LSTM Network for Prediction of Hemorrhagic Transformation in Acute Stroke

Yannan Yu[1], Bhargav Parsi[2], William Speier[3], Corey Arnold[3], Min Lou[4], and Fabien Scalzo[2(✉)]

[1] Department of Radiology, Stanford University, Stanford, CA 94305, USA
[2] Department of Neurology, UCLA, Los Angeles, CA 90095, USA
fab@cs.ucla.edu
[3] Department of Radiology, UCLA, Los Angeles, CA 90095, USA
[4] Second Affiliated Hospital, Zhejiang University, Hangzhou, Zhejiang, China

Abstract. Hemorrhagic transformation (HT) is one of the most devastating complications of reperfusion therapy in acute ischemic stroke. Prediction of an upcoming HT remains beyond current techniques in routine clinical practice. If made available, such information would benefit the management of acute ischemic stroke patients and help to tailor therapeutic strategies. This study aims at providing a machine learning framework for predicting occurrence and extent of HT from source perfusion-weighted magnetic resonance imaging (PWI) combined with diffusion weighted imaging (DWI). The model relies on a LSTM network based on PWI combined with DWI imaging features into a fully connected neural network. A retrospective comparative analysis performed on 155 acute stroke patients demonstrate the efficacy of the LSTM model (AUC-ROC: 89.4%) against state-of-the-art machine learning models. Predicted likelihood of HT at the voxel level was evaluated against HT annotations of stroke neurologists obtained from follow-up gradient recalled echo (GRE) imaging.

1 Introduction

Acute ischemic stroke (AIS) has a lifetime risk of 25% worldwide. It is often associated with significant disability, including motor and cognitive deterioration. Recent advances in treatment of acute stroke have demonstrated that reperfusion therapy, which dissolves or mechanically retrieves the clot, can significantly improve the outcome of the patients [14]. However, reperfusion therapy is also associated with complications; the most critical of which being intracerebral bleeding, also referred to as hemorrhagic transformation (HT). The risks of HT are particularly challenging to be assessed prior to reperfusion therapy. As HT may occur in eloquent brain areas, the consequences of HT have been shown to be associated with deteriorating symptoms, delayed neurological improvement, and poor outcome. The development of a computational model that could predict the occurrence and spatial extent of an upcoming HT before the reperfusion

D. Shen et al. (Eds.): MICCAI 2019, LNCS 11767, pp. 177–185, 2019.
https://doi.org/10.1007/978-3-030-32251-9_20

therapy would be very useful to refine eligibility criteria for reperfusion therapy. In this study, we introduce and evaluate a computational model based on a Long Short-Term Memory (LSTM) network that uses source perfusion-weighted magnetic resonance imaging (PWI) (i.e. after reconstruction from k-space but before feature extraction).

PWI imaging of the brain is obtained via a series of T2*-weighted MRI after the venous injection of a contrast bolus and is represented as a 4D dataset characterizing blood flow through the vasculature and tissue. In the context of stroke, it is common to process the contrast concentration time curve obtained for each voxel and extract specific features related to physiological changes of the tissue, such as cerebral blood volume (CBV) and time-to-maximum of the residue function (Tmax) [3]. In addition, pre-established permeability metrics can also be computed [11], including contrast slope, final contrast, and maximum concentration, reflecting brain-blood-barrier permeability. Some of these perfusion parameters, such as low CBV, prolonged Tmax, increased permeability, and lesion size on diffusion-weighted imaging (DWI), are known predictors of parenchymal hemorrhage. Yet, most of studies have limited their predictions to the simple occurrence of HT [15]. The severity, cerebral territory involved, and eloquence have not been addressed. Recent works have demonstrated that the use of source/native PWI has demonstrated an advantage over pre-defined maps in the context of tissue fate prediction in acute stroke [1,7] and time from stroke onset [8]. In these frameworks, the relevant features of the PWI are learned by a machine learning algorithm rather than being pre-defined. The main rationale is that the source PWI may contain additional information that is not captured by pre-defined features and that could improve the prediction of the target variable.

The key contribution of this work is a predictive model of HT that explicitly models the temporal features of the PWI signal using an LSTM network. LSTMs have outperformed standard machine learning models on a wide variety of applications related to temporal signals. Inspired by these findings, our framework goes beyond the prediction of occurrence of HT and aims at estimating the spatial extent of the injury by identifying the voxels that will undergo HT. Such predictions could bring valuable insights to stroke neurologists about the affected territory and the chances of good outcome. We provide a comparative analysis with standard models, including linear regression, kernel spectral regression, SVM, and random forests.

2 Methods

MRI data was collected from patients identified with AIS within six hours of symptom onset and admitted at a Chinese University Hospital from 06/2009 to 10/2016. Inclusion criteria were as follows: (1) acute ischemic lesions confirmed on DWI; (2) baseline perfusion MRI and DWI performed before reperfusion therapy; and (3) Enhanced 3D gradient recalled echo (GRE) T2*-weighted angiography performed 24 h after reperfusion therapy. Patients with significant motion artifacts were excluded. The ethics committee of the Hospital approved the study protocol.

All patients underwent MRI on 3.0 T systems equipped with 8-channel head coils. The routine MRI protocol including PWI, DWI, and GRE was performed in AIS patients at baseline and 24 h after reperfusion therapy. Although the acquisition parameters slightly varied during the seven-year period, the median parameters of PWI were as follows: field of view (FOV) = 240 mm, repetitive time (TR) = 1800 ms, echo time (TE) = 30 ms, acquisition matrix = 128 × 128, repetitive scanning times = 60, gadolinium dose = 15 ml, contrast speed = 4–5 ml/s, average duration = 1 min 48 s. The parameters of DWI b1000 and b0 were: FOV = 240 mm, TR = 4000 ms, TE = 80 ms, slice sickness = 5 mm, acquisition matrix = 160 × 160. The parameters of GRE: TE = 4.5 ms (first echo), matrix size = 256 × 256, flip angle = 20°, slice thickness = 2.0 mm.

Baseline PWI, DWI, and 24 h follow-up GRE images were co-registered automatically using SPM12. The arterial input function (AIF) was automatically detected using Olea Sphere and computed from the average of the time-intensity curve from several voxel locations. Because of the difference in acquisition settings across subjects, AIF and PWI values were interpolated temporally using bilinear interpolation to 60 time points with a 1.8s time interval. Presence of HT at 24 h on GRE images was assessed by a stroke neurologist and delineated on each patient using Osirix software.

2.1 Predictive Model

The predictive model of HT makes use of an LSTM architecture which captures the temporal information stored in the PWI signal. The model aims at predicting the occurrence of HT at the voxel level. The target output $y_i \in Y$ of the classifier is a binary variable indicating the presence of bleeding at follow-up in the pixel i. As a standard pre-processing, the concentration of the contrast agent $C_i(t)$ at time t in a voxel i is obtained from the pre-intervention PWI signal I by:

$$C_i(t) = -TE^{-1} \log\left(\frac{I_i(t)}{I_i(0)}\right) \tag{1}$$

where TE is the echo time, and I_0, I_t are the image intensity measured before bolus arrival and time t, respectively. The contrast concentration time curve C over a 3×3 local patch is used as the input data to an LSTM model and is therefore composed of 9 features over 60 time points. The LSTM is described by 60 output values that are connected to the output y_i using a fully connected neural network.

Long Short-Term Memory Network. LSTM [9] is a variation of the recurrent neural network (RNN) which allows information to persist inside the network via a loopy architecture. LSTMs are particularly well suited to represent time series and are used in our framework to model the relationship between a PWI image patch captured over time and the occurrence of HT post-intervention. An LSTM cell is defined by a state that changes according to three types of gates:

- Input Gates $i_t \in \mathcal{R}^N$ update the state of the cell and decide which values should be updated.
- Forget Gates $f_t \in \mathcal{R}^N$ are used to select relevant information with respect to a previous state.
- Output Gates $o_t \in \mathcal{R}^N$ determine the final cell state and the output value.

Given an input sequence $x = \{x_1, x_2, \ldots, x_T\}$ of length T with corresponding memory cell unit $C_t \in \mathcal{R}^N$ and hidden unit $h_t \in \mathcal{R}^N$ at time t, the parameters of the model are updated sequentially, as follows:

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \tag{2}$$
$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{3}$$
$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{4}$$
$$\tilde{C}_t = \tanh(W_c.[h_{t-1}, x_t] + b_C) \tag{5}$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$
$$h_t = o_t * \tanh C_t \tag{7}$$

The function $\sigma(x) = (1 + e^{-x})^{-1}$ used to compute f_t, i_t, o_t is a sigmoid whose values lie within the range $[0, 1]$. In addition to input, forget, and output gates previously described, the LSTM makes use of a memory cell unit C_t obtained from the sum of the previous memory cell unit C_{t-1} modulated by f_t, and a function of the current input x_t and previous hidden state h_{t-1} modulated by the input gate i_t. The output gate o_t is then used to determine what parts should be considered and then multiplied with the tanh of the memory cell state C_t to produce the hidden unit h_t. By learning how much of the memory cell state C_t should be transferred to the hidden state h_t based on the input x_t and previous state, this structure allows the LSTM to capture complex temporal dynamics.

Operating Modes. In addition to building an LSTM model from the contrast concentration C_i, we allow the framework to utilize additional inputs, including: the contrast concentration of the AIF $C_{aif} \in \mathcal{R}^T$, and the value of the DWI at b0 and b1000 $X_{dwi} \in \mathcal{R}^S$. A fully connected layer is used to combine the output of the LSTM and additional inputs. The binary output is obtained with a Softmax operator. We distinguish between several modes of operation depending on which input is used:

Mode 1: LSTM PWI. The 3×3 PWI patches for each time-point are used as input to the LSTM model, corresponding to 9 features over 60 time-points. The LSTM is defined by 60 output values.

Mode 2: LSTM PWI+AIF. Because the AIF is represented as a vector of 60 time-points, the input data now corresponds to 10 features by 60 time-points.

Mode 3: LSTM PWI+DWI. The outputs of the LSTM are combined with two 3×3 patches obtained from DWI b0 and b1000 to form a 78 input vector to the fully connected layer.

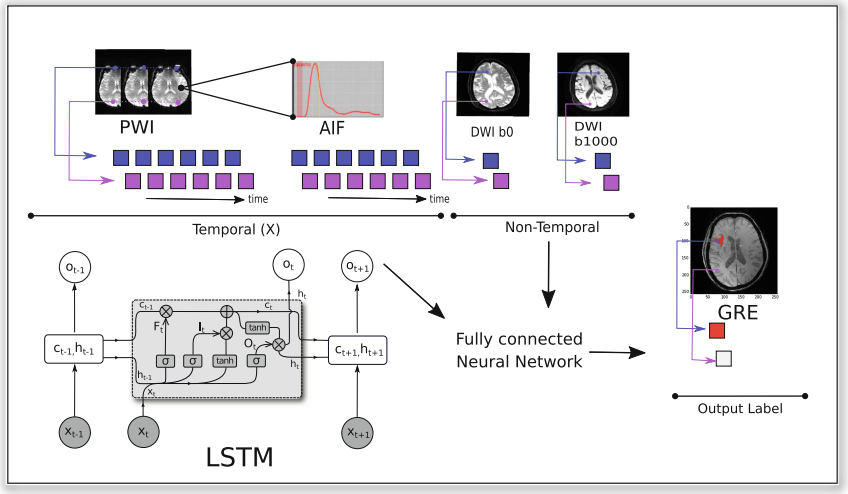


Fig. 1. Illustration of our LSTM-NeuralNet Framework. Local patches extracted on perfusion-weighted MRI (PWI) are combined together with the arterial input function (AIF) to train a LSTM model (bottom left, diagram modified from [13]). The output of the LSTM model is combined with local DWI image patches through a fully connected neural network that maps the features to the presence of HT as observed on gradient recalled echo (GRE) at followup.

Mode 4: LSTM PWI+DWI+AIF. The most comprehensive mode of operation combines the LSTM output (60) with the DWI b0 and b1000 patches, leading to 78 features combined in the fully connected layer (Fig. 1).

2.2 Experiments

We evaluate the framework using a 10-fold cross-validation performed at the voxel level. The accuracy of the model trained under 4 different modes is compared to Linear Regression, SR-KDA [2], SVM [4], and Random Forests. The LSTM training setup was defined with 60 cells of LSTM, 20 epochs, and trained with Adam optimizer [10] using binary cross-entropy as loss function. The outputs of the LSTM cells are converted to a binary output by using a fully connected neural network with a softmax optimization. The final output represents the probabilities of the HT class or the Non-HT class. The training of our machine learning model took place on 50,000 input samples equally distributed between HT and Non-HT voxels from various subjects as it has previously been shown that balanced classes provides a more representative estimation of the performance. The sampling method used in these experiments is similar to the one described in previous work [16]. As part of the cross-validation, we ensure that the data from a given patient were not included both in the training set and the test set. The accuracy is calculated on the basis of the area under

the curve of the receiver operating characteristic curve (AUC-ROC) and the precision-recall curve (AUC-PR). The 95% confidence interval associated with each result is obtained using the Bootstrapping method [5]. Optimization of the hyperparameters was performed using a nested cross-validation.

3 Results

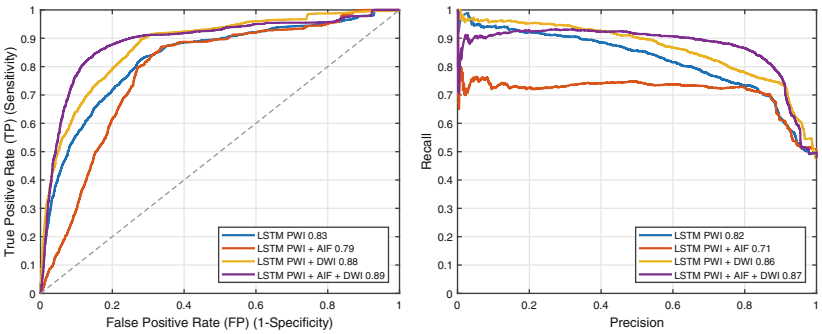
A total of 155 AIS patients satisfied the inclusion criteria and were included in this study, among whom 41 patients were diagnosed with HT. The results of each machine learning algorithm evaluated in this study are listed in Table 1 and illustrated in Fig. 2. For all models, the best performing configuration was the one that combines all inputs available (PWI, AIF and DWI), thus demonstrating the complementary information contained in these input variables. The LSTM model with PWI, DWI and AIF as inputs reached an AUC-ROC $= 89.4 \pm 4\%$, AUC-PR $= 87.4 \pm 6\%$. In addition, we also reported the predictive accuracy of manually defined ROI on Ktrans parametric map that was obtained in another study [12] (64%). It should be noted, however, that it is not directly comparable as the study was performed on another patient population and a different cross validation setup.

Table 1. Accuracy of the models in predicting voxel-wise HT occurrence.

Model	Input (s)	AUC-ROC	AUC-PR
LSTM	PWI	$83.1 \pm 2.9\%$	$82.3 \pm 5.8\%$
LSTM-NeuralNet	PWI+AIF	$79.6 \pm 2.6\%$	$71.9 \pm 5.8\%$
LSTM-NeuralNet	PWI+DWI	$88.3 \pm 3.6\%$	$86.4 \pm 6.4\%$
LSTM-NeuralNet	PWI+AIF+DWI	$\mathbf{89.4 \pm 4.3\%}$	$\mathbf{87.4 \pm 5.9\%}$
Linear regression	PWI+AIF+DWI	$58.5 \pm 7.5\%$	$44.4 \pm 9.8\%$
Random forests	PWI+AIF+DWI	$79.8 \pm 3.1\%$	$67.1 \pm 6.0\%$
SR-KDA [2]	PWI+AIF+DWI	$83.7 \pm 2.6\%$	$70.4 \pm 5.8\%$
SVM [4]	PWI+AIF+DWI	$82.1 \pm 2.9\%$	$69.9 \pm 5.8\%$
ROI-based [12]	K^{trans}	$64.1 \pm 6.0\%$	NA

4 Discussion

The aim of this paper is to introduce a predictive model for HT in acute ischemic stroke. The multi-input LSTM model achieved an AUC-ROC accuracy of 89% which is very promising considering the challenging nature of the problem. A significant finding of this paper is that the use of an LSTM improves the results over standard models.



(a) ROC curves

(b) PR curves

Fig. 2. Illustration of the accuracy in terms of ROC and PR curves for various predictive models of HT based on LSTM-NeuralNet architecture.

The development of hemorrhagic transformation in AIS is a complex pathophysiological process, which is influenced by multiple factors such as reperfusion, age, serum glucose level, stroke severity, and blood-brain barrier (BBB) damage. Perfusion-weighted imaging is rich in information about brain tissue and blood flow. It is particularly useful in detecting BBB permeability disruptions that are linked to ongoing or future HT. PWI remains largely under-utilized in the context of AIS due to the non-standard way of detecting these BBB impairments. The relationship and causal influence of PWI imaging markers remains poorly understood. In this context, the fact that machine learning models can predict HT from imaging alone is particularly encouraging and could be considered as part of new therapeutic strategies.

We can improve our present model by giving it additional features which may or may not be related to MRI features. Similar to the multi-input LSTM model, we can have these features fed to a different layer and the output of that layer can be concatenated with the output of the LSTM.

While the prediction of future occurrence of HT has been successful on a multi-center study [15], it is not clear that the presence of a small hemorrhage would prevent the patient from being treated with endovascular clot-retrieval therapy. In this paper, we are predicting the risk of HT at the voxel level, which can be used to infer volumetric and eloquence measures that would be more helpful in the clinical setting to rule out clot-retrieval therapy.

A limitation of this study is to include data with various degrees of revascularization which is typically assessed using the Thrombolysis in Cerebral Infarction (TICI) score [6]. The TICI score, which varies from 0 (no reperfusion) to 3 (complete reperfusion), is useful to quantify the degree of success in restoring blood flow after clot-retrieval intervention. A multi-center evaluation of the model on a more representative set of cases would be beneficial. A specific predictive model could be trained for various degrees of revascularization. During diagnosis, predicted maps of likelihood of HT development could help clinicians

to visualize the risks/benefits of the intervention and the dependency to the degree of revascularization.

5 Conclusion

This paper introduces a computational framework for the prediction of HT in acute ischemic stroke. The contribution of the model is two-fold; first it utilizes the source/native PWI signal and maps it to the development of HT using an LSTM model which captures the temporal signature of tissue voxels at risk of HT, second it does not only produce a prediction about overall presence of HT in the follow-up images but rather provides predictions at the voxel level that can be used to predict the severity of the HT and therefore better characterize the risks associated with an endovascular intervention.

Acknowledgements. This work was partially supported by a NIH award R01 NS100806. The GPU used for this research was donated by the NVIDIA Corporation.

References

1. Bertels, J., Robben, D., Vandermeulen, D., Suetens, P.: Contra-lateral information CNN for core lesion segmentation based on native CTP in acute stroke. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11383, pp. 263–270. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11723-8_26
2. Cai, D., He, X., Han, J.: Spectral regression for efficient regularized subspace learning. In: ICCV (2007). https://doi.org/10.1109/ICCV.2007.4408855
3. Calamante, F., Christensen, S., Desmond, P.M., Ostergaard, L., Davis, S.M., Connelly, A.: The physiological significance of the time-to-maximum (Tmax) parameter in perfusion MRI. Stroke **41**(6), 1169–1174 (2010)
4. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001). http://www.csie.ntu.edu.tw/~cjlin/libsvm
5. DiCiccio, T.J., Efron, B.: Bootstrap confidence intervals. Stat. Sci. **11**, 189–212 (1996)
6. Higashida, R.T., et al.: Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. Stroke **34**(8), e109–137 (2003)
7. Ho, K.C., Scalzo, F., Sarma, K.V., Speier, W., El-Saden, S., Arnold, C.: Predicting ischemic stroke tissue fate using a deep convolutional neural network on source magnetic resonance perfusion images. J. Med. Imaging (Bellingham) **6**(2), 026001 (2019)
8. Ho, K.C., Speier, W., Zhang, H., Scalzo, F., El-Saden, S., Arnold, C.W.: A machine learning approach for classifying ischemic stroke onset time from imaging. IEEE Trans. Med. Imaging **38**(7), 1666–1676 (2019)
9. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**, 1735–80 (1997)
10. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

11. Leigh, R., et al.: Pretreatment blood-brain barrier damage and post-treatment intracranial hemorrhage in patients receiving intravenous tissue-type plasminogen activator. Stroke **45**(7), 2030–2035 (2014)
12. Li, Y., et al.: Focal low and global high permeability predict the possibility, risk, and location of hemorrhagic transformation following intra-arterial thrombolysis therapy in acute stroke. AJNR Am. J. Neuroradiol. **38**(9), 1730–1736 (2017)
13. Olah, C.: Understanding LSTM networks (2015). https://colah.github.io/posts/2015-08-Understanding-LSTMs/
14. Powers, W., et al.: 2015 AHA/American stroke association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment. Stroke **46**(10), 3020–3035 (2015)
15. Scalzo, F., et al.: Multi-center prediction of hemorrhagic transformation in acute ischemic stroke using permeability imaging features. Magn. Reson. Imaging **31**(6), 961–969 (2013)
16. Scalzo, F., Hao, Q., Alger, J.R., Hu, X., Liebeskind, D.S.: Regional prediction of tissue fate in acute ischemic stroke. Ann. Biomed. Eng. **40**(10), 2177–2187 (2012)