

Robust automatic brain vessel segmentation in 3D CTA scans using dynamic 4D-CTA data

Alberto Mario Ceballos Arroyo^{†,*,1}, Shrikanth M. Yadav^{†,2,3,5}, Chu-Hsuan Lin^{4,5}, Jisoo Kim^{4,5}, Geoffrey S. Young^{4,5}, Huaizu Jiang¹, and Lei Qin^{3,5}

¹ Northeastern University, Boston, USA

² Washington University in St. Louis, St. Louis, USA

³ Dana-Farber Cancer Institute, Boston, USA

⁴ Brigham and Women's Hospital, Boston, USA

⁵ Harvard Medical School, Boston, USA

Abstract. In this study, we develop a novel methodology for annotating the brain vasculature using dynamic 4D-CTA head scans. By using multiple time points from dynamic CTA acquisitions, we subtract bone and soft tissue to enhance the visualization of arteries and veins, reducing the effort required to obtain manual annotations of brain vessels. We then train deep learning models on our ground truth annotations by using the same segmentation for multiple phases from the dynamic 4D-CTA collection, effectively enlarging our dataset by 4 to 5 times and inducing robustness to contrast phases. In total, our dataset comprises 110 training images from 29 patients and 165 test images from 14 patients. In comparison with two similarly-sized datasets for CTA-based brain vessel segmentation, a nnUNet model trained on our dataset can achieve significantly better segmentations across all vascular regions, with an average mDC of 0.846 for arteries and 0.957 for veins in the TopBrain dataset. Furthermore, metrics such as average directed Hausdorff distance (adHD) and topology sensitivity (tSens) reflected similar trends: using our dataset resulted in low error margins (aDHD of 0.304 mm for arteries and 0.078 for veins) and high sensitivity (tSens of 0.877 for arteries and 0.974 for veins), indicating excellent accuracy in capturing vessel morphology. Our code and model weights are available [online](#).

Keywords: Vessel segmentation · Dynamic 4D-CTA · Deep learning

1 Introduction

Intracranial computed tomographic angiography (CTA) is an essential diagnostic tool that employs intravenous contrast agents to enhance vascular visualization, aiding in the detection of intracranial vascular abnormalities [22]. Since its introduction, CTA has advanced beyond single-phase acquisitions to

[†] co-first authors with equal contribution

* corresponding author: ceballosarroyo.a@northeastern.edu

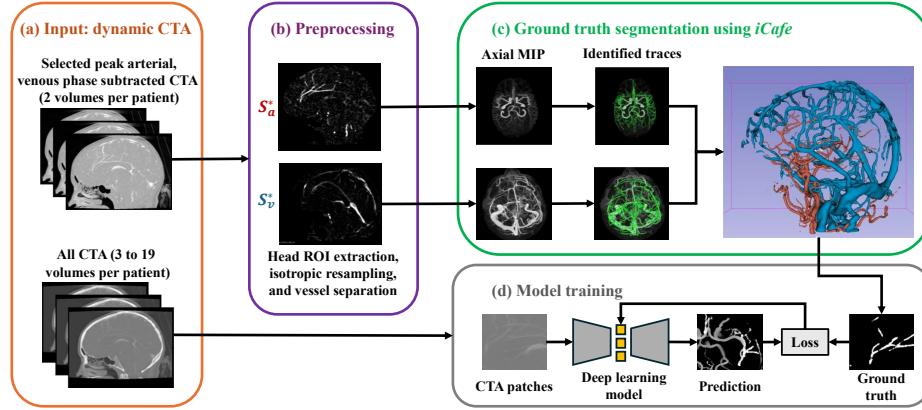


Fig. 1. Summary of our methodology. (a) Dynamic CTA are collected for all patients. (b) The subtracted images are processed to suppress voxels outside the head and resampled, then the artery-only and vein-only volumes are computed. (c) The vessel-separated images are processed using *iCafe*'s vessel tracing algorithm and then manually validated. (d) For model training, all available CTA data (*i.e.*, 3-19 volumes per patient) are paired with the GT segmentations.

include multiphase CTA [7], which captures arterial, mid-venous, and late venous phases, and dynamic 4D-CTA [16], which continuously or intermittently acquires images throughout the passage of contrast to provide detailed temporal vascular information. Despite these advances, single-phase CTA (which requires extensive post-processing to remove background bone/soft tissue and to distinguish arteries from veins) remains the predominant clinical protocol. Robust, automated methods to extract the brain vasculature in single-phase CTA could thus have substantial clinical impact, either by directly assisting radiologists during interpretation of scans or by serving as an auxiliary input to pipelines for tasks such as occlusion localization [11,14] and aneurysm detection [5,13].

Most efforts for brain vessel segmentation in CTA have centered on the automatic segmentation of the Circle of Willis (CoW), an important arterial structure. The TopCoW dataset [25] includes 125 CTA scans which were annotated using a manual, VR-based 3D workflow. The dataset has fine-grained labels for the arteries in the CoW and has drawn significant scientific interest [4,18]. Nevertheless, focusing only on the CoW ignores parts of the brain vasculature that are relevant for the management of various cerebrovascular diseases. As a result, two incremental efforts have sprung from TopCoW: a) TopBrain [1], which expands the labels in a subset of the scans from TopCoW to include both veins and arteries; and b) VesselVerse [2], which augments the labels for all of TopCoW with manual and automatic annotations of more arteries and some veins (albeit collapsed into a single vessel class). On the other hand, other researchers have attempted to address the limited availability of annotated CTA data by training

deep learning models on large, heterogeneous datasets from multiple modalities. One such example is VesselFM [24], which relies on annotations from many modalities (CTA, MRA, X-Ray, microscopy, etc.) combined with a huge number of synthetic images. However, the fraction of training CTA images is small (under 1%) and the model has not been evaluated for head CTA vessel segmentation.

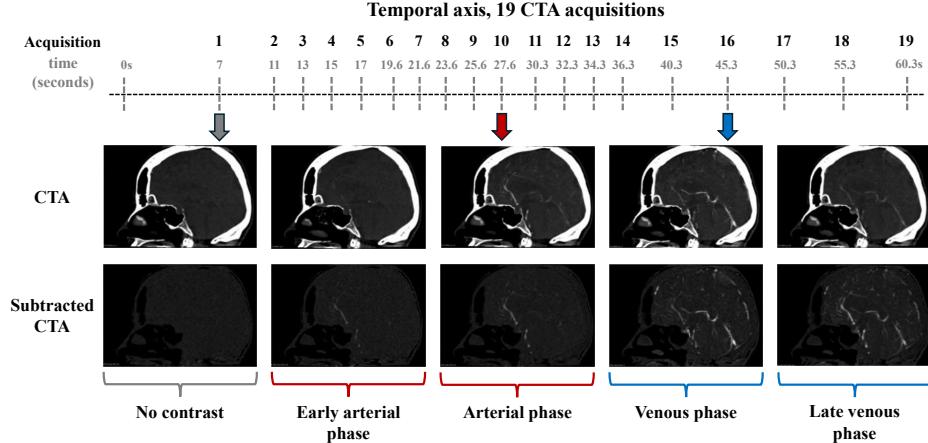


Fig. 2. Dynamic CTA from a sample patient. The top axis shows the acquisition order of the scans. The red and blue arrows point to the arterial and venous phase images, respectively. The *CTA* row depicts the original CTA images, while the *subtracted CTA* row displays the bone/soft tissue-subtracted versions.

Considering the above, there is a clear need for new annotation methodologies that minimize the time required to produce fine-grained labels for vessels in brain CTA. However, accurate annotation of CTA images requires specialized anatomical knowledge due to the presence of bone, soft tissue, and non-vascular contrast-enhanced structures. To address this challenge, we use dynamic 4D-CTA, which enables near-complete removal of bone and soft tissue through baseline subtraction. The high temporal resolution of dynamic CTA also minimizes interphase motion artifacts, resulting in cleaner subtracted images and more reliable vascular depiction. These enhancements overcome the longstanding difficulties in segmenting and distinguishing intracranial arteries and veins in angiographic CT—a process typically regarded as complex and labor-intensive. Using the proposed strategy (summarized in Fig. 1), we annotate a dataset comprising scans from 39 patients with artery and vein labels, and we use it to train an nnUNet [8] model, achieving remarkable artery and vein segmentation performance across contrast phases. In summary, our contributions are as follows:

- We document a pipeline for annotating arteries and veins in head CTA scans based on dynamic 4D-CTA data, which significantly reduces annotation time

- We present DynaVessel, an intracranial vessel segmentation dataset comprising 275 head CTA scans from 43 patients with artery and vein annotations
- We evaluate a model trained on DynaVessel against several baselines on various in-domain and out-of-domain benchmarks, achieving substantially better segmentation performance on both arteries and veins

2 Methods

2.1 Data

This retrospective study received IRB approval with a waiver of informed consent. The 4D-CTA scans obtained from our clinical research database were acquired using a 320-detector-row CT scanner (Canon-Toshiba Aquilion One, Toshiba Medical Systems, Tokyo, Japan). Imaging was conducted in axial mode using the following parameters: 80 kV, 150 mA, 0.75 s rotation time, 320 slices with 0.5 mm thickness, and a 512×512 matrix. For each patient, image acquisition began roughly 7 seconds after injecting 75–100 mL of iodinated contrast agent (Omnipaque 350, General Electric Healthcare, Chicago, USA) at a rate of 4–5 mL/s, continuing intermittently through the baseline, arterial, and venous phases. As shown in Fig. 2, images were captured at 19 time points over a 60-second interval. The peak arterial and venous phase images, which correspond to $t = 27.6\text{s}$ and $t = 45.3\text{s}$ respectively, are of primary interest. The scanner uses the baseline image—acquired before contrast reaches the brain—to subtract bone and soft tissue from subsequent images.

Of the 106 patients scanned in 2023, 38 had scans containing at least two phases of acceptable image quality. Although these were acquired using dynamic CTA protocols, complete phase datasets were not consistently archived in the PACS; in many cases, only thicker slice reconstructions, subtracted volumes, or MIP images were retained. After review, 29 patients were identified as having artifact-free arterial and venous phase images. These scans were annotated and used for model training. For evaluation, we retrieved data from 1910 patients scanned between 2016 to 2022, of whom 140 had dynamic CTA with at least two phases of acceptable resolution. Among these, 68 patients had both peak arterial and venous phase images (each comprising the original post-contrast CT image and the corresponding bone/soft tissue-subtracted image), with many also including more CTA phases. A subset of 14 patients were randomly selected to evaluate the performance of deep learning (DL) models trained using our dataset.

Pre-processing Fig. 3 illustrates the pre-processing steps for annotation. To suppress extracranial soft tissues, a binary ROI encompassing the skull and neck was delineated on a publicly available CT template [20]. The template was affine registered to the patient’s CTA image, and the resulting transformation matrix was applied to the ROI to create a binary mask of the skull and neck in the patient’s space. Voxels outside of this ROI were excluded from further analysis. All images were then resampled to achieve an isotropic resolution of

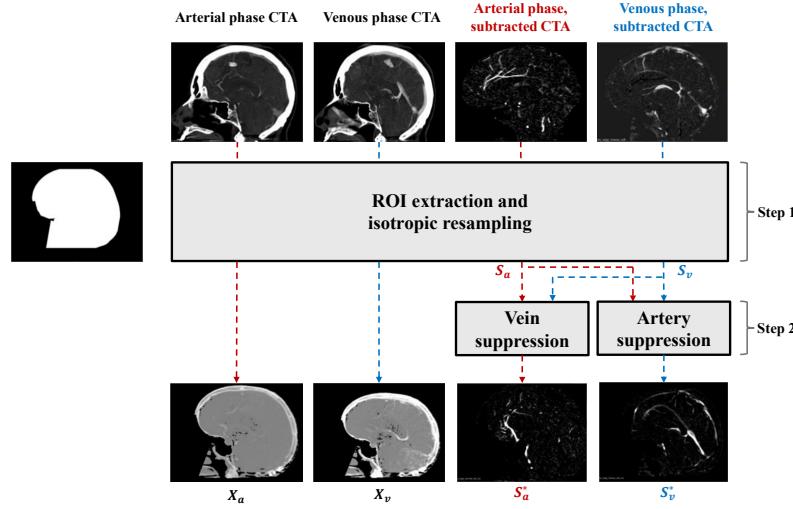


Fig. 3. Pre-processing steps for a dynamic CTA study. For both phases, a head ROI is used on the CTA and the subtracted CTA to suppress the background. Next, the subtracted images are processed using Algorithm 1 to create the vessel-separated images.

0.468 mm. Algorithm 1 was applied to the subtracted CTA images to separate arterial and venous structures into distinct volumes for each patient. The inputs to the algorithm include the bone/soft tissue-subtracted arterial (S_a) and venous (S_v) phase images (generated by the scanner), along with the corresponding CTA images X_a and X_v . We use affine transformations (steps 1 and 2) to align the images and minimize motion artifacts between acquisitions. Steps 5 and 6 enhance the contrast-to-background ratio for arteries and veins, respectively. The output of Algorithm 1 for each patient is two volumes: S_a^* (bone- and vein-suppressed arterial phase) and S_v^* (bone- and artery-suppressed venous phase).

Table 1. Train-test splits for the datasets used in this paper. Coverage for contrast phases (Contr.), proximal arteries (Pr.), first (1O) and second (2O) order arteries, distal arteries (Di.), arterial branches (Br.) and veins (V.) is indicated as ●: complete or near-complete, ▷: partial, or ○: absent.

Dataset	Num. train	Num. test	Labels	Coverage						
				Contr.	Pr.	1O	2O	Di.	Br.	V.
VesselFM [24]	510,902	—	Vessel-only	—	—	—	—	—	—	—
TopCoW [25]	112	13	13 arteries	○	●	●	○	○	○	○
ISLES [17,25]	—	26	13 arteries	○	●	●	○	○	○	○
VesselVerse [2]	112	13	Vessel-only	○	●	●	●	○	○	▷
TopBrain [1,25]	—	25	6 veins / 34 arteries	○	●	●	●	●	●	●
DynaVessel <i>train</i>	110	—	Vein / artery	●	●	●	●	●	●	●
DynaVessel ₁	—	14	Vein / 22 arteries	○	●	●	●	▷	▷	●
DynaVessel ₊	—	165	Vein / 22 arteries	●	●	●	●	▷	▷	●

Algorithm 1 Artery and vein suppression.**Input:** S_a, S_v, X_a, X_v **Output:** S_a^*, S_v^* 1: Compute rigid transformation G_{rv} such that X_a is moving and X_v is fixed2: Compute rigid transformation G_{ra} such that X_v is moving and X_a is fixed3: $S_{v \rightarrow a} \leftarrow G_{ra}(X_v)$, transform X_v so that it is in the same space as X_a 4: $S_{a \rightarrow v} \leftarrow G_{rv}(X_a)$, transform X_a so that it is in the same space as X_v 5: $s_a^* \leftarrow \begin{cases} 0 & \text{if } s_a < s_{v \rightarrow a}, \\ s_a & \text{if } s_a > s_{v \rightarrow a} \end{cases}$, where s_a^* , $s_{v \rightarrow a}$, and s_a are elements of matrices S_a^* , $S_{v \rightarrow a}$, and S_a 6: $s_v^* \leftarrow \begin{cases} 0 & \text{if } s_v < s_{a \rightarrow v}, \\ s_v & \text{if } s_v > s_{a \rightarrow v} \end{cases}$, where s_v^* , $s_{a \rightarrow v}$, and s_v are elements of matrices S_v^* , $S_{a \rightarrow v}$, and S_v

Ground truth (GT) segmentation Based on the vessel-separated volumes generated from the pre-processing steps in Figure 3, we label the arteries and veins using the *iCafe* [6] feature extraction pipeline. This process uses intensity normalization followed by segmentation using the Phansalkar threshold [15] (used for vessel tracking) and the Renyi entropy threshold [10] (used for visualization). A vessel tracing algorithm, based on an open-curve active contour model, is used to identify vessel-like structures [23]. Spurious vessel traces outside the skull and within the soft tissue of the head are then manually deleted. Finally, the vessel trace and its corresponding mask were exported as a 3D segmentation of the vascular structures. Fig. 1(c) shows the overall process of extracting the GT, which results in a two-class (artery, vein) segmentation dataset.

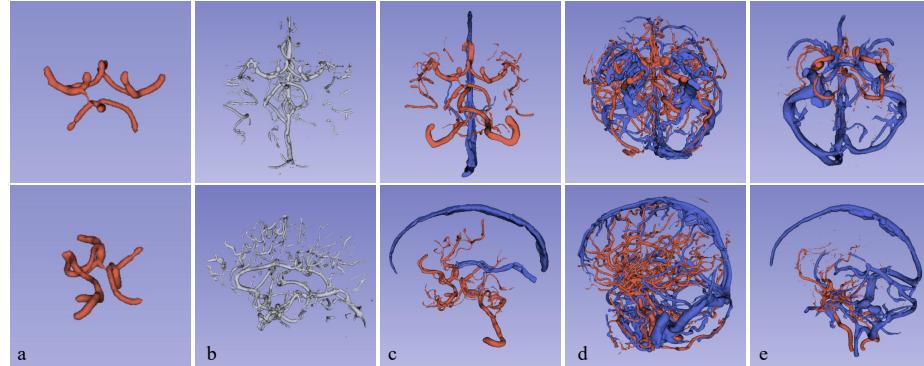


Fig. 4. Comparison of vascular structure coverage across datasets, with arteries in red and veins in blue: (a) TopCow, (b) VesselVerse, (c) TopBrain, (d) DynaVessel train, (e) DynaVessel test. Visualizations created on 3D Slicer.

For the evaluation dataset, 20 arteries were annotated using *iCafe*: ICA (Left), ICA (Right), VA (Left), VA (Right), BA, PComm (Left), PComm (Right), P1 (Left), P1 (Right), P2+ (Left), P2+ (Right), M1 (Left), M1 (Right), M2+ (Left), M2+ (Right), A1 (Left), A1 (Right), A2+ (Left), A2+ (Right), and Acomm.

A subspecialty-trained neuroradiologist validated the annotations by overlaying the segmentations with the subtracted images S_a in **3D Slicer** [12] and corrected inaccuracies related to anatomical variations in the arteries. *iCafe* lacks a dedicated venous annotation pipeline, so the GT vessel segmentations of the veins for the evaluation dataset were manually refined by using **3D Slicer** to remove distal branches from the venous segmentation. Fig. 4 depicts segmentations from our train (d) and test (e) data, compared with three other datasets (a, b, c).

2.2 Deep learning model training

Taking advantage of dynamic CTA images, our training and evaluation data includes multiple CTA phases from the same patient, each paired with a corresponding GT segmentation. All images were rigidly registered to the arterial phase CTA using Advanced Normalization Tools [21]. We split the training and test datasets at the patient level, to prevent any patient from appearing in both validation and training sets. To train the nnUNet model, we use the nnUNet v2 library [8,9]. Aside from specifying the model type to be a Residual Encoder UNet (large preset), all models were trained for 1000 epochs with a linearly decaying learning rate starting at 0.01, using the nnUNet planner’s automatically selected hyper-parameters. All model training and evaluations were performed on a workstation with an NVIDIA RTX 4090 GPU (24 GB VRAM), 8 CPU cores, and 64 GB of RAM.

2.3 Evaluation

Arterial and venous structures are evaluated separately to highlight potential differences in prediction accuracy across vessel types. Importantly, the DL models trained on our dataset output a volume with two classes (artery and vein), but the evaluation dataset contains 21 classes. Although we cannot provide a comprehensive semantic segmentation comparison, we analyze how the predictions align with different vascular structures to provide valuable insight into region-specific performance. Specifically, let volume A_i represent the ground truth vessel mask of interest, where $i = 1, 2 \dots 20$ corresponds to arterial labels and $i = 21$ corresponds to venous labels; let volume P denote the predicted mask for the arterial or venous region. In line with established guidelines [25], we use the following counting, distance, and centerline-based metrics:

1. **Modified Dice coefficient:** we propose a modified version of the Dice coefficient (mDC) focusing on sensitivity, defined as:

$$mDC(A_i, P) = \frac{|A_i \cap P|}{|A_i|}$$

2. **Average directed Hausdorff surface distance:** this variation of the Hausdorff distance was chosen based on previous work [3]. *adHD* is similar to the balanced Hausdorff distance, but we use only the first term in the

Table 2. Binary vessel segmentation results on three test sets.

Test set	Model	Train set	mDC \uparrow	tSens \uparrow	adHD \downarrow
CoW	DynUNet	VesselFM	0.721 \pm 0.17	0.849 \pm 0.14	0.566 \pm 0.58
	nnUNet v2	CoW	0.789 \pm 0.03	0.930 \pm 0.04	0.274 \pm 0.13
	nnUNet v2	VesselVerse	0.938 \pm 0.04	0.989 \pm 0.01	0.052 \pm 0.03
	nnUNet v2	DynaVessel	0.972 \pm 0.02	0.992 \pm 0.01	0.026 \pm 0.01
ISLES	DynUNet	VesselFM	0.744 \pm 0.18	0.880 \pm 0.16	0.572 \pm 1.11
	nnUNet v2	CoW	0.723 \pm 0.07	0.912 \pm 0.05	0.458 \pm 0.28
	nnUNet v2	VesselVerse	0.927 \pm 0.03	0.987 \pm 0.02	0.075 \pm 0.04
	nnUNet v2	DynaVessel	0.960 \pm 0.03	0.989 \pm 0.02	0.041 \pm 0.03
VesselVerse	DynUNet	VesselFM	0.320 \pm 0.09	0.322 \pm 0.08	3.266 \pm 0.94
	nnUNet v2	CoW	0.123 \pm 0.04	0.079 \pm 0.02	23.36 \pm 2.85
	nnUNet v2	VesselVerse	0.506 \pm 0.10	0.540 \pm 0.10	2.757 \pm 1.62
	nnUNet v2	DynaVessel	0.650 \pm 0.08	0.682 \pm 0.09	0.704 \pm 0.35

equation. This is because, in our application, the boundary of the predicted volume can be significantly larger than the GT. Thus, *adHD* is defined as:

$$adHD(A_{s,i}, P_s) = \frac{1}{|A_{s,i}|} \sum_{a \in A_{s,i}} \min_{p \in P_s} d(a, p)$$

where $A_{s,i}$ and P_s denote the surfaces of volumes A_i and P , respectively. $d(a, p)$ represents the Euclidean distance between surface points.

3. **Topology sensitivity:** finally, we compute topology sensitivity (*tSens*) to quantify the extent of centerlines that are captured by the vessel segmentation models. This metric is defined based on *clDice* [19]:

$$tSens(A_{c,i}, P) = \frac{|A_{c,i} \cap P|}{|A_{c,i}|}$$

where $A_{c,i}$ is the centerline volume of A_i .

3 Results

To assess the usefulness of DynaVessel, we compare it to two publicly available CTA vessel segmentation datasets of similar size: TopCoW [25], which contains 125 scans with semantic annotations of the arteries in the Circle of Willis; and VesselVerse [2], which expands the annotations in TopCoW to include veins and more arteries, although it uses a single vessel label. We thus train three nnUNet models as specified in Section 2.2 (from now on, we refer to the model trained on DynaVessel as M_{DV}). Likewise, we compare against the pre-trained VesselFM [24] DynUNet, a foundation vessel segmentation model optimized on a collection of 928 images (including TopCoW) from vessel segmentation datasets encompassing various modalities, 10,000 images sampled from a generative flow-matching model, and 500,000 randomly generated 3D images imitating blood vessels based on graph representations of corrosion casts. Similar to VesselVerse, the labels in VesselFM are binary.

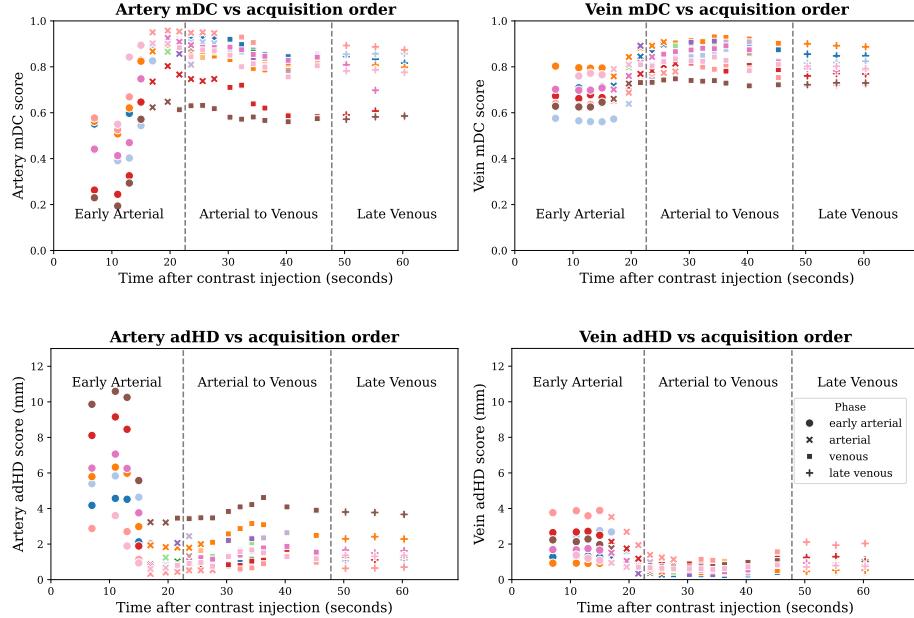


Fig. 5. mDC and adHD scores across phases. Each point represents a single scan from a patient’s Dynamic 4D-CTA, with color encoding the patient. Since the transition between arterial and venous phases is not discrete, we classify an scan’s phase as arterial vs venous by taking the average HU intensity on the centerline of the GT artery and vein masks; if the HU is greater for arteries, the scan goes in the arterial phase category, otherwise, it is put in the venous phase category.

We evaluate all four models on several test datasets, including held-out subsets of DynaVessel, TopCoW [25], and VesselVerse [2]. In addition, we use a subset of ISLES 2024 [17] annotated by the organizers of the TopCoW challenge [25], and TopBrain, a subset of TopCoW with detailed annotations of 34 arteries and 6 venous structures. Since DynaVessel includes images acquired at multiple contrasts for each patient, we define two evaluation variants: DynaVessel₁, which contains a single, arterial-phase image for each patient (14 images total), and DynaVessel₊, which contains all available contrast phases per patient (165 images in total), allowing us to measure model robustness to contrast variability. Table 1 offers more details about the datasets and the train/eval splits we used. Model performance is assessed using the metrics defined in Section 2.3 in three main ways, namely: a) the binary vessel segmentation setting, where classes in the GTs and the predictions are merged into 0 = background, 1 = vessel; b) the artery vs vein segmentation setting, where classes in the GTs are merged into 0 = background, 1 = artery, 2 = vein; and c) the semantic segmentation setting, where we evaluate performance on all available arterial anatomical locations.

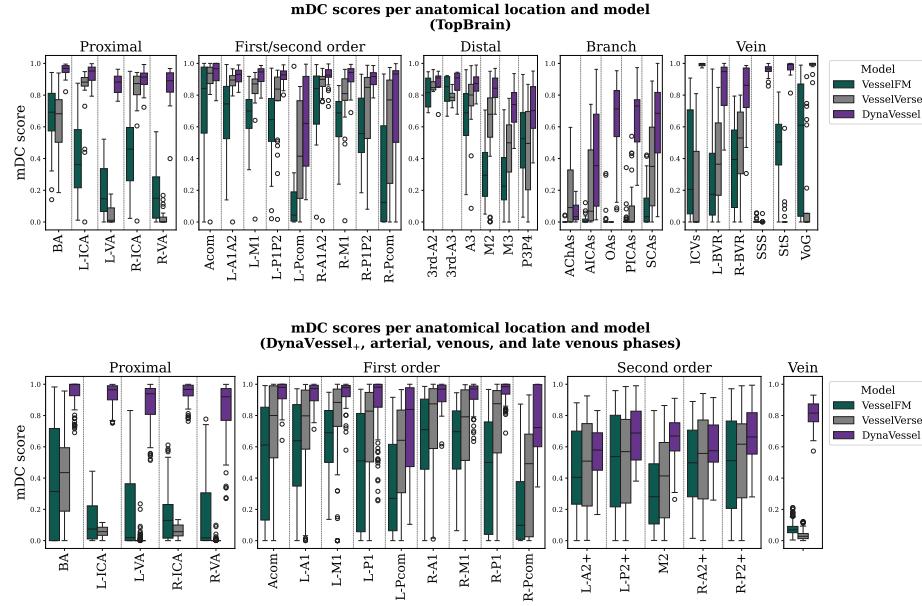


Fig. 6. mDC scores for semantic segmentation. For TopBrain, we merge annotations of right/left distal arteries and branches to improve readability. We omit the pre-contrast/early arterial phase images when calculating mDC on DynaVessel.

Importantly, for models that output a generic vessel segmentation, we use the entire vessel mask to separately measure artery and vein segmentation performance. Since our metrics prioritize sensitivity, this approach is biased toward models that output a single label, whereas models that output an artery vs vein mask, such as M_{DV} , will be penalized when they mistake arteries for veins and vice-versa. We follow a similar approach for semantic segmentation evaluation: we use the best available segmentation (artery, vein, or generic vessel) as the prediction to compute metrics when segmenting specific anatomical regions.

In Table 2, we present the results for binary vessel segmentation. Despite being trained on a similarly-sized dataset (or much smaller, when compared to VesselFM), M_{DV} achieves higher performance on the three datasets. Remarkably, training on DynaVessel leads to better results even on data that is *in-domain* for the other models⁶. When the results are disaggregated to arteries and veins separately (see Table 3), we see that M_{DV} achieves higher performance on arteries and much higher performance on veins on TopBrain and DynaVessel, underlining the positive impact of our annotation protocol across the board. The metrics for M_{DV} on artery segmentation are worse on DynaVessel+ compared to the single-phase test set, which is expected as the former includes scans taken

⁶ The TopCoW, TopBrain, and VesselVerse test sets are in-domain for VesselFM and the nnUNet models trained on TopCoW/VesselVerse, and out-of-domain for M_{DV} .

Table 3. Artery and vein segmentation results. For DynaVessel₁ we use a single, arterial phase scan. For DynaVessel₊ we evaluate all scans across phases.

Dataset	Region	Model	Train Set	mDC \uparrow	tSens \uparrow	adHD \downarrow
TopBrain	Artery	DynUNet	VesselFM	0.374 \pm 0.12	0.442 \pm 0.12	2.470 \pm 0.94
		nnUNet v2	CoW	0.192 \pm 0.04	0.158 \pm 0.03	21.15 \pm 1.92
		nnUNet v2	VesselVerse	0.455 \pm 0.12	0.591 \pm 0.17	8.070 \pm 4.29
		nnUNet v2	DynaVessel	0.846 \pm 0.05	0.877 \pm 0.06	0.304 \pm 0.15
	Vein	DynUNet	VesselFM	0.083 \pm 0.07	0.210 \pm 0.17	7.675 \pm 2.49
		nnUNet v2	CoW	0.000 \pm 0.00	0.000 \pm 0.00	50.27 \pm 8.95
		nnUNet v2	VesselVerse	0.031 \pm 0.04	0.132 \pm 0.15	17.30 \pm 7.44
		nnUNet v2	DynaVessel	0.957 \pm 0.03	0.974 \pm 0.02	0.078 \pm 0.06
DynaVessel ₁	Artery	DynUNet	VesselFM	0.414 \pm 0.09	0.494 \pm 0.08	1.598 \pm 0.44
		nnUNet v2	CoW	0.119 \pm 0.03	0.078 \pm 0.03	26.40 \pm 2.42
		nnUNet v2	VesselVerse	0.380 \pm 0.05	0.519 \pm 0.07	8.436 \pm 2.07
		nnUNet v2	DynaVessel	0.862 \pm 0.08	0.615 \pm 0.10	1.353 \pm 0.83
	Vein	DynUNet	VesselFM	0.036 \pm 0.02	0.082 \pm 0.05	5.465 \pm 0.75
		nnUNet v2	CoW	0.000 \pm 0.00	0.000 \pm 0.00	45.00 \pm 3.31
		nnUNet v2	VesselVerse	0.022 \pm 0.03	0.057 \pm 0.06	19.29 \pm 3.92
		nnUNet v2	DynaVessel	0.786 \pm 0.07	0.673 \pm 0.10	0.984 \pm 0.63
DynaVessel ₊	Artery	DynUNet	VesselFM	0.226 \pm 0.19	0.284 \pm 0.22	3.751 \pm 2.57
		nnUNet v2	CoW	0.072 \pm 0.05	0.054 \pm 0.04	40.42 \pm 40.38
		nnUNet v2	VesselVerse	0.207 \pm 0.15	0.276 \pm 0.21	17.71 \pm 19.23
		nnUNet v2	DynaVessel	0.757 \pm 0.17	0.510 \pm 0.16	2.281 \pm 2.02
	Vein	DynUNet	VesselFM	0.064 \pm 0.05	0.129 \pm 0.09	5.137 \pm 1.25
		nnUNet v2	CoW	0.001 \pm 0.00	0.003 \pm 0.01	56.49 \pm 32.56
		nnUNet v2	VesselVerse	0.028 \pm 0.03	0.071 \pm 0.06	29.12 \pm 19.98
		nnUNet v2	DynaVessel	0.788 \pm 0.09	0.683 \pm 0.12	1.059 \pm 0.78

before contrast and at the early arterial, venous, and late venous phases. Fig. 5 also reveals that the average scores are brought down by low performance on pre-contrast and early arterial phases, which is expected since at that point there is likely not enough contrast in the arteries.

Fig. 6 shows a significant gap between the performance of M_{DV} versus the two strongest baselines in the semantic segmentation evaluation, with great coverage of most anatomical regions annotated in TopBrain except for some distal arteries (the right and left anterior choroidal arteries). A similar trend is seen in the DynaVessel₊ test set, where the model generally performs worse on second-order arteries compared with the first-order ones. Meanwhile, in Fig. 7, we provide a qualitative visualization of the outputs produced by all three models. M_{DV} achieves almost complete coverage of arterial and venous structures, with very few false negatives. The difference in performance is particularly noticeable on veins, where VesselFM and the VesselVerse nnUNet fail to detect most veins. In VesselVerse’s case, this is due to the fact venous sinuses were not annotated to maintain consistency with the MRA annotations. Although a large number of false positive detections are visible in Fig. 7(b, c, e, f), the majority are correctly segmented arterial and venous structures that are not annotated in TopBrain. On the other hand, Fig. 7(a, d) shows that VesselFM generates a much larger number of spurious segmentations and unconnected elements, likely due to the limited amount of CTA it was trained on.

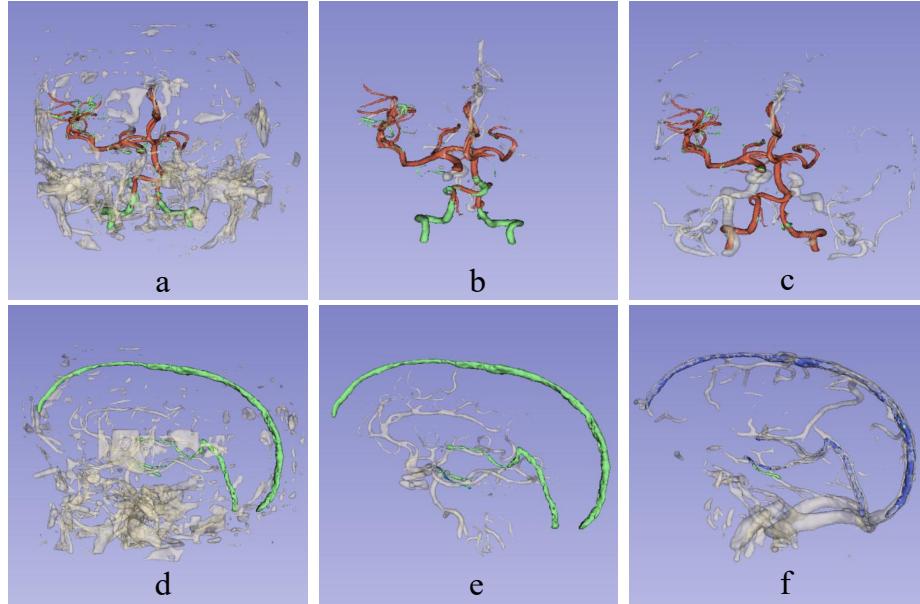


Fig. 7. Qualitative comparisons of the segmentations from several models on a Top-Brain sample. In the top row, we display results for artery segmentation using (a) VesselFM, (b) the VesselVerse nnUNet, and (c) our M_{DV} ; true positive (TP) areas are red, false positives (FPs) are gray, and false negatives (FNs) are green. The bottom row includes the results for vein segmentation using (d) VesselFM, (e) the VesselVerse nnUNet, and (f) our M_{DV} ; TP areas are blue, FPs are gray, and FNs are green. VesselFM and the VesselVerse nnUNet only output a single vessel class, so arterial predictions show up in the vein segmentation visualization as FPs, and vice-versa. We note that most FP areas segmented by M_{DV} (c, f) are actually TP arterial and venous structures that are not annotated in TopBrain. Visualizations created on 3D Slicer.

3.1 Ablation study

In Fig. 8, we measure the impact of removing the various contrast phases used for training M_{DV} by optimizing two new models: M_{art} , which is trained on only 29 peak arterial phase images, and M_{ven} , which is trained on only 29 peak venous phase images. It is important to note that M_{art} and M_{ven} have access to the same number of patients and GT annotations as M_{DV} . We make two observations: 1) that training on a single, peak phase per patient significantly reduces the models' capabilities to generalize across unseen phases; and 2) that the model trained on all available phases often surpasses the models trained on a single phase when evaluated on such phases. This once again confirms the usefulness of our methodology, especially in scenarios where the availability of peak arterial/venous phase images is not guaranteed.

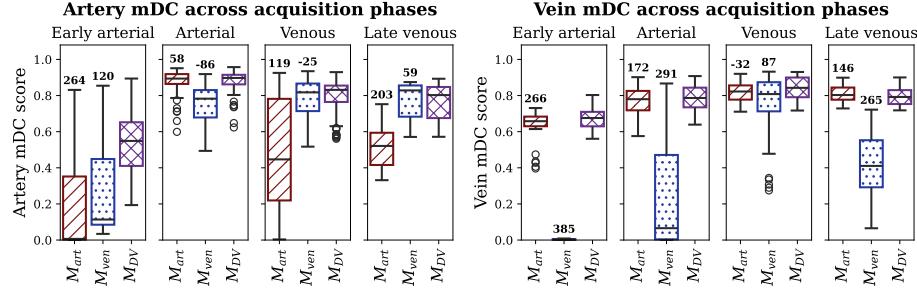


Fig. 8. Artery and vein mDC scores across contrast phases for our model (M_{DV} , violet, rightmost box in each subplot), compared with models trained either on the peak arterial phase images (M_{art} , red, leftmost on each box) or the peak venous phase images (M_{ven} , blue, middle box in each subplot). The number above the top whisker for each boxplot codifies the difference between the average Hounsfield unit (HU) intensity in arteries/veins during training vs during the phase on which the model is evaluated. The performance for single-phase training correlates negatively with the difference between mean training HU and mean test HU. The behavior in the late venous phase is explained by the circulation of part of the contrast agent back into arteries.

4 Discussion

We proposed an automatic data pre-processing method that utilizes dynamic CTA images acquired at multiple time points. This approach not only reduces annotation time but also enhances the diversity of the training data at no additional cost. Using this protocol, we created DynaVessel, a CTA intracranial vessel segmentation dataset with artery and vein labels comprising 110 images from 29 patients, covering a wide range of acquisition contrasts. Furthermore, we annotated a large evaluation set of 165 images from 14 patients with fine-grained semantic labels for veins and 20 arteries, and we evaluate the impact of training M_{DV} (a state-of-the-art nnUNet) on our data using modified metrics to emphasize sensitivity across different vessel segments. Overall, our evaluation pipeline provides detailed insight on the model’s ability to segment arteries of varying sizes and across multiple phases.

The M_{DV} performance trends across dynamic CTA phases highlight the dependence of deep learning-based vessel segmentation on image contrast. In our experiments training on a single phase, segmentation performance degrades severely during mismatched phases—such as arterial segmentation in late venous images or venous segmentation in early arterial images—where reduced contrast conspicuity introduces ambiguity. Notably, the difference in HU distributions between the training data and phase-specific test data correlates strongly with performance; for instance, M_{ven} is trained on images with a mean vein HU of 485, resulting in a delta of 87 HU with respect to the mean vein HU of the venous phase ($mDC = 0.81$) and a delta of 385 HU with respect to the mean HU

of the early arterial phase ($mDC = 0.01$). Together, these findings suggest that models trained on single-phase data leverage contrast-driven vessel appearance rather than anatomical priors. On the other hand, models trained on multi-phase data suffer far less performance degradation, although they are still sensitive to changes in contrast. Indeed, the artery segmentation performance of M_{DV} peaks during the arterial phase, when arterial enhancement is maximal and venous structures remain relatively unopacified, whereas vein segmentation achieves its highest accuracy during the venous phase, when venous enhancement dominates. Moreover, M_{DV} still requires the presence of some contrast agent to correctly segment brain vessels, as shown by its relatively poor performance during the early arterial phase.

Regardless, when evaluated across all vascular regions, M_{DV} achieved superior performance compared to models trained on other datasets of similar or much larger sizes. This performance gap was especially pronounced in veins, where M_{DV} achieved a modified DC of 0.75 to 0.95 depending on the dataset, compared to under 0.01 mDC for the other models. The topology Sensitivity and average directed Hausdorff Distance tell a similar story, with M_{DV} outperforming the other models by a significant margin. Although the performance gap was smaller for arteries, M_{DV} still came on top, with modified DC scores of 0.75 to 0.86 compared to scores ranging from 0.07 to 0.45 for the other models. Overall, our work constitutes a significant step toward the automatic segmentation of all brain vascular structures in CTA scans, with broad applicability as a primary or auxiliary input for the assessment of cerebrovascular pathologies. Nevertheless, despite the fact DynaVessel imbues M_{DV} with robust vessel segmentation performance, we note that our fine-grained semantic segmentation evaluations only serve to provide an upper-bound approximation of the performance of models trained using DynaVessel. Future work should address the semantic segmentation scenario by expanding the annotations in the training data to include all 20 artery (or more) sub-classes, establishing guidelines to annotate venous structures in a fine-grained manner, and designing deep learning architectures suited for such a complex task.

Acknowledgments. The authors acknowledge the financial support provided by NIH grant 1R01LM013891-01A1. Ceballos-Arroyo, A. is grateful for the funding provided by Colombia's Minciencias and Fulbright under the Fulbright Minciencias 2021 program.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. TopBrain Segmentation Challenge for Whole Brain Vessel Anatomy - Grand Challenge. <https://topbrain2025.grand-challenge.org/data/>
2. VesselVerse: A Dataset and Collaborative Framework for Vessel Annotation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. pp. 655–665. Springer Nature (2026). https://doi.org/10.1007/978-3-032-05169-1_63
3. Aydin, O.U., Taha, A.A., Hilbert, A., Khalil, A.A., Galinovic, I., Fiebach, J.B., Frey, D., Madai, V.I.: On the usage of average Hausdorff distance for segmentation performance assessment: hidden error when used for ranking. European Radiology Experimental **5**(1), 4 (Jan 2021), <https://doi.org/10.1186/s41747-020-00200-2>
4. Berger, A.H., Lux, L., Stucki, N., Bürgin, V., Shit, S., Banaszak, A., Rueckert, D., Bauer, U., Paetzold, J.C.: Topologically faithful multi-class segmentation in medical images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2024. pp. 721–731. Springer Nature (2024)
5. Ceballos-Arroyo, A.M., Nguyen, H.T., Zhu, F., Yadav, S.M., Kim, J., Qin, L., Young, G., Jiang, H.: Vessel-aware aneurysm detection using multi-scale deformable 3D attention. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2024 **15005**, 754–765 (Oct 2024), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11986933/>
6. Chen, L., Shaw, D.W.W., Dager, S.R., Corrigan, N.M., Chu, B., Kleinhans, N.M., Kuhl, P.K., Hwang, J.N., Yuan, C.: Quantitative assessment of the intracranial vasculature of infants and adults using iCafe (intracranial artery feature extraction). Frontiers in Neurology **Volume 12 - 2021** (2021), <https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2021.668298>
7. Dundamadappa, S., Iyer, K., Agrawal, A., Choi, D.: Multiphase CT Angiography: A Useful Technique in Acute Stroke Imaging—Collaterals and Beyond. AJNR: American Journal of Neuroradiology **42**(2), 221–227 (Jan 2021), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7872177/>
8. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021), <https://www.nature.com/articles/s41592-020-01008-z>
9. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jäger, P.F.: nnU-Net revisited: A call for rigorous validation in 3d medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. p. 488–498. Springer-Verlag, Berlin, Heidelberg (2024), https://doi.org/10.1007/978-3-031-72114-4_47
10. Kapur, J.N., Sahoo, P.K., Wong, A.K.C.: A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision, Graphics, and Image Processing **29**(3), 273–285 (Mar 1985), <https://www.sciencedirect.com/science/article/pii/0734189X85901252>
11. Kassam, J., Thamm, F., Rist, L., Taubmann, O., Maier, A.: Detecting Large Vessel Occlusions using Graph Deep Learning. In: Proceedings of the First International Workshop on Geometric Deep Learning in Medical Image Analysis. pp. 149–159. PMLR (Nov 2022)
12. Kikinis, R., Pieper, S.D., Vosburgh, K.G.: 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support, pp. 277–

289. Springer New York, New York, NY (2014), https://doi.org/10.1007/978-1-4614-7657-3_19
- 13. Kim, J., Ceballos-Arroyo, A., Lin, C.H., Liu, P., Jiang, H., Yadav, S., Wan, Q., Qin, L., Young, G.S.: Automated, anatomy-based, heuristic post-processing reduces false positives and improves interpretability of deep learning intracranial aneurysm detection models **16**(1), 3218, <https://www.nature.com/articles/s41598-025-33083-7>
 - 14. Lal-Trehan Estrada, U.M., Oliver, A., Sheth, S.A., Lladó, X., Giancardo, L.: Strategies to combine 3D vasculature and brain CTA with deep neural networks: Application to LVO. *iScience* **27**(2), 108881 (Jan 2024). <https://doi.org/10.1016/j.isci.2024.108881>
 - 15. Phansalkar, N., More, S., Sabale, A., Joshi, M.: Adaptive local thresholding for detection of nuclei in diversity stained cytology images. In: 2011 International Conference on Communications and Signal Processing. pp. 218–220 (Feb 2011), <https://ieeexplore.ieee.org/document/5739305>
 - 16. Rajiah, P.S., Weber, N., Loewen, J., Kasten, H., Williamson, E., Moore, A., Leng, S.: Dynamic CT Angiography in Vascular Imaging: Principles and Applications. *RadioGraphics* **42**(7), E224–E225 (Nov 2022), <https://pubs.rsna.org/doi/full/10.1148/rg.210177>, publisher: Radiological Society of North America
 - 17. Riedel, E.O., de la Rosa, E., Baran, T.A., *et al.*: ISLES'24 – a real-world longitudinal multimodal stroke dataset (2025), <https://arxiv.org/abs/2408.11142>
 - 18. Shi, P., Hu, J., Yang, Y., Gao, Z., Liu, W., Ma, T.: Centerline boundary dice loss for vascular segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. p. 46–56. Springer-Verlag, Berlin, Heidelberg (2024)
 - 19. Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P.W., Bauer, U., Menze, B.H.: cIDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16555–16564. IEEE, Nashville, TN, USA (Jun 2021), <https://ieeexplore.ieee.org/document/9578225/>
 - 20. Talou, G.D.M., Safaei, S., Hunter, P.J., Blanco, P.J.: Adaptive constrained constructive optimisation for complex vascularisation processes. *Scientific Reports* **11**(1), 6180 (Mar 2021), <https://www.nature.com/articles/s41598-021-85434-9>, publisher: Nature Publishing Group
 - 21. Tustison, N.J., Cook, P.A., Holbrook, A.J., *et al.*: The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports* **11**(1), 9068 (Apr 2021), <https://doi.org/10.1038/s41598-021-87564-6>
 - 22. Walluscheck, S., Canalini, L., Strohm, H., Diekmann, S., Klein, J., Heldmann, S.: MR-CT multi-atlas registration guided by fully automated brain structure segmentation with CNNs. *International Journal of Computer Assisted Radiology and Surgery* **18**(3), 483–491 (Mar 2023), <https://doi.org/10.1007/s11548-022-02786-x>
 - 23. Wang, Y., Narayanaswamy, A., Tsai, C.L., Roysam, B.: A broadly applicable 3-D neuron tracing method based on open-curve snake. *Neuroinformatics* **9**(2-3), 193–217 (Sep 2011). <https://doi.org/10.1007/s12021-011-9110-5>
 - 24. Wittmann, B., Wattberg, Y., Amiranashvili, T., Shit, S., Menze, B.: vesselFM: A foundation model for universal 3d blood vessel segmentation. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 20874–20884 (2025). <https://doi.org/10.1109/CVPR52734.2025.01944>
 - 25. Yang, K., Musio, F., Ma, Y., *et al.*: Benchmarking the CoW with the TopCoW Challenge: Topology-Aware Anatomical Segmentation of the Circle of Willis for CTA and MRA. ArXiv p. arXiv:2312.17670v4 (Jul 2025)

Supplementary material

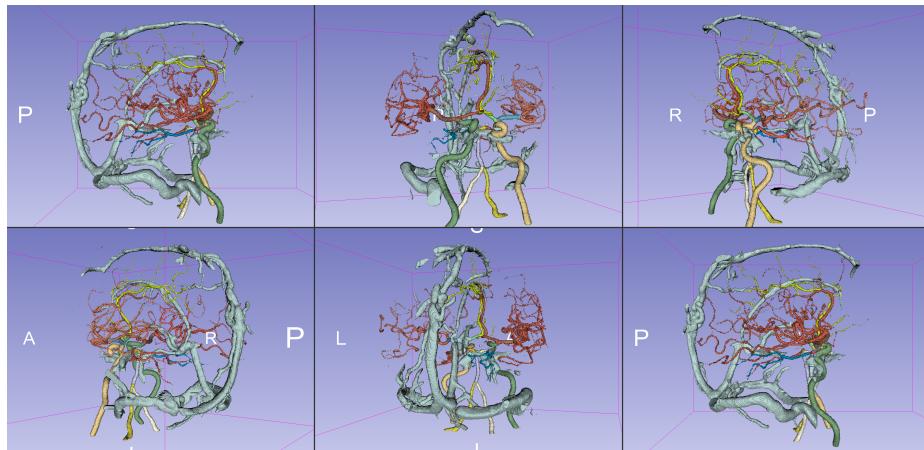


Fig. S1. Visualization of the semantic labels for the DynaVessel test data. Created using 3D Slicer.

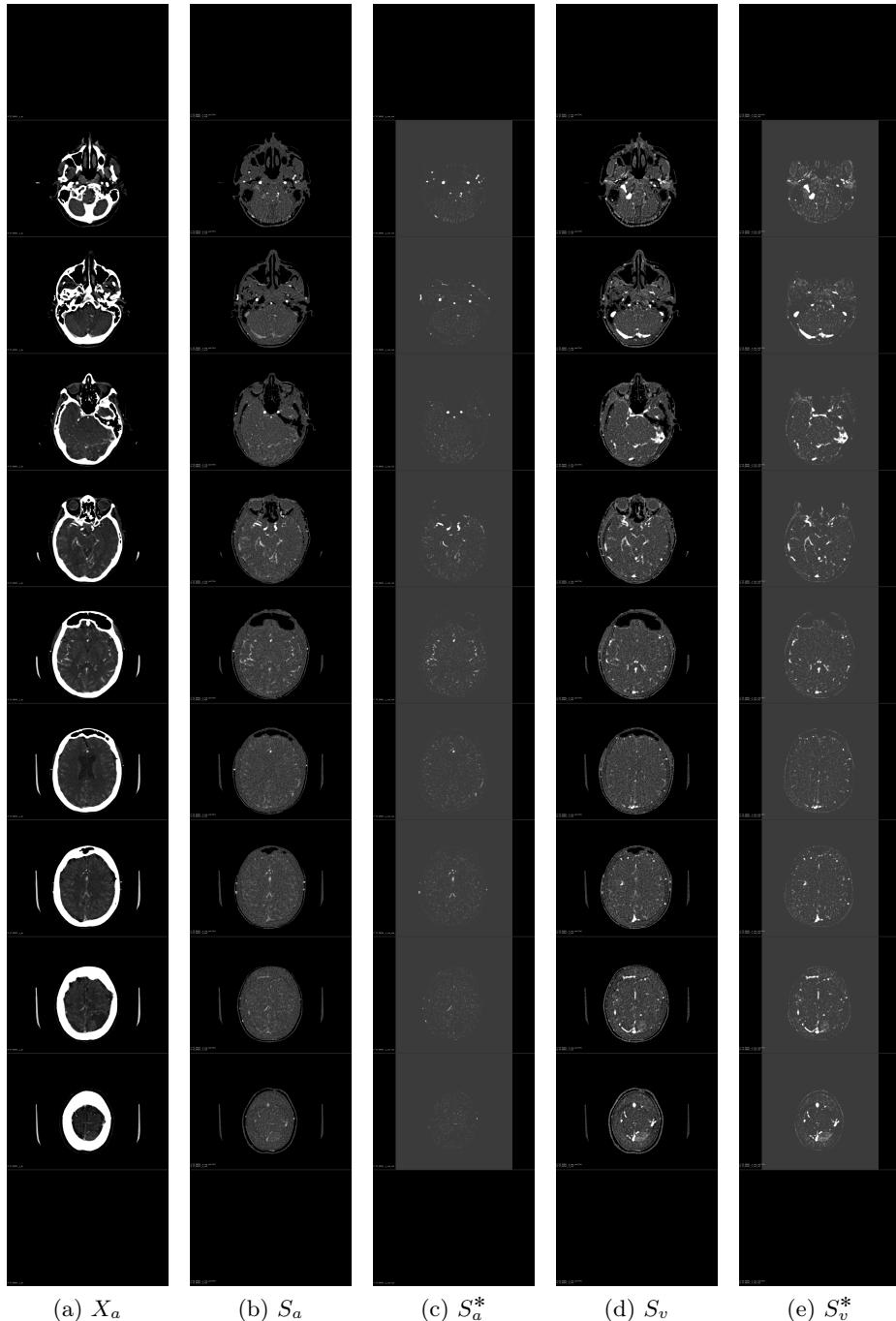


Fig. S2. Axial images of arterial phase CTA (X_a), bone subtracted arterial and venous phase (S_a and S_v), and the vessel separated arterial and venous phase (S_a^* and S_v^*) volumes from a sample patient.