EPISeg: Automated segmentation of the spinal cord on echo planar images using open-access multi-center data

January 27, 2025

Rohan Banerjee[1,2,27], Merve Kaptan[3], Alexandra Tinnermann[4], Ali Khatibi[5,6,7,8], Alice Dabbagh[10], Christian Büchel[4], Christian W. Kündig[11], Christine S.W. Law[3], Dario Pfyffer[3,11], David J. Lythgoe[16], Dimitra Tsivaka[16,34], Dimitri Van De Ville[19,20], Falk Eippert[10], Fauziyya Muhammad[21], Gary H. Glover[14], Gergely David[11], Grace Haynes[33], Jan Haaker[4], Jonathan C. W. Brooks[18], Jürgen Finsterbusch[4], Katherine T. Martucci[15], Kimberly J. Hemmerling[22,23], Mahdi Mobarak-Abadi[6], Mark A. Hoggarth[22,24], Matthew A. Howard[16], Molly G. Bright[22,23], Nawal Kinany[19,20], Olivia S. Kowalczyk[13,16], Patrick Freund[11,12,13], Robert L. Barry[25,26,30], Sean Mackey[3], Shahabeddin Vahdat[5,6,31], Simon Schading[11], Stephen B. McMahon[17], Todd Parish[23,32], Véronique Marchand-Pauvert[9], Yufen Chen[23], Zachary A. Smith[21], Kenneth A. Weber II[3], Benjamin De Leener[1,27,29†], Julien Cohen-Adad[2,27,28,29*†]

[1]Department of Computer Science, Polytechnique Montreal, Montreal, Quebec, Canada [2]Mila - Quebec AI Institute, Montreal, Quebec, Canada [3]Division of Pain Medicine, Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Palo Alto, CA, United States, [4]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [5]McConnell Brain Imaging Center, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada, [6]Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Quebec, Canada, [7]Centre of Precision Rehabilitation for Spinal Pain (CPR Spine), University of Birmingham, UK, [8]Centre for Human Brain Health, University of Birmingham, UK, [9]Sorbonne Université, Inserm, CNRS, Laboratoire d'Imagerie Biomédicale, LIB, F-75013, Paris, France, [10]Max Planck Research Group Pain Perception, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [11]Spinal Cord Injury Center, Balgrist University Hospital, University of Zurich, Zurich, Switzerland, [12]Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [13]Functional Imaging Laboratory, Department of Imaging Neuroscience, Queen Square Insti-

tute of Neurology, University College London, London, UK, [14]Radiological Sciences Laboratory, Department of Radiology, Stanford University School of Medicine, Palo Alto, CA, United States, [15]Center for Translational Pain Medicine, Department of Anesthesiology, Duke University Medical Center, Durham, NC, USA, [16]Department of Neuroimaging, Institute of Psychology, Psychiatry & Neuroscience, King's College London, London, UK, [17]Wolfson Centre for Age Related Diseases, King's College London, London UK, [18]Department of Psychology, University of Liverpool, Liverpool, United Kingdom, [19]Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland, [20]Neuro-X Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), Geneva, Switzerland, [21]Department of Neurosurgery, University of Oklahoma Health Sciences Center, OK, USA, [22]Department of Physical Therapy and Human Movement Sciences, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA, [23]Department of Biomedical Engineering, McCormick School of Engineering, Northwestern University, Evanston, Illinois, USA, [24]Department of Physical Therapy, North Central College, Naperville, Illinois, USA, [25]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA, [26]Harvard Medical School, Boston, MA, USA, [27]NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montreal, Montreal, Quebec, Canada, [28]Functional Neuroimaging Unit, CRIUGM, University of Montreal, Montreal, Quebec, Canada, [29]Research Center, Ste-Justine Hospital University Centre, Montreal, Quebec, Canada, [30]Harvard-Massachusetts Institute of Technology Health Sciences & Technology, Cambridge, MA, USA, [31]Department of Applied Physiology and Kinesiology, University of Florida, Gainesville, FL, USA, [32]Department of Radiology, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA, [33]Stephenson School of Biomedical Engineering at the University of Oklahoma in Norman, OK, USA, [34]Medical Physics Department, Medical School, University of Thessaly, Larisa, Greece

Keywords: Spinal Cord, Echo Planar Imaging, Functional Magnetic Resonance Imaging, Segmentation

Abstract: Functional magnetic resonance imaging (fMRI) of the spinal cord is relevant for studying sensation, movement, and autonomic function. Preprocessing of spinal cord fMRI data involves segmentation of the spinal cord on gradient-echo echo planar imaging (EPI) images. Current automated segmentation methods do not work well on these data, due to the low spatial resolution, susceptibility artifacts causing distortions and signal drop-out, ghosting, and motion-related artifacts. Consequently, this segmentation task demands a considerable amount of manual effort which takes time and is prone to user bias. In this work, we (i) gathered a multi-center dataset of spinal cord gradient-echo EPI with ground-truth segmentations and shared it on OpenNeuro https://openneuro.org/datasets/ds005143/versions/1.3.0, and (ii) developed a deep learning-based model, EPISeg, for the automatic segmentation of the spinal cord on gradient-echo EPI data. We observe a significant improvement in

terms of segmentation quality compared to other available spinal cord segmentation models. Our model is resilient to different acquisition protocols as well as commonly observed artifacts in fMRI data. The training code is available at https://github.com/sct-pipeline/fmri-segmentation/, and the model has been integrated into the Spinal Cord Toolbox as a command-line tool.

1 Introduction

Functional magnetic resonance imaging (fMRI) allows non-invasive investigation of the central nervous system, of which, the spinal cord (SC) is a crucial part [Kinany et al., 2022, Powers et al., 2018, Haynes et al., 2023]. SC fMRI has the potential for tracking disease progression and assessing treatment effectiveness by providing novel insights into various neurological conditions such as multiple sclerosis, neuropathic pain, and SC injury [Chen et al., 2015, Landelle et al., 2023, Staud et al., 2021, Conrad et al., 2018]. Although fMRI was introduced in the 1990s, the application of fMRI in the SC has been lagging behind brain imaging studies due to the technical challenges owing to data acquisition and processing [Stroman et al., 2014, Cohen-Adad, 2017]. Despite these difficulties and thanks to recent technical developments, SC fMRI is rapidly taking off, with applications to study sensorimotor systems, as well as autonomic processes in humans and animals [Barry et al., 2014, Kinany et al., 2020, Wu et al., 2019, Tinnermann et al., 2022, K. A. Weber et al., 2020, Sengupta et al., 2024, Hemmerling et al., 2023, Vahdat et al., 2015, Khatibi et al., 2022, K. A. Weber et al., 2016, K. Weber et al., 2016, K. A. Weber et al., 2018, Kündig et al., 2024].

FMRI data are commonly acquired with a gradient-echo echo planar imaging (EPI) sequence, due to its short acquisition times (about 2s per volume) and its heavy T2* weighting, which produces increased sensitivity to the blood oxygenation level contrast [Stehling et al., 1991]. The most notable challenge arises from the substantial changes in the magnetic susceptibility of tissues surrounding the SC, which introduces image distortions and signal drop-outs in the images [Powers et al., 2018, Kinany et al., 2022, Stroman et al., 2014, Cohen-Adad, 2017]. Additionally, physiological noise, motion artifacts, and poor contrast between the SC and surrounding pulsatile cerebrospinal fluid lead to degradation in the quality of EPI images [Giove et al., 2004, Fratini et al., 2014].

The analysis of fMRI data typically involves statistical inferences at the group level, which necessitates the valid and precise normalization of data into a template space which aims to establish a one-to-one correspondence between the data from different individuals. The emergence of SC templates [De Leener et al., 2018, Fonov et al., 2014, Azzarito et al., 2020, Taso et al., 2014, Bosma and Stroman, 2014] and

normalization procedures [De Leener et al., 2017, Stroman et al., 2008, Azzarito et al., 2020] make the group-level analysis accessible and reproducible across different SC fMRI datasets.

Currently, the gold standard for registering SC data to a template relies on the segmentation of the SC [Valosek and Cohen-Adad, 2024]. While there has been important development of automatic methods to segment the SC on MRI images [Blanc et al., 2023, De Leener et al., 2014, Gros et al., 2018, Bédard et al., 2023, Karthik et al., 2024, Horsfield et al., 2010, Hohenhaus et al., 2024, Tsagkas et al., 2023, Toufani et al., 2021, Lukas et al., 2021], these methods do not work well on gradient-echo EPI data due to aforementioned challenges. Moreover, the relatively low resolution of axial EPI scans (typically 1-1.5 mm in-plane, with 3-5mm thickness) makes the segmentation task even harder. An automated and robust segmentation method for the SC is therefore desirable to increase the reproducibility of SC fMRI studies [Hoggarth et al., 2022].

In this work, we introduce EPISeg, a deep learning (DL) based automatic SC segmentation method. Along with EPISeg, this paper also introduces open-access fMRI data of the SC from multiple sites (N=15). The dataset comes with ground-truth SC segmentations that were used to train the EPISeg model. We demonstrate the robustness of EPISeg amongst different sites, involving different scanners including data from healthy volunteers as well as patient groups, and against commonly observed artifacts such as partial volume effects, and ghosting. We compare our method with other state-of-the-art segmentation models like PropSeg [De Leener et al., 2014], DeepSeg [Gros et al., 2018] and Contrast-agnostic SC segmentation [Bédard et al., 2023], and show a significant improvement in segmentation performance using the Dice Score (DS) and Hausdorff Distance (HD). To our best knowledge, this is the first work regarding automatic segmentation of SC on fMRI data. The contributions of this paper are:

1. Creation of a new open-source SC gradient-echo EPI dataset (n=406 participants, 15 sites).

2. Introduction of an automated segmentation method for SC on fMRI data, thoroughly validated on diverse datasets spanning different populations and different regions of the SC including the lumbar and cervical SC.

3. Integration of this algorithm as a single-line command in SCT [De Leener et al., 2017], making it easily accessible to the scientific community.

2 Methods

In this section, we describe the data collection, manual labeling, model architecture, and model training using an active learning framework. An overview of the pipeline is shown in Figure 1.

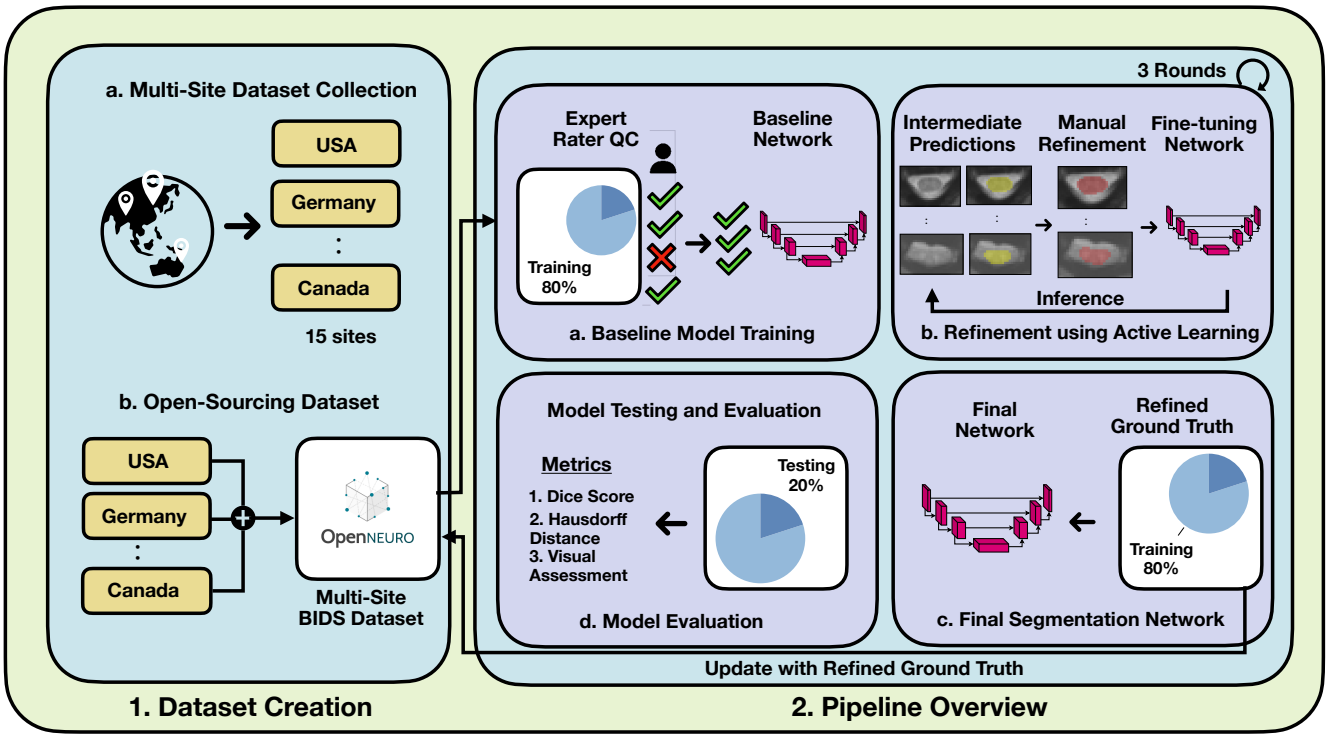

Figure 1: **EPISeg overall pipeline.** 1. Dataset creation and open-sourcing process. 1a. Initiation of dataset contribution call for EPISeg. 1b. The collected datasets are rearranged according to the brain imaging data structure (BIDS) convention and uploaded to OpenNeuro, forming a multi-site BIDS dataset. 2. Training and validation pipeline overview. 2.a. An expert selects reliable segmentations, which are used to train the baseline model. This model undergoes expert rater quality control (QC) to ensure high-quality training data. 2b. The baseline model generates initial predictions, which are then manually refined by experts. These refined predictions are used to fine-tune the baseline model. This process is repeated for three rounds, ensuring incremental improvement and robustness. 2c. Participants and refined ground truth from the refinement phase are used to train the final network. The refined ground truth is also updated in the open-source dataset. 2d. The final network is used for inference on test data, with metrics such as Dice Score (DS), Hausdorff Distance (HD), and visual assessments reported for comprehensive evaluation.

2.1 Data

We gathered gradient-echo EPI datasets from multiple sites (15 sites, 406 participants in total). All volunteers provided written informed consent following Institutional Review Board approval and the Declaration of Helsinki (see Table 1 for details of acquisitions for all sites). All participating sites were asked to share (i) the first 20 motion-corrected EPI volumes of the functional run (more volumes were unnecessary as the primary goal of these images is to perform segmentation, not fMRI analysis), (ii) the BIDS-compatible JSON file containing information about the acquisition, (iii) the mean image of the motion-corrected data, and (iv) the SC segmentation based on the mean motion-corrected image. Some sites manually created the segmentations from scratch, while other sites started off from a segmentation generated by automated methods such as PropSeg or DeepSeg, with manual correction. Out of 406 volunteers across all sites, the majority were healthy volunteers (n=359), 17 volunteers had degenerative cervical myelopathy, and 30 volunteers had fibromyalgia.

The dataset was put together into one multi-site dataset and organized according to the BIDS convention [Gorgolewski et al., 2016]. The segmentations produced at each site were reviewed by three raters overseeing the entire study (RB, MK, JCA) using the sct_qc quality control (QC) tool [Valosek and Cohen-Adad, 2024]. Segmentations were subsequently updated using EPISeg wherever we found erroneous segmentations. The dataset was uploaded on OpenNeuro (https://openneuro.org/datasets/ds005143/versions/1.3.0), except for one site due to the data sharing policy of their respective ethics committee.

Table 1: Dataset details

Site name	Institution	Sample	Scanner and sequence
Leipzig_rest★ [Kaptan et al., 2022 Kaptan, Horn, et al., 2023]	Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany	HV; N=45 (20F); Age: 27±3.8	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; Number of slices: 24; in-plane resolution: 1.0 × 1.0mm^2; slice thickness: 5.0mm; FoV = 128 × 128mm^2; TR: 2.312s; TE: 40 ms; FA = 84°

Continued on next page

Table 1: Dataset details (continued)

Site name	Institution	Sample	Scanner and sequence
Geneva_rest1 [Kinany et al., 2019]	École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland	HV; N=17 (11F); Age: 27±3.5	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 27; FoV = 48 × 144mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 3mm; TR = 2.5s; TE = 34 ms; FA = $80°$
Geneva_rest2 [Kinany et al., 2020]	École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland	HV; N=12 (6F); Age: 28.8±3.2	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 27; FoV = 48 × 144mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 3mm; TR = 2.5s; TE = 34 ms; FA = $80°$
Geneva_rest3 [Kinany et al., 2022]	École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland	HV; N=14 (7F); Age: 28.6±4.2	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 31; FoV = 48 × 144mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 3mm; TR = 2.5s; TE = 34 ms; FA = $80°$
KCL_rest★ [Kowalczyk et al., 2024]	King's College London, London, UK	HV; N=23 (13F); Age: 23.9±3.8	3T General Electric MR750 System; 2D gradient-echo EPI; number of slices: 38; FoV = 180mm; reconstruction matrix = 128 × 128mm^2; in-plane resolution = 1.875 × 1.875mm^2; slice thickness = 4mm (1mm gap); TR = 2.5s; TE = 30 ms; FA = $90°$

Continued on next page

Table 1: Dataset details (continued)

Site name	Institution	Sample	Scanner and sequence
NW_motor [Hemmerling et al., 2023]	Northwestern University, Chicago, Illinois, USA	HV; N=28 (18F); Age: 26±4.7	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 25; FoV = 128 × 44mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 3mm; TR = 2s; TE = 30ms; FA = $90°$
NW_motor-Weber [K. A. Weber et al., 2016]	Northwestern University, Chicago, Illinois, USA	HV; N=11 (6F); Age: 27.7±1.9	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 31; FoV = 128 × 44mm^2; in-plane resolution = 1 × 1mm^2, slice thickness = 3mm, TR = 2.5s, TE = 30ms, FA = $80°$
NW_thermal [K. Weber et al., 2016]	Northwestern University, Chicago, Illinois, USA	HV; N=12 (3F); Age: 28.8±2.5	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 31; FoV = 128 × 128mm^2; in-plane resolution = 1 × 1mm^2, slice thickness = 3mm, TR = 2.5s, TE = 30ms, FA = $80°$
NW_tactile [K. A. Weber et al., 2020]	Northwestern University, Chicago, Illinois, USA	HV; N=12 (3F); Age: 28.8±2.5 Degenerative cervical myelopathy N = 2; Age: 33.5 ± 2.1	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 25; FoV = 128 × 44mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 3mm; TR = 2s; TE = 30ms; FA = $80°$

Continued on next page

Table 1: Dataset details (continued)

Site name	Institution	Sample	Scanner and sequence
Zurich_Cerv _Bilatmotor	Balgrist University Hospital, University of Zurich, Zurich, Switzerland	HV; N=18 (3F); Age: 28.8±2.5	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 18; FoV = 128 × 128mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 5mm; TR = 3.26s; TE = 31ms; FA = 87°
Stanford_rest [Kaptan, Law, et al., 2023]	Stanford University, Palo Alto, California, USA	HV; N=29 (F); Age: 25.6±5.5	GE 3T Discovery 750; 2D gradient-echo EPI; number of slices: 43; FoV = 80mm; matrix = 64 × 64mm^2; in-plane resolution = 1.25 × 1.25mm^2; slice thickness = 5mm; TR = 2.5s; TE = 30ms; FA = 80°
Stanford_ restmartucci [Martucci et al., 2021, Martucci et al., 2019]	Stanford University, Palo Alto, California, USA	HV; N=15 (F); Age: 49±10.8 Fibromyalgia N=30 (F); Age: 50.2±8.8	3T General Electric MR750 System; number of slices: 14; FoV = 160mm; reconstruction matrix = 128 × 128mm^2; in-plane resolution = 1.25 × 1.25mm^2; slice thickness = 4mm^2; TR = 2.5s; TE = 30ms; FA = 75°
UNF_MSL [Khatibi et al., 2022, Mobarak-Abadi et al., 2024]	Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, QC, Canada	HV; N=27 (14F); Age: 24.9±3	3T whole-body Siemens TIM Trio MRI System; 2D gradient-echo; number of slices: 40-43 (8-10 SC); FoV = 132 × 132mm^2; in-plane resolution = 1.2 × 1.2mm^2; slice thickness = 5mm (9mm gap); TR = 3.14s; TE = 33ms; FA = 90°

Continued on next page

Table 1: Dataset details (continued)

Site name	Institution	Sample	Scanner and sequence
OUHSC_stim	University of Oklahoma Health Sciences Center, OK, USA	HV; N=13(9F); Age: 50.3±6.6 Cervical Spondylotic Myelopathy N=15(10F); Age: 57±9.9	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 10; FoV = 128 × 128mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 5mm; TR = 0.86s; TE = 30ms; FA = 50°
Hamburg _pain♦ [Tinnermann et al., 2021]	Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany	HV; N=30 (M); Age: 23.9±3.8	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 10; FoV = 128 × 128mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 5mm; TR = 0.86s; TE = 30ms; FA = 50°
Zurich _Lumb_rest [Kündig et al., 2024]	Balgrist University Hospital, University of Zurich, Zurich, Switzerland	HV; N=13 (4F); Age: 28.7±3.5	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 15; FoV = 144 × 48mm^2; in-plane resolution = 1 × 1mm^2; slice thickness = 5mm; TR = 1.29s; TE = 35ms; FA = 74°
Leipzig_pain♦ [Dabbagh et al., 2024]	Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany	HV; N=40 (20F); Age: 27.3±4.6	3T whole-body Siemens Prisma MRI System; 2D gradient-echo EPI; number of slices: 16; in-plane resolution:1.0 × 1.0mm^2; slice thickness: 5.0mm; FoV: 128 × 128mm^2; TR: 1.8s; TE: 40ms; FA = 70°

*HV = Healthy Volunteers; FoV = Field of View; TR = Repetition Time; TE = Echo Time; FA = Flip Angle; ★ = Semi-automatic segmentation; ♦ = Automatic segmentation

2.2 Data preprocessing

Our proposed model is designed to segment the SC on fMRI images. The input consists of a mean motion-corrected volume representing the fMRI scan. This mean volume condenses the motion-corrected time-series data into a single three-dimensional scan by averaging over all time points. Each input volume is matched with a ground truth binary segmentation that outlines the spinal cord's boundaries.

2.3 Criteria for selecting training data

When performing multi-step active learning training, the first training batch necessitates reliable ground truth segmentation. In this case, we aimed to identify data subsets, D_g, that had accurate ground truth SC segmentations. These data subsets were in general of high quality, notably thanks to the z-shimming method [Finsterbusch et al., 2012] used in some of the datasets like the Leipzig_pain, Leipzig_rest. It is important to note that all selection processes prioritized accurate segmentations over image quality (which is inherently related to segmentation quality), ensuring that the model was trained on the most precise delineations of the SC.

2.4 Model training protocol

Our training protocol involves a three-step model training procedure employing a human-in-the-loop active learning to iteratively remove errors and improve ground truth accuracy and then training a final segmentation model which is deployed in SCT. Our reasoning for using active learning in this setting was to develop a model resilient to the various types of image artifacts it encounters over the different rounds of training.

In the first step, as shown in Figure 2 we trained a baseline model M_b using the dataset D_g which had 96 participants. The baseline model is a 3D U-Net [Çiçek et al., 2016], trained using the nnUNetv2 framework [Isensee et al., 2020], which self-configures parameters for optimal performance. We chose nnUNetv2 to perform this task because it has shown state-of-the-art performance in a range of medical anatomies and imaging modalities [Isensee et al., 2020, Isensee et al., 2024]. For training M_b, we used the default parameters configured by nnUNetv2 for our dataset. The training details include a batch size of 5, learning rate of 0.01 with a polynomial learning rate scheduler and a stochastic gradient descent optimizer for a total of 1,000 epochs. The model was trained using the combination of Dice loss and cross-entropy loss.

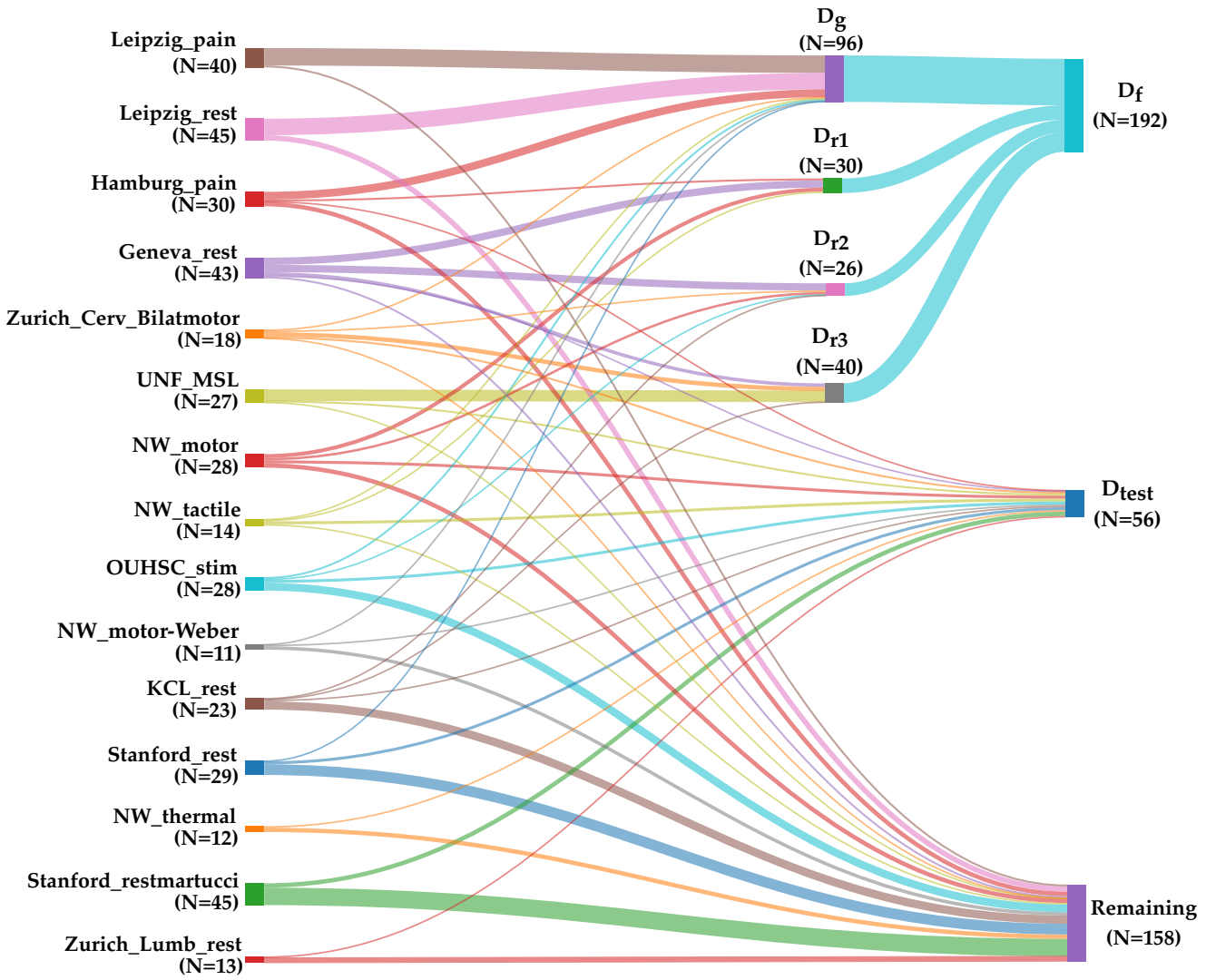

Figure 2: **Dataset split.** The sankey chart shows the different dataset splits created during the training process. The split D_g is for the baseline model, D_{r1} is for the first round of active learning training, D_{r2} is for the second round of active learning training, D_{r3} is for the third round of active learning training, D_f is the aggregated set of D_g, D_{r1}, D_{r2}. D_{r3} and D_{test} is the held-out test set to validate the EPISeg performance. The "remaining" set contains images with artifacts, erroneous segmentations.

The second step consisted of iterative refinement of training ground truth and model using active learning. Once the baseline model, M_b was trained, we used it to generate segmentations on the remaining dataset. For the first round of active learning, a subset of these segmentations, D_{r1}, consisting of 30 images were chosen for quality control assessment and manual corrections, if necessary. This subset D_{r1} was then used to fine-tune M_b, resulting in an updated model in round 1 of training, M_{r1}. To further improve the model's capability, two more rounds of active learning with the same procedure with variable dataset sizes were conducted which resulted in the creation of datasets D_{r2}, D_{r3} and corresponding models M_{r2} and M_{r3} respectively. These training rounds used the same default parameters used to train M_b.

For the third and final step of the training protocol, all the datasets D_g, D_{r1}, D_{r2} and D_{r3} were aggregated into one final dataset, D_f which had 192 3D images to train one model, M_f. Our reasoning behind training the M_f was to overcome the signs of forgetting that M_{r3} was exhibiting. Forgetting is a common phenomenon that is observed during active learning i.e. when networks are trained in a sequential fashion, the new learnings of the network tend to interfere with the old learning of the network [McCloskey and Cohen, 1989]. We discuss further about this observation in the "Effect of active learning on EPISeg" results section. Following a similar training procedure from the first step, a new set of nnUNetv2 parameters were configured for D_f, which included a batch size of 15 with a polynomial learning rate scheduler, trained for 1,000 epochs. Along with this, we use the latest nnUNetv2 ResEnc [Isensee et al., 2024] variant of the UNet with residual connections in the encoder since the authors mention that using residual connections along with the nnUNetv2 framework is a recipe for better segmentation performance.

2.5 Evaluation protocol

To evaluate our model performance, we created a multi-site held-out test set, D_{test}. The whole dataset (containing all subjects from all sites) was split in a 80:20 ratio where participants from the 80% of the dataset were used for training and 20% of the dataset were held out for testing. The D_{test} was not included in the training or validation sets. To have a fair representation of all the sites in D_{test}, we randomly chose 20% of the participants from each site and aggregated it into one dataset which in total had 56 participants. We evaluate our model's performance on the D_{test} and compare it with the SC segmentation models PropSeg, DeepSeg and Contrast-agnostic SC segmentation model. We discuss these results in the following sections.

We use Dice Score (DS) and Hausdorff Distance (HD) as validation metrics. For HD, we use the 95th percentile HD (HD95) since it reduces the impact of outlier errors, providing a more robust and

representative measure of segmentation accuracy. The DS quantifies the spatial agreement between the predicted and ground truth segmentations, with higher scores indicating a closer alignment. On the other hand, the HD measures the maximum separation between the predicted SC segmentation and the nearest point in the ground truth segmentation, providing insights into the model's ability to capture the geometric correspondence between predicted and true SC boundaries. We utilize the implementations provided by the MetricsReloaded framework [Maier-Hein et al., 2022] for our metrics. Additionally, we verified the model's performance to ensure that model was producing segmentations as expected [discussed in section 3.4]

3 Results

3.1 Automatic fMRI SC segmentation

Table 2 compares the performance of the EPISeg models trained on various datasets - M_{r3} trained on the final round of the active learning data subset D_{r3}; M_f trained on the aggregated set D_f from all rounds of active learning rounds; M_{ResEnc} is the model with ResNet connections trained on D_f. The best performance (in terms of DS) is seen with the M_f dataset, indicating the advantage of having a single model with reliable segmentations rather than having several rounds of fine-tuning.

Table 2: Performance of the proposed EPISeg on D_{test}. M_{r3} refers to the model we obtained after the 3rd round of active learning training, M_f refers to the model trained on D_f which is the final aggregated data from all the 4 rounds of active learning training and M_{ResEnc} refers to latest ResNet based nnUNet model trained on D_f.

Model	DS $\pm$ S.D.	HD95 $\pm$ S.D. (in mm)
EPISeg (M_{r3})	0.82 $\pm$ 0.06	1.73 $\pm$ 1.03
EPISeg (M_f)	0.88 $\pm$ 0.05	1.41 $\pm$ 0.86
EPISeg (M_{ResEnc})	0.87 $\pm$ 0.05	1.28 $\pm$ 0.73

Even though we see M_f performs a little better than M_{ResEnc} in terms of DS, M_{ResEnc} has a better HD95 score. When we investigated this qualitatively, we saw that M_{ResEnc} gave more complete segmentations, for example, it was able to segment the first and last slice more consistently and was more robust to slices with artifacts. We therefore used the M_{ResEnc} as our final EPISeg model.

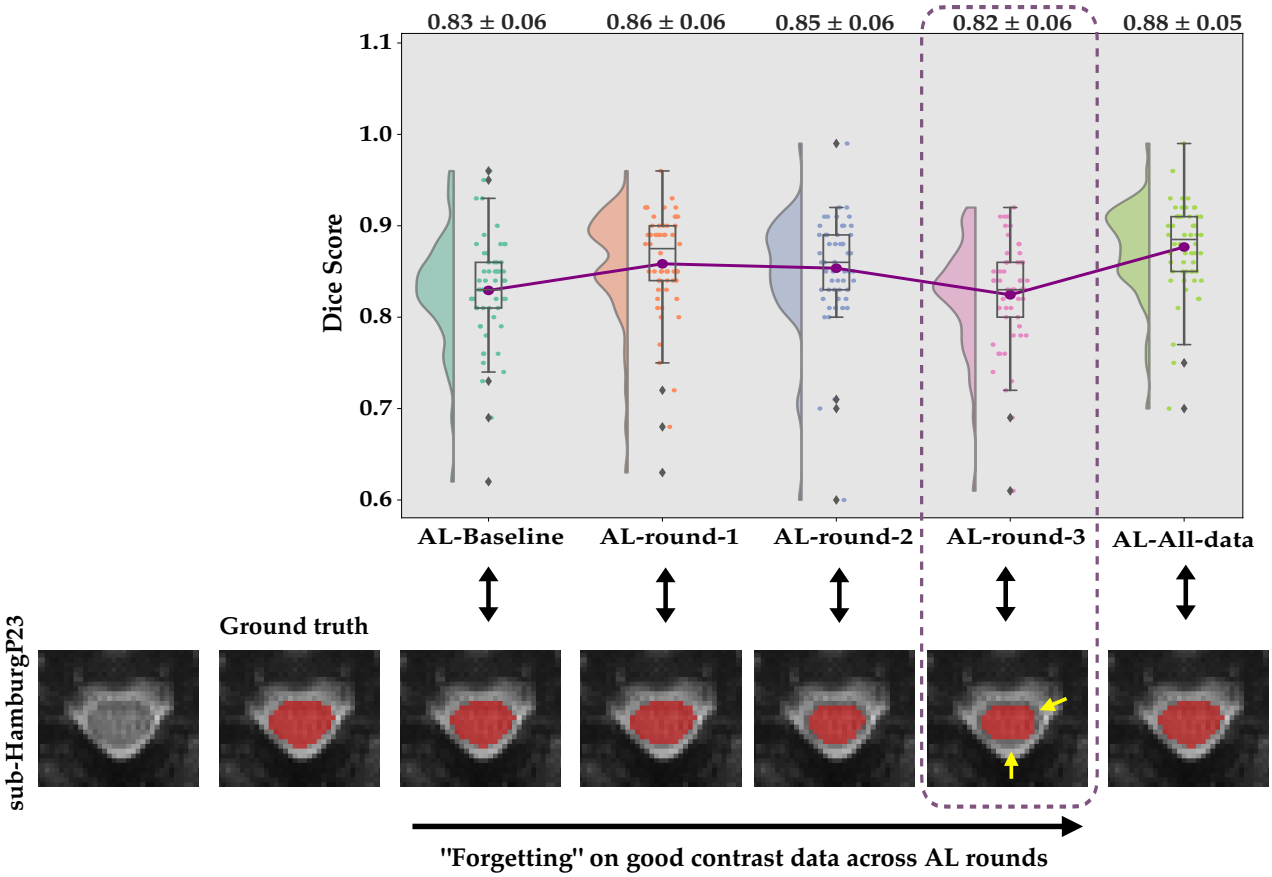

Figure 3: **Effects of Active Learning.** The raincloud plot shows the DS distribution of different models trained during the active learning training rounds. The images below show the axial slice from a subject and the red overlays show the model prediction of the SC. The dotted line and the yellow arrows show the performance degradation of the model over the training rounds.

3.2 Effect of active learning on EPISeg

We relied heavily on human-in-the-loop active learning for training EPISeg. Our selection of D_g was based on the quality of the segmentations rather than the image quality. However, since it was easier to achieve better segmentations on higher quality images, they were predominantly used in training M_b. More and more lower quality images were introduced in the later rounds of active learning. As we can observe in Figure 3, there is a noticeable dip in the performance of M_{r3}. When we investigated this phenomenon, we observed that, while the model's performance improved for lower quality images, it degraded the performance for high quality images similar to the ones used in training M_b. We refer to this as forgetting and therefore decided to train M_f to overcome this drawback of EPISeg.

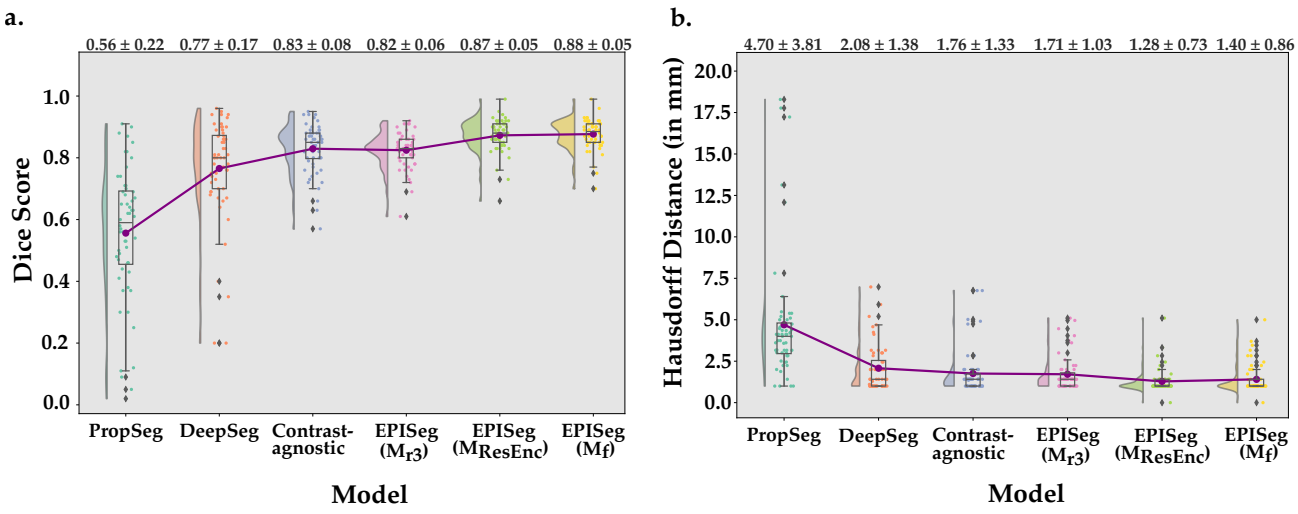

Figure 4: **Raincloud plots from method comparison.** 4a. The DS were computed between the ground truth and predictions. The raincloud plots show the DS distributions of the different models. The red dots and the line connect the mean values of the DS of the respective models. The mean DS along with the standard deviation is printed on the top of the plot. 4b. In a similar way to 4a., it shows the HD95 for the different models.

3.3 Comparison with other methods

We compared EPISeg with other methods, namely, PropSeg, DeepSeg and Contrast-agnostic (from SCT v6.3: https://github.com/spinalcordtoolbox/spinalcordtoolbox/releases/tag/6.3). PropSeg does not use DL for SC segmentation and has been tested only on T1, T2 and T2* contrasts, DeepSeg uses DL but no EPI data were used during the training. Similarly, Contrast-agnostic did not use gradient-echo EPI

data during training. However, both DeepSeg and Contrast-agnostic can segment EPI data. Our motive to compare our results with these methods was the fact that they are considered state-of-the-art, fully automatic 3D SC segmentation methods.

Table 3: Comparison of SC segmentation methods on D_{test}

Model	DS ± S.D.	HD95 ± S.D. (in mm)
Propseg	0.56 ± 0.22	4.70 ± 3.81
Deepseg	0.77 ± 0.17	2.08 ± 1.38
Contrast-agnostic	0.83 ± 0.08	1.76 ± 1.33
EPISeg (M_{r3})	0.82 ± 0.06	1.73 ± 1.03
EPISeg (M_f)	**0.88 ± 0.05**	1.41 ± 0.86
EPISeg (M_{ResEnc})	0.87 ± 0.05	**1.28 ± 0.73**

In Table 3, the various model performances on D_{test} are presented. The proposed model, EPISeg, exhibits superior performance in comparison to the alternative baselines. Both PropSeg and DeepSeg depict subpar performance when dealing with EPI data. Numerous instances exist where these models struggle to accurately segment the SC. PropSeg, for example, faces difficulties in identifying the SC edges in many scenarios, resulting in segmentation failure. Conversely, Deepseg initially locates the center of mass before proceeding with SC segmentation, a task it struggles with in EPI SC segmentation. The Contrast-agnostic model demonstrates satisfactory but lower performance on our dataset due to its exposure to diverse SC contrasts during the training process.

Figure 4 presents raincloud plots comparing the performance of different methods on D_{test}. The half-violin plots illustrate the distribution of DS, while individual points represent the performance of each data point in D_{test}. EPISeg's DS are tightly clustered around a single point, which coincides with both the mean and median of the scores. This concentration indicates that EPISeg delivers consistent performance without outliers. When it comes to comparison of HD95 in Figure 5, we observe a similar pattern as in Figure 4. To highlight the strong performance of EPISeg, we see a point where the HD95 value is 0.0mm which, means that 95% or more surface points had 0.0mm error, indicating a perfect segmentation performance.

3.4 Qualitative assessment

We evaluated our model's performance on D_{test} through a qualitative assessment involving two raters with 6 and 18 years of experience of SC MRI experience, respectively. These raters meticulously examined the segmentation results for each slice, as documented in the QC report generated by the sct_qc functionality in SCT. Figure 5 showcases the performance of various models on selected axial slices, highlighting the

quantitative performance of each. The participants represented in the figure include healthy controls as well as patients with degenerative cervical myelopathy and fibromyalgia.

In our evaluations, we observed that PropSeg frequently under-segments and sometimes fails to segment the SC entirely, especially when the SC is not clearly distinguishable from the cerebrospinal fluid. DeepSeg performs slightly better but still displays inconsistencies in SC segmentation. The Contrast-agnostic model offers more consistent segmentation boundaries but often over-segments the SC. EPISeg demonstrates superior performance, closely matching the ground truth. For instance, EPISeg successfully segmented the SC in the sub-ZurLumbC01, a dataset with low-quality lumbar cord images that were not included in the training set.

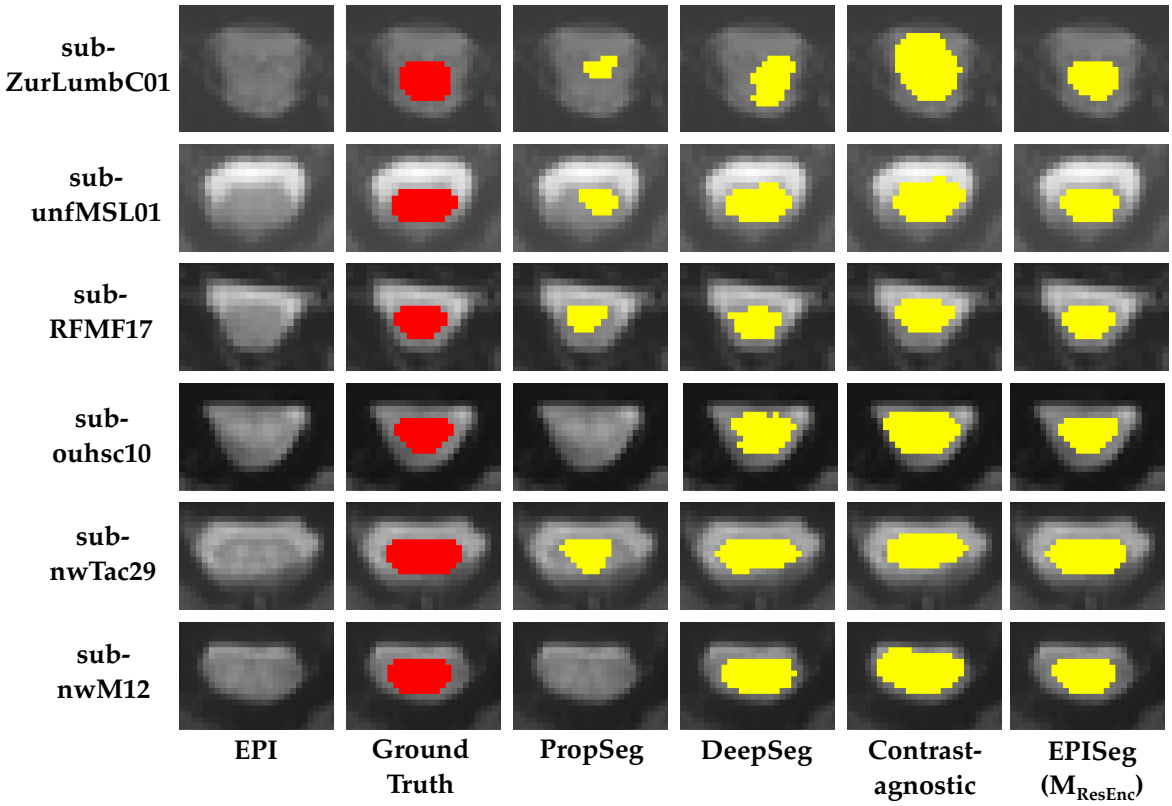

Figure 5: **Visual assessment.** Comparison of EPISeg (M_{ResEnc}) with other automatic SC segmentation methods. The colored overlay shows the SC on the images. We can observe that EPISeg provides the most precise segmentations.

4 Discussion

In this study, we introduce EPISeg, a DL-based method for automatic SC segmentation from gradient-echo EPI images, which are widely used in (fMRI). EPISeg is robust to heterogeneous resolutions, scanner settings, and motion artifacts. Consequently, it addresses challenges that have historically hindered SC segmentation on EPI data, requiring labor-intensive and bias-prone manual interventions.

In order to develop a robust EPI segmentation model, the first necessary step was to create a dataset that would provide a DL model with enough context to be able to segment the SC in a low-quality image setting. We curated a large multi-site EPI SC dataset consisting of spinal and corticospinal acquisitions, arranged it as per the BIDS convention, and made it publicly available in OpenNeuro - https://openneuro.org/datasets/ds005143/versions/1.3.0.

While creating the dataset, we observed that the provided ground truth was erroneous in some cases (e.g., over-segmented or under-segmented). This led us to the decision to train EPISeg sequentially. We used good-quality segmentations to train a strong baseline model which provided high-quality segmentations before successive models. At each model training step, a pool of best predictions was chosen along with manual correction wherever required, including verification by an expert. We obtained our final model using these steps iteratively, which was able to segment the SC reliably. We demonstrated the robustness of our model against low spatial resolution of EPIs, susceptibility artifacts, signal drop-outs, ghosting, and motion-related distortions and eliminated the need for manual interventions.

In terms of SC segmentation directly on EPI data, only [Giulietti et al., 2011] provides a semi-automatic method. This method consists of multiple preprocessing steps before the SC was segmented using k-means clustering. The authors claim that their method's performance efficiency declines with the increase in the distortion present in the image. Even though we cannot reproduce this method for comparison due to the lack of provided details regarding the parameters used, we show that, in addition to being completely automatic, EPIseg is highly robust to distortions present in EPI data.

Our method, EPISeg, demonstrates superior accuracy and robustness in spinal cord segmentation compared to state-of-the-art models like PropSeg, DeepSeg, and Contrast-agnostic segmentation. It generalizes well across different SC conditions and handles out-of-distribution data effectively. EPISeg achieves a mean Dice score of 0.88 with a standard deviation of 0.05, reflecting its consistent performance. PropSeg performs poorly as it was primarily optimized for T1 and T2 contrasts and is not well-suited for EPI data. Similarly, DeepSeg and the contrast-agnostic model, which were not specifically trained for spinal cord segmentation on EPI images, lack sufficient context to accurately handle the unique characteristics of EPI data. Although the Contrast-agnostic model performs better than PropSeg and DeepSeg due to its

broader training dataset, it still falls short of EPISeg's performance.

To potentially circumvent the need for segmenting the EPI for registration purposes, a high-resolution anatomical image can be acquired with the same field-of-view as the EPI images (in a way that registration between this image and the EPI is not necessary). However, this approach has also drawbacks. Mainly, the acquisition of a high-resolution image would require additional scan time. Especially considering the movement of a participant within and across multiple functional runs, it may be even necessary to acquire one high-resolution image which would significantly increase scan time.

Despite the superior performance of EPISeg for segmentation of spinal cord fMRI data against other automated methods, several limitations of the current work must be considered. Although this study includes a wide range of data from multiple sites and participants including data from different patient populations, it does not cover all possible data domains. Variations in acquisition parameters and spinal cord conditions across different populations may not be fully represented here, possibly limiting the generalizability of our findings. Additionally, our method has solely been tested on data acquired at 3T. Segmentation performance and accuracy could vary across other field strengths, due to the fact that, at lower (1.5T) and higher (7T) fields, differences in signal-to-noise ratio, contrast, and spatial resolution may influence the quality of the EPI data.

EPISeg here is optimized for EPI sequences as they are the most commonly used technique for fMRI acquisitions. However, spinal cord fMRI can also be performed using other sequence types (such as T2*-weighted fast gradient echo acquisitions). The performance of EPIseg on other acquisition types has not been evaluated in this study. Finally, although we gathered data from various sites that have employed different motion correction techniques, it is important to remember that effective motion correction is critical to the success of spinal cord fMRI segmentation, given the susceptibility of this region to motion artifacts from respiration, pulsation, and participant movement. Inadequate motion correction could impact the accuracy of segmentation by obscuring spinal cord boundaries. Future work should aim to broaden the data domains represented, including a wider range of acquisition parameters, populations, and magnetic field strengths (1.5T, 3T, and 7T). Moreover, validation of EPIseg on GRE T2*-weighted sequences and improvements in motion correction strategies will be necessary to enhance segmentation robustness and ensure that it performs effectively across diverse imaging conditions.

5 Conclusion

We provide an automatic SC segmentation method for gradient-echo EPI. We demonstrate the performance of our method on various spinal and also corticospinal fMRI datasets (including samples from clinical populations) from different sites with different acquisition parameters. The dataset is publicly available on OpenNeuro. This reliable segmentation method will provide a strong basis for fMRI data preprocessing, particularly for spatial normalization which is necessary for obtaining reliable group-level results. Our method is readily available as a part of SCT v6.3 or higher.

Data and Code Availability

All the data used for this study is open-sourced and can be found at: https://openneuro.org/datasets/ds005143/versions/1.3.0

All the scripts, instructions to reproduce results and the lastest model weights can be found here: https://github.com/sct-pipeline/fmri-segmentation/releases/tag/v0.2. This method is integrated into SCT v6.3 and can be used by using the *sct_deepseg* command with the *-task seg_sc_epi* parameter.

Author Contributions

Rohan Banerjee: Data Curation, Methodology, Formal Analysis, Software, Validation, Writing – Original Draft, Writing – Review Editing, Visualization; Merve Kaptan: Data Curation, Validation, Methodology, Writing – Original Draft, Writing – Review Editing; Kenneth A. Weber II: Supervision, Writing – Review Editing, Conceptualization, Funding Acquisition; Benjamin De Leener: Supervision, Methodology, Writing – Review Editing, Conceptualization, Funding Acquisition; Julien Cohen-Adad: Supervision, Methodology, Writing – Review Editing, Conceptualization, Funding Acquisition; Alexandra Tinnermann: Data Contribution; Ali Khatibi: Data Contribution; Alice Dabbagh: Data Contribution; Christian Büchel: Data Contribution; Christian W. Kündig: Data Contribution; Christine S.W. Law: Data Contribution; Dario Pfyffer: Data Contribution, Writing – Review Editing; David J. Lythgoe: Data Contribution, Writing – Review Editing; Dimitra Tsivaka: Data Contribution; Dimitri Van De Ville: Data Contribution; Falk Eippert: Data Contribution, Writing – Review Editing; Fauziyya Muhammad: Data Contribution; Gary H. Glover: Funding Acquisition, Data Contribution; Gergely David: Data Contribution, Writing – Review Editing; Grace Haynes: Data Contribution; Jan Haaker: Data Contribution, Writing – Review Editing;

Jonathan C. W. Brooks: Data Contribution, Writing – Review Editing; Jürgen Finsterbusch: Data Contribution, Writing – Review Editing; Katherine T. Martucci: Data Contribution, Funding Acquisition, Writing – Review Editing; Kimberly J. Hemmerling: Data Contribution, Mahdi Mobarak-Abadi: Data Contribution; Mark A. Hoggarth:Data Contribution; Matthew A. Howard: Data Contribution; Molly G. Bright: Data Contribution; Nawal Kinany: Data Contribution; Olivia S. Kowalczyk: Data Contribution; Patrick Freund: Data Contribution; Robert L. Barry: Data Contribution, Funding Acquisition, Writing – Review Editing; Sean Mackey: Funding Acquisition, Data Contribution; Simon Schading: Data Contribution; Shahabeddin Vahdat: Data Contribution; Todd Parish: Data Contribution; Yufen Chen: Data Contribution; Véronique Marchand-Pauvert: Data Contribution; Zachary A. Smith: Data Contribution

Funding

This study was funded by the Canada Research Chair in Quantitative Magnetic Resonance Imaging [CRC-2020-00179], the Canadian Institute of Health Research [PJT-190258], the Canada Foundation for Innovation [32454, 34824], the Fonds de Recherche du Québec - Santé [322736, 324636], the Natural Sciences and Engineering Research Council of Canada [RGPIN-2019-07244], the Canada First Research Excellence Fund (IVADO and TransMedTech), the Quebec BioImaging Network [5886, 35450], INSPIRED (Spinal Research, UK; Wings for Life, Austria; Craig H. Neilsen Foundation, USA), Mila - Tech Transfer Funding Program, the National Institute of Neurological Disorders and Stroke (K23NS104211, R01NS109450, K24NS126781, R01NS133305, and R01NS128478), the National Institute on Drug Abuse (K99/R00DA040154), the National Center for Complementary and Integrative Health (F32AT007800), and the National Institute of Biomedical Imaging and Bioengineering (R01EB027779 and R21EB031211), the Swiss National Science Foundation (SNSF; 32003B_204934). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Declaration of Competing Interests

Since January 2024, Dr. Barry has been employed by the National Institute of Biomedical Imaging and Bioengineering at the National Institutes of Health. This work was co-authored by Robert Barry in his personal capacity. The opinions expressed in this study are his own and do not necessarily reflect the views of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

The other authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements

We would like to thank Jan Valosek for their valuable feedback and review of the manuscript. We would also like to thank Joshua Newton and Mathieu Guay-Paquet for helping us integrate our EPISeg to SCT. We express our gratitude towards the members of the NeuroPoly Lab at Polytechnique Montreal for all the fruitful discussions.

References

Azzarito, M., Kyathanahally, S. P., Balbastre, Y., Seif, M., Blaiotta, C., Callaghan, M. F., Ashburner, J., & Freund, P. (2020). Simultaneous voxel-wise analysis of brain and spinal cord morphometry and microstructure within the spm framework. *Human Brain Mapping*, *42*, 220–232. https://api.semanticscholar.org/CorpusID:222157395

Barry, R. L., Smith, S. A., Dula, A. N., & Gore, J. C. (2014). Resting state functional connectivity in the human spinal cord. *eLife*, *3*. https://api.semanticscholar.org/CorpusID:6070136

Bédard, S., Karthik, E. N., Tsagkas, C., Pravatà, E., Granziera, C., Smith, A. C., Weber, K. A., & Cohen-Adad, J. (2023). Towards contrast-agnostic soft segmentation of the spinal cord. *ArXiv*, *abs/2310.15402*. https://api.semanticscholar.org/CorpusID:264438952

Blanc, C., Shahrampour, S., Mohamed, F. B., & Leener, B. D. (2023). Combining propseg and a convolutional neural network for automatic spinal cord segmentation in pediatric populations and patients with spinal cord injury. *International Journal of Imaging Systems and Technology*, *33*, 1396–1405. https://api.semanticscholar.org/CorpusID:256607698

Bosma, R. L., & Stroman, P. W. (2014). Assessment of data acquisition parameters, and analysis techniques for noise reduction in spinal cord fMRI data. *Magnetic Resonance Imaging*, *32*(5), 473–481. https://doi.org/10.1016/j.mri.2014.01.007

Chen, L. M., Mishra, A., Yang, P.-F., Wang, F., & Gore, J. C. (2015). Injury alters intrinsic functional connectivity within the primate spinal cord. *Proceedings of the National Academy of Sciences*, *112*, 5991–5996. https://api.semanticscholar.org/CorpusID:20499101

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3d u-net: Learning dense volumetric segmentation from sparse annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. https://api.semanticscholar.org/CorpusID:2164893

Cohen-Adad, J. (2017). Functional magnetic resonance imaging of the spinal cord: Current status and future developments. *Seminars in ultrasound, CT, and MR*, *38 2*, 176–186. https://api.semanticscholar.org/CorpusID:40687893

Conrad, B. N., Barry, R. L., Rogers, B. P., Maki, S., Mishra, A., Thukral, S., Sriram, S., Bhatia, A., Pawate, S., Gore, J. C., & Smith, S. A. (2018). Multiple sclerosis lesions affect intrinsic functional connectivity of the spinal cord. *Brain*, *141*, 1650–1664. https://api.semanticscholar.org/CorpusID:4931193

Dabbagh, A., Horn, U., Kaptan, M., Mildner, T., Müller, R., Lepsien, J., Weiskopf, N., Brooks, J. C., Finsterbusch, J., & Eippert, F. (2024). Reliability of task-based fMRI in the dorsal horn of the human spinal cord. *Imaging Neuroscience*, *2*, 1–27. https://doi.org/10.1162/imag_a_00273

De Leener, B., Fonov, V. S., Collins, D. L., Callot, V., Stikov, N., & Cohen-Adad, J. (2018). Pam50: Unbiased multimodal template of the brainstem and spinal cord aligned with the icbm152 space. *NeuroImage*, *165*, 170–179. https://api.semanticscholar.org/CorpusID:7337168

De Leener, B., Lévy, S., Dupont, S. M., Fonov, V. S., Stikov, N., Collins, D. L., Callot, V., & Cohen-Adad, J. (2017). Sct: Spinal cord toolbox, an open-source software for processing spinal cord mri data. *NeuroImage*, *145*, 24–43. https://api.semanticscholar.org/CorpusID:10981848

De Leener, B., Ullman, E., Kadoury, S., & Cohen-Adad, J. (2014). Propseg: Automatic spinal cord segmentation method for mr images using propagated deformable models. *F1000Research*, *5*. https://api.semanticscholar.org/CorpusID:74119587

Finsterbusch, J., Eippert, F., & Büchel, C. (2012). Single, slice-specific z-shim gradient pulses improve t2*-weighted imaging of the spinal cord. *NeuroImage*, *59*, 2307–2315. https://api.semanticscholar.org/CorpusID:5259447

Fonov, V. S., Troter, A. L., Taso, M., Leener, B. D., Lévêque, G., Benhamou, M., Sdika, M., Benali, H., Pradat, P.-F., Collins, D. L., Callot, V., & Cohen-Adad, J. (2014). Framework for integrated mri average of the spinal cord white and gray matter: The mni–poly–amu template. *NeuroImage*, *102*, 817–827. https://api.semanticscholar.org/CorpusID:1577725

Fratini, M., Moraschi, M., Maraviglia, B., & Giove, F. (2014). On the impact of physiological noise in spinal cord functional mri. *Journal of Magnetic Resonance Imaging*, *40*. https://api.semanticscholar.org/CorpusID:2063624

Giove, F., Garreffa, G., Giulietti, G., Mangia, S., Colonnese, C., & Maraviglia, B. (2004). Issues about the fmri of the human spinal cord. *Magnetic resonance imaging*, *22 10*, 1505–16. https://api.semanticscholar.org/CorpusID:9559516

Giulietti, G., Summers, P. E., Ferraro, D., Porro, C. A., Maraviglia, B., & Giove, F. (2011). Semiautomated segmentation of the human spine based on echoplanar images. *Magnetic resonance imaging*, *29 10*, 1429–36. https://api.semanticscholar.org/CorpusID:2224041

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D. B., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., . . . Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*. https://api.semanticscholar.org/CorpusID:795781

Gros, C., Leener, B. D., Badji, A., Maranzano, J., Eden, D., Dupont, S. M., Talbott, J. F., Zhuoquiong, R., Liu, Y., Granberg, T., Ouellette, R., Tachibana, Y., Hori, M., Kamiya, K., Chougar, L., Stawiarz, L., Hillert, J., Bannier, E., Kerbrat, A., . . . Cohen-Adad, J. (2018). Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *NeuroImage*, *184*, 901–915. https://api.semanticscholar.org/CorpusID:21699394

Haynes, G., Muhammad, F. Y., Khan, A. F., Mohammadi, E., Smith, Z. A., & Ding, L. (2023). The current state of spinal cord functional magnetic resonance imaging and its application in clinical research. *Journal of Neuroimaging*, *33*, 877–888. https://api.semanticscholar.org/CorpusID:262147906

Hemmerling, K. J., Hoggarth, M. A., Sandhu, M. S., Parrish, T. B., & Bright, M. G. (2023). Spatial distribution of hand-grasp motor task activity in spinal cord functional magnetic resonance imaging. *Human Brain Mapping*, *44*, 5567–5581. https://api.semanticscholar.org/CorpusID:261073944

Hoggarth, M. A., Wang, M. C., Hemmerling, K. J., Vigotsky, A. D., Smith, Z. A., Parrish, T. B., Weber, K. A., & Bright, M. G. (2022). Effects of variability in manually contoured spinal cord masks on fmri co-registration and interpretation. *Frontiers in Neurology*, *13*. https://api.semanticscholar.org/CorpusID:247780741

Hohenhaus, M., Klingler, J.-H., Scholz, C., Watzlawick, R., Hubbe, U., Beck, J., Reisert, M., Würtemberger, U., Kremers, N., & Wolf, K. (2024). Quantification of cervical spinal stenosis by automated 3D MRI segmentation of spinal cord and cerebrospinal fluid space. *Spinal Cord*.

Horsfield, M. A., Sala, S., Neema, M., Absinta, M., Bakshi, A., Sormani, M. P., Rocca, M. A., Bakshi, R., & Filippi, M. (2010). Rapid semi-automatic segmentation of the spinal cord from magnetic resonance images: Application in multiple sclerosis. *NeuroImage*, *50*, 446–455. https://api.semanticscholar.org/CorpusID:34258683

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. (2020). Nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, *18*, 203–211. https://api.semanticscholar.org/CorpusID:227947847

Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K. H., & Jaeger, P. F. (2024). Nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. *ArXiv*, *abs/2404.09556*. https://api.semanticscholar.org/CorpusID:269148560

Kaptan, M., Horn, U., Vannesjo, S. J., Mildner, T., Weiskopf, N., Finsterbusch, J., Brooks, J. C. W., & Eippert, F. (2023). Reliability of resting-state functional connectivity in the human spinal cord: Assessing the impact of distinct noise sources. *Neuroimage*, *275*. https://api.semanticscholar.org/CorpusID:255126115

Kaptan, M., Law, C. S. W., Weber, K. A., Pfyffer, D., Zhang, X., Maronesy, T., Glover, G., & Mackey, S. C. (2023). Exploring corticospinal functional connectome using resting-state functional magnetic resonance imaging. *The Journal of Pain*. https://api.semanticscholar.org/CorpusID:257974875

Kaptan, M., Vannesjo, S. J., Mildner, T., Horn, U., Hartley-Davies, R., Oliva, V., Brooks, J. C., Weiskopf, N., Finsterbusch, J., & Eippert, F. (2022). Automated slice-specific z-shimming for functional magnetic resonance imaging of the human spinal cord. *Human Brain Mapping*, *43*, 5389–5407. https://api.semanticscholar.org/CorpusID:251399718

Karthik, E. N., Valosek, J., Smith, A. C., Pfyffer, D., Schading-Sassenhausen, S., Farner, L., Weber, K. A., Freund, P., & Cohen-Adad, J. (2024). Sciseg: Automatic segmentation of t2-weighted intramedullary lesions in spinal cord injury. *medRxiv*. https://api.semanticscholar.org/CorpusID:266743496

Khatibi, A., Vahdat, S., Lungu, O. V., Finsterbusch, J., Büchel, C., Cohen-Adad, J., Marchand-Pauvert, V., & Doyon, J. (2022). Brain-spinal cord interaction in long-term motor sequence learning in human: An fmri study. *NeuroImage*, *253*. https://api.semanticscholar.org/CorpusID:247586476

Kinany, N., Pirondini, E., Martuzzi, R., Mattera, L., Micera, S., & Van De Ville, D. (2019). Functional imaging of rostrocaudal spinal activity during upper limb motor tasks. *NeuroImage*, *200*, 590–600. https://api.semanticscholar.org/CorpusID:156056105

Kinany, N., Pirondini, E., Micera, S., & Van De Ville, D. (2020). Dynamic functional connectivity of resting-state spinal cord fmri reveals fine-grained intrinsic architecture. *Neuron*, *108*, 424–435.e4. https://api.semanticscholar.org/CorpusID:221567643

Kinany, N., Pirondini, E., Micera, S., & Van De Ville, D. (2022). Spinal cord fmri: A new window into the central nervous system. *The Neuroscientist*, *29*, 715–731. https://api.semanticscholar.org/CorpusID:250489772

Kowalczyk, O. S., Medina, S., Tsivaka, D., McMahon, S. B., Williams, S. C. R., Brooks, J. C., Lythgoe, D. J., & Howard, M. A. (2024). Spinal fmri demonstrates segmental organisation of functionally connected networks in the cervical spinal cord: A test–retest reliability study. *Human Brain Mapping*, *45*. https://api.semanticscholar.org/CorpusID:267380960

Kündig, C. W., Finsterbusch, J., Freund, P., & David, G. (2024). Functional magnetic resonance imaging of the lumbosacral cord during a lower extremity motor task. *Imaging Neuroscience*, *2*, 1–19. https://doi.org/10.1162/imag_a_00227

Landelle, C., Dahlberg, L. S., Lungu, O. V., Miić, B., Leener, B. D., & Doyon, J. (2023). Altered spinal cord functional connectivity associated with parkinson's disease progression. *Movement Disorders*, *38*. https://api.semanticscholar.org/CorpusID:257077183

Lukas, C., Bellenberg, B., Prados, F., Valsasina, P., Parmar, K., Brouwer, I., Pareto, D., Rovira, À., Sastre-Garriga, J., Gandini Wheeler-Kingshott, C. A. M., Kappos, L., Rocca, M. A., Filippi, M., Yiannakas, M., Barkhof, F., & Vrenken, H. (2021). Quantification of cervical cord cross-sectional area: Which acquisition, vertebra level, and analysis software? a multicenter repeatability study on a traveling healthy volunteer. *Front. Neurol.*, *12*, 693333.

Maier-Hein, L., Reinke, A., Christodoulou, E., Glocker, B., Godau, P., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M., Wiesenfarth, M., Baumgartner, M., Eisenmann, M., Heckmann-Notzel, D., Kavur, A. E., Radsch, T., Tizabi, M. D., Ación, L., Antonelli, M., ... Jager, P. F. (2022). Metrics reloaded: Recommendations for image analysis validation. *Nature methods*. https://api.semanticscholar.org/CorpusID:249375598

Martucci, K. T., Weber, K. A., & Mackey, S. C. (2019). Altered cervical spinal cord resting-state activity in fibromyalgia. *Arthritis & Rheumatology*, *71*. https://api.semanticscholar.org/CorpusID:52911043

Martucci, K. T., Weber, K. A., & Mackey, S. C. (2021). Spinal cord resting state activity in individuals with fibromyalgia who take opioids. *Frontiers in Neurology*, *12*. https://api.semanticscholar.org/CorpusID:236900175

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation*, *24*, 109–165. https://api.semanticscholar.org/CorpusID:61019113

Mobarak-Abadi, M., Mahmoudi-Aznaveh, A., Dehghani, H., Zarei, M., Vahdat, S., Doyon, J., & Khatibi, A. (2024). DeepRetroMoCo: Deep neural network-based retrospective motion correction algorithm for spinal cord functional MRI [Publisher: Frontiers]. *Frontiers in Psychiatry*, *15*. https://doi.org/10.3389/fpsyt.2024.1323109

Powers, J. M., Ioachim, G., & Stroman, P. W. (2018). Ten key insights into the use of spinal cord fmri. *Brain Sciences*, *8*. https://api.semanticscholar.org/CorpusID:52185170

Sengupta, A., Mishra, A., Wang, F., Chen, L. M., & Gore, J. C. (2024). Characteristic bold signals are detectable in white matter of the spinal cord at rest and after a stimulus. *Proceedings of the National Academy of Sciences of the United States of America*, *121 22*, e2316117121. https://api.semanticscholar.org/CorpusID:269974582

Staud, R., Boissoneault, J., Lai, S., Mejia, M. S., Ramanlal, R., Godfrey, M. M., & Stroman, P. W. (2021). Spinal cord neural activity of fibromyalgia patients and healthy controls during temporal summation of pain: An fmri study. *Journal of neurophysiology*. https://api.semanticscholar.org/CorpusID:237214626

Stehling, M. K., Turner, R., & Mansfield, P. (1991). Echo-planar imaging: Magnetic resonance imaging in a fraction of a second. *Science*, *254 5028*, 43–50. https://api.semanticscholar.org/CorpusID:26600241

Stroman, P. W., Figley, C. R., & Cahill, C. M. (2008). Spatial normalization, bulk motion correction and coregistration for functional magnetic resonance imaging of the human cervical spinal cord and brainstem. *Magnetic resonance imaging*, *26 6*, 809–14. https://api.semanticscholar.org/CorpusID:207027755

Stroman, P. W., Wheeler-Kingshott, C. A. M., Bacon, M., Schwab, J. M., Bosma, R. L., Brooks, J. C. W., Cadotte, D. W. W., Carlstedt, T. P., Ciccarelli, O., Cohen-Adad, J., Curt, A., Evangelou, N., Fehlings, M. G., Filippi, M., Kelley, B. J., Kollias, S. S., MacKay, A. L., Porro, C. A., Smith, S. A., . . . Tracey, I. (2014). The current state-of-the-art of spinal cord imaging: Methods. *NeuroImage*, *84*, 1070–1081. https://api.semanticscholar.org/CorpusID:14443462

Taso, M., Troter, A. L., Sdika, M., Ranjeva, J.-P., Guye, M., Bernard, M., & Callot, V. (2014). Construction of an in vivo human spinal cord atlas based on high-resolution mr images at cervical and thoracic levels: Preliminary results. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *27*, 257–267. https://api.semanticscholar.org/CorpusID:24635366

Tinnermann, A., Büchel, C., & Haaker, J. (2021). Observation of others' painful heat stimulation involves responses in the spinal cord. *Science Advances*, *7*. https://api.semanticscholar.org/CorpusID:232480874

Tinnermann, A., Sprenger, C., & Büchel, C. (2022). Opioid analgesia alters corticospinal coupling along the descending pain system in healthy participants. *eLife*, *11*. https://api.semanticscholar.org/CorpusID:248390204

Toufani, H., Vard, A., & Adibi, I. (2021). A pipeline to quantify spinal cord atrophy with deep learning: Application to differentiation of MS and NMOSD patients. *Phys. Med.*, *89*, 51–62.

Tsagkas, C., Horvath-Huck, A., Haas, T., Amann, M., Todea, A., Altermatt, A., Müller, J., Cagol, A., Leimbacher, M., Barakovic, M., Weigel, M., Pezold, S., Sprenger, T., Kappos, L., Bieri, O.,

Granziera, C., Cattin, P., & Parmar, K. (2023). Fully automatic method for reliable spinal cord compartment segmentation in multiple sclerosis. *AJNR Am. J. Neuroradiol.*, *44*(2), 218–227.

Vahdat, S., Lungu, O. V., Cohen-Adad, J., Marchand-Pauvert, V., Benali, H., & Doyon, J. (2015). Simultaneous brain–cervical cord fmri reveals intrinsic spinal cord plasticity during motor sequence learning. *PLoS Biology*, *13*. https://api.semanticscholar.org/CorpusID:14823503

Valosek, J., & Cohen-Adad, J. (2024). Reproducible spinal cord quantitative mri analysis with the spinal cord toolbox. *Magnetic Resonance in Medical Sciences*, *23*, 307–315. https://api.semanticscholar.org/CorpusID:268382554

Weber, K., Chen, Y., Wang, X., Kahnt, T., & Parrish, T. (2016). Functional magnetic resonance imaging of the cervical spinal cord during thermal stimulation across consecutive runs. *NeuroImage*, *143*. https://doi.org/10.1016/j.neuroimage.2016.09.015

Weber, K. A., Chen, Y. J., Paliwal, M., Law, C. S. W., Hopkins, B. S., Mackey, S. C., Dhaher, Y. Y., Parrish, T. B., & Smith, Z. A. (2020). Assessing the spatial distribution of cervical spinal cord activity during tactile stimulation of the upper extremity in humans with functional magnetic resonance imaging. *NeuroImage*, *217*, 116905–116905. https://api.semanticscholar.org/CorpusID:218516504

Weber, K. A., Chen, Y. J., Wang, X., Kahnt, T., & Parrish, T. B. (2016). Lateralization of cervical spinal cord activity during an isometric upper extremity motor task with functional magnetic resonance imaging. *NeuroImage*, *125*, 233–243. https://api.semanticscholar.org/CorpusID:29015234

Weber, K. A., Sentis, A. I., Bernadel-Huey, O. N., Chen, Y. J., Wang, X., Parrish, T. B., & Mackey, S. C. (2018). Thermal stimulation alters cervical spinal cord functional connectivity in humans. *Neuroscience*, *369*, 40–50. https://api.semanticscholar.org/CorpusID:23340290

Wu, T.-L., Yang, P.-F., Wang, F., Shi, Z., Mishra, A., Wu, R., Chen, L. M., & Gore, J. C. (2019). Intrinsic functional architecture of the non-human primate spinal cord derived from fmri and electrophysiology. *Nature Communications*, *10*. https://api.semanticscholar.org/CorpusID:85567049