MSMilan2023
9th Joint ECTRIMS-ACTRIMS Meeting
11–13 October 2023 | Milan, Italy

Abstract Number: [1801]

Abstract Title: The Normalised Dice Similarity Coefficient for Multiple Sclerosis: tackling lesion load biases in white matter and cortical lesion segmentation

Abstract Category: Imaging and non-imaging biomarkers - 32 - Big data and artificial intelligence

Preferred Presentation Type: Oral or poster presentation

Vatsal Raina[1], Nataliia Molchanova[2, 3, 4], Mara Graziani[3], Andrey Malinin[5], Po-Jui Lu[6], Matthias Weigel[6, 7, 8, 9], Henning Muller[3], Cristina Granziera[6, 7, 8], Meritxell Bach Cuadra*[2, 4], Mark Gales[1]

[1]University of Cambridge, ALTA Institute, Cambridge, United Kingdom, [2]Lausanne University and University Hospital, Radiology Department, Lausanne, Switzerland, [3]University of Applied Sciences of Western Switzerland (HES-SO), Institute of Informatics, Sierre, Switzerland, [4]CIBM Center for Biomedical Imaging, Lausanne, Switzerland, [5]Isomorphic Labs, London, United Kingdom, [6]University Hospital Basel and University of Basel, Translational Imaging in Neurology (ThINk) Basel, Department of Biomedical Engineering, Faculty of Medicine, Basel, Switzerland, [7]University Hospital Basel, Department of Neurology, Basel, Switzerland, [8]University Hospital Basel and University of Basel, Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), Basel, Switzerland, [9]University Hospital Basel, Division of Radiological Physics, Department of Radiology, Basel, Switzerland

Introduction:
Deep learning (DL) methods have become widely adopted for automated lesion segmentation in multiple sclerosis (MS). The Dice similarity coefficient (DSC) is a common metric for evaluating segmentation quality, comparing the agreement between the predicted segmentation against a ground-truth mask. However, the DSC metric has been shown to be positively correlated to the occurrence rate of the positive class ("lesion" class) in the ground-truth, jeopardizing the direct comparison of DSC values among subjects with different lesion loads.

Objectives/Aims:
The recently proposed normalised Dice Similarity Coefficient (nDSC) (Raina et al ISBI 2023) removes the bias in the DSC metric by scaling the precision based on the subject lesion load. Our aim is to extend the validation of nDSC within the context of white matter lesion (WML) and cortical lesion (CL) automatic segmentation.

Methods:
We explore separately the segmentation of WML (1) on 3D FLAIR and CL on 3D MP2RAGE (2). Datasets 1) 33, 7, and 74 patients for training, validation and testing from the multi-centric dataset of the Shifts Project (5 hospitals, 1.5T and 3T, 6 vendors) 2) 84, 10, 92 patients for training, validation, and testing from a multi-centric privative dataset (2 hospitals, 3T, 2 scanner models). For both tasks, a 3D U-net model is used with a optimization loss the combination of focal Dice and cross entropy. The median WM lesion load is 2.2 (IQR: 1.0 - 4.0) 10^2 mm^3 and for CL load is 2.9 (IQR: 0.8 - 7.7) 10^-1 mm^3 in test sets 1 and 2, respectively. To quantify the

difference between DSC and nDSC, we compute the Spearman's rank correlation coefficient, ρ, between the lesion load and the metrics, where a score of 1.0 indicates positive bias and 0.0 indicates no correlation.

Results:
The average DSC was 0.48 and 0.23 while nDSC was 0.57 and 0.17 on the WML and CL tasks, respectively. For the WML task the correlation of the lesion load with DSC is ρ=0.60, and ρ=0.37 with nDSC. For the CL task the correlation of the lesion load with DSC is ρ=0.34, and ρ=0.20 with nDSC. It is evident from the results that nDSC makes an attempt to disentangle the performance metric bias with the lesion load by reducing the coefficient closer to 0.

Conclusion:
Our results show reduced correlation of nDSC with lesion load as compared with DSC. We believe the residual correlation in nDSC is due to the loss function optimised during training that indeed penalises less errors made on small lesions. Further investigation should include exploration of nDSC as a metric for neural network training.

Disclosures: VR: nothing to disclose
NM: nothing to disclose.
AM: Work done while at Yandex
MW has received research funding by Biogen for developing spinal cord MRI.
PJL: nothing to disclose
MG: IBM Research zurich
HM: mandates with Roche
CG: The University Hospital Basel (USB), as the employer of C.G., has received the following fees which were used exclusively for research support: (i) advisory boards and consultancy fees from Actelion, Novartis, Genzyme-Sanofi, GeNeuro, Hoffmann La Roche and Siemens; (ii) speaker fees from Biogen, Hoffmann La Roche, Teva, Novartis, Merck, Jannsen Pharmaceuticals and Genzyme-Sanofi; (iii) research grants: Biogen, Genzyme Sanofi, Hoffmann La Roche, GeNeuro.
MBC: nothing to disclose.
MG: nothing to disclose

Travel / Abstract Grant Application and Young Scientific Investigators' Session: I will not apply for Travel Grant or Young Scientific Investigators' Session

Which one would you like to apply to:

Date of Birth: