NeuroImage 152 (2017) 312–329

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

# Spinal cord grey matter segmentation challenge

Ferran Prados[a,b,*], John Ashburner[m], Claudia Blaiotta[m], Tom Brosch[h], Julio Carballido-Gamio[p], Manuel Jorge Cardoso[a,d], Benjamin N. Conrad[q], Esha Datta[p], Gergely Dávid[n], Benjamin De Leener[e], Sara M. Dupont[e], Patrick Freund[n], Claudia A.M. Gandini Wheeler-Kingshott[b,c,r], Francesco Grussu[b], Roland Henry[p], Bennett A. Landman[q], Emil Ljungberg[g], Bailey Lyttle[l], Sebastien Ourselin[a,d], Nico Papinutto[p], Salvatore Saporito[o], Regina Schlaeger[p], Seth A. Smith[k], Paul Summers[j], Roger Tam[i], Marios C. Yiannakas[b], Alyssa Zhu[p], Julien Cohen-Adad[e,f]

[a] *Translational Imaging Group, Centre for Medical Image Computing (CMIC), Department of Medical Physics and Bioengineering, University College London, Malet Place Engineering Building, London WC1E 6BT, UK*
[b] *NMR Research Unit, Queen Square MS Centre, Department of Neuroinflammation, UCL Institute of Neurology, University College London, Russell Square, London WC1B 5EH, UK*
[c] *Brain MRI 3T Centre, C. Mondino National Neurological Institute, Pavia, Italy*
[d] *Dementia Research Centre, Department of Neurodegenerative Disease, UCL Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK*
[e] *NeuroPoly Lab, Polytechnique Montreal, Montreal, QC, Canada*
[f] *Functional Neuroimaging Unit, CRIUGM, Université de Montréal, Montreal, QC, Canada*
[g] *Department of Medicine, University of British Columbia, Vancouver, BC, Canada V6T 2B5*
[h] *Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4*
[i] *Department of Radiology, UBC MS/MRI Research Group, University of British Columbia, Vancouver, BC, Canada V6T 2B5*
[j] *Department of Radiology, European Institute of Oncology, University of Modena and Reggio Emilia, 41121, Modena, MO, Italy*
[k] *Department of Radiology and Radiological Sciences, Biomedical Engineering, Ophthalmology, Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA*
[l] *Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA*
[m] *Wellcome Trust Centre for Neuroimaging, University College London, Queen Square, London WC1N 3BG, UK*
[n] *Spinal Cord Injury Center Balgrist, University Hospital Zurich, University of Zurich, Switzerland*
[o] *Eindhoven University of Technology, Netherlands*
[p] *Department of Neurology, University of California San Francisco, San Francisco, CA, USA*
[q] *Department of Electrical Engineering, Computer Science, Biomedical Engineering, Radiology and Radiological Sciences, Institute of Image Science at Vanderbilt University, Nashville, TN, USA*
[r] *Department of Brain and Behavioural Sciences, University of Pavia, Italy*

## ARTICLE INFO

*Keywords:*
Spinal cord
Grey matter
Segmentation
MRI
Challenge
Evaluation metrics

## ABSTRACT

An important image processing step in spinal cord magnetic resonance imaging is the ability to reliably and accurately segment grey and white matter for tissue specific analysis. There are several semi- or fully-automated segmentation methods for cervical cord cross-sectional area measurement with an excellent performance close or equal to the manual segmentation. However, grey matter segmentation is still challenging due to small cross-sectional size and shape, and active research is being conducted by several groups around the world in this field. Therefore a grey matter spinal cord segmentation challenge was organised to test different capabilities of various methods using the same multi-centre and multi-vendor dataset acquired with distinct 3D gradient-echo sequences. This challenge aimed to characterize the state-of-the-art in the field as well as identifying new opportunities for future improvements. Six different spinal cord grey matter segmentation methods developed independently by various research groups across the world and their performance were compared to manual segmentation outcomes, the present gold-standard. All algorithms provided good overall results for detecting the grey matter butterfly, albeit with variable performance in certain quality-of-segmentation metrics. The data have been made publicly available and the challenge web site remains open to new submissions. No

* Corresponding author at: Translational Imaging Group, Centre for Medical Image Computing (CMIC), Department of Medical Physics and Bioengineering, University College London, Malet Place Engineering Building, London WC1E 6BT, UK.
*E-mail addresses:* f.carrasco@ucl.ac.uk (F. Prados), jcohen@polymtl.ca (J. Cohen-Adad).

http://dx.doi.org/10.1016/j.neuroimage.2017.03.010
Received 25 November 2016; Accepted 6 March 2017
Available online 07 March 2017

modifications were introduced to any of the presented methods as a result of this challenge for the purposes of this publication.

## Introduction

A large spectrum of (non)-traumatic neurological disorders have been linked with spinal cord grey matter (GM) and white matter (WM) tissue changes (Amukotuwa and Cook, 2015). The spinal cord is a challenging area for magnetic resonance imaging (MRI) (Wheeler-Kingshott et al., 2014; Stroman et al., 2014) due to the small cross-sectional area dimension of the spinal cord, the presence of motion, susceptibility artifacts and, in particular, the complex shape and small area fraction of GM tissue. Recently, Yiannakas et al. (2012) demonstrated the feasibility to distinguish between the WM and GM by performing manual segmentation of the cervical cord using a T1-weighted fast field echo (FFE) data acquired in a 3 T scanner with reasonable acquisition times and an in-plane resolution of $0.5\times0.5$ mm$^2$. More recently, Schlaeger et al. (2014, 2015) also demonstrated that spinal cord GM area was the strongest correlate of disability in multiple sclerosis using multivariate models that included brain GM and WM volumes, fluid-attenuated inversion recovery lesion load, T1 lesion load, spinal cord cross-sectional area (CSA), T2 lesion load, age, sex, and disease duration.

Several semi- or fully-automated segmentation methods have been proposed in the last decade for cervical CSA estimation (Losseff et al., 1996; Hickman et al., 2004; Tench et al., 2005; Zivadinov et al., 2008; Horsfield et al., 2010; McIntosh et al., 2011; Bergo et al., 2012; Chen et al., 2013; de Leener et al., 2014, 2016; Asman and Bryan, 2014; Taso et al., 2015; El Mendili et al., 2015). While most methods present good performance, interpretation and comparison of results between different methods is seldom possible due to the use of different imaging datasets (usually in-house data), different MRI sequences, different ways to obtain gold standard segmentations (number of raters and consensus mask) and the use of various performance scores (2D/slice-wise or 3D/volumetric). Recent cervical cord CSA segmentation methods have reached a performance close to manual segmentation (de Leener et al., 2014; Asman and Bryan, 2014; El Mendili et al., 2015), but accurate GM segmentation remains a challenge. Moreover, there is a lack of publicly available datasets with GM/WM contrast and corresponding ground truth that facilitate a fair and reliable comparison across methods.

A GM spinal cord segmentation challenge was organised in conjunction by four internationally recognised spinal cord imaging research groups (University College London, Polytechnique Montreal, University of Zurich and Vanderbilt University) to test the different performances of various methods, with the aim of characterizing the state-of-the-art in the field according to a pre-defined set of assessment criteria as well as identifying opportunities for future improvement. Several GM spinal cord segmentation methods developed independently by various research groups across the world were compared. These methods were used to segment the same multi-centre and multi-vendor dataset acquired with distinct 3D gradient-echo sequences, which are available to the community at http://cmictig.cs.ucl.ac.uk/niftyweb/challenge, and the obtained results were compared to the manual segmentation performed by 4 raters.

## Material

Participating teams applied their automatic or semi-automatic segmentation algorithms to anatomical MR images of 40 healthy spinal cords. Challenge data was composed by 80 datasets, split in 40 training and 40 test datasets, 20 each acquired at 4 different sites (University College London, Polytechnique Montreal, University of Zurich and Vanderbilt University). See Table 1 for demographic data. Algorithms were evaluated against manual segmentations from four trained raters (one from each site who each analysed all data from all sites) in terms of segmentation accuracy and precision using several validation metrics.

### *Data*

A multi-centre, multi-vendor dataset of spinal cord anatomical images of healthy subjects was provided. Each site provided images from 20 healthy subjects along with WM/GM manual segmentation masks. The acquisition parameters for each site were the following:

- Site 1, University College London. Acquisition was performed using a 3 T Philips Achieva MRI system with dual-transmit technology enabled for all scans (Philips Healthcare, Best, Netherlands) and the manufacturer's product 16-channel neurovascular coil. All participants were immobilised using a MRI-compatible cervical collar (TalarMade Ltd, Chesterfield, UK). The cervical cord was imaged in the axial-oblique plane (i.e. slices perpendicular to the longitudinal axis of the cord) with the center of the imaging volume positioned at the level of C2-3 intervertebral disc. The MRI acquisition parameters were: fat-suppressed 3D slab-selective fast field echo (3D-FFE) with time of repetition (TR)=23 ms; time of echo (TE)=5 ms, flip angle $\alpha$=7°, field-of-view (FOV) =240×180 mm$^2$, voxel size=0.5×0.5×5 mm$^3$, NEX=8, 10 axial contiguous slices, scanning time 13:34 min. A 15 mm section of the high-resolution 3D-FFE volumetric scan (i.e. 3 slices) was extracted, with the middle slice passing through the C2/C3 intervertebral disc.
- Site 2, Polytechnique Montreal. Acquisition was performed using a 3 T Siemens TIM Trio, with the body coil used for RF transmission and the 12 channels head coil+4 channels neck coil for RF reception. All participants were immobilised with padding. Axial 2D spoiled gradient echo, TR=539 ms, TE=5.41, 12.56 and 19.16 ms (averaged off-line to create a single image with increased SNR), flip angle $\alpha$=35°, readout bandwidth (BW)=200 Hz per pixel, voxel size=0.5×0.5×5 mm$^3$, 10 slices, matrix size of 320×320, R=2 acceleration along RL direction with GRAPPA reconstruction, phase stabilization. Scanning time 4:38 min.
- Site 3, University of Zurich. Scanning was performed on a 3 T Siemens Skyra MRI scanner (Siemens Healthcare, Erlangen, Germany) using a 16-channel radio-frequency receive head and neck coil and radio-frequency body transmit coil. All participants wore an MRI-compatible neck collar (Laerdal Medicals, Stavanger, Norway). A 3D high-resolution optimized T2*-weighted multi-echo sequence (multiple echo data image combination; MEDIC) was applied to acquire five high-resolution 3D volumes of the cervical cord at C2/C3 level. Each volume consisted of twenty contiguous slices acquired in the axial-oblique plane and was obtained with a resolution of 0.5×0.5×2.5 mm$^3$ within 2:08 min for each of the five volumes. Following parameters were applied: TE=19 ms, TR=44 ms, FOV=192×162 mm$^2$, matrix size=384×324, flip angle

**Table 1**
Demographic data per site, first row: number of healthy controls per site, second row: gender - female (F):male (M); third row: mean age in years. Std: standard deviation.

| | Site 1 – UCL | Site 2 – Montreal | Site 3 – Zurich | Site 4 – Vanderbilt |
|---|---|---|---|---|
| *Subjects* | 20 | 20 | 20 | 20 |
| *Gender* | 14F:6M | 11F:9M | 6F:14M | 7F:13M |
| *Mean Age (Std)* | 44.3 (10.4) | 33.7 (17.4) | 40.6 (10.4) | 28.3 (8.2) |