Title

A deep learning–based method for improving reliability of multicenter diffusion kurtosis imaging with varied acquisition protocols

Author names and affiliations

Qiqi Tong[a,b,1], tongqq@zju.edu.cn
Ting Gong[a,1], gongting@zju.edu.cn
Hongjian He[a,*], hhezju@zju.edu.cn
Zheng Wang[c], zheng.wang@ion.ac.cn
Wenwen Yu[c], wwyu@ion.ac.cn
Jianjun Zhang[d], zhangyp31113@163.com
Lihao Zhai[d], zlh5721@163.com
Hongsheng Cui[e], chs01@sina.com
Xin Meng[e], mengxin777@sina.com
Chantal W.M. Tax[f], TaxC@cardiff.ac.uk
Jianhui Zhong[a,g], jzhong@zju.edu.cn

[a] Center for Brain Imaging Science and Technology, Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrumental Science, Zhejiang University, Hangzhou, Zhejiang, China
[b] Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, Zhejiang, China
[c] Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, State Key Laboratory of Neuroscience, CAS Key Laboratory of Primate Neurobiology, Chinese Academy of Sciences, Shanghai, China
[d] Department of Radiology, Zhejiang Hospital, Hangzhou, Zhejiang, China
[e] Department of Radiology, The Third Affiliated Hospital of Qiqihar Medical University, Qiqihar, Heilongjiang, China
[f] Cardiff University Brain Research Imaging Centre (CUBRIC), School of Physics and Astronomy, Cardiff University, Cardiff, United Kingdom
[g] Department of Imaging Sciences, University of Rochester, Rochester, NY, USA

[1] These authors contributed equally to this work.

***Corresponding author**

Hongjian He, hhezju@zju.edu.cn

Present address

Room 314, Zhouyiqing Building, Yuquan Campus, Zhejiang University, No. 38 Zheda Road, Hangzhou, China, 310027

Abstract

Multicenter magnetic resonance imaging is gaining more popularity in large-sample projects. Since both varying hardware and software across different centers cause unavoidable data heterogeneity across centers, its impact on reliability in study outcomes has also drawn much attention recently. One fundamental issue arises in how to derive model parameters reliably from image data of varying quality. This issue is even more challenging for advanced diffusion methods such as diffusion kurtosis imaging (DKI). Recently, deep learning–based methods have been demonstrated with their potential for robust and efficient computation of diffusion-derived measures. Inspired by these approaches, the current study specifically designed a framework based on a three-dimensional hierarchical convolutional neural network, to jointly reconstruct and harmonize DKI measures from multicenter acquisition to reformulate these to a state-of-the-art hardware using data from traveling subjects. The results from the harmonized data acquired with different protocols show that: 1) the inter-scanner variation of DKI measures within white matter was reduced by 51.5% in mean kurtosis, 65.9% in axial kurtosis, 53.7% in radial kurtosis, and 61.5% in kurtosis fractional anisotropy, respectively; 2) data reliability of each single scanner was enhanced and brought to the level of the reference scanner; and 3) the harmonization network was able to reconstruct reliable DKI values from high data variability. Overall the results demonstrate the feasibility of the proposed deep learning–based method for DKI harmonization and help to simplify the protocol setup procedure for multicenter scanners with different hardware and software configurations.

Keywords:
Deep learning
Diffusion magnetic resonance imaging
Diffusion kurtosis imaging
Multicenter harmonization

[1]1. Introduction

Diffusion magnetic resonance imaging (dMRI) is a promising technique for investigating tissue microstructure and neural networks in the brain. Recently, dMRI scans have been included in most national and international human brain projects[1–5]. In these projects, the MRI data are usually collected from multiple centers in considerations of needed sample size and diversity. For instance, the Adolescent Brain Cognitive Development study collected data on children and teenagers from varied racial and ethnic backgrounds in 21 sites[3], and the Alzheimer's Disease Neuroimaging Initiative created a longitudinal imaging repository on subjects with normal cognition, mild cognitive impairment, and Alzheimer's disease across more than 50 scanners[6]. However, the heterogeneity of MRI data among study centers became a critical issue[7]. In general, differences in both hardware and software can reduce the reliability of data across centers, and their effects have been observed in morphology[8,9], function[10–12], and diffusion studies. Especially in dMRI, scanner-induced biases can cause incoherence in the diffusion metrics[13–17], and our recent study revealed that between-scanner variations still exist even among very well-controlled hardware components and software protocols[18]. Although it is always possible to improve data reliability by minimizing differences in gradient and coil systems, as well as by using similar protocols, it is highly desirable to develop effective methods to harmonize multicenter dMRI data using widely varied settings.

There have been studies discussing the harmonization of multicenter diffusion tensor imaging (DTI) data, for example, by normalizing histograms of diffusion-derived metrics[19,20] and then correcting the scanner- or subject-specific factors[21,22], such as ComBat method[20]. Other studies have also applied harmonization on diffusion signal to minimize features including rotation invariant spherical harmonic (RISH) [23–25] or spherical moment[26] across scanners. Recently, the capability of deep learning–based methods was demonstrated to efficiently reconstruct diffusion parameters, such as predict multishell diffusion signal from single-shell[27], generate high order diffusion models[28,29] or fiber orientations from few diffusion weighted signals[30], and reveal untapped histology-liked fiber information from subjects scanned in multiple scanners[31,32]. These methods make it possible to save data of low quality. Inspired by these approaches, deep

[1] *Abbreviations:*

ACR - anterior corona radiata; AD - axial diffusivity; AK - axial kurtosis; ALIC - anterior limb of internal capsule; ANOVA - analysis of variance; BCC - body of the corpus callosum; CgC - cingulate gyrus cingulum; CP - cerebral peduncle; CSF - cerebrospinal fluid; CV - coefficient of variation; DKI - diffusion kurtosis imaging; dMRI - diffusion magnetic resonance imaging; DTI - diffusion tensor imaging; DWI - diffusion-weighted image; EC - external capsule; FA- fractional anisotropy; GCC - genu of corpus callosum; KFA - kurtosis fractional anisotropy; MD - mean diffusivity; MK - mean kurtosis; NRMSE - normalized root mean square error; PCR - posterior corona radiata; PLIC - posterior limb of the internal capsule; PTR - posterior thalamic radiation; RD - radial diffusivity; ReLU - rectified linear unit; RISH - rotation invariant spherical harmonic; RK - radial kurtosis; RLIC - retrolenticular part of internal capsule; ROI - region of interest; SCC - splenium of corpus callosum; SCR - superior corona radiata; SGM - subcortical grey matter; SLF - superior longitudinal fasciculus; SNR - signal-to-noise ratio; SS - sagittal stratum; TE - echo time; TR - repetition time; WM - white matter

learning networks were adopted to harmonize diffusion data from multiple scanners to a reference scanner[33,34]. At present, it remains to evaluate the harmonization performance on more high-order models.

Diffusion kurtosis imaging (DKI) is one of the advanced high-order diffusion models, utilizing second and fourth order tensors involves 21 independent components[35,36]. Recent literature has found DKI useful in many physiological changes, including in the early detection of diseases like gliomas[37–39] and neural disorders[40,41], as well as in the discovery of neuronal patterns in development[42] and aging[43–45]. However, the estimation from diffusion signals is sensitive to scanning-related factors such as acquisition protocol, image noise, and artifacts, which can induce anomalous deviations in the derived metrics[46–48]. Therefore, the harmonization of DKI measures faces challenges in accuracy, which can be reduced from fitting failures. The large sample projects of the above-mentioned studies demand a specific approach to efficiently merge DKI data from different centers.

In this study, a deep learning–based framework was designed for harmonization of DKI measures on two datasets with either different or identical protocols. For the reliability evaluation of harmonization, the DKI measures before and after harmonization were carefully compared in regard to the variance across different scanners and variance between each scanner and reference. This method was also compared with two existing harmonization methods of ComBat and RISH. Moreover, in consideration of the difference in data diversity between small training data and large-scale multicenter data, additional tests were conducted for data reliability on different tissues.

2. Methods

2.1. Datasets, subjects, and data preprocessing

The proposed method was evaluated in two independent datasets. In Dataset 1, five healthy traveling volunteers (25.4 ± 1.7 years of age, two males) were imaged using four 3T MRI scanners of different types (MAGNETOM Prisma/Skyra/Trio, Siemens Healthcare, Erlangen, Germany; Discovery MR 750, GE Healthcare, Milwaukee, WI) and with different protocols. For each scanner, the diffusion sequence was optimized for the shortest echo time (TE) with the same voxel size, similar total scan time, and a b-table containing a minimum of two b-values with 60 directions in total. The diffusion protocols of all scanners are listed in Table 1. The T1-weighted images were acquired using clinical sequences (magnetization-prepared rapid acquisition with gradient echo on Siemens scanners and brain volume imaging on the GE scanner). For each scanner, all data were collected on the same day, and all scans were completed within six months. Data acquisition was conducted with approval from the local institutional review boards, and all volunteers had signed informed consent forms beforehand.

Table 1. Diffusion protocols of the four scanners used in Dataset 1.

Scanner	Siemens Trio Scanner A	Siemens Skyra Scanner B	GE MR750 Scanner C	Siemens Prisma Scanner D (Reference)
Gradient Strength	40 mT/m	45 mT/m	50 mT/m	80 mT/m

Slew Rate	200 T/m/s	200 T/m/s	200 T/m/s	200 T/m/s
Head Coil	32ch	20ch	8ch	64ch
Sequence	SE EPI	SE EPI	SE EPI	Multi-band SE EPI
Matrix	110x110	110x110	110x110	110x110
Resolution [mm^3]	2x2x2	2x2x2	2x2x2	2x2x2
TR/TE [ms]	9200/99	9300/94	9000/80	3000/75
Bandwidth [Hz/Px]	1684	1818	2272	2272
Partial Fourier	6/8	6/8	4/5	6/8
Acceleration	GRAPPA 2	GRAPPA 2	none	Multiband 4
Repetitions	2	2	1	2
Scan Time	11:14	11:56	10:12	10:18
b-value [s/mm^2]	1000, 2000	1000, 2000	1000, 2000	1000, 2000, 3000
Directions	30, 30	20, 40	30, 30	30, 30, 30
b0 Frames	6, 1 opp. PE*	6, 1 opp. PE*	6	6, 1 opp. PE*

* Opposite phase-encoding (PE) direction.

For preprocessing, the diffusion-weighted images (DWIs) of Dataset 1 were firstly denoised[49] and Gibbs-ring removed[50]. Then the images acquired from scanner A, C, and D with reversed phase-encoding were further corrected for distortion and motion using the TOPUP and EDDY tools in the FSL software[51,52]. It should be noted that the repetition scans were not combined after the preprocessing and were seen as repeated scans.

Dataset 2 is an open benchmark dataset from the multi-shell diffusion MRI harmonisation challenge (MUSHAC) with an identical "standard" protocol on two scanners[33,53]. It consisted of diffusion- and T1-weighted data from two scanners of different models, the Siemens Prisma and the Siemens Connectom. On both scanners, the protocols were kept the same for a standard two-shell diffusion acquisition: isotropic resolution of 2.4 mm, TE = 89 ms, TR = 7200 ms, and 60 directions at b-value = 1200, 3000 s/mm^2. The released data were preprocessed with distortion and motion correction. Among the participating 10 traveling subjects, one (marked as "G") was excluded after manual data inspection due to the apparent lower image intensity on DWIs acquired in Connectom scanner.

2.2. DKI harmonization

The harmonization was designed based on a three-dimensional hierarchical convolutional neural network, which was specifically built for parametric DKI reconstruction from diffusion signals [29,54]. This artificial neural network was comprised of an input layer as DWIs from one scan, hidden layers consisting of a convolutional layer and several fully connected layers, and two output layers to generate eight DKI maps. The kernel size in convolutional layer was 3x3x3 and in fully connected layer was 1x1x1. All the hidden layers consisted of 150 epochs and rectified linear unit (ReLU) activation function. A dropout regularization size of 10% was considered before the output layer to avoid overfitting. Three diffusivity measures were then derived from the shallow output layer, including the mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD), whereas the fractional anisotropy (FA) together with the four DKI measures mean kurtosis (MK), axial kurtosis (AK), radial kurtosis (RK), and kurtosis fractional anisotropy (KFA) were derived

from a deeper output layer containing one more hidden layer.

A scanner to be harmonized is termed target scanner, and the one representing the harmonization goal is termed reference scanner. Fig. 1 illustrates the schematic framework for harmonizing DKI from a target scanner. First, the data from several subjects were randomly chosen for training. The training labels of DKI measures from reference scanner were estimated using an iteratively reweighted linear least square method[55,56], and were non-linearly registered to the DWIs from same subject acquired on target scanner using the FNIRT tool of the FSL library[57]. Second, the matched DWIs and DKI labels were utilized to train and optimize the neural network. Third, in the testing phase, the target scanner DWIs of previously unused test subjects were input individually into the trained neural network to predict the harmonized DKI measures. Additionally, if repeated scans were acquired, another set of harmonized DKI measures was generated from the same trained neural network. To keep all harmonized DKI measures consistent for the whole dataset, DWIs from the reference scanner also went through the same framework except for the registration of DKI labels before the harmonization.

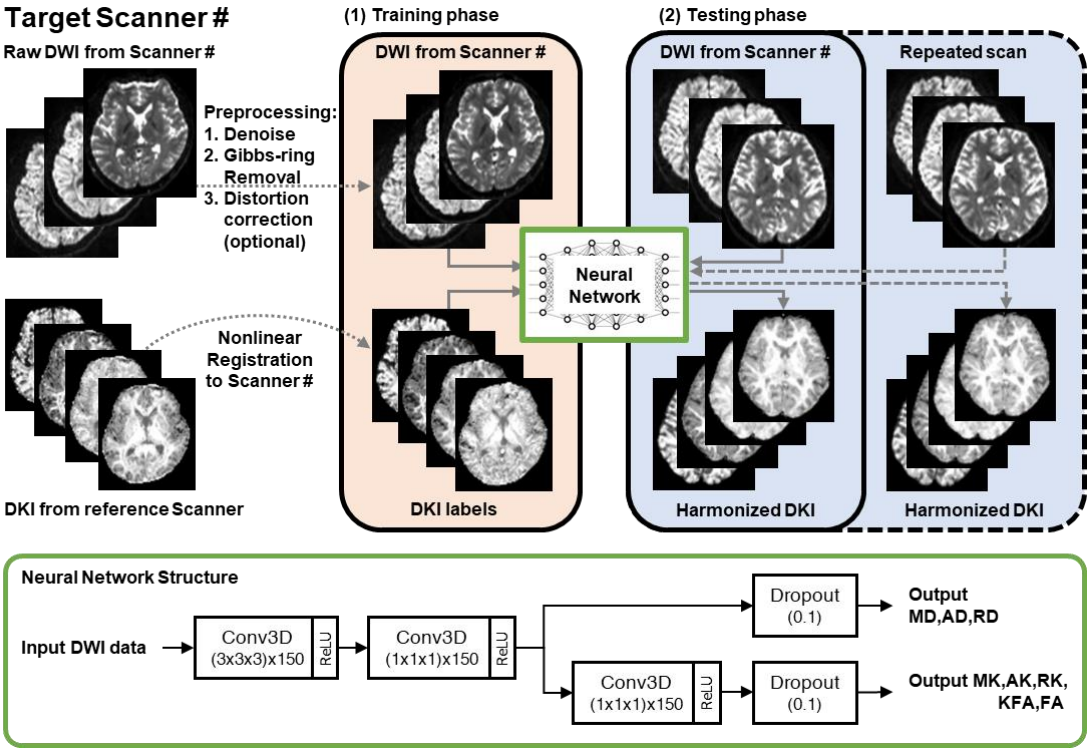


Fig. 1. Framework for the harmonization of one target scanner. The target scanner is denoted by the symbol "#".

In this study, reference scanner was chosen as the device with highest signal-to-noise ratio (SNR) in each dataset. In the training phase, in order to have enough number of effective voxels for training, we used 40% of the subjects from the whole dataset to ensure harmonization performance. In other words, two subjects from Dataset 1 and four from Dataset 2 were selected for the training. As parameters of interest, MK, AK, RK, and KFA within the white matter (WM) were selected in this study.

2.3. Experiments and evaluations

We conducted four types of evaluation experiments on Dataset 1, including harmonization performance, within-scanner repeatability, leave-p-out validation, and across-method comparison, as summarized in Table 2. On the other hand, on Dataset 2, the harmonization on two scanners was conducted with four subjects for training and the other five ones for testing, the effect of training size was examined by using a range of 1 to 8 randomly selected training subjects.

Table 2. Brief summary of designed experiments on Dataset 1.

Experiments		Evaluation criteria	Training subjects	Testing subjects
(1) Harmonization on 4 scanners		Inter-scanner variance Single scanner reliability	2	3
(2) Repeated scans on 3 scanners		Intra-scanner reproducibility	2	3 (Two repeated scans)
(3) Validation	Training size	Single scanner reliability	1/2/3/4	1
	Leave-p-subjects-out		2/4	3/1
	Leave-one-tissue-out		2 (Without SGM/CSF)	3 (Whole brain)
(4) Comparison	ComBat	Inter-scanner variance	5	5
	RISH		2	3
	DL		2	3

2.3.1. Inter-scanner data variance

Inter-scanner data variance was determined by variance of systematical bias across scanners. It was evaluated by inter-scanner coefficient of variation (CV) among all scanners on same subject for both datasets. The within-subject registration on DKI measures after harmonization was firstly processed using the FNIRT tool. Then the CVs within WM were calculated pixel-wisely on each DKI measure using the Matlab software.

2.3.2. Data reliability on single scanner

Data reliability on single scanner was determined for each target scanner by its variance of measurements to those from the reference scanner. It was evaluated within WM between each DKI measure from target scanner and the label from reference scanner by the within-subject normalized root mean square error (NRMSE), noted as NRMSE-to-reference. For Dataset 1, the first scans on three test subjects from all four scanners were selected for this evaluation, whereas for Dataset 2, the scans on five test subjects from both scanners were assessed. The NRMSEs-to-reference were calculated within 15 WM regions of interest (ROIs), which were labeled individually according to the ICBM-DTI-81 WM atlas[58].

For the repeated scans in Dataset 1, the intra-scanner reproducibility was evaluated by within-subject NRMSE between two repetitions. Comparison among 15 ROIs were also performed.

2.3.3. Validation on data reliability

The harmonization performance of the neural network was evaluated and validated by changing the training data. First experiment was designed by using different numbers and different combinations of subjects for the training data, meanwhile holding one subject in the testing phase for evaluation. The NRMSEs-to-reference were then calculated within WM between the harmonized data from each target scanner and labels from reference scanner on the testing subject. In Dataset 1, the training size was chosen from 1 to 4 subjects, and changed different combinations of subjects at each training size. In Dataset 2, the training size was chosen from 1 to 8 subjects.

Second experiment was designed according to the leave-p-out cross-validation[59]. Two strategies of training, leave-three-subjects-out and leave-one-subject-out, were traversed among five subjects in Dataset 1. The NRMSEs-to-reference within WM on the test subjects were compared among different trainings.

Moreover, to evaluate the feasibility of neural network's application on data with potential high diversity, third experiment was designed by manually decreasing the diversity of training data but holding the testing data. Thus, the leave-one-tissue-out strategy with one tissue masked out in the training was conducted on Dataset 1. Before the training, one kind of tissue was removed from the DWIs and DKI labels of all training subjects, and the DWIs for testing still involved the whole brain. After the harmonization, the NRMSEs-to-reference were calculated and compared with those from the whole brain training. We selected two tissues for removal like subcortical grey matter (SGM) of low diffusivity and cerebrospinal fluid (CSF) of high diffusivity separately for two training experiments. Moreover, each training was repeated for 10 times without losing any stability of the network.

2.3.4. Comparison with other harmonization methods

To compare the proposed deep learning–based method with the two methods, ComBat and RISH, data reliability of the three approaches were evaluated on Dataset 1. For applying the Combat and RISH methods, additional processing steps on DWIs were conducted and described in detail below. The RISH method requires training and testing, where the training data on two subjects were set the same with the proposed deep learning–based method. For comparison, all the approaches generated the inter-center CV on one same subject.

Since the ComBat method harmonizes voxel-wise DKI-derived data, the DKI metrics from all five subjects were firstly non-linearly registered into a template space using FSL. Then site effect was adjusted parametrically from each distribution, and four DKI measures were harmonized separately. Afterwards, the harmonized metrics were registered back to the individual space.

The RISH algorithm can only harmonize single-shell DWIs. Firstly, the multi-shell DWIs should be reorganized into two single-shell data of 1000 and 2000 s/mm^2. It should be noted that the data with b-value of 3000 s/mm^2 from scanner D were left out. Secondly, each single-shell DWIs were harmonized from target scanner to reference scanner D of same b-value. Finally, the two single-shell DWIs were assembled to generate the DKI measures, and the within-subject registration was performed before the CV calculation.

3. Results

3.1. Quality assessment

The SNR of a WM ROI the genu of corpus callosum (GCC) was measured for all scans and listed in table 3. The SNRs were generally highest on scanner D in Dataset 1, and on Connectom scanner in Dataset 2 respectively. The two scanners were then selected as the reference scanners.

Table 3. Signal-to-noise ratio (SNR)* of all scans in two datasets.

Subject	Dataset 1				Dataset 2	
No.	Scanner A	Scanner B	Scanner C	Scanner D	Prisma	Connectom
1	23.51	24.28	23.06	30.21	24.51	27.98
2	23.32	24.58	23.66	30.02	24.96	25.09
3	22.14	24.16	24.96	30.66	20.36	22.23
4	21.65	24.93	25.71	31.47	23.36	28.75
5	20.44	21.89	21.17	25.85	23.31	26.94
6	-	-	-	-	23.25	18.85
7	-	-	-	-	23.25	21.68
8	-	-	-	-	19.93	25.66
9	-	-	-	-	22.59	25.81

* All values are in unit of dB.

3.2. Improvement of inter-scanner reliability

Fig. 2 shows the parametric maps of four selected DKI measures from all scanners in two representative subjects of the two datasets. The variance among scanners can be seen before the harmonization, especially in the parameter KFA. After the harmonization, the parametric maps of all scanners were generally closer to each other as well as to that of the reference scanner.

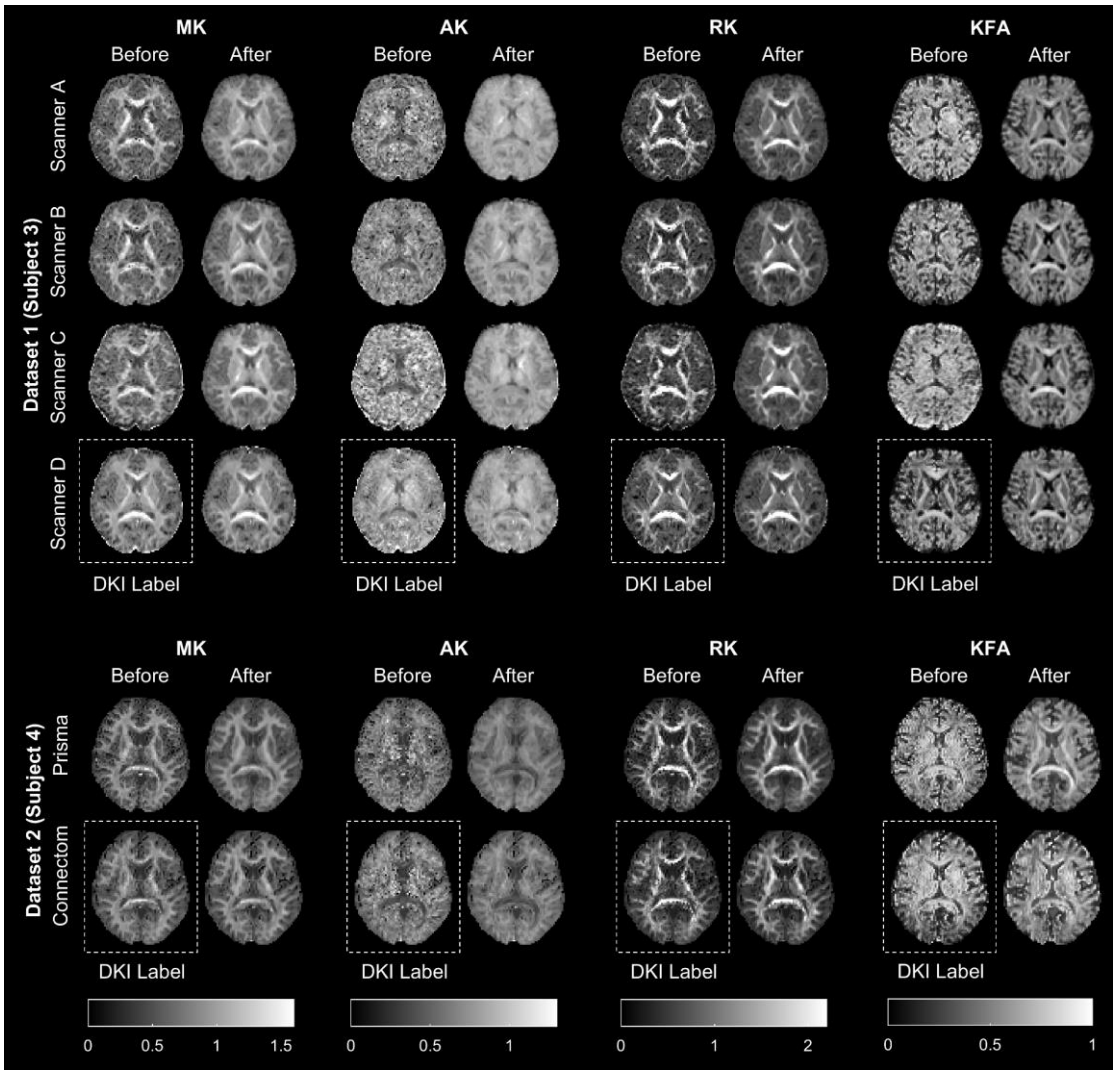


Fig. 2. Parametric diffusion kurtosis imaging (DKI) maps from all scanners before and after harmonization. The two representative subjects shown are subject 3 for Dataset 1 and subject 4 for Dataset 2. The outlined DKI labels of the two datasets are reference scans from scanner D and Connectom, respectively.

Fig. 3 presents the CV maps of the DKI measures from the same subjects shown in Fig. 2 before and after harmonization. In both datasets, the original scanner-related variances were considerable. These variances were substantially reduced after the proposed harmonization. In Dataset 1, the mean CVs within the WM were reduced from 13.6%, 17.3%, 24.4%, and 16.7% to 6.6%, 5.9%, 11.3%, and 9.4% in MK, AK, RK, and KFA, respectively. The CV changes in Dataset 2, which was generated using only one scanning protocol, were in general relatively moderate. The relevant reductions were 3.3%, 9.7%, 8.7%, and 5.4% for MK, AK, RK, and KFA, respectively. Similar reduction values were observed in the other subjects of the two datasets.

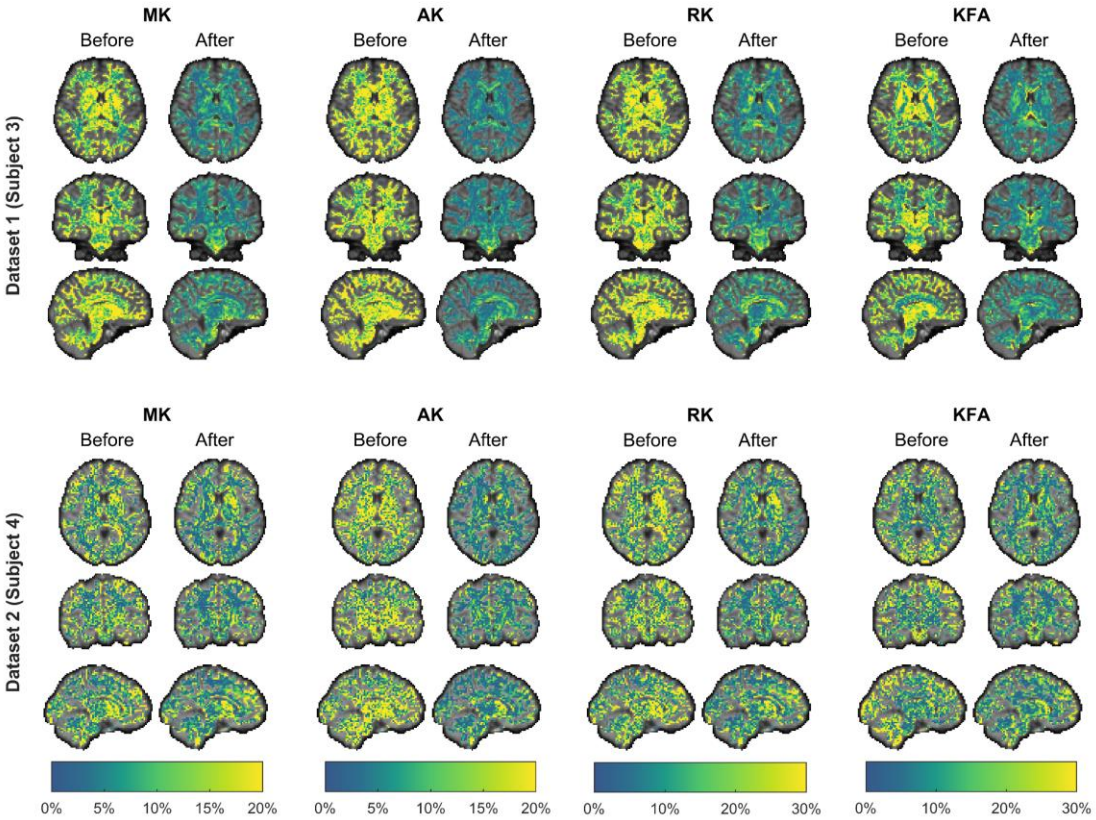


Fig. 3. Coefficient of variation (CV) maps of the DKI measures in two representative subjects before and after harmonization. Maps in the upper and lower rows are from subject 3 in Dataset 1 and from subject 4 in Dataset 2, respectively. The colored CV maps are overlaid on structural images of three different orientations.

3.3. Improvement of data reliability on single scanner

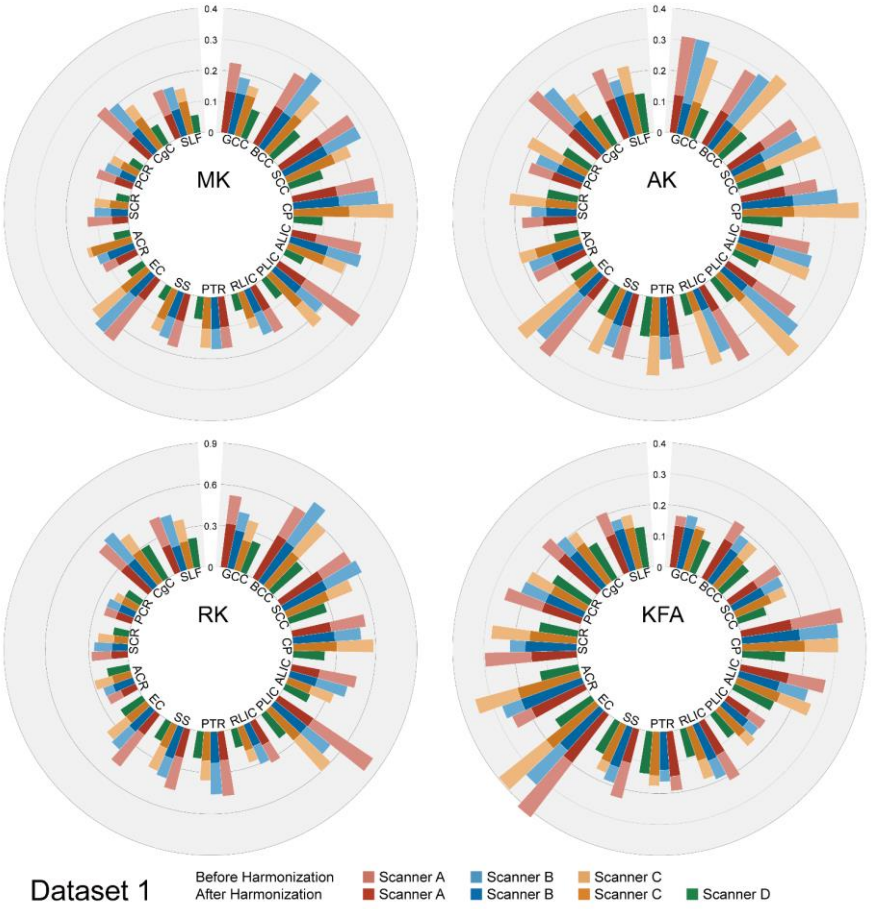
The data reliability of each target scanner was measured by the parameter NRMSE-to-reference (Table 4) within WM. After the harmonization, the NRMSEs-to-reference across the three scanners in Dataset 1 were mostly reduced by 40%. This reduction was found smaller in Dataset 2. It is notable that all test subjects in each dataset showed comparable NRMSEs-to-reference after the harmonization.

Table 4. NRMSEs-to-reference of DKI measures on each scanner in all test subjects.

Measure	Subject No.	Dataset 1							Dataset 2		
		Scanner A		Scanner B		Scanner C		Scanner D	Prisma		Connectom
		Before	After	Before	After	Before	After	After	Before	After	After
MK	1	0.19	0.11	0.18	0.10	0.22	0.14	0.09	0.20	0.18	0.08
	2	0.21	0.12	0.19	0.10	0.23	0.13	0.07	0.17	0.15	0.09
	3	0.22	0.11	0.19	0.11	0.26	0.17	0.08	0.18	0.14	0.09
	4								0.17	0.14	0.09
	5								0.17	0.14	0.08
AK	1	0.24	0.14	0.23	0.13	0.31	0.16	0.14	0.31	0.25	0.22
	2	0.27	0.14	0.22	0.12	0.29	0.15	0.12	0.27	0.21	0.21
	3	0.28	0.13	0.23	0.12	0.32	0.19	0.12	0.29	0.21	0.22

	4								0.27	0.21	0.22
	5								0.26	0.20	0.21
	1	0.44	0.23	0.40	0.22	0.39	0.24	0.22	0.34	0.28	0.24
	2	0.47	0.25	0.43	0.23	0.44	0.24	0.20	0.35	0.29	0.26
RK	3	0.45	0.24	0.42	0.24	0.42	0.24	0.21	0.36	0.26	0.24
	4								0.36	0.29	0.26
	5								0.33	0.27	0.22
	1	0.26	0.16	0.25	0.16	0.34	0.19	0.16	0.29	0.23	0.18
	2	0.27	0.17	0.24	0.16	0.34	0.20	0.14	0.24	0.21	0.16
KFA	3	0.30	0.18	0.25	0.17	0.37	0.20	0.14	0.27	0.22	0.17
	4								0.26	0.21	0.17
	5								0.26	0.21	0.16

Fig. 4 shows the NRMSEs-to-reference before and after harmonization in 15 ROIs for both datasets. In Dataset 1, the variance of DKI measures before harmonization varied much in different ROIs. For example, in scanners A, B, and C, their mean standard errors (SD) among all these ROIs in MK and KFA were 0.061 and 0.066, respectively. And after the harmonization, the NRMSEs-to-reference were comparable among these ROIs, and the SDs were respectively reduced to 0.032 and 0.030. The data reliabilities of three scanners were improved to be closer to those on the reference scanner, whose SDs were 0.024 and 0.017 in MK and KFA, respectively. In Dataset 2, the diversity of NRMSEs-to-reference in different ROIs from Prisma scanner was low before harmonization, the SDs in MK and KFA were 0.064 and 0.028, respectively. The data reliability of two scanners after harmonization were similar, which were respectively reduced to 0.050 and 0.026 in Prisma, and 0.045 and 0.023 in Connectom. respectively.


MK
AK
RK
KFA
GCC
BCC
SCC
CP
ALIC
PLIC
RLIC
PTR
SS
EC
ACR
SCR
PCR
CgC
SLF
Dataset 1
Before Harmonization
After Harmonization
Scanner A
Scanner A
Scanner B
Scanner B
Scanner C
Scanner C
Scanner D

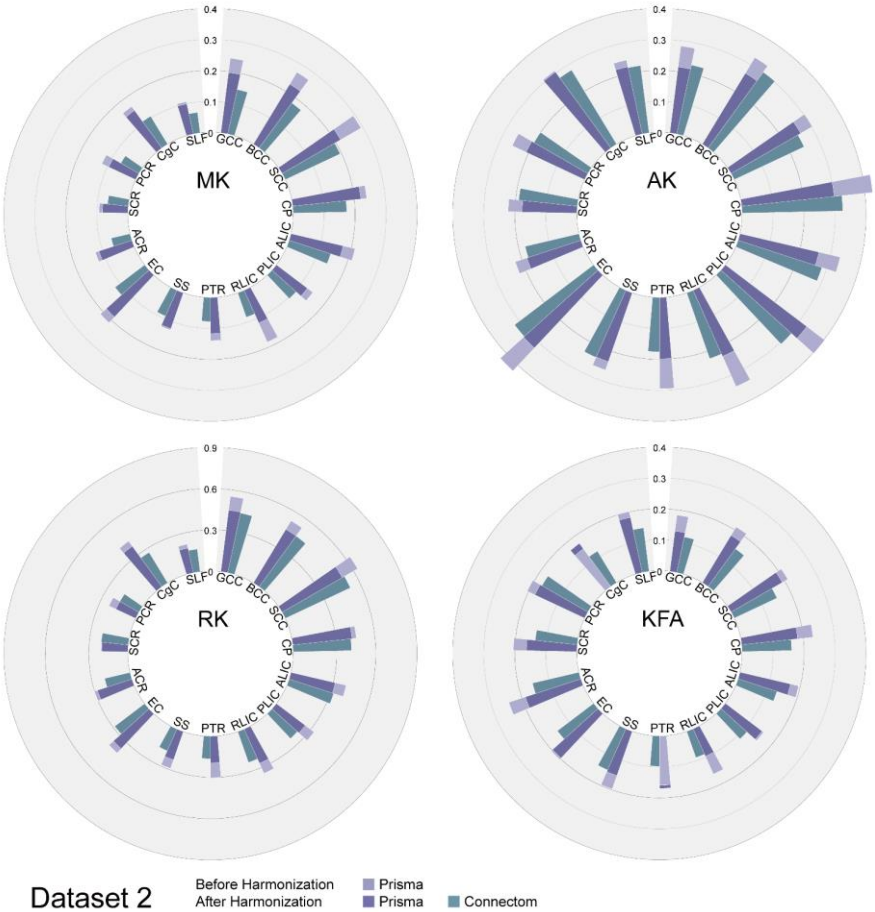

MK
AK
RK
KFA
GCC
BCC
SCC
CP
ALIC
PLIC
RLIC
PTR
SS
EC
ACR
SCR
PCR
CgC
SLF
Dataset 2
Before Harmonization
After Harmonization
Prisma
Prisma
Connectom

Fig. 4. Circular diagrams for normalized root mean square errors-to-reference (NRMSEs-to-reference) of DKI measures in 15 ROIs from representative subject 3 in Dataset 1 (upper group) and subject 4 in Dataset 2 (lower group) are shown here respectively. Each scanner is coded in different color, and the bar chart displays the value before (in light color) and after (in dark color) harmonization for selected regions. These ROIs include: GCC, body of corpus callosum (BCC), splenium of corpus callosum (SCC), cerebral peduncle (CP), anterior limb of internal capsule (ALIC), PLIC, retrolenticular part of internal capsule (RLIC), posterior thalamic radiation (PTR), sagittal stratum (SS), external capsule (EC), ACR, superior corona radiata (SCR), posterior corona radiata (PCR), cingulate gyrus cingulum (CgC) and superior longitudinal fasciculus (SLF).

The intra-scanner reliability of the scanners A, B, and D in Dataset 1 were estimated by NRMSE with a higher NRMSE value indicating a larger variance between two repeated scans on same subject, as shown in Table 5. For the scanners A and B, the harmonization greatly improved their reliability to a comparable level with that of the reference scanner D.

Table 5. NRMSEs of DKI measures between two repeated scans in all test subjects.

Measure	Subject No.	Scanner A		Scanner B		Scanner D	
		Before	After	Before	After	Before	After
MK	1	0.20	0.08	0.17	0.05	0.15	0.14
	2	0.21	0.08	0.17	0.04	0.09	0.08
	3	0.22	0.08	0.17	0.05	0.12	0.11
AK	1	0.29	0.07	0.24	0.04	0.16	0.12
	2	0.31	0.07	0.22	0.04	0.11	0.06
	3	0.32	0.06	0.23	0.04	0.12	0.09
RK	1	0.37	0.16	0.33	0.09	0.26	0.19
	2	0.41	0.15	0.33	0.08	0.21	0.15
	3	0.40	0.15	0.33	0.09	0.24	0.17
KFA	1	0.24	0.10	0.19	0.05	0.16	0.11
	2	0.24	0.10	0.19	0.04	0.13	0.09
	3	0.24	0.10	0.20	0.05	0.14	0.10

To intuitively investigate the change of DKI measures in different regions, Fig. 5 shows the scatter distributions of these measures before and after harmonization for three representative ROIs, the GCC, the posterior limb of internal capsule (PLIC), and the anterior corona radiata (ACR). The DKI measures of the scanners A and B before harmonization were distributed dispersedly, but the harmonization helped to normalize their distributions to that of the reference scanner D, resulting in improved intra-scanner reliability. It is also notable that the aggregation of each cluster was improved, which means that the repeated scans were closer to each other, even for the KFA on scanner D in the GCC. However, the ROIs still displayed dots that were after the harmonization not near the equality line.

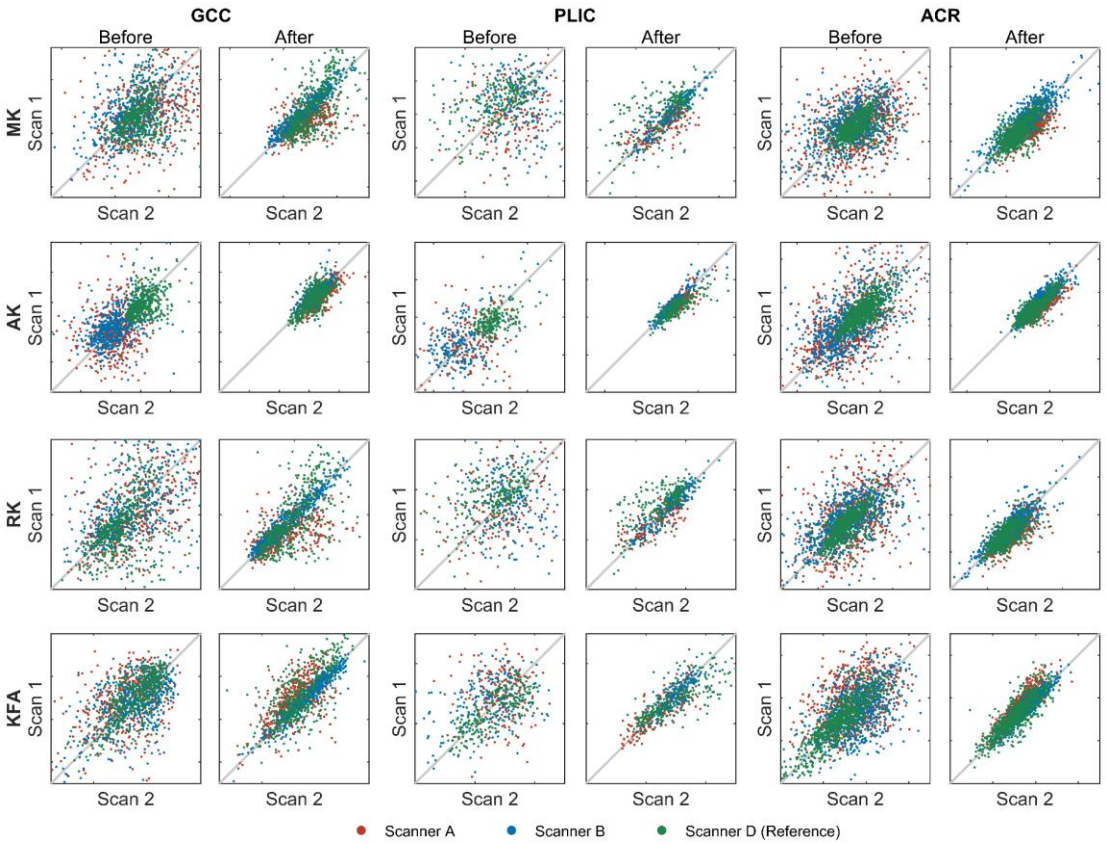


Fig. 5. Scatter distributions of diffusion measures in the regions of interest (ROIs) like genu of corpus callosum (GCC), posterior limb of internal capsule (PLIC), and anterior corona radiata (ACR) for the two repetition scans of subject 3. The auxiliary lines are the equality lines between the two scans. Each color represents one scanner type, and the reference is scanner D.

3.4. Validation for data reliability

In the first validation experiment, the NRMSEs-to-reference on subject 3 in Dataset 1 and on subject 4 in Dataset 2 over training size are shown in Fig. 6. Comparing to the original DKI difference, the NRMSEs-to-reference after harmonization on all scanners were reduced significantly, indicating increased data reliability using the deep learning harmonization method. Although the data reliability was increased further with the training size increased, the tendency of all DKI measures slowed down. For Dataset 1, increase of training subject numbers from 1 to 2, the NRMSEs-to-reference of MK on Scanner A was reduced by 7.6%, but only reduced by 5.2% further from 2 to 4 training subjects. It was similar in Dataset 2, the NRMSEs-to-reference of MK on Prisma was reduced by 11.2% from 1 to 4 training subjects, but only reduced by 4.9% further from 4 to 8 training subjects.

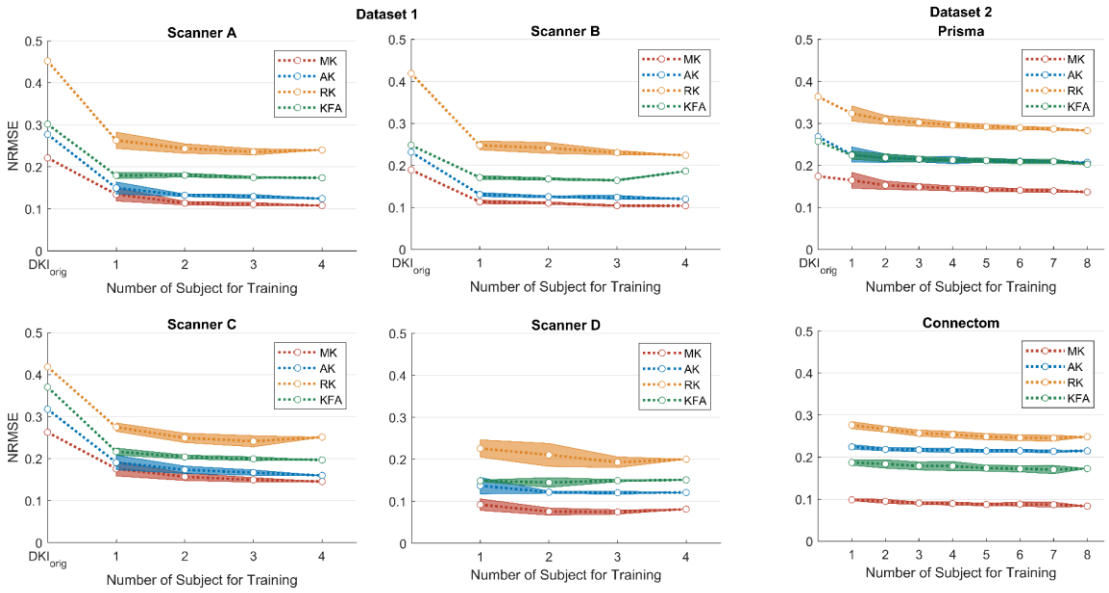


Fig. 6. NRMSEs-to-reference of DKI measures trained from different number of subjects on all scanners. The left and middle columns show the values from four scanners on subject 3 in Dataset 1 and the right column are from two scanners on subject 4 in Dataset 2. Mean values with SDs among all possible combinations are plotted in dots with error bars.

In the leave-p-subjects-out cross-validation experiment, the NRMSEs-to-reference on all scanners before harmonization, from leave-three-subjects-out training, and from leave-one-subjects-out training in Dataset 1 were shown in Fig.7. In all scanners, the NRMSEs-to-reference in two training strategies were comparable, and the values among different trainings were similar, indicating the robustness of neural network to different training data.

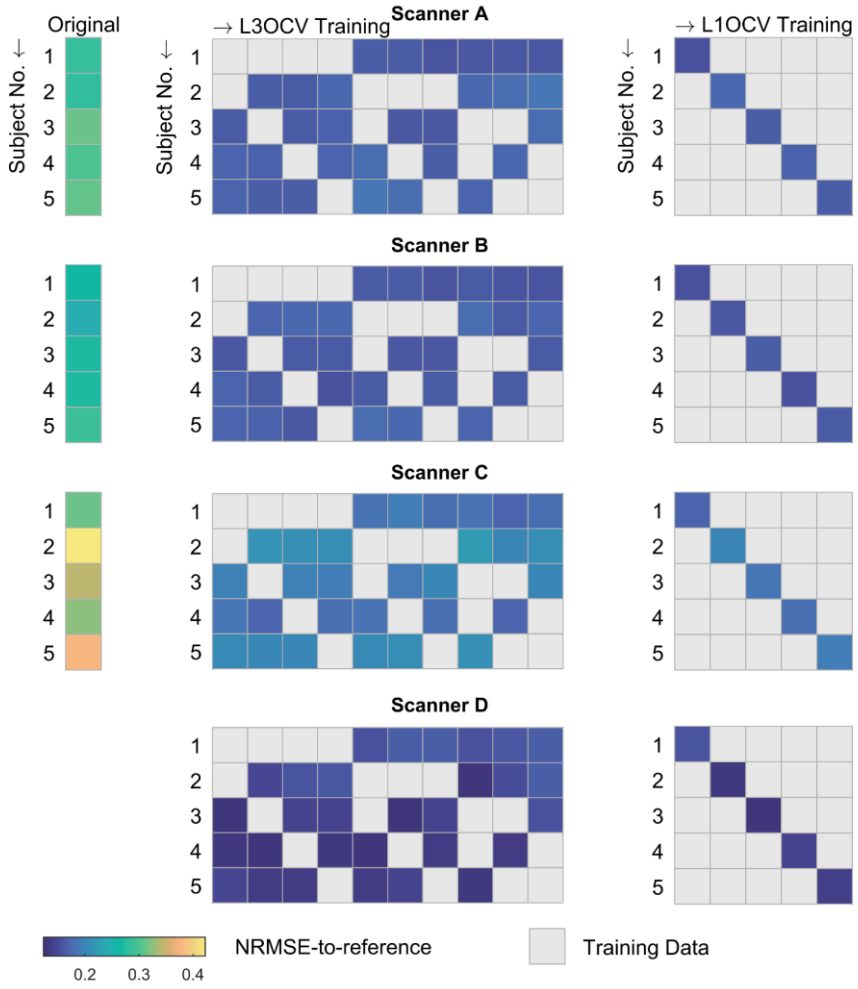


Fig. 7. NRMSEs-to-reference on different subjects from leave-p-subjects-out training on four scanners in Dataset 1. Three columns show the original values on all subjects, the harmonized values on testing subjects from leave-three-subjects-out cross-validation (L3OCV), and from leave-one-subjects-out cross-validation (L1OCV), respectively. The grey squares indicate the training subjects, and the squares with different colors indicate the averaged NRMSEs-to-reference among all DKI measures on the relevant subjects.

In the leave-one-tissue-out experiment, the predicted DKI values from the trainings with whole brain tissue, SGM removed, and CSF removed and their NRMSEs-to-reference on subject 3 from the Dataset 1 are shown in Fig. 8. The mean values and standard errors were calculated from 10 outputs of repeated trainings within three interested tissues of WM, SGM, and CSF. Comparing different tissues, the WM was trained in all the trainings and show lowest variance. The SGM showed slightly higher NRMSEs-to-reference in the SGM removal training than the other two, especially in KFA. Likewise, the CSF showed higher NRMSEs-to-reference when the training excluded CSF. However, the DKI measures varied greatly among scanners in CSF for its low kurtosis character.

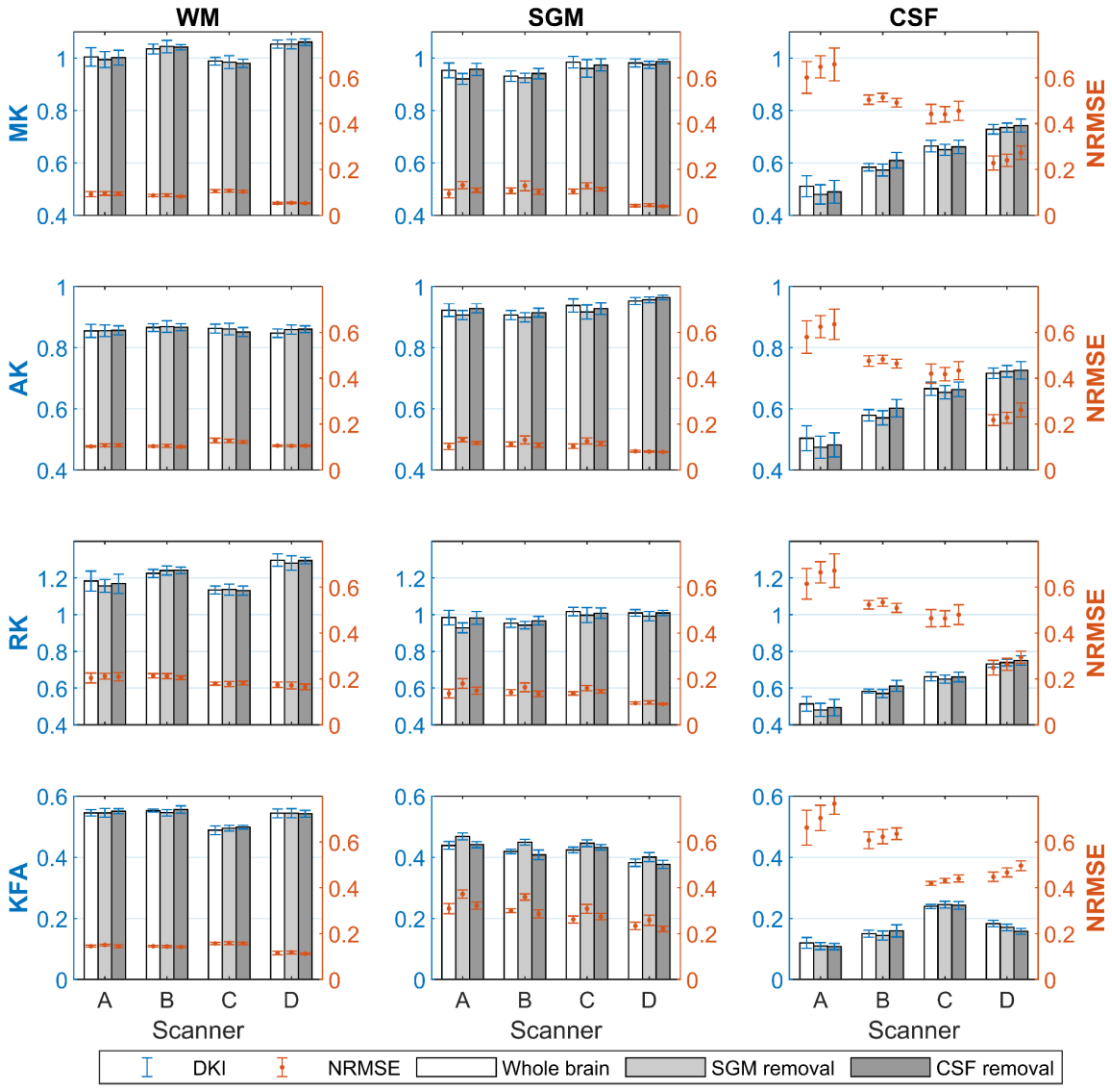


Fig. 8. The DKI measures and NRMSEs-to-reference within white matter (WM), subcortical grey matter (SGM), and cerebrospinal fluid (CSF) from three training experiments on subject 3 of Dataset 1. The mean values with standard errors are from 10 repeated trainings of each experiment. The blue ones denote the DKI measures according to left axes, and the red ones denote the NMRSEs-to-reference according to the right axes. The bars in white, light grey, and dark grey indicates three experiments that training with whole brain, with SGM removed, and with CSF removed, respectively.

3.5. Comparison with other harmonization methods

The original DKI measures from all scanners and the harmonized ones using three methods, together with the relevant histograms of inter-scanner CV are presented in Fig. 9. The DKI measures of each scanner harmonized by ComBat were generally smoother than the original ones, and the inter-scanner variance were reduced apparently. Whereas the harmonized DWIs via RISH method produced the DKI maps disturbed by large signal dropouts, the inter-scanner CV was influenced thus decreased, especially on KFA. The proposed deep learning–based method harmonized all the DKI measures to a similar intensity level of the reference scanner D, and the CV within WM was overall lower than those in the other two methods.

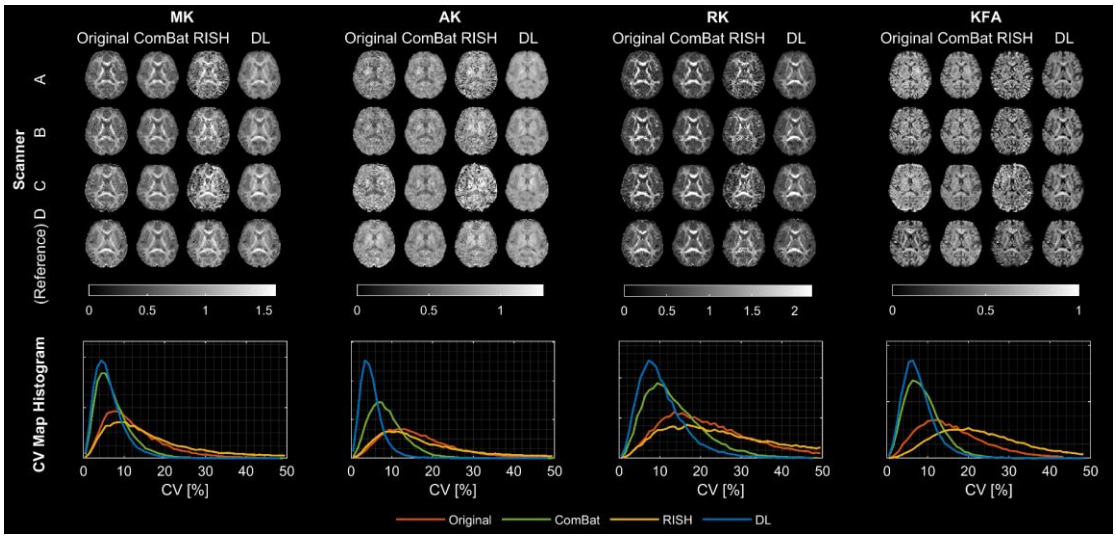


Fig. 9. Original parametric DKI maps and harmonized ones using ComBat, rotation invariant spherical harmonic (RISH), and the proposed method (noted as DL) in all scanners on subject 3 of Dataset 1. The histograms of inter-scanner CV within WM are shown in the lower row.

4. Discussion

In this study, multicenter harmonization of DKI measures was facilitated by a deep learning–based method applied to a relatively small training dataset of traveling subjects. It demonstrated the effectiveness of the proposed method in harmonizing a high-order diffusion model of DKI measures from different types of scanners or acquisition protocols. The harmonization process used high-quality images from one reference scanner and improved the reliability of diffusion kurtosis measures for all other scanners. This method jointly optimizes the network between the diffusion signals and the measures, consequently leading to reduced fitting errors in the derived DKI maps and increased noise tolerance in the DWIs.

The harmonization of multicenter data aims to generate reliable parametric results from images acquired under inconsistent conditions. It has been noticed that variance of DTI or other measures become more substantial when the data are acquired by different scanner models or varied protocols[16,60,61]. In this study, Dataset 1 reflects such a situation, and the suggested harmonization was able to substantially reduce the WM variability among scanners, as well as to decrease their deviations from the reference scanner. Our results showed that in general, the inter-scanner reliability after the harmonization was comparable to the original intra-scanner repeatability of the high-quality reference scanner. By contrast, in Dataset 2 with its consistent protocol between scanners, the WM variance of original DKI between the two scanners was already small, thus the reduction in variance after harmonization was less obvious. It is notable that the DKI measures in Dataset 1 and Dataset 2 were comparable after the harmonization, indicating the capability of the deep learning–based method for harmonizing data acquired with different protocols.

The deep learning–based method with joint reconstruction and harmonization can also improve the

reliability in DKI measurements on single scanner. The traditional DKI fitting methods either using the constrained weighted linear least square or the iteratively reweighted linear least square algorithms could fail in some regions that leads to signal loss in the DKI measures, examples of which were shown in Figure 2. In such cases, the data reliability would worsen in a multicenter project if scanners with high noise levels were involved. In this work, the deep learning–based method was trained to eliminate these outliers and learn the reliable features from majority normal signals and their neighborhoods. Moreover, to generate stable predictions from DWIs with lower SNR in some scanners, the convolution kernel in the neural network was applied on each voxel with its neighborhoods to improve the robustness against noise [54].

The deep learning–based harmonization helps to relax the requirements on hardware and protocol setting in existing multicenter studies. The only requirement on the protocols across scanners are the same voxel size in consideration of measuring similar microstructures and avoiding redundant spatial information from convolution network. It is more convenient and practical to choose one reliable scanner with advanced deployment as the reference rather than to keep all scanners and protocols strictly identical for a multicenter project. Other pulse sequence parameters for different scanners are recommended to be tuned individually for the shortest TE value and optimal image quality, such as high SNR and low Nyquist ghost. Our results indicate that the high-repeatability DKI measures of the reference scanner can be reproduced on other scanners via the suggested harmonization procedure. Hence, the setup of protocols for each scanner would be potentially more flexible in a multicenter project.

It should be noted that the DKI has optimized five b-values for standard measure [35] and fewer b-values for rapid measure [62], however, in multicenter studies, not all scanners can generate same b-values or finish same acquisition in similar time. Therefore, in Dataset 1, we used a compromised multi-shell acquisition with more b-values in reference scanner and reduced b-values in other scanners to achieve similar acquisition time. Besides, more diffusion directions would still produce better harmonization in some tissues, such as subcortical and crossing-fiber regions, and would be more flexible to reconstruct other high-order models as well.

The RISH method that harmonize the diffusion signal from high-order RISH features makes the output DWIs flexible for diverse diffusion postprocessing. From our findings, however, its limitation in harmonizing single-shell DWI could influences the data consistency between different b-values for multi-shell data. Furthermore, the DKI metrics after harmonization were easily influenced by the fitting errors using the traditional reconstruction methods. It was obvious that in the four scanners of Dataset 1, the DKI maps from deep learning–based method with less signal loss produced lower inter-scanner CV. Therefore, we believe the current deep learning–based harmonization is more robust and effective for the DKI analysis in the multicenter studies.

The ComBat and the proposed deep learning–based method can both harmonize the measurements from diffusion model to reduce the inter-scanner variance for the same purpose. For comparison, the two approaches exist several differences in application and the results. First, the ComBat as a retrospective method should be applied until the parameter maps of all scans have been reconstructed. While the proposed method can process data after being scanned, the training only

requires a small-sample of data on several subjects, then the trained network can be used to reconstruct the harmonized DKI maps immediately on subsequently scanned subjects. Second, the quality of harmonized DKI metrics via ComBat was limited by the original scanner. From our results, the DKI maps from scanners A and B were still noisy using ComBat and lead to higher CV. Although the image quality varied among scanners in Dataset 1, the outputs of all scanners from the deep learning–based harmonization could be 'improved' to be closer to the reference scanner of high-quality, and further reduce the inter-scanner variance. Generally, the proposed deep learning–based method introduced in this study is more suitable and effective for the DKI harmonization in some long-term and scanner varied multicenter studies.

The diffusion parameters for deep learning–based harmonization is not limited to the DKI-derived metrics used in this study; the neural network can be adjusted to predict diffusion properties for different requirements. A wide range of multi-component and high-order diffusion models have been proposed and extended that relate diffusion signals to microstructural features, such as the neurite orientation dispersion and density imaging (NODDI)[63], the composite hindered and restricted model of diffusion (CHARMED)[64], and ActiveAx[65,66]. Similar to DKI, these measures are also likely to be affected by different TEs, diffusion times, b-factors, and SNRs, resulting in varied measures in multicenter studies. However, unlike the strict diffusion schemes required by these models, the input DWIs of the neural network are not limited by the use of a fixed protocol; it can potentially harmonize these diffusion measures across multiple scanner types with inconsistent protocols. Since the complexity of the derived parameters would be different, the harmonization performance remains to be tested for each model in future studies.

It is worth noting that care should be taken in applying the deep learning–based harmonization in practice. Firstly, it is essential that the harmonization performance on a scanner depends on the quality of the reference scanner. In the current study, the DKI labels from reference scanner generated by the model fitting were of high reliability in the WM, and the harmonization results indicated high efficiency of the method. However, in the grey matter region of low kurtosis, the signal dropouts induced by model fitting were found distributed widely, especially in the AK and KFA. Therefore, the network would confuse these outliers with normal values, resulting in large variance in the grey matter after harmonization.

The second problem is that the image imperfections of the acquired DWIs would also potentially affect the data validity after harmonization, such as the image noise and distortion. First, the overall high noise of DWIs in each voxel together with its neighbors throughout all diffusion frames could slow down the reduction of the cost function within finite epochs during the training phase of the network, resulting in unstable weightings of the network. Subsequently in the testing phase, the outcomes would reflect this network's instability. Consequently, even with repeated training using the same subjects, the outcomes from the same test subject could still vary. In this study, the SNRs were high enough to produce results with low variance, and the variance caused by the network's instability was lower than that between repetition scans on the reference scanner. Second, when the image distortion appears differently between the target and the reference scanner, the data validity relies on the registration accuracy between the DWIs and DKI labels. Otherwise, the spatial mismatched between diffusion signal and DKI measures would be misled in the training phase and

impede the optimization of the neural network. For the Dataset 1 in this study, the image distortion from Scanner C was obvious without any preprocessing correction. Therefore, nonlinear registration was used to minimize the registration error between the Scanner C and the reference scanner D.

The last concern is that the size of the training dataset has an important impact on the effectiveness of the deep learning–based method. The outcome stability in the testing phase would improve if a large number of effective voxels were used in the training phase. Therefore, more traveling subjects should participate in the preparatory data collection of multicenter projects to obtain more training voxels. However, for the two datasets in the present study, the original coherence of the measures in Dataset 2 was higher than that in Dataset 1; however, the outcome stability of the harmonized measures was lower in Dataset 2. Moreover, the harmonization performance in four scanners in Dataset 1 varied even from same training data size. As a compromise between harmonization performance and collection costs, we used 40% of the subjects for training in this small-scale multi-scanner study. Nevertheless, the recommended number of traveling subjects is supposed to be identical among scanners for consistent harmonization outcomes, and it remains to be established in future multicenter studies.

5. Conclusions

A framework for harmonization of multicenter DKI measures was designed and evaluated for its feasibility on two datasets with different or identical protocols. The results demonstrated the apparent improvements in reliability among different scanners and for each scanner for both datasets. This approach of relatively low cost and high efficiency may provide practical guidance for the protocol setup and harmonization in future large-scale multicenter projects.

Acknowledgments

This work was supported by the National Key R&D Program of China [grant number 2017YFC0909200], National Natural Science Foundation of China [grant numbers 81871428, 91632109], Shanghai Key Laboratory of Psychotic Disorders [grant number 13dz2260500], Major Scientific Project of Zhejiang Lab [grant number 2018DG0ZX01], Fundamental Research Funds for the Central Universities [grant numbers 2019QNA5026, 2019XZZX001-01-08], and Zhejiang University Education Foundation Global Partnership Fund.

We acknowledge the great supports from the following institutions in collecting the Dataset 1 in this study: The Third Affiliated Hospital of Qiqihar Medical University in Qiqihar, Zhejiang Hospital in Hangzhou, and Chinese Academy of Sciences in Shanghai.

The Dataset 2 used in this study was acquired at the UK National Facility for In Vivo MR Imaging of Human Tissue Microstructure located in CUBRIC, funded by the EPSRC [grant number EP/M029778/1] and The Wolfson Foundation. Acquisition and processing of the data were supported by a Rubicon grant from the NWO [grant number 680-50-1527], a Wellcome Trust Investigator Award [grant number 096646/Z/11/Z), and a Wellcome Trust Strategic Award [grant number 104943/Z/14/Z]. This database was initiated by the 2017 and 2018 MICCAI Computational

Diffusion MRI committees (Chantal Tax, Francesco Grussu, Enrico Kaden, Lipeng Ning, Jelle Veraart, Elisenda Bonet-Carne, and Farshid Sepehrband) and CUBRIC, Cardiff University (Chantal Tax, Derek Jones, Umesh Rudrapatna, John Evans, Greg Parker, Slawomir Kusmia, Cyril Charron, and David Linden).

References

[1] Van Eijsden M, Vrijkotte TGM, Gemke RJBJ, Van der Wal MF. Cohort profile: The Amsterdam born children and their development (ABCD) study. Int J Epidemiol 2011;40:1176–86. https://doi.org/10.1093/ije/dyq128.

[2] Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, et al. The Human Connectome Project: A data acquisition perspective. Neuroimage 2012;62:2222–31. https://doi.org/10.1016/j.neuroimage.2012.02.018.

[3] Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Dev Cogn Neurosci 2018;32:43–54. https://doi.org/10.1016/j.dcn.2018.03.001.

[4] Sotiropoulos SN, Jbabdi S, Xu J, Andersson JL, Moeller S, Auerbach EJ, et al. Advances in diffusion MRI acquisition and processing in the Human Connectome Project. Neuroimage 2013;80:125–43. https://doi.org/10.1016/j.neuroimage.2013.05.057.

[5] Jernigan TL, Brown TT, Hagler DJ, Akshoomoff N, Bartsch H, Newman E, et al. The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. Neuroimage 2016;124:1149–54. https://doi.org/10.1016/j.neuroimage.2015.04.057.

[6] Gunter JL, Borowski BJ, Thostenson K, Arani A, Reid RI, Cash DM, et al. ADNI-3 MRI Acquisitions. Alzheimer's Dement 2017;13:P1368–9. https://doi.org/10.1016/j.jalz.2017.06.2110.

[7] Zuo XN, Xu T, Milham MP. Harnessing reliability for neuroscience research. Nat Hum Behav 2019;3. https://doi.org/10.1038/s41562-019-0655-x.

[8] Takao H, Hayashi N, Ohtomo K. Effects of study design in multi-scanner voxel-based morphometry studies. Neuroimage 2014;84:133–40. https://doi.org/10.1016/j.neuroimage.2013.08.046.

[9] Jovicich J, Marizzoni M, Sala-Llonch R, Bosch B, Bartrés-Faz D, Arnold J, et al. Brain morphometry reproducibility in multi-center 3T MRI studies: A comparison of cross-sectional and longitudinal segmentations. Neuroimage 2013;83:472–84. https://doi.org/10.1016/j.neuroimage.2013.05.007.

[10] Yuan LX, Wang JB, Zhao N, Li YY, Ma Y, Liu DQ, et al. Intra- and inter-scanner reliability of scaled subprofile model of principal component analysis on ALFF in Resting-State fMRI under eyes open and closed conditions. Front Neurosci 2018;12:311. https://doi.org/10.3389/fnins.2018.00311.

[11] Jovicich J, Minati L, Marizzoni M, Marchitelli R, Sala-Llonch R, Bartrés-Faz D, et al. Longitudinal reproducibility of default-mode network connectivity in healthy elderly participants: A multicentric resting-state fMRI study. Neuroimage 2016;124:442–54. https://doi.org/10.1016/j.neuroimage.2015.07.010.

[12] An HS, Moon WJ, Ryu JK, Park JY, Yun WS, Choi JW, et al. Inter-vender and test-retest reliabilities of resting-state functional magnetic resonance imaging: Implications for multi-center

imaging studies. Magn Reson Imaging 2017;44:125–30. https://doi.org/10.1016/j.mri.2017.09.001.

[13] Chuhutin A, Hansen B, Jespersen SN. Precision and accuracy of diffusion kurtosis estimation and the influence of b-value selection. NMR Biomed 2017;30:1–14. https://doi.org/10.1002/nbm.3777.

[14] Farrell JAD, Landman BA, Jones CK, Smith A, Prince JL, Zijl PCM Van, et al. Effects of SNR on the Accuracy and Reproducibility of DTI-derived Fractional Anisotropy, Mean Diffusivity, and Principal Eigenvector Measurements at 1.5T. J Magn Reson 2007;26:756–67. https://doi.org/10.1002/jmri.21053.Effects.

[15] Palacios EM, Martin AJ, Boss MA, Ezekiel F, Chang YS, Yuh EL, et al. Toward precision and reproducibility of diffusion tensor imaging: A multicenter diffusion phantom and traveling volunteer study. Am J Neuroradiol 2017;38:537–45. https://doi.org/10.3174/ajnr.A5025.

[16] Teipel SJ, Reuter S, Stieltjes B, Acosta-cabronero J, Ernemann U, Fellgiebel A, et al. Multicenter stability of diffusion tensor imaging measures: A European clinical and physical phantom study. Psychiatry Res - Neuroimaging 2011;194:363–71. https://doi.org/10.1016/j.pscychresns.2011.05.012.

[17] Zhou X, Sakaie KE, Debbins JP, Kirsch JE, Tatsuoka C, Fox RJ, et al. Quantitative quality assurance in a multicenter HARDI clinical trial at 3T. Magn Reson Imaging 2017;35:81–90. https://doi.org/10.1016/j.mri.2016.08.022.

[18] Tong Q, He H, Gong T, Li C, Liang P, Qian T, et al. Reproducibility of multi-shell diffusion tractography on traveling subjects: A multicenter study prospective. Magn Reson Imaging 2019;59:1–9. https://doi.org/10.1016/J.MRI.2019.02.011.

[19] Pohl KM, Sullivan E V., Rohlfing T, Chu W, Kwon D, Nichols BN, et al. Harmonizing DTI measurements across scanners to examine the development of white matter microstructure in 803 adolescents of the NCANDA study. Neuroimage 2016;130:194–213. https://doi.org/10.1016/j.neuroimage.2016.01.061.

[20] Fortin JP, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site diffusion tensor imaging data. Neuroimage 2017;161:149–70. https://doi.org/10.1016/j.neuroimage.2017.08.047.

[21] Salimi-khorshidi G, Smith SM, Keltner JR, Wager TD, Nichols TE. Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. Neuroimage 2009;45:810–23. https://doi.org/10.1016/j.neuroimage.2008.12.039.

[22] Venkatraman VK, Gonzalez CE, Landman B, Goh J, Reiter DA, An Y, et al. Region of interest correction factors improve reliability of diffusion imaging measures within and across scanners and field strengths. Neuroimage 2015;119:406–16. https://doi.org/10.1016/j.neuroimage.2015.06.078.

[23] Mirzaalian H, Ning L, Savadjiev P, Pasternak O, Bouix S, Michailovich O, et al. Multi-site harmonization of diffusion MRI data in a registration framework. Brain Imaging Behav 2016;12:284–95. https://doi.org/10.1007/s11682-016-9670-y.

[24] Mirzaalian H, Ning L, Savadjiev P, Pasternak O, Bouix S, Michailovich O, et al. Inter-site and inter-scanner diffusion MRI data harmonization. Neuroimage 2016;135:311–23. https://doi.org/10.1016/j.neuroimage.2016.04.041.

[25] Karayumak SC, Bouix S, Ning L, James A, Crow T, Shenton M, et al. Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters.

Neuroimage 2019;184:180–200. https://doi.org/10.1016/j.neuroimage.2018.08.073.

[26] Huynh KM, Chen G, Wu Y, Shen D, Yap P-T. Multi-Site Harmonization of Diffusion MRI Data via Method of Moments. IEEE Trans Med Imaging 2019;38:1599–609. https://doi.org/10.1109/TMI.2019.2895020.

[27] Koppers S, Haarburger C, Merhof D. Diffusion MRI Signal Augmentation: From Single Shell to Multi Shell with Deep Learning. Proc. MICCAI 2017 Comput. Diffus. MRI, Cham: 2017, p. 61–70. https://doi.org/10.1007/978-3-319-73839-0.

[28] Golkov V, Dosovitskiy A, Sperl JI, Menzel MI, Czisch M, Sämann P, et al. q-Space Deep Learning: Twelve-Fold Shorter and Model-Free Diffusion MRI Scans. IEEE Trans Med Imaging 2016;35:1344–51. https://doi.org/10.1109/TMI.2016.2551324.

[29] Gong T, He H, Li Z, Lin Z, Tong Q, Li C, et al. Efficient Reconstruction of Diffusion Kurtosis Imaging Based on a Hierarchical Convolutional Neural Network. Proc. Int. Soc. Magn. Reson. Med., 2018, p. 1653.

[30] Lin Z, Gong T, Wang K, Li Z, He H, Tong Q, et al. Fast learning of fiber orientation distribution function for MR tractography using convolutional neural network. Med Phys 2019:mp.13555. https://doi.org/10.1002/mp.13555.

[31] Nath V, Parvathaneni P, Hansen CB, Hainline AE, Bermudez C, Remedios S, et al. Inter-Scanner Harmonization of High Angular Resolution DW-MRI Using Null Space Deep Learning. Proc. MICCAI 2019 Comput. Diffus. MRI, Springer International Publishing; 2019, p. 173–82. https://doi.org/10.1007/978-3-030-05831-9.

[32] Nath V, Schilling KG, Parvathaneni P, Hansen CB, Hainline AE, Huo Y, et al. Deep learning reveals untapped information for local white-matter fiber reconstruction in diffusion-weighted MRI. Magn Reson Imaging 2019;62:220–7. https://doi.org/10.1016/j.mri.2019.07.012.

[33] Tax CM, Grussu F, Kaden E, Ning L, Rudrapatna U, John Evans C, et al. Cross-scanner and cross-protocol diffusion MRI data harmonisation: A benchmark database and evaluation of algorithms. Neuroimage 2019;195:285–99. https://doi.org/10.1016/j.neuroimage.2019.01.077.

[34] Koppers S, Bloy L, Berman JI, Tax CMW, Edgar JC, Merhof D. Spherical Harmonic Residual Network for Diffusion Signal Harmonization. Proc. MICCAI 2019 Comput. Diffus. MRI, Cham: 2019, p. 173–82.

[35] Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: The quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. Magn Reson Med 2005;53:1432–40. https://doi.org/10.1002/mrm.20508.

[36] Jensen JH, Helpern JA. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR Biomed 2010;23:698–710. https://doi.org/10.1002/nbm.1518.

[37] Jiang R, Jiang J, Zhao L, Zhang JJ, Zhang S, Yao Y, et al. Diffusion kurtosis imaging can efficiently assess the glioma grade and cellular proliferation. Oncotarget 2015;6:42380–93. https://doi.org/10.18632/oncotarget.5675.

[38] Raab P, Hattingen E, Franz K, Zanella FE, Lanfermann H. Cerebral Gliomas: Diffusional Kurtosis Imaging Analysis of Microstructural Differences. Radiology 2010;254:876–81. https://doi.org/10.1148/radiol.09090819.

[39] Cauter S Van, Veraart J, Sijbers J, Peeters RR, Himmelreich U, De Keyzer F, et al. Gliomas: Diffusion Kurtosis MR Imaging in Grading. Radiology 2012;263:492–501. https://doi.org/10.1148/radiol.12110927.

[40] Gao Y, Zhang Y, Wong CS, Wu PM, Zhang Z, Gao J, et al. Diffusion abnormalities in temporal

lobes of children with temporal lobe epilepsy: A preliminary diffusional kurtosis imaging study and comparison with diffusion tensor imaging. NMR Biomed 2012;25:1369–77. https://doi.org/10.1002/nbm.2809.

[41] Wang J, Lin W, Lu C, Ng S, Weng Y-H, Ng S, et al. Parkinson Disease: Diagnostic Utility of Diffusion Kurtosis Imaging. Radiology 2011;261:210–7. https://doi.org/10.1148/radiol.11102277/-/DC1.

[42] Grinberg F, Maximov II, Farrher E, Neuner I, Amort L, Thönneßen H, et al. Diffusion kurtosis metrics as biomarkers of microstructural development: A comparative study of a group of children and a group of adults. Neuroimage 2017;144:12–22. https://doi.org/10.1016/j.neuroimage.2016.08.033.

[43] Benitez A, Jensen JH, Falangola MF, Nietert PJ, Helpern JA. Modeling white matter tract integrity in aging with diffusional kurtosis imaging. Neurobiol Aging 2018;70:265–75. https://doi.org/10.1016/j.neurobiolaging.2018.07.006.

[44] Falangola MF, Jensen JH, Babb JS, Hu C, Castellanos FX, Di Martino A, et al. Age-related non-Gaussian diffusion patterns in the prefrontal brain. J Magn Reson Imaging 2008;28:1345–50. https://doi.org/10.1002/jmri.21604.

[45] Das SK, Wang JL, Bing L, Bhetuwal A, Yang HF. Regional Values of Diffusional Kurtosis Estimates in the Healthy Brain during Normal Aging. Clin Neuroradiol 2017;27:283–98. https://doi.org/10.1007/s00062-015-0490-z.

[46] Tabesh A, Jensen JH, Ardekani BA, Helpern JA. Estimation of tensors and tensor-derived measures in diffusional kurtosis imaging. Magn Reson Med 2011;65:823–36. https://doi.org/10.1002/mrm.22655.

[47] Koay CG, Carew JD, Alexander AL, Basser PJ, Meyerand ME. Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging. Magn Reson Med 2006;55:930–6. https://doi.org/10.1002/mrm.20832.

[48] Sprenger T, Sperl JI, Fernandez B, Golkov V, Eidner I, Sämann PG, et al. Bias and precision analysis of diffusional kurtosis imaging for different acquisition schemes. Magn Reson Med 2016;76:1684–96. https://doi.org/10.1002/mrm.26008.

[49] Veraart J, Novikov DS, Christiaens D, Ades-aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. Neuroimage 2016;142:394–406. https://doi.org/10.1016/j.neuroimage.2016.08.016.

[50] Kellner E, Dhital B, Kiselev VG, Reisert M. Gibbs-ringing artifact removal based on local subvoxel-shifts. Magn Reson Med 2016;76:1574–81. https://doi.org/10.1002/mrm.26054.

[51] Andersson JLR, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 2016;125:1063–78. https://doi.org/10.1016/j.neuroimage.2015.10.019.

[52] Andersson JLR, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. Neuroimage 2003;20:870–88. https://doi.org/10.1016/S1053-8119(03)00336-7.

[53] Ning L, Bonet-Carne E, Grussu F, Sepehrband F, Kaden E, Veraart J, et al. Muti-shell Diffusion MRI Harmonisation and Enhancement Challenge (MUSHAC): Progress and Results. Proc. MICCAI 2019 Comput. Diffus. MRI, Cham: 2019, p. 217–24. https://doi.org/10.1007/978-3-030-05831-9.

[54] Li Z, Gong T, Lin Z, He H, Tong Q, Li C, et al. Fast and robust diffusion kurtosis parametric

mapping using a three-dimensional convolutional neural network. IEEE Access 2019;7:71398–411. https://doi.org/10.1109/ACCESS.2019.2919241.

[55] Ades-Aron B, Veraart J, Kochunov P, McGuire S, Sherman P, Kellner E, et al. Evaluation of the accuracy and precision of the diffusion parameter EStImation with Gibbs and NoisE removal pipeline. Neuroimage 2018;183:532–43. https://doi.org/10.1016/j.neuroimage.2018.07.066.

[56] Collier Q, Veraart J, Jeurissen B, Den Dekker AJ, Sijbers J. Iterative reweighted linear least squares for accurate, fast, and robust estimation of diffusion magnetic resonance parameters. Magn Reson Med 2015;73:2174–84. https://doi.org/10.1002/mrm.25351.

[57] Andersson JLR, Jenkinson M, Smith S. Non-linear registration aka Spatial normalisation. 2007.

[58] Mori S, Oishi K, Jiang H, Jiang L, Li X, Akhter K, et al. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. Neuroimage 2008;40:570–82. https://doi.org/10.1016/j.neuroimage.2007.12.035.

[59] Celisse A, Robin S. Nonparametric density estimation by exact leave- p -out 2008;52:2350–68. https://doi.org/10.1016/j.csda.2007.10.002.

[60] Vollmar C, O'Muircheartaigh J, Barker GJ, Symms MR, Thompson P, Kumari V, et al. Identical, but not the same: Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3.0T scanners. Neuroimage 2010;51:1384–94. https://doi.org/10.1016/j.neuroimage.2010.03.046.

[61] Landman BA, Farrell JAD, Jones CK, Smith SA, Prince JL, Mori S. Effects of diffusion weighting schemes on the reproducibility of DTI-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. Neuroimage 2007;36:1123–38. https://doi.org/10.1016/j.neuroimage.2007.02.056.

[62] Hansen B, Lund TE, Sangill R, Jespersen SN. Experimentally and computationally fast method for estimation of a mean kurtosis. Magn Reson Med 2013;69:1754–60. https://doi.org/10.1002/mrm.24743.

[63] Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage 2012;61:1000–16. https://doi.org/10.1016/j.neuroimage.2012.03.072.

[64] Assaf Y, Basser PJ. Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. Neuroimage 2005;27:48–58. https://doi.org/10.1016/j.neuroimage.2005.03.042.

[65] Dyrby TB, Søgaard L V., Hall MG, Ptito M, Alexander DC. Contrast and stability of the axon diameter index from microstructure imaging with diffusion MRI. Magn Reson Med 2013;70:711–21. https://doi.org/10.1002/mrm.24501.

[66] Alexander DC, Hubbard PL, Hall MG, Moore EA, Ptito M, Parker GJM, et al. Orientationally invariant indices of axon diameter and density from diffusion MRI. Neuroimage 2010;52:1374–89. https://doi.org/10.1016/j.neuroimage.2010.05.043.