

Research Bank Book chapter

Acquiring and predicting multidimensional diffusion (MUDI) data : An open challenge

Pizzolato, Marco, Palombo, Marco, Bonet-Carne, Elisenda, Tax, Chantal M. W., Grussu, Francesco, Ianus, Andrada, Bogusz, Fabian, Pieciak, Tomasz, Ning, Lipeng, Larochelle, Hugo, Descoteaux, Maxime, Chamberland, Maxime, Blumberg, Stefano B., Mertzanidou, Thomy, Alexander, Daniel C., Afzali, Maryam, Aja-Fernández, Santiago, Jones, Derek K., Westin, Carl-Fredrik, Rathi, Yogesh, Baete, Steven H., Cordero-Grande, Lucilio, Ladner, Thilo, Slator, Paddy J., Hajnal, Joseph V., Thiran, Jean-Philippe, Price, Anthony N., Sepehrband, Farshid, Zhang, Fan and Hutter, Jana

This version of the book chapter has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <u>https://doi.org/10.1007/978-3-030-52893-5_17</u>

Acquiring and predicting multidimensional diffusion (MUDI) data: an open challenge

Marco Pizzolato^{1,2}, Marco Palombo³, Elisenda Bonet-Carne^{3,4}, Chantal M.W. Tax^{5,6}, Francesco Grussu^{3,7}, Andrada Ianus⁸, Fabian Bogusz⁹, Tomasz
Pieciak^{9,10}, Lipeng Ning¹¹, Hugo Larochelle¹², Maxime Descoteaux¹³, Maxime Chamberland¹⁴, Stefano B. Blumberg³, Thomy Mertzanidou³, Daniel C. Alexander³, Maryam Afzali¹⁴, Santiago Aja-Fernández¹⁰, Derek K. Jones^{14,15}, Carl-Fredrik Westin¹¹, Yogesh Rathi¹¹, Steven H. Baete^{16,17}, Lucilio Cordero-Grande¹⁸, Thilo Ladner¹⁹, Paddy J. Slator³, Joseph V Hajnal¹⁸, Jean-Philippe Thiran^{2,20}, Anthony N. Price¹⁸, Farshid Sepehrband²¹, Fan Zhang¹¹, and Jana Hutter¹⁸

¹Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark; ²Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; ³Centre for Medical Image Computing, Department of Computer Science, University College London, London, United Kingdom; ⁴BCNatal Fetal Medicine Research Center, Barcelona, Spain; ⁵Cardiff University Brain Research Imaging Centre (CUBRIC), School of Physics and Astronomy, Cardiff University, Cardiff, United Kingdom; ⁶University Medical Center Utrecht, Utrecht, The Netherlands; ⁷Queen Square MS Centre, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, United Kingdom; ⁸Champalimaud Research, Champalimaud Centre for the Unknown, Lisbon, Portugal; ⁹AGH University of Science and Technology, Kraków, Poland; ¹⁰Laboratorio de Procesado de Imagen (LPI), Universidad de Valladolid, Valladolid, Spain; ¹¹Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; ¹²Google Brain, Montréal, Canada; ¹³SCIL, Université de Sherbrooke, Sherbrooke, Canada; ¹⁴Cardiff University Brain Research Imaging Center (CUBRIC), School of Psychology, Cardiff University, Cardiff, United Kingdom; ¹⁵Mary MacKillop Institute for Health Research, Faculty of Health Sciences, Australian Catholic University, Melbourne, Australia; ¹⁶Center for Biomedical Imaging, Dept. of Radiology, New York University School of Medicine, New York, NY, United States; ¹⁷Center for Advanced Imaging Imaging, Innovation and Research, New York University School of Medicine, New York, NY, United States; ¹⁸Centre for Medical Engineering, Centre for the Developing Brain, King's College London, London, United Kingdom; ¹⁹Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland; ²⁰Radiology Department, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland; ²¹Laboratory of Neuro Imaging (LONI), USC Stevens Neuroimaging and Informatics Institute, University of Southern California, Los Angeles, CA, United States.

2 M. Pizzolato et al.

Abstract. In magnetic resonance imaging (MRI), the image contrast is the result of the subtle interaction between the physicochemical properties of the imaged living tissue and the parameters used for image acquisition. By varying parameters such as the echo time (TE) and the inversion time (TI), it is possible to collect images that capture different expressions of this sophisticated interaction. Sensitization to diffusion summarized by the b-value - constitutes yet another explorable "dimension" to modify the image contrast which reflects the degree of dispersion of water in various directions within the tissue microstructure. The full exploration of this multidimensional acquisition parameter space offers the promise of a more comprehensive description of the living tissue but at the expense of lengthy MRI acquisitions, often unfeasible in clinical practice. The harnessing of multidimensional information passes through the use of intelligent sampling strategies for reducing the amount of images to acquire, and the design of methods for exploiting the redundancy in such information. This chapter reports the results of the MUDI challenge, comparing different strategies for predicting the acquired densely sampled multidimensional data from sub-sampled versions of it.

Keywords: $MUDI \cdot diffusion \cdot relaxation \cdot Quantitative Imaging$

1 The multi-contrast nature of MRI images

The abundance of acquisition parameters available in magnetic resonance imaging (MRI), and the focus on the b-value (b) and gradient directions (dir) when performing a typical diffusion acquisition – for example a pulsed gradient spinecho experiment [1] – may lead one to consider parameters such as the echo time (TE) and the inversion time (TI) merely as indispensable yet irrelevant for the outcome of the acquisition. On the contrary, the choice of TE and TI is delicate as it attains to what component of the living tissue's physicochemical environment one would like to be the most dominant in the measured signal. Indeed, when the complexity of the tissue is summarized at the scale of an MRI voxel, the contributions from the different physicochemical environments add up in a distribution of longitudinal and transverse relaxation times T_1 and T_2 . Indicating with $\phi(TI, TE|T_1, T_2)$ the functional form of the contribution to the overall measured signal intensity, S(TI, TE), then

$$S(TI, TE) \propto \int_0^\infty \int_0^\infty P(T_1, T_2) \phi(TI, TE|T_1, T_2) dT_1 dT_2$$
 (1)

where $P(T_1, T_2)$ is the joint distribution of the T_1 and T_2 expressed within the voxel. From the equation it appears clear that the nature of the functional form ϕ determines the contribution of a specific (T_1, T_2) population in the overall measured signal S(TI, TE). For instance, focusing only on the dependency between TE and T_2 for the sake of simplification, in first approximation

$$\phi(TI, TE|T_1, T_2) \propto \exp(-TE/T_2) \tag{2}$$

3

which indicates that for a specific echo time TE a population having higher T_2 will have more contribution to the overall signal compared to an equivalently abundant population having a lower T_2 value being effectively less represented. This relaxation-weighted relative contribution is at the origin of an observability challenge in MRI: for instance, in order to acquire a signal that contains the contribution of water trapped within myelin sheaths in the nervous system, characterized by a short T_2 [2], it would be necessary to use a short echo time, according to eq. 2, which can pose technical challenges. Beyond observability, which attains to the technical issue of collecting data using the required values of the acquisition parameters for observing a particular population, the need for representing all the observable populations leads to a *sampling* problem. A viable strategy consists in collecting data with many different combinations of acquisition parameters, where a large coverage and dense sampling of the multidimensional parameter space is desirable. Following the example above, a naive implementation of such strategy would lead to a combinatorial explosion of all the possible feasible pairs (TI, TE) according to a convenient discretization of the corresponding bi-dimensional parameter space. However, as the number of dimensions increases a similar strategy is no longer a reasonable option for practical purposes related to the lengthening of the time required for acquiring the MRI data, and to the increasing capacity required for its storage. Additional dimensions come, for instance, from the use of diffusion sensitization in the acquisition – summarized here by the b-value and the gradient directions [1, 7] – leading to a parameter space defined by the space of (TI, TE, b, dir).

The design of an efficient sampling/sub-sampling strategy entails the adoption of two complementary approaches. The first one consists in developing accelerated MRI acquisition methods to reduce the overall acquisition time for a fixed sampling. These methods exploit acceleration opportunities on the engineering and physics side. They include a plurality of approaches like k-space sub-sampling, multi-band/multi-slice acquisition, parallel imaging, and others. An example of particular relevance for the parameter space (TI, TE, b, dir) are techniques combining the different contrasts in the same acquisition - thus effectively reducing the waiting time inserted into the sequence to achieve specific contrasts. One method among these is ZEBRA [3], which leverages ideas such as slice-shuffling [4] and multi-echo read-outs and adds diffusion-preparation changes on a slice level in order to efficiently acquire information related to the interplay of the diffusion process and the T_1 and T_2^* relaxation. Here, the usage of multiple gradient-echos leads to T_2^* weighting (the observed transverse relaxation time affected by the presence of magnetic field inhomogeneities). The second approach consists of designing a sampling strategy that allows the shortening of the acquisition time by reducing the required number of samples. Typically, a sampling strategy is associated with a corresponding reconstruction method that is capable of exploiting the information from the acquired samples. As different sampling and reconstruction designs have different performance it is important to rank them. This chapter reports the description of the MUltidimensional DIffusion (MUDI) challenge, organized within the MICCAI 2019 conference, with



Fig. 1. The organization of the data. The lines report the variation of the parameters (with arbitrary scale units) as a function of the volume number.

the purpose of evaluating the sampling and reconstruction designs on acquired *in vivo* data.

2 The challenge

In order to evaluate the performance of a particular design, comprising of sampling and reconstruction, it would be necessary to know a priori the underlying information of interest. This is however impossible in the case of acquired experimental data. This would indeed require knowing the biophysical properties of interest – the T_1 , T_2 , and the diffusion properties of each population – which would only be known through the use of the very reconstruction techniques to be ranked. Therefore, a viable paradigm for quantifying the performance of the various methods consists in assessing the power that each tested design has in predicting unseen data samples for which an actual ground truth is available. Indeed, if a design is capable of predicting missing signal samples from a subset of acquired ones then it fulfills a necessary condition for performing an efficient sampling and reconstruction. Following this criterion, designs can be ranked based on the capability of retrieving missing samples from a specified number of available ones.

2.1 Data

Five datasets were acquired from healthy human volunteers (3 f, 2 m, age=19-46 years), after informed consent was obtained (REC 12/LO/1247), on a clinical 3T Philips Achieva scanner (Best, Netherlands) with a 32-channel adult head coil. Each dataset includes 1344 volumes (Fig. 1) distributed over four b-shells, $b \in \{500, 1000, 2000, 3000\}$ s/mm², with 106 uniformly spread directions [5], three echo times $TE \in \{80, 105, 130\}$ ms, and 28 inversion times $TI \in [20, 7322]$ ms. Single-shot PGSE EPI with the modifications proposed in ZEBRA [3] sequence was employed. Other parameters include TR=7.5s, resolution=2.5mm isotropic, FOV=220x230x140mm, SENSE=1.9, halfscan=0.7, multiband factor 2, total acquisition time 52min (including preparation time).

Data was reconstructed and denoised in the complex domain [6]. Ad hoc affine volume registration was performed: collinear magnitude diffusion-weighted images (DWIs) acquired with different pairs (TI,TE) were first co-registered together using the highest TI and lowest TE volume as reference; the 106 reference volumes were then registered together based on a mutual information metric using Dipy [8], and the registrations were then propagated across the corresponding collinear DWIs. Data was collected with reversed phase-encode blips to allow for susceptibility-induced distortion correction using FSL [9, 10].

2.2 Tasks

Three fully-sampled datasets (1344 samples) were given to participants as training data, such that they could use them to propose a sub-sampling strategy based on 500 samples/volumes. In a second phase, participants were asked to provide the indexes of the desired 500 volumes, where each volume had a different set (TI,TE,b,dir). Then, only those selected 500 volumes for two unseen test subjects were provided to each participant. Finally, participants were asked to submit their prediction of the remaining 1344 – 500 volumes. Additionally, participants were asked to select X = 250, 100, and 50 volumes as subsets of the previously chosen 500 volumes subset, and again predict the 1344 – X remaining samples.

2.3 Evaluation

The capability of retrieving missing samples from each specified subset of available samples – of the test subjects – was ranked using the mean squared error (MSE) between the predicted volumes and the corresponding acquired ones that were unknown to the participant. A comparative analysis of the MSE on the whole brain (WB), white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) masks was also performed in order to characterize the behavior of the various designs. Tissue segmentation was performed using SPM12 [11] on the T₁-weighted image. The output included three probabilistic segmentation maps – WM, GM, and CSF – each subsequently down-sampled to match the resolution of the MUDI data. Only voxels with a probability over 0.9 were included. Although a ranking for each tissue type could be performed, the overall ranking accounts the MSE results over the whole brain.

3 Proposed methods

This report evaluates five submissions from different groups. Fig. 2 reports charts illustrating the characteristics of the submitted sub-sampling strategies with respect to those of the fully-sampled datasets. A description follows.

1. S1. The best subset of 500 samples was identified using an autoencoder neural network having a 'concrete selector layer' as first layer [12, 13]. A



Fig. 2. The sub-sampling strategies adopted in the five different submissions for the different number of samples used to make predictions. Each radar chart (except the last one on the right) reports four polygons each corresponding to the 500 (polygon with largest area), 250, 100, and 50 (polygon with smallest area) samples cases. Polygons are overlapped using transparency, therefore the most opaque region represents the intersection between the four cases. The charts report the mean TI, b, and TE, and the number of different directions "n_dir", and of TIs "n_TI" used in the sub-sampling. The dashed, black line pentagon "gt" (ground truth) reports the values of the fully-sampled dataset.

temperature associated to such layer was minimized during training [14] for 800 epochs; a linear decoder was employed to reconstruct the features (samples) from the selected ones; the procedure was used recursively for the other tasks i.e. selecting 250, 100, and 50 samples from the subset of 500 while every time predicting all 1344 features. After the selection process, five networks were trained to predict the full set of 1344 features based on leave-one-out cross-validation over the five training subjects. Each network had two hidden layers of 800 and 1000 nodes respectively, and Leaky ReLu as activation function, using a MSE loss. The average prediction from these networks was used for the test subjects.

- 2. S2. This is a multilayer, feedforward, fully-connected deep neural network that analyzes MRI data on a voxel-by-voxel basis. Two separate and sequential sub-networks, a selector and a predictor, work together to find the optimal subset of 500 samples and predict the remaining ones; as for S1, the 250, 100, and 50 samples tasks are managed recursively and samples extracted from the previously selected 500-measurement set. During training [14], the loss function was the MSE between the predicted and ground truth signals. 20% of the training data's voxels were used for validation, with the remaining 80% used for the actual learning via backpropagation.
- 3. S3. The submission is based on a representation of the diffusion and relaxation signal. For the relaxation part a single compartment was considered, i.e. a single pair (T_1, T_2) . Diffusion was represented through an analytical distribution of diffusivities leading to the inclusion of a Kurtosis term [15]. T_1, T_2 , proton density, inversion efficiency, diffusivity, and Kurtosis were estimated with a non-linear method. The prediction was based on extrapolating the relevant signal from the estimated parameters, the signal along different gradient directions being predicted with interpolation. Samples were selected by maximizing the number of included TIs while eventually penalizing the number of directions considered.

- 4. S4. This submission is based on a signal representation similarly to S3. In this case, however, the diffusion part has been represented using DTI [7]. In this case, the inversion efficiency was set to 2. The samples selection was based on the results of S2.
- 5. S5. This is a signal representation approach based on modeling the moments of the joint relaxation and diffusion probability distribution [16]. The 500 samples were selected to include 20 volumes with b = 0 and different values of TI and TE, plus 480 volumes along 30 gradient directions with $b \in \{1000, 3000\}$ s/mm². The directional signal for a given set (TI, TE, b)was represented using the spherical ridgelet functions [17]. No submissions to the 250, 100, and 50 samples cases were provided.

4 Preliminary results

Fig. 3 reports the MSE maps obtained by calculating the MSE values voxel-byvoxel for the four prediction cases and for all the submissions. The contrast of the images is therefore informative of the local performance of each proposed design, where a bright color indicates a larger error. It is therefore possible from these images to deduce that the submissions have different regional performance trends. Submissions 2 and 3, for instance, visually perform better in GM than in WM, whereas the opposite is true for submission 4. Submission 5 instead, reveals a substantially uniform performance across these two tissue types. A structured MSE map is visible also for submission 1 although less noticeable.

The differences across the different participant groups are attributed to both the proposed sub-sampling strategy and the reconstruction method. Regarding to the sub-sampling, all groups favored exploiting the redundancy in the diffusion gradient directions as illustrated in Fig. 2. Indeed, in the progressive reduction of the available samples illustrated by the reduced area of the radar charts (from more to less transparent due to the overlap), the "n_dir" entry corresponds to that displaying the largest decrease compared to the original sampling (dashed black line). In submissions 2 to 4 the signal is sub-sampled such that the mean TI in the dataset is kept substantially unaltered compared to the original sampling (this is marginally true for submission 1). However, submissions 1, 2, and 4 reduced of about one third the total number of unique TIs in the sub-sampling perhaps exploiting the redundancy of such a parameter space - whereas submission 3 included almost all available TI_s . To compensate for this, in submission 3 the mean b-value was much lower and slightly fewer gradient directions were selected. All these submissions kept the mean TE unaltered compared to the original sampling. Submission 5, on the other hand, reduced the mean TE in the sampling as well as the number of directions. This was likely a requirement due to the fitting procedure and signal representation they employed.

MSE values increased with the reduced number of samples used for the prediction, with submissions 1, 2, and 3 having overall the best three rankings. This is illustrated in the images of Fig. 5 and better quantified in summary results reported in the bar plots of Fig. 4. These results are calculated for the two test



Fig. 3. Voxel-wise MSE values for testsbj0002.



Fig. 4. Mean MSE per region of interest: white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), and for the whole brain (WB). Rows report results for the 500, 250, 100, and 50 samples cases.



Fig. 5. Progression of the MSE as the number of samples available for prediction decreases. Lines report the trend of the average MSE in the corresponding region (WM, GM, CSF, WB). Shadowed areas are proportional to the standard deviation.

subjects and for the various regions of interest. From these, it is possible to observe that submissions 3 and 4 suffered more the extreme reduction of available samples from 250 to 100 compared to submissions 1 and 2 (submission 5 was only applied to the 500 samples case). Indeed, the submissions based on machine learning report a superior stability and an overall lower MSE in all the regions and sub-sampling configurations. All methods revealed a performance loss, with different degrees, in CSF and cerebellar regions. A summary of the results is reported in Table 1.

5 Discussion

Results indicate that submission 1, based on machine learning, shows the highest prediction accuracy and stability among those submitted, followed by the other machine learning method submission 2. The challenge was deliberately designed in order for participants to be free in the choice of the samples such that this can best couple in synergy with the proposed reconstruction method. Because of this, however, it is impossible to separate the influence of the sampling strategy from the reconstruction method used. Nevertheless, it is possible that the proposed modeling-based reconstruction methods, which are primarily designed to enable more explicit descriptions of the physical phenomenon rather than signal reconstruction, might be affected by the difficulty of balancing model complexity and prediction performance. For instance, assuming one relaxation compartment (submissions 3 and 4) can oversimplify the fit while including more variability (submission 5) can destabilize it. Submissions 1 and 2, on the other hand, do not aim at an explicit representation of the physical parameters and for this reason they can rely on more flexible mathematical relationships to predict the signal. These reasons might contribute to an advantage of machine learning methods for tasks such as signal prediction. Moreover, a similar method can learn highly non-linear relationship between the reconstructions obtained from the different sub-sampling tasks thus obtaining a higher stability as the number of available

		S1	S2	S3	S4	$\mathbf{S5}$
WM	500	0.94, 0.56, 1.26	1.79, 1.45, 2.48	4.10, 2.98, 6.22	4.23, 1.67,3.85	5.74, 3.54, 7.38
		0.81, 0.53, 1.14	1.52, 1.28, 2.02	3.52, 2.83, 5.24	2.27, 1.5, 3.06	4.23, 3.28, 6.12
	250	1.12, 0.65, 1.46	3.15, 2.52, 4.72	4.2, 3.05, 6.36	4.46, 2.13, 5.24	-
		0.91, 0.61, 1.25	2.59, 2.20, 3.74	3.58, 2.90, 5.33	2.92, 1.92, 4.32	-
	100	1.15,0.71,1.53	3.44, 2.76, 5.15	4.56, 3.26, 6.88	6.32, 2.47, 5.93	-
		1.01,0.68,1.40	2.83, 2.39, 4.16	3.83, 3.09, 5.71	3.75, 2.23, 4.87	-
	50	1.3, 0.81, 1.74	2.81, 2.33, 4.04	4.55, 3.3, 6.88	6.16, 2.67, 6.41	-
		1.12,0.78,1.52	2.30, 1.98, 3.08	4.05, 3.3, 5.99	3.48, 2.35, 4.94	-
GM	500	1.22, 0.79, 1.96	1.48, 1.21, 2.49	2.99, 2.38, 4.92	6.86, 2.49, 9.69	6.8, 4.35, 10.44
		1.02, 0.64, 1.69	1.26, 0.97, 2.19	2.57, 2.09, 4.42	3.91, 1.91, 5.88	5.13, 3.78, 8.01
	250	1.41,0.92,2.28	2.26, 1.83, 3.91	3.14, 2.48, 5.12	5.87, 2.91, 10.71	-
		1.15, 0.74, 1.90	1.86, 1.50, 3.20	2.66, 2.16, 4.55	3.83, 2.18, 6.61	-
	100	1.43, 0.97, 2.29	2.29, 1.83, 3.95	3.46, 2.63, 5.55	8.08, 3.14, 12.96	-
		1.22, 0.80, 2.03	1.89, 1.49, 3.36	2.86, 2.28, 4.85	5.34, 2.37, 8.06	-
	50	1.58, 1.07, 2.54	2.44, 2.05, 4.07	3.38, 2.67, 5.40	7.23, 3.62, 12.69	-
		1.32, 0.87, 2.20	2.08, 1.74, 3.66	3.07, 2.5, 5.10	4.61, 2.63, 7.8	-
CSF	500	2.02, 1.17, 2.98	1.66, 1.28, 2.59	3.04, 2.38, 5.36	19.57, 7.93, 22.91	14.95, 8.17, 24.07
		1.31, 0.64, 1.88	1.00, 0.68, 1.70	1.78, 1.23, 3.31	8.26, 2.97, 10.26	6.82, 3.78, 10.81
	250	2.48, 1.46, 3.6	2.38, 1.88, 3.75	3.23, 2.52, 5.75	12.72, 8.4, 21.56	-
		1.52, 0.80, 2.17	1.48, 1.04, 2.62	1.86, 1.28, 3.46	6.13, 3.19, 10.93	-
	100	2.33, 1.42, 3.46	2.24, 1.77, 3.62	3.82, 2.84, 6.93	26.33, 13.57, 55.04	-
		1.52, 0.81, 2.21	1.42, 0.90, 2.54	2.22, 1.49, 4.1	11.23, 4.65, 19.94	-
	50	2.62, 1.63, 3.88	2.69, 2.05, 4.29	2.99, 2.44, 4.92	18.04, 10.34, 26.41	-
		1.58, 0.87, 2.33	1.67, 1.18, 2.9	2.07, 1.56, 3.71	8.34, 4.19, 13.39	-
WB	500	1.15, 0.69, 1.81	1.54, 1.23, 2.40	3.15, 2.37, 5.14	8.07, 2.26, 10.94	7.53, 4.03, 11.06
		0.89, 0.55, 1.44	1.24, 0.97, 1.96	2.58, 2.04, 4.37	4.05, 1.72, 5.81	4.88, 3.38, 7.43
	250	1.39, 0.82, 2.17	2.44, 1.92, 4.01	3.28, 2.47, 5.36	6.48, 2.75, 11.75	-
		1.02, 0.64, 1.64	1.94, 1.54, 3.24	2.66, 2.1, 4.48	3.88, 2.09, 6.62	-
	100	1.37, 0.85, 2.15	2.52, 1.94, 4.17	3.66, 2.65, 6.00	10.82, 3.15, 17.76	-
		1.09, 0.70, 1.74	2.01, 1.53, 3.49	2.91, 2.26, 4.86	5.86, 2.43, 9.03	-
	50	1.55, 0.96, 2.44	2.47, 2.02, 3.92	3.45, 2.61, 5.48	8.73, 3.43, 14.01	-
		1.18,0.78,1.88	1.98, 1.65, 3.19	3.04, 2.43, 5.00	4.89, 2.57, 7.91	-

Table 1. Mean, median, and 85th percentile of the MSE for each submission. For each sub-sampling case (500, 250, 100, and 50) the first raw corresponds to testsbj0001 and the second to testsbj0002.

12 M. Pizzolato et al.

samples for the prediction decreases: the proposed architectures in submissions 1 and 2 were flexible enough to interpret the 250, 100, and 50 samples predictions recursively. Among the modeling-based reconstruction methods, submissions 3 and 4 perform similarly when considering the whole brain. However, while submission 3 – modeling the diffusion kurtosis – seems more suitable for gray matter, submission 4 – based on DTI – outperforms it in white matter. Submission 5, although being the most comprehensive from a point of view of modeling as it describes the distribution of the parameters space (TI, TE, b, dir) within each voxel, has the highest MSE. This, however, could be mainly related to the chosen sub-sampling strategy which stands apart from the others.

Interestingly, the regional dependency of the performance of each submission suggests that some brain regions are more difficult to predict than others. While this might seem obvious, less intuitive is to grasp the reason behind such a regional trend. The cause of such regional performance differences is most likely connected with the reconstruction method, i.e. to the chosen type of modeling, although different brain regions might also require different sampling. With this regard, however, it is surprising to note similar regional trends also with submissions 1 and 2 as they do not involve an explicit modeling of the voxel' signal. Nevertheless, these regional performance differences underline the need for better descriptions of the MRI signal.

6 Conclusion

The MUDI challenge aimed at prompting the community to propose methods for exploiting the richness of information of multidimensional MRI acquisitions while at the same time suggesting sub-sampling strategies to allow for future clinical use. In this challenge, the sub-sampling and prediction tasks proposed were decoupled from the physical considerations connected to the MRI acquisition; indeed, every sample was considered to have the same weight in terms of acquisition time. Nevertheless, the results clearly point towards the direction to explore for future studies where the goal is to achieve MUDI data analysis within clinically feasible times. The machine learning methods indeed seem to provide greater prediction power and stability to sub-sampling compared to the signal representations. Perhaps the future will see the synergy between these two worlds where the former methods will provide the data redundancy necessary for using the latter methods as powerful and physically-informed tools for data analysis.

7 Acknowledgments

MPiz acknowledges support from European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754462. EB-C is partially supported by the Cerebra Foundation for the Brain Injured Child (Carmarthen, Wales, UK). CMWT is supported by a Veni grant (17331) from the Dutch Research Council (NWO) and a Sir Henry Wellcome Fellowship (215944/Z/19/Z). FB and TP acknowledge AGH Science and Technology, Kraków, Poland (16.16.120.773). MA and DKJ were supported by a Wellcome Trust Investigator Award (096646/Z/11/Z) and a Wellcome Trust Strategic Award (104943/Z/14/Z). SHB is supported in part by the National Institutes of Health (NIH, R01-CA111996, R01-NS082436, R01-MH00380 and P41EB017183). MPal, FG, DCA, TM and SBB acknowledge support from the UK EPSRC (EP/M020533/1, EP/N018702/1, EP/R006032/1), EU Horizon 2020 (ID 634541), NIH (Placenta imaging Project); Grant number: 1U01HD087202-01. AI acknowledges support from the Champalimaud Centre for the Unknown. LN is supported in part by NIH grants R21MH116352, R21MH115280, K01MH11 7346. SA-F's work was supported by Ministerio de Ciencia e Innovación of Spain with research grant RTI2018-094569-B-I00; FZ is supported by the following NIH grants: P41EB015898, R01MH108574, P41EB015902, R01MH 119222. JH was supported by the Wellcome Trust (Sir Henry Wellcome Fellowship, [201374/Z/16/Z] [201374/Z/16/B], and the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z].

References

- 1. Stejskal, E. O., Tanner, J. E. (1965). Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. The journal of chemical physics, 42(1), 288-292.
- MacKay, A. L., Laule, C. (2016). Magnetic resonance of myelin water: an in vivo marker for myelin. Brain plasticity, 2(1), 71-91.
- Hutter, J., Slator, P. J., Christiaens, D., Teixeira, R. P. A., Roberts, T., Jackson, L., Price, A. N., Malik, S., Hajnal, J. V. (2018). Integrated and efficient diffusionrelaxometry using ZEBRA. Scientific reports, 8(1), 1-13.
- Ordidge, R. J., Gibbs, P., Chapman, B., Stehling, M. K., Mansfield, P. (1990). High-speed multislice T1 mapping using inversion-recovery echo-planar imaging. Magnetic resonance in medicine, 16(2), 238-245.
- Jones, D. K., Horsfield, M. A., Simmons, A. (1999). Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 42(3), 515-525.
- Cordero-Grande, L., Christiaens, D., Hutter, J., Price, A. N., Hajnal, J. V. (2019). Complex diffusion-weighted image estimation via matrix recovery under general noise models. NeuroImage, 200, 391-404.
- Basser, P. J., Mattiello, J., LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. Biophysical journal, 66(1), 259-267.
- Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., Van Der Walt, S., Descoteaux, M., Nimmo-Smith, I. (2014). Dipy, a library for the analysis of diffusion MRI data. Frontiers in neuroinformatics, 8, 8.
- Andersson, J. L., Skare, S., Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage, 20(2), 870-888.
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Sefano, N., Brady, J. M.,

14 M. Pizzolato et al.

Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage, 23, S208-S219.

- Ashburner, J., Friston, K. J. (2000). Voxel-based morphometry—the methods. Neuroimage, 11(6), 805-821.
- 12. Abid, A., Balin, M. F., Zou, J. (2019). Concrete autoencoders for differentiable feature selection and reconstruction. arXiv preprint arXiv:1901.09346.
- 13. Maddison, C. J., Mnih, A., Teh, Y. W. (2016). The concrete distribution: A continuous relaxation of discrete random variables. arXiv preprint arXiv:1611.00712.
- Kingma, D.P., Ba, J.L., (2015). Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. International Conference on Learning Representations, ICLR.
- Jensen, J. H., Helpern, J. A. (2010). MRI quantification of non-Gaussian water diffusion by kurtosis analysis. NMR in Biomedicine, 23(7), 698-710.
- Ning, L., Gagoski, B., Szczepankiewicz, F., Westin, C. F., Rathi, Y. (2019). Joint RElaxation-Diffusion Imaging Moments (REDIM) to probe neurite microstructure. IEEE transactions on medical imaging.
- 17. Michailovich, O., Rathi, Y., Dolui, S. (2011). Spatially regularized compressed sensing for high angular resolution diffusion imaging. IEEE transactions on medical imaging, 30(5), 1100-1115.