

Automatic Segmentation of Cortical Superficial Siderosis: A Proof of Concept

Yuan Cao¹, Antonia Zachariadis², Seraphine Zübel³, Daniel Behme³, Kersten Villringer⁴, Stefanie Schreiber^{2,5}, Katja Neumann², Jose Bernal^{1,6,7,8}
¹Department of Artificial Intelligence in Biomedical Engineering (AIBE), Faculty of Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany; ²Department of Neurology, University Hospital Magdeburg, Magdeburg, Germany; ³Clinic for Neuroradiology, University Hospital Magdeburg, Magdeburg, Germany; ⁴Centre of Stroke Research, Charité - Universitätsmedizin Berlin, Berlin, Germany; ⁵German Centre for Neurodegenerative Diseases (DZNE), Magdeburg, Germany; ⁶Faculty of Medicine, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany; ⁷Institute of Cognitive Neurology and Dementia Research (IKND), Otto-von-Guericke University Magdeburg, Magdeburg, Germany; ⁸Institute for Neuroscience and Cardiovascular Research, Row Fogo Centre for Research into Ageing and The Brain, Department of Neuroimaging Sciences, The University of Edinburgh, Edinburgh, UK

Background

Cortical superficial siderosis (cSS)

- predicts future intracerebral haemorrhage
- contraindicates AD disease-modifying therapies (AD DMTs)
- during AD DMTs must be flagged as a potential treatment-related adverse event

Despite highly relevant:

- No fully automated tool currently exists for cSS quantification
- Scarcity of manually labelled cSS limits conventional approaches

Take away

Synthetic data generation unlocks robust, fully automated cSS quantification without the need for manual expert annotation

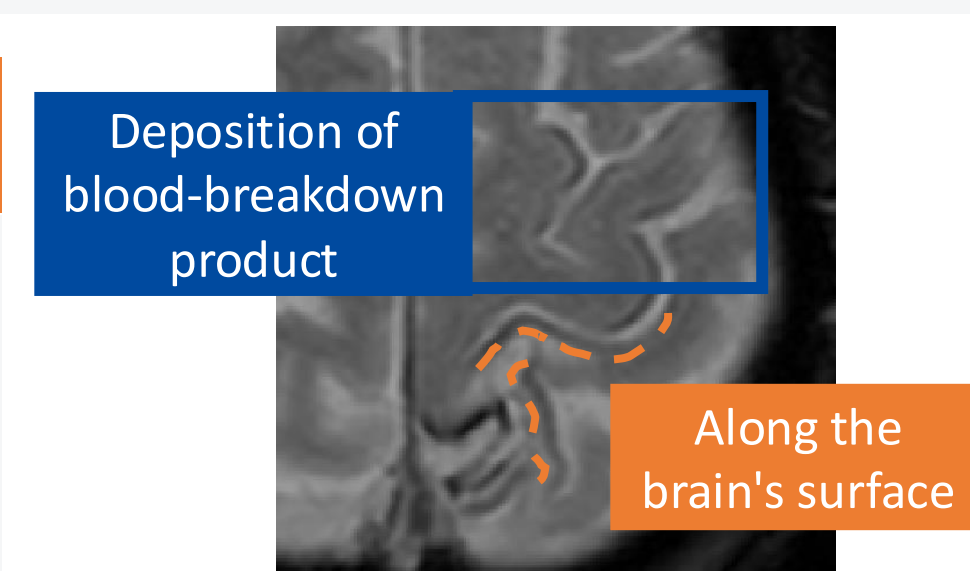


Figure 1. The clinical target: characteristic minute and irregular cSS lesions on MRI.

Materials and evaluation

Images: susceptibility- and T2*-weighted MRI scans from 10 individuals with cSS (6 female; median age 74.5 [IQR 67–78] years)

Ground truth: annotations by a trained investigator, reviewed by an experienced neuroradiologist

Metrics: area under precision-recall curve (AUPRC) and area under the receiver operating characteristic curve (AUROC)

Benchmarking: proposal compared against the 3D Frangi filter and random chance

Future work

Our framework represents a promising step towards fully automated cSS quantification. Further optimisation, large-scale, and multi-centre validation are needed, particularly in the context of AD DMTs.

Proposal

Hybrid framework: combination of stochastic synthetic image generation and deep learning

On-the-fly training: dynamic creation of image-label pairs through **domain randomisation**

Domain randomisation:

- Spatial augmentation: random non-linear deformation and resolution resampling
- CSS lesion simulation: variable 3D region growing anchored to multiple cortical seed points
- Intensity image synthesis: stochastic anatomical intensity assignment
- Artefact simulation: motion artefacts, Rician noise, and bias field inhomogeneity

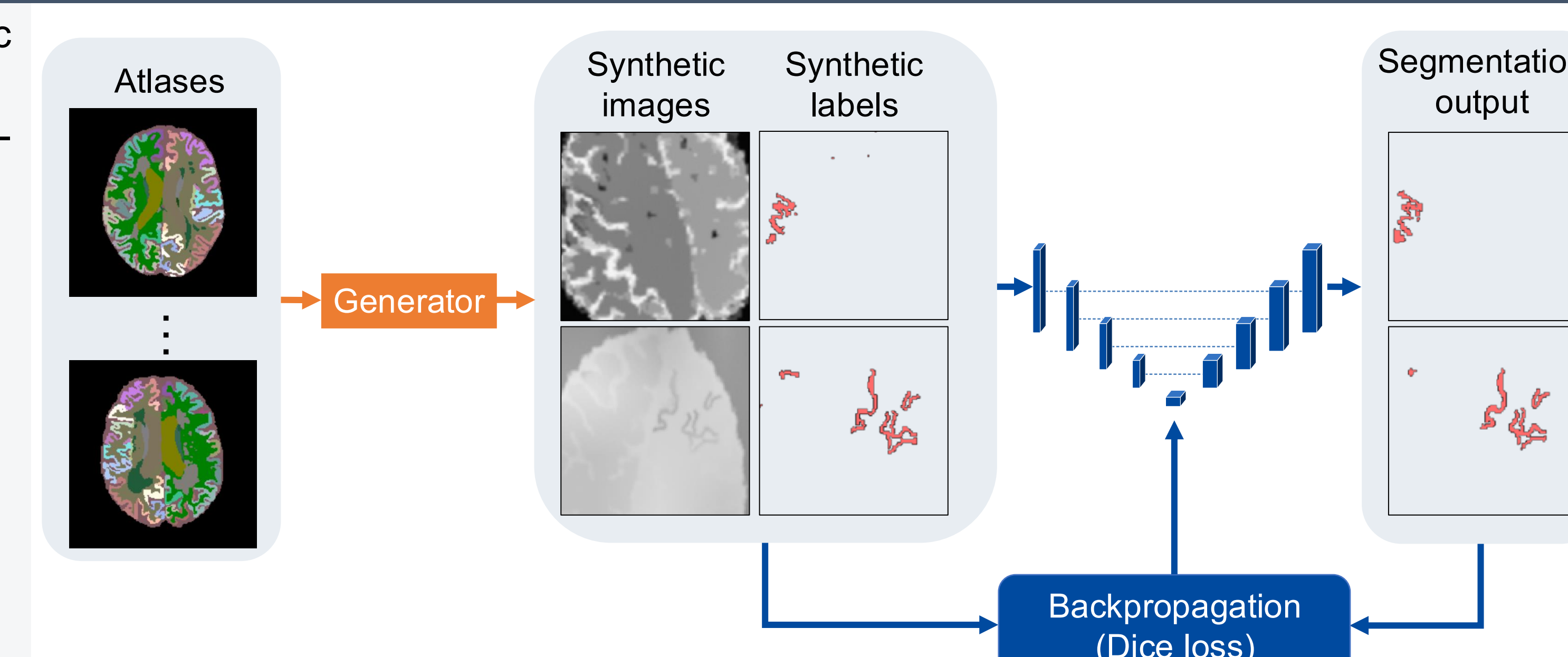


Figure 2. Proposed pipeline for synthetic image generation and CNN training.

Results

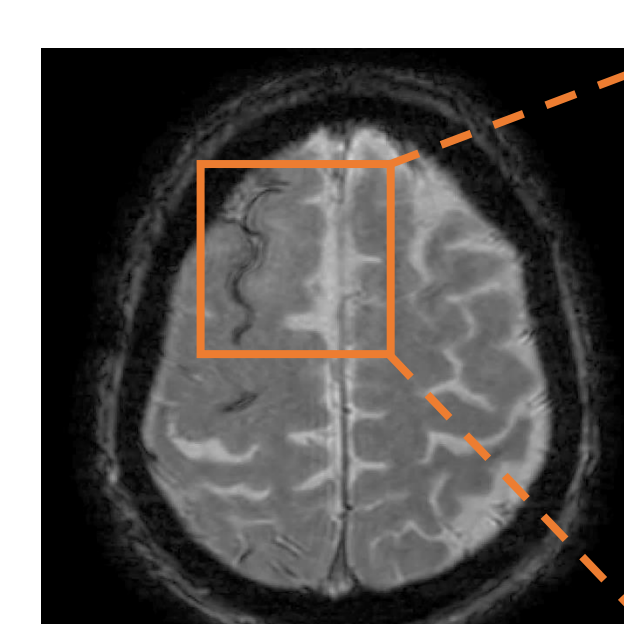
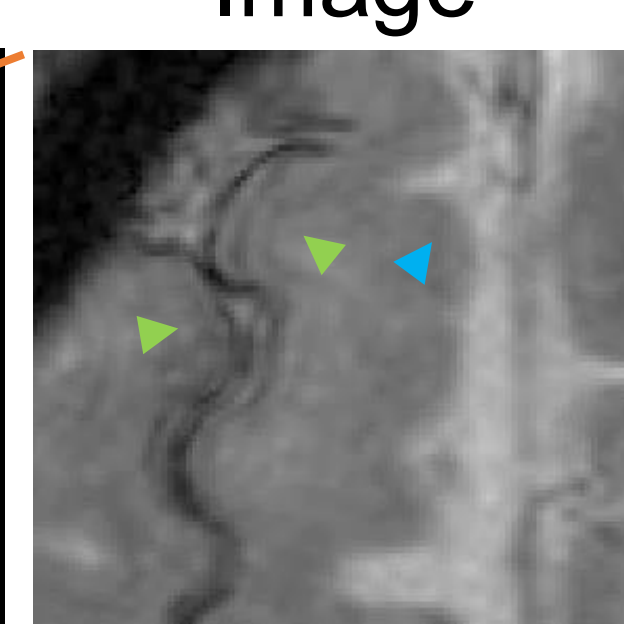
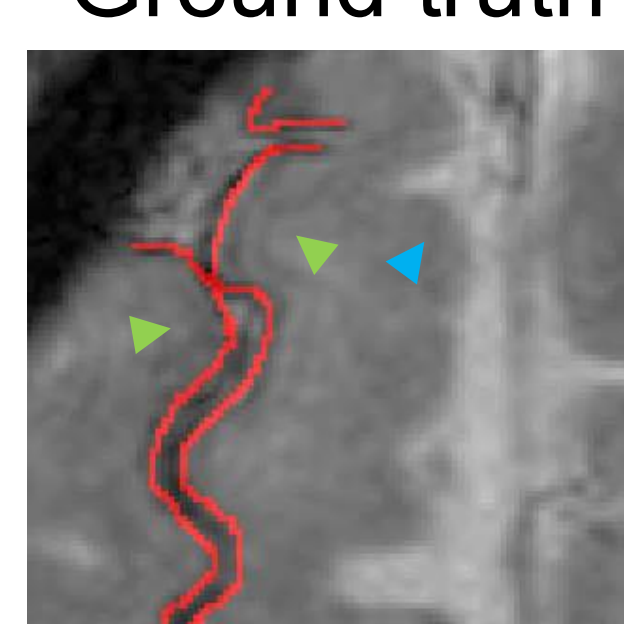
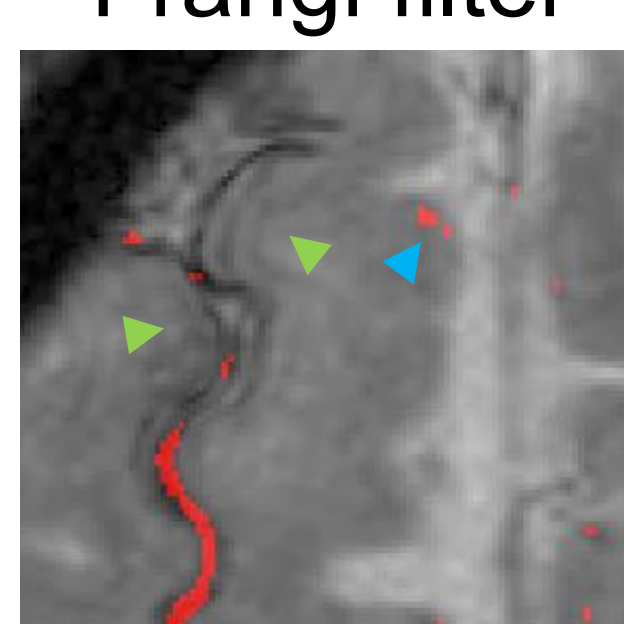
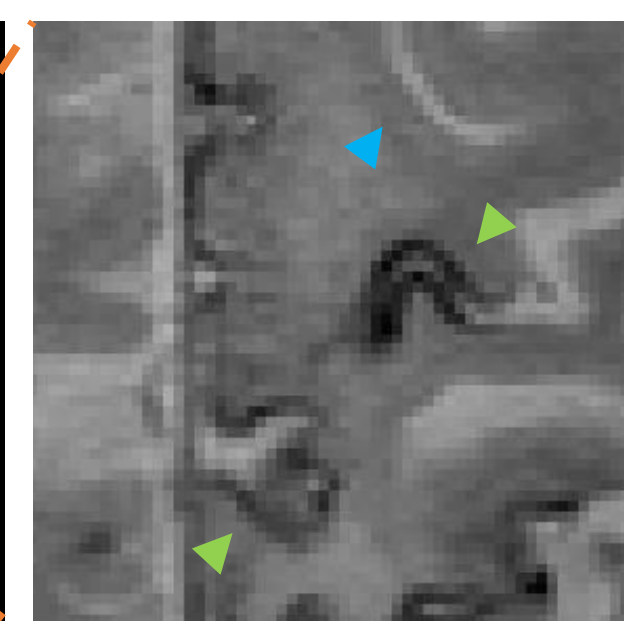
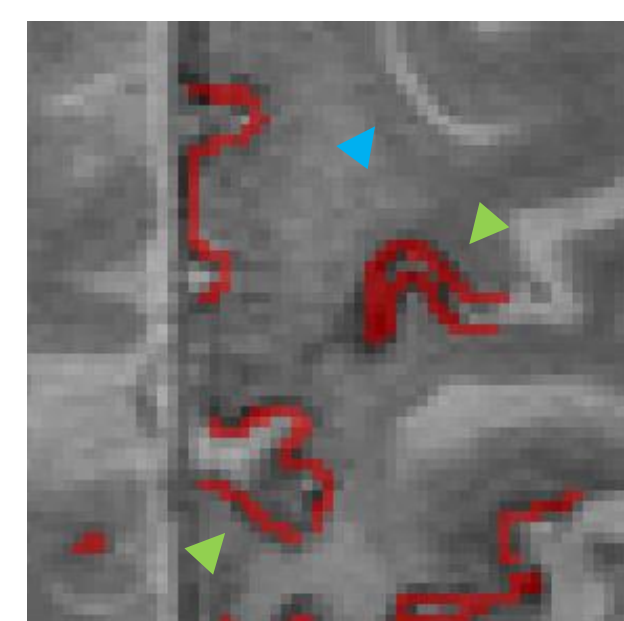
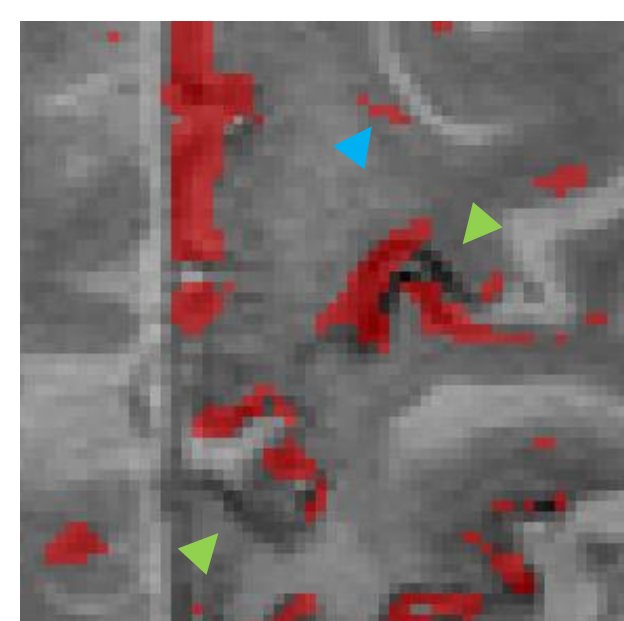
Reliable cSS Classification (AUROC): **proposed model: 0.906 [IQR 0.88–0.94]**

- Frangi baseline: 0.731 [0.68–0.81] | Random: 0.500

High precision & fewer false positives (AUPRC): **proposed model: 0.358 [IQR 0.15–0.37]**

- Frangi baseline: 0.083 [0.02–0.11] | Random: 0.010

(Takeaway: Over 4x improvement in AUPRC demonstrates robust detection in highly imbalanced data)

Image	Ground truth	Frangi filter	Our model
			
			

Performance comparison: proposed framework vs. 3D Frangi filter

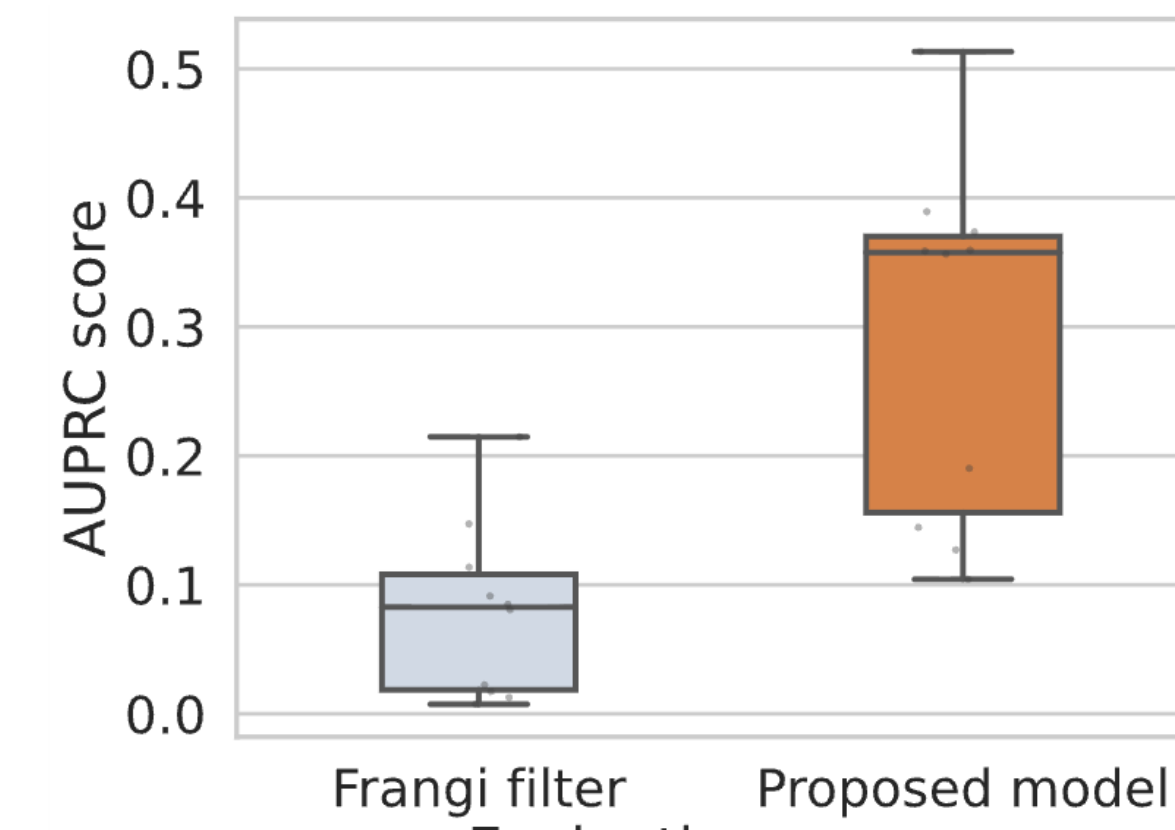


Figure 3. Visual comparison of segmentation performance. From left to right: original MRI, manual ground truth, Frangi filter output, and proposed model output. Our model shows higher sensitivity and specificity for cSS detection and outperforms the Frangi filter, which misses lesions (green arrows) and produces extensive false positives (blue arrows).

Figure 4. AUPRC comparison across 10 clinical test cases. Box plots show median, IQR, and range. The proposed framework achieved substantially higher median performance than the Frangi baseline.