FLAIR vs MPRAGE contribution to white matter lesion automatic segmentation in MS using localized saliency maps

Federico Spagnolo[a,b,*], Roger Schaer[b], Vincent Andrearczyk[b], Nataliia Molchanova[b,c,d], Mara Graziani[b], Henning Müller[b], Meritxell Bach Cuadra[c,d], Cristina Granziera[a,e], Adrien Depeursinge[b]

[a]Translational Imaging in Neurology (ThINk) Basel, Department of Biomedical Engineering, Faculty of Medicine, University Hospital Basel and University of Basel, Basel, Switzerland;
[b]Institute of Informatics, University of Applied Sciences and Arts Western Switzerland (HES-SO), Sierre, Switzerland;
[c]CIBM Center for Biomedical Imaging, Lausanne, Switzerland;
[d]Radiology Department, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland;
[e]Department of Neurology, University Hospital Basel, Switzerland, MS Center and Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), University Hospital Basel and University of Basel, Basel, Switzerland;
[*]Corresponding author. Email address: federico.spagnolo@unibas.ch

1 Introduction

One of the most important Multiple Sclerosis (MS) biomarkers is the hyper-intensity of white matter (WM) lesions reported on FLAIR magnetic resonance imaging (MRI). Despite many efforts to automate time-consuming lesion detection and segmentation with artificial intelligence (AI), their clinical adoption has been jeopardized by their black box nature [1]. To encourage AI integration into clinical practice, research in explainable AI (XAI) should at least provide lesion-wise evidence of models' decisions.

2 Methods

720 patients diagnosed with MS with a total of 4043 FLAIR and MPRAGE MRI scans (baseline, follow-ups and WM lesion masks annotated by three expert clinicians) were collected at the University Hospital of Basel, Switzerland. Data were randomly split into training, validation and test sets (containing respectively 585, 94 and 35 patients with 3369, 553 and 121 scans) to train a 3D U-Net network [5] for MS lesion segmentation, following [3] with a blob loss function [2]. We generated saliency maps for 5 test patients using SmoothGrad [4], injecting noise with std=0.05 to obtain 50 noisy versions. We developed a localized version where only voxels belonging to a specific lesion were cumulated. The maps were then normalized by the number of voxels within the lesion. In such a way, separate saliency maps were generated for each input modality, allowing to investigate their respective contributions. Then, we examined the distribution of values to check the contribution of each input, separating positive from negative gradients.

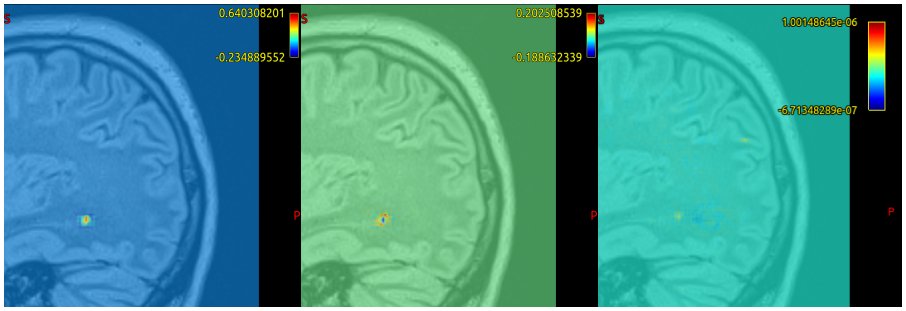
3 Results

Saliency maps generated with localized SmoothGrad for a true positive (TP) are shown in Fig 1a, 1b, while the output for the non-localized method is shown in Fig 1c. The saliency maps generated from an area with no lesions, Fig 2, consist of gradient values with orders of magnitude that are 10^5 and 10^6 times smaller than values obtained in TP cases. The distribution of gradient values for TPs and false negatives (FN) is shown in Fig 3.

4 Discussion

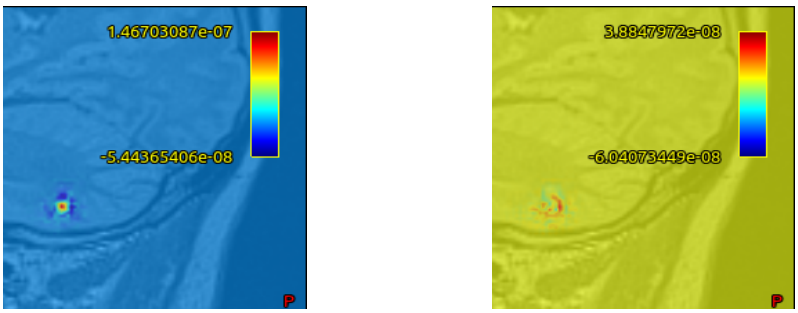
From preliminary results, gradients with respect to FLAIR appear positive inside the lesion and negative at its borders in both TP and FN cases. Similarly, gradients with respect to MPRAGE appear negative inside the lesion and positive at its borders. Both these findings could be explained by the fact that WM lesions appear as white in FLAIR and as black in MPRAGE. Thus, higher voxel values in FLAIR hyper-intense areas and lower voxel values in MPRAGE hypo-intense ones would reinforce the hypothesis that those areas contain lesions. Comparably, negative intensities around lesion edges in FLAIR or positive in MPRAGE would support the same concept. The distribution of gradient values shows significantly higher contribution when back-propagated to FLAIR compared to MPRAGE, and this is mainly due to positive gradients in FLAIR for both TP and FN cases. This resonates with the fact that WM lesions in MS are easily detected in FLAIR: expert clinicians usually rely on FLAIR to identify WM lesions, while using MPRAGE for structural or additional information. We conclude that the proposed localized saliency maps allow providing evidence supporting automatic segmentations.

Submitted to BIAS 2023



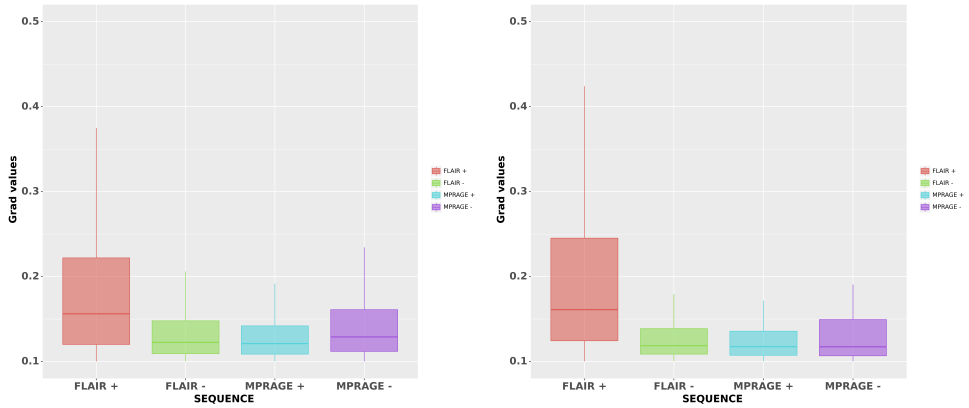
(a) (b) (c)

Figure 1: Saliency maps of a TP example w.r.t. FLAIR (a) and MPRAGE (b). Saliency map generated from a non-localized area in the prediction (c). Strong positive gradients are red, strong negatives are blue.



(a) (b)

Figure 2: Saliency map of a non-lesion area w.r.t. FLAIR (a) and MPRAGE (b). Strong positive gradients are red, strong negatives are blue.



(a) (b)

Figure 3: Boxplot representing the distribution of saliency maps values for TPs (a) and FNs (b). FLAIR + and MPRAGE + refer to positive gradient values while FLAIR - and MPRAGE - refer to negative gradient values.

References

[1] G. Baselli, M. Codari, and F. Sardanelli. Opening the black box of machine learning in radiology: can the proximity of annotated cases be a way? *European Radiology Experimental*, 4, 12 2020.

[2] F. Kofler, S. Shit, I. Ezhov, L. Fidon, I. Horvath, R. Al-Maskari, H. Li, H. Bhatia, T. Loehr, M. Piraud, A. Erturk, J. Kirschke, J. Peeken, T. Vercauteren, C. Zimmer, B. Wiestler, and B. Menze. Blob loss: instance imbalance aware loss functions for semantic segmentation. *arXiv*, 2022.

[3] A. Malinin, A. Athanasopoulos, M. Barakovic, M. B. Cuadra, M. J. F. Gales, C. Granziera, M. Graziani, N. Kartashev, K. Kyriakopoulos, P.-J. Lu, N. Molchanova, A. Nikitakis, V. Raina, F. La Rosa, E. Sivena, V. Tsarsitalidis, E. Tsompopoulou, and E. Volf. Shifts 2.0: Extending the dataset of real distributional shifts. *arXiv*, 2022.

[4] D. Smilkov, N. Thorat, B. Kim, F. Viégas, and M. Wattenberg. Smoothgrad: removing noise by adding noise. *CoRR*, 06 2017.

[5] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. *arXiv*, 2016.