University of Dundee

Perivascular Spaces Segmentation in Brain MRI Using Optimal 3D Filtering

Ballerini, Lucia; Lovreglio, Ruggiero; Valdés Hernández, Maria Del C.; Ramirez, Joel; MacIntosh, Bradley J.; Black, Sandra E.

SCIENTIFIC REPORTS

OPEN

Perivascular Spaces Segmentation in Brain MRI Using Optimal 3D Filtering

Lucia Ballerini[1], Ruggiero Lovreglio[2], Maria del C. Valdés Hernández[1], Joel Ramirez[3], Bradley J. MacIntosh[3], Sandra E. Black[3] & Joanna M. Wardlaw[1]

Perivascular Spaces (PVS) are a feature of Small Vessel Disease (SVD), and are an important part of the brain's circulation and glymphatic drainage system. Quantitative analysis of PVS on Magnetic Resonance Images (MRI) is important for understanding their relationship with neurological diseases. In this work, we propose a segmentation technique based on the 3D Frangi filtering for extraction of PVS from MRI. We used ordered logit models and visual rating scales as alternative ground truth for Frangi filter parameter optimization and evaluation. We optimized and validated our proposed models on two independent cohorts, a dementia sample (N = 20) and patients who previously had mild to moderate stroke (N = 48). Results demonstrate the robustness and generalisability of our segmentation method. Segmentation-based PVS burden estimates correlated well with neuroradiological assessments (Spearman's $\rho = 0.74$, p < 0.001), supporting the potential of our proposed method.

Perivascular spaces (PVS), also known as Virchow-Robin spaces, are fluid-filled spaces that follow the typical course of cerebral penetrating vessels. PVS have the same Magnetic Resonance Imaging (MRI) contrast characteristics as Cerebrospinal Fluid (CSF), that is they appear hypointense (dark) on T1-weighted (T1) and hyperintense (bright) on T2-weighted images (T2)[1,2]. They appear as small 3D tubular structures that, depending on the viewing plane, are linear or round, with a diameter generally smaller than 3 mm[3] (see Fig. 1).

Enlargement of perivascular spaces is associated with other morphological features of Small Vessel Disease (SVD) such as white matter hyperintensities and lacunes[4]; cognitive impairment[5] and inflammation[6]. Most studies use visual rating scales to assess PVS burden[7,8], but these are prone to inter-observer variability, particularly in the Centrum Semiovale, due to the coexistence of PVS with other neuroradiological features of SVD that confound their identification in this region[7].

Efforts have been made to computationally assess PVS[3,9]. Recent semi-automatic methods are based on thresholding and require user intervention either for the choice of parameters or for manual editing of the resulting masks, which, for small and frequent features such as PVS, risks introducing inter-observer variability and is very time consuming[10,11]. A promising approach proposed for PVS automatic segmentation uses the Frangi filter[12] parameterised through a Random Forest scheme[13] that learns discriminative PVS characteristics from manually segmented ground truth on MR images acquired at 7 T[14–16]. However, MRI in clinical research and practice is mostly performed in scanners with field strengths at 1.5 T or 3 T, and the reference standards available are visual ratings performed by neuroradiologists, which restricts the learning-based approach proposed by Park et al.[15] in practice. Moreover, it is difficult to assess enlarged PVS burden at high field 7 T MRI since normal PVS and deep medullary veins with similar intensities to PVS confound visualization and requires observer correction.

Our current goal is to present a segmentation approach for enlarged PVS that can be used widely in current clinical research studies, to further elucidate their pathological significance and assess their potential role in neurological disorders. The main innovation of this paper is a method for optimization and evaluation of the filter in absence of ground truth segmentation. In other words, this method allows to use labels requiring little annotation effort to derive much finer results (pixel-wise segmentation).

[1]Department of Neuroimaging Sciences, Centre for Clinical Brain Sciences, Centre for Cognitive Ageing and Cognitive Epidemiology and UK Dementia Research Institute Edinburgh Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [2]Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand. [3]Hurvitz Brain Sciences Program, LC Campbell Cognitive Neurology Research Unit, Heart and Stroke Foundation Canadian Partnership for Stroke Recovery, Sunnybrook Research Institute and the University of Toronto, Toronto, Ontario, Canada. Correspondence and requests for materials should be addressed to L.B. (email: lucia.ballerini@ed.ac.uk).
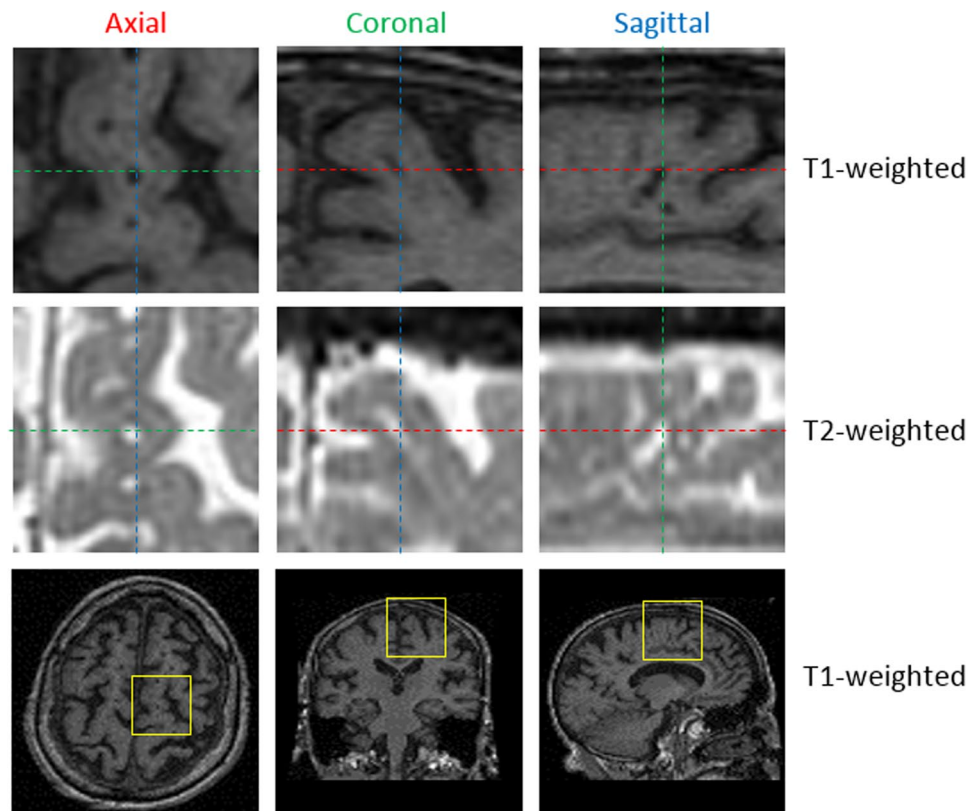
Figure 1. Magnified view of PVS in an axial, coronal and sagittal slice in T1-weighted and T2-weighted MR images. The position of these zooms in corresponding T1-weighted brain scans is highlighted with yellow squares (bottom row).

We propose a novel application of ordered logit models, usually used in statistics as a regression model for ordinal dependent variables, as this model provides a good estimate for capturing the sources of influence that explain the ordinal dependent variables (i.e. in this case the PVS visual rating scores) considering the uncertainty (i.e. subjectivity, inter-observer variability) in the measurement of such data[17]. We use this model to estimate the parameters of the Frangi filter[12] to obtain the maximum likelihood of a vessel-like structure to be a PVS in the Centrum Semiovale, by also estimating the count of PVS that most likely falls in the class corresponding to the category given by the neuroradiologist in this brain region.

We calibrated different ordered logit models, according to the rating scale available for every dataset. We optimized the parameters of the Frangi filter to deal with T1-weighted (T1W) and T2-weighted (T2W) modalities, and combined the resulting filtered images. Validation was carried out on different cohorts, using images acquired in 2 different sites, rated by 3 different raters.

Materials

Two datasets were used for developing, testing and validating the method:

1. *Sunnybrook Dementia Study* (SDS): a large registered ongoing longitudinal clinical trial conducted at Sunnybrook Health Science Centre, Toronto, Canada (ClinicalTrials.gov NCT01800214). The study has been approved by the Sunnybrook Research Ethics Board in accordance with the principles expressed in the Declaration of Helsinki. Each patient provided informed consent. Patients had an historical profile typical of Alzheimer's disease (AD). Full study details have been published previously[10].
2. *Mild Stroke Study* (MSS): a study conducted at Centre for Clinical Brain Science, Edinburgh, UK. Patients had clinical features of lacunar or mild cortical stroke. All experimental protocols were approved by the Lothian Ethics of Medical Research Committee (REC 09/81101/54) and the NHS Lothian R + D Office (2009/W/NEU/14) and conducted according to the principles expressed in the Declaration of Helsinki. All patients gave written informed consent. The MRI protocol has been published elsewhere[18].

The characteristics of the sequences relevant for PVS assessment are summarized in Table 1.

Methods

Observing the vessel-like structure of PVS, we propose a segmentation technique based on the 3D Frangi filtering[12], largely used for enhancing blood vessels, for instance in retinal images[19]. Given the absence of an accurate computational "ground truth" (i.e. manual labels of each PVS by experts), we propose a modelling technique to

Study/ parameters		Matrix	Voxel size (mm³)	TE (ms)	TR (ms)	flip angle	FOV (cm)	bandwidth (KHz)	Acq. time (min)
SDS	T1	256 × 256 × 124	0.86 × 0.86 × 1.4	5	35	35°	22	31.26	11
	T2/PD*	256 × 256 × 58	0.78 × 0.78 × 3	30/80	3000	90°	20	11.36	12
MSS	T1	256 × 216 × 256	1.02 × 0.9 × 1.02	14	400	—	24	15.63	0.54
	T2	384 × 224 × 28	0.47 × 0.47 × 6	90	6000	—	24	50	2.30
	T2cube	512 × 512 × 256	0.47 × 0.47 × 0.7	143	3000	—	24	31.25	3.55

Table 1. Characteristics of the relevant MRI sequences of (1) Sunnybrook Dementia Study (SDS)[10], and (2) Mild Stroke Study (MSS)[18]. *PD = proton density (interleaved), available in SDS.
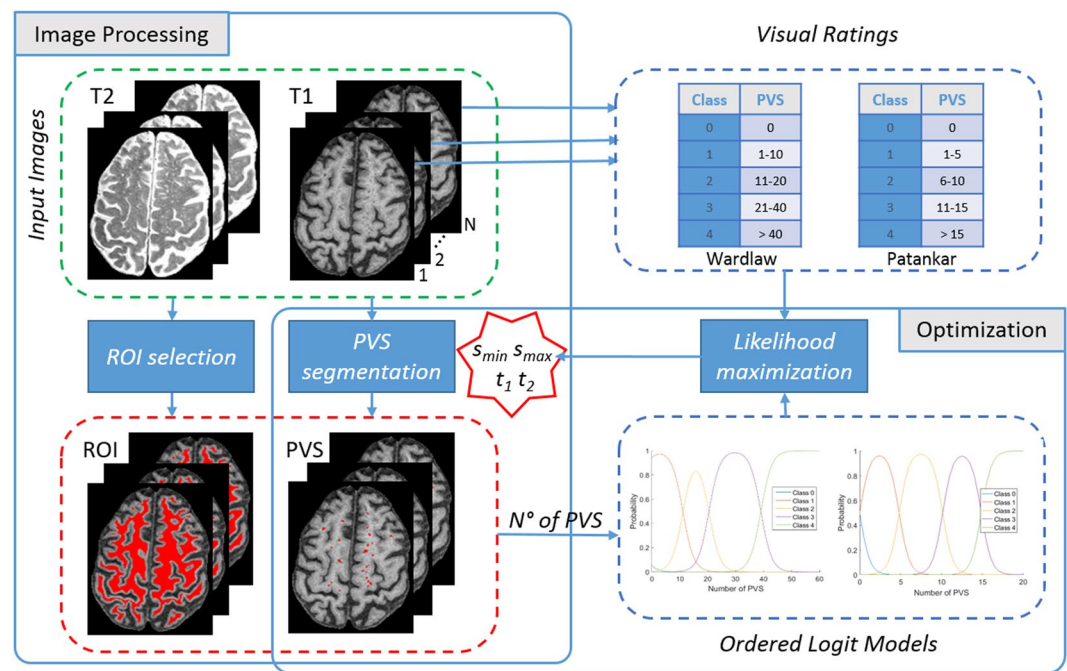

Figure 2. Framework of the proposed optimization approach: Frangi filter parameters (s_{min}, s_{max}) and thresholds (t_1, t_2) are optimized with order logit models and visual rating scales.

use the available information (i.e. PVS burden assessed using visual rating scales) to optimize the filter parameters. For this scope, an ordered logit model[17] has been used to simulate the relationship between the number of PVS and the rating categories, taking into account the uncertainty in the measurements. The framework of the proposed optimization process is illustrated in Fig. 2.

PVS assessment. PVS masks, obtained as described in Ramirez et al.[10], were available for the SDS dataset. These masks obtained using Lesion Explorer[20], which implements 2 false positive minimization strategies: (i) in order to reduce errors from minor imaging artifacts and improve differentiation from lacunar infarcts, candidate PVS are required to satisfy acceptance criteria from both T1W and T2W, and rejection criteria from PD, and (ii) to address potential registration errors and partial volume effects, the cortical Gray Matter segmentation was dilated by 1 voxel. This resulted in a relatively conservative estimate of the overall PVS burden and thus, limited its utility as a Ground Truth (GT) for segmentation optimization, as well as for pixel-wise evaluation of the results.

Two established visual rating scales for PVS severity were used in the present work. Previous work has demonstrated their comparability[7,10].

The visual rating scale developed by Potter et al.[7] (in the following called Wardlaw scale) required users to rate PVS burden on T2-weighted MRI in each of three major anatomical brain regions: midbrain, basal ganglia and centrum semiovale. According to the online user guide (http://www.sbirc.ed.ac.uk/documents/epvs-rating-scale-user-guide.pdf), PVS in the latter region should be assessed in the slice and hemisphere with the highest number, and rated as 0 (no PVS), 1 (mild; 1–10 PVS), 2 (moderate; 11–20 PVS), 3 (frequent; 21–40 PVS) or 4 (severe; >40 PVS).

The PVS scores proposed by Patankar et al.[21] were based principally on the appearances seen on T1W inversion recovery images. PVS should be scored in the centrum semiovale as 0 (none), 1 (less than five per side), 2 (more than five on one or both sides), reflecting the lesser visibility of PVS on T1W.

Two slightly modified versions of these rating methods, as previously described[10], were also used in this work. Coregistered MRIs were used for assessment, with T2W for primary identification, T1W for confirmation, and Proton Density (PD) for rejection as required. To reduce ceiling effects and account for a greater range of PVS, the Patankar scale was standardized: 0 (none), 1 (one to five), 2 (six to ten), 3 (eleven to fifteen), 4 (sixteen or more). To reduce double-counting, a slice increment of 3 was implemented as a standardized rating protocol. Centrum Semiovale was defined as the White Matter (WM) projections superior to the ventricles, present in each of the cerebral hemispheres under the cerebral cortex.

PVS were assessed in 20 representative cases of the SDS dataset by three raters: two experienced neuroradiologists using the two modified Wardlaw and Patankar visual rating scales, and a third rater strictly following the guideline of the original Wardlaw[7] and Patankar[21] rating methods. The two ratings (modified Wardlaw and Patankar) of the first raters were close to the conservative estimate of PVS burden obtained as described above. Inter-rater reliability was high (ICC = 0.99, <0.001) as previously discussed[10]. The third rater counted all visible PVS in the slice with the highest number in T1W and T2W, including the very small ones discarded by the first raters. All raters were blind to each other.

Frangi filter. Frangi[12] analyses the second order derivatives of an image I, defined in the Hessian matrix $H_s(v)$ as:

$$H_s(v) = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix}$$

(1)

to describe the "vesselness" $F(v)$ of a voxel v at scale s as:

$$F_s(v) = \begin{cases} 0 & \text{if } \lambda_2 \geq 0 \\ & \text{or } \lambda_3 \geq 0, \\ (1 - e^{-\frac{R_A^2}{2\alpha^2}}) \cdot e^{-\frac{R_B^2}{2\beta^2}} \cdot (1 - e^{-\frac{S^2}{2c^2}}) & \text{otherwise,} \end{cases}$$

(2)

where λ_1, λ_2 and λ_3 are the ordered eigenvalues ($|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$) of the Hessian matrix, $R_A = |\lambda_2|/|\lambda_3|$, $R_B = |\lambda_1|/(|\lambda_2\lambda_3|)^{1/2}$, $S = (\lambda_1^2 + \lambda_2^2 + \lambda_3^2)^{1/2}$, and α, β, c are thresholds which control the sensitivity of the filter to the measures R_A, R_B and S.

For a bright tubular structure in a 3D image we expect: $|\lambda_1| \leq |\lambda_2|, |\lambda_3|$ and $|\lambda_2| \sim |\lambda_3|; |\lambda_1| \sim 0$ and $\lambda_2, \lambda_3 \leq 0$. For a dark structure $\lambda_2, \lambda_3 \geq 0$ and the conditions in Eq. 2 should be reversed.

Given a set of scales $s \in [s_{min}, s_{max}]$, the responses are combined as:

$$F(v) = \max_s F_s(v)$$

(3)

where s_{min} and s_{max} are the minimum and maximum scales at which relevant structures are expected to be found[12].

Ordered Logit Model. An ordered logit model defines the relationship between an ordinal variable (y) which can vary between 0 and $m (m \in N^+)$, and the vector of independent variables (x) by using a latent continuous variable (y^*) defined in an one-dimensional space characterized by threshold points ($\mu_0, \ldots, \mu_{m-1}$) as described in equation:

$$y^* = \beta x + \varepsilon, \ \varepsilon \sim G(\mu|\sigma), \ \mu = 0, \ \sigma = \pi/\sqrt{3}$$

(4)

$$\begin{aligned} y_i &= 0 \quad \text{if} \quad -\infty < y_i^* \leq \mu_0 \\ y_i &= 1 \quad \text{if} \quad \mu_0 < y_i^* \leq \mu_1 \\ &\cdots \\ y_i &= m \quad \text{if} \quad \mu_{m-1} < y_i^* \leq \infty \end{aligned}$$

(5)

where β and μ_i are parameters to be estimated, ε is the error component which has a logistic random distribution with expected value equal to 0 and variance equal to $\pi/\sqrt{3}$, that accounts for the measurement error. This modelling approach provides a relevant methodology for capturing the sources of influence (independent variables) that explain an ordinal variable (dependent variable) taking into account the measurement uncertainty of such data[17].

Since y^* is not a deterministic quantity, it is only possible to define the probability to belong to each class:

$$\begin{aligned} P(y = j|\bar{x}) &= P(\mu_{j-1} < \bar{y}^* \leq \mu_j) \\ &= L(\mu_j - \beta\bar{x}) - L(\mu_{j-1} - \beta\bar{x}), \ j = 0 \div m \end{aligned}$$

(6)

where L is the logistic cumulative distribution function.

In our work, the ordinal variable (y) is the rating class (from 0 to 4) and the independent variable (x) is the number of PVS.
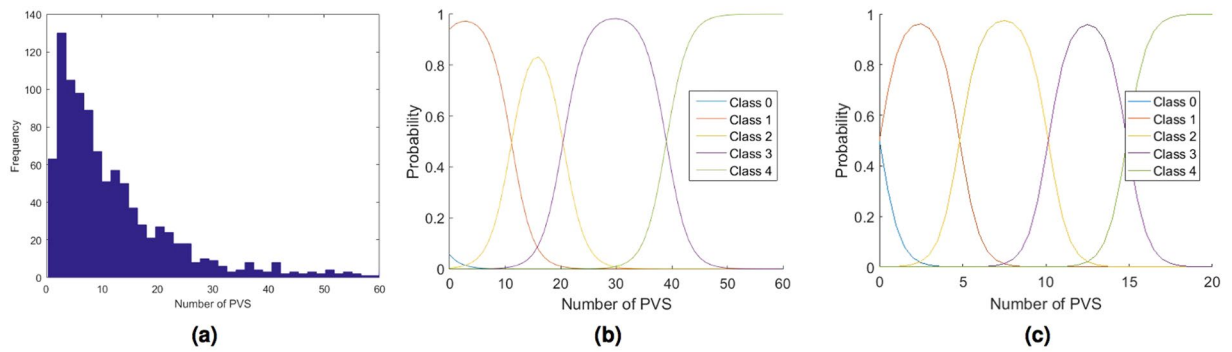
Figure 3. PVS distribution (**a**) of the synthetic dataset generated to calibrate the ordered logit model. Estimated ordered logit model for the Wardlaw (**b**) and the Patankar (**c**) rating scale.

Model Calibration. The ordered logit model has been calibrated by maximizing a likelihood function based on a synthetic dataset generated in 3 steps. In the first step 1000 numbers of PVS Count (PC_i, $i = 1, ..., 1000$) have been generated using a log-normal distribution (see Fig. 3a), that reflects the observed PVS distribution in known datasets[11]. In the second step, the uncertainty has been simulated for each PC_i casting a New value of PVS Count (NPC_i) using a normal distribution with mean equal to PC_i and standard deviation equal to one. Therefore, the probability that NPC_i is included between $PC_i - 3$ and $PC_i + 3$ is 0.997. These values reflect our measurements uncertainty[11]. In the third step, a Rating Class (RC_{ij}) has been assigned to each generated NPC_i.

Assuming m classes, the log-likelihood function can be written as:

$$LogL(\mu, \beta) = \sum_{i=1}^{1000}\sum_{j=1}^{m} P(y = j|NPC_i)RC_{ij}$$

(7)

where RC_{ij} is equal to one if the ith generated number belong to the jth rating class and it is equal to zero otherwise. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm has been used to estimate the ordered logit parameters.

For the Wardlaw scale[7], a rating class from 0 to 4, being 0(*none*), 1(1–10), 2(11–20), 3(21–40), 4(>40) PVS, has been assigned to each generated number. The estimated parameters are $\beta = 0.514$, $\mu_0 = -2.840$, $\mu_1 = 5.708$, $\mu_2 = 10.497$, $\mu_3 = 20.040$, and the model is illustrated in Fig. 3b.

For the Patankar scale[10] a rating class from 0 to 4, being 0(*none*), 1(1–5), 2(6–10), 3(11–15), 4(>15) PVS, has been assigned to each generated number. The estimated parameters for the Patankar rating scale are $\beta = 1.906$, $\mu_0 = 2.269$, $\mu_1 = 9.569$, $\mu_2 = 18.995$, $\mu_3 = 28.639$, and the model is illustrated in Fig. 3c.

Image Preprocessing. Images were preprocessed to generate the Region-of-Interest (ROI) masks. A fuzzy C-means clustering algorithm was applied to T1 images[22]. This is an unsupervised iterative clustering technique that effectively assigns each voxel to one of 4 membership classes: background, Cerebrospinal Fluid (CSF), Gray Matter (GM), and White Matter (WM). After a series of morphological and thresholding operations, the CSF and GM re-labelled voxels were combined to generate the final CSFGM mask which was used for false positive minimization. To avoid PVS mislabelled as GM, an hole filling procedure was used. The Centrum Semiovale (CS) was automatically identified as the region of WM, superior to the lateral ventricles previously obtained using Lesion Explorer[20]. In this paper we focused on the CS rather than the Basal Ganglia (BG), due to the availability of these ROI masks.

Parameter Optimization. In order to apply the 3D Frangi filtering, the coregistered MRI volumes were first resliced to make 1 *mm* isotropic voxels using linear interpolation. Then volumes have been filtered according to Eqs (2) and (3) and voxels having F(v) larger than a threshold t were kept. The two segmentations from T1W and T2W modalities were combined using an AND operation. PVS were identified as the tubular structures with lengths between 3 and 50 *mm*[3,11], using 3D connected component analysis with 18-neighbourhood rule. This provided the initial PVS binary masks. For each slice we calculated the PVS density as the area of the PVS mask divided by the area of the CS mask. We automatically selected the slice in the CS with highest density of PVS. This slice corresponded to the representative slice having the highest number of PVS selected by the radiologist for assessing the Wardlaw visual ratings[7]. The count of PVS in this slice was derived automatically with 2D connected component labelling. Similarly, the total number of PVS in the entire CS was obtained with 3D connected component labelling. This count of PVS corresponded to the count performed by the radiologist for the Patankar ratings[21].

A log-likelihood function has been defined to optimize the segmentation parameters: Frangi filter scales s_{min}, s_{max} and threshold t. In this work, we used the default configuration for the other Frangi filter parameters ($\alpha = 0.5$, $\beta = 0.5$, $c = 500$), as in our previous work[23] we noted that optimizing these parameters produced essentially similar results, at the cost of a much higher computational time.

	rater 1	rater 2	rater 3	
modified Wardlaw	◯	✓		exp 1
modified Patankar	◯	✓		exp 2
original Wardlaw			◯	exp 3
original Patankar			✓	exp 3

Table 2. Ratings available for the Sunnybrook Dementia Study (SDS): ◯ optimization scale/rater, ✓ validation scale/rater.

	s_{min}	s_{max}	t_1	t_2
min value	0.2	2.0	0.90	0.05
max value	2.0	4.0	0.99	0.50
increment	0.2	0.2	0.01	0.05

Table 3. Range of the segmentation parameters to optimize.

Based on the count of PVS ($x_i(s_{min}, s_{max}, t)$) for each case i we obtained the probabilities of each case i to belong to the five rating classes ($P(y = j|x_i), j = 0, \ldots, 4$) using the ordered logit model. The PVS visual rating category provided by an expert radiologist was then used to select a probability for each i case ($\overline{P_i}$). The sum of the logarithms of these selected probabilities is the log-likelihood function to maximize:

$$LogL(s_{min}, s_{max}, t) = \sum_{i=1}^{N} log(\overline{P_i})$$

(8)

where N is the number of cases.

Model Validation. Segmentation procedures are commonly evaluated by assessing the voxel-wise spatial agreement between two binary masks, one obtained by the automatic method and a manual one. In our case, the manual segmentation of PVS was not available, as it would have been a very tedious and time consuming task to manually annotate these tiny structures in a reasonable size dataset. Therefore the true number of PVS was also not available. Quantitative comparison with other methods[9,11] was unfeasible as they have been applied to MR images having different resolution, acquired using different protocols in different cohorts.

The performance of the models was therefore evaluated comparing single-slice PVS automatic counting on segmented images vs validated visual ratings using Spearman's ρ (statistical analysis were performed using MATLAB Robust correlation toolbox[24]). Correspondence of PVS total count and volume vs. visual ratings was also assessed to test generalizability.

Experiments and Results

For developing and optimizing the segmentation approach, the imaging datasets of 20 representative subjects were selected from a sample of the Sunnybrook Dementia Study (SDS)[10]. These 20 subjects had visual ratings assessed by three raters as summarized in Table 2.

The optimization procedure has been applied to T1W and T2W MRI sequences. Frangi filter scales (s_{min} and s_{max}) and two thresholds (t_1 and t_2, one for each modality) have been simultaneously optimized. The 2 binary masks obtained were combined using an AND operation. The range of the parameters that undergo the optimization process has been defined as in Table 3.

The high computational time needed to simultaneously optimize multiple parameters is a common drawback of optimization processes. Indeed, each log-likelihood function evaluation implies filtering all the training samples using Eq. (2), which may become critical in this 3D case. To keep a reasonable computational time, in this research contribution, we limited the search space to a subsets of parameters and roamed through this space using a systematic grid search.

Three sets of experiments were performed as indicated in Table 2. The symbol ◯ indicates the scale/rater used for optimization, while ✓ specifies those used for validation.

For illustration Fig. 4a and b show the surface plots for the parameter optimization using the modified Wardlaw rating scale for a range of examined s_{min} and s_{max} scales and t_1 values. Figure 4c show the trend of the log-likelihood function (LogL) for a range of examined threshold t_2 values with the the best combination of s_{min}, s_{max} and t_1. Figure 5 show the surface plots for the parameter optimization using the modified Patankar rating scale. The optimal parameters obtained with the 2 models are very similar ($s_{min} = 1.4$, $s_{max} = 3.2$, $t_1 = 0.96$, $t_2 = 0.35$ for the first model, $s_{min} = 1.4$, $s_{max} = 3.2$, $t_1 = 0.95$, $t_2 = 0.35$ for the second one). From the plots we can observe that the most significant parameter of the Frangi filter is the minimum scale (s_{min}). From these plots it is also clear that the Frangi filter was needed. Indeed for any combination of s_{min} and s_{max} the threshold values play a smaller role.

The optimal parameters obtained with the Wardlaw model using PVS assessed by the third rater were slightly different from the previous ones ($s_{min} = 0.2$, $s_{max} = 2$, $t_1 = 0.96$, $t_2 = 0.1$). The plots of the log-likelihood function are shown in Fig. 6. The trend of plots confirms the validity of the model demonstrating that the model was able to adapt to the rater, and finds the best parameters to segment the PVS accounted by that rater.
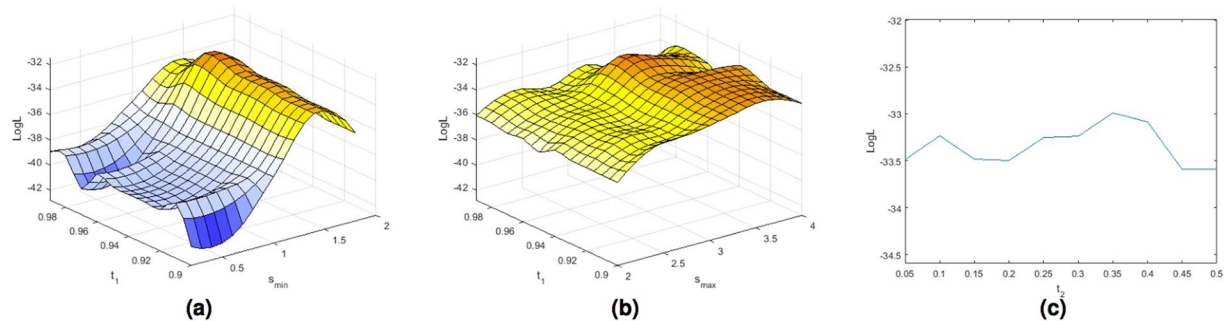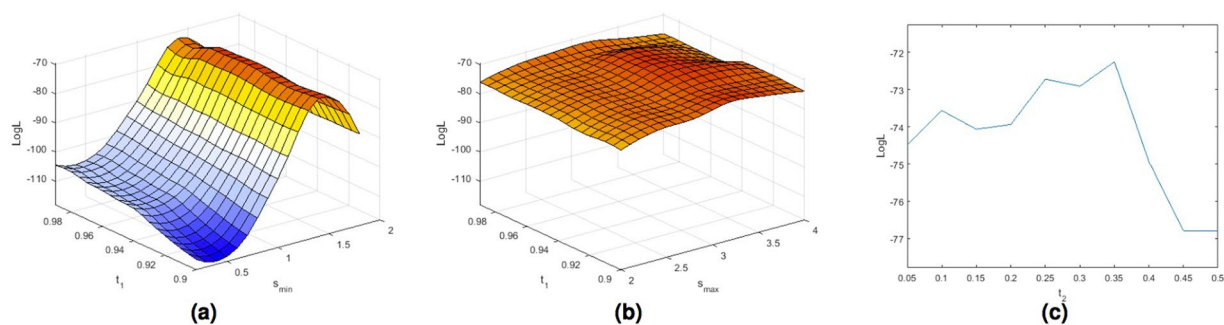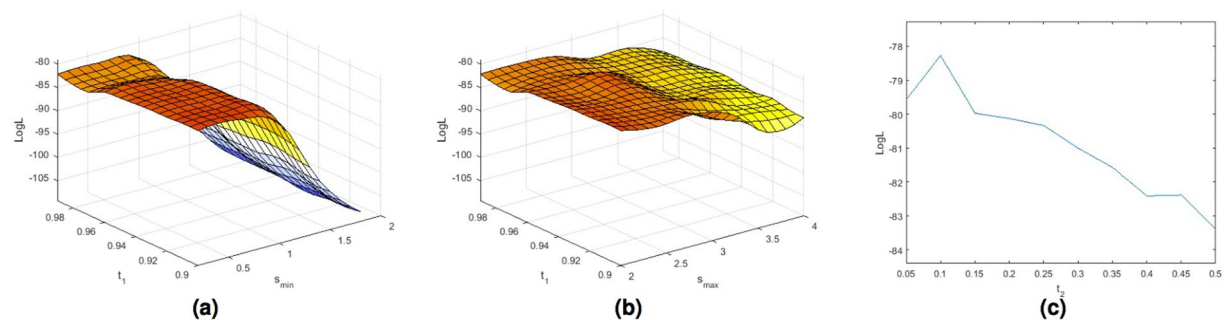
Figure 4. Plots of the log-likelihood function (LogL) obtained using the ordered logit model shown in Fig. 3b estimated with the Wardlaw ratings for a range of examined s_{min} (**a**) and s_{max} (**b**) scales and thresholds t_1 values, and threshold t_2 (**c**) values with the best combination of s_{min}, s_{max} and t_1.

Figure 5. Plots of the log-likelihood function (LogL) obtained using the ordered logit model shown in Fig. 3c estimated with the Patankar ratings for a range of examined s_{min} (**a**) and s_{max} (**b**) scales and thresholds t_1 values, and threshold t_2 (**c**) values with the best combination of s_{min}, s_{max} and t_1.

Figure 6. Plots of the log-likelihood function (LogL) obtained using the ordered logit model shown in Fig. 3b estimated with the Wardlaw ratings for a range of examined s_{min} (**a**) and s_{max} (**b**) scales and thresholds t_1 values, and threshold t_2 (**c**) values with the best combination of s_{min}, s_{max} and t_1.

Qualitative Evaluation.   Magnified views of PVS segmentation using the threshold-based method previously described[10] and the proposed method are shown in Fig. 7. It is clear that the proposed method detected most of the PVS, including the tiny ones, thanks to the enhancement of tubular structure performed by the Frangi filtering using the appropriate scale. The threshold based method missed them, as it was forced to be conservative in order to distinguish PVS from confounding tissue boundaries.

Examples of segmented PVS for two representative SDS cases having few and many PVS are shown in Figs 8 and 9. For each case, we show T1W, T2W and the PVS overlay in red. Volume rendering of the segmented PVS for two cases having few and many PVS are shown in Fig. 10 for visual qualitative evaluation.

Quantitative Evaluation.   When comparing single-slice PVS count obtained from segmented images with the modified Wardlaw and Patankar visual ratings of the second rater, a fair correlation was found for both methods (Spearman's $\rho = 0.58$, $p = 0.006$ and $\rho = 0.71$, $p = 0.0004$ respectively). However, low and no significant
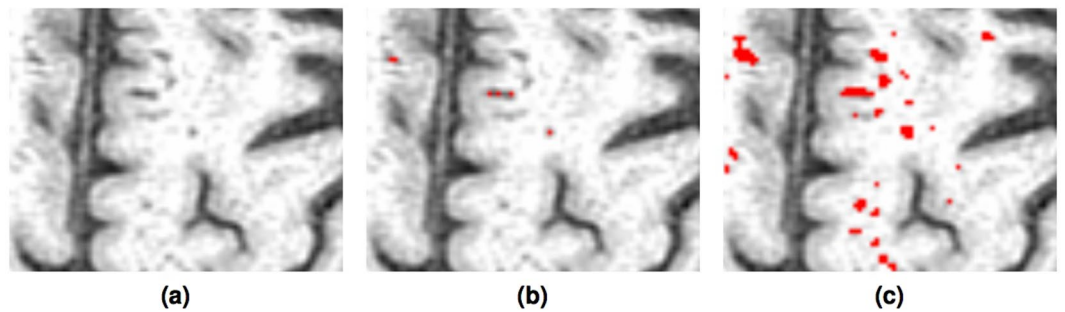
Figure 7. Visual comparison of the PVS segmentation overlaid on T1 (**a**) using the conservative threshold based[10] method (**b**) and the proposed Frangi filtered (**c**) method.
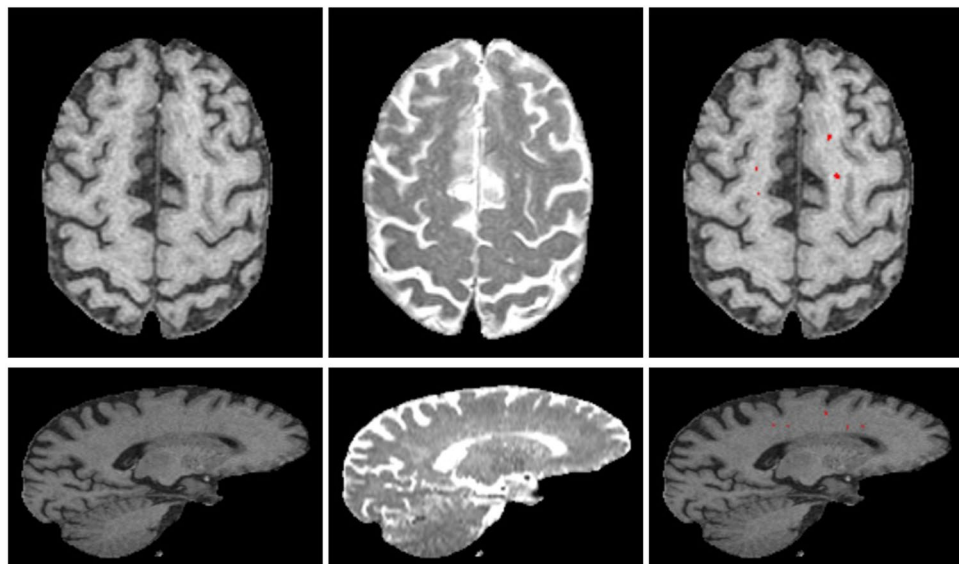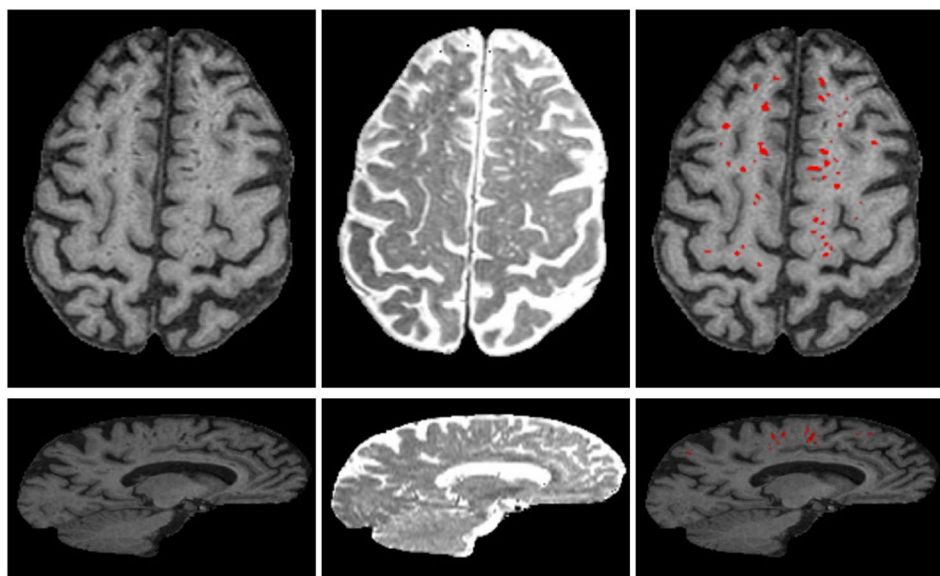

**Figure 8.** Examples of the final PVS segmentation a case of SDS dataset having few PVS. Axial (top row) and sagittal (bottom) slice of T1W, T2W and PVS overlay (red) on T1W. For illustration, T1W is shown in its native space ($256 \times 256 \times 124$) and T2W is shown registered to T1W.

Figure 9. Examples of the final PVS segmentation for a case of SDS dataset having many PVS. Axial (top row) and sagittal (bottom) slice of T1W, T2W and PVS overlay (red) on T1W.
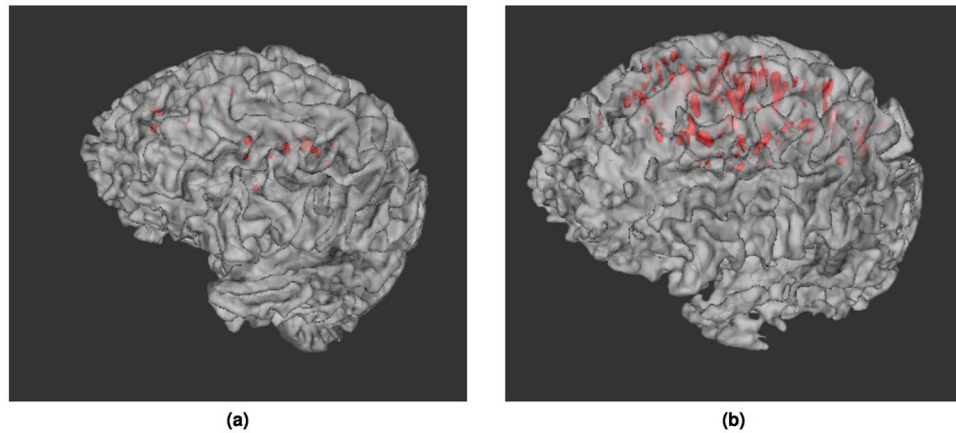
Figure 10. Volume rendering of segmented PVS (red) for two SDS cases having few (**a**) and many (**b**) PVS. PVS volumes overlayed onto a surface render of the brain.

correlation was found with total PVS number in volume in Centrum Semiovale, suggesting low generalizability. This replicates our previous analysis[10].

For the segmentation results obtained with the optimal parameters of the model optimized with the original Wardlaw scale a stronger correlation between single-slice PVS count vs visual ratings was found (Spearman's $\rho = 0.74$, $p = 0.0002$). In addition PVS total count and volume correlates with visual rating scores (Spearman's $\rho = 0.67$, $p = 0.001$ and $\rho = 0.53$, $p = 0.015$, respectively).

Application to alternative acquisitions. To validate the new PVS segmentation method we applied it to MRI of cases of the Mild Stroke Study (MSS). Visual ratings using the Wardlaw rating[7] were available for all the cases[4].

Automatic brain, cerebrospinal fluid (CSF) and normal-appearing white matter extraction were performed on T1W MRI using optiBET[25] and FSL-FAST[26] respectively. All subcortical structures were segmented, also automatically, using other tools from the FMRIB Software Library (FSL) and an age-relevant template as per the pipeline described elsewhere[18]. After identifying the lateral ventricles as the CSF-filled structures with boundaries with the subcortical structures, the CS was identified as the region of normal-appearing white matter, superior to the lateral ventricles, present in each of the cerebral hemispheres under the cerebral cortex. T1W sequence and CS region were linearly registered to the T2W-cube images[27]. This preprocessing differs from the one used for the SDS dataset due to the pipelines available at the two research groups[10,18]. The optimization procedure has been applied to T2-cube MRI sequences of 20 patients, and tested on 48 patients of the same study. The optimal parameters obtained for this dataset ($s_{min} = 0.4$, $s_{max} = 3.6$, $t_2 = 0.4$) were different from those for the SDS dataset. This confirm the method was able to adapt the parameters to the different voxel-size.

PVS total count and volume correlates with visual rating scores (Spearman's $\rho = 0.47$, $p < 0.001$ and $\rho = 0.57$, $p < 0.001$, respectively). Scatter plots of these associations are shown in Fig. 11a and b. Condensed raw PVS computational count into the same categories of the visual rating scale has a similar distribution of the visual rating scores, as shown in Fig. 11c.

The results of this experiment suggest fair generalizability of the output of the segmentation method vs validated visual rating scores.

Discussion

The 3D Frangi filter enhances and captures the 3D geometrical shape of PVS, thus this method shows promise for identifying and quantifying PVS that run both longitudinally and transversally in the Centrum Semiovale, avoiding the double-counting limitations of slice-based methods. Centrum Semiovale is more difficult to rate visually than Basal Ganglia, so future application of this method to Basal Ganglia may be more straightforward. The ordered logit model could deal with the measurement uncertainty and the unequal class intervals of the rating scores.

One limitation of this method is that it relies on the image preprocessing step for the ROI masks. If the masks provided by this step are not accurate, the method can detect as PVS boundary of grey matter and gyri. Another limitation of this method is that it requires high resolution and quasi isotropic structural MRI. Very noisy images have been excluded for this study, otherwise any noise spot of tubular shape can be wrongly segmented as PVS. This can be overcome by a learning method. However, learning methods require GT, and not just visual ratings assessment.

The method is fully automatic and therefore free from inter- and intra-rater variability. However, much more testing is required in a wider range of subjects including those with high burden of other ageing and neuroinflammation features. Visual checking and editing is likely to be needed in complex cases, but this remains to be defined.
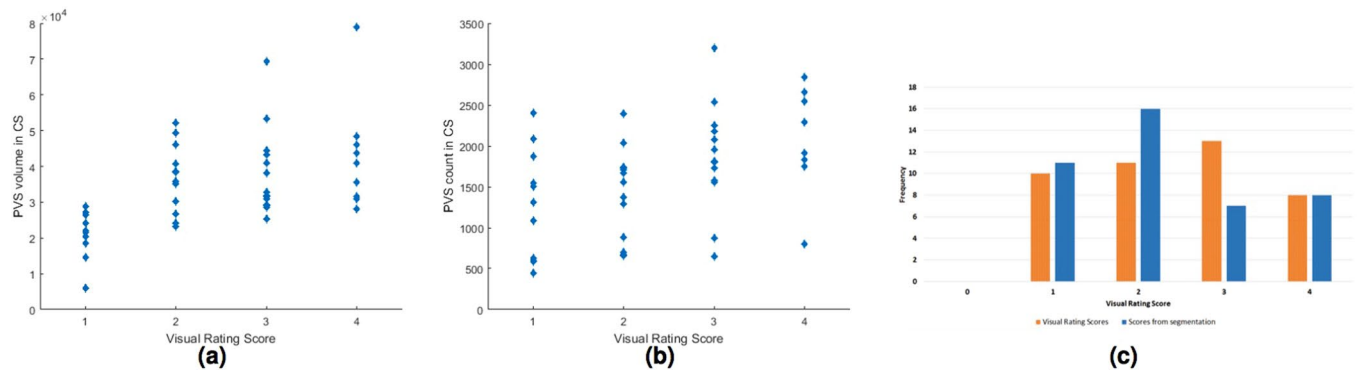
Figure 11. Associations between PVS computational total volume (**a**) and count (**b**) vs. PVS visual rating scores in centrum semiovale (CS) region for the 48 test cases of the MSS dataset. Comparison of PVS computational count condensed into a score of similar range to the visual rating categories (**c**).

The quantitative assessment of PVS volume and count is more suitable for longitudinal studies than visual ratings, that tend to be susceptible to ceiling/flooring effects. The accurate segmentation of PVS will allow the analysis of their spatial distribution, orientation and density. The resulting PVS masks could be used, in combinations with other quantitative sequences, to assess other tissue characteristics in adjacent tissue. Moreover, this method will enable the study of the spatial and volumetric relationships of PVS with other markers of SVD, e.g. acute lacunar infarcts, white matter hyperintensities, lacunes, and microbleeds. Additionally, this method shows promise for use in longitudinal studies where PVS burden can be assessed in relation to measures of cerebral blood brain barrier permeability, perfusion and cerebrovascular reactivity.

Conclusions

We presented an automatic method for 3D segmentation of PVS in conventional brain MRI. The novelty of this work is the fact that the ordered logit model allows use of the visual ratings for Frangi filter parameter optimization in absence of alternative computational ground truth. The automatically segmented PVS count and volume agree with visual ratings. Quantitative measurements will better characterize the severity of PVS in ageing people and their associations with dementia, stroke and vascular diseases. This is the first work to propose a multicentre study of PVS segmentation. It shows excellent multi-centre reproducibility.

References

1. Wardlaw, J. M. *et al*. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *The Lancet Neurology* **12**, 822–838 (2013).
2. Ramirez, J. *et al*. Imaging the Perivascular Space as a Potential Biomarker of Neurovascular and Neurodegenerative Diseases. *Cellular and Molecular Neurobiology* **36**, 289–299 (2016).
3. Valdés Hernández, M. d. C., Piper, R. J., Wang, X., Deary, I. J. & Wardlaw, J. M. 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 (2013).
4. Doubal, F. N., Maclullich, A. M. J., Ferguson, K. J., Dennis, M. S. & Wardlaw, J. M. Enlarged Perivascular Spaces on MRI Are a Feature of Cerebral Small Vessel Disease. *Stroke* **41**, 450–454 (2010).
5. Arba, F. *et al*. Enlarged perivascular spaces and cognitive impairment after stroke and transient ischemic attack. *International Journal of Stroke* **0**, 1–10 (2016).
6. Potter, G. M. *et al*. Enlarged perivascular spaces and cerebral small vessel disease. *International Journal of Stroke* **10**, 376–381 (2015).
7. Potter, G. M., Chappell, F. M., Morris, Z. & Wardlaw, J. M. Cerebral Perivascular Spaces Visible on Magnetic Resonance Imaging: Development of a Qualitative Rating Scale and its Observer Reliability. *Cerebrovascular Diseases* **39**, 224–231 (2015).
8. MacLullich, A. M. J. *et al*. Enlarged perivascular spaces are associated with cognitive function in healthy elderly men. *Journal of neurology, neurosurgery, and psychiatry* **75**, 1519–1523 (2004).
9. Descombes, X., Kruggel, F., Wollny, G. & Gertz, H. J. An Object-Based Approach for Detecting Small Brain Lesions: Application to Virchow-Robin Spaces. *IEEE Transactions on Medical Imaging* **23**, 246–255 (2004).
10. Ramirez, J. *et al*. Visible Virchow-Robin spaces on magnetic resonance imaging of Alzheimer's disease patients and normal elderly from the Sunnybrook Dementia study. *Journal of Alzheimer's Disease* **43**, 415–24 (2015).
11. Wang, X. *et al*. Development and initial evaluation of a semi-automatic approach to assess perivascular spaces on conventional magnetic resonance images. *Journal of Neuroscience Methods* **257**, 34–44 (2016).
12. Frangi, A., Niessen, W., Vincken, K. & Viergever, M. Multiscale vessel enhancement filtering. *Medical Image Computing and Computer-Assisted Intervention (MICCAI98)* 130–137 (1998).
13. Criminisi, A., Shotton, J. & Konukoglu, E. Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. *Microsoft Research Cambridge, Tech. Rep. MSRTR-2011-114* **5**, 12 (2011).
14. Zong, X., Park, S. H., Shen, D. & Lin, W. Visualization of perivascular spaces in the human brain at 7T: Sequence optimization and morphology characterization. *NeuroImage* **125**, 895–902 (2015).
15. Park, S. H., Zong, X., Gao, Y., Lin, W. & Shen, D. Segmentation of Perivascular Spaces in 7T MR Image using Auto-Context Model with Orientation-Normalized Features. *NeuroImage* **134**, 223–235 (2016).
16. Hou, Y. *et al*. Enhancement of Perivascular Spaces in 7T MR Image using Haar Transform of Non-local Cubes and Block-matching Filtering. *Scientific Reports* **7**, 8569 (2017).
17. Greene, W. H. & Hensher, D. A. *Modeling ordered choices: A primer* (Cambridge University Press, 2010).
18. Valdés Hernández, M. d. C. *et al*. Rationale, design and methodology of the image analysis protocol for studies of patients with cerebral small vessel disease and mild stroke. *Brain and Behavior* **415**, 1–18 (2015).

19. Lupascu, C. A., Tegolo, D. & Trucco, E. Fabc: Retinal vessel segmentation using adaboost. *IEEE Transactions on Information Technology in Biomedicine* **14**, 1267–1274 (2010).
20. Ramirez, J. *et al.* Lesion Explorer: A comprehensive segmentation and parcellation package to obtain regional volumetrics for subcortical hyperintensities and intracranial tissue. *NeuroImage* **54**, 963–973, https://doi.org/10.1016/j.neuroimage.2010.09.013 (2011).
21. Patankar, T. F. *et al.* Dilatation of the Virchow-Robin space is a sensitive indicator of cerebral microvascular disease: Study in elderly patients with dementia. *American Journal of Neuroradiology* **26**, 1512–1520 (2005).
22. Bezdek, J., Hall, L., Clark, M., Goldgof, D. B. & Clarke, L. Medical image analysis with fuzzy models. *Statistical Methods in Medical Research* **6**, 191–214 (1997).
23. Ballerini, L. *et al.* Application of the Ordered Logit Model to Optimising Frangi Filter Parameters for Segmentation of Perivascular Spaces. *Procedia Computer Science* **90**, 61–67 (2016).
24. Pernet, C., Rand, W. & Rousselet, G. Robust correlation analyses: false positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology* **3** (2013).
25. Lutkenhoff, E. S. *et al.* Optimized Brain Extraction for Pathological Brains (optiBET). *PLoS ONE* **9**, e115551 (2014).
26. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* **20**, 45–57 (2001).
27. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**, 825–841 (2002).

Acknowledgements

Author Contributions

L.B. wrote the manuscript, conducted the experiments and analysed the results. R.L. helped with the order logit models, M.V.H. and J.R. helped with image preprocessing and editing the manuscript. L.B., R.L., M.V.H., J.R., B.J.M., S.E.B., J.M.W. reviewed the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.