NEJM AI

NEJM AI 2026;3(8)
DOI: 10.1056/AIdbp2500994

DATASETS, BENCHMARKS, AND PROTOCOLS

The TopCoW Challenge — Topology-Aware Circle of Willis Segmentation for CT and MR Angiography

K. Yang,[1] F. Musio,[1,2] Y. Ma,[3,4] N. Juchler,[2] J. C. Paetzold,[5] R. Al-Maskari,[6,7] L. Höher,[6] H. B. Li,[1,8] I. E. Hamamci,[1] A. Sekuboyina,[1] S. Shit,[1] H. Huang,[1] C. Prabhakar,[1] E. de la Rosa,[1] B. Wittmann,[1] D. Waldmannstetter,[1,7] F. Kofler,[1,7,9,10] F. Navarro,[1,7,9] M. J. Menten,[7,11,12] I. Ezhov,[7] D. Rueckert,[7,9,11,12] I. N. Vos,[13] Y. M. Ruigrok,[14] B. K. Velthuis,[15] H. J. Kuijf,[13] P. Shi,[16] W. Liu,[16] T. Ma,[17] M. R. Rokuss,[18,19] Y. Kirchhoff,[18,19,20] F. Isensee,[18,21] K. Maier-Hein,[18,22] C. Zhu,[23] H. Zhao,[24] P. Bijlenga,[25] J. Hämmerli,[25] C. Wurster,[25,*], L. Westphal,[26] J. Bisschop,[27] E. Colombo,[28] H. Baazaoui,[26] H.-L. Handelsmann,[26] A. Makmur,[29] J. Hallinan,[29] A. Soundararajan,[30] B. Wiestler,[9] J. S. Kirschke,[9] E. O. Riedel,[9] R. Wiest,[31] E. Montagnon,[32] L. Letourneau-Guillon,[32] K. Oh,[33,34] D. Lee,[33] O. U. Aydin,[35] A. Hilbert,[35] J. Rieger,[35] D. Rallios,[35] S. Tanioka,[35] A. Koch,[35] D. Frey,[35] A. Qayyum,[36] M. Mazher,[37] S. Niederer,[36] N. Disch,[18,19,20] J. C. Holzschuh,[18] D. LaBella,[38] F. Galati,[39] D. Falcetta,[39] M. A. Zuluaga,[39] C. Lin,[40] H. Zhao,[40] Z. Zhang,[41] M. Zhang,[42,43] X. You,[42,43] H. Zhang,[42] G.-Z. Yang,[42] Y. Gu,[42,43] S. Ra,[44] J. Hwang,[44] H. Park,[45] J. Chen,[46] M. Wodzinski,[47,48] H. Müller,[47] N. Mansouri,[49,50] F. Autrusseau,[49,50] C. Yalcin,[51] R. E. Hamadache,[51] C. Lisazo,[51] J. Salvi,[51] A. Casamitjana,[51] X. Lladó,[51] U. M. L.-T. Estrada,[51] V. Abramova,[51] L. Giancardo,[52] A. Oliver,[51] P. Casademunt,[53] A. Galdran,[53] M. Delucchi,[2,54] O. Camara,[53] J. Liu,[55,56] H. Huang,[55,56] Y. Cui,[55,56] Z. Lin,[57] Y. Liu,[43] S. Zhu,[57] T. R. Patel,[58,59] A. H. Siddiqui,[58,59] V. M. Tutino,[58,60] M. Orouskhani,[23] H. Wang,[23] M. Mossa-Basha,[23] Y. Sato,[61] S. Hirsch,[2] S. Wegener,[26] and B. Menze,[1]

Received: September 1, 2025; Revised: May 17, 2026; Accepted: May 27, 2026; Published: July 23, 2026

Abstract

BACKGROUND The circle of Willis (CoW) is an important network of arteries connecting major circulations of the brain. Its vascular architecture is believed to influence the risk, severity, and outcome of serious neurovascular diseases. However, characterizing the highly variable CoW anatomy remains a manual and time-consuming expert task. The CoW is commonly imaged by two noninvasive angiographic imaging modalities — magnetic resonance angiography (MRA) and computed tomography angiography (CTA) — yet few datasets with annotated CoW anatomy exist, and there have been no established benchmarks for comparing CoW segmentation algorithms.

METHODS We organized the Topology-Aware Anatomical Segmentation of the Circle of Willis for Computed Tomography Angiography and Magnetic Resonance Angiography (TopCoW) benchmark challenge alongside the release of an annotated CoW dataset with 125 paired MRA and CTA scans from the same patients. Voxel-level annotations for 13 vessel components were created using virtual reality technology and verified by clinical experts. Participants submitted algorithms for CoW segmentation and variant classification, which we evaluated on internal and external test sets comprising 226 scans from over five centers. The benchmark includes voxel-level segmentation, CoW component detection, CoW variant classification, and two clinical application tasks.

Kaiyuan Yang, Fabio Musio, and Yihui Ma contributed equally to this article.

Sven Hirsch, Susanne Wegener, and Bjoern Menze contributed equally to this article.

The authors' full names and academic degrees are listed in the Appendix.

** Deceased.*

The author affiliations are listed at the end of the article.

Bjoern Menze can be contacted at bjoern.menze@uzh.ch or at the Department of Quantitative Biomedicine, University of Zurich, Plattenstrasse 10, 8032 Zürich, Switzerland; Susanne Wegener can be contacted at susanne.wegener@usz.ch or at the Department of Neurology, University Hospital of Zurich, Frauenklinikstrasse 26, 8091 Zürich, Switzerland; and Sven Hirsch can be contacted at sven.hirsch@zhaw.ch or at the Institute of Computational Life Sciences, Zurich University of Applied Sciences, Schloss 4, 8820 Waedenswil, Switzerland.

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

RESULTS We received submissions from over 250 participants across six continents. Top-performing teams achieved over 90% Dice similarity coefficient scores for CoW segmentation, over 80% F1 scores for detecting key vessel components, and over 70% balanced accuracy in CoW variant classification across nearly all test sets. The best algorithms also supported clinically relevant downstream tasks by accurately classifying fetal-type posterior cerebral arteries and localizing aneurysms in relation to CoW anatomy. Notably, modality-agnostic models and topological optimization emerged as key factors in improving segmentation performance.

CONCLUSIONS This benchmark demonstrated the utility of CoW segmentation algorithms for some downstream clinical applications with explainability. The annotated datasets and top-performing algorithms are published alongside this benchmark to foster methodological advances and clinical tool development. (Funded by the Digitalization Initiative of the Zurich Higher Education Institutions and the Helmut Horten Foundation.)

Introduction

The circle of Willis (CoW) is an important anastomotic network of arteries connecting the anterior and posterior circulations of the brain, as well as the left and right cerebral hemispheres.[1] Owing to its centrality, the CoW is commonly involved in pathologies such as aneurysms and stroke. Clinically, the vascular architecture of the CoW is believed to impact the occurrence and severity of stroke,[2-5] pose a potential risk for aneurysm formation,[6,7] and affect the neurologic events and clinical outcomes of neurosurgeries.[8,9] Accurate CoW characterization is therefore of great clinical relevance.

Clinicians have articulated an unmet demand for efficient software tools to analyze the angio-architecture of the CoW. Assessing the anatomy of the CoW from angiographic images remains a manual, time-consuming task requiring specialist judgment. The CoW anatomy involves vessels that vary in diameter from around 1–4 mm.[10] The vessels are often tortuous and difficult to identify in isolation; anatomical identification frequently depends on their spatial relationships. Furthermore, the CoW exhibits numerous natural variants, with key arteries being frequently hypoplastic or absent.[10,11] Such complexity and heterogeneity of the anatomy make characterizing CoW vasculature challenging.

The CoW vessels are commonly diagnosed by two noninvasive angiographic imaging modalities, namely magnetic resonance angiography (MRA) and computed tomography angiography (CTA). Although there have been recent releases of public MRA datasets,[12-16] vessel annotations were binary with no anatomical labeling of the CoW. Annotated datasets on the other important modality, CTA, did not exist. Prior work on the CoW characterization was developed mainly as a labeling task built upon binary vessel masks, skeletons, or graphs,[17-22] with two recent studies tackling it as a multiclass segmentation task.[23,24] However, only private CoW annotations were used, and the modality was restricted to MRA only. There were no benchmarks to compare CoW segmentation algorithms. Furthermore, past studies have not sufficiently explored the difficulties associated with the complex and variable CoW anatomies in clinical settings, leaving their clinical relevance uncertain.

To this end, we organized a benchmark on Topology-Aware Anatomical Segmentation of the Circle of Willis for Computed Tomography Angiography and Magnetic Resonance Angiography (TopCoW) in 2023 and 2024. TopCoW was the first public challenge on CoW anatomical segmentation featuring a paired dataset with voxel-level vessel annotations on both MRA and CTA. The main tasks benchmarked were CoW segmentation and CoW variant classification. Algorithms were crowdsourced from global participants and evaluated on both internal and external multicenter test datasets. CoW segmentation algorithms were additionally evaluated on two clinical downstream tasks: classifying fetal-type posterior cerebral artery (PCA) and locating intracranial aneurysms using CoW segmentation. For transparency and continued development, we publicly released our annotated datasets and top-performing algorithms on Zenodo (see links in data and code availability section at the end of the article).

Methods

DATASETS

The TopCoW challenge cohort was composed of patients admitted to the Stroke Center of University Hospital Zurich in 2018 and 2019 with suspicion of stroke-related neurologic conditions. The inclusion criteria for the TopCoW data were (1) both MRA and CTA scans were available and of good quality for the patient; (2) at least one of the MRA or CTA images allowed for an assessment of the CoW anatomy; (3) no large aneurysms were identified inside the CoW region of interest (ROI); and (4) certain rare CoW variants that could not be characterized by our 13 annotated CoW components were excluded. All TopCoW data were anonymized, defaced, reoriented, and cropped to the braincase

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

region. More information on the TopCoW dataset can be found in Sections S1–S4 in the Supplementary Appendix.

In addition to the internal test sets from the TopCoW dataset, we gathered and annotated 86 MRA and CTA scans from four external multicenter test datasets to evaluate the robustness of the algorithms. The external test datasets were sourced from existing public datasets and annotated in the same way as the TopCoW dataset. Two external CTA test sets were from the Ischemic Stroke Lesion Segmentation Challenge 2024 (ISLES'24) training set[25,26] and the 2021 Large Intracranial Aneurysm Segmentation (LargeIA) dataset.[27,28] Two external MRA test sets were from the 2022 Lausanne OpenNeuro aneurysm dataset[29,30] and the Information Extraction from Images Hammersmith Hospital dataset (IXI-HH).[13] The external data comprised scans from multiple manufacturers. The statistical summary of the image information for the training, internal test, and external test data is shown in Section S5.

DATA ANNOTATION

For each 3D angiography image, we provided three types of annotations regarding the CoW: the voxel-level multiclass segmentation mask of the CoW, a 3D bounding box for the CoW ROI, and the CoW variant graph. Virtual reality (VR) was used to efficiently annotate and verify CoW anatomy in 3D. Figure 1A shows the workflow and view from VR. VR annotations were performed using syGlass,[31] as previously described.[32] Figure 1A, far right panel, shows an MRA example with all 13 CoW vessel classes of multiclass segmentation annotation: left and right internal carotid artery (ICA), left and right anterior cerebral artery (ACA), left and right middle cerebral artery, anterior communicating artery (Acom), left and right posterior communicating artery (Pcom), left and right PCA, and basilar artery (BA). Occasionally, the anterior part of the CoW can have a third A2 artery arising from the Acom, and we labeled it with class third A2. Figure 1B shows an example segmentation mask and ROI annotations for both MRA and CTA from a TopCoW patient. The CoW annotation protocol was developed by a senior neurosurgeon (Y.M., over 10 years of experience) and reviewed by another senior neurosurgeon (P.B., over 15 years of experience) and a senior neurologist (S.W., over 15 years of experience). Two doctoral researchers (K.Y. and F.M.) performed voxel-level annotations after training by Y.M. The annotators each labeled half of the dataset partitioned by patient. Uncertain cases were adjudicated by Y.M., and a subset was resolved through expert consensus involving additional neurosurgeons and neurologists. More information on the annotation protocol is in Section S6.

The CoW variant graph for classification was derived from the segmentation mask and encompasses the anterior variant (AV) and posterior variant (PV) graphs. These variants were defined based on the presence of left A1, Acom, third A2, and right A1 vessels for AV; and left Pcom (L-Pcom), left P1, right P1, and right Pcom (R-Pcom) vessels for PV. Hypoplastic vessels were categorized as present. Figure 1C shows the AV and PV graph composition. Distributions of the CoW variants for AV and PV in all our datasets are shown in Section S8.

To estimate the variability of the annotations used for benchmarking and the upper bounds of algorithmic performance, we analyzed interrater agreement for CoW variant classification and voxel-level annotation. For CoW variant classification, senior neurosurgeon P.B. independently relabeled the AV and PV classes for 40 CTA patients from the TopCoW test set. These cases were selected using stratified sampling across all available CoW variants in the internal test data, prioritizing rare anatomical variants. As shown in Figure 1D, the selected sample included four AV and six PV classes. Labels from P.B. were compared with the annotations used in the benchmark. Balanced accuracies between the raters were 88% for AV and 78% for PV. Cohen's kappa scores were 83% for AV and 72% for PV, suggesting good agreement. Detailed results on the voxel-level segmentation agreement can be found in Section S7.

CHALLENGE STRUCTURE

The featured task of TopCoW across both iterations was multiclass segmentation. The first 2023 challenge had a task of merged binary segmentation, which was deemed sufficiently solved and replaced by CoW ROI detection and CoW variant graph classification tasks in the 2024 iteration. The dataset expanded considerably in 2024 (125 training and 70 internal test patients). Teams were ranked by average metric performance on internal data, with top performers validated on external datasets. Our challenge had two modality tracks, MRA and CTA, for algorithm evaluations. For the internal test set, algorithm input was an MRA and CTA pair, and teams could choose one or both modalities. The input during the external test phase was single-modality, which was compatible with all top submissions. (See Fig. S1 and Sections S9–S11 for details.)

EVALUATION OF SEGMENTATION ALGORITHMS

The submitted algorithms were in the form of isolated Docker containers and were evaluated within the CoW ROI using the following metrics.

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

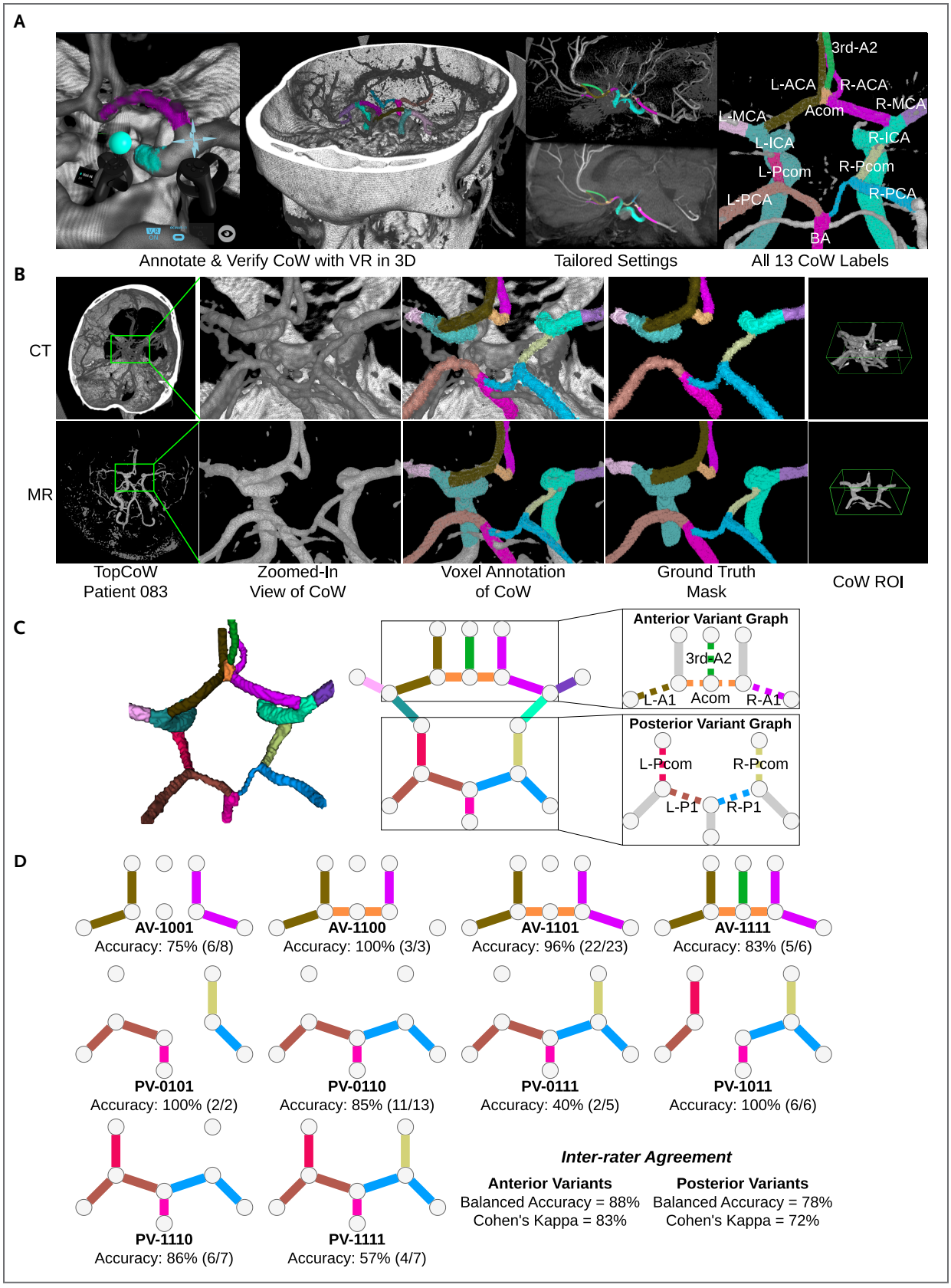


Figure 1. (*Continued*)

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

Voxel-Level Multiclass Segmentation

For voxel-level metrics, the multiclass CoW segmentation predictions were evaluated using Dice similarity coefficient (Dice score),[33] centerline Dice,[34] Hausdorff distance at 95th percentile,[35,36] and connected component error.[37,38]

Beyond Segmentation I: Key CoW Component Detection

The first beyond-segmentation metric was the average F1 score for detection of Acom, Pcom, and third A2. Positive detection was defined as at least 25% intersection over union between the predicted and ground truth masks. This relatively lenient cutoff is consistent with prior studies on small object detection in medical imaging with a similar range (10%–30%).[39,40]

Beyond Segmentation II: CoW Variant Classification

The second beyond-segmentation metric was variant balanced accuracy (VarBalAcc) for CoW variant graph classification. The VarBalAcc was calculated for both AVs and PVs. The variant class was determined using the AV and PV edge list of the variant graph, as shown in Figure 1C. The segmentation mask was converted to an edge list based on presence of the Acom, Pcom, and third A2 labels, and whether the ACA and PCA were connected to their respective neighboring labels — ICA and BA — for the A1 and P1 edges.

Clinical Application I: Fetal PCA Classification

We extracted centerlines and diameters of CoW segmentation masks using a prior workflow[41,42] (see Section S12 for more details). The diameters along the Pcom and P1 segments were used to determine the CoW anatomical variant called the fetal PCA variant.[1] We compared the diameters of the Pcom and P1 at the 25th percentile. The CoW was classified as having a fetal PCA variant if the Pcom was slightly larger in diameter (>1.05× the diameter of P1). Fetal PCA was assessed separately for the left and right sides. The same set of 40 TopCoW CTA cases used in interrater agreement for variant classification was labeled for fetal PCA class by a senior neurosurgeon (P.B). The fetal PCA labels from P.B. were treated as ground truth.

Clinical Application II: Locating Aneurysm

We selected 12 patients with intracranial aneurysms along their CoW vessels from the aforementioned external LargeIA dataset, which included aneurysm ground truth annotations. The aneurysm locations were then labeled by a senior neurosurgeon (Y.M.). CoW segmentation algorithms were applied to the images of the aneurysm patients, and the resulting CoW predictions were overlaid with the provided aneurysm ground truth. The predicted location of the aneurysm was determined by identifying the CoW labels adjacent to or overlapping with the aneurysm mask.

Results

THE TOPCOW DATASET

In total, 200 paired MRA and CTA scans from unique patients were curated and randomly split by patient into 125 training, 5 validation, and 70 test subjects. The cohort was 40% female with a median age of 74 years (interquartile range, 61–82; Table 1). The primary imaging indication was ischemic stroke (61%), followed by stroke or transient ischemic attack (TIA) mimic (15.5%) and TIA (14.0%).

Figure 1. Circle of Willis Annotation Overview.

In Panel A, the left image shows data annotation using virtual reality. The middle image shows tailored settings by adjusting opacity, threshold, and window dynamically for suitable visualization during annotation. The right image shows the 13 anatomical labels for the circle of Willis (CoW) anatomy. In Panel B, the Topology-Aware Anatomical Segmentation of the Circle of Willis for Computed Tomography Angiography and Magnetic Resonance Angiography (TopCoW) dataset provides paired images from two modalities, computed tomography angiography (CTA) and magnetic resonance angiography, for each patient. Voxel annotations of the CoW vessels and a 3D bounding box of the CoW region of interest are labeled for each modality. In Panel C, the CoW segmentation mask is converted into a CoW variant graph annotation. Each CoW can be classified by an anterior variant (AV) and a posterior variant (PV) graph. Both AV and PV are identified by a four-edge graph, with 0 being absent and 1 being present in the edge list. Panel D shows interrater agreement for CoW variant classification on 40 TopCoW CTA test cases. Accuracy for each variant is shown along with the balanced accuracy and Cohen's kappa score. Acom denotes anterior communicating artery; AV, anterior variant; BA, basilar artery; CoW, circle of Willis; CT, computed tomography; L-ACA, left anterior cerebral artery; L-ICA, left internal carotid artery; L-MCA, left middle cerebral artery; L-PCA, left posterior cerebral artery; L-Pcom, left posterior communicating artery; MR, magnetic resonance; PV, posterior variant; R-ACA, right anterior cerebral artery; R-ICA, right internal carotid artery; R-MCA, right middle cerebral artery; ROI, region of interest; R-Pcom, right posterior communicating artery; R-PCA, right posterior cerebral artery; TopCoW, Topology-Aware Anatomical Segmentation of the Circle of Willis for Computed Tomography Angiography and Magnetic Resonance Angiography; and VR, virtual reality.

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

Table 1. Clinical Characteristics of the TopCoW Cohort, Including Age, Sex, and Type of Event.*

Variable	Values (n = 200)
Age, years	
Mean (±SD)	71.27 (14.08)
Median (IQR)	74 (61–82)
Min–max	24–92
Sex — n (%)	
Male	119 (59.5%)
Female	80 (40.0%)
Type of event — n (%)	
Ischemic stroke	122 (61.0%)
Stroke or TIA mimic	31 (15.5%)
Transient ischemic attack	28 (14.0%)
Intracerebral hemorrhage	9 (4.5%)
Retinal infarct	6 (3.0%)
Cerebral sinus vein thrombosis	2 (1.0%)
Amaurosis fugax	2 (1.0%)

*Age statistics exclude one case with an implausible age of 0 years. Sex statistics exclude one case with no recorded sex information. IQR denotes interquartile range; max, maximum; min, minimum; SD, standard deviation; TIA, transient ischemic attack; and TopCoW, Topology-Aware Anatomical Segmentation of the Circle of Willis for Computed Tomography Angiography and Magnetic Resonance Angiography.

WINNING ALGORITHM DESIGN

Twenty-five teams out of over 250 registrants from six continents submitted to the challenge. As shown in Figure 2A, the performance improved for the five teams participating in both years, particularly for the CTA modality. Figure 2B summarizes key design choices of the segmentation algorithms from the top teams in 2024. Most teams utilized a unified architecture with shared parameters to process single-modality inputs from a mixed-modality training pool, making their models modality-agnostic. Only one team used additional independently prepared training data, which included some of the external MRA test images but without our ground truth labels. Over half adopted a two-stage pipeline, such as localization followed by segmentation, which allowed the model to localize segmentation training to a subvolume, reduce false positives in the background, and conserve computational resources for ensembling. All top teams based their architectures on nnUNet.[43] From 2023 to 2024, two methodological shifts emerged: broader adoption of mixed-modality training and topological optimization as shown in Figure 2C.

VOXEL-LEVEL MULTICLASS SEGMENTATION

Figure 3A shows qualitative segmentation results on the TopCoW internal test sets from a top team. Two patients per modality were selected, representing a wide range of CoW variants and class-average Dice scores. The predictions accurately segmented various complex CoW anatomies, capturing vessel curvatures and class boundaries.

Figure 3B shows class-average Dice scores from the top six teams on internal and external test sets. All top teams showed good generalization across the external test sets, with median Dice scores above 80% across modalities and test sets. The IXI-HH performance drop was unique to team IMR owing to their 0–700 intensity truncation for MRA training data preprocessing, which reduced generalization to a subset of IXI-HH images with high mean intensities. Other voxel-level metrics showed similarly strong performance and generalization trends (see Sections S14 and S15).

BEYOND SEGMENTATION I: KEY COW COMPONENT DETECTION

The presence or absence of four key CoW components — the communicating arteries (Acom, R-Pcom, L-Pcom) and the third A2 segment — directly determines many CoW variant types. Accurate detection of these four components is a clinically relevant feature of segmentation algorithms. We assessed detection performance across all test sets in Figure 3C, showing F1 scores from the top six teams of each modality in boxplots. Detection of the communicating arteries was consistent across test sets and modalities, with top teams achieving over 75% F1 scores on most sets.

BEYOND SEGMENTATION II: COW VARIANT CLASSIFICATION

Notably, four teams took part in both the segmentation task and the classification task with segmentation-based and classification-based algorithms respectively. As shown in Figure 3D, segmentation-based methods outperformed classification-based ones by at least a factor of two on the internal test sets.

Figure 3E shows the VarBalAcc from the top six segmentation teams for AVs and PVs across all test sets. The top teams were able to generalize well to external MRA datasets, with both anterior and posterior VarBalAcc averaging around 80%, and to external CTA datasets, with above 70% VarBalAcc in general, at near human-level performance.

CLINICAL APPLICATION I: FETAL PCA CLASSIFICATION

One potential clinical use of CoW segmentation models is fetal PCA classification, which is important for surgical

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

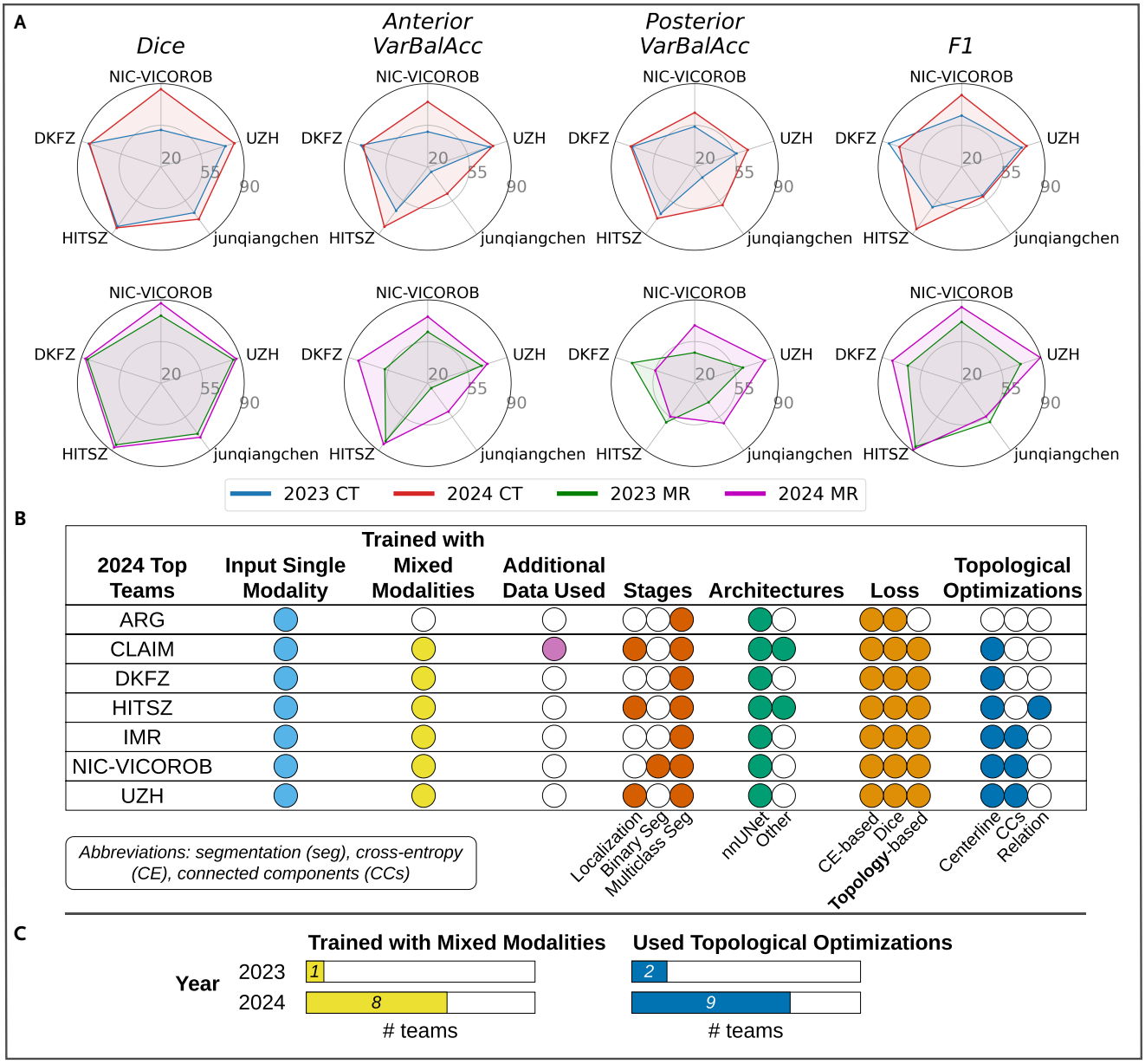


Figure 2. Progression of TopCoW Submissions and Winning Strategies.

Panel A shows the performance of the five teams that participated in both years on the same 34 patients in each year's test sets. The metrics shown are class-average Dice scores, variant-balanced accuracy of the anterior and posterior circle of Willis variant classifications, and average F1 score for detection of anterior communicating arteries, posterior communicating arteries, and third A2 arteries. The upper row shows computed tomography angiography performance; the lower shows magnetic resonance angiography performance. Panel B shows key characteristics of the segmentation algorithms from the top six teams from both tracks in alphabetical team name order. Panel C shows two methodological breakthroughs for segmentation in 2023 that got picked up by many more teams in the following year. CC denotes connected component; CE, cross-entropy; CT, computed tomography; MR, magnetic resonance; TopCoW, Topology-Aware Anatomical Segmentation of the Circle of Willis for Computed Tomography Angiography and Magnetic Resonance Angiography; and VarBalAcc, variant-balanced accuracy.

planning and interpretation of perfusion imaging;[1] neurosurgeons and neuroradiologists routinely assess fetal PCA status from CTA and MRA images. As shown in Figure 4A, diameters along the ipsilateral Pcom and P1 masks were used to predict the fetal PCA class. This allowed us to convert the top teams' segmentation outputs into fetal PCA

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

labels. Figure 4B shows that these outputs achieved around 80% or higher precision and recall for both fetal left PCA and fetal right PCA classification.

CLINICAL APPLICATION II: LOCATING ANEURYSM

Another important potential clinical use is locating aneurysms using CoW segmentation as reference. Clinicians routinely need to locate intracranial aneurysms relative to CoW anatomy in MRA and CTA images, as aneurysms most often occur along CoW vessels.[1] A total of 12 patients with aneurysms at various locations were segmented by the top four teams. Figure 4C shows the aneurysm ground truth overlaid with the CoW segmentations from a top team for four representative cases. CoW vessel labels overlapping or adjacent to the aneurysm were used to detect its location. Team UZH correctly located the aneurysms in all 12 patients, teams NIC-VICOROB and CLAIM in 11 out of 12, and team IMR in 10 out of 12 patients.

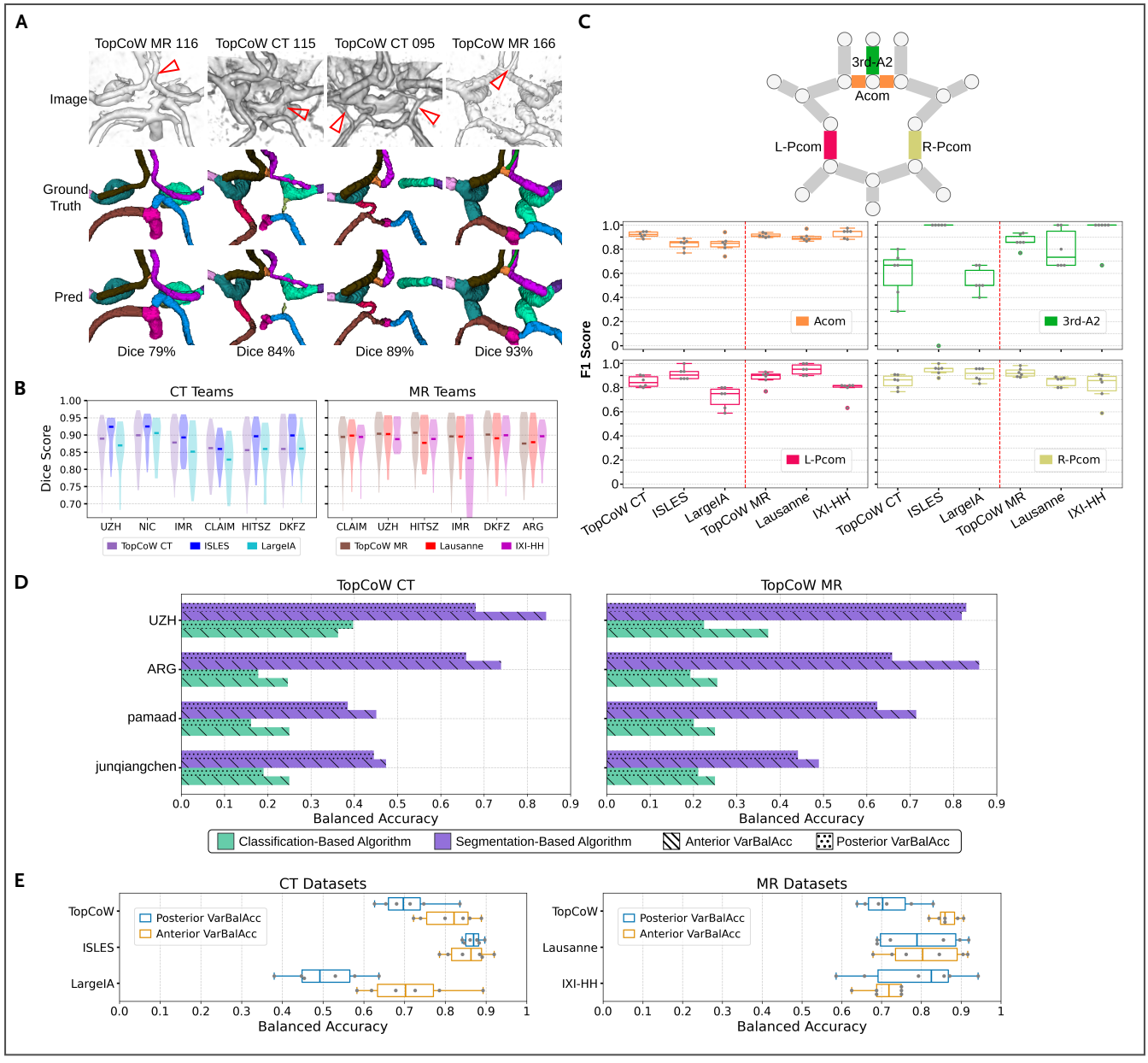


Figure 3. (*Continued*)

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

Discussion

BENEFITS OF MIXED-MODALITY TRAINING FOR CTA

Training a unified model with mixed modalities was an effective strategy, especially for the more difficult CTA modality. Between MRA and CTA, CTA consistently showed lower metric scores. This may be on account of veins and bones near the CoW being seen in CTA, and less detailed soft tissue being in the background compared with MRA. CTA benefited significantly more from mixed-modality training, showing greater performance gains than MRA. This is a key finding that may inform other CTA segmentation tasks.

IMPORTANCE OF TOPOLOGICAL OPTIMIZATIONS

Topological optimizations enabled CoW segmentation algorithms to be used in topology-dependent downstream clinical tasks, such as CoW variant and fetal PCA classification. These applications require segmented vessels to preserve key topological properties. Top submissions applied a range of topological optimizations, such as topology-preserving loss functions,[44-46] connected component postprocessing,[47,48] and class neighborhood relations.[49,50] Collectively, these methods improved connectivity, centerline accuracy, and overall topological correctness, making the segmentations suitable for topology-sensitive downstream tasks.

COMPARISON WITH CROWN RESULTS

Concurrent with the TopCoW challenge in 2023, the Circle of Willis Intracranial Artery Classification and Quantification (CROWN) challenge[22] also benchmarked the CoW variant classification task. The key difference was that CROWN did not provide any CoW segmentation annotations and instead formulated the task purely as image classification. We compared the performance of our algorithms with those reported in CROWN. Based on the merged common set of variants, the top two CROWN teams achieved 24%–30% anterior and 28%–50% posterior balanced accuracy, whereas the top two TopCoW MRA teams reached 82%–89% anterior and 75%–82% posterior balanced accuracy. Similar findings were observed in our analysis (Fig. 3D), where segmentation-based algorithms — trained and optimized for the CoW multiclass segmentation task — outperformed classification-based methods on the variant classification task by at least twofold.

EXPLAINABILITY VIA SEGMENTATION

CoW segmentation algorithms also provide explainability — a key feature in clinical settings. We showed that top-performing models could effectively detect key CoW components, classify CoW variants and fetal-type PCA, and locate intracranial aneurysms. Importantly, the conversion of segmentation masks into downstream outputs was fully transparent and interpretable, unlike black-box approaches. When clinicians require a confidence score or rationale behind a prediction — whether for detection,

Figure 3. Segmentation Performance.

Panel A shows qualitative results for the multiclass segmentation task. Ground truth masks are shown alongside predictions from team UZH for four selected cases from the internal test set, ordered by increasing class-average Dice score. Red arrowheads in the image indicate challenging features for each case (magnetic resonance [MR] 116: indeterminate anterior communicating artery vessel; computed tomography [CT] 115: hypoplastic right posterior communicating artery; CT 095: venous contamination; MR 166: third A2 artery variant). Panel B shows class-average Dice scores from the top six teams on CT angiography (CTA) and MR angiography (MRA) internal and external test sets. Team NIC-VICOROB is abbreviated as NIC. Team CLAIM used additional training data, which included some external MRA test images from Lausanne OpenNeuro aneurysm dataset and Information Extraction from Images Hammersmith Hospital dataset (IXI-HH) without our ground truth labels. Team IMR had five cases in IXI-HH with Dice scores below the shown range. The charts show violin plots with the middle bar indicating the median. Panel C shows detection of key circle of Willis (CoW) components for all test datasets by the top six teams from each modality. F1 scores from the six teams were displayed in boxplots. Panel D shows that four teams submitted both classification-based and segmentation-based algorithms for classifying the CoW variants. Classification-based algorithms were trained to predict CoW variant classes directly. Segmentation-based algorithms were purely for CoW segmentation and later evaluated for CoW variant classification performance. Panel E shows CoW variant classification performance from the top six segmentation teams on internal and external test sets. The boxplots show the anterior and posterior variant-balanced accuracy scores from the top six teams. Acom denotes anterior communicating artery; CT, computed tomography; ISLES, Ischemic Stroke Lesion Segmentation Challenge; IXI-HH, Information Extraction from Images Hammersmith Hospital; LargeIA, Large Intracranial Aneurysm Segmentation; L-Pcom, left posterior communicating artery; MR, magnetic resonance; R-Pcom, right posterior communicating artery; TopCoW, Topology-Aware Anatomical Segmentation of the Circle of Willis for Computed Tomography Angiography and Magnetic Resonance Angiography; and VarBalAcc, variant-balanced accuracy.

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

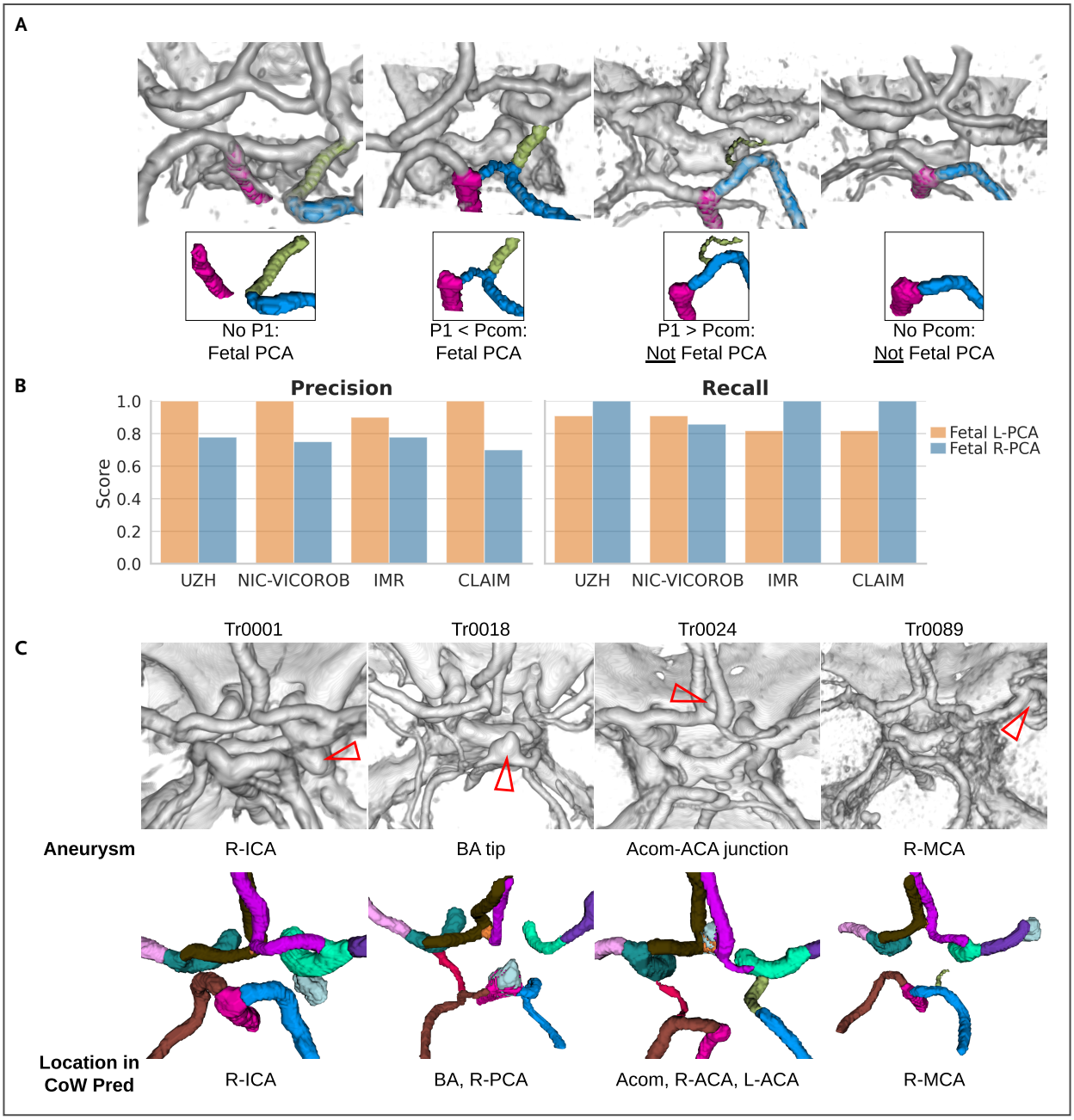


Figure 4. Two Clinical Applications of CoW Segmentation.

In Panel A, segmentation masks of the ipsilateral posterior communicating artery and P1 segments were used to determine whether a fetal posterior cerebral artery (PCA) class was present on that side. The upper row shows computed tomography angiography images overlaid with segmentation masks related to fetal PCA. The lower row shows a zoomed-in view of the segmentation masks. Examples shown illustrate the fetal right PCA classification. Panel B shows fetal PCA classification performance in terms of precision and recall scores from the top four teams. In Panel C, the upper row shows CTA images with aneurysms indicated by red arrowheads. The lower row shows the aneurysm masks in a silver color overlaid with predicted CoW segmentation masks from team UZH. The "Location in CoW Pred" indicates the predicted CoW labels that the aneurysm overlapped with or was adjacent to. ACA denotes anterior cerebral artery; Acom, anterior communicating artery; BA, basilar artery; CoW, circle of Willis; L-ACA, left anterior cerebral artery; L-PCA, left posterior cerebral artery; PCA, posterior cerebral artery; Pcom, posterior communicating artery; R-ACA, right anterior cerebral artery; R-ICA, right internal carotid artery; R-MCA, right middle cerebral artery; and R-PCA, right posterior cerebral artery.

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

classification, or localization — they can better interpret and contextualize the results by inspecting the CoW segmentation output.

VR TO HANDLE COMPLEX ANATOMY

As one of the first challenges to use VR-generated annotations at scale, our VR-based expert-in-the-loop annotation protocol proved very effective and can overcome the otherwise time-consuming annotation process for a complex multiclass anatomical segmentation problem. The 3D depth view offered efficient annotation and verification capabilities that are well suited for curvilinear structures such as the CoW vessels, enabling fast and accurate annotation of complex CoW anatomies and rare variants. This allowed us to prepare a densely annotated large-scale dataset that covered many clinically relevant CoW anatomies. This VR-driven efficiency mirrors recent findings where VR increased the speed and accuracy of 3D annotations.[32]

CLINICAL IMPLICATIONS, LIMITATIONS, AND FUTURE WORK

Accurate CoW variant characterization is relevant for neurovascular practice because variant patterns may influence collateral circulation assessment and procedural planning. In this benchmark, the segmentation-based approach provided voxel-level vessel delineation and transparent derivation of key components, variant classes, fetal PCA classification, and aneurysm localization. These outputs can inform endovascular catheter navigation and support evaluation of flow redistribution during microsurgical interventions, including bypass strategies and temporary vessel occlusion. The benchmarked algorithms may therefore help automate CoW analysis and enable more standardized preprocedural planning, although this benchmark does not directly evaluate effects on procedural outcomes or safety. Nevertheless, several limitations of our study should be acknowledged. Owing to the large heterogeneity of CoW anatomy, not all CoW variants were included in our annotation scheme. Expanding labels to include rarer variants could improve model robustness and applicability. The current CoW variant graph could also be refined to include more detailed types involving hypoplasia and vessel diameters using the same workflow applied in our fetal PCA classification, where diameters were extracted along the centerlines of P1 and Pcom. Our dataset exhibited class imbalance, with certain vessel variants represented by only two to three training examples, which may reduce prediction reliability for uncommon but clinically important variants. External validation was restricted to the top six teams per modality owing to resource constraints; however, as this represents approximately half of all submissions in the 2024 iteration, the subset provides a representative assessment of the state of the art. Finally, while this work solely focuses on the CoW, future efforts could focus on segmenting the full brain vessel anatomy, including all major arteries and veins visible in angiograms, enabling a more comprehensive vascular analysis.

Disclosures

Author disclosures and other supplementary materials are available at ai.nejm.org.

The Topology-Aware Anatomical Segmentation of the Circle of Willis for Computed Tomography Angiography and Magnetic Resonance Angiography (TopCoW) training dataset, consisting of 250 annotated images, and our multicenter test sets of 86 annotated images are available in our public Zenodo repository (15 GB) at https://zenodo.org/records/15692630.

The Docker images from the best-performing teams, along with the scripts to facilitate local execution, are available in our public Zenodo repository (45 GB) at https://zenodo.org/records/15665435.

For code availability of individual team submissions, please see Section S13 of the Supplementary Appendix.

The implementation of our evaluation metric code is available open-source at https://github.com/CoWBenchmark/TopCoW_Eval_Metrics.

The challenge is supported by the Digitalization Initiative of the Zurich Higher Education Institutions and the Helmut Horten Foundation. Dr. Ruigrok has received funding from the European Research Council under the European Union's Horizon 2020 Research and Innovation program (grant agreement number 852173). Dr. Baazaoui received funding from the Betty and David Koetser Foundation and the Young Talents in Clinical Research program of the Swiss Academy of Medical Sciences and of the Gottfried and Julia Bangerter-Rhyner Foundation. Ms. Casademunt, Dr. Camara, Dr. Bijlenga, Dr. Hämmerli, Dr. Hirsch, Dr. Juchler, Mr. Musio, and Dr. Delucchi received funding from GEMINI (funded by EU Horizon Europe Research and Innovation program, grant number 101136438). Dr. Ting Ma received funding from the National Natural Science Foundation of China (grant number 62276081). Team 2i_mtl was supported by grants from the Quebec Bio-Imaging Network (project number 21.24) and start-up funds from the University of Montreal Hospital Research Centre and the Department of Radiology, Radiation Oncology, and Nuclear Medicine, University of Montreal/Bayer. Dr. Letourneau-Guillon is supported by a Clinical Research Scholarship Junior 1 Salary Award (grant number 311203) from the Quebec Health Research Fund and the Quebec Association of Radiologists Foundation. Team ARG was supported by the Institute of Information and Communications Technology Planning and Evaluation grant funded by the South Korean government (Ministry of Science and ICT) (grant number RS-2022-II220959 [Part 2] Few-Shot Learning of Causal Inference in Vision and Language for Decision Making). Team

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

CLAIM acknowledges funding from the German Federal Ministry of Education and Research (Anonymised Data for the Digitalised Medicine of the Future [ANONYMED] Project, no. 16KISA042K) and from the European Unions Horizon 2020 research and innovation programme (VALIDATE Project, no. 777107, coordinator Dietmar Frey). Computation has been performed on the high-performance computing for research cluster of the Berlin Institute of Health. They also acknowledge the contribution of MRCLEAN investigators by providing access to data from the MRCLEAN trial. Team DKFZ was supported by the Helmholtz Association under the joint research school Helmholtz Information and Data Science School for Health and part of their work was funded by Helmholtz Imaging, a platform of the Helmholtz Incubator on Information and Data Science. Dr. Holzschuh has been funded by a fellowship of the German Cancer Research Center Clinician Scientist Program, supported by the Dieter Morszeck Foundation. Team EURECOM was partially funded by the French government through the 3IA Cote d'Azur Investments in the Future project managed by the French National Research Agency (ANR) (grant number ANR-19-P3IA-0002) and by the ANR Young Researchers Program project Interactive and Collaborative Learning for Vessel Segmentation (grant number 22-CE45-0015-01). Team IMR was supported in part by the National Key Research and Development Program of China (grant number 2022ZD0212400), National Natural Science Foundation of China (grant number 62373243), and the Science and Technology Commission of Shanghai Municipality, China (grant number 20DZ2220400), Shanghai Municipal Science and Technology Major Project (grant number 2021SHZDZX0102). Team IWantToGoToCanada was supported by the National Research Foundation of Korea (grant number NRF-2020M3E5D2A01084892), the Institute for Basic Science (grant number IBS-R015-D1), the Information Technology Research Center support program (grant number IITP-2023-2018-0-01798), the AI Graduate School Support Program (grant number 2019-0-00421), the ICT Creative Consilience program (grant number IITP-2023-2020-0-01821), and the Artificial Intelligence Innovation Hub program (grant number 2021-0-02068). Team lWM wants to acknowledge the Polish high-performance computing infrastructure PL-Grid support (grant number PLG/2023/016239). Team NantesU was partially supported by the ANR project eCAN and INSERM CoPoC #MAT-PI-22155-A-01 (grant number RVF23037NSA). Team NIC-VICOROB was supported by the Ministry of Science, Innovation, and Universities (grant number DPI2020-114769RB-I00) as well as by the Catalan Institution for Research and Advanced Studies (ICREA) under the ICREA Academia program. Members of the 2024 team received support from the PID2020-114769RBI00 and PID2023-146187OB-I00 projects funded by the Ministry of Science, Innovation, and Universities. Team Pamaad Ms. P. Casademunt was supported by the European Union's Horizon 2020 (grant agreement number 101136438) (GEMINI project), the Agency for Management of University and Research Grants (grant number 2024 FI-1 00419), and the Maria de Maeztu grant of excellence. Dr. Galdran is supported by grant number RYC2022-037144-I, funded by MCIN/AEI/10.13039/501100011033 and by European Social Fund Plus. They would like to thank the high-performance computing team from Zurich University of Applied Sciences, particularly Pascal Häussler and Stefan Weber, for their generous allocation of computational resources and technical assistance. Team UB-VTL wants to acknowledge the computational resources provided by the Center of Computational Research at University at Buffalo. Team UW was supported by the United States National Institutes of Health (grant numbers R01HL162743 and R00HL136883).

We thank Hrvoje Bogunović for helpful discussions and suggestions during the early planning stages. We also thank Nathan Spencer and Michael Morehead from syGlass for their technical assistance for the virtual reality setup, and James Meakin and Chris van Run from grand-challenge.org for their technical support for the challenge infrastructure.

Dr. Catherine Wurster contributed to this work during its early stages and passed away unexpectedly before publication.

Author Affiliations

[1] Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland

[2] Institute of Computational Life Sciences, Zurich University of Applied Sciences, Waedenswil, Switzerland

[3] Department of Neuroradiology, University Hospital of Zurich, Zurich, Switzerland

[4] Department of Neurosurgery, Zhongnan Hospital of Wuhan University, Wuhan, China

[5] Department of Radiology, Weill Cornell Medicine, Cornell University, New York, NY, USA

[6] Institute for Tissue Engineering and Regenerative Medicine, Helmholtz Munich, Neuherberg, Germany

[7] School of Computation, Information and Technology, Technical University of Munich, Munich, Germany

[8] Athinoula A. Martinos Center for Biomedical Imaging, Harvard Medical School, Boston, MA, USA

[9] School of Medicine and Health, TUM University Hospital, Technical University of Munich, Munich, Germany

[10] Helmholtz AI, Helmholtz Munich, Neuherberg, Germany

[11] Munich Center for Machine Learning, Munich, Germany

[12] Department of Computing, Imperial College London, London, UK

[13] Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

[14] Department of Neurology and Neurosurgery, University Medical Center Utrecht, Utrecht, The Netherlands

[15] Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands

[16] School of Electronics and Information Engineering, Harbin Institute of Technology, Shenzhen, China

[17] School of Biomedical Engineering, Harbin Institute of Technology, Shenzhen, China

[18] Division of Medical Image Computing, German Cancer Research Center, Heidelberg, Germany

[19] Faculty of Mathematics and Computer Science, Heidelberg University, Heidelberg, Germany

[20] Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany

[21] Helmholtz Imaging, German Cancer Research Center, Heidelberg, Germany

[22] Pattern Analysis and Learning Group, Department of Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

[23] Department of Radiology, University of Washington, Seattle, USA
[24] Department of Radiology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[25] Department of Clinical Neurosciences, Division of Neurosurgery, Geneva University Hospitals, Geneva, Switzerland
[26] Department of Neurology, University Hospital of Zurich, Zurich, Switzerland
[27] Department of Physiology, University of Toronto, Toronto, ON
[28] Department of Neurosurgery, University Hospital of Zurich, Zurich, Switzerland
[29] Department of Diagnostic Imaging, National University Hospital, Singapore, Singapore
[30] University of Chicago, Chicago, IL, USA
[31] Department of Diagnostic and Interventional Neuroradiology, University Hospital Berne and University of Bern, Bern, Switzerland
[32] University of Montreal Hospital Research Centre, Montreal, QC
[33] DEEPNOID, Seoul, South Korea
[34] Department of Artificial Intelligence, Korea University, Seoul, South Korea
[35] Charité Lab for Artificial Intelligence in Medicine, Charité — Berlin University Medicine, Berlin, Germany
[36] National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, UK
[37] Centre for Medical Image Computing, Department of Computer Science, University College London, London, UK
[38] Department of Radiation Oncology, Duke University Medical Center, Durham, NC, USA
[39] EURECOM, Biot, France
[40] Institute of Medical Technology, Peking University Health Science Center, Beijing, China
[41] Genlight MedTech, Hangzhou, China
[42] Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China
[43] School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China
[44] Department of Artificial Intelligence, Sungkyunkwan University, Seoul, South Korea
[45] Department of Electrical and Computer Engineering, Sungkyunkwan University, Seoul, South Korea
[46] Shanghai MediWorks Precision Instruments, Shanghai, China
[47] Institute of Informatics, HES-SO Valais-Wallis, Sierre, Switzerland
[48] Department of Measurement and Electronics, AGH University of Kraków, Kraków, Poland
[49] Institute of Thorax, Nantes University, Nantes, France
[50] Nantes Thermal and Energy Laboratory, Polytech Nantes, Nantes University, Nantes, France
[51] Research Institute of Computer Vision and Robotics, University of Girona, Girona, Spain
[52] Center for Precision Health, McWilliams School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, USA
[53] Physense, BCN-Medtech, Department of Communication and Information Technologies, Pompeu Fabra University, Barcelona, Spain
[54] Department of Mathematical Modeling and Machine Learning, University of Zurich, Zurich, Switzerland
[55] Laboratory of Brain Atlas and Brain-Inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[56] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
[57] School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China
[58] Canon Stroke and Vascular Research Center, University at Buffalo, Buffalo, NY, USA
[59] Department of Neurosurgery, University at Buffalo, Buffalo, NY, USA
[60] Department of Pathology and Anatomical Sciences, University at Buffalo, Buffalo, NY, USA
[61] LPIXEL, Tokyo, Japan

The authors' full names and academic degrees are as follows: Kaiyuan Yang, M.Sc., Fabio Musio, M.Sc., Yihui Ma, M.D., Ph.D., Norman Juchler, Ph.D., Johannes C. Paetzold, Ph.D., Rami Al-Maskari, M.Sc., Luciano Höher, Hongwei Bran Li, Ph.D., Ibrahim Ethem Hamamci, M.D., Anjany Sekuboyina, Ph.D., Suprosanna Shit, Ph.D., Houjing Huang, Ph.D., Chinmay Prabhakar, M.Sc., Ezequiel de la Rosa, Ph.D., Bastian Wittmann, M.Sc., Diana Waldmannstetter, Ph.D., Florian Kofler, Ph.D., Fernando Navarro, Ph.D., Martin J. Menten, Ph.D., Ivan Ezhov, Ph.D., Daniel Rueckert, Ph.D., Iris N. Vos, Ph.D., Ynte M. Ruigrok, M.D., Birgitta K. Velthuis, Ph.D., Hugo J. Kuijf, Ph.D., Pengcheng Shi, M.Sc., Wei Liu, M.Sc., Ting Ma, Ph.D., Maximilian R. Rokuss, M.Sc., Yannick Kirchhoff, M.Sc., Fabian Isensee, Ph.D., Klaus Maier-Hein, Ph.D., Chengcheng Zhu, Ph.D., Huilin Zhao, M.D., Ph.D., Philippe Bijlenga, M.D., Ph.D., Julien Hämmerli, M.D., Catherine Wurster, M.D., Laura Westphal, M.D., Jeroen Bisschop, M.D., Ph.D., Elisa Colombo, M.D., Hakim Baazaoui, M.D., Hannah-Lea Handelsmann, M.D., Andrew Makmur, M.B.B.S., M.Med., F.R.C.R., James Hallinan, F.R.C.R., Amrish Soundararajan, M.B.B.S., M.B.A., Benedikt Wiestler, M.D., Jan S. Kirschke, M.D., Evamaria O. Riedel, M.D., Roland Wiest, M.D., Emmanuel Montagnon, Ph.D., Laurent Letourneau-Guillon, M.D., M.Sc., Kwanseok Oh, Dahye Lee, M.Sc., Orhun Utku Aydin, M.D., Adam Hilbert, M.Sc., Jana Rieger, M.Sc., Dimitrios Rallios, M.D., Satoru Tanioka, M.D., Ph.D., Alexander Koch, M.Sc., Dietmar Frey, M.D., Abdul Qayyum, Ph.D., Moona Mazher, Ph.D., Steven Niederer, Ph.D., Nico Disch, Ph.D., Julius C. Holzschuh, M.D., M.Sc., Dominic LaBella, M.D., Francesco Galati, Ph.D., Daniele Falcetta, Ph.D., Maria A. Zuluaga, Ph.D., Chaolong Lin, M.Sc., Haoran Zhao, M.Sc., Zehan Zhang, M.Sc., Minghui Zhang, Ph.D., Xin You, Ph.D., Hanxiao Zhang, Ph.D., Guang-Zhong Yang, Ph.D., Yun Gu, Ph.D., Sinyoung Ra, M.Sc., Jongyun Hwang, M.Sc., Hyunjin Park, Ph.D., Junqiang Chen, M.Sc., Marek Wodzinski, Ph.D., Henning Müller, Ph.D., Nesrin Mansouri, Ph.D., Florent Autrusseau, Ph.D., Cansu Yalcin, M.Sc., Rachika E. Hamadache, M.Sc., Clara Lisazo, M.Sc., Joaquim Salvi, Ph.D., Adrià Casamitjana, Ph.D., Xavier Lladó, Ph.D., Uma Maria Lal-Trehan Estrada, M.Sc., Valeriia Abramova, M.Sc., Luca Giancardo, Ph.D., Arnau Oliver, Ph.D., Paula Casademunt, M.Sc., Adrian Galdran, Ph.D., Matteo Delucchi, Ph.D., Oscar Camara, Ph.D., Jialu Liu, M.Sc., Haibin Huang, M.Sc., Yue Cui, Ph.D., Zehang Lin, M.Sc., Yusheng Liu, Ph.D., Shunzhi Zhu, Ph.D., Tatsat R. Patel, Ph.D., Adnan H. Siddiqui, M.D., Ph.D., Vincent M. Tutino, Ph.D., Maysam Orouskhani, Ph.D., Huayu Wang, M.Sc., Mahmud Mossa-Basha, M.D., Yuki Sato, M.Sc., Sven Hirsch, Ph.D., Susanne Wegener, M.D., and Bjoern Menze, Ph.D.

References

1. Osborn AG. Osborn's brain: imaging, pathology, and anatomy. 1st ed. Salt Lake City, UT: Amirsys, 2012.

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

2. Liebeskind DS. Collateral circulation. Stroke 2003;34:2279-2284. DOI: 10.1161/01.STR.0000086465.41263.06.

3. Chuang YM, Chan L, Lai YJ, et al. Configuration of the circle of Willis is associated with less symptomatic intracerebral hemorrhage in ischemic stroke patients treated with intravenous thrombolysis. J Crit Care 2013;28:166-172. DOI: 10.1016/j.jcrc.2012.08.018.

4. van Seeters T, Hendrikse J, Biessels GJ, et al. Completeness of the circle of Willis and risk of ischemic stroke in patients without cerebrovascular disease. Neuroradiology 2015;57:1247-1251. DOI: 10.1007/s00234-015-1589-2.

5. Kim KM, Kang HS, Lee WJ, Cho YD, Kim JE, Han MH. Clinical significance of the circle of Willis in intracranial atherosclerotic stenosis. J Neurointerv Surg 2016;8:251-255. DOI: 10.1136/neurintsurg-2014-011439.

6. Rinaldo L, McCutcheon BA, Murphy ME, Bydon M, Rabinstein AA, Lanzino G. Relationship of A1 segment hypoplasia to anterior communicating artery aneurysm morphology and risk factors for aneurysm formation. J Neurosurg 2017;127:89-95. DOI: 10.3171/2016.7.JNS16736.

7. Hindenes LB, Ingebrigtsen T, Isaksen JG, et al. Anatomical variations in the circle of Willis are associated with increased odds of intracranial aneurysms: the Tromsø study. J Neurol Sci 2023;452:120740. DOI: 10.1016/j.jns.2023.120740.

8. Yang F, Li H, Wu J, et al. Relationship of A1 segment hypoplasia with the radiologic and clinical outcomes of surgical clipping of anterior communicating artery aneurysms. World Neurosurg 2017;106:806-812. DOI: 10.1016/j.wneu.2017.07.122.

9. Banga PV, Varga A, Csobay-Novák C, et al. Incomplete circle of Willis is associated with a higher incidence of neurologic events during carotid eversion endarterectomy without shunting. J Vasc Surg 2018;68:1764-1771. DOI: 10.1016/j.jvs.2018.03.429.

10. Krabbe-Hartkamp MJ, Van der Grond J, De Leeuw F, et al. Circle of Willis: morphologic variation on three-dimensional time-of-flight MR angiograms. Radiology 1998;207:103-111. DOI: 10.1148/radiology.207.1.9530305.

11. Iqbal S. A comprehensive study of the anatomical variations of the circle of Willis in adult human brains. J Clin Diagn Res 2013;7:2423-2427. DOI: 10.7860/JCDR/2013/6580.3563.

12. Bullitt E, Zeng D, Gerig G, et al. Vessel tortuosity and brain tumor malignancy: a blinded study. Acad Radiol 2005;12:1232-1240. DOI: 10.1016/j.acra.2005.05.027.

13. IXI. IXI dataset — brain development. 2022 (https://brain-development.org/ixi-dataset/).

14. CAS2023. CodaLab. Cerebral artery segmentation challenge (CAS) 2023. May 23, 2023 (https://codalab.lisn.upsaclay.fr/competitions/9804).

15. Chatterjee S, Mattern H, Dörner M, et al. SMILE-UHURA challenge–small vessel segmentation at mesoscopic scale from ultra-high resolution 7T magnetic resonance angiograms. November 14, 2024 (https://arxiv.org/abs/2411.09593v1). Preprint.

16. Mou L, Yan Q, Lin J, et al. COSTA: a multi-center TOF-MRA dataset and a style self-consistency network for cerebrovascular segmentation. IEEE Trans Med Imaging 2024;43:4442-4456. DOI: 10.1109/TMI.2024.3424976.

17. Bogunović H, Pozo JM, Cárdenes R, San Román L, Frangi AF. Anatomical labeling of the circle of Willis using maximum a posteriori probability estimation. IEEE Trans Med Imaging 2013;32:1587-1599. DOI: 10.1109/TMI.2013.2259595.

18. Robben D, Sunaert S, Thijs V, Wilms G, Maes F, Suetens P. Anatomical labeling of the circle of Willis using maximum a posteriori graph matching. In: Mori K, Sakuma I, Sato Y, Barillot C, Navab N, eds. Medical Image Computing and Computer-Assisted Intervention — MICCAI 2013. MICCAI 2013. Lecture Notes in Computer Science, Vol 8149. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-642-40811-3_71.

19. Robben D, Türetken E, Sunaert S, et al. Simultaneous segmentation and anatomical labeling of the cerebral vasculature. Med Image Anal 2016;32:201-215. DOI: 10.1016/j.media.2016.03.006.

20. Chen L, Hatsukami T, Hwang JN, Yuan C. Automated intracranial artery labeling using a graph neural network and hierarchical refinement. In: AL Martel, P Abolmaesumi, D Stoyanov, et al., eds. Proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020. September 29, 2020 (https://doi.org/10.1007/978-3-030-59725-2_8).

21. Hong SW, Song HN, Choi JU, et al. Automated in-depth cerebral arterial labelling using cerebrovascular vasculature reframing and deep neural networks. Sci Rep 2023;13:3255. DOI: 10.1038/s41598-023-30234-6.

22. Vos IN, Ruigrok YM, Bennink E, et al. Evaluation of techniques for automated classification and artery quantification of the circle of Willis on TOF-MRA images: the CROWN challenge. Med Image Anal 2025;105:103650. DOI: 10.1016/j.media.2025.103650.

23. Dumais F, Caceres MP, Janelle F, et al. eICAB: a novel deep learning pipeline for circle of Willis multiclass segmentation and analysis. Neuroimage 2022;260:119425. DOI: 10.1016/j.neuroimage.2022.119425.

24. Hilbert A, Rieger J, Madai VI, et al. Anatomical labeling of intracranial arteries with deep learning in patients with cerebrovascular disease. Front Neurol 2022;13:1000914. DOI: 10.3389/fneur.2022.1000914.

25. de la Rosa E, Su R, Reyes M, et al. ISLES'24: final infarct prediction with multimodal imaging and clinical data. Where do we stand? July 7, 2025 (https://arxiv.org/abs/2408.10966v2). Preprint.

26. Riedel EO, de la Rosa E, Baran TA, et al. The ischemic stroke lesion segmentation challenge (ISLES)'24 dataset: a multimodal stroke imaging dataset with hyperacute CT, acute postinterventional MRI, and 3-month clinical outcomes. Radiology AI 2026;8(3). DOI: 10.1148/ryai.250603.

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.

27. Bo ZH. Large IA segmentation dataset. Zenodo. February 12, 2021 (https://zenodo.org/records/6801398).

28. Bo ZH, Qiao H, Tian C, et al. Toward human intervention-free clinical diagnosis of intracranial aneurysm via deep neural network. Patterns 2021;2:100197. DOI: 10.1016/j.patter.2020.100197.

29. Di Noto T, Marie G, Tourbier S, et al. Lausanne TOF-MRA aneurysm cohort. OpenNeuro. 2022. DOI: 10.18112/openneuro.ds003949.v1.0.1.

30. Di Noto T, Marie G, Tourbier S, et al. Towards automated brain aneurysm detection in TOF-MRA: open data, weak labels, and anatomical knowledge. Neuroinformatics 2023;21:21-34. DOI: 10.1007/s12021-022-09597-0.

31. Pidhorskyi S, Morehead M, Jones Q, Spirou G, Doretto G. syGlass: interactive exploration of multidimensional images using virtual reality head-mounted displays. August 22, 2018 (http://arxiv.org/abs/1804.08197). Preprint.

32. Kaltenecker D, Al-Maskari R, Negwer M, et al. Virtual reality-empowered deep-learning analysis of brain cells. Nat Methods 2024;21:1306-1315. DOI: 10.1038/s41592-024-02245-2.

33. Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297-302. DOI: 10.2307/1932409.

34. Shit S, Paetzold JC, Sekuboyina A, et al. clDice – a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the 2021 IEEE/Computer Vision Foundation Conference on Computer Vision and Pattern Recognition. Nashville, TN: IEEE, 2021:16555-16564. DOI: 10.1109/CVPR46437.2021.01629.

35. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015;15:29. DOI: 10.1186/s12880-015-0068-x.

36. Maier-Hein L, Reinke A, Godau P, et al. Metrics reloaded: recommendations for image analysis validation. Nat Methods 2024;21:195-212. DOI: 10.1038/s41592-023-02151-z.

37. Hu X, Li F, Samaras D, Chen C. Topology-preserving deep image segmentation. In: Proceedings of the 33rd Conference on Neural Information Processing Systems. Red Hook, NY: NeurIPS: 2019:32 (https://proceedings.neurips.cc/paper_files/paper/2019/file/2d95666e2649fcfc6e3af75e09f5adb9-Paper.pdf).

38. Menten MJ, Paetzold JC, Zimmer VA, et al. A skeletonization algorithm for gradient-based optimization. In: Proceedings of the 2023 IEEE/Computer Vision Foundation International Conference on Computer Vision. Paris, France: IEEE, 2023:21337-21346. DOI: 10.1109/ICCV51070.2023.01956.

39. Baumgartner M, Full PM, Maier-Hein KH. Accurate detection of mediastinal lesions with nnDetection. In: Y Xiao, G Yang, S Song, eds. Lesion Segmentation in Surgical and Diagnostic Applications. CuRIOUS KiPA MELA 2022. March 14, 2023 (https://doi.org/10.1007/978-3-031-27324-7_10).

40. Ceballos-Arroyo AM, Nguyen HT, Zhu F, et al. Vessel-aware aneurysm detection using multi-scale deformable 3D attention. In: MG Linguraru, et al. Proceedings of Medical Image Computing and Computer Assisted Intervention–MICCAI 2024. MICCAI 2024. October 4, 2024 (https://doi.org/10.1007/978-3-031-72086-4_71).

41. Musio F, Yang K, Shit S, et al. Quantitative evaluation of the circle of Willis vascular architecture in 3D CT and MR angiography. In: Proceedings of the 8th International Conference on Computational and Mathematical Biomedical Engineering. Arlington, VA: CMBE24, 2024:563-566 (https://digitalcollection.zhaw.ch/server/api/core/bitstreams/2cc5a326-4302-44eb-babf-134ad8e9a868/content).

42. Musio F, Juchler N, Yang K, et al. Circle of Willis centerline graphs: a dataset and baseline algorithm. Neurosci Inform 2026;6:100265. DOI: 10.1016/j.neuri.2026.100265.

43. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203-211. DOI: 10.1038/s41592-020-01008-z.

44. Kirchhoff Y, Rokuss MR, Roy S, et al. Skeleton recall loss for connectivity conserving and resource efficient segmentation of thin tubular structures. In: Proceedings of the European Conference on Computer Vision 2024. Milan, Italy: ECCV, 2024: 218-234 (https://dl.acm.org/doi/10.1007/978-3-031-72980-5_13).

45. Shi P, Hu J, Yang Y, Gao Z, Liu W, Ma T. Centerline boundary dice loss for vascular segmentation. In: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention 2024. Marrakesh, Morocco: MICCAI, 2024: 46-56 (https://dl.acm.org/doi/10.1007/978-3-031-72111-3_5).

46. Zhang M, Gu Y. Towards connectivity-aware pulmonary airway segmentation. IEEE J Biomed Health Inform 2023;28:321-332. DOI: 10.1109/JBHI.2023.3324080.

47. Zhang M, You X, Zhang H, Gu Y. Topology-aware exploration of circle of Willis for CTA and MRA: segmentation, detection, and classification. October 21, 2024 (https://doi.org/10.48550/arXiv.2410.15614). Preprint.

48. Hamadache RE, Lisazo C, Yalcin C, et al. Topology-aware multiclass segmentation of the circle of Willis from MRA and CTA images. Comput Biol Med 2026;204:111516. DOI: 10.1016/j.compbiomed.2026.111516.

49. Gupta S, Hu X, Kaan J, et al. Learning topological interactions for multi-class medical image segmentation. In: S Avidan, G Brostow, M Cissé, GM Farinella, T Hassner, eds. Computer Vision–ECCV 2022. ECCV 2022. October 22, 2022 (https://doi.org/10.1007/978-3-031-19818-2_40).

50. Shi P, Guo X, Yang Y, Ye C, Ma T. NexToU: efficient topology-aware U-Net for medical image segmentation. May 25, 2023 (https://arxiv.org/abs/2305.15911). Preprint.

NEJM AI is produced by NEJM Group, a division of the Massachusetts Medical Society.
Downloaded from ai.nejm.org on November 16, 2026.