Leveraging a Foundation Model for Artificial Intelligence-Based Follow-up Infarct Segmentation in Acute Ischemic Stroke

Davide Carone[1,2,*], Giacomo Benvenuti[1,*], James Garrard[1,2], Olivier Joly[1], Petr Mikulenka[3], Ivana Stetkarova[3], George Harston[1,2], Ansaar Rai[4]

1. Brainomix Ltd, Oxford, United Kingdom. 2. Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom. 3. Department of Neurology, University Hospital Královské Vinohrady, Prague, Czechia. 4. Neuroradiology, Rockefeller Neuroscience Institute, Morgantown, WV, United States. *Contributed Equally.

Rationale

Follow-up infarct volume (FIV) is a proposed surrogate endpoint for proof-of-concept clinical studies in acute ischemic stroke.[1]

Manual annotation of infarction on diffusion-weighted magnetic resonance imaging (DWI) is labor-intensive, costly, subject to high intra- and inter-observer variability. Automated tools are being developed to address this issue, with varying accuracy.[2]

However, models trained in a purely supervised manner on task-specific datasets learn a limited representation of medical image statistics, which may restrict their ability to generalize.

To address this limitation, we leveraged an imaging foundation model that learns rich and transferable representations from large-scale medical imaging data using unsupervised pre-training. We developed an original fine-tuning strategy to enable automatic FIV segmentation without prompting.

Methods

Imaging Analysis:

BrainomixSAM2 is an artificial intelligence tool for automated FIV segmentation on diffusion-weighted imaging (b1000). BrainomixSAM2 was developed based on MedSAM2, an adaptation of Meta's foundation model "Segment Anything Model 2" (SAM2) optimized for medical images. The development of this tool included a two-stage fine-tuning to allow automated segmentation without prompting, trained on 344 patients with acute ischemic stroke who had follow-up DW MRI.

Validation Was Performed on Two Independent Stroke Imaging Registries:

Cohort #1: 61 patients with suspected anterior circulation large-vessel occlusion (West Virginia University, US)

Cohort #2: 59 patients with mixed anterior and posterior circulation stroke (University Hospital Královské Vinohrady, Czechia)

Statistical Analysis:

The performance of BrainomixSAM2 performance with ground-truth expert raters and the current state-of-the-art automated approach (DeepISLES[2]). Examples are provided (Figure 1).

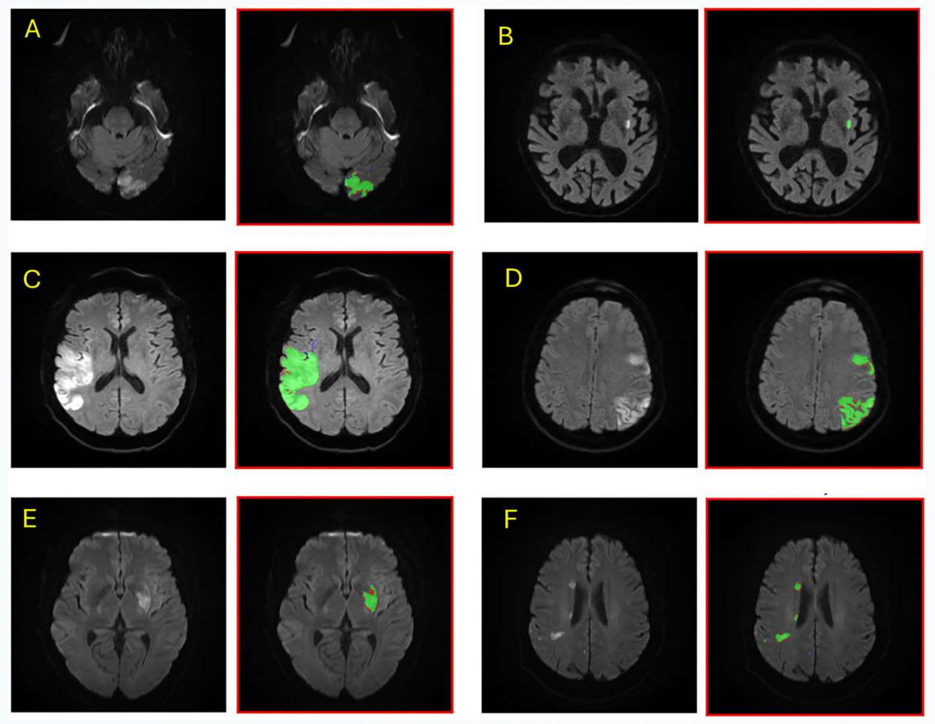

Figure 1: Six example segmentations using BrainomixSAM2. True positive (green), false positive (red) and false negative (blue) areas are identified.

Performance metrics included volumetric concordance (intraclass correlation coefficient - ICC), and segmentation accuracy (Dice Similarity Index - DICE).

Results

1. **BrainomixSAM2 demonstrates excellent volumetric agreement with ground-truth annotations across diverse stroke cohorts.**
2. **Segmentation accuracy was comparable to the current benchmark, DeepISLES, supporting its feasibility and reliability for automated FIV assessment.**

The demographic information for the two validation cohorts are provided in Table 1.

Domain	Description	Cohort 1 (WVU)	Cohort 2 (FHKV)
Age	Median [IQR]	66 [25]	66 [16]
Female Sex	%	31 (50.8%)	23 (39.0%)
NIHSS	Median [IQR]	16 [7]	9 [8.5]
Follow-up Infarct Volume	Median [IQR] (ml)	30.5 [37.6]	3.9 [14.6]

Table 1: Summary demographics of patients in the two validation cohorts.

The volumetric agreement (ICC) and DICE in Cohorts 1 and 2 are provided in Table 2. For comparison, the results using DeepISLES are provided on Cohort 2.

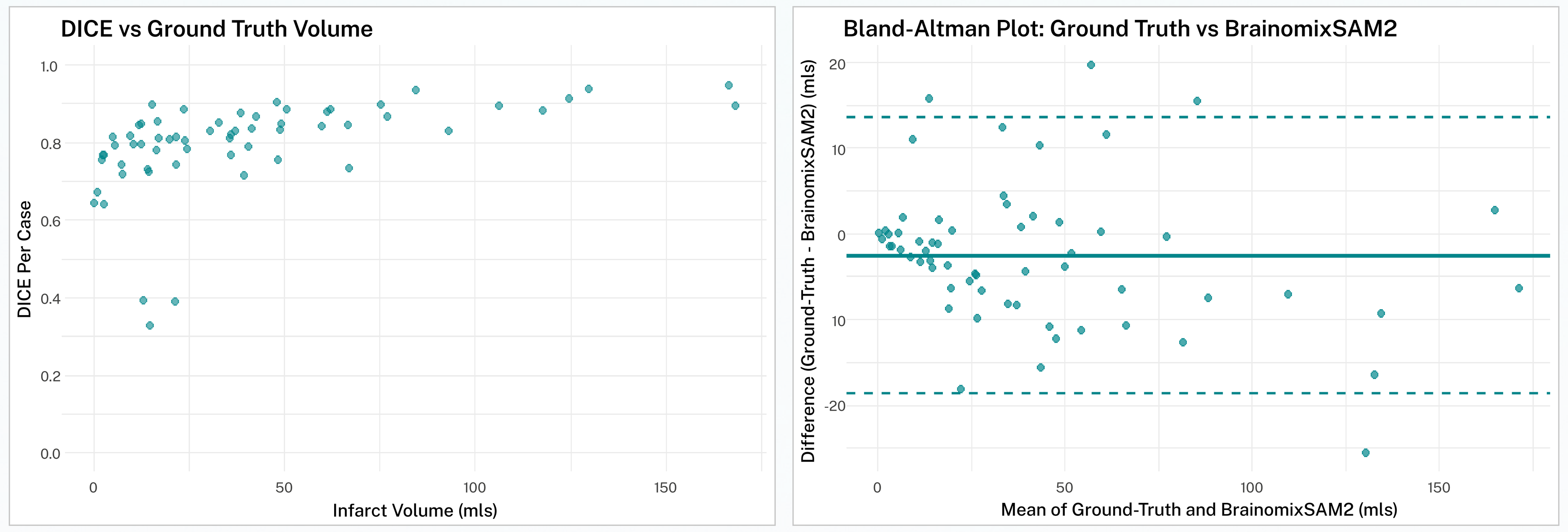

Figure 2. DICE values per case (left) and Bland-Altman plot of BrainomixSAM2 vs. ground truth (right) in Cohort 1.

Domain	Cohort 1 (WVU)	Cohort 2 (FHKV)	Cohort 2 DeepISLES
ICC	0.98 [0.96-0.99]	0.98 [0.97-0.99]	0.99 [0.97-0.99]
DICE mean (SD)	80% (10.2%)	71.9% (20.5%)	71.5% (20.4%)

Table 2. Volumentric agreement (ICC) and DICE of BrainomixSAM2 in Cohorts 1 and 2, and using DeepISLES in Cohort 2.

In Cohort 2, there were a mix of anterior (76%) and posterior circulation (24%) ischemic strokes.

When compared to ground-truth, BrainomixSAM2 performed similarly to DeepISLES (Table 2, Figure 3):

ICC = 0.98 [0.97-0.99] vs. 0.99 [0.97-0.99]

DICE = 71.9% [20.5] vs. 71.5% [20.4]; p=0.22 (Figure 3).

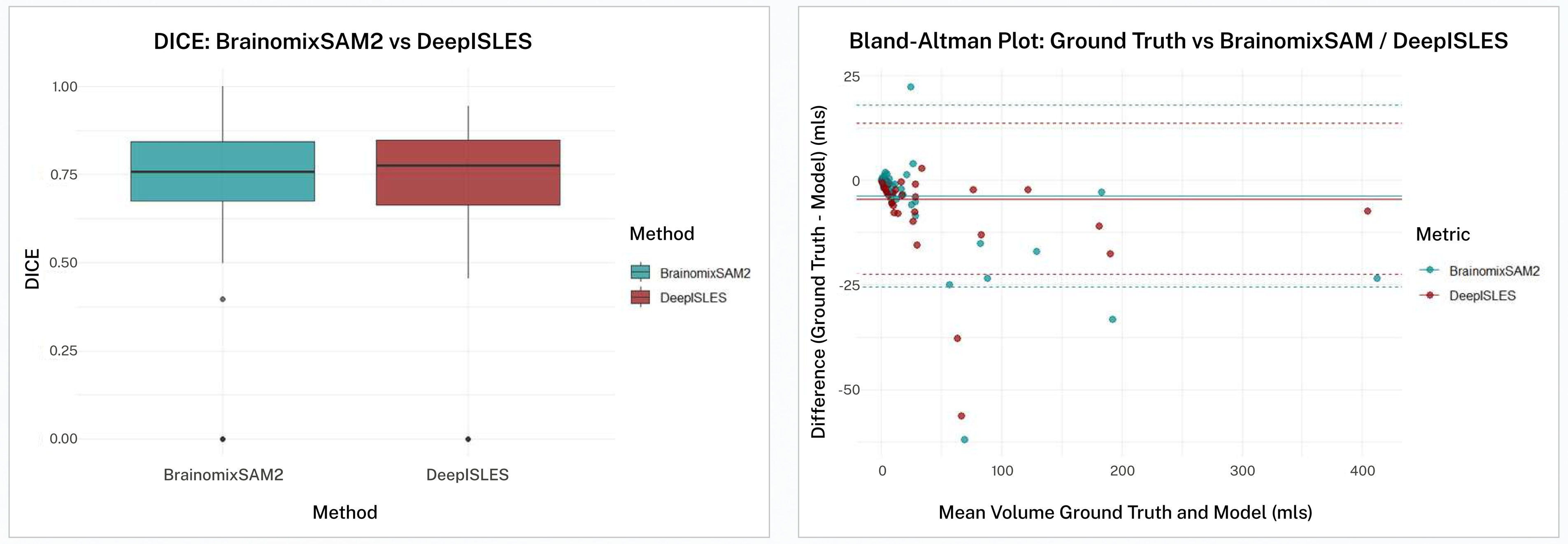

Figure 3: Agreement (DICE) between ground truth and two models (left) and Bland-Altman of BrainomixSAM2 vs ground truth (right) in Cohort 2.

Conclusions

- BrainomixSAM2 shows that state-of-the-art segmentation performance is attainable with minimal training data, enabling faster and more cost-effective development of segmentation tools in data-limited clinical settings.
- Wider validation and further head-to-head evaluations are warranted.

References

1. Warach, et al. Stroke. (2016) 47:1389–98. doi: 10.1161/STROKEAHA.115.012364
2. de la Rosa, et al. Nat Commun 16, 7357 (2025). https://doi.org/10.1038/s41467-025-62373-x

WestVirginiaUniversity BRAINOMIX