IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 December 2023, accepted 14 December 2023, date of publication 18 December 2023,
date of current version 27 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3344456

RESEARCH ARTICLE

Deep Learning-Based Meniscus Tear Detection From Accelerated MRI

FATMA HARMAN[1], MUSTAFA ALPER SELVER[2,5], MUSTAFA MAHMUT BARIS[3], ALI CANTURK[3], AND ILKAY OKSUZ[4]

[1]Dokuz Eylül University The Graduate School of Natural and Applied Sciences, 35160 İzmir, Turkey
[2]Department of Electrical and Electronics Engineering, Dokuz Eylül University, 35160 İzmir, Turkey
[3]Department of Radiology, Faculty of Medicine, Dokuz Eylül University, Balcova, 35320 İzmir, Turkey
[4]Department of Computer Engineering, İstanbul Technical University, 34485 İstanbul, Turkey
[5]İzmir Health Technologies Development and Accelerator (BioIzmir), Dokuz Eylül University, Balcova, 35330 İzmir, Turkey

Corresponding author: Fatma Harman (fatmaharman89@gmail.com)

ABSTRACT Magnetic Resonance Imaging (MRI) is an exceptional diagnostic tool known for its ability to provide superior soft tissue contrast. Despite its proven efficacy, traditional MRI methods have some inherent limitations, including relatively longer scan times and the requirement for specialized expertise in data analysis. These factors can, at times, create challenges in the widespread adoption of MRI in certain research and clinical scenarios. Deep learning (DL) methods for MRI reconstruction and analysis offer a promising solution to address this issue. While these DL techniques have been validated using standard image quality metrics, they fall short in assessing clinically relevant details. This hinders their clinical reliability and practical application. Although radiologists have previously compared the diagnostic equivalence of accelerated DL-reconstructed images to conventional ones for evaluating knee internal derangement, it remains uncertain whether DL detectors can accurately identify clinically important details when reconstruction models are used. Furthermore, the artifacts or hallucinations generated by DL reconstructions in knee MRI have not been examined in a clinical setting. Thus, this study aims to determine the performance of DL detectors on retrospectively accelerated DL-reconstructed knee MRI, comparing them to conventional imaging and expert evaluations for detecting meniscal tears. Our investigation yields three significant contributions. First, an in-depth analysis of DL reconstruction highlights the presence of hallucinations in the femur, tibia, and false positive artifacts, indicating that the overall reconstruction quality does not directly affect pathological features. Second, the results of DL detectors demonstrate that their performance aligns well with image quality assessment metrics and expert scores. This finding validates the reliability of the detection outcomes. Finally, we propose an integrated (i.e. reconstruction + detection) process for meniscal tears on fastMRI+ data and achieved state-of-the-art results with average precision scores of 0.69 and 0.67 at 4− and 8-fold accelerations, respectively.

INDEX TERMS MRI reconstruction, meniscal tear detection, and deep learning.

I. INTRODUCTION

Magnetic Resonance Imaging (MRI) is a powerful diagnostic tool known for its exceptional soft tissue contrast, making it invaluable in musculoskeletal diagnosis [1]. 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 [2]. Despite their reported success based

The associate editor coordinating the review of this manuscript and approving it for publication was Vishal Srivastava.

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. [3] 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. [4] 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 [5]. Although Machine Learning (ML), especially DL models, shows promise, their outcomes still fall short when compared to the inter-expert variations [6]. Furthermore, the prospective implementation of these models in clinical settings is very rare [7].

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 [8]. 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 [9]. 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 [10]. 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 [11].

In the realm of fastMRI reconstruction, various DL models with diverse network topologies have been proposed and evaluated [12], [13], [14], [15], [16], [17], [18], [19], [20]. These models primarily aim for high-quality reconstruction from accelerated MRI acquisition, often incorporating thorough evaluations involving expert readings [8], [21]. Recent advancements showcase the potential of DL reconstruction in significantly reducing scan time while enhancing image quality for knee MRI [22]. However, these methods predominantly focus on image reconstruction and lack integration for automatic disease detection.

Among these architectures, the UNET model [9] 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) [23] 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) [12], [15], 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 [24]. This tool has demonstrated outstanding performance across various medical object detection tasks on well-known public datasets, including LUNA16 [25], ADAM [26], CADA [27], RibFrac [28], LIDC-IDRI [29], Kits19 [30], and others. Notably, it has excelled in Decathlon tasks like Pancreas, Prostate, and Hepatic Vessel [31]. 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 [9], [12], [23]. 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 3). Building on the analyses conducted, we assess the performance of a state-of-the-art DL detector, nnDetection [24], on retrospectively accelerated 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 paper is structured as follows. In Section II, we discuss the fastMRI dataset and its constraints. Section III presents the background of MRI reconstruction and reviews related studies. The materials and methods employed for reconstruction and analysis are described in Section IV. In Section V, we present the image quality metrics and expert evaluations. The results and analysis are presented in Section VI. Finally, in Section 7, we draw conclusions and highlight avenues for future research.

II. 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 [32]. Despite their success, these models often overlook important factors like acquisition variability, orientation, pathology, and artifacts. For instance, recent research by Patsanis et al. [33] 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. [34] demonstrated superior performance in fetal brain MRI reconstruction, detection of pathological brains, and expert reader quality assessments. Baalbaki et al. [35] proposed the identification of atherosclerosis in reconstructed images.

In the realm of musculoskeletal imaging, several studies have addressed specific pathologies. Dung et al. [36] and Ni et al. [37] proposed approaches for classifying anterior cruciate ligament tears and diagnosing anterior talofibular ligament tears, respectively, in 2023. In a different study, Zhao et al. [11] 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. [38] proposed an adaptive Convolutional Neural Network (CNN) model for k-space data interpolation by integrating complementary information of contiguous slices, while Huang et al. [39] 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 [40], [41], [42].

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 [8], [9], [43]. This dataset played a pivotal role in advancing ML-based MRI reconstruction tasks [21]. 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. [44] 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 [4].

In parallel with these developments, the Osteoarthritis Initiative (OAI) [45] and MRNet [46] have been instrumental in classifying knee abnormalities in 2006 and 2018, respectively. More recently, Nasser et al. [47] introduced a discriminative model to enhance classification performance, while Hu et al. [48] explored the severity grade of longitudinal osteoarthritis using an adversarial evolved neural network in the OAI dataset. Dai et al. [49] combined transformer and CNN for multi-modal classification, and Kara and Hardalaç [50] 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 [10]. 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. [51] focusing on determining the optimal k-space undersampling pattern for effective localization or segmentation of pathology in a downstream manner.

TABLE 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

III. 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 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.

A. DATASET

The fastMRI dataset, as extensively described in [9] (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 [10], 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.

FIGURE 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).

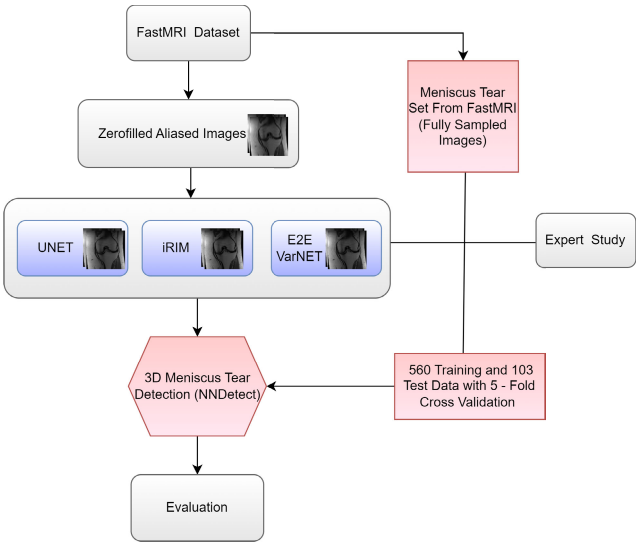


FIGURE 2. The study design through the model training and testing phases. Initially, we undersampled k-space and generated zero-filled aliased images. Subsequently, three distinct reconstruction models were employed to obtain high-quality images. For the evaluation of meniscus tear, expert study and 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.

FastMRI+ [10], 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 1 details the categorization of meniscal tears and artifacts, with a visual representation in Figure 1.

B. 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 [9], [52], [53]. 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 3×3 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 layer with two strides in the down-sampling path and a bilinear upsampling layer in the up-sampling path, doubling the resolution between blocks [9].
  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 [23], [54]. 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 [12].

Variational networks were originally designed using generalized compressed sensing methods within an unrolled gradient descent scheme [15]. In these methods, all parameters, including prior models specified by kernel filters and activation functions, were learned offline during training [15]. 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.

C. DETECTION

FastMRI+ [10], 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 [24], 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 [55]. 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 [56], where domain knowledge is divided into fixed, rule-based, and empirical parameter forms.

1) 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 [55]. The key elements of the architecture, such as pool strides, pooling, and kernel sizes, were adapted through rule-based parameters.

2) 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.

3) 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 [55] 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 [25], ADAM [26], CADA [27], RibFrac [28], LIDC-IDRI [29], Kits19 [30], and 11 other tasks such as Decathlon [31]-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

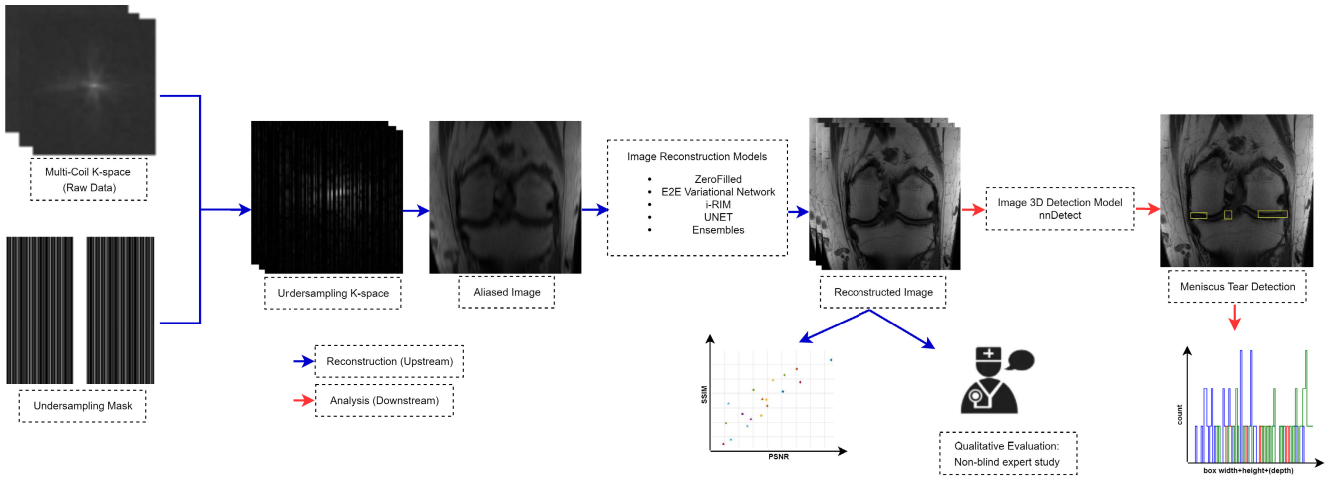


FIGURE 3. An overview of the comprehensive multi-coil MRI reconstruction, detection, and analysis pipeline. Initially, the acquired data are undersampled using cartesian trajectories, resulting in an aliased image. Subsequently, the aliased image is processed through three different reconstruction models (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).

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 3. The corresponding figure represents each step of the comprehensive multi-coil MRI reconstruction and detection pipeline together with expert analysis.

IV. EVALUATION METRICS

A. RECONSTRUCTION METRICS

Two well-established reconstruction metrics were computed for evaluation. One of these metrics is the Peak Signal-to-Noise Ratio (PSNR), which assesses the local differences. A higher PSNR indicates better image quality in the reconstructed image. It can be defined in Equation 1 as:

$$PSNR(x, y) = 10 \log_{10}[(Max.(x))^2 / MSE(x, y)] \quad (1)$$

The second quality metric used is the Structural Similarity Index (SSIM), which quantifies the similarity between two images by considering the interdependencies between neighboring pixels [9]. The SSIM value was computed using the Equation 2:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

μ_x and μ_y represent the average pixel intensities in the images x and y, while μ_x^2 and μ_y^2 denote their respective variances. The covariance between x and y is represented by σ_{xy}, and C_1 and C_2 are constants used to stabilize the division in the SSIM calculation. The resulting SSIM values range from 0 to 1, with a value of 1 indicating a perfect match between the reconstructed image and the ground truth.

B. 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, IOU(gt, pd) = area(gt $\cap$ pd)/area(gt $\cup$ 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 [57].

V. 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.

A. 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 2.

TABLE 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 (Acq.) accelerated (Acc.) at 4x and 8x.

Acc.	Modality	Acq.	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 [58] and segmentation competitions [59], [60], we investigated their potential in the context of MRI reconstruction. Employing three fundamental combination methods [61] (Min/Max Combiner, Averaging Combiner, and Weighted Averaging), we evaluated their efficacy in improving the reconstruction process. The results, as presented in Table 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).

B. 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. 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 showcases the qualitative results of this detector at both acceleration rates. It becomes

TABLE 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 (Acq.) accelerated (Acc.) at 4x and 8x.

Acc.	Modality	Acq.	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)

evident that at a 4-fold acceleration, the detector successfully identifies the meniscus tear across all reconstruction outputs except ZeroFilled. However, at 8-fold acceleration, the detector exclusively detects the tear within the E2E VarNET reconstructions. 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 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 5, incontrovertibly establish our proposed method's superior performance compared to Zhao et al.'s approach in meniscus tear detection.

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).

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

Acc.	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**

TABLE 5. The proposed framework for meniscus tear detection compared with existing methods available for this task.

Author	Acc.	Modality	Precision	Recall
Zhao et al. [11] (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 et al. [11] (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**

The conclusive results of this extensive assessment, presented in Table 5, undeniably demonstrate the superior performance achieved by the proposed method in meniscus tear detection.

C. 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. [21] introduced

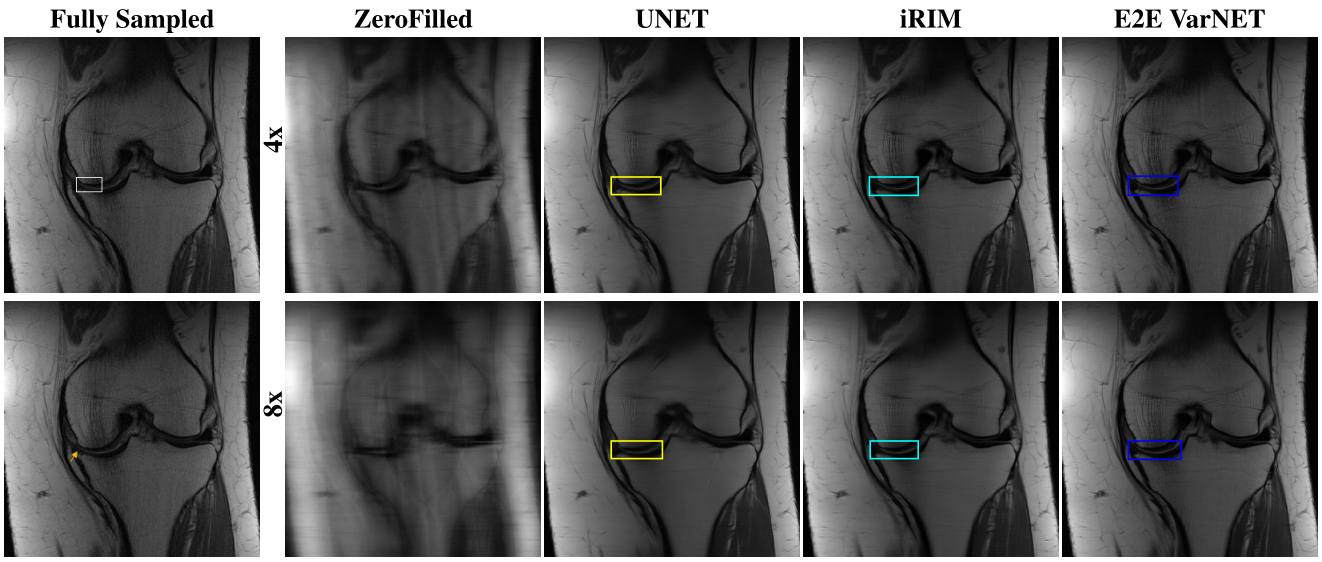


FIGURE 4. A demonstration of the nnDetection outcomes represented by bounding boxes for different reconstructions. At 4-fold acceleration (top row), the detector successfully identified the meniscus tear across all reconstruction methods except for ZeroFilled. However, at 8-fold acceleration (bottom row), the detector only recognized the meniscus tear through E2E VarNET. 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.

a radiologists' ranking system based on the quality of pathology depiction and identified hallucinations in the fastMRI brain dataset challenge. Knoll et al. [8], on the other hand, ranked radiologists' submissions in the fastMRI knee challenge by considering both artifacts and Image 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.

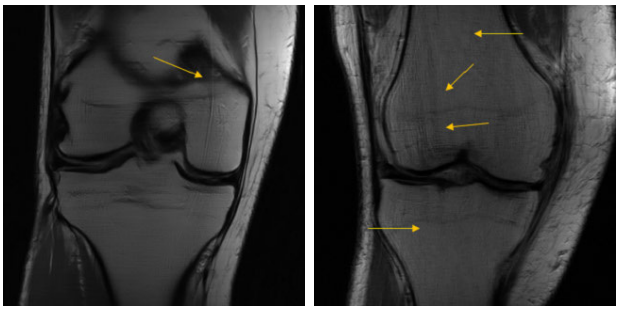
The expert study involves qualitative evaluation of the original and DL-reconstructed images. Two radiologists, one with with 20-years of experience (referred as M.M.B.) and the other with 5-years of experience (referred as A.C.), participated to our study. They meticulously observed and documented the hallucinatory effects and artifacts introduced by the reconstruction models. The final conclusions are established through their mutual agreement. Examples of these hallucinations are presented in Figure 5, 6, and 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 5 illustrates scattered linear artifacts surrounding the femur, with some appearing in a fogging-style Figure 5 (b). Moreover, linear artifacts were also evident in the tibia, proximal tibia, and fibula, as shown in Figure 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 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.

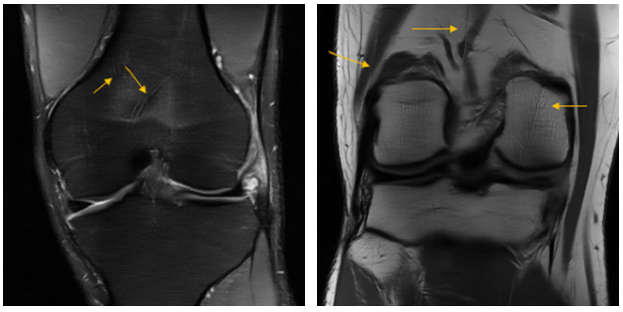
These indicate that, DL may not be fully optimized to handle hallucinations despite their high performance in terms quantitative metrics. Moreover, adversarial perturbations have demonstrated the instability of DL models [62]. These hallucinations and artifacts persist as significant challenges across 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 8, showing the total number of artifacts.

Figure 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 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,

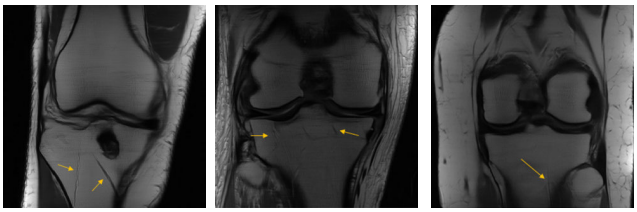


(a) Artifacts in the medial femur (b) Fogging-style artifact in the distal femur

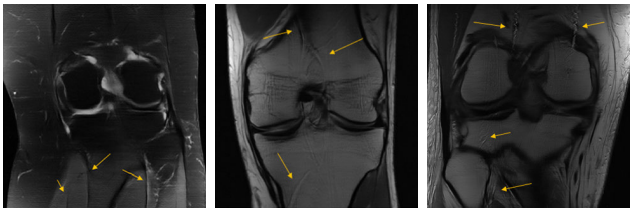


(c) Artifacts in the distal femur (d) Lineal artifacts in the femur

FIGURE 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.



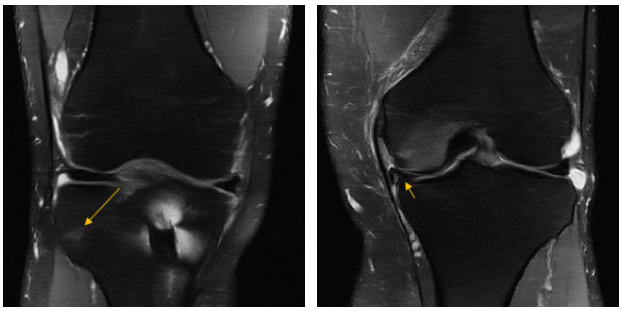
(a) Lineal artifacts in the tibia (b) Minimal hypointense lineal artifacts in the proximal tibia (c) Artifacts in the fibula.



(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 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.

(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 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.

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 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 9, indicate that E2E VarNET showed the best performance at 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

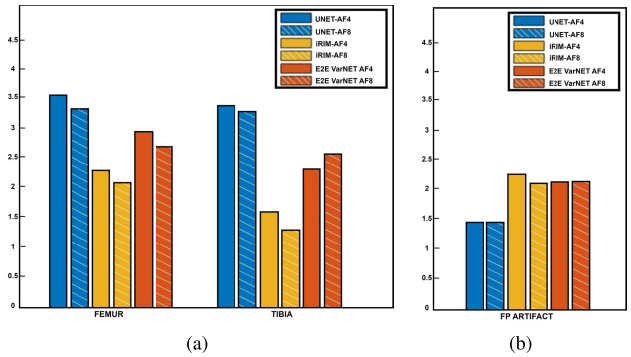


FIGURE 8. The hallucinations observed in the femur, tibia, and FP artifacts were carefully diagnosed and counted slice by slice by an expert. The counts were then scaled logarithmically using the formula (1+log10(count)). The resulting counts are presented as bar plots for both acceleration rates. The 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.

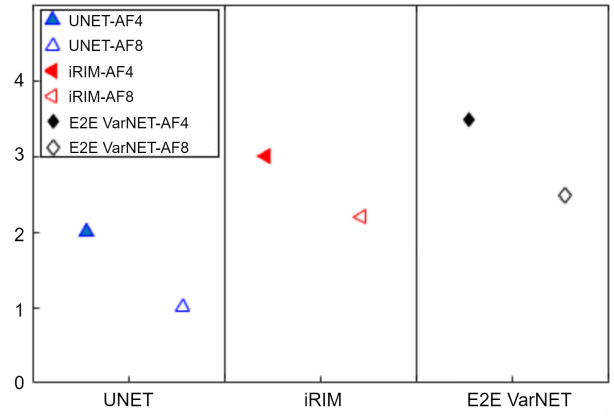


FIGURE 9. The plot illustrates the average diagnostic scores from radiological assessments for the reconstructed images. The scoring system is summarized as follows: 1 for far worse, 2 for worse, 3 for alike, 4 for better, and 5 for far better compared to the reference image. The signs on the plot represent the calculated average of expert scores.

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 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 evidenced by both the FROC and mAP scores. The expert evaluations of meniscal tear pathology are consistent with the results obtained from

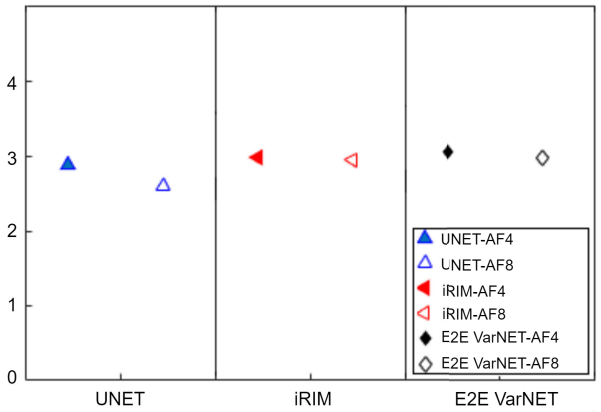


FIGURE 10. The plot displays the average scores obtained from radiological evaluations of pathological features in the reconstructed images. The scoring system used is as follows: 0 indicates that the pathology is not observable, 1 for far worse, 2 for worse, 3 for similar, and 4 for better compared to the reference image.

the DL models, validating the reliability of the detection outcomes.

VI. 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 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 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.

REFERENCES

[1] E. J. van Beek, C. Kuhl, Y. Anzai, P. Desmond, R. L. Ehman, Q. Gong, G. Gold, V. Gulani, M. Hall-Craggs, and T. Leiner, "Value of MRI in medicine: More than just another test?" *J. Magn. Reson. Imag.*, vol. 49, no. 7, pp. 14–25, 2019.

[2] P. M. Johnson, A. Tong, A. Donthireddy, K. Melamud, R. Petrocelli, P. Smereka, K. Qian, M. B. Keerthivasan, H. Chandarana, and F. Knoll, "Deep learning reconstruction enables highly accelerated biparametric MR imaging of the prostate," *J. Magn. Reson. Imag.*, vol. 56, no. 1, pp. 184–195, Jul. 2022.

[3] K.-L. Lee, D. A. Kessler, S. Dezonie, W. Chishaya, C. Shepherd, B. Carmo, M. J. Graves, and T. Barrett, "Assessment of deep learning-based reconstruction on T2-weighted and diffusion-weighted prostate MRI image quality," *Eur. J. Radiol.*, vol. 166, Sep. 2023, Art. no. 111017.

[4] A. D. Desai, A. M. Schmidt, E. B. Rubin, C. M. Sandino, M. S. Black, V. Mazzoli, K. J. Stevens, R. Boutin, C. Re, and G. E. Gold, "SKM-tea: A dataset for accelerated MRI reconstruction with dense image labels for quantitative clinical evaluation," 2022, *arXiv:2203.06823*.

[5] S. Gassenmaier, T. Küstner, D. Nickel, J. Herrmann, R. Hoffmann, H. Almansour, S. Afat, K. Nikolaou, and A. E. Othman, "Deep learning applications in magnetic resonance imaging: Has the future become present?" *Diagnostics*, vol. 11, no. 12, p. 2181, Nov. 2021.

[6] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444.

[7] A. S. Chaudhari, C. M. Sandino, E. K. Cole, D. B. Larson, G. E. Gold, S. S. Vasanawala, M. P. Lungren, B. A. Hargreaves, and C. P. Langlotz, "Prospective deployment of deep learning in MRI: A framework for important considerations, challenges, and recommendations for best practices," *J. Magn. Reson. Imag.*, vol. 54, no. 2, pp. 357–371, Aug. 2021.

[8] F. Knoll, T. Murrell, A. Sriram, N. Yakubova, J. Zbontar, M. Rabbat, A. Defazio, M. J. Muckley, D. K. Sodickson, C. L. Zitnick, and M. P. Recht, "Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge," *Magn. Reson. Med.*, vol. 84, no. 6, pp. 3054–3070, Dec. 2020.

[9] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, and M. Bruno, "fastMRI: An open dataset and benchmarks for accelerated MRI," 2018, *arXiv:1811.08839*.

[10] R. Zhao, B. Yaman, Y. Zhang, R. Stewart, A. Dixon, F. Knoll, Z. Huang, Y. W. Lui, M. S. Hansen, and M. P. Lungren, "fastMRI+: Clinical pathology annotations for knee and brain fully sampled multi-coil MRI data," 2021, *arXiv:2109.03812*.

[11] R. Zhao, Y. Zhang, B. Yaman, M. P. Lungren, and M. S. Hansen, "End-to-end AI-based MRI reconstruction and lesion detection pipeline for evaluation of deep learning image reconstruction," 2021, *arXiv:2109.11524*.

[12] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson, "End-to-end variational networks for accelerated MRI reconstruction," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Lima, Peru: Springer, 2020.

[13] A. Sriram, J. Zbontar, T. Murrell, C. L. Zitnick, A. Defazio, and D. K. Sodickson, "GrappaNet: Combining parallel imaging with deep learning for multi-coil MRI reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14303–14310.

[14] H. Kerstin, S. Jo, Q. Chen, D. Jingming, and R. Daniel, "Σ-Net systematic evaluation of iterative deep neural networks for fast parallel MR image reconstruction," 2019, *arXiv:1912.09278*.

[15] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, Jun. 2018.

[16] N. Pezzotti, S. Yousefi, M. S. Elmahdy, J. Van Gemert, C. Schülke, M. Doneva, T. Nielsen, S. Kastryulin, B. P. Lelieveldt, and M. J. Van Osch, 'An adaptive intelligence algorithm for undersampled knee MRI reconstruction: Application to the 2019 fastMRI challenge," 2020, *arXiv:2004.07339*.

[17] Z. Ramzi, P. Ciuciu, and J.-L. Starck, "Benchmarking MRI reconstruction neural networks on large public datasets," *Appl. Sci.*, vol. 10, no. 5, p. 1816, Mar. 2020.

[18] X. Liu, Y. Pang, R. Jin, Y. Liu, and Z. Wang, "Dual-domain reconstruction network with V-Net and K-Net for fast MRI," *Magn. Reson. Med.*, vol. 88, no. 6, pp. 2694–2708, Dec. 2022.

[19] C. Tong, Y. Pang, and Y. Wang, "HIWDNet: A hybrid image-wavelet domain network for fast magnetic resonance image reconstruction," *Comput. Biol. Med.*, vol. 151, Dec. 2022, Art. no. 105947.

[20] A. Pal and Y. Rathi, "A review and experimental evaluation of deep learning methods for MRI reconstruction," *Mach. Learn. Biomed. Imag.*, vol. 1, pp. 1–50, Mar. 2022.

[21] M. J. Muckley, B. Riemenschneider, A. Radmanesh, S. Kim, G. Jeong, J. Ko, Y. Jun, H. Shin, D. Hwang, and M. Mostapha, "Results of the 2020 fastMRI challenge for machine learning MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2306–2317, Sep. 2021.

[22] P. M. Johnson, D. J. Lin, J. Zbontar, C. L. Zitnick, A. Sriram, M. Muckley, J. S. Babb, M. Kline, G. Ciavarra, E. Alaia, M. Samim, W. R. Walter, L. Calderon, T. Pock, D. K. Sodickson, M. P. Recht, and F. Knoll, "Deep learning reconstruction enables prospectively accelerated clinical knee MRI," *Radiology*, vol. 307, no. 2, Apr. 2023, Art. no. 220425.

[23] P. Putzky and M. Welling, "Invert to learn to invert," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.

[24] M. Baumgartner, P. F. Jäger, F. Isensee, and K. H. Maier-Hein, "nnDetection: A self-configuring method for medical object detection," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Strasbourg, France: Springer, 2021, pp. 530–539.

[25] A. A. A. Setio, A. Traverso, T. De Bel, M. S. Berens, C. Van Den Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, and B. Geurts, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, Dec. 2017.

[26] K. Timmins, E. Bennink, I. van der Schaaf, B. Velthuis, Y. Ruigrok, and H. Kuijf, "Intracranial aneurysm detection and segmentation challenge," in *Proc. 23rd Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*. Lima, Peru: Zenodo, Oct. 2020, doi: 10.5281/zenodo.3715848.

[27] A. Hennemuth, L. Goubergrits, M. Ivantsits, and J.-M. Kuhnigk, *Cerebral Aneurysm Detection and Analysis*. Berlin, Germany: Springer, 2021.

[28] L. Jin, J. Yang, K. Kuang, B. Ni, Y. Gao, Y. Sun, P. Gao, W. Ma, M. Tan, H. Kang, J. Chen, and M. Li, "Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet," *eBioMedicine*, vol. 62, Dec. 2020, Art. no. 103106.

[29] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, and E. A. Hoffman, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, Feb. 2011.

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

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

[32] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. D. Laak, B. Van Ginneken, and C. I. Sanchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[33] A. Patsanis, M. R. S. Sunoqrot, S. Langørgen, H. Wang, K. M. Selnæs, H. Bertilsson, T. F. Bathen, and M. Elschot, "A comparison of generative adversarial networks for automated prostate cancer detection on T2-weighted MRI," *Informat. Med. Unlocked*, vol. 39, 2023, Art. no. 101234.

[34] M. Ebner, G. Wang, W. Li, M. Aertsen, P. A. Patel, R. Aughwane, A. Melbourne, T. Doel, S. Dymarkowski, P. De Coppi, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI," *NeuroImage*, vol. 206, Feb. 2020, Art. no. 116324.

[35] J. M. Baalbaki, S. S. El Kadri, K. A. Awad, C. H. Al Aridy, M. S. Aridi, M. A. El Zein, M. O. Diab, M. M. Sabbah, R. N. Zantout, and Y. A. A. Adla, "An end-to-end deep learning approach for atherosclerosis detection using autoencoders," in *Proc. 5th Int. Conf. Bio-Eng. Smart Technol. (BioSMART)*, 2023, pp. 1–4.

[36] N. T. Dung, N. H. Thuan, T. Van Dung, L. Van Nho, N. M. Tri, V. P. T. Vy, L. N. Hoang, N. T. Phat, D. A. Chuong, and L. H. Dang, "End-to-end deep learning model for segmentation and severity staging of anterior cruciate ligament injuries from MRI," *Diagnostic Interventional Imag.*, vol. 104, no. 3, pp. 133–141, Mar. 2023.

[37] M. Ni, W. Chen, Q. Zhao, Y. Zhao, and H. Yuan, "Deep learning approach for MRI in the classification of anterior talofibular ligament injuries," *J. Magn. Reson. Imag.*, vol. 58, no. 5, pp. 1544–1556, Nov. 2023.

[38] T. Du, H. Zhang, Y. Li, S. Pickup, M. Rosen, R. Zhou, H. K. Song, and Y. Fan, "Adaptive convolutional neural networks for accelerating magnetic resonance imaging via *k*-space data interpolation," *Med. Image Anal.*, vol. 72, Aug. 2021, Art. no. 102098.

[39] J. Huang, S. Wang, G. Zhou, W. Hu, and G. Yu, "Evaluation on the generalization of a learned convolutional neural network for MRI reconstruction," *Magn. Reson. Imag.*, vol. 87, pp. 38–46, Apr. 2022.

[40] F. Ong, S. Amin, S. Vasanawala, and M. Lustig, "mridata.org: An open archive for sharing MRI raw data," in *Proc. Int. Soc. Magn. Reson. Med.*, vol. 26, 2018.

[41] M. Z. Darestani, A. S. Chaudhari, and R. Heckel, "Measuring robustness in deep learning based compressive sensing," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 2433–2444.

[42] CUBE FSE'XL, PD PD, and FAT FAT, "Creation of fully sampled MR data repository for compressed sensing of the knee," in *Proc. SMRT Conf.*, Salt Lake City, UT, USA, 2013.

[43] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, "FastMRI: A publicly available raw K-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning," *Radiol., Artif. Intell.*, vol. 2, no. 1, Jan. 2020, Art. no. e190007.

[44] C. Peng, P. Guo, S. K. Zhou, V. M. Patel, and R. Chellappa, "Towards performant and reliable undersampled mr reconstruction via diffusion model sampling," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2022, pp. 623–633.

[45] M. Nevitt, D. Felson, and G. Lester, "The osteoarthritis initiative," *Protocol Cohort Study*, vol. 1, pp. 1–155, 2006.

[46] N. Bien, P. Rajpurkar, R. L. Ball, J. Irvin, A. Park, E. Jones, M. Bereket, and B. N. Patel, "Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet," *PLOS Med.*, vol. 15, no. 11, Nov. 2018, Art. no. e1002699.

[47] Y. Nasser, R. Jennane, A. Chetouani, E. Lespessailles, and M. E. Hassouni, "Discriminative regularized auto-encoder for early detection of knee OsteoArthritis: Data from the osteoarthritis initiative," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2976–2984, Sep. 2020.

[48] K. Hu, W. Wu, W. Li, M. Simic, A. Zomaya, and Z. Wang, "Adversarial evolving neural network for longitudinal knee osteoarthritis prediction," *IEEE Trans. Med. Imag.*, vol. 41, no. 11, pp. 3207–3217, Nov. 2022.

[49] Y. Dai, Y. Gao, and F. Liu, "TransMed: Transformers advance multi-modal medical image classification," *Diagnostics*, vol. 11, no. 8, p. 1384, Jul. 2021.

[50] A. C. Kara and F. Hardalaç, "Detection and classification of knee injuries from MR images using the MRNet dataset with progressively operating deep learning methods," *Mach. Learn. Knowl. Extraction*, vol. 3, no. 4, pp. 1009–1029, Dec. 2021.

[51] A. Razumov, O. Y. Rogov, and D. V. Dylov, "Optimal MRI undersampling patterns for pathology localization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2022, pp. 768–779.

[52] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Munich, Germany: Springer, 2015, pp. 234–241.
[53] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
[54] P. Putzky, D. Karkalousos, J. Teuwen, N. Miriakov, B. Bakker, M. Caan, and M. Welling, "i-RIM applied to the fastMRI challenge," 2019, *arXiv:1910.08952*.
[55] P. F. Jaeger, S. A. Kohl, S. Bickelhaupt, F. Isensee, T. A. Kuder, H.-P. Schlemmer, and K. H. Maier-Hein, "Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection," in *Proc. Mach. Learn. Health Workshop*, 2020, pp. 171–183.
[56] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021.
[57] A. I. Bandos, H. E. Rockette, T. Song, and D. Gur, "Area under the free-response ROC curve (FROC) and a related summary index," *Biometrics*, vol. 65, no. 1, pp. 247–256, Mar. 2009.
[58] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2014.
[59] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, and D. Rueckert, "Ensembles of multiple models and architectures for robust brain tumour segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Quebec City, QC, Canada: Springer, 2018, pp. 450–462.
[60] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, and S. Wirkert, "nnU-Net: Self-adapting framework for U-Net-based medical image segmentation," 2018, *arXiv:1809.10486*.
[61] A. E. Kavur, L. I. Kuncheva, and M. A. Selver, "Basic ensembles of vanilla-style deep learning models improve liver segmentation from ct images," in *Convolutional Neural Networks for Medical Image Processing Applications*. Boca Raton, FL, USA: CRC Press, 2022, pp. 52–74.
[62] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 48, pp. 30088–30095, Dec. 2020.

FATMA HARMAN received the B.Sc. degree in electrical and electronics engineering from Balikesir University, in 2011, and the M.Sc. degree in electrical and electronics engineering from Manisa Celal Bayar University, in 2016. She is currently pursuing the Ph.D. degree in electrical and electronics engineering with Dokuz Eylül University, İzmir. Her current research interests include machine learning, magnetic resonance imaging, reconstruction, and analysis.

MUSTAFA ALPER SELVER received the B.Sc. degree in electrical and electronics engineering (EEE) from Gazi University, Ankara, Turkey, in 2002, and the M.Sc. and Ph.D. degrees in EEE from Dokuz Eylül University (DEU), İzmir, Turkey, in 2005 and 2010, respectively. He has studied with the FH-Jülich, Medical Informatics Laboratory, Aachen, Germany, from 2004 to 2005, and the Heffner Biomedical Imaging Laboratory, Columbia University, New York, NY, USA, from January 2009 to August 2009. He was a Coordinator/a Researcher for several TUBITAK Research Projects and an Advisor for various TUBITAK-TEYDEB Industrial Projects, including the Turkeys First European Union Project having Euripidies2 Label and the SIL Certified Advanced Integrated Data Recording and Analysis System for Railway Transportation Systems (ADORAS), which was conducted in partnership with South Korea, Czech Republic, and Spain. He is currently an Associate Professor of EEE with DEU. His research interests include multiscale and hierarchical data analysis and the development of ensemble models for industrial and medical, particularly radiological, and signal/image processing applications.

MUSTAFA MAHMUT BARIS received the M.D. degree from Ege University, İzmir, Turkey, in 2004, the Radiology Specialist degree from Dokuz Eylü University (DEU), İzmir, the European Board of Cardiovascular Radiology (EBCR) Diploma degree, in 2016, and the European Diploma in Radiology (EDiR) degree, in 2017. He studied cardiac imaging with the University of Vienna, Vienna, Austria, in 2008. He finished his Radiology Residency, in 2010. He studied head and neck tumors with Columbia University, New York, NY, USA, from 2018 to 2019. He was a Researcher in many TUBITAK research projects. He is currently an Associate Professor with the DEU Radiology Department. His research interests include the use of artificial intelligence in radiology, researches with laboratory animals, and the use of microbubbles.

ALI CANTURK received the M.D. degree from İstanbul University Çapa Medicine Faculty, İstanbul, in 2018, and the Radiology Residency degree from Dokuz Eylül University (DEU), İzmir, Turkey, in 2023. His research interests include musculoskeletal radiology, interventional radiology, and artificial intelligence.

ILKAY OKSUZ received the B.Sc. degree in electronics engineering from İstanbul Technical University, the M.Sc. degree in electrical and electronics engineering from Bahcesehir University, and the Ph.D. degree from the Department of Computer, Decision, and Systems Science, IMT School for Advanced Studies Lucca. He was a Research Associate with the Biomedical Engineering Department, King's College London, from 2017 to 2019. He has been an Assistant Professor with the Computer Engineering Department, İstanbul Technical University, since 2020. His work focuses on machine learning with a particular interests in medical image quality assessment, reconstruction, and analysis.