



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

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

► To cite this version:

Burhan Rashid Hussein, Cédric Meurée, 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

Burhan Rashid Hussein¹, Cédric Meurée¹, Malo Gaubert^{1,3}, Arthur Masson¹, Anne Kerbrat^{1,2}, Benoit Combès^{1*}, Francesca Galassi^{1*}

¹ Univ Rennes, Inria, CNRS, Inserm IRISA UMR 6074, Empenn ERL U 1228, Rennes, France

² Rennes University Hospital (CHU), Department of Neurology, Rennes, France

³ Rennes University 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).

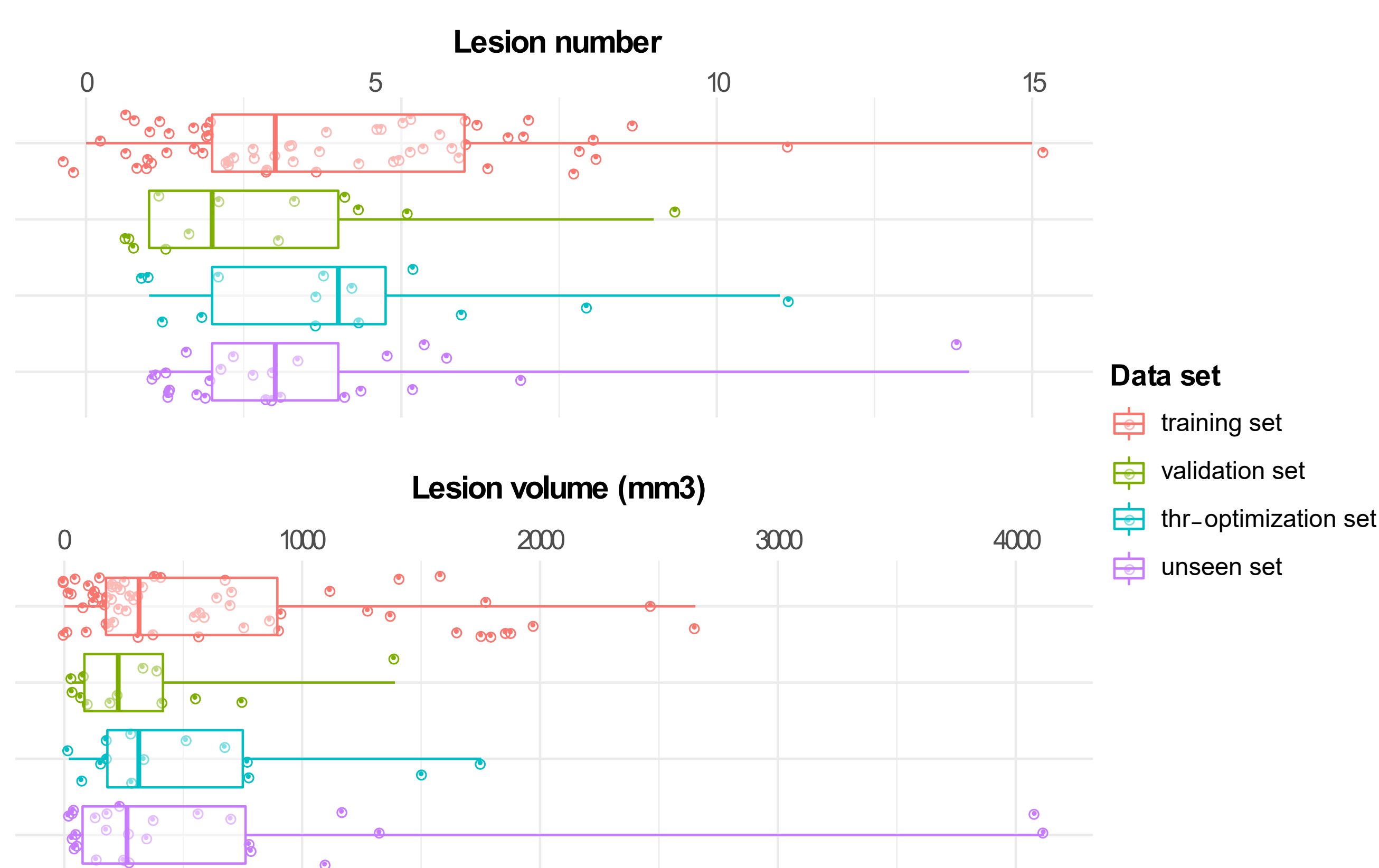


Fig. 1 Boxplots for lesion number and volume in the training (62 MRI scans), validation (13), thr-optimization (14) and unseen set (26, 46 with no MS lesions are not included here). Points are jittered to improve visualization.

Methods

- Images were **reoriented** and **resampled** to 0.5mm isotropic using linear interpolation.
- The **Spinal Cord Toolbox¹ (SCT)** was used for detecting spinal cord centerline.
- 3D patches** (48x48x48) were extracted along the centerline for model training using a **3D-Unet** architecture^{2,3}.
- 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).

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

| Loss | L-F1 | V-F1 (Dice) | P-sensitivity | P-specificity |
|------------|---------------------------------------|--|---------------|---------------|
| BCE | 0.27 (0.28 \pm 0.3) | 0.33 (0.29 \pm 0.24) | 0.58 | 0.24 |
| SCT | 0.38 (0.34 \pm 0.32) | 0.39 (0.35 \pm 0.25) | 0.62 | 0.33 |
| F. Tversky | 0.26 (0.28 \pm 0.28) | 0.31 (0.31 \pm 0.25) | 0.65 | 0.22 |
| Tversky | 0.4 (0.34\pm0.31) | 0.41 (0.33\pm0.24) | 0.65 | 0.26 |

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).

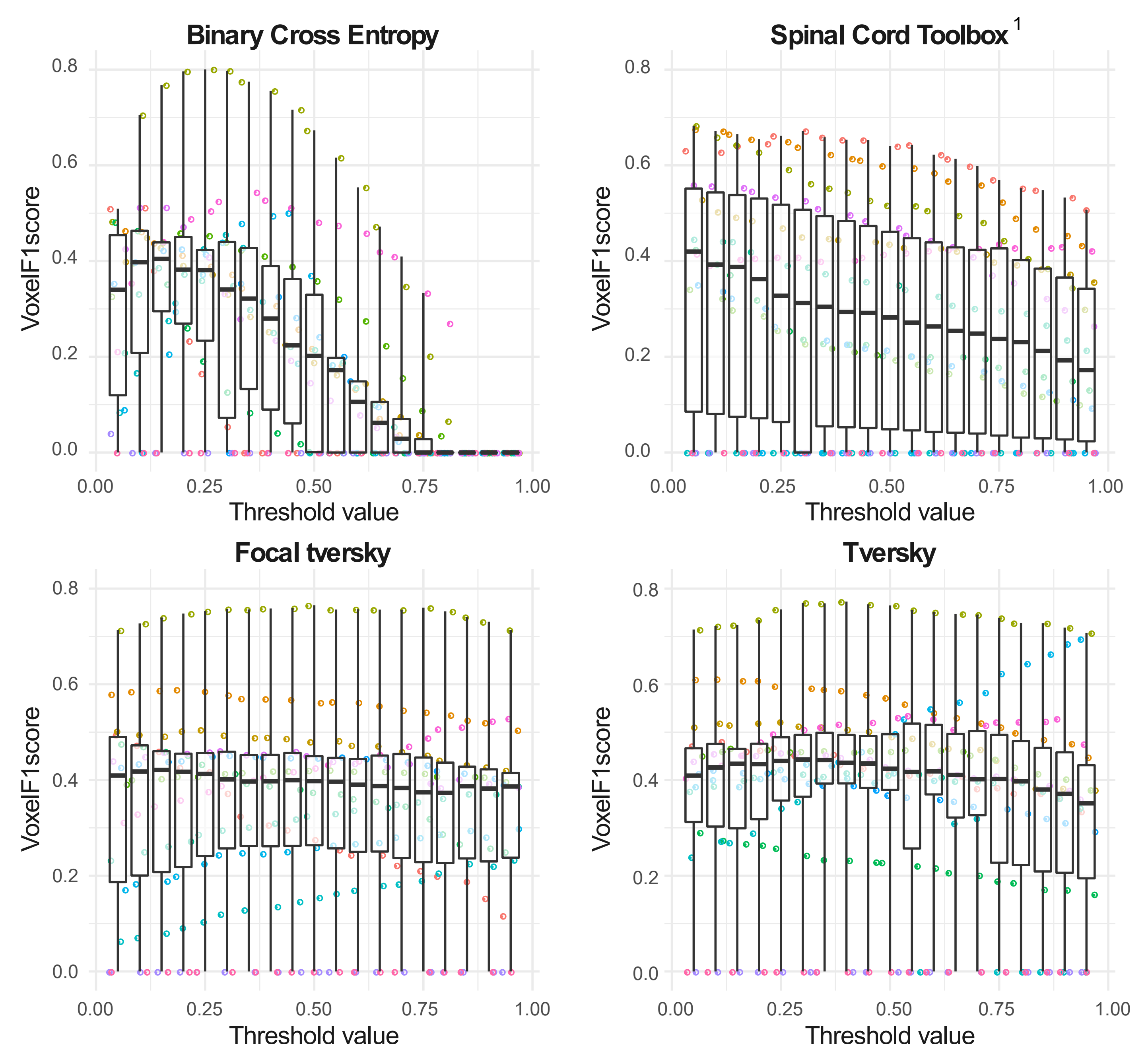


Fig. 2. Boxplots of voxel-wise F1 score for BCE, SCT, Focal Tversky and Tversky losses at each decision threshold value. One point corresponds to one image.

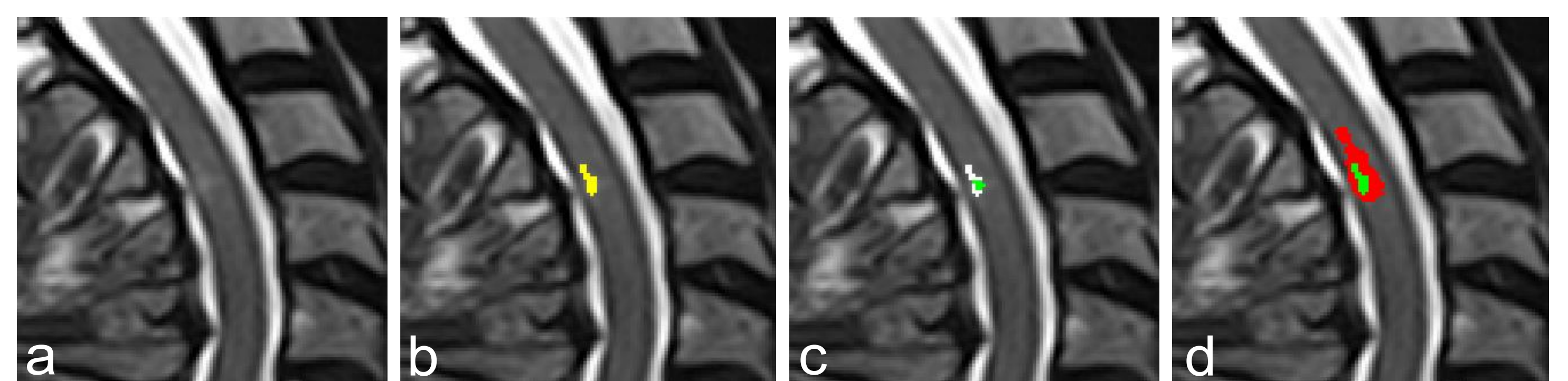


Fig. 3. Original image (a), overlaid by a reference lesion mask (b), and segmentations with the Tversky (c) and Dice (d) losses. FN, FP and TP voxels appear in white, red and green respectively.

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

¹ De Leener et al., "SCT: Spinal Cord Toolbox, an open source software for processing spinal cord MRI data", NeuroImage, 2017

² C. Gros et al., "Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks," NeuroImage, 2019

³ O. Çiçek et al., "3d u-net: Learning dense volumetric segmentation from sparse annotation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI, 2016

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-COHO-002 OFSEP.