



OPEN

DATA DESCRIPTOR

# A Multi-view Open-access Dataset of Paired Knee MRI for Motion Artifact Removal

Yu Xi<sup>1,4</sup>, Fang Wang<sup>2,3,4</sup>, Feng Shi<sup>2</sup>, Qing Zhou<sup>2</sup>, Ran Li<sup>1</sup>, Yingchun Li<sup>1</sup> & Yuting Wang<sup>1</sup>

Magnetic resonance imaging (MRI) has become a standard examination method for the knee, facilitating the identification of a range of knee-related issues, including injuries, arthritis, and other conditions. The lengthy image acquisition time inherent to MRI results in the generation of motion artifacts, which in turn impairs the efficiency of MRI applications. To address this challenge, we present a multi-view, multi-sequence knee joint paired MRI dataset (image with motion artifact vs. Ground Truth obtained after rescanning), named Knee MRI for Artifact Removal (KMAR)-50K, which includes 1,190 patients, 1,444 pairs of MRI sequences, and 62,506 scan images. The dataset comprises images of anonymous paired NIFTI files that have undergone bias field correction, maximum minimum normalization, and paired image spatial registration in sequence. The objective of our data-sharing program is to facilitate the benchmark testing of methods of knee MRI motion artifact removal. Benchmarking three models revealed U-Net's superior transverse plane performance (PSNR = 28.468, SSIM = 0.927) with fastest inference (0.5 s/volume), highlighting its clinical value in accuracy-efficiency balance.

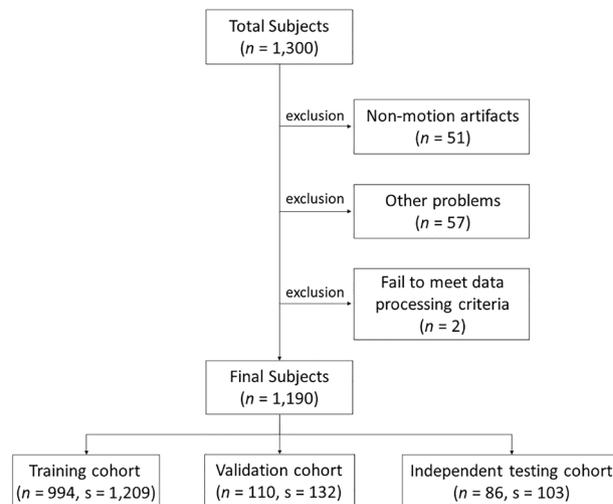
## Background & Summary

Magnetic Resonance Imaging (MRI) is a non-invasive imaging method, which has been widely applied in clinical practice<sup>1</sup>. Compared to X-ray and computerized tomography (CT), MRI stands out for its high spatial resolution, excellent soft tissue contrast, and lack of radiation<sup>2</sup>. It enables accurate examination of soft tissue structures such as ligaments and menisci of the knee, showcasing advantages such as a high detection rate, low misdiagnosis rate, and high specificity. Consequently, MRI has become the standard imaging modality for assessing the knee joint disorders and provides valuable support for the clinical diagnosis and treatment of knee joint diseases<sup>3,4</sup>.

Due to its prolonged image acquisition period, MRI is more prone to motion artifacts in comparison to other imaging modalities, such as CT and ultrasound<sup>5</sup>. Although rescanning can to some extent reduce the impact of motion artifacts, it also results in a greater investment of valuable machine time, which ultimately diminishes clinical efficiency. Moreover, in specific clinical scenarios, such as with infants, patients experiencing involuntary tremors, individuals with traumatic injuries, and those in pain, a repeated scan may not necessarily yield clear images. Consequently, images with motion artifacts have the potential to impede accurate diagnosis by physicians, which could result in misdiagnosis. This has the effect of undermining the applicability of MRI technology. Therefore, effectively overcoming motion artifacts caused by movement is one of the key areas of research in the field of magnetic resonance technology.

Despite the advent of several proposed techniques aimed at preventing, mitigating, and correcting motion artifacts, including physically restricting patient movement<sup>6,7</sup>, advanced sampling techniques<sup>8–10</sup>, faster imaging sequences<sup>11,12</sup>, and the implementation of some prospective<sup>13,14</sup> and retrospective correction methods<sup>15–17</sup>, motion correction remains an evolving field. It is not feasible to eliminate motion artifacts by applying a single method<sup>5</sup>.

<sup>1</sup>Department of Radiology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China. <sup>2</sup>Department of Research and Development, Shanghai United Imaging Intelligence, Shanghai, China. <sup>3</sup>School of Biomedical Engineering, ShanghaiTech University, Shanghai, China. <sup>4</sup>These authors contributed equally: Yu Xi, Fang Wang. e-mail: [anicespringspring@163.com](mailto:anicespringspring@163.com); [wangyuting\\_330@163.com](mailto:wangyuting_330@163.com)



**Fig. 1** Study flowchart of the enrolled patients. Non-motion artifacts include fold over artifacts, vascular pulsation artifacts and metal artifacts. Other problems include inconsistencies in scope before and after rescanning, changes in PD sequence parameters, uneven magnetic fields, etc. Failing to meet data processing criteria include missing frames, inconsistencies before and after rescanning. Roughly divided into training cohort and validation cohort by a ratio of approximately 9:1. Additionally, an independent testing cohort was collected for model evaluation.

At present, numerous retrospective artifact correction techniques are based on deep learning methodologies<sup>18–20</sup>, which frequently necessitate a substantial quantity of paired artifact and artifact-free image data to substantiate their efficacy. It is noteworthy that there are existing public knee MRI image datasets, such as fastMRI<sup>21</sup>, MRNet<sup>22</sup>, and kneeMRI<sup>23</sup>, whose primary focus is on utilizing machine learning models to aid in diagnosis instead of artifact removal. Although the fastMRI project incorporates certain aspects of image reconstruction, its principal objective is to expedite the MRI process. These publicly available databases have an obvious limitation: the absence of critical paired image data, namely knee MRI images with motion artifacts and without artifacts after rescanning from the same patient. In light of this, the importance of our dataset comes to the fore, as it provides paired data not provided in the aforementioned datasets and fills a gap in the field of image quality improvement of knee MRI.

In this study, we propose KMAR-50K dataset, which includes 1,444 pairs of knee joint scans from 1,190 patients, for a total of 62,506 slices. These images are from two different machines with field strengths of 3.0 T and 1.5 T, respectively, ensuring the diversity and breadth of the data. The dataset includes the mostly frequently and routinely used MRI sequences in medical diagnosis: Proton Density-weighted Turbo Spin Echo (PD-TSE), Proton Density-weighted Fast Spin Echo (PD-FSE), T1-weighted Turbo Spin Echo (T1-TSE), T1-weighted Fast Spin Echo (T1-FSE), T2-weighted Turbo Spin Echo (T2-TSE), and T2-weighted Fast Spin Echo (T2-FSE). Moreover, we registered the corresponding paired images from the coronal, sagittal, and transverse planes, respectively, which allows for more precise adjustment of image accuracy and benefits the generalization ability of deep learning models.

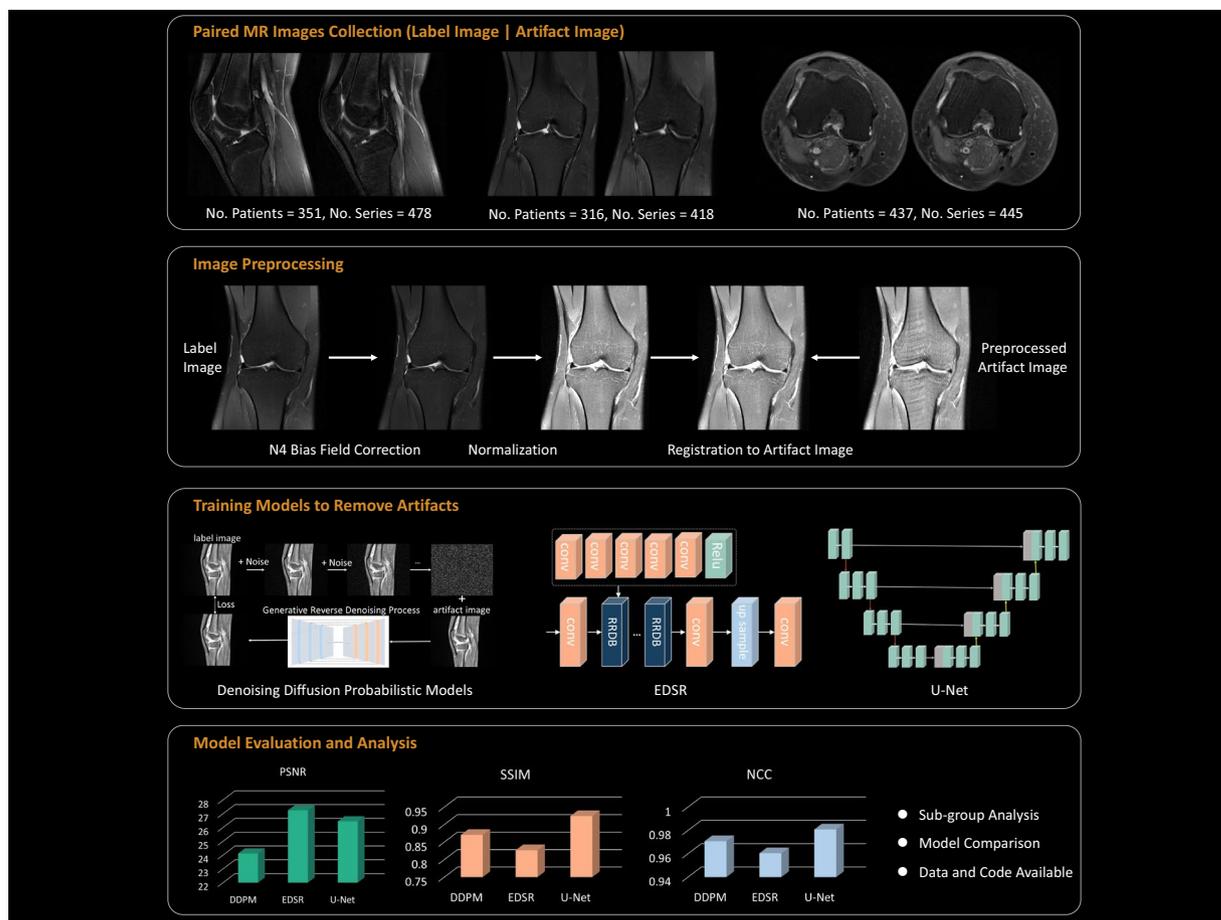
In summary, this dataset is highly comprehensive in terms of size, device source, and sequence type for paired data. It went thorough quality control for inclusion and professional scoring by experienced radiologists. It is represented in a ready-to-use format. In addition, its usability was validated by three commonly used models and their results of artifact removal can be compared using Peak Signal-to-Noise Ratio (PSNR), structural similarity index matrix (SSIM), and error metrics. Benchmark validation across validation and independent testing cohorts revealed U-Net's consistent superiority, achieving peak transverse-plane performance (PSNR = 28.683, 28.468, SSIM = 0.933, 0.927) with 0.5 s/volume inference speed – 18 times faster than Enhanced Deep Residual Networks for Single Image Super-Solution (EDSR) while maintaining <0.06 in 3 error metrics across all anatomical planes. The standardized paired data structure facilitates direct comparison of artifact removal methods, establishing KMAR-50K as a foundational resource for developing clinically deployable MRI reconstruction solutions.

## Methods

This section provides a comprehensive account of the sources and processing procedures utilized in the provided dataset (Fig. 1). In order to prove the usability of the dataset for the purpose of motion artifact removal, we have presented a detailed introduction to three categories of frequently-used image motion-artifact removal models, accompanied by an evaluation of the models and the results they generate (Fig. 2).

**Ethics approval and consent to participate.** This study was reviewed and approved by the Ethics Committee of Sichuan Provincial People's Hospital (Approval No. (Research) 3, 2024). The committee granted a waiver of informed consent for the following reasons:

- (1) This is a retrospective study using previously acquired clinical imaging data;
- (2) All personal identifiers and protected health information were completely removed from the DICOM files and associated data;



**Fig. 2** Workflow of this study. This study was conducted in five steps. Firstly, pairs of ground truth images and noise images were collected for a total of 1,104 patients (1,341 pairs of MR images). Subsequently, the ground truth and noise images were subjected to preprocessing, including N4 bias field correction, normalization and image registration. Thirdly, the de-artifact model was trained based on the pre-processed images of 90% of the patients. The network included in this study comprised DDPM (Denoising Diffusion Probabilistic Models), EDSR (Enhanced Deep Residual Networks for Single Image Super-Resolution) and U-Net. Fourthly, the performance of models were evaluated in the remaining 10% of patients and independent testing cohort by comparing the difference between the generated image and the pre-processed ground truth images. Furthermore, subgroup analyses were conducted to assess the performance of images with varying axes.

- (3) The research involves no more than minimal risk to participants. The ethics committee explicitly approved the open publication of this fully anonymized dataset. All DICOM tags containing personal information (including patient names, IDs, birth dates, and examination dates) were systematically removed prior to data sharing.

**KMAR-50K dataset.** The multi-parametric knee MRI dataset of this study comprises 1,190 patients who underwent knee joint scans twice, providing images with motion artifact and ground truth images without artifact (after rescanning while staying in that scanner). The imaging data was acquired in three different views: coronal, sagittal, and transverse. The following three data sources are provided:

- (1) Anonymous but unprocessed data: these data from two field strengths (1.5 T and 3.0 T), with a layer thickness range of 3 mm to 5 mm, and a median thickness of 3.5 mm. It includes six sequences: PD-TSE, PD-FSE, T1-TSE, T1-FSE, T2-FSE, and T2-TSE.
- (2) Preprocessed data: Images generated after N4 bias field correction and maximum minimum normalization for both artifact and ground-truth images. In addition, this dataset also provides ground-truth images mapped to the artifact images by spatial registration.
- (3) Radiologist's subjective scoring data: In evaluating the artifact and ground truth images, junior and senior radiologists conducted subjective assessments, including the degree of artifact, target structure, and overall image quality. Each of the three indicators was scored on a five-point scale, with higher scores indicating lower severity of image artifacts, superior diagnostic ability of target structures, and higher overall image quality.

View	Sequence	FOV	Slice thickness (mm, range)	TR (ms)	TE (ms)
<b>Aera (1.5 T)</b>					
Coronal	PD-TSE	170 × 170	[3.0, 4.5]	3360	47
	T1-TSE	170 × 170	[3.0~3.5]	680	13
	T2-TSE		3.5		
Sagittal	PD-TSE	170 × 170	[2.5, 4.0]	3360	47
	T1-TSE	170 × 170	[3.0, 4.0]	957	12
Transection	PD-TSE	170 × 170	[3.5, 4.0]	3360	47
	T2-TSE		4.0		
<b>Verio (3.0 T)</b>					
Coronal	PD-TSE	170 × 170	[3.0, 3.5]	3000	34
	T1-TSE	170 × 170	[3.0, 3.5]	516	17
Sagittal	PD-TSE	170 × 170	[3.0, 4.0]	3000	34
	T1-TSE	170 × 170	3.5	537	17
Transection	PD-TSE	180 × 180	[3.5, 5.0]	3300	34
<b>Alltech Medical Systems (1.5 T)</b>					
Coronal	PD-FSE		3.5		
	T2-FSE		[3.5, 5.0]		
Sagittal	PD-FSE		[3.0, 4.5]		
	T1-FSE		[3.0, 3.5]		
	T2-FSE		3.5		
Transection	PD-FSE		5.0		
	T2-FSE		5.0		

**Table 1.** Detailed parameter information. mm: millimeter. ms: millisecond. PD-FSE: Proton Density-weighted Fast Spin Echo. PD-TSE: Proton Density-weighted Turbo Spin Echo. T1-FSE: T1-weighted Fast Spin Echo. T1-TSE: T1-weighted Turbo Spin Echo. T2-FSE: T2-weighted Fast Spin Echo. T2-TSE: T2-weighted Turbo Spin Echo.

**Image acquisition.** The MRI scan was obtained from two manufacturers - Siemens (87.0% for Aera 1.5 T, 11.8% for Verio 3.0 T) and Alltech Medical Systems (1.2%) at the Department of Radiology, Sichuan Provincial People's Hospital, Chengdu, China. The quartiles of slice thickness (unit: mm) are 3.5 (3.5~4.0) and 3.5 (3~5.0), respectively. The specific parameters are shown in the Table 1.

**Image inclusion.** The study initially included 330, 445, 439 and 86 patients who underwent knee MRI at Sichuan Provincial People's Hospital from 2020 to 2023, resulting in a total of 1,300 cases. In 51 cases, non-motion artifacts were excluded. In 57 cases, other problems were identified, including inconsistent ranges before and after rescanning, altered PD sequence parameters, magnetic field inhomogeneity, and so forth. Two additional cases were excluded due to missing frames and inconsistencies before and after rescanning. These exclusions resulted in a total of 110 cases being excluded from further analysis. Ultimately, a total of 1,104 patients were included in the study from 2020 to 2022, which were randomly divided into a training cohort (n = 994) and a validation cohort (n = 110) in a ratio of 9:1<sup>24</sup>. Patients starting from 2023 were included as an independent testing cohort (n = 86). This partitioning strategy is designed to provide the model with sufficient data for generalization while also ensuring an unbiased evaluation of its performance. The process flowchart is shown in Fig. 1.

**Scanning image quality evaluation.** A junior and a senior radiologist independently evaluated the quality of the images before and after rescanning. The criteria and basis for scoring the degree of artifact, target structure, and overall image quality are shown in Table 2. All features were evaluated using a 5-point scale, with lower scores indicating the presence of lighter artifacts, less obvious target structures, and poorer image quality. Higher scores indicate superior image quality, which can be utilized for clinical diagnosis. The intra-class correlation (ICC) method was employed to assess the inter-observer agreement in the evaluation of two radiologists<sup>25</sup>.

**Pre-processing and quality control.** Image preprocessing consists of the following steps: (1) The raw DICOM files were sorted by sequence and converted to NIFTI format; (2) All images underwent N4 bias field correction in order to eliminate intensity inhomogeneity correction of MR images<sup>26</sup>; (3) The adaptive normalizer was employed to remove voxels with MR image intensity values exceeding 95% and falling below 5%<sup>27</sup>; (4) Max-min Normalization was employed to standardize the image intensity values between 0 and 1; (5) By employing the motion artifact image as a fixed image and leveraging the synchronous normalization algorithm of the Advanced Normalization Tools (ANTs) toolkit<sup>28</sup>, which is a form of intensity-based image registration, the ground truth image is mapped onto a fixed image. This process achieves spatial alignment of paired images, ensuring that subsequent analyses account for any spatial variations introduced by motion artifacts.

**Model construction.** In order to investigate the efficacy of different baseline models for the removal of motion artifacts, this study explored three frequently used baseline models on the KMAR-50K dataset.

Type	Score	Explanation
Artifact score	1 (very poor)	severe artifacts and cannot be used for diagnosis
	2 (poor)	lots of artifacts
	3 (moderate)	slight artifacts
	4 (good)	minimal artifacts, clear images
	5 (very good)	no artificial artifacts
Target structure score	1 (very poor)	the target structure cannot be displayed
	2 (poor)	the target structure is blurred
	3 (moderate)	the target structure is slightly blurred
	4 (good)	the meniscus, ligaments, and cartilage are displayed clearly
	5 (very good)	the meniscus, ligaments, and cartilage are displayed clearly, with a higher signal-to-noise ratio
Image quality score	1 (very poor)	cannot be used for clinical diagnosis and needs to be rescanned
	2 (poor)	do not need to be rescanned
	3 (moderate)	clinically acceptable
	4 (good)	no obvious artifacts
	5 (very good)	no artifacts and high signal-to-noise ratio

**Table 2.** Criteria and basis for scoring the degree of artifacts, target structure and overall image quality.

- (1) **DDPM:** The Denoising Diffusion Probabilistic Model (DDPM) is introduced and developed in the work of Jonathan Ho<sup>29</sup>, which are essential probabilistic models for generating data distributions gradually. In particular, the forward diffusion process defines an operation where the data structure is gradually destroyed, for example, by corrupting it with noise<sup>30</sup>. The reverse diffusion process is an iterative de-noising method whereby a U-Net is trained to model a target distribution, gradually recovering the ground truth image data structure.
- (2) **EDSR:** The Enhanced Deep Residual Networks for Single Image Super-Solution (EDSR) model represents an advanced deep learning framework that has been specifically designed for the purpose of single image super-resolution reconstruction. The EDSR model was proposed by Ming Hui Zhang *et al.* and effectively addresses the issues of blur and artifacts present in traditional super-resolution methods through the use of a Residual in Residual structure<sup>31</sup>. This model is capable of generating high-quality, high-resolution images by learning the mapping relationship between low-resolution and high-resolution images.
- (3) **U-Net:** We have implemented novel enhancements to the EDSR model by substituting the initial deep residual network configuration with the widely acknowledged U-Net architectural framework. The U-Net, as proposed by Ronneberger *et al.*, is an efficient convolutional neural network that achieves super-resolution reconstruction of images through an encoder-decoder structure<sup>32</sup>. The primary advantage of U-Net is its symmetrical network design, which enables the network to capture comprehensive feature information during the encoding phase and accurately restore the details of high-resolution images during the decoding phase.

The three models above were trained in a computing environment equipped with a NVIDIA A40 GPU, thereby achieving dual improvements in image quality and resolution. The input for model training is uniformly a cropping size of  $256 \times 256$  pixels, which simulate an imaging spacing of  $0.5 \times 0.5$ . This ensures consistency and comparability of the input data. During the training process, the EDSR and U-Net models were further enhanced in their ability to generalize through the utilization of data augmentation techniques, including shift, rotation, and flip, which enabled the simulation of various image deformations that may be encountered in practical applications. To ensure precise optimization of the predictive performance of the models, all three models employ the L1 loss function, which enhances the accuracy of reconstructed images by minimizing the absolute error between the predicted results and the ground truth images. Additionally, the Adam optimizer was selected, as it is an adaptive learning rate optimization algorithm that can effectively adjust the learning rate and accelerate the learning process of the model. The experimental environment is based on Python 3.7.13 and utilizes Torch 1.13.1 and torch-vision 0.14.1 libraries, which provide robust support for these models.

**Evaluation methods.** This study employed six key indicators to evaluate image quality, which are categorized into three distinct groups: similarity, error, and signal-to-noise ratio metrics. The similarity metrics include the structural similarity index matrix (SSIM) and normalized cross-correlation (NCC)<sup>33</sup>, which evaluate the structural and correlation similarities between images. The error metrics consist of the mean absolute error (MAE)<sup>34</sup>, normalized root mean square error (NRMSE)<sup>35</sup> and L1 loss, which quantify the discrepancies between predicted and actual values. Additionally, the Peak Signal-to-Noise Ratio (PSNR) is employed as a signal-to-noise ratio metric, reflecting the image's contrast and the efficacy of noise suppression. A variance analysis was conducted to identify significant performance differences among the models. Additionally, subgroup analyses were performed for transverse, sagittal, and coronal view to evaluate the impact of direction on image quality. Computational efficiency was also compared, encompassing parameters, floating point operations (FLOPS), memory, convolutional layers, maximum stride, and forward time. This provides a comprehensive reference for model selection and evaluation.

Characteristics	Value
Gender <sup>a</sup>	
Female	632 (53%)
Male	558 (47%)
Age <sup>a</sup>	46(18~83)
Field Strength <sup>a</sup>	
1.5 T	1034 (87%)
3.0 T	156 (13%)
Series <sup>b</sup>	
PD-FSE	5 (0.35%)
PD-TSE	1191 (82.48)
T1-FSE	4 (0.28%)
T1-TSE	227 (15.72%)
T2-FSE	6 (0.42%)
T2-TSE	11 (0.76%)
Axis <sup>b</sup>	
Transverse	492 (34%)
Coronal	441 (31%)
Sagittal	511 (35%)

**Table 3.** Demographic information. <sup>a</sup>represents patient-level statistics and <sup>b</sup>represents sequence-level statistics. A total of 1,444 MR Sequences were obtained from 1,190 patients. PD-FSE: Proton Density-weighted Fast Spin Echo. PD-TSE: Proton Density-weighted Turbo Spin Echo. T1-FSE: T1-weighted Fast Spin Echo. T1-TSE: T1-weighted Turbo Spin Echo. T2-FSE: T2-weighted Fast Spin Echo. T2-TSE: T2-weighted Turbo Spin Echo.

## Data Records

The dataset is publicly available via Mendeley Data repositories<sup>36</sup>: Repository 1 (<https://data.mendeley.com/datasets/xw7mrg7ntg>) and Repository 2 (<https://data.mendeley.com/datasets/95w9f5tzz8>)<sup>37</sup>.

The Training Cohort (1,341 sequences from 1,104 patients) is stored in folders ArtifactData\_part1, GroundTruthData\_part1 (Repository 1) and ArtifactData\_part2, GroundTruthData\_part2 (Repository 2), while the Independent Testing Cohort (103 sequences from 83 patients) is in folders Testing\_ArtifactData (Repository 2) and Testing\_GroundTruthData (Repository 1). All files are in NIfTI format (.nii.gz), with filenames denoting processing stages: “N4” indicates N4 bias field correction, “Norm” signifies subsequent intensity normalization, and “SegSyN” (exclusive to ground truth folders) denotes ground truth images registered to artifact space via SyN algorithm of ANTs (<https://antspy.readthedocs.io>).

Metadata tables include TrainingCohort.csv and TestingCohort.csv, with columns: Patient (ID format: Year\_Order\_Direction\_SeriesNumber, matching filenames), StudyDate, SeriesDescription, MagneticFieldStrength, SequenceName, RepetitionTime, EchoTime, FlipAngle, SliceThickness, PixelSpacing, Manufacturer, ManufacturerModel.

## Technical Validation

**Subject characteristics.** The data set was derived from 1,190 patients who underwent knee scans, with a total of 1,444 pairs of MRI sequences (Table 3) in training, validation and independent testing cohort. The population consisted of 558 males (47%) and 632 females (53%), with a median age of 46 years (quartile range: 18–83 years). The PD-TSE, T1-TSE, and T2-TSE accounted for 82.83% (n = 1,196), 16.00% (n = 231), and 1.17% (n = 17), respectively. The distribution of sequences across the transverse, sagittal, and coronal planes was as follows: 34% (n = 492), 31% (n = 441), and 35% (n = 511), respectively.

**Consistency analysis of subjective evaluation of scanned images.** The inter-rater reliability for the degree of artifacts, target structure display, and overall image quality was strong. The ICC results demonstrated that both radiologists assigned a score of 2 (1~3) to the degree of artifacts in the first scan, with an ICC value of 0.882 (0.866~0.895). Following the rescanning, the distribution of artifact degree scores shifted to 5 (4~5), with an ICC value of 0.871 (0.855~0.886). Similarly, the target structure score distribution after the first scan and rescans was 2 (2~3) and 4 (3~4), with ICC values of 0.837 (0.816~0.856) and 0.869 (0.852~0.884), respectively. The overall quality of the image after the first scan and the rescan was rated 2 (1~2) and 4 (3~4), respectively. The radiologists score consistency was 0.864 (0.847~0.88) and 0.891 (0.877~0.903), respectively. The average scores of the rescanned images were found to be significantly higher than those of the first scanned images for all three of these metrics (paired Wilcoxon-test,  $p < 0.001$ ).

**Motion artifacts removal performance of different models.** The evaluation results of three different motion artifact removal models are presented in Table 4 and Fig. 3. In the independent testing cohort, the ESDR method exhibited the most optimal performance in terms of PSNR and NRSME, with average values of 27.704 and 0.046, respectively. With regard to SSIM, NCC, MSA, and L1 Loss, the U-Net demonstrated the most optimal performance, with average values of 0.921, 0.979, 0.011 and 0.026, respectively. The subgroup analysis of

**Table 4-1. Evaluation results of the three networks on three views in the training cohort**

Models	PSNR	SSIM	NCC	NRMSE	MSA	L1 Loss
DDPM	24.496 (24.124–24.868)	0.879 (0.871–0.887)	0.966 (0.961–0.971)	0.065 (0.061–0.069)	0.014 (0.013–0.015)	0.036 (0.034–0.038)
Transverse	23.268 (22.645–23.891)	0.877 (0.864–0.890)	0.963 (0.957–0.969)	0.074 (0.068–0.080)	0.017 (0.015–0.019)	0.041 (0.037–0.045)
Coronal	23.151 (22.547–23.755)	0.861 (0.844–0.878)	0.955 (0.943–0.967)	0.076 (0.069–0.083)	0.015 (0.013–0.017)	0.041 (0.037–0.045)
Sagittal	26.829 (26.385–27.273)	0.897 (0.888–0.906)	0.980 (0.977–0.983)	0.047 (0.044–0.050)	0.010 (0.009–0.011)	0.026 (0.024–0.028)
EDSR	<b>27.490 (27.329–27.651)</b>	0.820 (0.816–0.824)	0.955 (0.952–0.958)	<b>0.047 (0.046–0.048)</b>	0.016 (0.015–0.017)	0.046 (0.045–0.047)
Transverse	26.624 (26.304–26.944)	0.820 (0.812–0.828)	0.963 (0.959–0.967)	0.053 (0.051–0.055)	0.021 (0.020–0.022)	0.051 (0.049–0.053)
Coronal	27.308 (27.019–27.597)	0.812 (0.805–0.819)	0.926 (0.920–0.932)	0.048 (0.046–0.050)	0.012 (0.011–0.013)	0.054 (0.052–0.056)
Sagittal	28.356 (28.163–28.549)	0.826 (0.819–0.833)	0.979 (0.976–0.982)	0.042 (0.041–0.043)	0.016 (0.015–0.017)	0.035 (0.034–0.036)
U-Net	27.110 (26.880–27.340)	<b>0.922 (0.918–0.926)</b>	<b>0.980 (0.978–0.982)</b>	0.048 (0.046–0.050)	<b>0.011 (0.011–0.011)</b>	<b>0.026</b> <b>(0.025–0.027)</b>
Transverse	25.949 (25.539–26.359)	0.913 (0.906–0.920)	0.977 (0.974–0.980)	0.055 (0.052–0.058)	0.014 (0.013–0.015)	0.031 (0.029–0.033)
Coronal	25.755 (25.370–26.140)	0.914 (0.906–0.922)	0.974 (0.969–0.979)	0.057 (0.053–0.061)	0.011 (0.010–0.012)	0.029 (0.027–0.031)
Sagittal	29.136 (28.854–29.418)	0.937 (0.933–0.941)	0.988 (0.987–0.989)	0.037 (0.036–0.038)	0.008 (0.008–0.008)	0.019 (0.018–0.020)

**Table 4-2. Evaluation results of the three networks on three views in the validation cohort**

DDPM	24.541 (24.089–24.993)	0.888 (0.882–0.894)	0.972 (0.969–0.975)	0.062 (0.059–0.065)	0.013 (0.012–0.014)	0.033 (0.031–0.035)
Transverse	26.689 (26.049–27.329)	0.902 (0.893–0.911)	0.980 (0.977–0.983)	0.048 (0.045–0.051)	0.009 (0.008–0.010)	0.026 (0.024–0.028)
Coronal	23.740 (23.118–24.362)	0.887 (0.875–0.899)	0.971 (0.967–0.975)	0.067 (0.062–0.072)	0.015 (0.013–0.017)	0.038 (0.035–0.041)
Sagittal	23.000 (22.426–23.574)	0.873 (0.862–0.884)	0.965 (0.960–0.970)	0.072 (0.067–0.077)	0.013 (0.012–0.014)	0.038 (0.036–0.04)
EDSR	<b>27.275 (26.958–27.592)</b>	0.827 (0.817–0.837)	0.960 (0.954–0.966)	0.047 (0.045–0.049)	0.016 (0.015–0.017)	0.045 (0.043–0.047)
Transverse	27.829 (27.232–28.426)	0.816 (0.799–0.833)	0.977 (0.973–0.981)	0.045 (0.042–0.048)	0.017 (0.015–0.019)	0.036 (0.033–0.039)
Coronal	26.847 (26.436–27.258)	0.839 (0.822–0.856)	0.971 (0.967–0.975)	0.050 (0.048–0.052)	0.019 (0.017–0.021)	0.046 (0.044–0.048)
Sagittal	27.103 (26.523–27.683)	0.826 (0.809–0.843)	0.929 (0.917–0.941)	0.048 (0.045–0.051)	0.011 (0.010–0.012)	0.052 (0.049–0.055)
U-Net	26.720 (26.244–27.196)	<b>0.927 (0.921–0.933)</b>	<b>0.982 (0.980–0.984)</b>	<b>0.046 (0.038–0.057)</b>	<b>0.011 (0.01–0.012)</b>	<b>0.025</b> <b>(0.024–0.026)</b>
Transverse	<i>28.683 (27.985–29.381)</i>	<i>0.933 (0.924–0.942)</i>	<i>0.986 (0.983–0.989)</i>	<i>0.038 (0.035–0.041)</i>	<i>0.008 (0.007–0.009)</i>	<i>0.020</i> <i>(0.018–0.022)</i>
Coronal	26.002 (25.414–26.590)	0.927 (0.917–0.937)	0.982 (0.979–0.985)	0.051 (0.048–0.054)	0.013 (0.011–0.015)	0.029 (0.027–0.031)
Sagittal	25.298 (24.505–26.091)	0.920 (0.908–0.932)	0.978 (0.973–0.983)	0.057 (0.052–0.062)	0.011 (0.01–0.012)	0.028 (0.026–0.03)

**Table 4-3. Evaluation results of the three networks on three views in the Independent testing cohort**

DDPM	24.948 (24.381~25.514)	0.882 (0.873~0.891)	0.970 (0.965~0.974)	0.060 (0.056~0.063)	0.013 (0.011~0.014)	0.033 (0.031~0.035)
Transverse	27.006 (26.278~27.734)	0.896 (0.880~0.912)	0.979 (0.976~0.983)	0.046 (0.042~0.051)	0.009 (0.008~0.011)	0.025 (0.023~0.028)
Coronal	23.437 (22.559~24.316)	0.877 (0.861~0.892)	0.965 (0.956~0.974)	0.069 (0.062~0.077)	0.016 (0.013~0.018)	0.039 (0.035~0.043)
Sagittal	23.203 (22.502~23.905)	0.867 (0.854~0.880)	0.960 (0.951~0.969)	0.071 (0.065~0.077)	0.015 (0.013~0.017)	0.039 (0.036~0.043)
EDSR	<b>27.704 (27.311~28.098)</b>	0.815 (0.800~0.829)	0.959 (0.953~0.964)	<b>0.046 (0.044~0.048)</b>	0.015 (0.014~0.017)	0.044 (0.042~0.047)
Transverse	28.307 (27.725~28.889)	0.815 (0.790~0.841)	0.975 (0.971~0.978)	0.043 (0.040~0.047)	0.014 (0.012~0.016)	0.034 (0.031~0.038)
Coronal	26.693 (25.894~27.492)	0.825 (0.797~0.853)	0.964 (0.954~0.975)	0.051 (0.046~0.056)	0.020 (0.017~0.023)	0.049 (0.045~0.054)
Sagittal	27.603 (26.925~28.281)	0.806 (0.785~0.827)	0.932 (0.920~0.944)	0.045 (0.041~0.049)	0.014 (0.012~0.016)	0.055 (0.051~0.059)
U-Net	26.695 (26.064~27.326)	<b>0.921 (0.913~0.928)</b>	<b>0.979 (0.976~0.983)</b>	0.050 (0.046~0.053)	<b>0.011 (0.010~0.012)</b>	<b>0.026</b> <b>(0.024~0.028)</b>
Transverse	28.468 (27.562~29.374)	0.927 (0.914~0.941)	0.985 (0.981~0.988)	0.040 (0.036~0.045)	0.008 (0.007~0.010)	0.021 (0.018~0.023)
Coronal	25.350 (24.270~26.430)	0.917 (0.902~0.932)	0.975 (0.967~0.984)	0.056 (0.049~0.064)	0.013 (0.011~0.015)	0.031 (0.027~0.034)
Sagittal	25.224 (24.283~26.165)	0.914 (0.902~0.926)	0.974 (0.967~0.981)	0.057 (0.051~0.064)	0.013 (0.011~0.015)	0.030 (0.027~0.033)

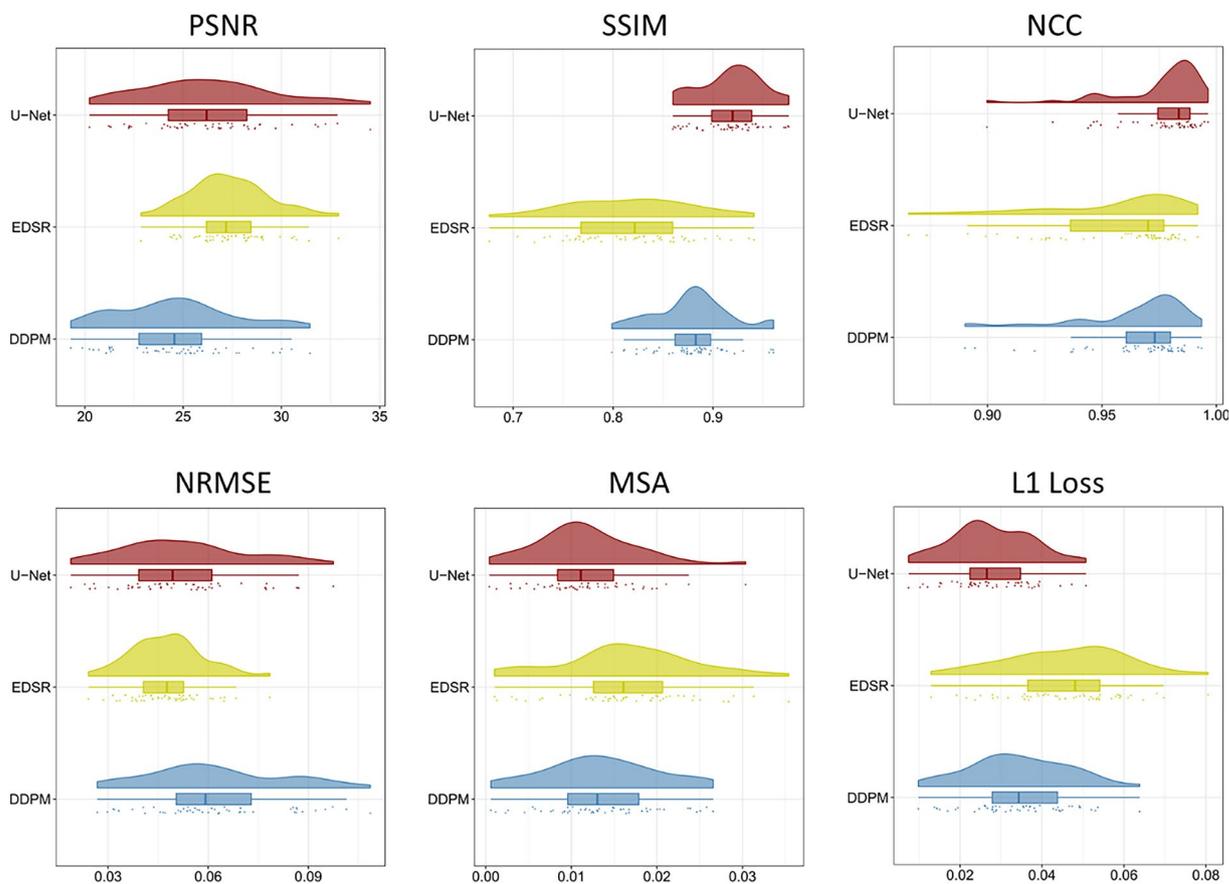
**Table 4.** The bold value represents the best overall performance on three views for each metric. The subgroup analysis indicates that transverse images perform optimally in the U-Net (italics). DDPM: Denoising Diffusion Probabilistic Models; EDSR: Enhanced Deep Residual Networks for Single Image Super-Resolution; PSNR: Peak Signal to Noise Ratio; SSIM: Structural Similarity Index; NCC: Neighborhood Cross-Correlation; NRMSE: Normalized Root Mean Square Error; MSA: Median Symmetric Accuracy.

transverse, sagittal and coronal images revealed that the performance of transverse images was superior, whereas that of sagittal images was the least effective. Figure 4 illustrates the comparison of generated images and ground truth for three instances under different models. While the EDSR and U-Net excel in specific metrics, their performance trade-offs suggest that no single method universally addresses all aspects of motion artifact removal. The persistent underperformance on sagittal images across all models may indicate inherent challenges in capturing anisotropic motion patterns or limited representation of sagittal-plane artifacts in training data.

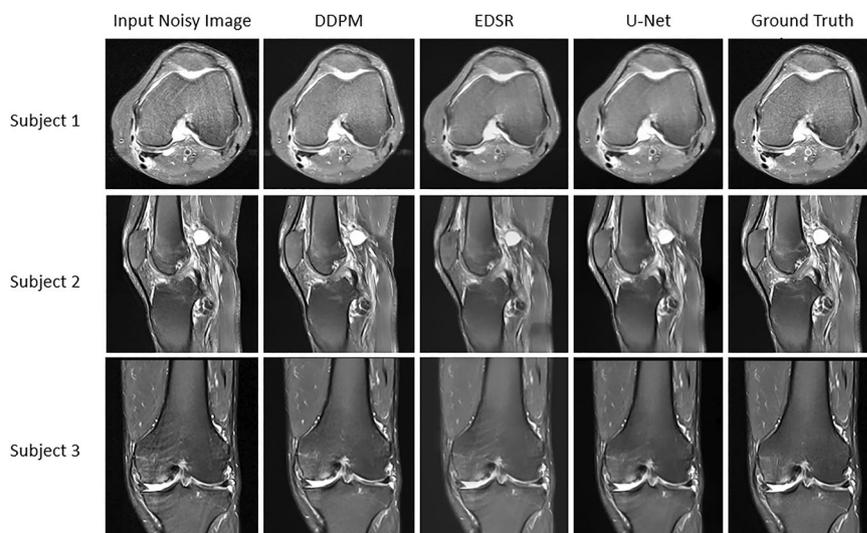
**Comparison results of different models.** According to ANOVA test, there were statistically significant differences among the DDPM, EDSR and U-Net with regard to the six evaluation metrics (all  $p < 0.001$ ). A pairwise comparison of the models using a paired t-test or a paired Wilcoxon test revealed that the EDSR exhibited significantly superior performance compared to the DDPM and U-Net for the PSNR metric ( $p < 0.001$ ). Conversely, the U-Net demonstrated significantly enhanced performance compared to the DDPM and EDSR for the SSIM, NCC, NRMSE, MSA and L1 loss (all  $p < 0.001$ ). The results of the subgroup analyses of the three views were found to be consistent. Subgroup analyses (transverse, sagittal, and coronal) yielded consistent results.

A comparative analysis was conducted on the parameters, memory consumption, FLOPs, forward time (average of 132 experiments from validation cohort conducted on NVIDIA A40), number of convolutional layers, and maximum stride of different networks. In comparison to DDPM, U-Net exhibits a markedly reduced parameter count (17.3 M), memory consumption (11.41MB), FLOPs (39991), and convolutional layers (20). In contrast to EDSR, the forward time has been observed to decrease from 9 s to 0.5 s (for further details, please refer to Table 5). The computational efficiency of U-Net (0.5 seconds inference time) comes at the cost of a lower PSNR, while the high accuracy of EDSR requires 18 times the computational time of U-Net, limiting its real-time clinical applicability. While DDPM's 3315-second inference time is clinically impractical, its visually plausible motion artifact removal motivates future work on accelerated diffusion models (e.g., latent diffusion) to reduce computational costs while preserving perceptual fidelity.

**Limits of the dataset.** Our dataset has several limitations. Firstly, the dataset is constrained by its single-center retrospective design, which may not fully encompass the diversity of practices and patient



**Fig. 3** Raincloud plot of each matrix on different models in testing cohort. The larger the three indicators, PSNR, SSIM, and NCC (the first row), the superior the model performance. Conversely, the smaller the NRSME, MSA, and L1 Loss (the second row), the more optimal the result. The dots represent the testing samples. The width of the violin plot represents the concentration trend of the various matrices under consideration. In a box-and-whisker plot, the box represents the inter-quartile range (IOR), the central line represents the median, the whisker boundaries extend to a distance of 1.5 IOR, and points outside the whisker boundaries represent outliers.



**Fig. 4** The de-motion-artifact results of various algorithms are compared in three subjects with different axial views. The image input to the model was preprocessed. In DDPM, EDSR and U-Net models, the PSNR values for the subject 1 (the first row) were 30.80, 27.11 and 31.91, while the SSIM values were 0.93, 0.92 and 0.96. The corresponding values for the subjects 2 (the second row) and 3 (the third row) were 26.93, 22.737, 28.66 and 0.91, 0.84, 0.95; 28.62, 25.63, 30.72 and 0.95, 0.91, 0.97.

Networks	DDPM	EDSR	U-Net
Parameters (M)	926	16.6	17.3
FLOPS (GMac)	200887	1088447	39991
Memory (M)	3594	85	88
Convolutional layers	186	19	20
Max stride	16	1	16
Forward time (s)	3315	9	0.5

**Table 5.** Comparison of parameters, FLOPS, memory consumption, forward time (averaged over 132 trials on an NVIDIA A40), number of convolutional layers, and the coarsest stride of different networks in validation cohort. The M and GMac represent Mega and Giga Multiply-Accumulate Operations per Second, respectively. The parameters serve to reflect the complexity and capacity of the model. An increase in the number of parameters results in a greater potential for learning, but also an elevated risk of overfitting. FLOPS represents the number of floating-point operations performed by the model in the forward propagation phase. An increase in FLOPS results in a proportional increase in the computational cost of the model, which in turn leads to a longer inference time. Memory is used to describe the quantity of memory utilized by the model throughout the processes of training and inference. The number of convolutional layers can be used as an indicator of the depth and complexity of the model.

demographics across different centers. Secondly, the overwhelming majority of the data in the dataset (98.9%) originates from Siemens scanners, which introduces vendor-specific biases that may impact the generalizability of our results on other manufacturer platforms. Additionally, the data from the 1.5 T scanner (88%) is markedly larger than that from 3.0 T and higher field strengths, indicating that our method may not be fully evaluated under a more diverse range of imaging conditions. Moreover, the retrospective nature of our study may introduce selection bias, which may not accurately reflect the characteristics of a more inclusive patient population.

### Code availability

In this study, U-Net is available at <https://github.com/milesial/Pytorch-UNet>, EDSR can be assessed at <https://github.com/yulunzhang/EDSR-PyTorch>, and DDPM can be found at [https://gitcode.com/gh\\_mirrors/me/med-seg-diff-pytorch](https://gitcode.com/gh_mirrors/me/med-seg-diff-pytorch). We gratefully acknowledge the contributions of these open-source projects and encourage readers to consult the respective repositories for detailed implementation and usage instructions.

Received: 26 December 2024; Accepted: 20 June 2025;

Published online: 09 July 2025

### References

- Carr, M. W. & Grey, M. L. Magnetic resonance imaging. *Am J Nurs* **102**, <https://doi.org/10.1097/0000446-200212000-00012> (2002).
- Kijowski, R. A.-O. & Fritz, J. Emerging Technology in Musculoskeletal MRI and CT. *Radiology* **306**, <https://doi.org/10.1148/radiol.220634> (2023).
- Del Grande, F. *et al.* Five-Minute Five-Sequence Knee MRI Using Combined Simultaneous Multislice and Parallel Imaging Acceleration: Comparison with 10-Minute Parallel Imaging Knee MRI. *Radiology* **299**, <https://doi.org/10.1148/radiol.2021203655> (2021).
- Chalian, M. A.-O. *et al.* The QIBA Profile for MRI-based Compositional Imaging of Knee Cartilage. *Radiology* **301**, <https://doi.org/10.1148/radiol.2021204587> (2021).
- Spieker, V. *et al.* Deep Learning for Retrospective Motion Correction in MRI: A Comprehensive Review. *IEEE Trans Med Imaging* **43**, <https://doi.org/10.1109/TMI.2023.3323215> (2024).
- McClelland, J. R. *et al.* Respiratory motion models: a review. *Med Image Anal* **17**, <https://doi.org/10.1016/j.media.2012.09.005> (2013).
- Ringe, K. A.-O. X. & Yoon, J. A.-O. Strategies and Techniques for Liver Magnetic Resonance Imaging: New and Pending Applications for Routine Clinical Practice. *Korean J Radiol* **24**, <https://doi.org/10.3348/kjr.2022.0838> (2023).
- Sander, J., de Vos, B. D. & Išgum, I. Autoencoding low-resolution MRI for semantically smooth interpolation of anisotropic MRI. *Med Image Anal* **78**, <https://doi.org/10.1016/j.media.2022.102393> (2022).
- Yang, J. *et al.* Fast Multi-Contrast MRI Acquisition by Optimal Sampling of Information Complementary to Pre-Acquired MRI Contrast. *IEEE Trans Med Imaging* **42**, <https://doi.org/10.1109/TMI.2022.3227262> (2023).
- Lei, P. *et al.* Joint Under-Sampling Pattern and Dual-Domain Reconstruction for Accelerating Multi-Contrast MRI. *IEEE Trans Image Process* **33**, <https://doi.org/10.1109/TIP.2024.3445729> (2024).
- Friedrich, M. A.-O. Steps and Leaps on the Path toward Simpler and Faster Cardiac MRI Scanning. *Radiology* **298**, <https://doi.org/10.1148/radiol.2021204084> (2021).
- Haldar J. P. & Setsompop, K. Linear Predictability in MRI Reconstruction: Leveraging Shift-Invariant Fourier Structure for Faster and Better Imaging. *IEEE Signal Process Mag* **37**, <https://doi.org/10.1109/msp.2019.2949570> (2020).
- Pietsch, M., Christiaens, D., Hajnal, J. V. & Tournier, J. D. dStripe: Slice artifact correction in diffusion MRI via constrained neural network. *Med Image Anal* **74**, <https://doi.org/10.1016/j.media.2021.102255> (2021).
- An, H. J. *et al.* MRI-Based Attenuation Correction for PET/MRI Using Multiphase Level-Set Method. *J Nucl Med* **57**, <https://doi.org/10.2967/jnumed.115.163550> (2016).
- Klodowski, K. A.-O., Sengupta, A., Dragonu, I. & Rodgers, C. A.-O. Prospective 3D Fat Navigator (FatNav) motion correction for 7T Terra MRI. *NMR Biomed* **38**, <https://doi.org/10.1002/nbm.5283> (2024).
- Hewlett, M. A.-O. X., Oran, O. A.-O., Liu, J. A.-O. & Drangova, M. A.-O. Prospective motion correction for brain MRI using spherical navigators. *Magn Reson Med* **91**, <https://doi.org/10.1002/mrm.29961> (2024).
- Ariyurek, C. A.-O., Wallace, T. A.-O., Kober, T. A.-O., Kurugol, S. A.-O. & Afacan, O. A.-O. Prospective motion correction in kidney MRI using FID navigators. *Magn Reson Med* **89**, <https://doi.org/10.1002/mrm.29424> (2023).
- Zheng, S. *et al.* A Deep Learning Method for Motion Artifact Correction in Intravascular Photoacoustic Image Sequence. *IEEE Trans Med Imaging* **42**, <https://doi.org/10.1109/TMI.2022.3202910> (2023).

19. Zhu, Y. *et al.* Sinogram domain metal artifact correction of CT via deep learning. *Comput Biol Med* **155**, <https://doi.org/10.1016/j.combiomed.2023.106710> (2023).
20. Luo, L. *et al.* Scale-Aware Super-Resolution Network With Dual Affinity Learning for Lesion Segmentation From Medical Images. *IEEE Trans Neural Netw Learn Syst* **28**, <https://doi.org/10.1109/TNNLS.2024.3477947> (2024).
21. Zbontar, J. *et al.* fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *arXiv:1811.08839* <https://doi.org/10.48550/arXiv.1811.08839> (2018).
22. Bien, N. A.-O. *et al.* Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Med* **27**, <https://doi.org/10.1371/journal.pmed.1002699> (2018).
23. Štajduhar, I., Mamula, M., Miletic, D. & Unal, G. Semi-automated detection of anterior cruciate ligament injury from MRI. *Comput Methods Programs Biomed* **140**, <https://doi.org/10.1016/j.cmpb.2016.12.006> (2017).
24. Alvarez Villela, M. *et al.* Feasibility of high-intensity interval training in patients with left ventricular assist devices: a pilot study. *ESC Heart Fail* **8**, 498–507, <https://doi.org/10.1002/ehf2.13106> (2021).
25. Yen, M. & Lo, L. H. Examining test-retest reliability: an intra-class correlation approach. *Nurs Res* **51**, <https://doi.org/10.1097/00006199-200201000-00009> (2002).
26. He, X. *et al.* Test Retest Reproducibility of Organ Volume Measurements in ADPKD Using 3D Multimodality Deep Learning. *Acad Radiol* **31**, <https://doi.org/10.1016/j.acra.2023.09.009> (2024).
27. Wang, C. *et al.* The Association Between Tumor Radiomic Analysis and Peritumor Habitat-Derived Radiomic Analysis on Gadodotate Disodium-Enhanced MRI With Microvascular Invasion in Hepatocellular Carcinoma. *J Magn Reson Imaging* <https://doi.org/10.1002/jmri.29523> (2024).
28. Pilutti, D., Strumia, M. & Hadjidemetriou, S. Bi-modal Non-rigid Registration of Brain MRI Data with Deconvolution of Joint Statistics. *IEEE Trans Image Process* **23**, <https://doi.org/10.1109/TIP.2014.2336546> (2014).
29. Ho, J., Jain, A. & Abbeel, P. Denoising Diffusion Probabilistic Models. *arXiv:2006.11239* <https://doi.org/10.48550/arXiv.2006.11239> (2020).
30. Zhao, K. *et al.* MRI Super-Resolution with Partial Diffusion Models. *IEEE Trans Med Imaging* **17**, <https://doi.org/10.1109/TMI.2024.3483109> (2024).
31. Lee, B. L. A. S. S. A. H. K. A. S. N. A. K. M. Enhanced Deep Residual Networks for Single Image Super-Solution. *arXiv:1707.02921* <https://doi.org/10.48550/arXiv.1707.02921> (2017).
32. Ronneberger, O., Fischer, P. & Brox, T. U-Net Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention* **9351**, <https://doi.org/10.48550/arXiv.1505.04597> (2015).
33. Chen, X., Fan, X., Meng, Y. & Zheng, Y. AI-driven generalized polynomial transformation models for unsupervised fundus image registration. *Front Med (Lausanne)* **16**, <https://doi.org/10.3389/fmed.2024.1421439> (2024).
34. Haubold, J. A.-O. *et al.* Contrast agent dose reduction in computed tomography with deep learning using a conditional generative adversarial network. *Eur Radiol* **31**, <https://doi.org/10.1007/s00330-021-07714-2> (2021).
35. Robeson, S. A.-O. & Willmott, C. J. Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *PLoS One* **17**, <https://doi.org/10.1371/journal.pone.0279774> (2023).
36. Wang, Y. & Shi, F. KMAR-50K-part1. *Mendeley Data* <https://doi.org/10.17632/xw7mrg7ntg.5> (2025).
37. Wang, Y. & Shi, F. KMAR-50K-part2. *Mendeley Data* <https://doi.org/10.17632/95w9f5tzz8.6>.

## Acknowledgements

This work was supported by the Talent Program of Sichuan Provincial People's Hospital (No.30420220145).

## Author contributions

Y.W. worked on the concept of the dataset. Y.X. and R.L. worked on the image inclusion. Y.X. and Y.L. performed the quality control and subjective evaluation of the data. F.W. Q.Z. and F.S. wrote the simulation scripts and performed the model-based technical validation. Y.X. and F.W. wrote the main parts of the manuscript; Y.W. and F.S. revised the manuscript. All authors reviewed and approved the manuscript.

## Competing interests

F.W. and F.S. are employees of United Imaging Intelligence. The company has no role in designing and performing the surveillances and analyzing and interpreting the data. All other authors report no conflicts of interest relevant to this article.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-05439-1>.

**Correspondence** and requests for materials should be addressed to Y.L. or Y.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025