

UCSF

UC San Francisco Previously Published Works

Title

Region of interest-specific loss functions improve T2 quantification with ultrafast T2 mapping MRI sequences in knee, hip and lumbar spine

Permalink

<https://escholarship.org/uc/item/6mm1f6q4>

Journal

Scientific Reports, 12(1)

ISSN

2045-2322

Authors

Tolpadi, Aniket A

Han, Misung

Calivà, Francesco

et al.

Publication Date

2022

DOI

10.1038/s41598-022-26266-z

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



OPEN

Region of interest-specific loss functions improve T_2 quantification with ultrafast T_2 mapping MRI sequences in knee, hip and lumbar spine

Aniket A. Tolpadi[✉], Misung Han, Francesco Calivà, Valentina Pedoia & Sharmila Majumdar

MRI T_2 mapping sequences quantitatively assess tissue health and depict early degenerative changes in musculoskeletal (MSK) tissues like cartilage and intervertebral discs (IVDs) but require long acquisition times. In MSK imaging, small features in cartilage and IVDs are crucial for diagnoses and must be preserved when reconstructing accelerated data. To these ends, we propose region of interest-specific postprocessing of accelerated acquisitions: a recurrent UNet deep learning architecture that provides T_2 maps in knee cartilage, hip cartilage, and lumbar spine IVDs from accelerated T_2 -prepared snapshot gradient-echo acquisitions, optimizing for cartilage and IVD performance with a multi-component loss function that most heavily penalizes errors in those regions. Quantification errors in knee and hip cartilage were under 10% and 9% from acceleration factors $R = 2$ through 10, respectively, with bias for both under 3 ms for most of $R = 2$ through 12. In IVDs, mean quantification errors were under 12% from $R = 2$ through 6. A Gray Level Co-Occurrence Matrix-based scheme showed knee and hip pipelines outperformed state-of-the-art models, retaining smooth textures for most R and sharper ones through moderate R . Our methodology yields robust T_2 maps while offering new approaches for optimizing and evaluating reconstruction algorithms to facilitate better preservation of small, clinically relevant features.

Magnetic Resonance Imaging (MRI) has emerged as a crucial part of diagnosing pathologies such as osteoarthritis, ligament damage, tumors, and others^{1–3}. Within MRI, several sequences can be deployed that exploit intrinsic tissue properties, providing images of varying weightings that effectively visualize tissues such as muscle, ligaments, bone marrow, and others⁴. In musculoskeletal (MSK) applications, clinical imaging protocols consist mostly of 2D fast spin echo (FSE) acquisitions with T_1 or T_2 weighting in various acquisition planes, which do well in depicting the structure and morphology of the underlying anatomy⁵. However, compositional MRI (cMRI) techniques to assess actual tissue parameters are gaining more attention as a complement of qualitative imaging.

cMRI techniques like T_2 relaxometry can provide maps of T_2 values (or another intrinsic MR parameter) across an imaging volume rather than a morphological image. For MSK applications, T_2 relaxometry offers sensitivity to water content, collagen content, and collagen fiber orientation in cartilage⁶, making it sensitive to biochemical changes that can precede morphological changes across several tissues and anatomies^{7,8}. Premorphological change sensitivity has been best characterized in the knee, where T_2 values are significantly higher across most cartilage compartments in healthy patients that later develop osteoarthritis (OA) compared to controls^{9,10}. Additionally, T_2 relaxometry offers quantitative MSK tissue health assessments, correlating with measures of hip cartilage and intervertebral disc (IVD) health^{11,14–16}, whereas in conventional clinical imaging, only semiquantitative tissue health assessments are obtainable with expert annotation^{12,13}. All of this makes cMRI a promising potential addition to clinical imaging protocols.

A major challenge facing clinical adoption of cMRI, however, is acquisition time: while mapping sequences like the magnetization-prepared angle-modulated partitioned k-space spoiled gradient echo snapshots (MAPSS) can provide robust MR parameter maps, their acquisition times can exceed 5–6 minutes, making their addition to a clinical scan protocol difficult¹⁷. Acquisitions can be accelerated by sampling fewer points in k-space, inducing

Department of Radiology and Biomedical Imaging, University of California, 1700, 4th Street, San Francisco, CA 94158, USA. ✉email: Aniket.Tolpadi@ucsf.edu

aliasing artifacts in resulting images that must be removed through subsequent postprocessing. Some proposed approaches to these ends are reconstruction strategies such as parallel imaging (PI), compressed sensing (CS), model-based reconstructions, deep learning (DL), low-rank and sparse modeling methods, and MR Fingerprinting (MRF). Most of these approaches design an algorithm or exploit the redundancy of k-space acquisition across multiple coils to predict the appearance of the fully-sampled reconstructed image.

PI was one of the earliest techniques to accelerate MRI acquisition and has seen clinical adoption. Here, the redundancy of a multiple coil acquisition is leveraged to mitigate aliasing artifacts^{18–20}, reducing clinical scan time up to acceleration factor $R = 3$ for MSK applications^{21,22}. CS²³ has also shown promise, where aliased images are iteratively reconstructed by minimizing an objective function, retaining fidelity to acquired k-space and imposing sparsity on the reconstructed image in another domain. CS has attained clinically acceptable MSK image quality through roughly $R = 4$ ^{21,22,24,25}, and up to $R = 8$ in research settings for knee cartilage $T_{1\rho}$ mapping²⁶. Similarly, PI and CS have also been applied sequentially (and simultaneously) for further acceleration²⁷.

For cMRI acceleration, model-based reconstructions have gained traction, integrating the physics of T_2/T_2^* decay and T_1 recovery into an objective function iteratively optimized to reconstruct maps, showing promise in brain and lumbar spine T_2 mapping^{28–30}. More generally, incorporation of the physics of MRI parameter recovery/decay has seen applications not just in model-based approaches, but in various aspects of other methodologies as well³¹. DL approaches have gained prominence in solving inverse problems such as reconstruction, allowing for cMRI reconstructions at higher R than other methods. Standalone DL approaches have seen promising results in knee MAPSS acceleration, T_1 mapping, and T_2 mapping sequences^{32–36}. In other methodologies, DL has been integrated with model-based approaches while introducing loss functions to maintain fidelity to acquired k-space, seeing promise up to $R = 8$ in knee and brain T_1 and T_2 mapping^{37–39}. DL has been applied to accelerate T_2 mapping in MR Fingerprinting, where DL can remove aliasing artifacts from undersampled acquisitions and/or replacing time-consuming dictionary lookup steps to predict MR parameter maps, and exploiting spatial correlations within maps to improve reconstructions^{40,41}. Lastly, aside from DL, low-rank and sparse modeling methods have emerged as a means of accelerating acquisitions, where several MRI images acquired at different echo times are decomposed into temporal basis functions and spatial coefficients to model an MRI parameter, showing promise through $R = 8$ ⁴².

These works represent great progress, although avenues for improvement remain. Above all, these methods have optimized reconstructed images for full-volume performance; however, in MSK applications, clinical assessment relies on the inspection of precise anatomic features in specific anatomic regions, and consequently, the reconstruction quality cannot be compromised within these regions. Put differently, given clinical context, strong image quality may be most important in specific regions of an image, leaving room for algorithm optimization. Furthermore, most recent published approaches leverage k-space data in formal reconstruction approaches, but for niche applications such as region of interest (ROI)-focused optimization, such approaches may be outperformed by DL-based post-processing algorithms that denoise and fit undersampled T_2 -weighted images without using raw k-space. Moreover, performance of standard reconstruction algorithms is typically evaluated using metrics such as structural similarity index (SSIM), normalized root mean square error (NRMSE), and peak signal-to-noise ratio (PSNR), but recent works show these metrics may not provide the best correspondence with radiologist annotations^{43,44}, leading other groups to propose alternate metrics to fill this niche⁴⁵.

To these ends, this study proposes a recurrent UNet pipeline to postprocess undersampled coil-combined T_2 -weighted echo images, fitting and predicting T_2 maps from accelerated MAPSS acquisitions in the knee, hip and lumbar spine^{46,47}. These algorithms are trained with multi-component, ROI-specific losses that optimize predicted maps for T_2 value and textural retention in cartilage and IVDs. In doing so, our approach allows for ROI-specific optimization, facilitating retention of small, crucial clinical features in tissues of interest while building on past applications of weighted loss functions for image processing tasks⁴⁸.

To summarize, the contributions of this work are as follows:

- By using a 4-component loss function in network training, we introduce the concept of “ROI-specific optimization” of cMRI accelerated acquisition pipelines.
- We conduct a thorough ablation study of these 4 loss function components, proving the value of all in retaining textures in predicted maps while retaining high fidelity to ground truth T_2 values.
- Acknowledging that standard evaluation metrics such as SSIM and NRMSE provide suboptimal sensitivity to clinically relevant metrics, we conduct a thorough Gray Level Co-Occurrence Matrix (GLCM)-metric-based analysis of smooth and sharp textural retention in predicted maps, with an eye towards better evaluation of retention of small features crucial to clinical diagnoses^{49,50}.
- We build on limited literature in hip and lumbar spine cMRI accelerated acquisition schemes by developing and evaluating our pipeline not only in knee cartilage, as several other works have done, but also for hip cartilage and lumbar spine IVD in ultrafast acquisitions.

Methods

MAPSS acquisitions. Retrospective datasets including MAPSS in the knee ($n = 244$ patients, 446 scans), hip ($n = 67$ patients, 89 scans), and lumbar spine ($n = 21$ patients, 24 scans) acquired from clinical 3 T MRI scanners was used. Patients were scanned in accordance with all pertinent guidelines, including approval from the University of California, San Francisco Institutional Review Board (Human Research Protection Program), and informed consent was obtained from all study participants. MAPSS simultaneously acquired multiple $T_{1\rho}$ and T_2 weighted images, using $T_{1\rho}$ or T_2 preparation followed by 3D RF-spoiled gradient-echo Cartesian acquisition in a segmented radial centric view ordering during a transient state. A fat-selective inversion pulse was applied before either $T_{1\rho}$ ^{51,52} or T_2 preparation⁵³. Each acquisition included $T_{1\rho}$ -prepared images at four spin-lock

		Knee	Hip	Lumbar spine
Acquisition parameters	Scanner (s)	GE Discovery MR750w (GE Healthcare, Waukesha, WI), GE Discovery MR750 (GE Healthcare, Waukesha, WI)	GE Discovery MR750w (GE Healthcare, Waukesha, WI), GE Discovery MR750 (GE Healthcare, Waukesha, WI)	GE Discovery MR750w (GE Healthcare, Waukesha, WI), GE Signa PET/MR (GE Healthcare, Waukesha, WI)
	Coil(s)	8-channel T/R knee array (Invivo, Gainesville, FL)	32-channel cardiac array (Invivo, Gainesville, FL)	Geometry embracing method (GEM) posterior array (GE Healthcare, Aurora, OH)
	FOV	14 × 14 cm ²	14 × 14 cm ²	20 × 20 cm ²
	acquisition matrix	256 × 128	256 × 128	256 × 128
	Slice thickness	4.0 mm	4.0 mm	8.0 mm
	Slices	22	28	12
	TEs	0 ms, 12.9 ms, 25.7 ms, 51.4 ms	0 ms, 10.4 ms, 20.8 ms, 41.7 ms	0 ms, 12.9 ms, 25.7 ms, 51.4 ms
	Readout BW	± 62.5 kHz	± 62.5 kHz	± 62.5 kHz
	Magnetization Recovery time	1.3 s	1.2 s	1.5 s
	ARC	2X	2X	None
	no phase wrap	None	2X k _y oversampling	None
Other	64-view acquisition/T ₂ preparation	64-view acquisition/T ₂ preparation	64-view acquisition/T ₂ preparation	
Demographics information	Sex (M/F)	140/104	35/32	10/11
	Age	29.7 ± 12.9	48.9 ± 13.2	45.3 ± 14.7
	Weight	74.3 ± 12.7 kg	69.8 ± 12.4 kg	69.6 ± 11.0 kg
Training information details	Learning rate	0.001	0.001	0.001
	Batch size	1	1	1
	Number of batches used for training	R=2: 40,390 R=3: 36,351 R=4: 36,351 R=6: 40,390 R=8: 60,585 R=10: 36,351 R=12: 40,390	R=2: 39,030 R=3: 28,622 R=4: 26,020 R=6: 39,030 R=8: 36,428 R=10: 26,020 R=12: 39,030	R=2: 1120 R=3: 1344 R=4: 1120 R=6: 1120 R=8: 2016 R=10: 1344 R=12: 3360
Training	Patients	144	39	13
	Scans	265	59	14
	Slices	5591	1533	112
Validation	Patients	50	15	4
	Scans	91	15	5
	Slices	1952	390	42
Test	Patients	50	13	4
	Scans	90	15	5
	Slices	1928	390	40
Total	Patients	244	67	21
	Scans	446	89	24
	Slices	9471	2313	194

Table 1. Knee, hip and lumbar spine datasets and splits. MAPSS acquisition parameters for all datasets, with corresponding training, validation and test splits. ARC refers to Auto-calibrating Reconstruction for Cartesian Imaging⁷⁰. For hip acquisitions, no phase wrap was applied; k_y was oversampled by a factor of 2X, with space outside the prescribed γ -FOV eliminated after reconstruction. In some cases, multiple acquisitions were taken per patient due to having multiple knees/hips scanned, or due to having follow-up scans for the same patient. Age and weight are reported mean ± 1 standard deviation (s.d.). Datasets were split into training, validation and test, ensuring all scans of a particular patient were only placed into one of the three datasets. Unless otherwise noted, all results are reported on the test set as described by this table; to ensure robustness of trained pipeline to data splits, additional versions were trained on 2 more splits detailed in Supplementary Tables S2 and S13, with results on those splits described in Supplementary Tables S14, S15, S16.

times (TSLs) for T_{1 ρ} quantification, and three additional T₂-prepared images for T₂ quantification (TSL = 0 ms images were shared for TE = 0 ms images). In this study, only T₂-prepared images at four different TEs and corresponding T₂ maps from the MAPSS sequence were used. k_y-k_z space was acquired within an elliptical coverage (area = 0.7 compared to rectangular k_y-k_z, not acquiring corner space). Knee images were acquired from patients having ACL injuries, with scans taken at baseline and 3 years post-reconstruction. Hip images were acquired from patients having hip OA. Lumbar spine images were acquired from healthy subjects or patients with low back pain. Table 1 shows acquisition parameters.

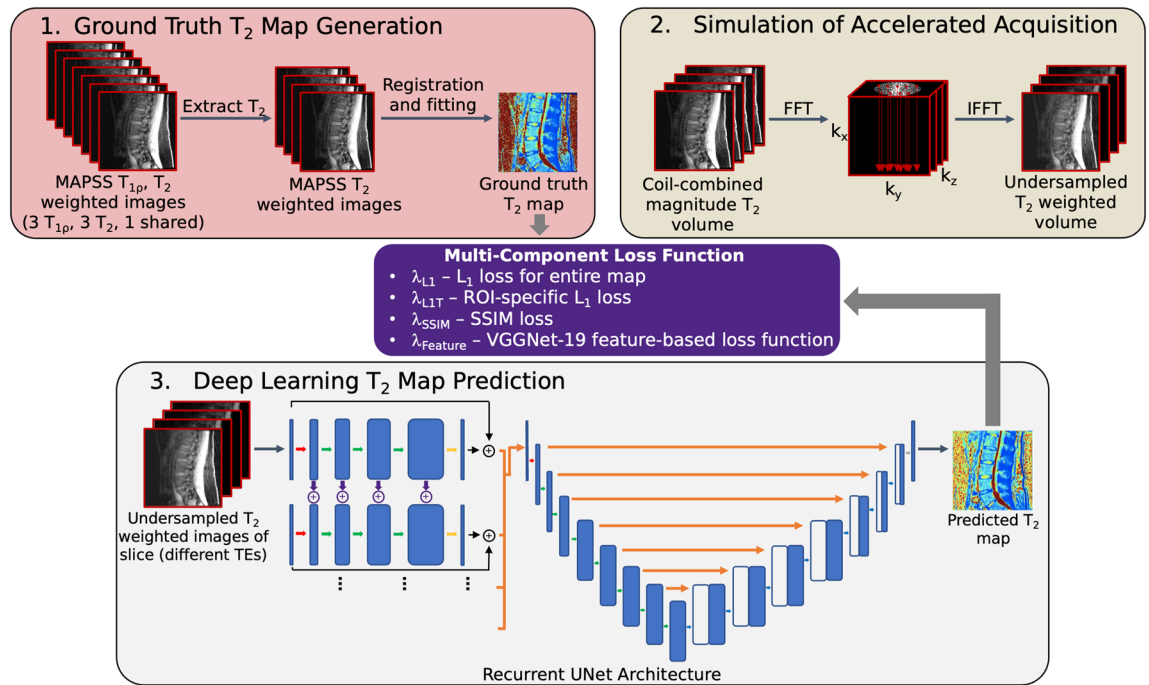


Figure 1. Proposed pipeline. Experiments in proposed study entail generating ground truth T_2 maps from MAPSS, simulating accelerated acquisition of T_2 -weighted MAPSS images, and training a network to predict T_2 maps from undersampled images. (1) MAPSS contains 7 images, 3 that are T_2 weighted, 3 $T_{1\rho}$ weighted, and 1 shared; the T_2 and shared image weightings are extracted, registered, and fitted slice-wise to yield ground truth T_2 maps. To simulate accelerated acquisition, each volume of coil-combined magnitude T_2 weighted images acquired at a given echo time are Fourier transformed, undersampled along the k_y - k_z plane with a center-weighted Poisson disc pattern, and inverse Fourier transformed to yield a simulated accelerated acquisition of a volume. Finally, undersampled T_2 weighted images acquired at all echo times for the same anatomic slice are concatenated and fed to the proposed recurrent UNet architecture, which predicts the T_2 map appearance for the slice. Training is done slice-wise with a multi-component loss function that includes a novel ROI-specific L_1 loss that optimizes predicted T_2 maps in cartilage and IVD ROIs, with other components that improve training stability and encourage retention of textures.

T_2 Fitting and spatial undersampling. Later T_2 weighted echo time images for each slice were registered to corresponding TE = 0 ms images using a 3D rigid registration algorithm with a normalized mutual information criterion⁵⁴. Levenberg–Marquardt fitting of registered T_2 weighted images yielded ground truth T_2 maps⁵⁵.

To simulate accelerated acquisition, coil-combined T_2 weighted magnitude images after reconstruction (ARC for knee and hip) were Fourier transformed and retrospectively undersampled using a center-weighted Poisson disc pattern, fully sampling a central 5% square in k_y - k_z ($R=2, 3, 4, 6, 8, 10, 12$). Acquisition times associated with ground truth and accelerated MAPSS acquisitions in each body part can be found in Supplementary Table S1. As MAPSS acquires phase-encode lines with elliptical coverage in k_y - k_z (relative area of 0.7 compared to rectangular coverage), phase encoding lines solely within the sampling ellipse were undersampled. Although working with synthesized k -space data generated from coil-combined magnitude images, retrospective undersampling was done and R reported with respect to elliptical coverage in k_y - k_z to accurately simulate an actual undersampling pattern and not overstate model performance⁵⁶. However, for hip acquisitions, reconstructed space outside the y -FOV had already been discarded; thus, simulating acquisitions with application of ‘no phase wrap’ was not possible and undersampling patterns would differ from those implemented on a scanner. T_2 weighted images from each echo time were undersampled with a unique pattern. For k_y - k_z lines not sampled at a given echo time, those k_y - k_z lines were initialized with the corresponding k_y - k_z from the image with the temporally closest echo time for which that k_y - k_z was sampled. Only k_y - k_z lines not sampled in images acquired at all echo times were zero-filled. k -Space was subsequently inverse Fourier transformed, yielding undersampled, aliased images.

DL pipeline training. DL architecture. An overview of the data processing and training schemes is shown in Fig. 1, while a detailed diagram depicting our proposed network architecture is in Supplementary Fig. S1 (“Full Model”; 39,808,710 trainable parameters). Magnitude images from data undersampled as specified were fed into a recurrent UNet network. The network contains an initial recurrent portion: aliased images from each T_2 echo time have a 5-layer processing stream of 2D 3×3 convolutions with stride 1, yielding layers of depth 64, 128, 256, 512, and 1. Residual connections connect input aliased images with processing stream outputs. 2D 3×3 convolutions with stride 1 and residual connections transfer information between temporally adjacent corresponding hidden echo time processing layers with weighting parameter $\lambda_w = 0.2$ ⁵⁷. This soft-weighted view-

sharing of neighboring T_2 weighted echo time images facilitated sharing of feature map information between temporally adjacent echo time images, which can augment sharing of k_y - k_z initializations to improve network image predictions. Outputs of all 4 echo time image processing streams were concatenated and fed to a UNet that predicted T_2 maps. 2D 3×3 convolutions with stride 2 were used for the encoder, and 2D 4×4 transpose convolutions with stride 2 for the decoder. Two additional architecture versions were also trained: one UNet with no recurrent portion (“No RNN”; 35,116,037 trainable parameters) and a second in which all layers apart from inputs to the recurrent portion and UNet had half the depth listed in Supplementary Fig. S1 (“Reduced Parameters”; 9,958,246 trainable parameters).

Loss function. Networks were trained with the multi-part loss function shown in Eq. (1):

$$L_{network} = \lambda_{L_1} L_{L_1} + \lambda_{L_{1,\phi}} L_{L_{1,\phi}} + \lambda_{SSIM} L_{SSIM} + \lambda_{Feature} L_{Feature} \quad (1)$$

in which L_{L_1} is a scaled global L_1 loss detailed in Eq. (2):

$$L_{L_1} = \left| S(T_2) - S(\hat{T}_2) \right| \quad (2)$$

where T_2 represents ground truth T_2 , \hat{T}_2 represents predicted T_2 , and $S(x)$ is a translated and scaled sigmoid operator that assigns more weight to higher T_2 values. Sharp contrasts and high T_2 values can easily be lost in accelerated acquisition schemes, so $S(x)$ proved useful through empirical testing in focusing networks to preserve these details. $S(x)$ is defined below in Eq. (3):

$$S(x) = y_l + (y_h - y_l) \left(1 + \exp(-10/(x_l - x_i)(x - (x_l + x_h)/2)) \right)^{-1} \quad (3)$$

where x_l , x_h were the low and high T_2 value limits where the sigmoid operator weighting will transition from y_l to y_h . Parameters selected for the knee were as follows: $x_l = 0$ ms, $x_h = 100$ ms, $y_l = 0.1$, $y_h = 1.0$. In the hip: $x_l = 0$ ms, $x_h = 60$ ms, $y_l = 0.5$, $y_h = 1.0$. In the lumbar spine: $x_l = 0$ ms, $x_h = 150$ ms, $y_l = 0.25$, $y_h = 1.0$. A schematic of the operator that results from parameters of all three anatomies can be found as Supplementary Fig. S2.

$L_{L_{1,\phi}}$ is the ROI-specific L_1 loss, and is described in Eq. (4):

$$L_{L_{1,\phi}} = \left| S(T_{2,\phi}) - S(\hat{T}_{2,\phi}) \right| \quad (4)$$

where $T_{2,\phi}$ were ground truth T_2 values in the tissue of interest ϕ (IVD or cartilage), scaled by $S(x)$ (Eq. (3)), and $\hat{T}_{2,\phi}$ is the same for predicted T_2 . Pixels corresponding to ϕ are obtained from segmentation masks, the generation of which is described in “Training and Segmentation Details”. For both L_{L_1} and $L_{L_{1,\phi}}$, L_1 norms were used instead of L_2 due to reduced sensitivity to outliers, leading to more stable trainings.

L_{SSIM} is an SSIM loss, described in Eq. (5):

$$L_{SSIM} = 1 - SSIM \quad (5)$$

where SSIM was the structural similarity index between predicted and target maps.

$L_{Feature}$ is a feature-based loss function designed to retain sharper textures, calculated as in Eq. (6):

$$L_{Feature} = \left| VGG_{T_2} - VGG_{\hat{T}_2} \right| \quad (6)$$

where VGG_{T_2} and $VGG_{\hat{T}_2}$ were the outputs of the 21st layer of a VGG-19⁵⁸ network pretrained on ImageNet when fed resized and normalized target and predicted T_2 maps, respectively. Maps were resized to $224 \times 224 \times 1$, concatenated with themselves along the channel axis to yield $224 \times 224 \times 3$ inputs, and normalized such that the channels had mean pixel values of 0.485, 0.456 and 0.406, with standard deviations of 0.229, 0.224, and 0.225, respectively.

λ_{L_1} , $\lambda_{L_{1,\phi}}$, λ_{SSIM} , $\lambda_{Feature}$ were loss component weightings. All were positive-valued and optimized through constrained random hyperparameter searches with the following ranges:

- Knee: $\lambda_{L_1} = 1$, $\lambda_{L_{1,\phi}} = 50 - 150$, $\lambda_{SSIM} = 0 - 2$, $\lambda_{Feature} = 0 - 0.5$
- Hip: $\lambda_{L_1} = 1$, $\lambda_{L_{1,\phi}} = 0 - 3$, $\lambda_{SSIM} = 0 - 2$, $\lambda_{Feature} = 0 - 1$.
- Spine: $\lambda_{L_1} = 1$, $\lambda_{L_{1,\phi}} = 1 - 10$, $\lambda_{SSIM} = 10 - 100$, $\lambda_{Feature} = 5 - 55$.

Training and segmentation details. Scans of all three anatomies were split into training, validation and test sets as shown in Table 1. In the knee, cartilage was segmented manually. In the hip, cartilage was segmented manually for 4 central slices per volume. Segmentation in both was performed by research assistants trained by radiologists with over 20 years of experience. Since the hip dataset had substantially fewer segmented than unsegmented slices, the hip training set was bootstrapped to equalize the number of slices with and without segmentations (1068 bootstrapped slices). Finally, in the lumbar spine, IVDs were segmented with an ensemble of coarse-to-fine context memory (CFCM) networks⁵⁹. To calculate performance metrics and implement ROI-specific training losses, these segmentation masks were leveraged to identify pixels in tissues of interest (cartilage or IVD).

Signal values were scaled per slice for the middle 95% of pixel values to fall between 0 and 500 for the knee and lumbar spine, and 0 and 100 for the hip; these ranges were optimized empirically. During training, imaging volumes were augmented with random translation (± 10 pixels across phase and frequency directions) and random rotation (± 5 degrees about slice direction). All models were trained with learning rate 0.001 and Adam

optimizer on an NVIDIA Titan Xp 12 GB GPU with batch size of 1 so the model would fit on a single GPU. Separate pipelines were trained for all 3 anatomies at R=2, 3, 4, 6, 8, 10, and 12. For each pipeline, and at each trained R, a constrained random hyperparameter search was done for 15 iterations at 10 epochs per iteration to optimize λ_{L_1} , $\lambda_{L_1, \phi}$, λ_{SSIM} , and $\lambda_{Feature}$ for visual fidelity of predicted maps to ground truth. Visual fidelity was assessed in the search using NRMSE (calculated as shown in Eq. (7)) and Pearson's r in the tissue of interest⁶⁰.

$$NRMSE = \|T_2 - \widehat{T}_2\|_{2, \phi} (\|T_2\|_{2, \phi})^{-1} \quad (7)$$

Final pipelines across all anatomies and R were trained using optimized parameter sets until validation loss did not decrease for 10 epochs. Key training details are summarized as part of Table 1.

Experiments

Loss function ablation study. An ablation study is key to understand contributions of loss components. Given optimized loss function weights, every combination of loss components was ablated and corresponding models were retrained until validation loss no longer decreased. “No RNN” and “Reduced Parameters” networks were also trained while maintaining loss function components at optimized values to assess the utility of simpler architectures. NRMSE and Pearson's correlation coefficient (r) were calculated in tissues of interest across the test set for original and ablated models to determine loss component contributions to performance. Pearson's r was deemed an appropriate statistical test for this and subsequent experiments, as it is useful in assessing the linear relationship between related pairs of interval data. While no formal NRMSE test was done, it nonetheless allows for quantitative assessment of T_2 quantification quality and easy comparison with results from other approaches. NRMSE is reported ± 1 standard deviation (s.d.); Pearson's r was deemed significant in accordance with corresponding *P* values, $\alpha = 0.001$, 0.01, and 0.05. NRMSEs within tissues of interest of a given scan were also multiplied by mean T_2 values within the tissue of interest of that patient, generating T_2 value equivalents of error rates.

To more specifically evaluate the utility of the ROI-specific loss component, two loss function configurations from the ablation study were further analyzed at all R: no ROI-specific loss component ($\lambda_{L_1, \phi} = 0$; λ_{L_1} , λ_{SSIM} , $\lambda_{Feature} \neq 0$) and no ROI-specific or feature-based components ($\lambda_{L_1, \phi}$, $\lambda_{Feature} = 0$; λ_{L_1} , $\lambda_{SSIM} \neq 0$). These models were intended to represent baselines in which all loss functions were preserved except the ROI-specific component, and a standard reconstruction loss function of pixel and SSIM-based loss components, respectively. Pearson's r—evaluated in tissues of interest and globally—was calculated to determine the degree and significance of correlation between predicted maps and ground truth, both globally and within tissues of interest, $\alpha = 0.001$, 0.01, and 0.05.

Evaluation of accelerated acquisition scheme performance. Three versions of our pipeline (full pipeline, “No RNN,” and “Reduced Parameters”) were compared to state-of-the-art CS, DL, and DL/model-based solutions. At each R, MANTIS (54,413,056 trainable parameters) and MANTIS-GAN (54,413,056 [Generator] and 2,763,648 [Discriminator] trainable parameters) pipelines were trained using published network architectures, loss functions and undersampling strategies^{42,43}. Loss function weightings for both were optimized through grid hyperparameter searches yielding the following: (MANTIS) $\lambda_{data} = 0.1$, $\lambda_{cnn} = 1$; (MANTIS-GAN) $\lambda_{data} = 0.1$, $\lambda_{cnn} = 1$, $\lambda_{GAN} = 0.01$. To apply CS reconstruction, original MAPSS T_2 -prepared images were Fourier transformed into coil-combined k-space, 1D-inverse Fourier transformed along the readout direction, and individual slices in $k_y - k_z$ reconstructed using an L_1 wavelet-based algorithm with regularization coefficient 0.001⁶¹. CS reconstructed images were registered to the TE=0 ms echo time image using a 3D rigid registration algorithm with a normalized mutual information criterion and fitted using Levenberg–Marquardt fitting to yield T_2 maps. Performance of these approaches and our proposed methods was evaluated through the following:

Comparison of global and ROI-specific performance. To test for completeness of training, performance of our proposed pipelines was compared against state-of-the-art models that did not use ROI-specific components in predicting T_2 maps. Pearson's r ($\alpha = 0.001$, 0.01, and 0.05) was used to compare model performances and assess strength of correlations to ground truth T_2 .

Standard reconstruction metrics. Performance was reported in tissues of interest with standard reconstruction metrics: NRMSE (mean ± 1 s.d.) and Pearson's r ($\alpha = 0.001$, 0.01, and 0.05). NRMSEs were also converted into T_2 value equivalents by tissue compartment as in the ablation study.

T_2 value retention. Fidelity of predicted maps to ground truth T_2 was also assessed. First, predicted and ground truth T_2 values were compared across tissues of interest within the test set (mean ± 1 s.d.), generating violin plots for all three anatomies with overlaid boxplots for T_2 value distribution comparison. T_2 agreement was also assessed through Bland–Altman analysis.

Texture retention. Gray Level Co-Occurrence Matrix (GLCM)⁶² metrics were used to assess texture retention within tissues of interest. GLCM contrast and dissimilarity are maximized by large local pixel value changes and thus by sharper textures. GLCM homogeneity is maximized by small local pixel value changes, while GLCM energy and angular second moment (ASM) are maximized by few total pixel values within an image; hence, all three are maximized by smoothness. For each anatomy and R, we calculated these texture metrics at 4 orientations ($\theta = 0^\circ, 45^\circ, 90^\circ$ and 135° ; $d = 1$ pixel) and averaged across all orientations. Finally, we calculated intraclass correlation coefficients (ICCs) for all metrics with respect to ground truth (two-way mixed effects, single rater⁶³)

and reported 95% ICC confidence intervals ($\alpha=0.001, 0.01, \text{ and } 0.05$). These tests were chosen as appropriate, as they assess both reliability and agreement of associated metrics, and in this use case, individual GLCM metric values themselves are considered the only rater, justifying the ICC test type selected.

Repeatability study. To assess the robustness of pipelines to different datasets, two additional splits of the knee, hip and spine datasets were made, ensuring no patient was part of multiple validation and/or test datasets and that all scans from a given patient were only in one of training, validation and test for each split (folds 2 and 3 in Supplementary Table S2, where fold 1 is the original split). Additional hyperparameters searches optimized loss function weights on the two new splits. Optimized loss weights and corresponding T_2 quantification and texture retention performance for each splits is presented at all tested R in the same manner as for the primary split.

Raw multicoil data assessment. An in-house pipeline was developed that leveraged GE Orchestra 1.10 and other postprocessing tools to reconstruct coil-combined images from raw k-space data. As a proof of concept, knee MAPSS scans were performed on 3 volunteers, hip scans for 2, and lumbar spine for 2, all using the acquisition parameters listed for the retrospective datasets used for algorithm training, with raw k-space data saved for all. Multicoil k-space data (after ARC for knee and hip) was undersampled with the same center-weighted Poisson disc pattern described earlier, with each coil seeing the same undersampling pattern and k_y - k_z lines being shared across different T_2 weighted echo time k-spaces as previously described. Coil-combined images resulting from undersampled multi-coil data at all tested R were fed through corresponding post-processing pipelines to predict T_2 map appearance. A radiologist with 2 years of experience segmented knee cartilage, hip cartilage, and intervertebral discs from these acquisitions, allowing for visualizations of predicted T_2 maps and NRMSE calculations in ROIs.

Results

Ablation study results. Voxel-wise performance metrics for ablation study models at $R=8$ are shown in Supplementary Table S3, with T_2 value NRMSE equivalents in Supplementary Table S4. Within the knee and hip, all loss components were necessary to obtain the optimal combination of high Pearson's r and low NRMSE in cartilage. For the lumbar spine, while all loss components proved vital in maximizing Pearson's r and minimizing NRMSE in IVDs, performance improved when the initial recurrent network was omitted. Though quantitative analysis is shown for all three pipeline versions in subsequent experiments, the full model is designated as best for knee and hip, and the no RNN for the spine.

ROI-specific and global assessments of best models and corresponding models trained without an ROI-specific loss ($\lambda_{1,\phi}=0$) and models trained with a generic loss ($\lambda_{1,\phi}=0, \lambda_{\text{Feat}}=0$) are shown in Supplementary Table S5. In the knee and hip, across nearly all R, ROI-specific loss addition leads to improved correlations between predicted and ground truth cartilage T_2 , with diminished performance globally. In the lumbar spine, which was trained with a substantially fewer batches than the knee and hip pipelines, these trends were inconsistent across tested R. Example predictions and ground truth for one slice of a patient in each pipeline are shown in Supplementary Fig. S3, showing that patterns of local T_2 value elevations in cartilage and IVDs are better preserved with an ROI-specific loss as opposed to pipelines trained without the loss component.

Visuals of network performance and comparison with state-of-the-art models. Predicted T_2 maps are displayed at select R for knee, hip and lumbar spine models in Fig. 2 for our three pipelines and three methods from the literature. In knee, hip, and lumbar spine, T_2 quantification performance is strongest with our proposed methods, maintaining low error rates, showing promising results compared with state-of-the-art methods through $R=10$. Optimal architecture performances are further explored in Figs. 3–5. As shown in Fig. 3a, predicted T_2 knee maps retained strong fidelity to ground truth within tibiofemoral joint cartilage. Patterns within predicted maps became slightly more diffuse as R increased to 10, as indicated by a slight rise in NRMSE for cartilage in the slice, but visually, T_2 values and map patterns are preserved. As seen in Fig. 4a, hip predicted maps preserve T_2 values well in femoral and acetabular cartilage through $R=10$, although T_2 patterns become more diffuse by $R=10$. Figure 5a shows T_2 map predictions in the lumbar spine. The L4-L5 IVD is shown in more detail, where T_2 quantification performance was acceptable at $R=3$, moderate at $R=6$, and worse at $R=10$, as indicated by rising IVD NRMSEs.

ROI and global performance comparisons of our selected pipelines against state-of-the-art approaches are in Supplementary Table S5. Across pipelines trained with relatively large dataset (knee and hip), DL and model-based approaches (MANTIS and MANTIS-GAN) outperformed our proposed pipeline globally, but within cartilage ROIs, our pipeline exhibited stronger Pearson's r at each tested R. These trends were not as strong in the lumbar spine pipelines, possibly owing to the randomness of training with a smaller dataset. Global and ROI-specific T_2 predictions are further visualized in Supplementary Fig. S4, showing predicted T_2 values exhibit substantially more visual fidelity to ground truth and lower NRMSE in state-of-the-art models compared to our pipeline, but a reversal of that trend in cartilage. In the lumbar spine, at some but not all R, those trends held, yielding similar conclusions to the Pearson's r analysis.

Evaluation of T_2 quantification performance and comparison with state-of-the-art models. *Voxel-wise T_2 evaluation fidelity.* Pearson's r and NRMSE across all anatomies and R for our approaches and state-of-the-art methods are in Table 2. T_2 value NRMSE equivalents are in Supplementary Table S6. For all anatomies and across nearly all R, T_2 quantification performance is strongest in our methods, particularly in the No RNN and full model pipelines, compared to state-of-the-art models.

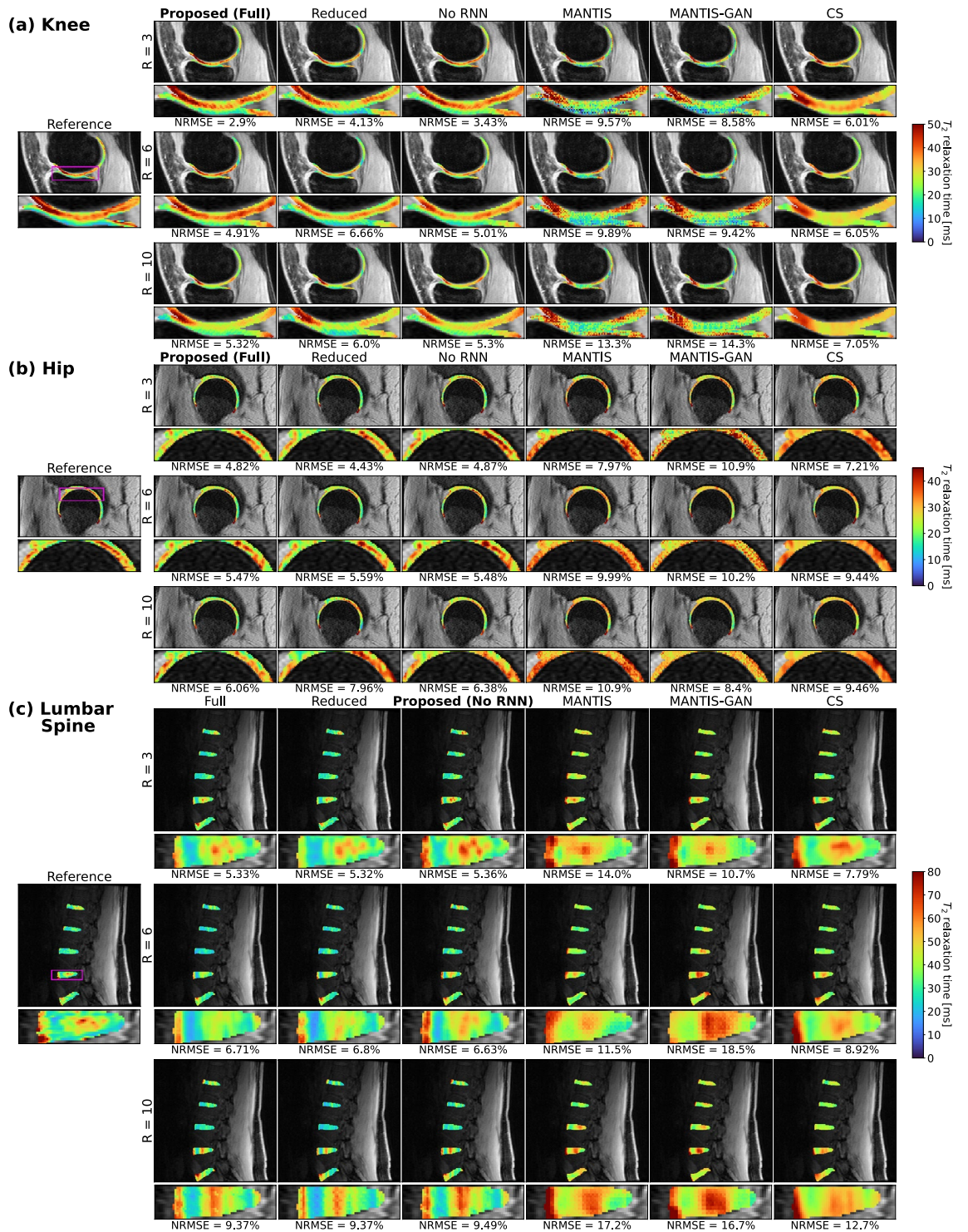


Figure 2. Comparison of predicted T_2 maps with ROI-specific methodologies to past approaches. **(a)** Predicted T_2 maps in knee cartilage for a representative patient within test set. T_2 quantification performance was best in pipelines trained with ROI-specific losses (Full Model, Reduced Parameters, and No RNN), where strong fidelity to T_2 values and patterns of local elevations within cartilage were maintained through $R=10$, while other tested approaches did a poorer job in predicting T_2 values in these maps. **(b)** Predicted hip cartilage T_2 maps showed similar trends, where performance of the full model was especially strong, showing low T_2 quantification error and better retaining local T_2 elevations through $R=10$ than other approaches. **(c)** Predicted T_2 maps in lumbar spine IVDs show higher T_2 quantification errors than in hip and knee cartilage, but ROI-specific loss pipelines best preserved map textures and values.

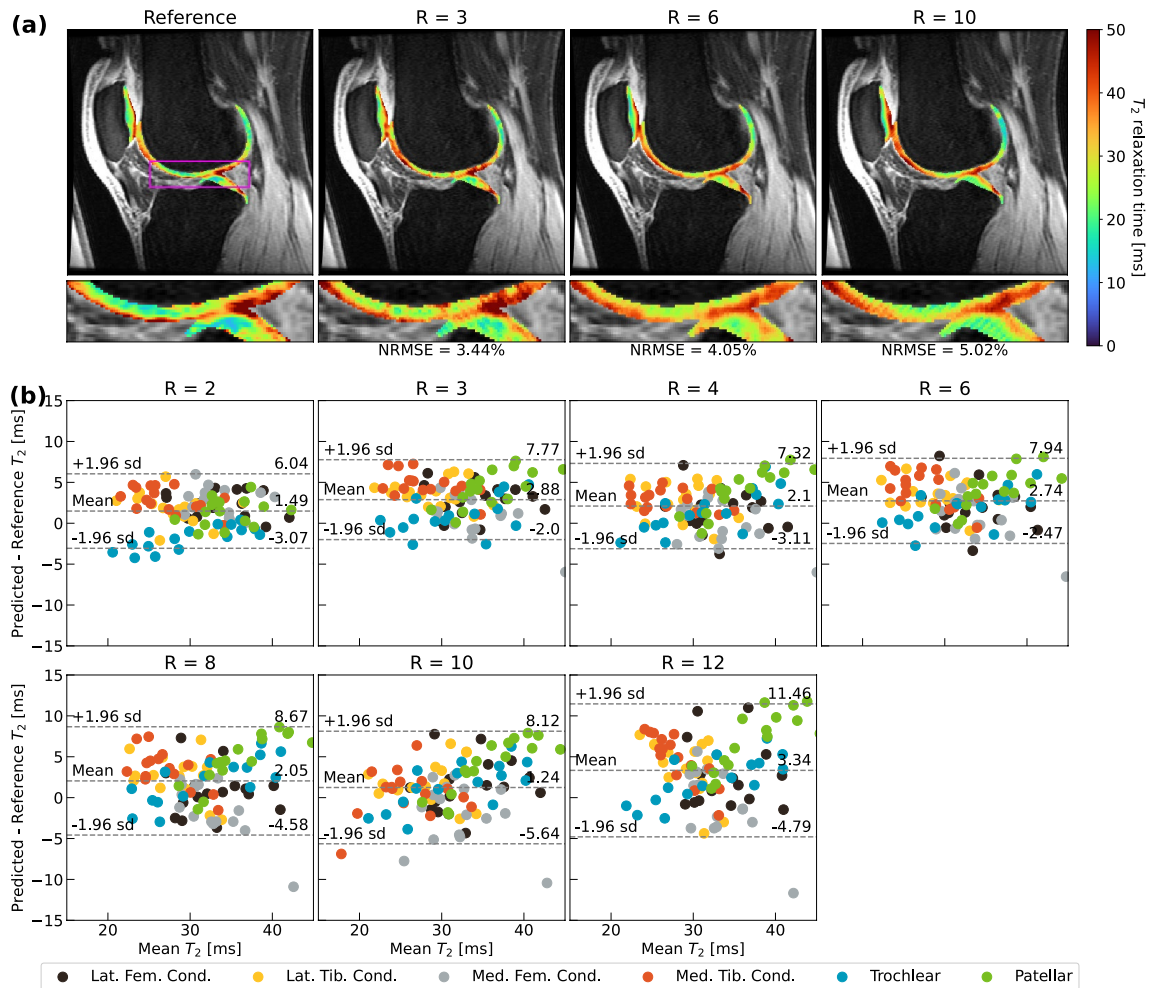


Figure 3. T_2 quantification performance of optimal ROI-specific pipeline in knee cartilage. **(a)** Visual pipeline performance within the knee for a representative patient, with corresponding NRMSEs for cartilage in the predicted T_2 map slice. Performance remains strong through $R = 10$, maintaining T_2 patterns in the medial tibiofemoral cartilage, indicating pipeline utility. Predicted maps generated by the network are masked using a cartilage segmentation mask and superimposed on the ground truth, fully sampled $TE = 0$ ms MAPSS echo time image. **(b)** Bland–Altman plots for all scans within test set for which multiclass cartilage compartment segmentations were available ($n = 16$, 6 cartilage compartments for each). Predicted T_2 values demonstrate minimal bias and tight limits of agreement across most tested R , with best performance coming from patellofemoral cartilage.

An exhaustive examination of knee T_2 quantification performance, stratified by cartilage compartments, is in Supplementary Tables S7 and S8. For the full model, across all cartilage compartments, T_2 estimation errors remained under 10% through $R = 10$ across all cartilage compartments while Pearson's r ranged from 0.748 at $R = 2$ to 0.491 at $R = 12$, indicating strong correlations⁶⁴ between predictions and ground truth at $R = 2$ and moderate correlations through $R = 12$. For some cartilage compartments and R , performance was stronger in the No RNN pipeline. Interestingly, quantification performance was strongest in patellofemoral joint cartilage, generally exhibiting lower NRMSE and stronger correlations. Our ROI-specific loss pipelines outperformed state-of-the-art models in each cartilage compartment.

Supplementary Tables S9 and S10 show hip T_2 quantification performance across cartilage compartments. As in the knee, quantification performance was strong, with error rates across all cartilage under 9% through $R = 12$ for the no RNN and full model pipelines. While the no RNN pipeline had stronger quantification errors, the full model had higher Pearson's r , which ranged from 0.794 at $R = 2$ to 0.517 at $R = 12$, showing strong correlations between predictions and ground truth through $R = 3$ and moderate correlations through $R = 12$. T_2 quantification performance was slightly stronger in femoral than acetabular cartilage. Our pipelines again outperformed state-of-the-art models in each cartilage compartment.

Supplementary Tables S11 and S12 show lumbar spine T_2 quantification performance, which was mixed. Pearson's r across all discs was very high, ranging from 0.884 at $R = 2$ to 0.643 at $R = 12$ for the no RNN model, indicating strong correlations through $R = 8$ and moderate correlations through $R = 12$ to ground truth. That said, IVD error rates were markedly higher across all R than in hip and knee cartilage, ranging from 4.86% to 18.8%. Though there was some volatility, error rates and Pearson's r generally showed poorest T_2 quantification in L1/

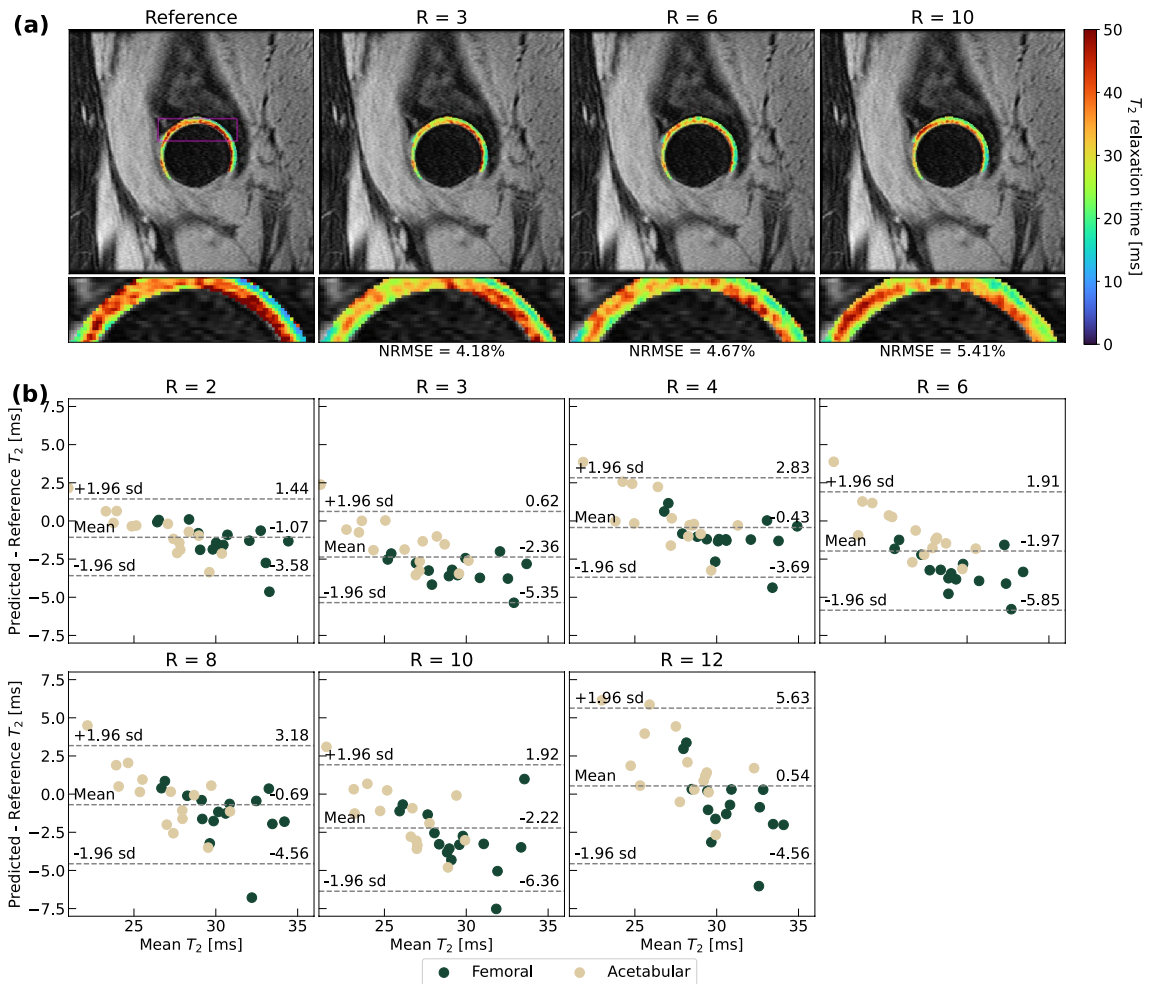


Figure 4. T_2 quantification performance of optimal ROI-specific pipeline in hip cartilage. (a) Visual pipeline performance within the hip for a representative patient, with corresponding NRMSEs for cartilage in the predicted T_2 map slice. Predicted maps are masked using a cartilage segmentation mask and superimposed on the ground truth, fully sampled TE = 0 ms MAPSS echo time image. For this patient, T_2 patterns maintain through R = 10, although local T_2 elevations are more diffusely predicted at higher R. (b) Bland–Altman plots for all scans within test set ($n = 15$, 2 cartilage compartments for each). Plots demonstrate very limited bias and even tighter limits of agreement from R = 2 through R = 12 than knee pipeline, showing hip pipeline effectiveness in reproducing T_2 values from accelerated MAPSS acquisitions.

L2 and L2/L3 discs. Through R = 8, ROI-specific loss pipelines outperformed state-of-the-art models at nearly all disc levels, with stronger Pearson's r in most IVD levels through R = 12.

T_2 Value retention on region of interest averages. Bland–Altman plots are provided for the knee, hip and lumbar spine in Figs. 3b, 4b, and 5b. In knee and hip, T_2 values are predicted with minimal bias with respect to ground truth. The ± 1.96 s.d. limits of agreement were less than approximately ± 6 ms with mean biases under ± 3 ms through R = 8 for knee cartilage (Fig. 3b). Among cartilage compartments, predictions in trochlear and patellar cartilage showed the least bias, while tibiofemoral cartilage T_2 was generally slightly overestimated. In the hip (Fig. 4b), ± 1.96 s.d. limits of agreement were less than approximately ± 5 ms with mean biases under ± 3 ms through R = 12, although T_2 quantification performance was similar across femoral and acetabular cartilage. In the lumbar spine (Fig. 5b), limits of agreement were considerably wider than the hip and knee pipelines, particularly above R = 4. While the line of equality was contained in these limits at all R, spine pipelines generally overestimated T_2 values. While at some particular R, a disc level saw poorer T_2 quantification than others (i.e. L2/L3 at R = 6), on balance, predicted maps yielded similar bias and error across all discs.

Supplementary Fig. S5 shows T_2 value distributions in violin and boxplots. Plots reveal minimal bias in hip cartilage predicted T_2 maps and slight but limited bias towards overestimating T_2 in knee cartilage. In the lumbar spine, more volatility was observed in predicted T_2 distributions, likely due to small test set size ($n = 5$), but at least through R = 6, these deviations had limited magnitude.

Texture retention. ICCs ± 1 s.d. for GLCM metrics are in Table 3 for our best performing pipelines: no RNN and full model. In knee cartilage, ICCs showed significant correlations between predicted and ground truth

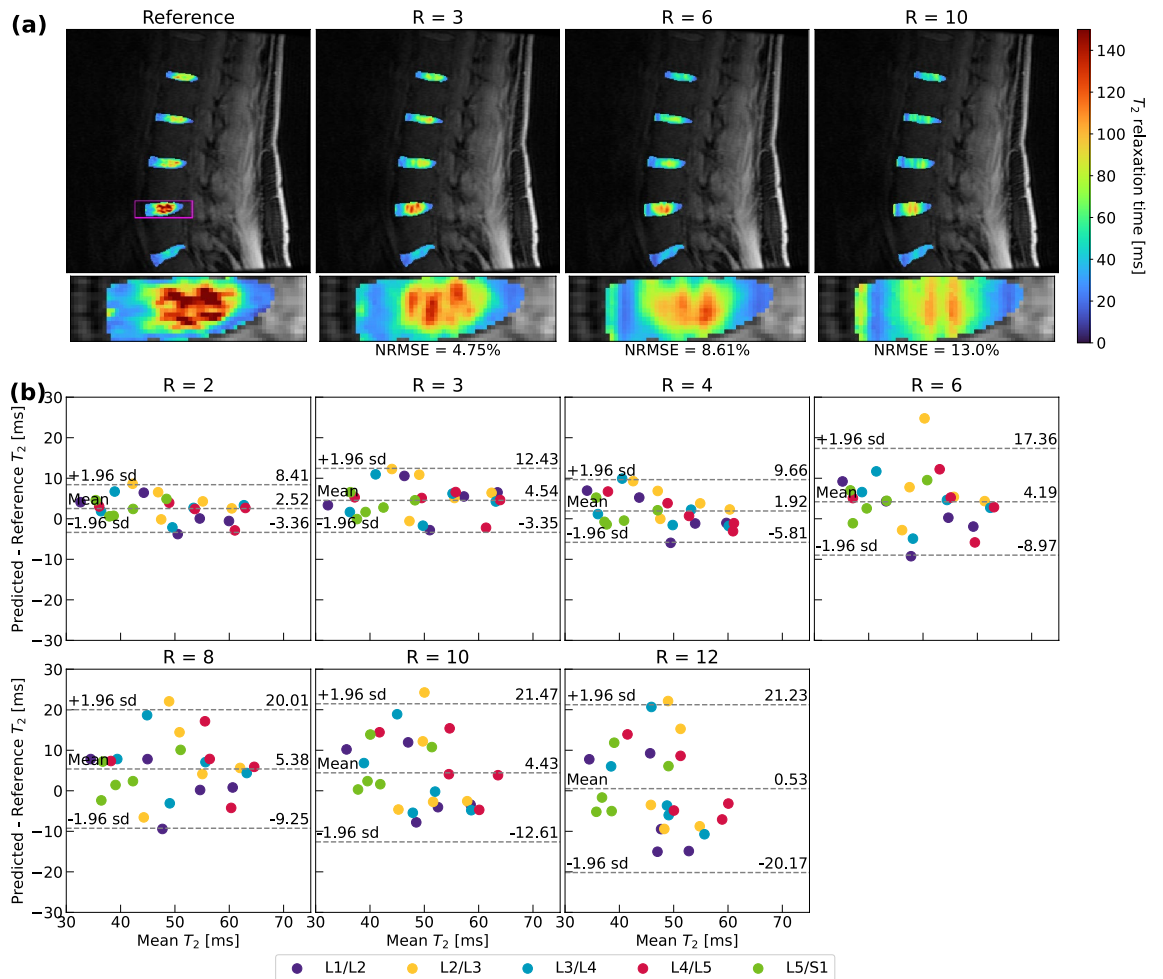


Figure 5. T₂ quantification performance of optimal ROI-specific pipeline in lumbar spine intervertebral discs. (a) Visual pipeline performance within the lumbar spine IVDs for a representative patient, with corresponding NRMSEs for IVDs in the predicted T₂ map slice. Predicted maps are masked using an IVD segmentation mask and superimposed on the ground truth, fully sampled TE = 0 ms MAPSS echo time image. Network performance is best through R = 6, after which local T₂ elevations are diffuse and underestimated. (b) Bland–Altman plots for all scans within test set (n = 5, 5 IVDs plotted for each if segmentation of disc available). T₂ value predictions reflect some bias and fairly wide limits of agreement, particularly above R = 4. These results indicate progress but the need for improvement. Smaller lumbar spine dataset and test set size are likely responsible for poorer model when compared to hip and knee performance, as well as the relatively smaller number of slices in k_z, which exacerbates undersampling effects.

GLCM metrics at all R for smooth textures and many R for sharp textures, indicating good to excellent reliability in preserving smooth textures (ASM and energy) at all R and moderate reliability in preserving sharper textures at low R (dissimilarity). In hip cartilage, ICCs showed significant correlations across all R in preserving smooth textures, and at low to moderate R for sharper textures. Reliability in smooth texture preservation ranged from good to excellent for all R and moderate for sharper textures at low to medium R. In both knee and hip cartilage, the full pipeline saw substantially higher GLCM ICCs for smooth and sharper texture across nearly all R. Within the lumbar spine, ICCs were significant across nearly all R for smoother textures. While ICCs were reasonable high for some R in contrast metrics, confidence intervals were wide, limiting findings of significant correlations. ICCs showed moderate to excellent reliability in preserving smoother textures, and poor to moderate reliability for sharper textures. For the spine, the No RNN model yielded optimal texture retention.

Repeatability study. Optimal loss weightings from hyperparameter searches on the two additional splits are in Supplementary Table S13. Results of trainings on additional splits in T₂ quantification error, Pearson's r, and texture metrics are in Supplementary Tables S14, S15, S16. In the knee and hip pipelines, experiments show comparable results across all folds for these metrics. In the lumbar spine, Pearson's r exhibited similar values across all folds, but in some cases, mean texture metric ICCs and NRMSEs exhibited substantial differences. However, confidence intervals were very wide for ICCs and NRMSEs in the lumbar spine, likely due to limited test set size (n = 5).

Tissue	R	Metric	Full model	Reduced parameters	No RNN	MANTIS	MANTIS-GAN	CS
Knee cartilage	2	NRMSE	5.52 ± 1.25	6.07 ± 3.21	4.76 ± 1.78	14.4 ± 2.85	13.5 ± 3.3	8.92 ± 3.2
		Pearson's r	0.748***	0.736***	0.807***	0.587***	0.611***	0.620***
	3	NRMSE	6.52 ± 2.17	7.18 ± 3.08	6.39 ± 2.59	16.5 ± 3.43	15.1 ± 2.89	9.92 ± 3.23
		Pearson's r	0.695***	0.668***	0.722***	0.467***	0.502***	0.559***
	4	NRMSE	7.54 ± 2.96	9.56 ± 5.47	7.56 ± 3.19	16.6 ± 3.73	15.7 ± 4.5	11.8 ± 3.73
		Pearson's r	0.651***	0.637***	0.677***	0.451***	0.467***	0.486***
	6	NRMSE	8.09 ± 2.65	10.7 ± 6.67	8.44 ± 3.49	15.2 ± 2.33	16.3 ± 3.91	12.4 ± 4.1
		Pearson's r	0.612***	0.610***	0.629***	0.397***	0.378*	0.445***
	8	NRMSE	8.94 ± 2.66	9.59 ± 3.83	10.1 ± 4.42	16.7 ± 2.73	17.3 ± 2.39	12.9 ± 3.93
		Pearson's r	0.585***	0.574***	0.609***	0.352***	0.364**	0.410***
	10	NRMSE	9.77 ± 3.44	10.2 ± 3.61	9.35 ± 3.5	17.6 ± 2.48	16.7 ± 3.44	13.4 ± 3.76
		Pearson's r	0.555***	0.514***	0.565***	0.327***	0.333***	0.386***
12	NRMSE	10.7 ± 2.32	9.93 ± 3.76	10.5 ± 3.37	18.2 ± 4.5	20.5 ± 5.58	13.4 ± 3.96	
	Pearson's r	0.491***	0.545***	0.511***	0.290***	0.287***	0.381***	
Hip cartilage	2	NRMSE	3.97 ± 1.03	4.1 ± 1.1	3.79 ± 0.807	4.58 ± 0.993	8.21 ± 1.42	14.8 ± 2.78
		Pearson's r	0.794***	0.782***	0.770***	0.716***	0.514***	0.310***
	3	NRMSE	6.53 ± 1.63	5.63 ± 1.68	5.25 ± 1.13	6.41 ± 1.31	10.0 ± 1.57	12.9 ± 3.15
		Pearson's r	0.705***	0.726***	0.703***	0.596***	0.372***	0.332***
	4	NRMSE	6.15 ± 1.01	6.17 ± 1.47	5.84 ± 0.891	7.33 ± 1.67	9.97 ± 1.74	11.8 ± 2.03
		Pearson's r	0.646***	0.665***	0.648***	0.510***	0.333***	0.339***
	6	NRMSE	8.1 ± 1.85	8.22 ± 2.06	7.48 ± 1.52	8.63 ± 2.32	9.68 ± 1.92	11.8 ± 2.14
		Pearson's r	0.587***	0.597***	0.570***	0.382***	0.321***	0.334***
	8	NRMSE	6.97 ± 1.93	6.33 ± 1.33	6.98 ± 1.45	10.2 ± 2.72	12.0 ± 2.64	10.5 ± 2.3
		Pearson's r	0.598***	0.588***	0.558***	0.334***	0.237***	0.347***
	10	NRMSE	8.99 ± 2.65	8.12 ± 1.28	8.7 ± 3.46	9.74 ± 2.24	10.5 ± 1.91	10.2 ± 2.4
		Pearson's r	0.558***	0.534***	0.522***	0.279***	0.268***	0.335***
12	NRMSE	7.75 ± 1.5	8.27 ± 2.19	7.34 ± 1.38	9.74 ± 2.23	11.5 ± 2.36	10.3 ± 2.52	
	Pearson's r	0.517***	0.566***	0.512***	0.280***	0.228***	0.349***	
Lumbar spine IVD	2	NRMSE	6.71 ± 1.7	6.86 ± 1.57	4.86 ± 1.16	8.78 ± 2.08	8.95 ± 1.91	10.1 ± 3.06
		Pearson's r	0.865***	0.866***	0.884***	0.784***	0.785***	0.802***
	3	NRMSE	9.92 ± 2.39	8.76 ± 2.16	7.13 ± 1.69	11.0 ± 1.17	11.3 ± 1.74	9.48 ± 1.4
		Pearson's r	0.836***	0.823***	0.832***	0.717***	0.712***	0.777***
	4	NRMSE	10.3 ± 3.02	9.73 ± 3.07	7.42 ± 1.1	12.1 ± 1.24	12.6 ± 1.35	11.3 ± 2.31
		Pearson's r	0.799***	0.813***	0.819***	0.680***	0.671***	0.723***
	6	NRMSE	12.1 ± 3.58	12.2 ± 4.11	10.3 ± 3.31	15.6 ± 2.65	12.1 ± 1.9	12.0 ± 2.76
		Pearson's r	0.776***	0.764***	0.771***	0.660***	0.658***	0.728***
	8	NRMSE	13.4 ± 3.89	13.0 ± 2.63	12.0 ± 3.07	13.2 ± 1.42	12.7 ± 1.7	12.8 ± 2.53
		Pearson's r	0.742***	0.723***	0.742***	0.631***	0.645***	0.695***
	10	NRMSE	15.3 ± 3.22	14.8 ± 2.78	14.8 ± 2.26	13.8 ± 1.57	13.2 ± 1.81	15.0 ± 3.77
		Pearson's r	0.695***	0.700***	0.672***	0.647***	0.636***	0.648***
12	NRMSE	18.1 ± 1.95	23.1 ± 2.71	18.8 ± 2.76	14.8 ± 3.03	14.1 ± 1.88	24.8 ± 11.2	
	Pearson's r	0.664***	0.320***	0.643***	0.651***	0.614***	0.586***	

Table 2. ROI-specific model performance in standard metrics from R = 2 through R = 12. Top performing pipeline for each metric, at each R, is shown in bold. Performances of pipelines trained with ROI-specific losses and other state-of-the-art methods in T₂ quantification error rates in knee cartilage, hip cartilage, and lumbar spine IVDs. NRMSEs are reported ± 1 s.d., and Pearson's r is reported with significances as follows: *P < 0.05, **P < 0.01, ***P < 0.001 (knee: n = 90; hip: n = 15; lumbar spine: n = 5). Across all anatomies, performances were strongest in ROI-specific loss pipelines (Full Model, Reduced Parameters, and No RNN): in the knee, the No RNN and Full Model pipelines particularly excelled across all tested R; in the hip, the No RNN pipeline was strong in maintaining minimal T₂ quantification errors, while the Full Model and Reduced Parameters models had strongest correlations between predicted maps and ground truth; in the lumbar spine, the No RNN pipeline especially had strong T₂ quantification performance. Performance in the knee and hip pipelines is strong and below clinically significant T₂ changes across nearly all tested R, while Pearson's r indicates strong T₂ value preservation in the lumbar spine through R = 6. T₂ quantification performance is thus promising in all three pipelines, but particularly for the knee and hip.

	R	GLCM texture metric					
		Contrast	Dissimilarity	Homogeneity	ASM	Energy	
Knee	Full model	2	0.307 ± 0.18**	0.638 ± 0.12***	0.734 ± 0.09***	0.966 ± 0.015***	0.954 ± 0.02***
		3	0.153 ± 0.2	0.521 ± 0.15***	0.735 ± 0.09***	0.962 ± 0.015***	0.95 ± 0.02***
		4	0.11 ± 0.2	0.387 ± 0.17***	0.61 ± 0.12***	0.973 ± 0.01***	0.95 ± 0.02***
		6	0.0667 ± 0.2	0.22 ± 0.19*	0.382 ± 0.17***	0.97 ± 0.015***	0.94 ± 0.025***
		8	0.061 ± 0.2	0.111 ± 0.2	0.0615 ± 0.2	0.952 ± 0.02***	0.9 ± 0.04***
		10	0.0594 ± 0.2	0.218 ± 0.19*	0.307 ± 0.18**	0.961 ± 0.015***	0.928 ± 0.03***
		12	0.0032 ± 0.2	-0.066 ± 0.2	-0.178 ± 0.19	0.927 ± 0.03***	0.861 ± 0.055***
	No RNN	2	0.455 ± 0.16***	0.599 ± 0.13***	0.32 ± 0.18***	0.898 ± 0.04***	0.904 ± 0.04***
		3	0.394 ± 0.17***	0.523 ± 0.15***	0.383 ± 0.17***	0.709 ± 0.11***	0.802 ± 0.07***
		4	0.262 ± 0.18**	0.305 ± 0.18**	0.244 ± 0.18**	0.646 ± 0.12***	0.754 ± 0.09***
		6	0.103 ± 0.2	0.0574 ± 0.2	0.061 ± 0.2	0.874 ± 0.045***	0.869 ± 0.05***
		8	0.0645 ± 0.2	0.0411 ± 0.2	0.0435 ± 0.2	0.922 ± 0.03***	0.911 ± 0.035***
		10	0.0474 ± 0.2	0.0382 ± 0.2	0.093 ± 0.2	0.92 ± 0.035***	0.913 ± 0.035***
		12	0.0568 ± 0.2	0.0315 ± 0.2	0.0885 ± 0.2	0.818 ± 0.065***	0.862 ± 0.055***
Hip	Full model	2	0.312 ± 0.34*	0.633 ± 0.23***	0.837 ± 0.12***	0.945 ± 0.04***	0.957 ± 0.035***
		3	0.369 ± 0.32*	0.671 ± 0.21***	0.816 ± 0.14***	0.976 ± 0.02***	0.98 ± 0.015***
		4	0.328 ± 0.33*	0.597 ± 0.25***	0.801 ± 0.15***	0.957 ± 0.035***	0.954 ± 0.04***
		6	0.235 ± 0.35	0.475 ± 0.3**	0.645 ± 0.23***	0.939 ± 0.05***	0.941 ± 0.045***
		8	0.199 ± 0.36	0.487 ± 0.28**	0.823 ± 0.13***	0.923 ± 0.06***	0.933 ± 0.055***
		10	0.127 ± 0.36	0.308 ± 0.34	0.48 ± 0.29**	0.862 ± 0.11***	0.855 ± 0.11***
		12	0.198 ± 0.36	0.38 ± 0.32*	0.523 ± 0.28**	0.927 ± 0.06***	0.914 ± 0.07***
	No RNN	2	0.285 ± 0.34	0.399 ± 0.32*	0.406 ± 0.31*	0.855 ± 0.11***	0.841 ± 0.12***
		3	0.15 ± 0.36	0.241 ± 0.35	0.292 ± 0.34	0.867 ± 0.1***	0.85 ± 0.12***
		4	0.113 ± 0.36	0.202 ± 0.36	0.282 ± 0.34	0.836 ± 0.12***	0.813 ± 0.14***
		6	0.0394 ± 0.36	0.0504 ± 0.36	0.0785 ± 0.36	0.793 ± 0.15***	0.767 ± 0.16***
		8	0.0229 ± 0.37	0.000593 ± 0.37	-0.0583 ± 0.37	0.682 ± 0.21***	0.653 ± 0.22***
		10	-0.00292 ± 0.36	-0.0328 ± 0.37	-0.196 ± 0.36	0.644 ± 0.23***	0.621 ± 0.24***
		12	-0.00208 ± 0.37	-0.0312 ± 0.36	-0.0646 ± 0.36	0.712 ± 0.2***	0.687 ± 0.2***
Lumbar spine	Full model	2	0.557 ± 0.7	0.695 ± 0.62	0.744 ± 0.57*	0.892 ± 0.35**	0.923 ± 0.27**
		3	0.499 ± 0.73	0.615 ± 0.67	0.644 ± 0.66	0.819 ± 0.48*	0.872 ± 0.39*
		4	0.236 ± 0.8	0.421 ± 0.76	0.497 ± 0.73	0.67 ± 0.64	0.775 ± 0.54*
		6	0.341 ± 0.78	0.428 ± 0.76	0.262 ± 0.8	0.566 ± 0.7	0.67 ± 0.64*
		8	0.0633 ± 0.81	0.152 ± 0.8	0.276 ± 0.79	0.685 ± 0.62	0.728 ± 0.58*
		10	-0.0393 ± 0.81	-0.0631 ± 0.81	-0.0699 ± 0.81	0.403 ± 0.76	0.479 ± 0.74*
		12	-0.0697 ± 0.81	-0.156 ± 0.8	-0.424 ± 0.76	0.16 ± 0.8	0.198 ± 0.8*
	No RNN	2	0.496 ± 0.73	0.731 ± 0.58*	0.883 ± 0.37**	0.967 ± 0.14***	0.975 ± 0.11***
		3	0.357 ± 0.78	0.615 ± 0.67	0.807 ± 0.5*	0.909 ± 0.31**	0.934 ± 0.24**
		4	0.336 ± 0.78	0.607 ± 0.68	0.771 ± 0.54*	0.874 ± 0.38*	0.91 ± 0.31**
		6	0.307 ± 0.78	0.53 ± 0.72	0.604 ± 0.68	0.903 ± 0.32**	0.916 ± 0.29**
		8	0.2 ± 0.8	0.4 ± 0.76	0.59 ± 0.68	0.847 ± 0.44*	0.871 ± 0.39*
		10	0.0696 ± 0.81	0.184 ± 0.8	0.386 ± 0.76	0.692 ± 0.62	0.726 ± 0.59*
		12	0.0157 ± 0.82	0.0858 ± 0.81	0.325 ± 0.78	0.561 ± 0.7	0.591 ± 0.68*

Table 3. Texture retention analysis in No RNN and Full Model pipelines. Intra-class correlation coefficients (ICCs) of Gray Level Co-Occurrence Matrix (GLCM)-based metrics. Contrast and dissimilarity are most sensitive to sharper image textures, while homogeneity, ASM, and energy are most sensitive to smoother image textures. Significance in correlations is noted as follows: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (knee: $n = 16$; hip: $n = 15$; lumbar spine: $n = 5$). In the knee and hip, Full Model pipelines outperformed No RNN versions in retention of smooth and sharp textures. In the lumbar spine, the No RNN pipeline outperformed the Full Model version, possibly because the smaller lumbar spine dataset size made training a larger network with a multi-component loss more difficult. In conjunction with standard reconstruction metrics, the Full Model pipeline was selected as the best knee and hip model, whereas the No RNN pipeline was selected as the best lumbar spine model. Top models in all anatomies preserved smoother textures at nearly all tested R, while dissimilarity texture metrics showed sharper textures were significantly correlated with ground truth and preserved in the knee and hip at low to medium R. In the lumbar spine, mean ICCs for sharper textures at many tested R also were high, but small dataset size likely led to wide standard deviations that prevented significant conclusions from being reached. All told, many textures are preserved in T_2 maps by all pipelines, particularly in the knee and hip.

Raw multicoil data assessment. Supplementary Fig. S6 shows T_2 maps predicted from our proposed pipelines on retrospectively undersampled raw k-space data. In the knee, T_2 quantification errors were low through $R=12$, with local T_2 elevations preserved and little dip in performance compared to corresponding retrospectively undersampled coil-combined knee data. In the hip, T_2 quantification errors were low, with local T_2 elevations reproduced at most R ; while performance at higher R matched expected performance from coil-combined experiments, lower R quantification errors were slightly higher. Performance was more volatile in the lumbar spine, where through $R=4$, T_2 quantification errors matched expected results and local T_2 patterns were generally preserved, but performance degraded substantially above $R=4$.

Discussion and conclusions. In this work, we present data-driven pipelines that leverage recurrent UNet architectures and multi-component losses to accelerate MAPSS T_2 mapping for anatomies where a subset of tissues is of particular clinical interest. By image processing and standard reconstruction metrics, through $R=10$, our knee pipelines retained fidelity to T_2 values with tight limits of agreement, preserving smooth textures with good to excellent reliability and sharper ones with moderate reliability for most tested R . While the no RNN pipeline delivered lower NRMSEs and higher Pearson's r across many cartilage compartments and R than full model, its texture retention was poorer, making the full model better suited to preserve small, key diagnostic features. In hip cartilage, predicted maps retained T_2 fidelity through $R=12$ with tight limits of agreement, preserved smooth textures with good to excellent agreement across tested R , and maintained sharper textures at low to moderate R . As with the knee, texture retention was strongest in the full pipeline despite lower no RNN NRMSEs. In IVDs, the no RNN pipeline delivered best standard reconstruction metric and texture retention performance. Despite maintaining smoother textures with moderate to excellent agreement across tested R and preserving sharper textures at lower R , the IVD pipeline revealed biases and fairly wide limits of agreement in T_2 preservation, particularly at $R=6$ and higher. When assessed on retrospectively undersampled multicoil raw k-space data, the knee and hip pipelines saw minimal degradation in performance as compared to results from images undersampled via synthetic k-space, whereas the lumbar spine pipeline exhibited similar performance through $R=4$. Furthermore, repeatability studies indicated that, particularly for the hip and knee, performance was stable with respect to datasets. All told, these metrics indicate promise for the knee and hip pipelines in MAPSS T_2 mapping acceleration, and progress but room for improvement in IVDs.

Assessments of ROI-specific loss component utility showed its potential for improving predictions in accelerated acquisition schemes. When trained with sufficiently large datasets, as our knee and hip pipelines were, its inclusion saw stronger fidelity to local T_2 patterns in cartilage ROIs and reduced T_2 quantification errors compared to analogous pipelines trained without the ROI-specific loss component. Compared to state-of-the-art DL pipelines, knee and hip pipelines saw improved Pearson's r in cartilage ROIs but poorer global Pearson's r , as expected from the focused training approach. Interestingly, CS approaches exhibit relatively strong NRMSEs while generating relatively smooth predicted T_2 maps; this is possibly because in training, DL-based approaches simultaneously removed aliasing artifacts and performed T_2 fitting, and could attempt to preserve finer details than a CS approach performing those steps sequentially. While our approaches outperformed state-of-the-art methods at many R and tissue compartments in the lumbar spine, global Pearson's r indicated this may have been partially due to some models being more completely trained than others. These results may have been different with a larger lumbar spine training set. Nonetheless, the value of ROI-specific loss functions in accelerated acquisition pipelines is clear: with sufficiently large datasets, they can optimize for ROIs and outperform state-of-the-art approaches at high R , as existing approaches are optimized for global and not ROI-specific performance.

We can contextualize performance by comparing quantification errors to clinically significant T_2 changes. In the knee, T_2 increases 13.4% in lateral femoral condyle (LFC) cartilage, 12.3% in medial femoral condyle (MFC) cartilage, and 8.1% in medial tibial condyle (MTC) cartilage among patients with mild OA compared to controls⁶⁵. Our top-performing knee pipeline saw errors below this benchmark through $R=12$ in the LFC and at $R=2$ in the MTC. In IVDs, T_2 decreases 36.3% in the nucleus pulposus and 24.2% in the annulus fibrosus from healthy to degenerative discs⁶⁶. Our top-performing pipeline saw quantification errors for each disc below the more stringent 24.2% through $R=12$. In the hip, T_2 values among healthy patients that progress to OA within 18 months are 7.3% higher in femoral and 5.2% higher in acetabular cartilage compared to controls⁶⁷. Our top-performing hip pipeline had errors below these benchmarks at all R in femoral cartilage and at $R=2$ in acetabular cartilage. Clinical metrics thus depict promise for pipelines in all three anatomies in maintaining sub-clinical-significance quantification errors.

Clinical and standard metrics show knee and hip pipeline performances to be particularly promising—the T_2 values, map texture preservation, and error rates relative to clinical benchmarks all mark meaningful progress towards reducing cMRI acquisition time for eventual clinical use. That said, while lumbar spine performance was strong by clinical metrics, it lagged the knee and hip by standard reconstruction metrics. One explanation is dataset size: the lumbar spine dataset had substantially fewer scans and imaging slices than the knee and hip. This has twofold impact: (1) the strength of a model trained from a smaller dataset is inherently limited, and (2) having only 5 test set scans limits statistical power and induces wide standard deviations of metrics, preventing significant conclusions from being reached. The effects of this small dataset size particularly surface in repeatability studies. Furthermore, lumbar spine acquisitions were more susceptible to breathing artifacts and had fewer slices than the hip and knee; undersampling therefore left fewer lumbar spine k_y - k_z lines sampled compared to the hip and knee, inducing worse initializations and possibly poorer performance. Nonetheless, to our knowledge, this is the first DL application to accelerate lumbar spine cMRI, marking progress that must be furthered with additional data procurement and algorithm development for clinical utility.

The GLCM-based textural retention evaluation demonstrated a framework through which reconstruction performance can be better evaluated than through standard metrics like SSIM, NRMSE, and PSNR. ICCs of

GLCM metrics between predicted and ground truth T_2 maps allow for intuitive, scaled measurements that can reflect how well a particular texture was preserved: for example, visual inspection of predicted T_2 maps in knee and hip cartilage in Figs. 3, 4 indicate that sharp textures are preserved better by the hip pipeline. This qualitative observation is confirmed by the GLCM Dissimilarity ICCs observed for the full model in the hip and knee pipelines in Table 3 at several tested R. This work could be furthered by extending this analysis to additional GLCM metrics for an even more thorough assessment of textural feature retention. Additional future improvements could also include pre-processing cartilage and IVD tissues prior to GLCM metric calculation to improve stability of these metrics, as other groups have started to do⁶⁸.

Moreover, by showing results at 7 acceleration factors instead of the 2–3 typical in the literature, we found performance did not always degrade steadily as R increased. Networks therefore may be sensitive not just to general undersampling patterns, but also the specific nature of the pattern. Thus, when future DL reconstruction pipelines are trained, a library of undersampling patterns may be advisable to encourage robustness to sampling patterns⁶⁹.

This study has limitations. First, we used retrospectively undersampled coil-combined magnitude echo time images that, in the knee and hip, had undergone ARC processing in their reconstruction, with 4 edge slices discarded for all data. Due to coil combination and post-processing, the k-space being undersampled would not match the acquisition's multi-coil k-space. Additionally, while we undersampled the MAPSS acquisition ellipse for each anatomy, the hip acquisitions had 'no phase wrap' applied, meaning that tested undersampling patterns would differ from those implemented on the scanner. While our raw k-space experiments show performance degradation was limited compared to coil-combined magnitude image experiments, models would be stronger if trained with a similarly sized multicoil k-space dataset. Second, this network is specific to our sampling patterns and acquisition parameters, and new pipelines would need to be trained should parameters like MAPSS T_2 echo times be substantially changed. Finally, the lumbar spine dataset size is rather small, limiting the power of conclusions.

To conclude, this study shows a novel means of training DL pipelines to accelerate cMRI in anatomies where specific tissues are of heightened clinical importance. In knee and hip, pipelines were effective at high R in maintaining textures, keeping fidelity to T_2 values, and minimizing T_2 quantification errors, whereas in the lumbar spine, the pipeline performed reasonably by those same criteria, but poorer in T_2 value fidelity and quantification errors. This reflects progress towards clinically useful pipelines that specialize in MSK T_2 mapping. The GLCM-based textural retention analysis elucidates an alternate to standard reconstruction metrics, allowing for intuitive measures of the types of features best preserved by an accelerated acquisition schemes, potentially allowing for better quantitative assessment of model performance. Future directions include multicoil k-space training, simultaneous MAPSS $T_{1\rho}$ and T_2 acceleration, and temporal undersampling of T_2 weighted echo time images.

Data availability

The datasets analyzed during the current study have been collected as part of multi-year studies or volunteers scans at the UCSF and their public release is not currently possible due to data privacy concerns. Codes to reproduce the results of this work are available upon reasonable request from the corresponding author (A. Tolpadi).

Received: 22 July 2022; Accepted: 13 December 2022

Published online: 23 December 2022

References

- Braun, H. J. & Gold, G. E. Diagnosis of osteoarthritis: Imaging. *Bone* **51**, 278–288 (2012).
- Wang, X., Oo, W. M. & Linklater, J. M. What is the role of imaging in the clinical diagnosis of osteoarthritis and disease management?. *Rheumatology* **57**, iv51–iv60 (2018).
- Abd-Allah, M. K., Awad, A. I., Khalaf, A. A. M. & Hamed, H. F. A. A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned. *Magn. Reson. Imaging* **61**, 300–318 (2019).
- Widmann, G., Henninger, B., Kremser, C. & Jaschke, W. MRI sequences in head & neck radiology—state of the art. *Rofo* **189**, 413–22 (2017).
- Chaudhari, A. S. et al. Rapid knee MRI acquisition and analysis techniques for imaging osteoarthritis. *J. Magn. Reson. Imaging* **52**, 1321–1339. <https://doi.org/10.1002/jmri.26991> (2019).
- Crema, M. D. et al. Articular cartilage in the knee: Current MR imaging techniques and applications in clinical practice and research. *Radiographics* **31**, 37–61. <https://doi.org/10.1148/rg.311105084> (2011).
- Soellner, S. T., Goldmann, A., Muelheims, D., Welsch, G. H. & Pachowsky, M. L. Intraoperative validation of quantitative T_2 mapping in patients with articular cartilage lesions of the knee. *Osteoarthr. Cartil.* **25**, 1841–1849. <https://doi.org/10.1016/j.joca.2017.07.021> (2017).
- Roemer, F. W., Kijowski, R. & Guermazi, A. Editorial: From theory to practice—the challenges of compositional MRI in osteoarthritis research. *Osteoarthr. Cartil.* **25**, 1923–1925. <https://doi.org/10.1016/j.joca.2017.08.007> (2017).
- Leibl, H. et al. Early T_2 changes predict onset of radiographic knee osteoarthritis—data from the osteoarthritis initiative. *Ann. Rheum. Dis.* **74**, 1353–1359 (2015).
- Atkinson, H. F. et al. MRI T_2 and $T_{1\rho}$ relaxation in patients at risk for knee osteoarthritis: A systematic review and meta-analysis. *BMC Musculoskelet. Disord.* **20**, 182. <https://doi.org/10.1186/s12891-019-2547-7> (2019).
- Friedrich, K. M. et al. T_2 measurements of cartilage in osteoarthritis patients with meniscal tears. *AJR AM. J Roentgenol.* **193**, W411–W415. <https://doi.org/10.2214/AJR.08.2256> (2009).
- Pfirmsmann, C. W., Metzendorf, A., Zanetti, M., Hodler, J. & Boos, N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine* **26**, 1873–1878. <https://doi.org/10.1097/00007632-200109010-00011> (2001).
- Kellgren, J. H. & Lawrence, J. S. Radiological assessment of osteo-arthrosis. *Ann. Rheum. Dis.* **16**, 494–502. <https://doi.org/10.1136/ard.16.4.494> (1957).
- Fernquest, S. et al. Compositional MRI of the hip: Reproducibility, effect of joint unloading, and comparison of T_2 relaxometry with delayed gadolinium-enhanced magnetic resonance imaging of cartilage. *Cartilage* **12**, 418–430. <https://doi.org/10.1177/1947603519841670> (2021).
- Bashir, A., Gray, M. L. & Burstein, D. Gd-DTPA₂- as a measure of cartilage degradation. *Magn. Reson. Med.* **36**, 665–673. <https://doi.org/10.1002/mrm.1910360504> (1996).

16. Wang, Y. X. J. *et al.* T₁rho and T₂ relaxation times for lumbar disc degeneration: An in vivo comparative study at 3.0-Tesla MRI. *Eur. Radiol.* **23**, 228–234. <https://doi.org/10.1007/s00330-012-2591-2> (2013).
17. Li, X. *et al.* Simultaneous acquisition of T₁ρ and T₂ quantification in knee cartilage—reproducibility and diurnal variation. *J. Magn. Reson. Imaging* **39**, 1287–1293. <https://doi.org/10.1002/jmri.24253> (2014).
18. Glockner, J. F., Hu, H. H., Stanley, D. W., Angelos, L. & King, K. Parallel MR imaging: A user's guide. *Radiographics* **25**, 1279–1297. <https://doi.org/10.1148/rg.255045202> (2005).
19. Pruessmann, K. P., Weiger, M., Scheidegger, M. B. & Boesiger, P. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Imaging* **42**, 952–962 (1999).
20. Griswold, M. A. *et al.* Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med.* **47**, 1202–1210. <https://doi.org/10.1002/mrm.10171> (2002).
21. Niitsu, M. & Ikeda, K. Routine MR examination of the knee using parallel imaging. *Clin. Radiol.* **58**, 801–807. [https://doi.org/10.1016/s0009-9260\(03\)00260-5](https://doi.org/10.1016/s0009-9260(03)00260-5) (2003).
22. Althawi, F. F., Blount, K. J., Morley, N. P., Raithe, E. & Omar, I. M. Comparing an accelerated 3D fast spin-echo sequence (CS-SPACE) for knee 3-T magnetic resonance imaging with traditional 3D fast spin-echo (SPACE) and routine 2D sequences. *Skeletal Radiol.* **46**, 7–15. <https://doi.org/10.1007/s00256-016-2490-8> (2017).
23. Lustig, M., Donoho, D. & Pauly, J. M. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195. <https://doi.org/10.1002/mrm.21391> (2007).
24. Iuga, A. I. *et al.* Accelerated MRI of the knee Quality and efficiency of compressed sensing. *Eur. J. Radiol.* <https://doi.org/10.1016/j.ejrad.2020.109273> (2020).
25. Ye, J. C. Compressed sensing MRI: A review from signal processing perspective. *BMC Biomed. Eng.* <https://doi.org/10.1186/s42490-019-0006-z> (2019).
26. Zibetti, M. V. W., Sharafi, A., Otazo, R. & Regatte, R. R. Accelerating 3D-T₁ρ mapping of cartilage using compressed sensing with different sparse and low rank models. *Magn. Reson. Med.* **80**, 1475–1491. <https://doi.org/10.1002/mrm.27138> (2018).
27. Pandit, P., Rivoire, J., King, K. & Li, X. Accelerated T₁ρ acquisition for knee cartilage quantification using compressed sensing and data-driven parallel imaging: A feasibility study. *Magn. Reson. Med.* **75**, 1256–1261. <https://doi.org/10.1002/mrm.25702> (2016).
28. Dipl-Ing, T. J. S., Uecker, M., Boretius, S. & Frahm, J. Model-based nonlinear inverse reconstruction for T₂ mapping using highly undersampled spin-echo MRI. *J. Magn. Reson. Imaging* **34**, 420–428. <https://doi.org/10.1002/jmri.22634> (2011).
29. Hilbert, T. *et al.* Accelerated T₂ mapping combining parallel MRI and model-based reconstruction: GRAPPATINI. *J. Magn. Reson. Imaging* **48**, 359–368. <https://doi.org/10.1002/jmri.25972> (2018).
30. Raudner, M. *et al.* Clinical implementation of accelerated T₂ mapping: Quantitative magnetic resonance imaging as a biomarker for annular tear and lumbar disc herniation. *Eur. Radiol.* **31**, 3590–3599. <https://doi.org/10.1007/s00330-020-07538-6> (2021).
31. Desai, A. D. *et al.* VORTEX: Physics-driven data augmentations for consistency training for robust accelerated MRI reconstruction. Preprint at <https://arxiv.org/abs/2111.02549>.
32. Li, H. *et al.* Ultra-fast Simultaneous T₁rho and T₂ mapping using deep learning. In *Proceedings of the 28th Annual Meeting of ISMRM, Virtual Conference and Exhibition*. 2669 (2020).
33. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. Preprint at <https://arxiv.org/abs/1505.04597>.
34. Jeelani, H. *et al.* A myocardial T₁-mapping framework with recurrent And U-Net convolutional neural networks. In *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. 1941–1944 (2020). doi: <https://doi.org/10.1109/ISBI45749.2020.9098459>.
35. Cai, C. *et al.* Single-shot T₂ mapping using overlapping-echo detachment planar imaging and a deep convolutional neural network. *Magn. Reson. Med.* **80**, 2202–2214. <https://doi.org/10.1002/mrm.27205> (2018).
36. Chaudhari, A. S., Fang, Z., Lee, J. H., Gold, G. & Hargreaves, B. Deep learning super-resolution enables rapid simultaneous morphological and quantitative magnetic resonance imaging. Preprint at <https://arxiv.org/abs/1808.04447>.
37. Liu, F., Feng, L. & Kijowski, R. MANTIS: Model-augmented neural network with Incoherent k-space sampling for efficient MR parameter mapping. *Magn. Reson. Med.* **82**, 174–188. <https://doi.org/10.1002/mrm.27707> (2019).
38. Liu, F., Kijowski, R., Feng, L. & El Fakhri, G. High-performance rapid MR parameter mapping using model-based deep adversarial learning. *Magn. Reson. Imaging* **74**, 152–160. <https://doi.org/10.1016/j.mri.2020.09.021> (2020).
39. Liu, F., Kijowski, R., El Fakhri, G. & Feng, L. Magnetic resonance parameter mapping using model-guided self-supervised deep learning. *Magn. Reson. Med.* **85**, 3211–3226. <https://doi.org/10.1002/mrm.28659> (2021).
40. Cohen, O., Zhu, B. & Rosen, M. S. MR fingerprinting deep reconstruction network (DRONE). *Magn. Reson. Med.* **80**, 885–894. <https://doi.org/10.1002/mrm.27198> (2018).
41. Fang, Z. *et al.* Deep learning for fast and spatially-constrained tissue quantification from highly-accelerated data in magnetic resonance fingerprinting. In *IEEE Trans. Med. Imaging* **38**, 2364–2374. <https://doi.org/10.1109/TMI.2019.2899328> (2019).
42. Meng, Z. *et al.* Accelerating T₂ mapping of the brain by integrating deep learning priors with low-rank and sparse modeling. *Magn. Reson. Med.* **85**, 1455–1467. <https://doi.org/10.1002/mrm.28526> (2020).
43. Mason, A. *et al.* Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. *IEEE Trans. Med. Imaging* **39**, 1064–1072. <https://doi.org/10.1109/TMI.2019.2930338> (2020).
44. Knoll F. *et al.* Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. Preprint at <https://arxiv.org/abs/2001.02518>
45. Adamson, P. M. *et al.* SSFD: Self-supervised feature distance as an MR image reconstruction quality metric. In *NeurIPS 2021: Workshop on Deep Learning and Inverse Problems*.
46. Tolpadi, A. A., Calivà, F., Han, M., Pedoia, V. & Majumdar, S. 8X accelerated intervertebral disc compositional evaluation with recurrent encoder-decoder deep learning network. In *Proceedings of the 29th Annual Meeting of ISMRM, Virtual Conference and Exhibition*. 2179 (2021).
47. Tolpadi, A. A., Calivà, F., Han, M., Pedoia, V. & Majumdar, S. A. Recurrent encoder-decoder network accelerates T₂ mapping in the knee, hip and lumbar spine. In *Proceedings of the 30th Annual Meeting of ISMRM, London, England, United Kingdom*. 1339 (2022).
48. Calivà, F., Iriondo, C., Martinez, A. M., Majumdar, S. & Pedoia, V. Distance map loss penalty term for semantic segmentation. Preprint at <https://arxiv.org/abs/1908.03679>.
49. Joseph, G. B. *et al.* Texture analysis of cartilage T₂ maps: Individuals with risk factors for OA have higher and more heterogenous knee cartilage MR T₂ compared to normal controls—data from the osteoarthritis initiative. *Arthritis Res. Ther.* **13**, R153. <https://doi.org/10.1186/ar3469> (2011).
50. Urish, K. L. *et al.* T₂ texture index of cartilage can predict early symptomatic OA progression: Data from the osteoarthritis initiative. *Osteoarthr. Cartil.* **21**, 1550–1557. <https://doi.org/10.1016/j.joca.2013.06.007> (2013).
51. Dixon, W. T., Oshinski, J. N., Trudeau, J. D., Arnold, B. C. & Pettigrew, R. I. Myocardial suppression in vivo by spin locking with composite pulses. *Magn. Reson. Med.* **36**, 90–94. <https://doi.org/10.1002/mrm.1910360116> (1996).
52. Chen, W., Takahashi, A. & Han, E. Quantitative T₁ρ imaging using phase cycling for B0 and B1 field inhomogeneity compensation. *Magn. Reson. Imaging* **29**, 608–619. <https://doi.org/10.1016/j.mri.2011.02.002> (2011).
53. Foltz, W. D., Stainsby, J. A. & Wright, G. A. T₂ accuracy on a whole-body imager. *Magn. Reson. Med.* **38**, 759–768. <https://doi.org/10.1002/mrm.1910380512> (2005).
54. Weingärtner, S. Manual for RView (2010). <https://biomedic.doc.ic.ac.uk/uploads/Internal/rviewManual>. Accessed July 2, 2021.

55. More, J. J. The Levenberg–Marquardt algorithm: Implementation and theory. In *Numerical Analysis Lecture Notes in Mathematics* (ed. Watson, G. A.) (Springer, 2006).
56. Shimron, E., Tamir, J. I., Wang, K. & Lustig, M. Implicit data crimes: Machine learning bias arising from misuse of public data. *PNAS* **119**, e2117203119. <https://doi.org/10.1073/pnas.2117203119> (2021).
57. Srivastava, R. K., Greff, K. & Schmidhuber, J. Highway networks. Preprint at <https://arxiv.org/abs/1505.00387>.
58. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. Preprint at <https://arxiv.org/abs/1409.1556>.
59. Iriondo, C., Padoia, V. & Majumdar, S. Lumbar intervertebral disc characterization through quantitative MRI analysis: An automatic voxel-based relaxometry approach. *Magn. Reson. Med.* **84**, 1376–1390. <https://doi.org/10.1002/mrm.28210> (2020).
60. Schober, P., Boer, C. & Schwarte, L. A. Correlation coefficients: Appropriate use and interpretation. *Anesth. Analg.* **126**, 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864> (2018).
61. Dwork, N. *et al.* Utilizing the wavelet transform's structure in compressed sensing. *Signal Image Video Process.* **15**, 1407–1414. <https://doi.org/10.1007/s11760-021-01872-y> (2021).
62. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern. Syst.* **3**, 610–621. <https://doi.org/10.1109/TSMC.1973.4309314> (1973).
63. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012> (2016).
64. Ratner, B. The correlation coefficient: Its values range between +1/–1, or do they?. *J. Target Meas. Anal. Market.* **17**, 139–142 (2009).
65. Dunn, T. C., Lu, Y., Jin, H., Ries, M. D. & Majumdar, S. T₂ Relaxation time of cartilage at MR imaging: Comparison with severity of knee osteoarthritis. *Radiology* **232**, 592–598. <https://doi.org/10.1148/radiol.2322030976> (2004).
66. Cui, Y. Z., Yang, X. H., Liu, P. F., Wang, B. & Chen, C. Preliminary study on diagnosis of lumbar disc degeneration with magnetic resonance T_{1ρ}, T₂ mapping and DWI quantitative detection technologies. *Eur. Rev. Med. Pharmacol. Sci.* **20**, 3344–3350 (2016).
67. Gallo, M. C. *et al.* T_{1ρ} and T₂ Relaxation times are associated with progression of hip osteoarthritis. *Osteoarthr. Cartilage* **24**, 1399–1407. <https://doi.org/10.1016/j.joca.2016.03.005> (2016).
68. Janacova, V., Szomolanyi, P., Vilimek D., Trattig, S. & Juras, V. GLCM texture analysis of knee cartilage T2 maps: Machine learning based selection of important features. In *Proceedings of the 30th Annual Meeting of ISMRM, London, England, United Kingdom*. 1498 (2022).
69. Liu, F., Samsonov, A., Chen, L., Kijowski, R. & Feng, L. SANTIS: Sampling-augmented neural network with incoherent structure for mr image reconstruction. *Magn. Reson. Med.* **82**, 1890–1904. <https://doi.org/10.1002/mrm.27827> (2019).
70. Brau, A. C., Beatty, P. J., Skare, S. & Bammer, R. Comparison of reconstruction accuracy and efficiency among autocalibrating data-driven parallel imaging methods. *Magn. Reson. Med.* **59**, 382–395. <https://doi.org/10.1002/mrm.21481> (2008).

Acknowledgements

We would like to acknowledge Prachi Pandit for help in curating the lumbar spine dataset. We would also like to knowledge Claudia Iriondo for additional help in lumbar spine and hip dataset curation and IVD segmentation. Additionally, we also acknowledge Alex Beltran for assistance in knee dataset curation. We also acknowledge Nicholas Dwork for his assistance in implementing the CS reconstruction algorithms reported in this paper. We acknowledge Johanna Luitjens for segmenting knee cartilage, hip cartilage, and IVDs from scans of patients for whom multicoil MAPSS acquisition k-space data was saved. This work was funded by the National Institutes of Health (UH3AR076724, R01AR078762).

Author contributions

A.T. designed and optimized the pipelines, performed the experiments, and wrote the manuscript. M.H. edited the manuscript and provided crucial feedback regarding correct implementations of undersampling patterns and general MR physics-based guidance. F.C. provided initial code to simulate undersampled acquisitions in early versions of the project. V.P. and S.M. guided the study design, originated the idea for accelerated T₂ map acquisition with MAPSS data, and provided valuable feedback throughout the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-26266-z>.

Correspondence and requests for materials should be addressed to A.A.T.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022