An ensemble of 3D convolutional neural networks for central vein detection in white matter lesions

Mário João Fartaria[1,2,3]*, Jonas Richiardi[1,2]*, João Jorge[4,5], Pietro Maggi[6], Pascal Sati[7], Daniel S. Reich[7], Reto Meuli[2], Cristina Granziera[8], Meritxell Bach Cuadra[2,5], Tobias Kober[1,2,3]

[1]ACIT, SIEMENS, [2]Dept. of Medical Radiology, Lausanne University Hospital, [3]LTS5, EPFL, [4]LIFMET, EPFL, [5]CIBM, UNIL, [6]Dept. of Neurology, Lausanne University Hospital, Lausanne, Switzerland. [7]Translational Neuroradiology Section, NINDS, NIH, MD, United States. [8]ThINk, USB, Basel, Switzerland.

Abstract

The presence of a vein inside white matter lesions was recently proposed as an imaging biomarker that can help in the differential diagnosis of Multiple Sclerosis (MS), potentially reducing the challenging clinical-radiological gap. Here, we propose a prototype based on ensembling small 3D convolutional networks to classify perivenular (P+) and non-perivenular (P-) lesions. Even without prior lesion masking, our approach reaches performance superior to imaging filters designed specifically to detect blood vessels, and that have access to a lesion mask.

1 Introduction

The presence of a vein whithin white matter (WM) lesions (the "central vein sign") is a recently-proposed biomarker that promises to improve MS diagnosis [5] by differentiating it from MS-mimic inflammatory diseases [3] that appear similar on magnetic resonance (MR) images. Classification of WM lesions that were formed around a vein (P+) are nowadays performed manually using an MR sequence specifically designed for this purpose, fluid-attenuated inversion recovery (FLAIR)* [4]. Despite the fact that vessels in P+ WM lesions are well-contrasted in FLAIR*, manual classifications are tedious and time-consuming, and potentially suffer from low inter-rater reliability.

Here, we propose an ensemble of small 3D convolutional networks, each trained on a different subset of ground truth data in cross-validation, to classify P+ and P- WM lesions. Our approach leverages the diversity inherent in the networks trained on different folds.

2 Data

Thirty-three patients (10 males, 23 females, median age 39 years, age range: 20-70) with MS, and other MS-mimic diseases, were scanned on a 3T Intera (Philips, Best, The Netherlands) MRI scanner. The MR imaging protocol included: i) 3-dimensional (3D) T2*-weighted echo-planar (EPI) and ii) 3D T2-FLAIR acquired during or after intravenous injection of gadolinium-based contrast. FLAIR* images were generated as follows: i) rigid registration between T2-FLAIR and T2* images, ii) upsampling of T2-FLAIR image to T2* resolution, and iii) voxelwise multiplication. P+ and P- lesions were identified and delineated manually in the entire cohort (see examples in Figure 1). The manual assessment was performed by consensus between two neurologists with expertise on MS imaging and blinded to the diagnosis. The raters were trained regarding the central vein assessment according to the consensus criteria as reported in [5]. A 3D patch with size $21 \times 21 \times 21$ for each

*The first two authors contributed equally to this work.

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

lesion was derived from FLAIR*. The intensity values were normalized using z-score and the center of mass of each lesion was considered the center of the respective patch.

3 Methods

To obtain a competitive baseline, we compare the neural network approach to a multi-scale vesselness filter that is specifically designed to detect blood vessels. The 3D FLAIR* patches were used as input for both methods. Based on the manual labelling of WM lesions, our cohort of patients showed a total number of 414 lesions (202 P+, 212 P-). The dataset was divided in three sub-datasets: i) training (315 samples) and ii) validation sets (36 samples) using 10-fold cross-validation and iii) a pure testing set (63 samples from a sub-cohort of 5 patients) that was untouched during the learning phase.

To obtain predictions on the test set, we combined the predictions from all fold classifiers via averaging or majority voting (see below), aiming to exploit variability due to training folds in order to provide approximate independence between classifier outputs. The Condorcet Jury theorem [1] states that performance increases monotonically with the number of (IID) classifiers, given they all provide above-chance predictions, suggesting that this approach should improve performance.

3.1 Classifying P+ lesions with a vesselness filter

Each FLAIR* patch was analysed with multiscale vessel enhancement filtering [2]. This technique explores the local second-order structure (curvature) information in the image, based on the eigenvalues of the local 3D Gaussian-based Hessian matrix. Six spatial scales (vessel thicknesses) were probed by adjusting the Gaussian FWHM by visual inspection of 2 training samples, and finally combined to yield a multiscale vessel likelihood map. As a final step, the respective lesion mask was used to mask out the classified vessel voxels in the regions outside the lesion. Finally, likelihood maps were binarised, and the ratio (RVL) between the number of voxels classified as vessel and the total lesion volume was obtained. To define the optimal threshold in order to binarise the likelihood maps, the ten folds from cross validation were analysed using the area under the receiver operating characteristic (ROC) curve (AUC). The optimal threshold was selected using the majority vote across all folds. Identical ROC analysis was used to find the RVL that maximized the separation between P+ and P- lesions. The code was implemented in Matlab R2013a and takes around 3 seconds to process one 3D patch.

3.2 Classifying P+ lesions with an ensemble of 3D convolutional neural networks

We used a small feedforward architecture broadly inspired by a 2D VGGNet. We have three convolutional layers, each with a 3D convolution, followed by a ReLU and dropout ($p = 0.5$). The convolution kernel sizes are $(3 \times 3 \times 3 \times 16), (3 \times 3 \times 3 \times 32), (3 \times 3 \times 3 \times 64)$. This is followed by a fully connected layer of size 32, then a fully connected layer of size 2 with sigmoid activation. The architecture comprises 71810 trainable parameters. Note that the CNN did not use a lesion mask to isolate relevant parts of the 3D patch, and was thus at a disadvantage compared to the vesselness filter. We used categorical cross-entropy loss, training with minibatch SGD (Adam) for 200 epochs with a minibatch size of 20, checking for approximate class balance for each resample. The weights were initialized with Xavier Gaussian initialization. Data augmentation used three versions of each patch rotated 90 degrees around one axis, and was applied only to the training and validation sets. For each of the ten cross-validation folds, we saved the network that had the lowest validation loss. While the validation loss was noisy over epochs, we have found this to be a reasonable heuristic, in the sense that it strikes a compromise between early stopping and allowing sufficient epochs to promote diversity for later ensembling. The code was implemented in Python, using Keras v2.1.5, scikit-learn v0.19.1, and Tensorflow 1.6.0.

4 Results

Figure 1 shows examples of the output from both algorithms for each type of lesion. Table 1 sumarizes the results of sensitivity, specificity, accuracy and balanced accuracy obtained in the validation and testing sets. Results from the vesselness filter without prior lesion masking were omitted due to it very low performance on this classification problem.

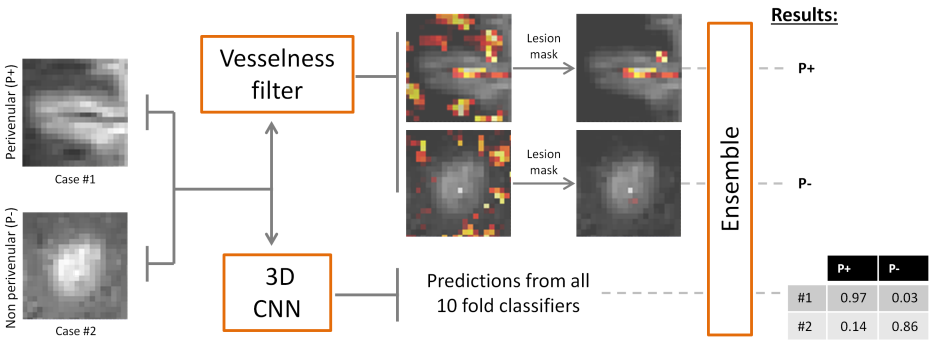

Figure 1: Examples of data and output from vesselness filter and ensemble 3D CNN.

Table 1: Performance evaluation of both methods on the validation and testing datasets. Average and standard deviation across ten folds are reported for the validation set. For the testing dataset, the final prediction was obtained by using the average or majority vote from the ten folds.

	Sensitivity	Specificity	Accuracy	Balanced Accuracy	AUC
Validation set					
Vesselness filter	57.0 ± 13.3	76.8 ± 17.3	67.3 ± 8.2	66.9 ± 7.3	79.4 ± 6.6
3D CNN	66.0 ± 8.3	76.5 ± 8.4	70.5 ± 4.0	71.3 ± 4.0	78.8 ± 4.7
Pure testing set					
Average decision					
Ensemble Vesselness filter	81.0	57.1	73.0	69.1	74.3
Ensemble 3D CNN	81.0	71.4	77.8	76.2	79.3
Majority vote decision					
Ensemble Vesselness filter	47.6	85.7	60.3	66.7	66.7
Ensemble 3D CNN	81.0	71.4	77.8	76.2	76.2

5 Discussion and conclusion

Our relatively simple and shallow architecture is competitive with one of the most commonly used filters for blood vessel detection, even when the filter is aided by a lesion mask to pinpoint the location of the lesion. This relatively good performance comes at the expense of training time and model complexity. Given that veins within lesions do not seem to have a preferential orientation, it is likely that performance can be further improved by increasing data augmentation to more rotations or other augmentation means. Simple architecture tweaks such as adding skip connections, and substituting our basic ReLUs with leaky, parametric, or random ReLUs, as well as using RMSProp rather than Adam for optimization is likely to result in further incremental gains.

References

[1] P. J. Boland. “Majority Systems and the Condorcet Jury Theorem”. In: *The Statistician* 38.3 (1989), p. 181.

[2] A. F. Frangi et al. “Multiscale vessel enhancement filtering”. In: *Medical Image Computing and Computer-Assisted Intervention — MICCAI’98*. 1998, pp. 130–137.

[3] P. Maggi et al. “Central vein sign differentiates Multiple Sclerosis from central nervous system inflammatory vasculopathies”. In: *Annals of Neurology* 83.2 (2018), pp. 283–294.

[4] P. Sati et al. “FLAIR*: A Combined MR Contrast Technique for Visualizing White Matter Lesions and Parenchymal Veins”. In: *Radiology* 265.3 (2012). PMID: 23074257, pp. 926–932.

[5] P. Sati et al. “The central vein sign and its clinical evaluation for the diagnosis of multiple sclerosis: a consensus statement from the North American Imaging in Multiple Sclerosis Cooperative”. In: *Nature Reviews Neurology* 12.12 (2016), pp. 714–722.