.

However, in all quantitative detection metrics, nnDetection results consistently show higher scores compared to blind expert results at both acceleration rates. This observation suggests that the automated detection method provides stable and reliable results even in challenging scenarios, such as higher acceleration rates and variations in dataset sizes. Furthermore, when using blind expert annotations as a reference, variations in the alignment with reconstruction outcomes are observed in nnDetection results. UNET exhibits a higher concordance with blind expert annotations at both acceleration rates, contrary to the previous assessment that considered FastMRI+ annotations. This difference may be attributed to lower annotated volumes and a higher degree of annotation box size overlapping.

Concluding the extensive investigations conducted in this study, our research has shed light on the intricate interplay between MRI reconstruction and deep learning-based detection methods for meniscus tears. Through a meticulous analysis involving diverse DL models, we have demonstrated the significant impact of reconstruction performance on detection outcomes, offering valuable insights for the medical imaging field.

Our comprehensive evaluations, encompassing blind and non-blind expert studies, along with the utilization of a well-established DL detector, nnDetection, have provided a holistic understanding of the strengths and limitations of various reconstruction models. These findings contribute to the existing body of knowledge and highlight the nuanced factors influencing the clinical reliability of meniscus tear detection.

Furthermore, the observed variations in detection performance, particularly in relation to different acceleration rates, underscore the importance of considering these factors in the deployment of deep learning models in clinical settings. The alignment of our proposed framework with expert assessments and the consistent performance of nnDetection across challenging scenarios emphasize the potential for reliable and stable results in real-world applications.

Looking ahead, our future research aims to enhance the clinical relevance of our

framework through prospective sampling strategies and a deeper exploration of parallel imaging approaches. Additionally, the incorporation of surgical background information into the dataset holds promise for further refining our model's adaptation to specific surgical requirements.

In conclusion, this study not only advances our understanding of the intricate dynamics between MRI reconstruction and deep learning-based detection but also provides a foundation for future research endeavors aimed at improving the accuracy, reliability, and clinical applicability of medical image analysis in meniscus tear detection.

CHAPTER SIX
INTEGRATED APPROACH TO MENISCAL TEAR RECONSTRUCTION AND CLASSIFICATION IN ACCELERATED MRI

6.1 Introduction

Meniscus tears, common injuries of the knee, can lead to enduring pain and cartilage degeneration, often necessitating surgical intervention. Magnetic Resonance Imaging (MRI) has emerged as a pivotal diagnostic tool, offering crucial insights before and after treatment in cases of meniscus tears. Given the wide prevalence of meniscus lesions, the demand for automated and quantitative diagnostic solutions delivering accurate and rapid results has grown significantly.

Efforts to automate meniscus tear detection have led to the development of computer-aided diagnosis systems that employ annotated coordinates and classification techniques (Lecouvet et al., 2018; Boniatis et al., 2008; Köse et al., 2009; Ramakrishna et al., 2009; Saygili & Albayrak, 2018). Recent studies have demonstrated the potential of Deep Learning (DL) in surpassing subjective assessments by radiologists in meniscus tear detection (Hung et al., 2023). Notably, DL has exhibited diagnostic performance comparable to that of musculoskeletal radiologists in evaluating various conditions, including anterior cruciate ligament tears, meniscus tears, and rotator cuff disorders (Fritz & Fritz, 2022). DL's applicability extends to other domains, encompassing cartilage lesion detection (Liu et al., 2018), detection of abnormalities and meniscal tears (Pandey et al., 2021), cartilage assessment (Siouras et al., 2022; Shakhovska & Pukach, 2022), recurrence prediction and tear classification (Shin et al., 2022), characterization of meniscal tears (Rizk et al., 2021), and broader abnormalities in knee MRI exams (Mangone et al., 2023).

However, the effectiveness of these automated systems, particularly those driven by deep learning, is contingent on the availability of large and diverse training datasets. To address this challenge, the fastMRI dataset has emerged as a valuable resource,

facilitating the acquisition of extensive MRI raw volumes for evaluation (Knoll et al., 2020). In contrast to traditional datasets, which are often limited in size and sourced from individual MRI scanners, the fastMRI dataset tackles the data scarcity inherent in MRI reconstruction research. Despite the potential of deep learning reconstruction techniques to yield quantitatively accurate outcomes, they often struggle to capture intricate details such as meniscus tears (Knoll et al., 2020).

To mitigate these limitations, the FastMRI+ initiative introduces clinically relevant annotations to the fastMRI dataset (Zhao et al., 2022). This strategic step resolves uncertainties concerning the locations of pathologies and fosters the development of deep learning approaches informed by clinical insights.

In the field of radiology, the primary focus has traditionally been on assessing the diagnostic comparability between images reconstructed through accelerated deep learning (DL) and conventional methods for evaluating internal knee derangements (Johnson et al., 2023). Additionally, researchers have delved into automated disease detection using DL-based reconstructions within the FastMRI+ dataset (Zhao et al., 2021). However, a significant research gap exists concerning the exploration of automated disease classification integrated with reconstruction models. This study aims to bridge this gap by introducing a comprehensive pipeline that involves the entirety of image reconstruction frameworks and meniscus tear classification. The thorough evaluation of this pipeline encompasses meticulous scrutiny at both 4-fold and 8-fold acceleration rates, using fully sampled scans as a benchmark for rigorous assessment. Figure 6.1 provides an illustrative overview of the entire pipeline.

This study has made three main contributions in the following key areas:

- **Diverse Evaluation of DL Models**: We applied MRI reconstruction and analysis using various DL models and regenerated results for fastMRI. Despite differences in model architecture and training, all methods consistently demonstrated closely aligned evaluation results (Putzky & Welling, 2019; Zbontar et al., 2018; Sriram et al., 2020). Here, our primary contribution is a comprehensive analysis that unveils the diverse outcomes produced by these DL

models as supported by qualitative expert analysis. These results indicate the potential to effectively preserve meniscus tears at both 4x and 8x acceleration rates.

- **Integrated Pipeline for Meniscal Tear Diagnosis**: Our second major contribution introduces an integrated pipeline encompassing both reconstruction and classification for meniscal tear diagnosis (see Figure 6.1). Building on the outcomes of the aforementioned analyses, we assessed the performance of transfer learning algorithms (including VGG-16, VGG-19, AlexNET, ResNet50, ResNet101, GoogleNET, NasNET Mobile) on retrospectively undersampled DL-reconstructed knee MRI. Through a quantitative comparison with conventional imaging for meniscus tears, we conducted a comprehensive evaluation of DL performance. Our findings demonstrate that VGG-16, when combined with the E2E Variational Network, possesses the ability to effectively classify clinically significant outcomes at both 4x and 8x acceleration rates. VGG-16 consistently achieved exceptional performance metrics, boosting Precision (0.99), Recall (0.99), Accuracy (0.99), Specificity (0.98), and F_1 Score (0.99) at these acceleration rates. Notably, qualitative analysis further highlights VGG-16's consistent accuracy, even when its primary focus, as assessed through activation maps, is directed at various image zones.

- **Analysis on Correlation between Reconstruction and Classification Performances**: The third contribution of our study involves a thorough comparative analysis of original and DL-reconstructed images in the classification task. Highly accelerated reconstructions can sometimes struggle to effectively classify clinically vital structures, despite exhibiting favorable global quality metrics like SSIM. Up to our knowledge, our study is the first to show that high SSIM scores do not necessarily translate into improved AI classifier performance in terms of True Positive (TP) or False Negative (FN) counts. This observation is validated for both acceleration rates and across all DL reconstruction models.

The remainder of this study is organized as follows: In Section 6.2, we delve into materials and methods, covering the dataset, preprocessing, reconstruction, and analysis, as well as introducing the gold-standard image quality metric. Section 6.3 presents our results, and finally, in Section 6.4, we conclude and discuss future research avenues.

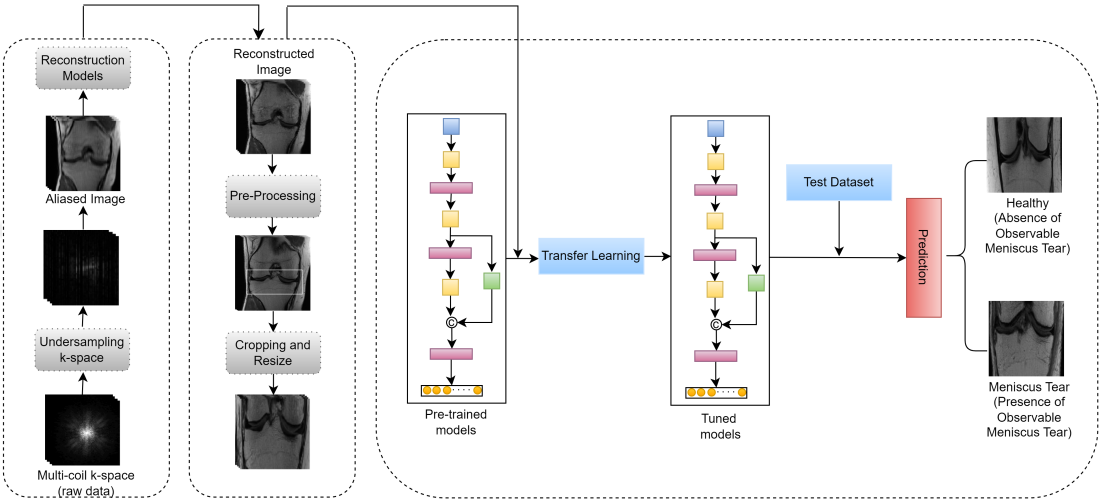


Figure 6.1 Unified multi-coil MRI reconstruction and analysis pipeline: (1) Acquired data is initially undersampled, resulting in an aliased image. (2) The aliased image is then processed through three different reconstruction models to transform it into a high-quality image. (3) These reconstructed outputs are further preprocessed through scaling, rotation, cropping, and resizing. (4) Following the preprocessing, the region of interest (in this case, the meniscus tear) is localized and trained using transfer learning models with fully sampled scans. (5) Finally, the trained models are tested with different reconstructed scans for comprehensive analysis.

6.2 Material and Methods

6.2.1 FastMRI Knee Dataset and FastMRI+ Annotations

The fastMRI dataset comprises an extensive collection of raw MRI data, with each fully sampled volume originating from either one of three clinical 3T systems or a clinical 1.5T system from the same vendor. Detailed information about acquisition, vendor details, protocols, orientation, slice, and volume distribution is available in Section 3.2. For specifics about this information and predefined undersampling

trajectories by fastMRI, please refer to Section 3.2.

In the fastMRI training and validation knee raw dataset, a total of 1,172 MRI volumes underwent initial reconstruction and subsequent annotation within the FastMRI+ clinical framework (Zhao et al., 2022). The detailed distribution of all pathologies for each set and artifacts, with a focus on meniscus tear as the most prominently annotated and classified as pathological, is explained in Sections 4.3.1 and 5.3.1, respectively. Please look for details in these sections.

The categorization and distribution of meniscus tears, as well as healthy volumes in each dataset, are presented in Table 6.1. For healthy volumes, the total number of slices was calculated by considering the slices with the highest information content, discarding the first and last 8 slices. Figure 6.2 presents a visual compilation of selected scans, organized based on their magnetic strength and contrast characteristics. This compilation includes images of both healthy instances and cases with meniscus tears. To enhance clarity, meniscus tears are indicated using white arrows.

Table 6.1 Summary of Meniscus Tear Labels and Healthy Subjects with Subset Divisions

Label	**Annotation Count**	**Slice Count**		**Subject Count**	
		Training	**Validation**	**Training**	**Validation**
Meniscus Tear	**5658**	**4158**	**700**	**560**	**103**
Healthy	**–**	**853**	**274**	**146**	**44**

* For healthy subjects, the total number of slices was calculated by considering the slices with the highest information content.

6.2.2 Preprocessing of the Images

In our methodology, preprocessing of the coronal fastMRI scans was a crucial initial step. The aim of this preprocessing was to standardize the images and prepare them for further analysis, particularly for accurate meniscus tear detection. First, we addressed the orientation of the scans. To ensure consistency, we flipped and arranged the scans based on the presence of meniscus tears. This step ensured that the femur consistently occupied the upper portion of the images, while the tibia was consistently



Figure 6.2 Illustrations of MRI scans depicting both healthy and meniscus tear cases. The top row showcases scans acquired with a 1.5T scanner, while the bottom row displays scans obtained using a 3T scanner. The first two columns represent healthy scans, and the subsequent two columns depict meniscus tear cases. In the first and third columns, Proton-Density (PD) contrast is utilized, while the second and fourth columns employ Proton-Density with Fat Suppression (PDFS) contrast. The presence of a meniscus tear is indicated by a white arrow.

positioned at the lower part, providing a standardized starting point for our analysis. Next, we standardized the dimensions of the image slices, resizing them to 320x320 pixels. Additionally, we adjusted the grayscale intensity range to a uniform scale of 0 to 255. However, it is important to note that achieving complete uniformity in contrast and brightness across all images proved challenging due to inherent variability in the dataset.

This preprocessing phase laid the foundation for our subsequent analysis, where our primary goal was the accurate identification of meniscus tears in MRI images. Meniscus tears typically manifest as thin, intense lines crossing the menisci, and they can exhibit various patterns that represent different types of tears and their positions within the menisci. These patterns were visually examined, as illustrated in Figure 6.3, to understand how tears may appear.

As part of our analysis, we utilized the top row of images to provide an overall view, while the bottom row allowed for a closer examination of tear characteristics. Tears were marked with white arrows for clarity. Additionally, the last two rows

displayed cropped views of the corresponding scans, reconstructed at 4-fold and 8-fold accelerations, enabling a more detailed assessment.

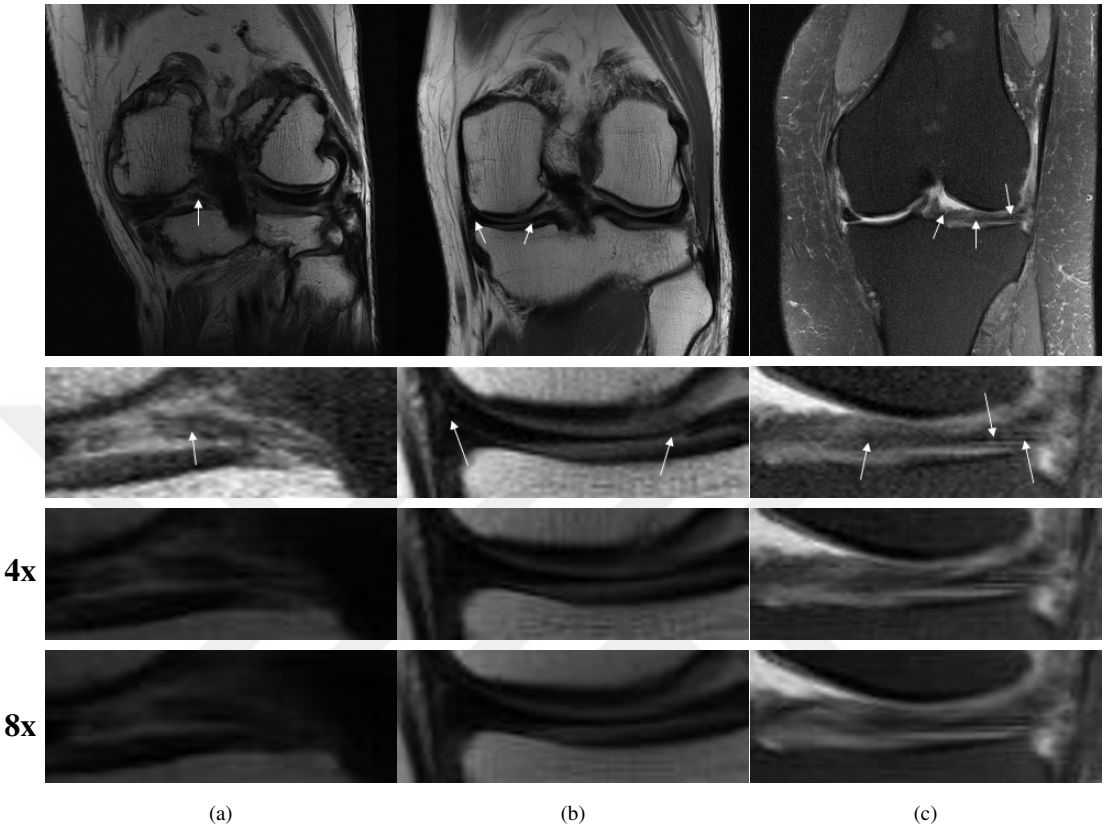
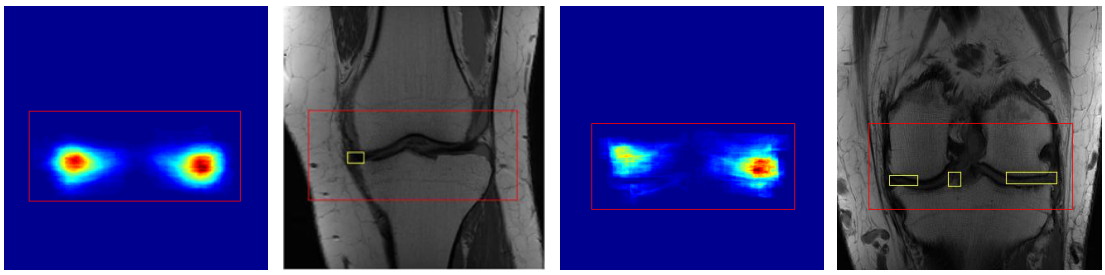


Figure 6.3 Different types and orientations of meniscus tears within the menisci. The top row displays fully sampled scans, and the rows just below illustrate specific tear instances with corresponding cropped views: (a) Bucket-handle meniscus tear shown in a cropped view. (b) Double tears on the left side observed in the cropped view. (c) Multiple tears showcased in the cropped view. The last two rows represent cropped scans of the corresponding slice reconstructed at 4-fold and 8-fold accelerations. Each tear is marked with a white arrow.

Despite the ability to identify meniscus tears within annotated boxes, challenges arose due to the limited resolution of these small regions. Precise tear detection and examination were hindered by the confined size of menisci within the knee joint and the nuanced appearance of tears. To address these challenges and enhance our classification process, we incorporated a preprocessing step involving image cropping based on the region of interest, which included annotated bounding boxes, as demonstrated in Figure 6.4. This practice allowed us to focus on relevant areas of interest, determine cropped coordinates, and avoid the inclusion of redundant regions.



(a) (b) (c) (d)

Figure 6.4 Visual representation of cropped scan coordinates (highlighted in red) utilized as a frame of reference. Both the training and validation datasets are cropped according to the designated red box coordinates, while yellow boxes indicate FastMRI+ annotations. (a) Heatmap presenting the training dataset with the selected red box coordinate, encompassing all bounding boxes. (b) An example scan extracted from the training set. (c) Heatmap showcasing the validation dataset, also including all bounding boxes in the validation set. (d) An illustrative scan from the validation set.

Our dataset was initially organized according to radiologists' evaluation criteria, and subsequent preprocessing steps, including cropping and resizing, were performed as detailed in (Fritz et al., 2020). The dataset was divided into a multicoil training dataset for model training and a validation dataset for testing purposes. To comprehensively evaluate the impact of accelerated k-space reconstruction on classification, we integrated 4-8 fold accelerated k-space data and their corresponding reconstructions. This comprehensive evaluation was conducted using a 5-fold cross-validation approach, as depicted in Figure 6.5.

After training AlexNET with cropped images, the network exhibited superior performance by exclusively focusing on meniscus tear features, even when they were rarely visible, as demonstrated in Figure 6.6 (d).

6.2.3 MRI Reconstruction

The initial focus of the fastMRI dataset was to address the reconstruction challenge, specifically emphasizing the reconstruction of clinically significant structures, including pathologies, at high acceleration factors (4 to 8-fold) while maintaining robust global quantitative metrics. Despite favorable outcomes, deep learning approaches encounter difficulties in effectively reconstructing intricate

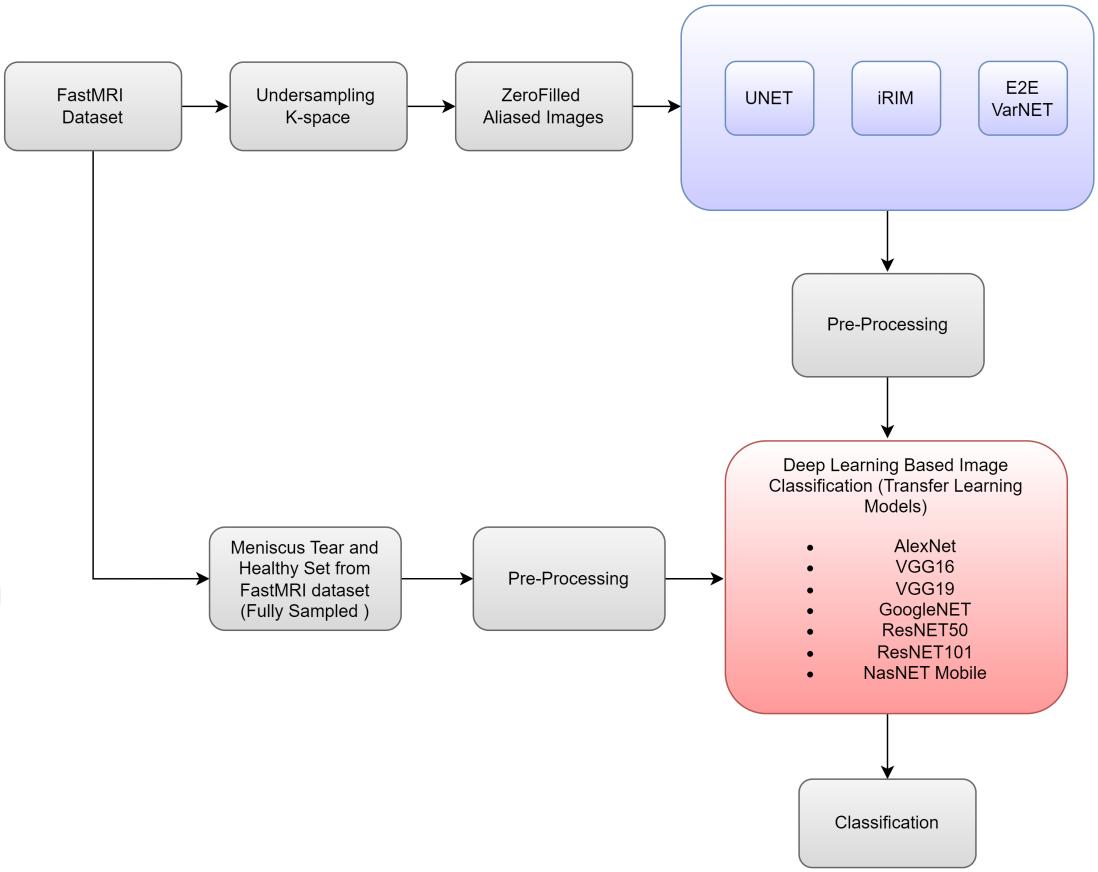
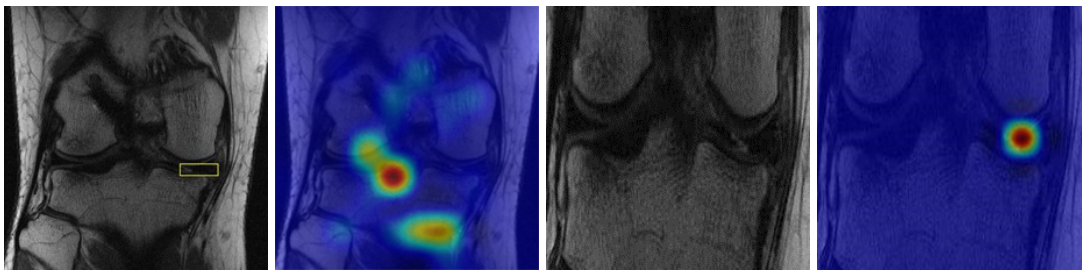


Figure 6.5 Study design encompassing model training and testing phases. (1) Undersampling of k-space and generation of zerofilled aliased images. (2) Utilization of three reconstruction models for high-quality image generation. (3) Collection and preprocessing of "Meniscus Tear" and "Healthy" data for transfer learning. (4) Training of models and testing with different reconstruction outcomes. (5) Classification results obtained.

features, such as meniscus tears. To address this challenge, our study proposes a sequential pipeline consisting of two main stages: image reconstruction at 4-fold and 8-fold accelerations, followed by the classification of meniscus tears based on these reconstructed results.

Within the domain of deep learning techniques, we employ three distinct models: UNET, Invertible Recurrent Inference Machine (iRIM), and End-to-End Variational Networks (E2E VarNET), each elucidated in Section 4.3.2 of Chapter 4. The UNET model (Zbontar et al., 2018) is notable for its widespread adoption in image reconstruction tasks, known for its exceptional performance in feature extraction and utilization of skip connections, making it a common choice in fastMRI research. The



(a) (b) (c) (d)

Figure 6.6 An illustration of a validation dataset image used for testing (a), where a previously unseen meniscus tear was accurately classified by AlexNET. Additionally, (b) presents the same image with a superimposed saliency mask, highlighting the network's focus on non-relevant areas, often overlooking torn menisci. Moreover, (c) displays the identical image cropped based on designated annotated coordinates, showcasing the targeted region of interest, while (d) shows the corresponding cropped image with an overlay of the saliency map, indicating areas of significance detected by the model.

Invertible Recurrent Inference Machine (iRIM) (Putzky & Welling, 2019) addresses memory challenges in learning by employing invertible networks, ensuring constant memory usage during training, making it particularly effective for deep models. Conversely, the End-to-End Variational Network (E2E VarNET) (Sriram et al., 2020; Hammernik et al., 2018) excels in achieving learning by incorporating sensitivity maps.

In summary, our study adopts a two-step approach in MRI reconstruction, involving the reconstruction of undersampled k-space data followed by the classification of resulting reconstructed images. This methodological strategy forms the basis of our comprehensive integrated pipeline, addressing the challenges posed in reconstructing intricate features within the fastMRI dataset.

6.2.4 Transfer Learning Methodologies for Meniscus Tear Detection

In the context of meniscus tear detection, we embarked on a comprehensive evaluation of various transfer learning techniques. These methodologies were devised to harness insights derived from natural image datasets, addressing the constraints posed by the relatively limited availability of clinical data (Sarkar et al., 2018). Our exploration revealed that these methodologies can be broadly classified into two main

categories: 'Feature Extractor' and 'Fine-tuning.'

The allocation of data for both the training and testing phases is delineated in Table 6.1. It is noteworthy that the validation set, alongside different reconstruction outcomes, played a pivotal role in shaping our testing framework. During the training phase, we utilized fully sampled scans, while testing encompassed a diverse range of reconstruction outcomes achieved at high acceleration rates. This comprehensive testing approach facilitated a robust assessment of the models, enabling a thorough evaluation of their performance on previously unseen data across a spectrum of reconstruction scenarios.

Now, let's delve into the two main strategies we explored in the realm of transfer learning: the 'Feature Extractor' strategy and the 'Fine-tuning' strategy.

- Feature Extractor Strategy: Under the 'Feature Extractor' approach, we leveraged a diverse array of pretrained models, including renowned architectures such as AlexNet, GoogleNet, VGG16, VGG19, ResNet50, ResNet101, and NasNet Mobile. By adopting this strategy, we harnessed the capabilities of these well-established models. We achieved this by stripping away the final fully connected layer (classifier layer) and treating the remaining layers as a dedicated feature extractor. This allowed the models to seamlessly adapt to the intricate intricacies of our meniscus tear classification task. Our strategic choice facilitated the effective utilization of high-level features learned from natural image datasets. However, it's important to recognize that this approach segregates the training process into two distinct stages: feature extraction and classifier training. While this division could potentially impact the classifier's discriminative performance, especially within the specialized context of medical imaging, we mitigated this challenge by integrating clinical labels and their corresponding bounding boxes. This comprehensive approach systematically structured the classification task, enhancing the network's proficiency in identifying meniscus tears within medical images.

- Fine-tuning Strategy: In parallel with the 'Feature Extractor' strategy, we also

embraced the 'Fine-tuning' methodology. This involved employing well-known networks like AlexNet, GoogleNet, VGG16, VGG19, ResNet50, ResNet101, and NasNet Mobile. Through fine-tuning, we not only replaced and retrained the classifiers using our dedicated meniscus tear dataset but also meticulously adjusted the pretrained network's weights, extending the back-propagation process. This dynamic approach enabled the network to adapt not only to the overarching high-level features but also to the distinct attributes unique to meniscus tears as observed in medical images. Aligning with the previous strategy, the integration of clinical labels and their corresponding bounding boxes ensured a structured classification task. The fine-tuning approach confers the distinct advantage of tailoring the pretrained models more closely to the intricacies of medical imaging, potentially enhancing the classifier's efficacy.

Through a systematic exploration and comparative analysis of these transfer learning methodologies, our objective was to ascertain the optimal approach for our meniscus tear detection task. This pursuit was driven by our recognition of the challenges posed by the scarcity of clinical data and the multifaceted nature of medical images.

6.2.5 Evaluation Metrics

For a fair evaluation, we chose parameters carefully and developed deep learning models using complete k-space data for both training and validation. This thorough approach helps ensure an unbiased assessment, making our study findings more reliable.

In evaluating the reconstruction task, we used a well-known metric called Structural Similarity Index (SSIM), chosen for its relevance. SSIM effectively measures the likeness between two images by considering relationships among neighboring pixels. This detailed assessment of image similarity provides a strong evaluation of our reconstructed images, giving us valuable insights into how well our reconstruction models perform. You can find more details in Section 3.5, and the

mathematical expression is shown in Equation 3.3. This metric is a powerful tool for quantitatively assessing the quality of the reconstructed images in our study.

For the classification task, predicted class probabilities vary between 0 and 1 for each image, representing the likelihood of belonging to a specific class. Utilizing a predefined threshold, predictions are categorized as positive if they exceed the threshold; otherwise, they are classified as negative. Among the predictions that surpass the threshold, we identify three key terms for further analysis:

- True Positives (TP): These are instances where the model correctly predicts the positive class. In the context of our study, it refers to cases where the model accurately identifies meniscus tears.
- False Positives (FP): These are instances where the model predicts a positive class, but the actual class is negative. In our case, it refers to situations where the model incorrectly identifies a meniscus tear when there isn't one present.
- False Negatives (FN): These are instances where the model predicts a negative class, but the actual class is positive. It signifies cases where the model misses identifying an actual meniscus tear.
- True Negatives (TN): These are instances where the model correctly predicts the negative class. In our study, it corresponds to cases where the model accurately identifies the absence of a meniscus tear.

Evaluation of commonly used classification metrics is based on a confusion matrix with a fixed threshold, in addition to counting-based metrics (Reinke et al., 2021). The classification metrics used in this study can be briefly summarized as follows:

- **Accuracy:** Accuracy straightforwardly computes the ratio of correct predictions to all predictions, defined as Equation 6.1

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{6.1}$$

- **Precision:** Precision quantifies the likelihood of positive class predictions actually being part of the positive class, noted as Equation 6.2

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6.2}$$

- **Recall:** Recall assesses the capability of retrieving positive class predictions among all actual positive instances, described by Equation 6.3

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6.3}$$

- **Specificity:** Also known as true negative rate, specificity evaluates the model's ability to correctly identify negative instances among all actual negative instances, denoted as Equation 6.4

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6.4}$$

- **F_1 Score:** F_1 Score balances recall and precision while penalizing extreme values, noted in Equation 6.5

$$F_1 \text{ Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{6.5}$$

All global evaluation metrics applied in the classification task yield values within the range of 0 to 1. A value of 1 indicates a perfect prediction, whereas 0 signifies a poor prediction outcome. However, the interpretation of these metrics hinges on the definition of the positive class, which may be either an arbitrary or evident selection. In essence, the outcomes of these metrics can considerably differ based on the specific choice of the positive class.

6.3 Results

In order to assess the performance of MRI reconstruction and classification, we conducted two distinct sets of experiments through a standardized evaluation pipeline. The primary objective of the first set was to acquire high-quality images using various reconstruction models, thus enhancing the analytical capabilities of our study. Meanwhile, the second set of experiments aimed to evaluate the classification proficiency of these models in detecting clinically significant outcomes, specifically meniscus tears. In this section, we present the outcomes derived from both of these experiment sets, shedding light on the effectiveness of our approach in addressing the challenges posed by meniscus tear detection.

The implementation of deep learning models has led to significant advancements in noise reduction, resulting in notable improvements in SSIM scores. To visually illustrate these enhancements, Figure 6.7 provides a representative example, including an image slice with corresponding zoomed-in regions. Annotated lesions are indicated by yellow boxes, while the presence of a tear is highlighted with a white arrow. This visual aid facilitates a comparative assessment of reconstruction quality and classification effectiveness across different reconstruction frameworks, specifically focusing on acceleration rates of 4 and 8. The visualization clearly reveals variations in reconstruction quality among different methods. Notably, the ZeroFilled approach demonstrates lower reconstruction quality when preserving meniscus tears compared to other techniques, followed by E2E VarNET, iRIM, and UNET, respectively. Furthermore, UNET, while successfully preserving the tear at both rates, displays more artifacts in the meniscus tear region.

For a more comprehensive evaluation, two radiologists, M.M.B. and A.C., each with 20 and 5 years of experience, conducted a joint assessment of the reconstructed scans alongside the original images. Their focus was on qualitative assessments, where they aimed to identify any hallucinatory effects and artifacts introduced by the reconstruction models. According to their findings, in the qualitative domain, E2E VarNET outperformed iRIM and UNET at both acceleration rates, excelling in terms

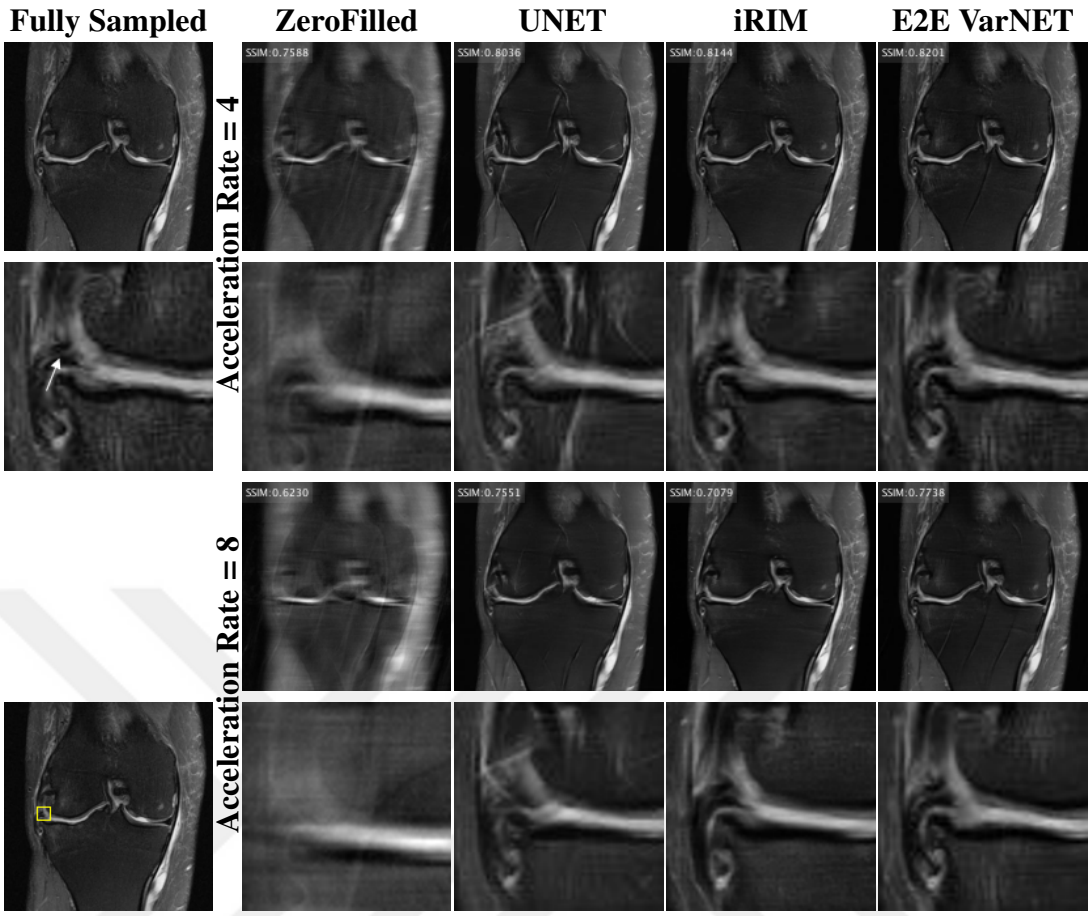


Figure 6.7 An example slice displaying a meniscus tear is showcased, along with the corresponding ground truth bounding box (highlighted in yellow). The zoomed-in region, indicated by a white arrow, specifically highlights the presence of the meniscus tear. Notably, the classification networks exhibited less favorable performance in cases of ZeroFilled FFT reconstruction (RSS), attributable to its low Signal-to-Noise Ratio (SNR) and pronounced artifacts at an acceleration rate of 8. A comparative analysis reveals that, in terms of reconstruction, UNET, iRIM, E2E VarNET, and their respective ensembles consistently yield intricate textures, with the exception of UNET at acceleration rates of 4 and 8.

of image quality and artifact reduction. UNET, on the other hand, exhibited a higher prevalence of artifacts (lineal or fogging-style) in the tibia and femur at both acceleration rates when compared to the other models. These artifacts had the potential to obscure or blend with the background anatomy, potentially masking the primary pathology. It is worth mentioning that some of these artifacts exhibited similarities to the appearance of an aged, non-displaced fracture line, which might present difficulties for radiologists with limited experience. E2E VarNET and iRIM showed higher occurrences of artifacts, particularly FP artifacts, generated by some

models, transforming certain structures into abnormalities. These artifacts have the potential to significantly impact clinical assessments during evaluations at both acceleration rates. For example, a falsely generated image of a meniscal tear can mislead radiologists into making incorrect diagnoses, as it may make the tear appear longer or more extensive than it actually is. Therefore, it is crucial to approach these artifacts with caution when interpreting the images.

Following our qualitative assessment, we proceeded with a detailed examination of classification networks to further explore the intricate relationship between deep learning models, reconstruction techniques, and the performance of classification networks, especially concerning meniscus tear classification. To provide a comprehensive overview of how various classification networks (AlexNet, GoogleNet, VGG16, VGG19, ResNet50, ResNet101, and NasNet Mobile) perform across these reconstruction models at acceleration rates of 4-8, we have compiled a comprehensive breakdown of the evaluation metrics in Table **??**. A closer look at the table reveals that the outcomes of meniscus tear classification assessments consistently align across different reconstruction methodologies. Notably, the E2E Variational Network consistently outperforms other models in terms of classification network performance, irrespective of acceleration rates. Several models, such as VGG16, VGG19, ResNET50, ResNET101, and NasNET Mobile, achieved exceptional accuracy scores exceeding 99%. More specifically, within the E2E VarNET context, ResNET50 and VGG16 exhibited slightly superior results in terms of precision, recall, and, consequently, the F_1 score. This comprehensive analysis enhances our understanding of how deep learning models, reconstruction methods, and classification networks interact in this context.

To delve deeper, a comprehensive analysis of class distribution in terms of image quality, considering classifier output, becomes imperative. The comparison of Structural Similarity Index (SSIM) between True Positives (TP) and False Negatives (FN) using diverse reconstruction techniques at acceleration rates of 4 and 8 facilitates the evaluation of the reconstruction quality's influence on the classifier's performance. This comparison is illustrated in Table **??**. The table provides a clear

inference that the average SSIM values between True Positives (TP) and False Negatives (FN) at rates 4 and 8 exhibit striking similarity across all reconstruction methods for each classifier. From this perspective, it can be deduced that the classifiers' performance was not compromised due to significant SSIM disparities in the reconstructed slices. The findings highlight that the pronounced significance of high Structural Similarity Index (SSIM) values, employed as a pivotal criterion in the reconstruction task, does not compromise the accuracy of meniscus tear classification.

After evaluating the impact of image quality on classifier performance through SSIM analysis, we shift our focus to the visualization of class activation map (CAM) results. These CAM results provide insights into the specific regions within the images that play a crucial role in the classifiers' decisions for accurate meniscus tear classification. Figures 6.8(a) and 6.8(b) present the CAM analyses for 4-fold and 8-fold accelerated and reconstructed scans with a probability score for each class, respectively. The CAM analysis highlights regions of high importance, represented in red, which correspond to the presence of meniscus tears. By visualizing the distinctive features utilized by the classifiers, the CAM analysis enhances our understanding of how these models effectively distinguish between healthy tissue and meniscus tears in the MRI scans under examination.

When we analyze these figures, especially Figures 6.8(a) and 6.8(b), it becomes clear that AI classifiers can accurately identify tears. However, they also make misclassifications in Figures 6.9(a) and 6.9(b) when it comes to the reconstructed scans at both acceleration rates. These misclassifications occur even though the AI classifiers show high probability scores. This indicates that while the AI classifiers pay attention to the meniscus region, they don't consistently achieve perfect tear localization.

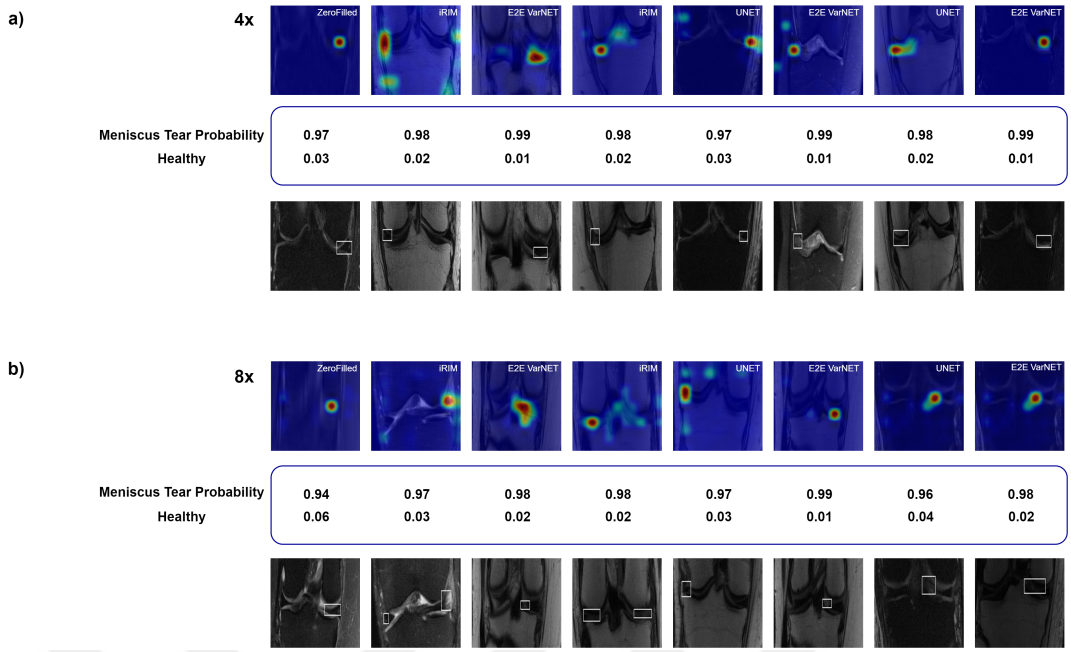


Figure 6.8 Meniscus Tear Detection Using Class Activation Map (CAM) with AI Classifiers Achieving High Probability Scores. Class activation maps (CAM) reveal the specific regions in images contributing to accurate meniscus tear classification. Images (a) and (b) depict 4-fold and 8-fold accelerated and reconstructed scans. Probability scores for each classifier, including ResNET50, AlexNET, VGG16, VGG19, ResNET50, ResNET101, GoogleNET, and NasNET Mobile, are presented from left to right. The CAM analysis effectively identifies meniscus tears, with high-importance areas highlighted in red.

6.4 Discussion and Conclusion

In this chapter, we introduced an innovative integrated deep learning framework that seamlessly integrates MRI reconstruction and Meniscus Tear classification within a sequential pipeline. This novel approach offers several advantages, including efficient testing and the ability to directly compare MRI reconstruction techniques based on global image quality metrics. Additionally, it allows for the evaluation of classifiers' performance in identifying clinically significant structures.

Specifically, our study has demonstrated the outstanding performance of deep learning models in meniscal tear classification from undersampled k-space reconstructions. Almost perfect results are obtained by VGG-16 within the E2E Variational Network compared to the clinically infeasible detection and segmentation performances reported for FastMRI+. Surprisingly, these results are almost equal for

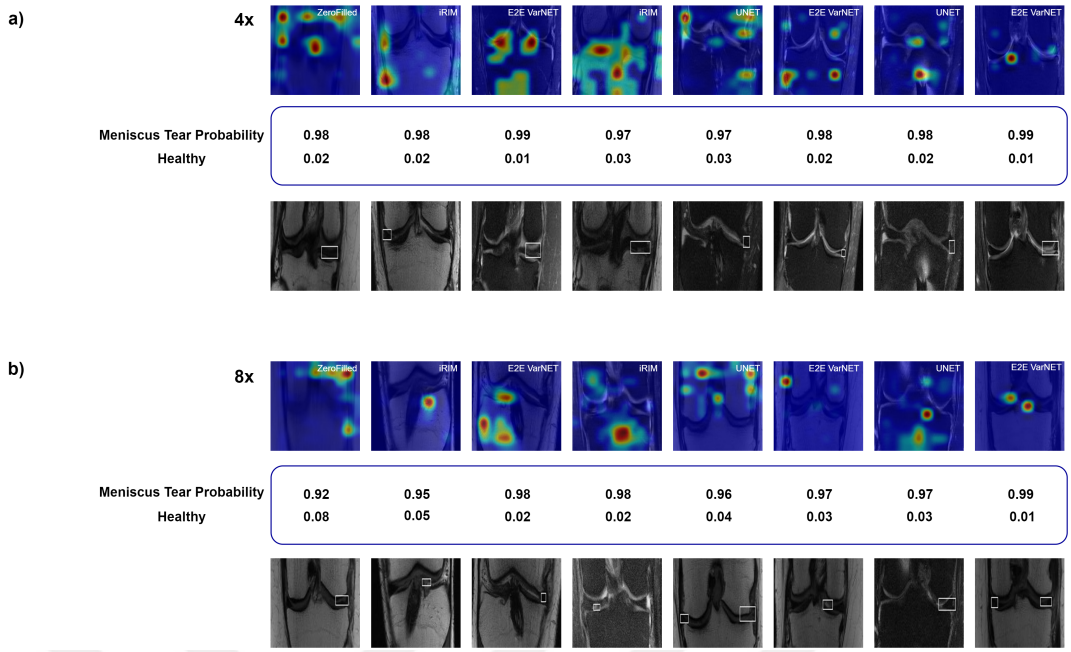


Figure 6.9 Meniscus Tear Localization with Class Activation Map (CAM) in Reconstructed Scans: AI-classifiers exhibit high probability scores for identifying meniscus tears but may mislocate them at 4-8 fold acceleration rates. Images in (a) and (b) depict 4-fold and 8-fold accelerated and reconstructed scans, respectively. Probability scores for each class are provided by various classifiers, including ResNET50, AlexNET, VGG16, VGG19, ResNET50, ResNET101, GoogleNET, and NasNET Mobile, ordered from left to right. The CAM analysis effectively highlights critical areas in red.

both 4x and 8x acceleration rates, which have significantly different characteristics. Notably, VGG-16 consistently achieved correct classifications, even when its primary focus was on different zones of the image. Thus, the exact reasons behind this success remain uncertain. This underscores the pivotal importance of refining the model's capacity for precise localization of meniscus tears, a vital aspect for accurate diagnosis and evaluation. To address this challenge, the implementation of a specialized detection loss function targeting the tear's specific patch during training is imperative. In addition to highlighting the need for improved tear localization, our study also signifies promising prospects for prospective applications. Despite encountering lower detection precision rates reported in the literature, our model consistently demonstrated high accuracy in the classification task. These results emphasize VGG-16's potential to enhance clinical diagnosis through its reliable metrics in both Meniscus Tear classification and image reconstruction tasks. Overall, our research contributes valuable insights to the advancement of deep learning in

enhancing MRI-based assessments.

Nevertheless, our results clearly demonstrate the substantial impact of DL reconstruction performance on classification accuracy, as evidenced by the high accuracy scores achieved on the fastMRI dataset. This underscores the importance of considering both aspects when developing and assessing AI-based diagnostic tools.

While deep learning methods have shown remarkable reconstruction quality at high acceleration rates, it is essential to acknowledge that certain clinically vital details, such as Meniscus Tear, may only be partially visible or compromised during the reconstruction process. This raises questions about the feasibility of expecting classifiers to accurately identify such lesions under these circumstances.

Our findings reveal that AI classifiers exhibit remarkable adaptability, consistently achieving high probability scores for meniscus tear identification, even when lesions are only partially visible or compromised during the reconstruction process. This adaptability, combined with the classifiers' focus on different regions of the menisci, highlights the significant potential of our approach in various clinical applications.

The framework proposed in this study represents a significant initial stride towards the automated assessment of reconstruction techniques and their impact on classification tasks, which is crucial for the advancement of deep learning methods in the context of accelerated MRI acquisition. While deep learning models have primarily focused on achieving high-quality reconstructions and have undergone thorough evaluations in the literature, it is essential to recognize that human raters' judgments may be influenced by variations in image quality, emphasizing the need for careful clinical evaluation.

Furthermore, it is worth noting that highly accelerated reconstructions can sometimes struggle to classify clinically vital structures effectively, despite exhibiting favorable global quality metrics such as SSIM. In this study, we revealed that high SSIM scores do not directly impact AI classifier performance in terms of TP or FN. This underscores the requirement for advanced clinically relevant metrics that can

capture and preserve crucial pathology information.

While our classification models with transfer learning have been dedicated to Meniscus Tear classification, it is important to acknowledge that the dataset encompasses various types of lesions. Future studies could expand the current classification models to accommodate multi-class categorization, enabling a comprehensive evaluation of other lesions present in the FastMRI+ dataset. However, it is crucial to consider the limitations of FastMRI+ in terms of the distribution and annotation of other pathologies.

For future directions, it would be beneficial to assess the effectiveness of this sequential framework using a prospective sampling priority for clinical application. This is crucial because retrospective undersampling does not account for signal relaxation differences during echo trains and lacks equivalence in a prospective manner. Additionally, we aim to thoroughly evaluate the performance of this framework with various acquisitions, orientations, and different vendors by incorporating more prospective cases to simulate real clinical setups.

In conclusion, this study underscores the importance of considering the interaction between image reconstruction and classification tasks. It also highlights the limitations of relying solely on traditional image quality metrics for evaluating the clinical suitability of accelerated MRI reconstructions. Future research should aim to develop and integrate more robust metrics that better align with clinical outcomes and expert assessments, ultimately advancing the field of accelerated MRI reconstruction and classification.

Table 6.2 Comparison of Meniscus Tear Classification Results on Different Reconstruction Methods

		ZeroFilled					UNET					iRIM					E2E VarNET				
		Ac.	**Prec.**	**Rec.**	**Spec.**	**F_1 Sc.**	**Ac.**	**Prec.**	**Rec.**	**Spec.**	**F_1 Sc.**	**Ac.**	**Prec.**	**Rec.**	**Spec.**	**F_1 Sc.**	**Ac.**	**Prec.**	**Rec.**	**Spec.**	**F_1 Sc.**
AlexNET	**Acceleration Rate = 4**	0.75	0.87	0.56	0.41	0.68	0.90	0.93	0.83	0.71	0.88	0.92	0.95	0.87	0.77	0.91	**0.95**	**0.96**	**0.92**	**0.80**	**0.94**
VGG16		0.72	0.86	0.50	0.47	0.64	0.98	0.99	0.97	0.80	0.98	0.99	0.99	0.99	0.97	0.99	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**
VGG19		0.72	0.86	0.51	0.43	0.64	0.95	0.97	0.92	0.94	0.94	0.99	0.99	0.98	0.98	0.98	**0.99**	**0.99**	**0.98**	**0.97**	**0.99**
GoogleNET		0.72	0.72	0.5	0.49	0.59	0.8	0.89	0.64	0.78	0.74	0.83	0.90	0.69	0.75	0.78	**0.88**	**0.93**	**0.79**	**0.97**	**0.86**
ResNET50		0.72	0.68	0.51	0.40	0.58	0.91	0.94	0.85	0.75	0.89	0.97	0.98	0.95	0.90	0.96	**0.98**	**0.99**	**0.97**	**0.95**	**0.98**
ResNET101		0.72	0.72	0.5	0.42	0.59	0.89	0.91	0.82	0.80	0.87	0.95	0.96	0.92	0.88	0.94	**0.97**	**0.97**	**0.95**	**0.96**	**0.96**
NasNET Mobile		0.72	0.72	0.5	0.47	0.59	0.87	0.93	0.77	0.86	0.84	0.93	0.96	0.88	0.93	0.92	**0.97**	**0.98**	**0.94**	**0.80**	**0.96**
AlexNET	**Acceleration Rate = 8**	0.74	0.87	0.54	0.37	0.66	0.83	0.91	0.70	0.63	0.79	0.88	0.92	0.79	0.68	0.85	**0.92**	**0.95**	**0.86**	**0.76**	**0.90**
VGG16		0.72	0.86	0.50	0.41	0.63	0.95	0.96	0.90	0.70	0.93	0.99	0.99	0.97	0.95	0.98	**0.99**	**0.99**	**0.99**	**0.97**	**0.99**
VGG19		0.72	0.86	0.50	0.39	0.64	0.90	0.94	0.82	0.82	0.87	0.96	0.97	0.92	0.93	0.95	**0.99**	**0.99**	**0.98**	**0.98**	**0.98**
GoogleNET		0.75	0.87	0.55	0.43	0.67	0.81	0.90	0.67	0.66	0.77	0.80	0.89	0.64	0.58	0.74	**0.85**	**0.91**	**0.73**	**0.71**	**0.81**
ResNET50		0.71	0.55	0.51	0.40	0.53	0.88	0.93	0.78	0.71	0.85	0.90	0.94	0.83	0.69	0.88	**0.95**	**0.97**	**0.92**	**0.85**	**0.94**
ResNET101		0.72	0.72	0.5	0.41	0.59	0.87	0.90	0.77	0.73	0.83	0.89	0.92	0.82	0.73	0.87	**0.94**	**0.95**	**0.90**	**0.89**	**0.92**
NasNET Mobile		0.72	0.36	0.50	0.40	0.42	0.81	0.90	0.67	0.68	0.77	0.85	0.91	0.73	0.82	0.81	**0.91**	**0.94**	**0.83**	**0.90**	**0.88**

* Ac., Prec, Rec., Spec. and F_1 Sc. are the abbreviation of the Accuracy, Precision, Recall, Specificity and F_1 Score, respectively and All values are truncated to two decimal with rounding, emphasized red values are higher in four decimal.

Table 6.3 Comparison of average SSIM between True Positives (TP) and False Negatives (FN) for reconstructed slices at 4x and 8x acceleration rates, across different classifiers. Scores in bold represent the highest SSIM values

Models	Acc.		AlexNET	VGG16	VGG19	ResNET50	ResNET101	GoogleNET	NasNET Mobile
ZeroFilled	**4x**	**TP**	0.71	0.71	0.71	0.71	0.71	0.71	0.71
		FN	0.68	0.68	0.68	0.68	0.68	0.68	0.68
UNET		**TP**	0.81	0.81	0.81	0.81	0.81	0.81	0.81
		FN	0.80	0.83	0.83	0.82	0.82	0.79	0.82
iRIM		**TP**	0.82	0.82	0.82	0.82	0.82	0.83	0.82
		FN	0.84	0.88	0.84	0.85	0.85	0.79	0.84
E2E VarNET		**TP**	**0.84**	**0.84**	**0.84**	**0.84**	**0.84**	**0.85**	**0.84**
		FN	**0.85**	**0.91**	**0.91**	**0.89**	**0.88**	**0.80**	**0.88**
ZeroFilled	**8x**	**TP**	0.66	0.66	0.66	0.66	0.66	0.66	0.66
		FN	0.58	0.59	0.59	0.59	0.59	0.59	0.59
UNET		**TP**	0.77	0.77	0.77	0.77	0.77	0.78	0.77
		FN	0.74	0.76	0.76	0.76	0.75	0.73	0.74
iRIM		**TP**	0.77	0.79	0.78	0.78	0.77	0.78	0.78
		FN	0.78	0.78	0.79	0.78	0.79	0.75	0.77
E2E VarNET		**TP**	**0.80**	**0.80**	**0.80**	**0.80**	**0.80**	**0.81**	**0.80**
		FN	**0.82**	**0.86**	**0.84**	**0.82**	**0.82**	**0.76**	**0.79**

CHAPTER SEVEN
CONCLUSION AND DISCUSSION

7.1 Conclusion

The broad aim of this thesis was to address the issues associated with MRI reconstruction and analysis, with a focus on enhancing efficiency, accuracy, and clinical applicability. Through a multi-faceted approach encompassing ensemble methods, DL-detected pathology identification, and innovative MRI classification techniques, the study aimed to contribute comprehensive solutions to existing challenges in the field of medical imaging. By bridging gaps, offering valuable insights, and proposing novel integrated frameworks, this thesis strives to pave the way for advancements that can significantly impact the efficacy and reliability of MRI-based diagnostics.

The main contributions of this thesis are as follows:

- **Advancements in Ensemble Methods:**
  - The exploration and application of ensemble methods in deep neural networks for MR image reconstruction, particularly on the fastMRI dataset, demonstrate promising advancements in enhancing reconstruction performance across various datasets and undersampling schemes. By leveraging the collective intelligence of multiple base learners, ensembles offer a good solution that outperforms individual models, even in challenging scenarios such as fat-suppressed scans and high acceleration rates. Specifically, averaging performs better across both single and multi-coil datasets at 4x acceleration, while weighted averaging achieves superior results across both datasets at 8x acceleration.
  - Through evaluating off-the-shelf models and established networks, combined with three non-trainable ensemble combiners, our study provides comprehensive insights into their impact on reconstruction performance. Our

analysis reveals a significant improvement in reconstruction quality, with an average increase of 0.94 dB in PSNR and 0.02 in SSIM scores for 4x acceleration, and 0.37 dB in PSNR, 0.005 in SSIM for 8x acceleration in single-coil datasets. In multi-coil datasets, we observe enhancements of 0.27 dB in PSNR, 0.001 in SSIM for 4x acceleration, and 0.43 dB in PSNR, 0.003 in SSIM for 8x acceleration across all datasets. These numerical improvements underscore the effectiveness of ensemble methods in enhancing reconstruction outcomes, indicating their potential for widespread adoption in clinical practice.

- **In-Depth Analysis of DL-Detected Pathologies:**
  - A detailed examination of DL-detected meniscal tears in retrospectively undersampled knee MRI revealed significant challenges, including the presence of hallucinations and false positive artifacts in DL reconstructions. UNET exhibited more artifacts in the femur and tibia at both acceleration rates compared to other models (iRIM and E2E VarNET). However, during evaluation at acceleration rates, E2E VarNET and iRIM showed more artifacts, particularly false positive artifacts.
  - The proposal and demonstration of an integrated pipeline for meniscal tear detection on FastMRI+ data presents state-of-the-art performance in clinical applications. Among the models, E2E VarNET achieves average precision scores of 0.69 and 0.67 at 4- and 8-fold accelerations, respectively. Notably, the UNET model demonstrates average precision values of 0.66 and 0.64, while the iRIM model shows values of 0.67 and 0.65 at 4-8 fold accelerated scans, further underscoring the superior performance of our proposed pipeline.

- **Incorporation of Blind and Non-Blind Evaluations:**
  - Significantly contributes to the field by adopting both blind and non-blind assessments, thereby providing a comprehensive understanding of diagnostic accuracy and the impact of subjective interpretation on outcomes.

– In the blind study, meniscal tears were meticulously annotated by expert radiologists without prior knowledge of acceleration levels or reconstruction methods, providing unbiased insights into the effectiveness of the DL reconstruction and detection pipeline. The study reveals that iRIM demonstrates higher performance with a mean Average Precision (mAP) of 0.44 at 8x acceleration, whereas E2E VarNET shows superior performance with an mAP of 0.6 at 4x acceleration across all datasets. However, the blind study results indicate underperformance compared to the nnDetection quantitative results when using FastMRI+ annotations as a reference. Surprisingly, when the blind study is taken as a reference, UNET exhibits mAP values of 0.91 at 4x acceleration and 0.93 at 8x acceleration across all datasets, respectively, contrary to expectations.

– The non-blind study evaluated reconstructed images in terms of diagnosis and the presence of pathological features, considering specific acceleration factors and reconstruction models, thereby enhancing our understanding of clinical applicability and robustness.

The diagnosis assessment results indicate that E2E VarNET exhibited the highest scores, averaging 3.48 at 4x acceleration and 2.48 at 8x acceleration, based on a structured 5-point scoring system. Scores ranged from "1," indicating a notable deterioration compared to the reference image, to "5," denoting a significant enhancement surpassing the reference image's quality. In comparison, iRIM scored 3 at 4x acceleration and 2.002 at 8x acceleration, while UNET scored 2.001 at 4x acceleration and 1 at 8x acceleration.

Regarding the pathological assessment, E2E VarNET demonstrated the highest average scores of 2.9862 at 4x acceleration and 2.9568 at 8x acceleration, utilizing a 0-4 ordinal scoring system for all slices. In this system, a score of "0" indicated that the meniscus tear was not observable, "1" and "2" indicated outcomes significantly worse and moderately worse compared to the reference, "3" indicated similarity to the reference, and "4" indicated outcomes better than the reference. Conversely, iRIM scored 2.9779 at 4x acceleration and 2.9202 at 8x acceleration, while UNET

exhibited scores of 2.9599 at 4x acceleration and 2.5839 at 8x acceleration.

- **Beyond Pathology Analysis:**
  - This surpasses beyond the conventional analysis of DL-detected pathologies by introducing an integrated pipeline for meniscal tear detection on FastMRI+ data, thereby advancing the scope of diagnostic capabilities in MRI imaging analysis.
  - The approach showcases state-of-the-art performance in clinical applications, demonstrating the innovative and effective nature of the proposed approach, thereby paving the way for enhanced diagnostic accuracy and improved patient care in knee MRI analysis.
- **Quantitative Concordance and Results:**
  - Highlights the quantitative concordance and results obtained from both blind and non-blind expert studies demonstrate the consistency and reliability of findings across different evaluation approaches. Both blind and non-blind expert evaluations underscore the distinct detection of meniscus tear pathology in the E2E VarNET model across all acceleration rates and datasets. In blind assessment, while iRIM shows a slightly higher mAP score for 8-fold acceleration, both E2E VarNET and iRIM annotate approximately the same volume count. Additionally, the consistency between nnDetection results, as measured by mean Average Precision (mAP), and reconstruction outcomes, as well as the outcomes from both blind and non-blind experts, highlights the robustness of the proposed approach.
  - Emphasizes the effectiveness of the comprehensive framework, underlining its potential to provide reliable and unbiased results in clinical applications, thereby enhancing diagnostic precision and contributing to more informed medical decisions.
- **Innovative Approaches to MRI Classification:**
  - Exploration of the untapped potential of DL methods in MRI, particularly

emphasizing the precise classification of meniscal tears through the utilization of undersampled k-space data and DL-generated reconstruction outputs, presents a promising avenue for enhancing diagnostic accuracy and efficiency in musculoskeletal imaging.

  - The introduction of an innovative sequential process has resulted in outstanding performance, achieving an accuracy of 0.99, precision of 0.99, recall of 0.99, specificity of 0.98, and F_1 score of 0.99 for E2E VarNET at both acceleration rates, employing VGG16, with similar outcomes observed across other transfer learning methodologies such as AlexNET, VGG19, ResNET50, ResNET 101, GoogleNET, and NasNET Mobile. These results challenge conventional assumptions concerning SSIM scores and classifier performance. Specifically, regarding FN and TP, E2E VarNET demonstrates an average SSIM of 0.84 for TP and 0.87 for FN at 4x acceleration rate, and an average of 0.80 for TP and 0.82 for FN at 8x acceleration across all classification models. The observed differences in SSIM scores between TP and FN consistently show FN SSIM scores to be generally higher than TP SSIM scores across all classification models and reconstruction outcomes at both acceleration rates.

- **Comprehensive Contributions to the Field:**

  - Bridging gaps in the current literature, providing comprehensive insights into the intricate relationship between MRI reconstruction and DL-based detection, allows us to develop more effective diagnostic tools and methodologies for enhanced medical imaging practices.

  - Addressing challenges in medical imaging and proposing novel integrated frameworks, aiming to improve the clinical utility and reliability of MRI-based diagnostics, is crucial for advancing patient care and medical decision-making in diverse healthcare settings.

7.1.1 Basic Ensembles For MRI Reconstruction Task

Ensemble methodologies, proven successful in classification and segmentation tasks, have been underexplored in Magnetic Resonance (MR) reconstruction, creating a notable gap in the literature. This study addresses this gap by introducing basic combination methods as the simplest ensemble combiners for MR reconstruction. Utilizing non-trainable ensemble combiners—Max/Min, Averaging, and Weighted Averaging Combiner—we conducted a comprehensive investigation on the fastMRI dataset. Results demonstrate the marked superiority of basic ensemble models, particularly Averaging and Weighted Averaging, over individual models at various acceleration rates, evident across evaluation metrics.

7.1.2 Meniscus Tear Detection in Accelerated MRI

Our study comprehensively evaluates MRI reconstruction and analysis, focusing on clinically relevant details related to meniscus tears. We assess the performance of deep learning (DL) detection for different DL reconstructions, previously analyzed separately. Findings emphasize the substantial impact of DL reconstruction performance on detection outcomes. The proposed comprehensive approach integrating reconstruction and detection strategies achieves state-of-the-art (SOTA) performance in detecting meniscal tears in the FastMRI+ dataset, enriching the field with valuable insights.

7.1.3 Improving DL Performance in Accelerated Knee MRI: Blind and Non-Blind Assessments

Our study significantly contributes to the field by incorporating both blind and non-blind evaluations. In the blind study, expert radiologists meticulously annotate meniscal tears without prior knowledge of acceleration levels or reconstruction methods, providing unbiased insights into the DL reconstruction and detection pipeline's effectiveness. In contrast, the non-blind study evaluates reconstructed

images, considering specific acceleration factors and reconstruction models, adding depth to our understanding of clinical applicability and robustness.

Our research goes beyond the analysis of DL-detected pathologies by introducing an integrated pipeline for meniscal tear detection on FastMRI+ data. This innovative approach not only showcases state-of-the-art performance in clinical applications but also highlights the quantitative concordance and results obtained from our blind and non-blind expert studies, emphasizing the effectiveness of our comprehensive framework.

7.1.4 Meniscus Tear Classification From Undersampled MRI

Introducing an integrated deep learning framework seamlessly combining MRI reconstruction and Meniscus Tear classification, our study offers advantages in testing efficiency and direct comparison of MRI reconstruction techniques. The outstanding performance of deep learning models in classifying meniscal tears from undersampled k-space reconstructions is highlighted. However, uncertainties persist regarding the reasons behind this success, emphasizing the need to enhance the model's tear localization capacity.

The study underscores the challenges of highly accelerated reconstructions in effectively classifying clinically vital structures, despite favorable global quality metrics. Adaptability of AI classifiers, achieving high probability scores for tear identification even with partially visible lesions, highlights the potential of our approach in various clinical applications.

7.2 Discussion and Future Endeavors

In the ongoing pursuit of advancing the clinical applicability of deep learning-based MRI reconstruction and detection methods, several key considerations and future directions emerge:

- **Comprehensive Performance Evaluation**: Ongoing efforts include a detailed evaluation of the proposed framework across various parameters. This involves systematically assessing its performance under diverse conditions, including different acquisition setups, orientations, and varying vendors. The aim is to gain a nuanced understanding of the framework's robustness and adaptability in real-world scenarios.

- **Exploration of End-to-End Models**: Another crucial avenue of exploration is the implementation and investigation of genuine end-to-end models. This involves developing neural networks that seamlessly integrate both image reconstruction and subsequent pathology detection. The goal is to enhance diagnostic efficiency by creating a unified pipeline that can optimize both tasks concurrently.

- **Enhanced Diagnostic Efficiency**: The exploration of end-to-end models specifically targets improved diagnostic efficiency. By streamlining the reconstruction and detection processes within a singular neural network, we anticipate a more seamless and efficient diagnostic pipeline. This approach holds the potential to elevate the overall performance of the proposed framework.

- **Adaptability Across Parameters**: The thorough evaluation across various parameters is designed to ensure the adaptability and generalizability of the framework. Understanding how the model performs under different acquisition conditions and orientations is crucial for its successful deployment in diverse clinical settings.

- **Advancing Clinical Relevance**: The overarching goal is to advance the clinical relevance of deep learning-based methods. This involves not only enhancing technical aspects but also ensuring that the proposed framework aligns with the practical needs of healthcare professionals. The continuous refinement of the model considers factors such as signal relaxation differences and aims to bridge the gap between research outcomes and clinical utility.

- **Prospective Sampling Strategy**: A key focus of future investigations is the

evaluation of the framework using a prospective sampling strategy. Recognizing the limitations of retrospective undersampling, particularly in accounting for signal relaxation differences during echo trains, emphasizes the importance of testing the model in a prospective setting for more accurate representation.

- **Comparison with Parallel Imaging Approaches**: The extended evaluation includes a comparative analysis with traditional parallel imaging approaches such as GRAPPA, SENSE, and PI. Understanding the relative strengths and weaknesses of these methods provides insights into how the proposed framework contributes to the landscape of MRI reconstruction techniques.

In summary, the ongoing and upcoming endeavors outlined above collectively aim to refine and optimize the proposed framework, ensuring its effectiveness, adaptability, and practical relevance in the dynamic and varied context of clinical MRI applications.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. et al. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, (265–283).

Adler, J., & Öktem, O. (2018). Learned primal-dual reconstruction. *IEEE transactions on medical imaging*, *37*(6), 1322–1332.

Afat, S., Herrmann, J., Almansour, H., Benkert, T., Weiland, E., Hölldobler, T., Nikolaou, K., & Gassenmaier, S. (2023). Acquisition time reduction of diffusion-weighted liver imaging using deep learning image reconstruction. *Diagnostic and Interventional Imaging*, *104*(4), 178–184.

Aggarwal, H. K., & Jacob, M. (2020). J-modl: Joint model-based deep learning for optimized sampling and reconstruction. *IEEE journal of selected topics in signal processing*, *14*(6), 1151–1162.

Aggarwal, H. K., Mani, M. P., & Jacob, M. (2018). Modl: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging*, *38*(2), 394–405.

Akçakaya, M., Moeller, S., Weingärtner, S., & Uğurbil, K. (2019). Scan-specific robust artificial-neural-networks for k-space interpolation (raki) reconstruction: Database-free deep learning for fast imaging. *Magnetic resonance in medicine*, *81*(1), 439–453.

Antun, V., Renna, F., Poon, C., Adcock, B., & Hansen, A. C. (2020). On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, *117*(48), 30088–30095.

Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A. et al. (2011). The lung image database consortium (lidc) and image database resource initiative

(idri): a completed reference database of lung nodules on ct scans. *Medical physics*, *38*(2), 915–931.

Baalbaki, J. M., El Kadri, S. S., Awad, K. A., Al Aridy, C. H., Aridi, M. S., El Zein, M. A., Diab, M. O., Sabbah, M. M., Zantout, R. N., & Adla, Y. A. A. (2023). An end-to-end deep learning approach for atherosclerosis detection using autoencoders. *BioSMART*, 1570902357.

Bae, S. H., Hwang, J., Hong, S. S., Lee, E. J., Jeong, J., Benkert, T., Sung, J., & Arberet, S. (2022). Clinical feasibility of accelerated diffusion weighted imaging of the abdomen with deep learning reconstruction: Comparison with conventional diffusion weighted imaging. *European Journal of Radiology*, *154*, 110428.

Bandos, A. I., Rockette, H. E., Song, T., & Gur, D. (2009). Area under the free-response roc curve (froc) and a related summary index. *Biometrics*, *65*(1), 247–256.

Baumgartner, M., Jäger, P. F., Isensee, F., & Maier-Hein, K. H. (2021). nndetection: a self-configuring method for medical object detection. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, (530–539). Springer.

Bazeley, P. (2017). Integrating analyses in mixed methods research. *Integrating Analyses in Mixed Methods Research*, 1–344.

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. (2018). Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of mrnet. *PLoS medicine*, *15*(11), e1002699.

Boniatis, I., Panayiotakis, G., & Panagiotopoulos, E. (2008). A computer-based system for the discrimination between normal and degenerated menisci from magnetic resonance images. In *2008 IEEE International Workshop on Imaging Systems and Techniques*, (335–339). IEEE.

Caballero, J., Price, A. N., Rueckert, D., & Hajnal, J. V. (2014). Dictionary learning and time sparsity for dynamic mr data reconstruction. *IEEE transactions on medical imaging*, *33*(4), 979–994.

Cao, C., Cui, Z.-X., Liu, S., Zheng, H., Liang, D., & Zhu, Y. (2022). High-frequency space diffusion models for accelerated mri. *arXiv preprint arXiv:2208.05481*.

Chambolle, A., & Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, *40*, 120–145.

Chaudhari, A., Fang, Z., Hyung Lee, J., Gold, G., & Hargreaves, B. (2018). Deep learning super-resolution enables rapid simultaneous morphological and quantitative magnetic resonance imaging. In *Machine Learning for Medical Image Reconstruction: First International Workshop, MLMIR 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1*, (3–11). Springer.

Chaudhari, A. S., Grissom, M. J., Fang, Z., Sveinsson, B., Lee, J. H., Gold, G. E., Hargreaves, B. A., & Stevens, K. J. (2021). Diagnostic accuracy of quantitative multicontrast 5-minute knee mri using prospective artificial intelligence image quality enhancement. *AJR. American journal of roentgenology*, *216*(6), 1614.

Chaudhari, A. S., Sandino, C. M., Cole, E. K., Larson, D. B., Gold, G. E., Vasanawala, S. S., Lungren, M. P., Hargreaves, B. A., & Langlotz, C. P. (2021). Prospective deployment of deep learning in mri: a framework for important considerations, challenges, and recommendations for best practices. *Journal of Magnetic Resonance Imaging*, *54*(2), 357–371.

Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis*, *79*, 102444.

Cheng, J., Wang, H., Ying, L., & Liang, D. (2019). Model learning: Primal dual networks for fast mr imaging. In *Medical Image Computing and Computer Assisted*

Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22, (21–29). Springer.

Cheng, J. Y., Chen, F., Alley, M. T., Pauly, J. M., & Vasanawala, S. S. (2018). Highly scalable image reconstruction using deep neural networks with bandpass filtering. *arXiv preprint arXiv:1805.03300*.

Cui, Z.-X., Jia, S., Cao, C., Zhu, Q., Liu, C., Qiu, Z., Liu, Y., Cheng, J., Wang, H., Zhu, Y. et al. (2023). K-unn: k-space interpolation with untrained neural network. *Medical Image Analysis*, *88*, 102877.

Dai, Y., Gao, Y., & Liu, F. (2021). Transmed: Transformers advance multi-modal medical image classification. *Diagnostics*, *11*(8), 1384.

Dar, S. U. H., Özbey, M., Çatlı, A. B., & Çukur, T. (2020). A transfer-learning approach for accelerated mri using deep neural networks. *Magnetic resonance in medicine*, *84*(2), 663–685.

Darestani, M. Z., Chaudhari, A. S., & Heckel, R. (2021). Measuring robustness in deep learning based compressive sensing. In *International Conference on Machine Learning*, (2433–2444). PMLR.

Darestani, M. Z., & Heckel, R. (2021). Accelerated mri with un-trained neural networks. *IEEE Transactions on Computational Imaging*, *7*, 724–733.

Desai, A. D., Schmidt, A. M., Rubin, E. B., Sandino, C. M., Black, M. S., Mazzoli, V., Stevens, K. J., Boutin, R., Ré, C., Gold, G. E. et al. (2022). Skm-tea: A dataset for accelerated mri reconstruction with dense image labels for quantitative clinical evaluation. *arXiv preprint arXiv:2203.06823*.

Du, T., Zhang, H., Li, Y., Pickup, S., Rosen, M., Zhou, R., Song, H. K., & Fan, Y. (2021). Adaptive convolutional neural networks for accelerating magnetic resonance imaging via k-space data interpolation. *Medical image analysis*, *72*, 102098.

Duff, M., Simpson, I. J., Ehrhardt, M. J., & Campbell, N. D. (2022). Compressed sensing mri reconstruction regularized by vaes with structured image covariance. *arXiv preprint arXiv:2210.14586*.

Dung, N. T., Thuan, N. H., Van Dung, T., Van Nho, L., Tri, N. M., Vy, V. P. T., Phat, N. T., Chuong, D. A., Dang, L. H. et al. (2023). End-to-end deep learning model for segmentation and severity staging of anterior cruciate ligament injuries from mri. *Diagnostic and Interventional Imaging*, *104*(3), 133–141.

Ebner, M., Wang, G., Li, W., Aertsen, M., Patel, P. A., Aughwane, R., Melbourne, A., Doel, T., Dymarkowski, S., De Coppi, P. et al. (2020). An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain mri. *NeuroImage*, *206*, 116324.

Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.-J., & Hwang, D. (2018). Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic resonance in medicine*, *80*(5), 2188–2201.

Fabian, Z., Tinaz, B., & Soltanolkotabi, M. (2022). Humus-net: Hybrid unrolled multi-scale network architecture for accelerated mri reconstruction. *Advances in Neural Information Processing Systems*, *35*, 25306–25319.

Feng, C.-M., Yan, Y., Chen, G., Xu, Y., Hu, Y., Shao, L., & Fu, H. (2022). Multi-modal transformer for accelerated mr imaging. *IEEE Transactions on Medical Imaging*.

Fritz, B., & Fritz, J. (2022). Artificial intelligence for mri diagnosis of joints: a scoping review of the current state-of-the-art of deep learning-based approaches. *Skeletal Radiology*, *51*(2), 315–329.

Fritz, B., Marbach, G., Civardi, F., Fucentese, S. F., & Pfirrmann, C. W. (2020). Deep convolutional neural network-based detection of meniscus tears: comparison with radiologists and surgery as standard of reference. *Skeletal radiology*, *49*(8), 1207–1217.

FSE'XL, C., PD, P., & FAT, F. Creation of fully sampled mr data repository for compressed sensing of the knee.

Fujima, N., Nakagawa, J., Kameda, H., Ikebe, Y., Harada, T., Shimizu, Y., Tsushima, N., Kano, S., Homma, A., Kwon, J. et al. (2023). Improvement of image quality in diffusion-weighted imaging with model-based deep learning reconstruction for

evaluations of the head and neck. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 1–9.

Ganaie, M. A., Hu, M., Malik, A., Tanveer, M., & Suganthan, P. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151.

Gao, Z., Guo, Y., Zhang, J., Zeng, T., & Yang, G. (2023). Hierarchical perception adversarial learning framework for compressed sensing mri. *IEEE Transactions on Medical Imaging*.

Gassenmaier, S., Küstner, T., Nickel, D., Herrmann, J., Hoffmann, R., Almansour, H., Afat, S., Nikolaou, K., & Othman, A. E. (2021). Deep learning applications in magnetic resonance imaging: has the future become present? *Diagnostics*, *11*(12), 2181.

Ghodrati, V., Shao, J., Bydder, M., Zhou, Z., Yin, W., Nguyen, K.-L., Yang, Y., & Hu, P. (2019). Mr image reconstruction using deep learning: evaluation of network structure and loss functions. *Quantitative imaging in medicine and surgery*, *9*(9), 1516.

Gu, H., Yaman, B., Ugurbil, K., Moeller, S., & Akçakaya, M. (2021). Compressed sensing mri with 1-wavelet reconstruction revisited using modern data science tools. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, (3596–3600). IEEE.

Haldar, J. P., Hernando, D., & Liang, Z.-P. (2010). Compressed-sensing mri with random encoding. *IEEE transactions on Medical Imaging*, *30*(4), 893–903.

Hammernik, K., Klatzer, T., Kobler, E., Recht, M. P., Sodickson, D. K., Pock, T., & Knoll, F. (2018). Learning a variational network for reconstruction of accelerated mri data. *Magnetic resonance in medicine*, *79*(6), 3055–3071.

Han, Y., Sunwoo, L., & Ye, J. C. (2019). *k*–space deep learning for accelerated mri. *IEEE transactions on medical imaging*, *39*(2), 377–386.

Han, Y., Yoo, J., Kim, H. H., Shin, H. J., Sung, K., & Ye, J. C. (2018). Deep learning with domain adaptation for accelerated projection-reconstruction mr. *Magnetic resonance in medicine*, *80*(3), 1189–1205.

Harman, F., Selver, M. A., Baris, M. M., Canturk, A., & Oksuz, I. (2023). Deep learning-based meniscus tear detection from accelerated mri. *IEEE Access*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, (1026–1034).

Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M. et al. (2019). The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. *arXiv preprint arXiv:1904.00445*.

Hennemuth, A., Goubergrits, L., Ivantsits, M., & Kuhnigk, J.-M. (2021). *Cerebral Aneurysm Detection and Analysis*. Springer.

Hou, H., Shao, Y., Geng, Y., Hou, Y., Ding, P., & Wei, B. (2023). Pncs: Pixel-level non-local method based compressed sensing undersampled mri image reconstruction. *IEEE Access*.

Hu, K., Wu, W., Li, W., Simic, M., Zomaya, A., & Wang, Z. (2022). Adversarial evolving neural network for longitudinal knee osteoarthritis prediction. *IEEE Transactions on Medical Imaging*, *41*(11), 3207–3217.

Huang, J., Wang, S., Zhou, G., Hu, W., & Yu, G. (2022). Evaluation on the generalization of a learned convolutional neural network for mri reconstruction. *Magnetic resonance imaging*, *87*, 38–46.

Huang, Z., & Ravishankar, S. (2022). Single-pass object-adaptive data undersampling and reconstruction for mri. *IEEE Transactions on Computational Imaging*, *8*, 333–345.

Hung, T. N. K., Vy, V. P. T., Tri, N. M., Hoang, L. N., Tuan, L. V., Ho, Q. T., Le, N. Q. K., & Kang, J.-H. (2023). Automatic detection of meniscus tears

using backbone convolutional neural networks on knee mri. *Journal of Magnetic Resonance Imaging*, *57*(3), 740–749.

Hyun, C. M., Kim, H. P., Lee, S. M., Lee, S., & Seo, J. K. (2018). Deep learning for undersampled mri reconstruction. *Physics in Medicine & Biology*, *63*(13), 135007.

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H. (2021). nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, *18*(2), 203–211.

Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S. et al. (2018). nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*.

Jaeger, P. F., Kohl, S. A., Bickelhaupt, S., Isensee, F., Kuder, T. A., Schlemmer, H.-P., & Maier-Hein, K. H. (2020). Retina u-net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. In *Machine Learning for Health Workshop*, (171–183). PMLR.

Jin, L., Yang, J., Kuang, K., Ni, B., Gao, Y., Sun, Y., Gao, P., Ma, W., Tan, M., Kang, H. et al. (2020). Deep-learning-assisted detection and segmentation of rib fractures from ct scans: Development and validation of fracnet. *EBioMedicine*, *62*.

Johnson, P. M., Lin, D. J., Zbontar, J., Zitnick, C. L., Sriram, A., Muckley, M., Babb, J. S., Kline, M., Ciavarra, G., Alaia, E. et al. (2023). Deep learning reconstruction enables prospectively accelerated clinical knee mri. *Radiology*, 220425.

Johnson, P. M., Tong, A., Donthireddy, A., Melamud, K., Petrocelli, R., Smereka, P., Qian, K., Keerthivasan, M. B., Chandarana, H., & Knoll, F. (2022). Deep learning reconstruction enables highly accelerated biparametric mr imaging of the prostate. *Journal of Magnetic Resonance Imaging*, *56*(1), 184–195.

Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D. et al. (2018). Ensembles of multiple models and architectures for robust brain tumour segmentation. In

Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 14, 2017, Revised Selected Papers 3, (450–462). Springer.

Kara, A. C., & Hardalaç, F. (2021). Detection and classification of knee injuries from mr images using the mrnet dataset with progressively operating deep learning methods. *Machine Learning and Knowledge Extraction*, *3*(4), 1009–1029.

Kavur, A. E., Kuncheva, L. I., & Selver, M. A. (2022). Basic ensembles of vanilla-style deep learning models improve liver segmentation from ct images. In *Convolutional Neural Networks for Medical Image Processing Applications* (52–74). CRC Press.

Kerstin, H., Jo, S., Chen, Q., Jingming, D., M., S. R., & Daniel, R. (2019). σ-net Systematic Evaluation of Iterative Deep Neural Networks for Fast Parallel MR Image Reconstruction. *arXiv preprint arXiv:1912.09278*.

Kim, D. H., Kim, B., Lee, H.-S., Benkert, T., Kim, H., Choi, J.-I., Oh, S. N., & Rha, S. E. (2023). Deep learning–accelerated liver diffusion-weighted imaging: Intraindividual comparison and additional phantom study of free-breathing and respiratory-triggering acquisitions. *Investigative Radiology*, *58*(11), 782–790.

Kim, D. K., Lee, S.-Y., Lee, J., Huh, Y. J., Lee, S., Lee, S., Jung, J.-Y., Lee, H.-S., Benkert, T., & Park, S.-H. (2024). Deep learning-based k-space-to-image reconstruction and super resolution for diffusion-weighted imaging in whole-spine mri. *Magnetic Resonance Imaging*, *105*, 82–91.

Knoll, F., Hammernik, K., Zhang, C., Moeller, S., Pock, T., Sodickson, D. K., & Akcakaya, M. (2020). Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE signal processing magazine*, *37*(1), 128–140.

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. (2020). Advancing machine

learning for mr image reconstruction with an open competition: Overview of the 2019 fastmri challenge. *Magnetic resonance in medicine*, *84*(6), 3054–3070.

Knoll, F., Zbontar, J., Sriram, A., Muckley, M. J., Bruno, M., Defazio, A., Parente, M., Geras, K. J., Katsnelson, J., Chandarana, H. et al. (2020). fastmri: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. *Radiology: Artificial Intelligence*, *2*(1), e190007.

Köse, C., Gençalioğlu, O., & Şevik, U. (2009). An automatic diagnosis method for the knee meniscus tears in mr images. *Expert systems with applications*, *36*(2), 1208–1216.

Kuncheva, L. I. (2014). *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons.

Kusakunniran, W., Karnjanapreechakorn, S., Siriapisith, T., & Saiviroonporn, P. (2023). Fast mri reconstruction using strainnet with dual-domain loss on spatial and frequency spaces. *Intelligent Systems with Applications*, *18*, 200203.

Lecouvet, F., Van Haver, T., Acid, S., Perlepe, V., Kirchgesner, T., Berg, B. V., Triqueneaux, P., Denis, M.-L., Thienpont, E., & Malghem, J. (2018). Magnetic resonance imaging (mri) of the knee: Identification of difficult-to-diagnose meniscal lesions. *Diagnostic and Interventional Imaging*, *99*(2), 55–64.

Lee, D., Yoo, J., Tak, S., & Ye, J. C. (2018). Deep residual learning for accelerated mri using magnitude and phase networks. *IEEE Transactions on Biomedical Engineering*, *65*(9), 1985–1995.

Lee, E. J., Chang, Y.-W., Sung, J. K., & Thomas, B. (2022). Feasibility of deep learning k-space-to-image reconstruction for diffusion weighted imaging in patients with breast cancers: Focus on image quality and reduced scan time. *European Journal of Radiology*, *157*, 110608.

Lee, K.-L., Kessler, D. A., Dezonie, S., Chishaya, W., Shepherd, C., Carmo, B., Graves, M. J., & Barrett, T. (2023). Assessment of deep learning-based

reconstruction on t2-weighted and diffusion-weighted prostate mri image quality. *European Journal of Radiology*, *166*, 111017.

Li, Y., Li, J., Ma, F., Du, S., & Liu, Y. (2021). High quality and fast compressed sensing mri reconstruction via edge-enhanced dual discriminator generative adversarial network. *Magnetic Resonance Imaging*, *77*, 124–136.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60–88.

Liu, F., Zhou, Z., Samsonov, A., Blankenbaker, D., Larison, W., Kanarek, A., Lian, K., Kambhampati, S., & Kijowski, R. (2018). Deep learning approach for evaluating knee mr images: achieving high diagnostic performance for cartilage lesion detection. *Radiology*, *289*(1), 160–169.

Liu, J., Qin, C., & Yaghoobi, M. (2023). High-fidelity mri reconstruction using adaptive spatial attention selection and deep data consistency prior. *IEEE Transactions on Computational Imaging*, *9*, 298–313.

Liu, T., Chaman, A., Belius, D., & Dokmanić, I. (2022). Learning multiscale convolutional dictionaries for image reconstruction. *IEEE Transactions on Computational Imaging*, *8*, 425–437.

Liu, X., Pang, Y., Jin, R., Liu, Y., & Wang, Z. (2022). Dual-domain reconstruction network with v-net and k-net for fast mri. *Magnetic Resonance in Medicine*, *88*(6), 2694–2708.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (3431–3440).

Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *58*(6), 1182–1195.

Lustig, M., Donoho, D. L., Santos, J. M., & Pauly, J. M. (2008). Compressed sensing mri. *IEEE signal processing magazine*, *25*(2), 72–82.

Ma, Q., Lai, Z., Wang, Z., Qiu, Y., Qu, B., Zhang, H., & Qu, X. (2023). Mri reconstruction with enhanced self-similarity using graph convolutional network.

Mangone, M., Diko, A., Giuliani, L., Agostini, F., Paoloni, M., Bernetti, A., Santilli, G., Conti, M., Savina, A., Iudicelli, G. et al. (2023). A machine learning approach for knee injury detection from magnetic resonance imaging. *International Journal of Environmental Research and Public Health*, *20*(12), 6059.

Montalt-Tordera, J., Muthurangu, V., Hauptmann, A., & Steeden, J. A. (2021). Machine learning in magnetic resonance imaging: image reconstruction. *Physica Medica*, *83*, 79–87.

Moreno López, M., Frederick, J. M., & Ventura, J. (2021). Evaluation of mri denoising methods using unsupervised learning. *Frontiers in Artificial Intelligence*, *4*, 642731.

Muckley, M. J., Riemenschneider, B., Radmanesh, A., Kim, S., Jeong, G., Ko, J., Jun, Y., Shin, H., Hwang, D., Mostapha, M. et al. (2021). Results of the 2020 fastmri challenge for machine learning mr image reconstruction. *IEEE transactions on medical imaging*, *40*(9), 2306–2317.

Nasser, Y., Jennane, R., Chetouani, A., Lespessailles, E., & El Hassouni, M. (2020). Discriminative regularized auto-encoder for early detection of knee osteoarthritis: data from the osteoarthritis initiative. *IEEE transactions on medical imaging*, *39*(9), 2976–2984.

Nevitt, M., Felson, D., & Lester, G. (2006). The osteoarthritis initiative. *Protocol for the cohort study*, *1*.

Ni, M., Chen, W., Zhao, Q., Zhao, Y., & Yuan, H. (2023). Deep learning approach for mri in the classification of anterior talofibular ligament injuries. *Journal of Magnetic Resonance Imaging*.

Ni, M., He, M., Yang, Y., Wen, X., Zhao, Y., Gao, L., Yan, R., Xu, J., Zhang, Y., Chen, W. et al. (2023). Application research of ai-assisted compressed sensing technology in mri scanning of the knee joint: 3d-mri perspective. *European Radiology*, 1–13.

Ong, F., Amin, S., Vasanawala, S., & Lustig, M. (2018). Mridata. org: An open archive for sharing mri raw data. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 26.

Pal, A., Ning, L., & Rathi, Y. (2023). A domain-agnostic mr reconstruction framework using a randomly weighted neural network. *bioRxiv*, 2023–03.

Pal, A., & Rathi, Y. (2022). A review and experimental evaluation of deep learning methods for mri reconstruction. *The journal of machine learning for biomedical imaging*, *1*.

Pandey, R., Mitra, S., Prajapati, A., Singh, A., & Shrivastava, V. (2021). Deep convolutional neural network-based knee injury classification using magnetic resonance imaging. In *Proceedings of International Conference on Communication, Circuits, and Systems: IC3S 2020*, (419–425). Springer.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Patsanis, A., Sunoqrot, M. R., Langørgen, S., Wang, H., Selnæs, K. M., Bertilsson, H., Bathen, T. F., & Elschot, M. (2023). A comparison of generative adversarial networks for automated prostate cancer detection on t2-weighted mri. *Informatics in Medicine Unlocked*, *39*, 101234.

Pawar, K., Egan, G. F., & Chen, Z. (2021). Domain knowledge augmentation of parallel mr image reconstruction using deep learning. *Computerized Medical Imaging and Graphics*, *92*, 101968.

Peng, C., Guo, P., Zhou, S. K., Patel, V. M., & Chellappa, R. (2022). Towards performant and reliable undersampled mr reconstruction via diffusion model

sampling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (623–633). Springer.

Pezzotti, N., Yousefi, S., Elmahdy, M. S., Van Gemert, J. H. F., Schuelke, C., Doneva, M., Nielsen, T., Kastryulin, S., Lelieveldt, B. P., Van Osch, M. J. et al. (2020). An adaptive intelligence algorithm for undersampled knee mri reconstruction. *IEEE Access*, *8*, 204825–204838.

Prince, J. L., & Links, J. M. (2023). *Medical imaging signals and systems*.

Putzky, P., Karkalousos, D., Teuwen, J., Miriakov, N., Bakker, B., Caan, M., & Welling, M. (2019). i-rim applied to the fastmri challenge. *arXiv preprint arXiv:1910.08952*.

Putzky, P., & Welling, M. (2019). Invert to learn to invert. *Advances in neural information processing systems*, *32*.

Quan, C., Zhou, J., Zhu, Y., Chen, Y., Wang, S., Liang, D., & Liu, Q. (2021). Homotopic gradients of generative density priors for mr image reconstruction. *IEEE Transactions on Medical Imaging*, *40*(12), 3265–3278.

Quan, T. M., Nguyen-Duc, T., & Jeong, W.-K. (2018). Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss. *IEEE transactions on medical imaging*, *37*(6), 1488–1497.

Ramakrishna, B., Liu, W., Saiprasad, G., Safdar, N., Chang, C.-I., Siddiqui, K., Kim, W., Siegel, E., Chai, J.-W., Chen, C. C.-C. et al. (2009). An automatic computer-aided detection system for meniscal tears on magnetic resonance images. *IEEE transactions on medical imaging*, *28*(8), 1308–1316.

Ramzi, Z., Ciuciu, P., & Starck, J.-L. (2020). Benchmarking mri reconstruction neural networks on large public datasets. *Applied Sciences*, *10*(5), 1816.

Rasti-Meymandi, A., Ghaffari, A., & Fatemizadeh, E. (2023). Plug and play augmented hqs: Convergence analysis and its application in mri reconstruction. *Neurocomputing*, *518*, 1–14.

Razumov, A., Rogov, O. Y., & Dylov, D. V. (2022). Optimal mri undersampling patterns for pathology localization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (768–779). Springer.

Reinke, A., Eisenmann, M., Tizabi, M. D., Sudre, C. H., Rädsch, T., Antonelli, M., Arbel, T., Bakas, S., Cardoso, M. J., Cheplygina, V. et al. (2021). Common limitations of image processing metrics: A picture story. *arXiv preprint arXiv:2104.05642*.

Rizk, B., Brat, H., Zille, P., Guillin, R., Pouchy, C., Adam, C., Ardon, R., & d'Assignies, G. (2021). Meniscal lesion detection and characterization in adult knee mri: a deep learning model approach with external validation. *Physica Medica*, *83*, 64–71.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, (234–241). Springer.

Sarkar, D., Bali, R., & Ghosh, T. (2018). *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras*. Packt Publishing Ltd.

Sauer, S. T., Christner, S. A., Lois, A.-M., Woznicki, P., Curtaz, C., Kunz, A. S., Weiland, E., Benkert, T., Bley, T. A., Baeßler, B. et al. (2023). Deep learning k-space-to-image reconstruction facilitates high spatial resolution and scan time reduction in diffusion-weighted imaging breast mri. *Journal of Magnetic Resonance Imaging*.

Saygili, A., & Albayrak, S. (2018). Meniscus tear classification using histogram of oriented gradients in knee mr images. In *2018 26th Signal Processing and Communications Applications Conference (SIU)*, (1–4). IEEE.

Schlemper, J., Caballero, J., Hajnal, J. V., Price, A., & Rueckert, D. (2017). A

deep cascade of convolutional neural networks for mr image reconstruction. In *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings 25*, (647–658). Springer.

Setio, A. A. A., Traverso, A., De Bel, T., Berens, M. S., Van Den Bogaard, C., Cerello, P., Chen, H., Dou, Q., Fantacci, M. E., Geurts, B. et al. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Medical image analysis*, *42*, 1–13.

Shakhovska, N., & Pukach, P. (2022). Comparative analysis of backbone networks for deep knee mri classification models. *Big Data and Cognitive Computing*, *6*(3), 69.

Shin, H., Choi, G. S., Shon, O.-J., Kim, G. B., & Chang, M. C. (2022). Development of convolutional neural network model for diagnosing meniscus tear using magnetic resonance image. *BMC Musculoskeletal Disorders*, *23*(1), 1–9.

Simpson, A. L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., Van Ginneken, B., Kopp-Schneider, A., Landman, B. A., Litjens, G., Menze, B. et al. (2019). A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*.

Siouras, A., Moustakidis, S., Giannakidis, A., Chalatsis, G., Liampas, I., Vlychou, M., Hantes, M., Tasoulis, S., & Tsaopoulos, D. (2022). Knee injury detection using deep learning on mri studies: A systematic review. *Diagnostics*, *12*(2), 537.

Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C. L., Yakubova, N., Knoll, F., & Johnson, P. (2020). End-to-end variational networks for accelerated mri reconstruction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, (64–73). Springer.

Sriram, A., Zbontar, J., Murrell, T., Zitnick, C. L., Defazio, A., & Sodickson, D. K. (2020). Grappanet: Combining parallel imaging with deep learning for multi-coil

mri reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (14315–14322).

Sun, J., Li, H., Xu, Z. et al. (2016). Deep admm-net for compressive sensing mri. *Advances in neural information processing systems*, *29*.

Sun, K., Wang, Q., & Shen, D. (2023). Joint cross-attention network with deep modality prior for fast mri reconstruction. *IEEE Transactions on Medical Imaging*.

Tajima, T., Akai, H., Sugawara, H., Furuta, T., Yasaka, K., Kunimatsu, A., Yoshioka, N., Akahane, M., Abe, O., Ohtomo, K. et al. (2022). Feasibility of accelerated whole-body diffusion-weighted imaging using a deep learning-based noise-reduction technique in patients with prostate cancer. *Magnetic Resonance Imaging*, *92*, 169–179.

Tian, Q., Bilgic, B., Fan, Q., Ngamsombat, C., Zaretskaya, N., Fultz, N. E., Ohringer, N. A., Chaudhari, A. S., Hu, Y., Witzel, T. et al. (2021). Improving in vivo human cerebral cortical surface reconstruction using data-driven super-resolution. *Cerebral Cortex*, *31*(1), 463–482.

Timmins, K., Bennink, E., van der Schaaf, I., Velthuis, B., Ruigrok, Y., & Kuijf, H. (2020). Intracranial aneurysm detection and segmentation challenge. *Zenodo*.

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. (2023). K2s challenge: From undersampled k-space to automatic segmentation. *Bioengineering*, *10*(2), 267.

Tong, C., Pang, Y., & Wang, Y. (2022). Hiwdnet: A hybrid image-wavelet domain network for fast magnetic resonance image reconstruction. *Computers in Biology and Medicine*, *151*, 105947.

van Amerom, J. F., Lloyd, D. F., Price, A. N., Kuklisova Murgasova, M., Aljabar, P., Malik, S. J., Lohezic, M., Rutherford, M. A., Pushparajah, K., Razavi, R. et al. (2018). Fetal cardiac cine imaging using highly accelerated dynamic mri with retrospective motion correction and outlier rejection. *Magnetic resonance in medicine*, *79*(1), 327–338.

van Beek, E. J., Kuhl, C., Anzai, Y., Desmond, P., Ehman, R. L., Gong, Q., Gold, G., Gulani, V., Hall-Craggs, M., Leiner, T. et al. (2019). Value of mri in medicine: more than just another test? *Journal of Magnetic Resonance Imaging*, *49*(7), e14–e25.

Van Nguyen, H., Zhou, K., & Vemulapalli, R. (2015). Cross-domain synthesis of medical images using efficient location-sensitive deep network. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part I 18*, (677–684). Springer.

Vasudeva, B., Deora, P., Bhattacharya, S., & Pradhan, P. M. (2022). Compressed sensing mri reconstruction with co-vegan: Complex-valued generative adversarial network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, (672–681).

Wang, S., Huang, N., Zhao, T., Yang, Y., Ying, L., & Liang, D. (2017). 1d partial fourier parallel mr imaging with deep convolutional neural network. In *Proceedings of the 25st annual meeting of ISMRM, Honolulu, HI, USA*, volume 1, (2).

Wilpert, C., Neubauer, C., Rau, A., Schneider, H., Benkert, T., Weiland, E., Strecker, R., Reisert, M., Benndorf, M., Weiss, J. et al. (2023). Accelerated diffusion-weighted imaging in 3 t breast mri using a deep learning reconstruction algorithm with superresolution processing: a prospective comparative study. *Investigative Radiology*, *58*(12), 842–852.

Xiao, Z., Du, N., Liu, J., & Zhang, W. (2021). Sr-net: a sequence offset fusion net and refine net for undersampled multislice mr image reconstruction. *Computer Methods and Programs in Biomedicine*, *202*, 105997.

Xin, B., Ye, M., Axel, L., & Metaxas, D. N. (2023). Fill the k-space and refine the image: Prompting for dynamic and multi-contrast mri reconstruction. *arXiv preprint arXiv:2309.13839*.

Yaman, B., Hosseini, S. A. H., & Akçakaya, M. (2021). Zero-shot self-supervised learning for mri reconstruction. *arXiv preprint arXiv:2102.07737*.

Yan, J., Chen, S., Zhang, Y., & Li, X. (2020). Neural architecture search for compressed sensing magnetic resonance image reconstruction. *Computerized Medical Imaging and Graphics*, *85*, 101784.

Yan, J., Yu, C., Chen, H., Xu, Z., Huang, J., Li, X., & Yao, J. (2022). Seeking common ground while reserving differences: Multiple anatomy collaborative framework for undersampled mri reconstruction. *arXiv preprint arXiv:2206.07364*.

Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P. L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y. et al. (2017). Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE transactions on medical imaging*, *37*(6), 1310–1321.

Yi, Q., Liu, J., Hu, L., Fang, F., & Zhang, G. (2021). Contrastive learning for local and global learning mri reconstruction. *arXiv preprint arXiv:2111.15200*.

Zabihi, S., Rahimian, E., Asif, A., & Mohammadi, A. (2021). Sepunet: depthwise separable convolution integrated u-net for mri reconstruction. In *2021 IEEE International Conference on Image Processing (ICIP)*, (3792–3796). IEEE.

Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M. J., Defazio, A., Stern, R., Johnson, P., Bruno, M. et al. (2018). fastmri: An open dataset and benchmarks for accelerated mri. *arXiv preprint arXiv:1811.08839*.

Zhang, C., Jia, J., Yaman, B., Moeller, S., Liu, S., Hong, M., & Akçakaya, M. (2021). Instabilities in conventional multi-coil mri reconstruction with small adversarial perturbations. In *2021 55th Asilomar Conference on Signals, Systems, and Computers*, (895–899). IEEE.

Zhang, L., Li, X., & Chen, W. (2023). Camp-net: Context-aware multi-prior network for accelerated mri reconstruction. *arXiv preprint arXiv:2306.11238*.

Zhao, R., Yaman, B., Zhang, Y., Stewart, R., Dixon, A., Knoll, F., Huang, Z., Lui, Y. W., Hansen, M. S., & Lungren, M. P. (2021). fastmri+: Clinical pathology annotations for knee and brain fully sampled multi-coil mri data. *arXiv preprint arXiv:2109.03812*.

Zhao, R., Yaman, B., Zhang, Y., Stewart, R., Dixon, A., Knoll, F., Huang, Z., Lui, Y. W., Hansen, M. S., & Lungren, M. P. (2022). fastmri+, clinical pathology annotations for knee and brain fully sampled magnetic resonance imaging data. *Scientific Data*, *9*(1), 152.

Zhao, R., Zhang, Y., Yaman, B., Lungren, M. P., & Hansen, M. S. (2021). End-to-end ai-based mri reconstruction and lesion detection pipeline for evaluation of deep learning image reconstruction. *arXiv preprint arXiv:2109.11524*.

Zhao, Z., Zhang, T., Xie, W., Wang, Y., & Zhang, Y. (2022). K-space transformer for fast mrireconstruction with implicit representation. *arXiv preprint arXiv:2206.06947*.

Zhou, B., Schlemper, J., Dey, N., Salehi, S. S. M., Sheth, K., Liu, C., Duncan, J. S., & Sofka, M. (2022). Dual-domain self-supervised learning for accelerated non-cartesian mri reconstruction. *Medical Image Analysis*, *81*, 102538.

Zhou, L., Zhu, M., Xiong, D., Ouyang, L., Ouyang, Y., Chen, Z., & Zhang, X. (2023). Rnlfnet: Residual non-local fourier network for undersampled mri reconstruction. *Biomedical Signal Processing and Control*, *83*, 104632.

Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R., & Rosen, M. S. (2018). Image reconstruction by domain-transform manifold learning. *Nature*, *555*(7697), 487–492.

Ziller, A., Usynin, D., Knolle, M., Hammernik, K., Rueckert, D., & Kaissis, G. (2021). Complex-valued deep learning with differential privacy. *arXiv preprint arXiv:2110.03478*.