AutoTICI: automatic brain tissue reperfusion scoring on 2D DSA images of acute ischemic stroke patients

Su, R.S.; Cornelissen, S.A.P.; Sluijs, M. van der; Es, A.C.G.M. van; Zwam, W.H. van; Dippel, D.W.J.; ... ; Walsum, T. van

Note: To cite this publication please use the final published version (if applicable).

autoTICI: Automatic Brain Tissue Reperfusion Scoring on 2D DSA Images of Acute Ischemic Stroke Patients

Ruisheng Su, Sandra A.P. Cornelissen, Matthijs van der Sluijs, Adriaan C.G.M. van Es, Wim H. van Zwam, Diederik W.J. Dippel, Geert Lycklama, Pieter Jan van Doormaal, Wiro J. Niessen, Aad van der Lugt, and Theo van Walsum

Abstract— The Thrombolysis in Cerebral Infarction (TICI) score is an important metric for reperfusion therapy assessment in acute ischemic stroke. It is commonly used as a technical outcome measure after endovascular treatment (EVT). Existing TICI scores are defined in coarse ordinal grades based on visual inspection, leading to inter- and intra-observer variation. In this work, we present autoTICI, an automatic and quantitative TICI scoring method. First, each digital subtraction angiography (DSA) acquisition is separated into four phases (non-contrast, arterial, parenchymal and venous phase) using a multi-path convolutional neural network (CNN), which exploits spatio-temporal features. The network also incorporates sequence level label dependencies in the form of a state-transition matrix. Next, a minimum intensity map (MINIP) is computed using the motion corrected arterial and parenchymal frames. On the MINIP image, vessel, perfusion and background pixels are segmented. Finally, we quantify the autoTICI score as the ratio of reperfused pixels after EVT. On a routinely acquired multi-center dataset, the proposed autoTICI shows good correlation with the extended TICI (eTICI) reference with an average area under the curve (AUC) score of 0.81. The AUC score is 0.90 with respect to the dichotomized eTICI. In terms of clinical outcome prediction, we demonstrate that autoTICI is overall comparable to eTICI.

Index Terms— Stroke, DSA, autoTICI, Deep Learning, Brain Tissue Perfusion, Phase Classification, MR CLEAN Registry

I. INTRODUCTION

A. Clinical Background

STROKE remains one of the worldwide leading causes of death and serious long-term disability [1]. Due to the ageing population, stroke incidence is rising, which is posing a large and still increasing public health burden to the society. Ischemic stroke, which is caused by an occluded artery of the brain, is the most common stroke type, accounting for about 88% of all strokes [2].

Recent studies have shown that endovascular therapy (EVT) improves outcome in patients with acute ischemic stroke caused by a large vessel occlusion (LVO) in the anterior circulation [3], [4]. This led to novel studies on maximizing patient benefits before and during EVT. Many studies have focused on pre-interventional patient selection and outcome prediction based on scores such as Alberta Stroke Program Early CT Score (ASPECTS) [5] and collateral score [6]. A recent study [7] demonstrated that peri-procedural imaging data can be valuable for predicting treatment outcome, which can in turn assist interventionalists to achieve better treatment quality.

DSA is the imaging modality to guide the EVT procedure. Despite rapid research progress in computational stroke related biomarker extraction from computed tomographic angiography (CTA) [8] and CT Perfusion (CTP) [9] images, automatic peri-procedural imaging biomarker extraction from DSA is yet to be further explored.

One of the most widely adopted metrics for evaluation of EVT quality and prediction of functional outcome on DSA images is the so-called Thrombolysis In Cerebral Infarction (TICI) score [2] and its variants, such as modified TICI (mTICI) [10], extended TICI (eTICI) [11] and expanded TICI [12]. The TICI scores define the extent of brain reperfusion. For example, the eTICI score is defined as follows:

- **eTICI 0**: no perfusion or antegrade flow in the target downstream territory (TDT). TDT refers to the occluded brain region that was supplied via antegrade blood flow prior to stroke onset [13];
- **eTICI 1**: blood flow past initial site of occlusion, but with minimal brain tissue perfusion;
- **eTICI 2A**: perfusion of $\leq 50\%$ of TDT;
- **eTICI 2B**: perfusion of $\geq 50\%$ of TDT;
- **eTICI 2C**: nearly complete perfusion except for slow flow or presence of small emboli in distal cortical vessels;
- **eTICI 3**: complete perfusion.

Above TICI score variants were introduced as attempts towards standardizing EVT treatment success scoring. However, these grading metrics suffer from several shortcomings:

- Inter- and intra-observer variation: TICI assessment by visual inspection is inherently error prone, as it is subject to various factors, such as experience level of the rater and inspection attentiveness. Moreover, a recent study showed that TICI scores are generally overestimated by operators during EVT compared to core-lab raters [14];
- Coarse ordinal scale: TICI-like scores are mostly defined in 4-6 grades. Several studies have indicated that EVT outcome is associated with greater degrees of reperfusion [15]–[19]. More granular perfusion grading systems would probably help to better assess treatment success;
- Conceptual confusion: in EVT terms, reperfusion is the antegrade restoration of a capillary blush, whereas recanalization generally refers to the restoration of blood flow past the arterial occlusion [13]. In other words, recanalization is a necessary but insufficient condition for reperfusion. Although TICI scores define reperfusion scales, the concept of recanalization and reperfusion is often interchanged during visual TICI assessment in clinical practice [13]. As a result, the cases of successful recanalization without adequate brain tissue reperfusion are usually overlooked, leading to overestimated TICI.

To mitigate the above limitations of existing TICI scoring mechanisms, we pursue an objective, robust, and quantitative TICI scoring algorithm in a fully automated manner.

B. Related Work

In recent years, the field of computer vision evolved rapidly. Triggered by the advances in deep learning particularly, automated medical imaging analysis has gained momentum as well. In the field of stroke imaging, various algorithms on automatic quantification of imaging biomarkers have been proposed lately. These algorithms can be categorized into two groups based on the target imaging modality: pre-treatment imaging and peri-procedural imaging.

For pre-treatment prognosis, MRI, CTA and CTP images are widely used. Robben *et al*. [9] exploited a multi-path CNN to predict the final infarct volume from CTP images and treatment metadata. Nielsen *et al*. [20] demonstrated state-of-the-art performance of a deep CNN on final lesion volume and treatment outcome prediction on MRI images. Su *et al*. [8] utilized a U-Net architecture to quantify collateral scores from 3D CTA images. Related to brain perfusion estimation,

McKinley *et al*. [21] developed an automated pipeline for penumbra volume estimation based on random forests using multi-modal MRI.

For peri-procedural evaluation of treatment effect, DSA is still considered the standard modality. Automatically quantified biomarkers on DSA have received less attention. As an example, Liebeskind *et al*. [22] recently proposed a machine learning based automatic arterial input function (AIF) extraction method. Concerning TICI scoring on DSA, to the best of our knowledge, no fully automatic algorithm has been proposed yet. Nevertheless, Prasetya *et al*. [23] recently developed a quantitative TICI (qTICI) scoring pipeline which demands manual annotation. First, given a DSA acquisition under assessment, venous phase frames need to be manually removed by a neuro/interventional radiologist so to exclude the retrograde perfusion (via pial collaterals), as only antegrade (via recanalization) perfusion defines the reperfusion. Second, an estimated TDT is delineated by an expert observer for each artery occlusion of the DSA under assessment. Lastly, qTICI is defined as the percentage of reperfused pixels within TDT based on a threshold relative to the maximum intensity. Prasetya *et al*. [23] demonstrated reliable correlation between qTICI and eTICI, revealing the potential value of quantitative TICI as an objective biomarker. Nevertheless, as qTICI relies on a series of manual steps, such a semi-automatic approach is more time consuming than traditional TICI grading procedures in practice. In addition, these manual steps are inevitably subjective, which can introduce new factors to inter- and intra-observer variability.

Beyond qTICI, we seek a fully automatic, yet explainable step by step approach for quantitative TICI scoring.

C. Contributions

The main contribution of this work is two-fold:

- we propose a fully automatic and quantitative TICI scoring algorithm, as an objective and robust alternative to existing visual inspection based TICI scoring;
- we assess the proposed methods on a large multi-center data set acquired in clinical routine.

The remainder of this paper is organized as follows: Section II provides a detailed overview of our proposed methods and Section III describes the data used for experiments. Section IV presents the experiment results and performance evaluations of the proposed methods, followed by further discussions in Section V. Finally, Section VI concludes the paper.

II. METHOD

Existing TICI scores are visually graded by estimating the extent of brain tissue-level reperfusion (post-EVT, numerator) in the initial TDT (pre-EVT, denominator). This would require a visual comparison between pre- and post-EVT acquisitions. Following this definition and the visual TICI scoring principles, we propose to automatically extract the numerator and denominator for quantitative TICI scoring.
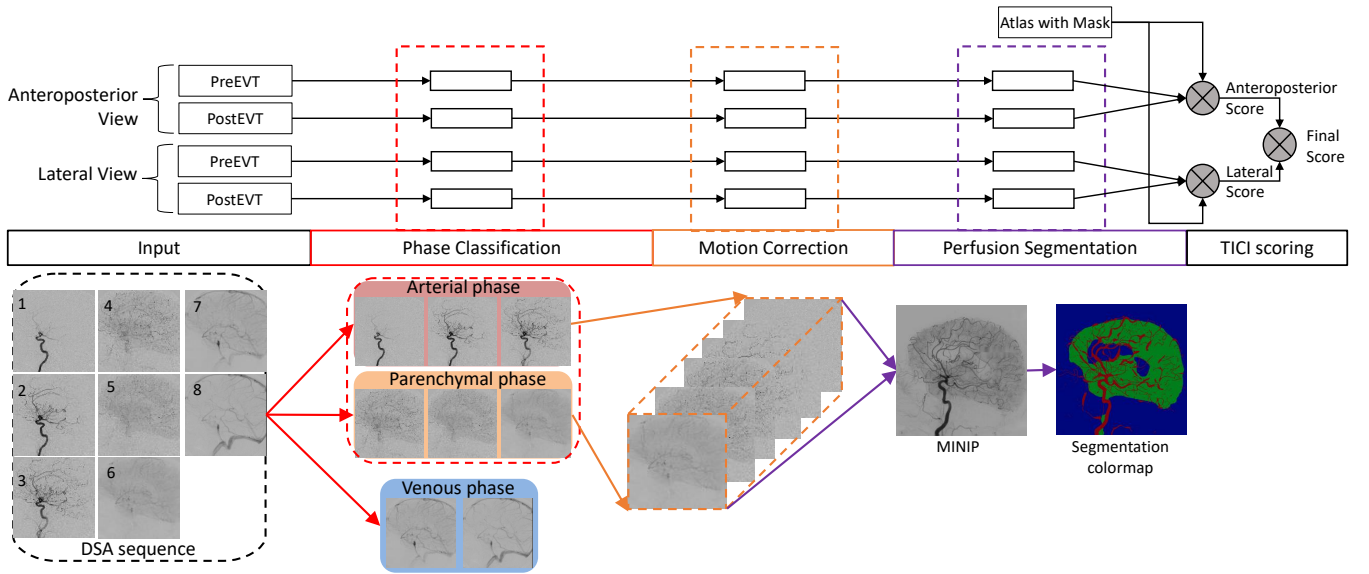
Fig. 1: Proposed pipeline for automatic TICI scoring.

As illustrated in Fig. 1, the proposed algorithm comprises four components: phase classification, motion correction, perfusion segmentation and TICI scoring. First, given four DSA acquisitions for each patient, both anteroposterior (AP) and lateral views for both pre-EVT and post-EVT, each of these acquisitions is separated into four phases, being non-contrast, arterial, parenchymal and venous phase. We trained a multi-channel CNN model to label image phases based on spatio-temporal textural features, as well as sequence label transition rules. Subsequently, venous phase frames are removed from the sequence. Non-contrast frames are also excluded, as they contain no valid information, but merely motion artifacts. The remaining sequence frames are aligned based on affine registration to correct motion artifacts. Third, a single 2D minimum intensity map (MINIP) is calculated from the aligned sequence, which is then segmented into vessels, perfused area and non-perfused (including background) area. Finally, by comparing the perfusion area before and after EVT treatment, a quantitative TICI score is deduced as the percentage of re-perfused area out of previously occluded downstream territory. Each of the components is further detailed below.

A. Phase Classification

The purpose of the phase classification step is to recognize and separate the venous phase frames; these frames need to be excluded from the series, because these may contain late perfusion attributed to retrograde flow via pial collaterals.

In a broad sense, phase classification can be seen as a video sequence labelling problem. Although 2D CNNs have been shown powerful in extracting representative features from images, temporal information is discarded when handling videos. Many recent studies have investigated how to effectively incorporate temporal frame dependencies in videos sequence labelling, various innovations have been proposed. Karpathy *et al.* [24] explored feature fusion of multiple frames. Adding recurrent layers [25], [26] have been shown effective

in capturing frame ordering and long-range dependencies. Alternatively, 3D CNNs have been studied several times [27], [28] for spatio-temporal feature extraction. More two-stream networks have been recently proposed to capture low-level motion based on optical flow [29], [30]. A summary of related works can be found in [30]. Above methods focus on learning a global spatio-temporal description which best represents a video sequence. In this work, we are especially interested in detecting the phase transitions between individual frames. On one hand, temporal dependencies do provide valuable hint for labelling current frame. One the other hand, such video descriptions are not necessarily sensitive to frame shifts.

Specifically on phase classification, existing methods can be grouped into two types: supervised and unsupervised. Lee *et al.* [31] used independent component analysis (ICA) to separate a DSA acquisition into three images (arterial, parenchymal and venous). Such unsupervised methods rely on the assumption of an expected number of components. For supervised methods, Schuldhaus *et al.* [32] utilized Rosenblatt perceptron-based classification using handcrafted positional intensity features.

In this work, we propose a deep learning framework using a CNN with a customized CRF for automatic phase classification. An overview of the framework is shown in Fig. 2. Provided a DSA sequence, a retrained ResNet18 takes each frame together with its direct neighbors as input and predicts a phase label. Next, on the sequence level, a tailored state-transition diagram is applied on the predicted label probabilities to ensure a logical label sequence based on maximum likelihood.

1) Phase Definition: A complete run of a DSA acquisition can be categorized into three phases based on the blood flow: the arterial, parenchymal and venous phase. Fig. 3 shows an example of each phase in AP and lateral views. In practice, the first or last few frames generally have no visible contrast. In this work, we consider non-contrast as an additional phase. The phases were defined according to the following criteria
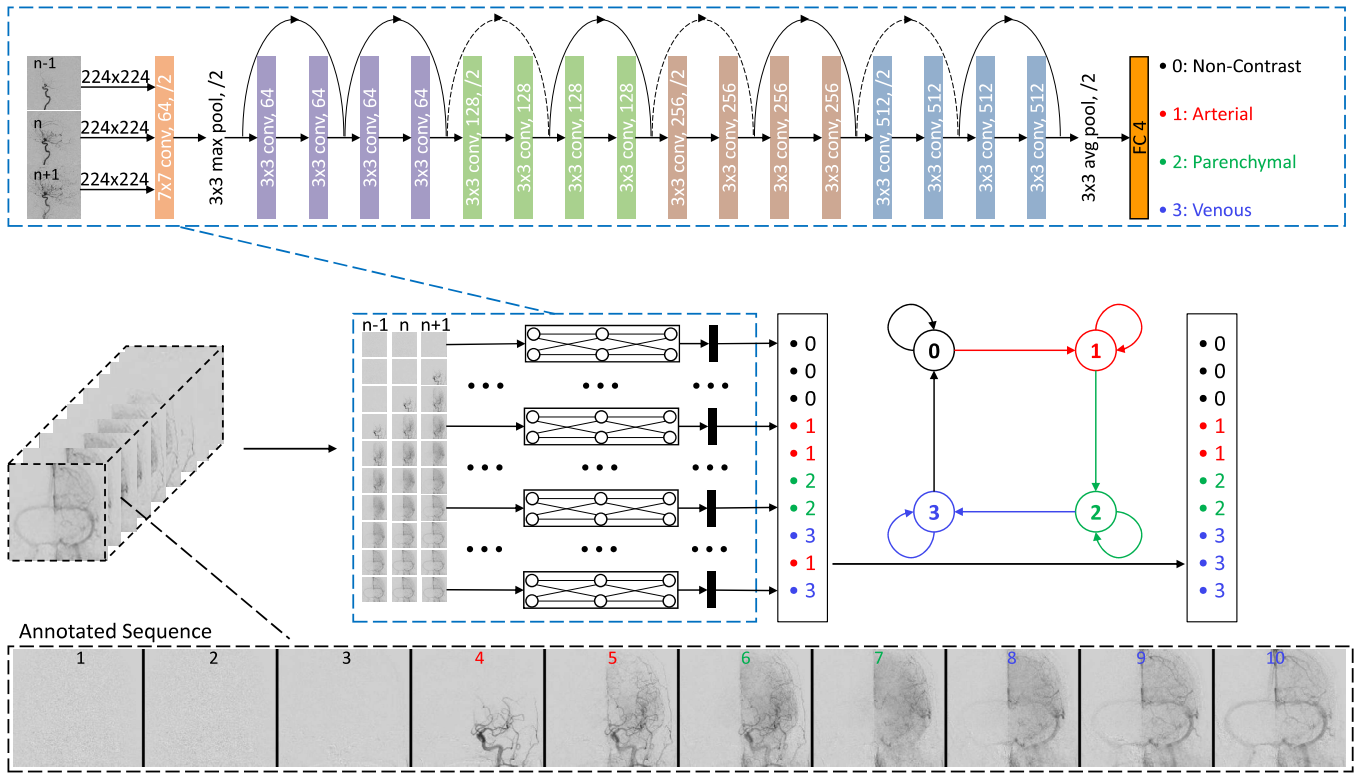
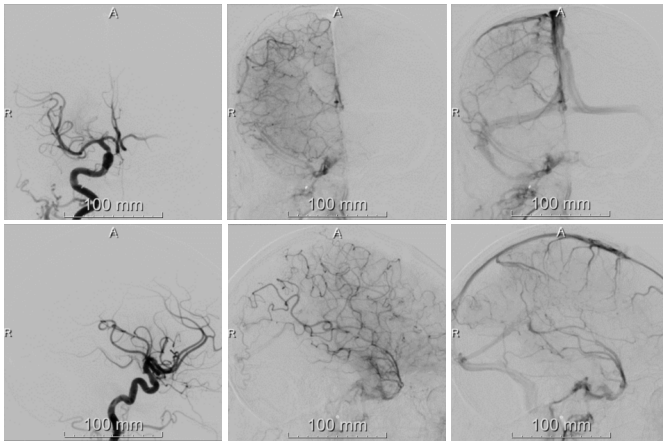Fig. 2: An overview of the phase classification framework.

Fig. 3: DSA examples of arterial (left column), parenchymal (middle column) and venous (right column) phases. Top row: frontal view; bottom row: AP view.

(determined by a clinical expert):

- Arterial phase:
  - First frame: appearance of contrast in image;
  - Last frame: appearance of contrast in cortical arterial branches. These are the peripherally located arterial branches which run on the cortex of the brain;
- Parenchymal phase:
  - First frame: right after last arterial frame.
  - Last frame: right before appearance of contrast in the superior sagittal sinus (SSS);
- Venous phase:
  - First frame: appearance of contrast in the SSS;

- Last frame: the last frame with clearly visible contrast thereafter.

It should be noted that a complete run of DSA does not necessarily contain all four phases. Whether these phases are all present depends on the level of the occlusion, the timing of the acquisition with respect to the contrast injection and the frame rate. In case of a carotid-T occlusion, mostly only an arterial phase is present because no contrast penetrates the affected hemisphere from the vessel where contrast was injected, and in some short acquisitions, the venous phase may not be present.

2) Network Architecture: The proposed network architecture is shown in Fig. 2. The network is based on ResNet-18 proposed by He *et al.* [33], to which we made modifications to adapt it to our task. In our case, the input is set to three consecutive 2D sequence frames, each for one input channel. In such a way, local features of three consecutive frames are fused. Our network ends with a fully connected (FC) layer with softmax normalization, which outputs an array of four elements with predicted per-phase probabilities.

3) Constrained Sequence Labelling: On the sequence level, a handcrafted state transition diagram (see Fig. 2) is applied to suppress invalid phase transitions from frame to frame. The rationale is that although the network fuses neighboring frames which possess temporal contrast flow information, the global sequence level information is not fully exploited. For instance, in a correctly ordered sequence, no arterial phase should appear during or after parenchymal phase. All sequence labels should follow a certain state transition logic. We embed this logic to bring further robustness to the network.

In this work, we adopted the idea of conditional random

fields (CRFs) [34] and tailored it according to our task. CRFs train a transition matrix between previous and current labels by maximum likelihood learning. Instead of training the conditional dependency between frames in a data-driven manner, we handcrafted the structural logic between labels directly into a transition matrix. The state transition diagram in Fig. 2 shows the allowed transitions between phases, which translates into the following transition matrix, where all allowed transitions are denoted as $T_{i,j} = 1$:

$$T = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad . \tag{1}$$

We intentionally assigned equal weights to all allowed transitions ($T_{i,j} = 1$) to avoid bias introduced by the training data itself. The remaining decoding and inference steps were kept the same as in CRF.

After phase classification, only arterial and parenchymal phase frames are carried on for further processing. Fig. 4 shows an example of the venous phase removal effect.
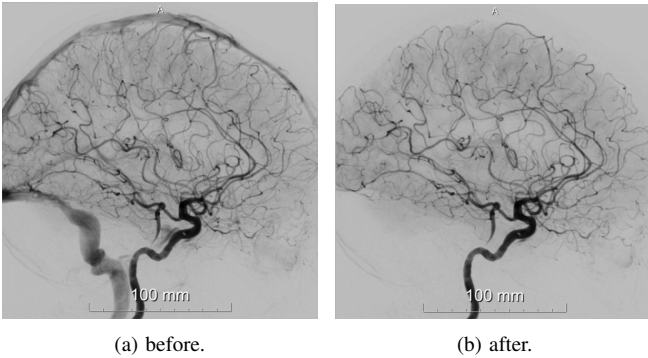

(a) before. (b) after.

Fig. 4: A MINIP before (a) and after (b) venous phase removal.

B. Motion Corrected MINIP

The image segmentation and quantification steps rely on the MINIP, which is a 2D image with each pixel being its minimum intensity value (which reflects the attenuation caused by iodine contrast agent) across the time axis of the sequence. Patient movements (voluntary or involuntary) during image acquisition may hamper this quantification, as they may cause vessels to be less apparent, and to overlap with brain tissue in the MINIP image. Therefore, the purpose of motion correction is to correct frame misalignments introduced by patient motion during image acquisition.

In this work, we used an affine registration that optimizes the Mattes mutual information [35] with the adaptive stochastic gradient descent (ASGD) [36] optimizer. Affine registration was chosen to handle possible scaling and sheer while avoiding additional artifacts that could be caused by B-spline registration. Mattes mutual information could allow us to handle contrast flow differences between frames. Motion correction was performed using the SimpleElastix toolbox [37]. The parameter file used in this work can be found at `https:`

`//elastix.lumc.nl/modelzoo/par0063/`. Fig. 5 visualizes the sequence motion compensation on the calculated MINIP image.

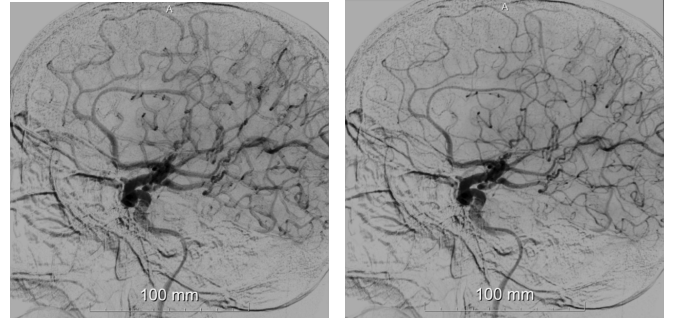(a) without motion correction. (b) with motion correction.

Fig. 5: A MINIP image without (a) and with (b) motion correction.

C. Perfusion Segmentation

The result of the previous steps is a MINIP of the motion corrected DSA sequence, that only contains arterial and parenchymal phases. From the constructed 2D MINIP image, (see Fig. 1), we segment blood vessels and perfused area from the image background, which serves as a prerequisite for the subsequent (re-)perfusion quantification. Firstly, the contrast-opacified blood vessels are segmented using the multi-scale Frangi filter [38], which is a common technique for enhancing vessel-like structures. In this work, the scale of Frangi vesselness filter was set to σ_{min} = 2 and σ_{max} = 12 according to the vessel diameter range from 5 to 35 pixels. We chose six different scales (scale step = 2) for detection of vessels of various sizes. Default value of α = 0.5 was used. For γ we chose 15 based on visual inspection on a set of representative images. An intensity threshold of 0.08 was applied on filter output to segment vessel structures. We opted for such a proven vessel segmentation approach instead of deep learning methods, as we do not rely on high precision vessel segmentation, rather we exclude the vessel pixels such that the perfusion area can be better segmented. Such a traditional algorithm satisfied our needs. Next, the remaining pixels are automatically clustered into two groups, perfused and non-perfused (including background), using Otsu's thresholding technique [39], which is an effective and parameter free method based on the image intensity histogram. Image background is included in the histogram so that perfused pixels can be recognized. Finally, a segmentation colormap is constructed (see Fig. 1), where vessels, perfused and non-perfused pixels are represented in red, green and blue respectively.

D. TICI Quantification

Finally, a quantitative TICI score can be computed from the resulted segmentation colormap. Both AP and lateral views are evaluated. Traditional visual TICI-like scores assess the change in the level of perfusion after EVT with respect to pre-EVT. Similarly, the proposed autoTICI score quantifies the

ratio of re-perfused area versus the initial TDT, which can be formulated as follows:

$$autoTICI = \frac{TDT_{preEVT} \cap P_{postEVT}}{TDT_{preEVT}} \quad , \qquad (2)$$

where $P_{postEVT}$ denotes perfused pixels after EVT (green and orange area in Fig. 6c) and TDT_{preEVT} is the initially occluded brain area prior to EVT (white area in Fig. 6). While $P_{postEVT}$ is available in the segmentation colormap, the distinction between non-perfused pixels and pure image background in the pre-EVT image is yet to be made. To this end, we introduce an atlas-based approach to exclude the out-of-skull pixels. More specifically, we utilize masked DSA atlases to align the brain area from atlas to the post-EVT MINIP of patients. 22 DSA acquisitions were selected as atlases from patients without stroke (Section III-B.5). We applied affine registration (similar to the registration used in motion correction in Section II-B) to spatially align the atlases with the post-EVT image. The atlas brain mask was used as the moving image mask during registration. The final registered atlas for each subject is the one with the maximum Mattes mutual information [35]. The pre-EVT MINIP is also registered to post-EVT MINIP. In such a way, the brain region mask on the registered atlas can be mapped to pre- and post-EVT MINIP. TDT_{preEVT} and $P_{postEVT}$ are determined within the aligned brain mask. As illustrated in Fig. 6, autoTICI represents the ratio of reperfused pixels (orange area) versus the pre-EVT TDT (white area).
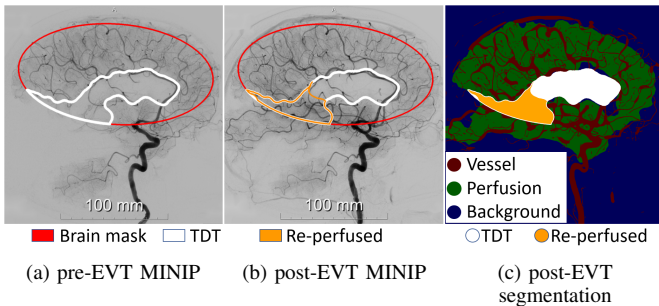

(a) pre-EVT MINIP   (b) post-EVT MINIP   (c) post-EVT segmentation

Fig. 6: An example of TICI quantification. Red: brain mask; white: TDT_{preEVT}; orange: reperfused area ($TDT_{preEVT} \cap P_{postEVT}$).

III. DATA AND ANNOTATION

A. Data Selection

In this work, we used the MR CLEAN Registry [40] dataset for training and evaluating the proposed methods. The MR CLEAN Registry is an on-going multi-center registry which contains all patients with acute ischemic stroke who underwent EVT in the Netherlands since March 2014 [40]. Due to the large variety in acquisition systems and imaging protocols in different centers as well as differences in patient condition, the acquired DSA images possess great variability in image appearance and quality. Therefore, it is necessary to perform data selection in order to obtain a set of qualified yet representative data for our study.

1) Phase Classification Dataset: For training and evaluating the proposed phase classification algorithm, the following selection criteria were applied:

- **Sequence length:** due to non-uniformed data collection procedures across intervention centers, not all DSA acquisitions were collected in the format of a sequence; often only a few snapshots were stored. Moreover, we expect a qualified image sequence to contain most of the contrast flow. Based on our observation, acquisitions with less than six frames are insufficient for our purpose, thus excluded;

- **Image quality:** for the data to be usable, a number of quality criteria must be met: (i) the acquisition must be a cerebral DSA; (ii) the acquisition must possess a sufficiently visible amount of contrast; (iii) the acquisition should not exhibit substantial motion artifacts, such as motion blur, which is not trivial to eliminate via post-processing;

- **Occlusion location:** in this study, we consider patients that had an occlusion of the intracranial internal carotid artery (ICA) or the M1 segment, as these are the most common indications for EVT.

The MR CLEAN Registry (part 1) contains data of 1488 patients, which are collected from March 16, 2014 till June 15, 2016. DSA acquisitions have been stored for 1479 patients, out of which 987 patients had an occlusion of either ICA or M1. After removing all short acquisitions (less than six frames), 872 patients remained. Subsequently, 192 patients were excluded due to lack of good quality image sequences, which led to 680 qualified patients. From each of those, we selected up to four acquisitions (pre-EVT/post-EVT, AP/lateral view) per subject which were of good image quality as described above.

For the purpose of phase classification, we ended up with 680 qualified patients, 1857 acquisitions and 30297 images. Non-subtracted, duplicated, inverted, corrupted frames were removed and finally all acquisitions were re-ordered according to frame acquisition time.

2) TICI Quantification Dataset: To obtain a qualified DSA dataset for TICI quantification experiments and the overall algorithm evaluation, we applied further data selection criteria on top of the selected dataset for phase classification as follows:

- **Reference eTICI score:** in this study, eTICI score was used as the reference standard for evaluating the proposed autoTICI. Therefore, it is required that both pre-EVT and post-EVT acquisitions were scored. In addition, a pre-EVT eTICI score of 2 or higher indicates that the occlusion has been reopened before EVT treatment, thus these images were excluded for further analysis;

- **Four acquisitions per patient:** in order to investigate the sensitivity of autoTICI on both AP and lateral views, pre-EVT and post-EVT DSA acquisitions for both views should be available;

- **Image quality:** for accurate TICI quantification, the acquisition should fulfill the following image quality criteria: (i) TDT of the brain should be fully visible in

TABLE I: eTICI and mRS score distribution on 141 test patients.

Training set: 539 patients, 648 acquisitions, 9440 images						
Test set: 141 patients, 564 acquisitions, 7421 images						
Statistics on test set						
post-EVT eTICI	Number of patients			mRS score	Number of patients	
0	14	(10%)	favorable	0 (no symptoms)	6	(4%)
1	4	(3%)	favorable	1	20	(14%)
2A	24	(17%)	favorable	2	37	(26%)
2B	27	(19%)	unfavorable	3	18	(12%)
2C	22	(16%)	unfavorable	4	12	(8%)
3	50	(35%)	unfavorable	5	6	(4%)
			unfavorable	6 (death)	42	(29%)

the image view; (ii) the view of the acquisition should not be substantially rotated. As our DSA images are 2D projections of the brain, affine transformations are unable to fix 3D rotation artifacts in its 2D projects.

Out of the 680 qualified patients from the previous selection, 34 patients were excluded due to no valid eTICI score. Following it, image quality criteria were applied and patients with four acquisitions available were selected, which leads to 141 patients (564 acquisitions) for TICI quantification and testing of phase classification.

3) Dataset Division: The statistics of selected patient information is summarized in Table I. Based on data quality and availability, a subset of 141 out of 680 patients were suitable for automatic TICI quantification. This subset was fully annotated and used for testing of phase classification, as well as evaluation of the entire pipeline. To ensure good model generalizability, this test set was thus excluded from the training set for phase classification. From the remaining 539 patients, 648 acquisitions (9440 images) were randomly selected (to alleviate manual annotation effort) for training and evaluating the phase classification model. In this way, the test set is maximized and kept unseen from the trained model.

B. Data Annotation

1) TICI Score Reference Standard: In the MR CLEAN registry, the eTICI [11] was visually scored by independent core-lab neuroradiologists. In this work, we compared the proposed autoTICI against eTICI.

2) Modified Rankin Scale (mRS): The mRS score assesses the neurological independence of patients at day 90 following an EVT. It is a seven-grade scale, running from no symptoms (score: 0) to death (score: 6). In this study, we used the mRS score as a reference to evaluate the outcome predictability of the proposed autoTICI. Table I describes the mRS score distribution on the test data set. Note that 20 patients had missing mRS records, which were imputed using multiple imputations by chained equations (MICE) [41]. We dichotomized the mRS scores to indicate favorable and unfavorable treatment outcomes.

3) National Institutes of Health Stroke Scale (NIHSS): NIHSS quantifies the impairment of stroke patients by evaluating 11 aspects. It is an ordinal score ranging from 0 (no

symtoms) to a maximum of 42. In the MR CLEAN registry, both the baseline (prior to EVT, $NIHSS_{BL}$) and follow-up (within 24 hours after EVT, $NIHSS_{FU}$) scores were assessed. In our experiments, we derived the NIHSS shift as follows:

$$NIHSS_{shift} = NIHSS_{BL} - NIHSS_{FU} \quad . \quad (3)$$

4) Sequence Phase Labelling: Manual phase labelling was performed independently by three annotators, one experienced clinician, one colleague researcher and the first author, using an in-house developed tool in MevisLab. According to the phase definition (Section II-A.1), the sequence labelling output is a per frame label sequence, with 0, 1, 2, 3 denoting non-contrast, arterial, parenchymal and venous phase respectively. 648 randomly selected training sequences (Section III-A.1) were randomly split into 3 parts, each annotated by one operator. 141 testing sequences (Section III-A.2) were labelled by all three operators independently, and consensuses were derived afterwards.

5) DSA Atlas: We selected 22 DSA acquisitions (12 AP and 10 lateral views) from 50 patients without stroke. Binary brain masks were delineated on those atlases for registration purposes as described in Section III-A.2.

IV. EXPERIMENTS AND RESULTS

A. Implementation

The proposed methods were implemented in Python. The deep learning model for phase classification was developed using PyTorch [42] on an NVIDIA 2080 Ti with 11 GB of memory. The deep learning network was trained with a batch size of 32 for 100 epochs. All training data were iterated once per epoch. To enrich the diversity of training set and prevent possible overfitting, the following augmentation techniques were randomly applied during data loading: horizontal flip, random rotation ($\in [-10°, 10°]$), random affine transformation (translation $\in [0, 10\%]$ of image width/height, scale $\in [0.8, 1.2]$). For this multi-class classification task, cross entropy was chosen as the loss function. We used the Adam optimizer [43] with an adaptive learning rate initialized at 0.001, halved every 10 epochs.

B. Evaluating Phase Classification

In the evaluation of phase classification, we assessed the added value of data augmentation, temporal information and the handcrafted constrained transition matrix module via an ablation study. Furthermore, we compared the method with inter-observer variability among three human annotators. Additionally, the impact of automatic phase classification accuracy on autoTICI scoring is presented in IV-C.3.

1) Evaluation metrics: The following metrics were adopted in this experiment:

- **average accuracy** represents the percentage of correctly classified images out of all the images;
- **weighted F1 score** refers to the harmonic mean of the precision and recall. In a multi-class problem of Nc classes, the weighted F1 ($F1_w$) is an average of F1 scores

TABLE II: Performance overview of phase classification. Ablation study was performed with 5-fold cross validation on the training and validation set. Performance comparison with human annotator was assessed on the test set.

	Method	Frame level metrics		Sequence level metrics								
		Average accuracy	Weighted F1	first arterial			last arterial			last parenchymal		
				acc	offset: frame(sec*) mean	std	acc	offset: frame(sec*) mean	std	acc	offset: frames(sec*) mean	std
Ablation	Proposed	**0.92**	**0.92**	**0.92**	**0.10**(0.05)	**0.41**(0.16)	**0.51**	**0.70**(0.53)	**1.20**(0.94)	**0.66**	**0.42**(0.39)	**0.84**(1.15)
	No augmentation	0.88	0.88	0.84	0.21(0.09)	0.67(0.25)	0.39	1.00(0.81)	1.56(1.65)	0.60	0.53(0.54)	0.98(1.32)
	No temporal info	0.91	0.91	0.91	0.12(0.06)	0.47(0.17)	0.46	0.76(0.54)	1.23(0.96)	0.66	0.43(0.44)	0.84(1.04)
	No state-transition matrix	0.91	0.91	0.91	0.14(0.06)	0.55(0.18)	0.51	0.89(0.55)	3.21(1.12)	0.66	0.49(0.42)	1.41(1.26)
Test	Method to human agreement	0.90	0.90	0.88	0.14(0.05)	0.48(0.14)	0.34	1.10(0.77)	1.54(1.56)	0.55	0.63(0.77)	1.11(1.31)
				>1 frame(sec)	1.4%(0.4%)		24.8%(21.3%)			10.9%(23.2%)		
	Human to human agreement	0.89	0.89	0.87	0.18(0.06)	0.59(0.17)	0.31	1.16(0.80)	1.56(1.58)	0.52	0.68(0.89)	1.10(1.47)
				>1 frame(sec)	1.9%(1.1%)		25.8%(21.0%)			12.3%(31.1%)		

*As the frame time increments of 103 (out of 648) DSA acquisitions are missing, the time offsets are the averages of the ones with frame time increments.
*The frame rate is inconstant in a DSA acquisition, it typically starts at 2-4 fps in arterial phase and gradually reduces to 0.5-1 fps in venous phase.

$(F1_i)$ weighted by the number of images (S_i) from each class, as in Eq. 4:

$$F1_w = \frac{\sum_{i=1}^{Nc} S_i F1_i}{\sum_{i=1}^{Nc} S_i} \quad ; \quad (4)$$

- **frame offset** measures the phase classification accuracy on sequence level, rather than image level. The phase border frame index offset between classified and ground truth can serve as an insightful metric for phase classification accuracy. In this experiment, the average absolute offset and standard deviation of offset for first, last arterial phase frame and last parenchymal phase frame were evaluated.

2) Ablation Study: Ablation experiments were performed to assess the contribution of separate components of the proposed algorithm. Table II summarizes the comparison under the aforementioned metrics. In all ablated cases, the model performance degraded. Without data augmentation, the generalizability of the model was reduced as a result of overfitting on the training data. Without incorporating neighboring frames as temporal information, the overall image level accuracy slightly dropped. The most marginal deterioration was observed at arterial/parenchymal border frame accuracy (51% vs 46%). Not surprisingly, by embedding human defined logic into the model, an overall performance improvement was evidenced across all metrics in the Table. More importantly, such a constrained transform matrix enforces phase transition logic within a sequence, suppressing unreasonable prediction errors.

3) Inter-observer Variability: Smooth transition between phases is often observed, which attributes to inter-observer variability during phase annotation. In this section, the algorithm variability is compared with three human annotators on the selected 141 acquisitions in test set. We computed the agreements in pairs for all combinations among the proposed method and three human annotators. For the agreement between the method and human annotators is defined as the average agreement between the method and each of the annotators. The inter-observer agreement is calculated as the average agreement among all human annotator pairs. As shown in Table II, the proposed method exhibits even better agreement than the inter-observer agreement across all metrics, achieving

TABLE III: Quantitative autoTICI reperfusion statistics.

eTICI	Mean area (pixels) TDT_preEVT	Mean area (pixels) TDT_postEVT	Reperfusion ratio mean	std
0	1.43×10^5	1.24×10^5	0.19	0.23
1	1.30×10^5	0.92×10^5	0.22	0.14
2A	1.49×10^5	0.70×10^5	0.57	0.25
2B	1.72×10^5	0.64×10^5	0.71	0.23
2C	1.31×10^5	0.26×10^5	0.83	0.16
3	1.29×10^5	0.28×10^5	0.82	0.18

TDT_preEVT: non-perfused pixels before EVT; TDT_postEVT: non-reperfused pixels after EVT; reperfusion ratio is based on Eq. 2

human level annotation precision on phase classification.

C. Evaluating autoTICI Quantification

On the test set (Table I), the relevance of autoTICI as a brain tissue reperfusion measure was evaluated based on the correlation with eTICI. Furthermore, we assessed the clinical value of autoTICI in comparison to eTICI with respect to mRS and NIHSS. Whereas the distribution of patients with respect to eTICI and mRS is imbalanced, we did not attempt to balance the distribution during data selection as this represents realistic patient statistics in clinical practice.

1) autoTICI vs eTICI: Fig. 7a shows the positive correlation between the proposed autoTICI and eTICI on the test dataset. According to eTICI definition, minimal tissue perfusion difference is expected between eTICI 0 and 1 or between eTICI 2C and 3. Therefore, eTICI 0/1 and eTICI 2C/3 are plotted to better visualize the correlation in brain tissue reperfusion. More quantitatively, Table III shows the reperfusion statistics of autoTICI versus eTICI grades. The heatmap in Fig. 7b illustrates the distribution of autoTICI over eTICI grades. Furthermore, spearman correlation test also showed that both AP and lateral view autoTICI scores were significantly associated with eTICI with $\rho = 0.54$, (P < 0.001) and $\rho = 0.65$, (P < 0.001), respectively. The stronger correlation with lateral view autoTICI scores matches the expectation of clinical experts because the brain area in AP views is about half as in lateral projections and occlusion effects are generally better visualized in lateral view. Reasons for the few outliers shown in Fig. 7a and 7b include questionable eTICI annotation (recanalization without reperfusion), insufficient contrast and suboptimal TDT registration.
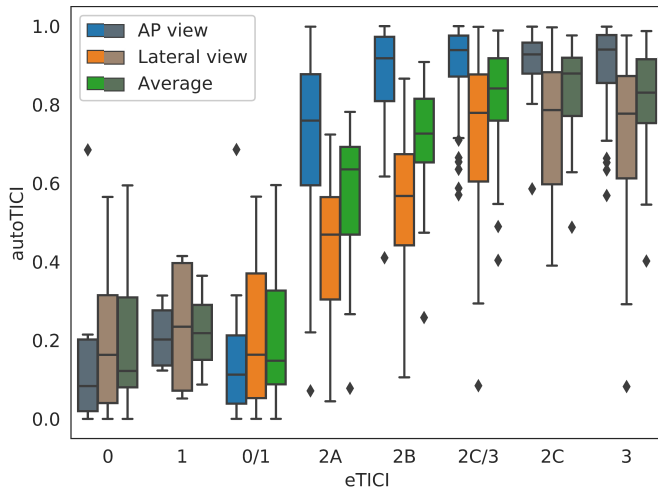
We performed multinomial logistic regression between autoTICI and eTICI on the test data with leave-one-out cross validation. As shown in Fig.7c, a micro-average (biased average by class frequency) AUC of 0.76 and 0.78 for AP and lateral view were achieved respectively. With both autoTICI scores as input features, the micro-average AUC reached 0.81. For dichotomized eTICI (failure: $\leq 2A$; success: $\geq 2B$), autoTICI achieved an AUC of 0.90.

2) Correlation to Treatment Outcome: We further compared autoTICI and eTICI with respect to their correlation to treatment outcome in terms of mRS (III-B.2) and NIHSS (III-B.3). We used both AP and lateral autoTICI scores as input features for logistic regression. As shown in Fig. 8, cross validated logistic regression between TICI scores and mRS showed that autoTICI had a slightly higher AUC of 0.63 than 0.60 for eTICI, though the difference was not statistically significant (P = 0.52). Similarly, the accuracy of autoTICI and eTICI were 0.66 and 0.62, respectively. Spearman correlation test was applied between TICI scores and NIHSS shift scores,
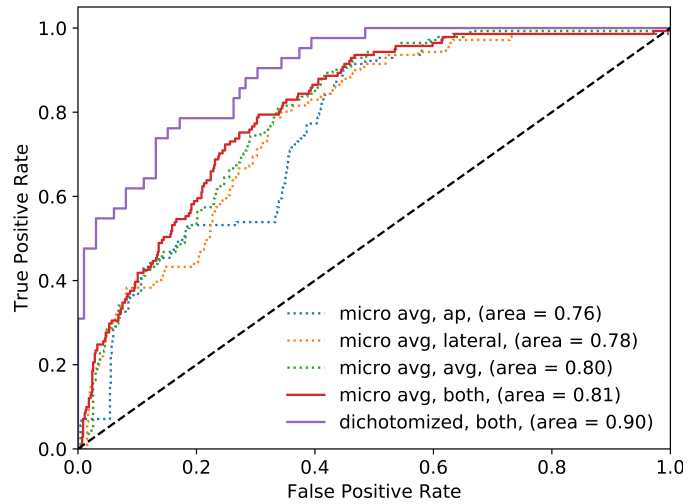
(c) Logistic regression ROC. Green, orange, blue, red: micro-average ROC of all eTICI grades; purple: dichotomized eTICI (failure: $\leq 2A$; success: $\geq 2B$).

Fig. 7: Correlation between autoTICI and eTICI.

which showed that autoTICI and eTICI were comparable with $\rho = 0.29$ (P < 0.01) and $\rho = 0.30$ (P < 0.01), respectively.
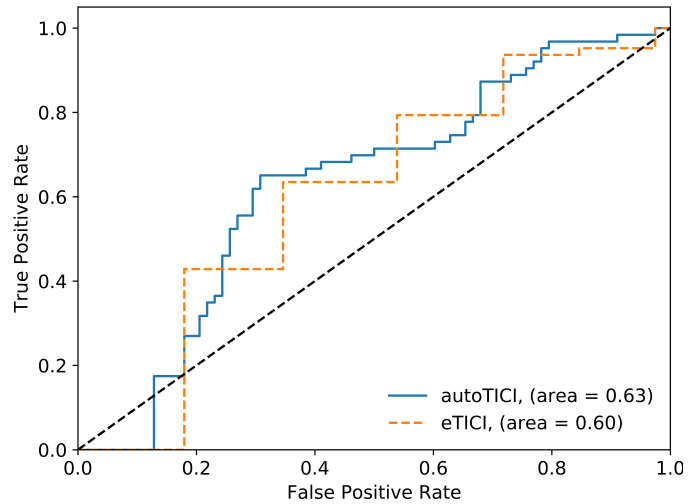
Fig. 8: Prediction of treatment outcome (mRS) using autoTICI.

Overall, both Fig. 8 and Table IV have demonstrated that autoTICI and eTICI possess comparable but limited capability in outcome prediction. It is also understandable that the relation between NIHSS (short term, at 24 hours after EVT) and mRS (long term, at 90 days after EVT) scores is not deterministic. Multiple recent studies [4], [40], [44] have shown that reperfusion scores play an irreplaceable role in outcome prediction, the long-term functional outcome is, however, multifactorial.

3) Impact of Phase Classification on autoTICI: As the correlation between autoTICI and eTICI has been established in Section IV-C.1, this correlation can be further employed to assess the value of accurate phase classification. We introduced offsets to the phase classification results and observed how autoTICI is impacted in terms of its correlation with eTICI. As shown in Table V, adding offsets of one or two frames

(a) autoTICI distribution with respect to eTICI. Blue: AP score; orange: lateral score; green: average of AP and lateral scores.
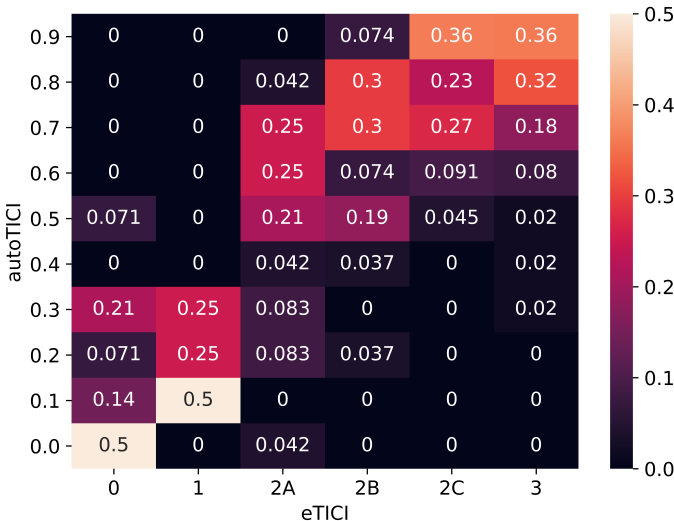
(b) Distribution heatmap of autoTICI over eTICI. Here, autoTICI (y-axis) refers to the average of AP and lateral scores binned in 0.1 intervals. The value in each cell is normalized per column.

TABLE IV: autoTICI vs eTICI in terms of correlation to outcome.

Method	logistic regression dichotomized mRS			spearman correlation			
	acc	AUC	odds ratio	ρ	p-value	ρ	p-value
				NIHSS_FU		NIHSS_shift	
autoTICI	0.66	0.63	1.27	-0.25	<0.01	0.29	<0.01
eTICI	0.62	0.60	1.02	-0.28	<0.01	0.30	<0.01

in classification results lead to noticeable deterioration in autoTICI statistics.

TABLE V: Correlation between autoTICI and eTICI correlation with respect to phase classification accuracy

Offset	Pearson correlation AP/Lateral	Logistic regression AUC	
		Micro-average both views	Dichotomized eTICI
PC	0.54/0.65	0.81	0.90
PC$\pm$1 frame	0.54/0.59	0.79	0.88
PC$\pm$2 frames	0.46/0.51	0.76	0.82

PC refers to the predicted results of the phase classification method.

V. DISCUSSION

We have presented a fully automatic, quantitative and truly perfusion based TICI scoring method that is inspired by traditional visual TICI assessment procedures. We exploited convolutional neural networks to tackle the phase classification challenge; it achieved performance on par with human experts. On the MR CLEAN registry, we demonstrated that the proposed autoTICI and eTICI are statistically significantly associated and they are both comparable predictors of functional outcome, revealing the potential value of autoTICI as a computer-aided biomarker for peri-procedural treatment assessment.

We have presented a deep learning based approach for DSA phase classification with human level precision. Apart from autoTICI, the proposed phase classification method can be generalized for many automated DSA analysis tasks, such as key frames selection, inter-acquisition temporal synchronization or slow arterial flow detection. The proposed network uses ResNet with a customized CRF. Although recurrent neural networks, such as LSTM [45] or GRU [46] can be alternatives in this task, we found that the proposed CRF handles inter-frame relations well in an explainable way, especially with limited number of labelled sequences (648 variable-length sequences in our case). When considering prior-art on phase classification, only Schuldhaus *et al*. [32] reported a quantitative performance results. On 14 DSA acquisitions, Schuldhaus *et al*. [32] reported 93% and 50% accuracy on first arterial border match and last arterial border match respectively. Accordingly, our proposed method achieved 92% and 51% respectively over 648 acquisitions. A pure numerical comparison seems to show comparable performance, but our study has different characteristics. It was tested on 648 instead of 14 acquisitions, these acquisitions originate from a multi-center registry, and the frame rates (0.5-4 fps) are higher than those (0.5-2 fps) reported by Schuldhaus *et al*. [32], and our approach also separates the venous phase.

The MR CLEAN Registry is an observational multi-center multi-year registry. This registry reflects daily practice in a wide variety of hospitals which on the one hand leads to the heterogeneity of data but on the other hand allows for broad translatability. It can be noticed that 141 out of 1488 patients end up suitable for autoTICI quantification. This is due to a large heterogeneity on image quality, annotation, acquisition process, and storage format as described in the data selection procedure, rather than algorithmic limitations. Main reasons for data exclusion in this study include missing DSA acquisitions, corrupted, incomplete, or too short acquisitions, bad image quality, occlusion location and image not showing the entire TDT area.

Affine registration was performed for sequence motion correction and atlas registration. In the MR CLEAN Registry, one of the limitations is that the original unsubtracted images were mostly discarded, leaving only the subtracted ones. In this case, motion correction does help in mitigating blur effects, it is nevertheless incapable of handling the subtraction artefacts. Besides, due to lacking of background texture, the atlas to patient registration relies on the skull outline and brain vasculature. The skull skeleton shown in a subtracted image is in fact the subtraction artifacts. Therefore, if the original non-subtracted images could be used instead, which encompass richer and more robust textural information, the registration robustness and accuracy could be potentially improved.

Aside from Otsu's thresholding, several alternatives for perfusion segmentation could be considered. One could quantify changes in vessels and tissue perfusion by directly subtracting the MINIP of parenchymal phase by the MINIP of arterial phase. This is however problematic, mainly due to inter-acquisition variations on contrast dilution and inject volume. It remains an interesting research topic on how to obtain contrast profile independent DSA perfusion parameters. The inter-acquisition variations could be alleviated if mechanical pumps are utilized during contrast injection with a fixed injection protocol, which however is not the case for MR CLEAN registry. Another alternative is to use the background statistics to derive a fixed threshold for TDT selection. We opted for image specific Otsu's thresholding instead of fixed thresholding for all patients. Other promising alternatives include deep learning based segmentation methods, such as U-Net [47], which is not studied in this work due to the massive amount of annotations required.

eTICI is one of the most comparable metrics to autoTICI as both are proposed as brain perfusion measures. However, it should be pointed out that the definition of autoTICI and eTICI are not fully identical. While autoTICI and eTICI focus on brain tissue antegrade reperfusion quantification, eTICI also considers evidence of contrast material penetration in vessels past initial occlusion and slow flow in distal. Both eTICI 0 and 1 define minimal reperfusion, they are distinguished based on whether contrast material has passed the initial occlusion. Analogically, in case of (nearly) complete perfusion, eTICI emphasizes the existence of slow flow in distal vessels in score 2C versus 3. Therefore, perfect correlation between autoTICI and eTICI cannot be achieved and is not the aim of autoTICI. In this work, a good correlation between the two (Fig. 7) serves as a proxy for demonstrating the relevance of autoTICI. The outcome prediction capability of autoTICI reported in

Section IV-C.2 further consolidates the conclusion and reveals the potential clinical value of autoTICI.

End-to-end deep learning based methods might be valid alternatives to the automatic TICI scoring method proposed in this study. The reasons that we opted for a step-by-step strategy are (i) data: end-to-end training approaches treat all four acquisitions of each patient as one input data sample. Such methods generally require larger number of patients, as well as more sophisticated GPU resources; (ii) interpretability: while aiming for minimizing prediction errors, end-to-end training approaches generally sacrifices causal interpretability. The proposed method, by contrast, provides intermediate visualizable outputs per step, offering enriched clinical insights; (iii) quantitativity: rather than categorical TICI grading, we seek quantitative brain reperfusion analysis methods. Training end-to-end networks for this purpose is not straightforward due to lacking of quantitative ground truth. Nevertheless, both phase classification and perfusion segmentation can benefit from deep learning methods in this work.

From a clinical perspective, the proposed autoTICI provides an objective, reproducible and quantitative measure of EVT quality, eliminating human errors and variations. As it is a true measure of tissue level reperfusion, human conceptual confusion is avoided. Therefore, autoTICI overcomes the aforementioned three shortcomings of existing TICI scores. Moreover, the extent and location of non-reperfused areas are real-time visualized, helping the operator to determine whether additional attempts for clot removal should be undertaken or additional drugs should be given to improve functional outcome.

VI. CONCLUSION

We have presented a robust and fully automatic perfusion quantification method, autoTICI. On a large routinely acquired multi-center dataset, we have demonstrated that autoTICI is significantly correlated with the eTICI reference with a dichotomized AUC of 0.90 and possesses comparable treatment outcome predictive capability with an AUC of 0.63 versus 0.60, revealing its potential in future studies and clinical practice.

ACKNOWLEDGMENT

REFERENCES

[1] G. WHO, "Global health estimates 2016: deaths by cause, age, sex, by country and by region, 2000–2016," 2018.

[2] R. T. Higashida and A. J. Furlan, "Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke," *Stroke*, vol. 34, no. 8, pp. e109–e137, 2003.

[3] M. Goyal, B. K. Menon, W. H. van Zwam, D. W. Dippel, P. J. Mitchell, A. M. Demchuk, A. Dávalos, C. B. Majoie, A. Van Der Lugt, M. A. De Miquel *et al.*, "Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials," *The Lancet*, vol. 387, no. 10029, pp. 1723–1731, 2016.

[4] O. A. Berkhemer, P. S. Fransen, D. Beumer, L. A. Van Den Berg, H. F. Lingsma, A. J. Yoo, W. J. Schonewille, J. A. Vos, P. J. Nederkoorn, M. J. Wermer *et al.*, "A randomized trial of intraarterial treatment for acute ischemic stroke," *n Engl J Med*, vol. 372, pp. 11–20, 2015.

[5] D. C. Haussen, S. Dehkharghani, S. Rangaraju, L. C. Rebello, M. Bouslama, J. A. Grossberg, A. Anderson, S. Belagaje, M. Frankel, and R. G. Nogueira, "Automated ct perfusion ischemic core volume and noncontrast ct aspects (alberta stroke program early ct score) correlation and clinical outcome prediction in large vessel stroke," *Stroke*, vol. 47, no. 9, pp. 2318–2322, 2016.

[6] E. Venema, M. J. Mulder, B. Roozenbeek, J. P. Broderick, S. D. Yeatts, P. Khatri, O. A. Berkhemer, B. J. Emmer, Y. B. Roos, C. B. Majoie *et al.*, "Selection of patients for intra-arterial treatment for acute ischaemic stroke: development and validation of a clinical decision tool in two randomised trials," *bmj*, vol. 357, 2017.

[7] C. Dargazanli, A. Consoli, M. Barral, J. Labreuche, H. Redjem, G. Ciccio, S. Smajda, J.-P. Desilles, G. Taylor, C. Preda *et al.*, "Impact of modified tici 3 versus modified tici 2b reperfusion score to predict good outcome following endovascular therapy," *American Journal of Neuroradiology*, vol. 38, no. 1, pp. 90–96, 2017.

[8] J. Su, L. Wolff, A. C. M. van Es, W. van Zwam, C. Majoie, D. W. Dippel, A. van der Lugt, W. J. Niessen, and T. Van Walsum, "Automatic collateral scoring from 3d cta images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2190–2200, 2020.

[9] D. Robben, A. M. Boers, H. A. Marquering, L. L. Langezaal, Y. B. Roos, R. J. van Oostenbrugge, W. H. van Zwam, D. W. Dippel, C. B. Majoie, A. van der Lugt *et al.*, "Prediction of final infarct volume from native ct perfusion and treatment parameters using deep learning," *Medical image analysis*, vol. 59, p. 101589, 2020.

[10] I. I. T. Investigators, "The interventional management of stroke (ims) ii study," *Stroke*, vol. 38, no. 7, pp. 2127–2135, 2007.

[11] M. A. Almekhlafi, S. Mishra, J. A. Desai, V. Nambiar, O. Volny, A. Goel, M. Eesa, A. M. Demchuk, B. K. Menon, and M. Goyal, "Not all "successful" angiographic reperfusion patients are an equal validation of a modified tici scoring system," *Interventional Neuroradiology*, vol. 20, no. 1, pp. 21–27, 2014.

[12] D. S. Liebeskind, S. Bracard, F. Guillemin, R. Jahan, T. G. Jovin, C. B. Majoie, P. J. Mitchell, A. van der Lugt, B. K. Menon, L. San Román *et al.*, "etici reperfusion: defining success in endovascular stroke therapy," *Journal of neurointerventional surgery*, vol. 11, no. 5, pp. 433–438, 2019.

[13] O. O. Zaidat, A. J. Yoo, P. Khatri, T. A. Tomsick, R. Von Kummer, J. L. Saver, M. P. Marks, S. Prabhakaran, D. F. Kallmes, B.-F. M. Fitzsimmons *et al.*, "Recommendations on angiographic revascularization grading standards for acute ischemic stroke: a consensus statement," *Stroke*, vol. 44, no. 9, pp. 2650–2663, 2013.

[14] G. Zhang, K. M. Treurniet, I. G. Jansen, B. J. Emmer, R. van den Berg, H. A. Marquering, M. Uyttenboogaart, S. F. Jenniskens, Y. B. Roos, P. J. van Doormaal *et al.*, "Operator versus core lab adjudication of reperfusion after endovascular treatment of acute ischemic stroke," *Stroke*, vol. 49, no. 10, pp. 2376–2382, 2018.

[15] M. G. Lansberg, M. Straka, S. Kemp, M. Mlynash, L. R. Wechsler, T. G. Jovin, M. J. Wilder, H. L. Lutsep, T. J. Czartoski, R. A. Bernstein *et al.*, "Mri profile and response to endovascular reperfusion after stroke (defuse 2): a prospective cohort study," *The Lancet Neurology*, vol. 11, no. 10, pp. 860–867, 2012.

[16] P. Khatri, J. Neff, J. P. Broderick, J. C. Khoury, J. Carrozzella, and T. Tomsick, "Revascularization end points in stroke interventional trials: recanalization versus reperfusion in ims-i," *Stroke*, vol. 36, no. 11, pp. 2400–2403, 2005.

[17] T. Tomsick, J. Broderick, J. Carrozella, P. Khatri, M. Hill, Y. Palesch, J. Khoury *et al.*, "Revascularization results in the interventional management of stroke ii trial," *American Journal of Neuroradiology*, vol. 29, no. 3, pp. 582–587, 2008.

[18] A. J. Yoo, C. Z. Simonsen, S. Prabhakaran, Z. A. Chaudhry, M. A. Issa, J. E. Fugate, I. Linfante, D. S. Liebeskind, P. Khatri, T. G. Jovin *et al.*, "Refining angiographic biomarkers of revascularization: improving outcome prediction after intra-arterial therapy," *Stroke*, vol. 44, no. 9, pp. 2509–2512, 2013.

[19] S. H. Suh, H. J. Cloft, J. E. Fugate, A. A. Rabinstein, D. S. Liebeskind, and D. F. Kallmes, "Clarifying differences among thrombolysis in cerebral infarction scale variants: is the artery half open or half closed?" *Stroke*, vol. 44, no. 4, pp. 1166–1168, 2013.

[20] A. Nielsen, M. B. Hansen, A. Tietze, and K. Mouridsen, "Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning," *Stroke*, vol. 49, no. 6, pp. 1394–1401, 2018.

[21] R. McKinley, L. Häni, J. Gralla, M. El-Koussy, S. Bauer, M. Arnold, U. Fischer, S. Jung, K. Mattmann, M. Reyes *et al.*, "Fully automated stroke tissue estimation using random forest classifiers (faster)," *Journal of Cerebral Blood Flow & Metabolism*, vol. 37, pp. 2728–2741, 2017.

[22] A. Liebeskind, A. Deshpande, J. Murakami, and F. Scalzo, "Automatic estimation of arterial input function in digital subtraction angiography," in *International Symposium on Visual Computing*. Springer, 2019, pp. 393–402.

[23] H. Prasetya, L. A. Ramos, T. Epema, K. M. Treurniet, B. J. Emmer, I. R. Van Den Wijngaard, G. Zhang, M. Kappelhof, O. A. Berkhemer, A. J. Yoo *et al.*, "qtici: Quantitative assessment of brain tissue reperfusion on digital subtraction angiograms of acute ischemic stroke patients," *International Journal of Stroke*, p. 1747493020909632, 2020.

[24] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

[25] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4694–4702.

[26] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.

[27] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler, "Convolutional learning of spatio-temporal features," in *European conference on computer vision*. Springer, 2010, pp. 140–153.

[28] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489–4497.

[29] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *Advances in neural information processing systems*, vol. 27, pp. 568–576, 2014.

[30] J. Carreira and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299–6308.

[31] H.-J. Lee, J.-S. Hong, C.-J. Lin, Y.-H. Kao, F.-C. Chang, C.-B. Luo, and W.-F. Chu, "Automatic flow analysis of digital subtraction angiography using independent component analysis in patients with carotid stenosis," *PloS one*, vol. 12, no. 9, p. e0185330, 2017.

[32] D. Schuldhaus, M. Spiegel, T. Redel, M. Polyanskaya, T. Struffert, J. Hornegger, and A. Doerfler, "Classification-based summation of cerebral digital subtraction angiography series for image post-processing algorithms," *Physics in Medicine & Biology*, vol. 56, p. 1791, 2011.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[34] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML '01, 2001, pp. 282–289.

[35] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "Pet-ct image registration in the chest using free-form deformations," *IEEE transactions on medical imaging*, vol. 22, no. 1, pp. 120–128, 2003.

[36] S. Klein, J. P. Pluim, M. Staring, and M. A. Viergever, "Adaptive stochastic gradient descent optimisation for image registration," *International journal of computer vision*, vol. 81, no. 3, p. 227, 2009.

[37] K. Marstal, F. Berendsen, M. Staring, and S. Klein, "Simpleelastix: A user-friendly, multi-lingual library for medical image registration," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 134–142.

[38] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 1998, pp. 130–137.

[39] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[40] I. G. Jansen, M. J. Mulder, and R.-J. B. Goldhoorn, "Endovascular treatment for acute ischaemic stroke in routine clinical practice: prospective, observational cohort study (mr clean registry)," *bmj*, vol. 360, 2018.

[41] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: what is it and how does it work?" *International journal of methods in psychiatric research*, vol. 20, no. 1, pp. 40–49, 2011.

[42] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in neural information processing systems*, 2019, pp. 8026–8037.

[43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[44] O. O. Zaidat, H. Bozorgchami, M. Ribó, J. L. Saver, H. P. Mattle, R. Chapot, A. P. Narata, O. Francois, A. P. Jadhav, J. A. Grossberg *et al.*, "Primary results of the multicenter arise ii study (analysis of revascularization in ischemic stroke with embotrap)," *Stroke*, vol. 49, no. 5, pp. 1107–1115, 2018.

[45] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[46] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[47] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.