



Copyright © Author(s) - Available online at dirjournal.org.  
Content of this journal is licensed under a Creative Commons  
Attribution-NonCommercial 4.0 International License.

# Deep learning reconstruction for brain diffusion-weighted imaging: efficacy for image quality improvement, apparent diffusion coefficient assessment, and intravoxel incoherent motion evaluation in *in vitro* and *in vivo* studies

Satomu Hanamatsu

Kazuhiro Murayama

Yoshiharu Ohno

Kaori Yamamoto

Masao Yui

Hiroshi Toyama

## PURPOSE

Deep learning reconstruction (DLR) to improve imaging quality has already been introduced, but no studies have evaluated the effect of DLR on diffusion-weighted imaging (DWI) or intravoxel incoherent motion (IVIM) in *in vitro* or *in vivo* studies. The purpose of this study was to determine the effect of DLR for magnetic resonance imaging (MRI) in terms of image quality improvement, apparent diffusion coefficient (ADC) assessment, and IVIM index evaluation on DWI through *in vitro* and *in vivo* studies.

## METHODS

For the *in vitro* study, a phantom recommended by the Quantitative Imaging Biomarkers Alliance was scanned and reconstructed with and without DLR, and 15 patients with brain tumors with normal-appearing gray and white matter examined using IVIM and reconstructed with and without DLR were included in the *in vivo* study. The ADCs of all phantoms for DWI with and without DLR, as well as the coefficient of variation percentage (CV%), and ADCs and IVIM indexes for each participant, were evaluated based on DWI with and without DLR by means of region-of-interest measurements. For the *in vitro* study, using the mean ADCs for all phantoms, a t-test was adopted to compare DWI with and without DLR. For the *in vivo* study, a Wilcoxon signed-rank test was used to compare the CV% between the two types of DWI. In addition, the Wilcoxon signed-rank test was used to compare the ADC, true diffusion coefficient ( $D$ ), pseudodiffusion coefficient ( $D^*$ ), and percentage of water molecules in micro perfusion within 1 voxel ( $f$ ) with and without DLR; the limits of agreement of each parameter were determined through a Bland-Altman analysis.

## RESULTS

The *in vitro* study identified no significant differences between the ADC values for DWI with and without DLR ( $P > 0.05$ ), and the CV% was significantly different for DWI with and without DLR ( $P < 0.05$ ) when  $b$  values  $\geq 250$  s/mm<sup>2</sup> were used. The *in vivo* study revealed that  $D^*$  and  $f$  with and without DLR were significantly different ( $P < 0.001$ ). The limits of agreement of the ADC,  $D$ , and  $D^*$  values for DWI with and without DLR were determined as  $0.00 \pm 0.51 \times 10^{-3}$ ,  $0.00 \pm 0.06 \times 10^{-3}$ , and  $1.13 \pm 4.04 \times 10^{-3}$  mm<sup>2</sup>/s, respectively. The limits of agreement of the  $f$  values for DWI with and without DLR were determined as  $-0.01 \pm 0.07$ .

## CONCLUSION

Deep learning reconstruction for MRI has the potential to significantly improve DWI quality at higher  $b$  values. It has some effect on  $D^*$  and  $f$  values in the IVIM index evaluation, but ADC and  $D$  values are less affected by DLR.

## KEYWORDS

Brain, magnetic resonance imaging, diffusion, intravoxel incoherent motion, deep learning reconstruction

From the Department of Radiology (S.H. ✉ shana@fujita-hu.ac.jp, K.M., Y.O., H.T.), Fujita Health University School of Medicine, Toyoake, Japan; Joint Research Laboratory of Advanced Medicine Imaging (Y.O.), Fujita Health University School of Medicine, Toyoake, Japan; Canon Medical Systems Corporation (K.Y., M.Y.), Otawara, Japan.

Received 21 February 2023; revision requested 05 April 2023; last revision received 29 May 2023; accepted 23 June 2023.



Epub: 09.08.2023

Publication date: 05.09.2023

DOI: 10.4274/dir.2023.232149

You may cite this article as: Hanamatsu S, Murayama K, Ohno Y, Yamamoto K, Yui M, Toyama H. Deep learning reconstruction for brain diffusion-weighted imaging: efficacy for image quality improvement, apparent diffusion coefficient assessment, and intravoxel incoherent motion evaluation in *in vitro* and *in vivo* studies. *Diagn Interv Radiol.* 2023;29(5):664-673.

The clinical application of artificial intelligence is expanding with a variety of targets not only for detection or diagnostics but also for image noise reduction for computed tomography (CT) and magnetic resonance imaging (MRI).<sup>1-10</sup> In recent years, some MRI suppliers have introduced deep learning reconstruction (DLR) for denoising and improving imaging quality, which has been tested for MRI of the central nervous system as well as of the body.<sup>1-10</sup> Moreover, in the late 1980s, Le Bihan et al.<sup>11</sup> developed intravoxel incoherent motion (IVIM), a non-invasive approach that measures perfusion-related parameters using diffusion-weighted imaging (DWI) for MR examinations.<sup>12-22</sup> This method exploits the fact that the signal acquired using a DWI sequence is affected by the incoherent motion of water resulting not only from thermal energy but also blood circulation in the microvasculature.<sup>11-22</sup> To date, however, no studies have evaluated the efficacy of DLR for apparent diffusion coefficient (ADC) evaluation or IVIM index assessments in *in vivo* or *in vitro* studies.

We hypothesized that DLR may affect IVIM index measurements, possibly without influencing ADC measurements, by denoising and improving DWI quality. The purpose of this study was therefore to determine the efficacy of DLR for MRI on image quality improvement, ADC assessment, and IVIM index evaluation in DWI through *in vitro* and *in vivo* studies using a 3 Tesla (T) MR system.

## Methods

### Research ethics standards compliance

The *in vivo* study was a retrospective study and was approved by the Institutional Review Board (IRB) of Fujita Health University, Japan (research registration: HM22-328; IRB-approval number: CI22-647); it is com-

#### Main points

- The *in vitro* study identified no significant differences in apparent diffusion coefficient values for diffusion-weighted imaging (DWI) with and without deep learning reconstruction (DLR) ( $P > 0.05$ ).
- There were significant differences in coefficient of variation percentages between the DWI with and without DLR ( $P < 0.05$ ) when b values of 250, 500, 750, 1000, and 1500 s/mm<sup>2</sup> were used.
- There were significant differences in the pseudodiffusion coefficient and percentage of water molecules in micro perfusion within 1 voxel values between the DWI with and without DLR ( $P < 0.001$ ).

pliant with the Health Insurance Portability and Accountability Act of Japan. Written informed consent was waived for each participant enrolled in this study. This study was also technically and financially supported by the Canon Medical Systems Corporation. Two of the authors are employees of the Canon Medical Systems Corporation (KY and MY) but did not have control over any of the data used in this study.

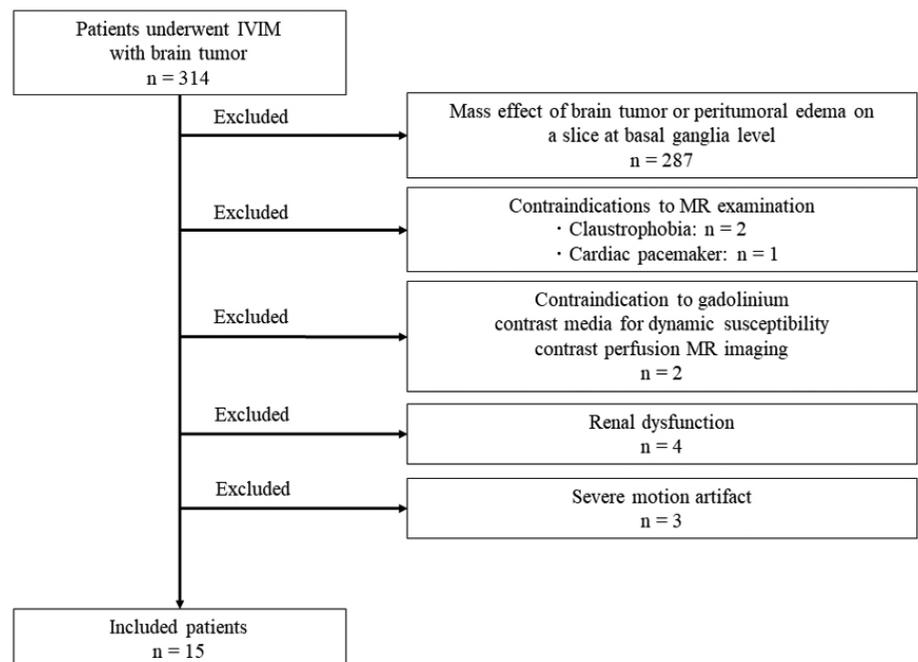
### Quantitative diffusion phantom for *in vitro* study

The *in vitro* study quantitatively assessed an ADC evaluation of DWI obtained with and without using the DLR method. For this study, the quantitative diffusion phantom (High Precision Devices, Boulder, CO, USA), which was developed by the National Institute of Standards and Technology/Quantitative Imaging Biomarker Alliance (QIBA) of the Radiological Society of North America and is commercially available, consisting of 13 vials filled with varying concentrations of polyvinylpyrrolidone (PVP) in an aqueous solution, was used to evaluate ADC measurement accuracy.<sup>23,24</sup> The phantom was specifically designed to quantitatively map the isotropic Gaussian diffusion of water molecules and generate physiologically relevant ADC values.<sup>25</sup> The distribution of PVP concentrations in the phantom was as follows: 0% (vials 1-3), 10% (vials 4 and 5), 20% (vials 6 and 7), 30% (vials 8 and 9), 40% (vials 10 and 11), and 50% (vials: 12 and 13) in an aqueous solution.<sup>24</sup> The vials were stored in an ice-water bath at 0°C to eliminate thermal variability across

scanner locations and timepoints for the ADC measurements.<sup>23</sup>

### Participants in the *in vivo* study

The *in vivo* study involved 314 consecutive patients (146 men, 168 women; mean age: 59.8 years; age range: 18-91 years) who had been diagnosed with a suspected brain tumor at nearby hospitals. The participants visited the outpatient clinic in our department of neurosurgery between March and August 2019, where they were examined using brain MRI with IVIM. The exclusion criteria were 1) mass effect of a brain tumor or peritumoral edema on a slice at the basal-ganglia level, 2) contraindications for MR examination, 3) contraindication for gadolinium (Gd) contrast media because of asthma, 4) renal dysfunction, and 5) severe motion artifact. Of the 314 patients originally included in this study, 299 were excluded because of the mass effect of brain tumor or peritumoral edema on a slice at basal-ganglia level ( $n = 287$ ), contraindications for MR examination because of claustrophobia ( $n = 2$ ) and having a cardiac pacemaker device ( $n = 1$ ), contraindication for Gd contrast media because of asthma ( $n = 2$ ) and renal dysfunction ( $n = 4$ ), and severe motion artifact ( $n = 3$ ). The remaining 15 patients with brain tumors with normal-appearing gray and white matter (8 men, 7 women; mean age: 49.6 years; age range: 31-82 years) were included in this study. The patient selection chart is presented in Figure 1, and details of patient characteristics are summarized in Table 1.



**Figure 1.** Patient selection chart. Patient selection for the final study population. IVIM, intravoxel incoherent motion; MR, magnetic resonance.

## Magnetic resonance examinations

All MR examinations for the *in vitro* and *in vivo* studies were performed using a 3T clinical MR scanner (Vantage Galan 3T/ZGO, Canon Medical Systems Corporation, Otawara, Tochigi, Japan) with a 32-channel phased-array surface coil (32 ch Head SPEED-

ER, Canon Medical Systems). The maximal gradient specifications were 100 mT/m for amplitude and 200 mT/m/msec for slew rate.

## In vitro study

For the *in vitro* study, DWI was acquired in the axial planes using a two-dimension-

al spin-echo (SE)-type echo-planar imaging (EPI) sequence with a parallel imaging technique (SPEEDER, Canon Medical Systems) and the following parameters: repetition time (TR)/echo time (TE), 4500/66 ms; field of view (FOV), 220 × 220 mm; acquisition matrix, 144 × 144; slice thickness, 4 mm; reduction factor (SPEEDER factor), 3; number of acquisition (NAQ), 1; *b* values, 0, 10, 25, 50, 75, 100, 250, 500, 750, 1000, 1500, 2000, and 3000 s/mm<sup>2</sup>. All MRI was then reconstructed with and without the DLR method (Advanced Intelligent Clear-IQ Engine, Canon Medical Systems), which is consistent with other studies,<sup>3,5,10</sup> and operated using an MRI system (version 6 SP0003, Canon Medical Systems).

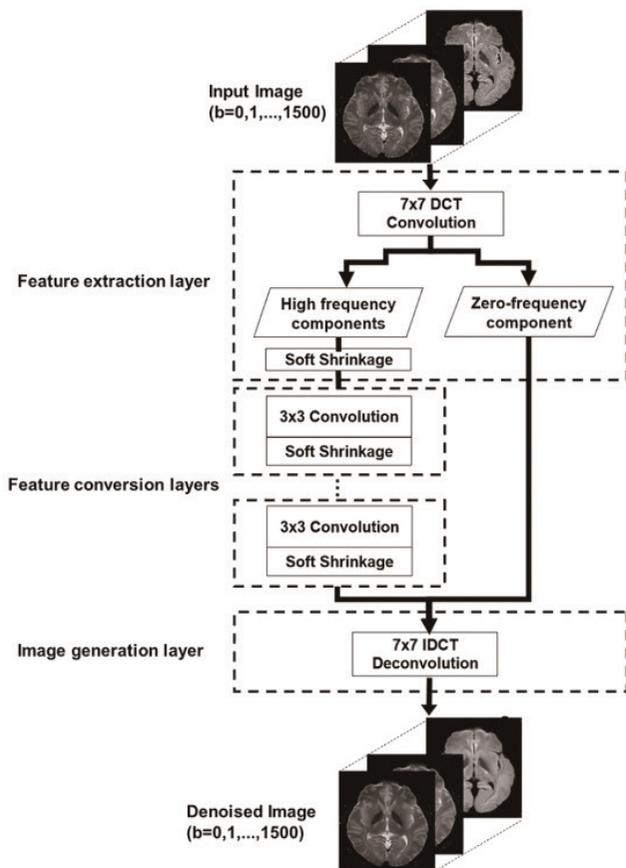
## In vivo study

For the *in vivo* study, DWI was acquired in the axial planes using the SE-EPI sequence with SPEEDER and the following parameters: TR/TE, 4500/72 ms; echo train spacing, 0.9 ms; number of slices, 30; slice thickness, 5 mm; FOV, 220 × 220 mm; acquisition matrix, 160 × 160; NAQ, 1; flip angle, 90/180; SPEEDER factor, 3; *b* values, 0, 5, 10, 20, 30, 50, 75, 100, 250, 500, 750, 1000, and 1500 s/mm<sup>2</sup>. All MRI data were then reconstructed with and without DLR, as in the *in vitro* study.<sup>3,5,10</sup> The examination time including reconstruction time with and without DLR was recorded for each participant.

## Deep learning reconstruction method for brain diffusion-weighted imaging

The DLR method used in this study is based on a convolutional neural network (CNN), and the details have been published in the literature.<sup>3</sup> Figure 2 provides a diagram of the DLR method. A study has proposed CNN denoising using soft shrinkage, which adapts to the amount of noise by introducing a variable threshold of an inactive section, as an activation function;<sup>26</sup> noise-adaptive soft shrinkage is also applied to the neural network in the DLR method.<sup>1</sup> The present study used the same trained DLR network described in the literature.<sup>3</sup> The network was trained and validated using conventional contrast images (T2 weighted, T1 weighted, etc.) of the brain and knees of several human volunteers.<sup>3</sup> The quality of the different contrast images reconstructed using the DLR method was clinically evaluated in several body regions, such as the brain,<sup>3</sup> pelvis,<sup>5</sup> and abdominal arteries.<sup>27</sup> The DLR details, including information on training and validation data sets, are provided in the literature.<sup>3,5,10</sup>

Variables		Values
Sex (count, %)	Male	8 (53.3%)
	Female	7 (46.7%)
Age as mean ± standard deviation (years, range)		49.6 ± 17.6 (28–82)
Diagnosis (count)	Low-grade gliomas (grade 1, 2)	5
	Schwannomas	4
	Meningiomas	2
	Central neurocytoma	1
	Arachnoid cyst	1
	Metastases	1
	Malignant peripheral nerve sheath tumor	1



**Figure 2.** Deep learning reconstruction network. In the first feature extraction layer, input images ( $b = 0, 1, \dots, 1500$ ) are convolved using a  $7 \times 7$  discrete cosine transform (DCT) kernel. After the initial soft shrinkage, 48 high frequency components undergo a  $3 \times 3$  convolution repeatedly and soft shrinkage in the feature conversion layers. Finally, the denoised output images ( $b = 0, 1, \dots, 1500$ ) are generated through the inverse DCT deconvolution of both the bypassed zero-frequency component and output data from the feature conversion layers.

## Image analysis

### In vitro study

First, an ADC map was generated from the DWI for all  $b$  values. Signal intensity data obtained from each voxel on the DWI for all  $b$  values were fitted to a mono-exponential model to calculate the ADC using a built-in Tensor application (System software version 6.0, Canon Medical Systems). The ADC for each phantom was then measured by a neuroradiologist (SH) with 3-years' experience using ImageJ version 1.52p (<https://imagej.nih.gov/ij/>). Five circular regions of interest (ROIs) with a diameter of 10 mm were placed on the center slice and two additional slices, one obtained 1 cm before and the other 1 cm after the center slice, as well as on each phantom, after which the mean ADC value within each phantom was calculated.

### In vivo study

An ADC map was generated for each patient from the DWI reconstructed with and without DLR for  $b$  values of 0 s/mm<sup>2</sup> and others (i.e.,  $b = 5, 10, 20, 30, 50, 75, 100, 250, 500, 750, 1000,$  and  $1500$  s/mm<sup>2</sup>) by means of commercially available software (IVIM) using a mono-exponential model on a Vitrea workstation (version 7.4, Canon Medical Systems). In addition, IVIM parameters were determined using commercially available software (IVIM) on the same workstation and based on the theory described in other studies.<sup>11-15</sup> Based on a bi-exponential model derived from DWI with different  $b$  values, the true diffusion coefficient ( $D$ ), pseudodiffusion coefficient ( $D^*$ ), and percentage of water molecules in micro perfusion within 1 voxel ( $f$ ) were determined using the following previously published formula:<sup>11-15</sup>

$$S(b)/S_0 = f_{ivim} \exp[-b(D^* + D_{blood})] + (1 - f_{ivim}) \exp(-bD_{tissue}) \quad [1]$$

where  $S(b)$  is the signal intensity for each  $b$  value and  $S_0$  is the signal intensity at a  $b$  value of zero.

To quantitatively evaluate the influence of DLR on the DWI obtained at each  $b$  value for all patients, ROIs were measured using the Vitrea workstation. A center line was first placed manually on each slice. Subsequently, ROIs with a diameter of 10 mm were automatically placed on the normal cortex and white matter of a slice at basal-ganglia level obtained from each brain hemisphere (total of 10 ROIs = 5 ROIs × right and left hemisphere) to determine the mean signal intensity and standard deviation for ROIs on each slice. An example of ROI placements is provided in Supplementary Figure 1. For a quantitative image quality comparison of DWI obtained for each  $b$  value and reconstructed with and without DLR, the coefficient of variation percentages (CVs%) of the DWI for each  $b$  value with and without DLR were calculated by means of the following previously published formula:<sup>10,28-30</sup>

$$CV\% = \text{standard deviation within ROI} / \text{mean signal intensity within ROI} \times 100\% \quad [2]$$

To determine the influence of DLR on all DWI parameter evaluations, ADC,  $D$ ,  $D^*$ , and  $f$  values from the automatically copied ROIs were measured on the same slices on ADC,  $D$ ,  $D^*$ , and  $f$  maps for each patient.

### Statistical analysis

#### In vitro study

To determine the influence of DLR on ADC measurements, mean ADCs measured within each phantom on DWIs with and without DLR were correlated with standard references and with each other using Spearman's correlation coefficient. To determine the effect of DLR on ADC evaluation, mean ADCs for each phantom were then compared in DWI with and without DLR by means of a t-test. A Bland-Altman analysis was then performed to determine the limits of agreement between the DWI with and without DLR.<sup>31,32</sup>

#### In vivo study

To compare the IVIM examination time, including reconstruction with and without DLR, the mean IVIM examination time with DLR and that without DLR was compared using Wilcoxon's signed-rank test.

To determine the utility of DLR for image quality improvement on the DWI at each  $b$  value, the Wilcoxon signed-rank test was used to compare CV% in the DWI with and without DLR. To determine the influence of DLR on ADC and IVIM index evaluations, ADC,  $D^*$ ,  $D$ , and  $f$  values were compared in the DWI with and without DLR by means of the Wilcoxon signed-rank test. Finally, the Bland-Altman analysis was used to evaluate the limits of agreement of the ADC and each IVIM index for DWI with and without DLR.<sup>31,32</sup>

## Results

### In vitro study

The correlations of ADC values for DWI with and without DLR with nominal ADC values as standard references are presented in Figure 3. The ADC values for DWI with and without DLR significantly and strongly correlated with those for standard references (with DLR:  $r = 0.99, P < 0.0001$ ; without DLR:  $r = 0.99, P < 0.0001$ ) and between DWI with and without DLR ( $r = 0.99, P < 0.0001$ ).

Table 2 provides a comparison of ADC values for DWI with and without DLR for the *in vitro* study. The ADC values for DWI with and without DLR in the *in vitro* study were not significantly different ( $P > 0.05$ ).

The results of the Bland-Altman analysis are presented in Figure 4. The limit of agreement between the ADC values for DWI with and without DLR and standard reference values was determined as  $-0.03 \pm 0.04 \times 10^{-3}$  mm<sup>2</sup>/s. In addition, the limit of agreement of ADC values for DWI with and without DLR was determined as  $-0.00 \pm 0.01 \times 10^{-3}$  mm<sup>2</sup>/s.

**Table 2.** Comparison of ADC values within each phantom for DWI with and without DLR

Concentration of phantom (%)	ADC with DLR ( $\times 10^{-3}$ mm <sup>2</sup> /s)	ADC without DLR ( $\times 10^{-3}$ mm <sup>2</sup> /s)	Upper and lower limits of agreement ( $\times 10^{-3}$ mm <sup>2</sup> /s)	$P$ value
0	1.18 ± 0.01	1.17 ± 0.01	0.00 ± 0.02	0.09
10	0.89 ± 0.01	0.88 ± 0.01	0.00 ± 0.01	0.43
20	0.64 ± 0.01	0.63 ± 0.01	0.00 ± 0.01	0.37
30	0.42 ± 0.01	0.42 ± 0.01	0.00 ± 0.01	0.99
40	0.26 ± 0.01	0.26 ± 0.01	0.00 ± 0.01	0.83
50	0.14 ± 0.01	0.14 ± 0.01	0.00 ± 0.01	0.75

ADC, apparent diffusion coefficient; DLR, deep learning reconstruction; DWI, diffusion-weighted imaging.

## In vivo study

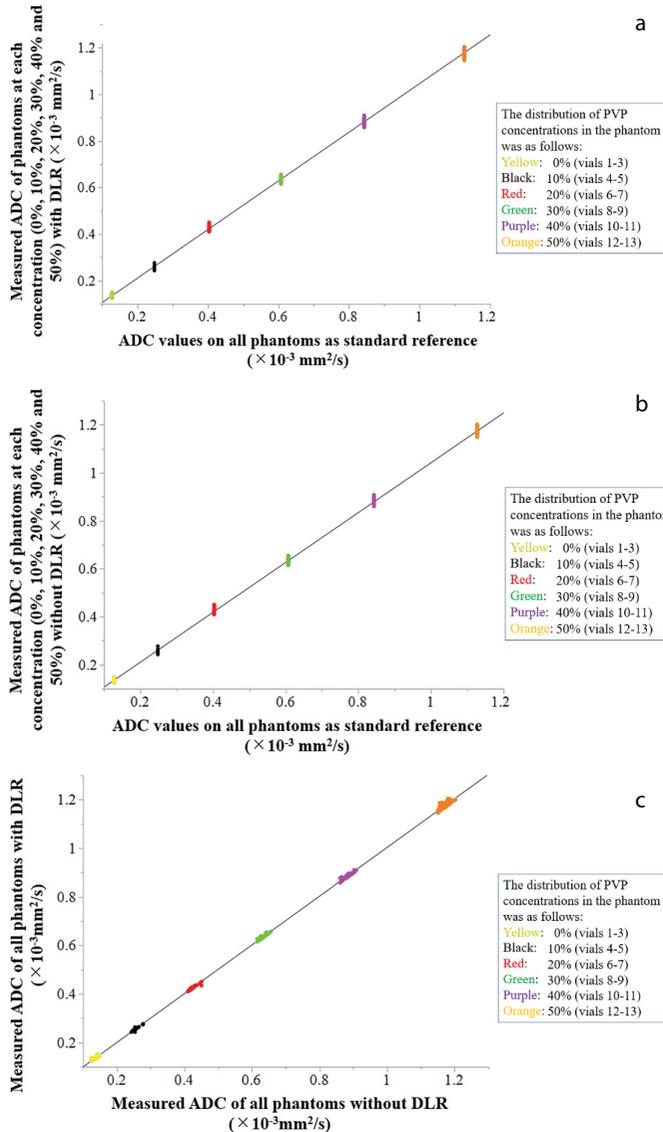
An example case is presented in Figure 5.

The mean examination time, including reconstruction time, of DLR (256 ± 4 s, range: 247–261 s) was significantly different from that without DLR (208 ± 4 s, range: 199–213 s,  $P < 0.001$ ).

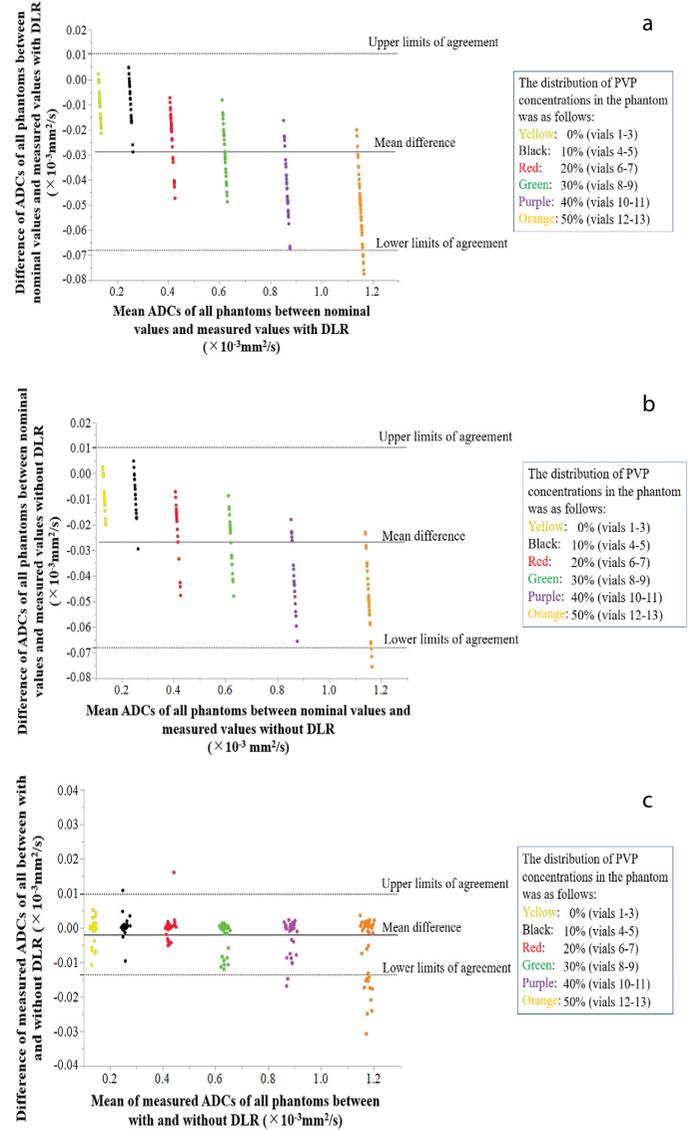
The results of the comparison of CV% of each phantom for DWI with and without DLR are summarized in Table 3. The CV% was significantly different for DWI with and without DLR ( $P < 0.05$ ) when  $b$  values equal to or higher than 250 s/mm<sup>2</sup> were used.

The results of the comparisons of ADC,  $D$ ,  $D^*$ , and  $f$  for DWI with and without DLR are presented in Table 4, and the results for the

limits of agreements for ADC,  $D$ ,  $D^*$ , and  $f$  for DWI with and without DLR are illustrated in Figure 6;  $D^*$  and  $f$  were significantly different for DWI with and without DLR ( $P < 0.001$ ). The limit of agreement of ADC values for DWI with and without DLR was determined as  $0.00 \pm 0.51 \times 10^{-3}$  mm<sup>2</sup>/s, the limit of agreement of  $D$  values for DWI with and without DLR was determined as  $0.00 \pm 0.06 \times 10^{-3}$  mm<sup>2</sup>/s, the



**Figure 3.** Correlations between apparent diffusion coefficient (ADC) values for diffusion-weighted imaging (DWI) with and without deep learning reconstruction (DLR) and standard reference values and between DWI with and without DLR [yellow: polyvinylpyrrolidone (PVP) concentration within the phantom of 0%, black: PVP concentration within the phantom of 10%, red: PVP concentration within the phantom of 20%, green: PVP concentration within the phantom of 30%, purple: PVP concentration within the phantom of 40%, and orange: PVP concentration within the phantom of 50%]. **(a)** ADC values for DWI with DLR and the standard reference values exhibit significant and strong correlations ( $r = 0.99$ ,  $P < 0.0001$ ). **(b)** ADC values for DWI without DLR and the standard reference values reveal significant and strong correlations ( $r = 0.99$ ,  $P < 0.0001$ ). **(c)** ADC values for DWI with and without DLR demonstrate significant and strong correlations ( $r = 0.99$ ,  $P < 0.0001$ ).



**Figure 4.** Mean differences and the limits of agreement of apparent diffusion coefficient (ADC) values for diffusion-weighted imaging (DWI) with and without deep learning reconstruction (DLR) and standard reference values and between DWI with and without DLR [yellow: polyvinylpyrrolidone (PVP) concentration within the phantom of 0%, black: PVP concentration within the phantom of 10%, red: PVP concentration within the phantom of 20%, green: PVP concentration within the phantom of 30%, purple: PVP concentration within the phantom of 40%, and orange: PVP concentration within the phantom of 50%]. Mean differences are indicated by solid lines; upper and lower limits of agreement are indicated as dotted lines at the top and bottom, whereas plotted dots indicate data points. **(a)** Limits of agreement of ADC values for DWI with DLR and standard reference values were assessed as between  $-0.07 \times 10^{-3}$  and  $-0.01 \times 10^{-3}$  mm<sup>2</sup>/s. **(b)** Limits of agreement of ADC values for DWI without DLR and standard reference values were assessed as between  $-0.07 \times 10^{-3}$  and  $-0.01 \times 10^{-3}$  mm<sup>2</sup>/s. **(c)** Limits of agreement of ADC values for DWI with and without DLR were assessed as between  $-0.01 \times 10^{-3}$  and  $0.01 \times 10^{-3}$  mm<sup>2</sup>/s.

limit of agreement of  $D^*$  values for DWI with and without DLR was determined as  $1.13 \pm 4.04 \times 10^{-3} \text{ mm}^2/\text{s}$ , and the limit of agreement of  $f$  values for DWI with and without DLR was determined as  $-0.01 \pm 0.07$ .

## Discussion

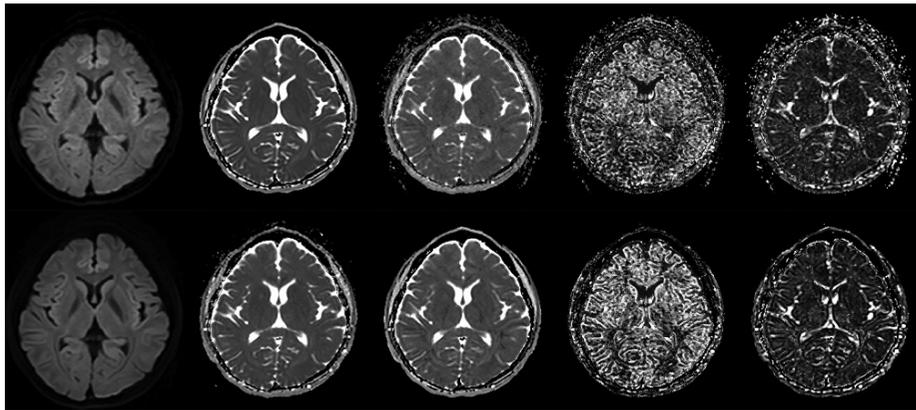
Our study used *in vitro* and *in vivo* studies to determine the effect of DLR on ADC or IVIM parameter evaluations. This study was

also the first to demonstrate that DLR had no effect on ADC evaluations in a QIBA-recommended diffusion phantom. In addition, the *in vivo* study was the first to determine that DLR had little effect on ADC and  $D$  evaluations using brain DWI or on brain IVIM examinations when  $b$  values were set at equal to or less than  $1500 \text{ s/mm}^2$ . However,  $D^*$  and  $f$  values were significantly different for IVIM with and without DLR when IVIM examinations applied the same  $b$  values. To the best of our knowledge, no other study has assessed the influence of DLR on ADC and IVIM parameter evaluations in *in vitro* or *in vivo* studies.

When the examination time, including reconstruction time, for IVIM with and without DLR was compared, IVIM with DLR exhibited a significantly longer mean examination time than that for IVIM without DLR; however, the acquisition time for IVIM was the same. Therefore, the prolongation of the mean examination time in IVIM with DLR was considered to mainly result from the significantly longer reconstruction time when compared with that of IVIM without DLR.

Our *in vitro* study demonstrated that correlations between ADC values assessed through DWI with and without DLR were significant and strong, whereas the differences between them were non-significant. Moreover, the limit of agreement for DWI with or without DLR compared with that for standard reference values can be considered negligible and small enough for clinical purposes. Therefore, DLR was determined to have little or no influence on ADC evaluations in this setting.

As for the *in vivo* study, we determined that DLR could significantly improve the CV% of DWI for  $b$  values set at equal to or more than 250, 500, 750, 1000, and  $1500 \text{ s/mm}^2$ . These results suggest that DWI at  $b$  values equal to or more than  $250 \text{ s/mm}^2$  may feature an increase in image noise level and be decreased by it. When considering the equation for CV%, the standard deviation within ROIs may have predominantly Gaussian noise but additionally include spatial variation resulting from the anatomy. More-



**Figure 5.** Diffusion-weighted imaging (DWI) ( $b = 1000 \text{ s/mm}^2$ ) and apparent diffusion coefficient (ADC),  $D$ ,  $D^*$ , and  $f$  maps reconstructed with and without deep learning reconstruction (DLR). Upper row, L to R: DLR. Lower row, L to R: DWI at  $b = 1000 \text{ s/mm}^2$  and ADC,  $D$ ,  $D^*$ , and  $f$  maps reconstructed with DLR. When DLR was applied, the contrast between gray and white matter was improved because of the image noise reduction on the DWI,  $D^*$ , and  $f$  maps. ADC and  $D$  maps exhibit only a slight reduction in noise.  $D$ , true diffusion coefficient;  $D^*$ , pseudodiffusion coefficient,  $f$ , percentage of water molecules in micro perfusion within 1 voxel.

**Table 3.** Comparison of the CV% for diffusion-weighted imaging with and without DLR at various  $b$  values for the *in vivo* study

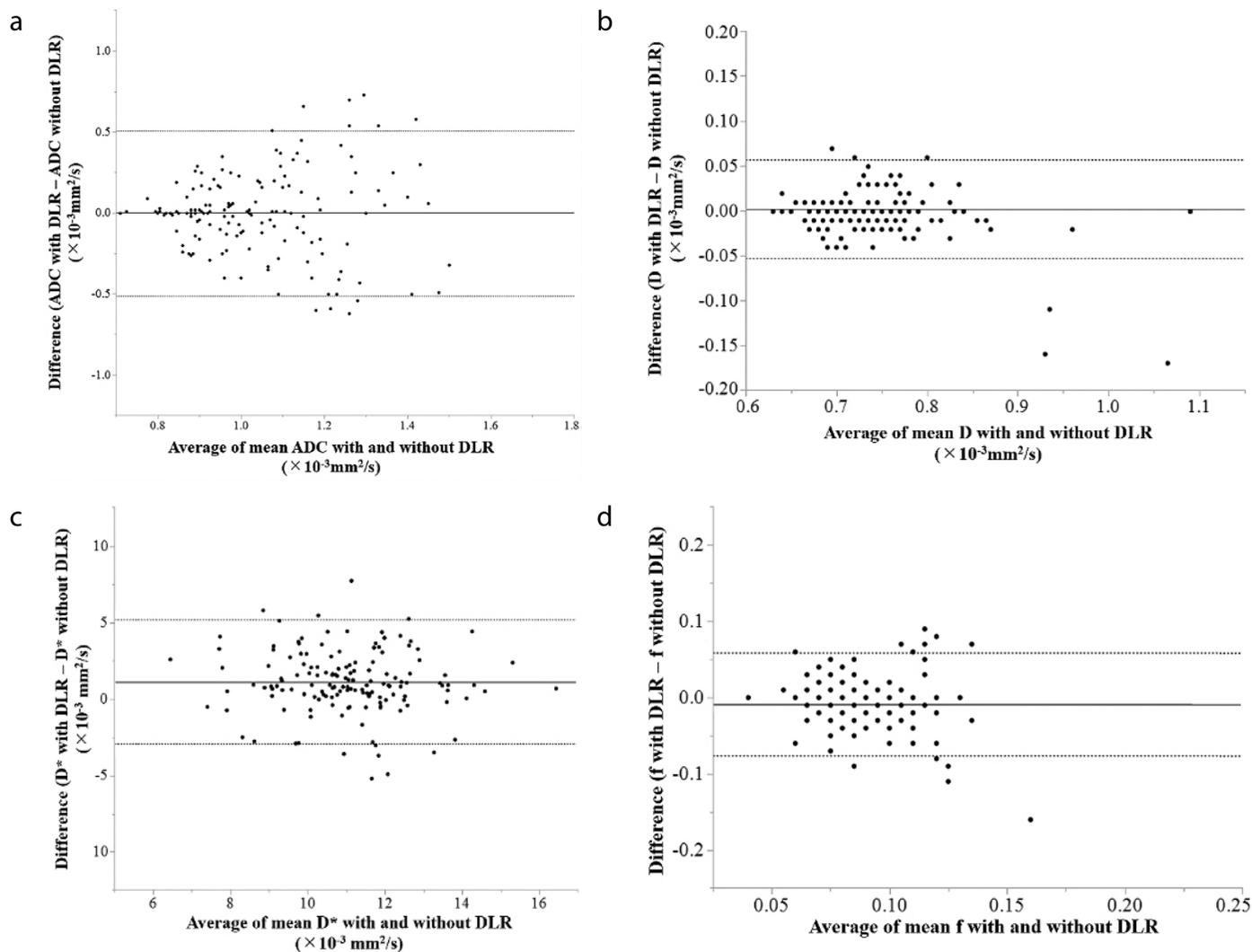
$b$ ( $\text{s/mm}^2$ )	CV% (mean $\pm$ standard deviation)		$P$ value
	With DLR	Without DLR	
0	$6.7 \pm 2.2$	$6.7 \pm 1.9$	0.34
5	$6.2 \pm 2.2$	$6.2 \pm 2.0$	0.37
10	$6.1 \pm 2.2$	$6.1 \pm 2.0$	0.95
20	$6.2 \pm 2.1$	$6.1 \pm 1.9$	0.97
30	$6.1 \pm 2.2$	$6.1 \pm 2.0$	0.84
50	$6.0 \pm 2.2$	$5.9 \pm 2.0$	0.36
75	$5.9 \pm 2.0$	$5.9 \pm 1.8$	0.10
100	$5.8 \pm 2.0$	$5.8 \pm 1.8$	0.09
250	$5.5 \pm 1.6$	$5.6 \pm 1.5$	<0.001
500	$5.3 \pm 1.5$	$5.4 \pm 1.4$	<0.001
750	$5.3 \pm 1.4$	$5.5 \pm 1.4$	<0.001
1000	$5.4 \pm 1.5$	$5.7 \pm 1.5$	<0.001
1500	$6.7 \pm 1.7$	$7.1 \pm 1.7$	<0.001

CV%, coefficient of variation percentage; DLR, deep learning reconstruction; DWI, diffusion-weighted imaging.

**Table 4.** Comparison of intravoxel incoherent motion measurements for DWI with and without DLR for the *in vivo* study

	With DLR (mean $\pm$ standard deviation)	Without DLR (mean $\pm$ standard deviation)	$P$ value
ADC ( $\times 10^{-3} \text{ mm}^2/\text{s}$ )	$1.04 \pm 0.22$	$1.04 \pm 0.21$	0.66
$D$ ( $\times 10^{-3} \text{ mm}^2/\text{s}$ )	$0.75 \pm 0.08$	$0.74 \pm 0.07$	0.30
$D^*$ ( $\times 10^{-3} \text{ mm}^2/\text{s}$ )	$11.62 \pm 1.84$	$10.49 \pm 1.98$	<0.001
$f$	$0.08 \pm 0.02$	$0.09 \pm 0.03$	<0.001

DWI, diffusion-weighted imaging; DLR, deep learning reconstruction;  $D$ , true diffusion coefficient;  $D^*$ , pseudodiffusion coefficient;  $f$ , percentage of water molecules in micro perfusion within 1 voxel.



**Figure 6.** Limits of agreement of apparent diffusion coefficient (ADC),  $D$ ,  $D^*$ , and  $f$  values between diffusion-weighted imaging (DWI) and intravoxel incoherent motion (IVIM) with and without deep learning reconstruction (DLR). Mean difference is denoted by a solid line; upper and lower limits of agreement are denoted by dashed lines at the top and bottom; circles denote data points. **(a)** Mean difference and the limits of agreement of ADC values for DWI with and without DLR. The limits of agreement are determined as  $0.00 \pm 0.51 \times 10^{-3} \text{mm}^2/\text{s}$ . **(b)** Mean difference and the limits of agreement of  $D$  values for IVIM with and without DLR. The limits of agreement are determined as  $0.00 \pm 0.06 \times 10^{-3} \text{mm}^2/\text{s}$ . **(c)** Mean difference and the limits of agreement of  $D^*$  values for IVIM with and without DLR. The limits of agreement are determined as  $1.13 \pm 4.04 \times 10^{-3} \text{mm}^2/\text{s}$ . **(d)** Mean difference and the limits of agreement of  $f$  values for IVIM with and without DLR. The limits of agreement are determined as  $-0.01 \pm 0.07$ .  $D$ , true diffusion coefficient;  $D^*$ , pseudodiffusion coefficient,  $f$ , percentage of water molecules in micro perfusion within 1 voxel.

over, DWI with higher  $b$  values tends to provide lower signal intensity in brain tissues than that of Gaussian noise; therefore, DLR can improve CV% more effectively when  $b$  values are higher. Thus, DLR is a viable choice for improving DWI with higher  $b$  values, such as  $\geq 250 \text{ s/mm}^2$ , which is often used in routine clinical practice. For lower  $b$  values, such as  $< 250 \text{ s/mm}^2$ , it is widely known that the signal of fluid components, such as cerebrospinal fluid and blood, remains. These components could provide additional spatial variation within ROIs, where reconstruction parameters with and without DLR affect the spatial variation of DWI differently. This situation may impact the statistical significance of the CV% with and without DLR. Our results therefore indicate that DLR should be

used for obtaining DWI at  $b$  values equal to or more than  $250 \text{ s/mm}^2$  to improve image quality in routine clinical practice. Moreover, DLR for denoising MRI has been tested not only for image quality improvement but also for a reduction in acquisition time using compressed sensing or other k-space data acquisition methods for various clinical aims on DWI as well as other MR sequences since 2021.<sup>5,8,10,33-38</sup> Therefore, it would be better for us to clinically apply DLR not only for denoising but also to reduce the examination time when using other techniques, although this study did not apply any techniques to reduce acquisition time.

A comparison of ADC and IVIM indexes for brain DWI with and without DLR in the *in*

*vivo* study revealed that the ADC of DWI for  $b$  values equal to or less than  $1500 \text{ s/mm}^2$  were not significantly different. This result is compatible with that for our *in vitro* study. Moreover, we identified no significant difference in  $D$  for DWI with and without DLR when routine  $b$  values of less than  $1500 \text{ s/mm}^2$  were used. However,  $D^*$  and  $f$  significantly influenced DLR when subjected to the same IVIM examination. The reasons for these results can be easily surmised when the similarity of the mechanisms underlying the models for those previously described are considered.<sup>11-22</sup> Our results for the *in vitro* and *in vivo* studies demonstrate that ADC and  $D$  measurements for DWI with  $b$  values equal to or less than  $1500 \text{ s/mm}^2$  can be assumed to have no effect on DLR in this setting. Howev-

er,  $b$  values of more than 1500 s/mm<sup>2</sup> should be used carefully because these values might have some effect on the DLR results when considering the results of DWI for detecting prostate cancer, for which  $b$  values equal to or more than 3000 s/mm<sup>2</sup> and DLR are used.<sup>10</sup>

In contrast to the ADC or  $D$  measurements obtained from DWI and IVIM examinations with and without DLR, we determined that  $D^*$  and  $f$  were significantly affected by DLR, even though there were no significant differences in the CV% of DWI with and without DLR at  $b$  values lower than 250 s/mm<sup>2</sup>. In this study, DWI with and without DLR was generated from the same DWI data obtained from the same sequence and reconstructed with and without DLR. Moreover, all IVIM indexes were measured by means of a commercially available IVIM model. These facts and findings lead us to consider that the differences in  $D^*$  and  $f$  in DWI with and without DLR might be the result of some interaction in the *in vivo* study between signal intensity and image noise within each voxel. Therefore, DLR may be useful for improving the quality of DWI as well as each DWI in the IVIM examinations and have limited influence on quantitative ADC and IVIM parameter evaluations in routine clinical practice.

This study has several limitations. First, we did not have an IVIM phantom or perform animal studies for IVIM in the *in vitro* study. Moreover, the scan parameters for the *in vitro* and *in vivo* studies were not fully matched because of different quantitative DWI index evaluations, and this study is a retrospective study, with no healthy volunteers included. Second, we applied commercially available software, using a mono-exponential model, for IVIM index calculations and assessed the influence of the reconstruction method on these calculations; other models, such as stretched exponential or tri-exponential models, were not tested in this study. Moreover, no comparison between DLR and non-DLR methods for denoising DWI or IVIM images was used, and no standard reference was determined, with only the differences in the ADC or each IVIM index between DWI and IVIM with and without DLR evaluated. To the best of our knowledge, no commercially available MRI phantom exists locally that contains multiple diffusion and circulation compartments and is suitable for IVIM quantification. This type of standardized phantom could be useful for clinicians to validate new advanced techniques in acquisition, reconstruction, and post-processing; therefore, we are now planning to study and develop this type of phantom in the near future. Further

investigations are also warranted to determine the effect of DLR on IVIM evaluations using different models for *in vitro* and *in vivo* studies. Third, the study population was too small to allow for evaluations of patients with a variety of brain diseases; the tumor types of the 15 patients involved in our study were highly heterogeneous, which may affect the study results. Further investigations are therefore warranted to determine the influence of DLR on IVIM parameters as well as on clinical outcomes. Fourth, the  $b$  values used in this study were equal to or less than 1500 s/mm<sup>2</sup> even though  $b$  values of more than 1500 s/mm<sup>2</sup> are currently and frequently used for brain DWI examinations for various purposes.<sup>10,39-41</sup> Further investigations are therefore warranted that use DLR for brain DWI examinations with  $b$  values higher than 1500 s/mm<sup>2</sup>. Fifth, no comparisons were made in this study of the  $D^*$  and  $f$  values of the DWI with and without DLR and perfusion parameters from other perfusion MR and CT techniques and nuclear medicine studies. Sixth, this study used DLR provided by a single supplier for IVIM calculations in the *in vivo* study; however, the clinical relevance of IVIM examinations is currently evaluated primarily for academic purposes rather than clinical aims. Finally, the IVIM sequence and software used here have not yet been standardized. Multi-center studies using DLR and IVIM software provided by different suppliers are thus warranted for standardization and the determination of clinical relevance for the brain and other organs. Large prospective cohort studies using a variety of MR scanners, DLR algorithms, and IVIM software provided by different suppliers are also warranted, and we will plan these studies to address these issues in the near future.

In conclusion, DLR for MRI has the potential to significantly improve the quality of DWI with higher  $b$  values. It also has some effect on  $D^*$  and  $f$  values for IVIM examination, whereas ADC and  $D$  values are less affected by DLR.

#### Conflict of interest disclosure

The authors declared no conflicts of interest.

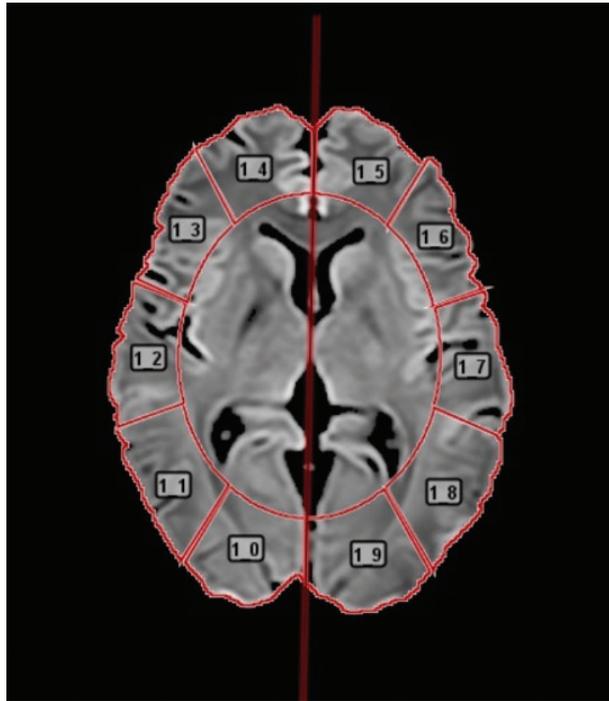
#### Funding

This study was technically and financially supported by the Canon Medical Systems Corporation. Drs Ohno, Murayama, and Toyama had or have research grants from Canon Medical Systems.

## References

1. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. *Jpn J Radiol.* 2019;37(1):73-80. [CrossRef]
2. Higaki T, Nakamura Y, Zhou J, et al. deep learning reconstruction at CT: phantom study of the image characteristics. *Acad Radiol.* 2020;27(1):82-87. [CrossRef]
3. Kidoh M, Shinoda K, Kitajima M, et al. Deep Learning Based Noise Reduction for Brain MR imaging: tests on phantoms and healthy volunteers. *Magn Reson Med Sci.* 2020;19(3):195-206. [CrossRef]
4. Gavazzi S, van den Berg CAT, Savenije MHF, et al. Deep learning-based reconstruction of *in vivo* pelvis conductivity with a 3D patch-based convolutional neural network trained on simulated MR data. *Magn Reson Med.* 2020;84(5):2772-2787. [CrossRef]
5. Ueda T, Ohno Y, Yamamoto K, et al. Compressed sensing and deep learning reconstruction for women's pelvic MRI denoising: utility for improving image quality and examination time in routine clinical practice. *Eur J Radiol.* 2021;134:109430. [CrossRef]
6. Lin DJ, Johnson PM, Knoll F, Lui YW. Artificial intelligence for MR image reconstruction: an overview for clinicians. *J Magn Reson Imaging.* 2021;53(4):1015-1028. [CrossRef]
7. Matsukiyo R, Ohno Y, Matsuyama T, et al. Deep learning-based and hybrid-type iterative reconstructions for CT: comparison of capability for quantitative and qualitative image quality improvements and small vessel evaluation at dynamic CE-abdominal CT with ultra-high and standard resolutions. *Jpn J Radiol.* 2021;39(2):186-197. [CrossRef]
8. Sagawa H, Fushimi Y, Nakajima S, et al. Deep learning-based noise reduction for fast volume diffusion tensor imaging: assessing the noise reduction effect and reliability of diffusion metrics. *Magn Reson Med Sci.* 2021;20(4):450-456. [CrossRef]
9. Johnson PM, Tong A, Donthireddy A, et al. Deep learning reconstruction enables highly accelerated biparametric MR imaging of the prostate. *J Magn Reson Imaging.* 2022;56(1):184-195. [CrossRef]
10. Ueda T, Ohno Y, Yamamoto K, et al. Deep learning reconstruction of diffusion-weighted MRI improves image quality for prostatic imaging. *Radiology.* 2022;303(2):373-381. [CrossRef]
11. Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology.* 1986;161(2):401-407. [CrossRef]
12. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel

- incoherent motion MR imaging. *Radiology*. 1988;168(2):497-505. [\[CrossRef\]](#)
13. Turner R, Le Bihan D, Maier J, Vavrek R, Hedges LK, Pekar J. Echo-planar imaging of intravoxel incoherent motion. *Radiology*. 1990;177(2):407-414. [\[CrossRef\]](#)
  14. Federau C. Intravoxel incoherent motion MRI as a means to measure in vivo perfusion: a review of the evidence. *NMR Biomed*. 2017;30(11). [\[CrossRef\]](#)
  15. Le Bihan D. What can we see with IVIM MRI? *Neuroimage*. 2019;187:56-67. [\[CrossRef\]](#)
  16. Iima M, Honda M, Sigmund EE, Ohno Kishimoto A, Kataoka M, Togashi K. Diffusion MRI of the breast: current status and future directions. *J Magn Reson Imaging*. 2020;52(1):70-90. [\[CrossRef\]](#)
  17. Zhang JL, Lee VS. Renal perfusion imaging by MRI. *J Magn Reson Imaging*. 2020;52(2):369-379. [\[CrossRef\]](#)
  18. Wang YXJ, Huang H, Zheng CJ, Xiao BH, Chevallier O, Wang W. Diffusion-weighted MRI of the liver: challenges and some solutions for the quantification of apparent diffusion coefficient and intravoxel incoherent motion. *Am J Nucl Med Mol Imaging*. 2021;11(2):107-142. [\[CrossRef\]](#)
  19. Federau C. Measuring perfusion: intravoxel incoherent motion mr imaging. *Magn Reson Imaging Clin N Am*. 2021;29(2):233-242. [\[CrossRef\]](#)
  20. Ma W, Mao J, Wang T, Huang Y, Zhao ZH. Distinguishing between benign and malignant breast lesions using diffusion weighted imaging and intravoxel incoherent motion: a systematic review and meta-analysis. *Eur J Radiol*. 2021;141:109809. [\[CrossRef\]](#)
  21. Wang DJJ, Le Bihan D, Krishnamurthy R, Smith M, Ho ML. Noncontrast pediatric brain perfusion: arterial spin labeling and intravoxel incoherent motion. *Magn Reson Imaging Clin N Am*. 2021;29(4):493-513. [\[CrossRef\]](#)
  22. Englund EK, Reiter DA, Shahidi B, Sigmund EE. Intravoxel incoherent motion magnetic resonance imaging in skeletal muscle: review and future directions. *J Magn Reson Imaging*. 2022;55(4):988-1012. [\[CrossRef\]](#)
  23. Yung JP, Ding Y, Hwang KP, et al. Quantitative Evaluation of apparent diffusion coefficient in a large multi-unit institution using the QIBA diffusion phantom. *medRxiv*. 2020. [\[CrossRef\]](#)
  24. Shukla-Dave A, Obuchowski NA, Chenevert TL, et al. Quantitative Imaging Biomarkers Alliance (QIBA) recommendations for improved precision of DWI and DCE-MRI derived biomarkers in multicenter oncology trials. *J Magn Reson Imaging*. 2019;49(7):e101-e121. [\[CrossRef\]](#)
  25. Federau C, O'Brien K, Meuli R, Hagmann P, Maeder P. Measuring brain perfusion with intravoxel incoherent motion (IVIM): initial clinical experience. *J Magn Reson Imaging*. 2014;39(3):624-632. [\[CrossRef\]](#)
  26. Isogawa K, Ida T, Shiodera T, Takeguchi T. Deep Shrinkage convolutional neural network for adaptive noise reduction. *IEEE Signal Processing Letters*. 2018;25(2):224-228. [\[CrossRef\]](#)
  27. Maubon AJ, Ferru JM, Berger V, et al. Effect of field strength on MR images: comparison of the same subject at 0.5, 1.0, and 1.5 T. *Radiographics*. 1999;19(4):1057-1067. [\[CrossRef\]](#)
  28. Voroney JP, Brock KK, Eccles C, Haider M, Dawson LA. Prospective comparison of computed tomography and magnetic resonance imaging for liver cancer delineation using deformable image registration. *Int J Radiat Oncol Biol Phys*. 2006;66(3):780-791. [\[CrossRef\]](#)
  29. Zierhut ML, Ozturk-Isik E, Chen AP, Park I, Vigneron DB, Nelson SJ. (1)H spectroscopic imaging of human brain at 3 Tesla: comparison of fast three-dimensional magnetic resonance spectroscopic imaging techniques. *J Magn Reson Imaging*. 2009;30(3):473-480. [\[CrossRef\]](#)
  30. Spinks TJ, Karia D, Leach MO, Flux G. Quantitative PET and SPECT performance characteristics of the Albira Trimodal pre-clinical tomograph. *Phys Med Biol*. 2014;59(3):715-731. [\[CrossRef\]](#)
  31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307-310. [\[CrossRef\]](#)
  32. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346(8982):1085-1087. [\[CrossRef\]](#)
  33. Park JC, Park KJ, Park MY, Kim MH, Kim JK. Fast T2-weighted imaging with deep learning-based reconstruction: evaluation of image quality and diagnostic performance in patients undergoing radical prostatectomy. *J Magn Reson Imaging*. 2022;55(6):1735-1744. [\[CrossRef\]](#)
  34. Bae SH, Hwang J, Hong SS, et al. Clinical feasibility of accelerated diffusion weighted imaging of the abdomen with deep learning reconstruction: Comparison with conventional diffusion weighted imaging. *Eur J Radiol*. 2022;154:110428. [\[CrossRef\]](#)
  35. Matsuyama T, Ohno Y, Yamamoto K, et al. Comparison of utility of deep learning reconstruction on 3D MRCPs obtained with three different k-space data acquisitions in patients with IPMN. *Eur Radiol*. 2022;32(10):6658-6667. [\[CrossRef\]](#)
  36. Tajima T, Akai H, Sugawara H, et al. Feasibility of accelerated whole-body diffusion-weighted imaging using a deep learning-based noise-reduction technique in patients with prostate cancer. *Magn Reson Imaging*. 2022;92:169-179. [\[CrossRef\]](#)
  37. Obama Y, Ohno Y, Yamamoto K, et al. MR imaging for shoulder diseases: Effect of compressed sensing and deep learning reconstruction on examination time and imaging quality compared with that of parallel imaging. *Magn Reson Imaging*. 2022;94:56-63. [\[CrossRef\]](#)
  38. Afat S, Herrmann J, Almansour H, et al. Acquisition time reduction of diffusion-weighted liver imaging using deep learning image reconstruction. *Diagn Interv Imaging*. 2023;104(4):178-184. [\[CrossRef\]](#)
  39. Mardor Y, Roth Y, Ochershvilli A, et al. Pretreatment prediction of brain tumors' response to radiation therapy using high b-value diffusion-weighted MRI. *Neoplasia*. 2004;6(2):136-142. [\[CrossRef\]](#)
  40. Kitajima K, Takahashi S, Ueno Y, et al. Clinical utility of apparent diffusion coefficient values obtained using high b-value when diagnosing prostate cancer using 3 tesla MRI: comparison between ultra-high b-value (2000 s/mm<sup>2</sup>) and standard high b-value (1000 s/mm<sup>2</sup>). *J Magn Reson Imaging*. 2012;36(1):198-205. [\[CrossRef\]](#)
  41. Tavakoli AA, Kuder TA, Tichy D, et al. Measured multipoint ultra-high b-value diffusion MRI in the assessment of MRI-detected prostate lesions. *Invest Radiol*. 2021;56(2):94-102. [\[CrossRef\]](#)



**Supplementary Figure 1.** Example of image analysis of diffusion-weighted imaging (DWI) with a  $b$  value of  $1000 \text{ s/mm}^2$ . To automatically place regions of interest (ROIs), a centerline is initially drawn manually on the DWI with a  $b$  value of  $1000 \text{ s/mm}^2$  per patient. The ROIs are then automatically placed on the normal cortex and white matter in each brain hemisphere on a slice at basal-ganglia level (total of 10 ROIs = 5 ROIs  $\times$  right and left hemisphere) using a commercially available workstation provided by Canon Medical Systems. The mean and standard deviation of the signal intensity within each ROI is automatically determined. To evaluate each quantitative index from the DWI and perform intravoxel incoherent motion evaluations, all ROIs are copied to the apparent diffusion coefficient (ADC),  $D$ ,  $D^*$ , and  $f$  maps of each patient. Finally, the ADC,  $D$ ,  $D^*$ , and  $f$  values within the ROIs are determined.  $D$ , true diffusion coefficient;  $D^*$ , pseudodiffusion coefficient,  $f$ , percentage of water molecules in micro perfusion within 1 voxel.