DRIPS: Domain Randomisation for Image-based Perivascular spaces Segmentation

Luna Bitar[1,2], Mario Díaz[3,4], Roberto Duarte Coello[5,6], Maria d.C. Valdés-Hernández[5,6], Hendrik Mattern[1,7,8], Katja Neumann[9], Malte Pfister[9], Carolin Beck[9], Huy Trong Mai[9], Erelle Fuchs[10], Serena Tang[11,12], Duygu Tosun[13], Bianca Besteher[14,15,16], Tonia Rocktäschel[14], Philipp A. Reuken[17], Andreas Stallmach[17], Nils Opel[24,25], Christian Gaser[14,15,16,18], Martin Walter[14,15,16], Marc Dörner[1,19], Philipp Arndt[1,9], Daniel Behme[10], Christiane Piechowiak[20], Yves Lading[20], Patrick Müller[20], Rüdiger Braun-Dullaeus[20], Sven G. Meuth[21], for the Alzheimer's Disease Neuroimaging Initiative[+], Joanna M. Wardlaw[5,6], Stefanie Schreiber[1,9], Maria Trujillo[3], Emrah Düzel[1,22], Gabriel Ziegler[1,22], Jose Bernal[1,5,22,23,*]

Author affiliations:

[1]German Centre for Neurodegenerative Diseases (DZNE), Magdeburg, Germany
[2]Institute of Medical Systems Bioinformatics (IMSB), Centre for Molecular Neurobiology (ZMNH), University Medical Centre Hamburg-Eppendorf (UKE), Hamburg, Germany
[3]Multimedia and Computer Vision (MCV) Lab, Universidad del Valle, Cali, Colombia
[4]Perception and Intelligent Systems (PSI) Research Group, School of Electrical and Electronics Engineering, Universidad del Valle, Cali, Colombia
[5]Institute 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
[6]UK Dementia Research Institute Centre at the University of Edinburgh, Edinburgh, UK
[7]Center for Behavioural Brain Sciences (CBBS), Magdeburg, Germany
[8]Biomedical Magnetic Resonance, Otto-von-Guericke University, Magdeburg, Germany
[9]Department of Neurology, University Hospital Magdeburg, Magdeburg, Germany
[10]Clinic for Neuroradiology, University of Magdeburg, Magdeburg, Germany
[11]Department of Bioengineering, University of California San Francisco, San Francisco, CA, USA
[12]Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA
[13]Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA, USA
[14]Centre for Intervention and Research on adaptive and maladaptive brain Circuits underlying mental health (C-I-R-C), Halle-Jena-Magdeburg
[15]Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany
[16]German Centre for Mental Health (DZPG), partner site Halle-Jena-Magdeburg, Germany
[17]Department of Internal Medicine IV, Gastroenterology, Hepatology and Infectious diseases, Jena University Hospital
[18]Department of Neurology, Jena University Hospital, Jena, Germany
[19]Department of Consultation-Liaison-Psychiatry and Psychosomatic Medicine, University Hospital Zurich, Zurich, Switzerland
[20]Division of Cardiology and Angiology, University Hospital Magdeburg, Magdeburg, Germany
[21]Department of Neurology, Heinrich-Heine-University, Düsseldorf, Germany
[22]Institute of Cognitive Neurology and Dementia Research, Otto-von-Guericke University Magdeburg, Magdeburg, Germany
[23]Cerebrovascular Imaging and Research Lab, Department Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany
[24]Department of Psychiatry & Neuroscience, Campus Benjamin Franklin, Charité Universitätsmedizin Berlin, Germany
[25]German Centre for Mental Health (DZPG), partner site Berlin-Potsdam, Germany

[+]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf
[*]Corresponding author

Correspondence to:
Jose Bernal

Full address: Cerebrovascular Research and Imaging Lab, Department of Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander Universität Erlangen-Nürnberg (FAU), Nürnberger Strasse 74, 91052 Erlangen, Germany
E-Mail: jose.bernal@fau.de

Abbreviations

AUPRC	Area under the precision-recall curve
BG ROI	Basal ganglia region of interest
CSF	Cerebrospinal fluid
CSO ROI	Centrum semiovale region of interest
DSC	Dice similarity coefficient
DRIPS	Domain Randomisation for Image-based PVS Segmentation
FFT	Fast Fourier transformation
IFFT	Inverse Fast Fourier transformation
MRI	Magnetic resonance imaging
PVS	Perivascular spaces
ROC	Receiver operating characteristic curve
ROI	Region of interest
RORPO	Ranking the orientation responses of path operators
SNR	Signal-to-noise ratio
SVF	Stationary velocity field
TE	Echo time
TR	Repetition time
WMH	White matter hyperintensities

Abstract

Perivascular spaces (PVS) are emerging as sensitive imaging markers of brain health. Yet, accurate out-of-sample PVS segmentation remains challenging since existing methods are modality-specific, require dataset-specific tuning, or rely on manual labels for (re-)training. We propose DRIPS (Domain Randomisation for Image-based PVS Segmentation), a physics-inspired framework that integrates anatomical and shape priors with a physics-based image generation process to produce synthetic brain images and labels for on-the-fly deep learning model training. By introducing variability through resampling, geometric and intensity transformations, and simulated artefacts, it generalises well to real-world data. We evaluated DRIPS on MRI data from five cohorts spanning diverse health conditions (N = 165; T1w and T2w, isotropic and anisotropic imaging) and on a 3D ex vivo brain model reconstructed from histology. We evaluated its performance using the area under the precision–recall curve (AUPRC) and Dice similarity coefficient (DSC) against manual segmentations and compared it with classical and deep learning methods, including Frangi, RORPO, SHIVA-PVS, and nnU-Net. Only DRIPS and Frangi achieved AUPRC values above chance across all cohorts and the ex vivo model. On isotropic data, DRIPS and nnU-Net performed comparably, outperforming the next-best method by a median of +0.17–0.39 AUPRC and +0.09–0.26 DSC. On anisotropic data, DRIPS outperformed all competitors by a median of +0.13–0.22 AUPRC and +0.07–0.14 DSC. Importantly, its performance was not associated with white matter hyperintensity burden. DRIPS delivers accurate, fully automated PVS segmentation across heterogeneous imaging settings, reducing the need for manual labels, modality-specific models, or cohort-dependent tuning.

Keywords: Perivascular spaces; Segmentation; Domain Randomisation; Deep Learning; Magnetic Resonance Imaging

1 Introduction

Perivascular spaces (PVS) are anatomical passageways that surround arterioles, capillaries, and venules in the brain and an integral part of the neurovascular unit (Gouveia-Freitas and Bastos-Leite, 2021; Wardlaw et al., 2020). Collectively, PVS form a brain-wide network of conduits for cerebrospinal fluid (CSF) circulation (Hirschler et al., 2025; Wardlaw et al., 2020, 2009; Yamamoto et al., 2024), a function that underlies the clearance of metabolic and neurotoxic waste products (Braun and Iliff, 2020; Hablitz and Nedergaard, 2021; Iliff et al., 2014, 2012; Mestre et al., 2018; Rasmussen et al., 2018; Wardlaw et al., 2020). These spaces are dynamic, with the capacity to shrink and enlarge, at times reaching a calibre that renders them visible *in vivo* on magnetic resonance imaging (MRI) at standard clinical field strengths (1.5 T / 3 T) (Kern et al., 2023; Kim et al., 2023; Lynch et al., 2023; Menze et al., 2024; Vikner et al., 2022). PVS enlargement is pathological (Bown et al., 2022; Francis et al., 2019; Okar et al., 2023; Wardlaw et al., 2020) and is considered an early structural change of impaired cerebrovascular and brain waste clearance function (Francis et al., 2019; Ineichen et al., 2022; Okar et al., 2023; Schreiber et al., 2023; Wardlaw et al., 2020; Waymont et al., 2024).

The growing recognition of PVS as a non-invasive imaging marker of compromised brain health function has prompted the development and large-scale deployment of computational methods for their quantification and monitoring (Smith et al., 2019; Waymont et al., 2024). Broadly, the literature describes two strategies: classical and machine learning based methods (Waymont et al., 2024). Classical methods use the morphology and CSF-like signal of PVS to distinguish them from other brain structures and, when multimodal data are available, from other concomitant lesions, such as

white matter hyperintensities (WMH) and lacunar infarcts (Ballerini et al., 2020, 2018; Barisano et al., 2025; Barnes et al., 2022; Bernal et al., 2021b, 2020; Boespflug et al., 2018; Duarte Coello et al., 2024; Menze et al., 2024; Schwartz et al., 2019; Valdés Hernández et al., 2024). These well-established methods offer high sensitivity (Bernal et al., 2022)—a double-edged sword that often necessitates careful parameter tuning and post-processing to minimise false positives (Ballerini et al., 2018; Bernal et al., 2022, 2020; Valdés Hernández et al., 2024). Machine learning methods, on the other hand, leverage supervised learning (Boutinaud et al., 2021a; Cai et al., 2024; Chai et al., 2025; Dubost et al., 2019a, 2019b; González-Castro et al., 2016; Hou et al., 2017; Lian et al., 2018; Park et al., 2016; Pham et al., 2024; Rashid et al., 2023; Zhang et al., 2017). Within this category, deep learning has emerged as the most widely adopted method (Waymont et al., 2024). The main advantage of deep learning is that, with sufficiently large, diverse, and well-annotated datasets, models are able to overcome some of the limitations of classical strategies. Nonetheless, the scarcity of such datasets (Sudre et al., 2024) generally hinders their ability to generalise effectively to unseen datasets (Billot et al., 2023a; Chalcroft et al., 2025). This, in turn, constrains their broader applicability beyond their training sets.

Domain randomisation has emerged as an alternative to address this generalisation problem (Tobin et al., 2017). In contrast to data augmentation—which applies predefined spatial and intensity transformations to existing images—domain randomisation uses procedural image generation models, conditioned on segmentations with fully randomised parameters, to create synthetic datasets for training deep learning models. The diversity of training samples enables models trained with domain randomisation to learn domain-independent features that characterise target structures well. SynthSeg is an example of a successful method

taking advantage of domain randomisation (Billot et al., 2023a). It is a model that segments brain structures on real MRI acquired with diverse sequences and modalities without retraining, despite being trained exclusively on synthetic data. Since its introduction in the early 2020s, approaches leveraging domain randomisation have been successfully applied to a variety of tasks, including skull-stripping (Hoopes et al., 2022), segmentation of brain structures (Billot et al., 2023a, 2023b), WMH (Laso et al., 2023), and stroke lesions (Chalcroft et al., 2025), as well as super-resolution (Iglesias et al., 2023) and image registration (Hoffmann et al., 2024).

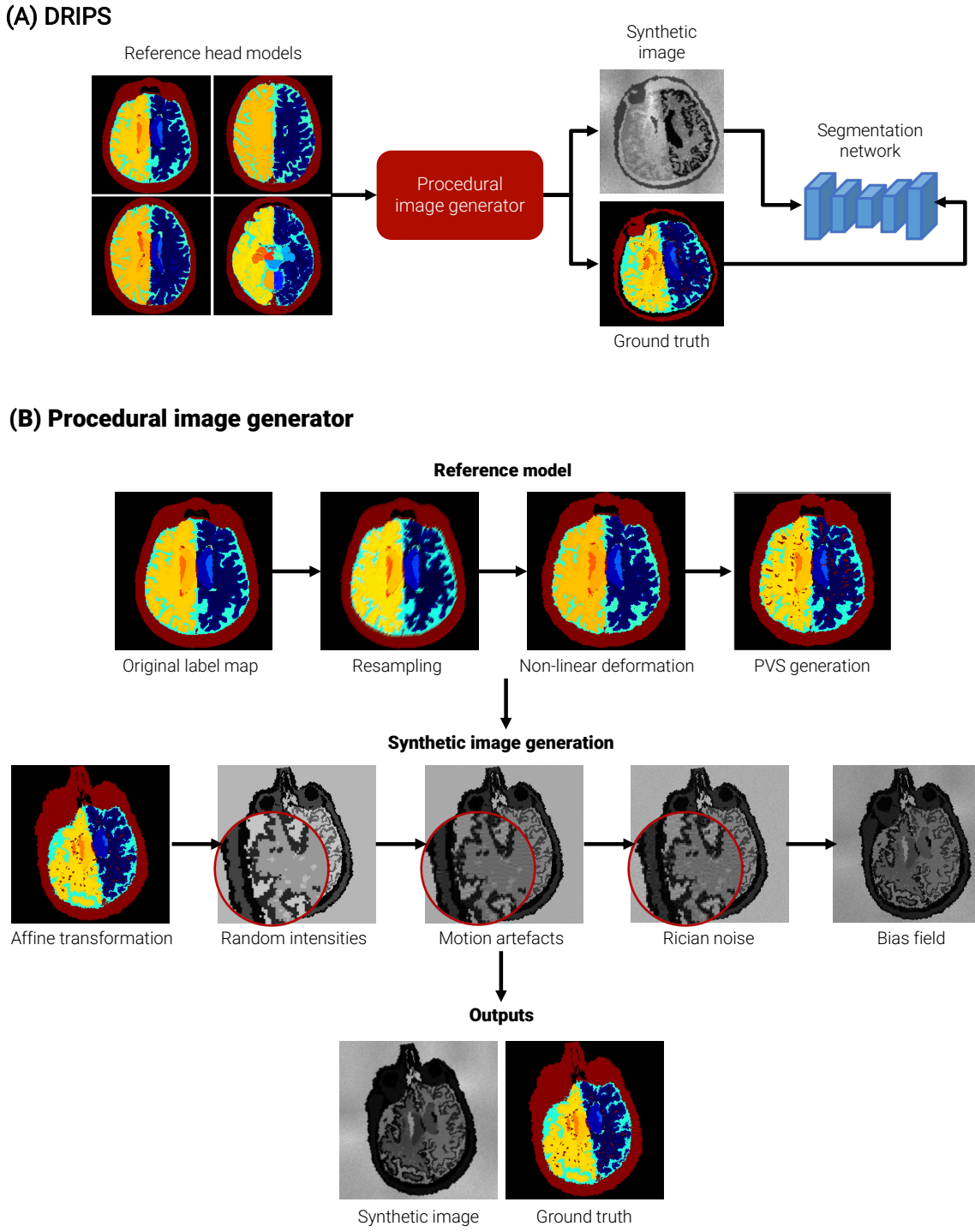
Realism in synthetic data generation is not essential; rather, it is crucial that generated data pose challenges comparable to real-world scenarios, enabling networks to learn robust and transferable features (Billot et al., 2023a). For synthetic PVS data generation, Bernal et al. (2022b) developed an open-source physics-inspired computational model that creates 3D digital reference objects containing PVS-like structures distributed throughout the brain. The generation process involved inserting randomly oriented tubular structures into a high-resolution head model, followed by k-space sampling, motion artefact simulation, and Rician noise corruption to produce low-resolution T2w-like images. Although it was originally conceived for method benchmarking, this computational model may serve as a basis for data generation and, when combined with domain randomisation, may facilitate training of deep learning algorithms with improved generalisability (Bernal et al., 2022).

Here, we introduce DRIPS (Domain Randomisation for Image-based PVS Segmentation), the first physics-inspired domain randomisation framework specifically developed for accurate out-of-sample PVS segmentation. DRIPS accurately segmented PVS in imaging data acquired with multiple imaging sequences and resolutions from patients with varying health conditions. It performed robustly across

all these settings and frequently surpassed both classical image-processing and deep learning methods.

2 DRIPS

DRIPS is a domain randomisation framework specifically designed for out-of-sample PVS segmentation (Figure 1). It integrates anatomical and shape priors of the human head and PVS with a physics-inspired procedural image generation process to create synthetic brain images and corresponding label maps. It then uses these synthetic datasets, generated on the fly, to train segmentation networks. By introducing variability through random resampling, geometric transformations, intensity sampling, and simulated MR artefacts, DRIPS produces models that achieve high segmentation accuracy and generalise effectively to real-world data. The following sections provide detailed descriptions of each step.



191

192 **Figure 1. Schematic of DRIPS.** (A) DRIPS is a domain randomisation framework that trains
193 segmentation networks for out-of-sample PVS segmentation. It combines anatomical and shape priors
194 of the human head and PVS with physics-inspired image generation to create synthetic brain images
195 and corresponding label maps containing PVS-like structures. It then trains segmentation networks—
196 here exemplified with a U-Net—on these synthetic image–label pairs. By exposing networks to broad
197 imaging variability during training, DRIPS achieves accurate PVS segmentation across diverse cohorts,
198 modalities, and acquisition settings. (B) Starting from anatomical head atlases with added synthetic,
199 tortuous PVS-like structures, DRIPS procedurally generates heterogeneous synthetic brain images
200 through random resampling, non-linear and affine transformations, intensity sampling, and typical MR
201 image corruptions and artefacts (motion artefacts, Rician noise, bias fields). Red-circled regions in the
202 procedural image generator correspond to zoomed-in views.

2.1 Reference model

2.1.1 Head model

We used 840 three-dimensional atlases derived from T1w and FLAIR scans of the ADNI database and CSVD Magdeburg cohorts as head models. Each atlas was a segmentation map with 1 mm³ resolution, in which every voxel was assigned to a specific class, including the lateral ventricles, white matter, WMH, cortical grey matter, cerebral white matter, cerebellar grey matter, brain stem, subcortical structures, or extracranial structures (Billot et al., 2023a). We used SynthSeg (Billot et al., 2023a) and LST-AI (Wiltgen et al., 2024) to obtain whole-brain parcellations and WMH masks, respectively. To introduce further anatomical variability, we applied random nonlinear diffeomorphic deformations to the original set of atlases. Specifically, we sampled a small stationary velocity field (SVF; 10 × 10 × 10 × 3) from a zero-mean Gaussian distribution, with standard deviation σ_{SVF} randomly drawn from a uniform distribution. The range of σ_{SVF} was set from 0 to 4 to allow for varying degrees of deformation. We then upsampled this field to full image resolution using trilinear interpolation to obtain a high-resolution SVF. Finally, we warped the original label map with this deformation field using nearest-neighbour interpolation to produce deformed brain atlases.

2.1.2 PVS model

We then added synthetic PVS-like structures to the generated head models. Although PVS are commonly described as tubular in clinical studies (Wardlaw et al., 2020), they do not conform to strictly Euclidean shapes and often exhibit tortuous geometries (Bernal et al., 2022). To capture this non-Euclidean morphology and have flexibility in

representing PVS-like structures, we modelled them as tortuous tubular structures using the following parametric equation:

$$x(t) = 0, y(t) = cos(\alpha t), z(t) = t,$$

where $t \sim \mathcal{U}(t_{low}, t_{high})$ and $\alpha \sim \mathcal{U}(\alpha_{low}, \alpha_{high})$ control the length and tortuosity of the generated PVS. Longer and more tortuous PVS structures are obtained by increasing t and decreasing α. We allowed t to vary between 2 and 10 voxels and α between 1/10 to 1/5. We placed these synthetic PVS in random locations within the white matter (normal-appearing and hyperintensities) and subcortical grey matter regions. We aligned each PVS towards the lateral ventricles, and to prevent clustering near the brain's centre, we used a stratified jittered sampling strategy.

2.2 Procedural synthetic image generation

We developed a procedural image generation model to create synthetic images for training the segmentation network. Using the head and PVS models, we generated synthetic images on the fly with fully randomised parameters, varying image intensities, contrasts, resolutions, and artefacts within each batch. The individual steps for synthetic data generation are illustrated in Figure 1 and described in detail below:

2.2.1 Resampling and voxel size variability

To enable the model to process scans acquired at different voxel sizes, we generated synthetic images and label maps with varying voxel sizes. We achieved this by resampling the input label maps to a randomly selected target voxel size. The target voxel size was randomly chosen on-the-fly during training, with each dimension varying between 0.5 mm and 4 mm to enable the processing of both research and

clinical scans. We resampled label maps using nearest-neighbour interpolation to preserve the original discrete voxel values.

2.2.2 Affine transformations

We applied random affine transformations to the label maps to increase anatomical variability and, at the same time, to preserve structural integrity. Rotation, scaling, shearing, and translation parameters were randomly selected, with all values sampled from uniform distributions (see (Billot et al., 2023a) for more information).

2.2.3 Random intensity generation

We assigned each anatomical structure a single random intensity, sampled from a standard uniform distribution $\mathcal{U}(0,1)$. This procedure varied structure intensities across images, eliminating consistent local patterns and forcing the model to rely on shape and spatial information for segmentation.

2.2.4 Motion artifacts

Motion artefacts are a common source of image degradation in MRI and can markedly affect the visibility and quantification of fine structures such as PVS (Bernal et al., 2022). Owing to their thin, elongated morphology, PVS are particularly susceptible to being mistaken for motion streaks, making artefact mitigation especially critical for this application. We simulated rotational motion during *k*-space acquisition using a composite *k*-space model (Bernal et al., 2022, 2021a; Shaw et al., 2020). We first rotated the original synthetic volume twice by random angles within [−15°, 15°] around random axes and compute the *k*-space of both the original and rotated volumes. We then generated a composite k-space by taking between 50% and 100% of the data from the original volume and replacing the remainder with data from the

rotated volumes along a randomly selected axis. Finally, we transformed the resulting composite *k*-space to image space to produce a motion-corrupted image. The level of displacement between consecutive frames and the time at which the motion occurs determines the severity and appearance of the motion artefacts in the resulting image (Figure 2).

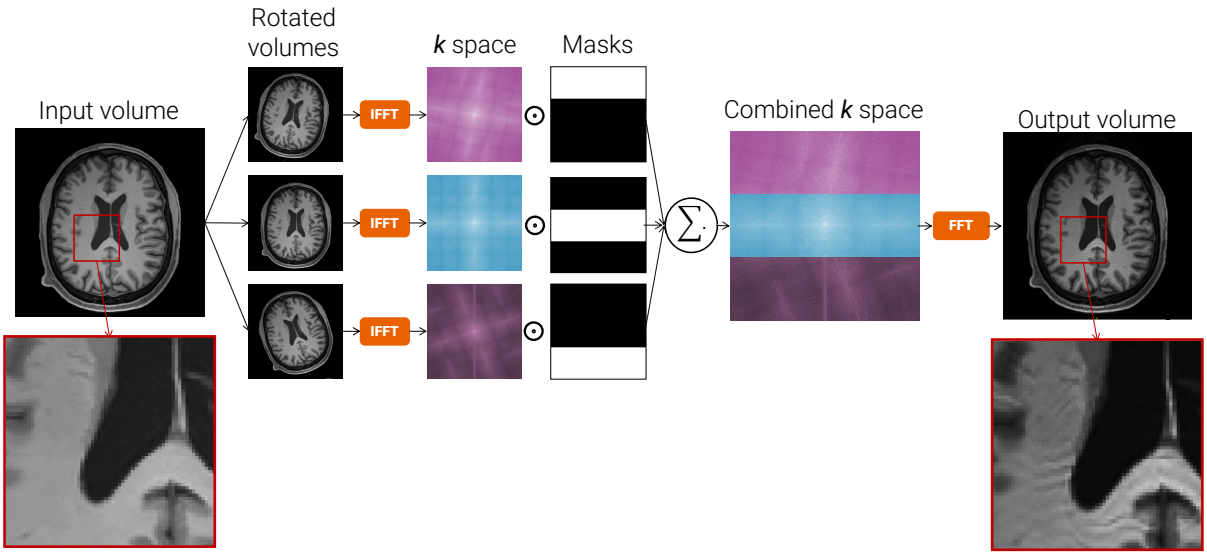


Figure 2. Simulation of motion artefacts in DRIPS. Motion was modelled in *k*-space by combining data from the original and randomly rotated volumes along a chosen axis. Varying the fraction of *k*-space segments taken from the original and rotated versions and the "timing" of motion yielded different levels of blurring and ghosting. Abbreviation: IFFT/FFT: (inverse) fast Fourier transformation

2.2.5 Rician noise

MRI data are inherently affected by noise originating during acquisition in *k*-space, where additive white Gaussian noise affects both the real and imaginary components of the complex signal (Gudbjartsson and Patz, 1995). Following transformation into the spatial domain and magnitude reconstruction, this noise takes a Rician distribution. To add Rician-distribution noise to the images, we added uncorrelated additive white Gaussian noise to the real and imaginary channels of the combined *k* space. The Gaussian noise standard deviation was computed as $\sigma_{noise} = \mu_{signal}/10^{SNR_{dB}/20}$, with

the SNR in decibels sampled from a uniform distribution $\mathcal{U}(SNR_{min},\ SNR_{max})$. We set SNR_{min} and SNR_{max} to 5 dB and 40 dB to simulate a broad spectrum of image noise.

2.2.6 Bias field inhomogeneity

We modelled bias field corruption to mimic MRI intensity inhomogeneities arising from B-field inhomogeneities and magnetic field variations. Following the approach in (Billot et al., 2023a), we sampled a 4 × 4 × 4 Gaussian random volume σ_B, upsampled it to image resolution for smooth variation, exponentiated to enforce positive multiplicative effects, and applied it to the synthetic image. We normalised intensities to [0, 1] and subjected the image to a random Gamma transformation to introduce additional non-linear signal variations.

2.2.7 Final training pair

Figure 3 presents four examples of synthetically generated images with their corresponding label maps, illustrating the variability in brain shape, structure, and intensity introduced by DRIPS. These DRIPS-generated pairs can be used as input and ground truth for training segmentation models.

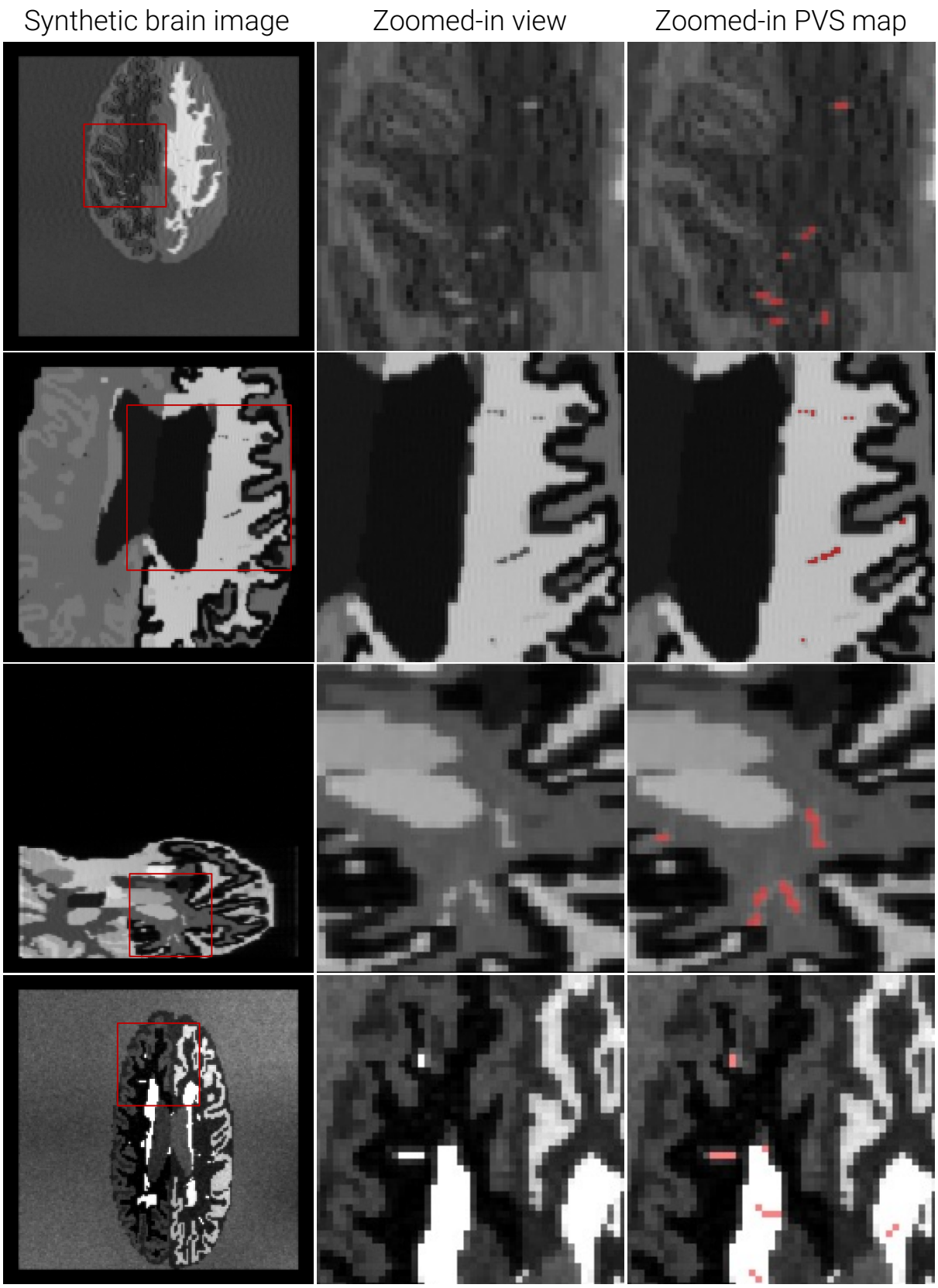


Figure 3. Synthetic brain images with corresponding ground truths, obtained using the proposed domain randomisation method. Note that synthetic images vary, among other aspects, in anatomy, orientation, intensity, PVS distribution, and levels of Rician noise, motion artefacts, and inhomogeneities.

2.3 DRIPS-based model training and testing

DRIPS provides a basis for training segmentation networks for out-of-sample PVS segmentation. In this work, we employed a 3D U-Net as the segmentation model, a well-established architecture capable of capturing both local and global spatial features critical for accurate segmentation. We note that the framework is architecture-agnostic and can be readily adapted to alternative architectures.

The 3D U-Net consisted of five encoding and five decoding levels. Each level had two convolutional layers with kernels of size 3 × 3 × 3, followed by a batch normalisation layer and max pooling or upsampling layers, depending on whether the level was part of the encoding or decoding part, respectively. All convolutional layers employed an exponential linear unit activation, except for the final layer, which had a softmax activation. The number of kernels per level doubled after each max pooling and halved after each upsampling layer. The first layer contained 24 feature maps. The network had skip connections to transfer feature maps from the encoding path to the decoding path.

2.3.1 Training on synthetic data

We trained the segmentation network in DRIPS for 50 epochs, each comprising 5000 batches of size 1, with each image–label pair generated on the fly by the procedural image generator. We used the Adam optimiser (learning rate of 10^{-4}) and a generalised Dice loss function for model optimisation. The generalised Dice loss function for multiple classes is given by (Milletari et al., 2016):

$$\text{Generalised Dice Loss}(\text{GT}, S) = 1 - \sum_{c\in\{0,1\}} \frac{2 \cdot \sum_{x,y,z} GT_c(x,y,z)\, S_c(x,y,z)}{\sum_{x,y,z} GT_c(x,y,z)^2 + S_c(x,y,z)^2},$$

where $c \in \{0,1\}$ denotes the considered classes (0: background, 1: PVS), and GT_c and S_c are the ground truth and soft probability map for class c, respectively. We implemented the segmentation model in Keras with a TensorFlow backend. Training took approximately twelve days on an NVIDIA A100 Tensor Core GPU.

2.3.2 Testing on real data

To evaluate the model on real data, we first normalised image intensities before feeding the scans into the network (Billot et al., 2023a). Inference took approximately ten seconds on an NVIDIA A100 Tensor Core GPU and 60 seconds on CPU.

Contrast agnosticism encourages models to prioritise shape over intensity. Though advantageous for generalisation, this technique makes models prone to detecting "tubular" structures regardless of whether their intensity profiles match those of PVS. For example, although PVS appear hypointense in T1w imaging, sections of the internal and external capsules—which are not hypointense in this modality—were sometimes flagged as potential PVS (non-zero response). To restrict detection to hypointense structures in T1w images and hyperintense structures in T2w images, we thus applied the Laplacian operator during post-processing, retaining regions with positive and negative Laplacian values, respectively.

3 Evaluation on real data

3.1 Cohorts and ground truth

We tested DRIPS on images and manual PVS segmentations from 165 participants from five cohorts: post-COVID Brain (PCB; N=42) (Besteher et al., 2022), EBBIVD (N=18), heart failure with preserved ejection fraction on cerebral microangiopathy

(HIM; N=39; DRKS00031583) (Müller et al., 2024), MagDeburger DrAinage-Reserve-Score (MD-DARS; N=6) (Neumann et al., 2022), and ADNI-3 (N=60). Further information can be found in Table 1. Ethical approval was granted by the Ethics Committees of the University Hospital Magdeburg for the EBBIVD, HIM, and MD-DARS cohorts, and by the Ethics Committee of Jena University Medical School for the PCB cohort, and by Institutional Review Boards of all participating centres for the ADNI-3 cohort. All participants provided written informed consent in accordance with the Declaration of Helsinki.

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership. The original goal of ADNI was to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease. The current goals include validating biomarkers for clinical trials, improving the generalizability of ADNI data by increasing diversity in the participant cohort, and to provide data concerning the diagnosis and progression of Alzheimer’s disease to the scientific community. For up-to-date information, see adni.loni.usc.edu

Table 1. Imaging protocols, and PVS and WMH burden across cohorts. The table summarises the imaging sequences, acquisition parameters, and scanner specifications used for manual PVS segmentation across the PCB, EBBIVD, HIM, MD-DARS, and ADNI cohorts. We also report PVS and WMH burden separately for the BG and CSO ROIs, presented as median counts, volumes, and fractional volumes, with interquartile ranges in brackets. Fractional volumes represent the volume of PVS within a region of interest relative to the volume of the region.

	PCB	EBBIVD	HIM	MD-DARS	ADNI
N	42	18	39	6	60
Clinical groups	Normal cognition post-COVID syndrome	Hypertensive arteriopathy Cerebral amyloid angiopathy	Individuals with heart failure with preserved ejection fraction	Individuals spanning the Alzheimer’s disease continuum	Normal cognition Mild cognitive impairment Alzheimer’s disease
Imaging					

Sequence for PVS segmentation	3D T1w MPRAGE	2D T2w TSE	2D T2w TSE	2D T2w TSE	3D T1w MPRAGE
Key parameters	TR = 2400 ms TE = 2.22 ms FA = 9°	TE = 73 ms TR = 6500 ms	TE = 73 ms TR = 6500 ms	TE = 73 ms TR = 6500 ms	TE = min full TR = 2300 ms TI = 900 ms FA = 9°
Voxel size [mm³]	0.8 × 0.8 × 0.8	0.5 × 0.5 × 2.0	0.5 × 0.5 × 2.0	0.5 × 0.5 × 2.0	1.0 × 1.0 × 1.0
Magnetic field strength [T]	3	3	3	3	3
Scanner model/vendor	Siemens Tim Trio (Siemens Healthineers, Erlangen, Germany)	Skyra (Siemens Healthineers, Erlangen, Germany)	Skyra (Siemens Healthineers, Erlangen, Germany)	Skyra (Siemens Healthineers, Erlangen, Germany)	Siemens, GE, and Philips (multi-vendor)
PVS burden					
BG PVS count	397 [311–506]	890 [643–1368]	656 [499–1053]	816 [526, 1033]	36 [17–67]
BG PVS volume [ml]	0.263 [0.212–0.342]	0.779 [0.592–1.255]	0.587 [0.435–0.928]	0.699 [0.459, 0.992]	0.051 [0.022–0.099]
Fractional BG PVS volumes (%)	0.587 [0.416–0.611]	1.865 [1.394–2.768]	1.172 [0.945–1.839]	1.585 [1.118–2.046]	0.117 [0.052–0.232]
CSO PVS count	1370 [782–2450]	7692 [5204–9056]	6379 [5012–8227]	6424 [3943, 7143]	510 [267, 858]
CSO PVS volume [ml]	0.976 [0.545–1.686]	8.471 [5.854–9.552]	6.853 [5.266–8.691]	6.9705 [4.275, 7.397]	0.900 [0.479, 1.4305]
Fractional CSO PVS volumes (%)	0.344 [0.192–0.652]	3.027 [1.852–3.801]	2.200 [1.678–2.822]	2.446 [1.468–2.579]	0.387 [0.216–0.598]
WMH burden					
BG WMH volume [ml]	No FLAIR imaging	No FLAIR imaging	0.218 [0.017–0.837]	0.293 [0.032–1.537]	0.408 [0.103–0.878]
CSO WMH volume [ml]	No FLAIR imaging	No FLAIR imaging	1.510 [0.243–7.573]	2.081 [0.155–9.819]	3.364 [1.374–11.382]

Under the guidance of experienced neuroradiologists, four medical residents and one neuroscientist segmented PVS manually using either Mango or ITK-SNAP. PVS segmentation was performed on T1w scans for PCB and ADNI, and on T2w scans for EBBIVD, HIM, and MD-DARS, following STRIVE criteria (Duering et al., 2023). The smallest available paint tool was used to manually delineate PVS across all axial slices throughout the entire brain. FLAIR sequences were taken into account, when available, to minimise the inclusion of WMH.

3.2 Evaluation metrics

We assessed PVS segmentation using voxel-wise and lesion-wise Dice similarity coefficients (DSC_{voxel} and DSC_{lesion}) and the area under the precision–recall curve (AUPRC). DSC_{voxel} quantifies spatial overlap between the predicted and ground-truth

388 binary maps within the ROI. DSC_{lesion} evaluates object-wise agreement after
389 connected-component labelling, measuring overlap between individual predicted and
390 reference PVS (e.g., one-inside-the-another criterion) (Maier-Hein et al., 2024).
391 AUPRC summarises segmentation performance across all possible thresholds. We
392 opted for precision–recall over receiver operating curves given the pronounced class
393 imbalance (Maier-Hein et al., 2024).

394 Since all of our evaluations are performed out-of-sample, discrepancies may arise
395 between how PVS were segmented in the training data and how they appear in an
396 unseen dataset (e.g. where PVS boundaries end). To mitigate this potential mismatch
397 and ensure a fair comparison across methods, we derived DSC values by thresholding
398 each output at the operating point on the precision–recall curve that maximised
399 segmentation performance. In practice, this corresponds to the threshold at which the
400 trade-off between sensitivity and precision yields the highest DSC_{voxel}.

401 **Generalisation criterion.** Although generalisation is inherently continuous, we
402 defined a practical criterion for it based on the expected performance under random
403 chance. Methods with performance overlapping with or below the chance-level
404 AUPRC were considered to have failed to generalise. The chance-level AUPRC value
405 is equivalent to the prevalence of the positive class within a given region of interest
406 (Saito and Rehmsmeier, 2015). In our case, this corresponds to the ratio between the
407 PVS volume in the ground truth and the total volume of the region of interest, i.e., the
408 fractional BG/CSO PVS volumes for each dataset (Table 1).

409 3.3 Regions of interest

410 We applied SynthSeg (Billot et al., 2023a) to T2w or T1w images to obtain
411 parcellations, which we then aggregated to generate masks for the basal ganglia and

the centrum semiovale region of interest (BG ROI and CSO ROI). The BG ROIs included the internal and external capsules, caudate, lentiform, and thalamic nuclei, while the CSO ROI covered the remaining supratentorial white matter. While these two ROIs do not precisely match anatomical structures, we adhered to the established nomenclature to maintain consistency with widely used visual rating methods in the field (Potter et al., 2015). We refined these masks to guarantee the exclusion of the ventricular atrium, choroid plexus, and posterior horns of the lateral ventricles via atlas registration (https://doi.org/10.7488/ds/1369). All regions of interest were kept identical across evaluated methods to ensure that observed differences arose from the methods themselves rather than from variations in ROI definition.

3.4 Competing methods

We compared DRIPS against four other methods: the Frangi filter (Frangi et al., 1998), RORPO (Ranking the Orientation Responses of Path Operators) (Merveille et al., 2018, 2014), SHIVA-PVS (Boutinaud et al., 2021b), and nnU-Net (Pham et al., 2024). Both SHIVA-PVS and nnU-Net were used as pretrained models, tested only in an out-of-sample setting, with no training performed on the cohorts used in this study.

Frangi and RORPO are classical strategies designed for enhancing tubular structures. Frangi relies on Hessian-based voxel analysis of shape features, while RORPO applies multi-orientation path opening to distinguish tubular from spherical structures. We employed a thoroughly validated pipeline developed at the University of Edinburgh that integrates both methods (more details can be found in (Ballerini et al., 2018; Bernal et al., 2022; Duarte Coello et al., 2024; Valdés Hernández et al., 2024); the step-by-step pipeline can be found in https://datashare.ed.ac.uk/handle/10283/8501). Unlike standard Frangi filter implementations, the pipeline modifies the Gaussian

filtering step to handle anisotropic voxel sizes. We employed the uint8 conversion step for RORPO provided in the pipeline and used parameter settings derived from earlier optimisation studies (Frangi: σ_{min} = 0.4, σ_{max} = 1.2, σ_{step} = 0.2, α = 0.5, β = 0.5, and c = 500; RORPO: scaleMin=1, nbscales = 9, factor=1.7, dilationSize=1) (Ballerini et al., 2018; Bernal et al., 2022; Duarte Coello et al., 2024). We did not use any other pre- or post-processing strategies.

SHIVA-PVS is a U-Net-based convolutional neural network designed to segment PVS on T1w MRI scans. It requires input images of size 160 × 214 × 176, with a 1 × 1 × 1 mm^3 isotropic resolution and intensity values normalised to [0,1]. Pre-processing involved rigid registration of all T1w images to MNI space, cropping to the required dimensions, and applying min–max normalisation. Following inference, the resulting segmentations were padded and transformed back to native space using the inverse rigid registration. The algorithm requires no parameter tuning and is publicly available on GitHub: https://github.com/pboutinaud/SHIVA_PVS.

nnU-Net is a convolutional neural network that extends the no-new-U-Net (nnU-Net) (Isensee et al., 2021) for PVS segmentation. Two modality-specific models were trained, one for T1w and one for T2w images. We refer to the models as nnU-Net (T1w) and nnU-Net (T2w), respectively. Models requires no manual parameter tuning, as all pre-processing, processing, and post-processing steps are automated and implemented in the publicly available codebase: https://github.com/wpham17/nnUNet-Perivascular-Spaces.

3.5 Generalisation to other imaging modalities

Since our aim was to assess the generalisation capabilities of models trained with DRIPS and the transferability of its learnt features, we also examined whether it could

extend to imaging modalities beyond MRI. As a proof of concept, we applied it to a 3D ex-vivo model of the human brain (Amunts et al., 2013). The Human Brain Histology dataset provides an ultrahigh-resolution 3D model of the human brain reconstructed from 7404 histological sections. For compatibility with our models and due to hardware constraints, the data were converted to greyscale and downsampled to 1 mm^3 resolution. Manual PVS segmentation was then performed on five axial slices in the BG ROI and five axial slices in the CSO ROI by an experienced image analyst using ITK-Snap with the smallest available drawing tools.

4 Results

4.1 Ablation study

We evaluated the impact of individual DRIPS modules by comparing a model incorporating them with one that did not. We conducted these assessments on real data. It should be noted that the real data were not modified in any way.

4.1.1 Effect of voxel size variation in DRIPS on model performance

To assess the effect of resampling and voxel size variation in DRIPS (Section 2.2.1), we compared the performance of two models: one with fixed and one with variable voxel sizes. We did this evaluation using data from the EBBIVD cohort (Figure 4). The use of variable voxel sizes led to a significant ($P<0.001$) and consistent improvement in segmentation performance. In the BG, median AUPRC improved from 0.325 [0.210–0.423] to 0.459 [0.358–0.541], DSC_{voxel} from 0.397 [0.276–0.459] to 0.499 [0.432–0.555], and DSC_{lesion} from 0.508 [0.322–0.552] to 0.635 [0.567–0.679]. In the CSO, median AUPRC rose from 0.256 [0.236–0.338] to 0.363 [0.304–0.439], DSC_{voxel}

from 0.323 [0.305–0.386] to 0.423 [0.348–0.463], and DSC_{lesion} from 0.435 [0.393–0.506] to 0.532 [0.461–0.570].

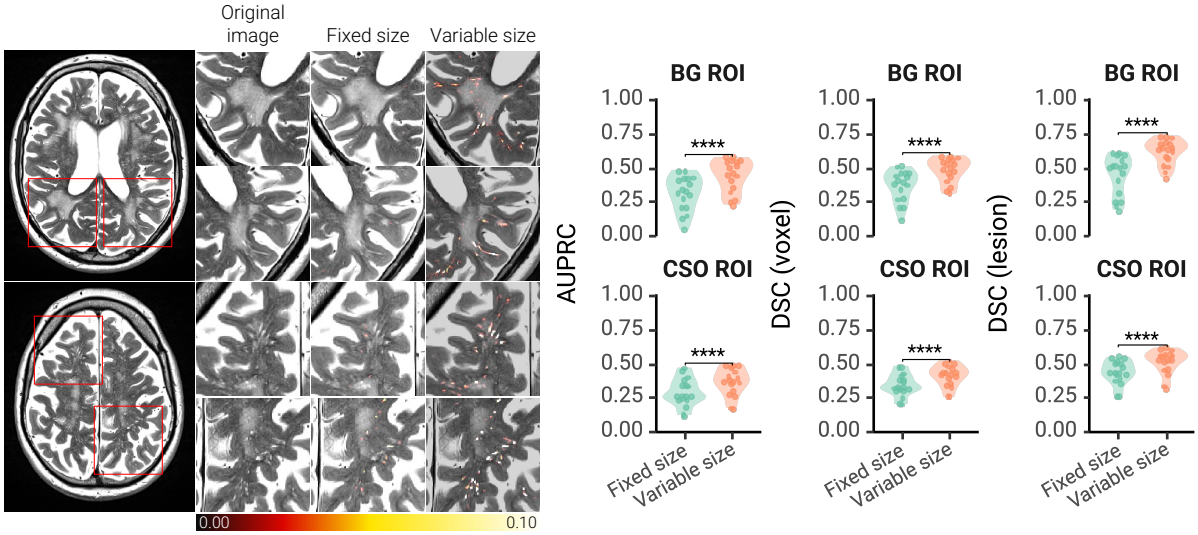


Figure 4. Allowing variable voxel sizes during image generation yielded better segmentation performance than using fixed voxel sizes. To illustrate this effect, we evaluated performance on EBBIVD, a cohort with highly anisotropic voxels (2.2 mm). Fewer PVS were segmented when models were trained with fixed voxel sizes compared to variable ones (left). With fixed voxel sizes, the model systematically missed multiple PVS in both normal-appearing white matter and WMH. For clarity, outputs were truncated to the 0.0–0.10 interval. At the cohort level, Wilcoxon signed-rank tests confirmed significant differences (P<0.0001) in both AUPRC and DSC across regions of interest (right).

4.1.2 Effect of motion simulation in DRIPS on model performance

We assessed how simulating motion in DRIPS influenced segmentation performance (Section 2.2.4), using data from the EBBIVD cohort. We compared the performance of two models: one incorporating motion simulation during training and one not (Figure 4). Motion simulation enhanced the model's ability to distinguish true PVS from motion-induced ghosting, as illustrated in case-level examples with and without visible motion. At the group level, where both motion-affected and unaffected images are present, performance in the BG was comparable between models: AUPRC 0.469 [0.336–0.545] without vs. 0.459 [0.358–0.541] with motion, DSC_{voxel} 0.509 [0.406–0.556] vs. 0.499 [0.432–0.555], and DSC_{lesion} 0.628 [0.517–0.701] vs. 0.635 [0.567–0.679]. In

the CSO, however, the motion-trained model was slightly more conservative voxel-wise, with AUPRC 0.390 [0.299–0.432] vs. 0.363 [0.304–0.439] (P=0.031), but achieved a higher DSC_{lesion}, 0.493 [0.430–0.510] vs. 0.532 [0.461–0.570] (P<0.001).

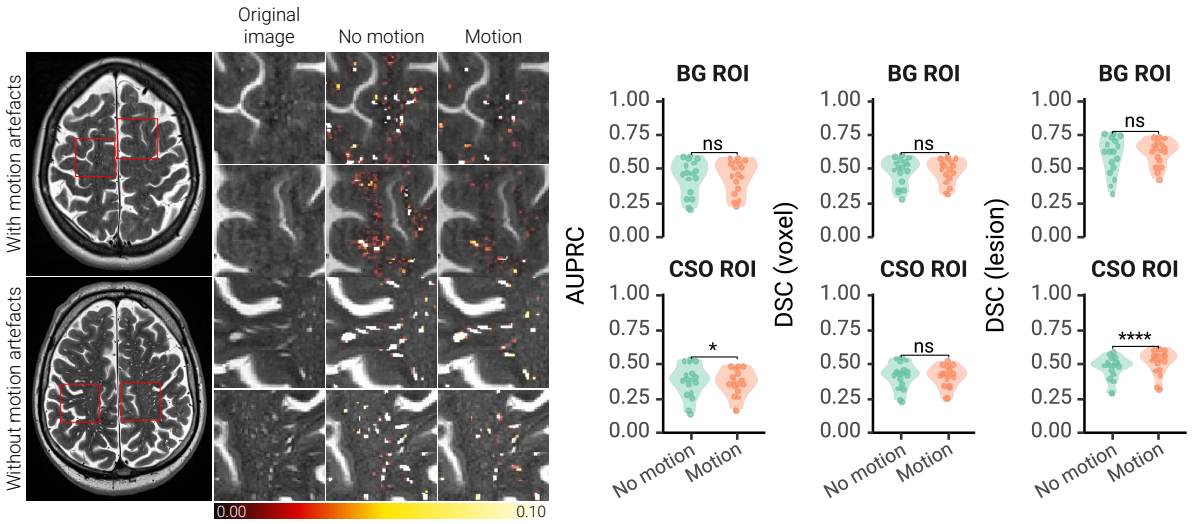


Figure 5. Incorporating motion artefacts during image generation results in a more conservative model with improved ability to separate motion artefacts from PVS. To illustrate this effect, we evaluated performance on EBBIVD and show probability maps for two cases: one with visible motion artefacts (left top row) and one without (left bottom row). When trained with motion artefacts, the model demonstrated improved ability to separate true PVS from motion-induced ghosting. In the motion case, the best DSC_{voxel} and DSC_{lesion} values without motion augmentation were 0.288 and 0.235, respectively, whereas training with motion artefacts increased them to 0.334 and 0.258. In cases without visible motion artefacts, both models yielded comparable results. At the group level, where images with and without motion artefacts are present, no differences were observed for BG PVS. For CSO PVS, however, the model trained with motion artefacts tended to be more conservative in detection but, once optimally thresholded, identified lesions better than the model trained without motion artefacts.

4.1.3 Effect of Laplacian constraint on model performance

We tested whether applying a Laplacian constraint to restrict detections to hypointense structures on T1w images and hyperintense structures on T2w images improved segmentations yielded by DRIPS (Section 2.3.2). We compared the performance of DRIPS with and without post-processing of its outputs using data from PCB (Figure 6). The Laplacian constraint significantly (P<0.0001) reduced the number of false positives, leading to overall improvements in PVS segmentation. In the BG ROI,

AUPRC increased from 0.416 [0.307–0.506] to 0.494 [0.409–0.567], DSC_{voxel} from 0.465 [0.373–0.528] to 0.509 [0.441–0.570], and DSC_{lesion} from 0.571 [0.406–0.635] to 0.590 [0.445–0.667]. In the CSO ROI, AUPRC improved from 0.454 [0.321–0.558] to 0.515 [0.380–0.611], DSC_{voxel} from 0.479 [0.397–0.555] to 0.522 [0.424–0.590], and DSC_{lesion} from 0.615 [0.477–0.694] to 0.635 [0.471–0.692].

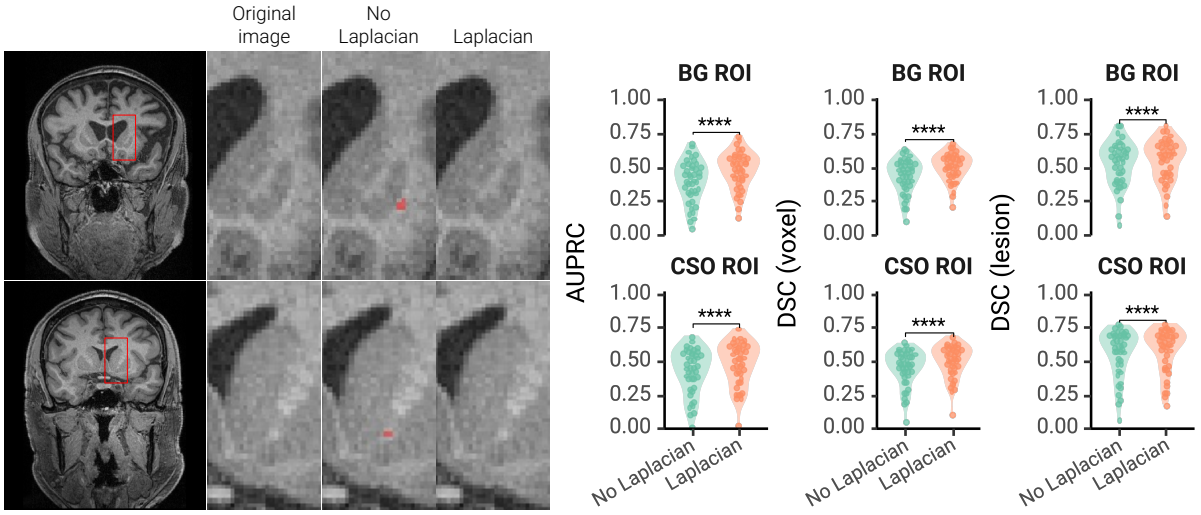


Figure 6. With the Laplacian constraint, only tubular structures matching the expected intensity profiles are detected. We assessed the effect of the Laplacian constraint on segmentation performance using the PCB cohort, which consists of T1w images where PVS appear hypointense. Because DRIPS is contrast-agnostic, it disregards intensity information. As a result, models trained with DRIPS may identify tubular structures regardless of whether they are hypo- or hyperintense, even though PVS present with a specific intensity profile. Such cases occurred most frequently within the internal and external capsules. By retaining regions with positive Laplacian values in T1w and negative values in T2w images, the Laplacian constraint reduced false positives and improved the quality of PVS segmentation overall.

4.2 Out-of-sample PVS segmentation

We compared DRIPS against Frangi, RORPO, SHIVA-PVS, and nnU-Net (Table 2). Below, we focus on two aspects: whether methods generalise out-of-sample and, if so, how they compare with one another.

4.2.1 Generalisation

DRIPS generalised across all cohorts, independent of voxel anisotropy or image modality (T1w/T2w). The other method that ran successfully on all datasets was the Frangi filter. RORPO was not able to segment any PVS on ADNI. The generalisation of SHIVA-PVS and the nnU-Net models was limited to their respective training modalities, with AUPRC values overlapping with or falling below those of a random classifier when applied to unseen modalities.

4.2.2 Segmentation performance

In cohorts with isotropic T1w imaging (PCB and ADNI), DRIPS and nnU-Net (T1w) were the top performers. Compared to the third-best method, they showed median improvements of +0.17–0.39 in AUPRC, +0.09–0.26 in DSC_{voxel}, and +0.14–0.25 in DSC_{lesion}.

In PCB, DRIPS and nnU-Net (T1w) performed similarly in the BG ROI, with no significant differences across AUPRC (0.504 [0.406–0.568] vs 0.445 [0.421–0.535], P=0.644), DSC_{voxel} (0.512 [0.440–0.570] vs 0.516 [0.483–0.571], P=0.163) or DSC_{lesion} (0.589 [0.442–0.667] vs 0.553 [0.473–0.619], P=0.396). In the CSO ROI, however, nnU-Net (T1w) achieved significantly higher scores, leaving DRIPS as the second-best performer (AUPRC: 0.515 [0.369–0.608] vs 0.549 [0.450–0.636], P<0.001; DSC_{voxel}: 0.521 [0.418–0.587] vs 0.543 [0.484–0.610], P<0.001; DSC_{lesion}: 0.630 [0.460–0.690] vs 0.649 [0.581–0.736], P<0.001).

In ADNI, DRIPS significantly outperformed nnU-Net (T1w) in the BG ROI across all metrics (AUPRC: 0.569 [0.417–0.662] vs 0.474 [0.340–0.535], P=0.003; DSC_{voxel}: 0.564 [0.460–0.651] vs 0.322 [0.175–0.421], P<0.001; DSC_{lesion}: 0.685 [0.571–0.823] vs 0.517 [0.437–0.588], P<0.001). In the CSO ROI, results were more balanced:

DRIPS had higher sensitivity, detecting more PVS (DSC_{lesion}: 0.680 [0.616–0.720] vs 0.593 [0.534–0.667], P<0.001), while nnU-Net (T1w) provided slightly more precise delineation (DSC_{voxel}: 0.636 [0.575–0.657] vs 0.652 [0.563–0.690], P=0.040).

In cohorts with anisotropic T2w imaging (EBBIVD, HIM, and MD-DARS), DRIPS had the best performance, followed generally by Frangi, RORPO, and the nnU-Net (T2w) in that order. The performance gap between DRIPS and the second-best method was most pronounced in the CSO ROI, with median gains of +0.17-0.22 in AUPRC, +0.12-0.14 in DSC_{voxel}, and +0.06-0.09 in DSC_{lesion}. In the BG ROI, the gap was smaller yet consistent, with AUPRC gains of +0.13-0.17, DSC_{voxel} gains of +0.07-0.12, and DSC_{lesion} gains of +0.03-0.09.

SHIVA-PVS typically underperformed (AUPRC<0.10; DSC<0.15), with the only exception in CSO PVS segmentation in ADNI, where it placed third above Frangi.

Table 2. Out-of-sample PVS segmentation performance across five cohorts. We assessed PVS segmentation in the basal ganglia (BG ROI) and the centrum semiovale (CSO ROI) using voxel- and lesion-wise Dice similarity coefficients (DSC_{voxel} and DSC_{lesion}) and the area under the precision–recall curve (AUPRC). We report medians with interquartile ranges, and "NA" where no PVS could be segmented. We identified the best-performing methods across regions and cohorts using the Wilcoxon signed-rank test and highlighted them in bold. Following the generalisability criterion described in Section 3.2, we marked with "NG" all AUPRC values that overlapped with or fell below the expected performance of a random classifier.

	Method	Metric	PCB (N=42)	EBBIVD (N=18)	HIM (N=39)	MD-DARS (N=6)	ADNI (N=60)
BG ROI	Frangi	AUPRC	0.263 [0.213–0.321]	0.334 [0.272–0.348]	0.331 [0.269–0.360]	0.294 [0.253–0.314]	0.198 [0.101–0.324]
		DSC_{voxel}	0.331 [0.292–0.391]	0.425 [0.379–0.439]	0.407 [0.372–0.443]	0.388 [0.345–0.404]	0.342 [0.247–0.445]
		DSC_{lesion}	0.447 [0.390–0.493]	0.592 [0.524–0.629]	0.556 [0.516–0.639]	0.556 [0.529–0.596]	0.467 [0.323–0.583]
	RORPO	AUPRC	0.203 [0.168–0.287]	0.324 [0.234–0.351]	0.308 [0.214–0.361]	0.264 [0.219–0.336]	NA
		DSC_{voxel}	0.303 [0.242–0.374]	0.417 [0.376–0.453]	0.415 [0.356–0.451]	0.381 [0.347–0.421]	NA
		DSC_{lesion}	0.252 [0.207–0.326]	0.424 [0.376–0.541]	0.432 [0.363–0.493]	0.421 [0.365–0.450]	NA
	SHIVA-PVS	AUPRC	0.060 [0.049–0.073]	0.018 [0.015–0.025] (NG)	0.013 [0.010–0.018] (NG)	0.013 [0.011–0.019] (NG)	0.101 [0.037–0.158]
		DSC_{voxel}	0.134 [0.114–0.154]	0.040 [0.032–0.048]	0.032 [0.028–0.039]	0.036 [0.033–0.042]	0.214 [0.108–0.285]
		DSC_{lesion}	0.242 [0.193–0.275]	0.039 [0.023–0.094]	0.078 [0.039–0.103]	0.094 [0.062–0.108]	0.426 [0.330–0.520]
	nn-Unet (T1w)	AUPRC	**0.445 [0.421–0.535]**	0.009 [0.006–0.012] (NG)	0.006 [0.005–0.008] (NG)	0.007 [0.004–0.008] (NG)	0.474 [0.340–0.535]
		DSC_{voxel}	**0.516 [0.483–0.571]**	0.031 [0.024–0.045]	0.021 [0.017–0.032]	0.026 [0.018–0.033]	0.322 [0.175–0.421]
		DSC_{lesion}	**0.553 [0.473–0.619]**	0.028 [0.023–0.035]	0.032 [0.027–0.041]	0.035 [0.025–0.044]	0.517 [0.437–0.588]
	nn-Unet (T2w)	AUPRC	0.009 [0.006–0.011] (NG)	0.100 [0.057–0.141]	0.107 [0.082–0.136]	0.087 [0.078–0.106]	0.002 [0.001–0.006] (NG)
		DSC_{voxel}	0.027 [0.022–0.037]	0.208 [0.121–0.251]	0.225 [0.173–0.254]	0.216 [0.180–0.220]	0.013 [0.005–0.027]
		DSC_{lesion}	0.050 [0.042–0.065]	0.278 [0.184–0.344]	0.282 [0.217–0.343]	0.283 [0.202–0.365]	0.041 [0.023–0.090]
	DRIPS	AUPRC	**0.504 [0.406–0.568]**	**0.459 [0.358–0.541]**	**0.503 [0.426–0.553]**	**0.424 [0.396–0.447]**	**0.569 [0.417–0.662]**
		DSC_{voxel}	**0.512 [0.440–0.570]**	**0.499 [0.432–0.555]**	**0.532 [0.488–0.567]**	**0.475 [0.471–0.478]**	**0.564 [0.460–0.651]**
		DSC_{lesion}	**0.589 [0.442–0.667]**	**0.635 [0.567–0.679]**	**0.646 [0.618–0.697]**	**0.581 [0.571–0.478]**	**0.685 [0.571–0.823]**
CSO ROI	Frangi	AUPRC	0.170 [0.119–0.276]	0.185 [0.129–0.253]	0.192 [0.156–0.250]	0.150 [0.121–0.180]	0.272 [0.157–0.362]
		DSC_{voxel}	0.311 [0.249–0.409]	0.286 [0.249–0.341]	0.314 [0.268–0.340]	0.260 [0.238–0.289]	0.381 [0.269–0.447]
		DSC_{lesion}	0.429 [0.333–0.546]	0.472 [0.383–0.515]	0.451 [0.394–0.514]	0.391 [0.340–0.440]	0.426 [0.362–0.509]
	RORPO	AUPRC	0.181 [0.103–0.309]	0.196 [0.119–0.244]	0.181 [0.127–0.263]	0.148 [0.112–0.179]	NA
		DSC_{voxel}	0.311 [0.219–0.429]	0.301 [0.220–0.354]	0.290 [0.227–0.360]	0.253 [0.208–0.287]	NA
		DSC_{lesion}	0.376 [0.252–0.484]	0.344 [0.267–0.408]	0.309 [0.264–0.388]	0.298 [0.217–0.325]	NA
	SHIVA-PVS	AUPRC	0.011 [0.005–0.022]	0.026 [0.018–0.032] (NG)	0.013 [0.010–0.018] (NG)	0.019 [0.013–0.020] (NG)	0.421 [0.319–0.481]
		DSC_{voxel}	0.035 [0.019–0.056]	0.049 [0.035–0.060]	0.039 [0.029–0.049]	0.038 [0.025–0.042]	0.469 [0.417–0.521]
		DSC_{lesion}	0.063 [0.038–0.110]	0.014 [0.001, 0.031]	0.014 [0.005, 0.023]	0.010 [0.005–0.018]	0.564 [0.487–0.652]
	nn-Unet (T1w)	AUPRC	**0.549 [0.450–0.636]**	0.013 [0.008–0.016] (NG)	0.010 [0.007–0.012] (NG)	0.009 [0.007–0.011] (NG)	**0.632 [0.564–0.697]**
		DSC_{voxel}	**0.543 [0.484–0.610]**	0.047 [0.029–0.059]	0.037 [0.028–0.047]	0.037 [0.025–0.040]	**0.652 [0.563–0.690]**
		DSC_{lesion}	**0.649 [0.581–0.736]**	0.010 [0.009, 0.014]	0.010 [0.008–0.015]	0.013 [0.007–0.015]	0.593 [0.534–0.667]
	nn-Unet (T2w)	AUPRC	0.003 [0.017–0.005] (NG)	0.140 [0.106–0.177]	0.160 [0.124–0.206]	0.116 [0.105–0.136]	0.007 [0.004–0.010] (NG)
		DSC_{voxel}	0.007 [0.005–0.012]	0.216 [0.181–0.246]	0.256 [0.210–0.302]	0.209 [0.200–0.224]	0.019 [0.013–0.029]
		DSC_{lesion}	0.014 [0.010–0.019]	0.296 [0.263–0.327]	0.348 [0.293–0.399]	0.284 [0.232–0.330]	0.046 [0.028–0.059]
	DRIPS	AUPRC	0.515 [0.369–0.608]	**0.363 [0.304–0.439]**	**0.409 [0.336–0.464]**	**0.323 [0.286–0.358]**	**0.665 [0.569–0.698]**
		DSC_{voxel}	0.521 [0.418–0.587]	**0.423 [0.348–0.463]**	**0.452 [0.399–0.482]**	**0.387 [0.352–0.412]**	0.636 [0.575–0.657]
		DSC_{lesion}	0.630 [0.460–0.690]	**0.532 [0.461–0.570]**	**0.545 [0.498–0.616]**	**0.467 [0.402–0.521]**	**0.680 [0.616–0.720]**

4.2.3 WMH and PVS segmentation

WMH can impair accurate PVS segmentation. To assess this effect, we examined the relationship between WMH volume and AUPRC using Spearman correlations (Figure 7). For this secondary analysis, we used data from ADNI, HIM, and MD-DARS (N = 105), all of which had WMH segmentations. We combined HIM and MD-DARS due to the small sample size of MD-DARS, which could otherwise lead to spurious correlations. The analysis focused on models that successfully generalised.

In the BG ROI, AUPRC values obtained by the Frangi filter in both T1w and T2w imaging, by RORPO in T2w imaging, and by SHIVA-PVS in T1w imaging increased with greater WMH volume ($P \leq 0.01$). The underlying reasons differed between SHIVA-PVS and the Frangi filter or RORPO. SHIVA-PVS performed better in cases with more visible BG PVS (Figure 8), which occurred more frequently in patients with greater WMH burden (Spearman correlation between BG PVS volume and BG WMH volume in ADNI: $\rho = 0.396$, $P = 0.002$). Both the Frangi filter and RORPO produced non-zero responses within WMH. In patients with higher BG WMH burden, many WMH voxels were adjacent or around to true PVS, causing false detections to overlap with true positives and artificially inflating recall rates and AUPRC (Figure 8). Unlike Frangi, RORPO, and SHIVA-PVS, AUPRC of the nnU-Net models (T1w and T2w) and DRIPS in the BG ROI did not relate to BG WMH volumes ($P > 0.10$).

In the CSO ROI, AUPRC values of the Frangi filter in both T1w and T2w imaging, as well as that of the nnU-Net (T2w) in T2w imaging, declined with increasing WMH volume in the same region ($P < 0.05$). The reasons behind these associations differed between methods. The Frangi filter generally marked WMH as potential PVS. As a result, higher WMH burden produced more false positives and consequently lower

AUPRC values (Figure 8). In contrast, the nnU-Net (T2w) more effectively disregarded WMH as potential PVS candidates, but this same ability also led to the omission of PVS located within WMH regions (Figure 8). The AUPRC values obtained by RORPO, SHIVA-PVS, nnU-Net (T1w), and DRIPS in the CSO ROI were not associated with CSO WMH volumes (P > 0.05).

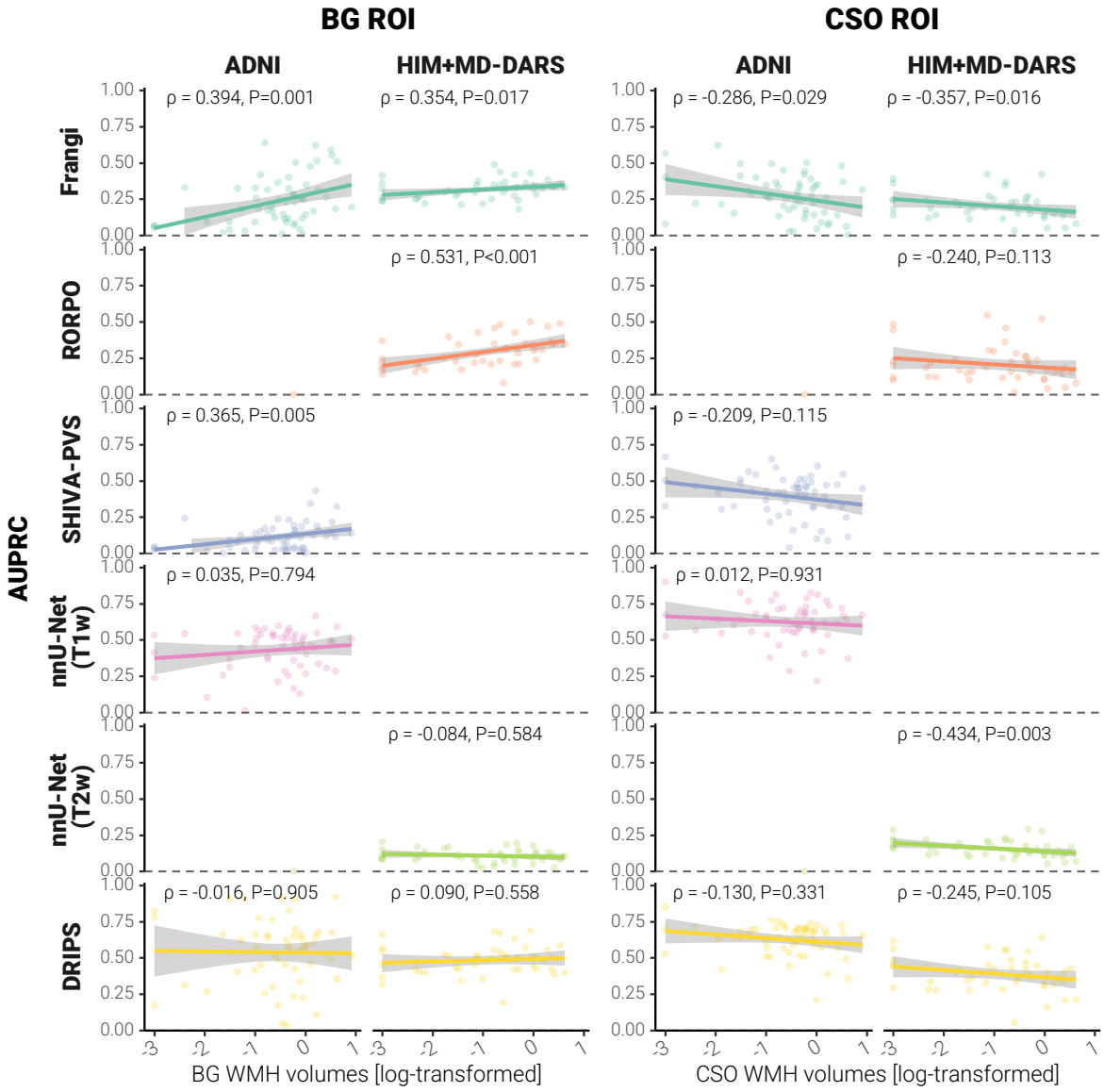


Figure 7. Relationship between segmentation performance (AUPRC) and regional WMH volume for each algorithm across the ADNI and HIM and MD-DARS. We studied these relationships using Spearman correlation coefficients (shown above each subplot). Algorithms that failed or showed limited generalisation within specific datasets were excluded from this secondary analysis (ADNI: RORPO, nnU-Net (T2w); HIM+MD-DARS: SHIVA-PVS, nnU-Net (T1w)). We used the Greek letter ρ to denote the Spearman correlation coefficient and P to denote its p-value.

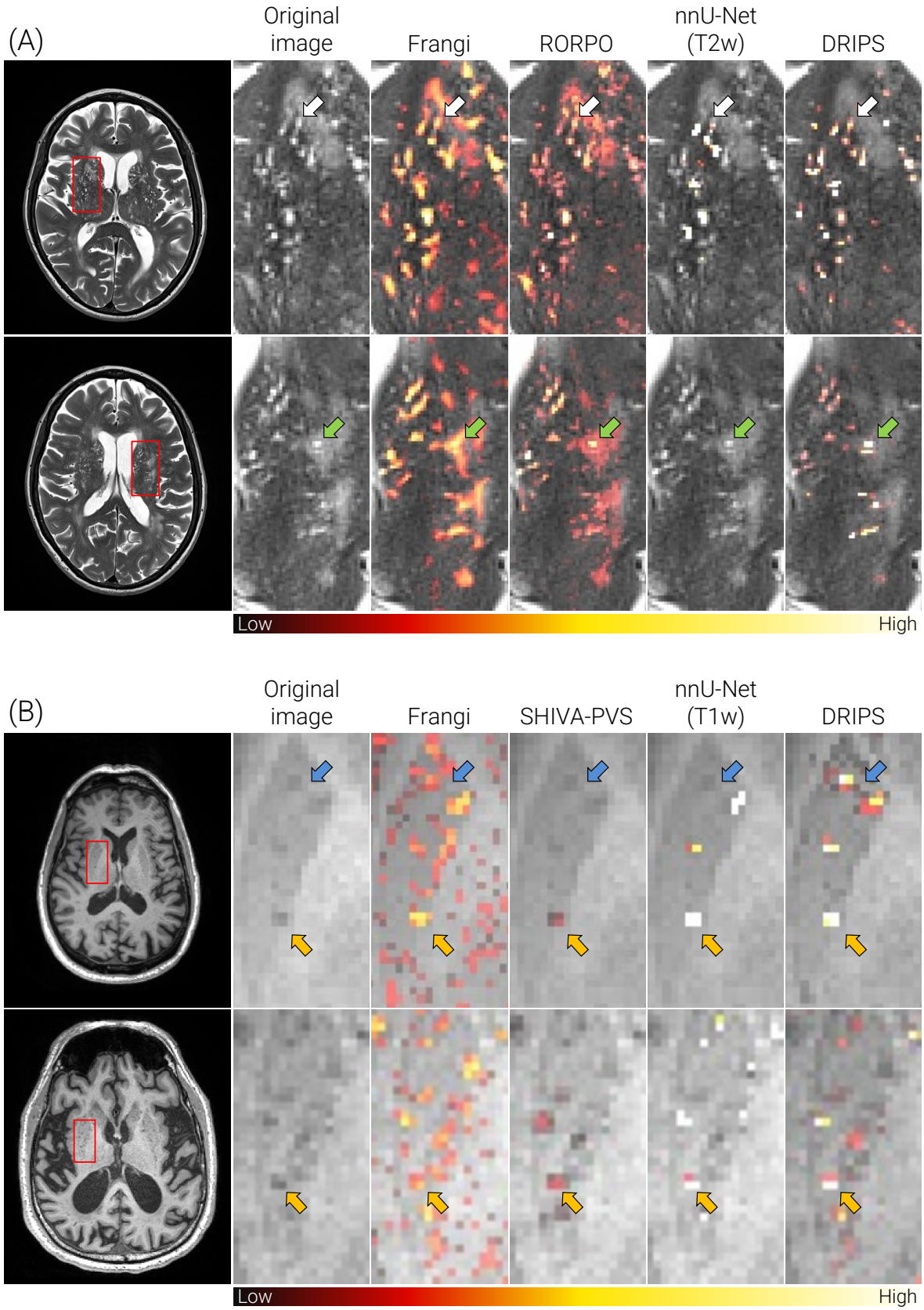


Figure 8. Response map yielded by PVS segmentation methods on T2w and T1 imaging. (A) Maps obtained from all methods that generalised to T2w imaging. The Frangi filter and RORPO produced non-zero responses within WMH. In patients with higher BG WMH burden, false detections

near true PVS often overlapped spatially, artificially inflating AUPRC (white and green arrows). On the other hand, the nnU-Net (T2w) tended to miss PVS located within WMH (green arrow). (B) Maps obtained from all methods that generalised to T1w imaging. SHIVA-PVS identified salient as opposed to subtle PVS (yellow vs blue arrows).

4.3 Generalisation to other imaging modalities

We evaluated the generalisation capacity of DRIPS and the four competing methods beyond MRI, with particular emphasis on their transferability to a 3D *ex vivo* brain model reconstructed from histology (Figure 9). Histology-to-MNI registration was unsuccessful with SynthMorph and ANTs, preventing SHIVA-PVS from being evaluated. DRIPS achieved the best performance across both BG and CSO ROIs. In the BG, it reached a DSC_{lesion} of 0.477, DSC_{voxel} of 0.482, and AUPRC of 0.512, clearly outperforming all other methods (next best DSC_{lesion} 0.373 with RORPO, DSC_{voxel} 0.260 with Frangi, and AUPRC 0.205 with RORPO). In the CSO, it again obtained the highest scores with DSC_{lesion} 0.629, DSC_{voxel} 0.592, and AUPRC 0.625, surpassing RORPO (0.607/0.466/0.475), nnU-Net (T1w; 0.542/0.517/0.450), and Frangi (0.564/0.492/0.493). nnU-Net (T2w) did not generate meaningful PVS segmentations.

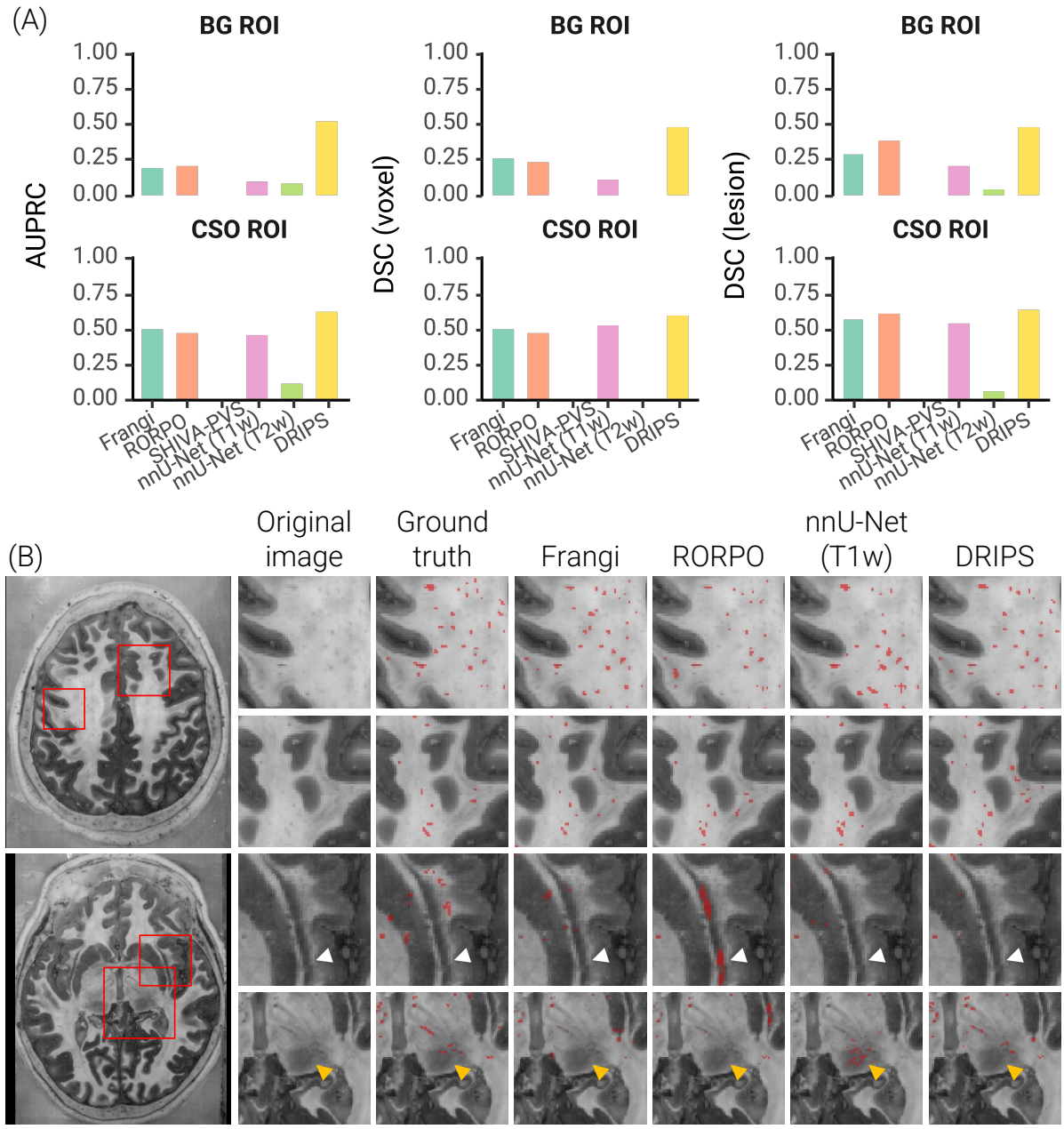


Figure 9. PVS segmentation on a 3D ex-vivo brain model reconstructed from histology images. (A) Segmentation performance, as measured by AUPRC and Dice at both voxel and lesion level. Registration of histology images to MNI space was unsuccessful with SynthMorph and ANTs, preventing SHIVA-PVS evaluation. The nnU-Net (T2w) could not segment PVS successfully. The classical methods, Frangi and RORPO, successfully segmented PVS as expected due to their modality-agnostic design. Both nnU-Net (T1w) and DRIPS produced valid segmentations. DRIPS outperformed all other methods across both regions of interest. RORPO and Frangi achieved the next best results, while nnU-Net (T1w) had the lowest performance. (B) Example segmentations in the CSO and BG ROIs of histology sections. Across algorithms, thresholds were chosen to yield the highest voxel-level Dice coefficients. Segmentation in the CSO ROI was successful across methods, whereas performance in the BG ROI was impaired by systematic errors, including misclassification of the claustrum as PVS (white arrow) and spurious segmentation of multiple thalamic structures as PVS (yellow arrow).

5 Discussion

In recent years, deep learning has become the dominant paradigm for PVS segmentation (Waymont et al., 2024) and for medical image analysis more broadly, primarily owing to its strong within-dataset performance. Yet, as also illustrated by our findings, such models often struggle to generalise when faced with data distributions or imaging modalities not represented during training. Unlike classical approaches—which may be less accurate in noisy settings but still provide a usable output—deep learning models can fail outright when applied outside their training domain, producing no meaningful segmentation. However, relying on constant manual labelling and fine-tuning for every new dataset is neither scalable nor sustainable.

Against this backdrop, our aim was not to develop a model narrowly optimised for a single dataset, but to propose a new PVS segmentation method that achieves high accuracy and robust generalisation across imaging sequences and cohorts. By leveraging physics-based image generation and domain randomisation, we demonstrated that it is possible to mitigate domain shifts and achieve accurate PVS segmentation under conditions seen during training. Across five independent cohorts, we show that DRIPS can (i) segment PVS on both isotropic and anisotropic T1- and T2w images, (ii) outperform classical and machine learning–based approaches, (iii) segment PVS independently of the overall WMH burden, and (iv) generalise even to other modalities, including histology. Taken together, these results position DRIPS as a robust and versatile framework for PVS segmentation.

5.1 Physics-inspired domain randomisation

DRIPS brings together two complementary research directions: domain randomisation and physics-inspired data augmentation. Domain randomisation tackles the challenge of generalisation by exposing models to synthetic data generated from segmentations with fully randomised parameters (Tobin et al., 2017), enabling the learning of robust and transferable features provided that the synthetic variability adequately reflects real-world conditions (Billot et al., 2023a). Physics-inspired data augmentation builds on this by modelling the image acquisition process and its artefacts, thereby enhancing realism and surpassing purely agnostic randomisation strategies (Adams et al., 2024). We observed that introducing voxel size variability through resampling and simulating motion artefacts both contributed positively to performance. Resampling proved essential for handling anisotropic scans. Overall, DRIPS achieved consistent improvements of approximately 0.10–0.13 across all evaluation metrics, with slightly greater gains in the BG compared to the CSO, when resampling was considered as opposed to when it was not. Simulating motion artefacts also helped models trained with DRIPS distinguish true PVS from motion-related ghosting, as seen in case-level examples. At the group level, performance in the BG was largely unaffected by motion training, whereas in the CSO it led to slightly lower voxel-wise precision–recall but significantly improved lesion-wise detection, suggesting a more conservative yet accurate segmentation strategy.

5.2 DRIPS segments PVS accurately on real MRI data

Conventional deep learning approaches to PVS segmentation have typically depended on small, carefully curated training datasets. While such models can achieve high accuracy within their training domain, they often fail to generalise well to

new datasets. SHIVA-PVS, a 3D U-Net trained solely on T1w images, exemplifies this limitation: it did not transfer to T2w scans and showed only limited sensitivity to PVS even within its training modality. A similar limitation was seen with the nnU-Net framework, where models trained on T1w images could only process T1w data, and likewise for T2w images, with little to no generalisation across modalities. Clearly, training separate models for each input modality offers a practical workaround, but it bypasses rather than addresses the fundamental issue of generalisability.

In contrast, our results highlight the utility of domain randomisation for bridging the generalisation gap. DRIPS had stable performance across both T1- and T2w images without the need for retraining, and importantly, the learned features also transferred to histological data—a modality entirely distinct from MRI. These findings reinforce the central premise of domain randomisation: that exposure to sufficiently diverse synthetic variation enables models to acquire representations that remain applicable beyond their original training domain.

5.3 DRIPS versus competing approaches

We compared DRIPS to both classical image-processing-based and machine learning–based methods, using scans and manual annotations from five cohorts (n = 165) that included healthy controls as well as individuals with Long-COVID, hypertensive arteriopathy, cerebral amyloid angiopathy, heart failure, mild cognitive impairment, and Alzheimer's disease. DRIPS outperformed all competing methods on anisotropic scans (EBBIVD, HIM, and MD-DARS) and ranked among the top two on isotropic scans (PCB and ADNI).

On anisotropic T2w scans, the conventional Frangi filter generally emerged as the second-best method. This finding carries important implications for prior studies: when

carefully tuned, Frangi can achieve accurate PVS segmentation, outperforming all machine learning–based methods aside from DRIPS. Its main drawback, as with any other classical PVS segmentation strategy, is the need for manual calibration on each new dataset to reach optimal performance.

On isotropic T1w scans, nnU-Net and DRIPS achieved the highest overall performance, surpassing all other methods by median margins of at least 0.17 in AUPRC, 0.09 in DSC_{voxel}, and 0.14 in DSC_{lesion}. The marked improvement in precision and recall over classical image-processing methods likely stems from the fact that, as shown in Figure 8 and Figure 9, regions such as the boundaries of the putamen, pallidum, and claustrum are often misidentified as PVS by these methods solely due to their “tubular” appearance. Note that, in general, signal intensity differences between the basal ganglia and the surrounding white matter on T1w imaging—particularly at higher field strengths—can also be erroneously highlighted as PVS. In these situations, post-processing strategies that analyse jointly location, length and shape become essential. Their impact on the segmentation performance of classical techniques was not evaluated in this study, as it lay outside the primary scope of our work.

5.4 Robustness against WMH

Previous studies have shown that the presence of WMH can substantially compromise the performance of PVS segmentation methods (Bernal et al., 2022; Pham et al., 2022; Valdes Hernandez et al., 2013; Waymont et al., 2024). Our findings align with this evidence, revealing that both classical and deep learning approaches are often dependent on the regional WMH burden. Traditionally, one of the most common ways to mitigate this issue has been to exclude WMH from analyses. However, as illustrated

by the nnU-Net (T2w), this approach introduces its own bias: by excluding WMH, the method inherently omits PVS that overlap with them, creating artificial correlations with WMH volume, since individuals with more WMH also tend to have more PVS within them.

Both extremes—erroneously labelling WMH as PVS or excluding WMH entirely—are suboptimal. The goal should instead be to develop models whose performance is independent of WMH burden. In this regard, our results indicate that DRIPS was able to segment PVS comparably accurately even in cases with high WMH volumes, without its performance being significantly compromised or biased, regardless of whether input data were T1w or T2w images. A similar pattern was observed for the nnU-Net (T1w) in T1w imaging. Although these findings are based on a limited sample (60 T1w and 45 T2w images), they represent a promising step towards developing segmentation methods that are more robust and less biased by co-occurring brain lesions.

5.5 Limitations and future work

Despite the demonstrated generalisability of our approach, four limitations merit consideration. First, we modelled PVS as tortuous tubular structures distributed throughout the brain. While effective for training and segmentation, this abstraction oversimplifies their biology. *In vivo*, PVS are closely aligned with the cerebral vasculature, following the trajectories of arterioles, capillaries, and venules, with their orientation, calibre, and spatial density shaped by vascular anatomy, regional blood supply, and vessel tortuosity. Second, we assumed a predominant orientation towards the lateral ventricles. This is a reasonable approximation for PVS in the centrum semiovale, which often follow medullary arteries radiating to the ventricles, but it does

not hold in other regions—for instance, PVS surrounding the lenticulostriate arteries in the basal ganglia, which typically run perpendicular to the axial plane. Looking ahead, these limitations highlight an opportunity: conditioning PVS generation on vascular maps could produce more physiologically plausible simulations, improving anatomical fidelity and reducing false positives in regions where tubular structures occur independently of vessels. Third, in this work, we implemented the segmentation network in DRIPS as a 3D U-Net and did not investigate alternative, more advanced architectures. While this represents a limitation, it was not the primary focus of our study. Our main objective was to demonstrate that DRIPS can achieve accurate and robust PVS segmentation across multiple cohorts, health conditions, and imaging settings, rather than to develop a new method optimised for peak performance. Exploring whether more sophisticated segmentations models—such as nnU-Net or transformers—can further improve performance represents an important direction for future work. Fourth, although our evaluation included data from multiple individuals across five cohorts spanning a wide range of conditions—from normal cognition to post-COVID syndrome, hypertensive arteriopathy, cerebral amyloid angiopathy, heart failure, mild cognitive impairment, and Alzheimer's disease—our assessment remains limited by the imaging protocols included in this study. As part of our future work, we plan to include patients spanning a broader range of disease severities—from very mild to advanced—scanned using multiple imaging sequences, and to conduct longitudinal assessments to evaluate the method's ability to track changes in PVS over extended periods of time.

6 Conclusion

We introduced DRIPS, the first physics-inspired domain randomisation framework for accurate out-of-sample PVS segmentation. DRIPS accurately segmented PVS in both T1w and T2w images, at isotropic and anisotropic resolutions, without requiring manual PVS segmentations, retraining, or fine-tuning. It outperformed all competing methods on anisotropic images and achieved performance comparable to nnU-Net on isotropic data. Unlike the segmentation performance of competing methods, its performance was not associated by the volume of WMH in the brain. DRIPS's out-of-sample capabilities extended beyond MRI, successfully segmenting PVS in 3D *ex vivo* brain models reconstructed from histology. Collectively, our findings demonstrate that DRIPS segments PVS accurately across diverse imaging settings and patient populations, enabling more accessible and reliable automated PVS quantification for both research and clinical use.

Declarations

Use of AI-assisted technologies in the manuscript preparation process

The authors used ChatGPT to assist with grammar correction during the preparation of this work. All content was in all instances reviewed and edited by the authors, who take full responsibility for the final published article.

Ethics approval and consent to participate

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Consent for publication

Not applicable.

Availability of data and materials

We provide a ready-to-use Docker image for running DRIPS, available at https://github.com/CIR-FAU/DRIPS.git, to ensure easy deployment, reproducibility, and consistent performance across systems.

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors report no competing interests.

Funding

This research was supported by the German Centre for Neurodegenerative Diseases (Deutsches Zentrum für Neurodegenerative Erkrankungen, DZNE; reference number BN012) and funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG; Project IDs 425899996 and 362321501/RTG 2413 SynAGE; CRC 1436, projects A05, B02, B04 and C01).

The post-COVID Brain project is partly funded by IZKF Jena (advanced clinician scientist grant to B.B.) and the Bundesministerium für Bildung und Forschung (BMBF) start-up funding for the Germany Centre for Mental Health (DZPG, BMBF 01EE2305D, JE2/TP5). Further, the project was supported in part by the Deutsche Forschungsgemeinschaft (DFG) (MA 9235/3-1/SCHR 1418/5-1 (501214112), CRC 1436 (B04, 425899996), and RTG 2413 (SynAGE, 362321501)) and by the Deutsche Alzheimer Gesellschaft e.V. (DAlzG) and Förderstiftung Dierichs (MD-DARS and BB-DARS project, respectively).

The HIM-Study was supported by the following grants: European Regional Development Fund (EFRE) (ZS/2024/02/184014, to RBD, SS & PM), European Regional Development Fund (EFRE) (ZS/2024/05/187256 to PM) and the Polycarp-Leporin-Program (PLP23/5, to PM).

Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH) including generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; BristolMyers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development

LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics.

Computational PVS quantification was supported by The Galen and Hilary Weston Foundation under the Novel Biomarkers 2019 scheme (ref UB190097) administered by the Weston Brain Institute and the Row Fogo Charitable Trust (ref no. BRO-D. FID3668413).

The funding bodies played no role in the design of the study or collection, analysis, or interpretation of data or in writing the manuscript.

CRediT authorship contribution statement

Conceptualisation: L.B., M.Di., G.Z., J.B.; Methodology: L.B., M.Di., J.B.; Software: L.B., M.Di., J.B.; Validation: L.B., M.Di., J.B.; Formal analysis: L.B., M.Di., J.B.; Investigation: L.B., M.Di., J.B.; Resources: H.M, K.N, M.P, C.B., H.T.M., E.F., S.T., D.T., B.B., T.R., P.A.R, A.S., N.O., C.G., M.W., P.A., D.B., C.P., Y.L., P.M., R.B.D, S.S., E.D., G.Z., J.B. Data Curation: L.B., J.B.; Writing - Original Draft: L.B., J.B.; Writing - Review & Editing: All authors; Visualisation: L.B., J.B.; Supervision: G.Z., J.B.; Project administration: J.B..

References

Adams, R., Huynh, K.M., Zhao, W., Hu, S., Lyu, W., Ahmad, S., Ma, D., Yap, P.-T., 2024. UltimateSynth: MRI Physics for Pan-Contrast AI. bioRxiv 1–31. https://doi.org/10.1101/2024.12.05.627056

882 Amunts, K., Lepage, C., Borgeat, L., Mohlberg, H., Dickscheid, T., Rousseau, M.-É.,
883 Bludau, S., Bazin, P.-L., Lewis, L.B., Oros-Peusquens, A.-M., Shah, N.J., Lippert,
884 T., Zilles, K., Evans, A.C., 2013. BigBrain: An Ultrahigh-Resolution 3D Human
885 Brain Model. Science (1979) 340. https://doi.org/10.1126/science.1235381

886 Ballerini, L., Lovreglio, R., Valdés Hernández, M.D.C., Ramirez, J., MacIntosh, B.J.,
887 Black, S.E., Wardlaw, J.M., 2018. Perivascular Spaces Segmentation in Brain
888 MRI Using Optimal 3D Filtering. Sci Rep 8, 1–11. https://doi.org/10.1038/s41598-
889 018-19781-5

890 Ballerini, L., McGrory, S., Valdés Hernández, M. del C., Lovreglio, R., Pellegrini, E.,
891 MacGillivray, T., Muñoz Maniega, S., Henderson, R., Taylor, A., Bastin, M.E.,
892 Doubal, F., Trucco, E., Deary, I.J., Wardlaw, J., 2020. Quantitative measurements
893 of enlarged perivascular spaces in the brain are associated with retinal
894 microvascular parameters in older community-dwelling subjects. Cereb Circ Cogn
895 Behav 1, 100002. https://doi.org/10.1016/j.cccb.2020.100002

896 Barisano, G., Iv, M., Choupan, J., Hayden-Gephart, M., 2025. Robust, fully-automated
897 assessment of cerebral perivascular spaces and white matter lesions: a
898 multicentre MRI longitudinal study of their evolution and association with risk of
899 dementia and accelerated brain atrophy. EBioMedicine 111.
900 https://doi.org/10.1016/j.ebiom.2024.105523

901 Barnes, A., Ballerini, L., Valdés Hernández, M. del C., Chappell, F.M., Muñoz
902 Maniega, S., Meijboom, R., Backhouse, E. V., Stringer, M.S., Duarte Coello, R.,
903 Brown, R., Bastin, M.E., Cox, S.R., Deary, I.J., Wardlaw, J.M., 2022. Topological
904 relationships between perivascular spaces and progression of white matter

905 hyperintensities: A pilot study in a sample of the Lothian Birth Cohort 1936. Front
906 Neurol 13. https://doi.org/10.3389/fneur.2022.889884

907 Bernal, J., Valdés-Hernández, M., Ballerini, L., Escudero, J., Jochems, A.C.C., Clancy,
908 U., Doubal, F.N., Stringer, M.S., Thrippleton, M.J., Touyz, R.M., Wardlaw, J.M.,
909 2020. A Framework for Jointly Assessing and Reducing Imaging Artefacts
910 Automatically Using Texture Analysis and Total Variation Optimisation for
911 Improving Perivascular Spaces Quantification in Brain Magnetic Resonance
912 Imaging. Communications in Computer and Information Science 1248 CCIS,
913 171–183. https://doi.org/10.1007/978-3-030-52791-4_14

914 Bernal, J., Valdés-Hernández, M. d. C., Escudero, J., Heye, A.K., Sakka, E., Armitage,
915 P.A., Makin, S., Touyz, R.M., Wardlaw, J.M., Thrippleton, M.J., 2021a. A four-
916 dimensional computational model of dynamic contrast-enhanced magnetic
917 resonance imaging measurement of subtle blood-brain barrier leakage.
918 Neuroimage 230, 117786. https://doi.org/10.1016/j.neuroimage.2021.117786

919 Bernal, J., Valdés-Hernández, M.D.C., Escudero, J., Duarte, R., Ballerini, L., Bastin,
920 M.E., Deary, I.J., Thrippleton, M.J., Touyz, R.M., Wardlaw, J.M., 2022.
921 Assessment of perivascular space filtering methods using a three-dimensional
922 computational model. Magn Reson Imaging 93, 33–51.
923 https://doi.org/10.1016/j.mri.2022.07.016

924 Bernal, J., Xu, W., Valdés-Hernández, M. d. C., Escudero, J., Jochems, A.C.C.,
925 Clancy, U., Doubal, F.N., Stringer, M.S., Thrippleton, M.J., Touyz, R.M., Wardlaw,
926 J.M., 2021b. Selective Motion Artefact Reduction via Radiomics and k-space
927 Reconstruction for Improving Perivascular Space Quantification in Brain Magnetic
928 Resonance Imaging, Lecture Notes in Computer Science (including subseries

Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer International Publishing. https://doi.org/10.1007/978-3-030-80432-9_12

Besteher, B., Machnik, M., Troll, M., Toepffer, A., Zerekidze, A., Rocktäschel, T., Heller, C., Kikinis, Z., Brodoehl, S., Finke, K., Reuken, P.A., Opel, N., Stallmach, A., Gaser, C., Walter, M., 2022. Larger gray matter volumes in neuropsychiatric long-COVID syndrome. Psychiatry Res 317. https://doi.org/10.1016/j.psychres.2022.114836

Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A. V., Iglesias, J.E., 2023a. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. Med Image Anal 86. https://doi.org/10.1016/j.media.2023.102789

Billot, B., Magdamo, C., Cheng, Y., Arnold ID, S.E., Das, S.I., Eugenio Iglesias, J., 2023b. Robust machine learning segmentation for large-scale analysis of heterogeneous clinical brain MRI datasets. PNAS 120, e2216399120. https://doi.org/https://doi.org/10.1073/pnas.2216399120

Boespflug, E.L., Schwartz, D.L., Lahna, D., Pollock, J., Iliff, J.J., Kaye, J.A., Rooney, W., Silbert, L.C., 2018. MR imaging-based multimodal autoidentification of perivascular spaces (mMAPS): Automated morphologic segmentation of enlarged perivascular spaces at clinical field strength. Radiology 286, 632–642. https://doi.org/10.1148/radiol.2017170205

Boutinaud, P., Tsuchida, A., Laurent, A., Adonias, F., Hanifehlou, Z., Nozais, V., Verrecchia, V., Lampe, L., Zhang, J., Zhu, Y.C., Tzourio, C., Mazoyer, B., Joliot, M., 2021a. 3D Segmentation of Perivascular Spaces on T1-Weighted 3 Tesla MR

952 Images With a Convolutional Autoencoder and a U-Shaped Neural Network. Front
953 Neuroinform 15. https://doi.org/10.3389/fninf.2021.641600

954 Boutinaud, P., Tsuchida, A., Laurent, A., Adonias, F., Hanifehlou, Z., Nozais, V.,
955 Verrecchia, V., Lampe, L., Zhang, J., Zhu, Y.C., Tzourio, C., Mazoyer, B., Joliot,
956 M., 2021b. 3D Segmentation of Perivascular Spaces on T1-Weighted 3 Tesla MR
957 Images With a Convolutional Autoencoder and a U-Shaped Neural Network. Front
958 Neuroinform 15, 1–21. https://doi.org/10.3389/fninf.2021.641600

959 Bown, C.W., Carare, R.O., Schrag, M.S., Jefferson, A.L., 2022. Physiology and
960 Clinical Relevance of Enlarged Perivascular Spaces in the Aging Brain. Neurology
961 98, 107–117. https://doi.org/10.1212/WNL.0000000000013077

962 Braun, M., Iliff, J.J., 2020. The impact of neurovascular, blood-brain barrier, and
963 glymphatic dysfunction in neurodegenerative and metabolic diseases, 1st ed,
964 International Review of Neurobiology. Elsevier Inc.
965 https://doi.org/10.1016/bs.irn.2020.02.006

966 Cai, D., Pan, M., Liu, C., He, W., Ge, X., Lin, J., Li, R., Liu, M., Xia, J., 2024. Deep-
967 learning-based segmentation of perivascular spaces on T2-Weighted 3T
968 magnetic resonance images. Front Aging Neurosci 16.
969 https://doi.org/10.3389/fnagi.2024.1457405

970 Chai, Y., Zhang, H., Robles, C., Kim, A.S., Janhanshad, N., Thompson, P.M., van der
971 Werf, Y., van Heese, E.M., Kim, J., Joo, E.Y., Aksman, L., Kang, K.-W., Shin, J.-
972 W., Trang, A., Ha, J., Lee, E., Moon, Y., Kim, H., 2025. Precise perivascular space
973 segmentation on magnetic resonance imaging from Human Connectome Project-
974 Aging. medRxiv 1–11. https://doi.org/10.1101/2025.03.19.25324269

Chalcroft, L., Pappas, I., Price, C.J., Ashburner, J., 2025. Synthetic Data for Robust Stroke Segmentation. Machine Learning for Biomedical Imaging 3, 317–347.

Duarte Coello, R., Valdés Hernández, M. del C., Zwanenburg, J.J.M., van der Velden, M., Kuijf, H.J., De Luca, A., Moyano, J.B., Ballerini, L., Chappell, F.M., Brown, R., Jan Biessels, G., Wardlaw, J.M., 2024. Detectability and accuracy of computational measurements of in-silico and physical representations of enlarged perivascular spaces from magnetic resonance images. J Neurosci Methods 403. https://doi.org/10.1016/j.jneumeth.2023.110039

Dubost, F., Adams, H., Bortsova, G., Ikram, M.A., Niessen, W., Vernooij, M., de Bruijne, M., 2019a. 3D regression neural network for the quantification of enlarged perivascular spaces in brain MRI. Med Image Anal 51, 89–100. https://doi.org/10.1016/j.media.2018.10.008

Dubost, F., Yilmaz, P., Adams, H., Bortsova, G., Ikram, M.A., Niessen, W., Vernooij, M., de Bruijne, M., 2019b. Enlarged perivascular spaces in brain MRI: Automated quantification in four regions. Neuroimage 185, 534–544. https://doi.org/10.1016/j.neuroimage.2018.10.026

Duering, M., Biessels, G.J., Brodtmann, A., Chen, C., Cordonnier, C., de Leeuw, F.-E., Debette, S., Frayne, R., Jouvent, E., Rost, N.S., ter Telgte, A., Al-Shahi Salman, R., Backes, W.H., Bae, H.-J., Brown, R., Chabriat, H., De Luca, A., DeCarli, C., Dewenter, A., Doubal, F.N., Ewers, M., Field, T.S., Ganesh, A., Greenberg, S., Helmer, K.G., Hilal, S., Jochems, A.C.C., Jokinen, H., Kuijf, H., Lam, B.Y.K., Lebenberg, J., MacIntosh, B.J., Maillard, P., Mok, V.C.T., Pantoni, L., Rudilosso, S., Satizabal, C.L., Schirmer, M.D., Schmidt, R., Smith, C., Staals, J., Thrippleton, M.J., van Veluw, S.J., Vemuri, P., Wang, Y., Werring, D., Zedde,

M., Akinyemi, R.O., Del Brutto, O.H., Markus, H.S., Zhu, Y.-C., Smith, E.E., Dichgans, M., Wardlaw, J.M., 2023. Neuroimaging standards for research into small vessel disease—advances since 2013. Lancet Neurol 22, 602–618. https://doi.org/10.1016/S1474-4422(23)00131-X

Francis, F., Ballerini, L., Wardlaw, J.M., 2019. Perivascular spaces and their associations with risk factors, clinical disorders and neuroimaging features: A systematic review and meta-analysis. International Journal of Stroke 14, 359–371. https://doi.org/10.1177/1747493019830321

Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale Vessel Enhancement Filtering, in: Lecture Notes in Computer Science. pp. 130–137.

González-Castro, V., Hernández, M.D.C.V., Armitage, P.A., Wardlaw, J.M., 2016. Texture-based Classification for the Automatic Rating of the Perivascular Spaces in Brain MRI, in: Procedia Computer Science. Elsevier B.V., pp. 9–14. https://doi.org/10.1016/j.procs.2016.07.003

Gouveia-Freitas, K., Bastos-Leite, A.J., 2021. Perivascular spaces and brain waste clearance systems: relevance for neurodegenerative and cerebrovascular pathology. Neuroradiology. https://doi.org/10.1007/s00234-021-02718-7

Gudbjartsson, H., Patz, S., 1995. The Rician Distribution of Noisy MRI Data. Magn Reson Med 34, 910–914.

Hablitz, L.M., Nedergaard, M., 2021. The glymphatic system: A novel component of fundamental neurobiology. Journal of Neuroscience 41, 7698–7711. https://doi.org/10.1523/JNEUROSCI.0619-21.2021

Hirschler, L., Runderkamp, B.A., Decker, A., van Harten, T.W., Scheyhing, P., Ehses, P., Petitclerc, L., Layer, J., Pracht, E., Coolen, B.F., van der Zwaag, W., Stöcker, T., Vollmuth, P., Paech, D., Effland, A., van Walderveen, M.A.A., Radbruch, A., van Buchem, M.A., Petzold, G.C., van Veluw, S.J., Caan, M.W.A., Deike, K., van Osch, M.J.P., 2025. Region-specific drivers of CSF mobility measured with MRI in humans. Nat Neurosci. https://doi.org/10.1038/s41593-025-02073-3

Hoffmann, M., Hoopes, A., Greve, D.N., Fischl, B., Dalca, A. V., 2024. Anatomy-aware and acquisition-agnostic joint registration with SynthMorph. Imaging Neuroscience 2, 1–33. https://doi.org/10.1162/imag_a_00197

Hoopes, A., Mora, J.S., Dalca, A. V., Fischl, B., Hoffmann, M., 2022. SynthStrip: skull-stripping for any brain image. Neuroimage 260. https://doi.org/10.1016/j.neuroimage.2022.119474

Hou, Y., Park, S.H., Wang, Q., Zhang, J., Zong, X., Lin, W., Shen, D., 2017. Enhancement of Perivascular Spaces in 7 T MR Image using Haar Transform of Non-local Cubes and Block-matching Filtering. Sci Rep 7. https://doi.org/10.1038/s41598-017-09336-5

Iglesias, J.E., Billot, B., Balbastre, Y., Magdamo, C., Arnold, S.E., Das, S., Edlow, B.L., Alexander, D.C., Golland, P., Fischl, B., 2023. SynthSR: A public AI tool to turn heterogeneous clinical brain scans into high-resolution T1-weighted images for 3D morphometry. Sci Adv eadd3607.

Iliff, J.J., Chen, M.J., Plog, B.A., Zeppenfeld, D.M., Soltero, M., Yang, L., Singh, I., Deane, R., Nedergaard, M., 2014. Impairment of glymphatic pathway function promotes tau pathology after traumatic brain injury. Journal of Neuroscience 34, 16180–16193. https://doi.org/10.1523/JNEUROSCI.3020-14.2014

Iliff, J.J., Wang, M., Liao, Y., Plogg, B.A., Peng, W., Gundersen, G.A., Benveniste, H., Vates, G.E., Deane, R., Goldman, S.A., Nagelhus, E.A., Nedergaard, M., 2012. A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β. Sci Transl Med 4, 1–12. https://doi.org/10.1126/scitranslmed.3003748

Ineichen, B. V., Okar, S. V., Proulx, S.T., Engelhardt, B., Lassmann, H., Reich, D.S., 2022. Perivascular spaces and their role in neuroinflammation. Neuron 110, 3566–3581. https://doi.org/10.1016/j.neuron.2022.10.024

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18, 203–211. https://doi.org/10.1038/s41592-020-01008-z

Kern, K.C., Nasrallah, I.M., Bryan, R.N., Reboussin, D.M., Wright, C.B., 2023. Intensive systolic blood pressure treatment remodels brain perivascular spaces: A secondary analysis of the Systolic Pressure Intervention Trial (SPRINT). Neuroimage Clin 40, 103513. https://doi.org/10.1016/j.nicl.2023.103513

Kim, H.G., Shin, N.-Y., Nam, Y., Yun, E., Yoon, U., Lee, H.S., Ahn, K.J., 2023. MRI-visible Dilated Perivascular Space in the Brain by Age: The Human Connectome Project. Radiology 306, 1–9. https://doi.org/10.1148/radiol.213254

Laso, P., Cerri, S., Sorby-Adams, A., Guo, J., Mateen, F., Goebl, P., Wu, J., Liu, P., Li, H., Young, S.I., Billot, B., Puonti, O., Sze, G., Payabavash, S., DeHavenon, A., Sheth, K.N., Rosen, M.S., Kirsch, J., Strisciuglio, N., Wolterink, J.M., Eshaghi, A., Barkhof, F., Kimberly, W.T., Iglesias, J.E., 2024. Quantifying white matter hyperintensity and brain volumes in heterogeneous clinical and low-field portable

MRI, in: IEEE International Symposium on Biomedical Imaging (ISBI). IEEE, Athens, Greece, pp. 1–5.

Lian, C., Zhang, J., Liu, M., Zong, X., Hung, S.C., Lin, W., Shen, D., 2018. Multi-channel multi-scale fully convolutional network for 3D perivascular spaces segmentation in 7T MR images. Med Image Anal 46, 106–117. https://doi.org/10.1016/j.media.2018.02.009

Lynch, K.M., Sepehrband, F., Toga, A.W., Choupan, J., 2023. Brain perivascular space imaging across the human lifespan. Neuroimage 271, 120009. https://doi.org/10.1016/j.neuroimage.2023.120009

Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M.D., Buettner, F., Christodoulou, E., Glocker, B., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M.A., Wiesenfarth, M., Kavur, A.E., Sudre, C.H., Baumgartner, M., Eisenmann, M., Heckmann-Nötzel, D., Rädsch, T., Acion, L., Antonelli, M., Arbel, T., Bakas, S., Benis, A., Blaschko, M.B., Cardoso, M.J., Cheplygina, V., Cimini, B.A., Collins, G.S., Farahani, K., Ferrer, L., Galdran, A., van Ginneken, B., Haase, R., Hashimoto, D.A., Hoffman, M.M., Huisman, M., Jannin, P., Kahn, C.E., Kainmueller, D., Kainz, B., Karargyris, A., Karthikesalingam, A., Kofler, F., Kopp-Schneider, A., Kreshuk, A., Kurc, T., Landman, B.A., Litjens, G., Madani, A., Maier-Hein, K., Martel, A.L., Mattson, P., Meijering, E., Menze, B., Moons, K.G.M., Müller, H., Nichyporuk, B., Nickel, F., Petersen, J., Rajpoot, N., Rieke, N., Saez-Rodriguez, J., Sánchez, C.I., Shetty, S., van Smeden, M., Summers, R.M., Taha, A.A., Tiulpin, A., Tsaftaris, S.A., Van Calster, B., Varoquaux, G., Jäger, P.F., 2024. Metrics reloaded: recommendations for image analysis validation. Nat Methods 21, 195–212. https://doi.org/10.1038/s41592-023-02151-z

Menze, I., Bernal, J., Kaya, P., Aki, Ç., Pfister, M., Geisendörfer, J., Yakupov, R., Coello, R.D., Valdés-Hernández, M. d. C., Heneka, M.T., Brosseron, F., Schmid, M.C., Glanz, W., Incesoy, E.I., Butryn, M., Rostamzadeh, A., Meiberth, D., Peters, O., Preis, L., Lammerding, D., Gref, D., Priller, J., Spruth, E.J., Altenstein, S., Lohse, A., Hetzer, S., Schneider, A., Fliessbach, K., Kimmich, O., Vogt, I.R., Wiltfang, J., Bartels, C., Schott, B.H., Hansen, N., Dechent, P., Buerger, K., Janowitz, D., Perneczky, R., Rauchmann, B.-S., Teipel, S., Kilimann, I., Goerss, D., Laske, C., Munk, M.H., Sanzenbacher, C., Hinderer, P., Scheffler, K., Spottke, A., Roy-Kluth, N., Lüsebrink, F., Neumann, K., Wardlaw, J., Jessen, F., Schreiber, S., Düzel, E., Ziegler, G., 2024. Perivascular space enlargement accelerates in ageing and Alzheimer's disease pathology: evidence from a three-year longitudinal multicentre study. Alzheimers Res Ther 16, 242. https://doi.org/10.1186/s13195-024-01603-8

Merveille, O., Talbot, H., Najman, L., Passat, N., 2018. Curvilinear Structure Analysis by Ranking the Orientation Responses of Path Operators. IEEE Trans Pattern Anal Mach Intell 40, 304–317. https://doi.org/10.1109/TPAMI.2017.2672972

Merveille, O., Talbot, H., Najman, L., Passat, N., 2014. Tubular structure filtering by ranking orientation responses of path operators. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8690 LNCS, 203–218. https://doi.org/10.1007/978-3-319-10605-2_14

Mestre, H., Tithof, J., Du, T., Song, W., Peng, W., Sweeney, A.M., Olveda, G., Thomas, J.H., Nedergaard, M., Kelley, D.H., 2018. Flow of cerebrospinal fluid is

driven by arterial pulsations and is reduced in hypertension. Nat Commun 9, 4878. https://doi.org/10.1038/s41467-018-07318-3

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571.

Müller, P., Horndasch, L., Neumann, K., Mattern, H., Cardace, S., Arndt, P., Pfister, M., Groscheck, T., Vielhaber, S., Meuth, S., Dunay, I., Schmeisser, A., Behme, D., Schreiber, S., Braun-dullaeus Rüdiger, 2024. CEREBRAL SMALL VESSEL DISEASE MEDIATES THE EFFECT OF ARTERIAL STIFFNESS ON COGNITIVE DECLINE IN PATIENTS WITH HEART FAILURE WITH PRESERVED EJECTION FRACTION. J Hypertens 42(Suppl 1), e92–e93.

Neumann, K., Günther, M., Düzel, E., Schreiber, S., 2022. Microvascular Impairment in Patients With Cerebral Small Vessel Disease Assessed With Arterial Spin Labeling Magnetic Resonance Imaging: A Pilot Study. Front Aging Neurosci 14. https://doi.org/10.3389/fnagi.2022.871612

Okar, S. V., Hu, F., Shinohara, R.T., Beck, E.S., Reich, D.S., Ineichen, B. V., 2023. The etiology and evolution of magnetic resonance imaging-visible perivascular spaces: Systematic review and meta-analysis. Front Neurosci 17, 1–13. https://doi.org/10.3389/fnins.2023.1038011

Park, S.H., Zong, X., Gao, Y., Lin, W., Shen, D., 2016. Segmentation of perivascular spaces in 7 T MR image using auto-context model with orientation-normalized features. Neuroimage 134, 223–235. https://doi.org/10.1016/j.neuroimage.2016.03.076

1139 Pham, W., Jarema, A., Rim, D., Chen, Z., Khlif, M.S., Macefield, V.G., Henderson,
1140 L.A., Brodtmann, A., 2024. A Comprehensive Framework for Automated
1141 Segmentation of Perivascular Spaces in Brain MRI with the nnU-Net. ArXiv 1–49.

1142 Pham, W., Lynch, M., Spitz, G., O'Brien, T., Vivash, L., Sinclair, B., Law, M., 2022. A
1143 critical guide to the automated quantification of perivascular spaces in magnetic
1144 resonance imaging. Front Neurosci 16.
1145 https://doi.org/10.3389/fnins.2022.1021311

1146 Potter, G.M., Chappell, F.M., Morris, Z., Wardlaw, J.M., 2015. Cerebral perivascular
1147 spaces visible on magnetic resonance imaging: Development of a qualitative
1148 rating scale and its observer reliability. Cerebrovascular Diseases 39, 224–231.
1149 https://doi.org/10.1159/000375153

1150 Rashid, T., Liu, H., Ware, J.B., Li, K., Romero, J.R., Fadaee, E., Nasrallah, I.M., Hilal,
1151 S., Bryan, R.N., Hughes, T.M., Davatzikos, C., Launer, L., Seshadri, S., Heckbert,
1152 S.R., Habes, M., 2023. Deep learning based detection of enlarged perivascular
1153 spaces on brain MRI. Neuroimage: Reports 3.
1154 https://doi.org/10.1016/j.ynirp.2023.100162

1155 Rasmussen, M.K., Mestre, H., Nedergaard, M., 2018. The glymphatic pathway in
1156 neurological disorders. Lancet Neurol 17, 1016–1024.
1157 https://doi.org/10.1016/S1474-4422(18)30318-1

1158 Saito, T., Rehmsmeier, M., 2015. The precision-recall plot is more informative than the
1159 ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One
1160 10. https://doi.org/10.1371/journal.pone.0118432

1161 Schreiber, S., Bernal, J., Arndt, P., Schreiber, F., Müller, P., Morton, L., Braun-
1162 Dullaeus, R.C., Valdés-Hernández, M.D.C., Duarte, R., Wardlaw, J.M., Meuth,

S.G., Mietzner, G., Vielhaber, S., Dunay, I.R., Dityatev, A., Jandke, S., Mattern, H., 2023. Brain Vascular Health in ALS Is Mediated through Motor Cortex Microvascular Integrity. Cells 12. https://doi.org/10.3390/cells12060957

Schwartz, D.L., Boespflug, E.L., Lahna, D.L., Pollock, J., Roese, N.E., Silbert, L.C., 2019. Autoidentification of perivascular spaces in white matter using clinical field strength T1 and FLAIR MR imaging. Neuroimage 202, 116126. https://doi.org/10.1016/j.neuroimage.2019.116126

Shaw, R., Sudre, C.H., Varsavsky, T., Ourselin, S., Cardoso, M.J., 2020. A k-Space Model of Movement Artefacts: Application to Segmentation Augmentation and Artefact Removal. IEEE Trans Med Imaging 39, 2881–2892. https://doi.org/10.1109/tmi.2020.2972547

Smith, E.E., Biessels, G.J., De Guio, F., de Leeuw, F.E., Duchesne, S., Düring, M., Frayne, R., Ikram, M.A., Jouvent, E., MacIntosh, B.J., Thrippleton, M.J., Vernooij, M.W., Adams, H., Backes, W.H., Ballerini, L., Black, S.E., Chen, C., Corriveau, R., DeCarli, C., Greenberg, S.M., Gurol, M.E., Ingrisch, M., Job, D., Lam, B.Y.K., Launer, L.J., Linn, J., McCreary, C.R., Mok, V.C.T., Pantoni, L., Pike, G.B., Ramirez, J., Reijmer, Y.D., Romero, J.R., Ropele, S., Rost, N.S., Sachdev, P.S., Scott, C.J.M., Seshadri, S., Sharma, M., Sourbron, S., Steketee, R.M.E., Swartz, R.H., van Oostenbrugge, R., van Osch, M., van Rooden, S., Viswanathan, A., Werring, D., Dichgans, M., Wardlaw, J.M., 2019. Harmonizing brain magnetic resonance imaging methods for vascular contributions to neurodegeneration. Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring 11, 191–204. https://doi.org/10.1016/j.dadm.2019.01.002

Sudre, C.H., Van Wijnen, K., Dubost, F., Adams, H., Atkinson, D., Barkhof, F., Birhanu, M.A., Bron, E.E., Camarasa, R., Chaturvedi, N., Chen, Y., Chen, Z., Chen, S., Dou, Q., Evans, T., Ezhov, I., Gao, H., Girones Sanguesa, M., Gispert, J.D., Gomez Anson, B., Hughes, A.D., Ikram, M.A., Ingala, S., Jaeger, H.R., Kofler, F., Kuijf, H.J., Kutnar, D., Lee, M., Li, B., Lorenzini, L., Menze, B., Molinuevo, J.L., Pan, Y., Puybareau, E., Rehwald, R., Su, R., Shi, P., Smith, L., Tillin, T., Tochon, G., Urien, H., van der Velden, B.H.M., van der Velpen, I.F., Wiestler, B., Wolters, F.J., Yilmaz, P., de Groot, M., Vernooij, M.W., de Bruijne, M., 2024. Where is VALDO? VAscular Lesions Detection and segmentatiOn challenge at MICCAI 2021. Med Image Anal 91. https://doi.org/10.1016/j.media.2023.103029

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P., 2017. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver, pp. 23–30. https://doi.org/10.1109/IROS.2017.8202133

Valdes Hernandez, M. d. C., Piper, R.J., Wang, X., Deary, I.J., Wardlaw, J.M., 2013. Towards the automatic computational assessment of enlarged perivascular spaces on brain magnetic resonance images: A systematic review. Journal of Magnetic Resonance Imaging 38, 774–785. https://doi.org/10.1002/jmri.24047

Valdés Hernández, M. del C., Duarte Coello, R., Xu, W., Bernal, J., Cheng, Y., Ballerini, L., Wiseman, S.J., Chappell, F.M., Clancy, U., Jaime García, D., Arteaga Reyes, C., Zhang, J.F., Liu, X., Hewins, W., Stringer, M., Doubal, F., Thrippleton, M.J., Jochems, A., Brown, R., Wardlaw, J.M., 2024. Influence of threshold selection and image sequence in in-vivo segmentation of enlarged perivascular

spaces. J Neurosci Methods 403. https://doi.org/10.1016/j.jneumeth.2023.110037

Vikner, T., Karalija, N., Eklund, A., Malm, J., Lundquist, A., Gallewicz, N., Dahlin, M., Lindenberger, U., Riklund, K., Bäckman, L., Nyberg, L., Wåhlin, A., 2022. 5-Year Associations among Cerebral Arterial Pulsatility, Perivascular Space Dilation, and White Matter Lesions. Ann Neurol 92, 871–881. https://doi.org/10.1002/ana.26475

Wardlaw, J.M., Benveniste, H., Nedergaard, M., Zlokovic, B. V., Mestre, H., Lee, H., Doubal, F.N., Brown, R., Ramirez, J., MacIntosh, B.J., Tannenbaum, A., Ballerini, L., Rungta, R.L., Boido, D., Sweeney, M., Montagne, A., Charpak, S., Joutel, A., Smith, K.J., Black, S.E., 2020. Perivascular spaces in the brain: anatomy, physiology and pathology. Nat Rev Neurol 16, 137–153. https://doi.org/10.1038/s41582-020-0312-z

Wardlaw, J.M., Doubal, F., Armitage, P., Chappell, F., Carpenter, T., Muñoz Maniega, S., Farrall, A., Sudlow, C., Dennis, M., Dhillon, B., 2009. Lacunar stroke is associated with diffuse Blood-Brain barrier dysfunction. Ann Neurol 65, 194–202. https://doi.org/10.1002/ana.21549

Waymont, J.M.J., Valdés Hernández, M. del C., Bernal, J., Duarte Coello, R., Brown, R., Chappell, F.M., Ballerini, L., Wardlaw, J.M., 2024. Systematic review and meta-analysis of automated methods for quantifying enlarged perivascular spaces in the brain. Neuroimage. https://doi.org/10.1016/j.neuroimage.2024.120685

Wiltgen, T., McGinnis, J., Schlaeger, S., Kofler, F., Voon, C.C., Berthele, A., Bischl, D., Grundl, L., Will, N., Metz, M., Schinz, D., Sepp, D., Prucker, P., Schmitz-Koep,

B., Zimmer, C., Menze, B., Rueckert, D., Hemmer, B., Kirschke, J., Mühlau, M., Wiestler, B., 2024. LST-AI: A deep learning ensemble for accurate MS lesion segmentation. Neuroimage Clin 42. https://doi.org/10.1016/j.nicl.2024.103611

Yamamoto, E.A., Bagley, J.H., Geltzeiler, M., Sanusi, O.R., Dogan, A., Liu, J.J., Piantino, J., 2024. The perivascular space is a conduit for cerebrospinal fluid flow in humans: A proof-of-principle report. Proc Natl Acad Sci U S A 121. https://doi.org/10.1073/pnas.2407246121

Zhang, J., Gao, Y., Park, S.H., Zong, X., Lin, W., Shen, D., 2017. Structured Learning for 3-D Perivascular Space Segmentation Using Vascular Features. IEEE Trans Biomed Eng 64, 2803–2812. https://doi.org/10.1109/TBME.2016.2638918