Check for updates

Editorial: Automatic methods for multiple sclerosis new lesions detection and segmentation

Olivier Commowick[1†], Benoît Combès[1], Frédéric Cervenansky[2] and Michel Dojat[3]*

[1]Empenn INSERM U1228, CNRS UMR6074, Inria, University of Rennes I, Rennes, France, [2]Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, Lyon, France, [3]Univ Grenoble Alpes, Inserm, U1216, CHU Grenoble Alpes, Grenoble Institut Neurosciences, GIN, Grenoble, France

Editorial on the Research Topic
Automatic methods for multiple sclerosis new lesions detection and segmentation

Multiple Sclerosis (MS) is a chronic inflammatory disease of the central nervous system (CNS) affecting more than half a million persons in Europe, with a prevalence rate of 83 per 100,000 with higher rates in northern countries and a female/male ratio around 2.0 (Pugliatti et al., 2006). Today, conventional MR imaging (MRI) is widely used for the patient follow-up, the monitoring of the therapy effects, and more generally in a perspective of personalized medicine, for the understanding of the individual MS progression (Thompson et al., 2018). One of the major challenges in using MRI for MS is the segmentation of lesions whose number, location and appearance at a given time point, are crucial indicators for diagnostic and to tailor treatment to the specific individual disease's evolution.

To cope with inter- and intra-observer variability and reduce the burden and complexity of lesions identification for clinicians, a large number of techniques have been proposed in the literature for the automatic segmentation of MS lesions (see Garcia-Lorenzo et al., 2013; Valverde et al., 2017; Danelakis et al., 2018 for reviews). Several challenges have been proposed to evaluate the performances of these methods (e.g., Carass et al., 2017; Commowick et al., 2021 to cite the most recent ones). Moreover, recently Bonacchi et al. (2022) proposed an overview of Artificial Intelligence applications for MS clinical practice.

A growing literature focuses on the delineation of new MS lesions on T2/FLAIR occurring between two consecutive exams. Detecting the apparition of new MS lesions is of central interest in clinical practice. Indeed, while the palette of Disease Modifying Drugs (DMDs) approved for MS has presently an unknown impact on the compartmentalized neurodegenerative process within the CNS, they aim to substantially reduce, or even stop, the accumulation of new lesions. Consequently, the assessment of such an accumulation allows the clinician to monitor the efficiency of a given DMD on each patient it follows, and therefore to consider a change of treatment in case of insufficient efficiency. Moreover, there is a direct link between accumulation of new lesions and increasing handicap (Sormani et al., 2013). Automating the detection of these new lesions or helping clinicians to identify them would therefore be a major advance for evaluating the patient disease progression and response to treatment.

In 2021, we launched a MICCAI challenge, MSSEG-II (see https://www.ofsep.org/fr/etudes/msseg-ii-challenge-miccai-2021), to compare automated solutions for this specific task i.e., the detection of new lesions appearing at the second time point of two T2/FLAIR images of the patient. For that purpose, we used a large database: 100 patients, each with two time points, the time between the two time points varying between 1 and 3 years. Data were extracted from the national OFSEP cohort (Vukusic et al., 2020), the national French MS registry (https://clinicaltrials.gov/ct2/show/NCT03603457), with 3D FLAIR images from different centers and scanners (15 different scanners in total) using the OFSEP specific protocol (Cotton et al., 2015; Brisset et al., 2020). Only 3D FLAIR images—that is the mostly used clinical sequence for MS brain—were considered. As in our previous challenge (Commowick et al., 2021), the evaluation of solutions was performed on the dedicated FLI-IAM infrastructure (https://www.francelifeimaging.fr/en/about/noeuds/iam/), which comprises Shanoir, a web-oriented solution for imaging data storage and sharing for preclinical and clinical research studies (Barillot et al., 2016; Kain et al., 2020); and the VIP platform (Glatard et al., 2013) for the execution of the corresponding docker of each image processing algorithm/pipeline on EGI infrastructures (https://www.egi.eu/). The use of FLI-IAM allows to automate the competition's process through a sustainable framework and remove the potential biases (e.g., challengers manually optimizing their parameters for each provided case). The ground truth was defined based on the manual delineation, using ITK Snap, of the 100 cases by four neuroradiologists with an MS expertise. Then, a consensus was formed in two steps: a senior expert neuroradiologist examined and confirmed (or declined) disputed lesions among the experts; then a fusion using the STAPLE (Warfield et al., 2004) algorithm was performed. This consensus was then the reference for the evaluation procedure. Forty cases were provided to challengers (e.g., for algorithm training) and 60 cases for algorithm testing. The manual segmentations were provided with the former and unknown to the challengers for the latter.

The present RT gathers 10 papers about solutions for the automatized detection of new lesions in MS subsequent images. All but one (Dufresne et al.) competed during MSSEG-II challenge and were executed on FLI-IAM infrastructure. They are based on a deep learning approach, the U-net architecture (Ronneberger et al., 2015) with its 2D or 3D versions. We may distinguish two classes of approaches, ones that use exclusively the examples provided by the Miccai challenge organizers and those which introduce additional real (Hitziger et al.) or synthetic (Andresen et al.; Kamraoui et al.; Valencia et al.) datasets. Finally, joint modeling, mixing both a registration and a segmentation task, have been investigated (Andresen et al.; Dufresne et al.; Salem et al.).

Then, Hitziger et al. train a 2D U-net with residual units with axial, coronal and sagittal slices. The corresponding slices from the two time-point volumes are paired and introduced to the system as a two-channel input. The predictions from each orientation are then merged with different strategies. The best performances are obtained for the unanimous voting strategy where lesions are confirmed in each orientation. The gain in performance by introducing additional datasets (25 supplementary patients to the initial 40 patients training set) seems weak.

In the same line, Sarica and Seker propose a 2D U-net solution where the standard plain blocks are replaced by residual units and attention gates are introduced to, respectively, enhance the model performances and focalize on new MS lesions on each 2D slice. A majority voting generates the final 3D binary output.

Similarly, Ashtari et al. introduce residual units, this time in a 3D U-net version and data augmentation methods to improve robustness and generalizability of the obtained model.

Basaran et al. consider the recent 3D U-Net version ("No-NewU-Net") combined with several image preprocessing step brain extraction, bias correction, registration and multiple data augmentation methods.

To overcome the difficulty of a supervised training based on scarce new lesion annotated examples, Kamraoui et al. interestingly propose to first pretrain a 3D U-Net on a large one time-point MS dataset (transfer learning), second to pretrain the model used for time-points by introducing realistic synthetic data, and finally to fine-tune the obtained network with the real two time-points data as provided by MSSEG-II.

To tackle class imbalance between voxels belonging to new lesions or not, Schmidt-Mengin et al. introduce a two-stage training strategy to iteratively define a fixed number of patches (30%) containing lesions. This "online hard example mining" strategy is implemented with two 3D U-Nets applied patch-wise in cascade. Such a strategy, applied for the first time on 3D brain scans, seems to emphasize false positive rate.

Instead of using a unique intensity-based approach, Andresen et al., Salem et al., and Dufresne et al. propose to consider a deformation-based approach. Maps of non-corresponding regions between subsequent images are generated during the registration process. In Andresen et al. such maps are then used by a fully convolutional network to segment new lesions that occur across time. Offset maps with baseline allow exploring morphology appearance of new lesions. New lesions are rare and similarly to the previous paper (Kamraoui et al.) the authors insert synthetic lesions during the network training. In Salem et al. the authors introduce a cascade of two 3D U-net patch-wise fully convolutional neural networks. The first registration network learns the deformation field to register the individual sequence of FLAIR images, while the second performs new lesions segmentation. The latter is fed by registered FLAIR images and the deformation maps. Indeed, the first network allows to filter the majority of non-lesion voxels and reveals the possible new lesion candidates, while the second refine the detection in reducing misclassified voxels. The simultaneous training of registration and segmentation modules improves the performances compared to a sequential learning. Valencia et al. propose to improve the previous results in adding synthetic images. The hypothesis is that the introduction of T1-weighted images (T1w), artificially generated, in addition to the FLAIR images improves new MS lesions detection. They use a generative adversarial network (GAN) with an additional MS FLAIR dataset (136 cases) in order to generate T1w corresponding images. The trained GAN is then used to generate the T1w corresponding to the provided MSSEG-II FLAIR images. They show an improvement of the sensitivity performance compared to the only use of FLAIR images.

TABLE 1 Averaged (patient-wise) score for the four experts.

	New lesion cases (n = 32)								No new lesion (n = 28)	
	DSC	F1	Sensitivity	Specificity	PPV	ηTP	ηFP	ηFN	ηFP	ηFP
Expert 1	0.629	0.709	0.650	1.000	0.707	6.063	1.281	1.094	0.036	1.453
Expert 2	0.597	0.601	0.526	1.000	0.813	4.500	0.844	2.375	0.000	0.000
Expert 3	0.535	0.637	0.580	1.000	0.760	4.313	1.094	2.500	0.107	3.981
Expert 4	0.459	0.519	0.407	1.000	0.801	4.469	0.594	2.375	0.036	0.623

DSC, dice score; PPV, positive predictive value; ηTP, mean number of true positives; ηFP, mean number of false positive; ηFN, mean number of false negative. The provided sensitivity, precision, and PPV are computed at the voxel scale.

Finally, in Dufresne et al., a different deformation-based approach is proposed where deformable registration and local intensity change detection are jointly estimated as a unified optimization problem solving. The joint method is evaluated on synthetic and real MS datasets and compared to the sequential version, where registration and change detection are performed successively, to demonstrate the performance improvement obtained by the former. Such an optimization approach cannot discriminate between new lesions from evolving lesions. It is interesting to note that this is the only non-Deep Learning-based method presented in this RT.

In Table 1, we provide several indexes for the readers in order to have a flavor of the current performances reached by the different solutions described in this RT compared to human experts.

To conclude, MS new lesions detection and segmentation remain very difficult tasks. Presently, automatic methods can be more sensitive for detecting new lesions, but produce more false positive compare to manual delineation by experts. Thus, in spite of slight persistent differences, performances between automatic solutions and human experts are closer than in the previous challenge (see Commowick et al., 2021). However, in order to be used in clinical routine, several steps need to be completed, such as the integration of computerized solutions in the hospital information flow and the quantification of the uncertainty associated to the automatic lesion detection, in place of the standard binary output, to leverage the clinician's work for obvious lesion and requiring his/her expertise only for difficult cases (Lambert et al., 2022). This will lead to the design of a new family of computerized medical assistants for care improvement.

Data from the MSSEG challenges are available here https://shanoir.irisa.fr/shanoir-ng/welcome and can be used to evaluate new solutions.

Author contributions

All authors listed have made a substantial, direct, intellectual contribution to the work. BC, FC, and MD approved it for publication.

Dedication

This editorial is dedicated to OC, our young and talented colleague who prematurely passed away in December 2022.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Barillot, C., Bannier, E., Commowick, O., Corouge, I., Baire, A., Fakhfakh, I., et al. (2016). Shanoir: applying the software as a service distribution model to manage brain imaging research repositories. *Front. ICT* 3, 25. doi: 10.3389/fict.2016.00025

Bonacchi, R., Filippi, M., and Rocca, M. A. (2022). Role of artificial intelligence in MS clinical practice. *Neuroimage Clin.* 35, 103065. doi: 10.1016/j.nicl.2022.103065

Brisset, J. C., Kremer, S., Hannoun, S., Bonneville, F., Durand-Dubief, F., Tourdias, T., et al. (2020). New OFSEP recommendations for MRI assessment of multiple sclerosis patients: special consideration for gadolinium deposition and frequent acquisitions. *J. Neuroradiol.* 47, 250–258. doi: 10.1016/j.neurad.2020.01.083

Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., et al. (2017). Longitudinal multiple sclerosis lesion segmentation data resource. *Data Brief* 12, 346–350. doi: 10.1016/j.dib.2017.04.004

Commowick, O., Kain, M., Casey, R., Ameli, R., Ferré, J. C., Kerbrat, A., et al. (2021). Multiple sclerosis lesions segmentation from multiple experts: the MICCAI 2016 challenge dataset. *Neuroimage* 244, 118589. doi: 10.1016/j.neuroimage.2021.118589

Cotton, F., Kremer, S., Hannoun, S., Vukusic, S., and Dousset, V. (2015). OFSEP, a nationwide cohort of people with multiple sclerosis: consensus minimal MRI protocol. *J. Neuroradiol.* 42, 133–140. doi: 10.1016/j.neurad.2014.12.001

Danelakis, A., Theoharis, T., and Verganelakis, D. A. (2018). Survey of automated multiple sclerosis lesion segmentation techniques on magnetic resonance imaging. *Comput. Med. Imaging Graph.* 70, 83–100. doi: 10.1016/j.compmedimag.2018.10.002

Garcia-Lorenzo, D., Francis, S., Narayanan, S., Arnold, D. L., and Collins, D. L. (2013). Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Med. Image Anal.* 17, 1–18. doi: 10.1016/j.media.2012.09.004

Glatard, T., Lartizien, C., Gibaud, B., da Silva, R. F., Forestier, G., Cervenansky, F., et al. (2013). A virtual imaging platform for multi-modality medical image simulation. *IEEE Trans. Med. Imaging* 32, 110–118. doi: 10.1109/TMI.2012.2220154

Kain, M., Bodin, M., Loury, S., Chi, Y., Louis, J., Simon, M., et al. (2020). Small Animal Shanoir (SAS): a cloud-based solution for managing preclinical MR brain imaging studies. *Front. Neuroinformat.* 14, 20. doi: 10.3389/fninf.2020.00020

Lambert, B., Forbes, F., Tucholka, A., Doyle, S., Dehaene, H., and Dojat, M. (2022). *Trustworthy Clinical AI Solutions: A Unified Review of Uncertainty Quantification in Deep-Learning Models for Medical Image Analysis*. Available online at: https://arxiv.org/abs/2210.03736 (accessed March 2023).

Pugliatti, M., Rosati, G., Carton, H., Riise, T., Drulovic, J., Vécsei, L., et al. (2006). The epidemiology of multiple sclerosis in Europe. *Eur. J. Neurol.* 13, 700–722. doi: 10.1111/j.1468-1331.2006.01342.x

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, eds N. Navab, J. Hornegger, W. Wells, and A. Frangi (Cham: Springer). doi: 10.1007/978-3-319-24574-4_28

Sormani, M. P., Rio, J., Tintorè, M., Signori, A., Li, D., Cornelisse, P., et al. (2013). Scoring treatment response in patients with relapsing multiple sclerosis. *Mult. Scler.* 19, 605–612. doi: 10.1177/1352458512460605

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., et al. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 17, 162–173. doi: 10.1016/S1474-4422(17)30470-2

Valverde, S., Cabezas, M., Roura, E., Gonzalez-Villa, S., Pareto, D., Vilanova, J. C., et al. (2017). Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *Neuroimage* 155, 159–168. doi: 10.1016/j.neuroimage.2017.04.034

Vukusic, S., Casey, R., Rollot, F., Brochet, B., Pelletier, J., Laplaud, D. A., et al. (2020). Observatoire Français de la Sclérose en Plaques (OFSEP): a unique multimodal nationwide MS registry in France. *Mult. Scler.* 26, 118–122. doi: 10.1177/1352458518815602

Warfield, S., Zou, K., and Wells, W. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23, 903–921. doi: 10.1109/TMI.2004.828354