Summarizing *A Machine Learning Approach For Classifying Ischemic Stroke Onset Time From Imaging*[1]

Daria Glaser

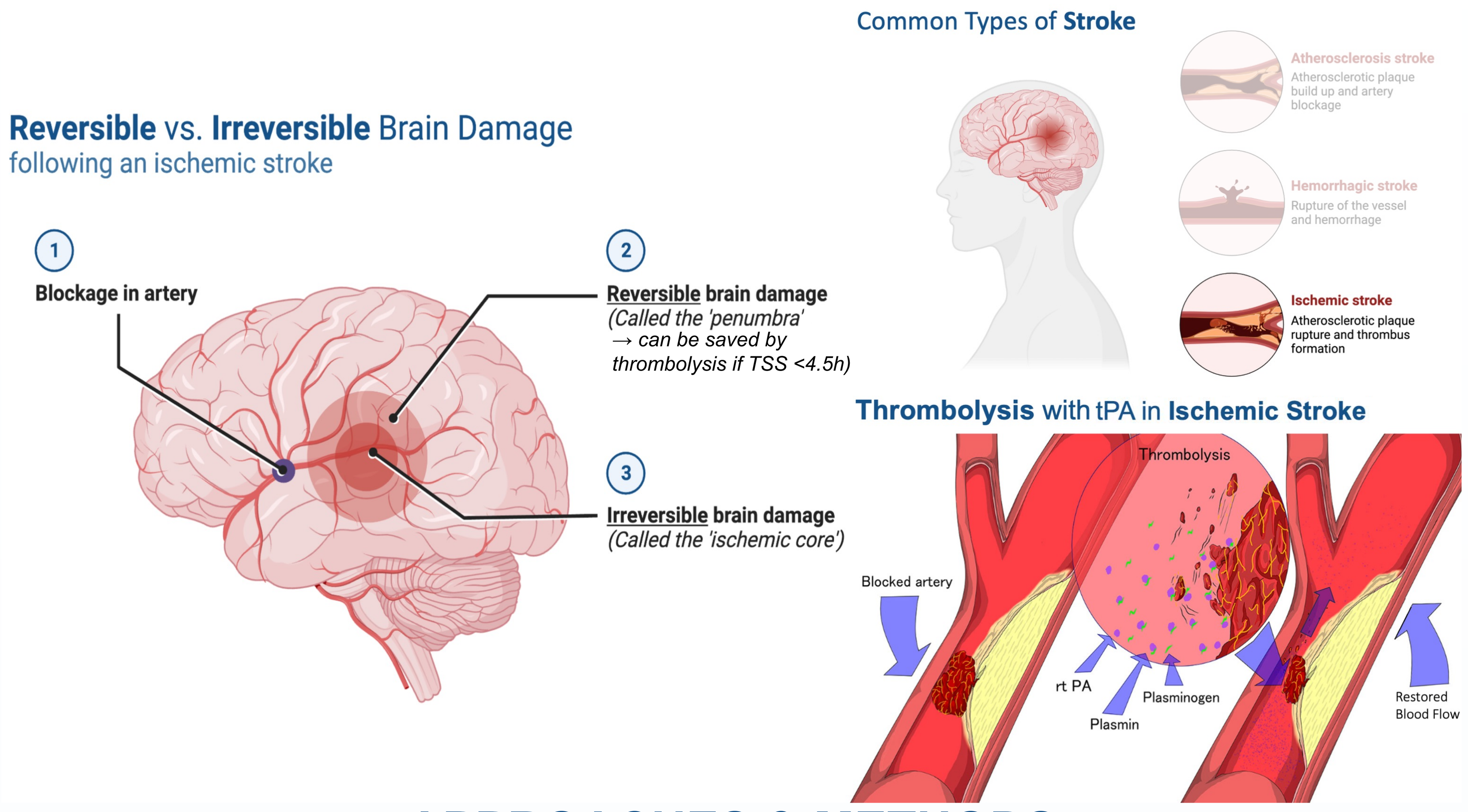
Clinical Department of Neurology, University Hospital Jena; daria.glaser@med.uni-jena.de

SUMMER SCHOOL 2023

ai.uni-jena.de

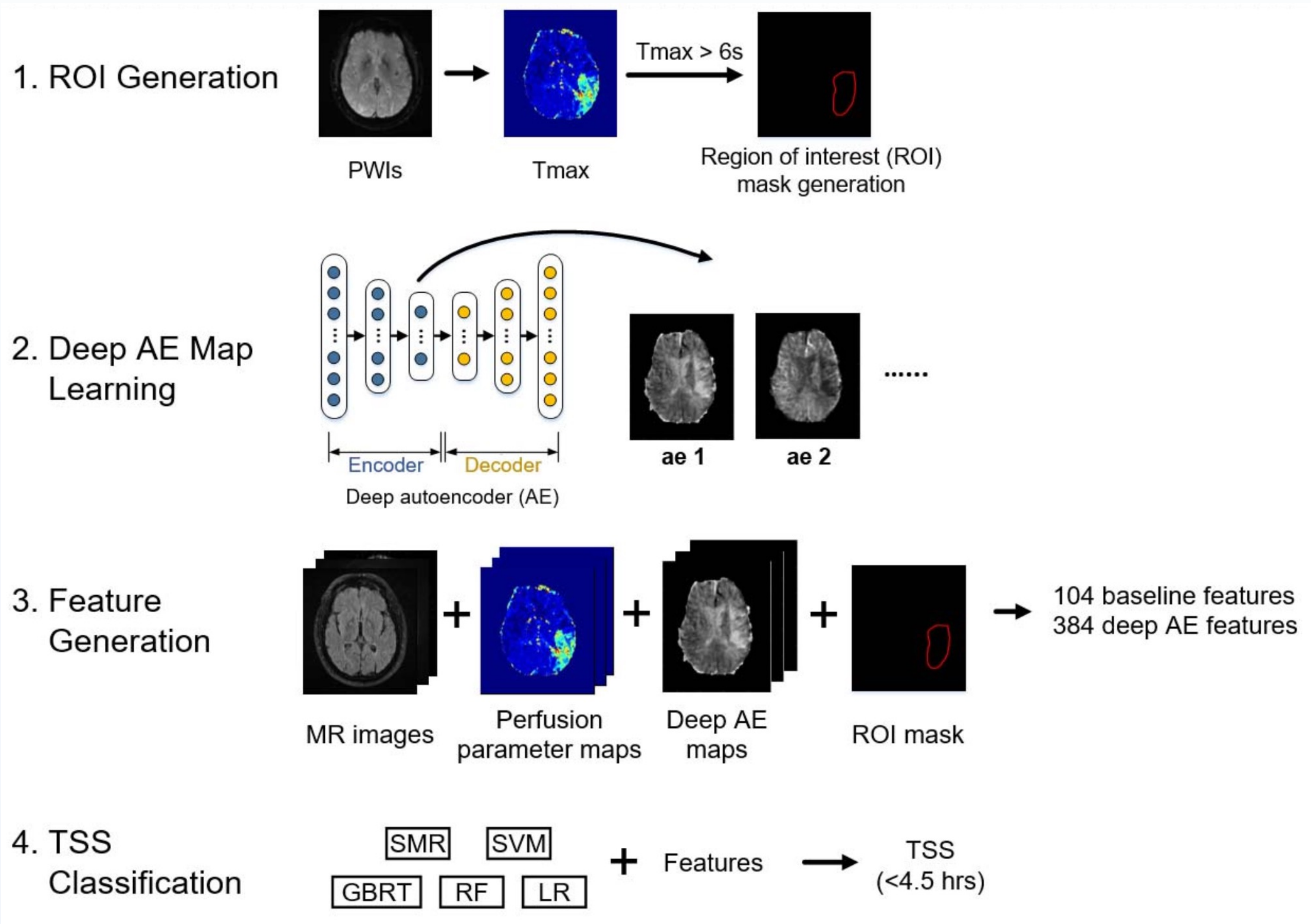
INTRODUCTION & OBJECTIVES

Stroke is the 2nd **most common cause of death** and the 3rd **most common cause of long-term disability** worldwide. The usual treatment consists of thrombolysis by administration of tPA (tissue plasminogen activator) within <4.5h after onset of symptoms (*time-since-stroke, TSS*). Approximately ⅓ **of patients have to be excluded** from thrombolysis treatment as TSS cannot be determined (e.g. wake-up strokes, unwitnessed strokes). These patients receive special magnetic resonance (MR) imaging (DWI-FLAIR) in order to determine if tissue with reversible brain damage is still left. However, DWI-FLAIR-mismatch method only has a moderate inter-observer-agreement as it requires extensive clinical training. Therefore, a machine learning model was introduced in order to improve patients' outcome and overcome the difficult DWI-FLAIR evaluation. The presented results are the intellectual property of Ho *et al.*



APPROACHES & METHODS

Imaging Feature Generation: Areas that take >6s in order for the contrast bolus to arrive (T_{max}), are marked as a region of interest (ROI). These areas consist of brain tissue with both reversible and irreversible damage (**ROI generation**). The given images are then used for generation of descriptive statistics and morphological features (**baseline imaging feature generation**). A deep autoencoder (AE) that is based on a stacked autoencoder was implemented to learn hidden features from perfusion weight images (PWI). The learning process is automatic and aggregated an "**AE feature map**" (**deep imaging features generation**).



Experimental Setup: The deep AE was implemented in Torch7. The model was trained with images based on 131 cases. The optimal number of hidden layers and hidden units was determined for different coupling patch types (see table below).

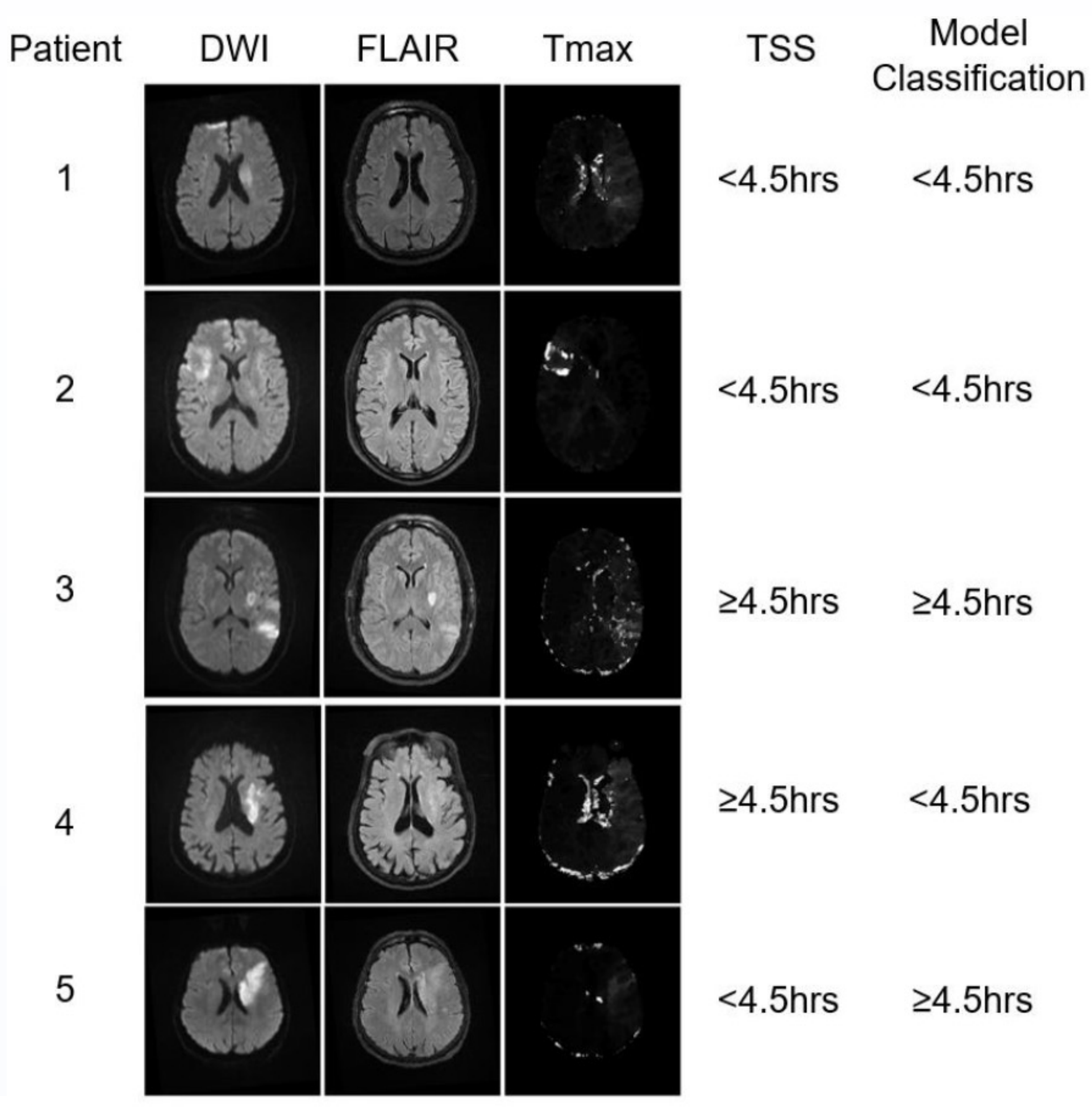
Coupling Patch Type	Optimal AE Architecture (# of hidden units/layer)	Optimal Mean Square Error (MSE; Average Deep AE MSE)
Arterial Input Function (AIF) patch only	1152-192-32-32-192-1152	0.606 (1.54)
Contralateral patch only	1152-288-32-32-288-1152	1.160 (1.95)
AIF + Contralateral patch	1728-288-32-32-288-1728	1.060 (4.49)

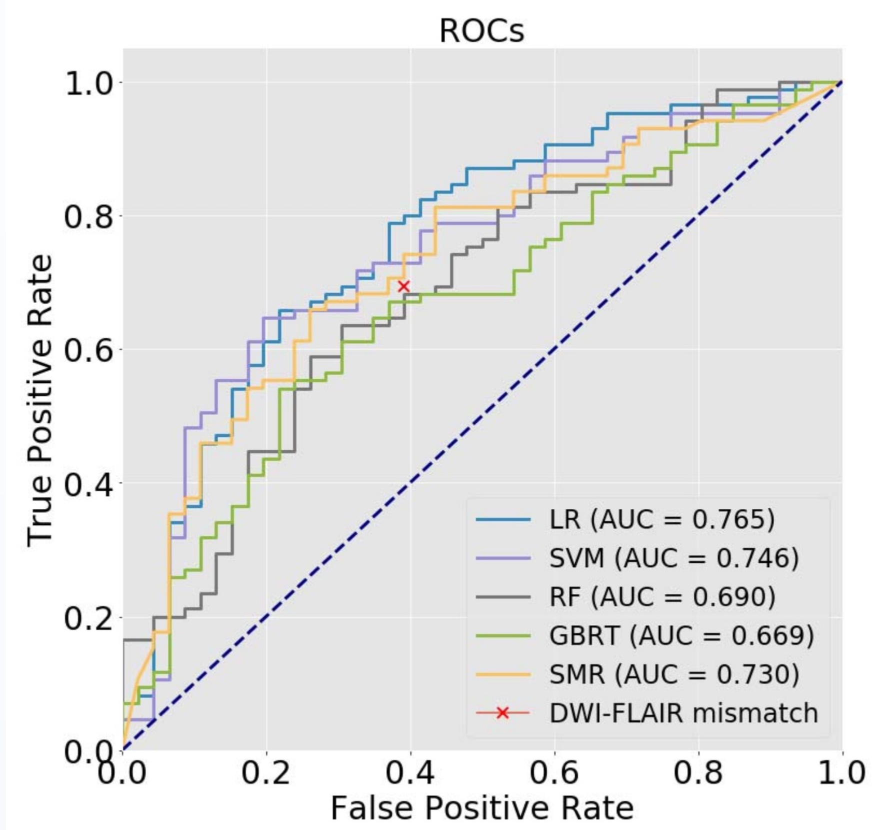
Evaluation: To understand what the deep features represented, correlations between baseline imaging features and deep features were evaluated (**feature correlation analysis**). **TSS subgroup classification analysis** showed consistency regarding the year of imaging (2011-2017) and different MR field strengths (e.g. 1.5T vs. 3.0T).

RESULTS

- T_{max} >6s provided ROIs that resulted in optimal performance in all classifiers with baseline and deep AE features.
- The correlation analysis demonstrates that the deep AE features capture a variety of complex representations (e.g. shape, morphology) that led to better TSS classification.
- The subgroup analysis shows that the classifiers were robust to both field strength and year of image acquisition.
- Deep AE features were able to improve the performance of almost every classifier and threshold combination. With the combination of baseline imaging features and deep features, all classifiers–except the GBRT trained with AIF coupling patch–showed an improvement of AUC.

Overall, the proposed model can potentially serve as an **alternative to the current standard** as it outperforms DWI-FLAIR-mismatch method in the given examples.





With the combination of **baseline imaging, deep features and AIF + contralateral coupling**, 3 classifiers–logistic regression (LR), stepwise multilinear regression (SMR) and support vector machine (SVM)–achieved higher sensitivity in classifying TSS than the DWI-FLAIR-mismatch method.

The highest AUC was achieved by LR with an AUC=0.765.

CONCLUSIONS

In this work a machine learning model was developed that was able to distinguish strokes of unknown TSS supposedly better than the current gold standard, the DWI-FLAIR-mismatch method. Compared to previous models, it bears the advantage that due to correlation analysis, it is possible to comprehend how the model–given the data–comes to its conclusions. Thereby, the deep learning approach does not have to be declared as an inapprehensible "black box", which previously engendered doubt in medical professionals leading to a low acceptance rate.

This shows, that machine learning and deep learning potentially set a **tool for clinical decision making**. In the future this could possibly be used in order to help doctors of every level of training to decide whether to lyse a patient or not even if the TSS is unclear. In addition to that, this machine learning model could not only be applied to neurologic cases: with modifications it could also be applied to other medical imaging data such as cardiac perfusion-weighted imaging.

DISCUSSION

Even though the results of this study sound promising, the question remains whether this method could hold up to daily clinical practice. I am wondering if the detection rate of the DWI-FLAIR-mismatch method would have been better if not only one but multiple doctors were allowed to evaluate the stroke images. As the authors themselves stated in the paper, the "inter-observer agreement" was only moderate. Therefore, other doctors would have potentially assessed the MR images differently, possibly outperforming the machine learning model. In addition to that, the amount of data that was used, was quite small.–Only image sets of 131 cases were used.

REFERENCES

[1] Ho KC, Speier W, Zhang H, Scalzo F, El-Saden S, Arnold CW. 2019. A Machine Learning Approach for Classifying Ischemic Stroke Onset Time From Imaging. IEEE Trans Med Imaging, 38 (7):1666-1676.