



Fully automated stroke tissue estimation using random forest classifiers (FASTER)

Richard McKinley^{1,*}, Levin Häni^{1,**}, Jan Gralla¹, M El-Koussy¹, S Bauer², M Arnold³, U Fischer³, S Jung³, Kaspar Mattmann¹, Mauricio Reyes^{2,**} and Roland Wiest^{1,**}

Abstract

Several clinical trials have recently proven the efficacy of mechanical thrombectomy for treating ischemic stroke, within a six-hour window for therapy. To move beyond treatment windows and toward personalized risk assessment, it is essential to accurately identify the extent of tissue-at-risk (“penumbra”). We introduce a fully automated method to estimate the penumbra volume using multimodal MRI (diffusion-weighted imaging, a T2w- and T1w contrast-enhanced sequence, and dynamic susceptibility contrast perfusion MRI). The method estimates tissue-at-risk by predicting tissue damage in the case of both persistent occlusion and of complete recanalization. When applied to 19 test cases with a thrombolysis in cerebral infarction grading of 1–2a, mean overestimation of final lesion volume was 30 ml, compared with 121 ml for manually corrected thresholding. Predicted tissue-at-risk volume was positively correlated with final lesion volume ($p < 0.05$). We conclude that prediction of tissue damage in the event of either persistent occlusion or immediate and complete recanalization, from spatial features derived from MRI, provides a substantial improvement beyond pre-defined thresholds. It may serve as an alternative method for identifying tissue-at-risk that may aid in treatment selection in ischemic stroke.

Keywords

Acute stroke, magnetic resonance perfusion, magnetic resonance diffusion imaging, endovascular therapy, mathematical modeling

Received 26 February 2016; Revised 5 August 2016; Accepted 9 August 2016

Introduction

During the last two decades, multiple attempts have been made to predict the fate of ischemic tissue under risk for infarction due to occlusion of the afferent vessel. Advanced neuroimaging techniques aim to identify the so-called “ischemic penumbra”: the severely hypoperfused, neurophysiologically silent brain tissue that is potentially salvageable if reperfused early enough.¹ Non-invasive measurements of tissue-at-risk after ischemic stroke have used H₂¹⁵O-PET and MRI as surrogate markers to define the thresholds between functional impairment and irreversible cell death.² However, varying perfusion and diffusion (ADC) thresholds have been suggested in the literature without consensus. A Tmax of >6 s has been shown to be a predictor of severely hypoperfused tissue,³ has been used in large-scale trials,^{4–6} and is considered the MR standard for estimating

tissue-at-risk in many centers.⁷ Similarly, thresholding of the apparent diffusion coefficient (ADC) is frequently used for semi-automated post-processing to identify the infarct core. A threshold of $600 \times 10^{-6} \text{ mm}^2/\text{s}$ has been

¹Support Center for Advanced Neuroimaging (SCAN), Institute for Diagnostic and Interventional Neuroradiology, Inselspital, University of Bern, Bern, Switzerland

²Institute for Surgical Technology and Biomechanics, University of Bern, Bern, Switzerland

³Department of Neurology, Inselspital, University of Bern, Bern, Switzerland

*These authors contributed equally to the study.

**These authors share senior authorship.

Corresponding author:

Levin Häni, Support Center for Advanced Neuroimaging (SCAN), Institute for Diagnostic and Interventional Neuroradiology, Inselspital, University of Bern, Freiburgstrasse, Bern 3010, Switzerland.
 Email: Levin.Haeni@insel.ch

suggested as optimal,⁸ but it remains a matter of debate which threshold to apply.

Fixed thresholds have been widely used, for example in the automated tool RAPID.⁸ However, voxel-wise thresholding of a single parameter map is an inaccurate method for identifying tissue-at-risk. Tissue indicated as unsalvageable may be salvaged (“pseudonormalization”), and tissue indicated as being at risk can survive even in the absence of adequate reperfusion (“benign oligemia”). Tissue-at-risk estimations are prone to error in about 25% of patients, with variations up to 100 ml, when varying the algorithm underlying the analysis.⁹ Initial attempts have been made to develop automated predictive stroke models that directly classify the tissue-at-risk beyond this threshold-based imaging model.¹⁰ Early studies established the increased ability of multimodal data to predict damage, compared to single parameter models.¹¹ Models incorporating information about the spatial arrangement of hypoperfused voxels outperform those based on voxel-by-voxel prediction.^{12,13} Very recently, neural networks have been employed to predict tissue fate, allowing visual features predicting stroke tissue fate to be learned directly from training data.¹⁴

All these studies omit a crucial variable predictive of stroke tissue fate: namely the degree of recanalization. This aspect of stroke tissue fate prediction is made more vital by recent class I evidence that mechanical thrombectomy is a safe and effective therapy. Seven prospective studies (MR-CLEAN,¹⁵ ESCAPE,¹⁶ EXTEND-IA,⁴ SWIFT-PRIME,¹⁷ REVASCAT,¹⁸ THRACE, and THERAPY) demonstrated the superiority of mechanical thrombectomy in proximal vessel occlusions within 6 h after stroke onset. The availability of safe, reproducible, and reliable information about expected tissue salvage would not only allow neuroradiologists to select patients that would benefit from mechanical thrombectomy, but also enable the selection of patients for revascularization in a time window that exceeds 6 h if sufficient collateral flow enables sustained tissue survival. In this case, it is crucial that a prediction of tissue damage can be made in both the presence and absence of successful reperfusion. The degree of success of the mechanical thrombectomy depends on a variety of clinical factors which cannot be inferred from imaging data (such as time to recanalization, experience of the interventionalist, or comorbidity) and therefore the outcome for a given patient cannot be reliably estimated. However, an estimate of the lesion in the event of favorable response to therapy / an estimate of the lesion in the event of unfavorable response to therapy can be given, and used as a tool to drive decision making.

An important advance in personalized stroke prediction was made in a recent study by Kemmling et al.¹⁹

A logistic regression model, incorporating recanalization status, time to recanalization and CT perfusion imaging data, was used to predict the extent of stroke tissue damage from CT imaging. The model parameters show an 83% reduction in the odds of infarction in the event of successful recanalization for each voxel, holding all other variables constant. By introducing an interaction variable between time to recanalization and recanalization status, the authors were able to show that the odds for infarction increased by 18.9% for every additional hour until thrombolysis in cerebral infarction (TICI) 0–2a recanalization, and by 33.2% for every additional hour until TICI 2b-3 recanalization.

This study raises new questions, the first of which is the added ability of similar methods to predict infarction in the case of successful or unsuccessful revascularization, using MR perfusion and MR diffusion-weighted imaging rather than CT perfusion. Secondly, the role of the interaction between recanalization status and imaging is not uniform across the affected tissue: one would expect that the effect of rapid reperfusion will be negligible on voxels lying within a large infarct core, and felt most in tissue lying inside the penumbra but sustained by sufficient collateral flow.

In light of these questions, we have developed a predictive framework for acute stroke based on multimodal MR imaging. Training of the multilinear model in the study mentioned above is based on a dichotomization of patients into “near complete” recanalization (TICI2b-3) versus “less than near complete” recanalization (TICI 0–2a). As a consequence, some voxels belonging to a patient with TICI 2a recanalization (classified as “less than near complete”) may experience re-established blood flow, leading to underestimation of the tissue-at-risk during testing. Following Wintermark et al.,²⁰ we instead dichotomize our training data not into complete recanalization (TICI 3) or permanent occlusion (TICI 0), omitting all intermediate TICI grades from training.

The basis of this framework is two predictive models, one predicting the outcome in the event of a good response to therapy, and the other predicting the natural course of the stroke. The prediction is made on a voxel-by-voxel basis but is made for each voxel considering features derived from the surrounding tissue. By training separate predictive models for these two cases, the effect of recanalization can depend, nonlinearly, on the perfusion and diffusion characteristics of the affected tissue.

It is our hypothesis that prediction of the extent of final infarction using compound spatial information from multimodal MRI can provide a more informative definition of the tissue-at-risk than approaches based solely on voxel-wise delineation of the infarct core and salvagable penumbra using linear thresholds,

and that these predictions can be used as a tool to aid treatment selection.

Materials and methods

Patients

The study utilizes anonymized data from the Bernese stroke registry, a prospectively collected database approved by the Kantonale Ethikkommission Bern, some aspects of which have been reported previously.^{21–24} All patients were treated for an acute ischemic stroke at the University Hospital of Berne between 2005 and 2013. The study was performed according to the ethical guidelines of the Canton of Bern (Swiss Humanforschungsgesetz) with approval of our institutional review board (Kantonale Ethikkommission Bern).

Patients were included in this analysis if: (i) a diagnosis of ischemic stroke was established by MR imaging with an identifiable lesion on DWI and perfusion imaging, (ii) a proximal occlusion of the middle cerebral artery (M1 or M2 segment) was documented on digital subtraction angiography, (iii) endovascular therapy was attempted, either by intra-arterial thrombolysis (before 2010) or by mechanical thrombectomy (since 2010), (iv) pre-treatment MRI was performed with sufficient quality (i.e. no motion artefacts), (v) the imaging data were recorded completely into the picture archiving and communication system, (vi) the patients had a minimum age of 18 years at the time of stroke. Patients were excluded if they received only purely diagnostic angiography. Patients with a stenosis or occlusion of the carotid artery were excluded as well. Revascularization success was stratified retrospectively according to the TICI score by two examiners blinded for clinical data.²⁵ Stroke severity for these patients was assessed at admission according to the National Institutes of Health Stroke Scale (NIHSS) scale. We aimed to identify all patients with a three-month axial T2-weighted follow-up image in order to define the final extent of infarction.

Imaging protocols

The stroke MRI was performed on either a 1.5T (Siemens Magnetom Avanto) or 3T MRI system (Siemens Magnetom Verio). The stroke protocol encompassed whole brain DWI, (24 slices, thickness 5 mm, repetition time 3200 ms, echo time 87 ms, number of averages 3, matrix 256×256 , flip angle 90°) yielding images for b values of 0 s/mm^2 and 1000 s/mm^2 as well as ADC maps that were calculated automatically. Standard dynamic susceptibility contrast-enhanced perfusion MRI (gradient-echo echo-planar imaging sequence, repetition time 1410 ms,

echo time 30 ms, field of view $230 \times 230 \text{ mm}$, voxel size: $1.8 \times 1.8 \times 5.0 \text{ mm}$, slice thickness 5.0 mm, 19 slices, 80 acquisitions, flip angle 90°) was acquired. Images were acquired during the first pass of a standard bolus of 0.1 mmol/kg gadobutrol (Gadovist, Bayer Healthcare). Contrast medium was injected at a rate of 5 ml/s followed by a 20 ml bolus of saline at a rate of 5 ml/s. In addition, an axial T2-weighted turbo-spin echo sequence (TR 3760–4100 ms, TE 85–100 ms, flip angle 150°) and contrast-enhanced T1-weighted sequence (1.5T system: spin-echo sequence (TR 663 ms, TE 17 ms, flip angle 90°), 3T system: gradient-echo sequence (TR 250 ms, TE 2.67 ms, flip angle 70°)), a time-of-flight angiography and a first pass Gd-MRA were acquired, with T2-weighted imaging and TOF angiography performed before contrast injection.

Data processing and ground truth generation

Perfusion maps (TTP, CBF, CBV, Tmax) were obtained by block-circulant singular value decomposition using the Perfusion Mismatch Analyzer (PMA, from Acute Stroke Imaging Standardization Group ASIST) Ver.3.4.0.6. The perfusion data were denoised both spatially and temporally using a Gaussian filter. The time-concentration-curves were generated for each pixel from the time-intensity-curve. The AIF location was determined automatically by the software. The AIF curve was not rescaled using the venous output function. A detailed description of the perfusion analysis is described in the publication by Kudo et al.²⁶ All MR sequences and their derived maps (T1contrast, T2w, ADC, CBF, CBV, TTP, Tmax, and the follow-up T2w) were registered to the pre-treatment T1 contrast image of the same patient. Calculated perfusion maps were registered by initially registering the first timepoint of the DSC perfusion sequence to the T1 post-contrast volume, and then using the resulting transformation to register the maps. Registration was performed using a rigid registration model with the mutual information similarity metric as implemented in ITK (“VersorRigid3DTransform” with “MattesMutualInformation” similarity metric and three multi resolution levels). All maps were automatically skull-stripped,²⁷ then resampled to 2 mm isotropic resolution in a standardized axial orientation with a linear interpolator. No attempt was made to put the individual patients in a common reference space.

Delineation of lesions was performed using Slicer 3D Version 4.3.1. For each patient, a manual segmentation of the diffusion restriction and the hypoperfused tissue was performed, beginning with a threshold ($\text{ADC} < 600 \times 10^{-6} \text{ mm}^2/\text{s}$ in the case of diffusion, and $\text{TMax} > 6 \text{ s}$ in the case of perfusion), which was adjusted on a patient-by-patient basis in accordance

with the opinion of the expert segmenter, to ensure that all affected tissue was correctly labeled. This was followed by a dilation and erosion. Due to considerable variation and overlap of radiologic features affecting the deep gray matter nuclei, resulting in incoherent water motion and restricted diffusion, a reduction of ADC values in white matter and basal ganglia is not always ischemic in its nature (e.g. due to calcification, systemic metabolic abnormalities or iron deposits). To avoid erroneous classifications arising from this phenomenon, in diffusion lesions adjacent to the caudate nucleus, tissue with ADC below $600 \times 10^{-6} \text{ mm}^2/\text{s}$ was excluded from the lesion if tissue in the contralateral caudate nucleus was also below $600 \times 10^{-6} \text{ mm}^2/\text{s}$. The rater also excluded sulci, ventricles, previous infarcts, and imaging artefacts.

For the determination of outcome, the “true infarct core” was segmented manually on the registered T2-weighted 90-day follow-up images. Manual regions of interest were drawn to the maximal extent of the final infarction, including areas with hemorrhagic transformation, but excluding regions already hyperintense on acute T2 imaging. The boundaries of the infarctions were manually delineated for every single transversal slice. The 90-day follow-up lesion was chosen as the definition of final infarction, rather than the lesion in the subacute phase of lesion evolution, since apparent lesion size in the subacute phase is known to overestimate final lesion volume.²⁸ T2 was chosen as the modality for identifying the final lesion, since it was more widely available than a FLAIR follow-up image in the retrospective data used.

All segmentations were done by two independent raters and inter-rater agreement was checked using Bland–Altman’s method.²⁹

Segmentation forest classifiers

We have recently introduced an algorithm called Segmentation forests,³⁰ a variation on the well-known Random Forests algorithm: in both algorithms, the resulting classifier is a decision forest, in which the final classification is derived from combining the votes of many decision trees, each built on random samples of the training data.^{31,32} In the Random Forests approach to medical image segmentation, a forest of decision trees is built, with each tree being trained on a random sample of the voxels contained in the image. This random sample is made over data from all the patient cases in the training data. By contrast, in Segmentation Forests, the random sampling occurs at two levels: first at the patient level, then at the voxel level. Each tree is therefore constructed on only a subset of the patient cases. This has two advantages: first, since voxel-wise data are clustered at the patient

level, this form of random sampling produces more representative random samples than sampling without patient-level stratification, reducing the variance of the classifier without increasing bias. The second advantage is that, since each tree is trained only on a subset of the patient cases, the remaining “out of bag” cases can be used to tune the classifier, by estimating the cutoff for the classifier at which a metric (for example, mean Dice coefficient) is optimized. In this study, the metrics used for establishing the cutoff were the mean Dice score, (used to establish the cutoff for the lesion prediction in the case of successful reperfusion) and the mean F2 score (defined analogously to the Dice score, but penalizing false negatives twice as much as false positives), for the lesion prediction in the case of unsuccessful reperfusion. By maximizing the F2 metric instead of the Dice, we increase the volume of the predicted infarction, to ensure that in most cases the final infarction lies within the predicted lesion. Further details of the Segmentation Forest Classifier are available in the Supplementary Material (Appendices A and B).

Segmentation forests was the algorithm used by our group to produce the top result in the MICCAI ISLES challenge (www.isles-challenge.org/), reproducing the segmentation of the penumbra produced by a manual rater. Here, we use the same framework to predict the final infarction directly.

Feature extraction

The features used were adapted from the BraTumIA tool^{33,34} and a pilot study on stroke lesions.³⁵ There are two categories of features: 1. Features extracted from local histograms, and from the gradient magnitude of the image. 2. Global features, extracted by registering the T1c volume to a T1-weighted atlas using ITK affine registration to extract atlas coordinates. Using keypoints in the atlas space, the mid-sagittal plane of the brain was extracted. The corresponding point lying on the other side of the mid-sagittal plane was found, and a symmetry feature calculated for each modality, by calculating the normal first smoothing (using a `SmoothingRecursiveGaussianImageFilter` from ITK) and then subtracting the value of the smoothed modality at the voxel from the value at the mirrored voxel.

In total, 274 features were calculated for each voxel. Since the extracted TMax maps contained outliers (extremely high biologically meaningless values), the Tmax maps were clipped to lie within a range [0,20s] (i.e. Tmax values above 20 were set to 20). Since CSF typically exhibits ADC in the range $3000\text{--}4000 \times 10^{-6} \text{ mm}^2/\text{s}$, the ADC maps were also clipped to be within a range $[0, 2600] \times 10^{-6} \text{ mm}^2/\text{s}$ to allow greater ability to discriminate ADC values within the parenchyma. To

ensure efficient calculation of the histograms used to calculate feature maps, all modalities were linearly scaled to integer values [0,255] (after clipping) as described in Meier et al.³⁴

Supervised learning models

In order to guarantee the quality of the training data available, only cases with clear response to endovascular therapy (TICI 0 or 3) were used to train the predictive classifiers. These classifiers, trained on relatively small (<20) cohorts, performed well at identifying what tissue would infarct within the penumbra. However, performed less well at identifying the site of the stroke itself, occasionally identifying tissue contralateral to the stroke as being at risk of infarction. Our contribution to the ISLES challenge, on the other hand, had previously been shown to provide a localization of the hypoperfused territory. We therefore developed a hybrid approach, in which segmenting classifiers were first used to locate the rough location of the stroke, trained on manual segmentations, after which predictive classifiers predict what tissue within that territory will go on to infarct, trained on the final lesion in a three-month follow-up. For a visual summary of the pipeline, see Figure 1.

An initial segmentation of the tissue is provided by the output of two segmenting classifiers, FASTER_{ADC} and FASTER_{Tmax}, which attempt to reproduce the manual delineation of the affected tissue, as described in the Data processing and ground truth generation Section. Two predictive classifiers are then used to predict which tissue will go on to infarct: a classifier FASTER₋, which predicts the lesion extent in the case of no reperfusion, and a classifier FASTER₊, which predicts the lesion extent in the case of full reperfusion.

Given a new case, voxels are classified by fully automated stroke tissue estimation using random forest classifiers (FASTER) as follows:

- If a voxel is outside of the lesions given by both FASTER_{ADC} and FASTER_{Tmax}, it is labeled “no risk.”
- If a voxel is inside the lesion defined either by FASTER_{ADC} or by FASTER_{Tmax}, it is considered within the hypoperfused area, and classified as follows:
 - If the voxel is outside the lesions indicated by both FASTER₋ and FASTER₊, it is labeled “no risk.”
 - If the voxel is inside the lesion defined by FASTER₊, it is labeled “unsalvageable.”
 - If the voxel is outside the lesion defined by FASTER₊, but outside the lesion defined by FASTER₋, it is labeled “salvageable.”

FASTER₋ was trained on cohort of patients who had little-to-no reperfusion (“no reperfusion”): those receiving a TICI grade of 0. FASTER₊ was trained on a cohort of patients who had “complete reperfusion” (TICI grade 3) and whose imaging features reflected a non-malignant ischemic profile. A malignant profile was defined as a patient presenting with one or more of: an ADC lesion of volume larger than 70 ml, a ratio of perfusion to diffusion lesion smaller than 1.8, or a region of extreme hypoperfusion (Tmax > 10 s) greater than 100 ml.¹⁷

Data not falling into the selection criteria for training were retained, and used to provide a large and representative testing set for the classifiers. Models were built using H2O v. 2.8.6.2, a predictive analytics tool developed by H2O.ai (Mountain View, California), on a machine running Windows 8.1 with 32 Gb of memory.

Data analysis

Area-under-curve analysis

The area under the receiver operating characteristic (ROC) curve (AUC) is a well-known measure of the discriminative power of classifiers. However, as noted by Jonsdottir et al.,³⁶ the large number of unaffected (i.e. neither diffusion – nor perfusion-restricted) voxels in stroke can lead to an artificially high AUC. To avoid a bias, Jonsdottir et al. suggest measuring also a quantity called AUC_R: the area under the ROC curve for the region consisting only of perfusion-restricted voxels. We computed ROC curves both globally (on voxels from all testing cases together) and on an individual level (for each patient in the testing set).

Analysis of the favorable/unfavorable case prediction

We used three standard measures of segmentation performance: sensitivity, specificity, and precision. Since the ratio of lesion to non-lesion tissue in the brain is rather low, specificity will be high even for rather simple techniques. Precision, on the other hand, is independent of class balance, and measures the probability that a voxel labeled as in the lesion class will be in the three-month follow-up lesion. It is therefore a good measure of the tendency of a method to overestimate the extent of likely tissue damage: a method with high precision labels fewer voxels incorrectly as being in the lesion class.

Results

Demographic data

By using the above-mentioned criteria, we identified 80 eligible patients from the Bernese stroke registry who

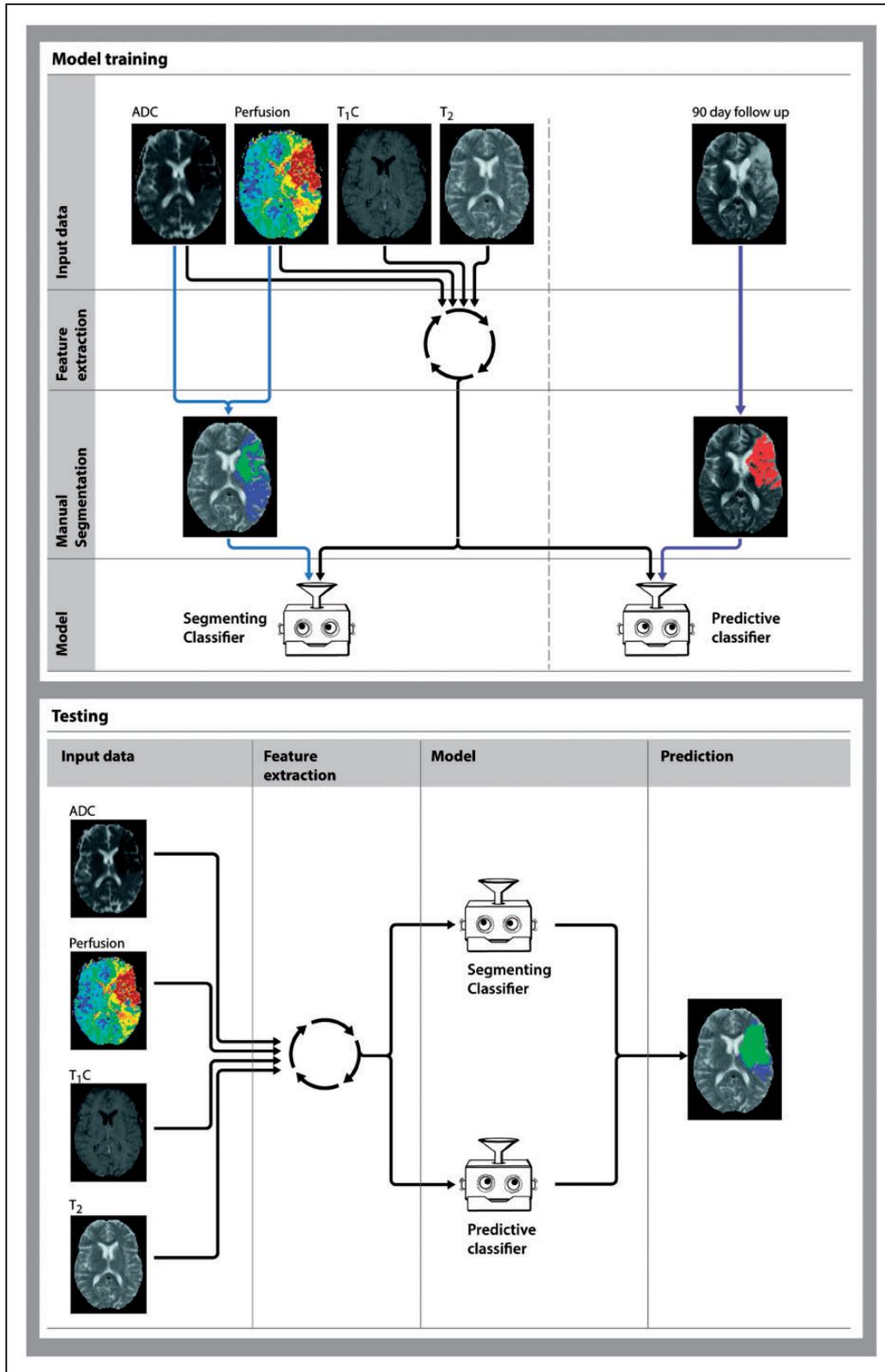


Figure 1. Workflow for final lesion prediction: In the training step, classifiers are built from training data to segment the ischemic penumbra and infarct core (as defined by a semi-manual segmentation on maps derived from perfusion and diffusion imaging) and to predict the fate of stroke tissue (as defined by a manually segmented follow-up image acquired 90 days after the manifestation of the stroke), using features extracted from the initial stroke MRI performed within 6 h of stroke onset. For every stroke patient, the output of segmentation and predictive models are fused into a single image, showing the predicted outcome in the event of favorable or unfavorable response to therapy in case of a successful vs. unsuccessful recanalization.

had a follow-up MRI scan at three months. We identified 20 further patients without a follow-up MRI scan, treated between 2011 and 2013, whose data were used exclusively to train the segmentation classifiers. Baseline and treatment characteristics of the different patient groups are given in Table 1.

Median initial NIHSS was 14, median time from symptom onset to MRI was 144 min and median time from symptom onset to intra-arterial revascularization attempt was 267 min. Among the 80 patients with follow-up MRI present, 10 were assigned a TIC1 grade of 0, 5 a TIC1 grade of 1, 14 a TIC1 grade of 2a, 31 a TIC1 grade of 2b, and 20 a TIC1 grade of 3. Defining successful revascularization as a TIC1 of 2b or 3, 63% of patients were successfully revascularized. Among those patients with a follow-up, 59 were treated with endovascular therapy before the 6 hour recommended window, and 20 at or beyond that window. In the 20 cases treated at or after 6h, 10 had a TIC1 grade of 0–2a, and 10 a TIC1 grade of 2b–3. In one case, the time to treatment was not available.

Median volumes for diffusion and perfusion restriction at initial MRI (as defined by semi-manual segmentation) were 12.52 ml and 138.70 ml, respectively. The median volume of the final infarction was 2.47 ml for those with TIC1 3 revascularization, 7.77 ml for those with intermediate TIC1 scores (1–2b), and 35.81 ml for those with a low TIC1 score of 0.

Interrater variability

Two raters manually segmented the lesion indicated on ADC and Tmax maps, calculated as described above. The results of the comparison, including inter-class correlation, and the mean Dice score between the two raters, are displayed in Table 2.

Prediction of tissue fate

Among the 20 cases having a three month follow-up and a TIC1 score of 3, five were found to have a malignant infarction profile. The 15 remaining patients in the three-month follow-up cohort who were completely revascularized (TIC1 score 3) and had initially a non-malignant infarction profile, and the 10 patients assessed as unsuccessful revascularization (TIC1 score 0) were used as training data. A favorable-outcome classifier (FASTER+) was trained on the data from patients experiencing total reperfusion, and an unfavorable-outcome classifier (FASTER–) was trained on the data from patients experiencing no revascularization.

In total, 25 cases out of the 80 with follow-up were used to train the classifiers FASTER+ and FASTER–. These classifiers were then applied to the remaining 55 cases with follow-up, of which 19 had a TIC1 grade

Table 1. Baseline and treatment characteristics of eligible patients.

	With follow-up (n = 80)	TIC1 0	TIC1 3	TIC1 1–2b	Without follow-up (n = 20)
Age, years	62.01	59.33	63.73	61.40	68.12
Female sex, n (%)	33 (41.25%)	3 (30%)	11 (55%)	17 (39.53%)	12 (60%)
Baseline NIHSS score, median (range)	14 (1–26)	12.5 (8–24)	14 (1–23)	14 (2–26)	15 (5–36)
Minutes from symptom onset to MRI, median (range)	151 (39–748)	203.5 (92–390)	114 (85–748)	177 (39–609)	139.5 (65–575)
Minutes from symptom onset to reperfusion, median (range)	262 (148–860)	318.5 (179–592)	213.5 (165–860)	278 (148–702)	N/A
Collaterals	1.27	1.00	1.40	1.30	N/A
Volume diffusion restriction on initial MR in ml, median (range)	11.67 (0.06–114.73)	17.38 (4.61–59.87)	8.64 (1.64–114.728)	11.70 (0.06–92.55)	14.90 (0.56–80.18)
Volume perfusion restriction on initial MR in ml, median (range)	140.78 (40.13–253.63)	139.20 (66.94–221.76)	184.16 (55.69–253.63)	138.45 (40.13–245.21)	134.38 (47.56–200.13)
Volume final lesion on follow-up MR in ml, median (range)	7.36 (0.23–141.20)	35.81 (4.49–114.30)	2.47 (0.23–141.2)	7.77 (0.37–121.25)	N/A

N/A: not applicable.

between 1 and 2a, and 36 had a TICI grade 2b or 3 (including among that 36 the malignant cases with TICI 3 which were excluded from training). To establish a baseline performance with which to compare the

Table 2. Interrater statistics for the manual segmentation of the DWI and PWI lesions.

	DWI	PWI
Inter-class correlation	0.999	0.983
95% confidence interval		
Lower	0.999	0.918
Upper	1	0.993
Significance	$p < 0.05$	$p < 0.05$
Dice score	Lesion in DWI	Lesion in PWI
Mean	0.96	0.89
Median	0.98	0.89
Standard deviation	0.04	0.04

classifiers underlying FASTER, we also built logistic regression models (as used by Wu et al.¹⁰ and more recently Kemmling et al.¹⁹) trained on same 25 cases used to train FASTER+ and FASTER-. The global ROC curves for these classifiers are displayed in Figure 2, plotted both over the whole brain, and over the region of interest defined by the hand-segmented tissue-at-risk (giving the AUC_R score of Jonsdottir et al.³⁶). The ROC curves were calculated using the R package pROC³⁷ using the roc command, with algorithm=2 and no smoothing. The mean individual AUC of FASTER+ over the testing set was 0.94 (sd=0.08), and the mean individual AUC of FASTER- over the testing set was 0.96 (sd=0.06). Meanwhile, the mean AUC of the logistic regression model was 0.85 (sd=0.08), the mean AUC of thresholding on voxel-wise ADC was 0.832 (sd=0.08), and the mean AUC of thresholding on Tmax was 0.86 (sd=0.09).

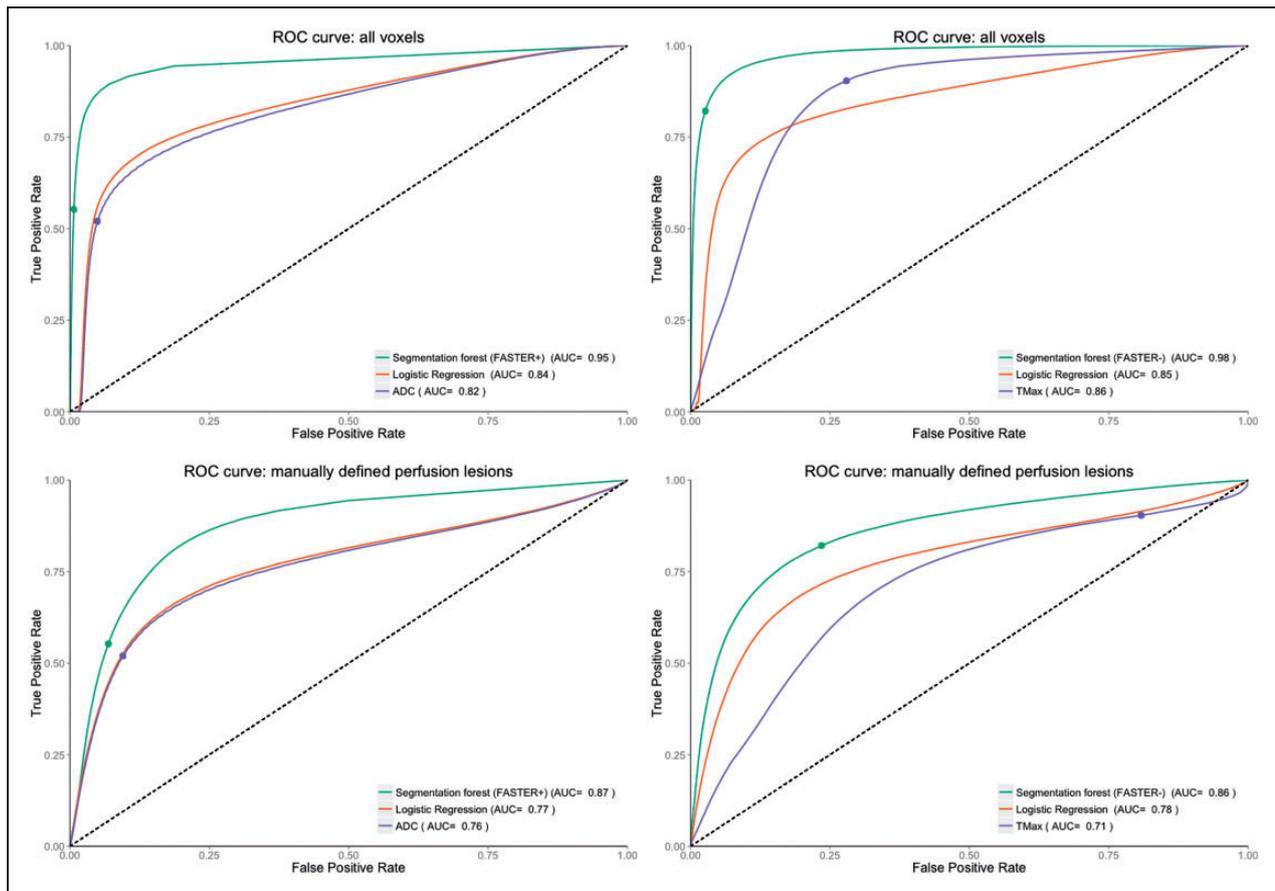


Figure 2. Receiver operating characteristic (ROC) curves for tissue fate prediction measured over all 55 testing cases. Upper row: ROC measured over all brain voxels. Lower Row: ROC measured in the manually defined perfusion lesion. Left-hand column: classifiers trained on successfully revascularized patients. Right-hand column: classifiers trained on unsuccessfully revascularized patients. Thresholds derived from the segmentation forest algorithm are indicated. For comparison, the ROC curve of thresholding on ADC (with a threshold of 600×10^{-6} mm²/s indicated) is shown on the left, and the ROC curve of thresholding on Tmax (with a Tmax threshold of 6 s indicated) is shown on the right.

Confidence intervals for the area under the ROC curve were also calculated, using DeLong's method in the pROC package, but these intervals were found to be unrealistically narrow: for example, the confidence interval of the AUC for FASTER+ was (0.915–0.928). This may be the result of applying a method assuming independent samples to datapoints from a relatively small number of highly correlated sources.

FASTER vs. linear thresholds

Accuracy of delineating the ischemic core

Over the 36 testing cases, we compared three-month follow-up lesion to the manually segmented DWI lesion and to the predicted lesion extent in the case of a favorable reperfusion. The manually segmented DWI lesion achieved a sensitivity of 0.52, a specificity of 0.99, and a precision of 0.47. FASTER had a sensitivity of 0.53, specificity of 0.99, and a precision of 0.56. The mean Dice score between the manually segmented lesion and the follow-up was 0.34 (sd=0.16), while the mean Dice score between the prediction of FASTER and the follow-up was 0.34 (sd=0.22).

Accuracy of delineating the tissue-at-risk

Over the 19 testing cases having a TICI 1–2a, we compared the three-month follow-up lesion to the manually segmented PWI lesion, and to the lesion extent predicted by our model given an unfavorable response to therapy. The manual delineation achieved a sensitivity of 0.84, a specificity of 0.998, and a precision of 0.14, while FASTER had a sensitivity of 0.77, a specificity of 0.998, and a precision of 0.33. The mean Dice score between the manually segmented lesion and the follow-up was 0.20 (sd=0.21), while the mean Dice score between the prediction of FASTER and the follow-up was 0.32 (sd=0.23).

Tissue-at-risk volumetry

The tissue classification generated by FASTER gives a worst case estimation of the volume of the final infarction. This yields in most cases an overestimation of the final lesion load, but the volume of tissue indicated as at risk is substantially lower than the lesion defined by Tmax>6s. To see that this lower volume estimate does not come at the expense of underestimating the eventual lesion size, we focused on those patients in our test-set having experienced an unsuccessful revascularization (TICI 1–2a). The mean difference in volume between the manually segmented perfusion lesion and the final lesion was 121 ml (± 55 ml), while the mean difference in volume between the automatically defined

tissue-at-risk and the final lesion was 30 ml (± 26 ml). Moreover, the automatically segmented lesion bears a much closer relationship to the final lesion volume than the manually defined lesion: there is no significant correlation between the volume of the manually delineated perfusion lesion and the final lesion volume ($r=0.16$, $p=0.51$). Meanwhile, the final lesion volume is significantly correlated to the predicted lesion volume as defined by FASTER, with a correlation coefficient of 0.72 ($p < 0.001$).

By contrast, automated and manual lesion prediction performs comparably well at volumetry in the case of good reperfusion, as tested on the 36 testing cases with TICI 2b-3. Both predictors of final lesion volume were positively linearly correlated with final lesion volume ($p < 0.05$ in both cases).

Scatter- and Bland–Altman plots showing the performance of FASTER can be seen in Figure 3. Example outputs of the FASTER pipeline can be viewed in Figure 4.

Performance beyond the 6-h time-to-treatment window

Mechanical thrombectomy is currently restricted to a 6-h time window from stroke onset to the initiation of the therapy, as defined by first symptom onset or last time being witnessed as asymptomatic as clinical surrogates. This window is established from demographic data, without consideration of the stability and extent of the collateral supply in individual cases, and ruling out the potential benefit of treatment in individual cases where onset time is unknown, e.g. in patients with wake-up stroke. Advanced neuroimaging techniques such as FASTER may add complementary information to structural imaging markers as FLAIR signal increase,³⁸ enabling personalized assessments of tissue risk.

We therefore performed a subanalysis of 14 test cases that have been treated beyond the 6-h treatment window. Of these, eight were successfully revascularized, and six unsuccessfully revascularized. In five of those six cases, the predicted lesion was larger than the final infarction volume, suggesting that FASTER remains a robust tool for estimating penumbral volume beyond the 6-h timeframe. No significant correlation was found between the extent of overestimation, defined as the ratio between estimated and actual lesion size, and the time to therapy (Spearman's rank correlation test, $p=0.81$). It was not possible to establish a significant correlation between the predicted and actual lesion size, but given the sample size (6 cases) and the p value obtained ($p=0.11$), a subsequent study with more may be sufficient to establish this connection: 11

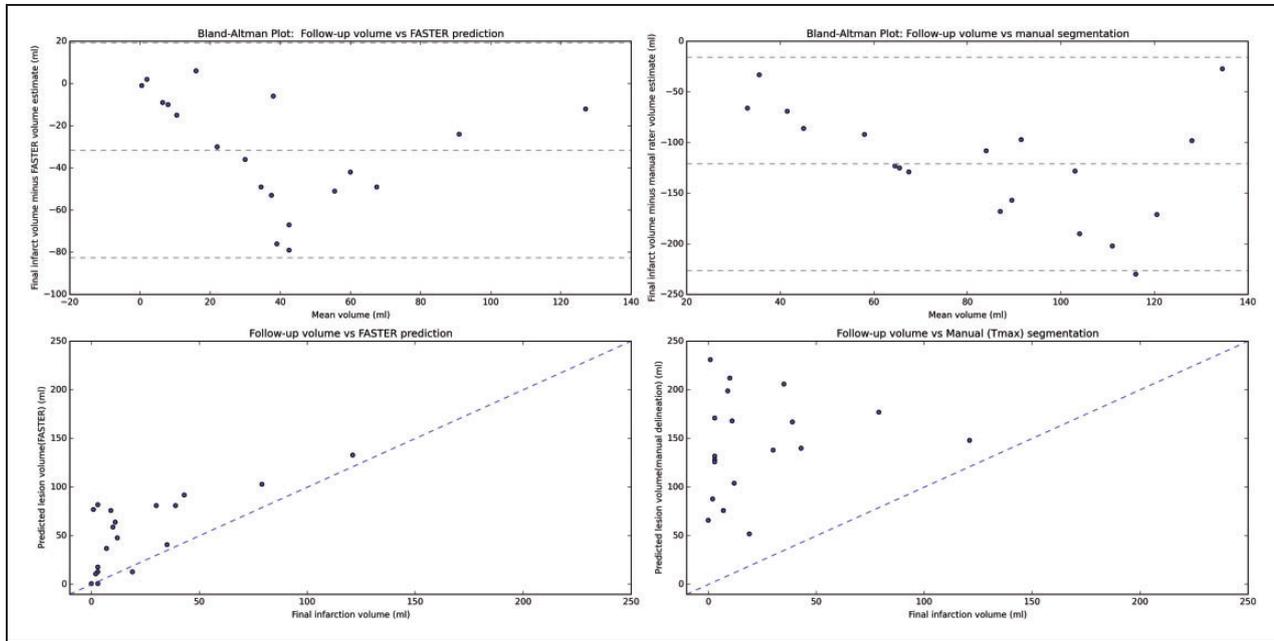


Figure 3. Volumetric performance of manual segmentation based on Tmax, vs. segmentation from FASTER, for 19 test cases with TIC1 1–2a.

cases would be sufficient to establish a correlation coefficient of 0.7 with an estimated power of 80%.

Variable importance analysis

The variable importance algorithm in H₂O calculates, for each split in the decision forest, the increase in classification accuracy obtained by splitting at that node, and then sums these gains in accuracy over each feature. The gains are then linearly scaled such that the most important feature is assigned an importance of 1. Results of this analysis are shown in Table 3.

While no features derived from structural imaging appeared in the top 10 imaging features used by either model, the point intensity of the T2-weighted scan was the 13th most important feature used in FASTER—, with a relative importance of 0.15. While the majority of strokes affect only a single hemisphere, bilateral strokes occur in roughly 9.4% of cases.³⁹ Since the most important feature of our model, Tmax symmetry, could be rendered useless by the presence of a bilateral stroke, we also examined the effect of removing the symmetry features from the available feature space. The resulting classifier had somewhat reduced sensitivity (0.75, vs. 0.79 for the full feature set), but performed comparably in terms of precision and AUC. In terms of feature importance, the most predictive features were again derived from the TTP and Tmax maps. The 75th percentile of the TTP was the most predictive feature, and all five most predictive features appeared

in the top 10 features of the full model (i.e. including symmetry features).

Discussion

In this study, we propose a fully automated framework for determining tissue-at-risk in ischemic stroke that extends beyond threshold-based perfusion/diffusion mismatch analysis, by directly predicting tissue fate ahead of therapy. The approach is an application of supervised learning, in which nonlinear classifiers are trained on stroke patient cases that underwent intra-arterial therapy, dichotomized by complete recanalization vs. permanent occlusion. By separating training data into these two cohorts, and training one model for each, we were able to predict the tissue risk for new patients with respect to the success of treatment. FASTER is able to provide, given a new patient case where the outcome is unknown, a comparison between the likely outcome in case of a successful or unsuccessful reperfusion, yielding a more accurate delineation of tissue-at-risk than that given by an expert rater using linear thresholded maps. Results are available 6–10 min after calculation of the perfusion maps. The lesion volume predicted by FASTER in the case of a poor response to therapy was significantly correlated with the final lesion volume in test cases having TIC1 score 1–2a. Manually segmented perfusion lesion volume was not found to correlate with final lesion volume in patients having a poor response to therapy. Since

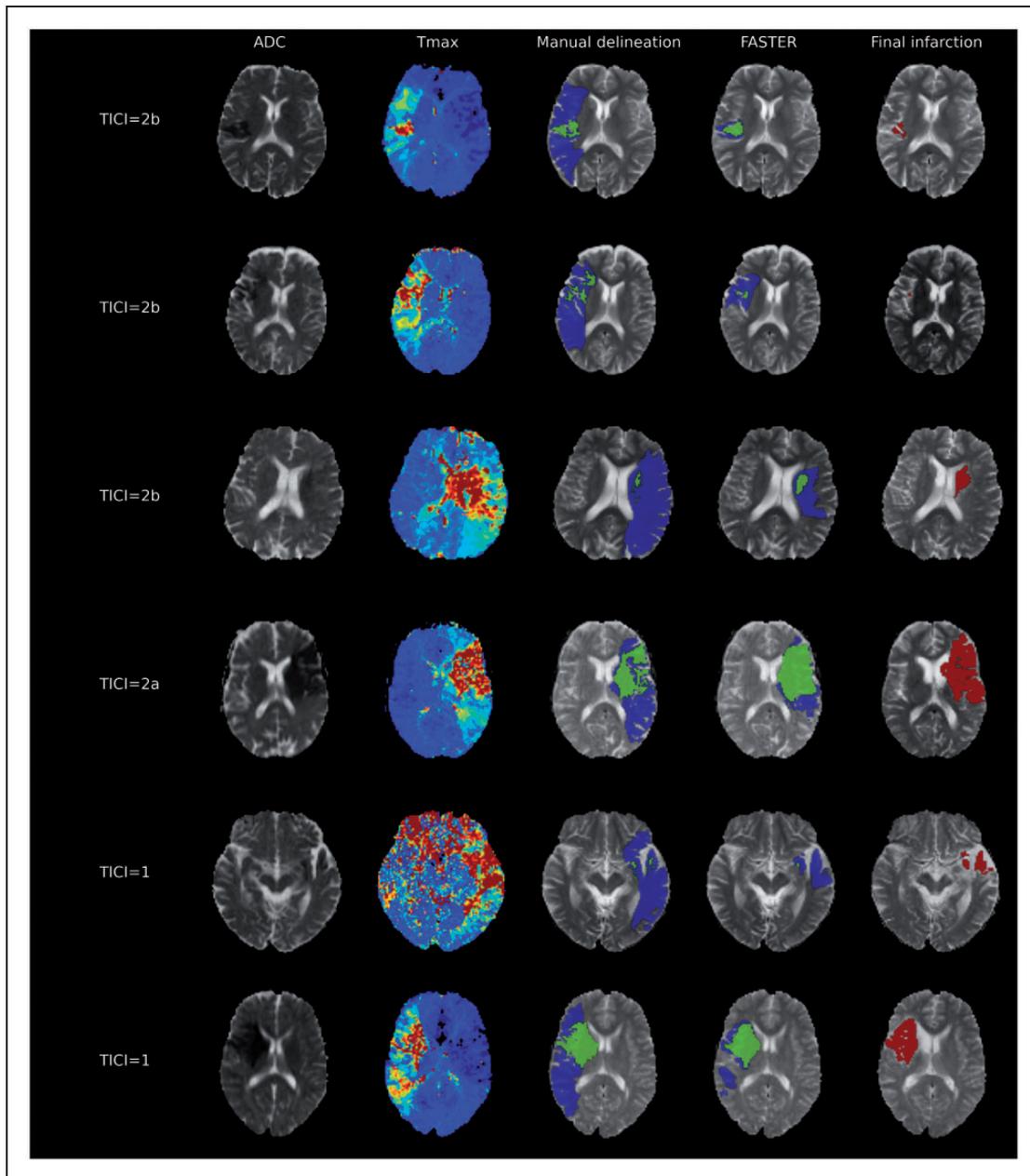


Figure 4. Six cases selected from the 55 test cases. In each case, from left to right: the ADC image, Tmax map, manually segmented tissue-at-risk assessment (infarct core in green, penumbra in blue), assessment of tissue-at-risk produced by FASTER (favorable outcome in green, unfavorable outcome in blue), and manually segmented 90-day outcome (in red). TICl grading is indicated for each case.

final infarction volume correlates with functional outcome and has been used as a marker for success of acute stroke treatment,⁴⁰ the output of FASTER may improve on threshold-based approaches to assessing the potential benefits of reperfusion.

Machine learning techniques have been criticized, from the perspective that they are a “black box,” insofar as it is difficult to directly interpret their output from a clinical perspective. However, the feature importance analysis allows one to see which features

contribute most to model: certain features were consistently selected for their ability to separate healthy tissue from tissue-at-risk, and salvageable tissue from unsalvageable tissue. The individual features are readily interpreted and could, themselves, be used to give a more robust definition of tissue-at-risk. Notably, no voxel-wise perfusion map was found to be useful in predicting infarction risk, suggesting that voxel-wise perfusion maps do not, themselves, characterize sufficient vs. insufficient vascular supply. Features that

Table 3. Predominant contributing features for FASTER+ and FASTER–.

Predominant contributing features for FASTER–	
Feature	Normalized importance
Tmax symmetry	1.0
75th percentile TTP (5 ³)	0.39
90th percentile TTP (5 ³)	0.28
Median Tmax (5 ³)	0.28
75th percentile Tmax (5 ³)	0.25
75th percentile TTP (3 ³)	0.23
CBF symmetry	0.21
Median TTP (5 ³)	0.21
Mean Tmax (5 ³)	0.19
25th percentile Tmax (5 ³)	0.19
Predominant contributing features for FASTER+	
Feature	Normalized importance
Tmax symmetry	1.0
Median ADC (5 ³)	0.65
TTP symmetry	0.63
25th percentile ADC (3 ³)	0.56
Median ADC (3 ³)	0.50
75th percentile ADC (3 ³)	0.46
25th percentile Tmax (5 ³)	0.43
Variance ADC (3 ³)	0.41
Median Tmax (5 ³)	0.41
CBF Symmetry	0.41

Feature importance normalized in the range [0, 1] (1= most important feature). Features derived from local statistical measures are shown with the volume in which they were calculated, (3³) indicating a three by three by three voxel volume, and (5³) indicating a five by five by five voxel volume.

consistently contributed to the determination of outcome in the absence of revascularization were derived predominantly from perfusion parameters that are also used for the linear estimation of the penumbra. The Tmax symmetry feature, which is calculated from a smoothed Tmax map by comparison across hemispheres was highly predictive in both FASTER– and FASTER+: this validates the use of manually corrected Tmax as the solitary feature of linear tissue-at-risk estimation. Features derived from upper percentiles (90th and 75th) of TTP also made substantial contributions to the classification. These features may therefore be more reliable markers of collateral supply than standard perfusion maps. The ADC played a major role in FASTER+ but made no major contribution to FASTER–. There was also less important contribution

from features derived from T2-weighted imaging in our classification than might be expected. Recent studies have demonstrated an added value of FLAIR instead of T2-weighted images to detect early and subtle inhomogeneities that indicate beginning of hypoxic damage.³⁸

We were able to demonstrate an improved precision compared to other predictive models; however, the extent of tissue damage has been overestimated (a volume difference of more than 10 ml) in eight patient cases. By comparison, the classical threshold-based assessment of tissue-at-risk yielded a more than 10 ml overestimation of final infarction volume in 14 cases. This effect may be explained by the effect of collateral flow in these patients: good collaterals halt the loss of penumbral tissue and have been shown to indicate reversal of DWI imaging lesions.^{41,42} Scar formation and lesion shrinkage may also lead to volume overestimation.

We have applied FASTER to a single center MRI data set acquired and trained on different scanners from the same vendor. In future, both training and validation require inclusion of a multicenter dataset to verify if a generalization of the models is feasible. Another limitation is the source of the data: by obtaining test data only from cases with that required 90-day follow-up scans to test the reliability of the model on final infarction volumes. Thus, the most severe cases of patients who subsequently died from stroke were excluded from analysis. Furthermore, patients treated with endovascular therapy outside of the six hour window in this study were selected using non-random treatment criteria, leading to potential bias in the response to therapy. Further prospective analyses are necessary to answer this unsolved question in future.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study has been supported by the Swiss Heart Foundation. SJ is supported by the Swiss National Science Foundation (SNSF SPUM-Grant 140340). RM thanks the BNF program of the University of Bern and the Swiss Heart Foundation for their support.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Authors' contributions

All authors have contributed substantially to drafting the article and revising it critically, and approved the version to be published. RM contributed to study design, image analysis, machine learning methodology, and statistical analysis. LH

contributed to study design, image processing, and statistical analysis. JG, ME-K, MA, UF, and SJ contributed to acquisition of data. SB contributed to study design. KM contributed to image analysis. MR and RW contributed to study design, analysis and interpretation.

Supplementary material

Supplementary material for this paper can be found at <http://jcbfm.sagepub.com/content/by/supplemental-data>

References

- Schlaug G, Benfield A, Baird AE, et al. The ischemic penumbra: operationally defined by diffusion and perfusion MRI. *Neurology* 1999; 53: 1528–1537.
- Heiss W-D, Sobesky J and Hesselmann V. Identifying thresholds for penumbra and irreversible tissue damage. *Stroke* 2004; 35: 2671–2674.
- Olivot JM, Mlynash M, Thijs VN, et al. Geography, structure, and evolution of diffusion and perfusion lesions in diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE). *Stroke* 2009; 40: 3245–3251.
- Campbell BC, Mitchell PJ, Kleinig TJ, et al. Endovascular therapy for ischemic stroke with perfusion-imaging selection. *N Engl J Med* 2015; 372: 1009–1018.
- Lansberg MG, Straka M, Kemp S, et al. MRI profile and response to endovascular reperfusion after stroke (DEFUSE 2): a prospective cohort study. *Lancet Neurol* 2012; 11: 860–867.
- Wintermark M, Sanelli PC, Albers GW, et al. Imaging recommendations for acute stroke and transient ischemic attack patients: a joint statement by the American Society of Neuroradiology, the American College of Radiology and the Society of NeuroInterventional Surgery. *J Am Coll Radiol* 2013; 10: 828–832.
- Albers GW, Thijs VN, Wechsler L, et al. Magnetic resonance imaging profiles predict clinical response to early reperfusion: the diffusion and perfusion imaging evaluation for understanding stroke evolution (DEFUSE) study. *Ann Neurol* 2006; 60: 508–517.
- Straka M, Albers GW and Bammer R. Real-time diffusion-perfusion mismatch analysis in acute stroke. *J Magn Reson Imag* 2010; 32: 1024–1037.
- Forkert ND, Kaesemann P, Treszl A, et al. Comparison of 10 TTP and tmax estimation techniques for MR perfusion-diffusion mismatch quantification in acute stroke. *Am J Neuroradiol* 2013; 34: 1697–1703.
- Rekik, et al. Medical image analysis methods in MR/CT-imaged acute-subacute ischemic stroke lesion: Segmentation, prediction and insights into dynamic evolution simulation models. A critical appraisal. *Neuroimage* 2012; 1: 164–178.
- Wu O, Sumii T, Asahi M, et al. Infarct prediction and treatment assessment with MRI-based algorithms in experimental stroke models. *J Cereb Blood Flow Metab* 2007; 27: 196–204.
- Nguyen HV, Cooperman G, Menezes N, et al. Stroke tissue outcome prediction using a spatially-correlated model. *PPIC (Pan-Pacific Imaging Conference)*, Tokyo, Japan. Vol. 8, 25–27 June 2008, pp.238–241.
- Scalzo F, Hao Q, Alger JR, et al. Regional prediction of tissue fate in acute ischemic stroke. *Ann Biomed Eng* 2012; 40: 2177–2187.
- Stier N, Vincent N, Liebeskind D, et al. Deep learning of tissue fate features in acute ischemic stroke. In: *IEEE international conference on bioinformatics and biomedicine (bIBM)*, 2–4 November 2015, pp.1316–1321. Washington DC, USA: IEEE.
- Berkhemer OA, Fransen PS, Beumer D, et al. A Randomized trial of intraarterial treatment for acute ischemic stroke. *N Engl J Med* 2015; 372: 11–20.
- Goyal M, Demchuk AM, Menon BK, et al. Randomized Assessment of Rapid Endovascular Treatment of Ischemic Stroke. *N Engl J Med* 2015; 372: 1019–1030.
- Saver JL, Goyal M, Bonafe A, et al. Stent-retriever thrombectomy after intravenous t-PA vs. t-PA alone in stroke. *N Engl J Med* 2015; 372: 2285–95.
- Jovin TG, Chamorro A, Cobo E, et al. Thrombectomy within 8 hours after symptom onset in ischemic stroke. *N Engl J Med* 2015; 372: 2296–2306.
- Kemmling A, Flottmann F, Forkert ND, et al. Multivariate dynamic prediction of ischemic infarction and tissue salvage as a function of time and degree of recanalization. *J Cereb Blood Flow Metab* 2015; 35: 1397–1405.
- Wintermark M, Flanders AE, Velthuis B, et al. Perfusion-CT assessment of infarct core and penumbra: Receiver operating characteristic curve analysis in 130 patients suspected of acute hemispheric stroke. *Stroke* 2006; 37: 979–985.
- Arnold M, Kappeler L, Nedeltchev K, et al. Recanalization and outcome after intra-arterial thrombolysis in middle cerebral artery and internal carotid artery occlusion: does sex matter? *Stroke* 2007; 38: 1281–1285.
- Jung S, Mono ML, Fischer U, et al. Three-month and long-term outcomes and their predictors in acute basilar artery occlusion treated with intra-arterial thrombolysis. *Stroke* 2011; 42: 1946–1951.
- Jung S, Schindler K, Findling O, et al. Adverse effect of early epileptic seizures in patients receiving endovascular therapy for acute stroke. *Stroke* 2012; 43: 1584–1590.
- Galimanis A, Jung S, Mono ML, et al. Endovascular therapy of 623 patients with anterior circulation stroke. *Stroke* 2012; 43: 1052–1057.
- Higashida RT, Furlan AJ, Roberts H, et al. Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. *Stroke* 2003; 34: e109–e137.
- Kudo K, Sasaki M, Ogasawara K, et al. Difference in tracer delay-induced effect among deconvolution algorithms in CT perfusion analysis: quantitative evaluation with digital phantoms. *Radiology* 2009; 251: 241–249.
- Bauer S, Fejes T and Reyes M. A skull-stripping filter for ITK. *Insight J* 2012; . Available at: <http://hdl.handle.net/10380/3353>.
- Gaudinski MR, Henning EC, Miracle A, et al. Establishing final infarct volume: Stroke lesion evolution past 30 days is insignificant. *Stroke* 2008; 39: 2765–2768.

29. Altman DG and Bland JM. Measurement in medicine: the analysis of method comparison studies. *The Statistician* 1983; 32: 307–317.
30. McKinley R, Haeni L, Wiest R, et al. Segmenting the ischemic penumbra: a decision forest approach with automatic threshold finding. LNCS brainlesion: Glioma, MS, Stroke and traumatic brain injuries – first international brainles workshop MICCAI 2015. *Lect Notes Comput Sc* 2016; 9556: 275–283.
31. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
32. Criminisi A, Shotton J and Konukoglu E. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and trends® in computer graphics and vision*. 2011; 7: 81–227.
33. Porz N, Bauer S, Pica A, et al. Multi-modal glioblastoma segmentation: man versus machine. *PloS One* 2014; 9: e96873.
34. Meier R, Bauer S, Slotboom J, et al. Patient-specific semi-supervised learning for postoperative brain tumor segmentation. *Med Image Comput Comput Assist Intervent* 2014; 17: 714–721.
35. Bauer S, Gratz PP, Gralla J, et al. Towards automatic MRI volumetry for treatment selection in acute ischemic stroke patients. In: *The annual international conference of the IEEE engineering in medicine and biology society*, 26–30 August 2014, pp.1521–1524. Chicago: IEEE.
36. Jonsdottir KY, Østergaard L and Mouridsen K. Predicting tissue outcome from acute stroke magnetic resonance imaging: improving model performance by optimal sampling of training data. *Stroke* 2009; 40: 3006–3011.
37. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform* 2011; 12: 77.
38. Thomalla G, Fiebich JB, Østergaard L, et al. A multi-center, randomized, double-blind, placebo-controlled trial to test efficacy and safety of magnetic resonance imaging-based thrombolysis in wake-up stroke (WAKE-UP). *Int J Stroke* 2014; 9: 829–836.
39. Michel P, Odier C, Rutgers M, et al. The Acute Stroke Registry and Analysis of Lausanne (ASTRAL): design and baseline analysis of an ischemic stroke registry including acute multimodal imaging. *Stroke* 2010; 41: 2491–2498.
40. Yoo AJ, Chaudhry ZA, Nogueira RG, et al. Infarct volume is a pivotal biomarker after intra-arterial stroke therapy. *Stroke* 2012; 43: 1323–1330.
41. Jung S, Gilgen M, Slotboom J, et al. Factors that determine penumbral tissue loss in acute ischaemic stroke. *Brain* 2013; 136: 3554–3560.
42. Kim SH, Kim EH, Lee BI, et al. Chronic cerebral hypoperfusion protects against acute focal ischemia, improves motor function, and results in vascular remodeling. *Curr Neurovasc Res* 2008; 5: 28–36.