LesionBrain: An Online Tool for White Matter Lesion Segmentation

Pierrick Coupé, Thomas Tourdias, Pierre Linck, Jose V Romero, Jose V. Manjon

LesionBrain: An Online Tool for White Matter Lesion Segmentation

Pierrick Coupé[1,2], Thomas Tourdias[3,4,5], Pierre Linck[4,5], José E. Romero[6] and José V. Manjón[6]

[1] CNRS, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France
[2] Univ. Bordeaux, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.
[3] Neurocentre Magendie, INSERM U1215 F-33077 Bordeaux, France
[4] Univ. Bordeaux, F-33000 Bordeaux, France
[5] CHU de Bordeaux, Services de Neurologie et Neuroradiologie, Bordeaux, France,
[6] 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, España

Abstract. In this paper, we present a new tool for white matter lesion segmentation called lesionBrain. Our method is based on a 3-stage strategy including multimodal patch-based segmentation, patch-based regularization of probability map and patch-based error correction using an ensemble of shallow neural networks. Its robustness and accuracy have been evaluated on the MSSEG challenge 2016 datasets. During our validation, the performance obtained by lesionBrain was competitive compared to recent deep learning methods. Moreover, lesionBrain proposes automatic lesion categorization according to location. Finally, complementary information on gray matter atrophy is included in the generated report. LesionBrain follows *a software as a service* model in full open access.

Keywords: White Matter Lesion Segmentation, Patch-based Segmentation, Service as a Software.

1   Introduction

The presence of white matter lesions (WML) is associated with different brain diseases such as multiple sclerosis (MS), small vessel disease or head injury among others, but it also occurs in normal aging. Magnetic resonance imaging (MRI), especially FLAIR images, has been found to be very sensitive in the detection of these WML. Therefore, MRI is the reference standard to identify WML and it plays a crucial role in the diagnosis and the monitoring of many neurological pathologies. Despite the importance of quantifying WML, this task remains mainly based on manual counting of lesions or semi quantitative scores such as Fazekas score. Manual delineation for volumetric analyses is extremely time-consuming and prone to errors due to inter- and intra-rater variability. As a result, the automation of WML segmentation has received a great deal of attention during the last decade and a wide range of methods have been

proposed [1]. These methods are usually classified into two categories, unsupervised and supervised. Unsupervised methods do not require a training dataset with manual segmentation of the lesions. These methods estimate lesions mainly using MRI intensities and some anatomical knowledge. They can be based on Bayesian models, Graph-cut [2] or thresholding approaches [3] among others. Supervised methods require a training dataset including manual segmentations of experts to learn from examples. Many different techniques have been proposed such as Random Forest [4], Patch-based methods [5, 6] and more recently deep learning methods [7-9]. Although automatic methods are becoming more and more accurate, manual segmentation remains used especially in clinical research or clinical trials in which very accurate quantification is needed to use lesion load as judgement criteria. Several factors can explain the difficulty to apply automatic methods in clinical context.

First, validating the accuracy of WML segmentation methods is challenging because of the difficulty to define a ground truth. Indeed, the high intra and inter-rater variability makes difficult to define a gold standard. Moreover, the lack of freely available annotated datasets leads to highly heterogeneous validation in the literature making methods comparison arduous. Therefore, it is difficult to appreciate the respective performances of automatic methods and their potential under clinical conditions. Recently, important efforts have been done to limit these aspects by sharing freely available datasets based on the consensus of several experts [10]. As a result, evaluation and comparison of methods become easier and more reliable. In this paper, we propose a new tool called lesionBrain which is an extension of the rotationally-invariant nonlocal means (RI-NLM) segmentation method [5]. To evaluate its performance compared to state-of-the-art methods, the validation is carried out on the MSSEG MICCAI Challenge 2016 dataset which is freely available providing a high quality ground truth based on the consensus of seven experts.

Second, few methods are freely available making their use in clinic research difficult. When available, these methods are usually distributed as packages that need to be downloaded, installed and configured. Installation steps can be complicated and thus may require experimented persons not always available in a research laboratory and especially in clinical context. In addition, users have to be trained to use the software and computational resources have to be allocated to run it. These requirements can make the use of these packages complex, especially the most recent and sophisticated ones requiring advanced hardware configuration (e.g., advanced GPU). To address this issue, lesionBrain is proposed as an online open access solution following the model of Software as a Service (SaaS). Our method works remotely through a web-interface and does not require any installation, resources or human interaction.

In addition, automatic methods generally provide the volume of WML as the sole output. However, complementary information can be relevant from a clinical point of view. Indeed, the location of lesions is useful to establish a diagnosis of multiple sclerosis after a first clinical episode according to the McDonald diagnosis criteria for MS[11]. To provide this information, lesionBrain proposes a lesion classification based on their proximity to lateral ventricles, cerebral cortex or cerebellum and brain stem. As a result, the lesion load in volume and also the number of lesions are provided for periventricular, juxtacortical, infratentorial and deep white matter areas.

Finally, most of the existing tools provide information focused on WML. However, complementary information from other structures might be needed to better study brain pathologies globally. For instance, gray matter (GM) atrophy can provide relevant information to investigate the neurodegenerative impact of MS or Alzheimer's Disease (AD). Therefore, lesionBrain not only provides volumetric measurement on WML but also a quantification of WM, GM and Cerebrospinal fluid (CSF). When age and gender of the subject are available, the volumes of these brain tissues are compared to reference values derived from lifespan models to detect abnormalities [12].

2 Material and Methods

2.1 Datasets

LesionBrain dataset: Our training dataset is composed of 43 patients who underwent 3T 3D-T1w MPRAGE and 3D-Fluid-Attenuated Inversion Recovery (FLAIR) MRI. The preprocessing steps described in the next subsection have been applied to all the images to align them into the MNI space and to normalize their intensities. Afterwards, a first expert performed manual segmentations in the MNI space for all the patients with ITKsnap [13] using T1w and FLAIR images. Then, a second expert validated and/or corrected all the manual segmentations. At the end, all the images were flipped as done in [14] to double the size of our training library (i.e., 86 training images).

MSSEG MICCAI Challenge 2016 dataset: To evaluate our tools, we used the dataset of the MSSEG MICCAI Challenge 2016 [10]. For this dataset, 15 patients underwent 3D-T1w MPRAGE, 3D-FLAIR, Gadolinium- enhanced T1w, Proton Density (PD), and T2w MRI. Only T1w and FLAIR MRI were used during our experiments. These 15 subjects consist in 3 groups of five subjects scanned with Philips Ingenia 3T, Siemens Aera 1.5T and Siemens Verio 3T. All the images have been manually delineated by seven experts. Finally, the experts' consensus is used as gold standard.

2.2 Pipeline description

Preprocessing: First, the images are preprocessed to normalize their intensity and to register them into the MNI space. A denoising step based on the adaptive nonlocal means filter is first applied to T1w and FLAIR images [15]. Both denoised MRI are then coarsely corrected for inhomogeneity [16]. Afterwards, the T1w is registered into the MNI space using an affine transform [17]. FLAIR is then registered to T1w in the MNI space. A fine inhomogeneity correction is performed on both images [18]. Finally, brain tissue maps (i.e., WM, GM and CSF) are obtained using [19]. These tissue maps are used to perform intensity normalization based on a piece-wise linear scaling of intensity where the median intensity of each tissue is set to a fixed value [20].

Structure segmentation: The T1w is used to segment several anatomical structures. First, the intracranial cavity (ICC) is extracted using [21] and brainstem and cerebellum using [22]. Finally, lateral ventricles are segmented using [23].

Candidate map: To reduce computational time, the segmentation is performed only on areas which potentially contain lesions as defined below. As done in [4, 6], the mean μ and the standard deviation σ of the GM FLAIR intensities are used to estimate a threshold (th $= \mu + \alpha\sigma$, with $\alpha = 0.5$). All voxels above this threshold and within the ICC mask are considered as lesion candidates. However, FLAIR intensity within lesion may sometimes be below this threshold. Therefore, an atlas of lesions (average of all the manual lesion maps of the lesionBrain dataset in the MNI space) is also used to look for lesions at the most probable location. Voxels at locations with probability higher than 20% to contain a lesion are added to the map of candidates obtained by thresholding.

Lesions segmentation: Lesions are segmented using an extension of the RI-NLM method proposed in [5]. On the one hand, such voxel-wise method may produce false positive detections especially in cortical areas while implicit regularization of multipoint/patch-wise frameworks demonstrated better performance than voxel-wise approaches [20]. On the other hand, using patch-wise methods for lesion segmentation does not enable to efficiently capture heterogeneity of shape, size and location of lesions [5]. Therefore, in lesionBrain, we propose to apply first the RI-NLM method on T1w and FLAIR images to obtain the probability map of lesions. Second, we achieve a regularization of the probability map using a patch-wise NLM denoising filter [24]. The weights of the NLM filter are estimated on the FLAIR and then used to average the probabilities. The RI-NLM takes advantage of inter-subject similarity while patch-wise NLM regularization (NLMr) takes advantage of intra-subject similarity. Finally, a systematic error correction step is performed to obtain the final segmentation. Automatic correction of systematic errors was first proposed in [25] with SegAdapter. In lesionBrain, we used the Patch-based Ensemble Corrector (PEC) proposed in [26]. Contrary to SegAdapter which is based on a voxel-wise Adaboost classifier, PEC involves patch-wise ensemble of multilayer perceptron classifiers. Recently, second-pass strategy such as cascade of Convolutional Neural Networks (CNN) [9] demonstrated high performance to limit false positive detection.

Lesions classification: Once the lesions are segmented, a last step is performed to classify them into the following categories: periventricular, juxtacortical, deep white and infratentorial. Such classification might be clinically relevant since some diagnose criteria of MS are based on it [11]. Therefore, all the lesions located within 3 voxels (i.e., 3 mm in the MNI space) from the lateral ventricles, the GM map, and the union of brainstem and cerebellum are classified respectively as periventricular, juxtacortical and infratentorial. The remaining lesions located in WM map are classified as deep white.

Report generation: At the end, a pdf report is automatically generated providing the lesion load, the number of lesions for each class and screenshots of the processed images. Moreover, in case the gender and the age of the patient are provided, the estimated volumes of WM, GM and CSF are compared to expected normal values based on lifespan models [12]. The proposed lesionBrain tool has been integrated into the volBrain[1] platform in full open access [20].

---

[1] http://volbrain.upv.es

2.3   Validation framework

First, the method parameters were validated using training lesionBrain dataset through a K-fold cross validation. For RI-NLM segmentation and NLMr of the probability map, the patch size was set to 3x3x3 voxels as proposed in the original papers [5, 24]. The search area was set to 9x9x9 voxels for RI-NLM and NLMr although 11x11x11voxels is suggested in [5, 24]. This enables to reduce computational time with marginal accuracy loss. The number of used training images was set to the maximum (i.e., 86 when testing on the MSSEG Challenge 2016 dataset). For PEC we used the default parameters [26]. Therefore, the number of networks was set to 10 and the two patch scales to 3x3x3 voxels and 7x7x7 voxels. During the validation, we first evaluate the improvement in terms of mean DICE coefficient provided by each component of the proposed segmentation pipeline – RI-NLM, RI-NLM+NLMr and RI-NLM+NLMr+PEC (i.e., lesionBrain). Then, lesionBrain is compared with six state-of-the-art methods. To this end, we used the mean DICE coefficient published by authors who have evaluated their method on the 15 MS patients of the training MSSEG Challenge 2016 dataset as we did here. First, lesionBrain is compared with two unsupervised methods based on graph-cut [2] and thresholding as implemented in LST-LPA [3]. In addition, the proposed method is compared with four supervised methods including Random Forest [4] and recent advanced DL methods such as U-Net [7], Nabla-Net [8] and Dense-Net [7]. Finally, the inter-expert variability estimated in [4] between the seven experts is provided for reference purposes.

3   Results

First, Table 1 presents the mean DICE coefficient obtained with RI-NLM, RI-NLM+NLMr and lesionBrain of the MSSEG Challenge 2016 dataset. These results show that each component of the pipeline improved the segmentation accuracy. The mean DICE increased from 66.59% to 69.27% with the NLMr of the probability map and from 69.27% to 72.49% with PEC. Both improvements were found to be significant when tested with a paired t-test. This demonstrates the advantage of combining methods based on inter-subject similarity, intra-subject self-similarity and correction of systematic errors. Table 1 also shows the comparison of lesionBrain with six state-of-the-art methods. First, lesionBrain obtained the best mean DICE coefficient with 72.49 followed by the Dense-Net proposed in [24] which obtained 70.30. It has to be noted that lesionBrain only requires 2 contrasts while Dense-Net uses 5 contrasts. Increasing the number of sequences has a negative impact on the acquisition time, the patient's comfort and the related costs. In addition, the Dense-Net has been trained using cross-validation which can introduce overfitting and thus overestimates the performance of the method. The Nabla-Net proposed in [8] requires only one contrast and has been trained on external dataset. This method obtained a DICE of 67% which is similar to the accuracy obtained by RI-NLM with 2 contrasts, but less than the accuracy obtained with RI-NLM+NLMr or lesionBrain.

Table 1. Methods comparison on the 15 MS patients of the MSSEG challenge 2016 dataset in term of mean DICE coefficient.

Methods	Mean DICE in %	Training	Modalities
Inter-expert variability [4]	63.02		
lesionBrain	72.49	External	T1w and FLAIR
Dense-Net [7]	70.30	Cross-validation	T1, T1Gd, T2, PD and FLAIR
RI- NLM [5] + NLMr	69.27	External	T1w and FLAIR
Nabla-Net [8]	67.00	External	FLAIR
RI- NLM [5]	66.59	External	T1w and FLAIR
Random Forest [4]	63.80	Cross-validation	T1w and FLAIR
LST-LPA [3]	61.00	Unsupervised	FLAIR
Graph-cut [2]	57.09	Unsupervised	T1, T2 and FLAIR
U-Net [7]	56.42	Cross-validation	T1, T1Gd, T2, PD and FLAIR

Compared to Random Forest [4] which obtained 63.80% of accuracy, RI-NLM, RI-NLM+NLMr and lesionBrain obtained higher accuracy while they require the same contrasts. All these methods obtained accuracy higher than inter-expert variability estimated at 63.02% contrary to the 3 remaining ones. The two unsupervised methods based on graph-cut [3] and LST-LPA [3] obtained a mean DICE of 57.09% and 61% respectively. Finally, the U-Net method proposed in [7] obtained the worst accuracy with 56.42%. These results indicate that supervised methods are ranked among the best, better than inter-expert variability, while unsupervised methods failed to reach inter-expert variability. However, the use of CNN does not necessarily ensure a good accuracy since the worst method is based on a U-Net using 5 contrasts. Finally, Figure 1 shows examples of WML segmentation obtained by lesionBrain for three patients of the MSSEG Challenge 2016 dataset (for best, median and worst DICE).

4   Conclusion

In this paper, we present a new tool for WML segmentation using T1w and FLAIR MRI. Our method combined several complementary patch-based approaches to accurately segment WML. We evaluated its accuracy on the MSSEG challenge 2016 datasets with a strong ground truth based on the consensus of seven experts. During our validation, the performance obtained by lesionBrain were competitive compared to Dense-Net [7], Nabla-Net [8] and U-Net [7]. Moreover, lesionBrain obtained a higher accuracy than the inter-expert variability. Finally, our tool is already integrated into a web-platform in open access.
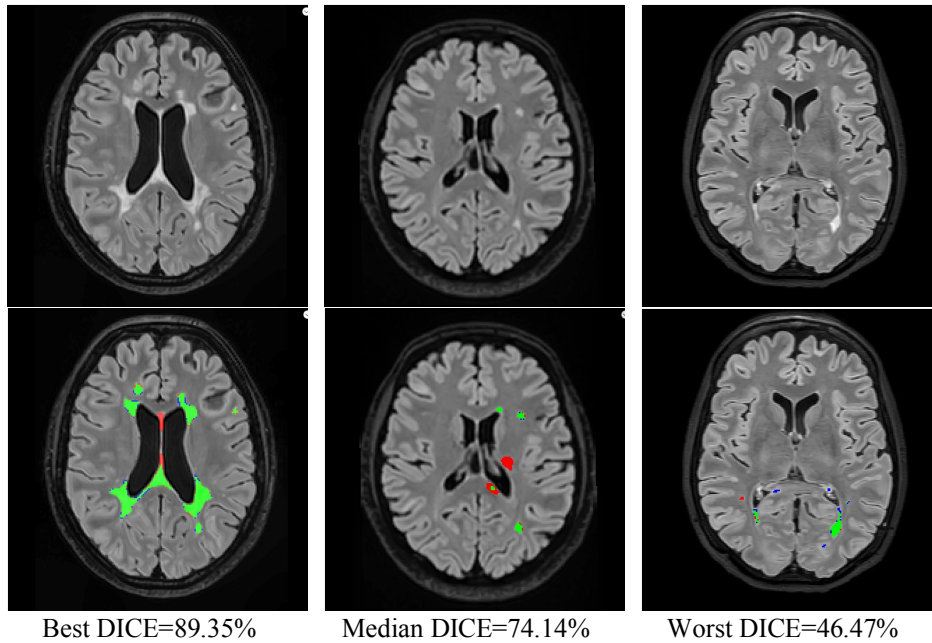
6

| Best DICE=89.35% | Median DICE=74.14% | Worst DICE=46.47% |

Fig. 1. Examples of WML segmentation produced by lesionBrain for best, median and worst DICE obtained on the MSSEG Challenge 2016 dataset. True positives are in green, False Negatives in red and False Positives in blue.

References

1. Garcia-Lorenzo, D., et al., *Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging.* Medical image analysis, 2013. **17**(1): p. 1-18.
2. Beaumont, J., O. Commowick, and C. Barillot. *Multiple Sclerosis lesion segmentation using an automated multimodal Graph Cut.* 2016.
3. Schmidt, P., et al., *An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis.* Neuroimage, 2012. **59**(4): p. 3774-3783.

4. Vera-Olmos, F.J., H. Melero, and N. Malpica, *Random Forest for Multiple Sclerosis Lesion Segmentation.* MSSEG Challenge Proceedings: Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure, 2016: p. 81.
5. Guizard, N., et al., *Rotation-invariant multi-contrast non-local means for MS lesion segmentation.* Neuroimage Clin, 2015. **8**: p. 376-89.
6. Mechrez, R., J. Goldberger, and H. Greenspan, *Patch-based segmentation with spatial consistency: application to MS lesions in brain MRI.* Journal of Biomedical Imaging, 2016. **2016**: p. 1687-4188.
7. Hashemi, S.R., et al., *Tversky as a Loss Function for Highly Unbalanced Image Segmentation using 3D Fully Convolutional Deep Networks.* arXiv preprint arXiv:1803.11078, 2018.
8. McKinley, R., et al., *Nabla-net: a deep dag-like convolutional architecture for biomedical image segmentation: application to white-matter lesion segmentation in multiple sclerosis.* MSSEG Challenge Proceedings: Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure, 2016: p. 37.
9. Valverde, S., et al., *Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach.* NeuroImage, 2017. **155**: p. 159-168.
10. Commowick, O., F. Cervenansky, and R. Ameli. *MSSEG Challenge Proceedings: Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure.* 2016.
11. Thompson, A.J., et al., *Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria.* The Lancet Neurology, 2017: p. 1474-4422.
12. Coupe, P., et al., *Towards a unified analysis of brain maturation and aging across the entire lifespan: A MRI analysis.* Hum Brain Mapp, 2017. **38**(11): p. 5501-5518.
13. Yushkevich, P.A., et al., *User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability.* Neuroimage, 2006. **31**(3): p. 1116-28.
14. Eskildsen, S.F., et al., *BEaST: brain extraction based on nonlocal segmentation technique.* Neuroimage, 2012. **59**(3): p. 2362-73.
15. Manjon, J.V., et al., *Adaptive non-local means denoising of MR images with spatially varying noise levels.* J Magn Reson Imaging, 2010. **31**(1): p. 192-203.
16. Tustison, N.J., et al., *N4ITK: improved N3 bias correction.* IEEE Trans Med Imaging, 2010. **29**(6): p. 1310-20.
17. Avants, B.B., et al., *A reproducible evaluation of ANTs similarity metric performance in brain image registration.* Neuroimage, 2011. **54**(3): p. 2033-2044.
18. Weiskopf, N., et al., *Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT).* Neuroimage, 2011. **54**(3): p. 2116-24.
19. Manjón, J.V., J. Tohka, and M. Robles, *Improved estimates of partial volume coefficients from noisy brain MRI using spatial context.* Neuroimage, 2010. **53**(2): p. 480-490.
20. Manjon, J.V. and P. Coupe, *volBrain: An Online MRI Brain Volumetry System.* Front Neuroinform, 2016. **10**: p. 30.
21. Manjon, J.V., et al., *Nonlocal intracranial cavity extraction.* Int J Biomed Imaging, 2014. **2014**: p. 820205.
22. Romero, J.E., et al., *NABS: non-local automatic brain hemisphere segmentation.* Magn Reson Imaging, 2015. **33**(4): p. 474-84.

23. Coupé, P., et al., *Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation.* Neuroimage, 2011. **54**(2): p. 940-54.

24. Coupe, P., et al., *An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images.* IEEE transactions on medical imaging, 2008. **27**(4): p. 425-441.

25. Wang, H., et al., *A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation.* Neuroimage, 2011. **55**(3): p. 968-85.

26. Romero, J.E., P. Coupé, and J.V. Manjón, *HIPS: A new hippocampus subfield segmentation method.* NeuroImage, 2017. **163**: p. 286-295.