

# Deep Learning Reconstruction Enables Prospectively Accelerated Clinical Knee MRI



Patricia M. Johnson, PhD • Dana J. Lin, MD • Jure Zbontar, PhD • C. Lawrence Zitnick, PhD • Anuroop Sriram, MS • Matthew Muckley, PhD • James S. Babb, PhD • Mitchell Kline, MD • Gina Ciavarra, MD • Erin Alaia, MD • Mohammad Samim, MD • William R. Walter, MD • Liz Calderon, MPH • Thomas Pock, PhD • Daniel K. Sodickson, MD, PhD • Michael P. Recht, MD • Florian Knoll, PhD

From the Department of Radiology, New York University Grossman School of Medicine, 650 1st Ave, New York, NY 10016 (P.M.J., D.J.L., J.S.B., M.K., G.C., E.A., M.S., W.R.W., L.C., D.K.S., M.P.R., F.K.); Meta AI Research (FAIR), Menlo Park, Calif (J.Z., C.L.Z., A.S.); Meta AI Research, New York, NY (M.M.); Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria (T.P.); and Faculty of Engineering, Friedrich Alexander University Erlangen-Nurnberg (FAU), Erlangen, Germany (F.K.). Received February 24, 2022; revision requested April 4; revision received October 20; accepted November 15. **Address correspondence to** P.M.J. (email: [patricia.johnson3@nyulangone.org](mailto:patricia.johnson3@nyulangone.org)).

Supported by the National Institutes of Health (NIH) (grant R01EB024532). D.K.S. is supported by Meta/Facebook Artificial Intelligence Research with a collaborative research agreement. M.P.R. is supported by Facebook Artificial Intelligence Research and Amazon Web Services. F.K. is supported by the NIH (grant P41EB017183).

Conflicts of interest are listed at the end of this article.

See also the editorial by Roemer in this issue.

Radiology 2023; 000:1–9 • <https://doi.org/10.1148/radiol.220425> • Content codes:  

**Background:** MRI is a powerful diagnostic tool with a long acquisition time. Recently, deep learning (DL) methods have provided accelerated high-quality image reconstructions from undersampled data, but it is unclear if DL image reconstruction can be reliably translated to everyday clinical practice.

**Purpose:** To determine the diagnostic equivalence of prospectively accelerated DL-reconstructed knee MRI compared with conventional accelerated MRI for evaluating internal derangement of the knee in a clinical setting.

**Materials and Methods:** A DL reconstruction model was trained with images from 298 clinical 3-T knee examinations. In a prospective analysis, patients clinically referred for knee MRI underwent a conventional accelerated knee MRI protocol at 3 T followed by an accelerated DL protocol between January 2020 and February 2021. The equivalence of the DL reconstruction of the images relative to the conventional images for the detection of an abnormality was assessed in terms of interchangeability. Each examination was reviewed by six musculoskeletal radiologists. Analyses pertaining to the detection of meniscal or ligament tears and bone marrow or cartilage abnormalities were based on four-point ordinal scores for the likelihood of an abnormality. Additionally, the protocols were compared with use of four-point ordinal scores for each aspect of image quality: overall image quality, presence of artifacts, sharpness, and signal-to-noise ratio.

**Results:** A total of 170 participants (mean age  $\pm$  SD, 45 years  $\pm$  16; 76 men) were evaluated. The DL-reconstructed images were determined to be of diagnostic equivalence with the conventional images for detection of abnormalities. The overall image quality score, averaged over six readers, was significantly better ( $P < .001$ ) for the DL than for the conventional images.

**Conclusion:** In a clinical setting, deep learning reconstruction enabled a nearly twofold reduction in scan time for a knee MRI and was diagnostically equivalent with the conventional protocol.

© RSNA, 2023

Supplemental material is available for this article.

Due to its superior soft-tissue contrast, absence of ionizing radiation, and flexibility of contrast mechanisms, MRI is a common imaging method to evaluate joints. Specifically, it is used for the detection and assessment of acute and chronic internal derangement injuries of the knee. MRI is useful in the characterization of injuries to the cruciate ligaments, collateral ligaments, menisci, and extensor mechanism, as well as abnormalities of the cartilage and bone marrow. One major challenge of MRI is that data acquisition is slow in comparison to other modalities. The scan times for a single patient often reach 20–30 minutes or more. Consequently, shortening of scan times is a major research focus.

The advent of compressed sensing in the late 2000s (1) extended the limits of acceleration by enabling reconstruction of diagnostic quality images from undersampled data.

Since 2016, deep learning (DL) has offered the prospect of still greater accelerations by allowing robust reconstruction of high-quality images from heavily undersampled data with use of suitably configured neural networks (2–8). DL reconstruction is an extension of compressed sensing, where a sparsity constraint regularizes the inverse problem that arises in the context of an accelerated (undersampled) MRI acquisition. DL opens up the possibility to learn this regularizer from a set of training data. The availability of public data sets has opened the field up to research groups without access to their own MRI data (9,10). Public research challenges by means of open competitions provided the first systematic comparisons of approaches on standardized data (11). As of today, a variety of algorithms and network topologies have been proposed and their performance explored retrospectively with available data.

This copy is for personal use only. To order printed copies, contact [reprints@rsna.org](mailto:reprints@rsna.org)

## Abbreviation

DL = deep learning

## Summary

Compared with conventional reconstruction, deep learning reconstruction of prospectively accelerated knee MRI enabled an almost twofold scan time reduction, improved image quality, and had equivalent diagnostic utility.

## Key Results

- In a prospective analysis of 170 patients evaluated with 3-T knee MRI, deep learning (DL)-reconstructed accelerated images were diagnostically equivalent to conventional accelerated images for detection of pathologic features.
- Mean scan times were reduced from 9 minutes 56 seconds  $\pm$  19 seconds for conventional imaging to 5 minutes 33 seconds  $\pm$  16 seconds for DL imaging.
- Image quality assessed by six readers was rated better ( $P < .001$ ) for the DL images.

However, the question remains whether DL-based image reconstruction can be translated to daily use in actual clinical practice. Studies to date have typically evaluated only perceived image quality instead of true diagnostic information content (3,11,12), and they have generally only considered pieces of the

full diagnostic imaging protocol, which is made up of multiple distinct acquisitions with use of various pulse sequences. Recently, diagnostic interchangeability of DL-accelerated MRI of the knee with a full suite of conventional images was demonstrated (13), but that study used retrospective simulations of accelerated acquisitions. The purpose of the current study was to determine the diagnostic equivalence of prospectively accelerated DL-reconstructed knee MRI compared with conventional MRI for evaluating internal derangement of the knee in a clinical setting.

## Materials and Methods

The authors affiliated with Meta AI Research provided technical support for training of the DL reconstruction model. They had access to anonymized training image data for this purpose. The code to run the model is available at: [https://github.com/facebookresearch/fastMRI/tree/main/fastmri\\_examples/RadiologyJohnson2022](https://github.com/facebookresearch/fastMRI/tree/main/fastmri_examples/RadiologyJohnson2022).

## Study Sample

This study was approved by our institutional review board and was compliant with the Health Insurance Portability and Accountability Act. The acquisition of training data and the prospective test data were approved by our institutional review board.

**Table 1: Conventional (Acquired for Training and Test Sets) and DL (Acquired for Test Set) MRI Sequence Parameters**

Sequence Parameters	Acquired Data				
Conventional accelerated protocol for knee MRI at 3 T					
Protocol	Ax T2 FS	Cor PD	Cor PD FS	Sag PD	Sag T2 FS
Voxel size (mm × mm)	0.4 × 0.4	0.4 × 0.4	0.4 × 0.4	0.4 × 0.4	0.4 × 0.4
FOV (mm × mm)	140 × 140	140 × 140	140 × 140	140 × 140	140 × 140
Section thickness (mm)	3	3	3	3	3
Matrix size	320 × 320	320 × 320	320 × 320	320 × 320	320 × 320
TR (msec)*	4000–5280	2300–2630	2600–3570	2300–2800	4300–5250
TE (msec)	65	27	33	22	46–50
Turbo factor	9	4	4	4	11
No. of sections*	36–50	31–48	31–48	31–40	31–40
Acceleration	2	2	2	2	2
Calibration lines	30	27	27	26	34
Scan time (min:sec)*	1:20–1:45	1:46–2:01	2:02–2:48	2:23–2:54	1:30–1:50
DL-reconstructed accelerated protocol					
Protocol	Ax T2 FS	Cor PD	Cor PD FS	Sag PD	Sag T2 FS
Voxel size (mm × mm)	0.4 × 0.4	0.4 × 0.4	0.4 × 0.4	0.4 × 0.4	0.4 × 0.4
FOV (mm × mm)	140 × 140	140 × 140	140 × 140	140 × 140	140 × 140
Section thickness (mm)	3	3	3	3	3
Matrix size	320 × 320	320 × 320	320 × 320	320 × 320	320 × 320
TR (msec)*	4000–5280	2300–2630	2600–3570	2300–2800	4300–5250
TE (msec)	65	27	33	22	46–50
Turbo factor	9	4	4	4	11
No. of sections*	36–50	31–48	31–48	31–40	31–40
Acceleration	4	4	4	4	4
Calibration lines	22	18	18	18	19
Scan time (min:sec)*	0:48–1:03	0:58–1:06	1:08–1:33	1:16–1:32	0:52–1:03

Note.—For parameters that varied between patients, the parameter range is given. Ax = axial, Cor = coronal, DL = deep learning, FOV = field of view, FS = fat saturated, PD = proton density weighted, Sag = sagittal, TE = echo time, TR = repetition time, T2 = T2 weighted.

\* Data are ranges.

**Retrospective training and validation set.**—We obtained the training data from 298 knee examinations acquired with commercial 3-T MRI systems (MAGNETOM Skyra and Biograph mMR, Siemens Healthineers) between February and March 2019. Given the retrospective nature of this data collection, it received a waiver of consent from our institutional review board. Examinations were performed with the conventional protocol (sequence parameters are listed in Table 1) and consisted of five two-dimensional turbo spin-echo pulse sequences performed in the sagittal, coronal, and axial planes. Each sequence in this protocol used GeneRalized Autocalibrating Partially Parallel Acquisitions (14) parallel imaging with an acceleration factor of 2.

**Prospective test set.**—Written informed consent was obtained from all participants. We prospectively enrolled the test set—a convenience sample of patients referred for diagnostic knee MRI between January 2020 and February 2021. Patients referred for diagnostic MRI of the knee who were 18 years of age or older and had no metal implants or contraindications for MRI were eligible for the study. An additional criterion was that the patient had to be able to fit in the knee coil. All examinations were performed with a single 3-T MRI system (MAGNETOM Skyra, Siemens Healthineers) with use of a 15-channel knee coil (sequence parameters listed in Table 1). Participants were evaluated for meniscal and ligament injuries as well as bone marrow and cartilage abnormalities and underwent the conventional protocol followed by our accelerated DL protocol.

The conventional protocol has an acceleration factor of 2, and the DL protocol an acceleration factor of 4. The center lines of k-space remain fully sampled in both protocols, as these are used for estimation of coil sensitivities. An acceleration factor of 4 was chosen based on results from a prior study that demonstrates a drop in image quality at higher accelerations (15).

### Model Training

We trained a single DL model with use of training data from all five individual sequences. This is a deviation from the original variational network results (3), where individual networks were trained for each particular sequence, image contrast, and orientation. We chose to train a single network due to the more streamlined setup of the reconstruction pipeline. The model architecture is shown in Figure S1, and further details about model training can be found in Appendix S1.

### Image Assessment

Each examination was reviewed by six fellowship-trained subspecialized musculoskeletal radiologists (D.J.L., M.K., G.C., E.A., M.S., W.R.W., all with between 2 and 20 years of subspecialty experience). All readers read all examinations and did so independently. The readers were blinded to all patient information and reconstruction methods. The interpretation scheme consisted of each reader reviewing one group of 10–15 examinations at a time. These groups were made up of both conventional and DL examinations. To limit the potential for recall bias, interpretations of the conventional and DL examinations for each participant were separated by a period of at

**Table 2: Participant Demographic and Referral Information for the 170 Participant Test Set**

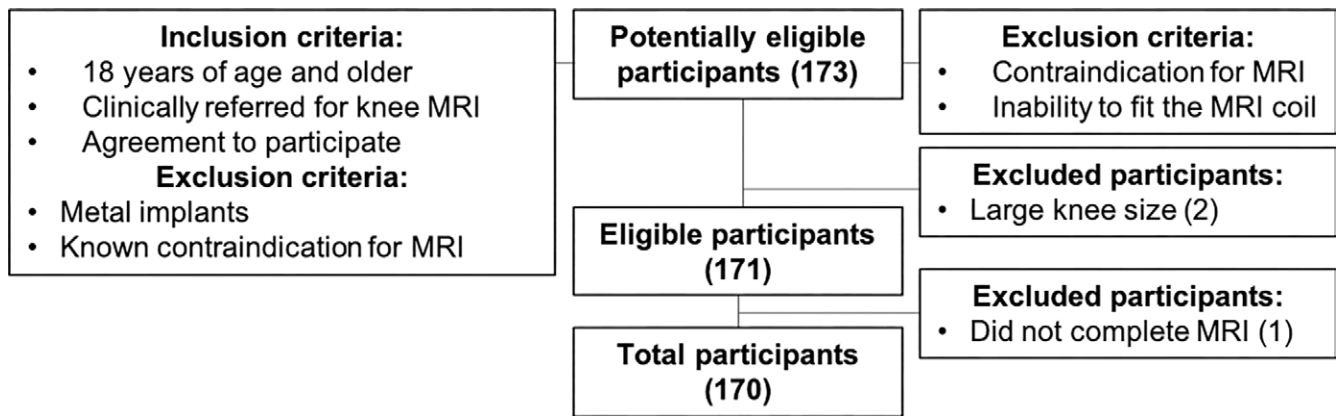
Participant Test Set	No. of Participants
Total participants	170
Demographics	
F	94
M	76
Age (y)*	18–76 (45 ± 16)
Referring department	
Orthopedic surgery	153
Sports medicine	4
Rehabilitation	10
Rheumatology	2
Internal medicine	1
Reason for referral	
Concern for meniscus tear	37
Concern for ACL tear	4
Concern for MCL tear	1
Concern for cartilage wear	3
Concern for Baker cyst	2
Concern for stress injury	2
Concern for meniscal tear, cartilage wear	4
Concern for lateral meniscal tear, iliotibial band pathologic feature	3
Concern for partial quadriceps tendon tear, chondromalacia	1
Concern for ACL and MCL tears	1
Concern for ACL tear, meniscus tear	1
Concern for lateral meniscal tear, iliotibial band pathologic feature, LCL sprain	1
Concern for distal biceps femoris tendon injury	1
Concern for thickened medial plica	1
Preoperative evaluation	1
Patellofemoral pain syndrome	1
Instability	2
Patellar instability	2
Trauma and/or injury	23
Fracture follow-up	1
Osteoarthritis	2
Prior knee surgery, new symptoms	11
Pain	64

Note.—Data are numbers of participants unless otherwise noted. ACL = anterior cruciate ligament, LCL = lateral collateral ligament, MCL = medial collateral ligament.

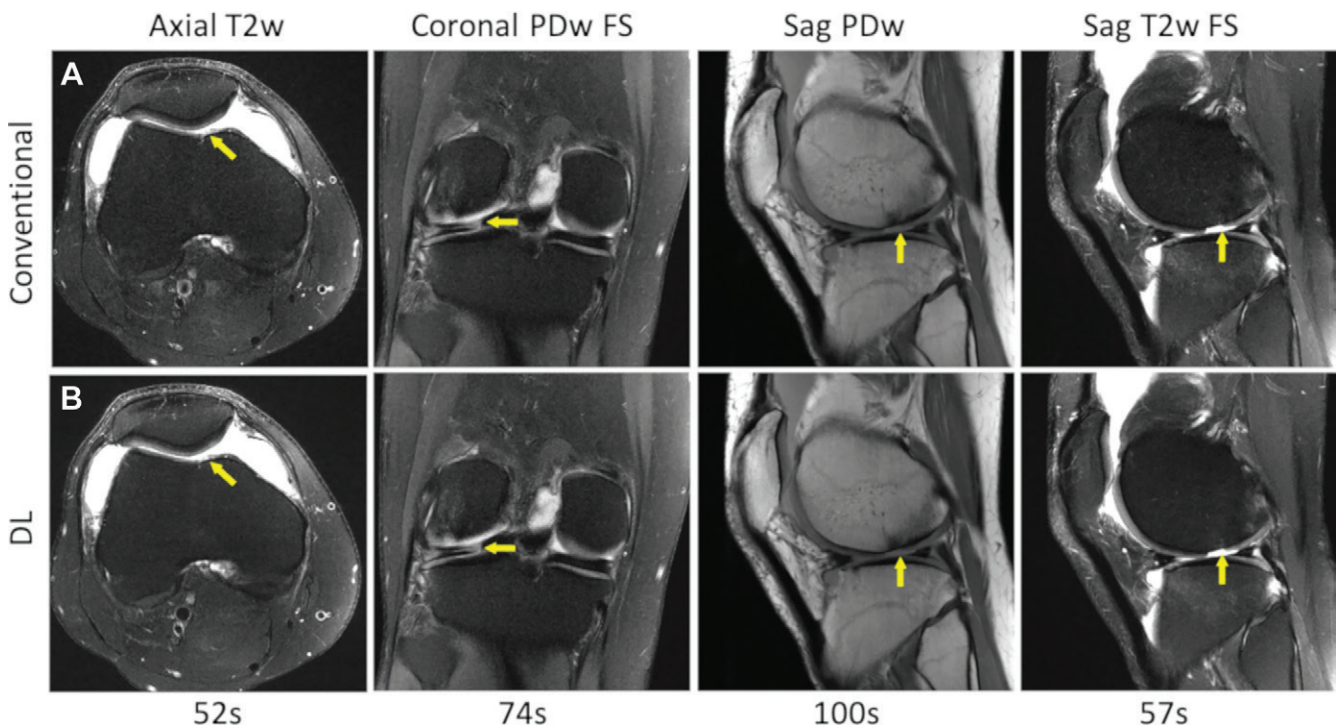
\* Data are ranges, with means ± SDs in parentheses.

least 4 weeks, and the readers were blinded to the other readers' evaluations.

Analyses pertaining to the detection of meniscal or ligament tears and bone marrow or cartilage abnormalities were based on four-point ordinal scores (1, definitely absent; 2, probably absent; 3, probably present; and 4, definitely present). A binary indicator of the presence versus absence of a tear or abnormality was derived by recoding the ordinal scores of 3 and 4 as 1 for present and scores of 1 and 2 as 0 for absent. Agreement between



**Figure 1:** Flow diagram shows participants in the study.



**Figure 2:** Image reconstruction in a 26-year-old man referred for clinical knee MRI with lateral femoral condyle, medial trochlear chondral defects, lateral femoral condyle bone marrow abnormality, lateral meniscal tear, and joint effusion. The pathologic feature (indicated by the yellow arrows) in this case can be seen equally well in (A) the conventional protocol and (B) the deep learning (DL) protocol. The scan planes and contrast are indicated above the images. DL scan times are shown below the images. FS = fat saturated, PDw = proton density weighted, Sag = sagittal, T2w = T2 weighted.

the two protocols in terms of detection was characterized by the percentage of times concordant opinions were obtained when a single reader assessed the same patient with use of both protocols. Agreement among readers was characterized in terms of the percentage of times concordant opinions were obtained when two readers independently assessed the same patient with use of the same protocol.

The images were also evaluated for image quality with use of a four-point Likert scale. For both artifacts and sharpness, the scoring system was: 1, none; 2, mild; 3, moderate; and 4, severe artifacts or blurring. Readers were asked to look for motion, aliasing, flow artifact, or any other type of artifact they observed.

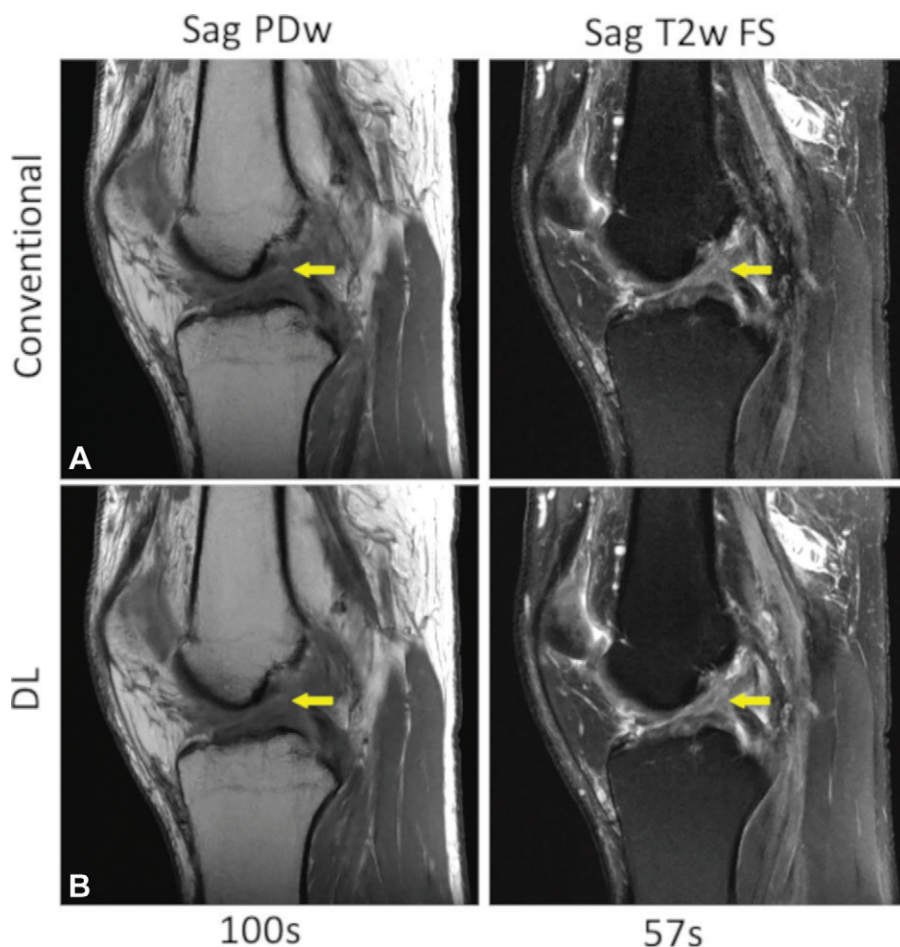
Perceived signal-to-noise ratio and overall image quality were scored as follows: 1, excellent; 2, good; 3, fair; and 4, poor.

### Statistical Analysis

Given the absence of a reference standard (surgical ground truth) for most patients, the utility of the DL protocol relative to the conventional protocol for the detection of an abnormality was assessed in terms of interchangeability (16). Specifically, interchangeability was assessed by comparing the percentage of times concordant opinions were given for the same patient by two independent readers with use of the conventional protocol ( $P_0$ ) and the percentage of times concordant opinions were rendered by the same two readers when one reader used the conventional protocol and the other reader used the DL protocol ( $P_1$ ).

The DL protocol was declared interchangeable with the conventional protocol if the upper limit of the 95% CI for the difference was less than 5%, as this would imply 95% confidence





**Figure 3:** Image reconstruction in a 52-year-old man referred for clinical knee MRI with an anterior cruciate ligament tear. The ligament tear (indicated by the yellow arrows) can be seen equally well in (A) the conventional protocol and (B) the deep learning (DL) protocol. The scan planes and contrast are indicated above the images. DL scan times are shown below the images. FS = fat saturated, PDw = proton density weighted, Sag = sagittal, T2w = T2 weighted.

that interchanging the protocols would result in a less than 5% decrease in reader agreement (16).

Statistical analysis was performed with use of SAS version 9.4 software (SAS Institute) by one author (J.S.B., with 26 years of experience). Generalized estimating equations based on binary logistic regression were used to test and derive a 95% CI for the difference of  $P_0 - P_1$ . Additional details for this test can be found in Appendix S1. Where  $P$  values are calculated, statistical significance was determined to be  $P < .05$ .

## Results

### Characteristics of Study Sample

Retrospective training and validation set included images from 298 clinical 3-T examinations that were used for model training. We divided the examinations randomly into a training set ( $n = 242$ ) and a validation set ( $n = 56$ ).

Prospective test set included a total of 170 participants (mean age  $\pm$  SD, 45 years  $\pm$  16; 76 men) who were enrolled for the prospective analysis. Demographic and referral information for the study participants are provided in Table 2. The majority of

participants were referred for MRI from orthopedic surgery (153 of 170 participants). A smaller number of participants were referred from sports medicine, rehabilitation, rheumatology, and internal medicine. The most common reason for referral was knee pain (64 of 170 participants). A total of 52 subjects were referred out of concern for specific abnormalities (eg, meniscus tear [37 of 170 participants]). Acute trauma and/or injury were identified as the reason for referral for 23 of 170 subjects. A flowchart illustrating participant enrollment and patient exclusion is shown in Figure 1.

### Scan Times

The total mean scan time for the conventional protocol was 9 minutes 56 seconds  $\pm$  19 seconds. The mean scan time for our DL protocol was 5 minutes 33 seconds  $\pm$  16 seconds.

### Image Results

The accelerated data were reconstructed with our DL model. Inference computation times for the DL model were between 145 msec and 255 msec per section, depending on the matrix size of the particular sequence. The reconstructed images were then sent to picture archiving and communication systems for clinical evaluation. Image results are shown in Figures 2 and 3. In

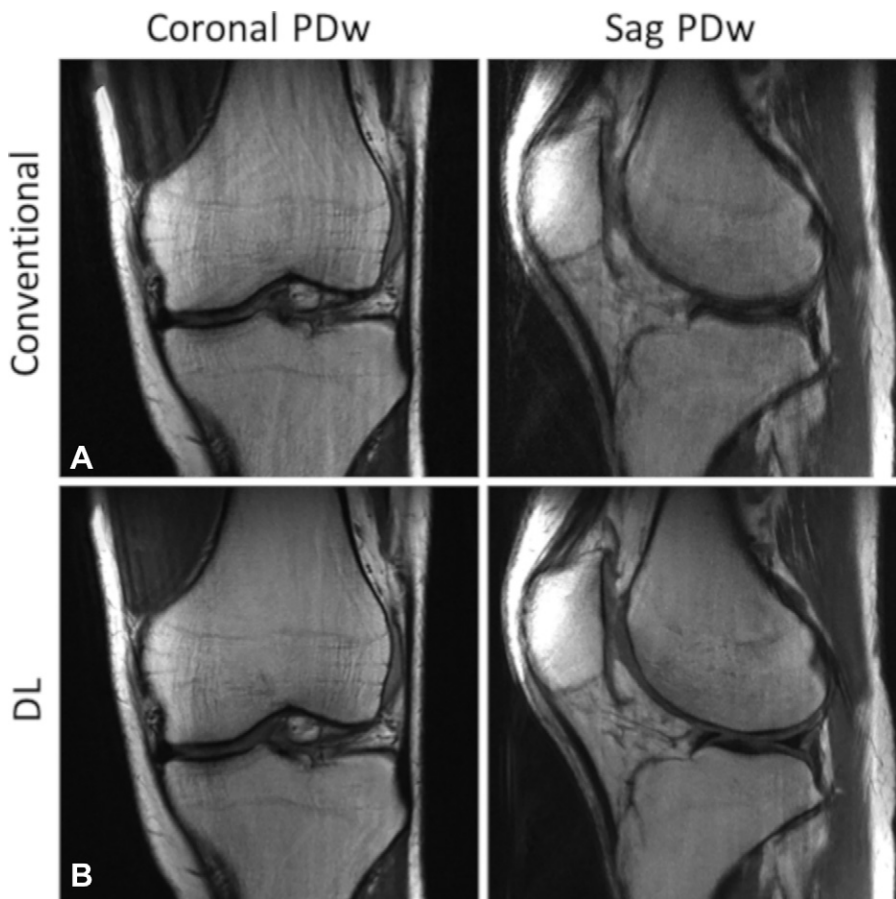
both figures, a pathologic feature is clearly visible in both the conventional and the DL protocols. Figure 4 illustrates an example in which the DL examination has fewer artifacts than the conventional protocol. This illustrates the additional benefit that shorter scans are less susceptible to patient motion and the resulting image artifacts. While image acceleration does not necessarily lead to decreased motion artifacts, fewer examinations will be degraded by motion due to the decreased likelihood of motion in a shorter examination.

### Clinical Reader Study: Imaging Findings

There were 19 pathologic features assessed and a total of 615 detected pathologic features among the 170 participants. The most common tear was a medial meniscal tear, detected in 64 participants. The most common chondral abnormality was in the patellar cartilage, detected in 85 participants. The total number of participants for which each pathologic feature was detected is reported in Table S1.

### Clinical Reader Study: Interchangeability

Interchangeability results for ordinal scoring of tears and abnormalities are presented in Table 3. Data regarding the



**Figure 4:** Image reconstruction in a 30-year-old man referred for clinical knee MRI with no reader-identified pathologic features illustrating greater motion artifacts in conventional versus deep learning (DL). **(A)** The conventional images in the top row have increased ghosting artifacts and blurring compared with **(B)** the DL reconstruction in the bottom row. The average artifact scores were 3.3 and 2.8 for the conventional and DL protocols, respectively. Longer scans are more susceptible to artifacts due to patient motion. PDw = proton density weighted, Sag = sagittal.

number of times the ordinal scores of 1–4 were reported for each pathologic feature is reported in Figure S2. It can be concluded that the conventional and DL protocols are clinically interchangeable because the upper limit of the 95% CI for the difference (see Table S2 for full results) was less than 5% for all combinations of outcome and feature.

#### Clinical Reader Study: Image Quality

The mean image quality scores for both conventional and DL examinations are presented in Figure 5. The full data for all individual readers are presented in Table S3. Note that lower scores mean better image quality. The mean overall image quality, sharpness, and signal-to-noise scores were better (lower) for the DL protocol than for the conventional protocol ( $P < .001$ ), and it may be concluded that the DL protocol had better image quality. The mean artifact scores were similar—2.10 and 2.11 ( $P = .76$ ) for the conventional and DL protocols, respectively.

#### Interreader Reliability

Measures of interreader reliability are presented in Table S4. Agreement is defined as the percentage of times concordant opinions were obtained when two readers independently as-

essed the same participant with use of the same protocol. The difference in agreement between the conventional and DL protocol was small and did not exceed 2% for any feature; however, it was statistically significant for binary assessment of five of the 19 features. These were for tears of the anterior cruciate ligament ( $P = .03$ ), medial collateral ligament ( $P = .02$ ), and posterior cruciate ligament ( $P = .004$ ), as well as for bone marrow abnormalities in the lateral femoral condyle ( $P = .02$ ) and lateral tibial plateau ( $P = .01$ ).

#### Discussion

In the past 5 years, many research studies have demonstrated that deep learning (DL) methods can outperform classic reconstruction approaches in terms of both quantitative image metrics like mean squared error to ground truth (3,8), as well as in qualitative reader studies (12,13). DL reconstruction has been explored for a range of MRI applications, such as abdominal (12), prostate (17,18), and cardiac (6,19) imaging. The majority of work evaluating DL reconstruction has done so by retrospectively undersampling conventional acquisitions. This approach—used in a prior knee imaging study (13)—does not account for improvements in image quality, such as reduced motion artifacts and blurring,

that may result from the reduced scan time. In this work, we demonstrate that DL reconstructions of prospectively accelerated DL acquisitions are clinically interchangeable with the conventional reconstructions for diagnosis of meniscal or ligament tears and bone marrow or cartilage abnormalities. Additionally, the average reader scores in this study rated the DL reconstructions as having higher overall image quality (DL, 1.44; conventional, 1.71;  $P < .001$ ), higher signal-to-noise ratio (DL, 1.29; conventional, 1.70;  $P < .001$ ), improved sharpness (DL, 1.48; conventional, 1.60;  $P < .001$ ), and similar artifacts (DL, 2.10; conventional, 2.11;  $P = .755$ ) compared with the conventional images. This work adds to a growing body of literature that shows diagnostic equivalence of conventional and DL reconstruction for musculoskeletal imaging (13,20). The improvement in overall image quality with use of DL reconstruction was also demonstrated by Herrmann et al (21) for musculoskeletal imaging.

Our DL-based image reconstruction method is an extension of classic image reconstruction methods that use iterative optimization to solve an ill-posed inverse problem. As with classic methods, our DL model enforces consistency with acquired data and imposes a constraint on the optimization in the form of a regularizer. With DL reconstruction, we use a more powerful

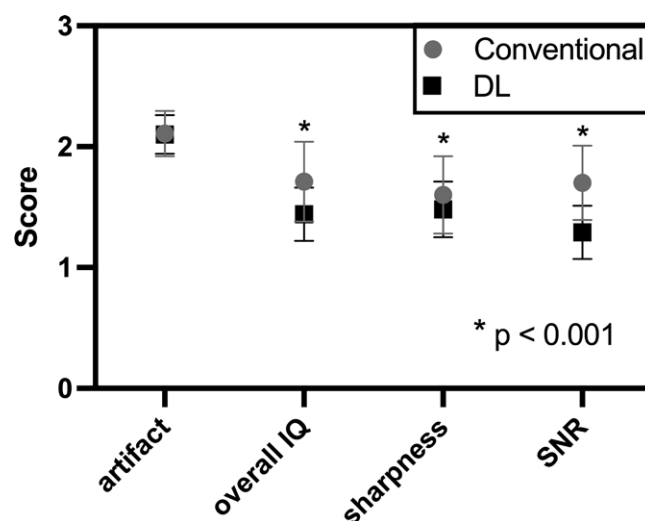
**Table 3: Concordant Opinions Obtained for Two Readers Who Independently Assessed the Same Participant with Use of the Conventional Protocol and Different Protocols**

Outcome	Feature	Same Protocol ( <i>n</i> = 2550)		Different Protocol ( <i>n</i> = 5100)		Interchangeable (Yes vs No)
		Patients (%)	No. of Patients	Patients (%)	No. of Patients	
Ordinal	Anterior cruciate ligament tear	95.8	2444	94.9	4841	Yes
Ordinal	Extensor mechanism tear	98.3	2507	98.2	5007	Yes
Ordinal	Lateral collateral ligament tear	99.3	2531	99.3	5065	Yes
Ordinal	Lateral femoral condyle bone marrow abnormality	95.6	2437	95.1	4851	Yes
Ordinal	Lateral femoral condyle cartilage abnormality	81.3	2074	80.2	4091	Yes
Ordinal	Lateral meniscus tear	88.2	2249	86.9	4432	Yes
Ordinal	Lateral tibial plateau bone marrow abnormality	90.8	2316	89.1	4543	Yes
Ordinal	Lateral tibial plateau cartilage abnormality	82.3	2098	81.1	4135	Yes
Ordinal	Medial collateral ligament tear	97.4	2483	96.7	4930	Yes
Ordinal	Medial femoral condyle bone marrow abnormality	92.9	2368	92.8	4735	Yes
Ordinal	Medial femoral condyle cartilage abnormality	80.0	2040	79.6	4059	Yes
Ordinal	Medial meniscus tear	87.3	2227	86.4	4408	Yes
Ordinal	Medial tibial plateau bone marrow abnormality	89.5	2283	89.1	4542	Yes
Ordinal	Medial tibial plateau cartilage abnormality	82.2	2095	81.5	4156	Yes
Ordinal	Patellar bone marrow abnormality	95.3	2429	95.4	4867	Yes
Ordinal	Patellar cartilage abnormality	88.5	2257	86.9	4431	Yes
Ordinal	Posterior cruciate ligament tear	99.2	2530	99.0	5050	Yes
Ordinal	Trochlear bone marrow abnormality	92.3	2354	91.4	4660	Yes
Ordinal	Trochlear cartilage abnormality	85.5	2180	85.2	4344	Yes

Note.—Test of interchangeability: The protocols were declared interchangeable if the upper limit of the 95% CI for the difference was less than 5%, as this would imply 95% confidence that interchanging the protocols would result in a less than 5% decrease in reader agreement (16). See Table S2 for full data. The percentage (%) and number of times concordant opinions were obtained when two readers independently assessed the same participant with use of the conventional protocol (same protocols) and when two readers independently assessed the same participant with use of different protocols. There were 170 examinations read by six readers. For same protocols, this results in 15 possible pairs (*n* = 2550 participants). For different protocols, there are 30 possible pairs (*n* = 5100 participants).

and more general regularizer that is learned from training data rather than using handcrafted regularizers, such as the sparsity constraints used in compressed sensing (1). Our method is a physics-based model that uses a learned convolutional neural network to regularize the reconstruction; this sets it apart from image synthesis and simulation methods that generate images without raw data. It has been shown that DL reconstructions may be susceptible to perturbations of image data that create large errors in the reconstructed images (5). However, these perturbations were intended to be “worst-case” anomalies in the acquired data and are not modeled off realistic scanner and acquisition errors that could occur in routine clinical imaging.

There are three limitations of our study design. The first is the absence of a surgical ground truth. While we are able to show that the conventional and DL protocols are clinically interchangeable, we cannot compare the true diagnostic accuracy of the two protocols. The second limitation is the binary or four-point ordinal scale assessment of pathologic features, which may not reflect the descriptive nature of typical clinical interpretations. This grading system was chosen specifically to evaluate conspicuity (definitely absent, probably absent, probably present, and definitely present) of pathologic features in the conventional and DL-accelerated protocols. A third limitation of the study design is that the sample sizes of 2550 and 5100 participants for the generalizing estimating



**Figure 5:** Plot shows means and SDs of image quality (IQ) scores for conventional and accelerated examinations. Scores are averaged over six readers. For both artifacts and sharpness, the scoring system was: 1, none; 2, mild; 3, moderate; and 4, severe artifacts or blurring. Perceived signal-to-noise ratio (SNR) and overall IQ were scored as follows: 1, excellent; 2, good; 3, fair; and 4, poor. DL = deep learning.



equations analysis are inflations of the true sample size of 170 cases, due to the inclusion of data from multiple reader pairs per case. This inflation was required to achieve adequate statistical power for the test of interchangeability. While the generalizing estimating equations analysis accounted for the inflation by assuming that results derived for the same case were positively correlated rather than independent, the analysis only achieved adequate power for the test pooled over all reader pairs, and the presented results are not informative regarding any individual pair of readers.

Our study also had two limitations that may limit the generalizability of the presented method. First, our DL model was trained for a single acceleration factor and a single anatomy and is unlikely to generalize beyond fourfold accelerated knee examinations. Also, it may not generalize to rare findings or unique cases that are not well represented by the training set (eg, when there is metal in or around the knee). Second, examinations were performed with a single 3-T MRI scanner from a single vendor.

Deep learning (DL) reconstruction enables a reduction in scan time for a knee examination from approximately 10 minutes to approximately 5.5 minutes, and the inference time is less than 10 seconds per volume. This reduction in scan time will reduce the overall room time required for a patient examination, thereby reducing examination costs and increasing patient throughput. Variational network reconstruction enables rapid acquisition and results in improved image quality, with equivalent diagnostic information. This work demonstrates that DL reconstruction of accelerated knee MRI at 3 T does work reliably in a real clinical setting with real patients. Important future steps include correlating reader assessments of pathologic findings with a surgical ground truth and establishing robustness across a variety of scanners and vendors.

**Acknowledgments:** We thank Brian Domin, Jessica Tuite, Shane Lugo, and Britany LoGrande for their help with subject recruitment and scanning.

**Author contributions:** Guarantors of integrity of entire study, P.M.J., J.Z., F.K.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, P.M.J., C.L.Z., M.M., T.P., D.K.S., M.P.R., F.K.; clinical studies, P.M.J., D.J.L., M.K., G.C., M.S., W.R.W., L.C., M.P.R., F.K.; experimental studies, P.M.J., D.J.L., J.Z., C.L.Z., A.S., M.K., T.P., D.K.S., F.K.; statistical analysis, J.Z., M.M., J.S.B., F.K.; and manuscript editing, P.M.J., D.J.L., A.S., M.M., J.S.B., M.K., G.C., E.A., M.S., W.R.W., T.P., D.K.S., M.P.R., F.K.

**Disclosures of conflicts of interest:** P.M.J. No relevant relationships. D.J.L. No relevant relationships. J.Z. Research engineer at Meta. C.L.Z. No relevant relationships. A.S. No relevant relationships. M.M. Stock from employment. J.S.B. Statistical consultant on the *Radiology* editorial board. M.K. Patents planned, issued, or pending with NYU. G.C. No relevant relationships. E.A. No relevant relationships. M.S. No relevant relationships. W.R.W. No relevant relationships. L.C. No relevant relationships. T.P. No relevant relationships. D.K.S. Patent royalties from General Electric; patent royalties and license fees from Bruker; consulting fees from Q Bio and Ezra; patent on image reconstruction using machine learning (US10671939B2); stock options in Ezra. M.P.R. No relevant relationships. F.K. Issued patent (US20170309019A1); stock or stock options in Subtle Medical.

## References

1. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med* 2007;58(6):1182–1195.
2. Wang G. A Perspective on Deep Imaging. *IEEE Access* 2016;4:8914–8924.
3. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med* 2018;79(6):3055–3071.
4. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans Med Imaging* 2018;37(2):491–503.
5. Antun V, Renna F, Poon C, Adcock B, Hansen AC. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc Natl Acad Sci U S A* 2020;117(48):30088–30095.
6. Vishnevskiy V, Walheim J, Kozerke S. Deep variational network for rapid 4D flow MRI reconstruction. *Nat Mach Intell* 2020;2(4):228–235.
7. Zhu B, Liu JZ, Cauley SE, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature* 2018;555(7697):487–492.
8. Aggarwal HK, Mani MP, Jacob M. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Trans Med Imaging* 2019;38(2):394–405.
9. Zbontar J, Knoll F, Sriram A, et al. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *arXiv 1811.08839 [preprint]* <https://arxiv.org/abs/1811.08839>. Posted November 21, 2018. Accessed November 15, 2021.
10. Knoll F, Zbontar J, Sriram A, et al. fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning. *Radiol Artif Intell* 2020;2(1):e190007.
11. Knoll F, Murrell T, Sriram A, et al. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. *Magn Reson Med* 2020;84(6):3054–3070.
12. Chen F, Taviani V, Malkiel I, et al. Variable-Density Single-Shot Fast Spin-Echo MRI with Deep Learning Reconstruction by Using Variational Networks. *Radiology* 2018;289(2):366–373.
13. Recht MP, Zbontar J, Sodickson DK, et al. Using Deep Learning to Accelerate Knee MRI at 3 T: Results of an Interchangeability Study. *AJR Am J Roentgenol* 2020;215(6):1421–1429.
14. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med* 2002;47(6):1202–1210.
15. Sriram A, Zbontar J, Tullie M, et al. End-to-End Variational Networks for Accelerated MRI Reconstruction. In: Martel AL, Abolmaesumi P, Stoyanov D, et al, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. MICCAI 2020. Lecture Notes in Computer Science, vol 12262. Springer, 2020; 64–73.
16. Obuchowski NA, Subhas N, Schoenhagen P. Testing for interchangeability of imaging tests. *Acad Radiol* 2014;21(11):1483–1489.
17. Gassenmaier S, Afat S, Nickel D, Mostapha M, Herrmann J, Othman AE. Deep learning-accelerated T2-weighted imaging of the prostate: Reduction of acquisition time and improvement of image quality. *Eur J Radiol* 2021;137:109600.
18. Johnson PM, Tong A, Donthireddy A, et al. Deep Learning Reconstruction Enables Highly Accelerated Biparametric MR Imaging of the Prostate. *J Magn Reson Imaging* 2022;56(1):184–195.
19. Küstner T, Fuin N, Hammernik K, et al. CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. *Sci Rep* 2020;10(1):13710.
20. Hahn S, Yi J, Lee HJ, et al. Image Quality and Diagnostic Performance of Accelerated Shoulder MRI With Deep Learning-Based Reconstruction. *AJR Am J Roentgenol* 2022;218(3):506–516.
21. Herrmann J, Koerzdoerfer G, Nickel D, et al. Feasibility and Implementation of a Deep Learning MR Reconstruction for TSE Sequences in Musculoskeletal Imaging. *Diagnostics (Basel)* 2021;11(8):1484.