

Scaling up fine-grained intracranial vessel annotations in computed tomography angiography

Chu-Hsuan Lin*

CLIN71@BWH.HARVARD.EDU

*Brigham and Women’s Hospital, Harvard Medical School
Boston, MA, USA*

Alberto Mario Ceballos-Arroyo*

CEBALLOSARROYO.A@NORTHEASTERN.EDU

*Khoury College of Computer Sciences, Northeastern University
Boston, MA, USA*

Jisoo Kim

JKIM@BWH.HARVARD.EDU

*Brigham and Women’s Hospital, Harvard Medical School
Boston, MA, USA*

Shrikanth M. Yadav

SHRIKANTH@WUSTL.EDU

*Washington University in St. Louis
St. Louis, MI, USA*

Huaizu Jiang

H.JIANG@NORTHEASTERN.EDU

*Khoury College of Computer Sciences, Northeastern University
Boston, MA, USA*

Lei Qin

LEI_QIN@DFCI.HARVARD.EDU

*Dana-Farber Cancer Institute, Harvard Medical School
Boston, MA, USA*

Geoffrey S. Young

GSYOUNG@BWH.HARVARD.EDU

*Brigham and Women’s Hospital, Harvard Medical School
Boston, MA, USA*

Abstract

In this work, we present SEMANTICVESSEL, a dataset for fine-grained brain vessel segmentation in computed tomography angiography scans. Based on the detailed contrast provided by dynamic 4D-CTA scans, we generate segmentation traces for arteries and veins. We then use intensity-guided region growing to obtain segmentations of the majority of vascular territories in the human brain, which are refined and annotated with 20 unique arterial classes by an expert radiologist. Unlike existing datasets, where minor arteries are discarded as background content, we merge these minor arteries into a generic arterial class. Due to the multiple-phase acquisition of dynamic 4D-CTA, labels for a single phase can be re-used for other phases in the same series, greatly increasing the size of our dataset with no additional annotation cost. The results show that models trained with the additional generic artery class produce better fine-grained segmentations across the board. Code and weights, as well as instructions to access the data, are available on <https://github.com/alceballosa/robust-vessel-segmentation>.

* Chu-Hsuan Lin and Alberto Mario Ceballos-Arroyo contributed equally to this work.

1. Introduction

Intracranial vessel segmentation is a critical task in neuroimaging with broad clinical relevance; the diagnosis and management of cerebrovascular diseases (CVDs) such as stroke, intracranial aneurysm, arteriovenous malformation, and cerebral small vessel disease. Accurate distinction of individual vessel segments, including major arteries such as the internal carotid artery (ICA), middle cerebral artery (MCA), anterior cerebral artery (ACA), and posterior cerebral artery (PCA), as well as the venous territories, provides clinicians with quantitative biomarkers for disease assessment, treatment planning, and longitudinal monitoring. Despite the clinical importance of this task, it remains largely performed manually by trained radiologists, a process that is time-consuming, expensive, and subject to inter-rater variability.

Today, intracranial vessels are imaged using various modalities, including computed tomography angiography (CTA), magnetic resonance angiography (MRA), and digital subtraction angiography (DSA). DSA is considered to be the gold standard for some cerebrovascular diseases (CVDs), but its invasive, procedural complication risk, and reliance on X-ray imaging have led to it being used less routinely than the other two (Bash et al., 2005). Compared with MRA, CTA is more cost-effective and widely used in clinical practice due to its accessibility, faster acquisition time, and compatibility with emergency settings. Although MRA offers higher soft-tissue contrast and can provide more detailed visualization of vascular structures without ionizing radiation, its longer scan time and higher cost limit its routine use, particularly in acute scenarios. In addition, CTA and MRA have been shown (Chen et al., 2018) to have comparable diagnostic performance for the detection of intracranial aneurysms, with CTA achieving a pooled sensitivity of 0.84 and a specificity of 0.85, and MRA a sensitivity of 0.80 and a specificity of 0.87. Furthermore, Phillips and Bubash (2002) reported that CTA can outperform MRA for carotid vascular disease detection, with the former showing substantially higher sensitivity than MRA for detecting intracranial stenosis (98% vs. 70%) and occlusion (100% vs. 87%). As a result, even in centers where MR scanners are readily available, a single patient will receive scans from both modalities at different points in time based on various criteria.

Despite the importance of both modalities in the clinic, most work on intracranial vessel segmentation has focused on MRA data (Hilbert et al., 2020; Min et al., 2024; Zhou et al., 2024). To address this gap, we propose a unified framework for intracranial vessel semantic segmentation in CTA that jointly delineates and labels both arterial and venous structures at the segment level. Our approach leverages time-resolved CT angiography (4D-CTA), where contrast dynamics naturally distinguish arterial and venous phases. We construct weakly supervised training targets by combining binary vessel segmentation with semantic labels through proximity-based propagation. To the best of our knowledge, this is the first time end-to-end semantic segmentation of intracranial cerebrovascular vessels is achieved, simultaneously identifying and labeling both arteries and veins within a single model.

Generalizable Insights about Machine Learning in the Context of Healthcare

Semantic segmentation is widely used in medical imaging, with applications such as organ and brain segmentation (Meijs et al., 2017; Wasserthal et al., 2023; Ma et al., 2024) playing an important role in clinical practice. In intracranial vessel segmentation, automated

semantic segmentation can assist in the diagnosis of cerebrovascular diseases and reduce radiologist workload. Automatically-generated brain vessel segmentations can also be very useful to contrastively pre-train vision-language models so that they better recognize the brain vasculature. However, many public datasets do not adequately capture very distal vessels due to limited image contrast and the small size of these structures. As a result, during model training, distal vessels are often lumped together with background parenchyma. Recent work (Saluja et al., 2025) has shown that combining many structures under a generic background label can hamper a model’s capabilities. Inspired by this finding, we annotate as many veins as possible and label the majority of important brain arteries while retaining an “undefined” class to account for arterial regions that would be too costly to annotate in detail. Our experiments indicate that dynamic 4D-CTA data can significantly increase the amount of brain CTA that can be annotated with fine-grained labels by a single neuro-radiologist, shifting the majority of the burden to the relatively simple task of assigning labels to segmented vessel segments. We hope that our work will inspire the development of new and more comprehensive datasets for intracranial vessel segmentation, as well as novel contrastive pre-training strategies for cerebrovascular data.

2. Related Work

2.1. Annotating and Segmenting Brain Vessels in CTA Scans

Most prior work approached intracranial vessel segmentation as a binary task, distinguishing vessels from background tissue (Meijs et al., 2017; Fu et al., 2020; Patel et al., 2023; van Voorst et al., 2026), resulting in a unified vascular mask without anatomical differentiation. Some studies aiming to incorporate anatomical specificity have followed two main approaches:

The first involves direct multiclass voxel-wise segmentation; the TopCoW challenge (Yang et al., 2025) established the first public benchmark for 13-class Circle of Willis (CoW) segmentation across both MRA and CTA modalities, with topology-aware approaches subsequently improving structural consistency within this region (Zhang et al., 2024; Hamadache et al., 2026). However, these efforts are confined to the Circle of Willis, limiting their applicability to other parts of the brain. The recently announced TopBrain challenge (Yang et al., 2026) extends the annotation scope to over 40 vessel labels spanning both arteries and veins across CTA and MRA, but only 25 scans per modality are available.

The second direction treats anatomical identification as a post-hoc labeling problem applied on top of a binary segmentation. Some of these approaches rely on graphs, assigning labels to bifurcation nodes or centerline segments via graph matching (Rist et al., 2023; Thamm et al., 2022) or deep learning on preprocessed vessel masks (Chen et al., 2024). Other approaches have attempted to assign labels to segments by registering arterial atlases to the target scans, but their use remains limited to MRA data (Falcetta et al., 2026). Overall, while covering a wider anatomical scope beyond the Circle of Willis, these methods are two-stage pipelines that depend on the quality of the upstream binary segmentation.

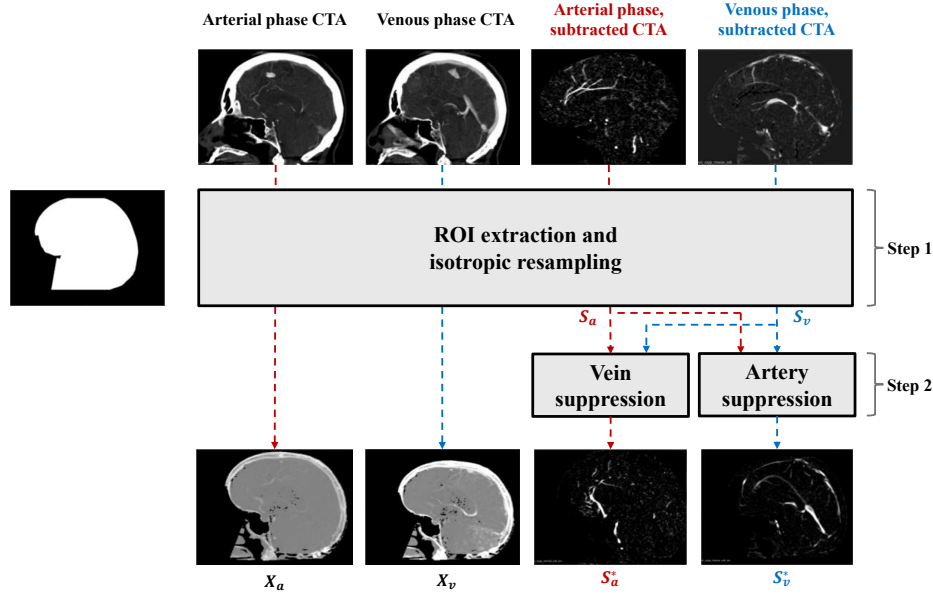


Figure 1: Pre-processing steps for a dynamic CTA study. For both phases, a head ROI is used on the CTA and the subtracted CTA to suppress the background. Next, the subtracted images are processed to create the vessel-separated images.

2.2. Gaps in existing annotation approaches

Existing large-scale annotation efforts for brain CTA usually produce annotations made from scratch, meaning experts have to review and annotate hundreds of individual slices. 3D augmented reality tools have been developed and used in some cases (Yang et al., 2025, 2026). However, due to the high workload of neuro-radiologists, the amount of data that can be annotated is limited, and in the case of TopBrain, the performance of published methods is constrained by the size of dataset. The VesselVerse (Falcetta et al., 2025) initiative tackles the annotation availability issue by proposing an integrative framework that combines annotations from multiple experts and models on various CTA and MRA datasets, but it remains limited to binary vessel-background segmentations.

Due to the superior contrast and background-vessel separability of MRA, comprehensive annotation of such datasets is far easier. For this modality, tools such as iCafe (Chen et al., 2019) provide a relatively fast annotation framework where the neuro-radiologist only has to assign labels to individual segments, but it is not directly usable with CTA scans and require significant contrast enhancement for this modality. Recently, it has been proposed to subtract baseline images from arterial- and venous-phase images in Dynamic 4D-CTA datasets to simulate the contrast provided by MRA, enabling the use of iCafe for CTA (Ceballos-Arroyo et al., 2026). However, iCafe introduces a complicated, non-open-source dependency, limiting the reproducibility of such an approach.

3. Methods

3.1. Dynamic CTA preprocessing

To obtain better contrast arterial and venous structures, a two-step pre-processing pipeline was applied to the dynamic CTA images, as illustrated in Fig. 1. The pipeline takes as input the arterial (X_a) and venous (X_v) CTA images with their scanner-generated subtracted counterparts (S_a , S_v). In the first step, each phase is registered to a common spatial reference frame to correct for inter-acquisition motion; taking the arterial phase as the reference, the venous CTA image X_v is affine registered to X_a , and the same transformation is applied to S_v to obtain $S_{v \rightarrow a}$, the subtracted venous image is warped into the arterial space. A skull-and-neck binary mask, derived from a CT atlas (Talou et al., 2021) via affine registration, is additionally applied to exclude extracranial voxels from all images. The same procedure is repeated for the venous phase to obtain $S_{a \rightarrow v}$.

In the second step, to enhance vascular contrast and eliminate erroneous highlights from the opposing phase, separate vein and artery suppression are performed. Specifically, vein suppression is applied to the arterial phase by retaining voxels where $S_a - S_{v \rightarrow a} > 0$ and zeroing out those where $S_a - S_{v \rightarrow a} \leq 0$, yielding the bone- and vein-suppressed volume S_a^* . Artery suppression is applied to the venous phase, retaining voxels where $S_v - S_{a \rightarrow v} > 0$ and zeroing out those where $S_v - S_{a \rightarrow v} \leq 0$, producing the bone- and artery-suppressed volume S_v^* .

3.2. Ground Truth Generation

Building on the pre-processing pipeline described above, the bone- and vein-suppressed arterial volume S_a^* and the bone- and artery-suppressed venous volume S_v^* were used as the arterial and venous images, respectively.

To generate seed candidates for region growing, each volume was first normalized using Nyul histogram normalization to standardize intensity distributions across cases. Renyi entropy-based auto-thresholding was then applied to generate an initial binary vessel mask. The multi-scale Hessian vesselness filter was subsequently applied to enhance tubular structures using sigma range $[0.5, 2.0]$ mm with 4 steps, $\alpha = 0.5$, $\beta = 1.0$, $\gamma = 5.0$. The vesselness response was normalized to $[0, 1]$ and thresholded at 0.2 to ensure vessel connectivity, yielding the binary vessel masks T_a^* and T_v^* with small connected components removed.

To recover missed thin vessels, intensity-guided region growing was applied to S_a^* and S_v^* . Seed voxels were selected from T_a^* and T_v^* and ranked by their local neighborhood maximum intensity. For each seed, a tolerance was defined as the difference between the local neighborhood maximum and the seed voxel intensity, and was constrained to lie within $[\frac{1}{5}, \frac{1}{2}]$ of the neighborhood maximum to prevent over- or under-segmentation. Seeds failing this criterion were skipped and the next candidate was evaluated. To ensure spatial diversity, candidate seeds located too close to previously accepted seeds were excluded from further consideration. Seed spacing was set to 5 voxels with neighborhood size of 7 voxels for arteries and 10 voxels for veins, reflecting the larger cross-sectional diameter of venous structures compared to arteries. The region growing process continued iteratively until the segmented region remained below 1.5% of the total volume for venous structures and 1% for arterial structures. In cases where the default region-growing parameters produced suboptimal

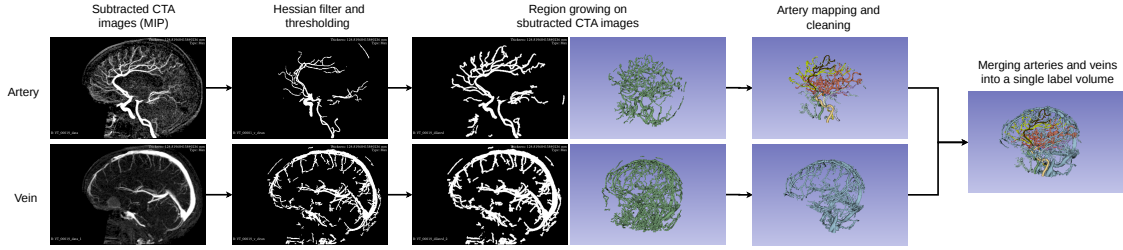


Figure 2: The proposed methodology for semi-automated brain vessel annotation.

results, assessed by visual inspection on the source CTA image, parameters were adjusted, and the pipeline was rerun.

The artery and vein masks were then merged, and a brain mask (Wasserthal et al., 2023) was applied to remove non-brain arterial components and retain only the largest venous component, with arteries prioritized in overlapping regions. The full process is illustrated in Figure 2.

For semantic labeling, a skeletonization algorithm was applied to the binary arterial masks to extract vessel centerlines. An experienced neuro-radiologist with 5 years of subspecialty experience then manually assigned 20 arterial class labels to the extracted centerlines using our in-house developed interactive GUI (Figure 3), which supports both 3D and 2D centerline-based label assignment as well as brush-based tools for adding or erasing vessel segmentations where necessary. The radiologist additionally reviewed and confirmed the quality of the underlying segmentation prior to finalizing the annotations. The classes are ICA (R/L), M1 (R/L), M2+ (R/L), A1 (R/L), A2+ (R/L), AComm, VA (R/L), BA, P1 (R/L), P2+ (R/L), and PComm (R/L). Labels were assigned to artery voxels based on the nearest labeled voxel within a 2.5 mm radius. Voxels beyond this threshold were assigned to an “other artery” class, while all venous voxels were grouped into a single “vein” class, resulting in a total of 22 classes. Finally, all phases of the dynamic CTA were rigidly registered to the arterial phase using ANTs (Tustison et al., 2021), and the same label map was propagated across phases. In total, we produced semi-automatic fine-grained annotations for 41 patients with a total 360 CTA volumes. In addition, the radiologist manually annotated the vascular tree for 5 held-out patients with a total of 50 scans using 3D Slicer (Kikinis et al., 2014), based on their peak arterial and venous phase scans.

3.3. Training

We trained a 3D segmentation network using the nnU-Net framework (Isensee et al., 2021) with the Residual Encoder Large (ResEncL) preset. The model was trained on all available cases for 1,000 epochs using stochastic gradient descent (SGD) with an initial learning rate of 0.01, momentum of 0.99, and weight decay of 3×10^{-5} . The network receives two input channels: the CTA image and a subject-specific left-right coordinate volume, in which each voxel encodes its normalized distance relative to the brain midline, ranging from -1 (left) to $+1$ (right). This coordinate volume provides the model with explicit spatial context regarding hemispheric laterality, enabling it to distinguish anatomically homologous structures on opposite sides of the brain, which is particularly important for

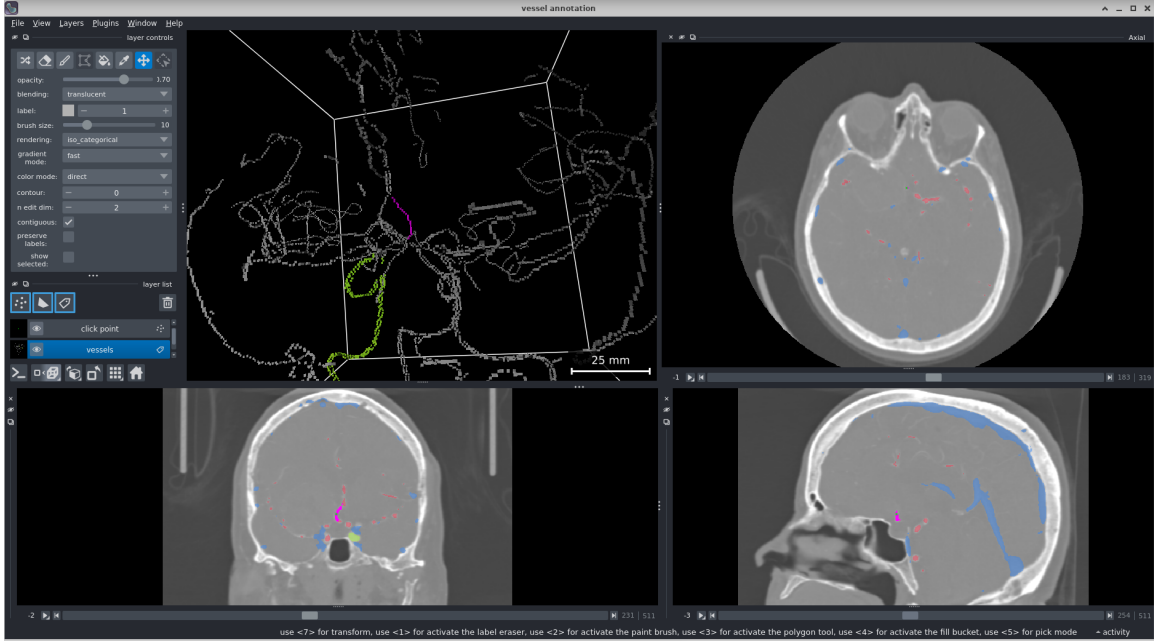


Figure 3: Annotation graphical user interface (GUI) for vessel labeling. Arteries and veins are displayed in **red** and **blue** by default, respectively. Once labeled, each vessel is assigned a unique color corresponding to its vessel name (e.g., **green**). Clicking on a vessel highlights it simultaneously in both the 3D rendering and 2D slice views in **bright magenta**, enabling consistent cross-view identification during the annotation process.

paired vessels such as the left and right middle cerebral arteries. The base training loss combines the Dice and cross-entropy losses. To improve the connectivity and topological completeness of predicted vessel segments, we also trained a second model with skeleton recall loss (Kirchhoff et al., 2024), which penalizes missed centerline voxels more heavily and thereby better preserves vessel connectivity.

3.4. Post Processing

For all model outputs, we apply a per-class threshold to remove residual segmentations with under 100 voxels. This helps remove noisy predictions for subjects with anatomical variations or occluded arteries.

4. Cohort

Internal dynamic CTA Dataset This retrospective study was approved by the Institutional Review Board with a waiver of informed consent (MGB HRC IRB Protocol #2022P000792). The 4D-CTA scans were retrieved from our clinical research database and acquired on a 320-detector-row CT scanner (Canon-Toshiba Aquilion One). Acquisition parameters were as follows: 80 kV tube voltage, 150 mA tube current, 0.75 s rotation time, 320 slices at 0.5 mm thickness, and a 512×512 matrix. For each patient, scanning commenced approximately 7 seconds following the injection of 75–100 mL of iodinated contrast agent

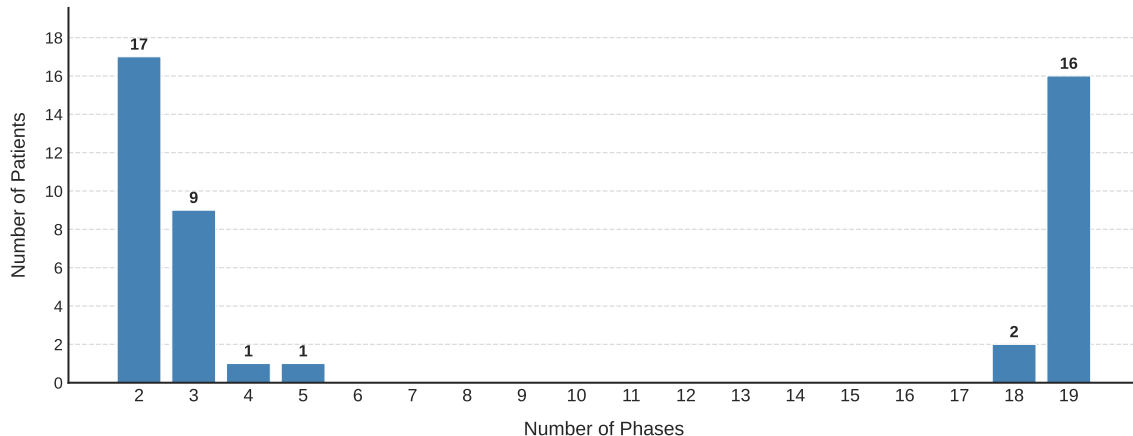


Figure 4: Number of dynamic CT phases per patient in the dataset, including test data (n=46).

(GE Omnipaque 350) administered at a flow rate of 4–5 mL/s, and continued intermittently across the baseline, arterial, and venous phases.

A total of 1,910 patients scanned between 2016 and 2022 were initially retrieved. Among these, 68 patients had dynamic CTA studies containing at least one peak arterial and one peak venous phase at acceptable resolution, including both original post-contrast images and corresponding subtracted images. Many cases additionally contained intermediate temporal phases, resulting in variable phase counts across the cohort. The distribution of phases is illustrated in Fig 4. From this cohort, 41 patients were randomly selected for semi-automatic annotation and model training. A further 5 patients scanned between 2023 and 2025 were chosen for our held-out test subset.

TopCoW and TopBrain Datasets The TopCoW (Yang et al., 2025) and TopBrain (Yang et al., 2026) datasets are publicly available neurovascular imaging datasets that provide paired MRA and CTA scans with annotated intracranial vessels. In this study, the 125 CTA images from TopCoW and the 25 CTA images from TopBrain are used as out-of-domain evaluation datasets¹. Table 1 and Figure 5 show the differences in coverage and size between the three datasets.

5. Results

5.1. Evaluation Approach/Study Design

To investigate the impact of label completeness on segmentation performance, we evaluate four label configurations: (1) All labels: training with all labeled structures, including arteries and veins; (2) All arteries: excluding the *vein* class; (3) Named arteries: excluding both the *other artery* and *vein* classes; and (4) Ignore labels: treating both the *other artery* and *vein* voxels as ignored labels that are not taken into consideration, such that these regions do not contribute to the loss during training. To isolate the effect of each variable,

1. The VesselVerse dataset contains no fine-grained vessel annotations, and thus is omitted from our evaluations.

Table 1: Description of existing datasets for brain vessel segmentation in CTA. Coverage for contrast phases (Ctrs.), proximal arteries (Pr.), first (1O) and second (2O) order arteries, distal arteries (Di.), arterial branches (Br.) and veins (V.) is indicated as $\bullet$: complete or near-complete, $\blacktriangleright$: partial, or $\circ$: absent. We only include training partitions when describing dataset size.

Dataset	# scans	Labels	Coverage						
			Ctrs.	Pr.	1O	2O	Di.	Br.	V.
TopCoW (Yang et al., 2025)	125	13 arteries	$\circ$	$\bullet$	$\bullet$	$\circ$	$\circ$	$\circ$	$\circ$
VesselVerse (Falcetta et al., 2025)	125	Vessel-only	$\circ$	$\bullet$	$\bullet$	$\bullet$	$\circ$	$\circ$	$\blacktriangleright$
TopBrain (Yang et al., 2026)	25	6 veins / 34 arteries	$\circ$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
DynaVessel (Ceballos-Arroyo et al., 2026)	110	Vein / Artery	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$
SEMANTICVESSEL	360	Vein / 22 arteries	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$	$\bullet$

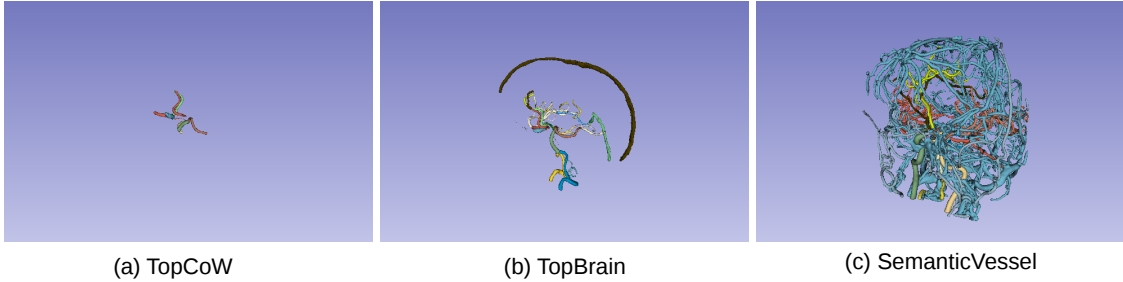


Figure 5: Example of our proposed brain vessel segmentations, compared with the public datasets TopCoW (Yang et al., 2025) and TopBrain (Yang et al., 2026)

we adopt a sequential evaluation strategy. First, we vary the label configuration while keeping the loss function fixed to the default nnU-Net compound loss (Dice and Cross-Entropy), training one model per configuration for a total of three models. Based on these results, we identify the best-performing label configuration and investigate the effect of the loss function by additionally training two models with Skeleton Recall Loss (Kirchhoff et al., 2024).

In addition to validating the generalizability of our annotation pipeline, we conduct two complementary experiments: (1) Arterial phase only: training an additional model using only the arterial phase images and comparing its performance against the multiphase model; and (2) Cross-dataset training: training on a subset of the TopBrain CTA scans and evaluating the TopBrain, SEMANTICVESSEL multiphase, and SEMANTICVESSEL arterial phase-only models on 5 held-out TopBrain cases and 5 held-out SEMANTICVESSEL cases.

5.2. Evaluation Metrics

As established before, we evaluate our trained models on TopCoW (Yang et al., 2025) and TopBrain (Yang et al., 2026). Since the label definitions in these datasets differ from those used in our training data, we establish a unified label scheme by merging semantically equivalent classes across datasets prior to evaluation. The full details of the label mapping are provided in the Appendix A1.

We assess segmentation performance using three complementary metrics designed to capture volumetric overlap, boundary accuracy, and topological completeness, respectively. Let A_i denote the ground truth volume, P the predicted volume, $A_{s,i}$ and P_s the corre-

sponding surfaces, and $A_{c,i}$ the centerline volume of A_i , we define the three metrics as follows:

(1) Modified Dice Coefficient (mDC) measures the sensitivity of the predicted segmentation with respect to the ground truth, and is defined as:

$$\text{mDC}(A_i, P) = \frac{|A_i \cap P|}{|A_i|} \quad (1)$$

Unlike the standard Dice coefficient, mDC is asymmetric and penalizes false negatives rather than false positives, making it more suitable for evaluating recall of small and distal vessels.

(2) Average Directed Hausdorff Distance (adHD) measures the mean surface distance from the ground truth boundary to the nearest predicted boundary point:

$$\text{adHD}(A_{s,i}, P_s) = \frac{1}{|A_{s,i}|} \sum_{a \in A_{s,i}} \min_{p \in P_s} d(a, p) \quad (2)$$

where $d(a, p)$ is the Euclidean distance between surface points. We adopt the directed variant rather than the symmetric Hausdorff distance, as the predicted segmentation boundary can be substantially larger than the ground truth in vessel segmentation. Compared with mDC, the adHD metric penalizes segmentations that are too thick compared with the ground truth.

(3) Topology Sensitivity (tSens) quantifies the fraction of the ground truth centerline captured by the predicted segmentation, following cDice (Shit et al., 2021):

$$\text{tSens}(A_{c,i}, P) = \frac{|A_{c,i} \cap P|}{|A_{c,i}|} \quad (3)$$

A high tSens indicates that the predicted segmentation preserves the topological connectivity of the vessel tree, which is critical for downstream clinical analysis.

5.3. Results on Fine-Grained Vessel Segmentation

All models were trained and tested on a workstation with an NVIDIA 4090 GPU (24 GB VRAM) and 8 CPU cores. Table 2 reports the aggregated results across three experimental settings, evaluating model performance trained with Dice Loss and Cross-Entropy (CE) Loss and Skeleton Recall Loss on the TopBrain dataset. For Dice Loss and Cross-Entropy (CE) Loss, across all three metrics, the *Named arteries* configuration establishes a baseline, while progressively incorporating additional structures — extra arteries and veins — leads to consistent improvements across all metrics. Notably, the *All arteries* configuration (named arteries + other artery) accounts for the majority of the performance gain, suggesting that the extra artery class provides the most significant supervisory benefit. **The *All arteries* configuration achieves the highest artery-only mDC of 0.8847 ± 0.0407 . When further incorporating the vein class in the *All labels* configuration, the artery mDC decreases slightly (0.8734 ± 0.0388), but we achieve the highest tSens (0.9387 ± 0.0257) and lowest adHD (0.2399 ± 0.1289) among all configurations, indicating that adding veins provides complementary benefit without sacrificing artery segmentation performance.** A detailed per-label breakdown result for the Skeleton Recall Loss can be found in Table A2. Figure 6 depicts a comparison between the *Named*

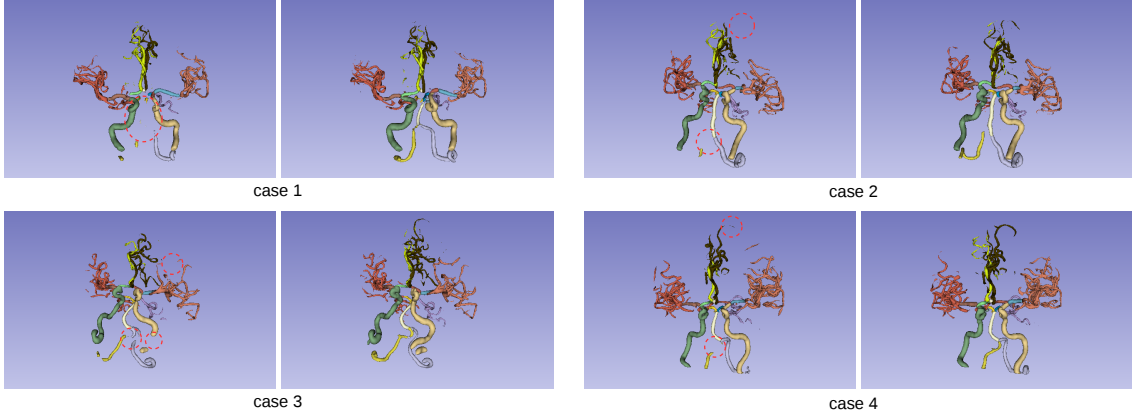


Figure 6: Comparison between the named arteries (left) and all labels (right) training configurations. Red dashed circles indicate regions of missing predictions in the model that was only trained on the named arteries.

arteries and *All labels* training configurations. While differences in distal vessels are subtle, the *Named arteries* model notably fails to segment major arteries such as the VA and ICA.

When using Skeleton Recall Loss, we see a slight improvement across the board when compared to the model trained using Dice and CE Loss, with the relative ranking of vessel classes and label configurations remaining stable across both training objectives. The vein class continues to be segmented very well, while large and well-opacified vessels such as the M2+ branches maintain near-ceiling scores across all metrics. Small communicating arteries remain the most challenging structures, with AComm and PComm exhibiting the lowest Dice scores and highest variance.

For the TopCoW dataset, consistent findings are observed across all evaluated configurations. The *All labels* configuration trained with Skeleton Recall Loss achieves the best overall performance, with the highest artery mDC of 0.9638 ± 0.0273 and tSens of 0.9909 ± 0.0153 , along with the lowest adHD of 0.0334 ± 0.0288 (Tables A3). Per-label results further confirm this trend (Table A4 and A5), where the *All labels* configuration consistently outperforms or matches *All arteries* across most vessel classes, particularly for smaller and more challenging vessels such as AComm and PComm.

Table A6 compares multiphase and arterial phase only training under the *All labels* configuration using Skeleton Recall Loss. Multiphase training consistently outperforms arterial phase only training in artery segmentation across both datasets, with mDC improvements of 0.0409 on TopBrain and 0.0166 on TopCoW. However, vein segmentation shows a slight decrease of 0.0305 in mDC compared to arterial phase only training. When compared against a model trained on a subset of TopBrain, the models trained on both the single-phase and multi-phase variants of SEMANTICVESSEL outperform the former on two held-out evaluation sets comprising 5 scans from TopBrain and 5 manually annotated arterial phase scans from SEMANTICVESSEL (Tables A7, A8, A9, and A10). Additional results are provided in the Appendix, including per-label results with Dice and CE Loss on TopBrain (Table A2) and full evaluations for TopCoW (Tables A3, A4, and A5).

Table 2: Aggregate segmentation performance by vessel mode across label configurations using Dice Loss, Cross-Entropy (CE) Loss, and Skeleton Recall Loss on the TopBrain dataset (mean $\pm$ std).

Loss	Configuration	Evaluated vessel	mDC $\uparrow$	tSens $\uparrow$	adHD $\downarrow$
Dice/CE	All labels	All vessel	0.9116 ± 0.0296	0.9600 ± 0.0175	0.1514 ± 0.0746
		Artery	0.8734 ± 0.0388	0.9387 ± 0.0257	0.2399 ± 0.1289
		Vein	0.9216 ± 0.0318	0.9738 ± 0.0254	0.1057 ± 0.0482
	All arteries	Artery	0.8847 ± 0.0407	0.9218 ± 0.0356	0.2748 ± 0.1663
	Names arteries	Artery	0.7279 ± 0.0839	0.7649 ± 0.0706	1.5167 ± 0.8808
	Ignore label	Artery	0.8395 ± 0.0464	0.8475 ± 0.0462	0.8896 ± 0.3535
Skeleton	All labels	All vessel	0.9381 ± 0.0217	0.9740 ± 0.0120	0.1008 ± 0.0511
		Artery	0.9025 ± 0.0373	0.9477 ± 0.0228	0.1945 ± 0.1176
		Vein	0.9422 ± 0.0334	0.9724 ± 0.0253	0.0758 ± 0.0435
	All arteries	Artery	0.8859 ± 0.0427	0.9396 ± 0.0276	0.2349 ± 0.1596

Table 3: Per-label segmentation results across label configurations trained with the Skeleton Recall Loss on the TopBrain dataset (mean $\pm$ std).

Label	mDC $\uparrow$		tSens $\uparrow$		adHD $\downarrow$	
	All labels	All arteries	All labels	All arteries	All labels	All arteries
ICA (R)	0.7768 ± 0.0799	0.7553 ± 0.0810	0.9029 ± 0.0705	0.8989 ± 0.0704	0.4256 ± 0.2191	0.4520 ± 0.2019
ICA (L)	0.7848 ± 0.0687	0.7703 ± 0.0645	0.9019 ± 0.1132	0.8996 ± 0.1122	0.3776 ± 0.1695	0.3955 ± 0.1512
M1 (R)	0.6607 ± 0.1540	0.6814 ± 0.1397	0.5840 ± 0.2040	0.6150 ± 0.1935	1.3845 ± 0.9958	1.2027 ± 0.7869
M1 (L)	0.6748 ± 0.1463	0.6844 ± 0.1506	0.5628 ± 0.2069	0.5683 ± 0.2027	1.5151 ± 0.9186	1.4746 ± 0.9152
M2+ (R)	0.9693 ± 0.0255	0.9583 ± 0.0297	0.9870 ± 0.0116	0.9807 ± 0.0197	0.0271 ± 0.0248	0.0383 ± 0.0315
M2+ (L)	0.9764 ± 0.0174	0.9669 ± 0.0213	0.9930 ± 0.0130	0.9895 ± 0.0133	0.0170 ± 0.0168	0.0258 ± 0.0192
ACA (R)	0.8923 ± 0.0914	0.8752 ± 0.0975	0.9517 ± 0.0795	0.9285 ± 0.0880	0.1085 ± 0.1208	0.1602 ± 0.2327
ACA (L)	0.8283 ± 0.0924	0.8564 ± 0.0829	0.8875 ± 0.1247	0.9097 ± 0.1012	0.2525 ± 0.3390	0.1872 ± 0.2718
AComm	0.6542 ± 0.2675	0.6244 ± 0.2871	0.7379 ± 0.4063	0.7359 ± 0.4272	0.2715 ± 0.2451	0.4666 ± 0.8098
VA (R)	0.7992 ± 0.1320	0.7620 ± 0.1299	0.9268 ± 0.1281	0.9060 ± 0.1377	0.4492 ± 0.6051	0.6371 ± 1.1562
VA (L)	0.8605 ± 0.0749	0.8294 ± 0.0912	0.9720 ± 0.0499	0.9529 ± 0.0634	0.1899 ± 0.1391	0.3368 ± 0.4004
BA	0.7561 ± 0.0851	0.7600 ± 0.0863	0.9287 ± 0.0615	0.9435 ± 0.0557	0.4480 ± 0.2122	0.4216 ± 0.2071
PCA (R)	0.8729 ± 0.0877	0.8551 ± 0.1118	0.8955 ± 0.0933	0.8758 ± 0.1324	0.2964 ± 0.4132	0.4201 ± 0.6317
PCA (L)	0.9097 ± 0.0589	0.8923 ± 0.0790	0.9368 ± 0.0629	0.9144 ± 0.0888	0.1346 ± 0.1204	0.1956 ± 0.2004
PComm (R)	0.6180 ± 0.3774	0.5117 ± 0.3772	0.6886 ± 0.3982	0.5565 ± 0.4134	0.2899 ± 0.4676	0.7263 ± 0.9697
PComm (L)	0.5736 ± 0.3614	0.5850 ± 0.3652	0.6033 ± 0.3899	0.6433 ± 0.4057	0.3327 ± 0.3471	0.5488 ± 0.9391
Other artery	0.6202 ± 0.1488	0.6199 ± 0.1360	0.6569 ± 0.1331	0.6484 ± 0.1142	2.1038 ± 1.8708	2.1669 ± 1.9989
Vein	0.9422 ± 0.0334	—	0.9724 ± 0.0253	—	0.0758 ± 0.0435	—

6. Discussion

In this paper, we proposed a semi-automatic methodology for fine-grained annotation of arteries in 3D CTA data. By relying on Dynamic 4D CTA, we are able to significantly reduce the annotation burden compared to existing approaches where radiologists need to annotate highly dense 3D volumes from scratch. Our pipeline instead requires only minimal expert intervention. By training our model with four different label configurations, we demonstrated that incorporating unlabeled vessels and complementary vascular structures provides a richer supervisory signal, leading to consistent performance gains. Central to this finding is our explicit treatment of unlabeled vessels. Rather than naively assigning them to the background class, which would introduce conflicting supervisory signal and

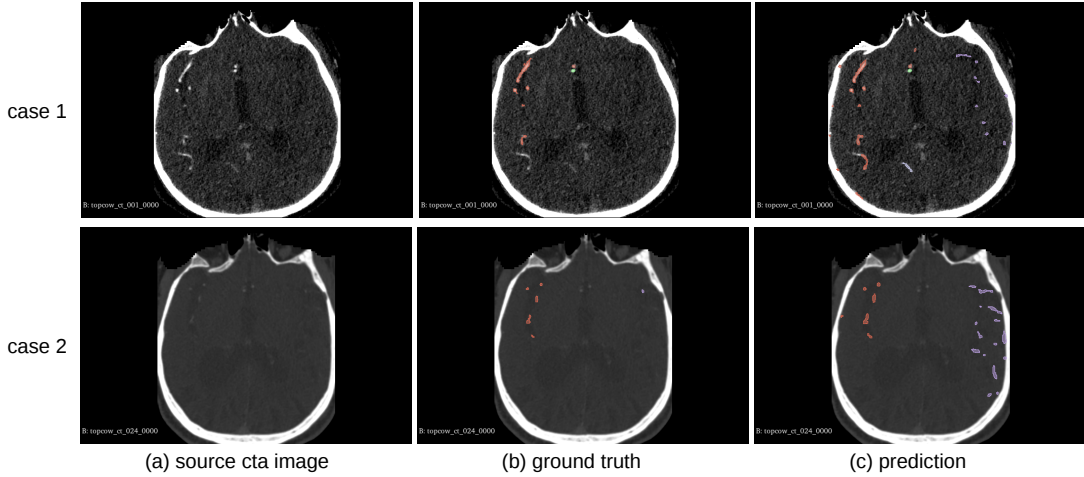


Figure 7: Comparison between ground truth annotations (b) and model predictions (c) on a CTA case from TopBrain. Case 1 (`topcow_ct_001`) The left ICA is occluded. The model predicts additional vessel structures on the left side M2 segment (purple, visible on the right half of column (c)). Case 2 (`topcow_ct_024`) demonstrates poor opacification from the M2 to M3 segment. The model predicts additional vessel structures in the left M3 segment (purple, visible on the right half of column (c)).

penalize the model for correctly detecting vessels outside the annotated set, we explicitly handle these regions to avoid false suppression of true vascular predictions. This distinction is particularly important in the context of cerebrovascular anatomy, where the vascular tree is highly interconnected and partially annotated datasets are the norm. By allowing the model to focus its learning on annotated structures without being misled by unannotated ones, our approach ultimately improves its ability to delineate individual arterial branches in fine-grained cerebrovascular segmentation.

Table 4: Aggregate segmentation performance with and without left-right coordinate information across label configurations using Dice Loss, Cross-Entropy (CE) Loss on the TopBrain dataset (mean $\pm$ std).

Configuration	Evaluated vessel	mDC $\uparrow$		tSens $\uparrow$		adHD $\downarrow$	
		With LR	Without LR	With LR	Without LR	With LR	Without LR
All labels	All vessels	0.9116 ± 0.0296	0.7876 ± 0.0556	0.9600 ± 0.0175	0.8468 ± 0.0379	0.1514 ± 0.0746	0.5159 ± 0.1965
	Artery	0.8734 ± 0.0388	0.6453 ± 0.0499	0.9387 ± 0.0257	0.7453 ± 0.0350	0.2399 ± 0.1289	1.3213 ± 0.5117
	Vein	0.9216 ± 0.0318	0.8545 ± 0.0713	0.9738 ± 0.0254	0.8949 ± 0.0694	0.1057 ± 0.0482	0.2986 ± 0.2165

In addition, we observed that our model predicts more vessel voxels than are present in the ground truth annotations. As illustrated in Figure 7, an expert radiologist verified that these additionally detected structures correspond to true arterial anatomy, suggesting that the model may be recovering anatomically valid vessel structures that were missed or incompletely annotated in the ground truth. These findings indicate that our model is robust and can detect vessels in lower contrast regions, where manual annotation is inherently difficult and prone to under-segmentation.

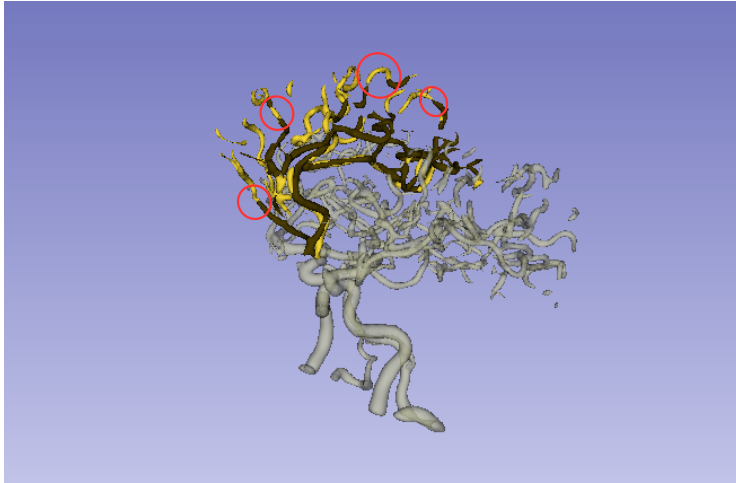


Figure 8: Representative failure cases of ACA segmentation with our best performing model (Skeleton Recall loss). Yellow: right ACA; Brown: left ACA. Red circle: misclassified regions.

Another observation from our experiments is the benefit of incorporating left-right coordinate information during model training. Cerebrovascular structures such as the ICA, M1, ACA, VA, PCA, and PComm are inherently paired, with corresponding branches on both hemispheres sharing similar morphology and spatial relationships. By explicitly encoding left-right coordinate information as an additional input, the model gains awareness of hemispheric context, enabling it to better distinguish between left and right counterparts and exploit their structural symmetry as a complementary supervisory signal. Table 4 presents the segmentation performance comparison between models trained with and without left-right coordinate information. However, certain structures, such as the ACA, especially in the A2 and more distal segments, remain challenging to separate due to their complex anatomy and close proximity to the midline of the brain (see Figure 8).

Limitations Although SEMANTICVESSEL is the largest dataset for fine-grained brain artery segmentation in CTA to date, key limitations have to be considered:

- Due to the presence of symmetrical structures on the left and right sides of the brain, we are forced to use an additional channel encoding laterality, which itself is derived from a brain mask produced by an off-the-shelf brain segmentation tool; in cases where the brain segmentation fails (*e.g.*, skull-stripped scans), performance can be expected to go down significantly.
- Our annotation scheme is more limited than that of TopBrain, which covers 34 arterial and 6 venous structures compared to the 21 arterial structures and a single venous label in our dataset; the lack of specificity could make models trained on SEMANTICVESSEL less useful in certain scenarios, such as grounded vision-language modeling.

- Our dataset has no coverage of occluded arteries or anatomical variants that are somewhat common in real-world scenarios; for instance, 9 out of 25 CTA scans in TopBrain are from patients without a right posterior communicating artery. Although the nnUNet trained on our data often does not make predictions for CTA with missing arteries, a comprehensive analysis with a larger number of patients with anatomical variants will be necessary.
- Since our annotations are not fully connected in all cases, predictions made by the nnUNet model sometimes display gaps; this limits the usability of our dataset for dynamic fluid simulations, which require fully accurate vessel trees and are critical tools in diagnosing patient risk under certain cerebrovascular diseases.
- The region growing step in our annotation pipeline involves heuristic parameters that may require case-specific tuning when applied to data from different scanners or acquisition protocols. While our interactive GUI tool allows radiologists to efficiently review and correct suboptimal segmentations, this represents an inherent limitation of the heuristic pipeline design that may increase the annotation burden for institutions with different imaging characteristics.

Acknowledgments

The authors acknowledge the financial support provided by NIH grant 1R01LM013891-01A1. Ceballos Arroyo, A. received funding from Colombia’s Minciencias and Fulbright under the Fulbright Minciencias 2021 program. This work was partially supported by the DeltaAI system at the National Center for Supercomputing Applications through allocation CIS260200 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

References

- Suzie Bash, J. Pablo Villablanca, Reza Jahan, Gary Duckwiler, Monica Tillis, Chelsea Kidwell, Jeffrey Saver, and James Sayre. Intracranial vascular stenosis and occlusive disease: evaluation with CT angiography, MR angiography, and digital subtraction angiography. *AJNR Am J Neuroradiol*, 26(5):1012–1021, May 2005. ISSN 0195-6108.
- Alberto M. Ceballos-Arroyo et al. Robust automatic brain vessel segmentation in 3D CTA scans using dynamic 4D-CTA data. *arXiv preprint arXiv:2602.00391*, 2026. URL <https://arxiv.org/abs/2602.00391>.
- Li Chen, Mahmud Mossa-Basha, Jie Sun, Daniel S. Hippe, Niranjana Balu, Quan Yuan, Kristi Pimentel, Thomas S. Hatsukami, Jenq-Neng Hwang, and Chun Yuan. Quantification of morphometry and intensity features of intracranial arteries from 3D TOF MRA using the intracranial artery feature extraction (iCafe): A reproducibility study. *Magn Reson Imaging*, 57:293–302, April 2019. ISSN 1873-5894. doi: 10.1016/j.mri.2018.12.007.

- Ting Chen, Wei You, Liyuan Zhang, Wanxing Ye, Junqiang Feng, Jing Lu, Jian Lv, Yudi Tang, Dachao Wei, Siming Gui, et al. Automated anatomical labeling of the intracranial arteries via deep learning in computed tomography angiography. *Frontiers in Physiology*, 14:1310357, 2024. doi: 10.3389/fphys.2023.1310357.
- Xiaodan Chen, Yun Liu, Huazhang Tong, Yonghai Dong, Dongyang Ma, Lei Xu, and Cheng Yang. Meta-analysis of computed tomography angiography versus magnetic resonance angiography for intracranial aneurysm. *Medicine (Baltimore)*, 97(20):e10771, May 2018. ISSN 0025-7974. doi: 10.1097/MD.00000000000010771. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC5976319/>.
- Daniele Falcetta, Vincenzo Marciano, Kaiyuan Yang, Jon Cleary, Loïc Legris, Massimiliano Domenico Rizzaro, Ioannis Pitsiorlas, Hava Chaptoukaev, Benjamin Lemasson, Björn Menze, and Maria A. Zuluaga. VesselVerse: A Dataset and Collaborative Framework for Vessel Annotation. In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume LNCS 15972. Springer Nature Switzerland, September 2025.
- Daniele Falcetta, Liane S. Canas, Lorenzo Suppa, Matteo Pentassuglia, Jon Cleary, Marc Modat, Sébastien Ourselin, and Maria A. Zuluaga. An automated framework for large-scale graph-based cerebrovascular analysis, 2026. URL <https://arxiv.org/abs/2512.03869>.
- Fan Fu, Jianyong Wei, Miao Zhang, Fan Yu, Yueting Xiao, Dongdong Rong, Yi Shan, Yan Li, Cheng Zhao, Fangzhou Liao, Zhenghan Yang, Yuehua Li, Yingmin Chen, Ximing Wang, and Jie Lu. Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network. *Nat Commun*, 11(1):4829, September 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-18606-2. URL <https://www.nature.com/articles/s41467-020-18606-2>.
- Rachika E. Hamadache, Clara Lisazo, Cansu Yalcin, Uma M. Lal-Trehan Estrada, Valeriia Abramova, Adrià Casamitjana, Arnau Oliver, and Xavier Lladó. Topology-aware multiclass segmentation of the Circle of Willis from MRA and CTA images. *Computers in Biology and Medicine*, 204:111516, 2026. doi: 10.1016/j.combiomed.2026.111516.
- Adam Hilbert, Vince I. Madai, Ela M. Akay, Orhun U. Aydin, Jonas Behland, Jan Sobesky, Ivana Galinovic, Ahmed A. Khalil, Abdel A. Taha, Jens Wuerfel, Petr Dusek, Thoralf Niendorf, Jochen B. Fiebach, Dietmar Frey, and Michelle Livne. BRAVE-NET: Fully Automated Arterial Brain Vessel Segmentation in Patients With Cerebrovascular Disease. *Front. Artif. Intell.*, 3, September 2020. ISSN 2624-8212. doi: 10.3389/frai.2020.552258. URL <https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2020.552258/full>.
- Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*, 18(2):203–211, February 2021. ISSN 1548-7105. doi: 10.1038/s41592-020-01008-z. URL <https://www.nature.com/articles/s41592-020-01008-z>.

- Ron Kikinis, Steve D. Pieper, and Kirby G. Vosburgh. *3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support*, pages 277–289. Springer New York, New York, NY, 2014. ISBN 978-1-4614-7657-3. URL https://doi.org/10.1007/978-1-4614-7657-3_19.
- Yannick Kirchhoff, Maximilian R. Rokuss, Saikat Roy, Balint Kovacs, Constantin Ulrich, Tassilo Wald, Maximilian Zenk, Philipp Vollmuth, Jens Kleesiek, Fabian Isensee, and Klaus Maier-Hein. Skeleton Recall Loss for Connectivity Conserving and Resource Efficient Segmentation of Thin Tubular Structures. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *Computer Vision – ECCV 2024*, pages 218–234, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-72980-5. doi: 10.1007/978-3-031-72980-5_13.
- Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nat Commun*, 15(1):654, January 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-44824-z. URL <https://www.nature.com/articles/s41467-024-44824-z>.
- Midas Meijs, Ajay Patel, Sil C. van de Leemput, Mathias Prokop, Ewoud J. van Dijk, Frank-Erik de Leeuw, Frederick J. A. Meijer, Bram van Ginneken, and Rashindra Maniying. Robust Segmentation of the Full Cerebral Vasculature in 4D CT of Suspected Stroke Patients. *Sci Rep*, 7(1):15622, November 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-15617-w. URL <https://www.nature.com/articles/s41598-017-15617-w>.
- Yueqin Min, Jing Li, Shouqiang Jia, Yuehua Li, and Shengdong Nie. Automated Cerebrovascular Segmentation and Visualization of Intracranial Time-of-Flight Magnetic Resonance Angiography Based on Deep Learning. *J Imaging Inform Med*, 38(2):703–716, August 2024. ISSN 2948-2925. doi: 10.1007/s10278-024-01215-6. URL <https://pmc.ncbi.nlm.nih.gov/articles/PMC11950609/>.
- Tatsat R. Patel, Aakash Patel, Sricharan S. Veeturi, Munjal Shah, Muhammad Waqas, Andre Monteiro, Ammad A. Baig, Nandor Pinter, Elad I. Levy, Adnan H. Siddiqui, and Vincent M. Tutino. Evaluating a 3D deep learning pipeline for cerebral vessel and intracranial aneurysm segmentation from computed tomography angiography-digital subtraction angiography image pairs. *Neurosurg Focus*, 54(6):E13, June 2023. ISSN 1092-0684. doi: 10.3171/2023.3.FOCUS2374.
- C. Douglas Phillips and Lori A. Bubash. CT angiography and MR angiography in the evaluation of extracranial carotid vascular disease. *Radiol Clin North Am*, 40(4):783–798, July 2002. ISSN 0033-8389. doi: 10.1016/s0033-8389(02)00017-9.
- Leonhard Rist, Oliver Taubmann, Florian Thamm, Hendrik Ditt, Michael Sühling, and Andreas Maier. Bifurcation matching for consistent cerebral vessel labeling in CTA of stroke patients. *International Journal of Computer Assisted Radiology and Surgery*, 18(3):509–516, 2023. doi: 10.1007/s11548-022-02750-9.
- Rachit Saluja, Asli Cihangir, Ruining Deng, Johannes C. Paetzold, Fengbei Liu, and Mert R. Sabuncu. BackSplit: The Importance of Sub-dividing the Background in Biomedical Lesion Segmentation, November 2025. URL <http://arxiv.org/abs/2511.19394>. arXiv:2511.19394 [cs].

- Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien P. W. Pluim, Ulrich Bauer, and Bjoern H. Menze. clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16555–16564, Nashville, TN, USA, June 2021. IEEE. ISBN 978-1-66544-509-2. URL <https://ieeexplore.ieee.org/document/9578225/>.
- Gonzalo Daniel Maso Talou, Soroush Safaei, Peter John Hunter, and Pablo Javier Blanco. Adaptive constrained constructive optimisation for complex vascularisation processes. *Scientific Reports*, 11(1):6180, March 2021. ISSN 2045-2322. URL <https://www.nature.com/articles/s41598-021-85434-9>. Publisher: Nature Publishing Group.
- Florian Thamm et al. An algorithm for the labeling and interactive visualization of the cerebrovascular system of ischemic strokes. *Biomedical Physics & Engineering Express*, 8(6):065016, 2022. doi: 10.1088/2057-1976/ac9107.
- Nicholas J. Tustison, Philip A. Cook, Andrew J. Holbrook, Hans J. Johnson, John Muschelli, Gabriel A. Devenyi, Jeffrey T. Duda, Sandhitsu R. Das, Nicholas C. Cullen, Daniel L. Gillen, Michael A. Yassa, James R. Stone, James C. Gee, and Brian B. Avants. The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports*, 11(1):9068, April 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-87564-6. URL <https://doi.org/10.1038/s41598-021-87564-6>.
- Henk van Voorst, Jiahang Su, Praneeta R. Konduri, Charles B. L. M. Majoie, Yvo B. W. E. M. Roos, Bart J. Emmer, Henk A. Marquering, Bob D. de Vos, Matthan W. A. Caan, Ivana Išgum, and MR CLEAN Registry collaborators. Deep generative models for vessel segmentation in CT angiography of the brain. *Comput Biol Med*, 202:111432, February 2026. ISSN 1879-0534. doi: 10.1016/j.combiomed.2025.111432.
- Jakob Wasserthal, Hanns-Christian Breit, Manfred T. Meyer, Maurice Pradella, Daniel Hinck, Alexander W. Sauter, Tobias Heye, Daniel T. Boll, Joshy Cyriac, Shan Yang, Michael Bach, and Martin Segeroth. TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. *Radiology: Artificial Intelligence*, 5(5):e230024, September 2023. doi: 10.1148/ryai.230024. URL <https://pubs.rsna.org/doi/full/10.1148/ryai.230024>.
- Kaiyuan Yang, Fabio Musio, Yihui Ma, Norman Juchler, et al. Benchmarking the CoW with the TopCoW challenge: Topology-aware anatomical segmentation of the circle of Willis for CTA and MRA, 2025.
- Kaiyuan Yang, Pengcheng Shi, Houjing Huang, Fabio Musio, Hakim Baazaoui, Orhun Utku Aydin, Adam Hilbert, Rachika E. Hamadache, Cansu Yalcin, Minghui Zhang, Daniele Falcetta, Ezequiel de la Rosa, Suprosanna Shit, Chinmay Prabhakar, Bastian Wittmann, Maximilian R. Rokuss, Yannick Kirchhoff, Rami Al-Maskari, Luciano Höher, Norman Juchler, Adrià Casamitjana, Jon Cleary, Anton Schmick, Philipp Baumgartner, Julian Deseö, Olafs Vandans, Dahye Lee, Kwanseok Oh, Dominic LaBella, Moona Mazher, Steven A. Niederer, Abdul Qayyum, Yaoyu Liu, Junqiang Chen, Wooseung Kim, Napasara Asawalertsak, Minjae Kim, Dongho Shin, Sung-Hong Park, Shunsuke Kikuchi,

- Yaqing Zhang, Jialu Liu, Yue Cui, Yuchen Qiu, Anouk Verschuur, Jiaxin Zhang, Irene van der Schaaf, Ruisheng Su, Chantal M. W. Tax, Yosuke Yamagishi, Nishta Letchumanan, Shouhei Hanaoka, Jesús González, Riccardo Tiberi, Clara Lisazo, Valeriia Abramova, Uma Maria Lal-Trehan Estrada, Agustin Cartaya Lathuliere, Micaela Rivas Díaz, Arnau Oliver, Menghan Zhang, Zhiqiang Bai, Klaus Maier-Hein, Yun Gu, Xavier Lladó, Maria A. Zuluaga, Dietmar Frey, Johannes C. Paetzold, Sven Hirsch, Susanne Wegener, Yihui Ma, and Bjoern Menze. TopBrain Segmentation Challenge for Whole Brain Vessel Anatomy, May 2026. URL <https://www.medrxiv.org/content/10.64898/2026.05.28.26354312v1>. ISSN: 3067-2007 Pages: 2026.05.28.26354312.
- Minghui Zhang et al. Topology-aware exploration of Circle of Willis for CTA and MRA: Segmentation, detection, and classification. *arXiv preprint arXiv:2410.15614*, 2024. URL <https://arxiv.org/abs/2410.15614>.
- Langtao Zhou, Huiting Wu, Guanghua Luo, and Hong Zhou. Deep learning-based 3D cerebrovascular segmentation workflow on bright and black blood sequences magnetic resonance angiography. *Insights Imaging*, 15(1):81, March 2024. ISSN 1869-4101. doi: 10.1186/s13244-024-01657-0.

Appendix

Table A1: Mapping of TopBrain and TopCoW vessel labels to the classes in SEMANTICVESSEL.

SemanticVessel	TopBrain	TopCoW
ICA (R), ICA (L)	R-ICA, L-ICA	R-ICA, L-ICA
M1 (R), M1 (L)	R-M1, L-M1	R-MCA, L-MCA
M2+ (R), M2+ (L)	R-M2, R-M3, L-M2, L-M3	—
ACA (R)	R-A1A2, R-A3	R-ACA
ACA (L)	L-A1A2, L-A3	L-ACA
PCA (R)	R-P1P2, R-P3P4	R-PCA
PCA (L)	L-P1P2, L-P3P4	L-PCA
PComm (R), PComm (L)	R-Pcom, L-Pcom	R-Pcom, L-Pcom
AComm	Acom	Acom
BA	BA	BA
VA (R), VA (L)	R-VA, L-VA	—
Other artery	3rd-A2, 3rd-A3, R-SCA, L-SCA, R-AICA, L-AICA, R-PICA, L-PICA, R-AChA, L-AChA, R-OA, L-OA	3rd-A2
Vein	VoG, StS, ICVs, R-BVR, L-BVR, SSS	—

Table A2: Per-label segmentation results across label configurations trained with Dice Loss and Cross-Entropy (CE) Loss (mean $\pm$ std) on the TopBrain dataset.

Label	mDC $\uparrow$			tSens $\uparrow$			adHD $\downarrow$		
	All labels	All arteries	Named arteries	All labels	All arteries	Named arteries	All labels	All arteries	Named arteries
ICA (R)	0.7668 $\pm$ 0.0745	0.7593 $\pm$ 0.1461	0.6849 $\pm$ 0.1037	0.9338 $\pm$ 0.0585	0.9080 $\pm$ 0.1443	0.8961 $\pm$ 0.1254	0.4091 $\pm$ 0.1777	0.4476 $\pm$ 0.3040	0.6087 $\pm$ 0.5091
ICA (L)	0.7824 $\pm$ 0.0699	0.7863 $\pm$ 0.0820	0.6776 $\pm$ 0.1435	0.9114 $\pm$ 0.0969	0.9121 $\pm$ 0.1057	0.8634 $\pm$ 0.1848	0.3511 $\pm$ 0.1454	0.3461 $\pm$ 0.1528	0.5555 $\pm$ 0.3857
M1 (R)	0.6438 $\pm$ 0.1382	0.6673 $\pm$ 0.1505	0.5971 $\pm$ 0.1256	0.5731 $\pm$ 0.2057	0.5875 $\pm$ 0.2009	0.5657 $\pm$ 0.2076	1.3758 $\pm$ 0.8946	1.2982 $\pm$ 0.9273	1.4593 $\pm$ 0.9582
M1 (L)	0.6351 $\pm$ 0.1490	0.6568 $\pm$ 0.1591	0.6520 $\pm$ 0.1477	0.5164 $\pm$ 0.2156	0.5497 $\pm$ 0.2192	0.5736 $\pm$ 0.2033	1.7690 $\pm$ 1.0639	1.5824 $\pm$ 1.0622	1.4142 $\pm$ 0.9039
M2+ (R)	0.9407 $\pm$ 0.0420	0.9165 $\pm$ 0.0463	0.8836 $\pm$ 0.0588	0.9768 $\pm$ 0.0253	0.9549 $\pm$ 0.0353	0.9214 $\pm$ 0.0514	0.0514 $\pm$ 0.0444	0.1181 $\pm$ 0.1068	0.1767 $\pm$ 0.1254
M2+ (L)	0.9592 $\pm$ 0.0260	0.9438 $\pm$ 0.0300	0.9203 $\pm$ 0.0515	0.9901 $\pm$ 0.0108	0.9711 $\pm$ 0.0244	0.9601 $\pm$ 0.0441	0.0285 $\pm$ 0.0198	0.0673 $\pm$ 0.0523	0.0955 $\pm$ 0.1080
ACA (R)	0.8830 $\pm$ 0.0883	0.8831 $\pm$ 0.0958	0.8698 $\pm$ 0.0527	0.9493 $\pm$ 0.0747	0.9339 $\pm$ 0.0813	0.9681 $\pm$ 0.0440	0.1282 $\pm$ 0.1435	0.1504 $\pm$ 0.1668	0.1222 $\pm$ 0.0792
ACA (L)	0.8137 $\pm$ 0.1090	0.8689 $\pm$ 0.0812	0.7734 $\pm$ 0.1106	0.8868 $\pm$ 0.1244	0.9290 $\pm$ 0.0945	0.8849 $\pm$ 0.1077	0.2425 $\pm$ 0.2312	0.1589 $\pm$ 0.1925	0.3111 $\pm$ 0.2817
AComm	0.4757 $\pm$ 0.3729	0.5342 $\pm$ 0.3544	0.3122 $\pm$ 0.3214	0.5286 $\pm$ 0.4752	0.5835 $\pm$ 0.4729	0.4929 $\pm$ 0.4792	0.5396 $\pm$ 0.9225	0.4809 $\pm$ 0.7639	0.4778 $\pm$ 0.3539
VA (R)	0.7246 $\pm$ 0.1513	0.8306 $\pm$ 0.0744	0.4415 $\pm$ 0.2579	0.8776 $\pm$ 0.1906	0.9480 $\pm$ 0.1106	0.5305 $\pm$ 0.3381	1.2630 $\pm$ 3.8663	0.3165 $\pm$ 0.4529	3.4846 $\pm$ 3.8562
VA (L)	0.8148 $\pm$ 0.0585	0.8443 $\pm$ 0.0653	0.6175 $\pm$ 0.2328	0.9601 $\pm$ 0.0556	0.9497 $\pm$ 0.0558	0.7246 $\pm$ 0.2703	0.2822 $\pm$ 0.1831	0.2522 $\pm$ 0.1907	1.6796 $\pm$ 2.4661
BA	0.7483 $\pm$ 0.0941	0.7480 $\pm$ 0.0948	0.6825 $\pm$ 0.2143	0.9461 $\pm$ 0.0478	0.9338 $\pm$ 0.0647	0.8704 $\pm$ 0.2601	0.4522 $\pm$ 0.2779	0.4895 $\pm$ 0.3258	0.5992 $\pm$ 0.8116
PCA (R)	0.8329 $\pm$ 0.1078	0.8333 $\pm$ 0.1330	0.8106 $\pm$ 0.1375	0.8849 $\pm$ 0.1160	0.8510 $\pm$ 0.1440	0.8677 $\pm$ 0.1580	0.5269 $\pm$ 0.8478	0.5880 $\pm$ 0.7596	0.7176 $\pm$ 1.1256
PCA (L)	0.8803 $\pm$ 0.0714	0.8763 $\pm$ 0.0857	0.8136 $\pm$ 0.1303	0.9334 $\pm$ 0.0763	0.9003 $\pm$ 0.1087	0.8800 $\pm$ 0.1204	0.2427 $\pm$ 0.4130	0.2883 $\pm$ 0.3569	0.5235 $\pm$ 0.6201
PComm (R)	0.5961 $\pm$ 0.3394	0.7269 $\pm$ 0.3181	0.2400 $\pm$ 0.3482	0.6453 $\pm$ 0.3853	0.7894 $\pm$ 0.3390	0.2998 $\pm$ 0.4174	0.7700 $\pm$ 1.0989	0.1989 $\pm$ 0.2544	0.2932 $\pm$ 0.2373
PComm (L)	0.6640 $\pm$ 0.2967	0.5773 $\pm$ 0.4203	0.2581 $\pm$ 0.3659	0.7301 $\pm$ 0.3148	0.6164 $\pm$ 0.4435	0.3137 $\pm$ 0.4463	0.4008 $\pm$ 0.4487	0.1290 $\pm$ 0.1379	0.2382 $\pm$ 0.0700
Other artery	0.6037 $\pm$ 0.1083	0.5963 $\pm$ 0.1123	—	0.6613 $\pm$ 0.0982	0.6452 $\pm$ 0.0998	—	2.2988 $\pm$ 2.1245	2.3860 $\pm$ 2.0281	—
Vein	0.9216 $\pm$ 0.0318	—	—	0.9738 $\pm$ 0.0254	—	—	0.1057 $\pm$ 0.0482	—	—

Table A3: Aggregate segmentation performance label configurations using Dice Loss/Cross-Entropy (CE) Loss, and Skeleton Recall Loss on the TopCoW dataset (mean $\pm$ std).

Loss	Configuration	Evaluated vessel	mDC $\uparrow$	tSens $\uparrow$	adHD $\downarrow$
Dice/CE	All labels	Artery	0.9457 $\pm$ 0.0359	0.9837 $\pm$ 0.0208	0.0533 $\pm$ 0.0401
	All arteries	Artery	0.9581 $\pm$ 0.0331	0.9846 $\pm$ 0.0220	0.0427 $\pm$ 0.0392
	Names arteries	Artery	0.8865 $\pm$ 0.0637	0.9510 $\pm$ 0.0457	0.1480 $\pm$ 0.1677
Skeleton	All labels	Artery	0.9638 $\pm$ 0.0273	0.9909 $\pm$ 0.0153	0.0334 $\pm$ 0.0288
	All arteries	Artery	0.9496 $\pm$ 0.0334	0.9836 $\pm$ 0.0218	0.0508 $\pm$ 0.0399

Table A4: Per-label segmentation results across label configurations trained with Dice Loss and Cross-Entropy (CE) Loss (mean $\pm$ std) on the TopCoW dataset.

Label	mDC $\uparrow$			tSens $\uparrow$			adHD $\downarrow$		
	All labels	All arteries	Named arteries	All labels	All arteries	Named arteries	All labels	All arteries	Named arteries
ICA (R)	0.7618 $\pm$ 0.0762	0.7449 $\pm$ 0.1509	0.7019 $\pm$ 0.0903	0.8904 $\pm$ 0.0874	0.8606 $\pm$ 0.1573	0.8703 $\pm$ 0.1062	0.4156 $\pm$ 0.1885	0.5784 $\pm$ 0.9936	0.5479 $\pm$ 0.3228
ICA (L)	0.7736 $\pm$ 0.0690	0.7659 $\pm$ 0.1504	0.7003 $\pm$ 0.0995	0.8910 $\pm$ 0.0954	0.8676 $\pm$ 0.1784	0.8681 $\pm$ 0.1202	0.3696 $\pm$ 0.1592	0.5735 $\pm$ 1.5193	0.4988 $\pm$ 0.2635
M1 (R)	0.8038 $\pm$ 0.1589	0.8345 $\pm$ 0.1550	0.7739 $\pm$ 0.1730	0.8143 $\pm$ 0.2082	0.8343 $\pm$ 0.1988	0.8143 $\pm$ 0.2189	0.6254 $\pm$ 0.9330	0.5308 $\pm$ 0.8611	0.7104 $\pm$ 1.1389
M1 (L)	0.8144 $\pm$ 0.1706	0.8251 $\pm$ 0.1718	0.8189 $\pm$ 0.1530	0.8207 $\pm$ 0.2163	0.8362 $\pm$ 0.2155	0.8585 $\pm$ 0.1992	0.6168 $\pm$ 0.8452	0.5641 $\pm$ 0.8045	0.4992 $\pm$ 0.8107
ACA (R)	0.9084 $\pm$ 0.1059	0.9014 $\pm$ 0.1168	0.8885 $\pm$ 0.1029	0.9524 $\pm$ 0.0963	0.9348 $\pm$ 0.1085	0.9532 $\pm$ 0.1035	0.1016 $\pm$ 0.1825	0.1343 $\pm$ 0.2467	0.1524 $\pm$ 0.3281
ACA (L)	0.8539 $\pm$ 0.1431	0.8718 $\pm$ 0.1342	0.7934 $\pm$ 0.1670	0.9068 $\pm$ 0.1406	0.9159 $\pm$ 0.1285	0.8769 $\pm$ 0.1705	0.1915 $\pm$ 0.2687	0.1691 $\pm$ 0.2683	0.3654 $\pm$ 0.9038
AComm	0.4951 $\pm$ 0.3619	0.5346 $\pm$ 0.3481	0.2729 $\pm$ 0.3298	0.5365 $\pm$ 0.4651	0.5755 $\pm$ 0.4547	0.3350 $\pm$ 0.4421	0.3485 $\pm$ 0.5930	0.3748 $\pm$ 0.5820	0.4066 $\pm$ 0.3597
BA	0.7027 $\pm$ 0.1116	0.7004 $\pm$ 0.0985	0.5834 $\pm$ 0.1715	0.8833 $\pm$ 0.1371	0.8806 $\pm$ 0.1384	0.7965 $\pm$ 0.2459	0.5042 $\pm$ 0.3781	0.5103 $\pm$ 0.3424	0.8704 $\pm$ 1.5262
PCA (R)	0.9274 $\pm$ 0.0859	0.9460 $\pm$ 0.0848	0.9064 $\pm$ 0.1024	0.9720 $\pm$ 0.0852	0.9774 $\pm$ 0.0807	0.9641 $\pm$ 0.1066	0.0729 $\pm$ 0.1738	0.0576 $\pm$ 0.1721	0.1248 $\pm$ 0.3163
PCA (L)	0.9359 $\pm$ 0.0623	0.9499 $\pm$ 0.0521	0.9025 $\pm$ 0.1082	0.9834 $\pm$ 0.0486	0.9834 $\pm$ 0.0431	0.9603 $\pm$ 0.1099	0.0580 $\pm$ 0.0912	0.0451 $\pm$ 0.0749	0.1489 $\pm$ 0.4548
PComm (R)	0.6771 $\pm$ 0.3220	0.7432 $\pm$ 0.2812	0.3084 $\pm$ 0.3515	0.7321 $\pm$ 0.3544	0.8093 $\pm$ 0.2960	0.3529 $\pm$ 0.4012	0.3435 $\pm$ 0.6273	0.3198 $\pm$ 0.6174	0.7525 $\pm$ 0.8806
PComm (L)	0.6632 $\pm$ 0.2779	0.6564 $\pm$ 0.3361	0.3491 $\pm$ 0.3409	0.7679 $\pm$ 0.2991	0.7373 $\pm$ 0.3626	0.4075 $\pm$ 0.3990	0.3936 $\pm$ 0.5131	0.4670 $\pm$ 0.9512	0.9355 $\pm$ 1.1715
Other artery	0.0000 $\pm$ 0.0000	0.0180 $\pm$ 0.0675	—	0.0000 $\pm$ 0.0000	0.0200 $\pm$ 0.0748	—	16.9444 $\pm$ 4.0454	15.4374 $\pm$ 5.3517	—

Table A5: Per-label segmentation results for the model trained with Skeleton Recall Loss on the TopCoW dataset (mean $\pm$ std).

Label	mDC $\uparrow$		tSens $\uparrow$		adHD $\downarrow$	
	All labels	All arteries	All labels	All arteries	All labels	All arteries
ICA (R)	0.7634 $\pm$ 0.0711	0.7488 $\pm$ 0.0721	0.8646 $\pm$ 0.0847	0.8662 $\pm$ 0.0858	0.4279 $\pm$ 0.1922	0.4596 $\pm$ 0.1893
ICA (L)	0.7770 $\pm$ 0.0648	0.7600 $\pm$ 0.0716	0.8689 $\pm$ 0.0939	0.8624 $\pm$ 0.1016	0.3807 $\pm$ 0.1606	0.4095 $\pm$ 0.1746
M1 (R)	0.8238 $\pm$ 0.1600	0.8292 $\pm$ 0.1560	0.8258 $\pm$ 0.2048	0.8384 $\pm$ 0.1998	0.6030 $\pm$ 0.9330	0.5573 $\pm$ 0.8716
M1 (L)	0.8442 $\pm$ 0.1561	0.8497 $\pm$ 0.1557	0.8558 $\pm$ 0.1955	0.8649 $\pm$ 0.1951	0.4811 $\pm$ 0.6766	0.4669 $\pm$ 0.7006
ACA (R)	0.8965 $\pm$ 0.1155	0.8867 $\pm$ 0.1211	0.9354 $\pm$ 0.1148	0.9314 $\pm$ 0.1184	0.1191 $\pm$ 0.1967	0.1401 $\pm$ 0.2409
ACA (L)	0.8536 $\pm$ 0.1273	0.8615 $\pm$ 0.1222	0.8951 $\pm$ 0.1265	0.9029 $\pm$ 0.1186	0.1729 $\pm$ 0.2370	0.1575 $\pm$ 0.2143
AComm	0.6460 $\pm$ 0.3190	0.6044 $\pm$ 0.3317	0.6890 $\pm$ 0.4303	0.6488 $\pm$ 0.4463	0.2728 $\pm$ 0.4477	0.2860 $\pm$ 0.4698
BA	0.7039 $\pm$ 0.1041	0.6854 $\pm$ 0.1059	0.8696 $\pm$ 0.1412	0.8631 $\pm$ 0.1369	0.4928 $\pm$ 0.2589	0.5285 $\pm$ 0.2666
PCA (R)	0.9411 $\pm$ 0.0801	0.9288 $\pm$ 0.0937	0.9733 $\pm$ 0.0822	0.9696 $\pm$ 0.0956	0.0613 $\pm$ 0.1462	0.0723 $\pm$ 0.1722
PCA (L)	0.9489 $\pm$ 0.0626	0.9478 $\pm$ 0.0544	0.9776 $\pm$ 0.0550	0.9842 $\pm$ 0.0458	0.0509 $\pm$ 0.1027	0.0496 $\pm$ 0.0966
PComm (R)	0.6623 $\pm$ 0.3184	0.6180 $\pm$ 0.3382	0.7404 $\pm$ 0.3450	0.6991 $\pm$ 0.3740	0.2589 $\pm$ 0.3190	0.3990 $\pm$ 0.6599
PComm (L)	0.6776 $\pm$ 0.2994	0.6670 $\pm$ 0.2957	0.7607 $\pm$ 0.3249	0.7608 $\pm$ 0.3206	0.4115 $\pm$ 0.6359	0.4382 $\pm$ 0.6836
Other artery	0.0000 $\pm$ 0.0000	0.0000 $\pm$ 0.0000	0.0000 $\pm$ 0.0000	0.0000 $\pm$ 0.0000	14.3025 $\pm$ 4.7048	14.0284 $\pm$ 3.1399

Table A6: Aggregate segmentation performance of multiphase and arterial phase only models trained with the Skeleton Recall Loss on the TopBrain and TopCoW datasets (mean $\pm$ std).

Dataset	Configuration	Evaluated vessel	mDC $\uparrow$	tSens $\uparrow$	adHD $\downarrow$
TopBrain	All labels	All vessel	0.9381 $\pm$ 0.0217	0.9740 $\pm$ 0.0120	0.1008 $\pm$ 0.0511
		Artery	0.9025 $\pm$ 0.0373	0.9477 $\pm$ 0.0228	0.1945 $\pm$ 0.1176
		Vein	0.9422 $\pm$ 0.0334	0.9724 $\pm$ 0.0253	0.0758 $\pm$ 0.0435
	All labels (arterial phase only)	All vessel	0.9292 $\pm$ 0.0232	0.9532 $\pm$ 0.0119	0.1261 $\pm$ 0.0422
		Artery	0.8616 $\pm$ 0.0489	0.9310 $\pm$ 0.0229	0.3380 $\pm$ 0.4021
		Vein	0.9727 $\pm$ 0.0338	0.9527 $\pm$ 0.0497	0.0568 $\pm$ 0.0624
TopCoW	All labels	Artery	0.9638 $\pm$ 0.0273	0.9909 $\pm$ 0.0153	0.0334 $\pm$ 0.0288
	All labels (arterial phase only)	Artery	0.9472 $\pm$ 0.0415	0.9760 $\pm$ 0.0365	0.0710 $\pm$ 0.1431

Table A7: Aggregate segmentation performance of three models evaluated on 5 held-out TopBrain cases (mean $\pm$ std). All models use Skeleton Recall Loss.

Training set	Evaluated vessel	mDC $\uparrow$	tSens $\uparrow$	adHD $\downarrow$
SemanticVessel (360 scans, multiphase)	All vessel	0.9545 $\pm$ 0.0049	0.9792 $\pm$ 0.0031	0.0754 $\pm$ 0.0186
	Artery	0.8966 $\pm$ 0.0553	0.9532 $\pm$ 0.0121	0.2036 $\pm$ 0.1431
	Vein	0.9659 $\pm$ 0.0039	0.9740 $\pm$ 0.0270	0.0489 $\pm$ 0.0144
SemanticVessel (41 scans, arterial phase only)	All vessel	0.9402 $\pm$ 0.0185	0.9485 $\pm$ 0.0129	0.1091 $\pm$ 0.0406
	Artery	0.8353 $\pm$ 0.0809	0.9197 $\pm$ 0.0289	0.6363 $\pm$ 0.8027
	Vein	0.9895 $\pm$ 0.0076	0.9516 $\pm$ 0.0406	0.0362 $\pm$ 0.0384
TopBrain (20 scans)	All vessel	0.8337 $\pm$ 0.0429	0.8292 $\pm$ 0.0325	0.3852 $\pm$ 0.1179
	Artery	0.7804 $\pm$ 0.0449	0.7980 $\pm$ 0.0396	0.5103 $\pm$ 0.1596
	Vein	0.8891 $\pm$ 0.0672	0.8734 $\pm$ 0.0676	0.3044 $\pm$ 0.2753

Table A8: Aggregate segmentation performance of three models evaluated on 5 held-out SemanticVessel arterial phase cases (mean $\pm$ std). All models use Skeleton Recall Loss.

Training set	Evaluated vessel	mDC $\uparrow$	tSens $\uparrow$	adHD $\downarrow$
SemanticVessel (360 scans, multiphase)	All vessel	0.7555 $\pm$ 0.1336	0.8158 $\pm$ 0.0784	0.4881 $\pm$ 0.2890
	Artery	0.8599 $\pm$ 0.0489	0.9166 $\pm$ 0.0161	0.2398 $\pm$ 0.0804
	Vein	0.7185 $\pm$ 0.1676	0.7426 $\pm$ 0.1235	0.6725 $\pm$ 0.4369
SemanticVessel (41 scans, arterial phase only)	All vessel	0.6550 $\pm$ 0.1676	0.6805 $\pm$ 0.1302	1.0330 $\pm$ 0.5532
	Artery	0.7427 $\pm$ 0.1735	0.7872 $\pm$ 0.1347	0.8806 $\pm$ 0.6583
	Vein	0.6023 $\pm$ 0.1767	0.5156 $\pm$ 0.1250	1.7257 $\pm$ 0.9209
TopBrain (20 scans)	All vessel	0.1197 $\pm$ 0.0440	0.1555 $\pm$ 0.0365	12.8017 $\pm$ 1.0111
	Artery	0.2447 $\pm$ 0.0595	0.3111 $\pm$ 0.0690	10.2010 $\pm$ 1.4143
	Vein	0.0660 $\pm$ 0.0305	0.0660 $\pm$ 0.0226	22.7483 $\pm$ 2.5208

Table A9: Per-label segmentation results of three models evaluated on 5 held-out SemanticVessel arterial phase cases (mean $\pm$ std). All models use Skeleton Recall Loss. Multi = SemanticVessel multiphase (360 scans); Single = SemanticVessel arterial phase only (41 scans); TopBrain = TopBrain trained (20 scans).

Label	mDC $\uparrow$			tSens $\uparrow$			adHD $\downarrow$		
	Multi	Single	TopBrain	Multi	Single	TopBrain	Multi	Single	TopBrain
ICA (R)	0.9611 $\pm$ 0.0294	0.8905 $\pm$ 0.1267	0.0806 $\pm$ 0.0203	1.0000 $\pm$ 0.0000	0.9842 $\pm$ 0.0265	0.1157 $\pm$ 0.0254	0.0715 $\pm$ 0.0514	0.2020 $\pm$ 0.2374	23.8944 $\pm$ 1.5023
ICA (L)	0.9646 $\pm$ 0.0275	0.8856 $\pm$ 0.1579	0.0930 $\pm$ 0.0193	1.0000 $\pm$ 0.0000	0.9565 $\pm$ 0.0870	0.1404 $\pm$ 0.0255	0.0623 $\pm$ 0.0488	0.2360 $\pm$ 0.3508	22.8023 $\pm$ 1.2483
M1 (R)	0.9004 $\pm$ 0.0343	0.8546 $\pm$ 0.0805	0.5299 $\pm$ 0.1626	0.9677 $\pm$ 0.0408	0.8956 $\pm$ 0.1127	0.7917 $\pm$ 0.1642	0.1442 $\pm$ 0.0898	0.2352 $\pm$ 0.1304	0.9265 $\pm$ 0.5380
M1 (L)	0.8763 $\pm$ 0.1171	0.9205 $\pm$ 0.0804	0.7054 $\pm$ 0.1326	0.9789 $\pm$ 0.0421	1.0000 $\pm$ 0.0000	0.9339 $\pm$ 0.0770	0.2166 $\pm$ 0.2898	0.1246 $\pm$ 0.1387	0.4916 $\pm$ 0.3388
M2+ (R)	0.8886 $\pm$ 0.0927	0.7516 $\pm$ 0.2398	0.2209 $\pm$ 0.0960	0.9733 $\pm$ 0.0298	0.8120 $\pm$ 0.1756	0.3170 $\pm$ 0.0976	0.1294 $\pm$ 0.1167	1.0920 $\pm$ 1.2461	7.7854 $\pm$ 2.1956
M2+ (L)	0.8966 $\pm$ 0.0832	0.7328 $\pm$ 0.2221	0.3166 $\pm$ 0.1200	0.9807 $\pm$ 0.0197	0.7899 $\pm$ 0.1795	0.4210 $\pm$ 0.1353	0.1064 $\pm$ 0.0922	1.1561 $\pm$ 1.3066	6.6268 $\pm$ 2.6073
ACA (R)	0.8939 $\pm$ 0.0478	0.7330 $\pm$ 0.1885	0.3645 $\pm$ 0.1131	0.9516 $\pm$ 0.0385	0.7840 $\pm$ 0.1610	0.3988 $\pm$ 0.1650	0.1273 $\pm$ 0.0667	0.7421 $\pm$ 0.8062	10.9800 $\pm$ 6.0148
ACA (L)	0.8117 $\pm$ 0.0239	0.7001 $\pm$ 0.1547	0.2541 $\pm$ 0.1537	0.8880 $\pm$ 0.0504	0.7770 $\pm$ 0.1151	0.2659 $\pm$ 0.1640	0.2782 $\pm$ 0.0936	0.6593 $\pm$ 0.5028	12.2836 $\pm$ 6.6502
AComm	0.5266 $\pm$ 0.4390	0.5620 $\pm$ 0.3562	0.0000 $\pm$ 0.0000	0.5714 $\pm$ 0.4695	0.6737 $\pm$ 0.4163	0.0000 $\pm$ 0.0000	0.0687 $\pm$ 0.0642	0.3258 $\pm$ 0.2847	—
VA (R)	0.9385 $\pm$ 0.0379	0.6345 $\pm$ 0.3716	0.4997 $\pm$ 0.3024	0.9875 $\pm$ 0.0125	0.7321 $\pm$ 0.4237	0.6209 $\pm$ 0.3736	0.0720 $\pm$ 0.0465	0.2253 $\pm$ 0.1405	0.9674 $\pm$ 0.6131
VA (L)	0.9327 $\pm$ 0.0431	0.6933 $\pm$ 0.2728	0.7704 $\pm$ 0.1274	0.9649 $\pm$ 0.0608	0.8184 $\pm$ 0.1928	0.8781 $\pm$ 0.0956	0.0975 $\pm$ 0.0741	0.7897 $\pm$ 0.9585	0.5998 $\pm$ 0.4842
BA	0.6888 $\pm$ 0.3458	0.5951 $\pm$ 0.3800	0.5915 $\pm$ 0.4750	0.7562 $\pm$ 0.3804	0.7226 $\pm$ 0.3839	0.5939 $\pm$ 0.4795	3.9406 $\pm$ 7.3457	4.5054 $\pm$ 7.3539	4.4118 $\pm$ 7.5983
PCA (R)	0.8397 $\pm$ 0.1303	0.7572 $\pm$ 0.2432	0.3783 $\pm$ 0.1619	0.9428 $\pm$ 0.0643	0.8475 $\pm$ 0.1915	0.4502 $\pm$ 0.1476	2.4845 $\pm$ 4.7551	3.1757 $\pm$ 4.5586	6.8951 $\pm$ 4.4078
PCA (L)	0.8457 $\pm$ 0.1063	0.7617 $\pm$ 0.1699	0.4458 $\pm$ 0.1642	0.9537 $\pm$ 0.0662	0.8463 $\pm$ 0.1067	0.5636 $\pm$ 0.0894	0.8524 $\pm$ 1.4972	1.1147 $\pm$ 1.3122	3.2699 $\pm$ 1.3291
PComm (R)	0.8529 $\pm$ 0.1524	0.7019 $\pm$ 0.1988	0.0000 $\pm$ 0.0000	0.8463 $\pm$ 0.1955	0.7581 $\pm$ 0.1811	0.0000 $\pm$ 0.0000	0.2131 $\pm$ 0.2726	0.3092 $\pm$ 0.2201	—
PComm (L)	0.9247 $\pm$ 0.0277	0.6797 $\pm$ 0.2042	0.0000 $\pm$ 0.0000	1.0000 $\pm$ 0.0000	0.9231 $\pm$ 0.0769	0.0000 $\pm$ 0.0000	0.0735 $\pm$ 0.0307	0.3169 $\pm$ 0.1855	—
Other artery	0.6713 $\pm$ 0.1411	0.4242 $\pm$ 0.2263	0.0753 $\pm$ 0.0438	0.7694 $\pm$ 0.0503	0.4634 $\pm$ 0.1963	0.0999 $\pm$ 0.0464	0.6596 $\pm$ 0.2351	2.8469 $\pm$ 1.4715	28.7639 $\pm$ 3.6114
Vein	0.7185 $\pm$ 0.1676	0.6023 $\pm$ 0.1767	0.0660 $\pm$ 0.0305	0.7426 $\pm$ 0.1235	0.5156 $\pm$ 0.1250	0.0660 $\pm$ 0.0226	0.6725 $\pm$ 0.4369	1.7257 $\pm$ 0.9209	22.7483 $\pm$ 2.5208

Table A10: Per-label segmentation results of three models evaluated on 5 held-out TopBrain cases (mean $\pm$ std). All models use Skeleton Recall Loss. Multi = SemanticVessel multiphase (360 scans); Single = SemanticVessel arterial phase only (41 scans); TopBrain = TopBrain trained (20 scans).

Label	mDC $\uparrow$			tSens $\uparrow$			adHD $\downarrow$		
	Multi	Single	TopBrain	Multi	Single	TopBrain	Multi	Single	TopBrain
ICA (R)	0.7670 $\pm$ 0.1069	0.6717 $\pm$ 0.0369	0.8355 $\pm$ 0.0813	0.9165 $\pm$ 0.0526	0.8645 $\pm$ 0.0556	0.9778 $\pm$ 0.0444	0.3959 $\pm$ 0.2262	0.5543 $\pm$ 0.1691	0.1989 $\pm$ 0.1194
ICA (L)	0.7543 $\pm$ 0.0557	0.7232 $\pm$ 0.0610	0.7981 $\pm$ 0.0630	0.8563 $\pm$ 0.1190	0.8245 $\pm$ 0.1143	0.9684 $\pm$ 0.0632	0.4499 $\pm$ 0.1643	0.4902 $\pm$ 0.1712	0.2877 $\pm$ 0.1504
M1 (R)	0.6354 $\pm$ 0.1527	0.6787 $\pm$ 0.1053	0.8455 $\pm$ 0.0734	0.5652 $\pm$ 0.2299	0.6694 $\pm$ 0.1842	0.9459 $\pm$ 0.0663	1.2723 $\pm$ 0.7771	0.9311 $\pm$ 0.4969	0.2591 $\pm$ 0.2354
M1 (L)	0.6351 $\pm$ 0.0883	0.7667 $\pm$ 0.0890	0.7979 $\pm$ 0.0644	0.4912 $\pm$ 0.1516	0.5856 $\pm$ 0.1214	0.7426 $\pm$ 0.1760	1.5722 $\pm$ 0.7867	0.7914 $\pm$ 0.3078	0.4586 $\pm$ 0.3618
M2+ (R)	0.9769 $\pm$ 0.0115	0.9244 $\pm$ 0.0449	0.6817 $\pm$ 0.1071	0.9853 $\pm$ 0.0100	0.8990 $\pm$ 0.0522	0.7043 $\pm$ 0.1121	0.0192 $\pm$ 0.0126	0.2005 $\pm$ 0.1091	0.7536 $\pm$ 0.2661
M2+ (L)	0.9830 $\pm$ 0.0055	0.9037 $\pm$ 0.0318	0.6793 $\pm$ 0.0483	0.9976 $\pm$ 0.0033	0.8943 $\pm$ 0.0323	0.7463 $\pm$ 0.0612	0.0106 $\pm$ 0.0043	0.3572 $\pm$ 0.1561	0.9618 $\pm$ 0.4028
ACA (R)	0.9025 $\pm$ 0.0553	0.8395 $\pm$ 0.1477	0.8132 $\pm$ 0.0641	0.9735 $\pm$ 0.0403	0.8858 $\pm$ 0.1378	0.8946 $\pm$ 0.0911	0.0774 $\pm$ 0.0469	0.1994 $\pm$ 0.2148	0.3378 $\pm$ 0.2517
ACA (L)	0.8006 $\pm$ 0.1046	0.8915 $\pm$ 0.0572	0.7618 $\pm$ 0.0953	0.8499 $\pm$ 0.1824	0.9229 $\pm$ 0.0520	0.8404 $\pm$ 0.1356	0.4261 $\pm$ 0.5664	0.1015 $\pm$ 0.0613	0.3511 $\pm$ 0.2236
AComm	0.7354 $\pm$ 0.0830	0.4755 $\pm$ 0.4756	0.0000 $\pm$ 0.0000	0.7500 $\pm$ 0.4330	0.5009 $\pm$ 0.5000	0.0000 $\pm$ 0.0000	0.2638 $\pm$ 0.1137	0.0319 $\pm$ 0.0085	—
VA (R)	0.7561 $\pm$ 0.1817	0.5610 $\pm$ 0.2515	0.7600 $\pm$ 0.0755	0.9197 $\pm$ 0.1125	0.7273 $\pm$ 0.2577	0.8886 $\pm$ 0.0662	0.5446 $\pm$ 0.6066	3.5044 $\pm$ 5.0344	0.3568 $\pm$ 0.1153
VA (L)	0.8566 $\pm$ 0.1024	0.6506 $\pm$ 0.2604	0.7349 $\pm$ 0.2192	0.9832 $\pm$ 0.0283	0.7659 $\pm$ 0.3125	0.8711 $\pm$ 0.2106	0.1519 $\pm$ 0.1175	3.6797 $\pm$ 6.4809	0.4349 $\pm$ 0.4903
BA	0.7959 $\pm$ 0.0852	0.8172 $\pm$ 0.0470	0.8907 $\pm$ 0.0448	0.9545 $\pm$ 0.0594	0.9683 $\pm$ 0.0308	1.0000 $\pm$ 0.0000	0.3014 $\pm$ 0.1633	0.2673 $\pm$ 0.0876	0.1344 $\pm$ 0.0625
PCA (R)	0.9122 $\pm$ 0.0371	0.9354 $\pm$ 0.0209	0.7515 $\pm$ 0.1693	0.9282 $\pm$ 0.0426	0.9425 $\pm$ 0.0444	0.8399 $\pm$ 0.1326	0.1288 $\pm$ 0.0911	0.1186 $\pm$ 0.0798	0.4676 $\pm$ 0.3581
PCA (L)	0.8952 $\pm$ 0.0407	0.9350 $\pm$ 0.0343	0.7323 $\pm$ 0.0511	0.9118 $\pm$ 0.0480	0.9517 $\pm$ 0.0659	0.8206 $\pm$ 0.1028	0.2070 $\pm$ 0.1148	0.1159 $\pm$ 0.1282	0.5586 $\pm$ 0.4773
PComm (R)	0.4914 $\pm$ 0.4914	0.6372 $\pm$ 0.3456	0.0000 $\pm$ 0.0000	0.5000 $\pm$ 0.5000	0.6522 $\pm$ 0.3478	0.0000 $\pm$ 0.0000	0.0113 $\pm$ 0.0000	0.5444 $\pm$ 0.5319	—
PComm (L)	0.8239 $\pm$ 0.0000	0.0000 $\pm$ 0.0000	0.0000 $\pm$ 0.0000	0.9500 $\pm$ 0.0000	0.0000 $\pm$ 0.0000	0.0000 $\pm$ 0.0000	0.1125 $\pm$ 0.0000	—	—
Other artery	0.6452 $\pm$ 0.2243	0.7257 $\pm$ 0.0588	0.6139 $\pm$ 0.1976	0.6541 $\pm$ 0.1802	0.7283 $\pm$ 0.0459	0.6729 $\pm$ 0.1698	2.1274 $\pm$ 2.5119	0.9189 $\pm$ 0.2169	1.1479 $\pm$ 0.6128
Vein	0.9659 $\pm$ 0.0039	0.9895 $\pm$ 0.0076	0.8891 $\pm$ 0.0672	0.9740 $\pm$ 0.0270	0.9516 $\pm$ 0.0406	0.8734 $\pm$ 0.0676	0.0489 $\pm$ 0.0144	0.0362 $\pm$ 0.0384	0.3044 $\pm$ 0.2753