HIST: HyperIntensity Segmentation Tool

Jose Manjón, Pierrick Coupé, Parnesh Raniga, Ying Xia, Jurgen Fripp, Olivier Salvado

HIST: HyperIntensity Segmentation Tool

Jose V. Manjón[1], Pierrick Coupé[2,3], Parnesh Raniga[4], Ying Xia[4], Jurgen Fripp[4], Olivier Salvado[4]

[1]Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain
[2] Univ. Bordeaux, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.
[3] CNRS, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.
[4] Australian e-Health Research Centre, CSIRO, Brisbane QLD 4029, Australia.

Abstract. Accurate quantification of white matter hyperintensities (WMH) from MRI is a valuable tool for studies on ageing and neurodegeneration. Reliable automatic extraction of WMH biomarkers is challenging, primarily due to their heterogeneous spatial occurrence, their small size and their diffuse nature. In this paper, we present an automatic and accurate method to segment these lesions that is based on the use of neural networks and an overcomplete strategy. The proposed method was compared to other related methods showing competitive and reliable results in two different neurodegenerative datasets.

1 Introduction

White matter hyperintensities (WMH) are regions of increased MR signal in T2-Weighted (T2W) and FLuid Attenuated Inversion Recovery (FLAIR) images that are distinct from cavitations. The role of WMH have been extensively studied in normal ageing, cerebrovascular disease and dementia. The presence, topography and volume of WMH is used as biomarkers for stroke, small vessel cerebrovascular disease (CVD), dementia [1] and in multiple sclerosis (MS) [2]. In order to utilize WMH as a biomarker of cerebrovascular health, they need to be segmented and quantified in terms of volume and localization.

Manual segmentation of WMH is a demanding process requiring trained observers and several hours per image for manual delineation by an expert. Moreover, manual segmentation is prone to inter and intra-observer variability. With many large clinical studies being conducted into ageing, cerebrovascular disease and dementia, the need for robust, repeatable and automated techniques for the segmentation of WMH is essential. Over the years, several methods have been proposed for the automated segmentation of WMH in cerebrovascular disease and in multiple sclerosis.

Broadly, these methods can be classified into supervised and unsupervised. Supervised methods require a training dataset where manual labels for the WMH are available. Unsupervised methods rely on features and/or domain knowledge together with clustering type approaches. Among them, the lesion growth algorithm (LGA) available in the lesion segmentation toolbox (LST) is widely used [3]. LGA requires both

T1W and FLAIR images to first compute a map of possible lesions and then to use these candidates as seeds to segment entire lesions. Also included in the LST toolbox, there is a newer method called lesion prediction algorithm (LPA) that only requires FLAIR images as input. Besides, the LTS toolbox, Weiss *et al.* [4] proposed a dictionary learning-based approach that segments lesions as outliers from a projection of the dataset onto a normative dictionary. Other popular unsupervised approaches use either clustering approaches such as fuzzy c-means [5] or heuristics on histograms of the T2W or FLAIR images [6]. While unsupervised methods should be preferred since not requiring a training set, in practice the free parameters need to be trained on each dataset to get optimum results.

Supervised methods for WMH segmentation typically involve machine learning methods at a voxel level with pre and post processing steps to improve the sensitivity and specificity of the results. Such machine learning approaches have included for example random forests [7], artificial neural networks [8] or multiatlas patch-based label fusion methods [9]. All these methods are trained on mono and multi-spectral normalized neighborhood voxel intensities within a standardized anatomical coordinates. Although the proposed feature involved neighborhood information (e.g., patch), these methods performed classification step at the voxel level. Therefore, by performing voxel-wise classification these methods ignore the inherent spatial correlation in lesions what results in lower segmentation performance.

To overcome the lack of local consistency of methods performing voxel-wise classification, we propose an automatic pipeline for FLAIR hyperintense lesion segmentation that performs patch-wise lesions segmentation in an over-complete manner. This pipeline benefits from some preprocessing steps aimed at improving the image quality and to locate it in a standardized geometrical and intensity space. Finally an over-complete Neural network classifier is used to segment the lesions. The proposed method is compared to related state-of-the-art methods and shows competitive results.

2 Material and Methods

The aim of the proposed method is to automatically segment the hyperintense lesions visible on MR FLAIR brain images. The proposed method first preprocesses the images to improve their quality and to locate them into a standarized geometric and intensity space to take benefit of the redundant information among the images. Then preprocessed images and the corresponding label maps are used to train a neural network classifier.

Preprocessing

Denoising: The Spatially Adaptive Non-local Means Filter was applied to reduce the noise in the images. This filter was chosen because it is able to automatically deal with both stationary and spatially varying noise levels [10]. *MNI affine registration*: All the images have to be translated into a common coordinate space so the anatomy within the case to be segmented and the library cases is consistent. To this end, the

images were linearly registered (affine transform) to the Montreal Neurological Institute (MNI) space using the MNI152 template. This was performed using the Advanced Normalization Tools (ANTs) [11]. *Inhomogeneity correction and Brain extraction*: SPM12 was used to remove the inhomogeneity of the images and to provide a rough segmentation of the brain tissues [12]. A binary brain mask was estimated by summing grey matter, white matter probability maps and thresholding at 0.5. The resulting binary mask was further refined by applying an opening morphological operation (using a 5x5x5 voxel kernel). *Intensity normalization*: The estimated brain mask was used to select only brain voxels. The resulting volume was intensity standardized by dividing all brain voxels by the median intensity of the brain region.

Supervised classification:

Once the images have been located in a common geometric and intensity space we train a neural network to automatically segment the lesions. This is done in two steps: 1) Region Of Interest (ROI) selection and 2) Network training.

ROI selection: After brain extraction, we can just focus on brain voxels to locate the lesions. However, the brain region has millions of voxels and only small amount of them belong to lesions. As lesions in FLAIR images are hyperintense a simple thresholding allows preselecting a ROI containing mainly voxels belonging to lesions. We used a global threshold of 1.25 as it was found to be optimal in our experiments.

Neural Network training: After ROI selection, the selected ROI contains a mixture of normal tissue and lesion voxels. To segment only lesion voxels a neural network was trained using a library of preprocessed FLAIR images and the corresponding labeled images only patches belonging to their corresponding ROIs were used as training set).

- *Features*: The features used to train the network are a 3D patch around the voxel to be classified, the x, y and z voxel coordinates in MNI space and a value representing the *apriori* lesion probability obtained from a map computed averaging all training label maps in the MNI space (convolved with a Gaussian kernel (5 mm)). We used squared patch intensities to enhance the contrast between the lesions and the surrounding white matter. In our experiments we used a patch size of 3x3x3 voxel that leads to a total number of 31 features.
- *Network topology*: A classic multilayer perceptron was used. Two different settings were tested, voxel-wise and patch-wise. In the first case the network that we used had 31x27x1 neurons (one hidden layer). In the second case, we used a 31x54x27 network (labeling the whole patch rather just the central voxel). In this case an overcomplete approach was used so each voxel has contributions from several adjacent patches. This was done to enforce the final label regularity.

As a final step, the resulting lesion mask is mapped back to the native space by applying the corresponding inverse affine transform. Given a new case to be segmented, we first preprocess it following the previously described steps. Then, we apply the trained

neural network over the ROI selection. Finally, the obtained the lesion mask is registered back to the image native space. The total processing time of the full pipeline takes around 3 minutes. We called the proposed method HIST (for Hyperintense Segmentation Tool.

3 Experiments and results

All experiments were performed using MATLAB 2015a and its neural network toolbox on a standard PC (intel i7 6700 and 16 GB RAM) running Windows 10. For our experiments we used two different datasets consisting in T1 and FLAIR cases with their corresponding manual lesion segmentations.

Data description

AIBL dataset: In this work, we used a set of 128 subjects (including a wide range of white matter lesion severity, aged 38.6-92.1, male/female: 60/68) selected from the Australian Imaging Biomarkers and Lifestyle (AIBL) study (www.aibl.csiro.au) [13]. FLAIR scans were acquired for all the subjects on a 3T Siemens Magnetom TrioTim scanner using the following parameters: TR/TE: 6000/421 ms, flip angle: 120º, TI: 2100 ms, slice thickness: 0.90 mm, image matrix: 256×240, in-plane spacing: 0.98 mm. The ground truth for training and evaluating the proposed method was generated by manual delineation of the hyperintense lesions from all the FLAIR images by PR using MRIcro. Lesion boundaries were delineated on axial slices after bias correction and anisotropic diffusion smoothing.

MICCAI dataset: To further validate our proposed method, we used a publically available clinical dataset provided by the MS lesion segmentation challenge at MICCAI 2008 [14]. As done in [3] we focused on the 20 available training cases. The data comes originally from the Children's Hospital Boston (CHB) and the University of North Carolina (UNC). Although there are available T1 and FLAIR images in this dataset we used just the FLAIR.

AIBL dataset results

The AIBL dataset (N=128) was split in two subsets, one for training (N=68) and one for testing (N=60). Several neural network configurations were trained using the training subset and later applied to the testing subset. To measure the quality of the proposed method we used the Dice coefficient, sensitivity, specificity and volume correlation coefficient.

First, we measured the results using only the selected ROI as final segmentation to figure out how many outliers were included in the initial ROI. The average dice coef-

ficient for our simple ROI selection procedure was 0.5960±0.1597, a sensitivity of 0.7237, a specificity of 0.9995 and volume correlation of 0.9828. As can be noticed just a simple thresholding gives a relatively high dice coefficient although at the expense of having a large number of false positives.

Second we optimized the voxel-wise neural network (selecting randomly 20000 samples out of the full almost 1.000.000 samples training set). In this case, the average dice coefficient was 0.6447±0.1477, a sensitivity of 0.9188, a specificity of 0.9995 and volume correlation of 0.9921. As seen, the neural network was able to leverage the results by removing many false positives (but also added some false negatives).

Finally, we also optimized the patch-wise neural network. In this case, the average dice coefficient was 0.7521±0.1201, a sensitivity of 0.8171, a specificity of 0.9997 and volume correlation of 0.99. This overcomplete patch-wise version was able to further improve the dice results providing also very regular masks.

Comparison with other methods:

We compared HIST method with related state-of-the-art methods to assess its quality. We compared the proposed method with the two methods included in the LST toolbox (http://www.applied-statistics.de/lst.html). The first was the LGA method that uses both T1W and FLAIR images [5] and the second was the LPA that only requires a FLAIR image to perform the lesion segmentation. In table 1 the dice coefficient for these methods and for different lesion sizes is presented. In table 2 the volume correlation is also presented showing that HIST method has a stronger volume correlation (0.99). Finally, Figure 1 shows the boxplot graphs of dice, sensitivity, specificity and an example of the segmentation results.

Table 1. Mean dice coefficient. Best results in bold. .

Method	Lesion size*			
	Small	Medium	Big	All
LST-LGA	0.4518±0.1531	0.6685±0.0696	0.7668±0.0406	0.6261±0.1596
LST-LPA	0.4973±0.1688	0.7102±0.0983	0.7886±0.0679	0.6637±0.1669
HIST	**0.6441±0.1449**	**0.7764±0.0587**	**0.8423±0.0399**	**0.7521±0.1201**

Table 2. Volume correlation. Best results in bold. .

Method	Lesion size*			
	Small	Medium	Big	All
LST-LGA	0.7712	0.8841	0.9732	0.9836
LST-LPA	**0.8178**	0.7690	0.9649	0.9736
HIST	0.7882	**0.8877**	**0.9849**	**0.9900**

**Small(<4 ml), medium(4 ml to 18 ml), big(>18 ml)*

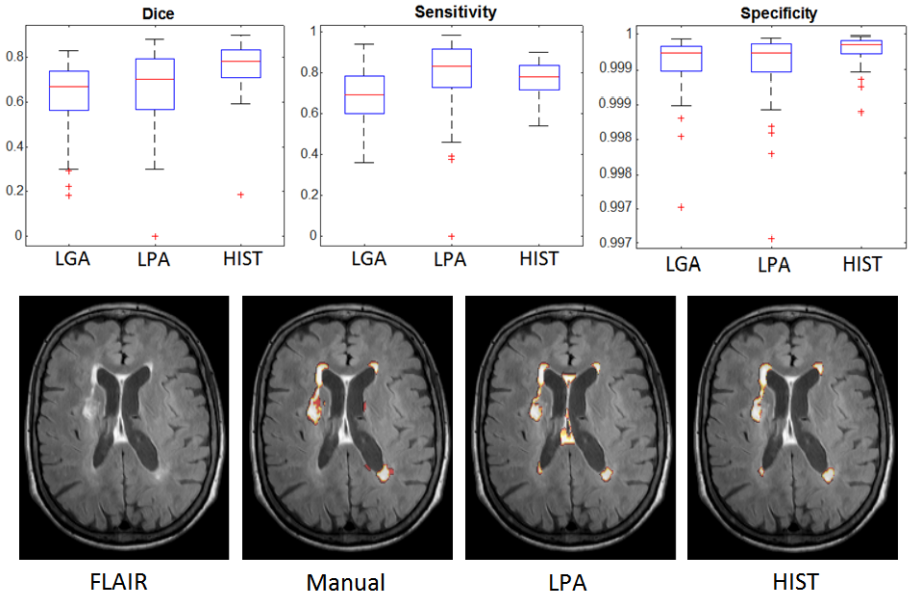

Fig. 1. AIBL dataset example results. In the upper row: dice, Sensitivity and specificity. In the lower row: a visual example of LPA and HIST segmentation results.

MICCAI 2008 dataset results

We compared the results of the proposed approach with a recent method that was applied to the same dataset [4]. In this case we used the same metrics used in Weiss paper (i.e. True Positive Rate (TPR), Positive Predictive Value (PPV) and Dice coefficient). In table 3 the results of this comparison are presented. As can be noticed, HIST method obtained the best results overall for the 3 metrics showing that the features learned on AIBL dataset were useful to segment lesions in other datasets.

4 Discussion

In this paper we presented a new method for hyperintense lesion segmentation based on an overcomplete patch-wise neural network strategy. We have shown that the overcomplete approach significantly improved the voxel-wise network by enforcing the regularity of the output masks and also by minimizing the variance of the classification error due to the aggregation of many patch contributions. The proposed method not only provided the best classification results in the comparative but also provided the higher volume correlation (0.99) which indicates that it can be used for fully automated lesion load assessment. Finally, HIST method was used on an independent dataset giving very competitive results demonstrating the generality of the proposed approach.

Table 3. MICCAI 2008 train dataset results. Best results in bold. .

Method	Weiss2013			HIST		
Case	TPR	PPV	Dice	TPR	PPV	Dice
UNC01	**0.33**	**0.29**	**0.31**	0.18	0.28	0.22
UNC02	**0.54**	0.51	**0.53**	0.39	**0.66**	0.49
UNC03	**0.64**	0.27	**0.38**	0.25	**0.32**	0.28
UNC04	**0.40**	0.51	0.45	**0.40**	**0.69**	**0.51**
UNC05	0.25	0.10	0.16	**0.62**	**0.31**	**0.41**
UNC06	**0.13**	**0.55**	**0.20**	0.09	0.20	0.12
UNC07	**0.44**	0.23	0.30	0.24	**0.72**	**0.36**
UNC08	**0.43**	0.13	0.20	0.30	**0.16**	**0.21**
UNC09	**0.69**	0.06	0.11	0.50	**0.37**	**0.42**
UNC10	0.43	0.23	0.30	**0.46**	**0.46**	**0.46**
CHB01	**0.60**	0.58	**0.59**	0.48	**0.68**	0.56
CHB02	0.27	**0.45**	**0.34**	**0.65**	0.20	0.31
CHB03	0.24	**0.56**	0.34	**0.48**	0.33	**0.39**
CHB04	0.27	**0.66**	0.38	**0.72**	0.49	**0.58**
CHB05	0.29	0.33	0.31	**0.39**	**0.50**	**0.44**
CHB06	0.10	0.36	0.16	**0.41**	**0.37**	**0.39**
CHB07	0.14	0.48	0.22	**0.51**	**0.50**	**0.51**
CHB08	0.21	**0.73**	0.32	**0.50**	0.47	**0.49**
CHB09	0.05	**0.22**	0.08	**0.40**	0.14	**0.21**
CHB010	0.15	0.12	0.13	**0.66**	**0.16**	**0.26**
All	0.33	0.37	0.29	**0.43**	**0.40**	**0.38**
All UNC	**0.43**	0.29	0.29	0.34	**0.42**	**0.35**
All CHB	0.23	**0.45**	0.29	**0.52**	0.39	**0.41**

5 Acknowledgements

This research has been done thanks to the Australian distinguished visiting professor grant and the Spanish “Programa de apoyo a la investigación y desarrollo (PAID-00-15)” of the Universidad Politécnica de Valencia. This study has been carried out with support from the French State, managed by the French National Research Ageny in the frame of the Investments for the future Program IdEx Bordeaux (ANR-10-IDEX-03-02, HL-MRI Project), Cluster of excellence CPU and TRAIL (HR-DTI ANR-10-LABX-57) and the CNRS multidisciplinary project "Défi imag'In".

6 References

1. Debette, S., Markus, H.S.: The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. BMJ. 341, c3666 (2010).
2. Filippi, M., Rocca, M.A. MR imaging of multiple sclerosis. Radiology. 259, 659–681 (2011).
3. Schmidt, P., Gaser, C., Arsic, M., Buck, D., Förschler, A., Berthele, A., Hoshi, M., Ilg, R., Schmid, V.J., Zimmer, C., Hemmer, B., Mühlau, M.: An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. Neuroimage. 59, 3774–3783 (2012).
4. Weiss, N., Rueckert, D., Rao, A.: Multiple sclerosis lesion segmentation using dictionary learning and sparse coding. Med Image Comput Comput Assist Interv. 16, 735–742 (2013).
5. Admiraal-Behloul, F et al. Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. NeuroImage. 28, 607–617 (2005).
6. Jack, C.R., O'Brien, P.C., Rettman, D.W., Shiung, M.M., Xu, Y., Muthupillai, R., Manduca, A., Avula, R., Erickson, B.J.: FLAIR histogram segmentation for measurement of leukoaraiosis volume. J Magn Reson Imaging. 14, 668–676 (2001).
7. Ithapu, V., Singh, V., Lindner, C., Austin, B.P., Hinrichs, C., Carlsson, C.M., Bendlin, B.B., Johnson, S.C.: Extracting and summarizing white matter hyperintensities using supervised segmentation methods in Alzheimer's disease risk and aging studies. Hum Brain Mapp. 35, 4219–4235 (2014).
8. Dyrby, T.B. et al. Segmentation of age-related white matter changes in a clinical multi-center study. NeuroImage. 41, 335–345 (2008).
9. Guizard N, Coupé P, Fonov V, Manjón J. V., Douglas A, Collins D. L. Rotation-invariant multi-contrast non-local means for MS lesion segmentation. *Neuroimage: Clinical*, 8: 376-389 (2015).
10. Manjón, J.V., Coupé, P., Martí-Bonmatí, L., Collins, D.L., Robles, M.: Adaptive non-local means denoising of MR images with spatially varying noise levels. J Magn Reson Imaging. 31, 192–203 (2010).
11. Avants, B., Tustison, N., Song, G.: Advanced Normalization Tools: V1.0. (2009).
12. Ashburner, J., Friston, K.J.: Unified segmentation. Neuroimage. 26, 839–851 (2005).
13. Ellis, K.A. et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr. 1–16 (2009).
14. Styner, M., Lee, J., Chin, B., Chin, M.S., Commowick, O., Tran, H.-H., Markovic-Plese, S., Jewells, V., Warfield, S.: 3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation. (2008).