Image Quality and Diagnostic Performance of Accelerated Shoulder MRI With Deep Learning–Based Reconstruction

Seok Hahn, MD[1], Jisook Yi, MD[1], Ho-Joon Lee, MD[1], Yedaun Lee, MD[1], Yun-Jung Lim, MD[1], Jin-Young Bang, MD[2], Hyunwoong Kim, MS[3], Joonsung Lee, PhD[4]

Musculoskeletal Imaging · Original Research

Keywords

deep learning, glenoid labrum, MRI, rotator cuff, shoulder

Submitted: Jul 14, 2021
Revision requested: Jul 26, 2021
Revision received: Aug 26, 2021
Accepted: Sep 8, 2021
First published online: Sep 15, 2021

J. Lee is an employee of GE Healthcare Korea. The remaining authors declare that they have no disclosures relevant to the subject matter of this article.

An electronic supplement is available online at doi.org/10.2214/AJR.21.26577.

BACKGROUND. Shoulder MRI using standard multiplanar sequences requires long scan times. Accelerated sequences have tradeoffs in noise and resolution. Deep learning–based reconstruction (DLR) may allow reduced scan time with preserved image quality.

OBJECTIVE. The purpose of this study was to compare standard shoulder MRI sequences and accelerated sequences without and with DLR in terms of image quality and diagnostic performance.

METHODS. This retrospective study included 105 patients (45 men, 60 women; mean age, 57.6 ± 10.9 [SD] years) who underwent a total of 110 3-T shoulder MRI examinations. Examinations included standard sequences (scan time, 9 minutes 23 seconds) and accelerated sequences (3 minutes 5 seconds; 67% reduction), both including fast spin-echo sequences in three planes. Standard sequences were reconstructed using the conventional pipeline; accelerated sequences were reconstructed using both the conventional pipeline and a commercially available DLR pipeline. Two radiologists independently assessed three image sets (standard sequence, accelerated sequence without DLR, and accelerated sequence with DLR) for subjective image quality and artifacts using 4-point scales (4 = highest quality) and identified pathologies of the subscapularis tendon, supraspinatus-infraspinatus tendon, long head of the biceps brachii tendon, and glenoid labrum. Interobserver agreement and agreement between image sets for the evaluated pathologies were assessed using weighted kappa statistics. In 27 patients who underwent arthroscopy, diagnostic performance was calculated using arthroscopic findings as a reference standard.

RESULTS. Mean subjective image quality scores for readers 1 and 2 were 10.6 ± 1.2 and 10.5 ± 1.4 for the standard sequence, 8.1 ± 1.3 and 7.2 ± 1.1 for the accelerated sequence without DLR, and 10.7 ± 1.2 and 10.5 ± 1.6 for the accelerated sequence with DLR. Mean artifact scores for readers 1 and 2 were 9.3 ± 1.2 and 10.0 ± 1.0 for the standard sequence, 7.3 ± 1.3 and 9.1 ± 0.8 for the accelerated sequence without DLR, and 9.4 ± 1.2 and 9.8 ± 0.8 for the accelerated sequence with DLR. Interobserver agreement ranged from kappa of 0.813–0.951 except for accelerated sequence without DLR for the supraspinatus-infraspinatus tendon (κ = 0.673). Agreement between image sets ranged from kappa of 0.809–0.957 except for reader 1 for supraspinatus-infraspinatus tendon (κ = 0.663–0.700). Accuracy, sensitivity, and specificity for tears of the four structures were not different ($p > .05$) among image sets.

CONCLUSION. Accelerated sequences with DLR provide 67% scan time reduction with similar subjective image quality, artifacts, and diagnostic performance to standard sequences.

CLINICAL IMPACT. Accelerated sequences with DLR may provide an alternative to standard sequences for clinical shoulder MRI.

MRI is recommended for the evaluation of shoulder pain owing to its excellent depiction of soft-tissue structures and is considered the reference standard diagnostic modality in various shoulder diseases [1]. MRI can be used to evaluate rotator cuff tears, glenoid labral tears and instability, and biceps tendon dislocations or tears, as well as to provide postoperative evaluation of the soft-tissue structures of the shoulder joint [1–4]. Shoulder MRI requires high-resolution sequences in multiple planes. This is usually achieved

[1]Department of Radiology, Inje University College of Medicine, Haeundae Paik Hospital, Haeundae-ro 875, Busan, 48108, South Korea. Address correspondence to J. Yi (jsyi2010@gmail.com).

[2]Department of Orthopedic Surgery, Inje University College of Medicine, Haeundae Paik Hospital, Busan, South Korea.

[3]Clinical Trial Center, Inje University College of Medicine, Haeundae Paik Hospital, Busan, South Korea.

[4]GE Healthcare Korea, Seoul, South Korea.

doi.org/10.2214/AJR.21.26577

AJR 2022; 218:506–516

ISSN-L 0361–803X/22/2183–506

in standard clinical protocols through multiplanar 2D fast spin-echo (FSE) sequences. Standard 2D FSE sequences provide sharp soft-tissue contrast and high in-plane spatial resolution, though require relatively long scan times and can be affected by motion artifacts. Alternative sequences apply various methods (e.g., k-space undersampling, decrease in number of averages, decrease in matrix size, parallel imaging) to reduce scan time and motion artifacts. However, these alternate sequences have tradeoffs in image quality. For example, use of parallel imaging or a decrease in averages results in a noisy image, and a decrease in matrix size results in a low-resolution image. Nonetheless, studies show that the image quality losses of more rapid sequences are tolerable and that diagnostic performance for assessment of shoulder pathologies is comparable to that of standard multiplanar 2D sequences [5–7].

A deep learning–based reconstruction (DLR) pipeline (AIR Recon DL, GE Healthcare) has recently become commercially available that overcomes the trade-off between scan time and image quality of conventional reconstruction pipelines, providing high-fidelity images with reduced noise levels [8]. This pipeline is readily applied to 2D sequences and uses a noise reduction factor that users may modify. A recent study using the prototype version of AIR Recon DL for pituitary MRI reported improved diagnostic performance compared with conventional reconstruction for residual pituitary adenoma [9]. DLR applied to cardiac MRI resulted in improved quality compared with conventional reconstruction of late gadolinium enhancement images [10].

A small number of studies have applied DLR to musculoskeletal MRI [11, 12], though to our knowledge no studies have investigated applications of DLR to shoulder MRI in a clinical setting. DLR could facilitate use of accelerated shoulder MRI protocols by maintaining such protocols' image quality in a reduced scan time. The aim of this study was to compare standard shoulder MRI sequences and accelerated sequences without and with DLR in terms of image quality and diagnostic performance.

Methods

Participants

This retrospective single-center study was approved by the institutional review board, which waived the requirement for obtaining informed consent. We searched the electronic medical record for adult patients who underwent 3-T MRI of the shoulder for shoulder pain between September 1, 2019, and June 30, 2020. During this time, the routine shoulder MRI protocol in place at our institution included both standard 2D FSE sequences in three planes and accelerated sequences using DLR in three planes. The MRI technologists were instructed to use PROPELLER sequences instead of standard images when patient artifact was observed during the scan [13]. A total of 276 examinations in 219 patients were identified from the initial search. Of these, 96 examinations were excluded because PROPELLER was used instead of standard 2D FSE for at least one plane, leaving 180 examinations in 164 patients. Seventy examinations were then excluded for the following reasons to identify a cohort of patients with pathologies that would be typically encountered in a primary sports medicine setting: prior shoulder surgery ($n = 57$), shoulder infection ($n = 6$), glenohumeral joint trauma ($n = 6$), and bony metastasis in the shoulder ($n = 1$). These exclusions resulted in a final study sample of 110 shoulder MRI examinations in 105 patients (45 men, 60 women; mean age, 57.6 ± 10.9 [SD] years; range, 22–86 years) (Fig. 1).

HIGHLIGHTS

Key Finding

- *For two independent readers of shoulder MRI, image quality and artifacts were worse ($p < .05$) for accelerated sequences without DLR compared with standard sequences, though not different ($p > .05$) between accelerated sequences with DLR and standard sequences. Accuracy, sensitivity, and specificity for tendon tears were not different ($p > .05$) among image sets.*

Importance

- *Accelerated sequences with DLR provide a 67% reduction in scan time with preserved image quality and diagnostic performance compared with standard sequences for shoulder MRI.*

MRI Protocol

Imaging was performed using a 3-T MRI system (Signa Architect, GE Healthcare) with a 16-channel dedicated shoulder coil. All examinations included standard and accelerated sets of sequences. Both sets of sequences included an axial intermediate-weighted fat-suppressed FSE sequence, an oblique coronal T2-weighted fat-suppressed FSE sequence, and an oblique sagittal T2-weighted fat-suppressed FSE sequence. The accelerated sequences used a parallel imaging factor of 4. Additional acquisition parameters are shown in Table 1. The scan time was 9 minutes 23 seconds for the three standard sequences and 3 minutes 5 seconds for the three accelerated sequences (a 67% reduction). The three standard sequences were reconstructed using the conventional reconstruction pipeline, whereas the three accelerated sequences were reconstructed using both the conventional pipeline and the DLR pipeline, generating a total of three image sets, each of which included sequences in three different planes.

The prototype DLR software used to reconstruct the accelerated sequences was supplied by the vendor and installed on the scanner. The DLR pipeline includes a deep convolutional neural network (CNN) component that replaces elements of the conventional pipeline. Further, the DLR uses a complex-valued floating-point data approach, which removes noise in the complex domain and estimates the truncated high-frequency k-space data to provide high-quality interpolation with sharp edges, while removing ringing artifacts. The CNN was trained in a supervised manner on more than 10,000 high-quality images to provide high-resolution data with minimal ringing artifact and very low noise levels that could be applied across a broad range of content. Low-quality data with a greater degree of truncation and an increase in noise were synthesized from the high-quality data. After additional augmentation, a training database of 4 million pairs of high- and low-quality data were generated. The CNN was trained to use these pairs to generate the high-quality images from the low-quality images. Training was performed in a single epoch, and the Adam optimization algorithm was used to minimize the loss. This pipeline can be applied to 2D sequences in multiple anatomic regions and for var-

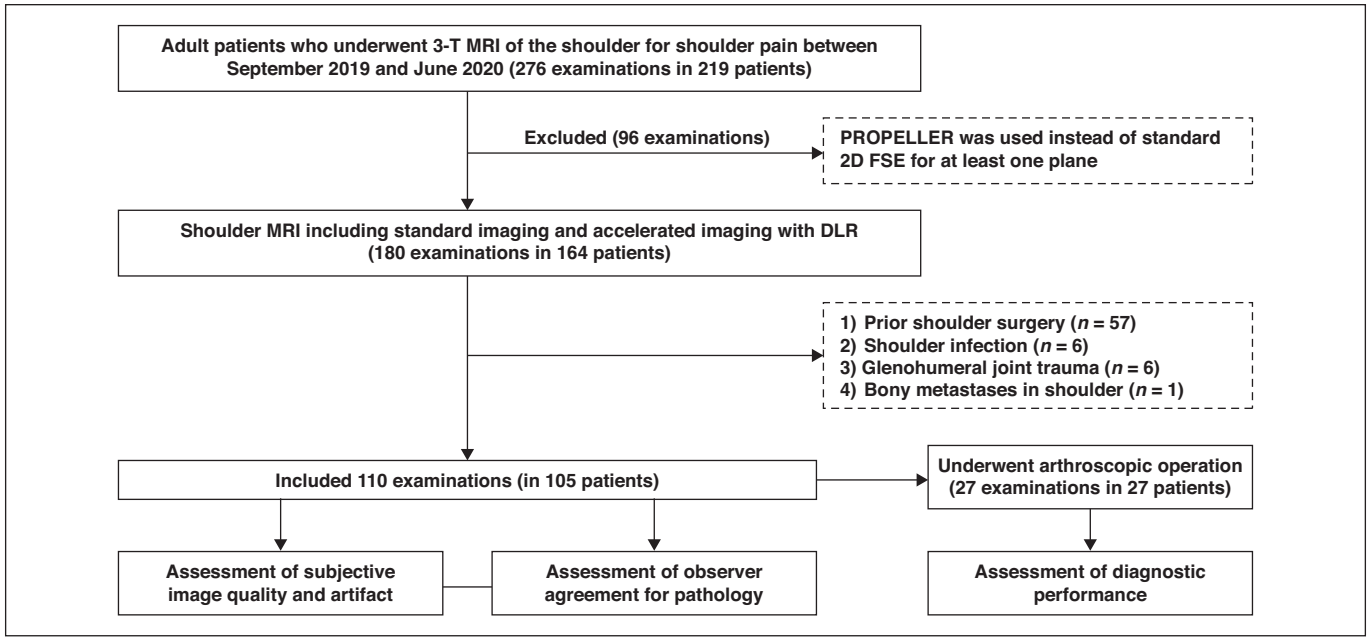

Fig. 1—Study flow diagram. FSE = fast spin-echo, DLR = deep learning–based reconstruction.

ious pulse sequences, contrast weightings, field strengths, and coil configurations [8]. The prototype software incorporates a tunable noise reduction factor between 0 and 1, which represents the fraction of the noise variance to be removed. Various noise reduction levels at 10% intervals from 50% to approximately 100% were tested to identify a threshold level that balances the decrease in noise and minimization of artificial texture. A reduction level of 60% was routinely used at our institution, because this level had been found to yield images that appear most similar to standard images.

Image Analysis

The images were reviewed independently by two musculoskeletal radiologists (J.Y. and S.H., with 5 and 6 years of subspecialty experience, respectively) who did not have access to the clinical radiology reports or clinical or any available pathologic information. The radiologists reviewed the three image sets in three separate sessions (first standard sequences, then accelerated sequences without DLR, and finally accelerated sequences with DLR), separated by 2 weeks between sessions. In each of the three sessions, the 110 examinations were reviewed in random order. When reviewing a given image set (e.g., standard sequences in a given patient), the three individual sequences in different planes were reviewed concurrently, though each plane was scored separately. Each plane was assessed subjectively for image quality and artifacts. The scores for each measure were based on the lowest-quality individual slice in that plane. Before

TABLE 1: MRI Parameters

Parameter	Standard			Accelerated		
	Axial	Coronal	Sagittal	Axial	Coronal	Sagittal
TR (ms)	3129	5080	5063	3129	5080	5063
TE (ms)	26.6	64.2	68.8	26.6	64.2	68.8
FOV (cm)	15 × 15	15 × 15	15 × 15	15 × 15	15 × 15	15 × 15
Section thickness/gap (mm)	3/0.3	3/0	3/0.3	3/0.3	3/0	3/0.3
Acquisition matrix	320 × 250	320 × 256	320 × 250	320 × 250	320 × 256	320 × 250
Echo-train length	7	14	15	7	14	15
Bandwidth (Hz/pixel)	162.8	162.8	195.3	162.8	162.8	195.3
Number of excitations	1	1	1	1	1	1
Parallel imaging factor (ARC)	1	1	1	4	4	4
Noise reduction factor (0–1)				0.6	0.6	0.6
Acquisition time	3 min 8 s	3 min 18 s	2 min 27 s	1 min 3 s	1 min 6 s	56 s

Note—Axial sequences were intermediate-weighted fat-suppressed sequences; coronal and sagittal sequences were T2-weighted fat-suppressed sequences. Total acquisition time was 9 min 23 s for standard sequences and 3 min 5 s for accelerated sequences. ARC = autocalibrating reconstruction for cartesian sampling.

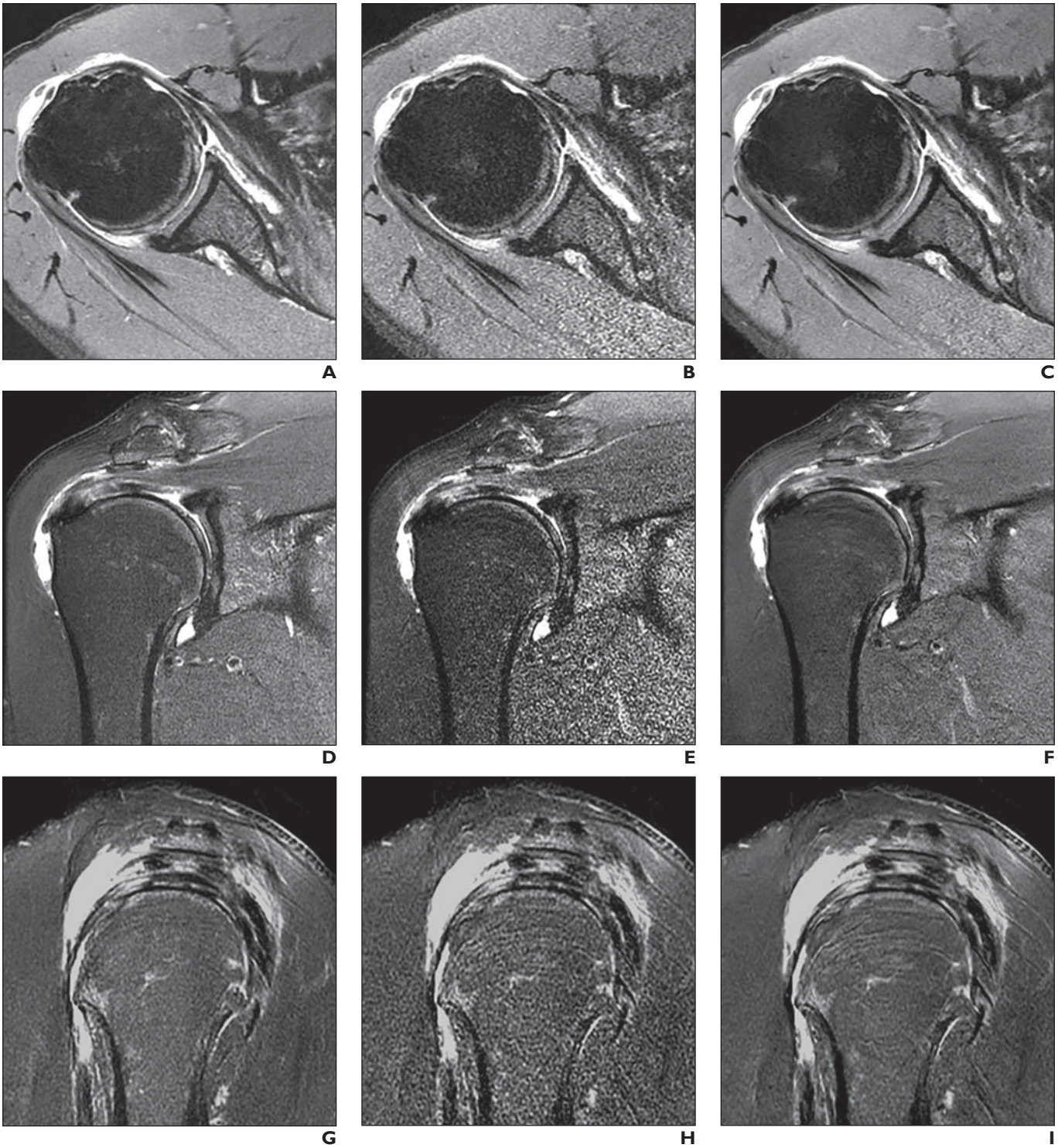

Fig. 2—61-year-old man with right shoulder pain who underwent MRI.
A–I, MR images include axial intermediate-weighted fat-saturated (FS) standard image (**A**), accelerated image without deep learning–based reconstruction (DLR) (**B**), and accelerated image with DLR (**C**); coronal oblique T2-weighted FS standard image (**D**), accelerated image without DLR (**E**), and accelerated image with DLR (**F**); and sagittal oblique T2-weighted FS standard image (**G**), accelerated image without DLR (**H**), and accelerated image with DLR (**I**). Accelerated images without DLR (**B**, **E, H**) show motion artifacts and noise. Accelerated images with DLR (**C, F, I**) show decreased noise and increased sharpness.

the independent readings, the two readers reviewed images for 10 examinations not included in the study sample and jointly discussed sources of discrepancy.

Subjective image quality was assessed using a 4-point scale, considering overall image quality and clarity of the rotator cuff tendon, bone marrow, and glenoid labrum (GL) (1 = poor: nondiagnostic and substantially limited evaluation of anatomic structures; 2 = sufficient: sufficient for most diagnoses but somewhat limited evaluation of anatomic structures; 3 = good: good for most diagnoses and evaluation of anatomic structure; 4 = excellent: optimal diagnostic utility with sharp depiction of the evaluated anatomic structure) (Fig. S1, which is available in the electronic supplement to this article at doi.org/10.2214/AJR.21.26577). Presence of artifacts was also assessed using a 4-point scale, considering motion, ringing, pulsation, and susceptibility, as well as subjective image noise (1 = severe artifacts; 2 = mild or moderate artifacts affecting the evaluation of anatomic structures; 3 = mild artifacts not affecting the evaluation of anatomic structures; 4 = no artifacts) [14] (Fig. S1). The scores from 1 to 4 for each of the three planes were summed to provide total scores from 3 to 12 for subjective image quality and artifacts for each image set.

The two readers also performed anatomic structure evaluation for each image set during the same sessions as when scoring subjective image quality and artifacts. This assessment was performed for the overall multiplanar image set, based on a composite assessment of the three different planes. The readers categorized the status of the subscapularis (SSC) tendon, the supraspinatus-infraspinatus tendon (SST-IST), and the long head of the biceps brachii tendon (LHBT) on each image set as normal, tendinopathy, or partial or complete tear. Tendinopathy was defined as hyperintensity in the tendon, with or without alteration in tendon diameter. Tendon tear was defined as complete or partial discontinuity of the tendon, with or without signal change. If the LHBT exhibited subluxation or dislocation, then the readers gave closer attention for possible SSC tendon tear given that LHBT subluxation or dislocation may be an indirect sign of SSC tendon tear [15]. The supraspinatus and infraspinatus tendons were regarded as a single entity because these tendons overlap and are intermingled on the footprint and thus difficult to separate [16, 17]; deeming these a single entity is also consistent with previous studies [16–18]. In addition, the GL was categorized as normal, normal variant, fraying, or tear [19]. A tear of the GL was defined as detachment of the labrum from the glenoid of the scapula [7]. Several MRI signs were used to differentiate a pathologic tear of the GL from normal variants (e.g., sublabral recess, sublabral foramen, and Buford complex) [20, 21]: irregular high-signal intensity line in the GL extending to the articular surface, flap tears of GL, a laterally extended high-signal intensity line, detached GL, paralabral cyst, and two high-signal intensity lines ("double Oreo" sign).

Arthroscopic Shoulder Surgery

Diagnostic accuracy of the anatomic structural evaluation was assessed in patients who underwent arthroscopic shoulder surgery after MRI, using arthroscopic findings as the reference standard. Arthroscopic surgery was performed in 27 shoulders in 27 patients (15 men, 12 women). The interval between MRI and arthoscopic surgery ranged from 1 to 247 days (mean, 40.2 ± 57.5 days). All surgeries were performed by a single orthopedic surgeon (J.B., with 8 years of experience in specialized shoulder surgeries), who had access to the MRI findings. From the surgical reports, we documented the presence or absence of SSC tendon, SST-IST, and LHBT tears, and of GL lesions. To compare the MRI and surgery findings, tendon pathology was dichotomized as no tear (normal and tendinopathy) versus tear, and the GL pathology was dichotomized as no tear (normal, normal variation, and fraying) versus tear. After completion of the independent readings and correlation with arthroscopic findings, the two readers performed in consensus a qualitative post hoc assessment of cases that were incorrectly interpreted by both readers on all three image sets to identify potential causes of the discrepancies.

Statistical Analyses

Scores for subjective image quality and artifact were compared among pairwise combinations of image sets for each reader using the Wilcoxon signed rank test. The frequencies of each pathology for each structure were summarized, stratified by image set and reader. Weighted kappa coefficients were calculated by the quadratic method to assess the interobserver agreement for intraarticular tendon pathology and GL lesions ($\kappa < 0$, no agreement; $0 < \kappa \leq 0.2$, slight agreement; $0.2 < \kappa \leq 0.4$, fair agreement; $0.4 < \kappa \leq 0.6$, moderate agreement; $0.6 < \kappa \leq 0.8$, substantial agreement; and $0.8 < \kappa \leq 1$, almost perfect agreement) [22]. The accuracy, sensitivity, and specificity for detecting tendon tears and GL tears on the three image sets were calculated using arthroscopic findings as the reference standard. The McNemar test was used to compare the diagnostic performance of image sets. Statistical significance was defined as $p < .05$. All statistical analyses were performed using SPSS Statistics software (version 25, IBM) and R software (ver. 4.0.2, R Core Team).

Results

Table 2 presents the scores for subjective image quality and artifacts for the three image sets for both readers. The mean subjective image quality score for reader 1 was 10.6 ± 1.2 for standard sequences, 8.1 ± 1.3 for accelerated sequences without DLR, and 10.7 ± 1.2 for accelerated sequences with DLR, and for reader 2 was 10.5 ± 1.4 for standard sequences, 7.2 ± 1.1 for accelerated sequences without DLR, and 10.5 ± 1.6 for accelerated sequences with DLR. The mean subjective image quality score was significantly higher for both standard sequences and accelerated sequences with DLR compared with accelerated sequences without DLR for both readers (all $p < .05$); the mean subjective image quality score was not significantly different between standard sequences and accelerated sequences with DLR for either reader (both $p > .05$).

The mean score for artifacts for reader 1 was 9.3 ± 1.2 for standard sequences, 7.3 ± 1.3 for accelerated sequences without DLR, and 9.4 ± 1.2 for accelerated sequences with DLR, and for reader 2 was 10.0 ± 1.0 for standard sequences, 9.1 ± 0.8 for accelerated sequences without DLR, and 9.8 ± 0.8 for accelerated sequences with DLR. The mean score for artifacts was significantly higher for both standard sequences and accelerated sequences with DLR compared with accelerated sequences without DLR for both readers (all $p < .05$); the mean score for artifacts was not significantly different between standard sequences and accelerated sequences with DLR for either reader (both $p > .05$). Figure 2 de-

TABLE 2: Subjective Image Quality Analysis of the Three Image Sets

Reader, Measure	Mean (± SD) Score			*p*		
	Standard	Accelerated Without DLR	Accelerated With DLR	Standard vs Accelerated Without DLR	Standard vs Accelerated With DLR	Accelerated Without DLR vs Accelerated With DLR
Reader 1						
Subjective image quality	10.6 ± 1.2	8.1 ± 1.3	10.7 ± 1.2	< .001	.10	< .001
Artifacts	9.3 ± 1.2	7.3 ± 1.3	9.4 ± 1.2	< .001	.30	< .001
Reader 2						
Subjective image quality	10.5 ± 1.4	7.2 ± 1.1	10.5 ± 1.6	< .001	.95	< .001
Artifacts	10.0 ± 1.0	9.1 ± 0.8	9.8 ± 0.8	< .001	.13	< .001

Note—DLR = deep learning–based reconstruction.

TABLE 3: Comparisons of Three MRI Sequences for Intraarticular Pathologies by Both Readers

Anatomic Structure, Status	Standard		Accelerated Without DLR		Accelerated With DLR	
	Reader 1	Reader 2	Reader 1	Reader 2	Reader 1	Reader 2
SSC tendon						
Normal	56 (50.9)	52 (47.3)	66 (60.0)	65 (59.1)	63 (57.3)	60 (54.5)
Tendinopathy	12 (10.9)	16 (14.5)	17 (15.5)	6 (5.5)	4 (3.6)	12 (10.9)
Tear	42 (38.2)	42 (38.2)	27 (24.5)	39 (35.5)	43 (39.1)	38 (34.5)
SST-IST						
Normal	2 (1.8)	1 (0.9)	7 (6.4)	2 (1.8)	2 (1.8)	1 (0.9)
Tendinopathy	50 (45.5)	49 (44.5)	58 (52.7)	50 (45.5)	47 (42.7)	48 (43.6)
Tear	58 (52.7)	60 (54.5)	45 (40.9)	58 (52.7)	61 (55.5)	61 (55.5)
LHBT						
Normal	77 (70.0)	63 (57.3)	71 (64.5)	70 (63.6)	66 (60.0)	58 (52.7)
Tendinopathy	17 (15.5)	27 (24.5)	25 (22.7)	21 (19.1)	25 (22.7)	32 (29.1)
Tear	16 (14.5)	20 (18.2)	14 (12.7)	19 (17.3)	19 (17.3)	20 (18.2)
GL						
Normal or normal variant	55 (50.0)	52 (47.3)	54 (49.1)	53 (48.2)	56 (50.9)	53 (48.2)
Fraying	14 (12.7)	15 (13.6)	26 (23.6)	16 (14.5)	14 (12.7)	14 (12.7)
Tear	41 (37.3)	43 (39.1)	30 (27.3)	41 (37.3)	40 (36.4)	43 (39.1)

Note—Values represent number of shoulder examinations with percentage in parentheses. DLR = deep learning–based reconstruction, SSC = subscapularis, SST-IST = supraspinatus tendon-infraspinatus tendon, LHBT = long head of biceps tendon, GL = glenoid labrum.

picts image quality and artifacts for the three image sets in a representative patient. Figure 3 depicts findings on the three image sets in a representative patient with a bursal-side SST tear on arthroscopic surgery.

Table 3 summarizes the frequency of pathology involving the four assessed anatomic structures for each image set for each reader. SSC tendon tear was visualized on standard sequences, accelerated sequences without DLR, and accelerated sequences with DLR in 38.2%, 24.5%, and 39.1% of examinations, respectively, for reader 1, and in 38.2%, 35.5%, and 34.5%, respectively, for reader 2. SST-IST tear was visualized on standard sequences, accelerated sequences without DLR, and accelerated sequences with DLR in 52.7%, 40.9%, and 55.5% of examinations, respectively, for reader 1, and in 54.5%, 52.7%, and 55.5%, respectively, for reader 2. LHBT tear was visualized on standard sequences, accelerated sequences without DLR, and accelerated sequences with DLR in 14.5%, 12.7%, and 17.3% of examinations, respectively, for reader 1, and in 18.2%, 17.3%, and 18.2%, respectively, for reader 2. GL tear was visualized on standard sequences, accelerated sequences without DLR, and accelerated sequences with DLR in 37.3%, 27.3%, and 36.4% of examinations, respectively, for reader 1, and in 39.1%, 37.3%, and 39.1%, respectively, for reader 2. Table 4 provides the interobserver agreement for intraarticular shoulder pathologies for each image set as well as the agreement between image sets for intraarticular shoulder pathologies for each reader. The interobserver agreement was almost perfect (κ = 0.813–0.951)

TABLE 4: Interobserver Agreement and Agreement Between Image Sets for Intraarticular Shoulder Pathologies

Anatomic Structure	Interobserver Agreement			Agreement Between Image Sets					
				Standard vs Accelerated Without DLR		Standard vs Accelerated With DLR		Accelerated Without DLR vs Accelerated With DLR	
	Standard	Accelerated Without DLR	Accelerated With DLR	Reader 1	Reader 2	Reader 1	Reader 2	Reader 1	Reader 2
SSC tendon	0.896	0.898	0.823	0.843	0.868	0.904	0.876	0.883	0.827
SST-IST	0.951	0.673	0.918	0.700	0.819	0.889	0.949	0.663	0.836
BT	0.813	0.917	0.916	0.861	0.864	0.809	0.886	0.852	0.886
GL	0.942	0.861	0.937	0.827	0.920	0.957	0.952	0.850	0.957

Note—Data expressed as kappa values. Kappa value < 0 indicates no agreement; 0 < kappa ≤ 0.2, slight agreement; 0.2 < kappa ≤ 0.4, fair agreement; 0.4 < kappa ≤ 0.6, moderate agreement; 0.6 < kappa ≤ 0.8, substantial agreement; and 0.8 < kappa ≤ 1, almost perfect agreement [20]. DLR = deep learning–based reconstruction, SSC = subscapularis, SST-IST = supraspinatus tendon-infraspinatus tendon, BT = biceps brachii long head tendon, GL = glenoid labrum.

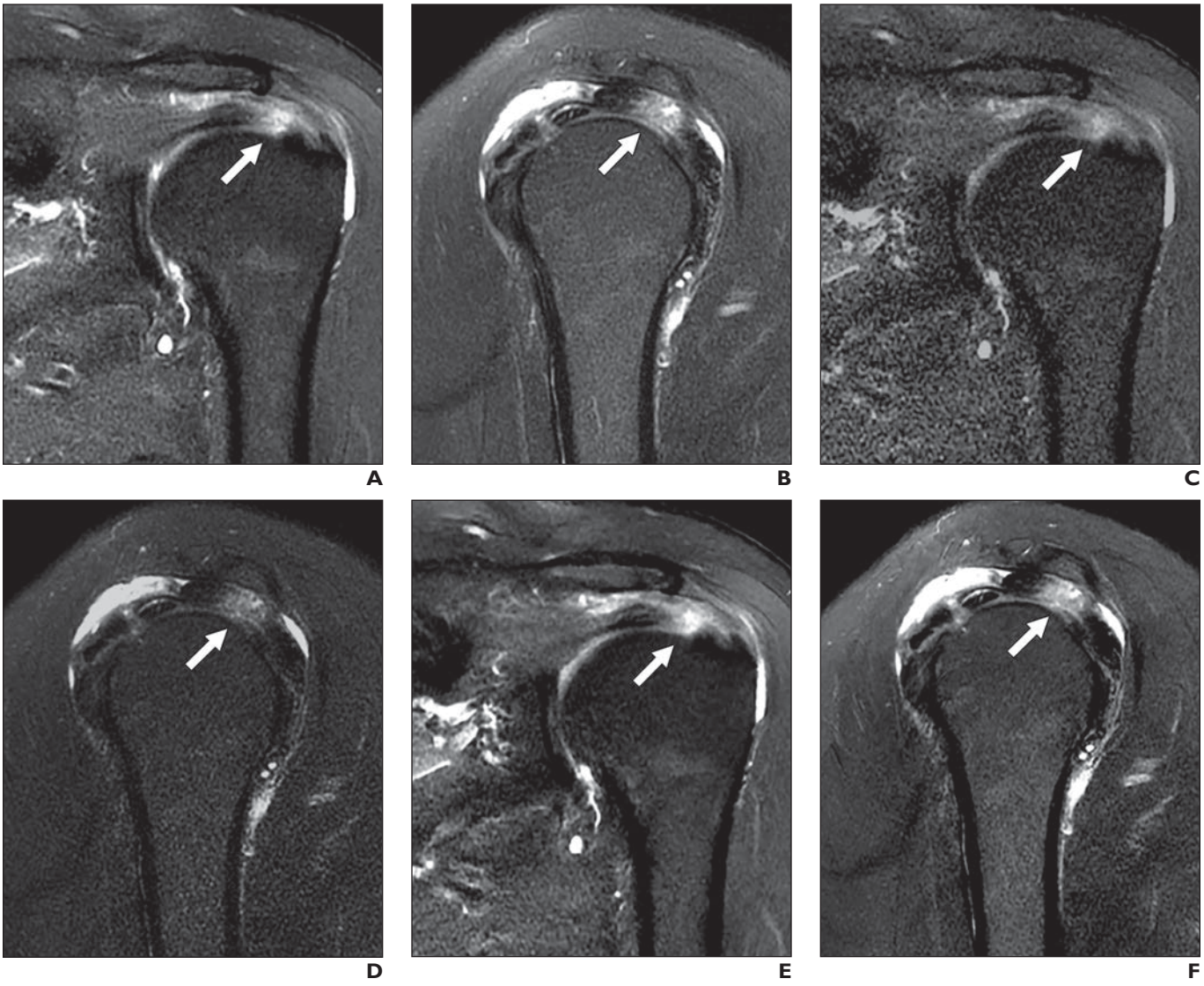

Fig. 3—65-year-old woman with left shoulder pain who underwent MRI.
A–F, Coronal (**A**) and sagittal (**B**) T2-weighted fat-saturated standard images, coronal (**C**) and sagittal (**D**) accelerated images without deep learning–based reconstruction (DLR), and coronal (**E**) and sagittal (**F**) accelerated images with DLR show T2 hyperintensity of supraspinatus tendon-infraspinatus tendon (SST-IST) (*arrow*). Both readers considered SST-IST to show tear on standard images and accelerated images with DLR. However, for accelerated MRI without DLR, one reader classified SST-IST as torn and one reader classified SST-IST as showing tendinopathy without tear. Arthroscopic surgery revealed bursal-side tear of SST-IST.

for all image sets for all pathologies except for the SST-IST on accelerated sequences without DLR (κ = 0.673, substantial agreement) (Fig. 3). The agreement between image sets was almost perfect (κ = 0.809–0.957) for all pairwise combinations of image sets for both readers for all pathologies except for the agreement between standard sequences and accelerated sequences without DLR for reader 1 for SST-IST tear (κ = 0.700, substantial agreement).

Of the 27 patients who underwent arthroscopic surgery, SSC tendon tear was present in 17 (63.0%), SST-IST tear in 23 (85.2%), LHBT tear in 11 (40.7%), and GL tear in 9 (33.3%). Table 5 presents the diagnostic performance of the three image sets for tears on arthroscopy. Across pathologies, image sets, and readers, the accuracy ranged from 81.5% to 92.6%, the sensitivity from 55.6% to 91.3%, and the specificity from 90.0% to 100% (aside from 75.0% for SST-IST tear for accelerated sequences without DLR for reader 2). There was no significant difference in accuracy, sensitivity, or specificity for any pairwise combination of image sets for either reader (all $p \geq .05$).

There were two false-negative cases each of SSC tendon tear, SST-IST tear, and LHBT tear by both readers for all three image sets. In the post hoc review, the two false-negative cases of SSC tendon tear were found to show fraying at the leading edge of the SSC tendon on MRI (Fig. S2, which is available in the electronic supplement to this article at doi.org/10.2214/AJR.21.26577). The two false-negative cases of SST-IST tear were found to exhibit calcific tendinitis without tear on MRI, in comparison with calcifications and tendon tear on MRI. No explanation was identified at post hoc assessment for the two false-negative LHBT tears. There were two false-negative GL tears and one false-positive GL tear by both readers on all three image sets. One of the false-negative GL tears was considered a labral recess by both readers. The other false-negative GL tear was a superior labral tear on arthroscopy with no abnormality identified on post hoc review of the MR image sets. In the case of a false-positive GL tear, arthroscopy exhibited medial dislocation and tear of LHBT and mild GL distortion, contributing to the misinterpretation on MRI as GL tear.

Discussion

In this study of patients who underwent shoulder MRI, we compared standard sequences and accelerated sequences performed without and with DLR in terms of subjective image quality, artifacts, and diagnostic accuracy using arthroscopy as the reference standard. Accelerated sequences had 67% shorter scan time than standard sequences and exhibited worse image quality and artifacts compared with standard sequences. The use of DLR with accelerated sequences resulted in improved image quality and artifacts compared with accelerated sequences without DLR, as well as similar image quality and artifacts compared with standard sequences. Further, interobserver agreement for tears of the SST-IST, as well as agreement between image sets for SST-IST tears, were better for both standard sequences and accelerated sequences with DLR compared with accelerated sequences without DLR. The three image sets exhibited similar diagnostic performance for tears of various anatomic structures in patients who underwent arthroscopy. The findings indicate a role for DLR to facilitate application of accelerated sequences for clinical shoulder MRI, providing substantial time savings with preserved image quality and diagnostic performance.

Previous investigations have assessed accelerated shoulder MRI using 2D or 3D sequences with or without arthrography [18, 23–25]. A recent study by Hou et al. [25] that compared the diagnostic accuracy of 2D images and 3D CAIPIRINHA (controlled aliasing in parallel imaging results in higher acceleration) SPACE reported that 3D CAIPIRINHA SPACE achieved comparable diagnostic accuracy for shoulder pathologies in shorter imaging time. In an additional study, Jung et al. [23] reported similar sensitivity and specificity for labral lesions for 3D MR arthrography compared with 2D MR arthrography, though with shorter scan time for the 3D approach. Further, a study by Subhas et al. [7] that applied a 5-minute shoulder MRI protocol consisting of 2D FSE sequences with parallel imaging reported similar accuracy in diagnosing shoulder pathology compared with a standard protocol. However, the fast sequences had slightly lower resolution than the standard sequences, and the fast protocol included one less sequence than the standard protocol.

For MRI of the knee, Recht et al. [11] compared standard images (acceleration factor of 2) and images obtained with DLR (acceleration factor of 3.49 compared with fully sampled images and of 1.88 compared with twofold-accelerated standard images) in terms of image quality and internal joint derangement. The authors reported that accelerated images with DLR were interchangeable with standard images and that subtle signal abnormalities were less conspicuous for DLR when using higher acceleration factors [11]. In our study, accelerated sequences used a parallel imaging reduction factor of 4 compared with a factor of 1 for standard sequences, similar to the ratio in the study by Recht et al.

DLR, as used in our study, improves signal-to-noise ratio (SNR) through the incorporation of a CNN within the MRI reconstruction pipeline [8]. The DLR pipeline achieves effective interpolation by estimating high-frequency k-space information, thereby suppressing ringing artifacts. The DLR approach is also able to omit software filters that are applied in the conventional pipeline to partially mitigate noise and truncation artifacts but that reduce effective spatial resolution; this difference ultimately leads to a sharper-appearing image when using DLR. These benefits of DLR in terms of both SNR and image sharpness can be used to improve image quality without an increase in scan time, as shown in the current study.

Shoulder MRI is susceptible to image degradation from various sources of patient motion, including respiration. A shorter scan time reduces the chance of motion artifact. Our study indicates a role for DLR to facilitate use of accelerated sequences as an alternative to standard sequences, which may in turn lead to reduced motion artifact. However, DLR is not itself intended to directly eliminate motion artifacts [8]. In fact, van der Velde et al. [10] reported that wrapping and ghosting artifacts were more prominent when DLR was used for late gadolinium enhancement imaging of the myocardium, which is in contrast to our observation of reduced artifacts after applying DLR. This difference may in part be explained by our institution's use of the PROPELLER sequence in patients who show motion artifact during shoulder MRI scanning; our exclusion of examinations that used PROPELLER would be expected to have resulted in exclusion of patients with severe motion artifact. As a future application, DLR may be applied in conjunction with techniques that serve to reduce motion artifact such as PROPELLER.

TABLE 5: Diagnostic Performance of the Three Image Sets for Detection of Intraarticular Shoulder Pathologies

Anatomic Structure, Reader, Sequence	Accuracy				Sensitivity				Specificity			
			p				*p*				*p*	
	Percentage (95% CI)	Raw Data	Std	Acc w/o DLR	Percentage (95% CI)	Raw Data	Std	Acc w/o DLR	Percentage (95% CI)	Raw Data	Std	Acc w/o DLR
SSC tendon												
Reader 1												
Std	85.2 (67.5–94.1)	23/27			82.4 (59.0–93.8)	14/17			90.0 (59.6–99.5)	9/10		
Acc w/o DLR	81.5 (63.3–91.8)	22/27	> .99[a]		76.5 (52.7–90.4)	13/17	> .99[a]		90.0 (59.6–99.5)	9/10	> .99[a]	
Acc with DLR	85.2 (67.5–94.1)	23/27	> .99[b]	> .99[c]	82.4 (59.0–93.8)	14/17	> .99[b]	> .99[c]	90.0 (60.0–99.0)	9/10	> .99[b]	> .99[c]
Reader 2												
Std	88.9 (71.9–96.1)	24/27			88.2 (65.7–96.7)	15/17			90.0 (59.6–99.5)	9/10		
Acc w/o DLR	88.9 (71.9–96.1)	24/27	> .99[a]		88.2 (65.7–96.7)	15/17	> .99[a]		90.0 69.6–99.5)	9/10	> .99[a]	
Acc with DLR	81.5 (63.3–91.8)	22/27	.50[b]	.50[c]	76.5 (52.7–90.4)	13/17	.50[b]	.50[c]	90.0 (59.6–99.5)	9/10	> .99[b]	> .99[c]
SST-IST												
Reader 1												
Std	92.6 (76.6–97.9)	25/27			91.3 (73.2–97.6)	21/23			100.0 (51.0–100.0)	4/4		
Acc w/o DLR	81.5 (63.3–91.8)	22/27	.25[b]		78.3 (58.1–90.3)	18/23	.25[a]		100.0 (51.0–100.0)	4/4	NA	
Acc with DLR	92.6 (76.6–97.9)	25/27	> .99[b]	.25[c]	91.3 (73.2–97.6)	21/23	> .99[b]	.25[c]	100.0 (51.0–100.0)	4/4	NA	NA
Reader 2												
Std	92.6 (76.6–97.9)	25/27			91.3 (73.2–97.6)	21/23			100.0 (51.0–100.0)	4/4		
Acc w/o DLR	88.9 (71.9–96.1)	24/27	> .99[a]		91.3 (73.2–97.6)	21/23	> .99[a]		75.0 (30.1–98.7)	3/4	> .99[a]	
Acc with DLR	92.6 (76.6–97.9)	25/27	> .99[b]	> .99[c]	91.3 (73.2–97.6)	21/23	> .99[b]	> .99[c]	100.0 (51.0–100.0)	4/4	NA	> .99[c]
LHBT												
Reader 1												
Std	85.2 (67.5–94.1)	23/27			63.6 (35.4–84.8)	7/11			100.0 (80.6–100.0)	16/16		
Acc w/o DLR	88.9 (71.9–96.1)	24/27	> .99[a]		72.7 (43.4–90.3)	8/11	> .99[a]		100.0 (80.6–100.0)	16/16	NA	
Acc with DLR	88.9 (71.9–96.1)	24/27	> .99[b]	> .99[c]	72.7 (43.4–90.3)	8/11	> .99[b]	> .99[c]	100.0 (80.6–100.0)	16/16	NA	NA
Reader 2												
Std	92.6 (76.6–97.9)	25/27			81.8 (52.3–94.9)	9/11			100.0 (81.0–100.0)	16/16		
Acc w/o DLR	92.6 (76.6–97.9)	25/27	> .99[a]		81.8 (52.3–94.9)	9/11	> .99[a]		100.0 (81.0–100.0)	16/16	NA	
Acc with DLR	92.6 (76.6–97.9)	25/27	> .99[b]	> .99[c]	81.8 (52.3–94.9)	9/11	> .99[b]	> .99[c]	100.0 (81.0–100.0)	16/16	NA	NA

(Table 5 continues on next page)

TABLE 5: Diagnostic Performance of the Three Image Sets for Detection of Intraarticular Shoulder Pathologies (continued)

Anatomic Structure, Reader, Sequence	Accuracy				Sensitivity				Specificity			
	Percentage (95% CI)	Raw Data	*p*		Percentage (95% CI)	Raw Data	*p*		Percentage (95% CI)	Raw Data	*p*	
			Std	Acc w/o DLR			Std	Acc w/o DLR			Std	Acc w/o DLR
GL												
Reader 1												
Std	88.9 (71.9–96.1)	24/27			77.8 (45.3–93.7)	7/9			94.4 (74.2–99.7)	17/18		
Acc w/o DLR	81.5 (63.3–91.8)	22/27	.50[a]		55.6 (26.7–81.1)	5/9	.50[a]		94.4 (74.2–99.7)	17/18	> .99[a]	
Acc with DLR	88.9 (71.9–96.1)	24/27	> .99[b]	.50[c]	77.8 (45.3–93.7)	7/9	> .99[b]	.50[c]	94.4 (74.2–99.7)	17/18	> .99[b]	> .99[c]
Reader 2												
Std	85.2 (67.5–94.1)	23/27			66.7 (35.4–87.9)	6/9			94.4 (74.2–99.7)	17/18		
Acc w/o DLR	85.2 (67.5–94.1)	23/27	> .99[a]		66.7 (35.4–87.9)	6/9	> .99[a]		94.4 (74.2–99.7)	17/18	> .99[a]	
Acc with DLR	85.2 (67.5–94.1)	23/27	> .99[b]	> .99[c]	66.7 (35.4–87.9)	6/9	> .99[b]	> .99[c]	94.4 (74.2–99.7)	17/18	> .99[b]	> .99[c]

Note—Std = standard, Acc = accelerated, w/o = without, DLR = deep learning–based reconstruction, SSC = subscapularis, SST-IST = supraspinatus tendon-infraspinatus tendon, NA = not applicable (p value not calculated for comparison of two percentages of 100.0%), LHBT = long head of biceps brachii tendon, GL = glenoid labrum.

[a]Comparison of accuracy between standard and accelerated without DLR, calculated using McNemar test.

[b]Comparison of accuracy between standard and accelerated with DLR, calculated using McNemar test.

[c]Comparison of accuracy between accelerated with and without DLR, calculated using McNemar test.

There were limitations to the current study. First, we did not compare quantitative image quality measures among the image sets. However, previous investigations with phantom experiments and clinical examinations have reported an increase in SNR and contrast-to-noise ratio using DLR [8, 9]. Second, the number of patients who underwent arthroscopic surgery was small. Nonetheless, the diagnostic performance for shoulder pathologies of the 2D FSE sequences was similar to prior studies [7, 18]. Third, we grouped partial and complete tendon tears for purposes of analysis. Fourth, although the readers were not informed whether the images were standard or accelerated, subtle differences in image texture and noise patterns may have allowed the readers to recognize the image type. Fifth, the interval between sessions was short at 2 weeks. Sixth, the maximum interval between MRI and surgery was long, and pathologies may have changed during the interval. Finally, the images were reviewed by only two radiologists with similar experience levels.

In conclusion, this is the first study to our knowledge to investigate DLR for accelerated shoulder MRI in a clinical setting. The accelerated sequences achieved a 67% reduction in scan time compared with standard sequences. Accelerated sequences without DLR showed significantly worse image quality and artifact compared with standard sequences, though use of DLR for accelerated sequences resulted in comparable image quality and artifact as standard sequences. Standard sequences as well as accelerated sequences with and without DLR all showed comparable diagnostic performance for shoulder pathologies. The findings indicate a role for DLR to substantially reduce shoulder MRI scan times by facilitating the clinical application of accelerated sequences.

References

1. Small KM, Adler RS, Shah SH, et al.; Expert Panel on Musculoskeletal Imaging. ACR Appropriateness Criteria® shoulder pain-atraumatic. *J Am Coll Radiol* 2018; 15(11S):S388–S402
2. Tuite MJ, De Smet AA, Norris MA, Orwin JF. Anteroinferior tears of the glenoid labrum: fat-suppressed fast spin-echo T2 versus gradient-recalled echo MR images. *Skeletal Radiol* 1997; 26:293–297
3. Spencer EE Jr, Dunn WR, Wright RW, et al.; Shoulder Multicenter Orthopaedic Outcomes Network. Interobserver agreement in the classification of rotator cuff tears using magnetic resonance imaging. *Am J Sports Med* 2008; 36:99–103
4. Malavolta EA, Assunção JH, Guglielmetti CL, de Souza FF, Gracitelli ME, Ferreira Neto AA. Accuracy of preoperative MRI in the diagnosis of disorders of the long head of the biceps tendon. *Eur J Radiol* 2015; 84:2250–2254
5. Kloth JK, Winterstein M, Akbar M, et al. Comparison of 3D turbo spin-echo SPACE sequences with conventional 2D MRI sequences to assess the shoulder joint. *Eur J Radiol* 2014; 83:1843–1849
6. Park HJ, Lee SY, Kim MS, et al. Evaluation of shoulder pathology: three-dimensional enhanced T1 high-resolution isotropic volume excitation MR vs two-dimensional fast spin echo T2 fat saturation MR. *Br J Radiol* 2015; 88:20140147
7. Subhas N, Benedick A, Obuchowski NA, et al. Comparison of a fast 5-minute shoulder MRI protocol with a standard shoulder MRI protocol: a multi-institutional multireader study. *AJR* 2017; 208:[web]W146–W154
8. Lebel RM. Performance characterization of a novel deep learning-based MR image reconstruction pipeline. arXiv website. arxiv.org/abs/2008.06559. Published August 14, 2020.
9. Kim M, Kim HS, Kim HJ, et al. Thin-slice pituitary MRI with deep learning–

based reconstruction: diagnostic performance in a postoperative setting. *Radiology* 2021; 298:114–122
10. van der Velde N, Hassing HC, Bakker BJ, et al. Improvement of late gadolinium enhancement image quality using a deep learning–based reconstruction algorithm and its influence on myocardial scar quantification. *Eur Radiol* 2021; 31:3846–3855
11. Recht MP, Zbontar J, Sodickson DK, et al. Using deep learning to accelerate knee MRI at 3 T: results of an interchangeability study. *AJR* 2020; 215:1421–1429
12. Chaudhari AS, Grissom MJ, Fang Z, et al. Diagnostic accuracy of quantitative multicontrast 5-minute knee MRI using prospective artificial intelligence image quality enhancement. *AJR* 2021; 216:1614–1625
13. Dietrich TJ, Ulbrich EJ, Zanetti M, Fucentese SF, Pfirrmann CW. PROPELLER technique to improve image quality of MRI of the shoulder. *AJR* 2011; 197:[web]W1093–W1100
14. Jung JY, Yoon YC, Kim HR, Choe BK, Wang JH, Jung JY. Knee derangements: comparison of isotropic 3D fast spin-echo, isotropic 3D balanced fast field-echo, and conventional 2D fast spin-echo MR imaging. *Radiology* 2013; 268:802–813
15. Habermeyer P, Magosch P, Lichtenberg S. *Classifications and scores of the shoulder*. Springer Science & Business Media, 2006
16. Mochizuki T, Sugaya H, Uomizu M, et al. Humeral insertion of the supraspinatus and infraspinatus. New anatomical findings regarding the footprint of the rotator cuff. *J Bone Joint Surg Am* 2008; 90:962–969
17. Minagawa H, Itoi E, Konno N, et al. Humeral attachment of the supraspinatus and infraspinatus tendons: an anatomic study. *Arthroscopy* 1998; 14:302–306
18. Lee JH, Yoon YC, Jee S, Kwon JW, Cha JG, Yoo JC. Comparison of three-dimensional isotropic and two-dimensional conventional indirect MR arthrography for the diagnosis of rotator cuff tears. *Korean J Radiol* 2014; 15:771–780
19. Wolf BR, Uribe B, Hettrich CM, et al.; MOON Shoulder Group. Shoulder instability: interobserver and intraobserver agreement in the assessment of labral tears. *Orthop J Sports Med* 2018; 6:2325967118793372
20. Tuite MJ, Rutkowski A, Enright T, Kaplan L, Fine JP, Orwin J. Width of high signal and extension posterior to biceps tendon as signs of superior labrum anterior to posterior tears on MRI and MR arthrography. *AJR* 2005; 185:1422–1428
21. De Coninck T, Ngai SS, Tafur M, Chung CB. Imaging the glenoid labrum and labral tears. *RadioGraphics* 2016; 36:1628–1647
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33:159–174
23. Jung JY, Yoon YC, Choi SH, Kwon JW, Yoo J, Choe BK. Three-dimensional isotropic shoulder MR arthrography: comparison with two-dimensional MR arthrography for the diagnosis of labral lesions at 3.0 T. *Radiology* 2009; 250:498–505
24. Magee T, Shapiro M, Williams D, Ramnath RR, Simon J. Usefulness of the simultaneous acquisition of spatial harmonics technique during MRI of the shoulder. *AJR* 2003; 181:961–964
25. Hou B, Li Y, Xiong Y, et al. Comparison of CAIPIRINHA-accelerated 3D fat-saturated-SPACE MRI with 2D MRI sequences for the assessment of shoulder pathology. *Eur Radiol* 2021 Jul 13 [published online]

Editorial Comment: Accelerated Joint MRI With Deep Learning–Based Reconstruction—A Promising Approach to Increasing Imaging Speed Without Compromising Image Quality

Two-dimensional fast spin-echo sequences have been and continue to be the workhorse for musculoskeletal joint MRI. A growing body of literature has demonstrated that it is possible to achieve significant time savings when performing these sequences using parallel imaging (PI) acceleration techniques, which are now available on most modern scanners. Although PI can reduce time, it comes at the cost of lower signal-to-noise ratio and increased artifacts. This image quality degradation has typically limited PI to acceleration factors of 2 or less and necessitated reductions in spatial resolution to maintain acceptable image quality. Although these accelerated sequences have been shown to have similar diagnostic performance to that of nonaccelerated sequences [1], most radiologists have not been willing to sacrifice image quality or resolution for gains in imaging speed. With the use of deep learning–based reconstruction (DLR) to reduce the noise and artifacts associated with PI, it may no longer be necessary to compromise quality for speed. In this article, the authors show that DLR applied to a shoulder MRI protocol acquired with fourfold acceleration using PI can generate images with no significant differences relative to nonaccelerated sequences in subjective image quality or degree of artifact, and with similar diagnostic accuracy in a small subset of patients with surgical correlation. These results build on another recent study [2] that showed that fourfold PI accelerated knee MRI examinations with DLR had interchangeable diagnostic performance and better subjective image quality compared with a twofold PI accelerated clinical protocol without DLR. These studies highlight the promising potential of DLR to perform comprehensive joint MRI examinations in 3–4 minutes without compromising spatial resolution or image quality. Further studies are needed to validate these results in other scanners and coils in use today, including 1.5-T systems, as well as in other joints.

Naveen Subhas, MD, MPH
Imaging Institute
Cleveland Clinic
Cleveland, OH
subhasn@ccf.org

The author declares that there are no disclosures relevant to the subject matter of this article.

doi.org/10.2214/AJR.21.26880

References

1. Subhas N, Benedick A, Obuchowski NA, et al. Comparison of a fast 5-minute shoulder MRI protocol with a standard shoulder MRI protocol: a multi-institutional multireader study. *AJR* 2017; 208:[web]W146–W154
2. Recht MP, Zbontar J, Sodickson DK, et al. Using deep learning to accelerate knee MRI at 3 T: results of an interchangeability study. *AJR* 2020; 215:1421–1429