Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2015 February ; 9413: . doi:10.1117/12.2082157.

Multi-Output Decision Trees for Lesion Segmentation in Multiple Sclerosis

Amod Jog^a, Aaron Carass^a, Dzung L. Pham^b, and Jerry L. Prince^a

^aImage Analysis and Communications Laboratory, The Johns Hopkins University

^bHenry M. Jackson Foundation for the Advancement of Military Medicine

Abstract

Multiple Sclerosis (MS) is a disease of the central nervous system in which the protective myelin sheath of the neurons is damaged. MS leads to the formation of lesions, predominantly in the white matter of the brain and the spinal cord. The number and volume of lesions visible in magnetic resonance (MR) imaging (MRI) are important criteria for diagnosing and tracking the progression of MS. Locating and delineating lesions manually requires the tedious and expensive efforts of highly trained raters. In this paper, we propose an automated algorithm to segment lesions in MR images using multi-output decision trees. We evaluated our algorithm on the publicly available MICCAI 2008 MS Lesion Segmentation Challenge training dataset of 20 subjects, and showed improved results in comparison to state-of-the-art methods. We also evaluated our algorithm on an in-house dataset of 49 subjects with a true positive rate of 0.41 and a positive predictive value 0.36.

Keywords

Multiple Sclerosis; lesion; segmentation; multi-output decision trees

1. INTRODUCTION

Multiple Sclerosis (MS) is a chronic, demyelinating disease affecting the central nervous system. The demyelination leads to formation of lesions or sclera, primarily in the white matter (WM) in the brain and spine. These lesions are visible as hyperintensities in fluid attenuated inversion recovery (FLAIR) sequence images and as hypointensities in T_1 -weighted (T_1 -w) images (see Fig. 1). Depending on the type of MS, lesions can increase or decrease in number and volume as a function of time. Lesion count is one of the criteria for diagnosing MS,¹ and lesion volumes can be used to track disease progression.² Lesions are highly variable in their appearance, location, and volume in MRI. In addition, the FLAIR sequence also suffers from artifacts,³ which make lesion segmentation a challenging task. Accurate delineation requires aligning multimodal data and visualizing the tissue boundaries and is a difficult task, even for experienced manual raters. Delineating the lesion-WM

Further author information: (Send correspondence to Amod Jog.), Amod Jog: amod@cs.jhu.edu.

*This work is not being submitted elsewhere for publication or presentation.

boundary can be especially subjective and variable. Manual segmentation also suffers from large intra- and inter-rater variability.⁴ Thus, an automated, fast, and accurate lesion segmentation method is crucial in order to segment and analyze large patient imaging studies.

There has been significant work in the area of MS lesion segmentation.^{5–14} Llado et al.¹⁵ in their survey analyze many recently published automated lesion segmentation techniques, both qualitatively and quantitatively. Some of the methods investigated in the survey were the best performing solutions as determined by the MICCAI 2008 MS Lesion Segmentation Challenge (MSLSC).¹⁶ The winning entry of the MSLSC estimated a threshold for FLAIR images based on the intensity distribution of the gray matter.⁵ Shiee et al.⁶ segmented MS lesions along with a topologically correct full brain segmentation. Recent work by Geremia et al.⁷ has surpassed the performance of the MSLSC winner.

2. NEW WORK TO BE PRESENTED

Our approach is similar to Geremia et al.,⁷ in that we also implement binary decision trees that are learned from intensity and context features. However, instead of predicting the label at a single voxel, as most methods do, we predict an entire neighborhood or patch for a given input feature vector. Lesion boundaries often seem to appear at slightly shifted locations in different modalities. Thus knowledge of predicted tissue classes in neighboring voxels is useful in determining the class of the central voxel. To achieve this, we have implemented a version of multi-output decision trees, which are similar in essence to the output kernel trees developed by Geurts et al.¹⁷ Predicting entire neighborhoods gives us further context information such as the presence of lesions predominantly inside white matter. We post-process the segmentations produced by our multi-output decision trees to reduce false positives.

3. METHOD

Co-registered T_1 -w, FLAIR, and T_2 -w images as well as expert manual segmentations are used as training data in our supervised learning approach. We denote this data collection as $\{\mathcal{I}_p^{(T_1)}, \mathcal{I}_p^{(F)}, \mathcal{I}_p^{(T_2)}, \mathcal{I}_p^{(L)}\}$, $p = 1, \dots, N$ for the images from the p^{th} training subject, corresponding to the T_1 -w, FLAIR, T_2 -w, and labeled image, respectively.

3.1 Preprocessing

All images undergo preprocessing including N4 bias correction,¹⁸ rigid registration to the 1 mm³ MNI space,¹⁹ skull stripping,²⁰ and finally intensity standardization. The T_1 -w images are intensity standardized by linear scaling of the WM peak of the intensity histogram to match the intensity of a reference image. The FLAIR and T_2 -w images are intensity standardized by fixing the histogram mode to a reference by a linear scaling.

3.2 Features

For the p^{th} training image set, $\mathcal{I}_p^{(T_1)}, \mathcal{I}_p^{(F)}, \mathcal{I}_p^{(T_2)}$, at each voxel i , we extract small $u \times v \times w$ -sized patches. Let these patches be $\mathbf{x}_i^{(T_1,p)}, \mathbf{x}_i^{(F,p)}, \mathbf{x}_i^{(T_2,p)}$, respectively for the input

modalities. In our experiments we use $u = v = w = 3$. Small patches provide us with local context of a particular voxel. However due to the variability of lesion appearance and variations in the FLAIR images due to artifacts, just local intensity information cannot distinguish between lesion and normal tissue. Thus we augment the local patches with features that provide global context for a voxel.

For each voxel i , we locate a voxel j , which is at a distance of r from i in a radial direction from i . By radial direction we mean the unit vector in the direction of voxel i from the center of the MNI space. We calculate the mean intensity of a large window of size $w_1 \times w_2 \times w_3$ ($11 \times 11 \times 3$) about the voxel j . This radial patch average provides a good indication of location in the brain relative to the cortex. Let these window averages be $y_i^{(T_1,p)}$, $y_i^{(F,p)}$, $y_i^{(T_2,p)}$, respectively for each modality. The final feature vector is created by concatenating $[\mathbf{x}_i^{(T_1,p)}, \mathbf{x}_i^{(F,p)}, \mathbf{x}_i^{(T_2,p)}, y_i^{(T_1,p)}, y_i^{(F,p)}, y_i^{(T_2,p)}] = \mathbf{X}_i$. The corresponding voxel i in the manually labeled training image $\mathcal{J}_p^{(L)}$ has a $u \times v \times w$ -sized local patch surrounding it. We denote it by $\mathbf{z}_i$. For a given training feature vector $\mathbf{X}_i$, we have a corresponding dependent label vector $\mathbf{z}_i$. The label vector is a binary vector as the classification is lesions vs normal tissue, but it can easily be used for a multi-class classification by assigning additional labels to the different tissues. $[\mathbf{X}_i; \mathbf{z}_i]$ are sampled from all the training samples to create the training set.

3.3 Learning Multi-output Decision Trees

Learning a multi-output decision tree is similar to the CART algorithm²¹ to learn a single output regression tree. Let $\Theta_q = \{[\mathbf{X}_1; \mathbf{z}_1], \dots, [\mathbf{X}_m; \mathbf{z}_m]\}$ be the set of all the training sample pairs at a node q in the tree. The output vectors $\mathbf{z}_i$ are also known dependent vectors. Let dimensionality of the dependent vectors be d and the node mean at node q be $\bar{\mathbf{z}}_q$. The squared distance (SD) from the mean for the dependent vectors at node q is,

$$SD_q = \sum_{k=1}^m \sum_{j=1}^d (\mathbf{z}_{ij} - \bar{\mathbf{z}}_{qj})^2. \quad (1)$$

If a particular feature, f , and a threshold for the feature, τ_f are chosen, then the data in q , Θ_q is separated into two disjoint subsets, $\Theta_{qL} = \{[\mathbf{X}_i; \mathbf{z}_i] | \forall i, \mathbf{X}_{if} < \tau_f\}$ and $\Theta_{qR} = \{[\mathbf{X}_i; \mathbf{z}_i] | \forall i, \mathbf{X}_{if} > \tau_f\}$. f and τ_f are chosen such that the combined SD of the two daughter nodes q_L and q_R of q , is minimized.

The splitting is done recursively until there are a minimum number of samples left in a node, which is then declared as a leaf node of the tree. In our experiments, we have used fifty as the minimum number of leaf samples. Since a single tree is considered as a weak learner, we train multiple (five, in our experiments) trees using bootstrap aggregation (bagging) to create an improved classifier.

3.4 Prediction

Given a test image set consisting of processed T_1 -w, T_2 -w and FLAIR images, we extract the aforementioned features at each of the voxels. Next we apply the trained multi-output tree ensemble to each extracted feature vector. The input vector travels through the tree as its features are evaluated against the ones in the tree nodes, until it lands in a leaf node. The leaf node consists of at least fifty training samples, each of dimensionality d , with a label at each dimension. This provides us with $50d$ voxel labels, from which we determine the percentage of lesion voxels. This gives a membership estimate of a voxel belonging to the lesion class vs the normal tissue class.

3.5 Postprocessing

The membership image acquired from the multi-output decision ensemble is smoothed using a Gaussian filter with $\sigma = 1$. The smoothed membership image is thresholded to create a binary lesion mask. The threshold was empirically set to 0.33 to ensure a balance between true positive rate (TPR) and positive predictive value (PPV). Since a large majority of the MS lesions occurs inside white matter, we mask out false positives detected in other tissues. We use a simple three class fuzzy k-means segmentation²² of the T_1 -w image to roughly identify a WM mask. Lesions, being hypointense in T_1 -w images, appear as holes in this WM mask, which are filled.

4. RESULTS

4.1 MS Lesion Segmentation Challenge

For our first experiment we used the MSLSC Training Dataset.¹⁶ We compared the results with the winner of MSLSC⁵ and one of the current top ranked results,⁷ published after the challenge. We report some of the challenge metrics and compare the estimated lesion volume with the true lesion volume. MSLSC consists of twenty subjects, ten each from the University of North Carolina (UNC) center and the Children's Hospital Boston (CHB), with T_1 -w, FLAIR, T_2 -w, and manual lesion segmentations. We randomly selected four subjects as training data and evaluated on the remainder. We also performed a leave-one-out classification of the four training data images to include them in the analysis.

Our algorithm performs generally better than the challenge winner.⁵ The results are in Table 1. In comparison to Geremia et al., our algorithm has a better TPR-PPV balance for the CHB data and performs worse for the UNC data. The large errors in the UNC data are of the type shown in the second row of Fig. 1, where the manual segmentation is debatable. Fig. 2(a) shows a scatter plot of the manual lesion volumes vs. those calculated by the algorithm. In an ideal scenario, the line fit on the scatter plot should match the identity line (black dashed line). Our algorithm slightly overestimates the lesion volume for the low lesion images, while underestimating for the high lesion volume images.

4.2 In-house MS Data Segmentation

We also used an internal dataset of MS patients that was annotated by an expert rater. Our In-house MS data consists of MS patients consisted of forty nine subjects with T_1 -w, P_D -w, T_2 -w and FLAIR images. T_1 -w images were acquired using the Magnetization Prepared

Gradient Echo (MPRAGE) sequence and were acquired at a $0.82 \times 0.82 \times 1.1 \text{ mm}^3$ resolution. Whereas the other scans were acquired at either $0.82 \times 0.82 \times 2.2 \text{ mm}^3$ for 36 subjects and at $0.82 \times 0.82 \times 4.4 \text{ mm}^3$ for 13 subjects. Thus this dataset has considerable internal variability.

All the images were preprocessed as before in the MNI space. We again used four random subjects as our training data. We have reported the average TPR and PPV value over all the subjects in Table 1 and the algorithm lesion volumes vs. manual lesion volumes in Fig 2(b). We can observe that our algorithm slightly overestimates the lesion volume for the low lesion load images, and underestimates for the high lesion load images. Results on two subjects with high and low lesion loads respectively are shown in Fig 3.

5. CONCLUSION AND FUTURE WORK

We have described a multi-output decision tree ensemble framework for MS lesion segmentation. Our approach produces results, which are comparable to the state-of-the-art. It is computationally very fast, taking about 1–2 minutes for a $181 \times 217 \times 181$ -sized image. We have evaluated our algorithm on the training dataset of the MS lesion segmentation challenge and also on our in-house collection of MS images. Currently the splitting criterion considers a squared distance from the mean for the dependent vectors $\mathbf{z}_i$ located at a node. However this criterion is not ideal as the $\mathbf{z}_i$ consist of discrete labels. We would like to improve our approach by considering a different metric, which is more akin to the Gini index developed for a typical single output decision tree for the purpose of classification. We would also like to explore more features in order to reduce the amount of false positives.

We have submitted the MS Challenge private test data for the MSLSC to the testing website and have not received the evaluation results by the time of submission of the final manuscript. These results will be made available on <http://www.iacel.ece.jhu.edu/> when we receive them.

Acknowledgments

This work was funded by NIH R01-NS070906, NIH/NIBIB 1R01EB017743 and NIH/NIBIB 1R21EB012765.

References

1. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, Fujihara K, Havrdova E, Hutchinson M, Kappos L, Lublin FD, Montalban X, O'Connor P, Sandberg-Wollheim M, Thompson AJ, Waubant E, Weinshenker B, Wolinsky JS. Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Annals of Neurology*. 2011; 69(2):292–302. [PubMed: 21387374]
2. Rovira A, Swanton J, Tintor M. A single, early magnetic resonance imaging study in the diagnosis of multiple sclerosis. *Archives of Neurology*. 2009; 66(5):587–592. [PubMed: 19433658]
3. Stuckey S, Goh T-D, Heffernan T, Rowan D. Hyperintensity in the Subarachnoid Space on FLAIR MRI. *American Journal of Roentgenology*. 2007; 189(4):913–921. [PubMed: 17885065]
4. Simon J, Li D, Traboulsee A, Coyle P, Arnold D, Barkhof F, Frank J, Grossman R, Paty D, Radue E, Wolinsky J. Standardized MR Imaging Protocol for Multiple Sclerosis: Consortium of MS Centers Consensus Guidelines. *American Journal of Neuroradiology*. 2006; 27(2):455–461. [PubMed: 16484429]

5. Souplet J, Lebrun C, Ayache N, Malandain G. An Automatic Segmentation of T2-FLAIR Multiple Sclerosis Lesions. 2008
6. Shiee N, Bazin PL, Ozturk A, Reich DS, Calabresi PA, Pham DL. A Topology-Preserving Approach to the Segmentation of Brain Images with Multiple Sclerosis Lesions. *NeuroImage*. 2010; 49(2): 1524–1535. [PubMed: 19766196]
7. Geremia E, Clatz O, Menze BH, Konukoglu E, Criminisi A, Ayache N. Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. *NeuroImage*. 2011; 57(2): 378–390. [PubMed: 21497655]
8. Roy S, He Q, Carass A, Jog A, Cuzzocreo JL, Reich DS, Prince J, Pham D. Example based lesion segmentation. *Proc SPIE*. 2014; 9034:90341Y–90341Y-8.
9. Roy S, Carass A, Shiee N, Pham D, Prince J. Mr contrast synthesis for lesion segmentation. *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*. Apr.2010 : 932–935.
10. Sweeney EM, Shinohara RT, Shiee N, Mateen FJ, Chudgar AA, Cuzzocreo JL, Calabresi PA, Pham DL, Reich DS, Crainiceanu CM. OASIS is Automated Statistical Inference for Segmentation, with applications to multiple sclerosis lesion segmentation in MRI. *NeuroImage: Clinical*. 2013; 2(0): 402–413. [PubMed: 24179794]
11. Shiee N, Bazin PL, Cuzzocreo JL, Ye C, Kishore B, Carass A, Calabresi PA, Reich DS, Prince JL, Pham DL. Robust Reconstruction of the Human Brain Cortex in the Presence of the WM Lesions: Method and Validation. *Human Brain Mapping*. 2014; 35(7):3385–3401. [PubMed: 24382742]
12. Lecoer J, Ferré J, Barillot C. Optimized Supervised Segmentation of MS Lesions from Multi-spectral MRIs. *MICCAI workshop on Med Image Anal on Multiple Sclerosis*. 2009:1–11.
13. Forbes F, Doyle S, García-Lorenzo D, Barillot C, Dojat M. Adaptive weighted fusion of multiple MR sequences for brain lesion segmentation. *7th International Symposium on Biomedical Imaging (ISBI 2010)*. 2010:69–72.
14. Van Leemput K, Maes F, Vandermeulen D, Colchester A, Suetens P. Automated segmentation of multiple sclerosis lesions by model outlier detection. *Medical Imaging, IEEE Transactions on*. Aug.2001 20:677–688.
15. Llado X, Oliver A, Cabezas M, Freixenet J, Vilanova JC, Quiles A, Valls L, Rami-Torrent L, Rovira lex. Segmentation of multiple sclerosis lesions in brain mri: A review of automated approaches. *Information Sciences*. 2012; 186(1):164–185.
16. Styner M, Lee J, Chin B, Chin M, Commowick O, Tran H, Markovic-Plese S, Jewells V, Warfield S. 3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation. 2008
17. Geurts P, Touleimat N, Dutreix M, d'Alche Buc F. Inferring biological networks with output kernel trees. *Bioinformatics, BMC*. May.2007 8
18. Tustison N, Avants B, Cook P, Zheng Y, Egan A, Yushkevich P, Gee J. N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imag*. Jun.2010 29:1310–1320.
19. Collins DL, Holmes CJ, Peters TM, Evans AC. Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*. 1995; 3(3):190–208.
20. Carass A, Cuzzocreo J, Wheeler MB, Bazin PL, Resnick SM, Prince JL. Simple Paradigm for Extra-cerebral Tissue Removal: Algorithm and Analysis. *NeuroImage*. 2011; 56(4):1982–1992. [PubMed: 21458576]
21. Breiman, L., et al. *Classification and Regression Trees*. Wadsworth Publishing Company; U.S.A: 1984.
22. Bezdek JC. A Convergence Theorem for the Fuzzy ISO-DATA Clustering Algorithms. *IEEE Trans on Pattern Anal Machine Intell*. 1980; 20(1):1–8.

Figure 1.

The first row shows an algorithm result with a high TPR and PPV from the CHB dataset.

The second row shows a low TPR and PPV result from the UNC dataset.

Figure 2.

(a) Volume scatter plot and linear fit for the MSLSC data and (b) our in-house MS dataset.

Figure 3.

In-house MS dataset images with corresponding segmentation results. The top row shows results on a subject with a high lesion load, while the bottom row shows the results of subject with a low lesion load.

Average TPR and PPV values across 10 subjects for each of the UNC and CHB datasets, reported for 3 methods. We also evaluated our algorithm on an In-House dataset of 49 subjects.

Table 1

Dataset	Souplet et al. ⁵		Geremia et al. ⁷		Our Algorithm	
	TPR	PPV	TPR	PPV	TPR	PPV
CHB	0.138	0.364	0.350	0.552	0.400	0.400
UNC	0.257	0.236	0.447	0.255	0.361	0.194
In-House MS Dataset					0.407	0.358