A study on loss functions and decision thresholds for the segmentation of multiple sclerosis lesions on spinal cord MRI

Burhan Rashid Hussein, Cédric Meurerée, Malo Gaubert, Arthur Masson, Anne Kerbrat, Benoît Combès, Francesca Galassi

To cite this version:

Burhan Rashid Hussein, Cédric Meurerée, Malo Gaubert, Arthur Masson, Anne Kerbrat, et al.. A study on loss functions and decision thresholds for the segmentation of multiple sclerosis lesions on spinal cord MRI. IABM 2023 - Colloque Français d’Intelligence Artificielle en Imagerie Biomédicale, Mar 2023, Paris, France. hal-04092786

HAL Id: hal-04092786

https://inria.hal.science/hal-04092786

Submitted on 9 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
A STUDY ON LOSS FUNCTIONS AND DECISION THRESHOLDS FOR THE SEGMENTATION OF MULTIPLE SCLEROSIS LESIONS ON SPINAL CORD MRI

Burhan Rashid Hussein 1, Cédric Meuré 1, Malo Gaubert 1,2, Arthur Masson 2, Anne Kerbrat 1,2, Benoît Combès 3, Francesca Galassi 1,4

1 Univ Rennes, Inria, CNRS, Inserm IRIS U 6074, Empenn ERL U 1228, Rennes, France
2 Rennes University Hospital (CHU), Department of Neurology, Rennes, France
3 Rennes University Hospital Hospital (CHU), Department of Neuroradiology, Rennes, France

Context and objectives

Multiple sclerosis patients often present hyper-intense T2-w lesions in the spinal cord. The severe imbalance between background and lesion classes poses a major challenge for Deep Learning segmentation methods. We aim at investigating the following strategies to help mitigating this issue:

• careful selection of the loss function
• adjustment of the conventional 0.5-thresholding

Dataset

• 161 T2-w cervical and thoracic MRI were acquired in 13 different sites from 108 subjects.
• Trained neurologists and radiologists performed manual segmentations, validated by a senior expert.
• Data splits were generated by assigning each subject to a single set while balancing the lesion load (Fig.1).

Methods

• Images were reoriented and resampled to 0.5mm isotropic using linear interpolation.
• The Spinal Cord Toolbox (SCT) was used for detecting spinal cord centerline.
• 3D patches (48x48x48) were extracted along the centerline for model training using a 3D-Unet architecture.5,6
• Data augmentation techniques such as patch overlapping, random mirroring, and 90° rotation were employed during the 300-epoch training.
• Overlapping patches were used for prediction with different thresholding values, and a thr-optimization subset was used to optimize binarization thresholds.
• The model was evaluated on an unseen test set and compared with the SOTA method (SCT).

Results

• Small changes in threshold lead to a larger change in voxel-F1 score for BCE and Dice (SCT) losses (Fig.2).
• Tversky-based losses seem more stable with threshold changes.
• Highest median L-F1 and V-F1 scores (40%) for Tversky loss compared to SOTA SCT with Dice loss (38%, Table 1).
• Example of segmentation output between Dice and Tversky loss on unseen test data with optimized threshold (Fig.3).

Conclusion

Tversky Index-based losses and a minimal adjustment of the decision threshold can yield less dispersed and higher median scores than more standard settings (Dice loss + 0.5 thr).1

Table 1. Lesion- and voxel-wise F1 scores, and patient-level sensitivity and specificity (median [mean ± std]) for models trained with the binary cross-entropy (BCE), Dice (SCT), Focal Tversky and Tversky losses. Metrics obtained on the unseen test set.

<table>
<thead>
<tr>
<th>Loss</th>
<th>L-F1</th>
<th>V-F1 (Dice)</th>
<th>P-sensitivity</th>
<th>P-specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCE</td>
<td>0.27 (0.28±0.3)</td>
<td>0.33 (0.29±0.24)</td>
<td>0.58</td>
<td>0.24</td>
</tr>
<tr>
<td>SCT</td>
<td>0.38 (0.34±0.32)</td>
<td>0.39 (0.35±0.25)</td>
<td>0.62</td>
<td>0.33</td>
</tr>
<tr>
<td>F. Tversky</td>
<td>0.26 (0.28±0.28)</td>
<td>0.31 (0.31±0.25)</td>
<td>0.65</td>
<td>0.22</td>
</tr>
<tr>
<td>Tversky</td>
<td>0.4 (0.34±0.31)</td>
<td>0.41 (0.33±0.24)</td>
<td>0.65</td>
<td>0.26</td>
</tr>
</tbody>
</table>


This work was supported by the French National Research Agency (Agence nationale de la recherche – ANR) as its 3rd PIA, integrated to France 2030 plan under reference ANR-21-RHUS-0014. Data collection was supported by a grant provided by the French State and handled by the French National Research Agency (Agence nationale de la recherche – ANR) within the framework of the France 2030 program under the reference ANR-10-COHD-002 OFSEP.