Enhanced-quickDWI: Achieving equivalent clinical quality by denoising heavily sub-sampled diffusion-weighted imaging data

Konstantinos Zormpas-Petridis[1], Antonio Candito[2], Christina Messiou[3], Dow-Mu Koh[2,3], and Matthew D. Blackledge[2]

[1] Department of Diagnostic Imaging and Radiation Oncology, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy
[2] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, UK
[3]Department of Radiology, The Royal Marsden NHS Foundation Trust, London, UK

Corresponding author: Matthew D. Blackledge; matthew.blackledge@icr.ac.uk

Abstract. Whole-body diffusion-weighted imaging (DWI) is a sensitive tool for assessing the spread of metastatic bone malignancies. It offers voxel-wise calculation of apparent diffusion coefficient (ADC) which correlates with tissue cellularity, providing a potential imaging biomarker for tumour response assessment. However, DWI is an inherently noisy technique requiring many signal averages over multiple b-values, leading to times of up to 30 minutes for a whole-body exam. We present a novel neural network implicitly designed to provide high-quality images from heavily sub-sampled diffusion data (only 1 signal average) which allow whole-body acquisitions of ~5 minutes. We demonstrate that our network can achieve equivalent quality to the clinical b-value and ADC images in a radiological multi-reader study of 100 patients for whole-body and abdomen-pelvis data. We also achieved good agreement to the quantitative values of clinical images within multi-lesion segmentations in 16 patients compared to a previous approach.

Keywords: Diffusion-weighted MRI, deep learning, image denoising

1 Introduction

1.1 Background

Diffusion-weighted MRI (DWI) is a non-invasive tool used for staging and response evaluation in oncologic practice. Whole-body DWI is at the core of emerging response criteria in advanced prostate and breast cancers [1-3] and has also been incorporated into the National Institute for Health and Care Excellence and International Myeloma Working group guidelines for assessing myeloma-related bone disease [4,5]. DWI is a sensitive tool that radiologists use to review the extent of disease and enables voxel-wise quantification of the change in the apparent diffusion coefficient (ADC), providing a potential marker for tumour response assessment [6].

Whole-body DWI is typically achieved using a series of sequential imaging stations from the head to the mid-thigh, with each station comprising 30–50 equally-spaced axial sections, with images acquired using two to three diffusion weightings and typically many signal averages [7]. Therefore, WBDWI accounts for more than 50% of the

acquisition time of conventional whole-body MRI studies with a 1-hour duration. In the context of the ever-increasing capacity pressures on MRI departments, reducing acquisition times would facilitate the wider adoption of clinical WBDWI, reduce costs, and improve patient acceptance. DWI is also embedded into consensus MRI protocols across almost all tumour types including primary prostate and breast cancers, metastatic liver disease, gynaecological and gastrointestinal cancers [8-12] where time savings would also be beneficial.

1.2 Contributions

We present a context-aware multi b-value deep learning neural network implicitly designed to denoise heavily sub-sampled diffusion-weighted imaging data, producing images with equivalent clinical image quality. We achieve this by retaining the network's 2D nature but introducing spatial context through slices neighboring to the slice of interest, as well as processing simultaneously all respective b-values. Additionally, we introduce a custom-made loss function based on calculating the ADC values during training.

Due to the quantitative nature of DWI and the potential of ADC to assess intra-lesion changes to monitor treatment response, the validation of the network-generated images needs to be meticulous going beyond the assessment of contrast. Therefore, we performed an exhaustive quantitative and qualitative analysis. We, also, compared our novelties against a published methodology which demonstrated that by using only 1 diffusion encoding direction and 1 signal average using b-values such as 50, 600 or 900 s/mm^2 it is possible to achieve up to 50% reduction in whole-body MRI acquisition times by reducing whole-body DWI to less than five minutes.

Overall, in this article, we achieve three objectives: *(i)* blindly compare radiological image quality of *Enhanced-quickDWI* with conventional DWI in a much larger patient cohort, *(ii)* compare quantitative results of our novel model trained with a custom-made loss function against a published methodology (henceforth called "*DNIF*"), and *(iii)* evaluate both algorithms in other cancer types acquired using a smaller field-of-view imaging protocol focused on the abdomen and pelvis.

2 Related Work

One of the most common approaches for DWI acceleration is based on supervised learning where the 2D "noisy" image acquired using fewer signal averages (NeX) is given as input and the "clinical" image reconstructed by all signal averages (NeX: ~9-16 for whole-body) is used as ground-truth. A convolutional neural network usually resembling the U-net architecture is used to denoise the image and is trained by a loss function to minimize the difference between the output and ground-truth. Wessling et al. utilized a variational network to accelerate a breast diffusion sequence by 40% [13] by reducing averages from 4 to 2 for low b-value and from 16 to 8 for high b-value images. Kaye et al. accelerated prostate DWI using a CNN with a mean-squared-error loss function to denoise sub-sampled data of two signal averages [14]. Afat et al., have used a variational network to accelerate live DWI by 40% by reducing the signal

averages from 12 to 6 for the high b-value images and from 2 to 1 for the low b-value images [15]. Self-supervised approaches, such as Noise2Noise [16] which model the noise and do not require labelled data could be incorporated. Kawamura et al used a supervised learning approach with a CNN to predict the residual noise instead of the denoised image directly [17]. Maosong et al., used a residual encoder-decoder Wasserstein generative adversarial network along with perceptual similarity loss function to denoise 3D MRI, but not DWI [18]. Instead of reducing NeX other approaches include shot-to-shot phase reconstruction [19] and k-space undersampling [20].

3 Materials and Methods

3.1 Image Acquisition

This retrospective study was approved by our local ethics committee and subsequently received national approval by the NHS health research authority. Patients were split into training (N = 84), validation (N = 16) and test (N = 84) groups. Validation data were used for quantitative evaluation of model losses/metrics during training, whilst test data were used for a qualitative radiological evaluation of the model which achieved the best quantitative performance. All images were acquired on one of three 1.5T scanners (two Aera and one Sola, Siemens Healthcare GmbH). We employed a multi-directional diffusion-weighted (MDDW) protocol for acquiring whole-body and abdomen-pelvis DWI examinations, the parameters for which are defined in **Appendix Table 1** (2D distortion correction was applied to all source DWI images).

3.2 Deep Learning Model

Core Architecture. Our *Enhanced-quickDWI* model consists of an encoder-decoder path using residual blocks and skip connections (**Fig. 1**): Each convolution layer consists of a 3x3 filter with ReLU activation and the weights incident to each hidden unit is constrained to have a norm value of less than or equal to 3 after hyperparameter tuning. The number of filters for each block were 64, 128, 256, 512 and 1024 respectively.

Spatial Context and Multi b-value Information. To give spatial context and random direction information to the network, we retain the 2D architecture but give as input a 9-channel image: 3 contiguous slices where the slice of interest is in the middle x all 3 corresponding b-values for every slice. Image voxels have been transformed using a log-transform (any zero values set to 1). Standardization is performed automatically as part of the model (**Eq. 1**). The output of core architecture is a 3-channel image containing the predicted 3 b-value images for the slice of interest. The novelty consists in the prediction of the slice of interest using the two neighboring slices as spatial context instead of predicting all three slices as a volume.

$$z_i = \frac{x_i - \mu}{\sigma} \quad \text{where} \quad \mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and.} \quad \sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad \textbf{[Eq. 1]}$$

Loss Function. We introduced a novel regularization to our model by forcing the network to make accurate predictions of the ADC map during training. We implemented a custom-made non-trainable layer which calculates the ADC map using a linear-log fit from the predicted b-value images, including it as part of the overall loss function. The final output of the model is a 4-channel image comprising all b-value trace-weighted images for the slice of interest plus the corresponding ADC map. The loss function for this model was the sum of mean-absolute-error (MAE) between log-transformed predicted images and ground truth images for all b-values and subsequently derived ADC (training parameters and data preprocessing are presented in **Appendix Table 2**). The network was trained separately on whole-body (b = 50, 600 and 900 s/mm^2) and subsequently on abdomen-pelvis (b = 100, 600 and 1050 s/mm^2) data. The *abdomen-pelvis* model was initialized using pretrained weights from the trained *whole-body* model.

Training Data. For each b-value, 1 diffusion direction and 1 signal average was randomly selected from the MDDW data. Three-slice slabs were created placing the slice of interest in the middle. This process resulted in a total of 16480 input / output data pairs for whole-body training images, 5469 for abdomen-pelvis training images, 2960 for whole-body validation images, and 1122 for abdomen-pelvis validation images.

Comparison. We compared our model against the *DNIF* [21] algorithm, where the input is a 1-channel stardardized (**Eq. 1**) image. The core architecture for both algorithms (e.g. the encoder-decoder path) were kept the same. The loss function for training this model was the mean-absolute error (MAE) between prediction and ground-truth images (all b-values treated equally). For the training data the same single direction and signal average was randomly selected for each slice. This process resulted in a total of 75198 image pairs for training, and 7068 available for validation.

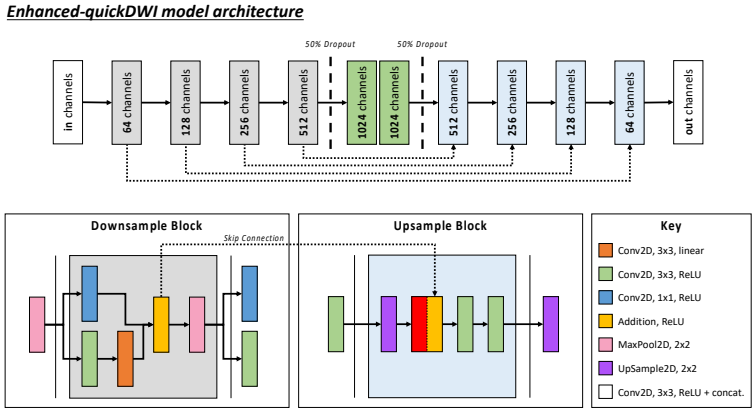

Fig 1. The core *Enhanced-quickDWI* deep-learning model with linear output activation. For the *DNIF* model, in channels = 1 and out channels = 1, whilst for both *whole-body* and *abdomen-pelvis* models, in channels = 9 and out channels = 3.

3.3 Experimental Design

Quantitative Model Evaluation. To quantitatively evaluate the precision of generated DWI and ADC values, we utilized an externally validated deep learning methodology [22] that automatically delineates the skeleton and soft-tissue organs within both whole-body and abdomen-pelvis validation patients (see **Fig 2** for an illustrative workflow, including a list of delineated body regions). For each model, the inferred voxel values (each b-value and derived ADC maps independently) were compared with the ground-truth voxel values from trace-weighted images. A quantitative comparison was made by comparing the voxel-wise root-mean-square-error (RMSE) within each region against the values of the clinical images, with the hypothesis that a better model will reduce the RMSE across all b-values and for ADC maps.

Qualitative Model Evaluation. To assess the clinical utility of our generated images, our model was evaluated by three expert radiologists using a combination of the validation and test datasets, comprising a total of 100 patients. For each patient study, the corresponding pairs of diffusion-weighted images and ADC maps for ground truth target datasets and *Enhanced-quickDWI* model output datasets were uploaded to a secure cloud-based GDPR-compliant radiological image viewing platform (Collective Minds Radiology, www.cmrad.com) for review by independent, external radiologists using a fully anonymized and randomized process. We ensured that at least two weeks elapsed between consecutive reads of the clinical averaged and *Enhanced-quickDWI* processed images (presented in random order) for the same patient study. We also re-anonymised (double-blinded) the images, so that reads of one image type could not inform the reading of the other. Readers scored overall image quality (OIQ) on a 5-point Likert scale for diffusion-weighted images and ADC maps separately.

Statistical Evaluation of Qualitative Analysis. Using a non-inferiority experimental design, we aimed to provide evidence for the null hypothesis that the average Likert score for image quality from *Enhanced-quickDWI* processed images is insignificantly than for clinical, trace-weighted scans (independently for b-value images and ADC maps). A non-inferiority margin of 0.5 was prospectively chosen by an internal expert panel of radiologists to define a non-inferiority threshold.

4 Results and Discussion

4.1 Model Training

All models achieved adequate convergence after 200 epochs for the whole-body model, and 40 epochs for the tuned abdomen-pelvis model. Upon visual inspection our *Enhanced-quickDWI* model successfully denoised the heavily sub-sampled noisy input data and improved their image quality. A representative patient example is presented in **Fig 3**. The *Enhanced-quickDWI*-generated images appear sharper and with more accurate contrast over the *DNIF*-generated. Additionally, in the whole-body projection there is a noticeably smoother transition between neighboring slices (**Fig. 3** left-bottom

and right: red arrows) which is more likely attributed to the inclusion of spatial context compared to the *DNIF* model which was trained only on single slices.

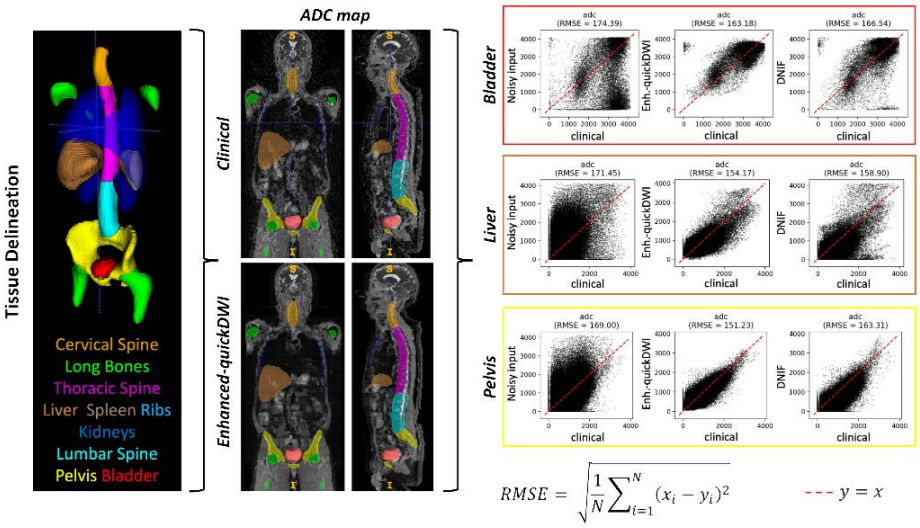

Fig 2. An illustrative workflow for quantifying improvement in image quality using our *Enhanced-quickDWI* model. Delineations for each body regions are presented in different colours, either as a surface rendering (**left**) or colour overlay on the multi-planar reformatted ADC maps (**centre**). The borders of the scatter plots (**right**) are colour-coded to match the body regions from which they were derived.

4.2 Quantitative Model Evaluation

Enhanced-quickDWI was superior to *DNIF* in more than 90% of cases as the RMSE value for whole-body or abdomen-pelvis images is reduced compared to the RMSE values for the *DNIF* model. Both cases show decreased RMSE values when compared to the noisy input MDDW data on every occasion. The same trend is observed across all b-values and ADC maps independently. An overview of the quantitative evaluation of the precision of predicted values for each of the delineated body regions are demonstrated in **Fig 4**. Based on these results the *Enhanced-quickDWI* model was selected over *DNIF* for the independent qualitative reader study.

Despite the same core architecture between the *Enhanced-quickDWI* and *DNIF* models this marked improvement in performance can be explained by the inclusion of the neighboring slices as spatial context which can make image transition smoother. The inclusion of the ADC calculation inside the loss function also acts as an additional regularization to the values produced by the network making them more realistic within the DWI context. Another decisive aspect is the simultaneous processing of all b-values. Other approaches utilize fewer averages or directions to accelerate the sequence but opt for more than 1 especially in the case of high b-values which result to a more modest time reduction [13-15]. Our approach essentially allows us to circumvent that

obstacle, as the b-value images are not only useful for the calculation of ADC and providing complementary information, but also essentially feed the network with different direction or signal average information within the same input.

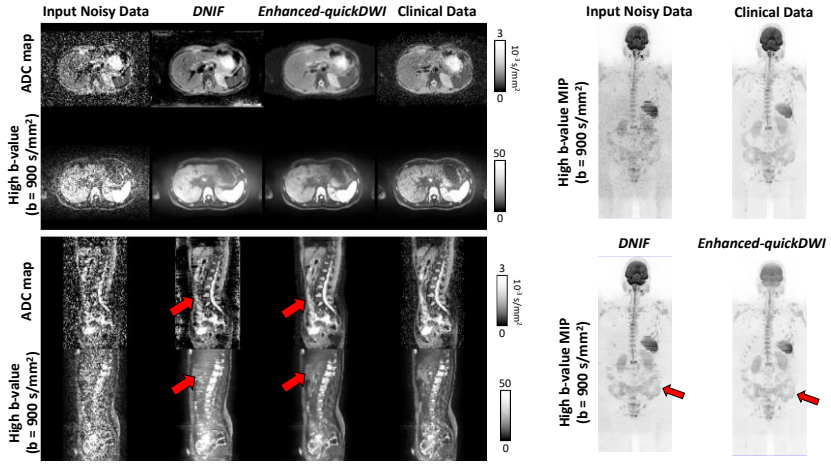

Fig 3. Results from an example whole-body validation patient. High b-value images (b = 900 s/mm^2) are presented along calculated ADC maps for the heavily sub-sampled noisy input data, the ground-truth clinical data, the *Enhanced-quickDWI*-generated and the *DNIF*-generated images [**Left-top**], with sagittal [**Left-bottom**], and Maximum intensity projection reconstructions [**Right**]. Note smoother transition between slices by *Enhanced-quickDWI* (**red arrows**).

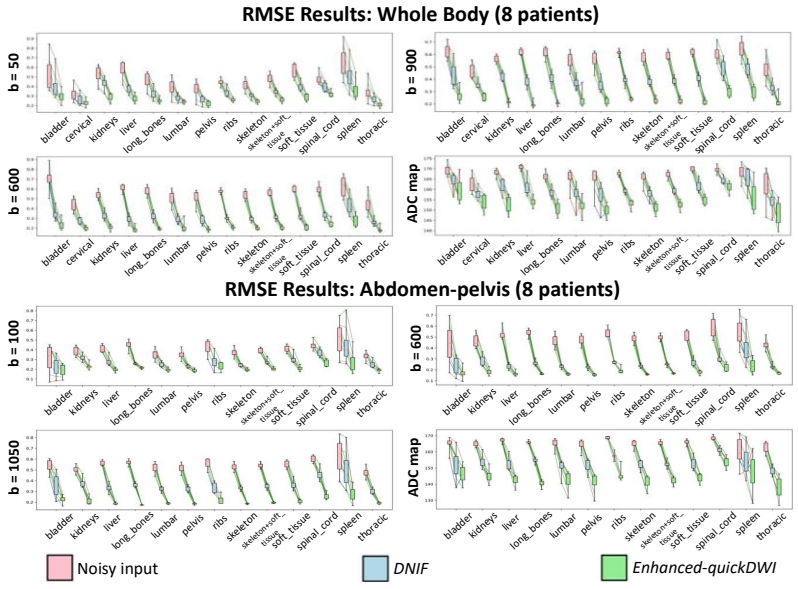

Fig 4. Quantitative comparison of value prediction for all 16 validation patients. Box plots indicate the range of RMSE calculated for each body region (Clinical values vs noisy input, *DNIF*-

generated and *Enhanced-quickDWI* generated respectively). In most cases a negative trend is observed indicating that image quality is better when using the *whole-body* [**top**] and *abdomen-pelvis* [**bottom**] *Enhanced-quickDWI* model versions compared to the *DNIF* model. Additionally, both models are more accurate when compared with the noisy input MDDW data.

4.3 Qualitative Model Evaluation

Enhanced-quickDWI demonstrated non-inferiority to the clinical images in all cases for both *whole-body* and *abdomen-pelvis* versions for the 100 patients of combined validation and test sets. This held true for all b-value images and calculated ADC maps. An overview of the reader overall image quality assessment is presented in **Fig 5**, which also demonstrates that sufficient data were available in this study warrant our non-inferiority claim (95% confidence interval in the mean Likert difference were within the non-inferiority margin). We plan to expand our validation analysis to an additional 200 patients.

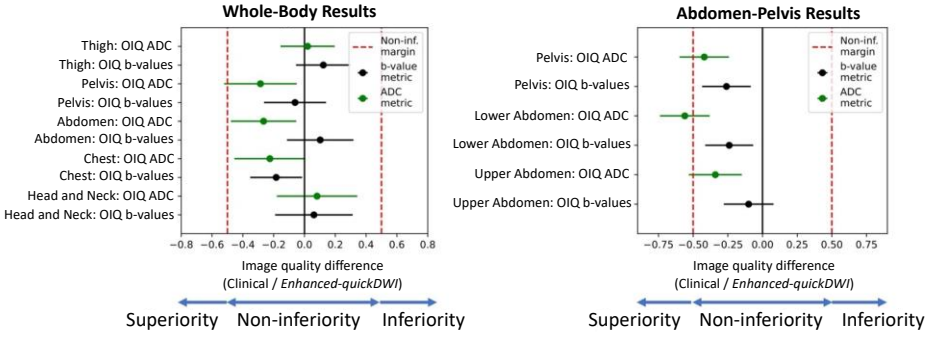

Fig 5. Results from the qualitative reader study (overall image quality). Results for the *whole-body* and *abdomen-pelvis Enhanced-quickDWI* model versions indicate that non-inferiority may be inferred for ADC maps (green bars), and for the for b-value images (black bars) as all cases lie to the left of the rightmost red dashed line (non-inferiority limit).

5 Conclusion

We propose *Enhanced-quickDWI*, a context-aware multi b-value deep learning neural network for the denoising of single average DWI data resulting in images of equivalent clinical quality. Our results indicate that the images produced by our model could be clinically used as a substitute to high-quality clinical images thereby significantly reducing acquisition time. We performed both quantitative and qualitative analysis in 100 patients providing the rigorous validation needed for the clinical translation of such quantitative imaging applications.

Acknowledgments. This study represents independent research funded by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and The Institute of Cancer Research, London, and by the Royal Marsden Cancer Charity and Cancer Research UK (CRUK) National Cancer Imaging Trials Accelerator (NCITA). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work uses data provided by patients and collected by the NHS as part of their care and support.

Disclosure of Interests. Konstantinos Zormpas-Petridis and Matthew D. Blackledge have applied for a patent on a technology broadly related to the one described in this manuscript.

References

1. Padhani AR et al. METastasis reporting and data system for prostate cancer: practical guidelines for acquisition, interpretation, and reporting of whole-body magnetic resonance imaging-based evaluations of multiorgan involvement in advanced prostate cancer. Eur Urol 2017;71(1):81–92
2. Eiber M, Holzapfel K, Ganter C, et al. Whole-body MRI including diffusion weighted imaging (DWI) for patients with recurring prostate cancer: technical feasibility and assessment of lesion conspicuity in DWI. J Magn Reson Imaging 2011;33(5):1160–1170
3. Padhani AR, Koh DM, Collins DJ. Whole-body diffusion-weighted MR imaging in cancer: current status and research directions. Radiology 2011;261(3):700–718
4. Chantry A, Kazmi M, Barrington S, et al. Guidelines for the use of imaging in the management of patients with myeloma. Br J Haematol 2017;178(3):380–393
5. Myeloma diagnosis and management: NICE guideline [NG35] and appendices. National Institute for Health and Care Excellence Web site. https:// www.nice.org.uk/guidance/ng35. Published February 2016. Last updated October 2018. Accessed October 2018
6. Evans R, Taylor S, Janes S, et al. Patient experience and perceived acceptability of whole-body magnetic resonance imaging for staging colorectal and lung cancer compared with current staging scans: a qualitative study. BMJ Open 2017;7(9):e016391
7. Evans RE, Taylor SA, Beare S, et al. Perceived patient burden and acceptability of whole body MRI for staging lung and colorectal cancer; comparison with standard staging investigations. Br J Radiol 2018;91(1086):20170731
8. Nougaret, Stephanie, et al. "Endometrial cancer MRI staging: updated guidelines of the European Society of Urogenital Radiology." European radiology 29 (2019): 792-805.
9. Chernyak, Victoria, et al. "Liver Imaging Reporting and Data System (LI-RADS) version 2018: imaging of hepatocellular carcinoma in at-risk patients." Radiology 289.3 (2018): 816-830.
10. Baltzer, Pascal, et al. "Diffusion-weighted imaging of the breast—a consensus and mission statement from the EUSOBI International Breast Diffusion-Weighted Imaging working group." European radiology 30 (2020): 1436-1450.
11. Turkbey, Baris, et al. "Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2." European urology 76.3 (2019): 340-351.
12. Beets-Tan, Regina GH, et al. "Magnetic resonance imaging for clinical management of rectal cancer: updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting." European radiology 28 (2018): 1465-1475.

13. Wessling, Daniel, et al. "Novel deep-learning-based diffusion weighted imaging sequence in 1.5 T breast MRI." European Journal of Radiology (2023): 110948
14. Kaye, Elena A., et al. "Accelerating prostate diffusion-weighted MRI using a guided denoising convolutional neural network: retrospective feasibility study." Radiology: Artificial Intelligence 2.5 (2020): e200007.
15. Afat, Saif, et al. "Acquisition time reduction of diffusion-weighted liver imaging using deep learning image reconstruction." Diagnostic and Interventional Imaging 104.4 (2023): 178-184.
16. Lehtinen, Jaakko, et al. "Noise2Noise: Learning image restoration without clean data." arXiv preprint arXiv:1803.04189 (2018).
17. Kawamura, Motohide, et al. "Accelerated Acquisition of High-resolution Diffusion-weighted Imaging of the Brain with a Multi-shot Echo-planar Sequence: Deep-learning-based Denoising." Magnetic Resonance in Medical Sciences 20.1 (2021): 99-105.
18. Ran, Maosong, et al. "Denoising of 3D magnetic resonance images using a residual encoder–decoder Wasserstein generative adversarial network." Medical image analysis 55 (2019): 165-180.
19. Hu, Yuxin, et al. "RUN-UP: Accelerated multishot diffusion-weighted MRI reconstruction using an unrolled network with U-Net as priors." Magnetic resonance in medicine 85.2 (2021): 709-720.
20. Aamir, Fariha, et al. "Accelerated diffusion-weighted MR image reconstruction using deep neural networks." Journal of Digital Imaging 36.1 (2023): 276-288.
21. Zormpas-Petridis, Konstantinos, et al. "Accelerating Whole-Body Diffusion-weighted MRI with Deep Learning–based Denoising Image Filters." Radiology: Artificial Intelligence 3.5 (2021): e200279.
22. Candito, Antonio, et al. "Deep learning assisted atlas-based delineation of the skeleton from Whole-Body Diffusion Weighted MRI in patients with malignant bone disease." Biomedical Signal Processing and Control 92 (2024): 106099.