



# Evaluation of deep learning reconstructed high-resolution 3D lumbar spine MRI

Simon Sun<sup>1</sup> · Ek Tsoon Tan<sup>1</sup> · Douglas N. Mintz<sup>1</sup> · Meghan Sahr<sup>1</sup> · Yoshimi Endo<sup>1</sup> · Joseph Nguyen<sup>1</sup> · R. Marc Lebel<sup>2</sup> · John A. Carrino<sup>1</sup> · Darryl B. Sneag<sup>1</sup>

Received: 2 December 2021 / Revised: 24 February 2022 / Accepted: 5 March 2022 / Published online: 24 March 2022  
© The Author(s), under exclusive licence to European Society of Radiology 2022

## Abstract

**Objectives** To compare interobserver agreement and image quality of 3D T2-weighted fast spin echo (T2w-FSE) L-spine MRI images processed with a deep learning reconstruction (DLRecon) against standard-of-care (SOC) reconstruction, as well as against 2D T2w-FSE images. The hypothesis was that DLRecon 3D T2w-FSE would afford improved image quality and similar interobserver agreement compared to both SOC 3D and 2D T2w-FSE.

**Methods** Under IRB approval, patients who underwent routine 3-T lumbar spine (L-spine) MRI from August 17 to September 17, 2020, with both isotropic 3D and 2D T2w-FSE sequences, were retrospectively included. A DLRecon algorithm, with denoising and sharpening properties was applied to SOC 3D k-space to generate 3D DLRecon images. Four musculoskeletal radiologists blinded to reconstruction status evaluated randomized images for motion artifact, image quality, central/foraminal stenosis, disc degeneration, annular fissure, disc herniation, and presence of facet joint cysts. Inter-rater agreement for each graded variable was evaluated using Conger's kappa ( $\kappa$ ).

**Results** Thirty-five patients (mean age  $58 \pm 19$ , 26 female) were evaluated. 3D DLRecon demonstrated statistically significant higher median image quality score (2.0/2) when compared to SOC 3D (1.0/2,  $p < 0.001$ ), 2D axial (1.0/2,  $p < 0.001$ ), and 2D sagittal sequences (1.0/2,  $p$  value  $< 0.001$ ).  $\kappa$  ranges (and 95% CI) for foraminal stenosis were 0.55–0.76 (0.32–0.86) for 3D DLRecon, 0.56–0.73 (0.35–0.84) for SOC 3D, and 0.58–0.71 (0.33–0.84) for 2D. Mean  $\kappa$  (and 95% CI) for central stenosis at L4-5 were 0.98 (0.96–0.99), 0.97 (0.95–0.99), and 0.98 (0.96–0.99) for 3D DLRecon, 3D SOC and 2D, respectively.

**Conclusions** DLRecon 3D T2w-FSE L-spine MRI demonstrated higher image quality and similar interobserver agreement for graded variables of interest when compared to 3D SOC and 2D imaging.

## Key Points

- 3D DLRecon T2w-FSE isotropic lumbar spine MRI provides improved image quality when compared to 2D MRI, with similar interobserver agreement for clinical evaluation of pathology.
- 3D DLRecon images demonstrated better image quality score (2.0/2) when compared to standard-of-care (SOC) 3D (1.0/2),  $p$  value  $< 0.001$ ; 2D axial (1.0/2),  $p$  value  $< 0.001$ ; and 2D sagittal sequences (1.0/2),  $p$  value  $< 0.001$ .
- Interobserver agreement for major variables of interest was similar among all sequences and reconstruction types. For foraminal stenosis,  $\kappa$  ranged from 0.55 to 0.76 (95% CI 0.32–0.86) for 3D DLRecon, 0.56–0.73 (95% CI 0.35–0.84) for standard-of-care (SOC) 3D, and 0.58–0.71 (95% CI 0.33–0.84) for 2D.

**Keywords** Deep learning · Imaging · Three-dimensional · Spine · Magnetic resonance imaging

## Abbreviations

DL	Deep learning	MPR	Multiplanar reformat
DLRecon	Deep learning reconstruction	SOC	Standard of care
L-spine	Lumbar spine	T2w-FSE	T2-weighted fast spin echo
		3D	Three-dimensional
		2D	Two-dimensional

✉ Darryl B. Sneag  
sneagd@hss.edu

<sup>1</sup> Department of Radiology and Imaging, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021, USA

<sup>2</sup> GE Healthcare, Waukesha, WI, USA

## Introduction

Deep learning (DL) applications in radiology have significantly evolved through recently developed convolutional neural

networks that facilitate automated classification, segmentation, and disease detection [1–3]. Uses of DL techniques in spine imaging include automating lumbar vertebrae numbering, classifying disc herniations, and grading spinal stenosis [4–6]. Additionally, DL applications in image reconstruction have been shown to improve image quality via noise and artifact reduction, and super-resolution reconstruction [7], particularly for knee MRI [8] and coronary CT angiography [9].

DL image reconstruction of 2D MRI has only recently been introduced on some MRI vendor platforms [10, 11]. DL image reconstructions for 3D MRI remain largely unexplored. Clinical lumbar spine (L-spine) protocols typically include multiplanar (sagittal, axial, and sometimes coronal), two-dimensional (2D) T2-weighted fast spin echo (T2w-FSE) acquisitions with total acquisition times nominally reaching 25 min [12–14]. In comparison, a single 3D T2w-FSE isotropic sequence can generate the same planes as 2D imaging via multiplanar reformations but with significantly shorter total acquisition time (6–8 min) and without interslice gaps, avoiding partial volume averaging effects [12, 13, 15]. The 3D acquisition also enables creation of “true” cross-sectional views orthogonal to the obliquely oriented neural foramina, which may facilitate more accurate assessment of neural foraminal stenosis, particularly in patients with scoliosis [12, 13]. However, due in part to signal-to-noise (SNR) limits, the achievable spatial resolution of 3D images has to date remained inferior to 2D T2w-FSE using standard reconstruction techniques [16].

Applying DL reconstruction to improve 3D image quality may potentially mitigate resolution and SNR drawbacks but requires comparisons of diagnostic performance prior to routine use. Observer performance in radiology is an important source of variability and has been previously investigated in interpretation of L-spine MRI, showing reasonable

performance for characterization of degenerative findings such as spinal stenosis, spondylolisthesis, disc degeneration, facet arthrosis, and Modic endplate changes [17–21].

This study’s objective was to evaluate the application of a DLRecon algorithm [22, 23] to enhance 3D T2w-FSE L-spine MRI. We hypothesized that in addition to overall decreased acquisition time compared to a 2D protocol, DLRecon 3D T2w-FSE would also demonstrate overall improved image quality and similar interobserver agreement compared to both standard-of-care (SOC) 3D reconstructed and SOC 2D images.

## Materials and methods

This retrospective study was approved by our institutional review board and informed subject consent was waived.

### Image acquisition

An L-spine MRI protocol was developed to include 3D and 2D T2-weighted FSE sequences optimized for spatial resolution, contrast, and scan time (Table 1). The new protocol was utilized for clinical imaging for 1 month (between August 17 and September 17, 2020) on all L-spine exams performed on either of two identical 3-T MRI scanners (Signa Premier, GE Healthcare) using combined 60-channel posterior table and 30-channel anterior arrays (GE AIR© Coil). The sequences evaluated included (1) isotropic 3D T2w-FSE (CUBE); (2) 2D T2w-FSE axial (parallel to L4-5 disc space); and (3) 2D T2w-FSE sagittal (Table 1). A vendor-supplied prototype of AIR Recon DL 3D [22–25] was used to reconstruct 3D SOC exams offline. The prototype reconstruction performs denoising, deringing, and interpolation in all three directions with the goal of providing a sharp, clean image volume that

**Table 1** Lumbar spine MRI parameters

Parameter	3D isotropic sagittal T2-weighted imaging	Axial 2D T2-weighted imaging	Sagittal 2D T2-weighted imaging
TR/TE	1550/90	2500–3500/110	2500–3500/110
Flip angle (°)	Variable	142	142
FOV (mm)	260	280	280
Matrix size	328 × 328 × 162	416 × 224	512 × 256
Slice thickness (mm)	0.8	3.5	3.5
Receiver bandwidth (kHz)	± 62.50	± 83.33	± 83.33
Echo train length	50	12	14
Number of signals averaged	1	1.5	1.5
Parallel imaging factor	2 × 1.75	-	-
Acquisition time	6.3–7.2 min <sup>a</sup>	4.2 min <sup>b</sup>	3.5 min <sup>b</sup>

<sup>a</sup> Total acquisition time for entire 3D MRI study: axial and coronal planes were generated with high-quality multiplanar reformations without requiring additional acquisition time

<sup>b</sup> Acquisition time for single planar sequence of 2D MRI study (entire study requires multiple 2D planar acquisitions to cover all planes)

can be effectively reformatted in any orientation. The network was trained using a supervised learning approach using pairs of pristine and typical datasets. Similar to the 2D version of AIR Recon DL [22–25], training data contained diverse content and feature contrasts. DLRecon was performed on a separate Unix workstation (Intel Xeon 12-core, 2.10 GHz, 64GB RAM), with an average reconstruction time of 221 min.

### Case inclusion

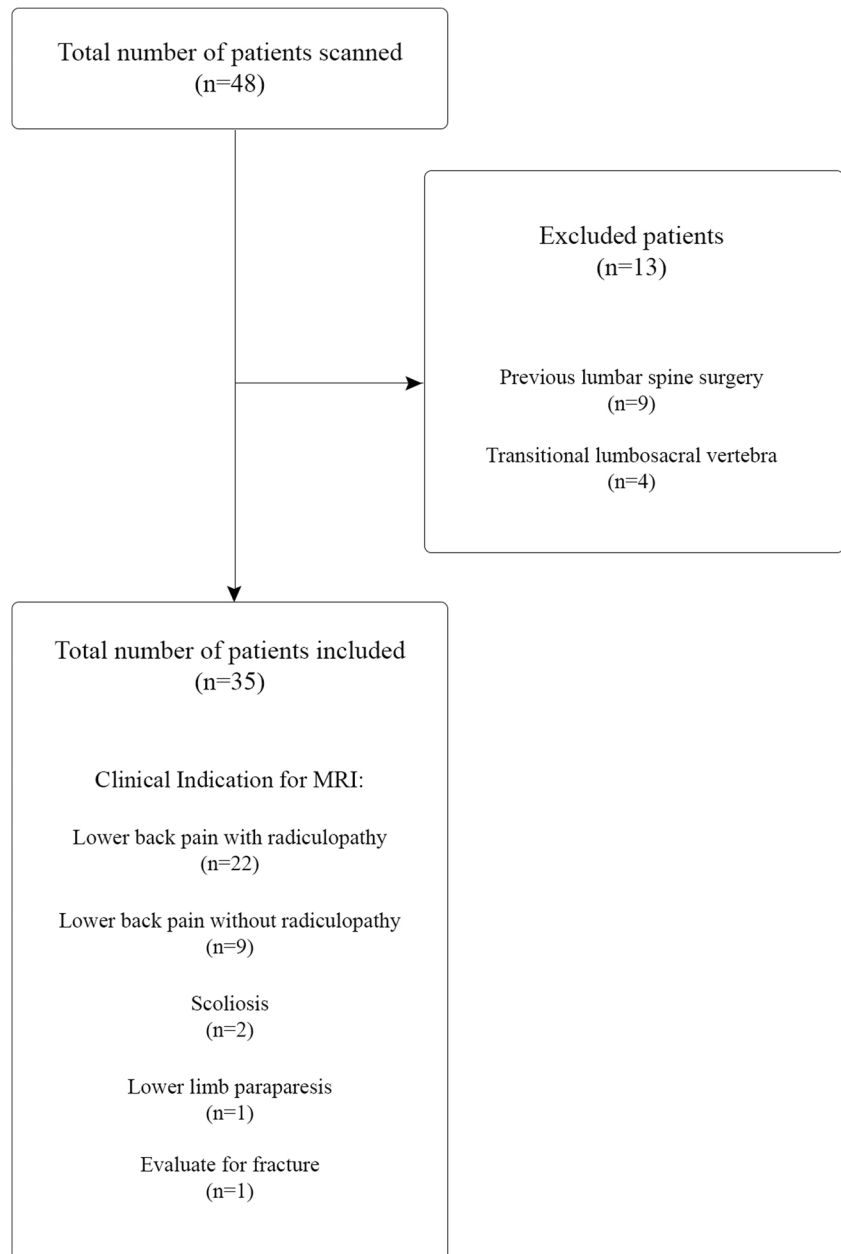
A total of 48 subjects were scanned and identified for analysis. Checks were performed to confirm offline storage of the 3D complex k-space data and consistency of scan parameters.

Exclusion criteria were absence of 2D sequences or 3D k-space data, history of previous L-spine surgery, and presence of transitional lumbosacral anatomy, giving  $n = 35$  subjects for final analysis (Fig. 1). This cohort size met the required sample size calculation based on an a priori power analysis for a hypothesized  $\kappa$  of 0.6 for a precision level of 0.15 (95% CI 0.45–0.75) for each variable of interest, based on reported prevalence from standardized grading scales used in the literature [20, 26–29].

### Observers

3D DLRecon, 3D SOC, and 2D SOC images for these 35 patients were anonymized and then randomized for blinded

**Fig. 1** Subject inclusion/exclusion diagram



**Table 2** Comparison of median and interquartile range [IQR] image quality (grades: 0–2) and motion artifact scores (grades: 0–3)

	3D DLRecon		3D SOC		2D axial		2D sagittal	
	Image quality	Motion artifact	Image quality	Motion artifact	Image quality	Motion artifact	Image quality	Motion artifact
Median [IQR]	2.0 [1.0–2.0]	0.0 [0.0–1.0]	1.0 [1.0–1.0]	1.0 [0.0–1.0]	1.0 [1.0–2.0]	1.0 [0.0–1.0]	1.0 [1.0–1.0]	1.0 [0.0–1.0]
Paired comparison <i>p</i> value								
vs. 3D SOC	< 0.001	0.001	-	-	-	-	-	-
vs. 2D axial	< 0.001	< 0.001	0.005	0.493	-	-	-	-
vs. 2D sagittal	< 0.001	< 0.001	0.822	0.123	-	-	-	-

DLRecon deep learning reconstruction, SOC standard-of-care, IQR interquartile range

evaluation by four readers: three musculoskeletal fellowship-trained attending radiologists (D.N.M., 23 years of experience; M.S., 4 years of experience; Y.E., 11 years of experience) and a musculoskeletal radiology fellow (S.S.). Prior to independent grading, each observer attended a training session led by a separate musculoskeletal radiologist (D.B.S. with 8 years of dedicated MRI experience) and was provided a handbook containing grading scales for each imaging variable of interest validated by current literature with pictorial examples.

### Image evaluation

Images were independently reviewed on dedicated picture archiving and communication system workstations (SECTRA IDS7, Sectra AB) with diagnostic quality monitors. Overall image quality was subjectively assessed on a 3-point grading scale (poor, average, or excellent). Motion artifact was subjectively assessed on a 4-point grading scale (absent, mild, moderate, or severe).

Central stenosis was determined via cross-sectional area measurement of the thecal sac at the L4-5 mid-disc level, categorized as absent ( $\geq 130$  mm<sup>2</sup>), mild ( $< 130$  mm<sup>2</sup> and  $\geq 100$  mm<sup>2</sup>), moderate ( $< 100$  mm<sup>2</sup> and  $\geq 65$  mm<sup>2</sup>), or severe ( $< 65$  mm<sup>2</sup>) [30]. Foraminal stenosis at L3-4, L4-5, and L5-S1 was graded using multiplanar reformations

(MPRs) to achieve orthogonal orientation to the foramen with level of severity based on degree of effacement of the fat in a 4-quadrant division of the neural foramina: none, mild, mild-to-moderate, moderate, moderate-to-severe, severe (0, 1, 1.5, 2–2.5, 3, 4 quadrants effaced, respectively) [31]. Disc degeneration at L3-4, L4-5, and L5-S1 was graded on a 4-point ordinal scale modified from Pfirrmann et al [20] to combine grades 1 and 2 due to the low prevalence of grade 1 [26]. Facet joint synovial cysts at L3-4 and L4-5 were documented only if ventral in location and causing impingement of intra- or extrathecal nerve roots. Annular fissures at L3-4 and L4-5 were considered present if demonstrating higher signal intensity than the nucleus pulposus on T2w imaging, and their laterality (right, left) and position (central, paracentral, subarticular, foraminal, extraforaminal) were recorded. Finally, the presence and position of a disc bulge or disc herniation and the latter's type (protrusion, extrusion, sequestration) at L3-4 and L4-5 were recorded.

### Statistical analysis

Analysis of interobserver reliability within image sequences was conducted using Conger's kappa ( $\kappa$ ) or intraclass correlation coefficient (central stenosis only). Mann-Whitney *U* tests were employed for comparison of summary measures between image sequences. Tukey's post hoc adjustments were

**Table 3** Interobserver agreement (Conger's  $\kappa$ ) for clinical variables of interest

	DLRecon 3D (95% CI)	Standard 3D (95% CI)	2D (95% CI)
Image quality	0.26 (0.08–0.47)	0.28 (0.09–0.50)	0.18–0.23 (0.04–0.43)
Motion artifact	0.44 (0.25–0.62)	0.32(0.15–0.51)	0.04–0.19 (–0.07 to 0.05)
Central stenosis (L4-5)	0.98 (0.96–0.99)	0.97 (0.95–0.99)	0.98 (0.96–0.99)
Neural foraminal stenosis (L3-4, L4-5, L5-S1)	0.55–0.76 (0.32–0.86)	0.56–0.73 (0.35–0.84)	0.58–0.71 (0.33–0.84)
Disc degeneration (L3-4, L4-5, L5-S1)	0.79–0.82 (0.65–0.90)	0.81–0.84 (0.68–0.92)	0.81–0.87 (0.68–0.93)
Disc herniation (L3-4, L4-5)	0.28–0.52 (0.12–0.69)	0.30–0.56 (0.13–0.72)	0.17–0.62 (0.03–0.76)
Annular fissure (L3-4, L4-5)	0.26–0.65 (0.11–0.78)	0.20–0.57 (0.06–0.72)	0.14–0.63 (0.004–0.76)
Facet synovial cyst (L3-4, L4-5)	0.22–0.80 (0.06–0.88)	0.22–0.67 (0.06–0.79)	0–1 (–0.11 to 1)

**Table 4** Comparison of median grading values and interquartile range for clinical variables of interest between each sequence

		Median [IQR]	Paired comparison <i>p</i> value		
			vs. 3D SOC	vs. 2D	
3D DLRecon	Central stenosis (L4-5)	0.0 [0.0–2.0]	0.70	0.87	
	Left neural foraminal stenosis	L3-4	1.0 [0.0–1.0]	0.95	0.59
		L4-5	1.0 [0.0–2.0]	> 0.99	0.96
		L5-S1	1.0 [0.0–2.0]	0.69	0.35
	Right neural foraminal stenosis	L3-4	1.0 [0.0–2.0]	0.96	0.66
		L4-5	1.0 [1.0–2.0]	0.83	0.68
		L5-S1	1.0 [0.0–2.0]	0.93	0.38
	Disc degeneration	L3-4	3.0 [1.0–4.0]	0.76	0.97
		L4-5	3.0 [3.0–4.0]	0.90	0.82
		L5-S1	3.0 [1.3–4.0]	0.87	0.61
	Left disc herniation	L3-4	0.0 [0.0–0.0]	0.95	0.38
		L4-5	0.0 [0.0–1.0]	0.79	0.11
	Right disc herniation	L3-4	0.0 [0.0–0.0]	0.94	> 0.99
		L4-5	0.0 [0.0–1.0]	> 0.99	<b>0.04</b>
	Central disc herniation	L3-4	1.0 [0.0–1.0]	> 0.99	0.47
		L4-5	1.0 [0.0–1.0]	0.87	0.80
	Left annular fissure	L3-4	0.0 [0.0–1.0]	0.57	0.58
		L4-5	0.0 [0.0–0.0]	> 0.99	0.17
	Right annular fissure	L3-4	0.0 [0.0–0.0]	0.70	<b>0.02</b>
		L4-5	0.0 [0.0–1.0]	0.10	<b>0.01</b>
	Central annular fissure	L3-4	0.0 [0.0–0.0]	0.69	0.25
		L4-5	0.0 [0.0–0.0]	0.61	0.78
	Left facet synovial cyst	L3-4	0.0 [0.0–0.0]	0.99	0.08
		L4-5	0.0 [0.0–0.0]	0.48	0.48
Right facet synovial cyst	L3-4	0.0 [0.0–0.0]	<b>&lt; 0.01</b>	<b>&lt; 0.01</b>	
	L4-5	0.0 [0.0–0.0]	0.66	0.36	
3D SOC	Central stenosis (L4-5)	0.0 [0.0–2.0]	-	0.57	
	Left neural foraminal stenosis	L3-4	1.0 [0.0–2.0]	-	0.63
		L4-5	1.0 [0.0–2.0]	-	0.96
		L5-S1	1.0 [0.0–2.0]	-	0.60
	Right neural foraminal stenosis	L3-4	1.0 [0.0–2.0]	-	0.62
		L4-5	1.0 [1.0–2.0]	-	0.52
		L5-S1	1.0 [0.0–2.0]	-	0.44
	Disc degeneration	L3-4	3.0 [1.0–4.0]	-	0.73
		L4-5	3.0 [3.0–4.0]	-	0.72
		L5-S1	3.0 [2.3–4.0]	-	0.50
	Left disc herniation	L3-4	0.0 [0.0–0.0]	-	0.35
		L4-5	0.0 [0.0–1.0]	-	0.06
	Right disc herniation	L3-4	0.0 [0.0–0.0]	-	0.95
		L4-5	0.0 [0.0–1.0]	-	<b>0.04</b>
	Central disc herniation	L3-4	1.0 [0.0–1.0]	-	0.48
		L4-5	1.0 [0.0–1.0]	-	0.94
	Left annular fissure	L3-4	0.0 [0.0–0.0]	-	> 0.99
		L4-5	0.0 [0.0–1.0]	-	0.17
	Right annular fissure	L3-4	0.0 [0.0–1.0]	-	<b>0.04</b>
		L4-5	0.0 [0.0–0.0]	-	0.31
	Central annular fissure	L3-4	0.0 [0.0–0.0]	-	0.12
		L4-5	0.0 [0.0–0.0]	-	> 0.99
	Left facet synovial cyst	L3-4	0.0 [0.0–0.0]	-	0.08
		L4-5	0.0 [0.0–0.0]	-	> 0.99
Right facet synovial cyst	L3-4	0.0 [0.0–0.0]	-	<b>&lt; 0.01</b>	
	L4-5	0.0 [0.0–0.0]	-	0.63	

**Table 4** (continued)

			Median [IQR]	Paired comparison <i>p</i> value	
				vs. 3D SOC	vs. 2D
2D	Central stenosis (L4-5)		1.0 [0.0–2.0]	-	-
	Left neural foraminal stenosis	L3-4	1.0 [0.0–1.0]	-	-
		L4-5	1.0 [0.0–2.0]	-	-
		L5-S1	1.0 [0.0–2.0]	-	-
	Right neural foraminal stenosis	L3-4	1.0 [0.0–2.0]	-	-
		L4-5	1.0 [1.0–2.0]	-	-
		L5-S1	1.0 [0.3–3.0]	-	-
	Disc degeneration	L3-4	3.0 [1.0–4.0]	-	-
		L4-5	3.0 [3.0–4.0]	-	-
		L5-S1	3.0 [1.0–4.0]	-	-
	Left disc herniation	L3-4	0.0 [0.0–0.0]	-	-
		L4-5	0.0 [0.0–0.0]	-	-
	Right disc herniation	L3-4	0.0 [0.0–0.0]	-	-
		L4-5	0.0 [0.0–0.0]	-	-
	Central disc herniation	L3-4	0.0 [0.0–1.0]	-	-
		L4-5	0.0 [0.0–1.0]	-	-
	Left annular fissure	L3-4	0.0 [0.0–0.0]	-	-
		L4-5	0.0 [0.0–0.0]	-	-
	Right annular fissure	L3-4	0.0 [0.0–0.0]	-	-
		L4-5	0.0 [0.0–0.0]	-	-
	Central annular fissure	L3-4	0.0 [0.0–0.0]	-	-
		L4-5	0.0 [0.0–0.0]	-	-
	Left facet synovial cyst	L3-4	0.0 [0.0–0.0]	-	-
		L4-5	0.0 [0.0–0.0]	-	-
	Right facet synovial cyst	L3-4	0.0 [0.0–0.0]	-	-
		L4-5	0.0 [0.0–0.0]	-	-

DLRecon deep learning reconstruction, SOC standard-of-care

Central stenosis (0 = none, 1 = mild, 2 = moderate, 3 = severe), neural foraminal stenosis (0 = none, 1 = mild, 2 = mild to moderate, 3 = moderate, 4 = moderate to severe, 5 = severe), disc degeneration (1 = Pfirmann grades 1 or 2, 3 = Pfirmann grade 3, 4 = Pfirmann grade 4, 5 = Pfirmann grade 5), disc herniation (0 = none, 1 = bulge, 2 = protrusion, 3 = extrusion, 4 = sequestration), annular fissure (0 = absent, 1 = present), facet synovial cyst (0 = absent, 1 = present); *p*-values reflect two significant figures

used for the adjustment of multiple comparisons. All analyses were performed using SAS version 9.4 (SAS Institute Inc.). *p* values of < 0.05 were deemed statistically significant.  $\kappa$  values indicated agreement on the basis of values suggested by Altman [32] and adapted from the method of Landis and Koch [33]: 0.81–1.00 indicated very good agreement; 0.61–0.80, good agreement; 0.41–0.60, moderate agreement; 0.21–0.40, fair agreement; and 0.20 or lower, poor agreement.

## Results

Of the 48 initial patient scans, 9 were excluded due to previous surgery, and 4 due to transitional lumbosacral anatomy, yielding a final number of  $n = 35$  (Fig. 1). Images from 35 patients (mean age  $58 \pm 19$  years, 26 female and 9 male) were evaluated by all 4 readers. Clinical indications for undergoing

lumbar spine MRI included lower back pain with radiculopathy ( $n = 22$ ), lower back pain without radiculopathy ( $n = 9$ ), scoliosis ( $n = 2$ ), lower limb paraparesis ( $n = 1$ ), and evaluation of fracture ( $n = 1$ ) (Fig. 1).

Median image quality was significantly higher for 3D DLRecon (2.0 out of a maximum score of 2) versus 3D SOC (1.0 out of 2), *p* value < 0.001, and for 3D DLRecon versus both 2D axial (1.0 out of 2), *p* value < 0.001 and 2D sagittal (1.0 out of 2), *p* value < 0.001 (Table 2). 3D SOC showed significantly lower image quality (1.0 out of a maximum score of 2) when compared to 2D axial (1 out of 2 as well), *p* value 0.005, but not when compared to 2D sagittal, *p* value = 0.822. Evaluation of motion artifact showed significantly lower median motion artifact for 3D DLRecon (0.0 out of a maximum score of 3) versus 3D SOC (1.0 out of 3), *p* value = 0.001, and also versus 2D axial (1.0 out of 3), *p* value < 0.001 and 2D sagittal (1.0 out of 3), *p* value < 0.001

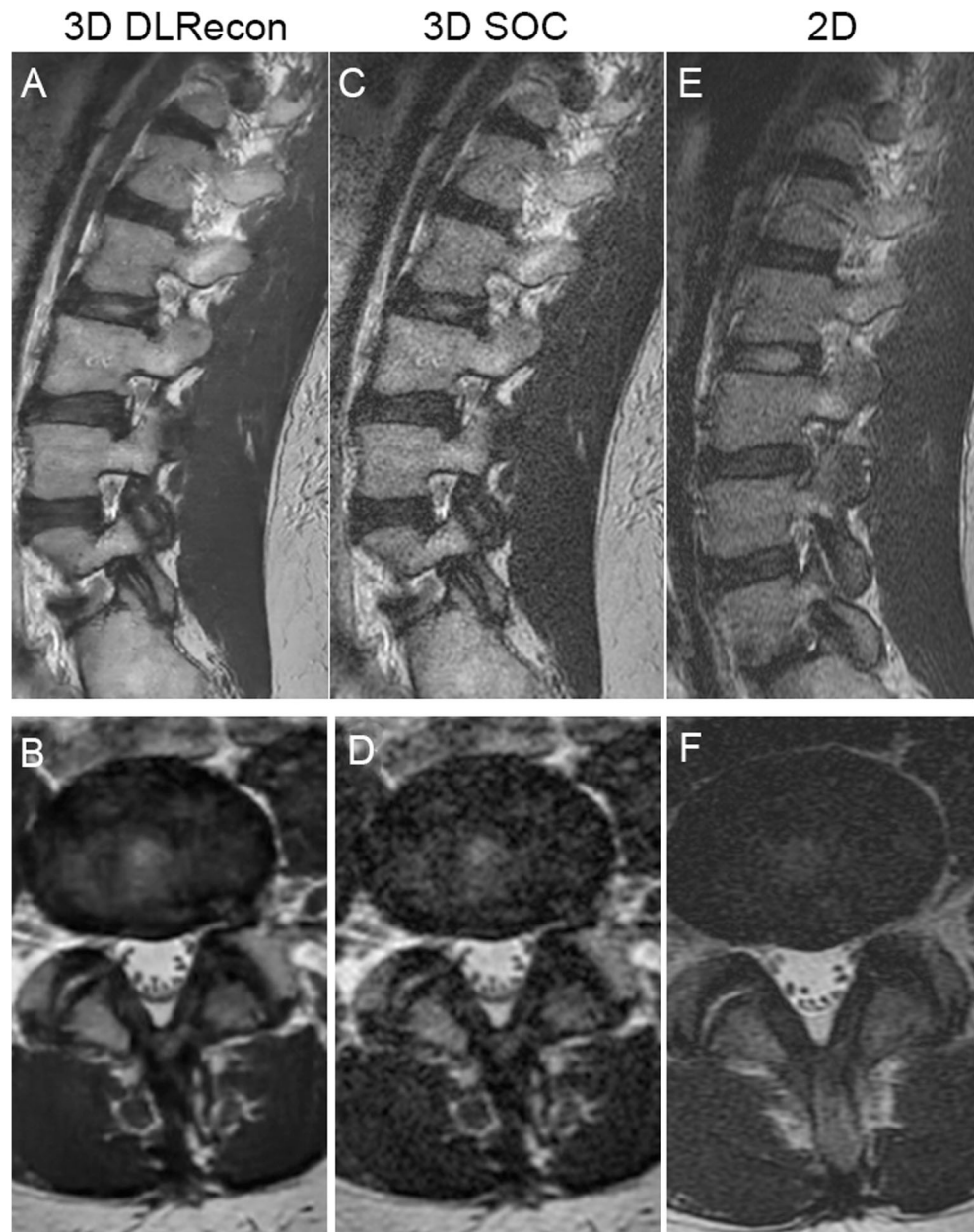
(Table 2). Motion artifact in 3D SOC was similar to 2D axial ( $p$  value = 0.493) and 2D sagittal ( $p$  value = 0.123).

Interobserver agreement for each variable of interest with corresponding 95% confidence intervals is shown in Table 3. There was very good agreement for central stenosis ( $\kappa = 0.97$ – $0.98$ ) for all sequences. Interobserver agreement of neural foraminal stenosis was moderate to good ( $\kappa = 0.55$ – $0.76$ ) for 3D DLRecon, comparable to both 3D SOC ( $\kappa = 0.56$ – $0.73$ ) and 2D ( $\kappa = 0.58$ – $0.71$ ); substantial for grading disc degeneration ( $\kappa = 0.79$ – $0.82$ ) for 3D DLRecon, comparable to both 3D SOC ( $\kappa = 0.81$ – $0.84$ ) and 2D ( $\kappa = 0.81$ – $0.87$ ); fair to moderate for grading disc herniation ( $\kappa = 0.28$ – $0.69$ ) for 3D DLRecon, comparable to 3D SOC ( $\kappa = 0.30$ – $0.56$ ), but with a one grade difference versus 2D ( $\kappa = 0.17$ – $0.62$ ); fair to moderate for

annular fissures ( $\kappa = 0.26$ – $0.65$ ) for 3D DLRecon, comparable to both 3D SOC ( $\kappa = 0.20$ – $0.57$ ), but with a one grade difference versus 2D ( $\kappa = 0.14$ – $0.53$ ); and fair to good for facet joint cysts ( $\kappa = 0.22$ – $0.80$ ) for 3D DLRecon comparable to 3D SOC ( $\kappa = 0.22$ – $0.67$ ), but different from 2D ( $\kappa = 0$ – $1$ ).  $\kappa$  values for image quality ranged from fair for 3D DLRecon (0.26) and 3D SOC (0.28) to poor for 2D (0.18–0.23), whereas motion artifact  $\kappa$  values were moderate for 3D DLRecon (0.44), fair for 3D SOC (0.32), and poor for 2D (0.04–0.19).

Comparison of the mean grading values for each variable of interest between all 3 sequences did not demonstrate any statistically significant difference for central or foraminal stenosis ( $p > 0.05$ ) (Table 4). Some occurrences of significant differences were noted for disc herniations for 3D DLRecon

**Fig. 2** Lumbar spine MRI of a 41-year-old female with lower back pain comparing 3D T2-weighted FSE with DLRecon (deep learning reconstruction) sagittal (A) and axial (B) images, and without (standard-of-care, SOC) sagittal (C) and axial (D) images, along with 2D T2-weighted FSE sagittal (E) and axial (F) images



vs. 2D (right L4-5), annular fissures for 3D DLRecon vs. 2D (right L3-4, right L4-5) and 3D SOC vs. 2D (right L3-4), and facet synovial cysts for 3D DLRecon and 3D SOC vs. 2D (right L3-4) and 3D DLRecon vs. 3D SOC (right L3-4).

## Discussion

Isotropic 3D MRI provides overall time savings and the ability to produce high-quality MPRs, but the associated SNR and spatial resolution loss, given equivalent scan times, are known trade-offs compared to 2D imaging [16]. In our study, we applied a 3D DLRecon algorithm to 3D T2w-FSE MRI to improve its image quality and evaluated its performance against SOC 3D and 2D imaging (Fig. 2).

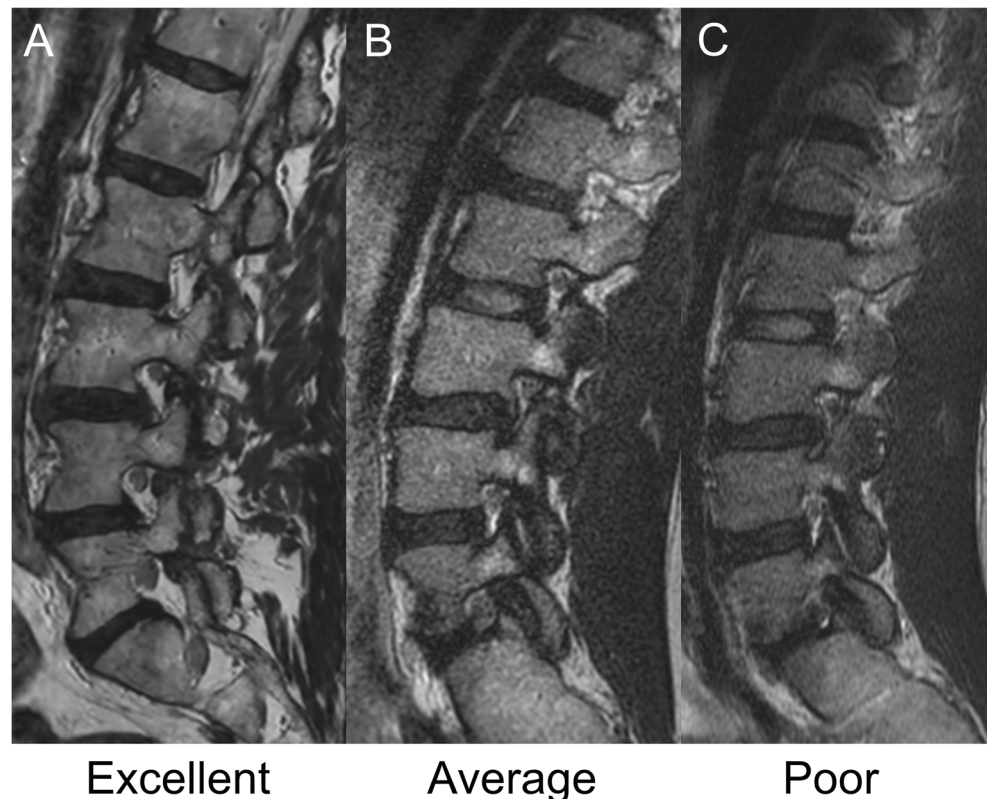
3D DLRecon had higher image quality than 3D SOC ( $p < 0.001$ ) and 2D imaging ( $p < 0.001$ ), despite only having fair agreement between observers; this supports the hypothesis that DLRecon may help address the image quality shortcomings of 3D SOC-reconstructed FSE images (Fig. 3). 3D SOC images were inferior to 2D axial, but not to 2D sagittal, which may be related to 3D view ordering that preferentially results in more FSE-related blurring in the axial plane. DLRecon also showed statistically significant less motion artifact when compared to 2D, despite having only fair interobserver agreement and increased chance of motion [12] as the scan time of 3D

acquisition was approximately 2–3 times longer compared to the 2D acquisition (Fig. 4). This may be explained by how motion artifacts more favorably manifest in 3D compared to 2D imaging [34], possibly due to slice-selection effects in 2D and differences in view-ordering. DLRecon also demonstrated significantly less motion compared to SOC images, which may be attributed to the reduced noise and increased sharpness with DLRecon.

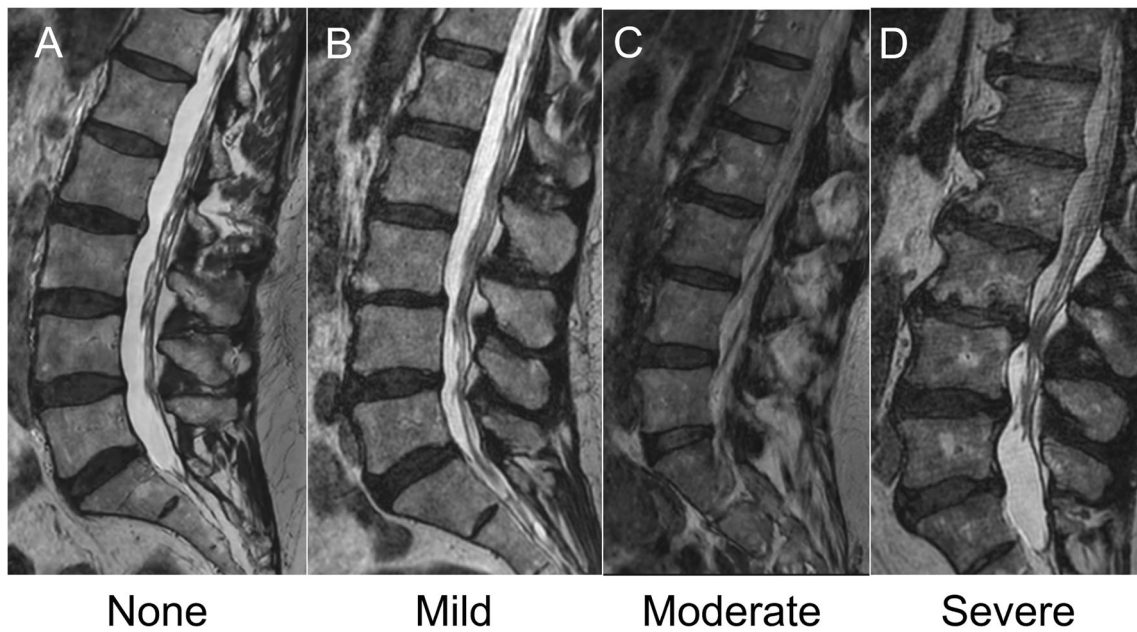
Overall, interobserver agreement for the variables of interest in our study was comparable between all 3 sequences and when compared to previous studies of interobserver variability from the literature. In evaluating central and neural foramina stenosis, moderate to very good agreement and no statistical difference were noted between 2D and 3D. Notably, the very good interobserver agreement between 2D and 3D for central stenosis ( $\kappa = 0.97$ – $0.98$ ) was higher than the  $\kappa = 0.809$  previously reported by Lee et al [12]; this result may be explained by either the shorter echo train length in our study (50 vs. 145) that improves image sharpness, or differences in grading systems, as we performed a quantitative measurement of thecal sac area instead of the qualitative grading method previously validated by Schizas et al [18].

Neural foraminal stenosis for 3D DLRecon demonstrated moderate to good agreement ( $\kappa = 0.55$ – $0.76$ ), which is slightly lower, but still comparable to what has been reported in the literature for 3D MRI ( $\kappa = 0.809$ ) [12, 35]. This may be due to the 6 levels of discrimination used in our grading scale [31],

**Fig. 3** Image quality grading scale and representative sagittal T2-weighted FSE images from 48-year-old (A) and 41-year-old (B, C) patients showing an image graded as excellent for image quality (A), average for image quality (B), and poor for image quality (C)







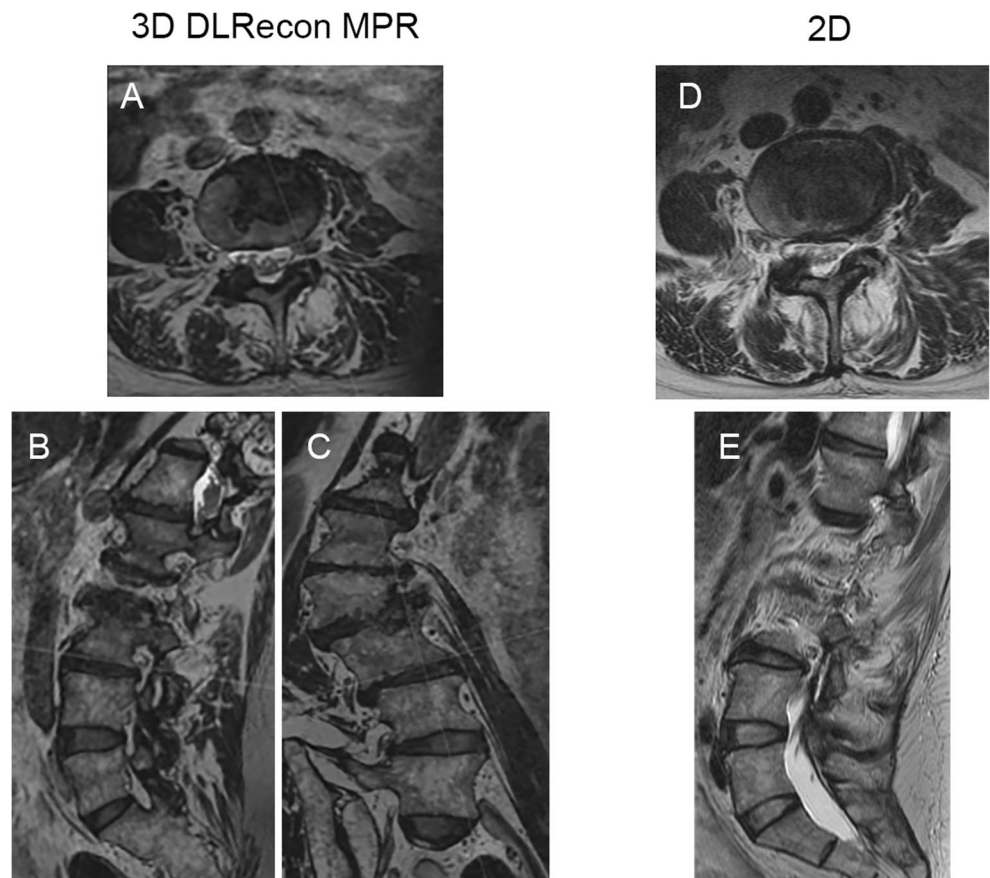
**Fig. 4** Motion artifact grading scale and representative sagittal T2-weighted FSE images from a 63-year-old female imaging graded as demonstrating no motion artifact (A), from a 32-year-old female graded as

mild motion artifact (B), from a 60-year-old graded as moderate motion artifact (C), and from a 64-year-old male graded as severe motion artifact (D)

and also used in clinical practice at the authors’ institution, versus previously a reported scale of 4 grades [35]. We did not observe improved  $\kappa$  for foraminal stenosis on 3D versus

2D MRI, as other studies have found as well (10, 11). The L-spine neural foramina may, in general, not demonstrate enough obliquity, compared to the cervical spine, to fully

**Fig. 5** Comparison of 3D deep learning reconstructed (DLRecon) images of the lumbar spine versus 2D images in a 53-year-old female with severe scoliosis. Axial (A), sagittal (B), and coronal (C) 3D DLRecon T2-weighted FSE multiplanar reformations (MPRs) facilitate the creation of optimal planes orthogonal to the exiting left L3 nerve root at the level of the L3-4 neural foramen in this patient with severe scoliosis. 2D T2-weighted FSE axial (D) and sagittal (E) acquisitions centered at the level of the L3-4 neural foramen in this patient with severe scoliosis demonstrate suboptimal evaluation of the exiting left L3 nerve root due to the inability to make use of MPRs



benefit from the use of 3D MPRs. By excluding cases with prior surgery, we may have also limited the proportion of patients with more complex anatomy, including those with significant scoliosis, in whom 3D MPR could have demonstrated superiority to conventional 2D planes (Fig. 5).

We acknowledge several limitations of this retrospective study. There was some variability in 2D imaging parameters (predominantly TR) due to heterogeneity in scan plane obliquity and specific absorption rate calculations, but 3D acquisition parameters were held constant. Imaging was performed only of the lumbar spine at a single field strength (3 T), and therefore protocols may not be generalizable to imaging at 1.5 T or to the cervical or thoracic spine without further optimization. While powered for this analysis, the study sample size was relatively small ( $n = 35$ ) and patients with previous lumbar spine surgery were excluded, which limits the applicability of our findings to patients with spinal hardware who may not be indicated for imaging at 3 T (vs. 1.5 T). Additionally, the DLRecon algorithm was not applied to 2D images, which could have improved image quality scores of 2D images but was not done as DLRecon is still not widely available across all vendors and institutions. Finally, the rather long current offline reconstruction time of the DLRecon algorithm (221 min) does not currently allow for practical use in an emergent clinical setting.

In summary, applying a 3D DLRecon image algorithm to isotropic 3D MRI of the L-spine improved image quality when compared to SOC 3D and 2D L-spine MRI without compromising interobserver agreement for evaluation of clinically relevant pathology. Our institution plans to routinely adopt the 3D DLRecon sequence as soon as the final version of the software becomes commercialized, which will hopefully also further improve reconstruction time. The resulting improved image quality and MPRs could provide additional diagnostic benefits over standard 2D imaging in other anatomic regions, such as the cervical spine, in which neural foramina are more obliquely oriented.

**Acknowledgements** We would like to acknowledge Yan Wen (GE Healthcare) and Maggie Fung (GE Healthcare) for their important contributions to the 3D deep learning reconstruction implementation. We would also like to acknowledge Jake Fiore for his assistance with data collection.

**Funding** The Hospital for Special Surgery receives institutional research support from GE Healthcare.

## Declarations

**Guarantor** The scientific guarantor of this publication is Darryl B. Sneag.

**Conflict of interest** Hospital for Special Surgery receives institutional research support from GE Healthcare. All authors acknowledge that they have no personal investment in the software evaluated in this study.

**Statistics and biometry** Joseph Nguyen kindly provided statistical advice for this manuscript. One of the authors has significant statistical expertise.

**Informed consent** Written informed consent was waived by the institutional review board.

**Ethical approval** Institutional review board approval was obtained.

## Methodology

- retrospective
- diagnostic study
- performed at one institution

## References

1. McBee MP, Awan OA, Colucci AT et al (2018) Deep learning in radiology. *Acad Radiol* 25:1472–1480
2. Mazurowski MA, Buda M, Saha A, Bashir MR (2019) Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI. *J Magn Reson Imaging* 49:939–954
3. Rajpurkar P, Irvin J, Zhu K et al (2017) CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv: 1711.05225
4. Castro-Mateos I, Hua R, Pozo JM, Lazary A, Frangi AF (2016) Intervertebral disc classification by its degree of degeneration from T2-weighted magnetic resonance images. *Eur Spine J* 25:2721–2727
5. Hallinan JTPD, Zhu L, Yang K et al (2021) Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. *Radiology* 300:130–138
6. Lewandrowski K-U, Muralledharan N, Eddy SA et al (2020) Feasibility of deep learning algorithms for reporting in routine spine magnetic resonance imaging. *Int J Spine Surg* 14:S86–S97
7. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K (2019) Improvement of image quality at CT and MRI using deep learning. *Jpn J Radiol* 37:73–80
8. Chaudhari AS, Fang Z, Kogan F et al (2018) Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med* 80:2139–2154
9. Tatsugami F, Higaki T, Nakamura Y et al (2019) Deep learning-based image restoration algorithm for coronary CT angiography. *Eur Radiol* 29:5322–5329
10. Kidoh M, Shinoda K, Kitajima M et al (2020) Deep learning based noise reduction for brain MR imaging: tests on phantoms and healthy volunteers. *Magn Reson Med Sci* 19:195–206
11. Hammernik K, Klatzer T, Kobler E et al (2018) Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med* 79:3055–3071
12. Lee S, Jee WH, Jung JY, Lee SY, Ryu KS, Ha KY (2015) MRI of the lumbar spine: comparison of 3D isotropic turbo spin-echo SPACE sequence versus conventional 2D sequences at 3.0 T. *Acta Radiol* 56:174–181
13. Blizzard DJ, Haims AH, Lischuk AW, Arunakul R, Hustedt JW, Grauer JN (2015) 3D-FSE isotropic MRI of the lumbar spine: novel application of an existing technology. *J Spinal Disord Tech* 28:152–157
14. Sayah A, Jay AK, Toaff JS, Makariou EV, Berkowitz F (2016) Effectiveness of a rapid lumbar spine MRI protocol using 3D T2-weighted SPACE imaging versus a standard protocol for evaluation

- of degenerative changes of the lumbar spine. *AJR Am J Roentgenol* 207:614–620
15. Rodegerdts EA, Boss A, Riemarzik K et al (2006) 3D imaging of the whole spine at 3T compared to 1.5T: initial experiences. *Acta Radiol* 47:488–493
  16. Ristow O, Steinbach L, Sabo G et al (2009) Isotropic 3D fast spin-echo imaging versus standard 2D imaging at 3.0 T of the knee—image quality and diagnostic performance. *Eur Radiol* 19:1263–1272
  17. Carrino JA, Lurie JD, Tosteson AN et al (2009) Lumbar spine: reliability of MR imaging findings. *Radiology* 250:161–170
  18. Schizas C, Theumann N, Burn A et al (2010) Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. *Spine (Phila Pa 1976)* 35:1919–1924
  19. Lurie JD, Tosteson AN, Tosteson TD et al (2008) Reliability of readings of magnetic resonance imaging features of lumbar spinal stenosis. *Spine (Phila Pa 1976)* 33:1605–1610
  20. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N (2001) Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine (Phila Pa 1976)* 26:1873–1878
  21. Weishaupt D, Zanetti M, Boos N, Hodler J (1999) MR imaging and CT in osteoarthritis of the lumbar facet joints. *Skelet Radiol* 28:215–219
  22. Lebel RM (2020) Performance characterization of a novel deep learning-based MR image reconstruction pipeline.
  23. Kim M, Kim HS, Kim HJ et al (2021) Thin-slice pituitary MRI with deep learning-based reconstruction: diagnostic performance in a postoperative setting. *Radiology* 298:114–122
  24. Zochowski KC, Tan ET, Argentieri EC et al (2021) Improvement of peripheral nerve visualization using a deep learning-based MR reconstruction algorithm. *Magn Reson Imaging* 85:186–192
  25. van der Velde N, Hassing HC, Bakker BJ et al (2021) Improvement of late gadolinium enhancement image quality using a deep learning-based reconstruction algorithm and its influence on myocardial scar quantification. *Eur Radiol* 31:3846–3855
  26. Oh CH, Yoon SH (2017) Whole spine disc degeneration survey according to the ages and sex using Pfirrmann disc degeneration grades. *Korean J Spine* 14:148–154
  27. Ishimoto Y, Yoshimura N, Muraki S et al (2013) Associations between radiographic lumbar spinal stenosis and clinical symptoms in the general population: the Wakayama Spine Study. *Osteoarthritis Cartilage* 21:783–788
  28. Stadnik TW, Lee RR, Coen HL, Neiryneck EC, Buisseret TS, Osteaux MJ (1998) Annular tears and disk herniation: prevalence and contrast enhancement on MR images in the absence of low back pain or sciatica. *Radiology* 206:49–55
  29. Weishaupt D, Zanetti M, Hodler J, Boos N (1998) MR imaging of the lumbar spine: prevalence of intervertebral disk extrusion and sequestration, nerve root compression, end plate abnormalities, and osteoarthritis of the facet joints in asymptomatic volunteers. *Radiology* 209:661–666
  30. Schönström N, Hansson T (1988) Pressure changes following constriction of the cauda equina. An experimental study in situ. *Spine (Phila Pa 1976)* 13:385–388
  31. Argentieri EC, Koff MF, Breighner RE, Endo Y, Shah PH, Sneag DB (2018) Diagnostic accuracy of zero-echo time MRI for the evaluation of cervical neural foraminal stenosis. *Spine (Phila Pa 1976)* 43:928–933
  32. Ashby D (1991) *Practical statistics for medical research*. Douglas G. Altman, Chapman and Hall, London, 1991. No. of pages: 611. *Stat Med* 10:1635–1636
  33. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
  34. Koontz NA, Wiggins RH 3rd, Mills MK et al (2017) Less is more: efficacy of rapid 3D-T2 SPACE in ED patients with acute atypical low back pain. *Acad Radiol* 24:988–994
  35. Lee S, Lee JW, Yeom JS et al (2010) A practical MRI grading system for lumbar foraminal stenosis. *AJR Am J Roentgenol* 194:1095–1098

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.