Simulated perfusion MRI data to boost training of convolutional neural networks for lesion fate prediction in acute stroke

Noelie Debs, Pejman Rasti, Léon Victor, Tae-Hee Cho, Carole Frindel, David Rousseau

Simulated perfusion MRI data to boost training of convolutional neural networks for lesion fate prediction in acute stroke

Noëlie DEBS[a], Pejman RASTI[b], Léon VICTOR[a], Tae-Hee CHO[a], Carole FRINDEL[a], David ROUSSEAU[b]

[a] *CREATIS, CNRS UMR-5220, INSERM U1206, Université Lyon 1, INSA Lyon Bât. Blaise Pascal, 7 avenue Jean Capelle, 69621 Villeurbanne, France*
[b] *LARIS, UMR IRHS INRA, Université d'Angers 62 avenue Notre Dame du Lac, 49000 Angers, France*

Abstract

The problem of final tissue outcome prediction of acute ischemic stroke is assessed from physically realistic simulated perfusion magnetic resonance images. Different types of simulations with a focus on the arterial input function are discussed. These simulated perfusion magnetic resonance images are fed to convolutional neural network to predict real patients. Performances close to the state-of-the-art performances are obtained with a patient specific approach. This approach consists in training a model only from simulated images tuned to the arterial input function of a tested real patient. This demonstrates the added value of physically realistic simulated images to predict the final infarct from perfusion.

Keywords: Stroke, Lesion Prediction, Perfusion Magnetic Resonance Imaging, Arterial Input Function, Simulation, Convolutional Neural Network

1. Introduction

Stroke is a major cause of mortality and disability in the world (Murray et al., 2015). Stroke is divided into ischemic (85%) and hemorrhagic types (15%) (Ca-

*Corresponding author
Email address: david.rousseau@univ-angers.fr (David ROUSSEAU)

plan, 2016). Ischemia occurs when a cerebral artery is occluded (Park, 2017). Neuroimaging in acute stoke aims to obtain rapid information on tissue and vessel status to aid acute stroke intervention (Sartor, 2006). Diagnosis obtained from modern neuroimaging modalities enables efficient management of ischemic stroke and decide wether patient may benefits from intravenous thrombolysis or mechanical thrombectomy (Goyal et al., 2015; Albers et al., 2018). The most common neuroimaging, due to its widespread immediate availability, is computed tomography (CT), which is used in the initial diagnosis to determine the type of stroke (ischemic or hemorrhagic) (Zerna et al., 2016). Magnetic resonance imaging (MRI) may be substituted for CT as it becomes more readily available and it provides greater physiological information on soft tissues. MRI imaging for acute stroke include diffusion-weighted imaging (DWI) and perfusion-weighted imaging (PWI)

Still, developing automated methods to predict the extent of the stroke lesion from MRI scans remains an open challenge (Maier et al., 2017). This prediction has been mainly addressed so far with thresholded hemodynamic biomarkers based on kinetic models (Christensen and Lansberg, 2018; Albers, 2018; Najm et al., 2018). However, given the high dimensionality of PWI, machine learning techniques (Zhang and Sejdić, 2019) have also been successfully proposed in recent years (Maier et al., 2015; Stier et al., 2015; Nielsen et al., 2018; Zhang et al., 2018; Lucas et al., 2018; Pedemonte et al., 2018; Livne et al., 2018; Subudhi et al., 2018; Praveen et al., 2018). A limitation to the use of supervised machine learning is the limited amount of data. This lack of data is mainly due to the poor quality of the clinical datasets (corrupted or missing images), the insufficient amount of labeled data (current datasets limited to a few hundred patients) and the imbalance between classes (more pixels healthy than pathological pixels). This can be considered as a bottleneck specially when using highly discriminating algorithms depending on a large number of parameters. A way to circumvent this limitation is to generate more data from simulation and image synthesis model (Shrivastava et al., 2017; Mahmood et al., 2018; Shin et al., 2018). Data augmentation is also a way to improve regularization and

reduce overfitting by injecting more prior information into the training dataset (Simard et al., 2003).

In this article, we assess the interest of simulation for the prediction of the fate of acute stroke lesion. This prediction is undertaken here with deep learning on convolutional neural networks (CNN) since they are known to have, by comparison with the classical shallow learning techniques (support vector machines, random forests ...), higher amount of parameters to be tuned and can produce the best performances but also are the most likely to benefit from data augmentation.

Deep learning has been applied in stroke in different contexts including prediction from perfusion imaging or with other MRI modalities, for whole tissue segmentation or voxel-based classification and also with more or less complex neural network architectures (Stier et al., 2015; Nielsen et al., 2018; Zhang et al., 2018; Lucas et al., 2018; Pedemonte et al., 2018). In our case, we apply a standard architecture of CNN learning at the voxel level for binary classification of the fate of the tissue (*i.e.* tissue expansion or regression) from spatio temporal data. Neural networks usually require some data augmentation which is conventionally done in deep learning with spatial distortions likely to occur in nature (Shorten and Khoshgoftaar, 2019). In our case, due to the temporal dimension of the data, usual spatial distortions would not correspond to realistic transformation which differentiate a patient from another. Instead, we propose to use a 3D plus time simulator recently developed for perfusion MRI (Giacalone et al., 2017b). In the use case of Giacalone et al. (2017b) the simulator served as a ground truth generator to evaluate the robustness of deconvolution algorithms (Frindel et al., 2014). We propose to extend here the use of the physical simulator of Giacalone et al. (2017b) to another problematic of more clinical importance in acute ischemic stroke management, with the prediction of the fate of the tissue from perfusion imaging.

As main novelty of our work, we demonstrate the possibility to boost the performance of final stroke prediction with help of synthetic perfusion MRI images produced by the physically and physiologically relevant simulator of

Giacalone et al. (2017b). Additionally, we further enrich this simulation by focusing on arterial input function which was stressed as the limiting factor of the simulator in Giacalone et al. (2017b). In the other deep learning studies applied to stroke (Stier et al., 2015; Nielsen et al., 2018; Zhang et al., 2018; Lucas et al., 2018; Pedemonte et al., 2018) the training dataset was based on a cohort of patients. By contrast here, we demonstrate the possibility, thanks to the use of simulation, to train on the perfusion MRI data of a single patient in acute stroke to predict the final infarct of this specific patient. Closest related work regarding synthetic data has been used to learn perfusion parameters from a relatively small number of training samples in CT perfusion (Robben and Suetens, 2018) with classical data augmentation techniques. By contrast, we predict the final fate of the tissue from raw (*i.e.* non deconvolved) data without help of perfusion parameters, we work on MRI perfusion images and we develop synthetic data from an MRI simulator. Recent studies have shown the benefit of learning from raw perfusion data of training cohorts for ischemic lesion prediction (Giacalone et al., 2018; Pinto et al., 2018; Robben et al., 2018). Here, as additional novelty, we show how synthetic data simulated from raw MRI perfusion data can be used to directly learn the final infarct of a given specific patient, without the need of a training cohort.

2. Material and Method

2.1. Clinical MRI data

We used clinical MRI data from the European I-Know multicenter database (Hermitte et al., 2013). All patients from the study gave their informed consent and the imaging protocol was approved by the regional ethics committee. In total, we had a cohort of 76 patients with acute ischemic hemispheric stroke at our disposal, including 40 patients who received a thrombolytic treatment while the remaining 36 patients received no treatment. None of the patients reperfused after stroke.

All patients underwent the following MRI protocol on admission: diffusion-weighted-imaging (DWI; repetition time 6000 ms, field of view 24 cm, matrix 128x128, slice thickness 5 mm), fluid-attenuated-inversion-recovery (FLAIR; repetition time, 8690 ms; echo time, 109 ms; inversion time 2500 ms; flip angle, 150; field of view, 21 cm; matrix, 224x256; 24 sections; section thickness, 5 mm), T2-weighted gradient echo , MR-angiography and dynamic susceptibility-contrast perfusion imaging (DSC-PWI; echo time 40 ms, repetition time 1500 ms, field of view 24 cm, matrix 128x128, 18 slices, slice thickness 5 mm; gadolinium contrast at 0.1 mmol/kg injected with a power injector). From DSC-PWI, we extracted the commonly used hemodynamic maps such as: the 3D maps of the cerebral blood flow (CBF), the cerebral blood volume (CBV), the mean transit time (MTT), the time to maximum (TMAX) and the time to peak (TTP). A follow-up FLAIR-MRI was performed at 1-month after admission time. Raw perfusion MRI were registered, for each slice, using the first time point as reference for all the other time points, with a maximum mutual information approach. Final lesion was segmented for each patient on the one-month follow-up FLAIR-MRI by 3 experts. The FLAIR-MRI were first co-registered to DSC-PWI by computing the average temporal signal before contrast-agent arrival. Raw perfusion MRI were registered, for each slice, using the first time point as reference for all the other time points, with a maximum mutual information approach. This was done by registering each temporal point (n+1) on its previous temporal point (n) and by then applying recursively the transformation matrices obtained until all time points were aligned with the first time point. All registrations were done using Elastix software (Klein et al., 2009). The transformation matrix obtained was then used to register the final ischemic lesion mask. After registration, final lesion was rebinarized by applying a 50% threshold correction to avoid possible partial volume effects introduced by registration.

2.2. MRI simulator

Simulated data were generated with the DSC-MRI perfusion simulator of Giacalone et al. (2017b) that is able to simulate contrast-agent concentration images. We briefly recall the parameters of this simulator which includes realistic brain and lesion shapes, distinct classes of tissues (healthy, infarcted, gray and white matter) and their associated hemodynamic parameters as well as arterial input function (AIF). In Giacalone et al. (2017b), the simulator was used to test the robustness of AIF deconvolution. The sensitivity to all parameters was systematically tested and the uncertainty of the AIF itself was demonstrated, as also shown in Calamante et al. (2000), to be the limiting factor. In this study, to extend the value of the simulator of Giacalone et al. (2017b) to machine learning, we decided to limit the investigation on the choice of the simulated AIF. Haemodynamic parameters were set to their default values presented in Giacalone et al. (2017b), that we recall in Table 1. These haemodyamic parameters correspond to average values from the values reported in the literature. Acquisition parameters were set to 200 a.u. for the baseline value, 60 s for the acquisition time, $0.030s$ s for the time echo and 21 dB for the SNR.

Hemodynamic Parameters	**BG**	**HGT**	**HWT**	**LT**
CBF ($\mu \pm \sigma$)	0 ± 0	60 ± 9	25 ± 2.1	10 ± 4.3
MTT ($\mu \pm \sigma$)	0 ± 0	4 ± 2.2	4.8 ± 3.2	10 ± 5

Table 1: Simulator default parameters for tissue variability of hemodynamic parameters Cerebral Blood Flow (CBF, in mL/g/s) and Mean Transit Time (MTT, in s). The distribution of each parameter is modeled by a Gaussian of average μ, standard deviation σ. We considered 3 tissue classes: Healthy Gray Tissues (HGT), Healthy White Tissues (HWT), and Lesional Tissues (LT). Background (BG) has a null distribution.

The AIF is modeled as a gamma distribution that can be expressed using the simplified formulation proposed by Madsen (1992):

$$f(t) = \begin{cases} 0, & \text{if } t \leq d \\ y_{max}.(\frac{t-d}{t_{max}})^{\alpha}.\exp\left(\alpha(1-\frac{t-d}{t_{max}})\right), & \text{if } t \geq d \end{cases}, \quad (1)$$

where y_{max} and t_{max} respectively correspond to the magnitude and the position of the maximum of the arterial input function, d is the arrival time of the contrast agent and α corresponds to the shape parameter of the gamma function.

AIF extraction. We characterized the AIF of each patients with a multiple AIFs selection method. To do so, AIFs were extracted for each patient from voxels located in the main cerebral arteries on their DSC-PWI (Waaijer et al., 2007). The voxel selection was performed with a manual ROI method by three different operators. Then raw perfusion signals for all selected voxels of each patient were averaged to produce mean contrast-agent concentration signal. The contrast agent curves were then fitted by a gamma function defined in Eq. (1).

AIF characterization. The estimation of AIF parameters are known to be critical for the prediction of the ischemic lesion fate (Calamante et al., 2000; Giacalone et al., 2017b). There are different possible origins of AIF variability. First, in a multicentric study the duration of the perfusion protocol may differ. In our case the duration was standardized to 1 minute for all centers. Also, the delay between beginning of acquisition and the injection may differ from one patient to another producing a temporal shift. However because CNN are translation invariant they are not sensitive to this possible time shift. Intra-patient variability accounts for the AIF variation depending on the location of the selected voxel. Inter-patient variability is attributed to the amount of blood coming to the brain that may of course vary from one patient to another. Both these biological sources of variability were present in our dataset. Intra-patient variability was reduced in our study by averaging AIF for each patient after arterial selection by three distinct experts. Concerning inter-patient variability, as observed in Meijs et al. (2016) and in our dataset, most AIFs present a narrow and hight distribution represented in blue in Fig. 1 (the transit time of the contrast agent is between 10 and 15 secondes). However, few patients have larger AIF (represented in red in Fig. 1), hence they are underrepresented in the cohort. These observations motivated the choice to investigate various

approaches to simulate AIF in this study along the various datasets described in the following.

2.3. Training and testing datasets

Three training datasets were created to predict ischemic lesion fate of real patients of the cohort as described in this subsection. The first training dataset corresponds to real patients from the cohort, while the two remaining training datasets are pure simulated data.

2.3.1. Training datasets

Training dataset with a selection of real patients. First, a training dataset of 6 patients from the cohort of Section 2.1 was created: all presented narrow AIFs as shown in blue in Fig. 1. This approach enables to test the predictive value of a biased training dataset of real patients presenting a very low AIF variability regarding our tested patients. This dataset from real patients is obviously very small. It will serve as reference to compare with the prediction result of dataset generated from simulated patients.

Training dataset with simulation from theoretical AIF found in the literature (dataset A). A simulated dataset, noted A, was created with concentration images generated with the AIF default settings of the simulator, which correspond to average AIF parameters from the literature (Kellner et al., 2013). In this configuration, AIF simulation parameters were set to a unique value, that is to say: $y_{max} = 0.61$, $t_{max} = 4.5$, $d = 3$, $\alpha = 3$. This approach enables to test the predictive value of synthetic perfusion images simulated from a theoretical AIF not tuned to the values of our tested patients.

Training dataset with simulation from patient-specific AIF (dataset B). A second simulated dataset, noted B, was created by setting the AIF input parameters to the clinical AIF fitting parameters extracted from each tested patient (see chosen values in Table 2). This approach enables to test the predictive added value of synthetic perfusion images simulated from an AIF tuned to the values of our tested patients.

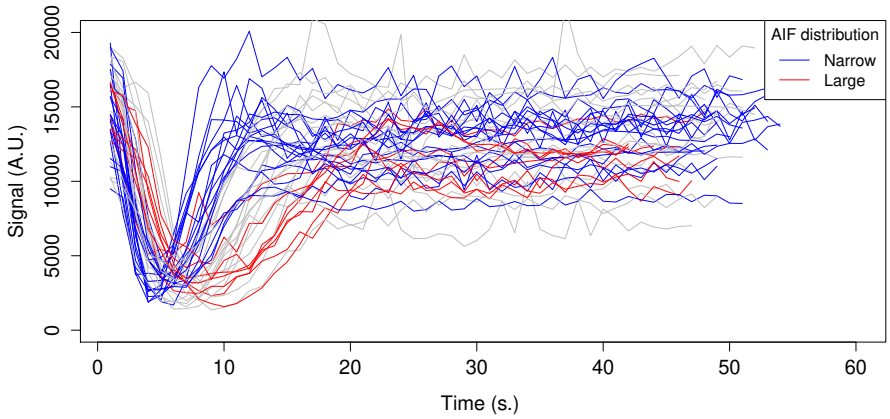

Figure 1: Mean AIF curves extracted on each of the 76 patients in the cohort. The blue and red lines correspond to AIFs with respectively narrow and large distributions. The grey lines correspond to the remaining AIFs. It shows the wide variability and inhomogeneity of AIF shapes.

2.3.2. Testing dataset

We chose 8 patients from the cohort of Section 2.1 to build the testing dataset. These were selected to cover the diversity of AIF shape observed in the cohort. Only 2 of these 8 patients received a thrombolitic treatment but none of them reperfused on their own. These 8 chosen patients, in addition to their representative AIF shapes, have been selected with sufficiently large final lesions. Indeed when final lesions are very small (typically smaller than some mLs), DSC are calculated on very few voxels, and each poorly predicted voxel very quickly penalizes the patient DSC. Also, lesions exploded in multiple non connected sub-lesions are also more difficult to predict as demonstrated in Frindel et al. (2015). This choice is justified to guarantee a controlled evolution of the lesion (stable or enlarged lesions) in order to focus on the AIF and the variability resulting from this parameter. As shown in Fig. 2, patients 1, 2, 3 and 6 present large AIF, whereas the other patients present relatively narrow AIF shapes. AIF of each individual patient to be tested were extracted. The

fitting parameters of each of the tested patients are presented in Table 2 and depicted in Fig. 2. In dataset B variants of the AIF of Table 2 were simulated.

Patient	y_{max}	t_{max}	d	α
1	0.90	6.32	13.14	4.23
2	0.86	10.53	11.99	8.54
3	0.60	3.41	18.67	0.99
4	0.92	5.26	19.64	0.94
5	0.74	2.35	11.62	1.01
6	0.60	4.58	14.01	2.30
7	0.77	3.45	9.40	2.48
8	0.85	7.08	8.79	8.03

Table 2: AIF parameters (y_{max}, t_{max}, d, α) values of the tested patients obtained from their mean contrast agent gamma curve.

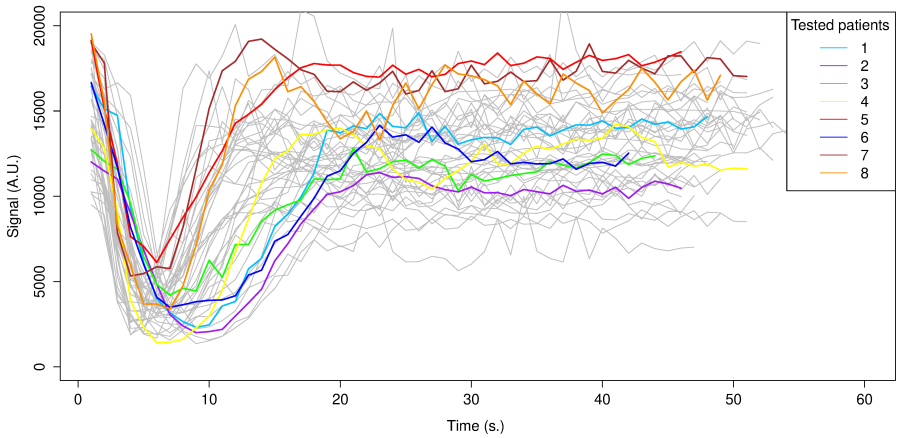

Figure 2: AIF curves from the 8 testing patients, shown in color, among all the AIF of the cohort. In order to better visualize them, we did not show the delay d. The chosen patients present wide AIF shape differences.

2.4. Voxel fate prediction models

The prediction of cerebral tissue fate from perfusion images was done after a spatio temporal encoding of the voxel environment fed to convolutional neural networks, as described in this subsection. Fig. 3 shows the proposed pipeline when learning from simulated data.

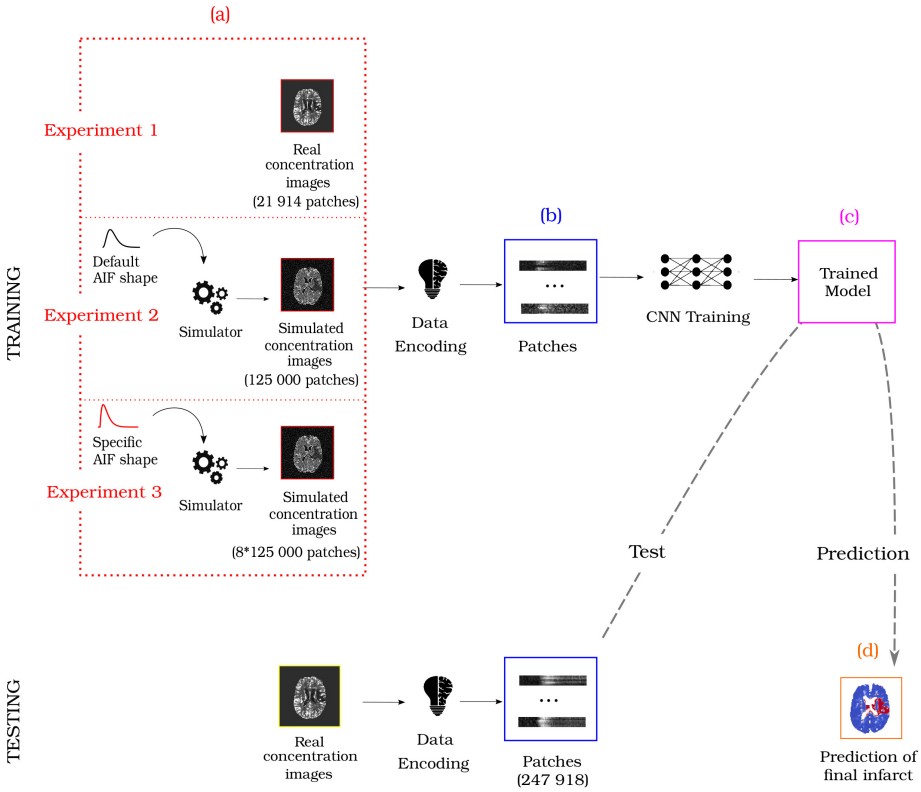

Figure 3: Overview of the proposed prediction pipeline. (a) The initial images are contrast-agent concentration images. In experiment 1, the training dataset consists in patches from real concentration images, whereas in experiment 2 and 3, the training dataset consists in synthetic patches obtained from the simulator. In experiment 2 AIF input parameter is set to a default value, and in experiment 3, AIF input parameter are the ones of the tested patient. (b) Concentration images are encoded into spatio-temporal patches. (c) A Convolutional Neural Network (CNN) model is trained from patches of the concentration images. (d) Each voxel from the tested concentration images is classified as healthy or infarcted.

2.4.1. Encoding of perfusion images

Perfusion images were converted into contrast-agent concentration images after a logarithmic transformation, under the assumption of a linear relationship between the contrast agent concentration and the change in transverse relaxation rate (Østergaard et al., 1996; Villringer et al., 1988). This transformation makes it possible to standardize the images between patients since the baseline is dispensed with. Then, concentration images were encoded in local spatio-temporal patches as recently described in Giacalone et al. (2018). Shortly, the time signal of a voxel of interest is deployed along a spatial direction. Its 8 voxels in the Moore neighborhood of order 1 are also deployed in the same direction, stacking the time signature of each neighboring voxel below each other. Thereby a patch of size 9 by Nt is created for each voxel, where Nt is the number of temporal acquisition points in the perfusion imaging sequence. In order to obtain patches independent of each other, we did not consider a Moore neighborhood of higher order.

It has been shown that patches for injured voxels present different patterns than patches for healthy voxels which can be discriminated in terms of texture (Giacalone et al., 2018). To go further than Giacalone et al. (2018), we can notice from Fig. 4 that this discriminability strongly depends on AIF. For instance in Fig. 4, patient 8, who had a narrow AIF, has a contrast-agent transit time much shorter than patient 2, who had a larger AIF. This supports the need for specific patient learning, taking into account the AIF of each patient to better predict their pathological voxels.

2.4.2. CNN classifier

A CNN was designed to directly take as input the spatio-temporal patches of dimension (9,60) to make a voxel-based prediction, as one patch represent the spatio-local environment of one voxel. The output for each patch was the predicted probability to belong to two classes (healthy tissue or infarcted tissue). We chose CNN as they are known to be translation invariant (Zeiler and Fergus, 2014; Fukushima, 1980). This property was a key point in learning, as we only

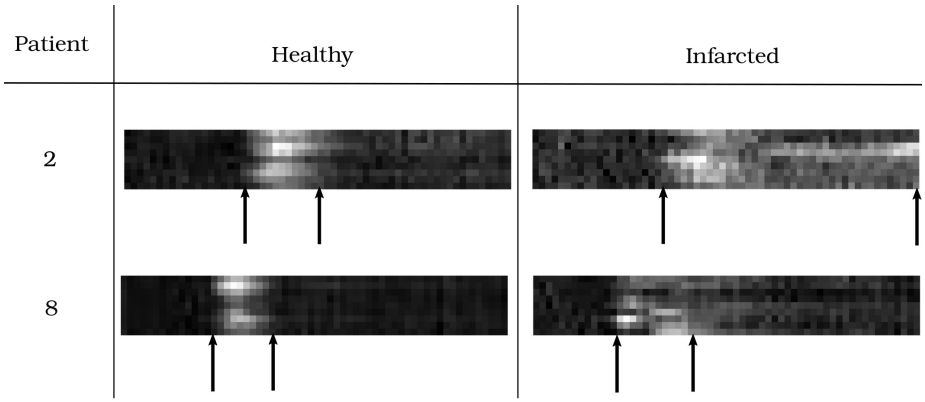

Figure 4: Illustration of a healthy patch (left) and a pathological patch (right) from concentration image for the tested patients 1 and 8. Healthy voxels exhibits a narrow hyperintensity segment resulting from the quick contrast-agent bolus passage. Pathological voxels exhibits a spread out hyperintensity segment, noisy and low contrasted, resulting from the difficult passage of the contrast-agent bolus. The two patients have different hemodynamic characteristics: the transit time of the contrast-agent, represented by the space between the two arrows, is faster for patient 8 than for patient 2. Patch intensities were converted into grayscale image for the representation.

wish to learn about the transit of the contrast agent in tissues and not about its arrival time. Thus, the network would not be sensitive to the delay d, but only to the white pattern of the spatio-temporal signature. A unique architecture described in Table 3 was designed. This architecture present a limited number of convolutional layers in order to avoid the patch size reduction and overfitting. Network weights were randomly initialized at the beginning of the training. Rectifier linear units (ReLU) function was used as activation function, known to perform better and faster than the sigmoïd or hyperbolic tangent functions (Krizhevsky et al., 2012; Jarrett et al., 2009). In the last fully connected layer, we used softmax, with 2 output units as our task approach a binary segmentation problem. As long as the patches input have small dimensions (9,60) and that convolution tends to reduce the output image dimension, we decided not to use any max-pooling to avoid further size reduction. We used dropout (Srivastava et al., 2014; Hinton et al., 2012) in the fully connected layers in order to avoid

overfitting. We used the categorical cross-entropy function as a loss function and a stochastic gradient descent to optimize the model. For all experiments, the total number of weights to train was 197 087, the dropout was set to 0.5, the number of epochs was set to 30, and the batch size to 32.

Layer	Type	Filter Size	Stride	# filters	FC units	Output Shape
1	Conv2D	2*2	1*1	16	-	(8, 59, 16)
2	Conv2D	2*2	1*1	32	-	(7, 58, 32)
3	FC	-	-	-	15	(15)
4	FC	-	-	-	2	(2)

Table 3: Convolutional neural network architecture proposed. The type of the two first layers are 2D Convolutional layers (Conv2D) and the two last layers are fully connected layers (FC).

Each model was trained 10 times in order to have an overview of their global performance and not only the best metric shot. To get a balanced training dataset we ensured that half of the patches contain voxels classified as lesion on the follow-up FLAIR-MRI. All CNNs were trained using Keras 2.1.3 with Python 3.6.3 interface. The training of the networks took globally less than 15 min on a standard work station with an NVIDIA GeForce GTX 1080 GPU with 8 GB memory.

2.4.3. Evaluation of the classification

We assessed our results using the Dice Similarity Coefficient (DSC) (Dice, 1945) and the Hausdorff Distance (HD) (Henrikson, 1999), as both were used for the international Ischemic Stroke Lesion Segmentation challenge of MICCAI (Winzeck et al., 2018). These metrics were computed between the predicted infarcted voxels and the mask of the final lesion provided by FLAIR-MRI. For comparison the prediction of the perfusion lesion was also computed from a TMAX perfusion map thresholded at 6 seconds with approach proposed by Frindel et al. (2014). This procedure is a standard approach in clinical research (Olivot et al., 2009; Cho et al., 2015). We computed DSC and HD between the voxels above this threshold and the mask of the final lesion as a clinical reference.

Using synthetic data for training enables to produce data on demand. We evaluated the minimum number of simulated patches required to obtain stable learning for the described CNN architecture. For each patient, we had 125 000 initial patches that we divided into 8 subsets by simple random sampling with replacement of different size: respectively 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% of the initial number of patches. For each patient, each set of patches was trained 10 times.

2.5. Experiment details

We classified the patches of the testing dataset through three experiments, each using a different training dataset. We give experimental details in the following subsection.

Experiment 1: training from a selection of real patients. In this experiment, the training dataset consisted of patches from 6 patients of the cohort presenting narrow AIF (see Section 2.3.1). From these 6 patients, we were able to get 21 914 patches with half of it healthy, and the other half infarcted. The validation dataset consisted of 17 766 patches from 3 independent patients presenting a different AIF shape. In this experiment, learning rate was set to 0.01.

Experiment 2: training with simulation from theoretical AIF. In this experiment, the training dataset consisted of 125 000 patches from the simulated images of dataset A with half of it healthy, and the other half infarcted. The validation dataset consisted of a set of 125 000 patches from other images obtained with the same simulation parameters. In this experiment, learning rate was set to 0.0001.

Experiment 3: training with simulation from patient-specific AIF. In this experiment, we learned from a synthetic specific patient. Dataset B was separated into several subsets, each set representing simulated images of a specific tested patient. Therefore, 8 different trainings were done. For each training, the corresponding training set consisted of 125 000 patches with half of it healthy, and

the other half infarcted, and the corresponding validation dataset consisted of a set of 125 000 patches from other images obtained with the same patient specific simulation parameters. In this experiment, learning rate was set to 0.0001.

3. Results

Table 4 reports the results obtained by the three conducted experiments in terms of mean DSC and HD values and their standard deviation for each tested patient. The DSC and HD values between the thresholded TMAX and the final lesion is also shown, as a clinical reference.

	Experiment 1		**Experiment 2**		**Experiment 3**		**TMAX thr**	
Patient	**HD**	**DSC**	**HD**	**DSC**	**HD**	**DSC**	**HD**	**DSC**
1	50.4 ± 0.0	0.17 ± 0.031	50.2 ± 0.0	0.12 ± 0.00	50.2 ± 0.0	0.086 ± 0.003	49.7	0.11
2	51.1 ± 0.0	0.18 ± 0.002	51.1 ± 0.0	0.18 ± 0.00	48.8 ± 0.829	**0.62 ± 0.008**	43.2	0.51
3	44.1 ± 0.0	0.054 ± 0.001	43.8 ± 0.0	0.053 ± 0.00	43.8 ± 0.0	**0.23 ± 0.010**	43.4	0.14
4	44.4 ± 0.0	0.47 ± 0.041	44.9 ± 0.0	0.17 ± 0.00	44.8 ± 0.085	**0.48 ± 0.008**	42.2	0.47
5	43.8 ± 1.08	0.36 ± 0.069	45.4 ± 0.0	0.11 ± 0.00	44.4 ± 0.0	**0.47 ± 0.005**	46.1	0.31
6	43.0 ± 0.0	0.11 ± 10^{-5}	43.0 ± 0.0	0.11 ± 0.00	**42.5 ± 0.0**	0.20 ± 0.006	44.7	0.28
7	44.0 ± 0.36	0.47 ± 0.031	45.0 ± 0.0	0.26 ± 0.00	45.0 ± 0.0	**0.53 ± 0.004**	46.3	0.36
8	40.1 ± 3.07	0.64 ± 0.026	47.1 ± 0.0	0.17 ± 0.00	44.0 ± 10^{-6}	0.48 ± 0.017	45.8	0.40
Total mean	45.1 ± 3.73	0.30 ± 0.22	46.3 ± 2.94	0.14 ± 0.074	45.4 ± 2.65	**0.40 ± 0.19**	45.18 ± 2.36	0.32 ± 0.14

Table 4: Hausdorff distance (HD) and similarity metrics (DSC) after performing 10 times experiment 1 (1st column), 2 (2nd column), and 3 (3rd column). All metrics are averaged over the 10 times, and shown for each tested patient (average ± standard deviation). DSC and HD between TMAX $\geq$ 6s and final lesion is shown (4rth column). We showed in bold when experiment 3 gave the best performance at the patient scale. For the three experiments, HD standard deviation is low or even zero because some outlier voxels were systematically mis-predicted. The metrics are also averaged over the test dataset (last raw).

The results from experiment 1 show the impact of a mismatch between training and testing in terms of AIF. Discrimination between healthy and pathological voxels was only trained from narrow AIF patients: it turns out to be impossible to correctly predict the voxels of the tested patients with different AIF shapes such as patient 1, 2, 3 and 6 who have mean DSC below 0.13. In contrast, tested patients with AIFs close to those of the learned patients, such as patients 4, 5, 7 and 8 have mean DSC greater than 0.49.

In experiment 2, training CNN with synthetic data obtained from a theoretical AIF, corresponding to average AIF parameters from the literature without any patient specific tuning, provides very poor results. Almost all voxels are predicted infarcted, so that DSC is very low and the Hausdorff distance high. This illustrates that a theoretical AIF is not able to capture the variability that exists between patients. AIF differs from subject to subject and since we work on raw perfusion imaging data, it is important to incorporate the specific AIF of each considered patient .

The variability brought by the AIF from one patient to another is well-known (Meijs et al., 2016). This is the reason why most classical approaches use deconvolution in order to compensate for this variability. In experiment 3, we did not solve this ill-positioned inverse problem and rather incorporate the specific arterial input function in the direct problem through simulation of the perfusion signals with a set of realistic arterial input functions. In experiment 3, the training dataset contains systematically simulated images where the AIF is tuned to those of the patient to be predicted (dataset B). It clearly appears that the adjustment of the AIF-related parameters in the simulator has a considerable impact on the learning performance. The average DSC in experiment 3 is 0.40 ($\pm$ 0.19), compared with 0.14 ($\pm$ 0.074) in experiment 2 and 0.30 ($\pm$ 0.22) in experiment 1. Learning from specific AIF in experiment 3 clearly improved the median DSC, which is 0.48 in experiment 3, compared with 0.27 in experiment 1 and 0.15 in experiment 2. It appears also that learning from raw data in experiment 3 gives better results than thresholding the deconvolved TMAX which present an average DSC of 0.32 ($\pm$ 0.14). As another reference, it is interesting to note that the best models so far in the stroke prediction ISLES 2017 challenge had an average DSC of 0.38 ($\pm$ 0.22) and an average Hausdorff distance of 29.21 ($\pm$ 15.04). The mean performance scores in experiment 3 is in the same order of magnitude. However an absolute comparison is not strictly possible because the two datasets are different. To test the transferability of our proposed method, we have run 10 times experiment 3 for patients with large lesions, patients number 7 and 20, from the ISLES dataset. The average DSC

was 0.39 (± 0.30) and the average Hausdorff distance was 46.9 (± 1.80).

Regarding our described CNN architecture in experiment 3, we also investigated the minimum patches required for stable training. Results are shown in Fig. 5. It appears that between 25 000 and 100 500 training patches, DSC increases almost linearly depending on the number of training patches. Beyond 100 000 patches, the curve seems to reach a plateau: all the diversity of information provided by the simulated images has been learned, and the supply of new images is redundant. Also the standard deviation of the DSC values is lower after training more than 100 000 patches. These observations indicate that given our CNN architecture and our dataset, the minimum number of training patches to obtain stable and optimal results is around 100 000.

As additional experiments, we compared experiment 1, 2 and 3 with other neural network architectures and metrics of merit. The results of this comparison are provided in the additional material section. The added value of the simulation used in experiment 3 over experiments 2 and 1 is robustly obtained.

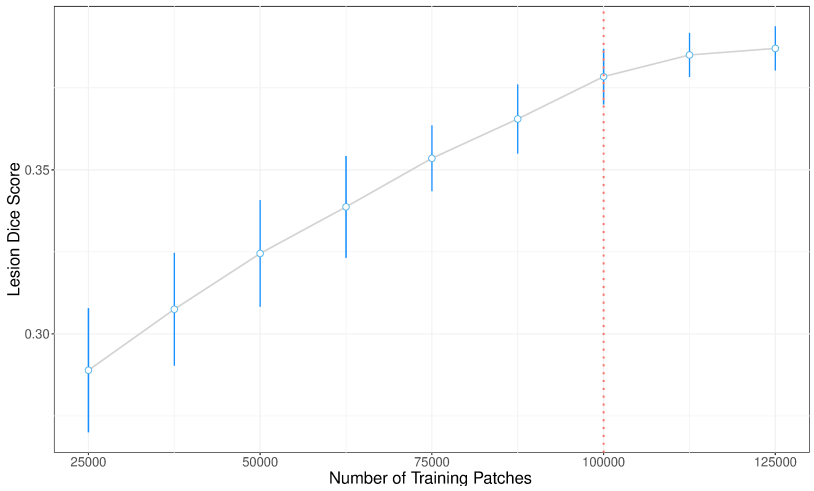

Figure 5: Resulting DSC according to the number of training patches in experiment 3. Each point corresponds to the mean DSC of 10 repeated experiments on each tested patient and its standard deviation. The red dotted line indicates the optimal number of training patches.

4. Discussion

The previous results have demonstrated the value of perfusion imaging simulation based on AIF for the prediction of lesion fate in stroke. In this section, we go beyond the sole observation of average performance and now discuss the limit of our experiments.

Faced with the problem of AIF representativeness in the training set underlined in experiments 1 and 2, we proposed through experiment 3 to learn directly from the AIF of the patient to be tested. Thanks to this type of learning, each patient was correctly predicted individually, even for patients whose AIF is under-represented in the overall cohort. It should be noted that for experiment 3, we simulated images from a single averaged AIF of the tested patient. However, some studies show that it may be beneficial to take into account the intra-patient AIF variability (Livne et al., 2017). Indeed, in experiment 3, the extracted AIF seems relevant for patients 2, 3, 4, 5, 6 and 7, but not enough for patients 1 and 8 as they show better performances in the experiment 1. Probably, these two patients cannot be summarized in one single AIF as they might present a large AIF intra-variability. This encourages us for further work to simulate images with several AIFs according to the intra-AIF variability of the tested patient, and therefore potentially better represent their hemodynamic characteristics.

At the voxel level, Fig. 6 shows the position of the badly classified voxels resulting from experiments 1, 2 and 3. We can see that many of our errors are in the ventricles. As contrast-agent does not pass into these areas, the voxels have a particularly noisy signal. Thus the model detects cerebral blood flow disturbance and directly links it with the final ischemic stroke. These errors could be easily filtered by limiting mask segmentation errors.

The current state-of-the-art in the prediction of lesion fate in stroke classically works on the deconvolved parametrics maps (CBV, CBF, Tmax, TTP and MTT) learned on cohorts of patient. With our approach it is actually not possible to learn directly from these maps since our encoding is based on the

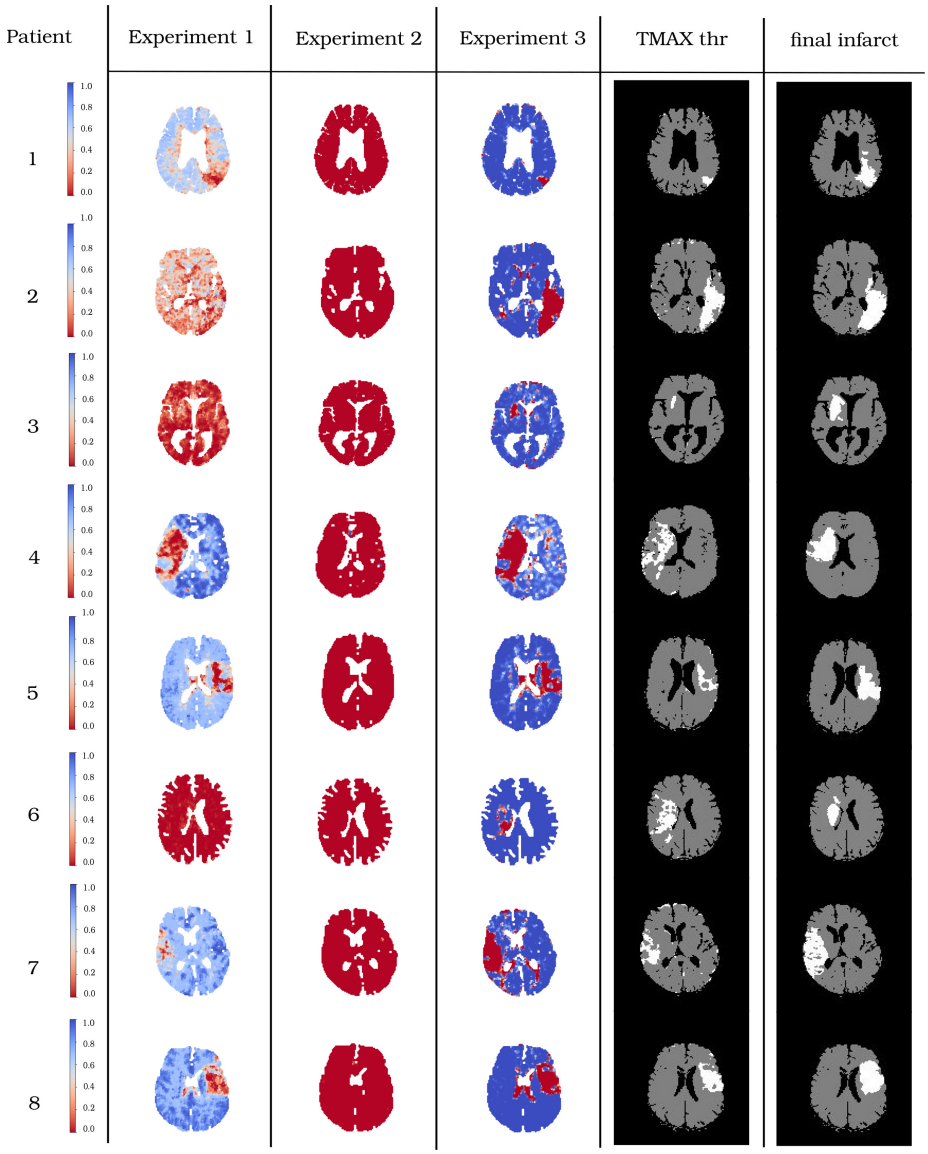

Figure 6: Output predictions of experiment 1 (1st column), experiment 2 (2nd column) and experiment 3 (3rd column) for the testing patients. The colorbar presents the probability for each voxel to be healthy. Voxels in blue shades were predicted healthy and voxels in red shades were predicted infarcted. The classical biomarker TMAX thresholded at 6 seconds is shown (4th column). Columns 1 to 4 should be compared to the final flair (5th column).

spatio-temporal signature of perfusion data. A possible comparison would be to compare learning from raw perfusion signals, encoded with our spatio-temporal approach, and learning from deconvolved signals (the one used to produce standard parametric maps CBV, CBF, Tmax, TTP, and MTT). Such a comparison was shown in Giacalone et al. (2018) and demonstrated the interest of the proposed spatio-temporal encoding of raw perfusion signal.

In this work, the CNN architecture exploited consists of a low number of layers. The minimum number of patches required for a specific patient learning was therefore limited to 100 000 (which is approximately 25 simulations needed to represent a patient with 2 000 infarcted voxels and corresponds to a computation time of 3 minutes on the work station of Section 2.4.2). However, with other encoding methods and more complex models, the number of data required for learning could be higher. The simulator we used in this paper (Giacalone et al., 2017b) would be able to overcome this problem, by allowing to produce a theoretically unlimited amount of labelled data.

Finally, it should be noted that this approach, although no instantaneous, seems fully compatible with the real-time management of stroke patient in clinical routine although including learning and simulation. Indeed, for performances compared with the state-of-the-art, the overall computation time (simulation and learning) for a patient-specific approach as developed in experiment 3 is about 20 minutes. This remains of the same order of magnitude as the time announced for a deconvolution approach as in Frindel et al. (2014).

5. Conclusion

In this article, we demonstrated how the simulation of hemodynamical signals can be used to increase the amount of data and boost the performance of convolutional neural networks for the prediction of the lesion evolution in acute ischemic stroke from DSC-PWI. This new demonstration of the value of simulation to train machine learning techniques in medical imaging enabled to obtain performances close to the ones of the literature for this important stroke

problem.

Several simulation approaches have been tested including the simulation from average AIF values found in the literature or simulation from the AIF of the specific tested patient. In this patient specific scenario, we have shown that the performance of prediction was higher to results from the state of the art methods applied on cohorts while learning here on a specific patient instead of a cohort. A limitation in machine learning based prediction for biomedical imaging is the limited size of cohorts. This is a priori especially true with highly expressive models such as the one based on convolutional neural networks. We have demonstrated on the specific disease of stroke that in fact it is possible to predict on an extremely limited cohort of a single patient from convolutional neural networks with help of simulation.

This work could be extended in several directions. For stroke, several works propose to predict the final infarction simply through the perfusion modality, in the framework of the challenge ISLES for example (Maier et al., 2017) or outside this challenge (Giacalone et al., 2018; Pinto et al., 2018; Lucas et al., 2018). However, acute DWI is known to be a highly predictive image for the final stroke lesion (Røhl et al., 2001; Giacalone et al., 2017a). Incorporating diffusion in the model could highly improve performance classification. This would require to integrate a diffusion simulator (Graham et al., 2016; Du et al., 2016) and redesign the architecture of the convolutional neural network used in this study. Also the simulation was here limited to a pixel-wise simulator. More spatial context could be integrated in the realism of the simulator by adding a physical layer at a spatial level of the DSC-MRI simulator. As another extension of this work for stroke, one can notice that the simulation of brain tissue was here based on global stationary binary class model (healthy versus infarcted) while from a clinical perspective, stroke appears as a more complex problem with importance on the spatial localisation of the tissue. It is however obvious that whatever the level of simulation complexity, some bias between real and simulated images will remain. A way to learn this bias and thus improve the performance of simple simulation-based training approaches, such as the

one introduced here, could be to add some domain adaptation techniques just before the last decision layers of the convolutional neural network used here (Courty et al., 2017). Finally, the use of simulation to train machine learning based model is of generic value and may be extended to any other disease and imaging modality for which some simulator are already available (Glatard et al., 2013).

Supplementary material

Additional metrics

Regarding the metrics presented in Table 4, we present in Table 5 additional metrics in order to further demonstrate the value of the tested CNN of Table 3 in experiments 1, 2 and 3. We can note that for all tested metrics, experiment 3 with training on simulated perfusion data, systematically present the best mean classification performances.

Additional architectures

Regarding our proposed architecture presented in 3 , we present additional fully convolutional architectures (FCN1, FCN2 and FCN3) described in Tables 6, 7, 8. Experiments 1, 2 and 3 were performed using these three architectures. For all experiments, we used the categorical cross-entropy function as a loss function and a stochastic gradient descent to optimize the model. The number of epochs was set to 30, the learning rate to 0.001 and the batch size to 32. The corresponding classification performances are reported on Table 9.

We can note that for all tested architectures, experience 3 systematically presents the best classification performance. The DSC and Hausdorff distance values obtained with the proposed convolutional neural network (CNN) are not significantly different from the presented fully convolutional networks (FCN1, FCN2 and FCN3 architectures) (p-value > 0.05 when performing Mann–Whitney test for paired data).

	Experiment 1				Experiment 2				Experiment 3			
Patient	**MCC**	**AUC**	**Recall**	**Precision**	**MCC**	**AUC**	**Recall**	**Precision**	**MCC**	**AUC**	**Recall**	**Precision**
1	0.12 ± 0.079	0.77 ± 0.055	0.86 ± 0.061	0.10 ± 0.060	0.0 +- 0.0	0.55 +- 0.076	1.0 +- 0.0	0.12 +- 0.0	0.040 ± 0.0039	0.57 ± 0.16	0.15 ± 0.17	0.11 ± 0.0096
2	-0.039 ± 0.043	0.77 ± 0.028	0.98 ± 0.0050	0.10 ± 0.0008	0.0 +- 0.0	0.51 +- 0.031	1.0 +- 0.0	0.10 +- 0.0	0.58 ± 0.0056	0.89 ± 0.0037	0.65 ± 0.21	0.48 ± 0.15
3	-0.013 ± 0.018	0.35 ± 0.020	0.98 ± 0.041	0.037 ± 0.026	0.0 +- 0.0	0.50 +- 0.0027	1.0 +- 0.0	0.022 +- 0.0	0.31 ± 0.0076	0.72 ± 0.0075	0.26 ± 0.017	0.20 ± 0.0065
4	0.46 ± 0.030	0.94 ± 0.0025	0.94 ± 0.021	0.30 ± 0.034	0.016 +- 0.00042	0.62 +- 0.036	1.0 +- 0.0	0.091 +- 1.29e-05	0.47 ± 0.0080	0.93 ± 0.0039	0.90 ± 0.0024	0.25 ± 0.089
5	0.35 ± 0.043	0.87 ± 0.17	0.91 ± 0.022	0.19 ± 0.037	0.0019 +- 0.0054	0.57 +- 0.048	1.0 +- 0.0	0.057 +- 0.0	0.47 ± 0.0044	0.91 ± 0.0080	0.78 ± 0.005	0.34 ± 0.0031
6	0.050 ± 0.13	0.45 ± 0.054	0.99 ± 0.0003	0.060 ± 1.0 e-05	0.0 +- 0.0	0.50 +- 0.00046	1.0 +- 0.0	0.060 +- 0.0	0.24 ± 0.0044	0.72 ± 0.014	0.22 ± 0.0054	0.28 ± 0.0039
7	0.35 ± 0.030	0.80 ± 0.011	0.62 ± 0.11	0.37 ± 0.072	0.0034 +- 0.00048	0.62 +- 0.039	1.0 +- 0.0	0.15 +- 2.71e-06	0.46 ± 0.0093	0.86 ± 0.0014	0.78 0.0023	0.41 ± 0.010
8	0.60 ± 0.042	0.88 ± 0.021	0.62 ± 0.050	0.62 ± 0.17	0.0042 +- 0.00048	0.70 +- 0.0398	1.0 +- 0.0	0.092 +- 3.47e-06	0.43 ± 0.028	0.71 ± 0.012	0.46 ± 0.0035	0.50 ± 0.050

Table 5: Mean Matthews Correlation Coefficient (MCC), mean Area Under the Curve (AUC), mean Recall and mean Precision after performing 10 repetitions in experiment 1, 2 and 3 using the convolutional neural network of Table 3. The metrics shown are averaged over 10 repetitions, and shown for each tested patient (average ± standard deviation).

Layer	Type	Filter Size	Pooling Size	Stride	# filters	Output Shape	activation function
1	Conv2D	2*2	-	1*1	8	(8, 59, 8)	relu
2	maxpooling	-	1*2	1*1	-	(8, 29, 8)	relu
3	Conv2D	2*2	-	1*1	8	(7, 28, 8)	relu
4	maxpooling	-	1*2	1*1	-	(7, 14, 8)	relu
5	Conv2D	2*2	-	1*1	16	(6, 13, 16)	relu
6	maxpooling	-	1*2	1*1	-	(6, 6, 16)	relu
7	Conv2D	2*2	-	1*1	16	(5, 5, 16)	relu
8	maxpooling	-	2*2	1*1	-	(2, 2, 16)	relu
9	Conv2D	2*2	-	1*1	32	(1, 1, 32)	relu
10	Conv2D	1*1	-	1*1	2	(1, 1, 2)	softmax

Table 6: Fully Convolutional Network 1 (FCN1) proposed. The type of the layers are either 2D Convolutional layers (Conv2D) or Maxpooling layer (maxpooling).

Layer	Type	Filter Size	Pooling Size	Stride	# filters	Output Shape	activation function
1	Conv2D	2*2	-	1*1	8	(8, 59, 8)	relu
2	maxpooling	-	1*2	1*1	-	(8, 29, 8)	relu
3	Conv2D	2*2	-	1*1	8	(7, 28, 8)	relu
4	maxpooling	-	1*2	1*1	-	(7, 14, 8)	relu
5	Conv2D	2*2	-	1*1	16	(6, 13, 16)	relu
6	maxpooling	-	1*2	1*1	-	(6, 6, 16)	relu
7	Conv2D	2*2	-	1*1	32	(5, 5, 32)	relu
8	Conv2D	2*2	-	1*1	32	(4, 4, 32)	relu
9	Conv2D	2*2	-	1*1	32	(3, 3, 32)	relu
10	Conv2D	2*2	-	1*1	64	(2, 2, 64)	relu
11	Conv2D	2*2	-	1*1	2	(1, 1, 2)	softmax

Table 7: Fully Convolutional Network 2 (FCN2) proposed. The type of the layers are either 2D Convolutional layers (Conv2D) or Maxpooling layer (maxpooling).

Mean validation softmax losses

In Table 10 we present the mean validation softmax losses for all the trained networks in all experiments. In experiment 2 and 3, the training and validation set are simulated data, while in experiment 1, the training and validation set are real data. It appears that the real data are more complex than the simulated data, since the convergence of the network on the validation set is better in experiment 2 and 3 than in experiment 1.

Layer	Type	Filter Size	Pooling Size	Stride	# filters	Output Shape	activation function
1	Conv2D	2*2	-	1*1	8	(8, 59, 8)	relu
2	maxpooling	-	1*2	1*1	-	(8, 29, 8)	relu
3	Conv2D	2*2	-	1*1	8	(7, 28, 8)	relu
4	maxpooling	-	1*2	1*1	-	(7, 14, 8)	relu
5	Conv2D	2*2	-	1*1	16	(6, 13, 16)	relu
6	Conv2D	2*2	-	1*1	16	(5, 12, 16)	relu
7	Conv2D	2*2	-	1*1	32	(4, 11, 32)	relu
8	Conv2D	2*2	-	1*1	32	(3, 10, 32)	relu
9	Conv2D	2*2	-	1*1	32	(2, 9, 32)	relu
10	Conv2D	2*2	-	1*1	64	(1, 8, 64)	relu
11	Conv2D	1*2	-	1*1	64	(1, 7, 64)	relu
12	Conv2D	1*2	-	1*1	64	(1, 6, 64)	relu
13	Conv2D	1*2	-	1*1	64	(1, 5, 64)	relu
14	Conv2D	1*2	-	1*1	64	(1, 4, 64)	relu
15	Conv2D	1*2	-	1*1	64	(1, 3, 64)	relu
16	Conv2D	1*2	-	1*1	64	(1, 2, 64)	relu
18	Conv2D	1*2	-	1*1	2	(1, 1, 2)	softmax

Table 8: Fully Convolutional Network 3 (FCN3) proposed. The type of the layers are either 2D Convolutional layers (Conv2D) or Maxpooling layer (maxpooling).

	FCN1			**FCN2**			**FCN3**		
Patient	**Exp. 1**	**Exp. 2**	**Exp. 3**	**Exp. 1**	**Exp. 2**	**Exp. 3**	**Exp. 1**	**Exp. 2**	**Exp. 3**
1	0.13 / 45.7	0.12 / 50.2	**0.14** / 48.3	0.16 / 49.7	0.12 / 50.2	**0.16 / 46.9**	0.21 / 49.7	0.12 / 50.2	0.13 / 46.4
2	0.17 / 49.7	0.18 / 51.1	**0.61 / 46.8**	0.21 / 51.1	0.18 / 51.1	**0.65 / 46.9**	0.18 / 51.1	0.18 / 51.1	**0.52 / 48.1**
3	0.055 / 49.6	0.053 / 43.8	**0.22 / 44.0**	0.055 / 44.1	0.053 / 43.8	**0.23 / 44.0**	0.055 / 44.1	0.053 / 43.8	**0.23 / 44.0**
4	0.48 / 51.1	0.17 / 44.9	**0.54 / 43.9**	0.48 / 44.4	0.17 / 44.9	**0.55 / 40.5**	0.42 / 44.4	0.17 / 44.9	**0.52 / 38.8**
5	0.37 / 44.1	0.11 / 45.4	**0.44 / 43.5**	0.40 / 43.5	0.11 / 45.4	**0.49 / 41.3**	0.39 / 43.8	0.11 / 45.4	**0.47 / 42.0**
6	0.11 / 44.4	0.11 / 43.0	**0.27 / 41.3**	0.11 / 43.0	0.11 / 43.0	**0.18 / 41.3**	0.11 / 43.0	0.11 / 43.0	**0.20 / 42.6**
7	0.57 / 44.4	0.26 / 45.0	0.54 / 43.2	0.61 / 42.6	0.26 / 45.0	0.59 / 42.3	0.55 / 43.1	0.26 / 45.0	0.55 / 43.3
8	0.65 / 43.0	0.17 / 47.1	0.48 / 34.9	0.66 / 40.7	0.17 / 47.1	0.55 / 33.9	0.65 / 40.9	0.17 / 47.1	0.51 / 38.2

Table 9: Mean dice similarity metrics (DSC) and Hausdorff distance (HD) after performing 10 repetitions experiment 1, 2 and 3 using three different fully convolutional networks: FCN1 (1st column), FCN2 (2nd column) and FCN3 (3rd column). The metrics shown are averaged over the 10 repetitions, and shown for each tested patient (average DSC/average HD). We highlight in bold when experiment 3 gave the best performance at the patient scale.

	Experiment 1	**Experiment 2**	**Experiment 3**
CNN	0.32	0.023	0.17
FCN1	0.39	0.010	0.14
FCN2	0.37	0.021	0.10
FCN3	0.36	0.010	0.08

Table 10: Mean validation softmax losses of all trained networks (CNN, FCN1, FCN2, FCN3)

References

Albers, G.W., 2018. Use of imaging to select patients for late window endovascular therapy. Stroke 49, 2256–2260.

Albers, G.W., Marks, M.P., Kemp, S., Christensen, S., Tsai, J.P., Ortega-Gutierrez, S., McTaggart, R.A., Torbey, M.T., Kim-Tenser, M., Leslie-Mazwi, T., et al., 2018. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. New England Journal of Medicine 378, 708–718.

Calamante, F., Gadian, D.G., Connelly, A., 2000. Delay and dispersion effects in dynamic susceptibility contrast MRI: Simulations using singular value decomposition. Magnetic Resonance in Medicine 44, 466–473.

Caplan, L.R., 2016. Caplan's stroke. Cambridge University Press.

Cho, T.H., Nighoghossian, N., Mikkelsen, I.K., Derex, L., Hermier, M., Pedraza, S., Fiehler, J., Østergaard, L., Berthezène, Y., Baron, J.C., 2015. Reperfusion within 6 hours outperforms recanalization in predicting penumbra salvage, lesion growth, final infarct, and clinical outcome. Stroke 46, 1582–1589.

Christensen, S., Lansberg, M.G., 2018. CT perfusion in acute stroke: Practical guidance for implementation in clinical practice. Journal of Cerebral Blood Flow & Metabolism .

Courty, N., Flamary, R., Tuia, D., Rakotomamonjy, A., 2017. Optimal transport for domain adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1853–1865.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Du, H.b., Wang, L.H., Liu, W.Y., Yang, F., Li, Z., Zhu, Y.M., 2016. Diffusion mri simulation for human brain based on the atlas, in: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 898–902.

Frindel, C., Robini, M.C., Rousseau, D., 2014. A 3-D spatio-temporal deconvolution approach for MR perfusion in the brain. Medical Image Analysis 18, 144–160.

Frindel, C., Rouanet, A., Giacalone, M., Cho, T.H., Østergaard, L., Fiehler, J., Pedraza, S., Baron, J.C., Wiart, M., Berthezène, Y., et al., 2015. Validity of shape as a predictive biomarker of final infarct volume in acute ischemic stroke. Stroke , STROKEAHA–114.

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36, 193–202.

Giacalone, M., Frindel, C., Grenier, E., Rousseau, D., 2017a. Multicomponent and longitudinal imaging seen as a communication channel - An application to stroke. Entropy 19, 187.

Giacalone, M., Frindel, C., Robini, M., Cervenansky, F., Grenier, E., Rousseau, D., 2017b. Robustness of spatio-temporal regularization in perfusion MRI deconvolution: An application to acute ischemic stroke. Magnetic Resonance in Medicine 78, 1981–1990.

Giacalone, M., Rasti, P., Debs, N., Frindel, C., Cho, T.H., Grenier, E., Rousseau, D., 2018. Local spatio-temporal encoding of raw perfusion MRI for the prediction of final lesion in stroke. Medical Image Analysis 50, 117–126.

Glatard, T., Lartizien, C., Gibaud, B., Da Silva, R.F., Forestier, G., Cervenansky, F., Alessandrini, M., Benoit-Cattin, H., Bernard, O., Camarasu-Pop, S., et al., 2013. A virtual imaging platform for multi-modality medical image simulation. IEEE Transactions on Medical Imaging 32, 110–118.

Goyal, M., Demchuk, A.M., Menon, B.K., Eesa, M., Rempel, J.L., Thornton, J., Roy, D., Jovin, T.G., Willinsky, R.A., Sapkota, B.L., et al., 2015. Randomized assessment of rapid endovascular treatment of ischemic stroke. New England Journal of Medicine 372, 1019–1030.

Graham, M.S., Drobnjak, I., Zhang, H., 2016. Realistic simulation of artefacts in diffusion MRI for validating post-processing correction techniques. NeuroImage 125, 1079–1094.

Henrikson, J., 1999. Completeness and total boundedness of the Hausdorff metric. MIT Undergraduate Journal of Mathematics 1, 69–80.

Hermitte, L., Cho, T.H., Ozenne, B., Nighoghossian, N., Mikkelsen, I.K., Ribe, L., Baron, J.C., Østergaard, L., Derex, L., Hjort, N., et al., 2013. Very low cerebral blood volume predicts parenchymal hematoma in acute ischemic stroke. Stroke , STROKEAHA–113.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 .

Jarrett, K., Kavukcuoglu, K., LeCun, Y., et al., 2009. What is the best multi-stage architecture for object recognition?, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE. pp. 2146–2153.

Kellner, E., Mader, I., Mix, M., Splitthoff, D.N., Reisert, M., Foerster, K., Nguyen-Thanh, T., Gall, P., Kiselev, V.G., 2013. Arterial input function measurements for bolus tracking perfusion imaging in the brain. Magnetic resonance in medicine 69, 771–780.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2009. Elastix: a toolbox for intensity-based medical image registration. IEEE transactions on medical imaging 29, 196–205.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, pp. 1097–1105.

Livne, M., Boldsen, J.K., Mikkelsen, I.K., Fiebach, J.B., Sobesky, J., Mouridsen, K., 2018. Boosted tree model reforms multimodal magnetic resonance imaging infarct prediction in acute stroke. Stroke 49, 912–918.

Livne, M., Madai, V.I., Brunecker, P., Zaro-Weber, O., Moeller-Hartmann, W., Heiss, W.D., Mouridsen, K., Sobesky, J., 2017. A PET-guided framework supports a multiple arterial input functions approach in DSC-MRI in acute stroke. Journal of Neuroimaging 27, 486–492.

Lucas, C., Kemmling, A., Mamlouk, A.M., Heinrich, M.P., 2018. Multi-scale neural network for automatic segmentation of ischemic strokes on acute perfusion images, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 1118–1121.

Madsen, M.T., 1992. A simplified formulation of the gamma variate function. Physics in Medicine & Biology 37, 1597.

Mahmood, F., Chen, R., Durr, N.J., 2018. Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. IEEE Transactions on Medical Imaging 37, 2572–2581.

Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al., 2017. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Medical Image Analysis 35, 250–269.

Maier, O., Schröder, C., Forkert, N.D., Martinetz, T., Handels, H., 2015. Classifiers for ischemic stroke lesion segmentation: A comparison study. PLOS ONE 10, 1–16.

Meijs, M., Christensen, S., Lansberg, M.G., Albers, G.W., Calamante, F., 2016. Analysis of perfusion MRI in stroke: To deconvolve, or not to deconvolve. Magnetic Resonance in Medicine 76, 1282–1290.

Murray, C.J., Barber, R.M., Foreman, K.J., Ozgoren, A.A., Abd-Allah, F., Abera, S.F., Aboyans, V., Abraham, J.P., Abubakar, I., Abu-Raddad, L.J., et al., 2015. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE)

for 188 countries, 1990-2013: Quantifying the epidemiological transition. The Lancet 386, 2145–2191.

Najm, M., Al-Ajlan, F.S., Boesen, M.E., Hur, L., Kim, C.K., Fainardi, E., Hill, M.D., Demchuk, A.M., Goyal, M., Lee, T.Y., et al., 2018. Defining CT perfusion thresholds for infarction in the golden hour and with ultra-early reperfusion. Canadian Journal of Neurological Sciences 45, 339–342.

Nielsen, A., Hansen, M.B., Tietze, A., Mouridsen, K., 2018. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. Stroke , STROKEAHA–117.

Olivot, J.M., Mlynash, M., Thijs, V.N., Kemp, S., Lansberg, M.G., Wechsler, L., Bammer, R., Marks, M.P., Albers, G.W., 2009. Optimal tmax threshold for predicting penumbral tissue in acute stroke. Stroke 40, 469–475.

Østergaard, L., Weisskoff, R.M., Chesler, D.A., Gyldensted, C., Rosen, B.R., 1996. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. Part I: Mathematical approach and statistical analysis. Magnetic Resonance in Medicine 36, 715–725.

Park, J., 2017. Acute Ischemic Stroke. Springer.

Pedemonte, S., Bizzo, B., Pomerantz, S., Tenenholtz, N., Wright, B., Walters, M., Doyle, S., McCarthy, A., De Almeida, R.R., Andriole, K., et al., 2018. Detection and delineation of acute cerebral infarct on DWI using weakly supervised machine learning, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer. pp. 81–88.

Pinto, A., Pereira, S., Meier, R., Alves, V., Wiest, R., Silva, C.A., Reyes, M., 2018. Enhancing clinical MRI perfusion maps with data-driven maps of complementary nature for lesion outcome prediction, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer. pp. 107–115.

Praveen, G., Agrawal, A., Sundaram, P., Sardesai, S., 2018. Ischemic stroke lesion segmentation using stacked sparse autoencoder. Computers in biology and medicine 99, 38–52.

Robben, D., Boers, A.M., Marquering, H.A., Langezaal, L.L., Roos, Y.B., van Oostenbrugge, R.J., van Zwam, W.H., Dippel, D.W., Majoie, C.B., van der Lugt, A., et al., 2018. Prediction of final infarct volume from native CT perfusion and treatment parameters using deep learning. arXiv preprint arXiv:1812.02496 .

Robben, D., Suetens, P., 2018. Perfusion parameter estimation using neural networks and data augmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer. pp. 439–446.

Røhl, L., Østergaard, L., Simonsen, C.Z., Vestergaard-Poulsen, P., Andersen, G., Sakoh, M., Le Bihan, D., Gyldensted, C., 2001. Viability thresholds of ischemic penumbra of hyperacute stroke defined by perfusion-weighted MRI and apparent diffusion coefficient. Stroke 32, 1140–1146.

Sartor, K., 2006. Magnetic resonance imaging in ischemic stroke. Springer Science & Business Media.

Shin, H.C., Tenenholtz, N.A., Rogers, J.K., Schwarz, C.G., Senjem, M.L., Gunter, J.L., Andriole, K.P., Michalski, M., 2018. Medical image synthesis for data augmentation and anonymization using generative adversarial networks, in: Simulation and Synthesis in Medical Imaging, Springer. pp. 1–11.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. Journal of Big Data 6, 60.

Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., Webb, R., 2017. Learning from simulated and unsupervised images through adversarial training, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 2107–2116.

Simard, P.Y., Steinkraus, D., Platt, J.C., 2003. Best practices for convolutional neural networks applied to visual document analysis, in: Proceedings of the Seventh International Conference on Document Analysis and Recognition, IEEE. pp. 958–962.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout : A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15, 1929–1958.

Stier, N., Vincent, N., Liebeskind, D., Scalzo, F., 2015. Deep learning of tissue fate features in acute ischemic stroke, in: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 1316–1321.

Subudhi, A., Acharya, U.R., Dash, M., Jena, S., Sabut, S., 2018. Automated approach for detection of ischemic stroke using delaunay triangulation in brain mri images. Computers in Biology and Medicine 103, 116–129.

Villringer, A., Rosen, B.R., Belliveau, J.W., Ackerman, J.L., Lauffer, R.B., Buxton, R.B., Chao, Y.S., Wedeenand, V.J., Brady, T.J., 1988. Dynamic imaging with lanthanide chelates in normal brain: Contrast due to magnetic susceptibility effects. Magnetic Resonance in Medicine 6, 164–174.

Waaijer, A., Van der Schaaf, I., Velthuis, B., Quist, M., Van Osch, M., Vonken, E., Van Leeuwen, M., Prokop, M., 2007. Reproducibility of quantitative CT brain perfusion measurements in patients with symptomatic unilateral carotid artery stenosis. American journal of neuroradiology 28, 927–932.

Winzeck, S., Hakim, A., McKinley, R., Pinto, J.A., Alves, V., Silva, C., Pisov, M., Krivov, E., Belyaev, M., Monteiro, M., et al., 2018. ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI. Frontiers in Neurology 9, 679.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks, in: Computer Vision – ECCV 2014, Springer. pp. 818–833.

Zerna, C., Hegedus, J., Hill, M.D., 2016. Evolving treatments for acute ischemic stroke. Circulation research 118, 1425–1442.

Zhang, R., Zhao, L., Lou, W., Abrigo, J.M., Mok, V.C., Chu, W.C., Wang, D., Shi, L., 2018. Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional densenets. IEEE Transactions on Medical Imaging 37, 2149–2160.

Zhang, Z., Sejdić, E., 2019. Radiological images and machine learning: trends, perspectives, and prospects. Computers in biology and medicine 108, 354–370.