Longitudinal automated detection of white-matter and cortical lesions in relapsing-remitting multiple sclerosis

Fartaria M. J.[1,2], Bonnier G.[1,2], Kober T[1,2,3], Roche A.[2,1,3], Maréchal B. [1,2,3], Rotzinger D.[2], Schluep M.[4], Du Pasquier R.[5], Thiran J.P.[3,2], Krueger G.[2,3,5], Meuli R.[2], Bach Cuadra M*.[2,3,6], Granziera C*. [1,4,7]

1. Advanced Clinical Imaging Technology (HC CMEA SUI DI BM PI), Siemens Healthcare AG, Lausanne, Switzerland.
2. Department of Radiology, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland.
3. Signal Processing Laboratory (LTS 5), Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
4. Neuroimmunology Unit, Neurology, Department of Clinical Neurosciences, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland
5. Siemens Medical Solutions USA, Inc., Boston, US
6. Signal Processing Core, Centre d'Imagerie BioMédicale (CIBM), Lausanne, Switzerland.
7. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, US

Synopsis: Magnetic Resonance Imaging (MRI) plays an important role for lesion assessment in early stages of Multiple Sclerosis (MS). This work aims at evaluating the performance of an automated tool for MS lesion detection, segmentation and tracking in longitudinal data, only for use in this research study. The method was tested with images acquired using both a "clinical" and an "advanced" imaging protocol for comparison. The validation was conducted in a cohort of thirty-two early MS patients through a ground truth obtained from manual segmentations by a neurologist and a radiologist. The use of the "advanced protocol" significantly improves lesion detection and classification in longitudinal analyses.

Introduction: Magnetic Resonance Imaging (MRI) plays a major role in Multiple Sclerosis (MS) diagnosis, follow-up and therapy monitoring. More specifically, the identification of new or resolved lesions and changes in lesion size due to the progression of inflammation and demyelination is important to perform early diagnosis[1], quantify ongoing disease activity and monitor treatment effects[2,3]. In this work, we assess the performance of an in-house automated tool[4], only for use in this research study, to detect and segment longitudinal changes in MS lesions based on a clinical and an "advanced" MRI protocol[5].

Material and Methods: 3T MR images were acquired on a MAGNETOM Trio a Tim system (Siemens Healthcare, Germany) using a 32-channel head coil. Thirty-two patients with relapsing-remitting MS and disease duration <5 years from diagnosis were enrolled in the study, and two MRIs were performed at enrolment (TP1) and at two-years (21.4 ± 2.5 months, range 16-27 months) follow-up (TP2). The patient cohort consisted of 13 males and 19 females, age range 20-60 years at TP1, with a median Expand Disability Status Scale (EDSS) of 1.5 at both time points. The MRI protocol included:

- Magnetization-Prepared Rapid Acquisition Gradient Echo (MPRAGE, TR/TI=2300/900ms, voxel size (vs)=1.0x1.0x1.2mm^3);
- Magnetization-Prepared 2 Rapid Acquisitions Gradient Echo (MP2RAGE, TR/TI1/TI2=5000/700/2500 ms, vs=1.0x1.0x1.2mm^3);
- 3D FLuid-Attenuated Inversion Recovery (FLAIR, TR/TE/TI=5000/394/1800, vs=1.0x1.0x1.2mm^3);
- 3D Double Inversion Recovery (DIR, TR/TE/TI1/TI2=10000/218/450/3650, vs=1.1x1.0x1.2mm^3).

At both time points, a supervised classifier based on the k-nearest-neighbour (k-NN) algorithm was used to determine the lesion probability of each image voxel using the following features: 1) image intensity in each contrast, 2) spatial coordinates in a reference space[6], and 3) tissue prior probabilities[7]. Manual detection of MS lesions (from a neurologist and a radiologist) was used as ground truth (GT) for both time points as well as to train the classifier. Minimum lesion size was set at 0.009 mL[8]. Grey- and white-matter (GM, WM) lesions were classified in 5 groups with the following criteria[9]:

(i) new: identifiable on the registered TP2 image but not on the registered TP1 image;
(ii) enlarged: increase in diameter by at least 50%;
(iii) resolved: clearly visible on the registered TP1 image but not on the registered TP2 image;
(iv) shrunken: decrease in diameter by at least 50%;
(v) unaltered: do not follow any of the above criteria.

The performance was evaluated through a “leave-one-out” cross-validation. Detection rate (DR, number of detected lesions/total GT lesions) per brain for the different types of lesions (i-v) was computed using: (1) clinical images (FLAIR and MPRAGE) and (2) clinical and research images (FLAIR, MP2RAGE and DIR, “advanced protocol”). The confusion table of average detection rate per lesion type was computed. Lesion detection performance was compared between (1) and (2) using the Wilcoxon signed-rank test. Automated

estimation of total lesion volume difference between TP1 and TP2 per patient was evaluated through a Bland-Altman plot[10].

Results: Detection rate for all types of lesions improved significantly when the "advanced protocol" (DR=81.3%) was used instead of the “clinical protocol” (DR=78.7%, $p<0.01$, Figure 2). Statistical differences in DR were short of significance for particular lesion types, except for lesion type (iii) (*resolved lesions*, $p<0.05$). However, the best median detection rates in all lesion types were consistently obtained using the "advanced protocol" (Figure 2): (i) 62.5%, (ii) 100%, (iii) 66.7%, (iv) 100%, and (v) 89.4%. Misclassification of different types of lesions was also significantly improved for resolved and unaltered lesions using the "advanced protocol" ($p<0.05$, Figure 3). Except for three patients, volume difference quantification lies within ±1.96 standard deviations, indicating the good agreement between manual and automated segmentations using any type of protocol (Figure 4).

Discussion & Conclusion: The method exhibits good performance in detection of enlarged, resolved and unaltered lesions. New and resolved lesions present the lowest DR compared to the other lesion types , possibly due to lower size (strongly affected by partial volume) and low contrast (low degree of tissue inflammation/demyelination). However, in some cases, new/resolved lesions are detected as enlarged, shrunken or unaltered in both time points, which was retrospectively confirmed as correct by an expert. This highlights the difficulty in distinguishing focal lesions from diffuse damage, even by experts, and demonstrates the ability of the algorithm to detect lesions in both time points that were missed by manual segmentation. When the "advanced protocol" is used, the variability of lesion volume differences (TP1-TP2) between manual and automated segmentation is lower. This can be explained by the higher sensitivity in detection of cortical lesions and the better lesion delineation when advanced sequences are used, as shown in previous work[4].
Future work will aim to reduce partial-volume effects and to optimize the weighting of contrasts in order to increase the sensitivity of both GM and WM lesion detection.

References:

[1] - Bonnier, G. *et al*. Advanced MRI unravels the nature of tissue alterations in early multiple sclerosis. Annals of Clinical and Translational Neurology. *Ann. Clin. Transl. Neurol.*2014; **1:**423–432.

[2] - Rovira, A. *et al*. Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis clinical implementation in the diagnostic process. *Nat Rev Neurol*. 2015; advance online publication.

[3] - Bonnier, G. *et al*. Longitudinal analysis of advanced and conventional magnetic resonance imaging measures of disease impact in multiple sclerosis. *ISMRM*. 2014;

[4] - Fartaria, M.J. *et al*. Automated Detection of White Matter and Cortical Lesions in Early Stages of Multiple Sclerosis. JMRI. 2015;

[5] - Kober, T. *et al*. MP2RAGE multiple sclerosis magnetic resonance imaging at 3T. *Invest Radiol*. 2012;**47**(6):346-52.

[6] - Anbeek, P. *et al*. Probabilistic segmentation of white matter lesions in MR imaging. *NeuroImage*. 2004;**21**(3):1037-44.

[7] - Steenwijk, M.D. *et al*. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). *NeuroImage:Clin*. 2013;**3**:462-9.

[8] - Guizard, N. *et al***.** Rotation-invariant multi-contrast non-local means for MS lesion segmentation. *NeuroImage:Clin***.** 2015; **8**:376-389.

[9] - Moraal, B. *et al*. Improved Detection of Active Multiple Sclerosis Lesions: 3D Subtraction Imaging. Radiology. 2010. **255**(1):154-163.

[10] - Bland, J.M. *et al*. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*. 1986;**327**(8476):307-10.

Acknowledgements: This work was supported by the Swiss National Science Foundation under grant PZ00P3_131914/11; The Swiss MS Society and the Societé Académique Vaudoise, the CIBM of the University of Lausanne (UNIL), the Swiss Federal Institute of Technology Lausanne (EPFL), the University of Geneva (UniGe), the Centre Hospitalier Universitaire Vaudois (CHUV), the Hôpitaux Universitaires de Genève (HUG) and the Leenaards and the Jeantet Foundations.

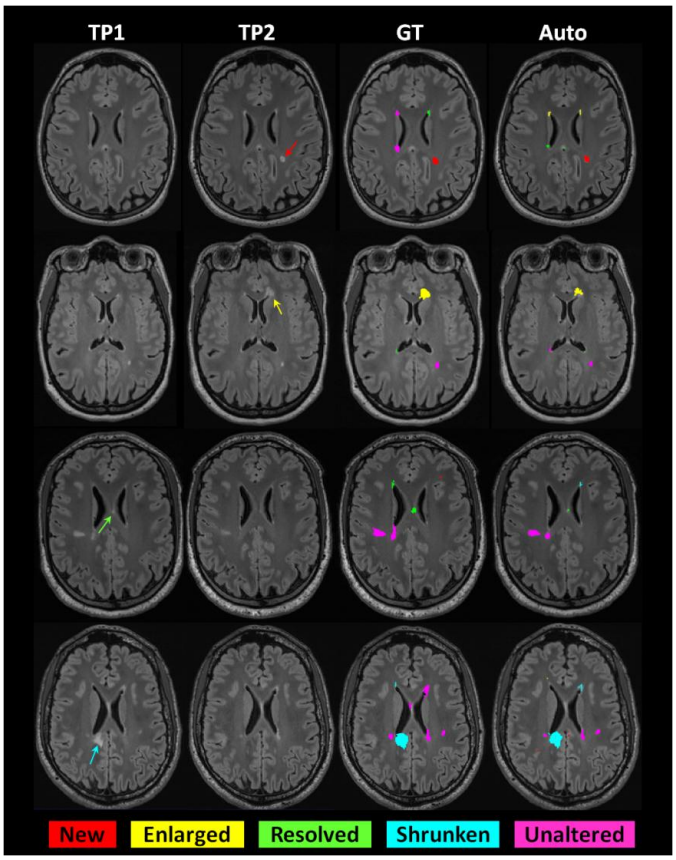

Figure 1 - Slices of FLAIR image from different patients at time point 1 (TP1) and respective time point 2 (TP2). Manual Segmentations (GT) and automated segmentations (AUTO) are shown for different types of lesions: new, enlarged, resolved, shrunken, and unaltered (red, yellow, green, blue and pink, respectively).

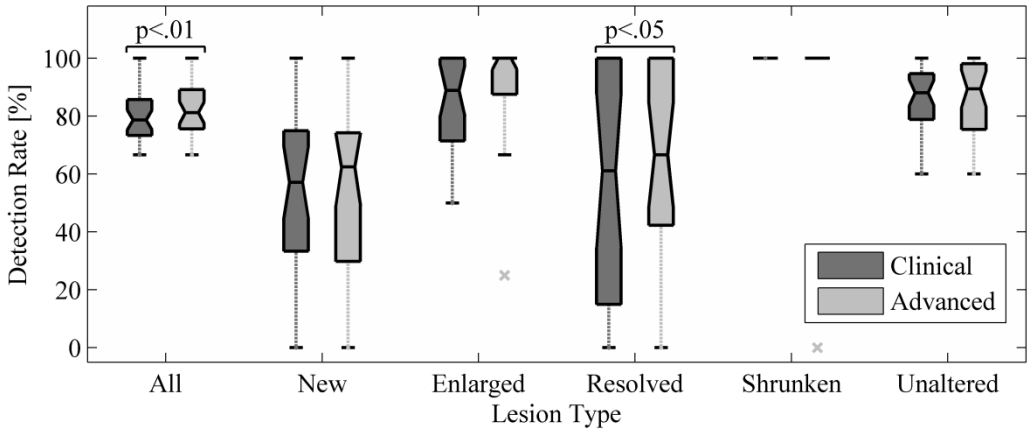

Figure 2 - Boxplots of detection rate from all patients across different types of lesions using the clinical (dark gray) and the advanced (light gray) protocols. The crosses in the plot represent outliers in the cohort.

Automated \ Manual	New	Enlarged	Resolved	Shrunken	Unaltered
New	38.0 39.5	6.0 7.7	0.0 0.5	0.0 3.0	3.2 2.3
Enlarged	10.6 11.2	6.8 9.9	35.0 23.1	6.3 6.8	7.8* 3.5
Resolved	12.8 5.0	14.8 14.7	20.7 34.3	12.5 0.0	13.8 9.7
Shrunken	4.8 7.1	5.3 5.4	9.4 2.9	24.3 10.6	5.0* 6.6
Unaltered	33.8 37.2	67.1 62.3	34.9 39.2	56.9 79.5	70.2* 77.8

Figure 3 - Confusion matrix showing the relation between the manual and automated classification of detected lesions. Values in each cell represent the percentage of average detection per lesion type across patients. Top value represents the results using a clinical protocol (FLAIR/MPRAGE) and the bottom value of the results using the advanced protocol (FLAIR/MP2RAGE/DIR). Significance of the results ($p<0.01$) are represented by the stars.

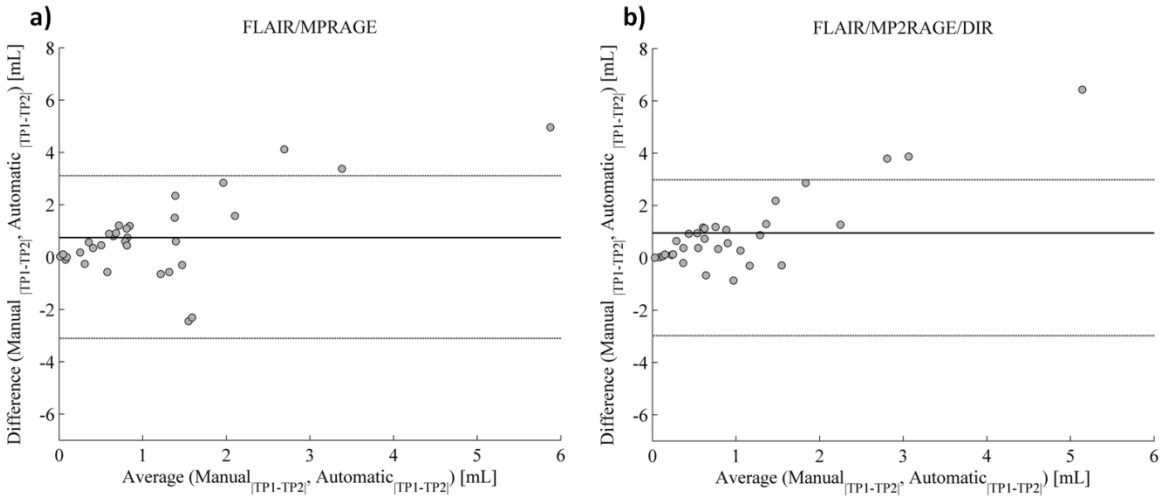

Figure 4 - Bland-Altman plot for total lesion difference volume TP1/TP2 between manual and automated segmentations. a) results using the clinical protocol (FLAIR/MPRAGE). b) results using the advanced protocol (FLAIR/MP2RAGE/DIR).