



Use of quantitative angiographic methods with a data-driven model to evaluate reperfusion status (mTICI) during thrombectomy

Mohammad Mahdi Shiraz Bhurwani^{1,2} · Kenneth V. Snyder^{2,3} · Muhammad Waqas^{2,3} · Maxim Mokin⁴ · Ryan A. Rava^{1,2} · Alexander R. Podgorsak^{1,2} · Felix Chin^{2,3} · Jason M. Davies^{2,3} · Elad I. Levy^{2,3} · Adnan H. Siddiqui^{2,3} · Ciprian N. Ionita^{1,2,3}

Received: 3 September 2020 / Accepted: 3 November 2020 / Published online: 7 January 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Purpose Intra-procedural assessment of reperfusion during mechanical thrombectomy (MT) for emergent large vessel occlusion (LVO) stroke is traditionally based on subjective evaluation of digital subtraction angiography (DSA). However, semi-quantitative diagnostic tools which encode hemodynamic properties in DSAs, such as angiographic parametric imaging (API), exist and may be used for evaluation of reperfusion during MT. The objective of this study was to use data-driven approaches, such as convolutional neural networks (CNNs) with API maps, to automatically assess reperfusion in the neuro-vasculature during MT procedures based on the modified thrombolysis in cerebral infarction (mTICI) scale.

Methods DSAs from patients undergoing MTs of anterior circulation LVOs were collected, temporally cropped to isolate late arterial and capillary phases, and quantified using API peak height (PH) maps. PH maps were normalized to reduce injection variability. A CNN was developed, trained, and tested to classify PH maps into 2 outcomes (mTICI 0,1,2a/mTICI 2b,2c,3) or 3 outcomes (mTICI 0,1,2a/mTICI 2b/mTICI 2c,3), respectively. Ensembled networks were used to combine information from multiple views (anteroposterior and lateral).

Results The study included 383 DSAs. For the 2-outcome classification, average accuracy was 81.0% (95% CI, 79.0–82.9%), and the area under the receiver operating characteristic curve (AUROC) was 0.86 (0.84–0.88). For the 3-outcome classification, average accuracy was 64.0% (62.0–66.0), and AUROC values were 0.85 (0.83–0.87), 0.74 (0.71–0.77), and 0.78 (0.76–0.81) for the mTICI 0,1,2a, mTICI 2b, and mTICI 2c,3 classes, respectively.

Conclusion This study demonstrated the feasibility of using hemodynamic information in API maps with data-driven models to autonomously assess intra-procedural reperfusion during MT.

Keywords Angiographic parametric imaging · Convolutional neural network · Mechanical thrombectomy · Acute ischemic stroke · Large vessel occlusion

✉ Ciprian N. Ionita
cnionita@buffalo.edu

¹ Department of Biomedical Engineering, University at Buffalo, Buffalo, NY 14228, USA

² Canon Stroke and Vascular Research Center, Buffalo, NY 14203, USA

³ Department of Neurosurgery, University at Buffalo, Buffalo, NY 14203, USA

⁴ Department of Neurosurgery and Brain Repair, University of South Florida, Tampa, FL 33606, USA

Abbreviations

AIS	Acute ischemic stroke
AP	Anteroposterior
API	Angiographic parametric imaging
AUROC	Area under the ROC curve
CAM	Class activation map
CNN	Convolutional neural network
CT	Computed tomography
DSA	Digital subtraction angiography
ICA	Internal carotid artery
LVO	Large vessel occlusion
MCA	Middle cerebral artery
MCCV	Monte Carlo cross-validation
MCC	Matthews correlation coefficient

MT	Mechanical thrombectomy
mTICI	Modified TICI
NIHSS	NIH stroke score
TDC	Time density curve
TICI	Thrombolysis in cerebral infarction
PH	Peak height
ROC	Receiver operating characteristic

Introduction

Patients with emergent large vessel occlusion (LVO) acute ischemic strokes (AIS) account for 24–46% of all AIS cases [1–3]. While there are many confounding factors, timely recanalization of such blockages is essential for optimal patient outcomes [4]. Mechanical thrombectomies (MTs) aim at endovascular clot retrieval to restore blood flow to ischemic territories [5, 6]. MTs are reported to have successful reperfusion rates between 75 and 80% [7] compared to the 30% early recanalization rate when using intravenous thrombolysis with recombinant tissue-type plasminogen activators (thrombus length < 7 mm) [8]. Thus, MTs have markedly reduced severe disability and mortality compared to intravenous thrombolysis [8] and have been established as the standard care for LVO AIS.

Patients with LVOs undergo computed tomography (CT), magnetic resonance imaging (MRI), CT angiogram, CT perfusion, or MR angiogram (MRA) imaging to determine eligibility for the MT procedure, followed by the procedure itself. While preprocedural AIS imaging has undergone major advancements [9], intra-procedural imaging still lags behind. Currently, intra-procedural MT success is assessed primarily by grading intracranial reperfusion using cerebral digital subtraction angiography (DSA). This is done using the thrombolysis in cerebral infarction (TICI) scale as proposed by Higashida and Furlan [10] or the modified TICI (mTICI) scale [11]. mTICI grading systems have received criticisms due to confusing internal inconsistencies [12] as well as the inclusion of bias due to grading being conducted solely through direct visual estimations of operators [13]. Clinical trials compared operator-scored mTICIs with core lab-scored mTICIs and found only a 56% proportion of agreement between the two. In 33% of these, the operator-scored mTICIs were overestimated compared to those from the core lab [14–16].

Intra-procedural assessment of such endovascular procedures could be improved using quantitative tools similar to CTP. However, such implementation is limited due to the 2D nature of DSA and variability caused by hand injection of contrast. Angiographic parametric imaging (API) has been proposed as an alternative solution. This form of image analysis uses a DSA sequence to semi-quantitatively analyze blood flow through the vasculature and angioarchitectures. Intensity at each pixel across the DSA sequence is measured,

resulting in a time density curve (TDC) at each pixel. TDCs' parameterization enables the extraction of various parameters such as mean transit time, time to peak, time to arrival, peak height (PH), and area under the TDC. This allows API map generation for each parameter which can be analyzed to understand the nature of flow through different vessels and phases in DSAs [17, 18]. Each map encodes one hemodynamic parameter derived from each X-ray pathway, thus making maps less sensitive to subtle flow differences. This suboptimal sensitivity could be improved using a hybrid approach where hemodynamics encoded in API are combined with a data-driven model.

Data-driven models such as convolutional neural networks (CNNs) and other machine learning tools using CT or MRI data have been clinically implemented for automated stroke assessment such as the ASPECT score [19]. Following similar trends, tools have been proposed for DSA using API to make predictions regarding treatment success [20]. In this work, we present a study to test the feasibility of using CNNs with quantitative angiographic information from API to classify cerebral reperfusion during MT procedures. For the data-driven classification, the reperfusion level was assessed using the mTICI scale. However, any other outcomes, including post-procedure MRI or neurologic evaluation, could be used.

Methods

Data collection

Retrospective collection and analysis of patient data was conducted at a single center and approved by our institutional review board. Inclusion criteria were any patient with an LVO undergoing a MT. For each patient, baseline, intra-procedural, and post-MT DSAs were collected. Anteroposterior (AP) and lateral view DSAs were collected for every scan. Patients with posterior circulation occlusions were excluded. DSAs with image artifacts caused by patient motion during the scan, mainly in cases treated under conscious sedation, were also excluded from the study.

We included 192 patients with 383 angiographic runs in our final analysis. Since angiographic runs from the same patients were taken at different time points during MT procedures, they have different levels of reperfusion and can be considered separate cases. Mean patient age was 68.75 years, initial NIH stroke score (NIHSS) was 12, post-procedure NIHSS was 4, and NIHSS shift was –7. Patient demographics, locations of LVOs, and summary of mTICI scores are displayed in Fig. 1.

Acquisition of DSA sequences for all patients was conducted using Canon Infinix biplane systems (Canon Medical

Systems Corporation, Otawara, Japan). DSAs were acquired at an average tube voltage of 84.3 ± 5.0 kVp (average \pm standard deviation), tube current of 149.4 ± 42.7 mA, pulse width of 84.0 ± 12.1 ms, and frame rate of 3 frames per second. Contrast used during acquisitions was iohexol (Omnipaque 350; GE Healthcare, Piscataway, NJ).

The overall study workflow is displayed in Fig. 2. For each DSA included in the analysis, an mTICI label was assigned by two experienced operators (qualified neuro-interventionalists) independently of each other. The operators graded every case and used the AP and lateral full DSA sequences. Operators were not involved in the procedure and were blinded to clinical outcomes and PH maps. Grading was performed according to the following 6 categories: no perfusion (grade 0), partial perfusion beyond initial occlusion but not in distal arteries (grade 1), partial perfusion less than 50% (grade 2a), partial perfusion more than 50% but less than full (grade 2b), complete but delayed perfusion (grade 2c), and complete perfusion (grade 3) (Fig. 2) [21]. Cases with disagreements in mTICI labels were resolved by consensus decision. This was done to remove any bias in labels used for training the network.

API map generation

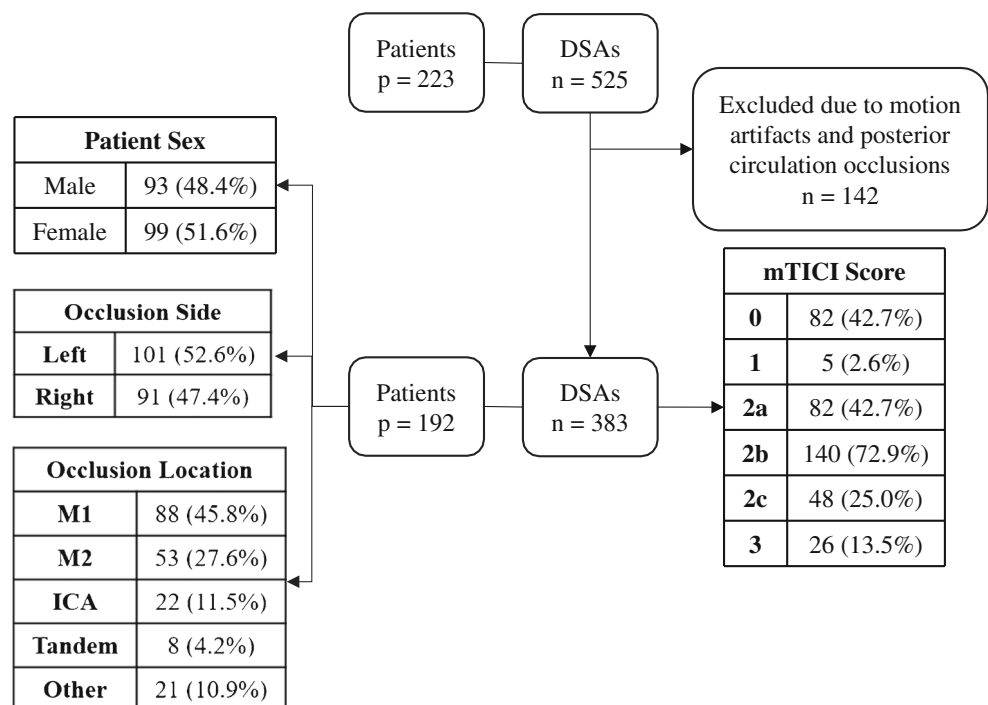
Reperfusion evaluation with mTICI scores was done based on the extent of tissue perfusion as represented by the capillary blush in DSAs [10, 22]. Each DSA was cropped to only

include frames where contrast was in the late arterial and capillary phases. Thus, API maps contained a limited number of overlapping structures from early arterial or venous phases. Given the three frames per second acquisitions, arterial structures were always present in the final API maps. The temporal cropping was done by an operator with 3 years of experience working with DSAs.

TDCs were extracted at each pixel by tracking the flow of contrast across frames in the cropped DSA sequence. PH maps were generated by calculating the maximum value from TDCs at each pixel. For the purpose of this feasibility study, only PH maps were considered as it reflects maximum contrast intensity in each pixel across all frames and is thus most reflective of perfusion.

Since hand contrast injection was used for these emergent cases, injection parameters such as concentration, volume, and injection rate were highly heterogeneous between DSA acquisitions. To account for this variability, every pixel value in the PH map was divided by the PH value in the main feeding artery, thus normalizing each map to contrast concentration in the respective inlet vessel. The location for normalization was manually conducted by an operator with 3 years of experience working with DSAs and API. The location was chosen on a straight portion of the main feeding artery before splitting into respective branches. Care was taken to avoid tortuous or overlapping structures that would affect the quantitative angiographic values as well as any regions of image artifacts.

Fig. 1 Patient demographic information. *p* refers to the number of patients, *n* refers to the number of DSAs. DSA digital subtraction angiography, ICA internal carotid artery



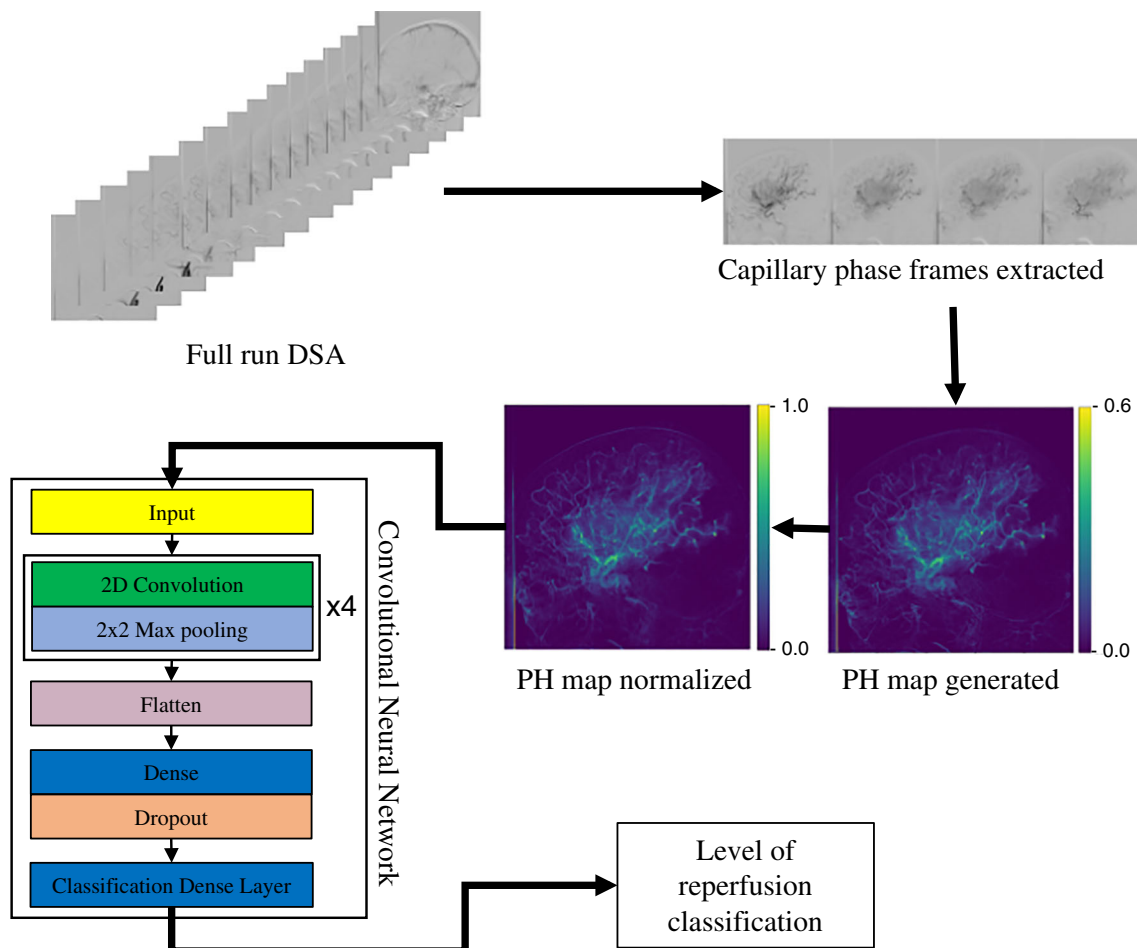


Fig. 2 Workflow of the study. *DSA* digital subtraction angiography, *PH* peak height

Network development

CNN was developed using Keras [23] to classify PH maps based on the reperfusion level. CNN architecture development was an iterative process based on the optimization of metrics such as classification accuracy and receiver operating characteristic (ROC) curves. The final architecture is displayed in Fig. 2. The optimizer used during training was the Adam optimizer with an initial learning rate of 10^{-3} . The loss function used was the categorical cross-entropy. Keras callbacks were used to reduce the learning rate as training progressed and automatically terminate training as loss plateaued. In addition, the class imbalance between classes was accommodated by implementing a balanced class weighting during training. Thus, the CNN balanced layer weights to ensure equal penalization of under- or overrepresented classes in the training set. CNNs were trained and tested on a single NVIDIA (Nvidia Corporation, Santa Clara, CA) Tesla V100 graphics processing unit.

Following guidelines proposed by *Radiology* [24] and in order to prevent network overfitting, we split the dataset with 70% (268 cases) reserved for training, 10% (39 cases) reserved as a validation set used for hyperparameter tuning during training, and 20% (77 cases) reserved for testing. Hyperparameter tuning during training was done by tracking loss on the validation set. To test network robustness and ensure that results were not based on specific training-testing splits, a 20-fold Monte Carlo cross-validation (MCCV) [25] was conducted. This approach involved randomly splitting the total dataset into training and testing sets 20 times.

In order to combine information contained in both the AP and lateral PH maps, we used ensembled networks. This method involves a combination of predictions from multiple networks to give a final prediction by assigning a weight to predictions from each network. One network was trained on AP PH maps, and another network with the same architecture was trained on lateral PH maps. Weights to assign to predictions from individual networks

were calculated using a differential evolution optimization algorithm [26] implemented using SciPy [27]. Once the weights were calculated, they were used to combine predictions from the AP and lateral networks to give weighted ensemble predictions.

CNNs can automate outcome predictions and quantitative assessment of lesions such as intracranial aneurysms [20]; however, as a self-controlling minimization technique, they do not allow users to oversee which image features are most important and thus how to improve network predictions. We investigated the use of class activation maps (CAMs) to visualize regions of PH maps that trigger the trained algorithm, thus lending insight to whether the CNN makes decisions based on flow or using some other portion of the image which may not be as predictive of reperfusion. CAMs were generated using a method described by Zhou et al. [28]. They are obtained by taking outputs from the final convolutional layer and passing it through a global average pooling layer. CAMs are heatmaps where high-intensity pixels are features highly weighted towards network classification output.

Statistical analysis

The CNN was evaluated using five quantitative metrics including classification accuracy, ROC curves, area under the ROC curves (AUROC), sensitivity, specificity, and Matthews correlation coefficient (MCC). Each of these metrics was averaged using results over the 20-fold MCCV. MCC is used in machine learning models to evaluate the quality of binary classifications [29]. It has proven to be advantageous as it takes into account class imbalance and uses every factor in the confusion matrix (true positives, false positives, true negatives, and false negatives) [30].

Currently, when classifying intra-procedural DSAs as having sufficient or insufficient reperfusion to determine the need for further treatment, clinicians use either a 2-outcome grouping where mTICI 0,1,2a is clinically insufficient reperfusion and mTICI 2b,2c,3 is sufficient reperfusion [31] or a 3-outcome grouping where mTICI 0,1,2a is insufficient reperfusion, mTICI 2c,3 is sufficient reperfusion, and mTICI 2b is either sufficient or insufficient reperfusion, and the need for further treatment is decided based on other factors [32]. Thus, in addition to a 2-outcome classification between mTICI 0,1,2a and mTICI 2b,2c,3, we also investigated a 3-outcome classification between mTICI 0,1,2a, mTICI 2b, and mTICI 2c,3. In addition, subgroup analysis was conducted for using AP and lateral view networks independently and for using both views combined using the ensemble network. Two-tailed McNemar's p test values were calculated in order to evaluate the significance of any performance

differences ($p < 0.05$). McNemar's p test values were also calculated between networks using temporally cropped DSAs and networks using uncropped DSAs in order to test the effect of temporally cropping out arterial and venous phases from the DSAs prior to PH map generation.

Results

Network performance

The algorithm takes 0.25 s to create and normalize each API map, the AP network took 9.2 min to train, the lateral network took 9.3 min to train, the ensemble weights were calculated in 0.65 s and a single case can be classified using the network in 0.6 ms. Average values for each evaluation metric along with their standard deviations and 95% confidence intervals are displayed in Table 1. Peak network performance was achieved when making a 2-outcome classification using an ensemble network that combined classifications from both the AP and lateral view networks. While better performance was achieved in terms of accuracy, AUROC, MCC, and sensitivity for 2-outcome classifications, better specificity was observed for 3-outcome classifications.

Performance was also evaluated with ROC curves that are displayed in Fig. 3. The highest AUROC values were achieved when using ensemble networks. In each plot, ROC curves for each subgroup are similar with the overlap of standard deviations, and McNemar's p test values indicate significant advantage towards using lateral view over AP view (p value < 0.05) and towards using ensemble networks over AP or lateral view networks independently ($p < 0.05$) for both the 2-outcome and 3-outcome classifications.

McNemar's p test values between the temporally cropped and uncropped AP, lateral, and ensemble networks were 0.16, 0.23, and 0.05. Since these values are above or equal to the 0.05 threshold, there is no significant advantage towards using temporally cropped DSAs.

Class activation maps

CAMs for two different cases were generated to visualize how the CNN makes its classifications and are displayed in Fig. 4. The CAMs were able to display what PH map regions were being used by the CNN to make the classification decision. The contrast in the vasculature activated the network, with larger vessels having a higher activation. Thus, the network is looking at the presence of the contrast and vasculature to make classification decisions. The internal carotid artery (ICA) terminus, middle cerebral artery (MCA) presence, and MCA territory seemed to be the greatest contributors towards network classification tasks.

Table 1 Convolutional neural network performance in classifying DSAs based on their level of reperfusion based on the mTICI scale. Performance is displayed in the form of average accuracies, area under the receiver operating characteristic curves (AUROC), Matthews correlation coefficients (MCC), sensitivities, and specificities along with their standard deviations and 95% confidence intervals (CI). (A) Two-outcome classification (mTICI grade 0,1,2a versus mTICI grade 2b,2c,3).

(B) Three-outcome classification (mTICI grade 0,1,2a versus mTICI grade 2b versus mTICI grade 2c,3). The 3-outcome classification requires a ROC curve for each outcome; thus, there is an AUROC for each outcome in (B). The best results are in italic. The results indicate that the best performance is achieved when making a 2-outcome classification using an ensembled network

Metric	(A) 2 class: mTICI 0,1,2a versus mTICI 2b,2c,3		
	Only AP	Only lateral	AP and lateral ensembled
Accuracy (%)	74.2 ± 3.3 (72.8, 75.7)	76.9 ± 5.9 (74.4, 79.5)	<i>81.0 ± 4.5 (79.0, 82.9)</i>
AUROC	0.83 ± 0.04 (0.81, 0.84)	0.84 ± 0.05 (0.82, 0.87)	<i>0.86 ± 0.04 (0.84, 0.88)</i>
MCC	0.49 ± 0.07 (0.46, 0.52)	0.54 ± 0.11 (0.49, 0.59)	<i>0.62 ± 0.09 (0.58, 0.66)</i>
Sensitivity	0.78 ± 0.1 (0.74, 0.83)	0.84 ± 0.1 (0.79, 0.88)	<i>0.88 ± 0.08 (0.85, 0.92)</i>
Specificity	0.69 ± 0.12 (0.64, 0.75)	0.68 ± 0.14 (0.62, 0.75)	0.72 ± 0.11 (0.67, 0.76)
Metric	(B) 3 class: mTICI 0,1,2a versus mTICI 2b versus mTICI 2c,3		
	Only AP	Only lateral	AP and lateral ensembled
Accuracy (%)	56.0 ± 4.9 (53.9, 58.2)	61.8 ± 4.1 (60, 63.6)	64.0 ± 5.0 (62.0, 66.0)
AUROC-mTICI 0,1,2a	0.81 ± 0.04 (0.79, 0.83)	0.84 ± 0.05 (0.82, 0.86)	0.85 ± 0.04 (0.83, 0.87)
AUROC-mTICI 2b	0.67 ± 0.06 (0.64, 0.70)	0.74 ± 0.06 (0.72, 0.77)	0.74 ± 0.06 (0.71, 0.77)
AUROC-mTICI 2c,3	0.74 ± 0.05 (0.71, 0.76)	0.75 ± 0.08 (0.72, 0.79)	0.78 ± 0.06 (0.76, 0.81)
MCC	0.30 ± 0.09 (0.26, 0.34)	0.40 ± 0.07 (0.37, 0.43)	0.43 ± 0.08 (0.40, 0.47)
Sensitivity	0.66 ± 0.16 (0.59, 0.73)	0.74 ± 0.08 (0.70, 0.77)	0.80 ± 0.1 (0.75, 0.84)
Specificity	0.73 ± 0.08 (0.69, 0.76)	0.74 ± 0.08 (0.70, 0.77)	<i>0.76 ± 0.07 (0.73, 0.79)</i>

Discussion

An objective and unbiased assessment of reperfusion status during MT is critical for the estimation of clinical prognosis and documentation for research purposes. In this study, we investigated the technical feasibility of using a CNN with quantitative angiographic information to assess the level of cerebral perfusion for patients undergoing a MT to treat an LVO AIS. We successfully classified PH maps generated from DSAs during an MT into 2-outcome categories (mTICI 0,1,2a and mTICI 2b,2c,3) or 3-outcome categories (mTICI 0,1,2a, mTICI 2b, and mTICI 2c,3). This indicates that data-driven models such as CNNs can be used to derive hemodynamic information encoded in API and make decisions regarding the nature of cerebral blood flow.

Numerical results for five evaluation metrics used are displayed in Table 1, and ROC curves are displayed in Fig. 3. Peak performance was observed when making a 2-outcome classification using ensembled networks where information from both the AP and lateral view networks were used to provide a final classification. This was observed considering high average values for each metric along with small standard deviations and tight confidence intervals. The numerical results, intersecting ROC curves, and overlapping standard deviations may indicate

similar performance between each subgroup; however, McNemar's *t* test *p* values ($p < 0.05$) indicate significant advantage to using ensembled networks over AP and lateral networks independently. In addition to commonly used evaluation metrics, we also calculated MCC. Since MCC is an application of the Pearson correlation coefficient [33], it follows the same patterns in terms of inferring correlation strength between classifications and ground truth [34]. MCC values indicate strong positive relationships for 2-outcome classifications and moderate positive relationships for 3-outcome classifications. Network performance on 3-outcome classifications is lower for each subgroup; however, it is still within an acceptable range given this is a feasibility study. The lower performance on 3-outcome classifications can be attributed to the lower number of cases in each outcome (169 cases:140 cases:74 cases) compared to 2-outcome classification (169 cases:214 cases) and to the increased task complexity of creating a finer classification. Increasing dataset size and increasing the number of cases in each specific class will allow us to achieve higher performance on 3-outcome classifications.

In order to understand which PH map features the CNN uses to make decisions, we generated CAMs. Two specific cases, including input PH maps, CAMs, and final classification

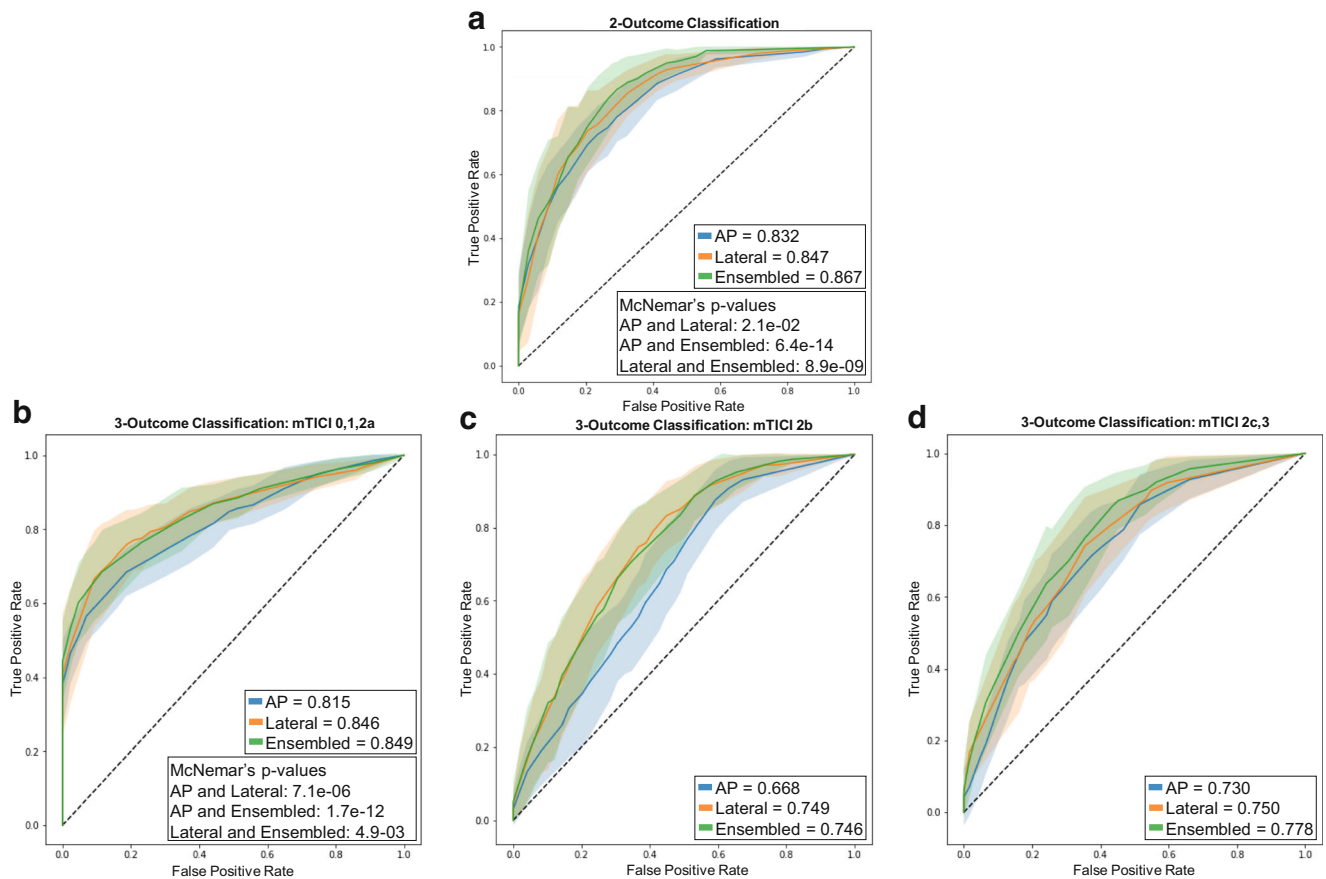


Fig. 3 Receiver operating characteristic (ROC) curves generated from the classifications of the convolutional neural network (CNN). **a** The ROC curves obtained for each subgroup when making a 2-class classification. **b** The ROC curves for each subgroup for the mTICI 0,1,2a class when making a 3-class classification. **c** The ROC curves for each subgroup for the mTICI 2b class when making a 3-class classification. **d** The ROC curves for each subgroup for the mTICI 2c,3 class when making a 3-class

classification. The shaded region around the ROC curve depicts the standard deviations at each point. High AUROC values and thin spread of standard deviations in **a** indicate that the best performance is achieved when making a 2-class classification. In all 4 subplots, while the ROC curves and the standard deviations overlap between the three subgroups, McNemar's *p* test values indicate significant improvement in the performance of ensembled networks over AP and lateral view networks

probabilities from the CNN for those cases, are analyzed and displayed in Fig. 4. In both cases, the network was able to correctly classify the input map into mTICI 0,1,2a or mTICI 2b,2c,3 outcomes. In all 4 maps, we observed that image regions that were activated were vessels, with higher activations in larger vessels. Thus, the network is making its decision based on image intensities in the vasculature. Using this method, we are able to interpret the CNN output and accept or reject the result if the salient features do not match the clinical experience. CAMs could also be used to optimize input data to improve network performance [35]. For instance, in Fig. 4, the activation of some extracranial regions is observed. Since common sense dictates that those regions should have no contribution to classification, cropping the regions should improve the data-driven model performance. Figure 4 also displays the advantages of using an ensembled network which combines information from both the AP and lateral view networks. The figure shows that misclassifications can occur when either the AP or lateral view networks are used independently; however, when information from both

networks are combined using an ensembled method, the PH map is correctly classified into the appropriate group.

The level of reperfusion was classified based on the mTICI scale which has its drawbacks [12, 13]. In addition, neuro-interventionalists can currently perform the classification themselves by visual assessment of the DSAs. However, an automated process trained using labels provided by a core lab or experts in the field could provide an objective tool across many institutions and users. This study is useful as it can be replicated for any other outcome scale such as post-op MRI or neurological evaluations; however, these are not intra-procedurally acquired. Further investigations need to be conducted using other outcome scales to classify reperfusion levels.

There are some limitations to this study. First, we are only using 383 angiograms (268 for training, 39 for in-training hyperparameter tuning or validation, and 77 for testing) that were all collected from the same center. Thus, we are currently

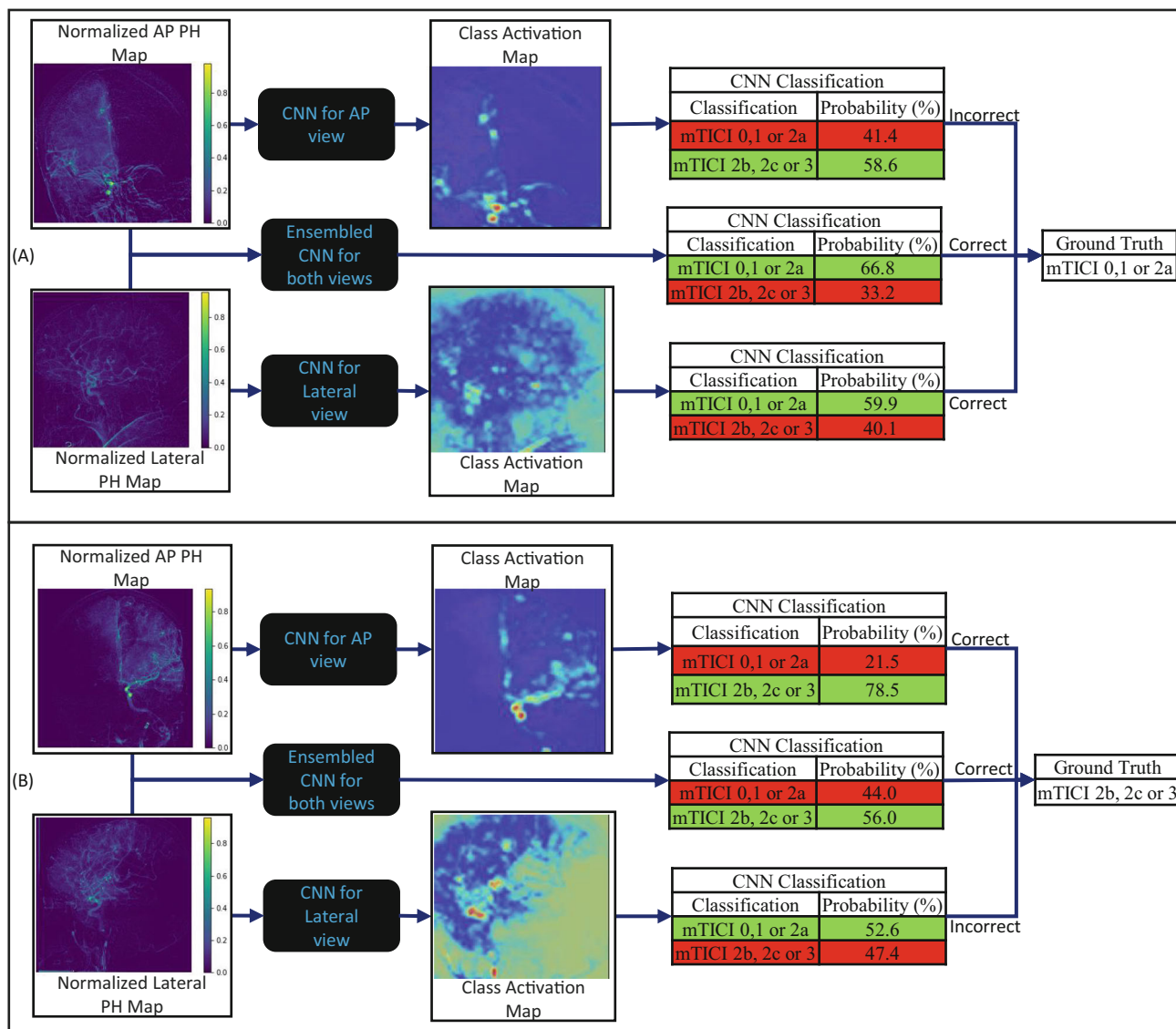


Fig. 4 a, b Two examples showing the anteroposterior (AP) and lateral digital subtraction angiography (DSA) sequences, the normalized peak height (PH) maps generated from those sequences, the class activation maps (CAMs), and the classifications from the convolutional neural network (CNN) for each view and the ensemble CNN for both views. In **a**, the AP view CNN incorrectly classifies the PH map as being mTICI 2b,2c,3 while the lateral view CNN and ensemble CNN both correctly classify the PH map as mTICI 0,1,2a. In **b**, the lateral view CNN incorrectly classifies the PH map as being mTICI 0,1,2a while the AP view

CNN and ensemble CNN both correctly classify the PH map as mTICI 2b,2c,3. This shows that misclassifications can occur when either the AP or lateral views are used independently; however, when information from both views are combined using an ensemble network, the tool is able to correctly classify the DSA into the appropriate group. In each CNN classification table, the green highlight indicates the network classification. All the CAMs show that the activation occurs in the vessels with the larger vessels causing a higher activation

limited to demonstrating only a technical feasibility study of using CNNs to assess the reperfusion level. Second, we are currently only using PH maps for this feasibility study; other API maps such as mean transit time, time to peak, and area under the TDC can also be derived and may be used in conjunction with PH maps to boost the performance of the network. Third, we are currently doing only a 2- or 3-outcome classification instead of a full-range mTICI scale. This is due to the low number of cases per outcome (mTICI 0, 82; mTICI 1, 5; mTICI 2a, 82; mTICI 2b, 140; mTICI 2c, 48; and mTICI

3, 26) which leads to a decrease in performance, as seen when going from a 2- to 3-outcome classification. Fourth, we are currently not identifying the location of LVOs, rather just the reperfusion status. Fifth, preprocessing methods such as cropping of DSAs to exclude early arterial and venous phases and identifying inlet vessels for normalization of PH maps are currently not automated. Lastly, the current normalization process only uses one point from the main feeding artery; we will investigate averaging over multiple points from the main feeding artery; this may provide a more effective normalization.

This study proves the feasibility of using CNNs to extract encoded hemodynamic information from API by assessing the level of reperfusion during an MT in patients with an LVO AIS. While this study provides neuro-interventionalists with a more robust tool to evaluate the level of reperfusion during MTs rather than relying solely on subjective assessment of DSAs, it also proved the feasibility of using CNNs with API maps and can thus be possibly used for other endovascular interventions.

Conclusion

This is a novel attempt at using a data-driven approach to classify DSAs based on the nature of flow in the neurovasculature by extracting hemodynamic information encoded in quantitative angiographic maps. In this study, we proved the feasibility of this approach to make decisions regarding the reperfusion status of patients undergoing a MT to treat an emergent LVO AIS. The CNN succeeded in making this assessment with an accuracy of 81.0%, AUROC of 0.86, and MCC of 0.62 when making a 2-outcome classification. When making a 3-outcome classification, the network succeeded with an accuracy of 64.0%, AUROC of 0.85 for mTICI 0,1,2a, AUROC of 0.74 for mTICI 2b, AUROC of 0.78 for mTICI 2c,3, and an MCC of 0.43.

Contributors MMSB, KVS, MM, AHS, and CNI conceived and designed the research.

JMD, EIL, KVS, and AHS performed all the clinical procedures.

MMSB, MW, RAR, and CNI collected the data.

MMSB analyzed the data.

MMSB performed the statistical analysis.

CNI handled funding and supervision.

MMSB drafted the manuscript.

All authors made critical revisions of the manuscript and reviewed the final version.

Code availability Additional data may be made available by contacting the corresponding author.

Data availability Additional data may be made available by contacting the corresponding author.

Funding This work is supported by the James H. Cummings Foundation and the University at Buffalo Clinical and Translational Science Institute (CTSI).

Compliance with ethical standards

Competing interests Maxim Mokin—Grants: Principal investigator NIH R21NS109575. Consultant: Medtronic, Cerenovus. Stock options: Serenity Medical, Synchron, Endostream, VICIS.

Ciprian Ionita—Equipment grant from Canon Medical Systems, support from the Cummings Foundation, NIH R21 grant.

Jason Davies—Research grant: National Center for Advancing Translational Sciences of the National Institutes of Health under award number KL2TR001413 to the University at Buffalo. Speakers' bureau:

Penumbra; Honoraria: Neurotrauma Science, LLC. Shareholder/ownership interests: RIST Neurovascular.

Kenneth Snyder—Consulting and teaching for Canon Medical Systems Corporation, Penumbra Inc., Medtronic, and Jacobs Institute. Co-Founder: Neurovascular Diagnostics, Inc.

Elad Levy—Shareholder/ownership interests: NeXtGen Biologics, RAPID Medical, Claret Medical, Cognition Medical, Imperative Care (formerly the Stroke Project), Rebound Therapeutics, StimMed, Three Rivers Medical. National Principal Investigator/Steering Committees: Medtronic (merged with Covidien Neurovascular) SWIFT Prime and SWIFT Direct Trials. Honoraria: Medtronic (training and lectures). Consultant: Claret Medical, GLG Consulting, Guidepoint Global, Imperative Care, Medtronic, Rebound, StimMed. Advisory Board: Stryker (AIS Clinical Advisory Board), NeXtGen Biologics, MEDX, Cognition Medical, Endostream Medical. Site Principal Investigator: CONFIDENCE study (MicroVention), STRATIS Study—Sub I (Medtronic).

Adnan Siddiqui—Research grant: NIH/NINDS 1R01NS091075 as a co-investigator for “Virtual Intervention of Intracranial Aneurysms.” Financial interest/investor/stock options/ownership: Amnis Therapeutics, Apama Medical, Blink TBI Inc., Buffalo Technology Partners Inc., Cardinal Consultants, Cerebrotech Medical Systems, Inc., Cognition Medical, Endostream Medical Ltd., Imperative Care, International Medical Distribution Partners, Neurovascular Diagnostics Inc., Q’Apel Medical Inc., Rebound Therapeutics Corp., Rist Neurovascular Inc., Serenity Medical Inc., Silk Road Medical, StimMed, Synchron, Three Rivers Medical Inc., Viseon Spine Inc. Consultant/advisory board: Amnis Therapeutics, Boston Scientific, Canon Medical Systems USA Inc., Cerebrotech Medical Systems Inc., Cerenovus, Corindus Inc., Endostream Medical Ltd., Guidepoint Global Consulting, Imperative Care, Integra LifeSciences Corp., Medtronic, MicroVention, Northwest University—DSMB Chair for HEAT Trial, Penumbra, Q’Apel Medical Inc., Rapid Medical, Rebound Therapeutics Corp., Serenity Medical Inc., Silk Road Medical, StimMed, Stryker, Three Rivers Medical, Inc., VasSol, W.L. Gore & Associates. Principal investigator/steering comment of the following trials: Cerenovus LARGE and ARISE II; Medtronic SWIFT PRIME and SWIFT DIRECT; MicroVention FRED & CONFIDENCE; MUSC POSITIVE; and Penumbra 3D Separator, COMPASS, and INVEST.

Ethics approval Institutional review board approval was obtained, and informed consent was waived for this Health Insurance Portability and Accountability Act-compliant retrospective study (IRB ID: STUDY00001529).

Informed consent This is a retrospective study; hence, no informed consent is needed.

References

1. Rennert RC et al (2019) Epidemiology, natural history, and clinical presentation of large vessel ischemic stroke. *Neurosurgery* 85(suppl_1):S4–S8
2. Waqas M, Mokin M, Primiani CT, Gong AD, Rai HH, Chin F, Rai AT, Levy EI, Siddiqui AH (2020) Large vessel occlusion in acute ischemic stroke patients: a dual-center estimate based on a broad definition of occlusion site. *J Stroke Cerebrovasc Dis* 29(2):104504
3. Waqas M, Rai AT, Vakharia K, Chin F, Siddiqui AH (2020) Effect of definition and methods on estimates of prevalence of large vessel occlusion in acute ischemic stroke: a systematic review and meta-analysis. *J Neurointerv Surg* 12(3):260–265
4. Smith WS, Sung G, Saver J, Budzik R, Duckwiler G, Liebeskind DS, Lutsep HL, Rymer MM, Higashida RT, Starkman S, Gobin

- YP, Multi MERCI Investigators, Frei D, Grobelny T, Hellinger F, Huddle D, Kidwell C, Koroshetz W, Marks M, Nesbit G, Silverman IE (2008) Mechanical thrombectomy for acute ischemic stroke: final results of the multi MERCI trial. *Stroke* 39(4):1205–1212
5. Berlis A, Lutsep H, Barnwell S, Norbash A, Wechsler L, Jungreis CA, Woolfenden A, Redekop G, Hartmann M, Schumacher M (2004) Mechanical thrombolysis in acute ischemic stroke with endovascular photoacoustic recanalization. *Stroke* 35(5):1112–1116
 6. Smith WS, Sung G, Starkman S, Saver JL, Kidwell CS, Gobin YP, Lutsep HL, Nesbit GM, Grobelny T, Rymer MM, Silverman IE, Higashida RT, Budzik RF, Marks MP (2005) Safety and efficacy of mechanical embolectomy in acute ischemic stroke: results of the MERCI trial. *Stroke* 36(7):1432–1438
 7. Bourcier R, Goyal M, Liebeskind DS, Muir KW, Desal H, Siddiqui AH, Dippel DWJ, Majoie CB, van Zwam WH, Jovin TG, Levy EI, Mitchell PJ, Berkhemer OA, Davis SM, Derraz I, Donnan GA, Demchuk AM, van Oostenbrugge RJ, Kelly M, Roos YB, Jahan R, van der Lugt A, Sprengers M, Velasco S, Lycklama à Nijeholt GJ, Ben Hassen W, Burns P, Brown S, Chabert E, Krings T, Choe H, Weimar C, Campbell BCV, Ford GA, Ribo M, White P, Cloud GC, San Roman L, Davalos A, Naggara O, Hill MD, Bracard S, for the HERMES Trialists Collaboration (2019) Association of time from stroke onset to groin puncture with quality of reperfusion after mechanical thrombectomy: a meta-analysis of individual patient data from 7 randomized clinical trials. *JAMA Neurology* 76(4):405–411
 8. Evans MR et al (2017) Revolution in acute ischaemic stroke care: a practical guide to mechanical thrombectomy. *Pract Neurol* 17(4):252–265
 9. Rava RA et al (2020) Assessment of computed tomography perfusion software in predicting spatial location and volume of infarct in acute ischemic stroke patients: a comparison of Sphere, Vitrea, and RAPID. *J Neurointerv Surg*
 10. Higashida RT, Furlan AJ (2003) Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke. *Stroke* 34(8):e109–e137
 11. Almekhlafi MA, Mishra S, Desai JA, Nambiar V, Volny O, Goel A, Eesa M, Demchuk AM, Menon BK, Goyal M (2014) Not all “successful” angiographic reperfusion patients are an equal validation of a modified TICI scoring system. *Interv Neuroradiol* 20(1):21–27
 12. Kallmes DF (2012) TICI: if you are not confused, then you are not paying attention. *Am J Neuroradiol* 33(5):975–976
 13. Sattar A et al (2014) E-011 TICI quantified: automated cerebral revascularization grading in acute ischemic stroke. *J NeuroInterv Surg* 6(Suppl 1):A43–A43
 14. De Leacy RA et al (2019) Wide-neck bifurcation aneurysms of the middle cerebral artery and basilar apex treated by endovascular techniques: a multicentre, core lab adjudicated study evaluating safety and durability of occlusion (BRANCH). *Journal of NeuroInterventional Surgery* 11(1):31–36
 15. Rezek I, Mousan G, Wang Z, Murad MH, Kallmes DF (2013) Effect of core laboratory and multiple-reader interpretation of angiographic images on follow-up outcomes of coiled cerebral aneurysms: a systematic review and meta-analysis. *Am J Neuroradiol* 34(7):1380–1384
 16. Zhang G, Treurniet KM, Jansen IGH, Emmer BJ, van den Berg R, Marquering HA, Uyttenboogaart M, Jenniskens SFM, Roos YBWEM, van Doormaal PJ, van Es ACGM, van der Lugt A, Vos JA, Lycklama à Nijeholt GJ, van Zwam WH, Shi H, Yoo AJ, Dippel DWJ, Majoie CB, for the MR CLEAN Registry Investigators (2018) Operator versus core lab adjudication of reperfusion after endovascular treatment of acute ischemic stroke. *Stroke* 49(10):2376–2382
 17. Bhurwani, M.M.S., et al. *Feasibility study of deep neural networks to classify intracranial aneurysms using angiographic parametric imaging*. In *Medical Imaging 2019: Computer-Aided Diagnosis*. 2019. International Society for Optics and Photonics
 18. Thompson HK Jr et al (1964) Indicator transit time considered as a gamma variate. *Circ Res* 14(6):502–515
 19. Albers GW, Wald MJ, Mlynash M, Endres J, Bammer R, Straka M, Maier A, Hinson HE, Sheth KN, Taylor Kimberly W, Molyneaux BJ (2019) Automated calculation of Alberta stroke program early CT score: validation in patients with large hemispheric infarct. *Stroke* 50(11):3277–3279
 20. Bhurwani, M.M.S., et al., Feasibility study for use of angiographic parametric imaging and deep neural networks for intracranial aneurysm occlusion prediction. *Journal of NeuroInterventional Surgery*, 2019
 21. Mokin M, Khalessi AA, Mocco J, Lanzino G, Dumont TM, Hanel RA, Lopes DK, Fessler RD, Ringer AJ, Bendok BR, Veznedaroglu E, Siddiqui AH, Hopkins LN, Levy EI (2014) Endovascular treatment of acute ischemic stroke: the end or just the beginning? *Neurosurg Focus* 36(1):E5
 22. Zaidat OO, Yoo AJ, Khatri P, Tomsick TA, von Kummer R, Saver JL, Marks MP, Prabhakaran S, Kallmes DF, Fitzsimmons BF, Mocco J, Wardlaw JM, Barnwell SL, Jovin TG, Linfante I, Siddiqui AH, Alexander MJ, Hirsch JA, Wintermark M, Albers G, Woo HH, Heck DV, Lev M, Aviv R, Hacke W, Warach S, Broderick J, Derdeyn CP, Furlan A, Nogueira RG, Yavagal DR, Goyal M, Demchuk AM, Bendszus M, Liebeskind DS, Cerebral Angiographic Revascularization Grading (CARG) Collaborators, STIR Revascularization working group, STIR Thrombolysis in Cerebral Infarction (TICI) Task Force (2013) Recommendations on angiographic revascularization grading standards for acute ischemic stroke: a consensus statement. *Stroke* 44(9):2650–2663
 23. Chollet, F., *Keras*. 2015
 24. Bluemke DA et al (2019) Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers-from the Radiology Editorial Board. *Radiology* 294:192515
 25. Xu Q-S, Liang Y-Z (2001) Monte Carlo cross validation. *Chemom Intell Lab Syst* 56(1):1–11
 26. Storn R, Price K (1997) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J Glob Optim* 11(4):341–359
 27. Virtanen P et al (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 17(3):261–272
 28. Zhou, B., et al. Learning deep features for discriminative localization. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016
 29. Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta (BBA)-Protein Struct* 405(2):442–451
 30. Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21(1):6
 31. Goyal M, Menon BK, van Zwam WH, Dippel DWJ, Mitchell PJ, Demchuk AM, Dávalos A, Majoie CB, van der Lugt A, de Miquel MA, Donnan GA, Roos YBWEM, Bonafe A, Jahan R, Diener HC, van den Berg LA, Levy EI, Berkhemer OA, Pereira VM, Rempel J, Millán M, Davis SM, Roy D, Thornton J, Román LS, Ribo M, Beumer D, Stouch B, Brown S, Campbell BCV, van Oostenbrugge RJ, Saver JL, Hill MD, Jovin TG (2016) Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet* 387(10029):1723–1731
 32. Dargazanli C, Fahed R, Blanc R, Gory B, Labreuche J, Duhamel A, Marnat G, Saleme S, Costalat V, Bracard S, Desal H, Mazighi M, Consoli A, Piotin M, Lapergue B, Redjem H, Ciccio G, Smajda S, Desilles JP, Rodesch G, Coskun O, di Maria F, Bourdain F, Decroix JP, Wang A, Tchikviladze M, Evrard S, Eker O, Turjman F, Labeyrie PE, Riva R, Mounayer C, Saleme S, Bonafé A, Gascou

- G, Tonnelet R, Derelle AL, Anxionnat R, Bourcier R, Dumas-Duport B, Berge J, Barreau X, Djemmane L (2018) Modified thrombolysis in cerebral infarction 2C/thrombolysis in cerebral infarction 3 reperfusion should be the aim of mechanical thrombectomy: insights from the ASTER trial (Contact Aspiration Versus Stent Retriever for Successful Revascularization). *Stroke* 49(5):1189–1196
33. Cramer H (1946) *Mathematical methods of statistics*. Princeton University Press, Princeton
34. Powers, D.M., *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*. 2011
35. Podgorsak, A., et al. Initial evaluation of the use of a convolutional neural network to determine coronary artery disease severity using computed tomography angiography. in *Medical physics*. 2019. WILEY 111 RIVER ST, HOBOKEN 07030–5774, NJ USA

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.