10.1148/ryai.2019190019

Appendix E1

In the following, we present the main technical aspects of our deep learning method for time-to-maximum of tissue residue function (Tmax) parameter regression from DSC-MRI data. The proposed methodology, network architecture, and experimental results of an ablation study of key model components corresponds to the material presented in a previous technical publication (19). The main motivation behind our methodology is to replace the two step procedure consisting of Arterial Input Function (AIF) detection and subsequent deconvolution to obtain time-to-maximum of tissue residue function (Tmax) values with a direct regression of Tmax values from the raw DSC-MRI perfusion data.

Preprocessing

The DSC-MRI raw data and the target Tmax perfusion maps were both preprocessed. First, all patient data were padded in the spatial as well as temporal domain to match the maximum size of any volume in the training dataset. The resulting Tmax perfusion maps were of size 24 × 256 × 256 and the DSC-MRI raw data of size 80 × 24 × 256 × 256, where 80 indicates the number of acquisitions over time. Second, the data were standardized such that it has zero mean and unit variance. Third, the DSC-MRI raw data in the training dataset was augmented by randomly offsetting the respective perfusion sequence by -5 to 30 frames. By doing so, the deep learning model becomes invariant to different bolus arrival times.

Deep Learning Architecture for Perfusion Regression

To perform the regression of Tmax values from DSC-MRI raw data, we employed a Convolutional Neural Network (CNN). An overview of the methodology is presented in Figure E4. The regression pipeline contains the following key components:

- Bolus characterization: An image patch located at the transition between the basilar artery and the posterior cerebral artery is fed to two 3D convolutional layers encoding each patch into a vector of size 16. This serves as an approximation of the AIF. The location of the bolus characterization region is globally fixed and not optimized for individual subjects.
- Sequence encoding: The sequence encoder processes individual voxels in a concurrent fashion. It is provided with three inputs, which are the output of the bolus characterization component, the temporal sequence of voxel intensity values (for 80 frames) and the sequence of frame times. The inputs are processed by three 1D convolutional layers producing an output vector of size 256, which captures the voxel-wise signal evolution over time.
- Spatial correlation and regression: A 2D convolution is applied on axial slices to spatially correlate the sequence data. The architecture used for the spatial correlation is presented in Figure E5. The output of the spatial correlation is then processed by a fully connected layer with two output neurons and identity activation providing the final estimate of the Tmax value in form of a mean value and corresponding variance.
- Loss weighting: The employed objective function is based on heteroscedastic aleatoric uncertainty modeling (35), in which the estimated value for Tmax corresponds to a mean and

is accompanied by a corresponding estimate of its variance. The loss function is a weighted version of the negative log-likelihood of a Laplace distribution, which penalizes large positive outliers as well as negative values that are physiologically not plausible. The target Tmax map, which used for loss computation, is provided by the oSVD method as implemented in Olea Sphere.

The method was trained on 76 patients using Adam optimizer (36) with an initial learning rate of $5e^{-4}$. Every four epochs the learning rate was divided by two. Furthermore, fully connected layers were regularized using dropout with a rate of 0.5. A Windows machine with an Intel Xeon E5–1630 v3 CPU with 3.7 GHz and an Nvidia GTX 1080, and an Ubuntu machine with an Intel i7–4790 K CPU with 4 GHz and two Nvidia GTX 1070 were used to perform training. All candidate models were trained for 30 epochs and the final model was selected based on the regression performance on the validation cohort (30 patients).

Regression Performance

In the following, the regression performance of the proposed model is presented in terms of the mean absolute error with clipping. The clipping was performed for values below 0 seconds and above 20 seconds. The reasoning behind the clipping is that values below 0 correspond to air and above 20 are largely noise.

In Table E2, the performance of the full model as well as an ablation study, for which one of the key components has been removed, is presented. After inspection of the perfusion data, we found that the standard deviation of bolus arrival times is 2.76 frames in the validation set and 2.09 frames in the test set. Therefore, the model without data augmentation reached a slightly smaller mean absolute error on the test cohort than the full model. To further investigate the benefits of the data augmentation, the mean absolute error was measured on the validation cohort while shifting the perfusion sequence by a number of frames. In Figure E6, it is evident that the model trained with data augmentation can compensate better for different bolus arrival times than a model trained without data augmentation. Additionally, the removal of the spatial correlation, bolus characterization or loss weighting led to increases in mean absolute error, which is demonstrated in Table E2.

Discussion

The proposed method is based on the simplifying assumption that values of the voxels in the Tmax map depend on the perfusion sequence of voxels at the same location (for the one-dimensional convolutions) and neighborhood (for the two-dimensional convolutions). This assumption does not hold for cases in which significant patient motion occurred. An obvious solution would be to first coregister the different image volumes of the DSC-MRI sequence. However, given the large slice thickness of the imaging data (6 to 6.5 mm) this can potentially introduce strong interpolation artifacts. Moreover, a coregistration step independent of the subsequent deep learning-based regression does not fit the concept of end-to-end learning. A more consistent way of solving this issue would be an integration of the coregistration step within the deep learning framework. Recently, a number of approaches using neural networks to perform image registration have been proposed (37) and their integration with our proposed method is subject to future work.

Table E1: Complete data for DEFUSE 3 eligibility criteria of patients (*n* = 45) in test cohort.

		Rater 1							Rater 2						
Patient #	DWI <70 mL	Ratio >1.8 (CNN)	Diff. >15 mL (CNN)	Eligibility (CNN)	Ratio >1.8 (oSVD)	Diff. >15 mL (oSVD)	Eligibility (oSVD)	Disagreement (CNN vs oSVD)	Ratio >1.8 (CNN)	Diff. >15 mL (CNN)	Eligibility (CNN)	Ratio >1.8 (oSVD)	Diff. >15 mL (oSVD)	Eligibility (oSVD)	Disagreement (CNN vs oSVD)
64	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
75	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
85	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
87	1	1	0	NO	1	1	YES	**YES**	1	0	NO	1	1	YES	**YES**
90	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
95	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
100	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
101	1	1	1	YES	1	0	NO	**YES**	1	1	YES	1	1	YES	NO
109	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
116	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
123	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
141	0	1	1	NO	1	1	NO	NO	1	1	NO	1	1	NO	NO
146	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
154	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
157	0	0	1	NO	0	1	NO	NO	0	1	NO	0	1	NO	NO
169	1	0	0	NO	1	1	YES	**YES**	0	0	NO	1	1	YES	**YES**
172	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
181	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
190	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
194	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
204	1	0	0	NO	0	0	NO	NO	0	0	NO	0	0	NO	NO
221	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
222	1	1	0	NO	1	1	YES	**YES**	1	0	NO	1	1	YES	**YES**
233	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
240	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
247	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
250	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
251	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
255	0	1	1	NO	1	1	NO	NO	1	1	NO	1	1	NO	NO
256	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
262	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO

266	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
288	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
296	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
299	1	0	1	NO	0	1	NO	NO	0	1	NO	0	1	NO	NO
302	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
303	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
316	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
323	1	0	1	NO	1	1	YES	**YES**	0	1	NO	1	1	YES	**YES**
376	1	0	0	NO	0	0	NO	NO	1	0	NO	1	1	YES	**YES**
383	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
426	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
431	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
439	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
448	1	1	1	YES	1	1	YES	NO	1	1	YES	1	1	YES	NO
Total:	**42**	**39**	**40**	**35**	**41**	**42**	**38**	**5**	**40**	**40**	**35**	**42**	**44**	**40**	**5**

The values 1 and 0 indicate if a particular DEFUSE 3 criteria is satisfied (= 1) or not (= 0). The total number of patients satisfying a particular criteria as well as DEFUSE 3 eligibility are reported in the last row. Disagreement between eligibility criteria derived from the CNN or oSVD method are reported for both raters as well. CNN = Convolutional Neural Network, oSVD = oscillation index singular value decomposition.

Table E2: Ablation study results for Tmax value regression on validation and test cohort in mean absolute error with clipping reported in seconds.

Model	MAEC on validation cohort (seconds)	MAEC on test cohort (seconds)
Full model	**0.513**	0.530
Model without Augmentation	0.531	**0.524**
Model without Spatial Correlation	0.562	0.629
Model without Bolus Characterization	0.632	0.680
Model without Loss Weighting	0.683	0.738

Tmax = Time to maximum of the tissue residue function, MAEC = Mean absolute error with clipping.