Development and Assessment of Learning-based Vessel Biomarkers from CTA in Ischemic Stroke

Jiahang Su

Development and Assessment of Learning-based Vessel Biomarkers from CTA in Ischemic Stroke

Ontwikkeling en evaluatie van op leren gebaseerde biomarkers van bloedvaten uit CT angiografie beelden voor ischemische beroerte

THESIS

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Wednesday 25 October 2023 at 15.30 hrs

by

Jiahang Su
born in Linyi, Shandong, China

Erasmus University Rotterdam

Doctoral Committee

Promotors	Dr. ir. T. van Walsum
	Prof. dr. A. van der Lugt
	Prof. dr. W. J. Niessen
Other members	Prof. dr. H. Lingsma
	Prof. dr. I. Išgum
	Dr. I. Smal

**Little drops of water, Little grains of sand,
Make the mighty ocean, And the beauteous land.
–T. J. Carney**

Contents

Chapter 1
Introduction

1

1.1   Stroke

Stroke is a medical condition which occurs when there is malfunction in the cerebral blood vessel network. The cerebral blood vessels supply the brain tissue with oxygen and nutrients. The brain is a complex organ that buffers, regulates and controls the body. The energy consumption of the brain itself accounts for 20% of total calorie use in general [1]. Malfunction of the cerebral blood vessel network can lead to the death of brain tissue. Depending on the affected region, dead brain tissue can result in neurological deficits like speech, motor or cognition impairment. Stroke is a common disease in the elderly population and the second cause of death and third cause of disability worldwide [2]. Moreover, the incidence rate of stroke increases yearly [2].

Depending on the cause of stroke, there are two main types of stroke: ischemic stroke and hemorrhagic stroke. Ischemic stroke is caused by an occlusion of arterial vessels. In this case, the blood flow is suppressed. and the affected region is the brain region that is supplied by the occluded vessels. This is also the most common stroke type, accounting for approximately 87% of the strokes of the total stroke population [3]. Hemorrhagic stroke is caused by bleeding of a blood vessel. This lead to the blood entering the surrounding tissue.Thus the affected region is the blood flooded brain region and the surrounding brain tissue. In my work, I focused on intracranial ischemic stroke, in which an intracranial vessel is occluded.

1.2   Acute Ischemic Stroke

Ischemic stroke is caused by a blood clot blocking an arterial vessel. Ischemic stroke can be classified according to the location of the occluded intracranial arteries. The intracranial arterial vessels are subdivided in an anterior circulation and a posterior circulation. The anterior circulation consists of the middle cerebral artery (MCA) tree and anterior cerebral artery (ACA) tree on the left and right side. The posterior circulation consists of the posterior cerebral artery (PCA) tree and the cerebellar arteries on the left and right side. An occlusion occurring in one of the the proximal segments, such as the first branch of an arterial tree, is a so-called large vessel occlusion (LVO). Approximately 30% of patients with an acute ischemic stroke has an LVO, and most of the LVO's occur in the anterior circulation [4]. Fig.1.1 shows the MCA and ACA trees highlighted with different color. When the occlusion occurs beyond the proximal segments of artery trees, it is called a distal occlusion. In current practice, patients with a more distal occlusion are often not treated yet by endovascular treatment.

1.3   Imaging in Ischemic Stroke Patient Diagnosis

Patient suspected of having a stroke generally undergo a standard imaging protocol for further examination and diagnosis. In the Netherlands, the standard imaging protocol consist of a series of computed tomography (CT) based acquisitions. CT is often preferred over MRI for logistic reasons (cheaper and faster). First, a non-contrast computed tomography (NCCT) image is acquired, to determine whether a stroke is ischemic or hemorrhagic, as the clinical symptoms of hemorrhagic and ischemic stroke
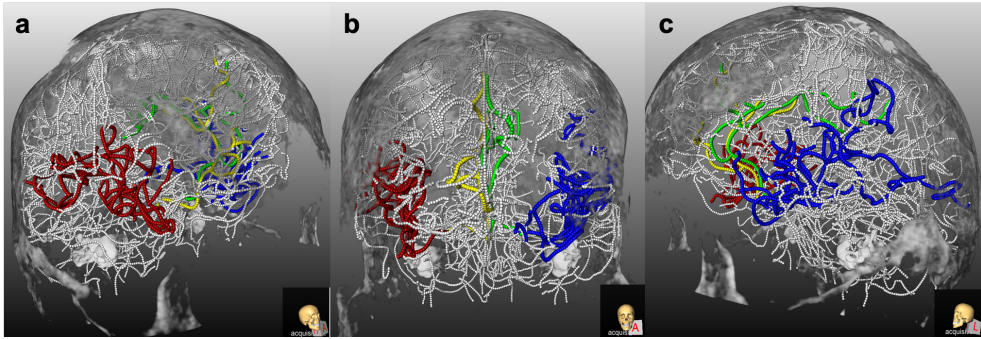
Figure 1.1: *Example of brain vessel annotation with colored anterior vessel trees for a subject with a large vessel occlusion in the left M2 segment. White represents veins and posterior cerebral artery trees, red is the right MCA tree, yellow is the right ACA tree, green is the left ACA tree, blue is the left MCA tree, a: the right sagittal view; b: the coronal view; c: the left sagittal view.*

can be similar; NCCT is currently the gold standard for hemorrhagic stroke diagnosis due to its high sensitivity [5]. Subsequently, in case of an ischemic stroke, computed tomography angiography (CTA) and computed tomography perfusion (CTP) images are acquired to demonstrate a vessel occlusion and its location, to assess the extent of collateral circulation, and to assess the extent of the brain perfusion abnormalities. Based on the imaging and clinical parameters (such as patient age, time since the stroke onset) a decision regarding the type of treatment is made.

In clinical practice, CTA images are acquired by injecting contrast material intravenously in the arm. The flow of the contrast agent is a dynamic process. Upon entering the brain, the contrast agent first flows through the arteries, then to the brain parenchyma, and subsequently leaves the brain via the veins. This implies that the visual appearance of a CTA image depends on the timing of the acquisition. For CTP images, many acquisitions are taken during the flow of contrast material through the brain. The contrast flow timing, in general, is divided over five phases [6] based on the Hounsfield unit (HU) of ICA-top and confluence sinus. The five phases include early arterial phase, peak arterial phase, equilibrium phase, peal venous phase, and late venous phase. Fig.1.2 shows a CTP image at each of these five contrast phases in the same subject. This shows that the CTA image appearance depends on the timing of the image acquisition.

1.4 Treatment Options for Acute Ischemic Stoke

The treatment options for AIS patients are intravenous tissue plasminogen activator (IV-tPA) for all patients with ischemic stroke who present within 4,5 hours after onset and endovascular treatment (EVT) for patients with a large vessel occlusion [7, 8, 9]. IV-tPA is a thrombolytic drug introduced into the veins to dissolve the blood clot. EVT

1
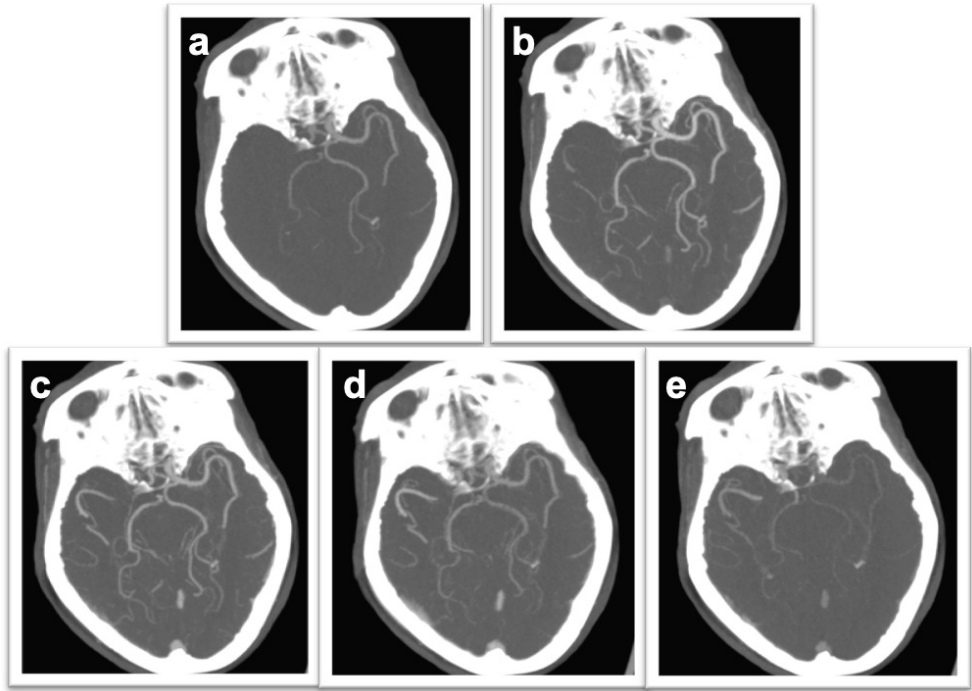
Figure 1.2: Example of CTP image in different contrast acquisition phase. a: the early arterial phase; b: the peak arterial phase; c:the equilibrium phase; d:the peak venous phase; e: the late venous phase.

is a minimally invasive procedure where microcatheters, inserted from the groin or the arm, are used to mechanically remove the blood clot and/or deliver the thrombolytic agent directly to the clot. Since several recent randomized controlled trials [10, 11, 12, 13, 14] demonstrated that EVT is an effective treatment option for patients with AIS and an LVO, EVT has become a standard procedure for patients with an AIS. However, not all patients benefit from the treatment. And as the treatment is not without risks, patient selection, though challenging, remains relevant. A recent treatment outcome prediction tool, MR PREDICT [15], used semi - automatic imaging biomarkers (visual scoring based) in combination with clinical parameters to determine the probability on good treatment outcome for an individual patient.

1.5   Imaging Biomarkers for Treatment Decision Making in Patients with Acute Ischemic Stoke

Image features that can be used for diagnostic or therapeutic purposes, are so-called imaging biomarkers [16]. In clinical practice, imaging biomarkers are primarily ex-

tracted visually and/or manually by radiologists. Empowered by the development of machine learning methods, including deep learning, automated quantification of imaging biomarkers has become trending in radiological research and clinical practice.

For the diagnosis and workup of patients with AIS, several baseline image visual scoring systems have been developed to provide semi-quantitative measures that may help in determining whether IVT and/or EVT would be beneficial for a specific patient. Two of the most used image based parameters in clinical practice are the Alberta stroke program early CT score (ASPECTS) and Collateral score (CS)[17, 18]. ASPECTS [19] is a measure to quantify the severity of brain tissue damage in NCCT. CS [20] is a measure to quantify the vessel status in CTA images. Once a main arterial tree is occluded, blood may still arrive in the affected region via alternative vessels. Those vessels are so called collateral vessels. In clinical practice, collaterals are scored visually, by comparing the amount of vessels in the occluded side with the contralateral side. The most used collateral scale in Netherlands is the four point scale proposed by Tan et al. [20]. Fig.1.3 shows this four point scale. Visual scoring methods may have large intraobserver variation [21]. In this case, an automatic biomarker modelling method might have better consistency in clinical practice.
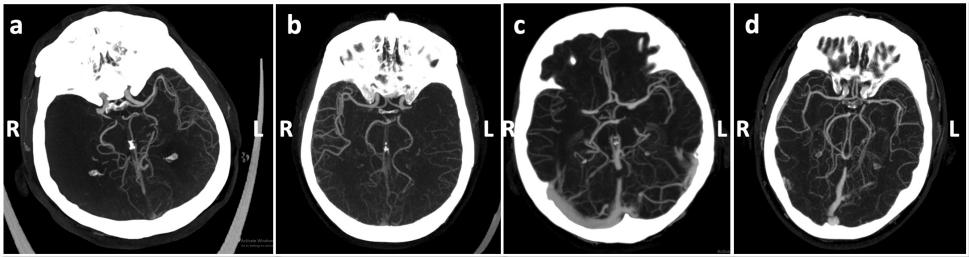
Figure 1.3: *Examples of four collateral score in Maximum Intensity Projection (MIP) from 3D CTA image acquired in four different patients. a: collateral score 0; b: collateral score 1; c: collateral score 2; d: collateral score 3.*

1.6 Goals & outline

The primary goal of this work is to develop and evaluate methods to extract quantitative imaging biomarkers of the cerebral vasculature from CTA images to assist therapeutic decision making in patients with AIS. The underlining research question is, to what extent learning based methods can be used in extracting vessel based biomarkers and beyond. Therefore, Chapter 2 till Chapter 4 investigated collateral scoring related matters. **??** and Chapter 5 focus on novel methods for extracting relevant image features via learning based methods.

Chapter 2 investigates a three step approach to modelling collateral score biomarker based on mincing how a human does the collateral scoring: by comparing the vessel structures of MCA territories in occluded versus contralateral side.

1

?? investigates, to what extent, an end-to-end approach performs well on the collateral score classification task. The end to end approach is also inherited to the human visual scoring by comparing the occluded hemisphere versus the non-occluded hemisphere.

Chapter 3 assesses the performance of the automatic collateral scoring method developed in Chapter 2 with carefully designed experiments in a random sampled subset of a randomized control trial.

Chapter 4 uses the automatic collateral scoring method introduced in Chapter 2 to investigate the optimal contrast acquisition time point for CTA images by computing the CS over all CTP images and associate this with the baseline clinical parameter.

?? investigates whether the general binary tubular structure output from segmentation task can be further improved using a post processing network.

Chapter 5 investigates first of all, to what extend the DRL based methods can be used in the cerebral vessel tracking; secondly, how well the CNN based bifurcation detection method performed in cerebral vessel network; at the last, to what extent, the anterior tree extraction methods works with the previous two models combined.

Chapter 6 provides a general discussion of the results from this thesis and presents possible future research directions.

Chapter 2
Automatic Collateral Scoring Using Vessel Features

2

Abstract

The collateral score is an important biomarker in decision making for endovascular treatment (EVT) of patients with ischemic stroke. The existing collateral grading systems are based on visual inspection and prone to subjective interpretation and interobserver variation. The purpose of our work is the development of an automatic collateral scoring method. In this work, we present a method that is inspired by human collateral scoring. Firstly, we define an anatomical region by atlas-based registration and extract vessel structures using a deep convolutional neural network. From this, high-level features based on the ratios of vessel length and volume of the occluded and the contralateral side are defined. Multi-class classification models are used to map the feature space to a four-grade collateral score and a quantitative score. The dataset used for training, validation and testing is from a registry of images acquired in clinical routine at multiple medical centers. The model performance is tested on 269 subjects, achieving an accuracy of 0.8. The dichotomized collateral score accuracy is 0.9. The error is comparable to the interobserver variation, the results are comparable to the performance of two radiologists with 10 to 30 years of experience.

2.1   Introduction

2.1.1   Clinical Background

Endovascular treatment (EVT) improves outcome in patients with acute ischemic stroke due to intracranial large vessel occlusion ([7]). Selection of eligible patients for EVT is important because not all patients benefit and the treatment is not without risk ([15]). Brain tissue at risk can survive longer in the presence of collateral circulation, which is a network of arterial anastomoses that provides blood flow to brain tissue when the principal conduits fail to meet demands ([22]).

The MR CLEAN trial, a multicenter, randomized trial of EVT versus no EVT, showed that baseline computed tomographic angiography (CTA) collateral status modified the treatment effect ([23]): patients with higher collateral score will most likely have better treatment outcome. A clinical decision tool based on multiple baseline clinical and imaging characteristics for individualized predictions of the effect of EVT has been developed and includes grade of collateral circulation as prognostic and predictive marker ([15]). Several collateral status grading systems exist, all based on visual scoring using coarse classification criteria ([24]). Such visual scoring systems suffer from subjective interpretations leading to inter- and also intra-observer variation. An automated scoring system could facilitate an objective and reproducible assessment of the cerebral collateral status. In our work, we use the four-grade score that was proposed by [20] as it has proven correlation with outcome and effect of EVT. The definition of this 4-grade score system is:

- 0: absent collaterals (0% filling in occluded territory)

- 1: poor collaterals ($>0\%$ and $\leq 50\%$ filling in occluded territory)

- 2: moderate collaterals ($>50\%$ and $<100\%$ in occluded territory)

- 3: good collaterals (100% filling in occluded territory)

Fig. 2.1 shows example images for four different collateral scores in Maximum Intensity Projection (MIP).
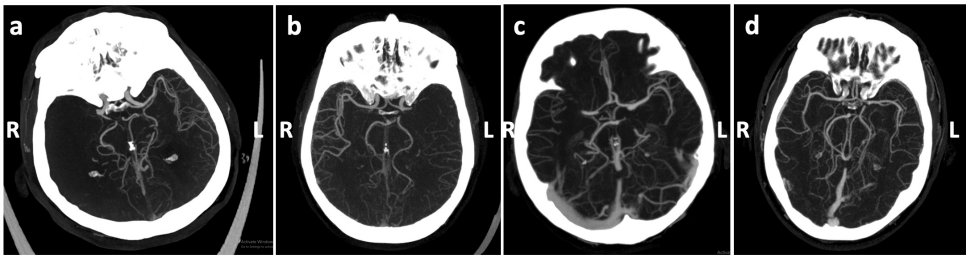

Figure 2.1: *Examples of four collateral score in Maximum Intensity Projection (MIP) from 3D CTA image. a: collateral score 0; b: collateral score 1; c: collateral score 2; d: collateral score 3.*

2.1.2   Related Work

Collateral status scoring relies on the difference between arterial trees in the middle cerebral artery (MCA) territory of the occluded side and its contralateral side. Therefore, vessel segmentation is an essential step in our application. [25] and [26] provide a review of many vessel segmentation approaches that have been developed over the past decades. For cerebral blood vessel segmentation, [27] summarized non-convolutional neural network based algorithms with respect to methods, image modality, and cerebral vessel segments. For example, [28] utilized a level set approach to detect the circle of Willis (CoW) in 3D CTA images, [29] utilized a Bayesian tracking approach to segment the internal carotid artery (ICA) in 3D CTA images, [30] utilized graph connectivity in combination with a tracking approach to obtain the label and vessel structure of the CoW in Magnetic Resonance Angiography (MRA) images, and [27] utilized a random forest classifier and histogram to extract the complete vessel structure from 4D CTA images.

Nowadays, convolutional neural networks (CNNs) have demonstrated in general superior performance for many medical image segmentation tasks. This trend is also witnessed in vessel segmentation, [31] summarized vessel segmentation using conventional and CNN based methods. [32], [33] and [34] employ 3D U-net based CNN model to extract the cerebral vessel structure from 3D Time-of-flight (TOF) MRA images, [35] use a 3D U-net based CNN model to extract cerebral vessel structure from 4D CTA images and further separate the arterial and vein structure by its spatial features, [36] utilize the 3D U-net frame work and replace the 3D convolution with a 2D cross-hair filter to segment the cerebral vessel structure and centerline in TOF MRA data. None of these approaches quantifies the collateral status.

To the best of our knowledge, only [37] published an automatic collateral scoring method. The region of interest (ROI) is defined by a probability density map that was generated from an atlas build from lesion segmentation from follow-up CT images. The 3D Frangi filter ([38]) with visually tuned parameters and a threshold of 200 Hounsfield units (HU) was applied to extract vessel structure in a pre-defined region of interest. The computed feature is the ratio of vessel volume with intensity between the occluded side and the contralateral side. The method was assessed on 59 subjects from which their follow-up non-contrast CT scan was used to construct the probability density map. The method was assessed on CTA images of patients with an occlusion in the M1 segment (for the detailed vessel segments in MCA territory please refer to Fig. 2.6) of the MCA territory, a maximal slice thickness of 1 mm and full coverage of the intracranial region.

2.1.3   Contributions and Organization of Our Work

In this work, we propose a three-stage algorithm to compute a collateral score. The collateral scoring method was assessed on 269 subjects.

Data preprocessing and vessel centerline segmentation are explained in Section 5.2, followed by feature design and a multi-label classification model for collateral scoring. The data set, collateral score reference standard and annotation strategy are described in Section 2.3, the experiments and results are detailed in Section 2.4, followed by discussion in Section 5.5, conclusions are drawn in Section 5.6.

2.2 Method

The proposed method starts with a pre-processing step to define the anatomical regions of interest: we use an atlas-based approach which was developed by [39] to obtain a 3D CTA brain image I_b, an MCA probability density map M and a hemisphere map H. This method takes the CTA image and an atlas image as input. This step is followed by a deep learning based segmentation of the brain vasculature (centerlines). In the final step, the output of the previous stages is transformed into a quantitative score, and a collateral class score. The algorithm overview is shown in Fig. 2.2. Each of the steps is detailed below.
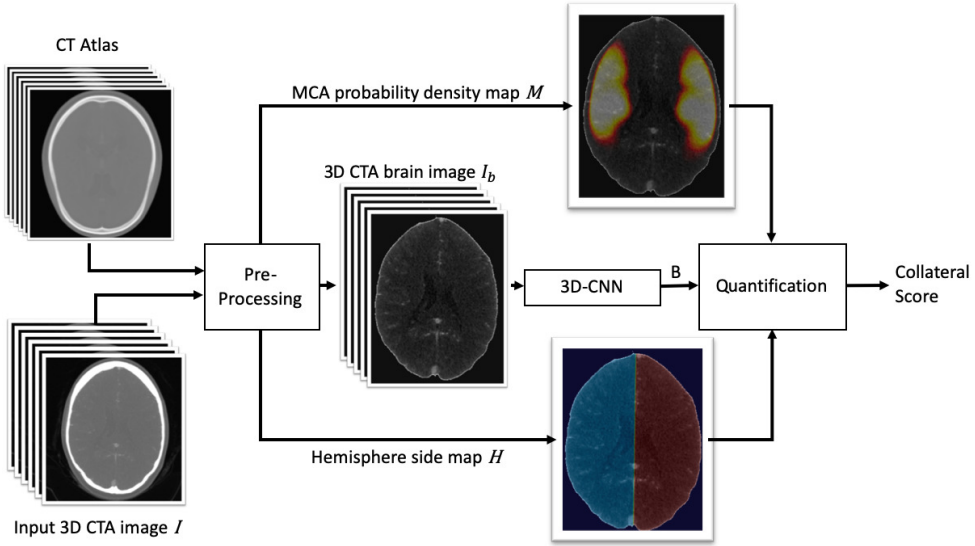
Figure 2.2: *Block diagram of proposed algorithms.*

2.2.1 Pre-processing: Anatomical Region Definition

The purpose of the first step is to define the relevant anatomical regions, i.e. the brain, the MCA region and both hemispheres (both left and right side of the brain). For the brain region, we use a CT atlas that was constructed from averaging high-resolution 3D CTA images of 30 healthy subjects ([40]) with a corresponding binary brain mask. This CT atlas was part of the symmetric CT-MR template of SPM toolbox ([40]). For the MCA region, we use a MCA probability density map (values ranging from 0 to 1) that represent the likelihood of MCA vessels present based on 50 healthy MCA vasculatures from the BraVa ([41] dataset. This was done by an affine and deformable symmetric diffeomorphic image registration of each subject from the BraVa dataset to the CT-MR template, following by a normalized sum of transformed individual MCA arterial trees. The hemisphere atlas is a three-value map that represents the left hemisphere, the right hemisphere and the background. All anatomical regions are defined in the space of the average CT atlas.

After registration of the CT atlas to the subject CTA image, the binary brain mask, the MCA probability density map and the hemisphere map are transformed to the subject CTA image space. The generated binary brain mask may fail to include vessel structures near the skull base. Therefore, to fine-tune the CT image space binary mask, an iterative morphological dilation with a 3D spherical structuring (radius equals 1 voxel) was applied to this binary mask. In each iteration, we remove dilated voxels if its corresponding 3D CTA voxel value exceed 850 Hounsfield unit (HU). The number of iterations is defined by the maximal gap, 5 mm, divided by the voxel dimension. A brain tissue image is constructed by multiplying the binary mask with the CTA image. After this masking step, the CTA brain image I_b is normalized with min-max normalization.

2.2.2   Deep Learning for Vessel Extraction

In the second step, the vasculature is extracted from the brain-masked and normalized CTA image. We opt for a deep learning approach, which trains a detector for vessel-like structures based on labeled training images; such approaches have been demonstrating excellent performance in the last years for many medical imaging tasks. More specifically, we intend to extract vessel structure with an encoder/-decoder architecture (the U-Net model proposed by [42]); this architecture is simple and still very effective ([43]). Note that for the final goal of obtaining collateral scores, we do not require a very precise segmentation of the vessels, rather we want to highlight the vessel centerlines such that the vasculature can be quantified in subsequent steps. For this 3D vessel extraction task, we use a standard 3D U-Net model as described in [44] to which we make modifications in the training process to tailor it to our vessel centerline segmentation task. Next, we will detail the network architecture, and the data preparation for the training.

Network Architecture

Our network architecture is shown in Fig. 2.3. The network utilizes the classic encoder and decoder architecture with a depth of 5 layers. The input data of the proposed model is a cube of 64x64x64 voxels with 28 convolutional filters in the first stage. The number of filters was determined according to a set of experiments we describe in Section 2.4.2. We apply instance normalization to each convolution layer. In the encoder path, activations are calculated using residual blocks. A residual block is a type-1 block in Fig. 2.3, including two cascaded 3x3x3 3D convolutional layers and the identity short connection (dashed lines in Fig. 2.3). This combination is similar to the context module described in [44], however, in our cases, we didn't use dropout layer in between. Two cascaded 3x3x3 3D convolutional layers with a stride of 2 (type-2 block in Fig. 2.3) are added in front of residual blocks in order to obtain more abstract feature maps as the encoder goes deeper. Leaky rectified linear units (ReLu) ([45]) are used as the activation function. In the concatenated decoder path, deconvolutional layers are constructed with an extra deep supervision ([46]) path. This can avoid information loss and vanishing gradients in each convolutional layer. The proposed model uses a sigmoid activation layer in the final step to output a 3D voxel-wise vessel probability map.
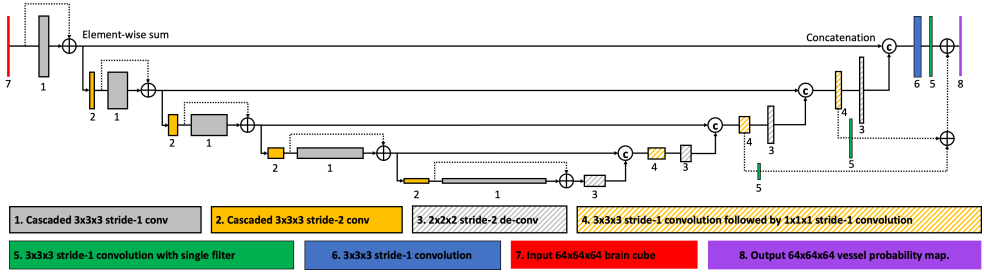
Figure 2.3: *Block diagram of CNN model.*

Data Preparation

For each brain, we construct ground truth vessel trees. The vessel tree is a 3D binary mask resulting from the vessel centerline annotation process described in Section 2.3.3. Then we split the whole brain and vessel tree into 3D cubes of 128x128x128 voxels for both training data and validation data. During the training process, we extract 64-voxel 3D cubes out of 128x128x128 voxels. In the validation process during training, we use 128x128x128 voxels as input data size for convenience. Data augmentation is applied to obtain different training images in each iteration. In the data preparation stage, we did not resample the image into common space as the training data is representative for the whole dataset and exhibits little variation in slice spacing and pixel size.

Deep Learning Post Processing

The proposed CNN model outputs a vessel probability map with values between 0 and 1. We threshold the probability map to obtain a binary vessel map B_0 (Fig. 2.4b). The threshold value is found by optimizing a Dice cost function on the deep learning validation dataset. There are some small isolated parts (mostly false positive parts) in the predicted vessel map B_0 and connectivity based noise removal ([47]) is applied to remove these small isolated parts. This results in a binary vessel tree map denoted by B (Fig. 2.4c).

2.2.3 Quantification

The purpose of the quantification step is to compute a collateral score from the results of the previous processing steps. Human collateral scoring is based on comparing the amount of vessels visible in the affected and non-affected side, and we follow a similar strategy: we compare the affected and non-affected hemisphere of subjects using a combination of the binary vessel structure from deep learning model output B, the corresponding MCA probability density map M and hemisphere map H as shown in Fig.2.2. We assume that it is known a priori (from clinical symptoms) which hemisphere (left or right) is affected. Based on this information and the hemisphere map, we generate an affected side binary map H_A and a non-affected side binary

2
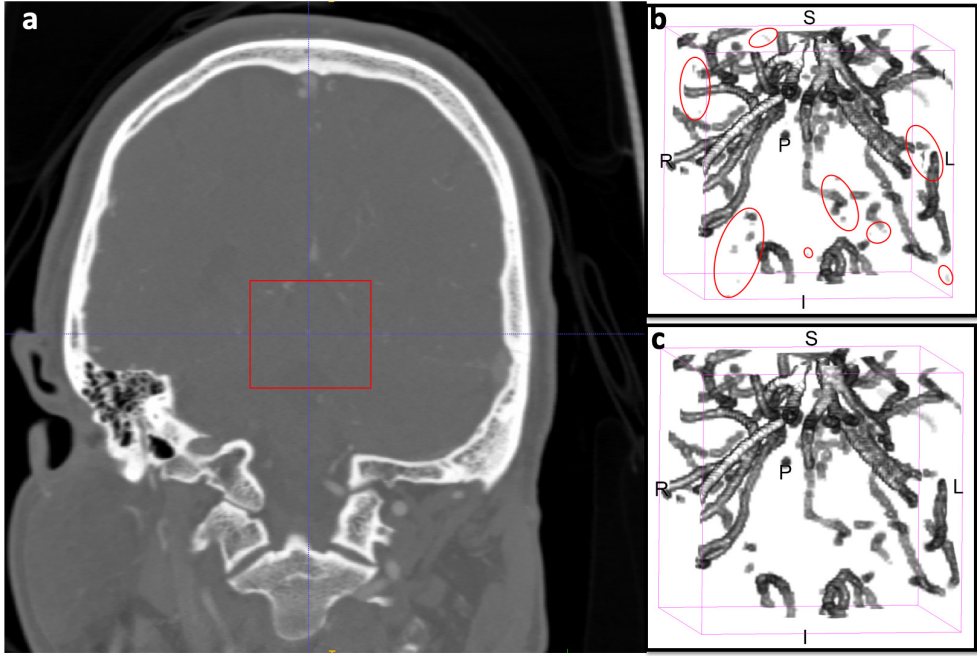
Figure 2.4: *The example image of deep learning post processing; a: shows the cube (100-pixel size) location in coronal view; b: shows an example binary vessel tree with some isolated parts B_0 in 100-pixel size cube; c: shows an example image of binary vessel tree B.*

map H_N. With this information, we compute four different ratios, each representing a different aspect of the vasculature, as detailed below.

Volume

An obvious quantification is difference in the number of vessels between the affected and the non-affected hemisphere. Assuming there are more vessels visible in the non-affected hemisphere, the ratio of vessel volumes (affected divided by non-affected) should give a number between 0 (no vessels visible in the affected side) to 1 (same amount of vessels visible in both sides). We propose two variants of this comparison: one where only the volume (number of voxels) is taken into account, and one where each voxel is weighted with its intensity r_v and r_{vi} respectively:

$$r_v \quad = \quad \frac{\sum_{p \in \mathcal{B}} M(p) \cdot H_A(p)}{\sum_{p \in \mathcal{B}} M(p) \cdot H_N(p)} \ , \qquad\qquad 2.1$$

$$r_{vi} \quad = \quad \frac{\sum_{p \in \mathcal{B}} I(p) \cdot M(p) \cdot H_A(p)}{\sum_{p \in \mathcal{B}} I(p) \cdot M(p) \cdot H_N(p)} \ , \qquad\qquad 2.2$$

where $\mathcal{B}$ is the set of voxels that are 1 in isolated part removed binary vessel tree B, p is a voxel and I is the image intensity.

Length

In addition to volume, we also consider vessel length. For the vessel length computation, a medial axis skeletonization approach that was developed by [48] is applied first. This step yields a pixel-wise skeleton structure S_0, a 26-connected structure (Fig. 2.5 left) which is a basic representation of the vessel network. Next, the vessel segments are determined from the vessel skeleton using a tree topology approach employed in the work of [47].

Subsequently those vessel segments were fitted by 3D spline curves and further smoothed ([49], [50]) and interpolated (Fig. 2.5 right), yielding a set of world coordinates S. For all points p from S we obtain the corresponding MCA probability value $M(p)$ from the 3D MCA probability density map. In same way, we obtain intensity values $H_A(p)$, $H_N(p)$, and $I(p)$. The ratios r_l and r_{li} are accumulated values over all points (weighted with the mean distance w between the point and its neighbours) in the affected hemisphere and the ones in the non-affected hemisphere:

$$r_l = \frac{\sum_{p \in S} M(p) \cdot H_A(p) \cdot w(p)}{\sum_{p \in S} M(p) \cdot H_N(p) \cdot w(p)} \ , \qquad 2.3$$

$$r_{li} = \frac{\sum_{p \in S} I(p) \cdot M(p) \cdot H_A(p) \cdot w(p)}{\sum_{p \in S} I(p) \cdot M(p) \cdot H_N(p) \cdot w(p)} \ . \qquad 2.4$$

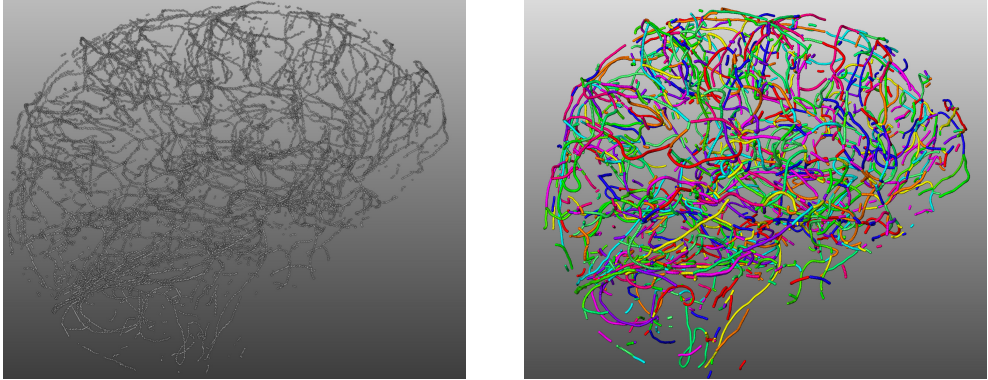***Figure 2.5:*** *An example of vessel length computation. Left: the vessel skeleton; Right: the smoothed and interpolated vessel segments, the vessel length computation are based on those segments. Different segments are shown in different colors.*

2

Multi-label Classification

In the last step, multi-class classification is used to predict collateral score (0,1,2,3) from an input feature vector $\boldsymbol{r} = [r_v, r_{vi}, r_l, r_{li}]$. We start from the baseline model, in which we take the median of feature vector $\boldsymbol{r}$ as input and define the threshold value by utilizing the clinical definition of collateral score. We then define our second model by using a support vector classifier (SVC) with linear kernel to find the optimal threshold value for the median of feature vector $\boldsymbol{r}$. For the third method, we use the complete feature vector $\boldsymbol{r} = [r_v, r_{vi}, r_l, r_{li}]$ as input to a random forest classifier. Finally, we use ordinal regression with the complete feature vector for ordered categorical prediction.

2.3   Data and Annotation

2.3.1   Data Overview

The images used for training and assessing the methods were obtained from the MR CLEAN Registry ([51]) and MR CLEAN trial ([10]). The MR CLEAN Registry is an on-going registry that contains patients who underwent endovascular treatment at a stroke intervention center in the Netherlands since March, 2014. The CTA images are acquired in clinical routine at several different sites, and thus there is large variation in image quality, as well as in imaging equipment and acquisition protocols (contrast phase, brain coverage).

Hence, in order to get a representative set of images, the following selection criteria have been used to select images in this study:

- **Spatial resolution:** The average diameter of the M1 segment of the MCA region is 3.1 ± 0.4 mm and of the M2 segment 2.4 ± 0.4 mm according to [52]. The spatial resolution should be sufficient to visualize the major arterial tree. Therefore, we only select images of which the slice thickness is smaller than 1.5 mm; additionally we require the slice spacing to be smaller than or equal to the slice thickness.

- **Contrast phase:** Five different contrast phases have been previously defined by [6]: early arterial, peak arterial, equilibrium, peak venous, and late venous, based on the image intensities in the contralateral ICA and the transverse sinus. In the early arterial phase, collateral vessels are likely not enhanced yet. In the late venous phase, the venous structures are more pronounced than arterial structures. Therefore, in this study we only select images with peak arterial, equilibrium or the early venous phase.

- **Image quality:** Image quality in MR CLEAN Registry is rated by the core lab into the following categories: good image quality, moderate image quality and bad image quality. A good quality image implies that the image is sufficiently informative for radiologist to rate. We similarly included good quality images in our analyses.

- **Brain coverage:** Brain coverage, the image should cover at least half of the vertical distance between skull base and vertex.

At the start of the current study, baseline CTA data of 1594 subjects had been collected in the MR CLEAN Registry. These data was acquired from 16 March 2014 till 15 June 2016. Of these 1594 subjects, the images of 1058 subjects had good image quality (based on MR CLEAN Registry core lab readings). 736 subjects fulfill both image quality and contrast phase criteria. At the end, 585 subjects fulfill all selection criteria. From this set of 585 subjects, 49 subjects were manually selected for annotation and training of the CNN. From the remaining 536 subjects, we randomly selected around half of these subjects (270 images) for our study. This number was assumed to be sufficient for our evaluation, and also reduced the amount of work for obtaining a consensus score compared to using the full set. Those 270 subjects were originally from 14 intervention centers with different vendors (mainly from Philips, Siemens, Toshiba and GE). For all cases, we obtained the occlusion side and occlusion position and initial collateral score from the registry information. From those 270 subjects, one subject was additionally excluded which was considered to have insufficient image quality by the expert readers. We did not have additional selection criteria on occlusion location. However, only 6 subjects with A1 or A2 occlusions were found among the 1574 MR CLEAN registry subjects, and our final set of images did not contain any A1 or A2 occlusions.

In addition, for training we added five collateral score 0 cases from the MR CLEAN Trial. Collateral score 0 is rare, and due to the random selection, only two images with collateral score 0 were in our initial selection. The additional five subjects were only used in the classifier training process. The accuracy of the algorithm was thus evaluated on the randomly selected 269 cases.

2.3.2   Collateral Score Reference Standard

To get a consistent and reliable collateral score reference standard for this study, collateral scoring was performed by three radiologists from different medical centers with 10 to 30 years of experience. Two radiologists rated the collateral score independently and the third radiologist independently rated the cases in which there was disagreement by the first two radiologists. The radiologists were asked to rate the collateral status according to the criteria of [20].

In this 269 subjects, the two independent raters had an interobserver agreement of 0.64, and their scores compared to the consensus score were 0.81 and 0.82. The details of the 269 test subjects are listed in the Table 2.1. Fig. 2.6 shows the location of occlusion in the vessel segments that are listed in Table 2.1.

2.3.3   Data Annotation

Training and assessment of the deep learning based vessel extraction requires (manually) annotated images. In total 9 whole brains and 40 cubes of 128x128x128 voxels were annotated. The 9 subjects were selected by a radiologist as being cases representative for daily clinical practice. The 40 subjects for cube annotation were selected to cover large variation of image quality and acquisition parameters, as well as collateral scores. The 40 cubes were randomly selected from the intracranial region of 40 subjects. Manual annotation was performed using an in-house developed tool build with MevisLab.

2
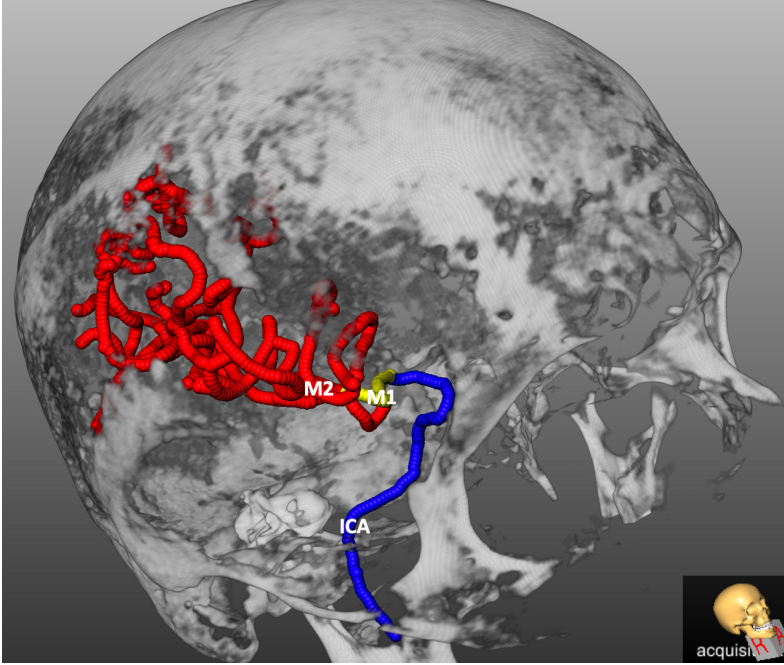
Figure 2.6: *An example of vessel segments in right hemisphere. The vessel segments include the ICA segment, M1, M2 and the more distal part of the MCA territory, for simplicity, we have combined M2 and the more distal part into M2.*

The annotation task was performed by 1 pre-med student, 3 experienced physicians and the first author of this paper. The purpose of the annotation task was to label centerlines of all intracranial vessels. Annotation points needed to be placed in the center of candidate vessel structures, after which semi-automated processing (shortest path connectivity, followed by a graph cut segmentation, yielding contours orthogonal to the centerline ([53]), after which the real centerline defined as the centers of the segmentation result) was applied. The labelled region also include vessels running through the skull base and isolated vessel structures. Fig. 2.7 shows the result of a whole brain annotation.

2.4   Experiments and Results

2.4.1   Implementation

The method consists of three parts: the pre-processing (atlas-based registration), deep learning based vessel centerline extraction, and classification. The deep learning model is written using the Keras and Tensorflow frameworks. Model training and validation are implemented on a local PC equipped with one NVIDIA Titan Xp GPU and 64 Gb

Property	Category	Number of subjects
Slice thickness	[0.5-0.75mm)	66
	[0.75-1mm)	94
	[1-1.5mm]	109
Acquisition phase	peak arterial phase	75
	equilibrium phase	115
	peak venous phase	79
Collateral score	score 0	2
	score 1	83
	score 2	124
	score 3	60
Occlusion location	ICA	63
	M1	155
	M2 and above	47
	No occlusion found	4

Table 2.1: *Data distribution of 269 test subjects.*
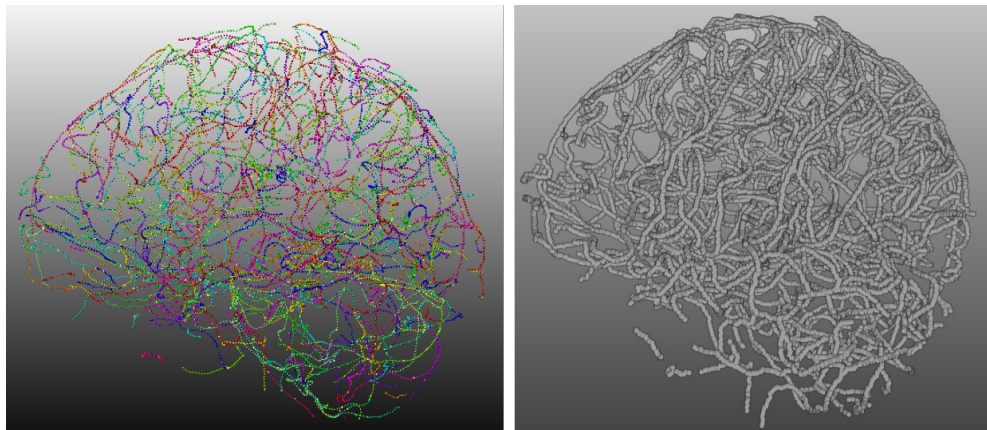
Figure 2.7: *The image in the left is the result of whole brain annotation; The image in the right is the ground truth image. Different vessel segments are shown in different colors.*

of RAM. The pre-processing, quantification and classification of proposed method were implemented in MATLAB 2018a.

Registration of the CT atlas to the CTA image was performed with ANTs ([54, 55]), following a previously described CT atlas based registration [39] that consists of a two-step approach: an initial rigid registration followed by a diffeomorphic non-rigid registration.

In the deep learning model training process, the network was trained with a batch size of 8 and 50 steps per epoch. In each epoch, we iterate twice over all 200 training images of 128x128x128 voxels: once with random shift and flipping along each axis and once with elastic deformation. We also introduced additive Gaussian noise at the input layer for regularization. The standard Dice score served as the loss function. We

chose the root mean square propagation (RMSProp) ([56]) optimizer with an initial learning rate of 0.1, and halved it every 10 epochs. We stopped training after 300 epochs, as the learning rate was approaching zero.

In the multi-label classification part, the baseline model has fixed threshold values. The three threshold values(θ_1, θ_2, θ_3) were determined by the clinical definition of collateral score ([20]). We introduced a small margin to the collateral 0 case, since, in practice, there were always some vessels in the occluded side. Therefore, we define the collateral 0 case as less than 10%. The other three threshold values follow the clinical definition, i.e. $\theta_1 = 10\%$; $\theta_2 = 50\%$; $\theta_3 = 100\%$. For the random forest classifier, we used five fold cross validation to optimize the maximal depth of the trees, the number of trees in forest and the number of features used.

2.4.2 Vessel Extraction Model Training and Hyperparameter Optimization

In the first experiment, we trained the convolutional neural network for vessel extraction. In this experiment, we first investigated the performance of 3D U-Net model with different hyperparameters and configurations. Compared to a standard 3D U-Net ([43]) (our baseline model), we first assessed the model performance enhanced by deep supervision and context modules, and subsequently also assessed the added value of varying the number of filters at convolutional layers.

The annotation dataset was randomly divided into a training and validation dataset. The training dataset consisted of 7 whole annotation brains and 20 cubes. The validation dataset consisted of the other 2 brains and 20 cubes. In order to guarantee the continuity and completeness of the centerline structure in the training process, we dilated every single-pixel centerline from manual annotation with a 3x3x3 square structuring element. The resulting ground truth image are shown in Figure. 2.7.

The Dice score on the validation dataset was used to measure model performance. In the first experiment, we assessed the added value of the various configurations. For this experiment, 28 input filters were used, as this was the maximal number that fitted into our GPU memory. The performance of the baseline model is shown in the first row of Table 2.2. Improvement in Dice score is observed when adding deep supervision and the context module to the baseline model. A combination of all components (last row of Table 2.2) shows the highest validation score. We also performed a paired T-test for two average values μ_1 and μ_i ($i = 2, 3, 4$), where μ_1 is the average of Dice score on all validation subjects with 3D U-net configuration and μ_i the average dice score corresponding to the network configuration in i-th entry of Table 2.2. Whereas the improvement of the final configuration is not statistical significant at the common 0.05 level, the trend in the Table 2.2 is clear and we attribute the lack of statistical significance to the limited validation set size. We therefore choose the 3D U-net with deep supervision and context module for the subsequent experiments.

Whereas the number of input filters increases model capacity, it also greatly affects the number of parameters of the model, and thus may impact the training process. In the next experiment, we therefore vary the number of input filters from 4 to 28 for the configuration in the last row of Table 2.2. The learning curves are shown in Fig. 2.8. As the number increases, both training and validation scores increase. This trend ends

Network Configuration	Dice mean	Dice std	p-value
3D U-net	0.53	0.19	-
3D U-net + deep supervision	0.54	0.19	0.69
3D U-net + context module	0.55	0.17	0.17
3D U-net +deep supervision + context module	**0.56**	**0.16**	**0.07**

Table 2.2: Deep learning test result for different configurations.

when the filter number goes beyond 24, where the improvement to the test score is marginal. Going from 24 filters to 28 filters, the test score only improves with less than 1%, compared to an improved of 4% when going from 4 filters to 8 filters.
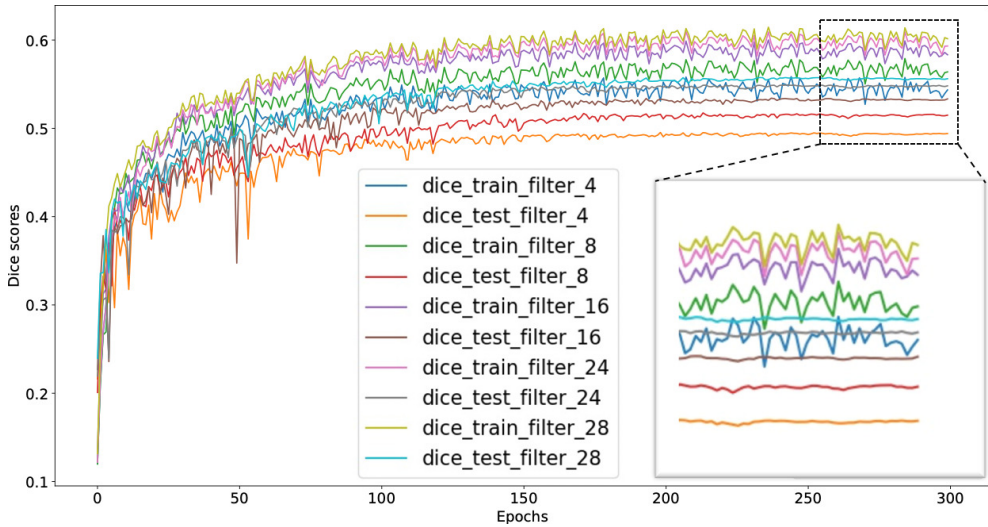
Figure 2.8: Learning curve with varies filter number.

The average Dice is 0.56 with a minimum of 0.42 and a maximum of 0.67. There are two outliers (0.27 and 0.28). Fig. 2.9, shows two example vessel trees extracted by the proposed deep learning model.

2.4.3 Collateral Score Quantification

Next, we evaluated the proposed collateral scoring method. First, we assessed the accuracy of the three proposed models. Then, we assessed the accuracy of proposed methods applied to a dichotomized decision based on collateral score.

In the accuracy test for collateral scoring we evaluated the baseline model, a linear support vector classifier (SVC) with single feature, ordinal regression with four features, and a random forest model as describe in Section 2.2.3. We used the consensus score as ground truth label.
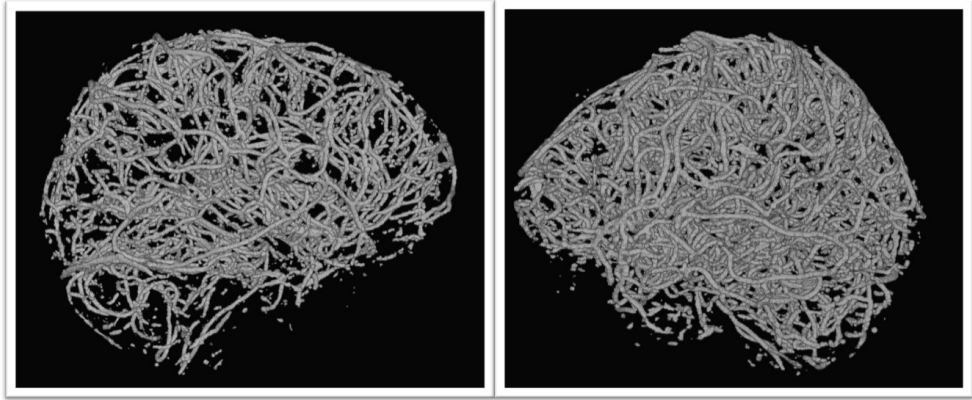
Figure 2.9: *Two extracted vessel trees in sagittal view.*

In the baseline model, we applied threshold values θ_1, θ_2, θ_3 to the median of $\boldsymbol{r} = [r_v, r_{vi}, r_l, r_{li}]$ and derived the collateral scores 0, 1, 2, and 3. Similarly, we use the median of $\boldsymbol{r} = [r_v, r_{vi}, r_l, r_{li}]$ with linear SVC model to find another set of threshold values that maximize the collateral score test accuracy on 269 subjects. The average threshold values for the SVC model were: $\theta_1 = 7\%$; $\theta_2 = 55\%$; $\theta_3 = 99\%$. We use ordinal regression with $\boldsymbol{r} = [r_v, r_{vi}, r_l, r_{li}]$. The averaged odds ratios of four features are [1.17, 0.9, 0.92, 1.15]. We further evaluated the added value of random forest in terms of accuracy. We explored the feature vector $\boldsymbol{r}$ starting from a single feature towards combined features and tested all 16 possible combinations. We performed a nested cross-validation with 20 splits at the outer level and 5-fold cross-validation for parameter tuning. More specifically, we first randomly splitted the 274 subjects into 20 subsets with a stratified sampler. Then we trained a model with data from 19 subsets and tested on 1 left-over subset. During training, we use 5-fold cross validation for hyper-parameter optimization of feature vector $\boldsymbol{r}$. In the end, we have 20 models with similar performance but with different parameter settings. In this way, we could fully utilize 269 subjects for testing and reduce the possible bias caused by a smaller test dataset.

Feature vector $[r_v, r_{li}]$ shows the highest prediction accuracy on test subjects. The results of baseline model, linear SVC, ordinal regression and the random forest with optimal feature vector are shown in Table 2.3. The proposed baseline model (*median* with fixed threshold value) and ordinal regression model have the same performance (accuracy=0.75), marginally different from SVC (accuracy = 0.76). In terms of accuracy, the random forest classifier outperforms the baseline model by 5%.

In the study of [20], a dichotomized collateral score that was proposed by [57] was used to assess the correlation with infarct volume before recanalization. Furthermore, in the MR CLEAN trial, substantial treatment effects were reached in patients with collateral scores of 2 or 3, whereas in patients with scores of 0 and 1, outcomes were poor and treatment effects small. Therefore, we focus on the accuracy of dichotomized prediction, wherein scores 2 and 3 are merged into one class, and score 0 and score 1 are merged into another class. We follow the same experimental setup as in the previous

Methods	Accuracy	Di-chotomized score
median +fix (baseline)	0.75	0.87
median + SVC	0.76	0.87
ordinal regression	0.75	NA
logistic regression	NA	0.89
random forest	0.80	0.90

Table 2.3: The accuracy, dichotomized accuracy of three proposed methods.

accuracy test. We use logistic regression with $\boldsymbol{r} = [r_v, r_{vi}, r_l, r_{li}]$ for this dichotomized binary classification problem. The performance of all four models is similar. The random forest classifier with features r_v and r_{li} performs slightly better. The overall dichotomized accuracy is listed in the dichotomized score column of Table 2.3. For comparison, clinician 1 and clinician 2 have an accuracy of 0.91 and 0.90 respectively in this case. We further assessed the performance of this binary class classifier by plotting receiver operating characteristic (ROC) curve of our proposed methods. Fig. 2.10 shows the ROC of the baseline model, and the performance of the clinicians, linear SVC and random forest classifier.
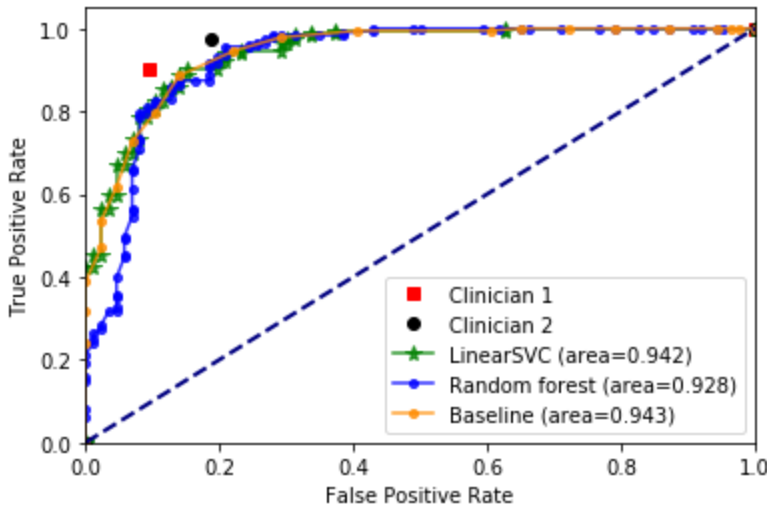
Figure 2.10: *The ROC curve of baseline model, linear SVC model and random forest model.*

Next, we investigated the misclassified subjects. Fig. 2.11 shows the confusion matrix test results on 269 subjects, for the two radiologists and the random forest classifier versus the consensus score. In total, the random forest classifier predicts an incorrect collateral score for 53 subjects; 26 out of these 53 subjects have predictions that are consistent with one of the two clinicians. The baseline model predicts an

incorrect collateral score for 66 subjects; 30 out of these 66 subjects have agreement with one clinician. Finally, in Table 2.4 we present a subgroup analysis.

The computation time for one brain with 0.5 mm slice thickness is about 15 min. Of this, the atlas to CTA space registration, MCA map, hemisphere map transformation with binary skull dilation takes 12 min, of which most time is spent on the Ants registration. It is likely that this can be optimized for clinical applicability. The vessel extraction is around 0.5 min. The quantification part in total is around 2.5 min. Those times were obtained using an implementation that was not optimized for computation time.

	0	2 / 2 / 2	0 / 0 / 0	0 / 0 / 0	0 / 0 / 0
Consensus score	1	6 / 0 / 0	74 / 65/ 63	3 / 19 / 21	1 / 0 / 0
	2	1 / 0 / 0	17 / 5 / 4	98 / 101 / 110	7 / 17 / 9
	3	0 / 0 / 0	1 / 2 / 1	14 / 7 / 18	45 / 51/ 41
		0	1	2	3

Predicted collateral score (C1 / C2 / RF)

Figure 2.11: *Confusion matrix of clinician 1 / clinician 2 / random forest versus consensus score.*

2.5   Discussion

In this work, we proposed an automatic collateral scoring method that is inspired by human visual collateral scoring. In the method, we compute the collateral score by comparing the difference of vessel structures in the occluded hemisphere and contralateral hemisphere. Vessel structure extraction is important for collateral status quantification. Therefore, we first investigated the performance of CNN and found a positive effect of context module and deep supervision on the performance of the 3D U-net vessel centerline segmentation model. We further evaluated network performance with an increased number of filters at convolutional layers. On average, we achieve a Dice score of 0.56 on the validation dataset. Whereas this might seem low, a value of 0.56 is reasonable for thin linear structures in 3D, as they contain a large proportion of boundary voxels. For such long thin structures, a single pixel shift in a direction orthogonal to the structure have a major impact on the Dice score.

In the collateral scoring assessment, the baseline model with using the median of feature vector $r = [r_v, r_{vi}, r_l, r_{li}]$ and fixed threshold values achieved an accuracy of 0.75 and a dichotomized accuracy of 0.87. This demonstrates that the features are relevant. We observed that the misclassified subjects are mostly at the border of decision boundary. The average error distance is 0.25. The error is computed from the floating point score, and represents the distance to the closest value of interval of the correct collateral score. For example, [0,1] corresponding to collateral score 0.

Property	Category	Number of subjects	Accuracy			
			Baseline	RF	Clinician 1	Clinician 2
Slice thickness	[0.5-0.75mm)	66	0.79	0.80	0.77	0.80
	[0.75-1mm)	94	0.79	0.87	0.82	0.84
	[1-1.5mm]	109	0.71	0.74	0.83	0.80
Acquisition phase	peak arterial phase	75	0.76	0.79	0.83	0.84
	equilibrium phase	115	0.77	0.8	0.79	0.79
	peak venous phase	79	0.73	0.81	0.84	0.82
Collateral score	score 0	2	1.00	1.00	1.00	1.00
	score 1	83	0.69	0.75	0.77	0.88
	score 2	124	0.89	0.90	0.82	0.80
	score 3	60	0.57	0.68	0.85	0.75
Occlusion position	ICA	64	0.81	0.86	0.86	0.87
	M1	154	0.75	0.81	0.86	0.77
	M2 and above	45	0.67	0.71	0.69	0.71
	no occlusion found	6	0.83	0.83	1.00	1.00

Table 2.4: Subgroup analysis.

On this scale, the error made in classification is the distance to the closest border of the correct class. The random forest model on average achieves an accuracy of 0.8, which outperforms the baseline model on the border cases. With two principle features $\boldsymbol{r}' = [r_v, r_{li}]$, the random forest model achieves a dichotomized accuracy of 0.9, which is comparable to two clinicians (0.91 and 0.9).

In Table 2.4, a subgroup analysis is presented for the acquisition phase, collateral score, slice thickness and occlusion position. The accuracy does not seem to depend on acquisition phase or collateral score. It also shows that collateral scoring is less accurate for more distal occlusions. This can be explained by the fact that a smaller region is affected in more distal occlusion. This trend can also be observed in the human scoring. In terms of slice thickness, both error rates increase along the slice thickness. When the slice thickness increases, the vessel structure is less pronounced in the 3D CTA image. The data used in this work was acquired 3 to 5 years ago. The latest images mostly have slice thicknesses less than 1 mm.

The anatomical regions are defined by a conventional atlas-based registration method due to the fact that the MCA probability density map is an essential element of the feature computation in the quantification step. In order to align the MCA probability mask, we need to first register the CT template to the CTA space. The registration result then can similarly be used to bring the hemisphere and brain segmentation to the patient space.

Collateral scoring involves only the arterial tree. In our approach, we do not discriminate between arteries and veins in the centerline segmentation. Application of the MCA territory mask, which was build from arterial trees of 50 subjects, removes some of the venous structures from the segmentation. Any remaining venous structures may affect the final quantification. As the amount of veins in the remaining region generally is small in the region of interest especially in peak arterial, equilibrium and early venous phase, and as their presence is expected to be symmetric, the remaining veins may have a minor effect on the subsequent quantification, which is demonstrated by our current results. Still, including an artery-vein separation may be a an interesting direction for future research.

End-to-end training might have been an alternative approach to computing collateral scores. We chose for a slightly more conventional approach for four reasons. First, the data and corresponding ground truth required for training may be need to be larger than the set we are using now. This would require additional expert radiology screening. Second, such approach generally require a network that is trained with full-size CTA images whereas our current vessel segmentation is trained with 3D patches. Such a training would be challenging to commonly available GPUs. Thirdly, the result of such end-to-end training is difficult to interpret, whereas our approach also gives insight in the vessel segmentation and quantification on which the scoring is based. Finaly, we aim for a more quantitative analysis of collaterals with a tool that provides a continuous output instead of a semiquantitative scale with 4 items.

In comparison with [37], the dataset and the level of ground truth are different: [37] used the data from MR CLEAN Trial ([10]) and we used dataset from MR CLEAN Registry ([51]). Their collateral score was derived by direct use of feature r_{vi}, and there was no direct assessment over collateral score; instead they perform a correlation test between their single feature and the manually obtained collateral score with a

Spearman correlation test. The result on 59 subjects was a Spearmen $\rho = 0.68$, p $<$ 0.001. For our method, a Spearmen correlation $\rho = 0.80$ was obtained on a test set of 269 subjects.

During this study, a collateral scoring product (e-Stroke Suite) became available from Brainomix. This software was evaluated by [58] recently. Ninety-eight subjects were used in their work. Their selection criteria were more restrictive than ours (1 mm slice thickness), and no information is provided on the occlusion location or number of excluded scans. Also, their consensus score may be biased towards the software performance, as the consensus score was determined after knowing the software score, which makes it hard to directly compare their result with ours. Reported accuracy of software compared to their reference standard is 90%, and they also demonstrate, similar to our work, that the errors of the software are within the interobserver variation.

All patients from the MR CLEAN registry that were included in the study had an acute large vessel occlusion, which was assessed on CTA. In all patients a large vessel occlusion was detected in one hemisphere only. Although we cannot rule out that small peripheral emboli were present in the contralateral hemisphere (which were not visible on CTA) in patients with a cardioembolic etiology, in general, the symptoms caused by a large vessel occlusion are more prominent than the symptoms caused by potential small peripheral emboli. Based on that we assume that the symptoms indicate the hemisphere in which the large vessel occlusion is present. However, outside the MR CLEAN Registry, indeed cardioembolic stroke accounts for 14-30% of ischemic stroke population ([59]), and for those cases, the assumption of occluded side as prior might not be valid. Our method is not intended for those cases.

2.6 Conclusions

We proposed a robust and automated collateral scoring method and evaluated it on a large set of images acquired in clinical routine as demonstrated in the MR CLEAN Registry. The proposed method achieves 80% accuracy on the 4 score prediction. In a dichotomized test, the proposed method achieves 90% accuracy. The result are comparable to two radiologists with 10 to 30 years of experience.

Acknowledgment

Chapter 4
Evaluation of Automatic Collateral Scoring

3

Abstract

Purpose: *Outcome of endovascular treatment in acute is-chemic stroke patients is depending on the collateral circulation maintaining blood flow to the ischemic territory. We evaluated the inter-rater reliability and accuracy of raters and an automated algorithm for assessing the collateral score (CS, range: 0–3) in acute ischemic stroke patients.*

Methods: *Baseline CTA scans with an intracranial anterior occlusion from the MR CLEAN study (n=500) were used. For each core lab CS, ten CTA scans with sufficient quality were randomly selected. After a training session in collateral scoring, all selected CTA scans were individually evaluated for a visual CS by three groups: 7 radiologists, 13 junior and 9 senior radiology residents. Two additional radiologists scored CS to be used as reference, with a third providing a CS to produce a 2 out of 3 consensus CS in case of disagreement. An automated algorithm was also used to compute CS. Inter-rater agreement was reported with intraclass correlation coefficient (ICC). Accuracy of visual and automated CS were calculated.*

Results: *39 CTA scans were assessed (1 corrupt CTA-scan excluded). All groups showed a moderate ICC (0.689-0.780) in comparison to the reference standard. Overall human accuracy was 65± 7% and increased to 88± 5% for dichotomized CS (0–1, 2–3). Automated CS accuracy was 62%, and 90% for dichotomized CS. No significant difference in accuracy was found between groups with different levels of expertise.*

Conclusion: *After training, inter-rater reliability in collat-eral scoring was not influenced by experience. Automated CS performs similar to residents and radiologists in determining a collateral score.*

3.1 Introduction

Acute ischemic stroke is mainly caused by the occlusion of one or more brain arteries, which leads to an inadequate supply of oxygen to a region of the brain [60]. Globally, ischemic stroke is the second leading cause of death, and a major contributor to disability-adjusted life years in the population [61]. The most effective treatment in ischemic stroke is timely reperfusion of the occluded arteries [62]. Patients with an intracranial large vessel occlusion could be eligible for endovascular thrombectomy (EVT). The thrombus is thereby removed from the vessel with a stent retriever and/or aspiration device, restoring the original blood flow and oxygen supply [63]. However, EVT is not an intervention without risks, and the effect of treatment will vary between individuals [15]. The effect of EVT is dependent on different clinical and imaging parameters. These parameters can be assessed pre-operatively to determine the chance of a treatment benefit and for patient selection, especially in the late time window [10]. Therefore, it is important to investigate how consistent and reliable these parameters can be obtained pre-operatively.

One of the parameters relevant to determine treatment effect is the collateral score (CS) [23]. The CS quantifies the contrast filling of the distal MCA branches through the arterial collateral circulation in the affected hemisphere. The collaterals are secondary pathways which can function as a back-up when the primary arteries fail to deliver an adequate blood supply [22]. Brain tissue at risk due to an occlusion is more likely to survive a period with insufficient blood supply through primary pathways if oxygen supply is ensured through collateral vessels [22].

Tan et al. developed a 4-point categorical grading system for assessment of collateral status in the occluded middle cerebral artery (MCA) territory on a computed tomography angiography (CTA) scan [18]. A score of 0 is given for absent collaterals, 1 for $> 0\%$ and $\leq 50\%$ collateral supply filling, 2 for $> 50\%$, and 3 for 100% filling of the occluded MCA territory [18]. The collateral score for a CTA scan is generally obtained by visual scoring, which is operator dependent with potential interobserver variation. Machine learning-based approaches to produce an automated quantitative collateral score (qCS) showed similar performance to that of experienced radiologists [21].

In this study, we aim to assess the interobserver variability for the CS and whether variability is influenced by years of experience [20]. Secondly, we compare the visual CS given by physicians with the previously mentioned qCS and a reference CS.

3.2 Method

3.2.1 Imaging Data

Baseline CTA scans were acquired from the Multicenter Randomized Clinical Trial of Endovascular Treatment for Acute Ischemic Stroke in The Netherlands (MR CLEAN, MR CLEAN Netherlands Trial Registry number: NTR1804. Current Controlled Trials number, ISRCTN10888758), a prospective, consecutive study which was performed in 16 stroke centers in the Netherlands [10]. The MR CLEAN study protocol was approved by the central medical ethics committee of the Erasmus MC and the research

board of each participating center. All patients or their legal representatives provided written informed consent before randomization.

The MR CLEAN database contains data from 500 patients with acute ischemic stroke caused by an occlusion in the anterior circulation. Pre-interventional CTA scans were rated for CS and occlusion location by a core lab of radiologists without access to other imaging data or any clinical information. CTA scans with good/moderate image quality; adequate head coverage; axial series; slice thickness <1.0 millimetres; and slice increment equal to or smaller than slice thickness were selected. From those CTA scans, ten scans were randomly selected for each CS. Axial and coronal maximum intensity projections (MIPs) with slice thickness of 8 millimetres were reconstructed.

3.2.2   Visual Collateral Scoring

If applicable, year of residency was recorded. Seven radiologists and 22 radiology residents in the Netherlands attended a 1-h training session in collateral scoring. The rationale and method of scoring were explained, and CS examples were shown and discussed.

After the training, attendees were asked to score all cases based on the axial CTA scan and 8 mm MIP reconstructions combined. The occluded vessel (ICA-T/M1/M2) and affected hemisphere (left/right) were given for each case.

3.2.3   Automated Collateral Scoring

qCS were produced with the model reported by Su et al. [21]. CTAs acquired in the MR CLEAN study were not used for the development of this model. Processing of the CTA scan started with an atlas-based registration and segmentation of the vessel centrelines using a neural network. After this, the relative amount of vessels in the MCA territory was quantified by comparing the affected hemisphere with the unaffected hemisphere in terms of vessel volume and vessel length, both weighted and unweighted for pixel intensity. The four ratios were used to compute a qCS. The qCS was converted to a collateral score using a modified definition of Tan et al.: collateral score 0 was defined as equal or less than 10% filling instead of 0% filling of the affected MCA territory [18].

3.2.4   Reference standard

The CTAs which were evaluated by the imaging core lab were re-evaluated for CS by two independent and experienced interventional neuroradiologists (A.v.E, P.J.v.D), who were not part of the group of raters. In case of disagreement, the core lab observer rating was used as third CS to provide a two-observer consensus.

3.2.5   Statistical Analysis

The results were analysed after grouping the respondents as follows: first and second-year radiology residents (junior residents, n=13), radiology residents in years 3–5 (senior residents, n=9), all radiology residents (n=22), radiologists (n=7), and all physicians combined (n=29). The mean and standard deviation of the visual CS were
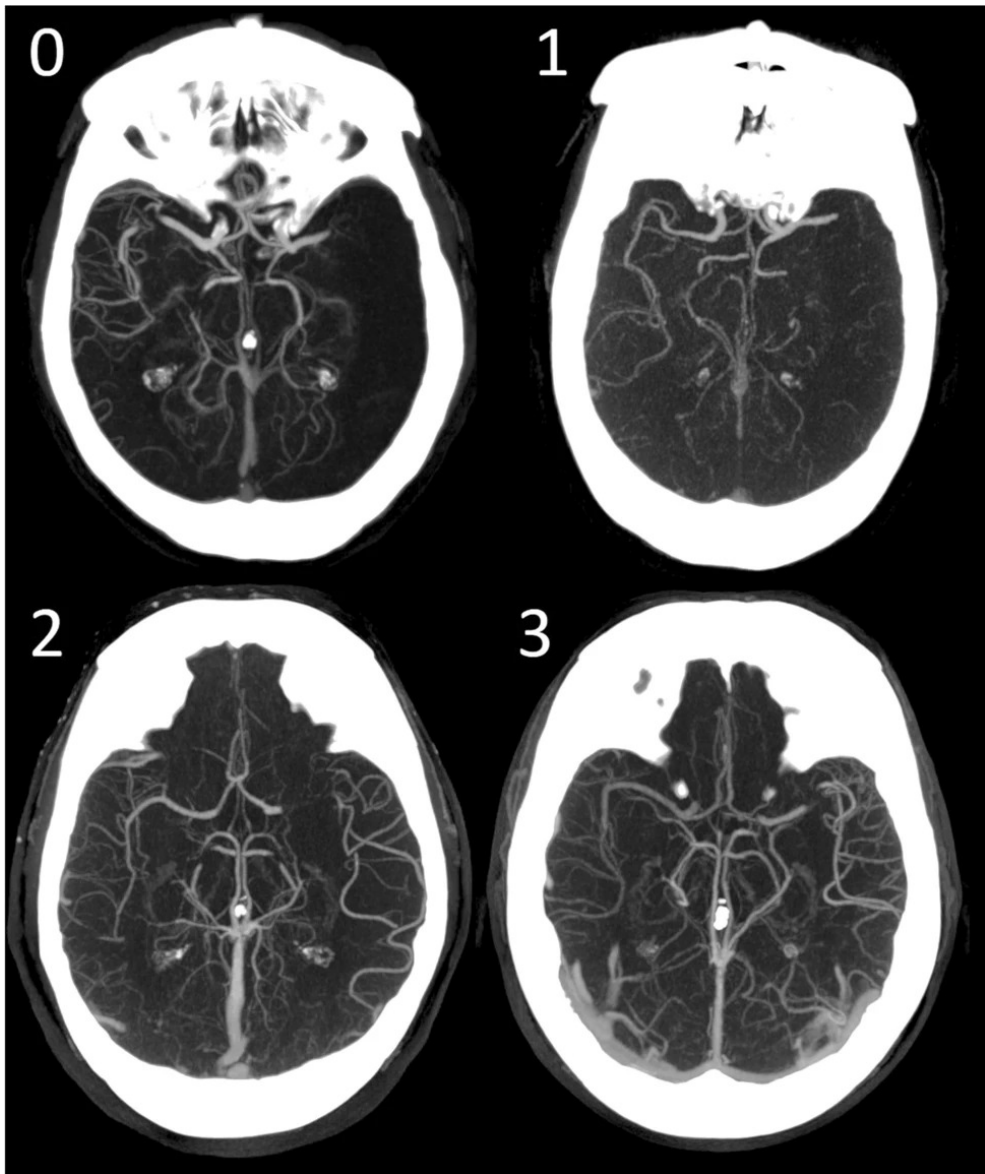
Figure 3.1: *Visual collateral score grading in patients with an M1 occlusion. 0—absent collaterals, 0% filling of the occluded territory. 1—poor collaterals, >0% and ≤50% filling of the occluded territory. 2—moderate collaterals, >50% and <100% filling of the occluded territory. 3—good collaterals, 100% filling of the occluded territory .*

calculated. Analysis was performed on the 4-point collateral score and on dichotomized assessments (CS 0–1: poor collaterals; CS 2–3: good collaterals). Dichotomisation was performed since treatment effects in MR CLEAN patients with good collaterals were substantial, whereas treatment effects were small in patients with poor collaterals [23].

Observer variability was reported using an intraclass correlation coefficient (ICC) using a two-way random, single measures, absolute agreement model [64]. ICC was calculated for the 4-point collateral score and a dichotomized score for all observers and for subgroups based on experience in radiological readings: junior residents, senior residents and radiologists. An ICC below 0.50 indicates poor, > 0.50 and ≤ 0.75 moderate, > 0.75 and ≤ 0.90 good, and > 0.90 excellent correlation. Accuracy for full CS and dichotomized CS was calculated for each group compared to the reference standard and qCS. Significant differences were calculated using One-way ANOVA. Statistical analyses were performed using SPSS Statistics Version 25.

3.3 Results

3.3.1 Selection of Dataset and Reference

From the 500 patients, 148 patients met inclusion criteria. For CS 0–3 (Fig. 4.1), based on image core lab evaluation, 10 CTA scans were randomly selected per collateral score. One scan could not be processed to create MIPs, which resulted in a test set with 39 cases. Fig. 4.2 shows a schematic visualization of the patient selection. For the 39 selected cases, a reference CS was obtained after rereading the scans: CS 0 (n=5), CS 1 (n=13), CS 2 (n = 10), and CS 3 (n = 11) (Table 3.1). In 31% of the cases (12/39), a third radiologist was needed to provide consensus.

3.3.2 Inter-rater Variability

The ICC is reported for the 4-point CS (Table 3.2) and for dichotomized CS (Table 3.3). All groups showed a moderate to good ICC with an ICC of 0.751 (95% CI: 0.665–0.835) for the combined results (Table 3.2). When dichotomizing CS, ICC for all observers combined decreased to 0.682 (95% CI: 0.585–0.783) (Table 3.3). No differences in ICC were demonstrated between the subgroups.

***Table 3.1:** Reference collateral score and occlusion location.*

Reference collateral score	Occlusion location			Total
	ICA-T	M1	M2	
0	1	4	0	5
1	5	7	1	13
2	5	3	2	10
3	3	7	1	11
Total	14	21	4	39

```
┌─────────────────────────────────┐
│ Patients from MR CLEAN trial     │
│ (n=500)                          │
└─────────────────────────────────┘
                                      ┌──────────────────────────────────────┐
                                      │ Excluded (n=352)                       │
                                      │ • No axial baseline CTA (n=16)         │
┌─────────────────────────────────┐  │ • Poor scan quality (n=79)             │
│ Axial scans with moderate/good   │  │ • Slice thickness > 1.0mm (n=257)      │
│ image quality and slice thickness│  └──────────────────────────────────────┘
│ ≤ 1.0mm (n=148)                  │
└─────────────────────────────────┘  ┌──────────────────────────────────────┐
                                      │ Excluded (n=108)                       │
                                      └──────────────────────────────────────┘
┌─────────────────────────────────┐
│ 10 randomly selected scans for   │
│ every collateral score (n=40)    │
└─────────────────────────────────┘  ┌──────────────────────────────────────┐
                                      │ Excluded (n=1)                         │
                                      │ • Corrupt dataset (n=1)                │
┌─────────────────────────────────┐  └──────────────────────────────────────┘
│ Multiplanar reconstructed axial  │
│ CTA scans with 8.0 mm axial and  │
│ coronal maximum intensity        │
│ projections (n=39)               │
└─────────────────────────────────┘
```

Figure 3.2: *Selection of CTA scans. CTA, computed tomography angiography; mm, millimetres .*

Table 3.2: *Correlation with reference collateral score for physician groups.*

Group (n)	4-point collateral score, intraclass correlation coefficient (95% CI)		
1st year residents (10)	0.780 (0.696–0.857)	0.775 (0.693–0.853)	0.751 (0.665–0.835)
2nd year residents (3)			
3rd year residents (4)	0.766 (0.676–0.849)		
4th year residents (4)			
5th year residents (1)			
Radiologists (7)	0.689 (0.577–0.795)		

3.3.3   Accuracy

The mean accuracy for rating CS by the 29 raters was $65\pm 7\%$, (Table 3.4). No significant differences in accuracy were found between the subgroups (Table 3.4). Accuracy increased to $88\pm 5\%$ when a dichotomized scale was used; however, the differences between subgroups remained statistically insignificant (Table 3.5). When using qCS (categorized, 0–3) as reference score, the mean overall accuracy was $67 \pm 8\%$, which increased to $88 \pm 5\%$, after dichotomization (CS 0–1; poor collaterals, CS 2–3; good collaterals). The accuracy for scoring CS was 62% for qCS, which increased

Table 3.3: *Correlation with dichotomized reference collateral score for physician groups.*

Group (n)	Dichotomized collateral score[1], intraclass correlation coefficient (95% CI)		
1st year residents (10)	0.722 (0.626–0.816)	0.698 (0.602–0.796)	0.682 (0.585–0.783)
2nd year residents (3)			
3rd year residents (4)	0.675 (0.566–0.783)		
4th year residents (4)			
5th year residents (1)			
Radiologists (7)	0.636 (0.516–0.755)		

to 90% after dichotomization of CS.

3.3.4   Individual Case Analysis

The mean visual CS (range: 0–3) per case ranged from 0.03 to 2.90. Full agreement in visual CS occurred in 2 cases (5%). Observers appointed 2 different CS in 20 cases (51%) and 3 different CS in the remaining 17 cases (44%). The qCS (range: 0–100%) ranged from 3.79 to 100%. Individual cases were sorted by ascending mean visual CS and visualized in Fig. 4.3.

3.4   Discussion

In this study, we evaluated the observer variability for visual collateral scoring and compared scores given by respondents after a 1-h training session and scores from automated software with reference scores. No difference was found between the different radiology resident groups and radiologists for performance in scoring CS. Accuracy in comparison with the reference was similar for all groups. Automated CS performs similar to residents and radiologists in determining a collateral score.

The inter-rater variability of scoring collateral circulation status has been reported before, but often this is done with Cohen's kappa [23, 65, 66]. However, using Cohen's kappa for a not-dichotomized score is harder to interpret because the differences

Table 3.4: *Collateral score accuracy for physician groups.*

Group (n)	4-point scale collateral score, accuracy (standard deviation)[1]		
1st year residents (10)	62±8%	64± 7%	65± 7%
2nd year residents (3)			
3rd year residents (4)	66± 7%		
4th year residents (4)			
5th year residents (1)			
Radiologists (7)	67±7%		
qCS(1)	62% (n.a.)		

Table 3.5: *Dichotomized collateral score accuracy for physician groups.*

Group (n)	Dichotomized collateral score[1], accuracy (standard deviation)[2]		
1st year residents (10)	87± 5%	87± 5%	88± 5%
2nd year residents (3)			
3rd year residents (4)	89± 4%		
4th year residents (4)			
5th year residents (1)			
Radiologists (7)	88± 4%		
qCS(1)	90% (n.a.)		

3
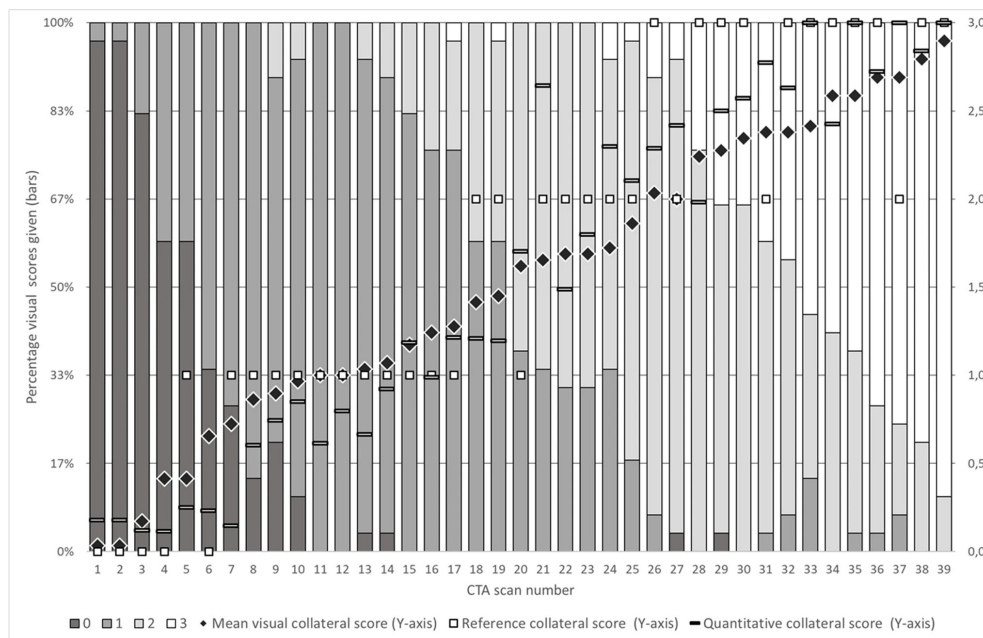
Figure 3.3: *Collateral scores separated by case and ranked on average collateral score. Black vertical bars: visual collateral score 0. Dark grey vertical bars: visual collateral score 1. Grey vertical bars: visual collateral score 2. Light grey vertical bars: collateral score 3. Black tilted square: mean visual score. White box: reference score. White bar: quantitative score.*

between scores must be weighted based on the distance between categories, which happens when using weighted Kappa or ICC. In a study by Weiss et al., weighted Kappa was provided for inter-rater reliability in scoring collateral status, but they

used only 2 readers [67]. An ICC of 0.87 for 4-point CS was given by Tan et al., but this was also based on 2 readers [18]. Using more observers would be preferable when assessing inter-rater reliability.

The automated CS shows comparable accuracy with visual raters for both the 4-point CS and the dichotomized score. There are other algorithms reported in literature for automatic CS. The research from Boers et al. presents a quantitative model which calculates the percentage of vascular presence of the occluded territory in comparison to the unaffected hemisphere [68]. A different approach was used for vessel recognition. It also showed a significant correlation (Spearman ρ: 0.75, P < .001) with the categorical visual CS (0–3) as defined by Tan et al., but an ICC, accuracy, or error matrix has not been reported [18].

Research from Grunwald et al. evaluates automated collateral score software from Brainomix Ltd. in clinical practice [58]. This software uses basic image segmentation and machine learning, both not further specified. The output is a 4-point scale collateral score. They reported an agreement of 90% and a non-specified ICC of 0.93 (95% CI 0.90–0.95) for the automated collateral score software in comparison to the reference score. However, the reference score was constructed with information on the automated collateral scores.

Collateral status can be used to predict outcome [69]. It is important to use the correct imaging to assess the collateral status for predicting outcome. Assessment of collateral status on multiphase CTA instead of single phase CTA showed a better performance in predicting outcome [70]. Optimal collateral assessment is after the peak arterial phase [71]. A limitation in the accuracy calculations is the categorical reference CS. It is difficult to reach consensus, even among experienced neuroradiologists. In 31% of the cases (12/39), a third radiologist was needed to provide consensus. Incorporating software to aid in determining collateral status can help to minimize the interobserver variability while maintaining high accuracy. For the quantitative model, the calculated CS can be given over the full range from 0 to 100% to add nuance to the score.

Radiological experience differs vastly between radiology residents and senior radiologists; however, no differences were found in accuracy of CS assessment. Based on those results, we expect other radiologists to perform similar in this setting. For the same reason, we do not expect the results from residents and radiologists to improve. Furthermore, the overall performance (ICC: 0.751, 95% CI: 0.665–0.835) is comparable to previously reported interobserver agreement for scoring collateral circulation status [18, 23, 66, 72].

We believe that achieving 65% accuracy and an ICC of 0.751 is possible for all radiology residents and radiologists after basic collateral score training. The definition of the score requires categorization of a quantitative value based on visual inspection. A difference between 90 and 20% will be clearly visible, but a difference between 45 and 55% is hard to distinguish. The difference in both situations is 1 point and questions the use of categorical scoring for collateral status. Using a quantitative scale rather than a categorical one may result in better treatment decisions. Quantitative automated CS software could be a solution, but validation is needed before integrating the software in clinical practice. The next step for validating the automatic CS would be to investigate the predictive performance: can they predict the functional outcome of patients based on the baseline scans? The modified Rankin Scale (mRS) 3 months

after acute ischemic stroke is commonly used to assess functional outcome [73]. The number of investigated scans (n=39) limits the predictive power of this study, and therefore the feasibility of analysing predictive performance. Future research using a larger sample size is recommended to investigate the correlation between automated CS and mRS in comparison to visual CS. Ideally, a study should be conducted in which both the visual CS and the automated quantitative CS are used for the same acute ischemic stroke patients to determine the performance of those collateral scores in predicting outcome and treatment benefit. Furthermore, not only the collateral score should be taken into account, also other proven predictors for outcome and/or treatment benefit should be included, combining those in a large prediction model, such as the MR PREDICTS decision tool [74]. Then, the performance of the two CS types can be compared in the prediction model.

3.5 Conclusion

On the individual rater level, there is considerable variability in rating collateral status. After a 1-h training, the accuracy of scored CS with a reference standard is not influenced by rater experience. Automated CS shows a similar performance as experienced radiologists and radiology residents. Automated CS can be an aid for physicians, especially for cases with borderline collateral scores.

Acknowledgements

Chapter 5
Time Dependency of Automatic Collateral Scores

4

Abstract

Background and Purpose: *The assessment of collateral status may depend on the timing of image acquisition. The purpose of this study is to investigate whether there are optimal time points in CT Perfusion (CTP) for collateral status assessment, and compare collaterals scores at these time points with collateral scores from multiphase CT angiography (mCTA).*

Materials and Methods: *Patients with an acute intracranial occlusion who underwent baseline non-contrast CT, mCTA and CT perfusion were selected. Collateral status was assessed using an automatically computed Collateral Ratio (CR) score in mCTA, and pre-defined time points in CTP acquisition. CRs extracted from CTP were correlated with CRs from mCTA. In addition, all CRs were related to baseline National Institutes of Health Stroke Scale (NIHSS) and Alberta Stoke Program Early CT Score (ASPECTS) with linear regression analysis to find the optimal CR.*

Results: *In total 58 subjects (median age 74 years; interquartile range 61-83 years; 33 male) were included. When comparing the CRs from the CTP vs. mCTA acquisition, the strongest correlations were found between CR from baseline mCTA and the CR at the maximal intensity projection of time resolved CTP ($r = 0.81$) and the CR at the peak of arterial enhancement point ($r = 0.78$). Baseline mCTA-derived CR had the highest correlation with ASPECTS ($\beta = 0.36$ (95%CI 0.11, 0.61)) and NIHSS ($\beta = -0.48$ (95%CI -0.72,-0.16)).*

Conclusion: *Collateral status assessment strongly depends on the timing of acquisition. Collateral scores obtained from CTP imaging do not correlate better with baseline clinical variables than collateral scores obtained from baseline mCTA imaging.*

4.1   Introduction

Ischemic stroke caused by a proximal occlusion in the intracranial cerebral arteries normally results in poor clinical outcome [75]. Endovascular treatment (EVT), in which the intracranial thrombus is removed, improves functional outcome [7]. The collateral circulation in the presence of an intracranial occlusion is an important determinant of clinical outcome and might also influence the beneficial effect of EVT [15, 23, 69, 72, 76]. However, accurate assessment of collateral status can be challenging from two perspectives. First, the current existing semi-quantitative collateral grading systems are based on visual scoring using coarse classification criteria [24, 77] which are subject to inter-observer variation. Automatic analysis could potentially overcome this limitation [21, 37, 58]. Secondly, the collateral grading is affected by the timing of image acquisition in relation to the start of contrast media injection [78]. For an accurate prediction of treatment outcome and clinical decision making it is relevant to know which acquisition time is optimal for assessment of collateral score.

For collateral status assessment, single-phase Computed Tomography Angiography (sCTA) is commonly used [77, 79, 80]. sCTA comes with the risk of incorrect collateral grading due to inappropriate scan timing [76] . To avoid this potential bias caused by incorrect timing, additional temporal CTA acquisitions can be added to the sCTA. The two most well-known imaging techniques are multi-phase CTA (mCTA) and CT perfusion (CTP). Multiphase CTA consists of two extra acquisitions with short delays of 6-8s between the acquisitions [70]. CTP consists of a series of acquisitions with 1.5 to 2 seconds interval and a total length of approximately 1 minute. In previous studies, a visual collateral scoring system based on mCTA was proven to be superior to sCTA in clinical outcome prediction [70, 78].

In this study, we applied a previously developed automated collateral scoring algorithm for CTA data also to CTP data and investigated how collateral status assessment depends on the timing of image acquisition. Additionally, we extracted collateral scores at specific time points from the CTP data, and investigated how these CTP collateral scores relate to mCTA collateral scores. Finally, we assessed the association of collateral scores from CTP and mCTA with baseline ASPECTS [19] and NIHSS [81].

4.2   Materials and Methods

4.2.1   Patient Selection

We used data from acute ischemic stroke patients with a large vessel occlusion from Erasmus MC, who were included in the Multicenter Randomized Clinical Trials of Endovascular Treatment of Acute Ischemic Stroke (MR CLEAN LATE, MR CLEAN NOIV or MR CLEAN MED studies from January, 2018 to March, 2020). Detailed information on inclusion criteria, description of variables and the methods of the three trials has been reported previously [82, 83, 84]. For this study, we selected patients with the following inclusion criteria: an intracranial occlusion on one side only, baseline imaging that contains both non-contrast enhanced CT (NCCT), mCTA and CTP; slice thickness $\leq$ 1.5 mm for both mCTA and CTP; complete CTP acquisition ($>$ 55

seconds), and absence of large motion artifact(s) in both mCTA and CTP. Neurologic deficit at baseline was assessed with the National Institutes of Health Stroke Scale (NIHSS; range 0-42, with higher scores indicating more severe neurologic deficits).

4.2.2 Imaging Protocol

All included subjects underwent a standardized imaging protocol for acute ischemic stroke. This protocol consists of NCCT, CTP and mCTA (Siemens SOMATOM Definition Edge). The CTP image acquisition was performed with a fixed tube voltage of 70 kVp and an injection of 40 mL of Iomeron 400 with a saline bolus chaser at a flow rate of 6 mL/s. The CTP images were reconstructed with a slice thickness of 1-1.5 mm. A CTP acquisition had a total duration of 55s to 60s and consisted of $30(\pm 1)$ volumes. After the CTP, the mCTA acquisition was performed with an automatic tube voltage of 80-100 kVp, injection of 60 ml of Iomeron 400 with a 40 ml saline bolus chaser at a flow rate of 4 ml/s. Timing of scanning was based on bolus tracking. The mCTA consists of three image volumes acquired with delays between the volumes of 8 seconds on average. The first acquisition of mCTA was defined as the mCTA baseline. All CTA images were reconstructed with a slice thickness of 0.75 mm and an increment of 0.4 mm.

4.2.3 Automatic Collateral Scoring and Arterial and Venous Enhancement Curves

To obtain a consistent location for arterial and venous structures and enhancement measurements over time, all CTP image volumes were rigidly aligned to the first acquisition using registration software ANTs [54, 55]. The automatic collateral scoring algorithm was developed by [21] with collateral ratio (CR) as output (expressed in percentage %). The proposed CR ranges from 0 to 150. In this method, the middle cerebral artery (MCA) region was defined with a two-step atlas based registration using ANTs [54, 55] with initial rigid registration and a subsequent diffeomorphic non-rigid registration. Vessels were identified using a U-net based vessel segmentation algorithm. The enhanced vessel volume in the occluded and contralateral MCA region was computed. In the last step, the CR was calculated as enhanced vessel volume in the MCA region of the occluded side divided by the contrast enhanced vessel volume in the healthy side.

Arterial and venous enhancement in the CTP and mCTA images was obtained by annotating a small region of interest at the top of the intracranial carotid artery (ICA) (arterial) of the non-occluded side and the confluence sinus (venous). The arterial and venous enhancement value were determined using the median image intensity of a spherical region at the locations mentioned above.

4.2.4 NCCT Analysis

Baseline Alberta Stoke Program Early CT Scores (ASPECTS) [19] were assessed on baseline CT images. ASPECTS was rated by radiologists with 5 to 20 years of

experience. The raters were blinded to all patient information during ASPECTS evaluation, except occlusion side on baseline CTA.

4.2.5 CTP Time Intensity Curve and Feature Extraction

To investigate the temporal relationship of enhanced vessel volume, CR, and arterial and venous enhancement in CTP images, four types of time-intensity curves (TIC) were generated: contrast enhanced vessel volume in the occluded and contralateral side (Fig. 4.1a), CR (Fig. 4.1b), arterial and venous enhancement (Fig. 4.1c) and arterial to venous (AV) enhancement ratio (Fig. 4.1d). In the AV enhancement ratio curve generation, an average background density value of 40 HU was subtracted from the arterial and venous density values before calculating the ratio.

From these curves, the following time points were defined to extract a CR: the peak of vessel volume in the non-occluded side (CR_{vol_max}); the peak of vessel volume in the occluded side ($CR_{vol\_occluded\_max}$); the peak of arterial enhancement ($CR_{arterial\_max}$); the peak of venous enhancement (CR_{venous_max}); the point where the arterial enhancement equals venous enhancement (CR_{equi}) and the peak of AV enhancement ratio (CR_{av_max}). Also, the time between contrast arrival time and these time points was calculated. In addition, in CTP images, a time resolved variable CR_{mip} was generated using the maximal intensity projection (MIP) over time.

4.2.6 mCTA Image Feature Extraction

For the three mCTA volumes, the CR was computed using the automatic collateral scoring method whereas the arterial and venous enhancement values were obtained with manual annotation. The AV enhancement ratio was obtained using the same approach as described for the CTP images.

4.2.7 Alignment of mCTA Acquisitions to the CTP and Across Subjects

To be able to relate the CR in mCTA to the CRs derived from CTP, and to investigate the contrast acquisition phase of mCTA volumes, the mCTA images were temporally aligned to time points in the CTP acquisition in the same patient. This temporal alignment method was based on the AV enhancement ratios from the mCTA and the CTP images, using the timing information on the delays between the three mCTA volumes. We determined the optimal temporal alignment by minimizing the overall absolute difference of AV enhancement ratio in the mCTA images and the three corresponding time points in the CTP images. In addition, normalized AV enhancement ratio curves for all subjects were aligned at the arterial peak and the mCTA acquisitions were aligned accordingly. Subsequently, the acquisition phase of the each mCTA image was determined based on the CTP-based acquisition phase definition as described below. Fig. 4.2 shows the image features of paired mCTA and CTP acquisitions
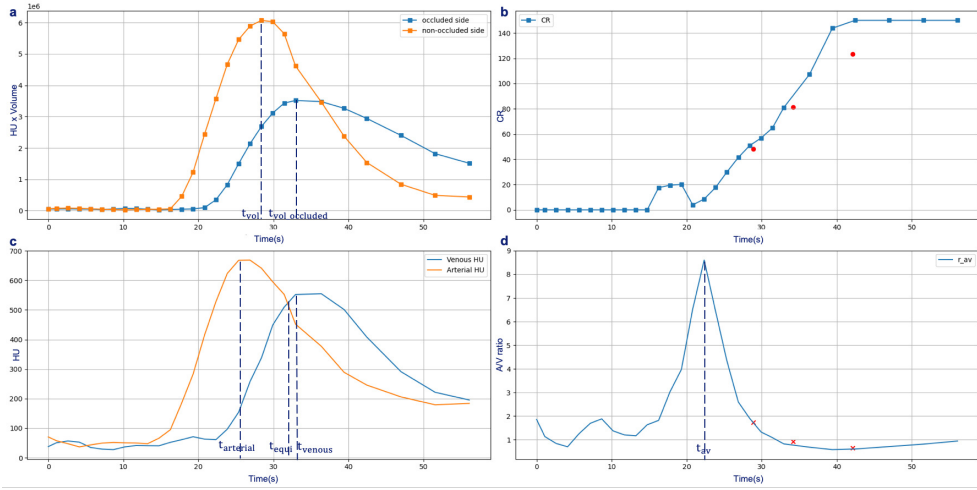
Figure 4.1: *An example feature plot. a. the contrast volume changing over time for both occluded side and contralateral side; b. the collateral ratio changing over time, three red dots are the corresponding collateral ratios of the mCTA; c. the arterial and venous enhancement curve; d. the arterial to venous (AV) ratio curve, the three crosses stand for the alignment points based on AV ratios of the three mCTA acquisitions.*

4.2.8   CTP-based Contrast Acquisition Phase Definition

A contrast acquisition phase was defined based on the patient specific normalized AV enhancement ratio curve. The normalization was done by offsetting the equilibrium point (where AV enhancement ratio = 1) to 0 and dividing the result by the maximum value. Fig. 4.3 shows the effect of this normalization for all included subjects, and also shows the positions of the temporarily aligned mCTA images. Subsequently, the contrast acquisition phase was defined using this curve as follows: early arterial phase: upslope in the ratio curve from 0.2 to 1.0; peak arterial phase: downslope from 1.0 to 0.5; equilibrium phase: downslope from 0.5 to 0; early venous phase: 0 to minimum ratio; late venous phase: minimum ratio to 0.

4.2.9   Statistical Tests

All values are reported as median and interquartile range (IQR). The Spearman correlation coefficient was used to analyze the relationship between the different CRs. To test the clinical relevance of choosing the optimal CR we assessed the association between the CRs and baseline ASPECTS and NIHSS with Spearman correlation coefficients and linear regression analysis. All statistical tests were performed using SPSS, IBM (version 26).
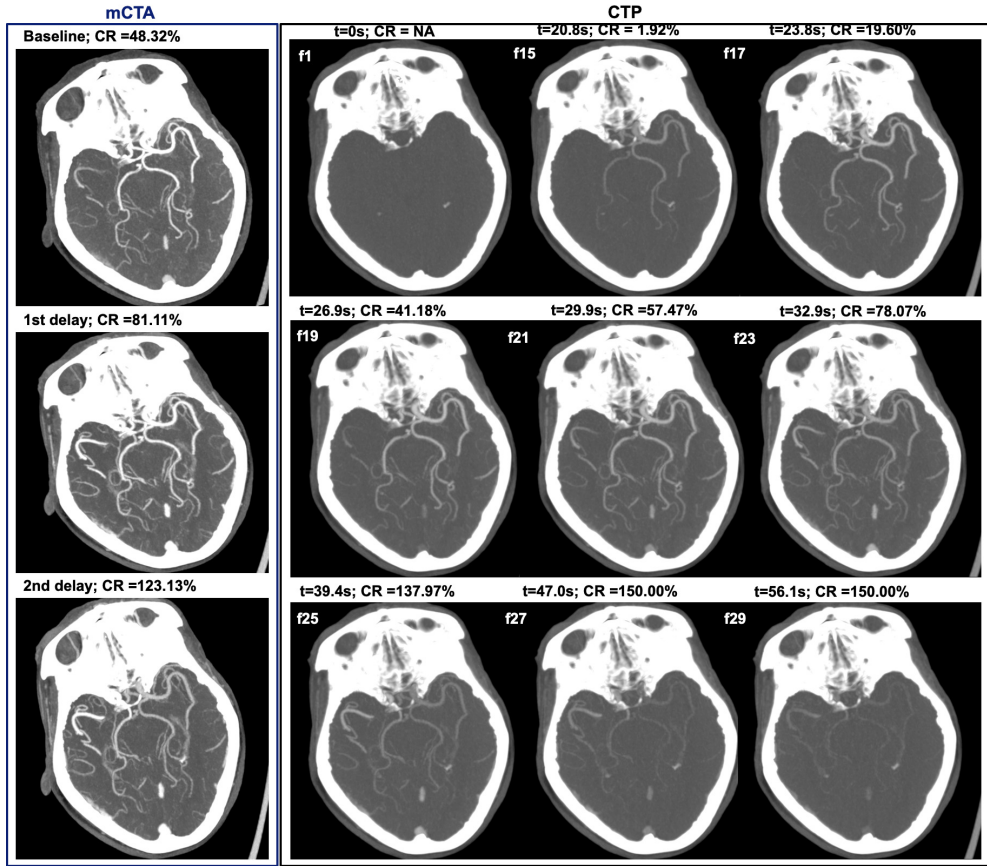
Figure 4.2: *The example image of CR changing over time on both mCTA and CTP. On CTP part, 'f' indicates the frame number of CTP images, 't' indicates the relative acquisition time. In this example, mCTA baseline corresponds approximately to f19 in CTP; mCTA delay 1 corresponding to f23; mCTA delay 2 corresponding to f25.*

4.3   Results

4.3.1   Patients Characteristics

In total, 335 patients were enrolled in Erasmus MC since January 2018 to March 2020. Baseline imaging was performed in the Erasmus MC on 106 patients. A CTP with slice thickness less or equal to 1.5 mm was available for 81 patients. Twenty-three patients were subsequently excluded because of vessel occlusions on both sides (n=1), an incomplete CTP acquisition (n=1), no baseline mCTA images (n=7), large motion artifacts in the CTP acquisitions (n=12), an NIHSS of 42 (n=1) and baseline mCTA
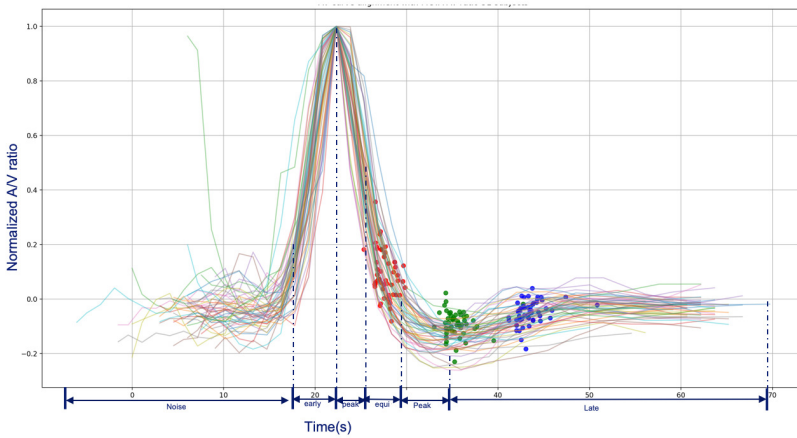
Figure 4.3: *The result of mCTA to CTP alignment and contrast phase definition. Red dots represent the mapped mCTA baseline; green dots represent the mapped mCTA delay 1; blue dots represent the mapped mCTA delay 2. Each line represents a CTP image series.*

with DICOM header error (n=1). The remaining 58 patients were included in this study. The median age was 74 years (IQR, 61-83 years), 33 were male, the median ASPECT score was 9, (IQR, 8-10) and the median baseline NIHSS was 16 (IQR 10-20). Fifteen patients had an occlusion in the intracranial ICA. The other patients had a middle cerebral artery (MCA) occlusion:27 patients had a M1 segment occlusion, 15 patients had a M2 occlusion and the remaining patient had an M3 occlusion.

4.3.2 Timing of Acquisition and Collateral Status

The CR of the mCTA and CTP images at the predefined time points, and the corresponding time delay since contrast arrival are shown in Table 4.1. Huge differences exist between CR from CTP images obtained early after contrast arrival at the peak of CTP AV enhancement ratio (median 26%, IQR 12-52) and late in the venous phase at the CTP peak of venous enhancement (median 128%, IQR 88-171). The CR in the second delay mCTA volume had a median of 160% (IQR 119-241), indicating arrival of contrast via collateral vessels on the occluded side and diminished contrast in the healthy side due to normal blood flow.

The Spearman correlation coefficients between all CRs are shown in Table 4.2. The highest correlation between the CRs extracted from the CTP acquisition exists between the CR at the peak of contrast volume in the non-occluded side and the CR at the peak of arterial enhancement (r=0.96; Fig. 4.4a). No correlations exist between the CRs in the mCTA acquisitions (Table 4.2 and Fig. 4.4b-c). When comparing the CRs from the CTP vs. mCTA acquisition, the highest correlations were found

Table 4.1: *The characteristic of predefined Collateral Ratios (CR) and the corresponding time points in the aligned A/V curve.*

	Timepoint(in sec) Median [IQR]	CR(%) Median [IQR]	Observations in the acquision phases (n)				
			EA[1]	PA[2]	EQ[3]	PV[4]	LV[5]
Vessel volume peak at non-occluded side (*vol_max*)	7.6 [6.1, 7.6]	53 [39, 74]	-	5	53	-	-
Vessel volume peak at occluded side (*vol_occluded_max*)	9.1 [7.6, 10.6]	83 [66, 102]	-	-	26	31	1
Arterial peak (*arterial_max*)	6.4 [4.5, 6.1]	44 [31, 67]	-	27	31	-	-
Equilibrium (*equi*)	9.2 [8.3, 10.2]	78 [57, 92]	-	-	58	-	-
Peak venous (*venous_max*)	12.1 [10.6, 13.6]	121 [88, 171]	-	-	-	56	2
r_{av} peak (*av_max*)	3.0 [1.5, 3.0]	26 [12, 52]	7	51	-	-	-
mCTA baseline	8.1 [7.3, 9.0]	71 [46, 90]	2	8	48	-	-
mCTA delay 1	15.7 [14.9, 16.7]	125 [86, 183]	-	-	2	23	33
mCTA delay 2	23.8 [22.9, 24.8]	160, [119, 241]	-	-	-	3	55

between CR from baseline mCTA and the time resolved CR (r=0.81) and the CR at the peak of arterial enhancement point (r=0.78). The median time difference between baseline mCTA and arterial peak was 2.7s, IQR [1.7, 3.2].

4.3.3 mCTA to CTP Time Point Alignment and Phase Classification

Fifty-six mCTAs out of 58 subjects were automatically temporarily aligned to the CTP acquisition based on the AV enhancement ratio curve. The two cases that failed were aligned manually. Fig. 4.3 shows the aligned AV enhancement ratio curves with the temporally aligned mCTAs. The median of arterial to venous enhancement ratio difference from three acquisitions of mCTA to three time points of CTP was 0.2 (AV ratio ranged from 0 to 20). The median of the CR difference between baseline mCTA and temporally aligned CTP was 6.4% (IQR 2.4-12). The median of absolute CR difference between 1st delay mCTA and temporally aligned CTP was 16.5% (IQR 0-30). The median of absolute CR difference between 2nd delay mCTA and temporally aligned CTP was 0.2% (IQR 0-19). The acquisition phase of all predefined CRs are reported in Table 4.1. The acquisition phase of the CRs differ significantly from each other. Multiphase CTA baseline images were mainly acquired in the equilibrium phase (n=48, 83%).

4.3.4 Association of CRs and Baseline Clinical Parameters

Table 4.3 and Table 4.4 list the correlations between the CRs at the proposed time points and baseline ASPECTS and NIHSS. The strongest correlation with baseline ASPECTS was found for CR determined from baseline mCTA ($\beta = 0.36$ (95%CI 0.11, 0.61)) and on time resolved CTP ($\beta = 0.33$ (95%CI 0.07, 0.58)), see (Fig. 4.5a-b). The strongest association with baseline NIHSS was found for CR at baseline mCTA ($\beta = -0.48$ (95%CI -0.72, -0.16)) and CR at volume peak of non-occluded side ($\beta = -0.40$ (95%CI -0.65, -0.16)), see (Fig. 4.5c-d).

4

Table 4.2: Correlation matrix of 10 proposed Collateral Ratios.

	$CR_{vol\_occluded\_max}$	$CR_{arterial\_max}$	CR_{mip}	CR_{equi}	$CR_{venous_{max}}$	CR_{av_max}	$CR_{mCTA\_baseline}$	$CR_{mCTA\_delay\_1}$	$CR_{mCTA\_delay\_2}$
CR_{vol_max}	0.55	0.96	0.86	0.83	0.15	0.82	0.73	0.18	-0.15
$CR_{vol\_occluded\_max}$	-	0.56	0.80	0.63	0.36	0.33	0.71	0.44	-0.23
$CR_{arterial\_max}$	-	-	0.85	0.81	0.14	0.85	0.78	0.20	-0.19
CR_{mip}	-	-	-	0.85	0.30	0.61	0.81	0.41	0.13
CR_{equi}	-	-	-	-	0.42	0.61	0.76	0.46	0.08
$CR_{venous_{max}}$	-	-	-	-	-	-0.06	0.36	0.77	0.38
CR_{av_max}	-	-	-	-	-	-	0.60	0.03	-0.31
$CR_{mCTA\_baseline}$	-	-	-	-	-	-	-	0.59	-0.04
$CR_{mCTA\_delay\_1}$	-	-	-	-	-	-	-	-	0.45
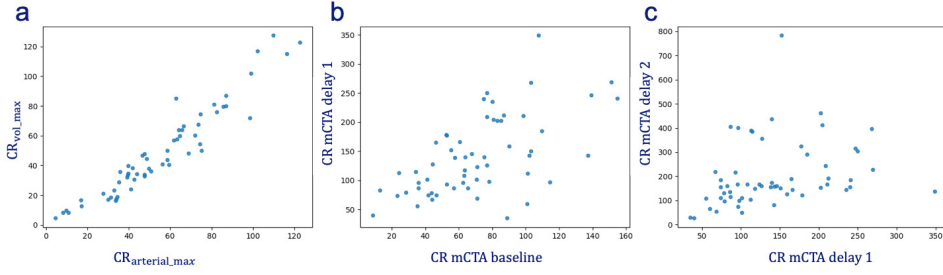
Figure 4.4: *The correlation scatter plot. a.shows the correlation between the CR at the peak of contrast volume in the non-occluded side (CR_{vol_max}) and the peak of arterial enhancement ($CR_{arterial\_max}$); b.The correlation between CR of mCTA baseline and delay 1. c. The correlation between CR of mCTA delay1 and mCTA delay 2.*
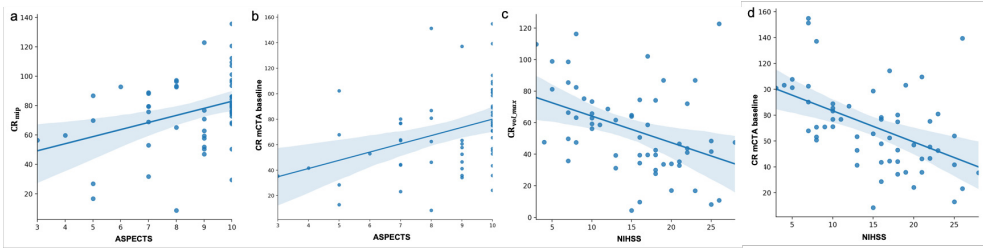
Figure 4.5: *The linear regression plots. a. shows the linear association between time-resolved CR (CR_{mip}) and ASPECTS. b. shows the linear association between CR of baseline mCTA and ASPECTS. c.shows the linear association between CR at the volume peak of the non-occluded side (CR_{vol_max}) and baseline NIHSS. d. shows the linear association between CR of baseline mCTA and baseline NIHSS..*

4.4 Discussion

In this study, we investigated the time dependency of collateral assessment in patients with an intracranial arterial occlusion. We defined objective CTP time points for CR extraction based on temporal curves of contrast enhanced vessel volume in the occluded and contralateral side, arterial enhancement and venous enhancement. We demonstrated that these CRs were significantly different from each other with an increase in CR in subsequent image volumes acquired during CTP. We subsequently compared the mCTA derived CRs with the CRs obtained at predefined time points in the CTP and demonstrated that baseline mCTA correlates best with the predefined CTP CR at maximum arterial enhancement. To evaluate the potential significance

Table 4.3: *The association of the predefined collateral ratios with ASPECT score.*

CR variable (without threshold)	Spearman Correlation r (p-value)	Linear Regression β (95%CI)
CR_{vol_max}	0.37 (<0.01)	0.32 (0.07, 0.58)
$CR_{vol\_occluded\_max}$	0.31 (<0.01)	0.24 (-0.03,0.50)
CR_{mip}	0.36 (<0.01)	0.35 (0.10, 0.60)
$CR_{arterial\_max}$	0.38 (<0.01)	0.33 (0.07, 0.58)
CR_{equi}	0.34 (<0.01)	0.28 (0.02, 0.53)
CR_{venous_max}	0.01 (0.46)	0.06 (-0.21,0.33)
CR_{av_max}	0.30 (0.01)	0.26 (0.05, 0.52)
$CR_{mCTA\_baseline}$	0.45 (<0.01)	0.36 (0.11, 0.61)
$CR_{mCTA\_delay\_1}$	0.17 (0.1)	0.20 (-0.06,0.46)
$CR_{mCTA\_delay\_s}$	-0.11 (0.22)	0.10 (-0.17,0.36)

Table 4.4: *The association of the predefined collateral ratios variable with NIHSS baseline.*

CR variable (without threshold)	Spearman Correlation r (p-value)	Linear Regression β (95%CI))
CR_{vol_max}	-0.40 (<0.01)	-0.40 (-0.65, -0.16)
$CR_{vol\_occluded\_max}$	-0.24 (0.17)	-0.12 (-0.39, 0.14)
CR_{mip}	-0.42 (<0.01)	-0.37 (-0.62, -0.12)
$CR_{arterial\_max}$	-0.44 (<0.01)	-0.39 (-0.64, -0.15)
CR_{equi}	-0.37 (<0.01)	-0.29 (-0.55, -0.04)
CR_{venous_max}	-0.07 (0.29)	-0.02 (-0.29, 0.25)
CR_{av_max}	-0.41 (<0.01)	-0.32 (-0.58, -0.07)
$CR_{mCTA\_baseline}$	-0.53 (<0.01)	-0.48 (-0.72, -0.25)
$CR_{mCTA\_delay\_1}$	-0.18 (0.10)	-0.18 (-0.44, 0.09)
$CR_{mCTA\_delay\_2}$	0.36 (<0.01)	0.35 (0.09, 0.60)

of optimal CR assessment we related all CRs to baseline ASPECTS and NIHSS and demonstrated that baseline mCTA-derived CR had the highest associations.

Collateral status assessed on baseline CTA is an important parameter in clinical decision making in patients with ischemic stroke due to intracranial large vessel occlusion and is an important predictor of outcome. Accurate, robust and reproducible assessment of CR is therefore crucial. The dependency of CR on the timing of image acquisition in relation to the contrast injection could be worrisome. This is especially true when automated analysis tools are able to provide a CR on a continuous scale instead of a semi-quantitative scale with a restricted number of grades (n=4) as is the case with the current visual analysis. A CR could be two times larger when the CTA is acquired several seconds later. A CR could even be > 100% when the CTA is acquired in the venous phase. A previous study reported a higher prevalence of lower collateral grades in CTA scans acquired in the early arterial phase[4]. The lack of temporal resolution in sCTA, resulting in a highly susceptible to erroneous assessment of CR, has been identified before [72] and multiphase CTA has been proposed as reliable tool to provide a time-resolved collateral assessment [70]. With three CTAs in three

different phases one may wonder which acquisition should be used for CR assessment, the authors solved this issue by proposing a new composite semi-quantitative scale with 6 grades which integrates information from the three phases in the mCTA. This collateral score for mCTA is different from the CR in our study, in which we assessed the quantitative CR per mCTA phase. Tong et al. have extracted one CR from CTP data after temporal fusion of the CTP images creating a time resolved CR. This approach does not take into account the between patient variation in collateral recruitment which could have clinical consequences [85]. In our study, we reproduced this approach by generating an extra time-resolved CR_{mip}.

The relevance of optimal timing of CTA for extraction of diagnostic information has also been reported for spot-sign detection in patients with intracerebral hemorrhage [6]. The authors proposed a classification for different acquisition phases, ranging from early arterial to late venous phase. The phase definition was based on absolute enhancement, expressed in Hounsfield units (HU), in arterial and venous structure. The dependence of HU on the voltage of the CT X-ray tube, and the dependence of the iodine concentration on the injection protocol (amount of contrast, concentration and injection speed) as well as hemodynamic factors [86], makes this definition of the phases rather arbitrarily. We proposed a phase classification based on a normalized arteriovenous enhancement ratio curve and demonstrated that most patients have identical normalized curves allowing a more robust assessment of acquisition phase for sCTA, mCTA and CTP image volumes.

We investigated in which phases the mCTAs were acquired and temporarily aligned the three mCTA volume to the CTP volumes based on the arteriovenous enhancement ratio. Baseline mCTA was acquired mainly in the equilibrium phase, close to maximum arterial enhancement and therefore, the CR of baseline mCTA was highly correlated to the CR at maximum arterial enhancement. This consistent acquisition in the equilibrium phase, close to the maximum arterial enhancement is probably caused by the use of bolus tracking for timing of CTA in relation to contrast arrival and means that CR can be assessed reproducible from sCTA or mCTA, timed with bolus tracking.

The remaining question is which CR correlates best with the real status of brain perfusion and is optimal for clinical decision making and outcome prediction. We demonstrated that several CRs are related to each other and could replace eventually each other. Others are less well correlated indicating that potentially more predictive information is available in the CTP derived CRs. To evaluate the potential significance of optimal CR assessment we determined the correlation between the CRs at the chosen time points to baseline ASPECTS and NIHSS and demonstrated that baseline mCTA-derived CR had the highest associations and was not outperformed by the other CRs. This could be explained by consistent acquisition of mCTA close to the peak of arterial enhancement.

Previous studies have compared the predictive power of mCTA-derived CR with sCTA-derived CR. Schreder et al., who extracted mCTA volumes from a CTP acquisition could not demonstrate a better outcome prediction compared to sCTA [87]. However, Wang et al. [78] investigated outcome prediction using single phase and multiphase CTA images, also extracted from CTP acquisition. Considering the potential variation in hemodynamics between patients, the baseline mCTA and first delay

mCTA was selected at the peak arterial phase and peak venous phase. The second delayed mCTA phase was defined by a time interval of 5.5 seconds after peak venous phase. They concluded that the mCTA composite collateral score performed better than single-phase collateral score in determining clinical outcome for patients with acute ischemic stroke. However, it is difficult to compare since the acquisition time of the single phase CTA in comparison with mCTA images was unknown.

This study has several limitations. First, the dataset size is relatively small, and was collected in a single center. Further assessment using multi-center data may be performed to validate the conclusions from this study. Second, we use an automatic collateral scoring method that was developed for CTA on CTP images. The scoring method yields a relative value, where the quantification of the occluded side is normalized with the non-occluded side. We have assumed that this method is equally valid on CTP images, which may have a lower signal to noise ratio and contrast intensity compared to mCTA images. Last, we used CTP images to investigate the best time-point for collateral scoring, based on associations with other baseline parameters. This study could not demonstrate a better association with baseline clinical (NIHSS) and imaging parameters (ASPECTS) for one of the CTP derived CRs over the mCTA derived CR. The small size of the study hampers strong conclusions. Finally, the choice for a specific CR should be based on an improved prognostic or predictive value, which should take into account also other predictors. Such an evaluation requires a large clinical study.

4.5   Conclusion

We conclude that collateral status assessment is dependent on the timing of acquisition. mCTA with bolus tracking results in a consistent acquisition in the same phase close to the peak of arterial enhancement. This consistent acquisition in the equilibrium phase makes CTA derived CR assessment a robust and reproducible approach. Although CTP can provide other CRs, the clinical value and better predictive value has not been demonstrated yet and should be assessed in larger studies focusing on the prediction of outcome.

Acknowledgments

Chapter 7

Cerebral Anterior Vessel Tree Extraction

Abstract

Extracting the cerebral anterior vessel tree of patients with an intracranial large vessel occlusion (LVO) is relevant to investigate potential biomarkers that can contribute to treatment decision making. The purpose of our work is to develop a method that can achieve this from routinely acquired computed tomography angiography (CTA) and computed tomography perfusion (CTP) images.

To this end, we regard the anterior vessel tree as a set of bifurcations and connected centerlines. The method consists of a proximal policy optimization (PPO) based deep reinforcement learning (DRL) approach for tracking centerlines, a convolutional neural network based bifurcation detector, and a breadth-first vessel tree construction approach taking the tracking and bifurcation detection results as input. We experimentally determine the added values of various components of the tracker. Both DRL vessel tracking and CNN bifurcation detection were assessed in a cross validation experiment using 115 subjects. The anterior vessel tree formation was evaluated on an independent test set of 25 subjects, and compared to interobserver variation on a small subset of images.

The DRL tracking result achieves a median overlapping rate until the first error (1.8mm off the reference standard) of 100, [46, 100] % on 8032 vessels over 115 subjects. The bifurcation detector reaches an average recall and precision of 76% and 87% respectively during the vessel tree formation process. The final vessel tree formation achieves a median recall of 68% and precision of 70%, which is in line with the interobserver agreement.

5.1   Introduction

5.1.1   Clinical Background

The cerebral vessel network is a complex network that feeds the brain tissue. Diseases affecting the cerebral vessel network, such as an intracranial ischemic stroke caused by a large vessel occlusion, may have severe consequences. In patients with an intracranial large vessel occlusion (LVO), the occlusion often occurs in the anterior circulation, which consists of the middle cerebral artery (MCA) vessel tree and anterior cerebral artery (ACA) vessel tree. Knowing the precise anterior vessel tree with its topology could provide additional information for the treatment planning of LVO and potentially contribute to the treatment decision making process. The complete cerebral vessel centerline annotation with anterior vessel trees is shown in Fig. 5.1.

In clinical practice, 3D modalities such as magnetic resonance angiography (MRA) and computed tomography angiography (CTA) are commonly used to visualize cerebral vasculature. In this manuscript, we are focusing on CT-based imaging, which is a common modality for the workup of stroke patients. CT imaging techniques for cerebral vessels in clinical practice are CTA (both single phase and multiphase) and computed tomography perfusion (CTP) ([88]). CTP is an imaging protocol where a series of 3D CTA images is acquired.

The purpose of our work is to extract the anterior vessel trees as distal as possible given the initial direction vectors of each tree.

5.1.2   Related Work

Traditionally, vessel tree extraction starts with obtaining a vessel segmentation or geometric model ([89]). Extensive reviews on vessel segmentation and vessel shape extraction methods have been presented by [25], [26], and [31]. After obtaining a vessel segmentation from the input image, various methods can be used to obtain the vessel tree. [48] used a thinning algorithm, [90] used geodesics in combination with a minimum spanning tree, [30] used a graph-based tracking approach to obtain the Circle of Willis and the proximal segment of each arterial trees from 3D MRA images and [91] utilize the general confluence constraint with minimal arborescence on a directed graph. Others build the tree in a more explicit way from a set of bifurcations and connected centerlines, an approach followed by [92] and [93]. In these approaches, vessel tracking and bifurcation detection are two essential ingredients.

Tracking

In conventional methods, vessel segmentation approaches start with centerline tracking. Many tracking methods have been described in the previously mentioned reviews. Nowadays, deep learning based methods have replaced conventional approaches for many image processing problems. Also, combinations of convolutional neural networks (CNNs) and conventional approaches have become popular. For example, [94] extract the coronary artery centerline with an iterative tracking approach utilizing a 3D CNN to estimate the orientation and radius from cardiac CT angiography, [95] use a multi-task CNN method to obtain an estimate of a centerline distance map and
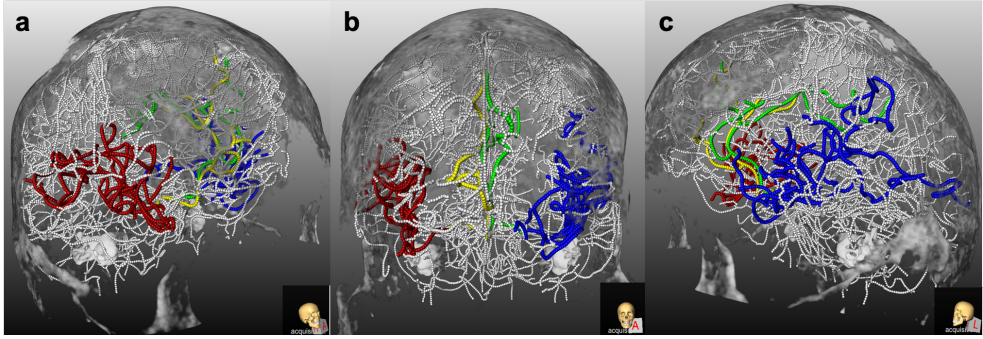
Figure 5.1: *Example of brain vessel annotation with colored anterior vessel trees for a subject with a large vessel occlusion in the left M2 segment. White represents veins and posterior cerebral artery trees, red is the right MCA tree, yellow is the right ACA tree, green is the left ACA tree, blue is the left MCA tree, a: the right sagittal view; b: the coronal view; c: the left sagittal view.*

an endpoint confidence map from coronary CTA images and subsequently a minimal path method is used to compute the centerline.

Reinforcement learning is a branch of machine learning algorithms that have been studied in domains such as gaming and control. Deep reinforcement learning (DRL), which is the fusion of reinforcement learning and deep learning, has been shown to be able to solve complex image processing challenges by exploiting image features ([96]). Such DRL based methods are gradually entering the medical imaging field ([97]). Tracking is a preferred application direction for DRL approaches, as DRL learns the optimal policy by maximizing the accumulated reward over time, which maps nicely to the problem of finding a path through an nD image. The optimal policy determines the best actions for an agent in each state over time.

There are two main streams in DRL methods, value-based and policy gradient based. In value-based methods (e.g. deep Q network ([98]) the policy π is directly obtained by approximating the state-action value $Q_\theta(s, a)$ as a function of state s and action a or state value $V_\theta(s)$ as a function of a state s. Here, θ represents the network parameters. For example, [99] apply a deep Q network ([98]) (DQN) approach with a discrete action space given by a six-connected neighborhood as action space and a point-to-curve reward function for thoracic aorta centerline tracking in 3D contrast-enhanced and none contrast-enhanced CT images. [92] similarly utilizes a double-DQN ([100]) with a discrete 26-connected neighborhood as the action space to track the coronary centerline. In this work, the bifurcations were detected using a standalone detector. [101] utilize the DQN approach with predefined orientation as discrete action space to track coronary centerlines. In addition, the value-based methods are known to often lead to an overestimation of the action value ([100]).

In policy gradient methods, a parameterized policy is learned directly using a stochastic optimization algorithm ([102]). Trust region policy optimization (TRPO) ([103])

and proximal policy optimization (PPO) ([104]) are two well-known methods for such optimization. [105] utilize PPO for learning the policy for neuronal tracking in 2D neuronal microscopy images. The above-mentioned DRL-based tracking approach has demonstrated superior performance in comparison with a benchmark using CNN based methods.

Bifurcation Detection

Bifurcation detection plays an important role in determining the vessel tree. Some feature based machine learning methods have been presented in the past. For example, [106] utilized the scale-space features of vessel structures to construct a bifurcation detector on 2D and 3D synthetic images. [107] used Gaussian profiles in cross sections of bifurcation points to discriminate vessels from bifurcation points in 2D retinal images. [93] performs bifurcation detection using intensity based spatial clustering methods for 3D coronary arteries extraction from CTA images. Recently, deep learning methods have been presented by [108], who utilized a two-stage CNN approach to detect carotid artery bifurcations from 3D CTA images, and [109], who used a multi-instance heat map with a U-net architecture ([42]) to estimate vessel crossings and bifurcation points from 2D retina images. The latter application was also addressed by [110] utilizing a two-stage refinement approach, in which an RCNN ([111]) based network first obtained a rough estimation of bifurcations and in a 2nd stage, the same network architecture was applied to address the errors from the first network.

5.1.3   Contribution and Organization of Our Work

Our aim is to extract the anterior vessel tree from 3D CT images, independent of the acquisition protocol. For such a task, to the best of our knowledge, most vessel tree extraction methods and the subsequent tracking and bifurcation steps have focused on coronary artery tree extraction from 3D cardiac CTA images and vessel tree extraction from 2D retina scanning laser ophthalmoscope (SLO) images. None of the existing works presents a method to extract the cerebral anterior vessel tree from 3D CTA images or methods that perform tree extraction beyond the proximal anterior vessel tree.

Our work has several contributions. First, we assess a DRL approach that, to the best of our knowledge, has not been applied yet for cerebral vessel tracking approaches. Second, the DRL method was adapted to this application. More specifically, we introduce a novel curve-to-curve distance based reward function and a network architecture (CNN with recurrent neural network (RNN)) for cerebral vessel tracking, and we assess the impact of various features of the methods in an ablation study. In addition, we performed a comparison with an existing baseline method. Finally, all experiments have been performed with a large set of clinical data.

The remainder of this manuscript is organized as follows. In Section 5.2, we first present a DRL PPO based directed vessel centerline tracking model, followed by a CNN-based bifurcation detection, and breadth-first tree formation methods. The data set, data annotations, and prepossessing are described in Section 5.3. The implementation, hyperparameter optimizations and experimental results can be found in Section 5.4, followed by discussion in Section 5.5 and conclusions in Section 5.6.

5.2   Method

The proposed cerebral anterior vessel tree extraction method starts with a DRL centerline tracking approach starting from the root of the tree (with an initial direction vector). Along the tracked path (agent path), a CNN based bifurcation detector is used to identify the bifurcation points. New tracks are generated from the bifurcation points detected along the tracks. In the tree formation, stopping criteria are applied to remove the spurious parts of the tracked paths, and a breadth-first approach is used to construct the tree from the tracked paths and the bifurcation points.

5.2.1   Directed Vessel Tracking

A reinforcement learning problem can be modeled as a Markov Decision Process (MDP) ([112]). We, therefore, introduce the key ingredients of our DRL tracker as the components of an MDP. The environment in our case is the normalized 3D CTA image. The other components, such as state and action, are introduced below.

State and Action

The agent at time t has an associated state, denoted as s_t. Next to the current position in the image, the agent state contains image information from previous steps and the corresponding displacement vector array. Let p be any discrete 3D position in a CTA volume. The image information in the state consists of three $21 \times 21 \times 21$ sub-volumes centered at respectively the agent position at p_t and two previous agent positions p_{t-1}, p_{t-2}. This size is sufficient to capture the local neighborhood, even for large vessels and bifurcations. In addition, the state contains a sub-volume of the binary agent path centered at p_t, which represents the voxels the agent went through. As a vessel can be viewed as a tubular structure with varying radius, any independent point in the track through a vessel can be tracked in two directions. Using the image information from the three previous steps and the binary agent path can help the agent track the vessel in a directed way. The dynamic displacement vector array at time t consists of the directional vectors (steps of the agent) from the current position till the initial position $[a_t, a_{t-1}, ..., a_0]$. The dynamic displacement vector provides additional directional information on previous steps.

   The agent discrete action space is defined as the 26 connected neighboring voxels of the current position p_t in an agent path.

   We represent the set of actions by scaled displacement vectors as follows:

$$\Delta p_t \quad = \alpha \times \{x, y, z \mid x, y, z \in \{-1, 0, 1\}\}, \qquad\qquad 5.1$$

As a consequence, an agent action a_t can make a step of at least two voxels per episode step if $\alpha = 2$.

Reward

The reward function is an essential component for the network to achieve the optimal policy. The ultimate goal of the vessel tracking training task is to learn the agent

tracking the vessel centerline by maximizing R, the accumulated discounted instant reward r_t:

$$R = \sum_{t=0}^{T-1} \gamma^t r_{t+1} \qquad\qquad 5.2$$

The instant reward r_t is used to measure the policy performance of the agent at time t, and thus is relevant for policy optimization. We would like the agent to follow the reference path G_t, which in our case is a voxelized representation of the annotated vessel track.

We, therefore, use the difference of the curve-to-curve distance between the agent path and reference standard between time t and $t - 1$ to measure the reward of the corresponding action. The curve-to-curve distance was introduced in [113] to quantify the distance between coronary centerlines. The curve-to-curve distance is computed by approximating the surface area spanned between two curves via the summation of the Euclidean length of all corresponding point-to-point connections. The optimal (minimal area) correspondence between two curves was determined via a Dijkstra minimum cost algorithm ([114]). Fig. 5.2 shows examples of surface area computation (the summation of red lines) in various scenarios.

In addition, we would like to enforce the agent to follow the reference standard G_t. Therefore, a binary overlap metric, which measures the overlap between the reference path G_t and the agent path, is part of the instant reward.

We propose an instant reward function that combines binary overlap and the curve-to-curve distance. The binary overlap metric provides extra positive $(+1)$ feedback when the agent path and ground truth are overlapping at position p_t. Denoting the surface distance in position p_t as L_t, and the binary overlap at position p_t as B_t, the proposed instant reward is defined as follows:

$$r_t = \begin{cases} B_t - \log(\epsilon + |L_t - L_{t-1}|), & \text{if } L_t - L_{t-1} < 0; \\ B_t + \log(\epsilon + |L_t - L_{t-1}|), & \text{otherwise.} \end{cases} \qquad 5.3$$

Policy

In our vessel tracking application, the optimal policy is obtained using the PPO method with the advantage actor critic (A2C) ([115]) framework. The A2C architecture consists of two networks: an actor network and a critic network. The critic network $V_v(s)$ estimates the state value and actor network $Q_w(s, a)$ learns the state-action value suggested by the critic. The advantage function $A(s_t, a_t)$ (Eq. (5.4)) is introduced to measure the state action pair at time t:

$$A(s_t, a_t) = Q_w(s_t, a_t) - V_v(s_t) \quad . \qquad\qquad 5.4$$

It determines how much better a selected action is compared to the expected value of all possible actions. In this application, generalized advantage estimation (GAE) ([116]) is applied to further reduce the variance of the temporal difference error of the advantage function.

In policy gradient DRL methods, stochastic optimization algorithms such as gradient ascent are sensitive to the gradient update step size. To accommodate for the
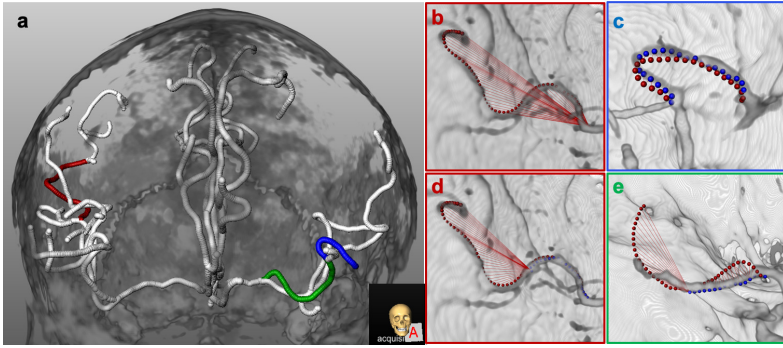
Figure 5.2: *Example of curve-to-curve distance and CCS metric, the red dotted line is the reference standard and the blue dotted line is the actual track, the summation of lengths of the red lines between the reference standard and track is the computed surface distance; a: the anterior vessel tree with highlighted color-coded vessel centerline (corresponding to the colored frame). b: example of L_0. c: example of success tracked case, CCS = 0.87, OR =1. d: example track at time t, CCS = 0.58, OR = 0.5. e: example of miss tracked case, CCS = 0.58, OR = 0.09;*

step size issue, trust region policy optimization (TRPO) ([103]) was introduced using KL-divergence ([117]) constraints to stabilize the gradient update step between old and new policies. However, the use of KL-divergence yields costly computations on multiple Hessian-vector products ([118]). The PPO method ([104]) which was later developed to approximate the KL-divergence constraints by a simple but efficient regularization mechanism in objective function as follows:

$$J(\theta) = \mathbb{E}[\min(\rho(t)A(s_t, a_t), \text{clip}(\rho(t), 1 - \varepsilon, 1 + \varepsilon)A(s_t, a_t))] \quad , \qquad 5.5$$

where $\rho(t)$ is defined as the likelihood ratio between policies at two time points:

$$\rho(t) = \frac{\pi_t(s_t, a_t)}{\pi_{t-1}(s_{t-1}, a_{t-1})} \quad . \qquad 5.6$$

Network Architecture

The proposed network architecture is shown in Fig. 5.3. The architecture consists of an actor network and a critic network. The critic network helps the agent (actor) with learning the optimal state-action value during the training process. Assuming that the policy is optimal after training, only actor network is required for obtaining the tracks in the inference stage.

The agent state consists of sub-volumes sampled from three consecutive points along the agent path and one sub-volume containing the binary agent path. In the CNN part, the three sub-volumes generally can not cover the whole path of the agent.
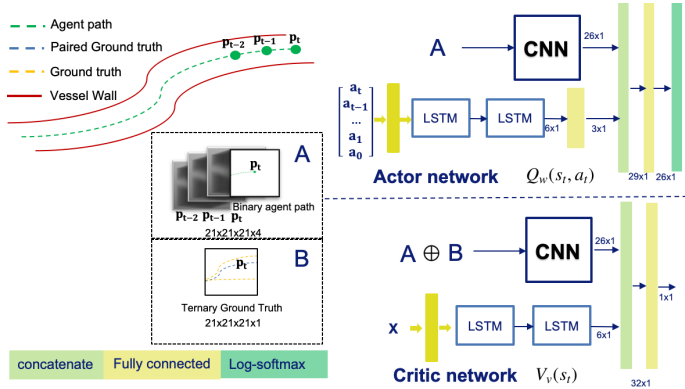
Figure 5.3: *The architecture of the advantage actor critic (A2C) network.
Both actor and critic networks consist of CNN and an RNN. The
RNN part consists of two LSTM cells. The input of the actor
CNN (A) consists of three CTA cubes and one binary agent path
cube ($21 \times 21 \times 21$ voxels). The three CTA cubes are centered
at the current position p_t and two previous positions p_{t-1}, p_{t-2}.
The binary agent path cube is centered in the current position p_t .
The CNN input of the critic network (A ⊕ B) consists of A and
a ternary ground truth cube (B) that is centered at the current
position p_t. The RNN input of the critic network is a vector ar-
ray (x) that consists of paired local features along the track. The
RNN input of the actor network is a vector $[a_t, a_{t-1}, ..., a_1, a_0]$
that contains the action history.*

A recurrent neural network (RNN)([119]) was therefore used to provide extra temporal
information along the agent path.

The architecture of the actor and critic networks are similar, they both consist of
a CNN and an RNN. The CNN model is similar to the architecture proposed by [94]
but with a parametric rectified linear activation function (PReLU) ([120]) instead of
a rectified linear activation function (ReLU) ([45]). The proposed CNN architecture
is based on a dilated network([121]) architecture. Each CNN block consists of a
series of dilated convolutions, with dilation kernels of size 1. For each convolutional
layer, instance normalization was applied. The detailed CNN architecture of the actor
and critic can be found in Table 5.1. The critic CNN takes a ternary ground truth
sub-volume centered at p_t as additional input. The dimension of this sub-volume is
$21 \times 21 \times 21$ voxels, and each voxel has a value of -1, 0, or 1: the paired ground truth
centerline for the current episode has value 1, other centerlines are labeled as -1, and
the remaining voxels have value 0. This setup ensures a decrease in instant reward if
the agent moves to the wrong vessel.

The RNN in both the actor and critic network consists of two LSTM cells. The
input of the RNN in the actor network is the action history vector; the input vector
x of the RNN in the critic network at time point t consists of the paired instant

Layer	1	2	3	4	5	6	7
Kernel width	3	3	3	3	1	1	7
Dilation rate	1	1	2	3	4	1	1
Channels	32	32	32	32	64	64	26

Table 5.1: *The network architecture of actor/critic CNN. The actor's input is $21 \times 21 \times 21 \times 4$ and the critic's input is $21 \times 21 \times 21 \times 5$. Each convolutional layer is followed by an instance normalization and PReLU..*

rewards (r_t), the log of instant reward $(log(r_t))$, the intensity value of agent path (I_t) and the mean intensity value of a two voxel wide region around the agent path $(\bar{I}_t)$:

$$x = \begin{pmatrix} r_t & log(r_t) & I_t & \bar{I}_t \\ r_{t-1} & log(r_{t-1}) & I_{t-1} & \bar{I}_{t-1} \\ \vdots & \vdots & \vdots & \vdots \\ r_0 & log(r_0) & I_0 & \bar{I}_0 \end{pmatrix} \qquad 5.7$$

The outputs of the CNN and RNN are concatenated and fed into a fully connected layer. The final output of the actor network is a (log) softmax layer, which determines the direction of the agent in the next state. The final output of the critic network is a scalar value, which approximates the value of the input state.

The final fully connected layer in the actor network is absent in the critic network. As the latter network needs to learn the state value, the final fully connected layer is not necessary.

5.2.2   Bifurcation Detection

To be able to build a vascular tree, a bifurcation detector was developed that can be applied to each path that was generated by the tracker; from bifurcations, new paths can be tracked.

We approach bifurcation detection as a binary classification problem. The structure of the model used for bifurcation detection is similar to the dilated network architecture in the CNN blocks of the A2C networks, but with only one image cube p_t as input. Each convolutional layer is followed by switchable normalization ([122]) and a dropout rate of 0.2; we use binary cross-entropy as the loss function. The detailed architecture of the bifurcation detector can be found in Table 5.2.

Layer	1	2	3	4	5	6	7
Kernel width	3	3	3	3	1	1	7
Dilation rate	1	1	2	3	4	1	1
Channels	32	32	32	32	64	64	26

Table 5.2: *The network architecture of bifurcation CNN. The input size is $21 \times 21 \times 21$. Each convolutional layer is followed by a switchable normalization, a dropout rate of 0.2, and a PReLU activation function.*

5.2.3 Tree Formation

The anterior vessel tree can be viewed as an arborescence with a given root. We apply a breadth-first tree formation for the anterior vessel tree extraction. The tracking agent and bifurcation detector described in the previous paragraph can be applied to obtain a series of tracks and bifurcation points. Starting from the root, we obtain the first track. Subsequently, we prune the tracked path based on stopping criteria, and recursively and in a breadth-first manner, find bifurcations along this pruned tracked path, and start new tracks from these bifurcations. In this latter step, new paths are tracked from the bifurcation point in several directions, and then failed or overlapping tracks are removed.

In an attempt to get a more robust approach, increasing the true positive tracks, while reducing the missed ones and false positives, we also introduce an ensemble method with five different vessel tracking models.

Stopping Criteria

Stopping criteria are used to determine the termination point of a track; i.e. the track will be pruned if one of the following conditions is met:

- the track runs into the brain tissue

- the track runs into the dilated skull,

- the length of the track exceeds a maximal length.

- the track reaches the border of images in z-axis (for half brain coverage cases).

- the track runs into a different arterial territory. e.g. forming ACA tree but running to MCA territory.

Fig. 5.4b shows the example of the relationship between the anterior vessel trees and the arterial territory map.

Bifurcation Inference

Application of the bifurcation detection model (Sect. 5.2.2) for all the voxels of a tracked path yields a probability of a bifurcation being present at each voxel of the track. The detector may already give some probability to voxels that are close to a bifurcation, The probabilities in voxels from near a bifurcation to the bifurcation increase, and have a peak at the bifurcation. We, therefore, extract the bifurcation points by taking the max probability voxel from a consecutive series of voxels that consists of at least two voxels with a probability larger than 0.5. The threshold allows for smaller bifurcations to be detected.

Bifurcation Path Tracking

After bifurcation detection, new paths must be tracked from the bifurcation. This may be done using an estimate of the candidate directions based on the vessel features. We take a two-step approach. In the first step, we start a tracker in all candidate directions regardless of the presence of vessels. Then, we remove unwanted tracks that do not satisfy the eligibility criteria. The candidate directions are determined from 26 connected neighborhoods. In order to prevent the tracked paths from following the original track backward, tracks are only started in the forward direction with regard to the current position (i.e. if the inner product of the current path and direction vector of the new path is positive). After tracking these candidate directions, spurious tracks will be removed based on the amount of overlap with existing tracks and the length of the track. If the track has an overlap with a previously tracked path that is larger than or equal to 95 %, the tracked path will be ignored. Short tracked paths (less than 4.5 mm) will also be ignored.

Ensemble Method

In addition to tracking with one model, the combined policy and models ensemble method aims to further improve the performance by combining results from multiple tracking models in two different ways. The first track of each tree is essential for tree formation. To ensure a good initial track, we use majority voting for the first track: at each tracking step, we initiate five agents to generate the candidate actions and select the action (i.e. next voxel) that is closest to the average of the actions of each of the models. After the first track, the output of all models is used to reduce missed vessels in bifurcations. The trackers in the ensemble method use the same stopping criteria as the other trackers (see Section 5.2.3).

5.3   Data

5.3.1   Data Overview

In our study, we used CTA and CTP image data from two sources: MR CLEAN registry ([51]) and Erasmus MC. The MR CLEAN registry is an ongoing multi-center registry for stroke patients that underwent endovascular treatment for LVO in The Netherlands since March 2014. Seventeen centers are involved in this study. Data selection criteria applied when selecting the images were: (*i*) slice thickness ≤ 1.5 mm; (*ii*) slice spacing ≤ 1.5 mm; (*iii*) for CTA, the contrast acquisition phase has to be peak arterial, equilibrium or early venous phase according to the definition of [6]; (*iv*) brain coverage has to be at least half of the brain, the insula region needs to present in both hemispheres; (*v*) no large motion artifact along the time axis for CTP data; (*vi*) CTP data has to contain full cardiac cycle from early arterial phase to late venous phase.

1594 subjects were included in the MR CLEAN registry from March 2014 to June 2016. In a previous study ([21]), we selected 270 images from this set that matched the inclusion criteria. Of these, 49 were manually selected such that there is variation

in vendor, image quality, and acquisition phase. In our study, we used these 49 images (with annotations), and we randomly selected another 26 from the 270 images.

From the 63 registry subjects with CTP images matching our selection criteria, we randomly selected 35 subjects. In addition, from 58 stroke patients with an LVO that were admitted from Jan. 2018 to March 2020 at the Erasmus MC, we randomly selected 15 subjects from the 58 subjects (out of a total of 335) that matched our selection criteria.

More detailed information on the data selected is presented in Table 5.3.

5.3.2   Data Annotation

The tracks of the anterior vessel tree for training the DRL were manually annotated by the first author of this manuscript (JS) for all 125 subjects. This annotation was done under the supervision of an experienced radiologist (20 years of experience) and a physician (5 years of experience). The annotation task was defined as labeling the track of every anterior vessel from the ICA-top to the most distal part of the MCA and ACA vessel tree in the CTA or CTP maximal intensity projection (MIP) images. Each tree always starts from the ICA-top of the corresponding side, if there is no ICA-top occlusion. The annotation of a track was discontinued when the artery was not clearly visible anymore. In our study, the vessel diameter varies from 3.6 ± 0.45 mm at the ICA top ([52]) to 0.45 mm in the most distal vessels. For this annotation task, we used an in-house developed annotation tool used in [21]. In order to increase the number of segments near the skull (which are difficult vessels to track), veins near the skull were annotated in 10 out of 75 subjects (randomly selected). These veins have an appearance that is similar to arteries.

5.3.3   Interobserver Variation on Annotation

For such a difficult annotation task, it is relevant to assess the interobserver variation in the annotation. To this end, a second observer, a medical student (M) was asked to perform the same annotation. For this, three subjects from the CTP category were selected, two subjects with middle complexity and one with high complexity. The tree overlap was calculated using a dilated binary spherical shape with a radius of four voxels with observer M as the reference standard, in the same way as the overlap is computed for the method. For the three subjects, the tree overlap was 0.50, 0.70, and 0.88 respectively, with an average overlap of 0.69. The vessels where the annotation differed were mainly low intensity vessels, and vessels close to the skull.

5.3.4   Data Preparation

In this study, we use both the CTA and the frame with maximal vessel volume from the CTP series. In order to minimize the variation in spatial resolution, all CTA and CTP images were resampled to a 0.45 mm isotropic grid using cubic B-spline interpolation and were normalized to a range of 0 to 1 using min-max normalization.

The annotated vessel tracks were transformed into an arborescence structure from the root based on the connectivity as shown in Fig. 5.4a. In this way, all the bifurcations

5

Properties		Data division (n=125)						
		Set A (n=20)	Set B (n=20)	Set C (n=20)	Set D (n=20)	Set E (n=20)	Validation (n=10)	Test (n=15)
Image Modalities	Numbers of CTA (n=75)	13	10	14	15	12	2	9
	Numbers of CTP (n=50)	7	10	6	5	8	8	6
Brain Coverage	More than half (n=25)	2	6	3	5	5	2	2
	Complete (n=100)	18	14	17	15	15	8	13
Slice thickness	[0.50-0.75)mm (n=59)	10	9	9	11	8	6	6
	[0.75-1.00)mm (n=32)	5	5	7	3	6	1	5
	[1.00-1.50]mm (n=34)	5	6	4	6	6	3	4
Acquisition phase	Early arterial phase (n=19)	4	3	3	1	3	0	5
	Peak arterial phase (n =37)	6	6	4	11	6	1	3
	Equilibrium phase (n=69)	10	11	13	8	11	9	7
Occlusion location	ICA (n=20)	4	4	5	2	2	2	1
	M1 (n=60)	9	12	11	7	10	4	7
	M2 and above (n=38)	6	3	3	9	8	3	6
	Others (n=7)	1	1	1	2	0	1	1
Complexity	Low: < 50 (n=40)	7	5	8	7	7	1	5
	Medium:[50-100] (n=65)	12	12	10	9	9	6	7
	High: $\geq$ 100 (n=20)	1	3	2	4	4	3	3
Proximal segments	Numbers (n=417)	67	65	65	70	64	34	52
	Intensity (HU)	311 ± 94	306 ± 84	280 ± 87	312 ± 92	321 ± 120	256 ± 105	274 ± 77
	Length (mm)	26 ± 17	24 ± 15	26 ± 14	25 ± 15	24 ± 13	21 ± 12	24 ± 16
Middle segments	Number (n=3814)	607	693	577	650	564	288	435
	Intensity (HU)	161 ± 76	161 ± 70	152 ± 68	153 ± 68	167 ± 80	153 ± 73	143 ± 61
	Length (mm)	29 ± 23	29 ± 26	29 ± 26	31 ± 25	31 ± 24	29 ± 25	31 ± 24
Distal segments	Number (n=2436)	343	390	336	432	410	200	325
	Intensity (HU)	129 ± 66	134 ± 61	130 ± 64	121 ± 54	148 ± 84	116 ± 51	127 ± 65
	Length (mm)	27 ± 22	29 ± 26	25 ± 22	28 ± 24	28 ± 23	27 ± 25	28 ± 27
Near skull segments	Number (n=1404)	170	237	216	220	255	112	194
	Intensity (HU)	153 ± 60	149 ± 61	144 ± 47	139 ± 45	149 ± 56	128 ± 46	127 ± 38
	Length (mm)	63 ± 40	64 ± 40	61 ± 39	62 ± 38	61 ± 37	61 ± 40	57 ± 39
Near skull segments (Added extra)	Numbers (n=595)	121	171	147	94	62	0	0
	Intensity(HU)	204 ± 81	140 ± 56	173 ± 83	148 ± 63	181 ± 95	NA	NA
	Length(mm)	65 ± 30	60 ± 30	67 ± 39	63 ± 30	57 ± 26	NA	NA

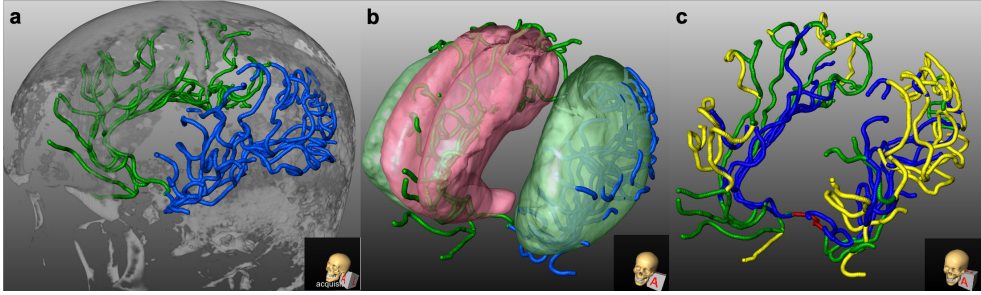***Table 5.3:*** *The data distribution of 125 subjects.*

Figure 5.4: *Processing of the anterior vessel trees data preparation; a. shows the example of the anterior vessel tree structure, the MCA tree is presented in blue and ACA trees are presented in green; b. shows the anterior vessel trees and thresholded (0.5) MCA and ACA mask, the green mask is the MCA mask, the red mask is the ACA mask; c. shows the segment labels. The proximal segments are in red, the middle segments are in blue, the distal segments are in green and the skull segments are in yellow.*

and the connected segments (vessel segment) are known. During this transformation process, two adjacent bifurcations with a distance of less than 1.8 mm were merged into one trifurcation. The vessel segment representation was subsequently transformed into a voxelized 26-connected representation for use in the training process of the directed vessel tracking model. The bifurcations were used for the bifurcation models. For the tree formation, the initial direction vectors were generated from two manually annotated starting points from the ICA top to the proximal points of each tree. The user has to click two points for the initial direction vector.

5.3.5 Segment Label Generation

To permit an analysis of the results (i.e. distal versus proximal), we define four categories (labels) for the vessels. proximal, near skull segments, middle segments (majority of the segment inside the MCA / ACA mask), and distal segments (remaining segments). The proximal segments consisted of M1 and A1, which are defined as the segments between the given root till the first bifurcation. Near skull segments were defined as the vessel segments falling into the dilated skull (thresholded at 1000 HU, dilated with a $5 \times 5 \times 5$ kernel) for consecutive five points. The middle segments were defined using the MCA and ACA mask, i.e. those segments where the majority of the points in the segment were inside these masks. The above-mentioned segments labels and the thresholded MCA and ACA maps are shown in Fig. 5.4b, Fig. 5.4c. The segment labels are only used in the analysis of the proposed method.

5.3.6   Experimental Setup

The data in this study is used for training and testing DRL directed vessel tracking and bifurcation detection model, and for assessing tree formation. In the directed vessel tracking and bifurcation detection experiments, we use the same data distribution: in both cases, we use 100 images in a five-fold cross validation setup, where the data was stratified on complexity (defined as the number of segments per subject). The ten subjects with extra annotations of segments near the skull were evenly distributed over each fold. The validation set in the cross-validation consisted of another ten subjects (from the remaining 25), and this set was kept the same for all cross-validation experiments. The remaining fifteen subjects were used as an independent test set. The tree formation assessments were performed on the validation and test sets. 1.8 mm means 4 voxels, and with a scaling (stepsize) of 2, we thus allow an error of maximal two steps.

The image characteristics and segment labels of each fold, validation, and test set are listed in Table 5.3. In the following, we will name the models based on the fold. For instance, Set A is the test set of tracking model Tr-A and Bifurcation model (Bf-A) in 5 fold cross validation setting. The tracking model Tr-S and bifurcation model (Bf-S) use set A to E for training, and a test set (n=15) for testing. The validation set (n=10) is the same for all models.

5.4   Experiments and Results

5.4.1   Implementation

Directed Vessel Centerline Tracking

The proposed PPO based directed vessel tracking approach was implemented in PyTorch ([123]). The model training and validation were done on NVIDIA A40 GPUs.

The PPO training uses episodic learning with a random start position from the first position to five voxels before a bifurcation point of a segment. The length of each episode was therefore arbitrary with a minimal length of five voxels. Each mini batch consists of 8 episodes, in which one episode corresponds to one segment. The training was stopped when one of the following criteria was met: 1) the total length of the track exceeds $1.5 \times$ the length of the reference standard, 2) the agent is off the reference standard segment for a distance of 1.8mm, 3) the target was reached 4) at the beginning of the training, if the agent has gone into the direction that is opposite to the direction of the initial vector.

The hyper-parameters of the PPO were: discount factor $\gamma = 0.9$; PPO clip value $\epsilon = 0.2$; GAE parameter $\lambda = 0.95$. The model weights were updated 10 times per mini batch. The initial learning rate for Adam was 1e−5. The learning rate is halved when the validation score does not improve for five epochs in a row. The lower bound of the learning rate was 1e−6. During training, the only augmentation applied is the random flipping of both 3D CTA image and centerline segments along the x- and y-axis with a probability of 0.5. We use a two-stage training: first, the data from the anterior vessel tree for learning the general anterior vessel tree track is used until the model stops converging; in this stage, the target of a track was a bifurcation. In the second stage,

we added the extra segments close to the skull to better learn to track vessels running in the vicinity of the skull. Also, the target is set to five points beyond the bifurcation with both branches with equal probability. In this way, the agent could learn to track beyond bifurcations.

To monitor the episodic training process, we use a curve-to-curve similarity (CCS) metric, defined as follows:

$$CCS = 1 - \frac{L_t}{L_0} \qquad 5.8$$

with L_0 as the max surface distance at the initial point of the reference standard curve, see also Fig. 5.2b. This metric is 1 in case of complete overlap and is negative when the track runs in the opposite position.

Bifurcation Detection

The model implementation and training for bifurcation detection were similar to the PPO, both using Pytorch and the same GPU. The training samples are sub-volumes that are at randomly shifted positions along the reference centerlines. This random offset is to simulate a track resulting from the vessel tracking. The random offsets were ranging from 1 to 3 voxels. The training label was obtained by dilating the ground-truth bifurcation points with a binary spherical kernel with a radius of four voxels, in order to increase the size of true positives and make the classification problem less imbalanced. During training, the number of true positive samples was equal to the number of true negative samples. In the training phase, the initial learning rate was 1e−3 for the Adam optimizer. The learning rate reduction scheme was identical to the PPO training. The learning rate lower bound was 1e−6 as well.

Tree Formation

The tree formation method was implemented in python. For the stopping criteria, tracking into brain tissue was defined as an intensity value along the track less than 50 HU for three consecutive points, and tracking into the skull is defined as tracking into a dilated skull (threshold at 1000 HU, dilated with $5 \times 5 \times 5$ kernel) for three consecutive points. The maximum length for a track was 330 mm, which is 1.2 times the maximal length of root-to-leaf distance in the training set of 100 subjects.

Probability density maps of the MCA and ACA regions, as well as a hemisphere map, were obtained using earlier described atlases ([39]) that were registered to the images. After transformation to the CTA or CTP images, the MCA and ACA arterial territory maps were thresholded at 0.5 to obtain a binary mask. The hemisphere map consists of three values, indicating the left and right brain hemispheres and the background.

5.4.2 Directed Vessel Tracking Assessment

The assessment of the directed vessel tracking consists of an ablation study, comparisons with a baseline DQN model and the baseline DQN with our own reward function. In addition, we assess the impact of using all data. For the final model, we also present an

analysis per vessel category. We use overlap rate and average distance as metrics, where the overlap rate (OR) is defined as the part of the track that matches the reference standard until the first error (distance larger than 1.8 mm). The average distance refers to the average curve-to-curve distance in world coordinates between the overlapping part of the tracking path and the reference standard. Statistical significance was assessed using a paired t-test on all metrics, and a value lower than $p = 0.05$ was considered statistically significant. The vessel tracking results are reported using a median with an interquartile range (IQR).

Ablation Study

In the ablation study, we investigate the importance of our network architecture design choices, training scheme, and reward function. In addition, we compare our reward function with an existing reward from [101]. We, therefore, divided our ablation study into two parts. In part one, the ablation study focused on investigating the added value of network architecture and training schemes. In part two, the ablation study investigates the impact of the reward functions.

The network architecture may have a substantial impact on the performance of DRL methods, and similarly, the activation function has been shown to have an impact on the performance ([124]). The ablation study part one focuses therefore on investigating the added value of:

- the RNN module, by comparing the complete model with a model that uses only a CNN part in the A2C network (*CNN only model*),

- the RNN module with less image information, by training a model that uses image information from only one time point with RNN module (*One time point model*),

- scaled action space, by removing the scaling factor, (*No scaling model*)

- PReLU instead of ReLU (*ReLU model*),

- two-stage training, by adding all data and bifurcation extension throughout the training process (*Single stage model*).

For reference, we also train a model that contains all components *model Tr-A*.

The ablation study part two investigates the impact of various reward functions with the same model configuration as model Tr-A, by comparing our model with the following rewards:

- curve-to-curve distance only, by using the log curve-to-curve distance function (*CC distance model*),

- the point to curve distance reward of [101] (*Li reward*)

Fig. 5.5 shows the learning curves with CCS metric and Table 5.4 contains the test results for the ablation study of network architecture and the training scheme. The median OR values in the ablation study are 100 %, but the variation in Q1 values

demonstrates that the distributions are different. The Q3 CCS of most models is above 90 %, except the *No scaling model* (83%). The difference between Tr-A and the other configurations is statistically significant for both CCS and OR metrics. In addition, statistically significant differences were found between the No scaling configuration and the other configurations in terms of average distance.

Fig. 5.6 shows the learning curves and Table 5.5 contains the results of the reward functions ablation study. Statistically significant differences were found between the Tr-A and the other configurations in all of the cases except for the average distance of *Li reward*.

Comparison with DQN Baseline

We also compared our method with a popular DQN method. Our baseline model was introduced by [101]. This method uses the same network architecture as [94], which is similar to the CNN architecture of our actor network but with one time point and ReLU activation function. To do a fair comparison, we limit the action space of the baseline DQN methods to the same action space (26 connected neighborhoods). The hyperparameters and optimizer settings are the same as for the PPO training. The data used is the same as in the ablation study. In addition, we also used a baseline architecture with our own reward function(*DQN + ours*) to further assess the added value of this reward function in a different DRL method. The performance of baseline DQN and (*DQN + ours*) are shown in Table 5.5 and the learning curves are shown in Fig. 5.6.

Model	CCS (%)	Average distance (mm)	Overlap rate (%)
CNN only	75 [29, 91]	0.64 [0.57, 0.74]	100 [36, 100]
One time point	78 [40, 93]	0.63 [0.56, 0.72]	100 [46, 100]
No scaling	38 [3, 80]	0.56 [0.49, 0.66]	72 [18, 100]
ReLU	79 [40, 93]	0.62 [0.55, 0.72]	100 [44, 100]
Single stage	80 [36, 93]	0.61 [0.54, 0.72]	100 [41, 100]
Model Tr-A	82 [44, 93]	0.65 [0.58, 0.75]	100 [50, 100]

Table 5.4: *Evaluation result of ablation in network architecture and training scheme (part one) and proposed method on set A.*

Model	CCS (%)	Average distance (mm)	Overlap rate (%)
CC distance	71 [34, 88]	0.75 [0.66, 0.85]	100 [40, 100]
Li reward	53 [21, 85]	0.64 [0.56, 0.73]	61 [28, 100]
Baseline DQN	26 [13, 53]	0.69 [0.59, 0.78]	29 [13, 58]
DQN + ours	44 [15, 79]	0.73 [0.63, 0.85]	44 [16, 100]
Model Tr-A	82 [44, 93]	0.65 [0.58, 0.75]	100 [50, 100]

Table 5.5: *Evaluation result of reward function ablation study(*CC distance and Li reward*), baseline DQN, DQN with our reward function(*DQN + ours*) and proposed method on set A.*
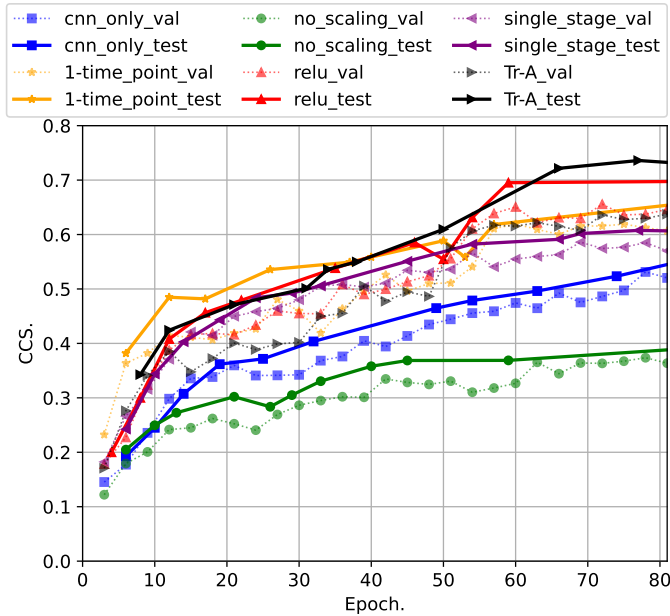
Figure 5.5: *Curve to curve similarity (CCS) metrics of the ablation study models on network architecture and training scheme (part one) and model Tr-A (All models are trained using set B to E, validated on set validation and tested on set A.).*

Amount of Training Data

We also investigated whether the current amount data training data is sufficient for our application. For this purpose, we trained a model (*model S*) with all 100 subjects from set A to E with the same training scheme as was used in the five fold cross validation setup. Table 5.6 shows the test results of model S and five models on the same independent test set. None of the differences is statistically significant.

Model Tr	CCS (%)	Average distance (mm)	Overlap rate (%)
Model Tr-A	83 [44, 93]	0.60 [0.54, 0.70]	100 [46, 100]
Model Tr-B	82 [40, 93]	0.60 [0.54, 0.72]	100 [46, 100]
Model Tr-C	84 [49, 93]	0.60 [0.54, 0.70]	100 [47, 100]
Model Tr-D	82 [34, 92]	0.60 [0.54, 0.70]	100 [41, 100]
Model Tr-E	84 [41, 93]	0.61 [0.55, 0.71]	100 [46, 100]
Model Tr-S	85 [50, 93]	0.59 [0.54, 0.70]	100 [46, 100]

Table 5.6: *The evaluation result of the five models and model S from bifurcation to bifurcation on 15 independent test subjects.*
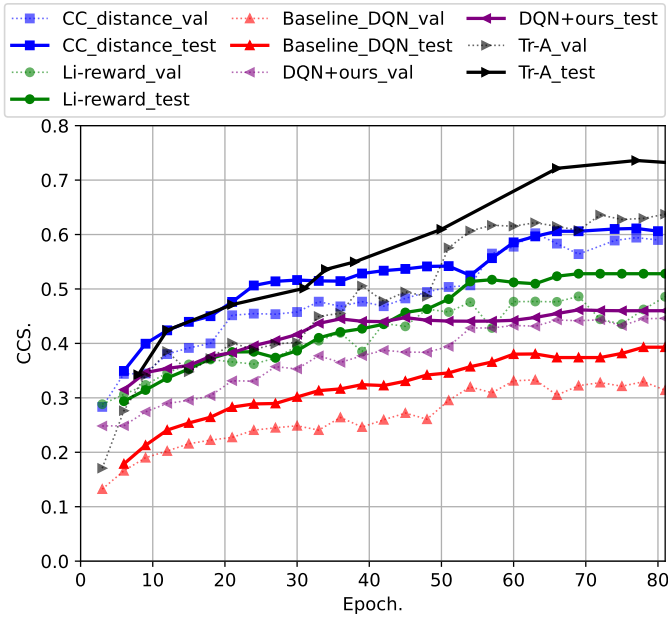
Figure 5.6: *Curve to curve similarity (CCS) metrics of the rewards ablation study, baseline DQN, DQN with our reward function(DQN + ours) and model Tr-A. All models are trained using set B to E, validated on set validation and tested on set A.*

Model Tr	CCS (%)	Average distance (mm)	Overlap rate (%)
Model Tr-A	82 [44, 93]	0.65 [0.58, 0.75]	100 [50, 100]
Model Tr-B	76 [32, 92]	0.65 [0.56, 0.76]	100 [38, 100]
Model Tr-C	81 [37, 93]	0.65 [0.56, 0.76]	100 [44, 100]
Model Tr-D	78 [35, 92]	0.66 [0.57, 0.75]	100 [44, 100]
Model Tr-E	85 [46, 94]	0.64 [0.56, 0.74]	100 [55, 100]

Table 5.7: *The evaluation result of the directed tracking from bifurcation to bifurcation using five fold cross validation over 100 subjects.*

Directed Vessel Tracking Performance

The last tracking experiment focuses on the generalizability of the tracking model and further analysis of the tracking performance. For this, we used 100 images from the five fold cross validation. The learning curves of the five fold cross validation are shown in Fig. 5.7. The test result of the models with their corresponding set is listed in Table 5.7. Variations in the results may be caused by differences in the validation datasets. Therefore, we compute the result for all models with the same independent test set. These results are listed in Table 5.6. It shows only minor differences between

the results, and there are no statistically significant differences between all listed models (including model S).

For the same test set and models, we also analyze the tracking performance with regard to the segment labels. The result can be found in Table 5.8. They show that the performance of directed vessel tracking depends on how distal the vessels are.
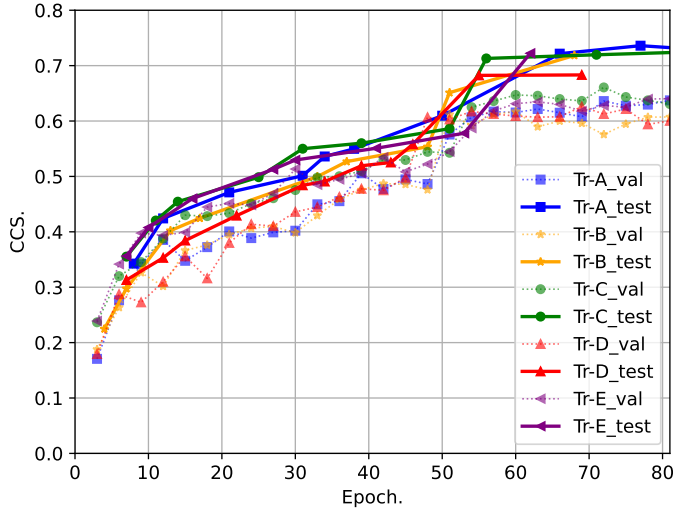
Figure 5.7: *Curve to curve similarity metrics for 5 fold cross validation.*

5.4.3   Bifurcation Detection Assessment

The bifurcation detection was tested in two ways. First, the performance was assessed in the training setup with a dilated reference standard, with random samples from the randomly shifted voxels along the annotated tracks (see. Sec. 5.4.1). This was done in a five fold cross validation setup (models Bf-A – Bf-E), and the results are shown in Fig. 5.8. The test accuracy of all five models with different sets converges to about 0.82. The accuracy on the validation set was 0.79, slightly lower than the test performance.

Segments (type)	CCS (%)	Average distance (mm)	Overlap rate (%)
Proximal	91 [84, 94]	0.61 [0.55, 0.69]	100 [100, 100]
Middle	88 [62, 94]	0.63 [0.55, 0.73]	100 [84, 100]
Distal	80 [44, 91]	0.65 [0.56, 0.74]	100 [59, 100]
Near skull	44 [12, 82]	0.68 [0.59, 0.79]	46 [18, 81]

Table 5.8: *Analysis based on the different segment labels on 115 independent subjects.*

Figure 5.8: *Learning curve of bifurcation model 5 fold cross validation with 100 subjects.*

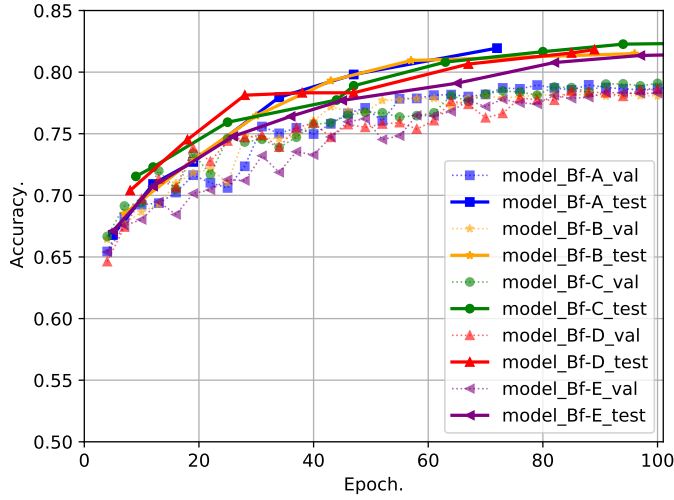The second assessment focused on bifurcation detection on paths that were obtained from the DRL-based tracker. For this, 100 subjects from the five fold cross validation and 15 subjects from the independent test set were used.

The directed vessel tracking model was combined with the corresponding bifurcation detection model. e.g. for the testing set A, tracking model A and bifurcation model A were used. In addition to the five fold cross validation models, we trained a bifurcation model Bf-S with 100 subjects. The average precision and recall for the different sets with corresponding models are listed in Table 5.9. The recall and precision of bifurcation detection during tree formation are around 76% and 87% respectively, and there is no statistically significant difference between the models.

5.4.4   Tree Formation

In the tree formation evaluation, we assess the performance of the combined tracker and bifurcation detector. In addition, we investigate whether the ensemble method would improve over the single model method, and the consistency of the tree formation method and we assess the impact of image characteristics. We performed a visual analysis

Model Bf	Bf-A	Bf-B	Bf-C	Bf-D	Bf-E	Bf-S
Precision (%)	86	87	88	87	86	87
Recall (%)	78	77	76	76	75	77

Table 5.9: *The evaluation result of bifurcation detection and bifurcation inference for tree formation over 115 subjects.*

of the false positive and false negative branches. Also, we performed interobserver analysis of tree formation between two annotators as shown in Table 5.10.

Two features are relevant for the tree formation method: 1) the completeness of the tree and 2) the topology of the tree (whether connectivity is correct). The completeness of the tree is measured with precision and recall. To this end, the reference standard and extracted tree were dilated using a binary spherical shape with a radius of four voxels (allowing a maximum distance of 1.8 mm). True positives are points along the extracted tree that are in the dilated reference standard, false positives are points of the extracted tree that are not being covered in dilated reference standard, and false negatives are points along the reference standard that are not in the dilated extracted tree.

The topology was assessed using the correct tree topology (CTT) ratio, which quantifies the fraction of points of the tree for which the path to the root is similar (i.e. within 1.8 mm distance everywhere) to the path to the root of the corresponding point in the reference standard.

Single Model vs Ensemble Method

For the comparison of the single model (Model Tr–S, trained on 100 images) and the ensemble method (using the five cross-validation models) both the 10 validation and 15 test images were used. The tree formation results are shown in the top row of Table 5.11. The precision and recall are reported with median and IQR. However, the CTT was reported using the mean and standard deviation since the median and IQR in almost all items are 100, [100, 100], except near skull segments. In general, the performance of the ensemble model is better than the performance of the single model, in terms of precision and recall. When compared with a single model, the true positive rate of the ensemble model increases by 14, [6, 30] %.

In addition, we determined the intensity distribution of false positive and false negative vessels with regard to our reference standard for the ensemble method over 25 subjects. The false positives have a mean and standard deviation of 134 ± 91 HU, and the false negatives have a mean and standard deviation of 114 ± 62 HU.

Consistancy Test of Ensemble Method

We also perform a consistency test with our final tree formation ensemble method, to assess the impact of changes in the starting positions. For this experiment, we used images with moderate complexity (n=13) from the test and validation set. The input of the methods was varied by randomly shifting the staring position within a range of five voxels along the given starting centerline. We ran the methods five times with a randomly shifted initial vector. The first result is used as the baseline of the tree formation consistency test. We then compute the precision and recall for the other four results. Both average precision and recall are above 99%.

Interobserver Analysis on Tree Formation

For the interobserver analysis part, we would like to compare the tree formation result for the three subjects that were annotated by both observers with the annotated
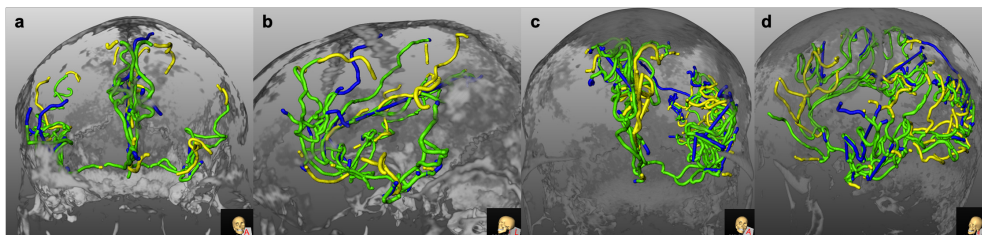
Figure 5.9: *Examples of ensemble method tree formation result in the test set. Those subjects were selected based on the median value of recall. Green denotes true positive segments, blue is false positive, and yellow is false negative. a&b: frontal and sagittal view of a subject with middle complexity, a precision of 0.85, and recall of 0.77. c&d: frontal and sagittal view of a subject with high complexity, a precision of 0.83, and recall of 0.76.*

5

versions. The precision and recall regarding to the different observers were shown in Table 5.10.

	Observer JS		Observer M		
	Precision	**Recall**	**Precision**	**Recall**	**Agreement**
	(%)	(%)	(%)	(%)	
Subject 1	0.78	0.57	0.45	0.49	0.50
Subject 2	0.59	0.72	0.53	0.68	0.70
Subject 3	0.71	0.76	0.71	0.54	0.88

Table 5.10: *The tree formation result using different reference standard.*

Impact of Imaging Characteristics on Tree Formation

To get more insight into the performance of the methods, and how these depend on the properties of the images (CTA vs CTP, acquisition phase, number of vessels, etc.), we present the tree formation results for each of these categories in Table 5.11. It shows that the method performs better in the CTA images compared to CTP and that proximal vessels are better detected than distal vessels. For the other categories (acquisition phase, tree type, or complexity), there are no apparent differences in the results.

5.5 Discussion

In this work, we developed and assessed a method to construct tree models of the anterior vessel arteries from 3D CTA and CTP images of the brain. The method consists of a deep reinforcement learning tracking approach, a CNN-based classifier to detect bifurcations along the tracks, and a breadth-first tree construction algorithm that uses the tracker and bifurcation detector to construct the tree from the tracking

Properties		Single Model			Ensemble		
		Precision (%)	Recall (%)	CTT (%)	Precision (%)	Recall (%)	CTT (%)
Overall		63 [52, 74]	59 [46, 70]	90 ± 31	70 [57, 81]	68 [54, 81]	87 ± 34
Image modalities	CTA	66 [52, 79]	64 [51, 73]	95 ± 22	72 [58, 83]	77 [63, 86]	92 ± 27
	CTP	62 [52, 71]	51 [41, 63]	85 ± 35	69 [57, 72]	61 [47, 78]	82 ± 38
Acquisition phase	Early arterial phase	77 [67, 87]	67 [64, 73]	94 ± 23	82 [77, 86]	77 [70, 81]	94 ± 23
	Peak arterial phase	66 [57, 78]	58 [49, 68]	95 ± 22	70 [58, 84]	80 [58, 85]	90 ± 30
	Equilibrium phase	60 [51, 69]	52 [42, 67]	87 ± 34	66 [55, 74]	62 [48, 81]	83 ± 38
Complexity	Low	57 [48, 71]	73 [62, 83]	95 ± 22	62 [56, 73]	77 [63, 87]	88 ± 33
	Middle	64 [56, 78]	62 [52, 81]	86 ± 35	72 [56, 82]	66 [48, 81]	85 ± 36
	High	68 [58, 72]	58 [41, 64]	94 ± 25	71 [67, 78]	66 [57, 76]	90 ± 31
Tree type	MCA	62 [51, 72]	57 [43, 67]	88 ± 33	73 [64, 83]	60 [52, 78]	84 ± 37
	ACA	66 [56, 79]	59 [48, 72]	92 ± 28	65 [55, 80]	70 [60, 86]	89 ± 31
Segment label	Proximal	99 [68, 100]	99 [68, 100]	100 ± 0	92 [71, 100]	90 [69, 100]	100 ± 0
	Middle	60 [42, 76]	59 [45, 70]	97 ± 16	69 [54, 79]	67 [57, 84]	98 ± 16
	Distal	61 [45, 82]	47 [38, 64]	80 ± 41	74 [59, 88]	67 [49, 79]	78 ± 42
	Near skull	39 [27, 59]	43 [18, 75]	63 ± 49	66 [48, 87]	64 [35, 75]	50 ± 51

Table 5.11: *The evaluation result of tree formation. The precision and recall are reported with median and IQR. However, the CTT were reported using the mean and standard deviation since the median and IQR in almost all items are 100, [100, 100], except near skull segments.*

and bifurcation detection results. When evaluated on a test set of 25 subjects, the median precision and recall w.r.t. the manual annotation is 68% and 69%.

For the tracking, we investigate the policy gradient based proximal policy optimization DRL approach to perform the directed tracking. The average tracking performance on directed tracking from bifurcation to bifurcation over 100 subjects with 7026 vessels was 100 [46, 100]. Bifurcation detection is another essential ingredient, it was tested both in the training scenario and during the single model tree formation with the same 100 subjects. The accuracy of bifurcation detection over 100 test subjects was 82%. During the single model tree formation setting, the precision and recall were on average 87% and 76% respectively. At the final ensemble tree formation method test on 25 independent subjects, the overall precision and recall were 68 [54, 81] and 70 [57, 81].

An ablation study was performed to assess the added value of various design choices made. From the results, it follows that most of the choices (including time information via RNN and multiple time points, activation function, and training strategy) have a small but statistically significant impact on the final tracking results. Both RNN and multiple time points serve as a means to include information from previous steps in the model. From the results it is clear that using RNN with one CNN time point has a better result than CNN only with more time points; apparently, these elements of the model permit to focus on different temporal aspects relevant for the tracking. Adding an RNN module would enhance the tracking ability.

Excluding the scaling of the step size (i.e. taking smaller steps) yields improved accumulated reward but lower CCS and OR. An explanation could be that, first, the frequency of calculating the instant reward is at least two times more than the rest of the configuration. Furthermore, in the no scaling variant, the distance between p_t and p_{t-2} is 0.9 mm. This leads to very similar image information at the three time points, which hampers the full exploitation of temporal information. The performance of this no scaling model is similar to the DQN with our reward function($DQN+ours$), which is the configuration using CNN with one time point.

In the reward ablation study, CC distance performed second best, by comparing the result of *CC distance* and our proposed method *Tr-A* model, adding binary overlap in the reward function would improve the average distance. The point-to-curve distance reward has a higher average distance but lower CCS and OR in comparison with our proposed reward. This result can be observed from both DQN methods and PPO methods. The curve-to-curve distance reward aims to optimize the surface area between the reference standard and agent path at t and $t-1$, whereas the point-to-curve distance used by the work of [99] and [101] aim to optimize the distance between the agent position and corresponding curves. Using the curve-to-curve distance reward function, the instant reward at t is approximated by the difference in surface area between the agent path at t and $t-1$ with regard to the complete reference standard. Therefore, the agent moves along the centerline because it minimizes the surface distance between the agent path and reference standard. This is particularly handy in the cases of sharp turning vessel structures, e.g helix shape, where the correspondence between agent position and corresponding reference standard point, such as used in other approaches, can not be found in an accurate way using the closest Euclidean distance (e.g using point to curve distance). Therefore, the curve-to-curve distance

fits better with our application.

Most existing DRL tracking methods use a DQN approach with a customized reward function; for instance, [101] utilize a DQN tracking approach with the same network architecture as the state-of-art method of [94] and a point-to-curve distance based instance reward for coronary tracking in a 3D CTA image. The tracking performance of the method by [101] outperforms the CNN tracking result of [94]. We, therefore, used this method as our baseline method. The baseline methods perform less well in our application. [99] utilize the CNN with one time point (different architecture) to track the thoracic aorta. The reward function is similar to [101]. [92] utilize the same CNN architecture with double-DQN methods and the dot product between the reference standard and agent path of t and $t-1$ as the reward function for coronary centerline tracking. The corresponding point between the reference standard and the corresponding agent location was determined based on the shortest distance. The above-mentioned architectures are less complex and are sufficient for tracking the coronary arteries and aorta. The brain vasculature, as shown in Fig. 5.1, is different from coronary arteries and aorta, e.g. there is more bending and there is a large variety in curvatures in cerebral vessels compared to coronary vessels, and they are much smaller than the aorta. As a consequence, during tracking, a CNN with one time point only might not be able to provide sufficient information for the network to make accurate direction estimation. In addition, the agent might have difficulty finding the correspondence between the agent's position and the reference standard.

While we were finalizing this manuscript, the work of [125] demonstrated the possibility of walking outside of the classical reinforcement learning problem formulation, which is either value-based or policy-based. In contrast, they regard the reinforcement learning problem as a sequence prediction problem. Similarly, we could use our trained agent and our reward with the sequential framework in an offline reinforcement learning fashion, which may improve our tracking performance. To what extent and whether the transformer would be the best sequential model for our reward function could be further investigated.

When training DL models, data is essential: both the amount of data and the quality. To investigate whether adding may further improve the results, we compared the results of the cross-validation models with a model trained on the full dataset. The results show that the additional 25% of data does not significantly improve the results, suggesting that the amount of data is sufficient for the task.

Annotation of anterior vessel trees in brain CT images is a tedious and difficult task, among others caused by the coexistence of arterial and venous structures in the whole brain. In Fig. 5.1, the anterior vessel tree is only a small part of the complete brain vessel annotation. Under such circumstances, kissing vessels, i.e. locations (mostly distally) where vessels (almost) touch are common. This explains the moderate average interobserver overlap of 0.69. The intensity distribution in the branches where the observers disagree is in the range of the low intensity vessels that consist of distal and near skull segments. Though the annotations are not perfect, we assume that for the training, because of the large number of vessels, the errors in the annotations would not greatly impact the models, which is also suggested by the results. Errors in the annotation also impact the quantitative results. We therefore also compare the tree formation results with the annotations from two different observers in three

datasets. The differences again are mainly in the vessels that have a low intensity, which is harder to annotate, and often are distal vessels. Also, there was a variation in vessels close to the skull. These errors are similar to the errors of the automated method, suggesting that these low-intensity vessels are difficult for both humans and the automated method.

From the analysis of the tree formation result with respect to several imaging characteristics, it followed that only the imaging modality has a significant impact on the quantitative results: the tree formation method on CTP performs less than on CTA. The difference between the validation and test set (the performance on the validation set was always worse than on the test set), may therefore be caused by the higher percentage of CTP images in the validation set. The worse performance may be caused by the use of the maximum volume image for the tracking, whereas the annotation was done in the MIP image, and the blurring was caused by the interpolation after the alignment to the first frame. After annotating the vessels of the CTP in the MIP image, we decided to use the maximum volume CTP image for training and tracking instead of the MIP. Whereas the MIP is convenient for annotation, its appearance is different from CTA images, because of the higher background intensities and more noisy appearance caused by the MIP. This is a limitation of our study, we however preferred including the (suboptimal) annotations over leaving out the CTP, or redoing the annotation.

The final result of the method is a tree representation of the brain vasculature, and this tree representation will be the basis for subsequent works. We intend to use this tree to e.g. find lesions (occlusions in more distal arteries, such as M2, which is still a challenging task [126]), better quantification of collateral status, and possibly also linking/registering the 3D vascular information from CTA and CTP to interventional DSA images for improving image guidance during endovascular treatment. The tree formation method performed reasonably well in the proximal and middle segments. In the distal and near skull segments, the performances were also in the interobserver ranges. For the DSA to CTA mapping and distal occlusion detection, it is likely that such tree formation performance is sufficient.

5.6   Conclusion

In this study, we developed and assessed a method to construct a brain vessel tree from a CTA or CTP image, using a starting point and direction vector. The method consists of a DRL-based tracker, a CNN-based bifurcation detector, and a breadth-first tree building. The tracker performance on segments from bifurcation to bifurcation has a median overlap of 100 %, and the bifurcation detector has a mean accuracy of 82 %. The combination of both components in an ensemble tree building algorithm yields trees that have a 69 % overlap with manual annotations, which is in the interobserver variation range.

Acknowledgments

Chapter 8
General Discussion

Ischemic stroke is a devastating event caused by the occlusion of intracranial arteries. Evaluation of cerebral vessels is an essential component for gaining knowledge about the severity of acute ischemic stroke, the success of treatment, and the outcome of a patient. The presence and density of cerebral vessels in an ischemic brain area is a reflection of the extent of collateral circulation which plays an important role in the survival of brain cells in the ischemic region.

The main methodological contributions of this thesis are aimed at the improvement of cerebral vessel extraction (Chapter 2 and **??**), and cerebral anterior trees extraction (Chapter 5). The extracted vessel structures in Chapter 2 were used to derive a collateral score. From Chapter 2 to Chapter 4, we focused on the design, modeling, and validation of automated collateral score algorithm using large multicenter clinical datasets.

In Chapter 2 and **??** contain two methods for collateral scoring with the intention to ultimately replace the repetitive laborious process of manual scoring, which is prone to inter- and intraobserver variability. In Chapter 2, I used a conventional U-net vessel segmentation approach to quantify the vessel features of the occluded side and non-occluded side. The output of this method is a floating point score. In **??**, an end-to-end Siamese model to compare the occluded hemisphere and the contralateral side was used. The output of this method is a categorical scale. Both methods showed performance comparable to experienced radiologists. Both methods have similar computation times due to the registration in the prepossessing step. The idea to develop the end-to-end approach was to let the network learn the relevant content by itself. However, in previous experiences of deep reinforcement learning(Chapter 5), using only image information performs less well than the local feature-guided image information in the tracking problem. Also, in **??**, the preprocessing step was designed to make the vessel structure more pronounced, in order for the network to more easily learn the difference between two hemispheres with 'attention' to vessel structure. At last, the term 'end-to-end' is a relative term. It depends on whether the pre-processing step was taken into account. For future work of image biomarker extraction, the automatic selection of 'attention' regions/features might be more relevant than focusing on 'end-to-end' approaches. In terms of performance, both methods achieve expert-level performance with sufficiently large evaluation datasets in a well-defined study. The end-to-end approach has slightly better but not statistically significant performance in the categorical collateral scores in comparison with the method proposed in Chapter 2. Other methods in the literature do not provide detailed explanations of the study design and data distribution. Therefore, it's difficult to compare.

The question remains, which method is best for implementation in clinical practice. My answer to this question is related to the current domain gaps between clinicians and machine learning experts. Machine learning based methods have been applied to clinical research for a long time, however, barriers do exist to bringing the machine learning based methods from research to the clinic. One of the barriers is the lack of explainability in many machine learning based methods [127]. People would rather use a tool they can understand. The end user of the machine learning based imaging biomarker methods is a health care professional. Therefore, it is essential for us to build tools that are understandable by clinical professionals. The automated vessel segmentation-based collateral scoring method developed in this thesis is a very

straightforward way to compute the collateral score. The intermediate steps, such as a segmentation of the vasculature, can be visualized. The vessel feature and the subsequent collateral score are then easy to interpret. In contrast to this, the end-to-end approach in ?? is more like a black-box approach. Only input and output are known. The only intermediate step are the feature maps for each categorical score. Such feature maps are difficult to interpret. Therefore, the vessel segmentation based method which is introduced in Chapter 2 would be more appropriate for clinical usage at this moment even if the performance would be worse.

One of the reasons for developing an automatic biomarker method was to investigate whether an automatic collateral score instead of a human score could better reflect the vessel status at the prognosis of a patient and, thus, improves the performance of a treatment outcome prediction model. To this end, we have replaced the manual collateral score with the automatic collateral score in the existing MR PREDICT model([128]). The result demonstrated that there is no additional value (i.e. improvement in outcome prediction) when using automatic collateral scoring. The model with the automatic collateral scoring method had a similar performance to the model with collateral score assessed by experienced radiologists. The visual scoring was evaluated by a corelab which consists of experienced radiologists. Therefore, the improvement of the model with automatic collateral scoring was not significant in comparison to the model with the visually assessed collateral scoring. Especially in the binary collateral score (good or bad collateral), both experienced radiologists and the automatic collateral score method achieve good accuracy. In the MR PREDICT, the strongest predictors were clinical variables such as age, baseline NIHSS, and systolic blood pressure. In Thrive-C, the collateral score was not used. It indicates that the collateral score itself is less important in the regression-based prediction model

One may thus conclude that automatic collateral scoring has added value in replacing human visual scoring, and the contribution to binary outcome prediction models is less. The role of automatically extracted radiological parameters is more like a tool that improves productivity.

In Chapter 3, the method introduced in Chapter 2 was compared with 29 radiologists and radiology residents. The performance of the automatic collateral scoring method was similar to the overall accuracy of these 29 raters. The performance of the rater did not depend on professional experience. Although the proposed automatic collateral scoring method achieved similar performance as the trained raters for visual collateral scoring, the proposed method is not used yet in routine clinical practice. The collateral score does however act as an inclusion criterion [83] or as one of the variables in clinical trials. The automatic collateral scoring method can therefore be used in clinical trials for screening of patients before inclusion to reduce labor costs and interobserver variability.

In Chapter 4 we demonstrated the usability of quantitative collateral scoring to assess the time dependency of assessing collateral scoring and to find the optimal acquisition time. In this work, we demonstrated that the collateral score is changing with the timing of data acquisition in relation to the contrast injection. The contrast acquisition phase has an impact on collateral scoring, as a CTA image is acquired at one time point in the contrast flow cycle. At this moment, both visual scoring and automatic collateral scoring do not take the contrast acquisition phase into account.

6

I used CT perfusion data to construct enhancement curves in order to evaluate the optimal time point for CTA data acquisition. Although the approximate optimal time was found using CT perfusion image, the collateral scores did not correlate better to baseline NIHSS or ASPECT than the collateral score from CTA. However, whether the currently used time point in CTA is optimal or not is still an open question. Further investigation into the acquisition timing in a larger data set is needed.

Supervised learning in general requires large sets of labeled data, and the labeling process is a costly and time-intensive procedure. In **??**, we investigated a method that can effectively use existing labels to improve segmentation results. The proposed method is easy to use as a standalone post-processing step for tubular structure segmentation tasks, such as brain vessels or lung airways. This method demonstrated the possibility of utilizing the explicitly designed synthetic structure error to further improve the tubular structure. The proposed method in **??** does statistically significantly improve lung airway segmentation results. However, the performance in the vessel centerline segmentation task is better but not statistically significantly better than the original result that was introduced in Chapter 2.

In Chapter 5, I demonstrated the feasibility of extracting anterior vessel tree from 3D CTA images. In this chapter, I explored a deep reinforcement learning proximal policy optimization method for cerebral vessel tracking. I provided insight into the model architecture, and the formulation of the reward function. I also provided a detailed description of the data distribution of the anterior tree, the bifurcation detection, and the tree formation itself. In this work, I learned that image information only yields sub-optimal results in a tracking approach, and image information with dynamic local features helps the model better follow the track path. Also, I proposed an elegant reward function that can be widely applied to any curvature tracking problems. This work might be of value for other researchers applying deep reinforcement learning to medical image tracking problems. In addition, the performance of deep reinforcement learning tracking approach achieved state of art performance and was shown to provide good tracking results in complex environment. This suggests that such a tracker can be used in the semi-automatic annotation task to generate vessel centerline in an efficient manner. The output of the anterior tree extraction method is a graph representation, such that the vessel feature information, such as, vessel length, number of bifurcations, and depth of vessel trees can be easily extracted.

Summarizing, the contribution of the thesis is in the demonstration of complete flow of design, feature engineering, modelling, and validation of learning based algorithms that may replace the intensive laborious task of manual labeling collateral score. In addition, it presents a data efficient approach to improve vessel segmentation results and a deep reinforcement learning approach to extract anterior trees from 3D CTA images.

Summary

In the last few years, the blooming of learning based methods has brought reincarnation to the medical image analysis field. Classical methods were rapidly replaced by many kinds of learning based methods. This thesis has explored biomarker extraction methods and applications in ischemic stroke. Such biomarkers may be used in treatment decision making. The first chapters of this thesis focuses on collateral scoring, including model design, the assessment of model performance and the clinical use of the automatic collateral scoring method. In of the last chapters this thesis, a general post-processing method that improves binary segmentation results for the tubular structure was discussed, and a method that extracts the anterior trees from CTA images was developed and thoroughly assessed.

In **Chapter 2**, we designed a method for collateral scoring that follows the human visual collateral scoring approach. The proposed method consists of three steps. First, the brain region is defined using atlas based registration. Next, a 3D U-net for vessel segmentation is applied. Finally, the collateral score is determined using the median ratio of vessel features on the occluded side versus the contralateral side. The obtained collateral score is a floating point score. The floating point scores were converted into categorical scales using a simple regression model or random forest classifier. The performance of the collateral scoring method was assessed in a randomly sampled subset of a large multicenter registry dataset (MR CLEAN Registry). The subset consist of 270 subjects and the collateral reference standard was a consensus score obtained from three experienced radiologists. The performance of the proposed method is comparable to an experienced radiologist. In addition, a manually labeled cerebral vessel centerline dataset was created for training the vessel segmentation model.

In **Chapter 3**, we validate the performance of the proposed method in Chapter 2 with 29 raters (either radiologist or radiology resident). There is no statistically significant difference between the accuracy of human raters and the accuracy of the automatic collateral scoring method.

In **Chapter 4**, we use the floating point automatic collateral scoring method from Chapter 2 to investigate the optimal contrast acquisition time point for CTA images by computing the CS over all timepoints of CTP and mCTA images. This information is used to temporarily align CTP and mCTA images of the same subjects. The study shows that collateral scores greatly depend on the timing of the acquisition, that mCTA images can be accurately timed, and that, when using baseline parameters as a metric, a CTP image does not have added value for collateral scoring; a good-timed

CTA image is sufficient.

In **Chapter 5**, a method for extracting the cerebral anterior vessel tree was presented. The method consists of three parts, a policy gradient based deep reinforcement learning (DRL) tracker, a CNN based bifurcation detection, and a classical breadth-first tree formation method. In the DRL tracking approach, we adopted a proximal policy optimization deep reinforcement learning method in an online fashion. A curve-to-curve distance metric reward function was proposed. We also investigated the network architecture configuration and the training scheme. In addition, we proposed a metric to better monitor the training process of the DRL tracker rather than relying on the accumulated reward. The bifurcation detection used a similar architecture but with switchable normalization after each convolutional layer. At last, a breath-first tree formation method with a tracking ensemble method was proposed to extract the anterior tree. The application of anterior tree extraction is novel in the field and the performance of DRL tracker achieves state of art performance. Moreover, a dataset of manually labeled cerebral anterior trees over 125 subjects (randomly selected from MR CLEAN Registry) was created for assessing and training the proposed method.

Samenvatting

In de afgelopen jaren heeft de bloei van op leren gebaseerde methoden een wed ergeboorte teweeggebracht in de medische beeldanalyse. Klassieke methode werden snel vervangen door allerlei op leren gebaseerde methoden. Dit proefschrift heeft methoden voor het bepalen van biomarkers in ischemische beroerten onderzocht. Zulke biomarkers kunnen bijv. gebruikt worden voor besluitvorming rond de behandeling. De eerste hoofdstukken van dit proefschrift gaan over het bepalen van een score voor collateralen; zowel het model ontwerp, het onderzoeken van de performance, en de klinische toepassing. In de laatste hoofdstukken van het proefschrift introduceren en evalueren we een generieke methode om de binaire segmentatie resultaten voor het segmenteren van buisvormige structuren, en een methode die de anterieure vaatbomen bepaalt in CTA beelden.

In **Hoofdstuk 2** hebben we een model om de score voor collateralen te bepalen, ontworpen. Het model volgt de menselijke manier van scoren. De voorgestelde methode bestaat uit drie stappen. Eerst wordt via een atlas registratie het brein bepaald. Vervolgens wordt een 3D U-Net voor vaatsegmentatie toegepast. En tenslotte wordt de score bepaald door de mediaan van de ratio's van vaat kenmerken van de geoccludeerde kant en de contra-laterale kant. Deze scores zijn reëel getallen, en conversie naar een categorische schaal werd gedaan via regressie of een 'random forest' model. De nauwkeurigheid van de scores is geëvalueerd in de beelden van 270 patiënten, die op een willekeurige manier gekozen zijn uit een grote, multi-centrische verzameling van beelden (MR Clean Registry). De referentie scores waren bepaald door drie ervaren radiologen, en de automatische methode heeft resultaten die vergelijkbaar zijn met de menselijke scores. Voor het trainen van het vaat segmentatie model was ook een handmatige annotatie van de vaten gemaakt.

In **Hoofdstuk ??** hebben we een end-to-end model voor het bepalen van een collaterale score ontwikkeld. De methode beschouwt het bepalen van een collaterale score als een classificatie probleem; een Siamees model wordt gebruikt om kenmerken uit het beeld te halen, en de kant met de occlusie te vergelijken met de andere kant. Om de vergelijking makkelijk te maken, werd een atlas gebruikt om de linker en rechter hersenhelft te bepalen. De methode bepaalt hiermee een score met vier mogelijke waarden. De methode is geëvalueerd met beelden van MR Clean studie. De nauwkeurigheid van deze methode is vergelijkbaar met de methode uit hoofdstuk 2.

In **Hoofdstuk 3** vergelijken we de resultaten van de methode uit hoofdstuk 2 met de scores die 29 radiologen en radiologen in opleiding hebben bepaald. Er was geen

statistisch significant verschil in nauwkeurigheid tussen de scores van de radiologen en radiologen in opleiding enerzijds, en de automatische methode anderzijds.

In **Hoofdstuk 4** onderzoeken we het beste tijdstip (na contrast toediening) voor het maken van een CTA afbeelden. We gebruiken hiervoor de methode uit hoofdstuk 2 en bepalen voor alle tijdstippen van een CTP beeld en van een mCTA beeld een collaterale score. Met behulp van deze scores worden de CTP en mCTA beelden langs de tijdsas naast elkaar gezet. De studie laat zien dat de collateralen score erg afhankelijk is van het moment waarop het beeld gemaakt wordt, dat dat moment voor mCTA beelden nauwkeurig gekozen kan worden, en dat, als we de collateralen score vergelijken met andere parameters van de patiënt voor een interventie, een CTP beeld geen toegevoegde waarde heeft: een CTA beeld genomen op het juiste moment is voldoende.

In **Hoofdstuk ??** wordt een generieke methode om het resultaat van segmentatie van buisvormige structuren (zoals luchtwegen in de longen, en vaten in het brein) te verbeteren geïntroduceerd. Om de methode te leren een eerste segmentatie te verbeteren, wordt de methode getraind met beelden met synthetische fouten, gegenereerd door een GAN. Deze methode kan gebruikt worden als een toevoeging op een eerste segmentatie aanpak. De buisvormige structuren bepaald door deze methode zijn statistisch significant completer en hebben een betere onderlinge verbinding. In deze studie is zijn de handmatige intekeningen gemaakt in hoofdstuk 2 gebruikt.

In **Hoofdstuk 5** wordt een methode voor de het bepalen van de anterieure vaatboom in de hersenen gepresenteerd. De methode bestaat uit drie delen: een deep reinforcement learning gebaseerde methode om een pad in een beeld te bepalen, een CNN gebaseerde methode om bifurcaties te detecteren, en een klassieke breedte-eerst methode om een boomstructuur te maken. Een maat om de afstand tussen twee krommen te bepalen is gebruikt als beloningsfunctie. We hebben de netwerk architectuur en het trainingsschema onderzocht. Daarnaast hebben we een methode voorgesteld om beter het trainingsproces van de DRL methode in de gaten te houden (en niet alleen te kijken naar de beloningsfunctie). De CNN gebaseerde bifurcatie detectie gebruikt een vergelijkbare architectuur, met een instelbare normalisatie na elke convolutie laag. De breedte-eerst boom constructie met een de resultaten van een aantal DRL resultaten is uiteindelijk gebruikt om de vaatboom te beplane. Deze toepassing was nieuw, en het resultaat is vergelijkbaar met andere methoden. Voor deze studie is een handmatige annotaties van de vaatbomen in 125 subjects (een willekeurige selectie uit de MR Clean registry) gemaakt om de methode te trainen, en te evalueren.

Acknowledgements

Pursuing a PhD was a unique chapter in my life. Looking back, many individuals have significantly influenced who am I today. In the acknowledgement section, which is considerably the most sentimental section in the whole thesis, I want to express my gratitude to those who have crossed my path and contributed to my growth.

Throughout my PhD journey, I was truly fortunate. I was mentored by the most exclusive supervisory team, entrusted with a compelling topic within a prominent clinical consortium. I had privilege of collaborating with exceptionally talented peers in the group, and the invaluable assistance of a clinical partner who bridged the gap in clinical terminology. First and foremost, I would like to express my deepest gratitude to my three supervisor and promoters.

Aad, I'm deeply grateful for your guidance and support over the years. You've consistently been there for me whenever I needed assistance. While I was once drawn to chasing the lasted tech trends, you enlightened me about the true essence of science. You taught me that the groundbreaking technologies are crafted to address real-world clinical issues, instead of numerical performance competition. I will always remember all those Wednesday morning meetings.

Theo, I think the best way I can complement a supervisor is by saying that I aspire to lead a group the way you do in the future. Beyond your vast expertise in medical image analysis, you've given me the freedom and trust to delve into my areas of interest. Your guidance has been patient and invaluable. I've learned so much from you, like questioning the clinical impact of a study whenever I review a paper, as well as influencing my future career decisions.

Wiro, I'm truly grateful for the time and attention you've dedicated to me. In the first year of my PhD, our regular meetings were invaluable. Whenever I sought feedback, you were always ready with insightful comments. Your journey as a distinguished scholar offers a wealth of lessons for us who follow in your footsteps. Moreover, your perspectives on AI are immensely valuable.

I would like to extend my gratitude to my doctoral committee members: **Hester Lingsma**, **Ivana Išgum**, **Ihor Smal**, **Wim van Zwam**, **Henk Marquering**. I appreciate your time spent reviewing my thesis and participating in my defense.

Lennard, you hold a unique spot in my journey. I am grateful for our past collaborations, your invaluable assistance in translating clinical terminology, and for being such a good friend. Lastly, thank you for being one of the paranymphs in my defense.

I'd also like to thank those who attended the stroke research meetings, delivering insightful presentations and offering invaluable feedback on my work: **Daniel**, **Rob**, **Sven**, **Martijne**, **Nadinda**, **Nikki**, **Noor**.

I also want to acknowledge the individuals I met in BIGR, thank you for being so kind to me in the past. **Bo**: Being my first office mate at BIGR, I feel we share similar views on responsibility and technical progression in academia. I truly admire the drive and courage that propels you forward. **Chaoping**: I've always appreciated your serene demeanor, as if nothing could ruffle you. Conversations with you have always brought a sense of calm to me. **Hua**: I value the moments we shared venting about work challenges. Your unique brand of sarcastic humor never failed to bring a smile to my face. **Kim**: Incredibly passionate and determined. I was genuinely impressed by your organizational prowess during the PhD dinner planning. **Riwaj**: I cherished the moments we shared as office mates and the casual conversations we had over time. **Rick**(previous Gerda): I enjoyed the moments when we were neighbors in the open area and when we shared the same office space. I must admit, our spot was probably the most cluttered corner in the entire open space! **Ruisheng**: I truly valued the time working with you. Your determination and knack for resource optimization never cease to impress me. **Shengnan**: I enjoyed the times we hung out together, especially when we delved into discussions about the complexities of data collection. **Shuai**(Chen): I like how you're easily captivated by the latest tech trends. Chatting with you is always a blast! **Sijie**: Even though you were here for just about a year, spending time with you during social events was truly delightful. **Xianjing** (Lau): You always have smile in your face. You have this magical way of brightening the mood. Though, I've yet to understand why people refer to you as Lau. **Yao**: I'm truly in awe of your determination. Balancing career, family, and a PhD is no small feat, yet you've managed to succeed in all three. **Zahra**: Your kindness is contagious, affecting everyone in your vicinity. From our last meeting, I noticed that your baby girl has already caught this wonderful "infection" from you.

I would also like to thank all my other BIGR colleagues: **Abdullah**, **Adriaan**, **Arno**, **Danilo**, **Dirk**, **Esther**, **Gennady**, **Gijs**, **Gokhan**, **Hakim**, **Jifke**, **Jose**, **Luisa**, **Marcel**, **Martijn**, **Mohamed**, **Pierre**, **Robin**, **Sebastian**, **Stefan**, **Thomas**, **Vikram**, **Wietzke**, **Yuanyuan**. Thank you for enhancing my PhD experience.

Many thanks to **Jeffrey**, **Laurens**, and **Mart** for the IT support.

Many thanks to **Andreas**, **Annemarijn**, **Desiree**, and **Petra** for all the administration support over the years.

Special gratitude goes to **Igor** and **Albert** for guiding me through my master's and bachelor's thesis. Those achievements were instrumental in my decision to delve into PhD research. **Albert**, you laid the groundwork for my robust coding practices and foundational project management skills. The success of the project aroused my interest in the field of engineering. **Igor**, you've imparted a wealth of knowledge to me, ranging from how to adeptly communicate with senior members of the group to the essential takeaway of always trusting one's intuition. Above all, the vital lesson of maintaining 'guts' stands out.

Being social creatures by nature, I find that friendships are essential to my well-being. I deeply value the companionship and support I've received along the journey. **Lisha**, distance has never weakened the bond of our friendship. **Xiaofei** and **Hin**

Shiang: thanks for having Windsor (biological brother and rival of Panda, as well as the admirer of Yuri) stay with us. **Siyu**, like a big sister to me. **Mei**, Loved all the hangout sessions we've had! **Ying** (Tutu), foodie buds who pack on the pounds together, and the proud parent of Yuri's sister. And many more. I'm really thankful for the time we've shared, chatting about both the highs and lows in our lives.

I also want to extend a heartfelt thanks to my parents for their rock-solid support. You've not only allowed me the freedom to follow my passions but also gave me the space to take breaks when necessary. Your unfailing backing has been the driving force propelling me onward!

Finally, **Shuai** (Bear), my co-author, my paranymph, and my partner in caring for Nini, Panda, Yuri, and our two goldfish—my deepest thanks go out to you. Your ultimate support has been invaluable. Thank you for giving me the freedom to pursue my interests. We've already co-authored several chapters of our lives together, and I look forward to co-writing many more over the next decades.

Nini (passed away)   **Panda**   **Yuri**
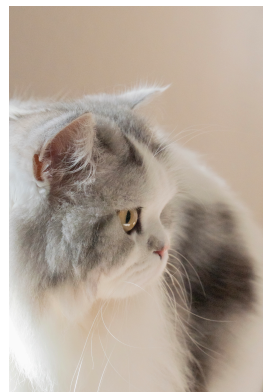
About the author

Jiahang Su, born on November 14, 1988, in Linyi, Shandong, China. In 2010, she relocated to the Netherlands to pursue her studies in Electrical Engineering at Fontys University of Applied Science, Eindhoven. During this study, she completed two internships at Philips Innovation Service and Philips Lighting where she was working on the wireless sensor networks and error correction algorithms with a commercialized FPGA platform.

In 2014, she went to the Eindhoven University of Technology, located just across the street, to specialize in Electrical Engineering with a focus on statistical signal processing. During master study, she was attracted by the medical image analysis. This new found passion lead her to an internship at Philips Research, where she focused on the vessel segmentation from 3D high resolution MR images. She obtained master of science diploma in 2017.

On November 15, 2017, Jiahang began her PhD at the Biomedical Imaging Group Rotterdam (BIGR) under the guidance of Dr. T. van Walsum, Prof. A. van der Lugt, and Prof. W.J. Niessen. Here, she focused on the vessel biomarker development and assessment in CONTRAST consortium. Throughout this academic journey, she developed board interest in understanding the disease progression using various machine learning techniques. And she will continue her journey in such domain. For more detail, please visit her Linkedin Page(`https://www.linkedin.com/in/jiahang-su-88918378/`).

Publications

Journal Papers

V. Fortunati, **J. Su**, L. Wolff, P. J. van Doormaal, J. Hofmeijer, R. P. Bokkers, W. van Zwam, A. van der Lugt, and T. van Walsum, "Siamese model for collateral score prediction from cta images in acute ischemic stroke," *Submitted*.

L. Wolff, E. Venema, **J. Su**, W. van Zwam, B. Roozenbeek, D. W. Dippel, C. B. Majoie, T. van Walsum, H. F. Lingsma, A. van der Lugt, and for the MR CLEAN Investigators, "Performance of automated imaging biomarkers in an outcome prediction model for patients with ischemic stroke," *Submitted*.

S. Liu, R. Su, **J. Su**, J. Xin, J. Wu, P. J. van Doormaal, A. van der Lugt, W. Niessen, N. Zheng, and T. van Walsum, "An automated framework for brain vessel centerline extraction from cta images," *Submitted*.

S. Chen[*], A. G. Uceda[*], **J. Su**[*], G. van Tulder, L. Wolff, T. van Walsum, and M. de Bruijne, "Label refinement network from synthetic error augmentation for medical image segmentation," *Submitted*. arXiv: `2209.06353`.

J. Su, S. Li, L. Wolff, W. van Zwam, W. J. Niessen, A. van der Lugt, and T. van Walsum, "Deep reinforcement learning for cerebral anterior vessel tree extraction from 3d cta images," *Medical Image Analysis*, p. 102 724, 2022.

J. Su, L. Wolff, P. J. van Doormaal, D. W. Dippel, W. van Zwam, W. J. Niessen, A. van der Lugt, and T. van Walsum, "Time dependency of automated collateral scores in computed tomography angiography and computed tomography perfusion images in patients with intracranial arterial occlusion," *Neuroradiology*, pp. 1–10, 2022.

L. Wolff, **J. Su**, D. Van Loon, A. van Es, P. J. van Doormaal, C. Majoie, W. van Zwam, D. Dippel, A. van der Lugt, and T. van Walsum, "Inter-rater reliability for assessing intracranial collaterals in patients with acute ischemic stroke: Comparing

29 raters and an artificial intelligence-based software," *Neuroradiology*, pp. 1–8, 2022.

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.

* indicates equal contributions

PhD portfolio

Courses	Year	ECTS
Advanced Pattern Recognition *Delft University of Technology, The Netherlands*	2018	4
Medical Image Summer School *Sicily, Italy*	2018	4
Computer Vision by Learning *University of Amsterdam, The Netherlands*	2019	4
Scientific Integrity *Erasmus MC, The Netherlands*	2019	0.3
English Biomedical Writing and Communication *Erasmus MC, The Netherlands*	2020	3
Study Design *Erasmus MC, The Netherlands*	2020	4.3
Advanced Decision Making In Vascular Care *Erasmus MC, The Netherlands*	2020	0.5
Machine Learning *Stanford University(online), USA*	2021	4
Probabilistic Graphical Models *Stanford University(online), USA*	2021	3
Total		27.1

International and local research meetings	Year	ECTS
Medical Imaging Symposium for PhD students *Erasmus MC, The Netherlands*	2020	0.3
European Congress of Radiology *Vienna, Austria*	2021	1
Dutch Bio-Medical Engineering Conference *Eindhoven, The Netherlands*	2021	1
Biomedical imaging group seminars (biweekly) *Erasmus MC, The Netherlands*	2017 – 2022	1
Medical Informatics research lunch meeting (biweekly) *Erasmus MC, The Netherlands*	2017 – 2020	1
Stroke research meetings (biweekly) *Erasmus MC, The Netherlands*	2017 – 2022	1
Total		5.3

Teaching activities	Year
Advanced Image Processing *Delft University of Technology, The Netherlands*	2020 – 2021

Grants & Awards	Year
NVidia GPU grant (together with Dr. Theo van Walsum)	2018
Pilot grant SURFsara (800,000 billing units)	2018 – 2022

Other Activities	Year
Radiology dinner Organizer *Organizing member*	2020

Bibliography

[1] M. E. Raichle and D. A. Gusnard, "Appraising the brain's energy budget," *Proceedings of the National Academy of Sciences*, vol. 99, no. 16, pp. 10 237–10 239, 2002.

[2] T. Vos, S. S. Lim, C. Abbafati, K. M. Abbas, M. Abbasi, M. Abbasifard, M. Abbasi-Kangevari, H. Abbastabar, F. Abd-Allah, A. Abdelalim, *et al.*, "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the global burden of disease study 2019," *The Lancet*, vol. 396, no. 10258, pp. 1204–1222, 2020.

[3] C. W. Tsao, A. W. Aday, Z. I. Almarzooq, A. Alonso, A. Z. Beaton, M. S. Bittencourt, A. K. Boehme, A. E. Buxton, A. P. Carson, Y. Commodore-Mensah, *et al.*, "Heart disease and stroke statistics—2022 update: A report from the american heart association," *Circulation*, vol. 145, no. 8, pp. e153–e639, 2022.

[4] N. Lakomkin, M. Dhamoon, K. Carroll, I. P. Singh, S. Tuhrim, J. Lee, J. T. Fifi, and J. Mocco, "Prevalence of large vessel occlusion in patients presenting with acute ischemic stroke: A 10-year systematic review of the literature," *Journal of neurointerventional surgery*, vol. 11, no. 3, pp. 241–245, 2019.

[5] J. C. Hemphill III, S. M. Greenberg, C. S. Anderson, K. Becker, B. R. Bendok, M. Cushman, G. L. Fung, J. N. Goldstein, R. L. Macdonald, P. H. Mitchell, *et al.*, "Guidelines for the management of spontaneous intracerebral hemorrhage: A guideline for healthcare professionals from the american heart association/american stroke association," *Stroke*, vol. 46, no. 7, pp. 2032–2060, 2015.

[6] D. Rodriguez-Luna, D. Dowlatshahi, R. I. Aviv, C. A. Molina, Y. Silva, I. Dzialowski, C. Lum, A. Czlonkowska, J.-M. Boulanger, C. S. Kase, *et al.*, "Venous phase of computed tomography angiography increases spot sign detection, but intracerebral hemorrhage expansion is greater in spot signs detected in arterial phase," *Stroke*, vol. 45, no. 3, pp. 734–739, 2014.

[7] 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.

[8]   G. W. Albers, M. P. Marks, S. Kemp, S. Christensen, J. P. Tsai, S. Ortega-Gutierrez, R. A. McTaggart, M. T. Torbey, M. Kim-Tenser, T. Leslie-Mazwi, *et al.*, "Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging," *New England Journal of Medicine*, vol. 378, no. 8, pp. 708–718, 2018.

[9]   R. G. Nogueira, A. P. Jadhav, D. C. Haussen, A. Bonafe, R. F. Budzik, P. Bhuva, D. R. Yavagal, M. Ribo, C. Cognard, R. A. Hanel, *et al.*, "Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct," *New England Journal of Medicine*, vol. 378, no. 1, pp. 11–21, 2018.

[10] 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.

[11] M. Goyal, A. M. Demchuk, B. K. Menon, M. Eesa, J. L. Rempel, J. Thornton, D. Roy, T. G. Jovin, R. A. Willinsky, B. L. Sapkota, *et al.*, "Randomized assessment of rapid endovascular treatment of ischemic stroke," *New England Journal of Medicine*, vol. 372, no. 11, pp. 1019–1030, 2015.

[12] J. L. Saver, M. Goyal, A. Bonafe, H.-C. Diener, E. I. Levy, V. M. Pereira, G. W. Albers, C. Cognard, D. J. Cohen, W. Hacke, *et al.*, "Stent-retriever thrombectomy after intravenous t-pa vs. t-pa alone in stroke," *New England Journal of Medicine*, vol. 372, no. 24, pp. 2285–2295, 2015.

[13] B. C. Campbell, P. J. Mitchell, T. J. Kleinig, H. M. Dewey, L. Churilov, N. Yassi, B. Yan, R. J. Dowling, M. W. Parsons, T. J. Oxley, *et al.*, "Endovascular therapy for ischemic stroke with perfusion-imaging selection," *New England Journal of Medicine*, vol. 372, no. 11, pp. 1009–1018, 2015.

[14] T. G. Jovin, A. Chamorro, E. Cobo, M. A. de Miquel, C. A. Molina, A. Rovira, L. San Román, J. Serena, S. Abilleira, M. Ribó, *et al.*, "Thrombectomy within 8 hours after symptom onset in ischemic stroke," *New England Journal of Medicine*, vol. 372, no. 24, pp. 2296–2306, 2015.

[15] 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.

[16] J. J. Smith, A. G. Sorensen, and J. H. Thrall, "Biomarkers in imaging: Realizing radiology's future," *Radiology*, vol. 227, no. 3, pp. 633–638, 2003.

[17] W. J. Powers, A. A. Rabinstein, T. Ackerson, O. M. Adeoye, N. C. Bambakidis, K. Becker, J. Biller, M. Brown, B. M. Demaerschalk, B. Hoh, *et al.*, "Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: A guideline for healthcare professionals from the american heart association/american stroke association," *Stroke*, vol. 50, no. 12, pp. e344–e418, 2019.

[18] I. Tan, A. Demchuk, J. Hopyan, L. Zhang, D. Gladstone, K. Wong, M. Martin, S. Symons, A. Fox, and R. Aviv, "Ct angiography clot burden score and collateral score: Correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct," *American Journal of Neuroradiology*, vol. 30, no. 3, pp. 525–531, 2009.

[19] P. A. Barber, A. M. Demchuk, J. Zhang, A. M. Buchan, A. S. Group, *et al.*, "Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy," *The Lancet*, vol. 355, no. 9216, pp. 1670–1674, 2000.

[20] J. C. Tan, W. P. Dillon, S. Liu, F. Adler, W. S. Smith, and M. Wintermark, "Systematic comparison of perfusion-ct and ct-angiography in acute stroke patients," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 61, no. 6, pp. 533–543, 2007.

[21] 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.

[22] D. S. Liebeskind, "Collateral circulation," *Stroke*, vol. 34, no. 9, pp. 2279–2284, 2003.

[23] O. A. Berkhemer, I. G. Jansen, D. Beumer, P. S. Fransen, L. A. Van Den Berg, A. J. Yoo, H. F. Lingsma, M. E. Sprengers, S. F. Jenniskens, G. J. Lycklama à Nijeholt, *et al.*, "Collateral status on baseline computed tomographic angiography and intra-arterial treatment effect in patients with proximal anterior circulation stroke," *Stroke*, vol. 47, no. 3, pp. 768–776, 2016.

[24] F. Seker, A. Potreck, M. Möhlenbruch, M. Bendszus, and M. Pham, "Comparison of four different collateral scores in acute ischemic stroke by ct angiography," *Journal of neurointerventional surgery*, vol. 8, no. 11, pp. 1116–1118, 2016.

[25] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys (CSUR)*, vol. 36, no. 2, pp. 81–121, 2004.

[26] D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3d vessel lumen segmentation techniques: Models, features and extraction schemes," *Medical image analysis*, vol. 13, no. 6, pp. 819–845, 2009.

[27] M. Meijs, A. Patel, S. C. van de Leemput, M. Prokop, E. J. van Dijk, F.-E. de Leeuw, F. J. Meijer, B. van Ginneken, and R. Manniesing, "Robust segmentation of the full cerebral vasculature in 4d ct of suspected stroke patients," *Scientific reports*, vol. 7, no. 1, pp. 1–12, 2017.

[28] R. Manniesing, B. K. Velthuis, M. S. van Leeuwen, I. C. van der Schaaf, P. Van Laar, and W. J. Niessen, "Level set based cerebral vasculature segmentation and diameter quantification in ct angiography," *Medical image analysis*, vol. 10, no. 2, pp. 200–214, 2006.

[29] M. Schaap, R. Manniesing, I. Smal, T. v. Walsum, A. v. d. Lugt, and W. Niessen, "Bayesian tracking of tubular structures and its application to carotid arteries in cta," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2007, pp. 562–570.

[30] D. Robben, E. Türetken, S. Sunaert, V. Thijs, G. Wilms, P. Fua, F. Maes, and P. Suetens, "Simultaneous segmentation and anatomical labeling of the cerebral vasculature," *Medical image analysis*, vol. 32, pp. 201–215, 2016.

[31] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics," *Computer methods and programs in biomedicine*, vol. 158, pp. 71–91, 2018.

[32] P. Sanchesa, C. Meyer, V. Vigon, and B. Naegel, "Cerebrovascular network segmentation of mra images with deep learning," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, 2019, pp. 768–771.

[33] M. Livne, J. Rieger, O. U. Aydin, A. A. Taha, E. M. Akay, T. Kossen, J. Sobesky, J. D. Kelleher, K. Hildebrand, D. Frey, *et al.*, "A u-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease," *Frontiers in neuroscience*, vol. 13, p. 97, 2019.

[34] H. Kandil, A. Soliman, F. Taher, A. Mahmoud, A. Elmaghraby, and A. El-Baz, "Using 3-d cnns and local blood flow information to segment cerebral vasculature," in *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, IEEE, 2018, pp. 701–705.

[35] M. Meijs and R. Manniesing, "Artery and vein segmentation of the cerebral vasculature in 4d ct using a 3d fully convolutional neural network," in *Medical Imaging 2018: Computer-Aided Diagnosis*, SPIE, vol. 10575, 2018, pp. 394–399.

[36] G. Tetteh, V. Efremov, N. D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, and B. H. Menze, "Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes," *Frontiers in Neuroscience*, p. 1285, 2020.

[37] A. M. Boers, R. Sales Barros, I. G. Jansen, C. H. Slump, D. W. Dippel, A. v. d. Lugt, W. H. v. Zwam, Y. B. Roos, R. J. v. Oostenbrugge, C. B. Majoie, *et al.*, "Quantitative collateral grading on ct angiography in patients with acute ischemic stroke," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, Springer, 2017, pp. 176–184.

[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] R. Peter, B. J. Emmer, A. C. van Es, and T. van Walsum, "Cortical and vascular probability maps for analysis of human brain in computed tomography images," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, IEEE, 2017, pp. 1141–1145.

[40] K. J. Friston, "Statistical parametric mapping," in *Neuroscience databases*, Springer, 2003, pp. 237–250.

[41] S. N. Wright, P. Kochunov, F. Mut, M. Bergamino, K. M. Brown, J. C. Mazziotta, A. W. Toga, J. R. Cebral, and G. A. Ascoli, "Digital reconstruction and morphometric analysis of human brain arterial vasculature from magnetic resonance angiography," *Neuroimage*, vol. 82, pp. 170–181, 2013.

[42] 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.

[43] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No new-net," in *International MICCAI Brainlesion Workshop*, Springer, 2018, pp. 234–244.

[44] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," in *International MICCAI Brainlesion Workshop*, Springer, 2017, pp. 287–297.

[45] A. L. Maas, A. Y. Hannun, A. Y. Ng, *et al.*, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, Atlanta, Georgia, USA, vol. 30, 2013, p. 3.

[46] B. Kayalibay, G. Jensen, and P. van der Smagt, "Cnn-based segmentation of medical imaging data," *arXiv preprint arXiv:1701.03056*, 2017.

[47] L. Risser, F. Plouraboué, and X. Descombes, "Gap filling of 3-d microvascular networks by tensor voting," *IEEE transactions on medical imaging*, vol. 27, no. 5, pp. 674–687, 2008.

[48] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, "Building skeleton models via 3-d medial surface axis thinning algorithms," *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, pp. 462–478, 1994.

[49] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Computational statistics & data analysis*, vol. 54, no. 4, pp. 1167–1178, 2010.

[50] D. Garcia, "A fast all-in-one method for automated post-processing of piv data," *Experiments in fluids*, vol. 50, no. 5, pp. 1247–1259, 2011.

[51] 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.

[52] A. T. Rai, J. P. Hogg, B. Cline, and G. Hobbs, "Cerebrovascular geometry in the anterior circulation: An analysis of diameter, length and the vessel taper," *Journal of neurointerventional surgery*, vol. 5, no. 4, pp. 371–375, 2013.

[53] M. Schaap, L. Neefjes, C. Metz, A. v. d. Giessen, A. Weustink, N. Mollet, J. Wentzel, T. v. Walsum, and W. Niessen, "Coronary lumen segmentation using graph cuts and robust kernel regression," in *International Conference on Information Processing in Medical Imaging*, Springer, 2009, pp. 528–539.

[54] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration," *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.

[55] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier, *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration," *Neuroimage*, vol. 46, no. 3, pp. 786–802, 2009.

[56] T. Tieleman, G. Hinton, *et al.*, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA: Neural networks for machine learning*, vol. 4, no. 2, pp. 26–31, 2012.

[57] P. Schramm, P. D. Schellinger, J. B. Fiebach, S. Heiland, O. Jansen, M. Knauth, W. Hacke, and K. Sartor, "Comparison of ct and ct angiography source images with diffusion-weighted imaging in patients with acute stroke within 6 hours after onset," *Stroke*, vol. 33, no. 10, pp. 2426–2432, 2002.

[58] I. Q. Grunwald, J. Kulikovski, W. Reith, S. Gerry, R. Namias, M. Politi, P. Papanagiotou, M. Essig, S. Mathur, O. Joly, *et al.*, "Collateral automation for triage in stroke: Evaluating automated scoring of collaterals in acute stroke on computed tomography scans," *Cerebrovascular Diseases*, vol. 47, no. 5-6, pp. 217–222, 2019.

[59] A. Arboix and J. Alió, "Acute cardioembolic cerebral infarction: Answers to clinical questions," *Current cardiology reviews*, vol. 8, no. 1, pp. 54–67, 2012.

[60] R. L. Sacco, S. E. Kasner, J. P. Broderick, L. R. Caplan, J. Connors, A. Culebras, M. S. Elkind, M. G. George, A. D. Hamdan, R. T. Higashida, *et al.*, "An updated definition of stroke for the 21st century: A statement for healthcare professionals from the american heart association/american stroke association," *Stroke*, vol. 44, no. 7, pp. 2064–2089, 2013.

[61] V. L. Feigin, M. H. Forouzanfar, R. Krishnamurthi, G. A. Mensah, M. Connor, D. A. Bennett, A. E. Moran, R. L. Sacco, L. Anderson, T. Truelsen, *et al.*, "Global and regional burden of stroke during 1990–2010: Findings from the global burden of disease study 2010," *The lancet*, vol. 383, no. 9913, pp. 245–255, 2014.

[62] A. A. Rabinstein, "Update on treatment of acute ischemic stroke," *Continuum: Lifelong Learning in Neurology*, vol. 26, no. 2, pp. 268–286, 2020.

[63] J. K. Holodinsky, A. Y. Yu, Z. A. Assis, A. S. Al Sultan, B. K. Menon, A. M. Demchuk, M. Goyal, and M. D. Hill, "History, evolution, and importance of emergency endovascular treatment of acute ischemic stroke," *Current neurology and neuroscience reports*, vol. 16, no. 5, pp. 1–14, 2016.

[64] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of chiropractic medicine*, vol. 15, no. 2, pp. 155–163, 2016.

[65] F. Kauw, J. W. Dankbaar, B. W. Martin, V. Y. Ding, D. B. Boothroyd, F. van Ommen, H. W. de Jong, L. J. Kappelle, B. K. Velthuis, J. J. Heit, *et al.*, "Collateral status in ischemic stroke: A comparison of computed tomography angiography, computed tomography perfusion, and digital subtraction angiography," *Journal of computer assisted tomography*, vol. 44, no. 6, p. 984, 2020.

[66] L. Souza, A. Yoo, Z. Chaudhry, S. Payabvash, A. Kemmling, P. Schaefer, J. Hirsch, K. Furie, R. González, R. Nogueira, *et al.*, "Malignant cta collateral profile is highly specific for large admission dwi infarct core and poor outcome in acute stroke," *American journal of neuroradiology*, vol. 33, no. 7, pp. 1331–1336, 2012.

[67] D. Weiss, B. Kraus, C. Rubbert, M. Kaschner, S. Jander, M. Gliem, J.-I. Lee, C.-A. Haensch, B. Turowski, and J. Caspers, "Systematic evaluation of computed tomography angiography collateral scores for estimation of long-term outcome after mechanical thrombectomy in acute ischaemic stroke," *The Neuroradiology Journal*, vol. 32, no. 4, pp. 277–286, 2019.

[68] A. Boers, R. S. Barros, I. Jansen, O. Berkhemer, L. Beenen, B. K. Menon, D. Dippel, A. van der Lugt, W. van Zwam, Y. Roos, *et al.*, "Value of quantitative collateral scoring on ct angiography in patients with acute ischemic stroke," *American journal of neuroradiology*, vol. 39, no. 6, pp. 1074–1082, 2018.

[69] D. S. Liebeskind, T. A. Tomsick, L. D. Foster, S. D. Yeatts, J. Carrozzella, A. M. Demchuk, T. G. Jovin, P. Khatri, R. von Kummer, R. M. Sugg, *et al.*, "Collaterals at angiography and outcomes in the interventional management of stroke (ims) iii trial," *Stroke*, vol. 45, no. 3, pp. 759–764, 2014.

[70] B. K. Menon, C. D. d'Esterre, E. M. Qazi, M. Almekhlafi, L. Hahn, A. M. Demchuk, and M. Goyal, "Multiphase ct angiography: A new tool for the imaging triage of patients with acute ischemic stroke," *Radiology*, vol. 275, no. 2, pp. 510–520, 2015.

[71] I. R. Van Den Wijngaard, G. Holswilder, M. Wermer, J. Boiten, A. Algra, D. Dippel, J. Dankbaar, B. Velthuis, A. Boers, C. Majoie, *et al.*, "Assessment of collateral status by dynamic ct angiography in acute mca stroke: Timing of acquisition and relationship with final infarct volume," *American Journal of Neuroradiology*, vol. 37, no. 7, pp. 1231–1236, 2016.

[72] B. Menon, E. Smith, J. Modi, S. Patel, R. Bhatia, T. Watson, M. Hill, A. Demchuk, and M. Goyal, "Regional leptomeningeal score on ct angiography predicts clinical and imaging outcomes in patients with acute anterior circulation occlusions," *American journal of neuroradiology*, vol. 32, no. 9, pp. 1640–1645, 2011.

[73] M. B. Maas, M. H. Lev, H. Ay, A. B. Singhal, D. M. Greer, W. S. Smith, G. J. Harris, E. Halpern, A. Kemmling, W. J. Koroshetz, *et al.*, "Collateral vessels on ct angiography predict outcome in acute ischemic stroke," *Stroke*, vol. 40, no. 9, pp. 3001–3005, 2009.

[74] E. Venema, B. Roozenbeek, M. J. Mulder, S. Brown, C. B. Majoie, E. W. Steyerberg, A. M. Demchuk, K. W. Muir, A. Dávalos, P. J. Mitchell, *et al.*, "Prediction of outcome and endovascular treatment benefit: Validation and update of the mr predicts decision tool," *Stroke*, vol. 52, no. 9, pp. 2764–2772, 2021.

[75] F. O. Lima, K. L. Furie, G. S. Silva, M. H. Lev, É. C. Camargo, A. B. Singhal,
      G. J. Harris, E. F. Halpern, W. J. Koroshetz, W. S. Smith, *et al.*, "Prognosis
      of untreated strokes due to anterior circulation proximal intracranial arterial
      occlusions detected by use of computed tomography angiography," *JAMA
      neurology*, vol. 71, no. 2, pp. 151–157, 2014.

[76] I. G. Jansen, M. J. Mulder, R.-J. B. Goldhoorn, A. M. Boers, A. C. van Es,
      S. Lonneke, J. Hofmeijer, J. M. Martens, M. A. van Walderveen, B. F. van
      der Kallen, *et al.*, "Impact of single phase ct angiography collateral status on
      functional outcome over time: Results from the mr clean registry," *Journal of
      neurointerventional surgery*, vol. 11, no. 9, pp. 866–873, 2019.

[77] S. B. Raymond and P. W. Schaefer, "Imaging brain collaterals: Quantification,
      scoring, and potential significance," *Topics in Magnetic Resonance Imaging*,
      vol. 26, no. 2, pp. 67–75, 2017.

[78] Z. Wang, J. Xie, T.-Y. Tang, C.-H. Zeng, Y. Zhang, Z. Zhao, D.-L. Zhao,
      L.-Y. Geng, G. Deng, Z.-J. Zhang, *et al.*, "Collateral status at single-phase and
      multiphase ct angiography versus ct perfusion for outcome prediction in anterior
      circulation acute ischemic stroke," *Radiology*, vol. 296, no. 2, pp. 393–400, 2020.

[79] E. Martinon, P. H. Lefevre, P. Thouant, G. V. Osseby, F. Ricolfi, and A.
      Chavent, "Collateral circulation in acute stroke: Assessing methods and impact:
      A literature review," *Journal of neuroradiology*, vol. 41, no. 2, pp. 97–107, 2014.

[80] F. McVerry, D. Liebeskind, and K. Muir, "Systematic review of methods for
      assessing leptomeningeal collateral flow," *American journal of neuroradiology*,
      vol. 33, no. 3, pp. 576–582, 2012.

[81] T. Brott, H. P. Adams Jr, C. P. Olinger, J. R. Marler, W. G. Barsan, J. Biller,
      J. Spilker, R. Holleran, R. Eberle, and V. Hertzberg, "Measurements of acute
      cerebral infarction: A clinical examination scale.," *Stroke*, vol. 20, no. 7, pp. 864–
      870, 1989.

[82] V. Chalos, R. A van de Graaf, B. Roozenbeek, A. CGM van Es, H. M den
      Hertog, J. Staals, L. van Dijk, S. FM Jenniskens, R. J van Oostenbrugge, W.
      H van Zwam, *et al.*, "Multicenter randomized clinical trial of endovascular
      treatment for acute ischemic stroke. the effect of periprocedural medication:
      Acetylsalicylic acid, unfractionated heparin, both, or neither (mr clean-med).
      rationale and study design," *Trials*, vol. 21, no. 1, pp. 1–17, 2020.

[83] F. Pirson, W. H. Hinsenveld, R.-J. B. Goldhoorn, J. Staals, I. R. de Ridder,
      W. H. van Zwam, M. A. van Walderveen, G. J. Lycklama à Nijeholt, M. Uytten-
      boogaart, W. J. Schonewille, *et al.*, "Mr clean-late, a multicenter randomized
      clinical trial of endovascular treatment of acute ischemic stroke in the nether-
      lands for late arrivals: Study protocol for a randomized controlled trial," *Trials*,
      vol. 22, no. 1, pp. 1–13, 2021.

[84] K. M. Treurniet, N. E. LeCouffe, M. Kappelhof, B. J. Emmer, A. C. van Es, J. Boiten, G. J. Lycklama, K. Keizer, L. S. Yo, H. F. Lingsma, *et al.*, "Mr clean-no iv: Intravenous treatment followed by endovascular treatment versus direct endovascular treatment for acute ischemic stroke caused by a proximal intracranial occlusion—study protocol for a randomized clinical trial," *Trials*, vol. 22, no. 1, pp. 1–15, 2021.

[85] E. Tong, J. Patrie, S. Tong, A. Evans, P. Michel, A. Eskandari, and M. Wintermark, "Time-resolved ct assessment of collaterals as imaging biomarkers to predict clinical outcomes in acute ischemic stroke," *Neuroradiology*, vol. 59, no. 11, pp. 1101–1109, 2017.

[86] F. Cademartiri, A. van der Lugt, G. Luccichenti, P. Pavone, and G. P. Krestin, "Parameters affecting bolus geometry in cta: A review," *Journal of computer assisted tomography*, vol. 26, no. 4, pp. 598–607, 2002.

[87] K. Schregel, I. Tsogkas, C. Peter, A. Zapf, D. Behme, M. Schnieder, I. L. Maier, J. Liman, M. Knauth, and M.-N. Psychogios, "Outcome prediction using perfusion parameters and collateral scores of multi-phase and single-phase ct angiography in acute stroke: Need for one, two, three, or thirty scans?" *Journal of Stroke*, vol. 20, no. 3, p. 362, 2018.

[88] C. A. Potter, A. S. Vagal, M. Goyal, D. B. Nunez, T. M. Leslie-Mazwi, and M. H. Lev, "Ct for treatment selection in acute ischemic stroke: A code stroke primer," *Radiographics*, vol. 39, no. 6, pp. 1717–1738, 2019.

[89] R. Damseh, P. Delafontaine-Martel, P. Pouliot, F. Cheriet, and F. Lesage, "Laplacian flow dynamics on geometric graphs for anatomical modeling of cerebrovascular networks," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 381–394, 2020.

[90] S. Moriconi, M. A. Zuluaga, H. R. Jäger, P. Nachev, S. Ourselin, and M. J. Cardoso, "Inference of cerebrovascular topology with geodesic minimum spanning trees," *IEEE transactions on medical imaging*, vol. 38, no. 1, pp. 225–239, 2018.

[91] Z. Zhang, D. Marin, M. Drangova, and Y. Boykov, "Confluent vessel trees with accurate bifurcations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9573–9582.

[92] Y. Zhang, G. Luo, W. Wang, and K. Wang, "Branch-aware double dqn for centerline extraction in coronary ct angiography," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2020, pp. 35–44.

[93] B. Jeon, "Deep recursive bayesian tracking for fully automatic centerline extraction of coronary arteries in ct images," *Sensors*, vol. 21, no. 18, p. 6087, 2021.

[94] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner, and I. Išgum, "Coronary artery centerline extraction in cardiac ct angiography using a cnn-based orientation classifier," *Medical image analysis*, vol. 51, pp. 46–60, 2019.

[95] Z. Guo, J. Bai, Y. Lu, X. Wang, K. Cao, Q. Song, M. Sonka, and Y. Yin, "Deep-centerline: A multi-task fully convolutional network for centerline extraction," in *International Conference on Information Processing in Medical Imaging*, Springer, 2019, pp. 441–453.

[96] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A brief survey of deep reinforcement learning," *arXiv preprint arXiv:1708.05866*, 2017.

[97] S. K. Zhou, H. N. Le, K. Luu, H. V. Nguyen, and N. Ayache, "Deep reinforcement learning in medical imaging: A literature review," *Medical image analysis*, vol. 73, p. 102 193, 2021.

[98] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

[99] P. Zhang, F. Wang, and Y. Zheng, "Deep reinforcement learning for vessel centerline tracing in multi-modality 3d volumes," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2018, pp. 755–763.

[100] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, 2016.

[101] Z. Li, Q. Xia, Z. Hu, W. Wang, L. Xu, and S. Zhang, "A deep reinforced tree-traversal agent for coronary artery centerline extraction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2021, pp. 418–428.

[102] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[103] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International conference on machine learning*, PMLR, 2015, pp. 1889–1897.

[104] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[105] T. Dai, M. Dubois, K. Arulkumaran, J. Campbell, C. Bass, B. Billot, F. Uslu, V. De Paola, C. Clopath, and A. A. Bharath, "Deep reinforcement learning for subpixel neural tracking," in *International conference on medical imaging with deep learning*, PMLR, 2019, pp. 130–150.

[106] D.-M. Baboiu and G. Hamarneh, "Vascular bifurcation detection in scale-space," in *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, IEEE, 2012, pp. 41–46.

[107] S. Kalaie and A. Gooya, "Vascular tree tracking and bifurcation points detection in retinal images using a hierarchical probabilistic model," *Computer methods and programs in biomedicine*, vol. 151, pp. 139–149, 2017.

[108] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, and D. Comaniciu, "3d deep learning for efficient and robust landmark detection in volumetric data," in *International conference on medical image computing and computer-assisted intervention*, Springer, 2015, pp. 565–572.

[109] Á. S. Hervella, J. Rouco, J. Novo, M. G. Penedo, and M. Ortega, "Deep multi-instance heatmap regression for the detection of retinal vessel crossings and bifurcations in eye fundus images," *Computer Methods and Programs in Biomedicine*, vol. 186, p. 105 201, 2020.

[110] H. Zhao, Y. Sun, and H. Li, "Retinal vascular junction detection and classification via deep neural networks," *Computer methods and programs in biomedicine*, vol. 183, p. 105 096, 2020.

[111] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[112] D. P. Bertsekas *et al.*, "Dynamic programming and optimal control 3rd edition, volume ii," *Belmont, MA: Athena Scientific*, 2011.

[113] T. v. Walsum, M. Schaap, C. T. Metz, A. G. Giessen, and W. J. Niessen, "Averaging centerlines: Mean shift on paths," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2008, pp. 900–907.

[114] E. W. Dijkstra *et al.*, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[115] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *International conference on machine learning*, PMLR, 2016, pp. 1928–1937.

[116] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, "High-dimensional continuous control using generalized advantage estimation," *arXiv preprint arXiv:1506.02438*, 2015.

[117] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[118] L. Engstrom, A. Ilyas, S. Santurkar, D. Tsipras, F. Janoos, L. Rudolph, and A. Madry, "Implementation matters in deep policy gradients: A case study on ppo and trpo," *arXiv preprint arXiv:2005.12729*, 2020.

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

[120] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[121] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[122] P. Luo, J. Ren, Z. Peng, R. Zhang, and J. Li, "Differentiable learning-to-normalize via switchable normalization," *arXiv preprint arXiv:1806.10779*, 2018.

[123] 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," *Advances in neural information processing systems*, vol. 32, 2019.

[124] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger, "Deep reinforcement learning that matters," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.

[125] L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, and I. Mordatch, "Decision transformer: Reinforcement learning via sequence modeling," *Advances in neural information processing systems*, vol. 34, pp. 15 084–15 097, 2021.

[126] S. P. Luijten, L. Wolff, M. H. Duvekot, P.-J. van Doormaal, W. Moudrous, H. Kerkhoff, G. J. L. a Nijeholt, R. P. Bokkers, S. Lonneke, J. Hofmeijer, *et al.*, "Diagnostic performance of an algorithm for automated large vessel occlusion detection on ct angiography," *Journal of neurointerventional surgery*, 2021.

[127] E. H. Weissler, T. Naumann, T. Andersson, R. Ranganath, O. Elemento, Y. Luo, D. F. Freitag, J. Benoit, M. C. Hughes, F. Khan, *et al.*, "The role of machine learning in clinical research: Transforming the future of evidence generation," *Trials*, vol. 22, no. 1, pp. 1–15, 2021.

[128] L. Wolff *et al.*, "Performance of automated imaging biomarkers in an outcome prediction model for patients with ischemic stroke," *Submitted*.