MS-LESION SEGMENTATION IN MRI WITH RANDOM FORESTS

Oskar Maier[⋆†] *Heinz Handels*[⋆]

[⋆] Institute for Medical Informatics, Universität zu Lübeck
[†]Graduate School for Computing in Medicine and Life Sciences, Universität zu Lübeck

ABSTRACT

Multiple sclerosis (MS) is a common autoimmune disorder, whose diagnosis and study often relies on the extraction of biomarkers from magnetic resonance imaging (MRI) scans. Manual segmentation of MS lesions suffers from large intra- and inter-rater differences, whereas automatic methods promise reproducibility and enhanced consistency, especially for tracking the disease progress over time. To test this claim, the ISBI 2015 Longitudinal MS Lesion Segmentation Challenge provides a platform to compare existing methods in a fair and consistent manner to each other and the manual approach. In this article, we present our challenge contribution, which is based on random forests and local context intensity features to segment MS lesions in multi-spectral MRI images.

1. INTRODUCTION

Multiple sclerosis (MS), the most common autoimmune disorder affecting the central nervous system, is an inflammatory and degenerative disease with pathology that can be observed in vivo through magnetic resonance imaging (MRI). Areas of demyelination (lesions) of characteristic form and distribution are primarily visible in white matter (WM) on conventional MRI. Numerous biomarkers have been proposed for the study of MS[1], some of which, such as lesion volume and whole brain volume, are widely used to monitor disease progress. Essential prerequisite is a reliable, reproducible and consistent lesion segmentation to track lesion evolution over time. For the *ISBI 2015 Longitudinal MS Lesion Segmentation Challenge*[1], we propose a method for MS lesion segmentation through voxel-wise classification with random forests (RF), which we have previously applied successfully to stroke lesion segmentation in multi-spectral MRI [2]. For simplicity, we neglect the longitudinal correspondence i.e. all timepoints are treated equally.

2. METHOD

2.1. The classifier

A RF (scikit-learn implementation [3]) is trained with supervised learning to infer the classification function underlying the training data. The classification of brain lesions in MRI is a complex task with high levels of noise[2], hence we train a total of 200 trees without any growth-restriction (e.g. limited depth). Contrary to observations reported, no overfitting occurred.

2.2. The features

From each MRI sequence, a number of intensity-based features are extracted: 1. voxel intensity value, 2. voxel intensity value after smoothing (Gaussian filter at $\sigma = 3, 5, 7mm$) and 3. three different local histogram configurations. These features supply information about gray-values at different scales as well as mean intensity distributions in small areas around each voxel. To provide the classifier with a rough estimation of spatial location, we additionally compute the center-distance, i.e. each voxels distance to the image center. See [2] for more details on these features. Since MS lesions appear primarily in WM, a probability based tissue segmentation is obtained with FSL-Fast [4] on the MPRAGE/T1w sequence, separating the brain tissue into WM, gray-matter (GM) and cerebral spinal fluid (CSF). From the resulting tissue probability maps we extract 1. voxel gray-value and 2. voxel gray value after smoothing (Gaussian filter at $\sigma = 3, 7, 15, 31mm$). Each voxel is thus described by a 161 element feature vector.

2.3. Training set sub-sampling

Stratified random sampling is employed to extract a representative sub-set from the training data, reducing the amount of training samples and thus training time. The original background-to-lesion ratio of each case is kept intact, leading to an unequal class representation, which we have found advantageous [2].

2.4. Pre-processing

Although provided already pre-processed, the training cases of the challenge display high intensity differences, a normal occurrence for MRI, where intensity ranges are not standardized. With a learning based intensity standardization method implemented in MedPy [5] we harmonize each sequences intensity profile.

[1] http://iacl.ece.jhu.edu/MSChallenge

2.5. Post-processing

To obtain a binary segmentation mask, the RFs probability output is thresholded at a value of 0.4, introducing a slight bias in favour of the lesion class to compensate the training sets unbalanced class ratio. Finally, single unconnected lesion voxels are removed as outliers, holes in binary lesion objects closed and a single-iteration closing operation with a 3D square-connected component is applied.

3. EVALUATION AND RESULTS

The challenges training data consists of multi-spectral (FLAIR, MPRAGE, T2w, PD) scans of 5 patients with $4-5$ time-points each (21 cases in total), with two sets of expert segmentations each (GT1 & GT2). Since the organizers have revealed neither the evaluation scenario nor the evaluation measures to date, we evaluate each time-point independently and borrow the measures from the MICCAI 2008 MS challenge[6], namely Dice's Coefficient (DC), lesion true positive rate (lTPR), lesion false positive rate (lFPR) and lesion average symmetric surface distance (lASSD). With these, we evaluate a number of scenarios highlighting different aspects of our method:

I. The inter-rater performance, (1) with GT1 as segmentation and GT2 as ground-truth and (2) reversed, reveals the manual approaches performance.

II. Leave-one-patient out cross-validation, where all time-points of one patient are evaluated with a RF trained on the remaining data. A total of 1 million training samples is extracted as described above. With four train→test scenarios: (1) GT1→GT1, (2) GT2→GT1, (3) G13→GT2, (4) GT2→GT2, the performance of our method is assessed.

III. The test data RF is trained on all cases with 2 million samples of each GT1 and GT2. Lacking the test set ground truth, we can only provide evaluation results from the training data, which gives our methods upper limit, as the test sets were used during training.

IV. Competitor For a weak comparison, we provided the results of the MICCAI 2008 MS challenge winner[7].

4. DISCUSSION AND CONCLUSION

The low inter-rater results reinforce the need for automatic MS lesion segmentation and scrutinise all conclusion drawn based on manual segmentations. Our method performs near inter-rater quality, dominating in some measures while succumbing in others. Remarkable is the independence of the training data sets used. Our approaches' upper limits reveals some room for improvement, e.g. through parameter optimization, but also caps the maximum performance to a level only slightly above inter-rater quality. Compared against the MICCAI 2008 competitors, we excel in all three measures. Given the different data sets, this finding is not conclusive, but raises hope for a favourable rank in this years ISBI 2015 Longitudinal MS Lesion Segmentation Challenge. The proposed method achieved good results on the training data. How it will stand against other state-of-the-art approaches remains to be seen when the testing set results are revealed. For the future, we plan to incorporate knowledge about longitudinal correspondence into our method, e.g. through semi-supervised RF classification.

	GT	DC [0,1]	lASSD (mm)	TPR (%)	FPR (%)
I	1	.73	6.9	74	52
	2	.73	6.9	48	26
II	1	.70	4.2	53	48
	2	.70	4.6	55	48
	3	.65	3.7	37	44
	4	.65	3.0	38	43
III	1	.82	5.8	72	46
	2	.80	7.1	58	33
IV	UNC		6.6	40	61
	CHB		6.7	47	51

Table 1. The inter-rater (I) and our methods (II) results, our methods upper limit (III) and the MICCAI 2008 MS challenge winner (IV).

5. REFERENCES

[1] Serafeim Katsavos and Maria Anagnostouli, "Biomarkers in multiple sclerosis: An up-to-date overview," *Multiple Sclerosis International*, 2013.

[2] Oskar Maier, Matthias Wilms, et al., "Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences," *Journal of Neuroscience Methods*, vol. 240, no. 0, pp. 89–100, 2015.

[3] F. Pedregosa, G. Varoquaux, et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[4] Yongyue Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm," *Medical Imaging, IEEE Transactions on*, vol. 20, no. 1, pp. 45–57, Jan 2001.

[5] Oskar Maier, "MedPy," `https://pypi.python.org/pypi/MedPy`, Accessed: 2015-03-29.

[6] M. Styner, J. Lee, et al., "3D segmentation in the clinic: A grand challenge II: MS lesion segmentation," *Midas Journal*, 11 2008.

[7] S. Bricq, S. Collet, and J. Armspach, "MS lesion segmentation based on hidden markov chains," *Midas Journal*, 07 2008.