The role of predictive algorithm selection on the accuracy of MRI-based prediction of tissue outcome after acute ischemic stroke

Mark JRJ Bouts[1], Elissa McIntosh[1], Raquel Bezerra[1], Izzuddin Diwan[1], Steven Mocking[1], Priya Garg[1], William T Kimberly[2], Ethem M Arsava[1], William A Copen[3], Pamela W Schaefer[3], Hakan Ay[1], Aneesh B Singhal[2], Bruce R Rosen[1], Rick M Dijkhuizen[4], and Ona Wu[1]

[1]Athinoula A. Martinos Center, Dept Radiology, Massachusetts General Hospital, Charlestown, Massachusetts, United States, [2]Dept Neurology, Massachusetts General Hospital, Boston, Massachusetts, United States, [3]Dept Radiology, Massachusetts General Hospital, Boston, Massachusetts, United States, [4]Biomedical imaging & Spectroscopy group, Image Sciences Institute, University Medical Center Utrecht, Utrecht, Utrecht, Netherlands

Target audience: Clinician research scientists, neuroimaging researchers

Purpose: Multiparametric MRI may play an important role in selection of patients that may benefit from thrombolytic treatment after an acute ischemic stroke.[1] Particularly measurements of differences in abnormal perfusion and diffusion tissue volumes (`perfusion-diffusion` mismatch) are increasingly applied, but may overestimate or oversimplify identifying tissue at risk of infarction.[2,3,4] It has therefore been postulated that predictive algorithms that on a voxel-wise basis integrate multiple MRI parameters within a single, quantitative index may more favorably assess tissue at infarction risk.[4,5,6,7] Nevertheless, few studies have actually evaluated predictive algorithms within a similar setting or directly compared their potential.[8] We therefore evaluated the efficacy of several algorithms in predicting tissue outcome after stroke, by comparing acute measures of infarct risk with subsequent outcome in a cohort of acute ischemic stroke patients.

Materials & Methods: Patients with acute ischemic stroke that received MRI within 12 hours after last known well and follow-up imaging (>=4 days) who did not receive revascularization intervention nor novel therapeutics were retrospectively analyzed for infarct prediction comparison. MRI included: diffusion-tensor imaging, to obtain trace diffusion-weighted images (DWI, b=1000 mm^2/s), T$_2$-weighted (T2-w, b=0 mm^2/s) images, and maps of the apparent diffusion coefficient (ADC), and dynamic susceptibility contrast-enhanced MRI, to obtain measures of cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT), and T$_{max}$ using oscillation index regularized deconvolution with an automatically selected arterial input function as reference.[9] All images were co-registered to one another using a semi-automated program (MNI Autoreg).[10] Measures of T$_2$-w, DWI, ADC, CBF, CBV, MTT, and T$_{max}$ were subsequently normalized (and expressed as relative) to unaffected contralateral white matter and used for tissue outcome prediction. Follow-up FLAIR images were used to determine tissue infarction (F/U). Infarct tissue volumes were manually outlined. Acute DWI lesion volumes were manually determined, for acute T$_{max}$ lesion volumes a threshold of 6s was used.[2] Four algorithms were selected for predictive comparisons. Prediction was either based on a separating hyperplane in parameter space, dichotomizing the data in non-infarcted and infarcted voxels, or ensembles of prediction trees which combined prediction estimated infarct probability. A separating hyperplane was either derived using linear (generalized linear model (GLM)) or non-linear (generalized additive model (GAM)) regression. Adaptive boosting (ADA) sequentially trained and adaptively weighted the input and output of a predefined number of regression trees (125 trees) using a logistic loss function to obtain an estimate of infarction risk. A fourth algorithm calculated multiple decision trees based on subsets of the original training data where each tree was created using bootstrap sampling with replacement and random feature selection. The combined effort of these trees then provided an estimate of infarction risk (random forest (RF)).[11] All algorithms were trained based on a balanced dataset of infarcted and non-infarcted voxels. Leave-one-out testing was used to compare predictive performance. Predictive performance was compared to F/U volumes in terms of correctly predicted infarcted (TP), or non-infarcted (TN) voxels, or falsely predicted infarcted (FP), or non-infarcted (FN) voxels and expressed in measures of sensitivity (sens=TP/TP+FN) specificity (spec=TN/(FP+TN)), area under receiver operator characteristic curve (AUC), and Dice's similarity index (DSI=(TP+TP)/(TP+FN+FP+TP)). Predicted lesion volumes (PLV) were determined using a ≥50% risk threshold. Repeated measures analysis of variance with post-hoc Tukey testing and intra class correlation coefficient (ICC) were used for statistical testing. P<0.05 was considered significant.

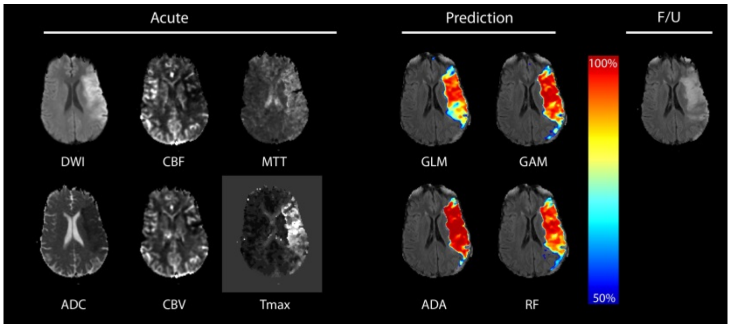

Figure 1: Acute diffusion and perfusion images used for GLM, GAM, ADA, and RF-based predictions overlaid on follow-up FLAIR. F/U FLAIR depicts eventual outcome (hyperintensity). Risk maps were thresholded ≥50%, with color-coding representing the level of infarct probability.

Results: Patients with a follow-up infarct volume ≥ 1cm^3 were selected for analysis (N=111; 66% Male; median 68 y IQR [55-77]). Figure 1 illustrates an example of GLM, GAM, ADA, and RF-based predictions, with corresponding tissue outcome at follow-up. Median [IQR] F/U volumes were larger than acute median DWI lesion volumes (F/U: 18.3 cm^3 [5.66-52.7] vs. DWI 12.7 cm^3 [3.1-33.2],P<0.001) but significantly smaller than acute median perfusion lesion volumes (40.5 [12.7-85.5],P<0.001). For all PLV, ICC with F/U volumes was higher than DWI (ICC=0.73) or perfusion based volumes (ICC=0.55). However GAM overestimated the tissue infarct volume (GAM: ICC=0.79; 31.9 cm^3 [17.1-69.7], P=0.02). Particularly ADA showed improved prediction of the extent of tissue infarction (ICC=0.80, 24.1 [14.7-63.8], P=0.78), also reflected by highest AUC and DSI, significantly higher than all other algorithms (Table 1).

Algorithm:	GLM	GAM	ADA	RF
ICC	0.80	0.79	0.80	0.80
Sensitivity	0.57 [.37-.78]	0.61 [.44-.77]	0.61 [.47-.75]	0.65 [.47-.79]
Specificity	0.94 [.89-.96]	0.92 [.87-.95]	0.94 [.90-.97]	0.93 [.87-.95]
AUC	0.85 [.78-.92]*	0.85 [.77-.91]*	0.87 [.80-.93]	0.86 [.78-.92]*
DSI	0.40 [.15-.53]*	0.36 [.15-.55]*	0.40 [.17-.56]	0.38 [.16-.55]*

Table 1. Predictive performance (median [IQR]) *P<0.01 versus ADA

Discussion: In this study we compared four different predictive algorithms to predict tissue infarction in patients with a `natural history` of stroke. Our results show that, in line with previous studies, predictive algorithms show improved prediction in the extent of infarction after stroke over single parameter based methods.[4] Yet where previous experimental stroke studies reported only minor differences in prediction performance,[6,7] we observed significantly improved predictions for more complex algorithms. These differences may originate from a more heterogeneous training set. Human stroke may develop more heterogeneously compared to development in well-defined experimental stroke models and may therefore be better described using more complex predictive algorithms. This study was limited by its retrospective and single center design, therefore future research should aim at further exploring applicability and generalizability of these algorithms.

References: 1. Donnan *et al* Lancet Neurol 2009; 8: 261-9. 2. Albers *et al* ANN Neurol. 2006; 60:508-17. 3. Sobesky *et al* J. Cereb. Blood Flow Metab. 2012; 32:1416-25. 4. Wu *et al* Brain. 2006; 129:2384-93. 5.Wu *et al* Stroke. 2001; 32: 933-42. 6. Huang *et al* Brain Res. 2011;1405:77-84 7. Bouts *et al* , J. Cereb. Blood Flow Metab. 2013; 33:1075-82. 8. Rekik *et al* Neuroimage Clin. 2012;1:164-78. 9. Wu *et al* Magn Reson Med. 2003; 50: 164-174. 10.Collins et al J. Comput Assist Tomo. 1994; 18:192-205. 11. Bishop.(ed) Pattern Recognition and Machine Learning. 2007 Springer: New York, NY, USA.