

**Enlarged perivascular spaces in brain MRI
Automated quantification in four regions**

Dubost, Florian; Yilmaz, Pinar; Adams, Heab; Bortsova, Gerda; Ikram, M. Arfan; Niessen, Wiro; Vernooij, Meike; de Bruijne, Marleen

DOI

[10.1016/j.neuroimage.2018.10.026](https://doi.org/10.1016/j.neuroimage.2018.10.026)

Publication date

2019

Document Version

Final published version

Published in

NeuroImage

Citation (APA)

Dubost, F., Yilmaz, P., Adams, H., Bortsova, G., Ikram, M. A., Niessen, W., Vernooij, M., & de Bruijne, M. (2019). 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>

Important note

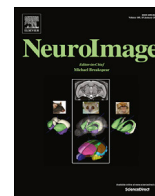
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Enlarged perivascular spaces in brain MRI: Automated quantification in four regions



Florian Dubost^{a,*}, Pinar Yilmaz^b, Hieab Adams^b, Gerda Bortsova^a, M. Arfan Ikram^c,
Wiro Niessen^{a,d}, Meike Vernooij^b, Marleen de Bruijne^{a,e,**}

^a Biomedical Imaging Group Rotterdam, Department of Radiology, Department of Medical Informatics, Erasmus MC - University Medical Center Rotterdam, the Netherlands

^b Department of Radiology, Department of Nuclear Medicine, Department of Epidemiology, Erasmus MC - University Medical Center Rotterdam, the Netherlands

^c Department of Radiology, Department of Epidemiology, Department of Neurology, Erasmus MC - University Medical Center Rotterdam, the Netherlands

^d Department of Imaging Physics, Faculty of Applied Science, TU Delft, Delft, the Netherlands

^e Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

ARTICLE INFO

Keywords:

Deep learning
Machine learning
Enlarged perivascular spaces
Virchow-Robin spaces
Perivascular spaces
Dementia

ABSTRACT

Enlarged perivascular spaces (PVS) are structural brain changes visible in MRI, are common in aging, and are considered a reflection of cerebral small vessel disease. As such, assessing the burden of PVS has promise as a brain imaging marker. Visual and manual scoring of PVS is a tedious and observer-dependent task. Automated methods would advance research into the etiology of PVS, could aid to assess what a “normal” burden is in aging, and could evaluate the potential of PVS as a biomarker of cerebral small vessel disease. In this work, we propose and evaluate an automated method to quantify PVS in the midbrain, hippocampi, basal ganglia and centrum semiovale. We also compare associations between (earlier established) determinants of PVS and visual PVS scores versus the automated PVS scores, to verify whether automated PVS scores could replace visual scoring of PVS in epidemiological and clinical studies. Our approach is a deep learning algorithm based on convolutional neural network regression, and is contingent on successful brain structure segmentation. In our work we used FreeSurfer segmentations. We trained and validated our method on T2-contrast MR images acquired from 2115 subjects participating in a population-based study. These scans were visually scored by an expert rater, who counted the number of PVS in each brain region. Agreement between visual and automated scores was found to be excellent for all four regions, with intraclass correlation coefficients (ICCs) between 0.75 and 0.88. These values were higher than the inter-observer agreement of visual scoring (ICCs between 0.62 and 0.80). Scan-rescan reproducibility was high (ICCs between 0.82 and 0.93). The association between 20 determinants of PVS, including aging, and the automated scores were similar to those between the same 20 determinants of PVS and visual scores. We conclude that this method may replace visual scoring and facilitate large epidemiological and clinical studies of PVS.

1. Introduction

This paper proposes and evaluates an algorithm for the automated quantification of enlarged perivascular spaces (PVS) in four brain regions. Perivascular spaces are fluid-filled areas surrounding cerebral arteries or veins. These spaces tend to enlarge locally in aging subjects (Wardlaw et al., 2013). Enlarged perivascular spaces can be identified as hyperintensities in T2-contrast MRI, as illustrated in Fig. 1. Though initially considered a strictly normal phenomenon, the presence of PVS is

increasingly thought to reflect the presence of cerebral small vessel disease and to function as a potential biomarker for various brain diseases such as dementia (Mills et al., 2007), stroke (Selvarajah et al., 2009), multiple sclerosis (Achiron and Faibel, 2002), and Parkinson (Zijlmans et al., 2004).

The progressive enlargement of PVS, their widespread occurrence in the brain, and presence of mimics with similar appearance on MRI make the manual annotation of individual PVS challenging and time consuming (see Fig. 2). Instead, current studies largely rely on visual

* Corresponding author.

** Corresponding author. Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC - University Medical Center Rotterdam, the Netherlands.

E-mail addresses: floriandubost1@gmail.com (F. Dubost), marleen.debruijne@erasmusmc.nl (M. de Bruijne).

<https://doi.org/10.1016/j.neuroimage.2018.10.026>

Received 28 July 2018; Received in revised form 20 September 2018; Accepted 9 October 2018

Available online 13 October 2018

1053-8119/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

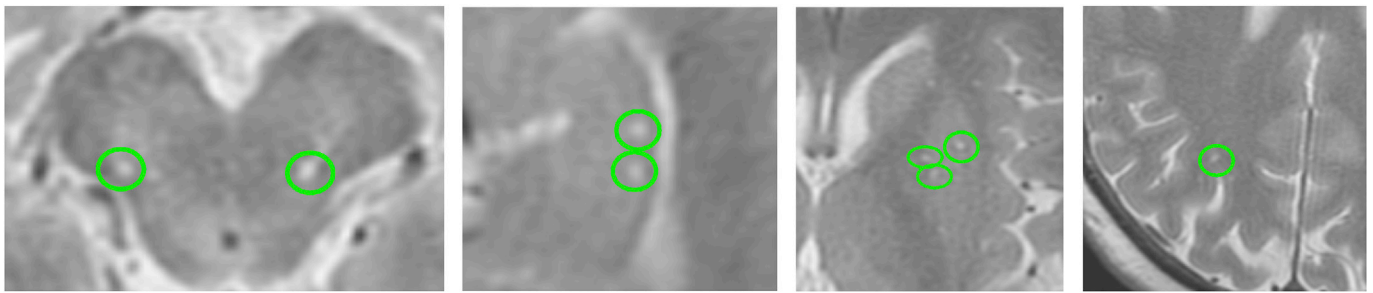


Fig. 1. Examples of enlarged perivascular spaces in different brain regions. T2-contrast MRI images in the axial view. PVS are circled in green. From left to right: midbrain, hippocampus, basal ganglia and centrum semiovale. On these images the PVS are relatively easy to detect for an expert rater, contrary to Fig. 2.

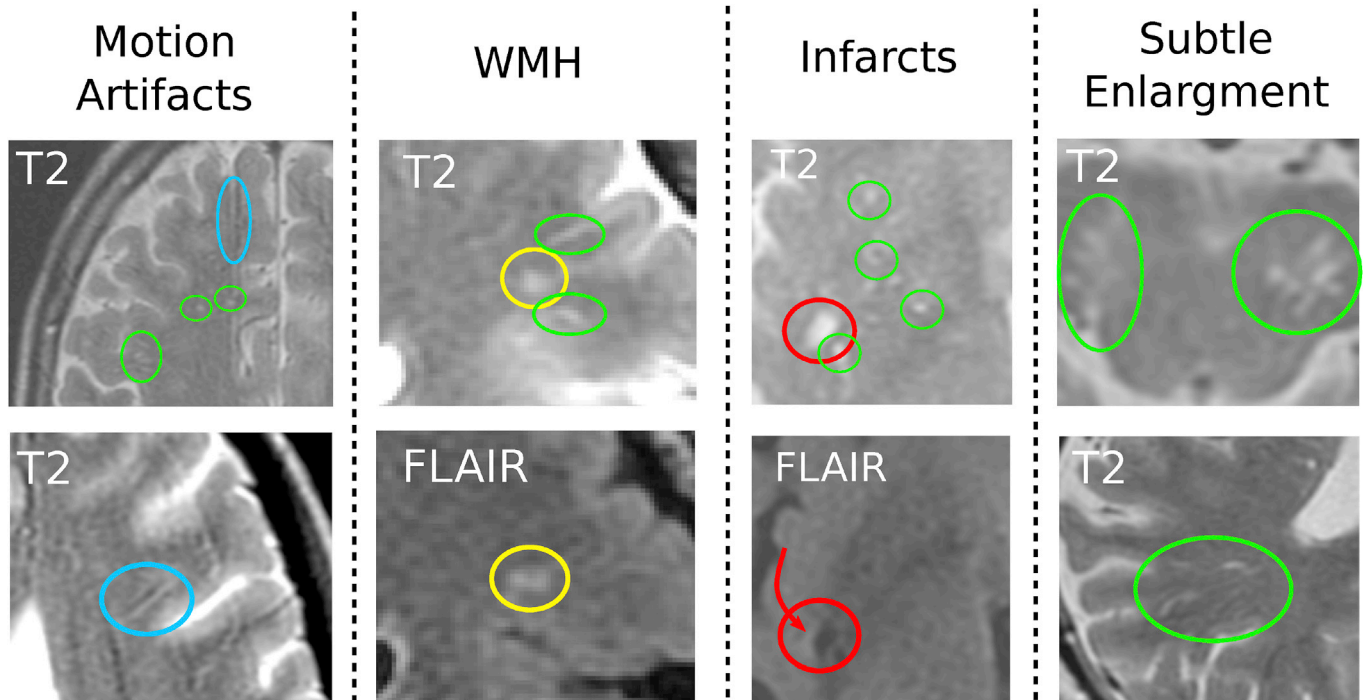


Fig. 2. Examples of enlarged perivascular spaces and their mimics in different brain regions. All images are in the axial view. PVS are circled in green, white matter hyperintensities (WMH) in yellow, lacunar infarcts in red and motion artifacts in blue. In the first column, motion artifact could be mistaken for an elongated PVS in the centrum semiovale. In the second column, the WMH could be mistaken for PVS. However on the FLAIR-weighted scan WMH are hyperintense, while PVS are hypointense and less visible (bottom image). In the third column, the lacunar infarct in the basal ganglia could be mistaken for a group of several PVS, which individual borders could not be seen because of partial volume effect (this lesion would unlikely be mistaken for a single EPVS because of its irregular shape). The FLAIR-weighted scan shows a hyperintense rim (red arrow) around the lesion, indicating the presence of a lacunar infarct. In the last column, the scans present several PVS, some of which are at the limit of being considered as enlarged. According to the visual scoring guidelines presented in Adams et al. (2013), to be considering enlarged, perivascular spaces should have a diameter larger than 1 mm. For many small perivascular spaces in these images, this is difficult to evaluate.

scoring systems. Two types of scoring systems have been proposed in the literature: expert raters either count the number of PVS within a region of interest (Adams et al., 2013, 2015) or categorize the PVS burden using a scale (e.g. Potter scores (Potter et al., 2015a) and Patankar scores (Patankar et al., 2005)). Adams PVS scores can be considered as a more finely graded version of Patankar and Potter scores, such that the number of categories in Adams PVS scores is equal to the number of PVS. Automated quantification of PVS would be preferred as it is more objective and faster than visual scoring. Furthermore, it would hold great potential to study burden of PVS as a continuous rather than a categorical measure, enabling to better disentangle “normal” structural brain changes in aging from a pathological load of PVS.

In a recent study (Dubost et al., 2018), we proposed a regression convolutional network to quantify PVS in the basal ganglia. In the present work, we extend this method to other brain regions most clinically

relevant for PVS quantification, namely the midbrain, hippocampi and centrum semiovale (Adams et al., 2013, 2015), and we provide a more elaborate evaluation.

Similar to the method described by Dubost et al. (2018), the input in our method is a T2-contrast brain scan, and the output is an automated PVS score. The aim of the method is to reproduce the visual scores of an expert rater, considered here as the reference standard. Our method uses a 3D convolutional neural network inspired by ResNet (He et al., 2016) and optimized with a mean squared error (MSE) loss function to minimize the difference between visual scores and predicted scores in a set of training images.

In all four brain regions, we compare the agreement between our automated PVS scores and the visual PVS score of the expert rater, with the level of inter-observer agreement. We assess scan-rescan reproducibility. Finally we check in a subset of 1485 scans whether the

associations between determinants of PVS and the automated scores are similar to those between the same determinants and visual scores. The determinants of PVS investigated here include demographics, cardiovascular risk factors, ApoE genotypes, and MRI markers.

1.1. Related work

Other researchers have published automated PVS quantification methods involving the use of the visual scores as ground truths.

Ballerini et al. (2018) proposed to enhance PVS in the centrum semiovale using multiscale vessel enhancement filtering (Frangi et al., 1998). The parameters of these filters are optimized with ordered logit models, using PVS category scores (Potter et al., 2015a; Patankar et al., 2005) as ground truth. To evaluate their methods, the authors compute correlations between the visual ratings and their segmentation-derived PVS count and PVS volume in two different datasets. This method has only been evaluated in the centrum semiovale. Results were mixed with correlations ranging from 0.47 to 0.74 in different datasets.

Gonzalez-Castro et al. (2017) addressed PVS quantification in the basal ganglia as a binary classification problem, where the objective is to discriminate between scans with few (≤ 10) or many (> 10) PVS. Their method uses support vector machines and bag-of-words descriptors. The agreement between their classifier and a human observer is similar to the inter-observer agreement. The authors also show associations between determinants of PVS (age, Fazekas scale, and presence of lacunar infarcts) and the binary score of the classifier. Our work extends this by proposing a continuous score indicating the number of PVS instead of a binary score, leading to a finer quantification. We evaluate our method in four brain regions, and investigate associations with a wider range of determinants.

Boespflug et al. (2017) proposed an automated quantification method based on the combination of image intensities and morphologic features (width, volume, and linearity) from several MRI sequences. They evaluate their method in the centrum semiovale.

Ramirez et al. (2015) used a semi-automated PVS segmentation method based on adaptive local intensity thresholding to study the difference of PVS burden in the centrum semiovale and basal ganglia between cognitively normal and Alzheimer subjects. The number of necessary user interactions can make this approach very time consuming.

2. Methods and materials

The objective of our method is to automatically predict the PVS visual scores. Our framework consists of two steps. We first extract the region of interest (ROI) (Section 2.2) and then apply a regression convolutional neural network (CNN) (Section 2.3) to compute the PVS score. The CNN is trained on an independent set of visually scored scans ($N = 400$ or $N = 1600$).

2.1. Data

In our experiments we used brain MRI scans from the Rotterdam Study. The Rotterdam Study is a prospective population study investigating - among others - neurological diseases in the middle aged and elderly, applying brain MRI in all participants (Ikram et al., 2017). In our experiments, we use 2115 scans of 2115 subjects, acquired between 2005 and 2011.

In addition, we used 60 other scans for which 30 study participants were scanned twice within a short period (19 ± 11 days). The 60 scans of this reproducibility set are not part of the 2115 scans mentioned above and were not visually scored for EPVS.

The Medical Ethics Committee of the Erasmus MC has approved the Rotterdam study, according to the Population Study Act, executed by the Ministry of Health, Welfare and Sports of the Netherlands. All participants provided written informed consent to participate in the study and for information to be obtained from their physicians.

2.1.1. MRI scan protocol

Brain MRI was performed on a 1.5-T MRI scanner (GE-Healthcare, Milwaukee, WI, USA) with an eight-channel head coil to obtain: T1-weighted (T1), T2-contrast (T2), fluid-attenuated inversion recovery (FLAIR) and T2*-weighted gradient-recalled-echo sequences.

To compute the automated PVS scores, we only used the T2 scans. These scans were acquired according to the following protocol: 12,300 ms repetition time, 17.3 ms echo time, 16.86 KHz bandwidth, 90–180° flip angle, 1.6 mm slice thickness, 25 cm² field of view, 416 × 256 matrix size. The images are reconstructed to a 512 × 512 × 192 matrix. The voxel resolution is 0.49 × 0.49 × 0.8mm³.

More details of the imaging protocol have been described elsewhere (Ikram et al., 2015).

2.1.2. Visual PVS scores

Visual PVS scores have been created, for each region, according to a standard procedure proposed in the international consortium UNIVRSE (Adams et al., 2015). PVS ratings are defined as linear, ovoid or round shaped hyperintensities on T2 scans and considered to be enlarged when their diameter is larger than 1 mm. For this study, we only use PVS with diameter smaller than 3 mm, since those larger than this cutoff have been suggested to be of potentially different origin. For this visual scoring, a trained observer counts the number of PVS in the midbrain, hippocampi, basal ganglia and centrum semiovale. For the midbrain and hippocampi, the PVS are counted in the whole volume. In the basal ganglia and centrum semiovale, PVS are counted in a single anatomically defined slice. For the basal ganglia, this is the slice showing the anterior commissure. For the centrum semiovale it is the slice 1 cm above the uppermost part of the lateral ventricles. The number of PVS in these slices correlates well with the number of PVS in the whole volume of the regions (Adams et al., 2013).

The inter-observer and intra-observer agreements of this scoring have previously been computed in the Rotterdam Study in every region

Table 1
Characteristics of the Study Population, N = 1485. The smallest volume of WMH was 0.53 mL.

| | |
|---|-------------------|
| Age, years, mean (SD) | 74.61 (7.32) |
| Men, n (%) | 679 (46) |
| Education, years, mean (SD) | 12.06 (3.76) |
| Ever-smoker, n (%) | 1001 (67) |
| Systolic blood pressure, mmHg, mean (SD) | 150.25 (21.05) |
| Diastolic blood pressure, mmHg, mean (SD) | 82.70 (11.07) |
| Diabetes mellitus, n (%) | 219 (15) |
| Glucose, mmol/L, mean (SD) | 5.71 (1.15) |
| Total Cholesterol, mmol/L, mean (SD) | 5.48 (1.02) |
| HDL Cholesterol, mmol/L, mean (SD) | 1.45 (0.39) |
| Body mass index, kg/m ² , mean (SD) | 27.31 (3.87) |
| ApoE ε2 allele carrier, n (%) | 241 (17) |
| ApoE ε4 allele carrier, n (%) | 385 (26) |
| Cortical Infarct on MRI, n (%) | 61 (4) |
| Lacunar Infarct on MRI, n (%) | 192 (13) |
| White matter hyperintensity volume, mL, median [interquartile range] | 6.72 [3.84–13.11] |
| Intracranial volume, mL, mean (SD) | 1144.26 (117.89) |
| White matter volume, mL, mean (SD) | 390.33 (58.47) |
| Gray matter volume, mL, mean (SD) | 518.84 (53.77) |
| Cerebrospinal fluid volume, mL, mean (SD) | 232.82 (53.83) |
| Visual PVS scores, midbrain, median [interquartile range] | 1 [0–3] |
| Visual PVS score, hippocampi, median [interquartile range] | 3 [1–5] |
| Visual PVS scores, basal ganglia, median [interquartile range] | 4 [2–6] |
| Visual PVS scores, centrum semiovale, median [interquartile range] | 7 [4–11] |
| Automated PVS scores, midbrain, median [interquartile range] | 1.39 [0.648–2.51] |
| Automated PVS score, hippocampi, median [interquartile range] | 2.92 [1.98–4.39] |
| Automated PVS scores, basal ganglia, median [interquartile range] | 3.33 [2.38–4.67] |
| Automated PVS scores, centrum semiovale, median [interquartile range] | 6.52 [4.57–9.6] |

Table 2
Agreement between automated and visual PVS scores for each brain region.

The metric reported is the intraclass correlation coefficient (ICC), computed on an independent set of 515 scans. These ICCs are compared to the inter-observer and intra-observer agreements reported by Adams et al. (2013). Note that the inter-observer and intra-observer agreements were computed on a different subset of the same dataset (Section 2.1).

| Region | Intra-observer Agreement | Inter-observer Agreement | Trained on 1600 scans | Trained on 400 scans |
|-------------------|--------------------------|--------------------------|-----------------------|----------------------|
| Midbrain | 0.82 | 0.75 | 0.75 | 0.74 |
| Hippocampi | 0.85 | 0.82 | 0.88 | 0.74 |
| Basal Ganglia | 0.80 | 0.62 | 0.82 | 0.73 |
| Centrum Semiovale | 0.88 | 0.80 | 0.86 | 0.80 |

(Adams et al., 2013). Inter-observer intraclass correlation coefficients (ICCs) have been computed with 105 MRI scans, and intra-observer ICCs with 85 scans (Table 2). The images in our dataset (2115 scans) were visual scored by a single expert rater (Dr. H. Adams).

2.1.3. Potential determinants of PVS

From the 2115 participants, we randomly selected 400 participants to optimize the parameters of our algorithm, and used the remaining 1715 participants to investigate associations between 20 determinants of PVS and automated and visual PVS scores. From these 1715 participants, we excluded participants without informed consent to access medical records and hospital discharge letters ($n = 8$), participants who already suffered stroke ($n = 98$) or were diagnosed with dementia ($n = 32$) or had incomplete information for stroke or dementia ($n = 1$) at time of MRI scan (de Bruijn et al., 2015; Wieberdink et al., 2012). We also excluded scans for which the brain region segmentation algorithm (FreeSurfer, Desikan et al. (2006)) failed for one or more regions ($n = 91$). Excluding these resulted in a set of 1485 participants, from which the highest number of missing values was 25 for cholesterol, HDL cholesterol and glucose. Table 1 lists the characteristics of the study population.

2.1.3.1. Assessments of determinants. Education was obtained from self-reported history and scaled in number of years according to the UNESCO classification.¹ Smoking behavior was assessed during home interviews and categorized as ever- and non-smokers. Blood pressure measurements were averaged over two readings with a random-zero sphygmomanometer at the right upper arm, in sitting position and a resting period of 5 min. Data on serum glucose, total serum cholesterol, serum high-density lipoprotein (HDL) cholesterol were obtained using an automated enzymatic procedure (Boehringer Mannheim System). Diabetes mellitus was defined as a fasting glucose level of ≥ 7.0 mmol/L, or the use of antidiabetic medication. Body mass index was calculated by dividing weight (in kilograms) by the height squared (in meters). ApoE genotyping on coded genomic DNA samples was performed for the $\epsilon 2$ and $\epsilon 4$ alleles of Apolipoprotein E (ApoE- $\epsilon 2$ and ApoE- $\epsilon 4$) carrier status, with a one-stage polymerase chain reaction and TaqMan assay (Wenham et al., 1991). Participants who were classified ApoE - $\epsilon 2\epsilon 4$ counted both as $\epsilon 2$ and $\epsilon 4$ carriers. The majority of samples (81.1%) were genotyped with the Illumina 610 K and 660 K chips, the remaining (18.9%) were imputed to the Haplotype Reference Consortium reference panel (version 1.0) with Minimac 3.

2.1.3.2. Assessment of MRI markers. Several focal and volumetric measures of subclinical brain damage were assessed. Cortical infarcts were defined as lesions involving cortical gray matter with tissue loss and lacunar infarcts as subcortical lesions ≥ 3 mm and < 15 mm on FLAIR, T1,

¹ United Nations Educational SaCOU. International Standard Classification of Education (ISCED) 1976. Available from: <http://unesdoc.unesco.org/images/0002/000209/020992eb.pdf>.

and T2 sequences. The presence of cortical and lacunar infarcts was visually rated by trained research physicians using binary scores (Ikram et al., 2017). White matter hyperintensities (WMH) were measured quantitatively using a validated automated segmentation method (de Boer et al., 2010). This method was also used to segment the brain into gray matter, white matter and cerebrospinal fluid. Total brain volume was defined as the sum of gray and white matter. And intracranial volume was defined as sum of gray and white matter, and cerebrospinal fluid. All WMH segmentations were visually checked by experts and corrected if needed.

2.2. Preprocessing

The first step of our method is to extract the target brain region from the scan and mask the surrounding structures. This preprocessing step is almost identical for all four regions.

N3 Bias field correction Sled et al. (1998) is applied prior to the extraction of the region of interest and prior to the network training. Then we apply the FreeSurfer multi-atlas segmentation algorithm (Desikan et al., 2006) to obtain a binary mask for each region: midbrain, hippocampi, basal ganglia and centrum semiovale. Note that the FreeSurfer segmentation is based on the T1-weighted sequence. All parameters are left as default, except for the skull stripping preflooding height threshold which is set to 10. These masks are then dilated (4 consecutive morphological binary dilations with a cube connectivity equal to one, i.e., 6-connected in 3D), with the exception of the mask of the midbrain, which is eroded (2 consecutive morphological binary erosions with a square connectivity equal to one). These morphological operations can correct segmentation errors and are especially important for the basal ganglia and hippocampi, as PVS can often be located on the border of these regions. On the contrary, for the midbrain, PVS are almost always located in the center and dilating the mask can make the optimization of the model more difficult.

For each region, the borders of the masks are smoothed with a Gaussian kernel of standard deviation $\sigma = 2$ voxel units, and multiplied pixel-wise with the image intensities. These masked images are then cropped around the center of mass of the mask to reduce the image size and memory requirements. The size in voxels of these cropped images for midbrain, hippocampi, basal ganglia and centrum semiovale are $88 \times 88 \times 11$, $168 \times 128 \times 84$, $168 \times 128 \times 84$ and $250 \times 290 \times 14$ respectively. The image values are then rescaled between zero and one to ease the learning process. The cropped volume of the centrum semiovale is relatively small in the craniocaudal direction (z-axis). Contrary to the other three brain regions, the complete volume of the centrum semiovale could not be fit in the memory of our graphics processing unit (GPU). Therefore, as input to our algorithm we kept only the slices surrounding the slice visually scored by the expert rater. We automatically identified this slice by segmenting the lateral ventricles with FreeSurfer, and selecting the slice 1 cm above, as defined by (Adams et al., 2013).

In the left column, Fig. 4 shows one example of the preprocessed images for each region.

2.3. 3D convolutional regression network

Once the images are preprocessed, they are given as input to a convolutional neural network (CNN) similar to the one proposed in our earlier work (Dubost et al., 2018) but with skip connections between layers. This network computes the automated PVS scores using a combination of learned filters.

We train a different network for each region. There are two reasons for this. PVS can have a different shape depending on their location in the brain. For instance, in the hippocampi, the shape of PVS is more round, while in the centrum semiovale, PVS are more elongated. Differentiating from mimics is also region specific. For instance, motion artifacts affect mostly the centrum semiovale and have a much lower influence in the midbrain, and lacunar infarcts are often located in the basal ganglia.

Our CNN architecture is similar to that of a small ResNet (He et al., 2016) adapted for regression in 3D image (see Fig. 3). Our CNN has two 3D $3 \times 3 \times 3$ convolutional layers, followed by a $2 \times 2 \times 2$ max-pooling layer, again two 3D $3 \times 3 \times 3$ convolutional layers, a global average pooling layer, and a fully connected layer, combining the contribution of the different features into a single score. The output of the network is hence a scalar and spans \mathbb{R} . The first two convolutional layers have 32 filters each, and the last two convolutional layers have 64 filters each. The convolutions are zero-padded, and are followed by a ReLU activation. We use skip connections between the input and output of two successive convolutional layers, to allow the network to skip unnecessary operations and adapt its complexity to the tasks, which can ease the learning process (He et al., 2016). For instance, we expect the quantification of PVS to be simpler in the midbrain than in the centrum semiovale. When using skip connections, we concatenate the features maps, instead of summing them as proposed in He et al. (2016). There is little evidence that using either one or the other strongly impacts the performance. However, the concatenation is easier to implement as it does not require to have the same number of feature maps. In total, our model has less than 200 000 parameters.

For the regularization, we use on-the-fly data augmentation (translation, rotation and flipping), and when training with smaller sets, we used dropout (Srivastava et al., 2014) (30%) after each convolutional layer and after the global pooling layer. See section 3.1 for details.

To train the network, we minimize the MSE loss function between the outputs of the network and the ground truth labels indicating the number of PVS in the given brain region.

The method proposed in our earlier work (Dubost et al., 2018) quantifies PVS in the basal ganglia also with regression CNN, but with a different architecture. There are three differences with the CNN we proposed in the current work. Firstly, the proposed network is simpler and lighter. Experiments on the parameters of the network (Dubost et al., 2018), indeed suggested that simpler models performed equally good with enough training data. In our experiments, the training of deeper models was also much longer with small training set (400 scans), especially for the centrum semiovale and hippocampi. The second change is the introduction of skip connections between blocks. The third and last change is the use of global pooling instead of two fully connected layers of 2000 neurons. Using global pooling does not harm the performance and saves large amounts of GPU memory. This change was also proposed by He et al. (2016) over the architecture proposed by Simonyan and Zisserman (2015), the preceding state-of-the-art neural network on the ImageNet challenge (Deng et al., 2009).

2.4. Model training

During training, a validation set is used to stop the optimization of network before over-fitting occurs. In most experiments the models were trained on a set of 1600 scans (1200 for training and 400 for validation). To demonstrate that reasonable results can still be achieved with less

training data, we also performed some experiments with a smaller subset of 400 scans (320 training and 80 validation).

As mentioned in section 2.3, a separate model is trained for each region. The training of such models can be unpredictably long for the hippocampi and centrum semiovale. To speed up the training, we first train the models in the basal ganglia, as the convergence is faster there. Then we fine-tune the networks with the target region only (hippocampi or centrum semiovale). The training in the midbrain converges quickly and no pre-training is needed. We chose to pre-train with the basal ganglia and not in the midbrain, as PVS in the basal ganglia are more similar to PVS in the hippocampi and centrum semiovale.

2.5. Statistical analyses

To evaluate associations between determinants of PVS and PVS scores, we used zero-inflated negative binomial regression models with the PVS score as outcome, as in the study of Adams et al. (2014). We used the ‘glmmADMB’ package for generalized linear mixed models in R. The models were corrected for age and sex (except for the associations of age, sex respectively) and additionally for intracranial volume when computing associations with volumetric measures (white matter, gray matter, and cerebrospinal fluid). To account for the skewed distribution of WMH, we log transformed the WMH volumes. Continuous determinants were normalized by computing z-scores. Bonferroni correction was used, therefore associations with a p-value below $0.05/(20 \text{ determinants} \times 4 \text{ brain regions}) = 6.25 \times 10^{-4}$ were considered significant.

3. Results

We evaluate the performance of the proposed model with three series of experiments. First, we inspect attention maps of the model, revealing that the model indeed focuses on PVS. Second, we measure the agreement between automated and visual scores, and show that this agreement is at least at the level of the human inter-observer agreement for each region. Then we verify the scan-rescan reproducibility of the automated PVS scores. Finally, we show that the associations between 20 determinants of PVS and the automated scores are similar to associations between the same determinants and visual scores.

3.1. Experimental settings

We initialize the weights of the CNN by sampling from a Gaussian distribution, use Adadelta (Zeiler, 2012) for optimization, and augment the training data on-the-fly with randomly transformed samples. The transformation parameters for augmentation are uniformly drawn from an interval of 0.2 radians for rotation, 2 pixels for translation and flipping in the x and y direction. During training, the images are augmented with a random combination of these parameters. The network is trained per sample (mini-batches of a single 3D image of the preprocessed region of interest). We implemented our algorithms in Python in Keras (Chollet

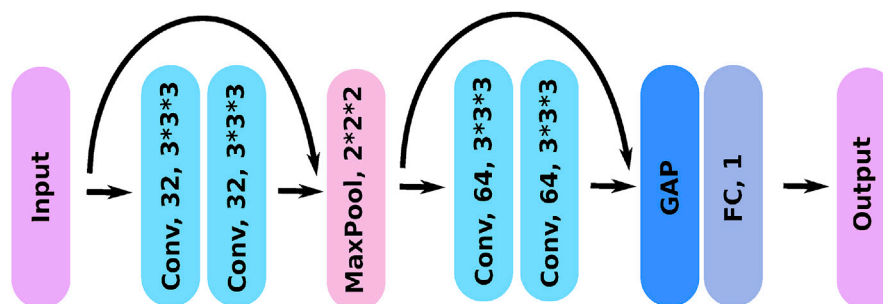


Fig. 3. Architecture of the neural network. The input is a 3D scan cropped around the region of interest, and the output is the automated PVS score. ‘Conv’ stands for convolutional layer, and is followed by the number of filters and the filter size; ‘MaxPool’ stands for max pooling layer; ‘GAP’ for global average pooling, ‘FC’ for fully connected layer; and the curved arrows represent skip connections with concatenation of feature maps.

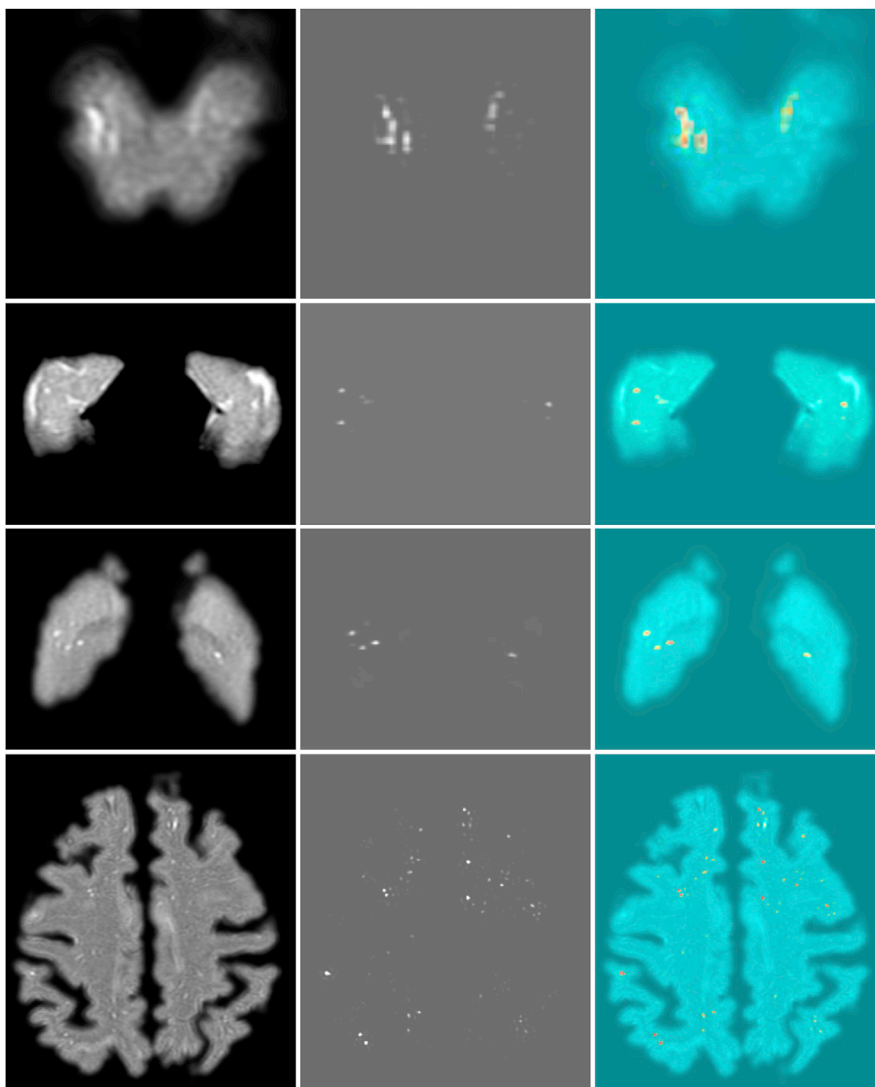


Fig. 4. Attention maps of the neural network. From left to right: preprocessed input image, attention map, overlay of the input image and the attention map. From top to bottom: midbrain, hippocampi, basal ganglia, and centrum semiovale. In the overlay, the heatmaps reflect the contribution of pixels to the prediction of the networks: red pixels contributed the most, while blue pixel did not contribute. One can notice that many slightly enlarged perivascular spaces appear in orange. The network detected these, but they influenced its prediction less than the larger PVS.

et al., 2015) with Tensorflow as backend, and ran the experiments on a Nvidia GeForce GTX 1070 GPU and Nvidia Tesla K40.² The average training time was one day. We stop the training after the validation loss converged to a stable value, or before over-fitting happens. The networks were trained for 450 epochs on average. The learning rate of the optimizer was set to 1, its default value in Keras. We chose Adadelta which is not sensitive to the initial setting of the learning rate. Once the CNN is trained, the automatic PVS scoring, given the segmented region of interest, takes in average 287 ms per region.

3.2. Attention maps

As first qualitative evaluation we check whether the neural networks learned to identify the structures of interest (PVS), or detected some other features that are correlated to the PVS. We use attention maps computed via “guided backpropagation” (Springenberg et al., 2015). These attention maps are computed as the derivative of the automated PVS scores (the output of the CNN) with respect to the input image. Springenberg et al. (2015) improved the original attention map computation proposed by Simonyan et al. (2014) by additionally masking out the values corresponding to negative entries of the top gradient in the

ReLU activations, which clears noise in the attention maps.

Fig. 4 shows examples of these attention maps for each of the four regions. We notice that the neural networks focus on the PVS, even though they are trained using global, image-wise labels only.

In Fig. 5, we verify with attention maps that the algorithm can discriminate between WMH, lacunar infarcts and PVS. We also investigate the limits of the method by showing an attention map for état criblé. For visualization, the PVS attention maps were thresholded to create a segmentation, and WMH segmentations were computed with another automated segmentation method (de Boer et al., 2010) and manually corrected by experts as described in Section 2.1.3. We selected scans with large volumes of WMH, and with lacunar infarcts in the basal ganglia. Segmented PVS and WMH rarely overlap, and when they do, it is often because of co-occurring PVS and WMH. The behavior is similar for lacunar infarcts. For the participant with état criblé, we notice that most of the PVS are detected.

Fig. 5 also shows the example of an attention map computed over a white matter slice lower in the brain, 1 cm below the uppermost part of the lateral ventricles. Although this region was not seen during training, PVS are identified as accurately as in the slice used for visual rating. Because the texture of the white matter is relatively uniform, PVS can be quantified in other slices of the white matter using the network optimized on a single slice of the CSO. For a full quantification we would recommend to train on other slices as well, as anatomical configurations differ with location.

² Nvidia Tesla K40 are GPUs available at Cartesius, a dutch national cluster.

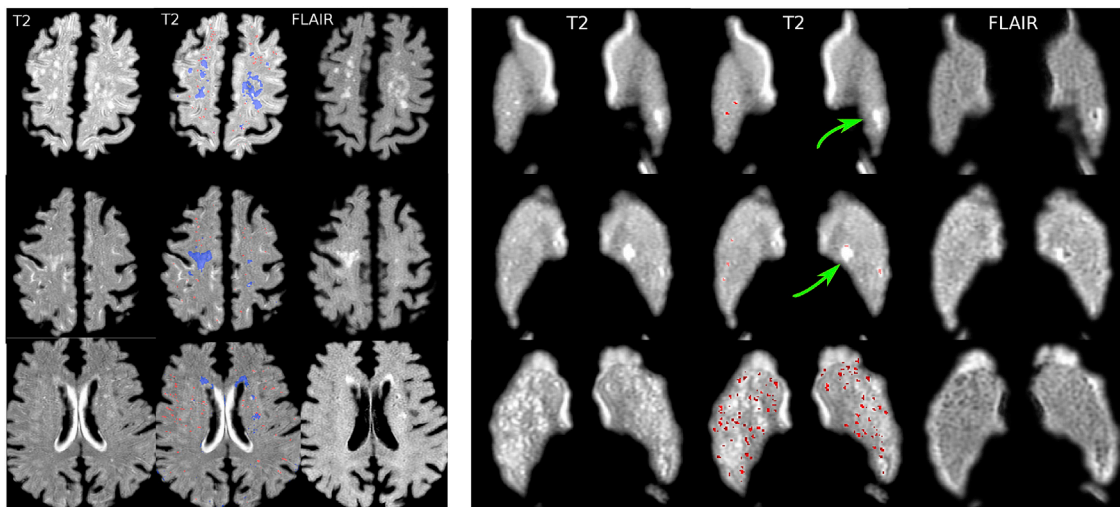


Fig. 5. PVS detection in challenging situations. In each row we display three images of the same region of interest for the same participant but with different modalities: first the T2-w scan, then the overlap between PVS and/or WMH segmentations and the T2-w scan, and finally the FLAIR-w scan. PVS are quantified in the centrum semiovale on the left side, and in the basal ganglia on the right side. WMH are indicated in blue, PVS in red, and lacunar infarcts with a green arrow. The PVS segmentations were obtained by thresholding the attention maps (Section 3.2) with a fixed threshold for all scans. In the last left row, the PVS attention map was computed in a white matter slice lower in the brain. White matter slices in that location were never used during training. In the last right row, we show the example of participant with état criblé. In the second row of the basal ganglia images, the algorithm detects a PVS at the border of the infarct. This can be verified by checking the upper slices.

3.3. Agreement between automated and visual scores

In this section, we evaluate the proposed automated scores by comparing with expert visual scores. We optimized the parameters of the CNN on a set of 1600 scans (1200 for training and 400 for validation). We also optimized the same model using only a subset of 400 scans (320 training and 80 validation), where we used dropout (Srivastava et al., 2014) after each convolution to avoid over-fitting. We evaluated both models on an independent set: the remaining 515 scans. The results are reported in Table 2. Fig. 6 shows Bland Altman plots for each region. Note that on the Bland Altman plots, the discrete nature of the distribution of the points, especially visible for the midbrain and hippocampi, is a consequence of the visual PVS scores being integer numbers.

When trained on 1600 scans, the ICC between the automated and visual scores were higher than the inter-observer agreement previously reported for each region. On the Bland Altman plots, one can notice that the largest errors usually occur for scans with many PVS, and for which there

are only few training examples. Also, even for expert raters the rating becomes more difficult and variable for scans with many PVS. This is due to the continuous nature of the enlargement of perivascular spaces: keeping a consistent threshold of enlargement becomes more challenging. In case of large disagreement, the automated scores mostly underestimate the number of PVS with respect to the visual scores. We also notice that the largest differences between automated and visual scores seem to scale linearly with the number of lesions before reaching a plateau. This is especially noticeable for the hippocampi and basal ganglia. The level of this plateau depends on the region of interest and is higher in the basal ganglia and centrum semiovale compared to midbrain and hippocampus.

3.4. Reproducibility

The reproducibility of the automated PVS scores is evaluated on a reproducibility set of 30 participants scanned twice (see Section 2.1). The ICC of the automated PVS scores between the first and second sets of scans is

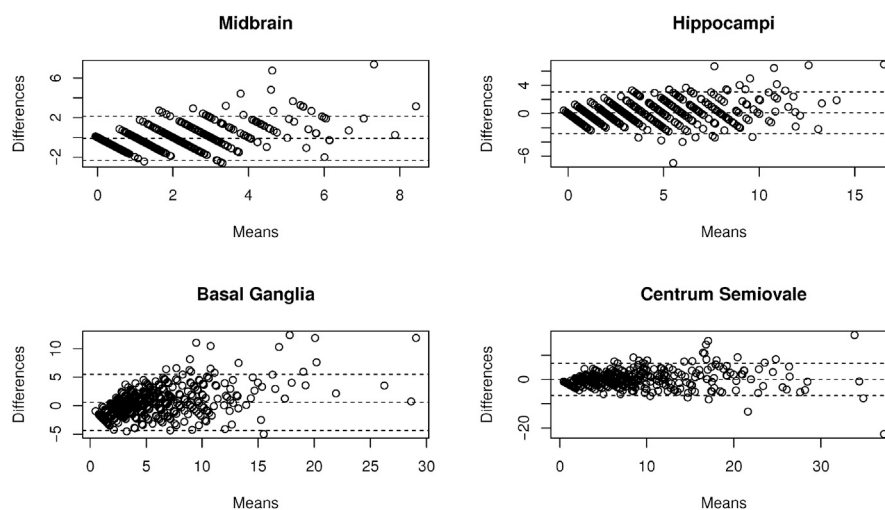


Fig. 6. Bland-Altman plots between the automated and visual PVS scores in the four regions. The algorithms were optimized with 1600 scans, and evaluated on 515 scans. For the differences, the automated scores were subtracted from the visual scores.

0.82 for the midbrain, 0.93 for the hippocampi, 0.92 for the basal ganglia, and 0.87 for the centrum semiovale. Except for the centrum semiovale, all values are higher than the intra-rater agreement computed on another subset of the same dataset and reported by Adams et al. (2013) (Table 2).

3.5. Associations with determinants of PVS

We investigate associations between 20 potential determinants of

PVS (characteristics in Table 1) and the automated PVS scores, and compare them with the associations between the same determinants and the visual PVS scores. The neural networks are first optimized using 400 scans for each region (we reuse the second model presented in section 3.3), and then applied to the remaining 1715 independent scans to produce the automated scores. We investigate associations on this set of 1715 scans. After excluding participants as described in section 2.1, this resulted in 1485 stroke-free and non-demented participants with

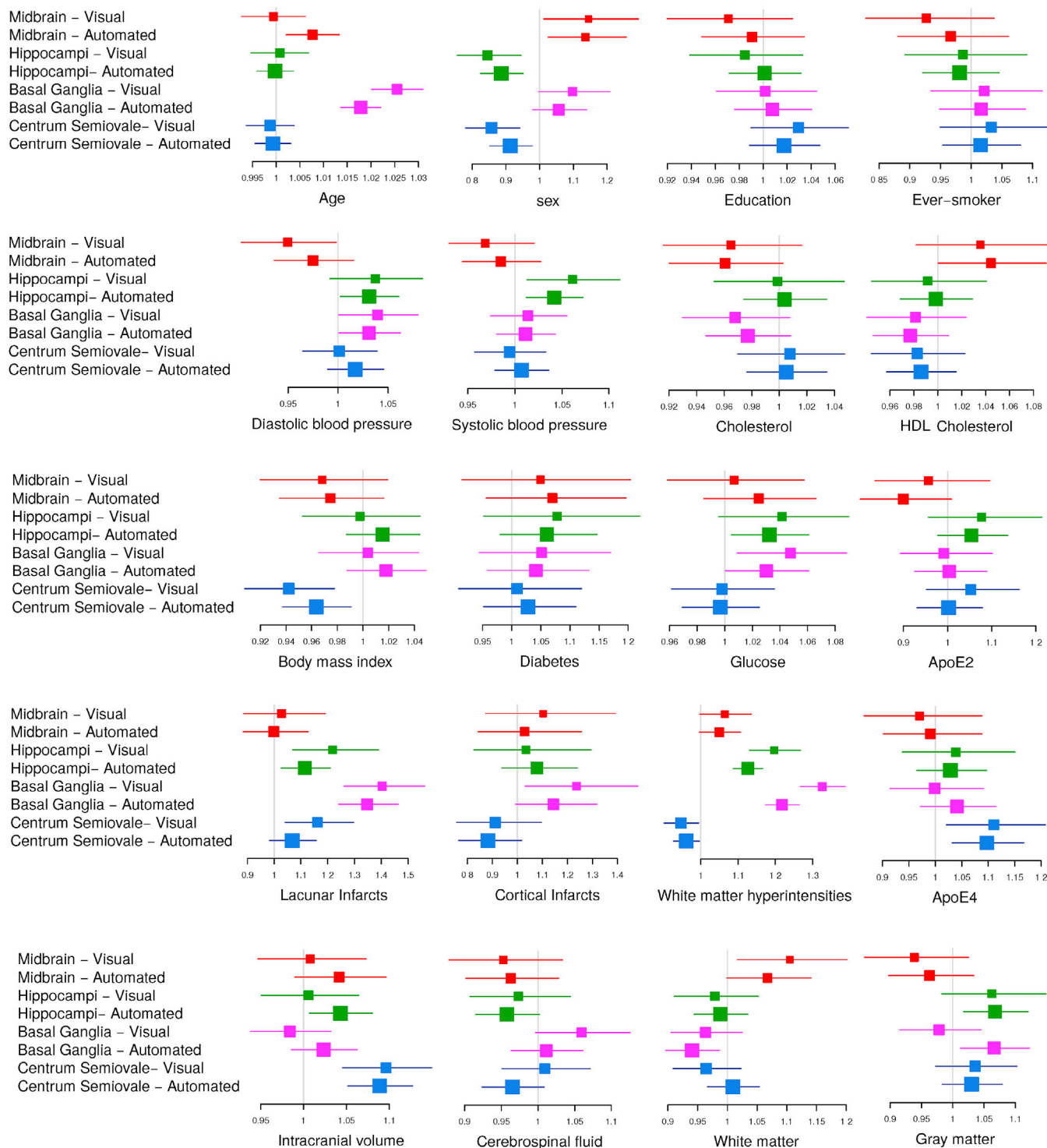


Fig. 7. Associations between determinants of PVS and PVS scores.

Odds ratio with 95% confidence intervals (non Bonferroni corrected). Characteristics of the study population are given in Table 1. The size of the colored boxes is inversely proportional to the size of the confidence intervals of the odds ratio.

available brain imaging.

Fig. 7 shows forest plots for each determinant, and a sorted list of all p-values can be found in supplementary materials. Overall, association patterns are very similar for visual and automated scores.

We found that white matter hyperintensity volume is associated with both visual and automated PVS scores in the basal ganglia and in the hippocampi. Age is associated with both visual and automated PVS scores in the basal ganglia. The presence of lacunar infarcts is also associated with both visual and automated PVS scores in the basal ganglia. And finally, intracranial volume is associated with both visual and automated PVS scores in the centrum semiovale. In all cases, determinants that are significantly associated with visual PVS scores, also show significant association with the automated PVS scores, and in almost the same order of p-values.

As the automated method takes as input the MRI scans, and is only optimized using global labels (the number of PVS), in the scans other information than PVS might be used to compute the automated PVS scores. This is an unwanted behavior. We did not notice any bias of the automated method towards more significant associations with imaging markers. For instance, for both visual and automated PVS scores, 9 of the 20 most significant associations were between imaging markers and PVS scores. However, computing the p-value of the difference of z-scores of the associations showed a significant difference for gray matter and PVS scores in the basal ganglia. In Fig. 7, we notice the same trend for the association between intracranial volume and PVS scores in the basal ganglia. There was also a significant difference (though with a higher p-value) for associations between intracranial volume and PVS scores in the hippocampi.

Computing the p-value of the difference of z-scores of the associations revealed a last significant difference: the association between age and the automated PVS scores in the midbrain (odds-ratio 1.008 [1.002–1.0013]) was significantly stronger than the association between age and the visual scores in the midbrain (odds-ratio 0.999 [0.992–1.006]).

4. Discussion

The algorithm developed in this work computes automated scores to quantify enlarged perivascular spaces (PVS) in the midbrain, hippocampi, basal ganglia and centrum semiovale - the four brain regions currently deemed most clinically relevant for PVS quantification. We demonstrated the performance of our algorithm using a set of 2115 MRI scans that were visually scored by an expert rater. For all four regions, the intraclass correlation coefficient between the automated scores and the visual scores was found to be higher than the inter-observer agreement, which was previously computed on a smaller subset of the same study population (Adams et al., 2013). Scan-rescan reproducibility was high (ICC 0.82–0.93). We also demonstrated the application of our automated scores by verifying the associations between determinants of PVS and our automated scores in a test set of 1485 scans, and comparing these associations to the visual scores. Based on these results, we believe that our automated scores could ultimately replace visual scores in future research projects studying the etiology and clinical relevance of PVS.

Automated PVS scores have two major advantages over visual scores: they are more objective (because the algorithm is deterministic), and can be computed more quickly. While a trained expert rater needs several minutes to score a scan, the computation of the automated PVS score on modern hardware (GPU) lasts less than a second. Quantifying PVS through all the white matter could also be achieved by applying the network optimized in the centrum semiovale in a sliding window in z. This process would last approximately 6 s on modern hardware after preprocessing. This makes our automated approach suited to be used in large scale studies, investigating for instance the etiology of PVS, their distribution in brain aging, their implications, and their potential as a biomarker for early diagnosis of cerebral small vessel disease. In addition, our method could be extended to fully quantify PVS by assessing

their volume with the attention maps produced by the neural networks (Fig. 4). These attention maps indeed provide a voxel-wise probability of PVS presence, which can for instance be summed over a region of interest, to yield a total volume or burden of PVS.

As the intensity histograms of WMH and PVS can overlap, using global intensity features to discriminate between WMH and PVS is impossible.

Using a multisequence input could ensure that the automated PVS quantification method does not misclassify other small vessel disease markers such as WMH or lacunar infarcts as PVS. Additional sequences could be added as additional input channels in the architecture of the networks. For the sake of simplicity, in this work we used only the T2-w sequence, and verified the performance of the algorithm in this scenario. It appeared that using other sequences was not necessary. Because they follow blood vessels, PVS have a characteristic shape, while the shape of WMH and lacunar infarcts is much more irregular. Finally, PVS are usually sharper delineated than WMH.

As mentioned in section 3.5, other imaging markers are not intended to be used in the computation of the automated PVS scores, because they would interfere with the detection of PVS and the explainability of the method. Fig. 5 illustrates that PVS in the attention maps do not overlap with WMH or with lacunar infarcts. In addition, apart from the associations between gray matter volume and PVS scores in the basal ganglia, we did not notice any strong trend of our method towards a stronger association with imaging markers. This difference of association in the basal ganglia most probably results from the automated PVS scores being computed across the complete volume of the basal ganglia, while visual EPVS scores are rated in a single slice (Section 2.1). The consequences of this difference have been thoroughly investigated by Dubost et al. (2018), and seem to favor the automated PVS scores, as they are less sensitive to perturbations, such as missed PVS. Although we found associations between WMH and PVS, and between lacunar infarcts and PVS, it should be noted that these associations are similar for automated and visual scores. This suggests that the cause is not a confusion between these types of lesions for the automated approaches but rather that WMH, lacunar infarcts and PVS are all markers of cerebral small vessel disease. Hurford et al. (2014), Adams et al. (2014), and Gonzalez-Castro et al. (2017) also found similar associations.

In none of the steps of our method did we model the anisotropy of the data, as we expect the network to be able to correct for anisotropy.

The Pearson correlation coefficients between visual PVS scores (Potter et al., 2015a,b) and automated PVS scores reported by Boespflug et al. (2017) are 0.65, 0.69, and 0.54 in the CSO for three different raters. Ramirez et al. (2015) measured the Pearson correlation coefficients between the segmented volumes of their semi-automated method and visual PVS scores (Patankar et al., 2005). The results were 0.84 in the CSO, and 0.75 in the basal ganglia. In comparison, in our work the ICC between the visual and automated PVS scores optimized on 1600 scans were 0.86 in the CSO, and 0.82 in the basal ganglia. Contrary to Ramirez et al. (2015), in cognitively normal participants we found significant associations between WMH and PVS, and between lacunar infarcts and PVS. Gonzalez-Castro et al. (2017) computed automatic binary scores of PVS burden in the basal ganglia and investigated associations with determinants of PVS. They found significant associations with higher age, Fazekas WMH scale, and the presence of lacunar infarcts, while there was no significant associations with brain atrophy, hypertension, or stroke subtype. In the current study, we found the same significant associations (age, WMH, and presence of lacunar infarcts) for the basal ganglia.

There is increasing evidence that ageing affects PVS, and putative mechanisms are dysfunction of the blood-brain barrier, or impaired perivascular drainage (Brown et al., 2018). Higher age was previously shown to be associated with higher visual PVS scores in the four regions investigated in the current study: midbrain, hippocampi, basal ganglia, and centrum semiovale (Adams et al., 2014). The study by Adams et al. (2014) has been carried out in a significantly larger population study (3146 participants against 1485 for our study). In the current study, age

was only associated with visual PVS scores in the basal ganglia. Higher age was also associated with higher automated PVS scores in the basal ganglia. Previous studies with visual PVS scoring have shown similar associations with age and basal ganglia PVS (Gutierrez et al., 2013; Martinez-Ramirez et al., 2013; Potter et al., 2015b). In the current study, in comparison with visual PVS scores, the automated PVS scores showed a significantly higher association power in the midbrain, which may suggest that they better capture the burden of PVS than visual scores. We did not find significant associations between age and PVS in the hippocampi or in the centrum semiovale (neither with visual PVS scores, nor with automated PVS scores). Similarly, in a recent study on a 7 T scanner by Bouvy et al. (2016), no association was found between age and PVS in centrum semiovale. While Adams et al. (2014) found the weakest association between age per decade and PVS to be in the hippocampi (odds ratio of 1.07 [1.02–1.12]), they also found the strongest association between age per decade and PVS to be in the centrum semiovale with an odds ratio of 1.24 [1.19–1.30]. While there seems still to be controversy in the detailed relationship between age and PVS, automated PVS scores could possibly be more powerful to better disentangle possible mechanisms of PVS which effect brain health in ageing.

The proposed PVS quantification method requires prior brain structure segmentation. We reused FreeSurfer Desikan et al. (2006) segmentations computed for previous projects. Other potentially more robust or faster methods could be used instead. For instance, Mehta and Sivaswamy (2017) and Roy et al. (2018) reported results at least as accurate as FreeSurfer, while being much faster. Including FreeSurfer segmentation time, the overall pipeline would last several hours. However, by replacing FreeSurfer with other brain structures segmentation methods such as Mehta and Sivaswamy (2017) or Roy et al. (2018), the overall pipeline could last less a minute on modern hardware.

The main limitation of this work is that, contrary to the UNIVRSE rating system (Adams et al., 2015), the method was evaluated using MRI scans acquired on a single scanner, precluding the assessment of performance on different datasets. However, we believe this method can easily be applied to other datasets by only fine-tuning the CNN parameters on a few scans (Yosinski et al., 2014). Besides the performance of the algorithm should also be evaluated in multi-center or multi-scanner data.

Another potential limitation of this work is that the models were trained and validated on a general population from which subjects with prevalent stroke or dementia were excluded, in order to focus on variability in the normal aging process. Our dataset thus may have included relatively fewer scans with exceptionally many PVS, such as e.g. in état criblé or neurological diseases (Fig. 5). These scans have a low prevalence and are rarely used during the optimization of our models. We expect that the good performance of our model in a population with more subtle brain changes will translate well to more extreme settings, though this has to be evaluated.

5. Conclusion

We present a regression method to automatically quantify the number of enlarged perivascular spaces in the midbrain, hippocampi, basal ganglia, and centrum semiovale. The automated scores are more objective than visual scores and less time consuming. We validated our approach on 1485 brain MRI scans, demonstrated that the automated PVS show good agreement with visual PVS scores, and showed that the automated PVS scores are associated with several determinants of PVS, in a similar fashion to the PVS visual scores. We believe that this method could replace visual scoring of PVS in epidemiological and clinical studies, and therefore advance research into the etiology of PVS and its potential as a risk indicator of small vessel disease.

Acknowledgment

This work was partly carried out on the Dutch national e-

infrastructure with the support of SURF Cooperative. This research was funded by The Netherlands Organisation for Health Research and Development (ZonMw) Project 104003005, with additional support of Netherlands Organisation for Scientific Research, project NWO-EW VIDI 639.022.010 and project NWO-TTW Perspectief Programme P15-26.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2018.10.026>.

References

- Achiron, A., Faibel, M., 2002. Sandlike appearance of Virchow-Robin spaces in early multiple sclerosis: a novel neuroradiologic marker. *Am. J. Neuroradiol.* 23, 376–380.
- Adams, H.H., Schwingschuh, P., van der Lugt, A., Katschnig-Winter, P., Hofman, A., Krestin, G., Vernooij, M., Schmidt, R., Ikram, M.A., 2014. Determinants of enlarged virchow-robin spaces: the univir consortium. *Alzheimer's & Dementia. The Journal of the Alzheimer's Association* 10, P408.
- Adams, H.H.H., Cavalieri, M., Verhaaren, B.F.J., Bos, D., Van Der Lugt, A., Enzinger, C., Vernooij, M.W., Schmidt, R., Ikram, M.A., 2013. Rating method for dilated virchow-robin spaces on magnetic resonance imaging. *Stroke* 44, 1732–1735.
- Adams, H.H.H., Hilal, S., Schwingschuh, P., Wittfeld, K., van der Lee, S.J., DeCarli, C., Vernooij, M.W., Katschnig-Winter, P., Habes, M., Chen, C., Seshadri, S., van Duijn, C.M., Ikram, M.K., Grabe, H.J., Schmidt, R., Ikram, M.A., 2015. A priori collaboration in population imaging: the Uniform Neuro-Imaging of Virchow-Robin Spaces Enlargement consortium. *Alzheimer's Dementia: Diagnosis, Assessment and Disease Monitoring* 1, 513–520.
- Ballerini, L., Lovreglio, R., Hernández, M.d.C.V., Ramirez, J., MacIntosh, B.J., Black, S.E., Wardlaw, J.M., 2018. Perivascular spaces segmentation in brain mri using optimal 3D filtering. *Sci. Rep.* 8, 2132.
- de Boer, R., Vrooman, H.A., Ikram, M.A., Vernooij, M.W., Breteler, M.M., van der Lugt, A., Niessen, W.J., 2010. Accuracy and reproducibility study of automatic mri brain tissue segmentation methods. *Neuroimage* 51, 1047–1056.
- Boespflug, E.L., Schwartz, D.L., Lahna, D., Pollock, J., Iliff, J.J., Kaye, J.A., Rooney, W., Silbert, L.C., 2017. Mr imaging-based multimodal autoidentification of perivascular spaces (mmaps): automated morphologic segmentation of enlarged perivascular spaces at clinical field strength. *Radiology* 286, 632–642.
- Bouvy, W.H., Zwambag, J.J., Reinink, R., Wisse, L.E., Luijten, P.R., Kappelle, L.J., Geerlings, M.L., Biessels, G.J., group, U.V.C.I.V.S., 2016. Perivascular spaces on 7 tesla brain mri are related to markers of small vessel disease but not to age or cardiovascular risk factors. *J. Cerebr. Blood Flow Metabol.* 36, 1708–1717.
- Brown, R., Benveniste, H., Black, S.E., Charpak, S., Dichgans, M., Joutel, A., Nedergaard, M., Smith, K.J., Zlokovic, B.V., Wardlaw, J.M., September 2018. Understanding the role of the perivascular space in cerebral small vessel disease. *Cardiovasc. Res.* 114 (11), 1462–1473.
- de Bruijn, R.F., Bos, M.J., Portegies, M.L., Hofman, A., Franco, O.H., Koudstaal, P.J., Ikram, M.A., 2015. The potential for prevention of dementia across two decades: the prospective, population-based rotterdam study. *BMC Med.* 13, 132.
- Chollet, F., et al., 2015. Keras. <https://github.com/fchollet/keras>.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE*, pp. 248–255.
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980.
- Dubost, F., Adams, H., Bortsova, G., Ikram, M.A., Niessen, W., Vernooij, M., de Bruijne, M., 2018. 3D Regression Neural Network for the Quantification of Enlarged Perivascular Spaces in Brain Mri arXiv preprint arXiv:1802.05914.
- Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, pp. 130–137.
- Gonzalez-Castro, V., Hernández, M.d.C.V., Chappell, F.M., Armitage, P.A., Makin, S., Wardlaw, J.M., 2017. Reliability of an automatic classifier for brain enlarged perivascular spaces burden and comparison with human performance. *Clin. Sci.* 131, 1465–1481.
- Gutierrez, J., Rundek, T., Ekind, M., Sacco, R., Wright, C., September 2013. Perivascular spaces are associated with atherosclerosis: an insight from the northern manhattan study. *Am. J. Neuroradiol.* 34 (9), 1711–1716.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
- Hurford, R., Charidimou, A., Fox, Z., Cicolotti, L., Jager, R., Werring, D.J., 2014. Mri-visible perivascular spaces: relationship to cognition and small vessel disease mri markers in ischaemic stroke and tia. *J. Neurol. Neurosurg. Psychiatry* 85 (5), 522–525.
- Ikram, M.A., Brusselle, G.G., Murad, S.D., van Duijn, C.M., Franco, O.H., Goedegebure, A., Klaver, C.C., Nijsten, T.E., Peeters, R.P., Stricker, B.H., et al., 2017. The rotterdam study: 2018 update on objectives, design and main results. *Eur. J. Epidemiol.* 32, 807–850.

- Ikram, M.A., van der Lugt, A., Niessen, W.J., Koudstaal, P.J., Krestin, G.P., Hofman, A., Bos, D., Vernooij, M.W., 2015. The Rotterdam Scan Study: design update 2016 and main findings. *Eur. J. Epidemiol.* 30, 1299–1315.
- Martinez-Ramirez, S., Pontes-Neto, O.M., Dumas, A.P., Auriel, E., Halpin, A., Quimby, M., Gurol, M.E., Greenberg, S.M., Viswanathan, A., 2013. Topography of dilated perivascular spaces in subjects from a memory clinic cohort. *Neurology* 10–1212.
- Mehta, R., Sivaswamy, J., 2017. M-net: a convolutional neural network for deep brain structure segmentation. In: *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, IEEE*, pp. 437–440.
- Mills, S., Cain, J., Purandare, N., Jackson, A., 2007. Biomarkers of cerebrovascular disease in dementia. *BJR (Br. J. Radiol.)* 80, S128–S145.
- Patankar, T.F., Mitra, D., Varma, A., Snowden, J., Neary, D., Jackson, A., 2005. Dilatation of the Virchow-Robin space is a sensitive indicator of cerebral microvascular disease: study in elderly patients with dementia. *Am. J. Neuroradiol.* 26, 1512–1520.
- Potter, G.M., Chappell, F.M., Morris, Z., Wardlaw, J.M., 2015a. Cerebral perivascular spaces visible on magnetic resonance imaging: Development of a qualitative rating scale and its observer reliability. *Cerebrovasc. Dis.* 39, 224–231.
- Potter, G.M., Doubal, F.N., Jackson, C.A., Chappell, F.M., Sudlow, C.L., Dennis, M.S., Wardlaw, J.M., 2015b. Enlarged perivascular spaces and cerebral small vessel disease. *Int. J. Stroke* 10, 376–381.
- Ramirez, J., Berezuk, C., McNeely, A.A., Scott, C.J., Gao, F., Black, S.E., 2015. Visible Virchow-Robin spaces on magnetic resonance imaging of Alzheimer's disease patients and normal elderly from the Sunnyside dementia study. *J. Alzheim. Dis.* 43, 415–424.
- Roy, A.G., Conjeti, S., Navab, N., Wachinger, C., 2018. Quicknat: Segmenting Mri Neuroanatomy in 20 Seconds arXiv preprint arXiv:1801.04161.
- Selvarajah, J., Scott, M., Stivaros, S., Hulme, S., Georgiou, R., Rothwell, N., Tyrrell, P., Jackson, A., 2009. Potential surrogate markers of cerebral microvascular angiopathy in asymptomatic subjects at risk of stroke. *Eur. Radiol.* 19, 1011–1018.
- Simonyan, K., Vedaldi, A., Zisserman, A., 2014. Deep inside convolutional networks: visualising image classification models and saliency maps. In: *International Conference for Learning Representations Workshop*.
- Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. In: *International Conference for Learning Representations*.
- Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Trans. Med. Imag.* 17, 87–97.
- Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M., 2015. Striving for simplicity: the all convolutional net. In: *International Conference for Learning Representations*.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Wardlaw, J.M., Smith, E.E., Biessels, G.J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R.I., O'Brien, J.T., Barkhof, F., Benavente, O.R., Black, S.E., Brayne, C., Breteler, M., Chabriat, H., DeCarli, C., de Leeuw, F.E., Doubal, F., Duering, M., Fox, N.C., Greenberg, S., Hachinski, V., Kilimann, I., Mok, V., van Oostenbrugge, R., Pantoni, L., Speck, O., Stephan, B.C., Teipel, S., Viswanathan, A., Werring, D., Chen, C., Smith, C., van Buchem, M., Norrving, B., Gorelick, P.B., Dichgans, M., 2013. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* 12, 822–838.
- Wenham, P., Price, W., Blundell, G., 1991. Apolipoprotein e genotyping by one-stage pcr. *Lancet* 337, 1158–1159.
- Wieberdink, R.G., Ikram, M.A., Hofman, A., Koudstaal, P.J., Breteler, M.M., 2012. Trends in stroke incidence rates and stroke risk factors in rotterdam, The Netherlands from 1990 to 2008. *EJE (Eur. J. Epidemiol.)* 27, 287–295.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks?. In: *Advances in Neural Information Processing Systems*, pp. 3320–3328.
- Zeiler, M.D., 2012. ADADELTA: an Adaptive Learning Rate Method arXiv:1212.5701.
- Zijlmans, J.C.M., Daniel, S.E., Hughes, A.J., Révész, T., Lees, A.J., 2004. Clinicopathological investigation of vascular parkinsonism, including clinical criteria for diagnosis. *Mov. Disord.* 19, 630–640.