Leveraging Ellipsoid Bounding Shapes and Fast R-CNN for Enlarged Perivascular Spaces Detection and Segmentation

Mariam Zabihi[1,2][0000−0002−7083−2318], Marleen de Bruijne[3,4], Silvia Ingala[5,6], Luigi Lorenzini[6], Jorge Cardoso[7], Robin Camarasa[5], Chayanin Tangwiriyasakul[7], Frederik Barkhof[6,8][0000−0003−3543−3706], and Carole Sudre[1][0000−0001−5753−428X]

[1] Department of Population Science and Experimental Medicine, University College London, UK
m.zabihi,c.sudre@ucl.ac.uk
[2] Donders Institue, Radboud University, The Netherlands
m.zabihi@donders.ru.nl
[3] Department of Radiology and Nuclear Medicine, Erasmus Medical Center Rotterdam, The Netherlands
marleen.debruijne,r.camarasa@erasmusmc.nl
[4] Department of Computer Science, University of Copenhagen, Denmark
[5] Department of Radiology, Copenhagen University Hospital Rigshopitalet, Denmark
[6] Department of Radiology and Nuclear Medicine, VU University Medical Center, The Netherlands
s.ingala,l.lorenzini @amsterdamumc.nl
[7] School of Biomedical Engineering and Imaging Sciences, King's College London, UK
m.jorge.cardoso,chayanin.tangwiriyasakul@kcl.ac.uk
[8] Department of Medical Physics and Biomedical Engineering, University College London, UK
f.barkhof@ucl.ac.uk

Abstract. Enlarged perivascular spaces (EPVS) are small fluid-filled spaces surrounding blood vessels in the brain. They have been found to be important in the development and progression of cerebrovascular disease, including stroke, dementia, and cerebral small vessel disease. Their accurate detection and quantification are crucial for early diagnosis and better management of these diseases.
In recent years, object detection techniques such as Mask R-CNN approach have been widely used to automate the detection and segmentation of small objects. To account for the tubular shape of these markers we use ellipsoid shapes instead of bounding boxes to express the location of individual elements in the implementation of the Fast R-CNN. We investigate the performance of this model under different modality combinations and find that the T2 modality alone, as well as the combination of T1+T2, deliver better performance.

Keywords: Ellipsoid bounding shapes · Ellipsoid bounding shapes · Fast R-CNN · Cerebrovascular diseases · enlarged perivascular spaces

1 Introduction

Enlarged perivascular spaces (EPVS) are minute, fluid-filled cavities that cradle the blood vessels coursing through the brain. Despite their minuscule size, their presence is associated with cognitive decline and dementia onset. They are also associated with the presence and progression of other markers of cerebral small vessel disease such as microbleeds and white matter hyperintensities (WMH)[1–3]. Accurate detection and quantification of these diminutive markers are crucial, as they pave the way for early diagnosis and improved management strategies for such ailments. Although mainly tubular, EPVS are highly diverse in spatial distribution. This variation in location adds a degree of difficulty in EPVS detection, that, combined with their small size renders their process arduous. In addition, while they are often visually graded on T2-weighed images, EPVS may be mimicked by that of white matter hyperintensities leading to possible misdetection of punctuate WMH as EPVS due to similarities in intensity signatures.

3D object detection models are revolutionizing medical imaging by precisely identifying and localizing anomalies in volumetric scans such as CT, MRI, and mammograms. These models excel in detecting larger structures like breast cancer lesions, brain tumors, or lung nodules, assisting in early diagnosis [4–6]. However, for smaller features like Enlarged Perivascular Spaces (EPVS), traditional 3D detection methods may not provide the same level of accuracy due to their small size and complex morphology. Despite these challenges, recent advancements in 3D object detection methods have significantly propelled the field of automated EPVS detection[7]. A larger scale study[8] offered an automated method for detecting and quantifying enlarged PVS across a broad population base and studying the relationship between EPVS and cognitive decline. In parallel, Rashid et al. [9] used a multi-scale 3D convolutional neural network incorporated deep learning to detect enlarged PVS in brain MRI scans accurately.[10] used distance transforms to detect EPVS, however, this is based on point annotations only and hence, it does not capture the shape. Despite these advances, common to many of these methods is the use of axis-oriented bounding boxes for the identification of EPVS. The tubular nature of PVS structures may lead to inaccuracies in this approach.

To bridge this gap, we developed a model in this paper, bringing together the strengths of advanced object detection methods and more representative geometric shapes. We delve into a thorough discussion of the implications and benefits of our approach, highlighting how it stands as a valuable contribution to the evolving field of automated EPVS detection. In light of its potential to provide a more accurate and efficient way to quantify and analyze EPVS, our approach opens new pathways for robust and reliable neurological research, contributing significantly to the ongoing efforts in understanding cerebrovascular disease.

2 Methods

2.1 Framework

In our study, we employed a 3D Region Convolutional Neural Network (RCNN) composed of four key stages for the detection and segmentation of Enlarged Perivascular Spaces (EPVS).

Feature Learning: The first step is to train a 3D convolutional U-Net architecture [11] using our input images. This U-Net serves as the backbone network and is tasked with learning the key features from these images. A distance map of the objects of interest forms the target for this network, enhancing its ability to extract relevant characteristics.

Region Proposal Network (RPN): The feature maps derived from the U-Net are then supplied to a Region Proposal Network (RPN). The RPN, essentially a 3D convolutional network, uses these feature maps to generate score maps. These score maps depict the probability of an object's presence at each voxel location. Hence, the RPN primarily serves as a classifier, identifying potential regions in the 3D space that are likely to contain target objects.

Non-Maximum Suppression: A non-maximum suppression (NMS) process is then applied to these score maps. This process includes score thresholding to filter out regions of interest (ROIs) with low confidence. The ROIs that survive the thresholding are extracted as ellipsoids and subjected to pairwise comparison. If the overlap between any two ellipsoids exceeds a certain threshold, the ellipsoid with the lower score is suppressed. This ensures that the remaining, or surviving, ellipsoids are those with high confidence and minimal overlap, optimizing the selection process for the ROIs. We used the Hellinger distance[12] to measure the overlap between ROIs (when overall > 0.2), and we used the Euclidean distance from the center of elements as an additional measure (when the distance > 2 voxels).

Object Detection and Segmentation: The selected ellipsoids are then provided as input to another convolutional network for detailed object detection. This network includes four layers that help to fine-tune the mean object classification and object shape.

Finally, to enhance segmentation results, the distance map generated from the U-Net is integrated with the convolutional network just before its final layer. This step aids a subsequent segmentation network in generating a segmented representation of the original 3D image.

2.2 Object Shape Encoding

To represent the shape of each candidate object, we adopted a seven-parameter simplified encompassing ellipsoid. Given the expected tubular shape of EPVS, we only considered the first eigenvalue to provide us with the scale of the investigated element. Specifically, the largest eigenvalue was used for the indication of scale along the main axis, the first two components of the associated eigenvector to indicate direction, and the fractional anisotropy value of the associated tensor

to represent the spread in the plane perpendicular to the main axis. The center of mass parameters (x,y,z) encoded the location of the elements of interest. From these ellipsoid characteristics, the smallest encompassing patches are iteratively fed into the RCNN component to confirm the presence of an element of interest and to perform the final segmentation. The method's pipeline is shown in Figure 1).

2.3 Data

The Where is VALDO? challenge was run in 2021 as satellite event of MICCAI[13]. It featured 3 tasks focusing on the detection and segmentation of small markers of cerebral small vessel disease namely EPVS, cerebral microbleeds, and lacunes. For this study, we used the six subjects from the SABRE dataset available for training. The elements were segmented manually using structural MR sequences coregistered to the 1mm3 isotropic T1-weighted sequence (T1-weighted, T2-weighted, T2 FLAIR) by two raters. The final label was generated by taking the union of the objects that have been annotated by both raters. Furthermore, only objects larger than 2 voxels were considered, resulting in a database comprising 1864 EPVS elements.

3 Experiments and Results

For performance assessment, the code made available for the VALDO challenge was applied providing 2 outputs related to detection (Absolute element difference and F1 score) and 2 outputs related to segmentation quality (Mean dice over true positive elements and absolute volume difference). An element is considered a true positive if the Intersection over Union with the ground truth is more than 0.10. We addressed the following aspects in our experiments: 1) Segmentation performance according to modality combination 2) Segmentation and detection of performance variability according to location. 3) Segmentation performance on the VALDO test set (SABRE component).

To ascertain which imaging modality is more informative for EPVS detection, we trained the model using different modalities, including T1, T2, and FLAIR, as well as all possible combinations of these modalities. This approach resulted in a total of seven distinct models for evaluation. Given our limited sample size, we employed leave-one-out cross-validation. Therefore, for each modality combination, we generated six different models.

During training, each model was fed with random 3D patches (size = 64) of the training images. Figure 2 shows an example of the gold standard and the model's output in T2 modality. Figure 3 displays the final segmented image. The out-of-sample results presented in Table 1 show the average performance across these cross-validated models for the different modality combinations. We applied the T1+T2 model to data from the testset of the VALDO challenge including only 3D segmentations (SABRE), which resulted in an F1 score of $.14 \pm .5$ and a Dice coefficient of $.25 \pm .3$.

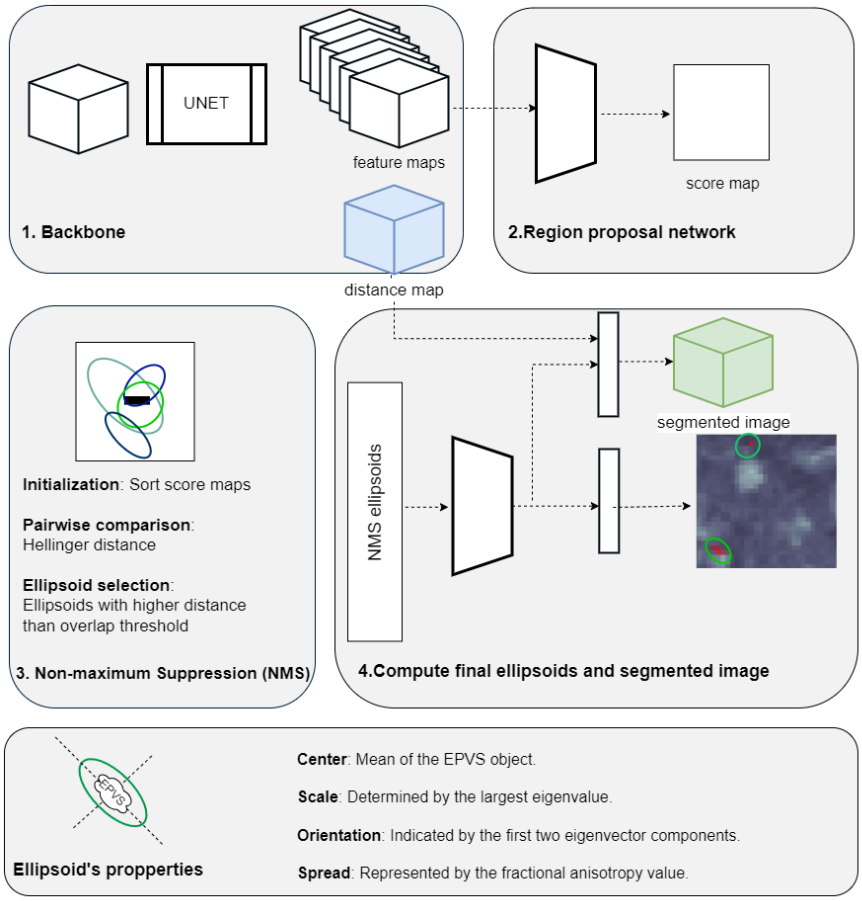

Fig. 1. Method's overview: (1) The 3D image undergoes processing via U-Net, yielding feature and distance maps. (2) These maps feed into the Region Proposal Network (RPN) to produce score maps. (3) Non-Maximum Suppression (NMS) sorts these score maps, filtering out low-confidence ROIs, leaving behind high-confidence, minimally overlapping ellipsoids. (4) These ellipsoids are used by a convolutional network, guided by the distance map from U-Net, for precise object detection.

3.1 EPVS detection according to location: centrum semi-ovale vs basal ganglia

EPVS can occur anywhere in the brain but regional considerations are often adopted separating in particular basal ganglia (BG) from the centrum semi-ovale (CSO) but also hippocampus and midbrain. While the EPVS load across regions is highly correlated and generally increases with age, differences in the association with risk factors are observed across regions. For instance, the relationship with hypertension is particularly notable in the basal ganglia [14]. Figure 4 shows the

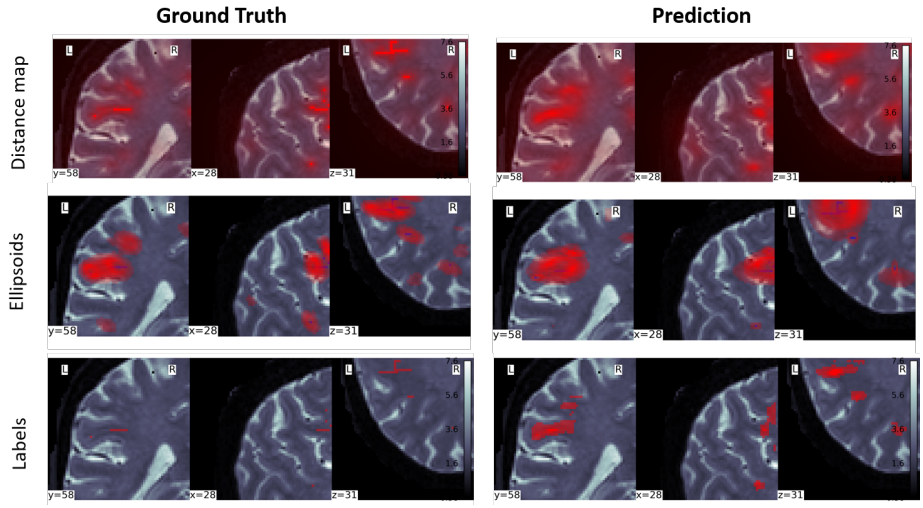

Fig. 2. Comparative visualization of gold standard and predictive model outputs on an example patch. The first column exhibits the 'gold standard', expert-annotated examples for comparison, the second column displays the corresponding outputs generated by our 3D-UNET (distance map) and 3D RCNN models. All images are overlaid on T2-weighted MR image patches.

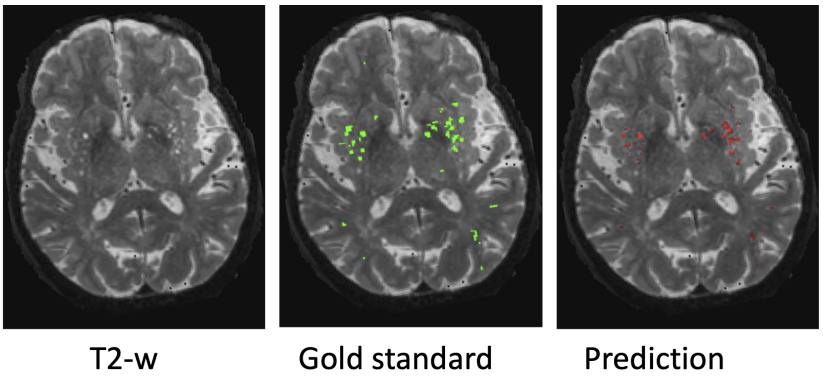

Fig. 3. Segmentation visualization Overlaid on a Full 3D T2-weighted MRI Scan.

comparative ratios of the voxel-wise appearance of EPVS and the corresponding total number of EPVS in the BG and CSO regions for the gold standard and prediction, true positive prediction (of modality: T1+T2). Table 2 provides a comprehensive summary of the performance of the model for different modality combinations, when tested on the left-out samples in a cross-validation setting.

Table 1. Model Performance Summary (mean and standard deviation across six subjects for each measure) for the different modality combinations on the left-out sample in the cross-validation setting.

Modality	Absolute difference	F1	Dice	Absolute volume difference
T1	167 ± 78	0.16 ± 0.04	0.39 ± 0.03	2014 ± 1815
T2	132 ± 80	0.21 ± 0.04	$\mathbf{0.43 \pm 0.04}$	$\mathbf{2107 \pm 1978}$
FLAIR	192 ± 105	0.02 ± 0.01	0.21 ± 0.1	2759 ± 225
T1+T2	$\mathbf{114 \pm 82}$	$\mathbf{0.22 \pm 0.03}$	$\mathbf{0.43 \pm 0.04}$	2180 ± 1810
T1+FLAIR	188 ± 91	0.09 ± 0.09	0.35 ± 0.05	2529 ± 2409
T2+FLAIR	151 ± 104	0.13 ± 0.06	0.37 ± 0.04	2366 ± 2109
T1+T2+FLAIR	140 ± 78	0.20 ± 0.05	$\mathbf{0.43 \pm 0.04}$	2642 ± 1435

4 Discussion

Our findings indicate that among the examined modalities, T2, and specifically the combination of T1 and T2, combination of modalities used by the annotators, showed superior performance. This suggests that T2-weighted imaging provides more relevant information for the identification of EPVS within the framework of our study and aligns with successful strategies used elsewhere (e.g winner of VALDO challenge).
Our results demonstrate that the model's performance in detecting EPVS elements was slightly superior in ganglionic regions compared to others. This may be attributed to the higher prevalence of EPVS voxels in these regions, thereby affecting the balance of examples during training. The difference in performance may be related to 1) the smaller amount of training examples in the CSO compared to the BG region and 2) the higher variability of the background tissue when comparing CSO to BG. This could be addressed in the future by either optimizing the patch sampling or separating the training for the two regions.
Given our multi-stage approach, tracking down the model's performance across different stages provides insights into where improvements can be made. For instance, engineering the sampling during training to ensure that regions like the CSO are adequately represented could be a beneficial adaptation. While our method currently achieves a lower performance than the contenders of the VALDO challenge overall for the SABRE dataset (winner with F1 around 0.35), it was similar to inter-rater F1 and performance in the BG appears higher. In addition, the drop in performance we observed between cross-validation testing and hold-out data testing may be due to our choice to only train on elements larger than 3 voxels and on examples where raters agreed.

Furthermore, this model, which takes into account the unique geometric morphology of EPVS, uses four parameters to encapsulate the shape of each candidate object. This offers a good geometrical fit to the general presentation of EPVS, a factor that could be clinically relevant when capturing the shape of EPVS.
Future direction: Despite the promising results, there are still opportunities

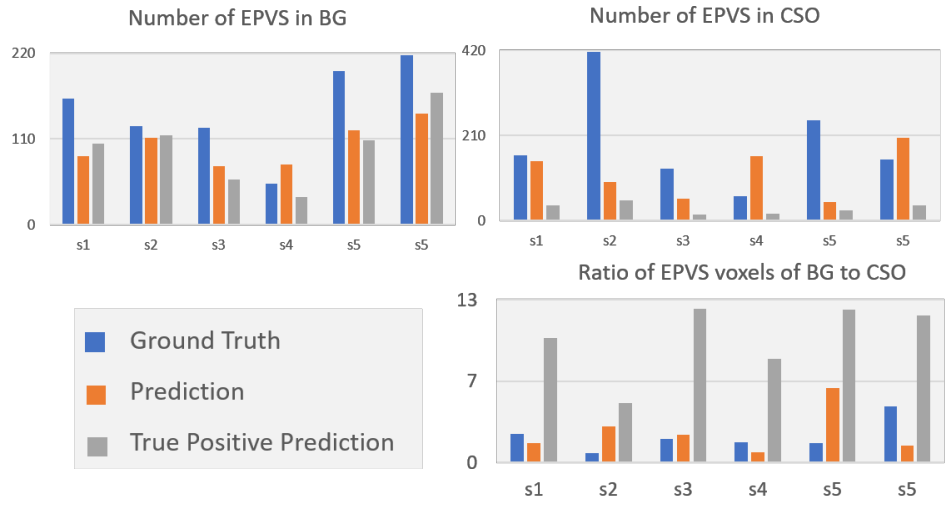

Fig. 4. Top-left: the number of EPVS in the BG, Top-right: the number of EPVS in CSO, Bottom: voxel-wise ratios of EPVS in BG to CSO for the gold standard, prediction, and true positive prediction (T1+T2 model).

Table 2. Model Performance Summary (mean and standard deviation across six subjects for each measure) for the different modality combinations on the left-out sample in the cross-validation setting for Basel Ganglia and CSO regions.

Basal Ganglia				
Modality	Absolute difference	F1	Dice	Absolute volume difference
T1	53 ± 21	0.29 ± 0.08	0.39 ± 0.04	872 ± 777
T2	37 ± 15	0.35 ± 0.05	$\mathbf{0.46 \pm 0.05}$	923 ± 737
FLAIR	79 ± 34	0.05 ± 0.04	0.18 ± 0.14	1453 ± 1411
T1+T2	$\mathbf{28 \pm 16}$	$\mathbf{0.37 \pm 0.05}$	$\mathbf{0.47 \pm 0.03}$	1003 ± 353
T1+FLAIR	69 ± 28	0.14 ± 0.14	0.31 ± 0.09	1070 ± 476
T2+FLAIR	50 ± 22	0.22 ± 0.08	0.40 ± 0.04	1284 ± 750
T1+T2+FLAIR	39 ± 18	0.32 ± 0.09	0.43 ± 0.05	1294 ± 296
CSO				
Modality	Absolute difference	F1	Dice	Absolute volume difference
T1	118 ± 82	0.07 ± 0.01	$\mathbf{0.43 \pm 0.04}$	1069 ± 1120
T2	101 ± 88	0.09 ± 0.04	0.34 ± 0.00	992 ± 1156
FLAIR	123 ± 96	0.01 ± 0.01	0.14 ± 0.18	1096 ± 1173
T1+T2	$\mathbf{98 \pm 86}$	$\mathbf{0.1 \pm 0.03}$	0.36 ± 0.06	1186 ± 1053
T1+FLAIR	123 ± 88	0.03 ± 0.04	0.39 ± 0.33	1057 ± 958
T2+FLAIR	113 ± 98	0.05 ± 0.02	0.41 ± 0.15	1092 ± 1171
T1+T2+FLAIR	110 ± 82	0.09 ± 0.03	0.41 ± 0.04	1100 ± 1045

for further optimization. Refinements could be made through fine-tuning model parameters, such as the number and types of features extracted, to enhance the precision of EPVS detection. Another potential improvement could involve implementing smart sampling based on the frequency of EPVS appearance in specific regions. By incorporating a prior probability map of EPVS occurrences, the overall process could be significantly improved. Lastly, we noted a tendency for oversegmentation in most of our models which could be alleviated by some restriction on the expected individual element volume.
Moreover, our approach could be extended beyond the detection of EPVS. Given the prevalence of elongated or irregularly shaped structures in various biological contexts, our model could be useful for the detection and analysis of other biomarkers. This positions our model as a potentially powerful tool in the broader field of medical imaging analysis.

References

1. Bown C.W: Physiology and Clinical Relevance of Enlarged Perivascular Spaces in the Aging Brain. Neurology **98**(3) 107-117 (2022)
2. Paradise M.: Association of Dilated Perivascular Spaces With Cognitive Decline and Incident Dementia. Neurology **96**(11) 1501-1511 (2021)
3. Ding J. : Large Perivascular Spaces Visible on Magnetic Resonance Imaging, Cerebral Small Vessel Disease Progression, and Risk of Dementia: The Age, Gene/Environment Susceptibility–Reykjavik Study. JAMA Neurology**74**(9) 1105–1112 (2017)
4. Asgari Taghanaki S.: Deep semantic segmentation of natural and medical images: a review. Artificial Intelligence Review**54** 137–178 (2021)
5. Ranjbarzadeh R. :Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools. Computers in Biology and Medicine **152**(2023)
6. Ribli D. : Detecting and classifying lesions in mammograms with Deep Learning. Scientific Reports**8**(1: 4165 (2018)
7. Williamson, B.: Automated Grading of Enlarged Perivascular Spaces in Clinical Imaging Data of an Acute Stroke Cohort Using an Interpretable, 3D Deep Learning Framework. Scientific Reports **12** (1) 1-7 (2023)
8. Dubost, F.: Enlarged perivascular spaces in brain MRI: Automated quantification in four regions. NeuroImage **185** 534-544(2019)
9. Rashid, T.: Deep Learning Based Detection of Enlarged Perivascular Spaces on Brain MRI. Neuroimage: Reports **3**(1) 100162(2023)
10. van Wijnen, K.M.H.: Automated Lesion Detection by Regressing Intensity-Based Distance with a Neural Network. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019
11. Çiçek, Ö.: 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016.
12. Fu, GH.: Hellinger distance-based stable sparse feature selection for high-dimensional class-imbalanced data. BMC Bioinformatics **21**(121) (2020).
13. Sudre, Carole H.: Where is VALDO? VAscular lesions detection and segmentation challenge at MICCAI 2021. arXiv preprint arXiv:2208.07167 (2022).

14. Wardlaw, J.M.: Perivascular spaces in the brain: anatomy, physiology and pathology. Nature Reviews Neurology **16**, 137–153 (2020).