

A convolutional neural network-based framework for imaging biomarkers in multiple sclerosis: white matter hyperintensity and brain region volumes



Kinnunen KM¹, De Blasi B¹, Parker R¹, Peraza L¹, Wolz R^{1,2}.
1. IXICO, London, UK | 2. Imperial College London, London, UK

Introduction

- White matter (WM) lesions are a hallmark of multiple sclerosis (MS) and reflect the inflammatory activity.
- Automatic detection of WM hyperintensities (WMH) from magnetic resonance imaging (MRI) T2-fluid attenuated inversion recovery (FLAIR) scans can support MS diagnosis and help to monitor treatment effectiveness in clinical trials.
- Brain volumetric measures can be used to study the neurodegenerative component of MS, to predict disability progression, and to evaluate potential anti-inflammatory, remyelinating or neuroprotective therapies.¹⁻³
- IXICO's IXIQ.Ai is a framework for MRI segmentation workflows.⁴⁻⁶

Objectives/Aims

To validate convolutional neural networks (CNNs) for:

- segmentation of WMH from 3D FLAIR scans
- segmentation of MS-relevant brain regions from 3D T1-weighted (T1W) scans
- volumetric analysis

Methods

- CNNs trained to segment WMH from 3D FLAIR scans and brain regions, incl. whole-brain (WB) and thalamus, from 3D T1W scans
- Jacobian integration used to estimate WB volume change
- Dataset for WMH workflow validation: a publicly available MS dataset⁷ with 3D FLAIR scans, manual segmentations by seven expert raters, and a rater majority voting consensus
- Datasets for brain region workflow validation: ADNI (<https://adni.loni.usc.edu/>), OASIS (<https://www.oasis-brains.org/>), Huntington's disease (two, internal), multiple system atrophy (internal), healthy controls aged 75-86 and 19-25 (<https://brain-development.org/ixi-dataset/>)
- WMH ground truth = majority voting consensus
- Brain region ground truth = semi-automated segmentations (with manual edits)
- Accuracy/robustness from dice score coefficient (DSC), visual assessment, volume correlation/volume error, group separation
- Test-retest performance (ADNI BTB and OASIS datasets)

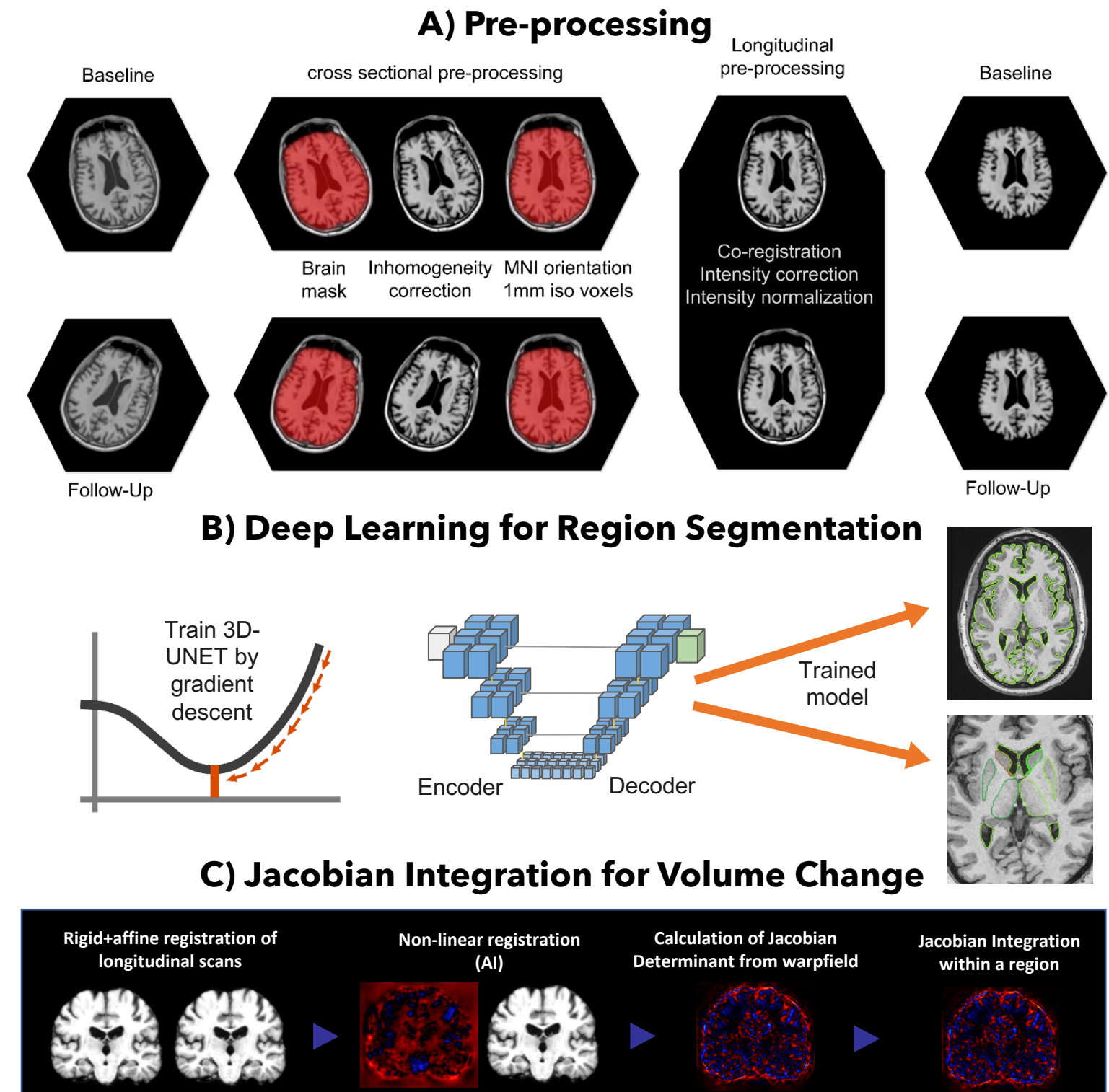


Figure 1: IXIQ.Ai workflows for A) pre-processing for cross-sectional and longitudinal analyses, B) region segmentation, and C) Jacobian Integration for volume change analysis

Results

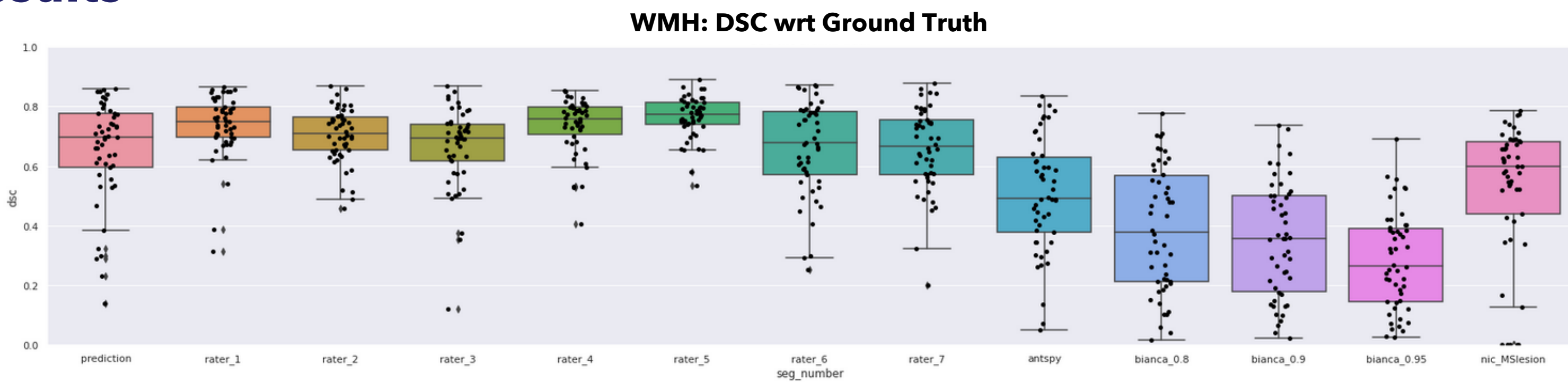


Figure 2: Boxplots showing the distribution of DSCs: IXIQ.Ai WMH workflow (prediction), each individual expert rater (raters 1-7) and other WMH segmentation methods compared to the WMH ground truth (consensus). (antspsy, <https://github.com/ANTsX/ANTs/wiki>; BIANCA, <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BIANCA/Userguide> - lesion probability maps thresholded at 0.8/0.9/0.95; nic_MSlesion, <https://github.com/sergivalverde/nicMSlesions>)

The WMH workflow's ground-truth overlap (mean±SD DSC = 0.65±0.17) was comparable to the individual raters' and higher than for other automated methods (antspsy 0.51±0.19; bianca_0.9 0.35±0.19; nic_MSlesion 0.52±0.23) (also see Figure 2). Its average volume correlation with the ground truth ($r = 0.96$) was higher than for the other methods (antspsy $r = 0.85$; bianca_0.9 $r = 0.69$; nic_MSlesion $r = 0.87$) and its average volume error (AVE) relative to the ground truth (mean±SD AVE, mL = 3.1±3.3) was the lowest (antspsy 6.5±9.1; bianca_0.9 8.3±11.0; nic_MSlesion 5.1±6.9).

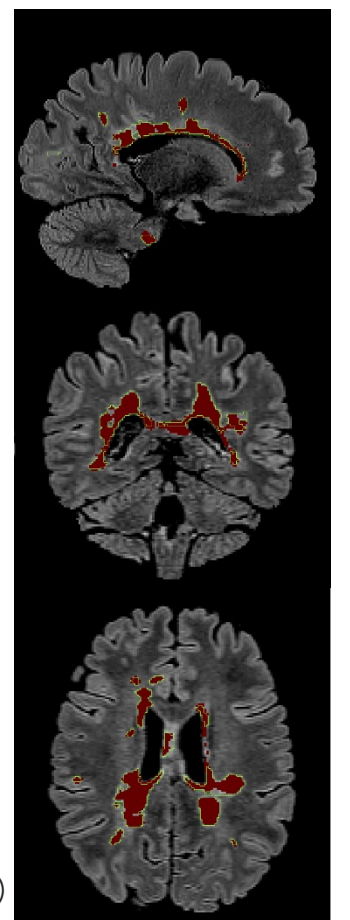


Figure 3: WMH segmentation from 3D FLAIR (yellow = ground truth; red = prediction)

Region Segmentation: WB and Thalamus

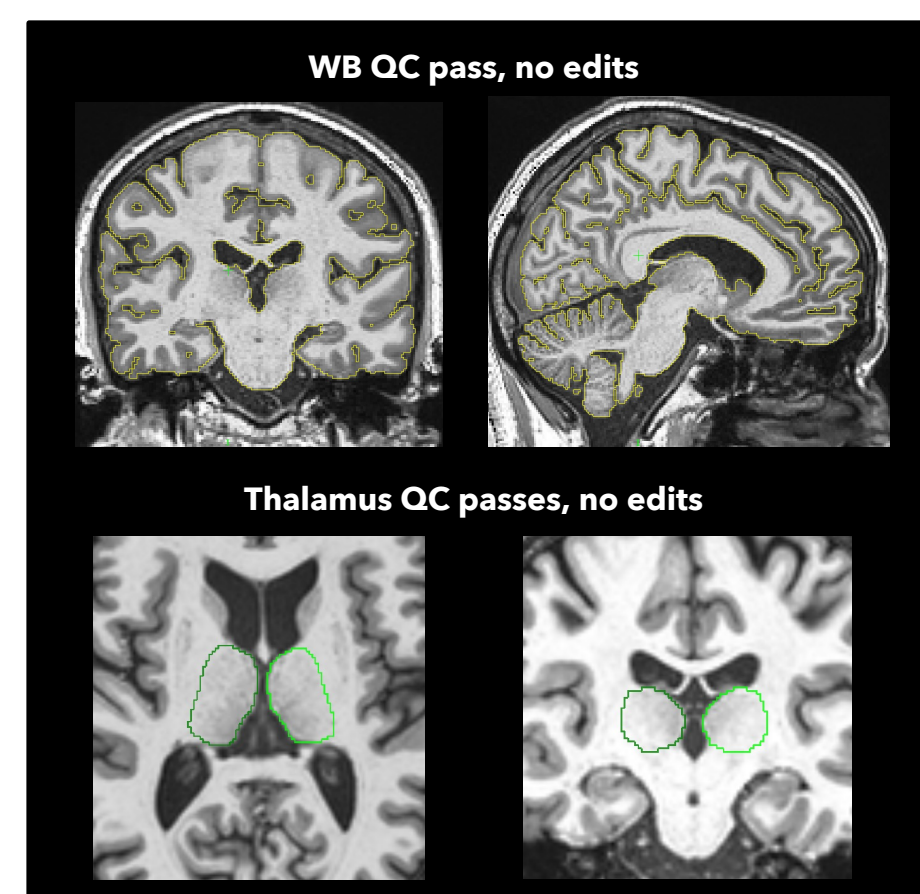


Figure 4: Examples of WB and thalamus segmentations from IXIQ.Ai brain region workflow (CNN) that passed QC without manual edits.

Volume Change: WB

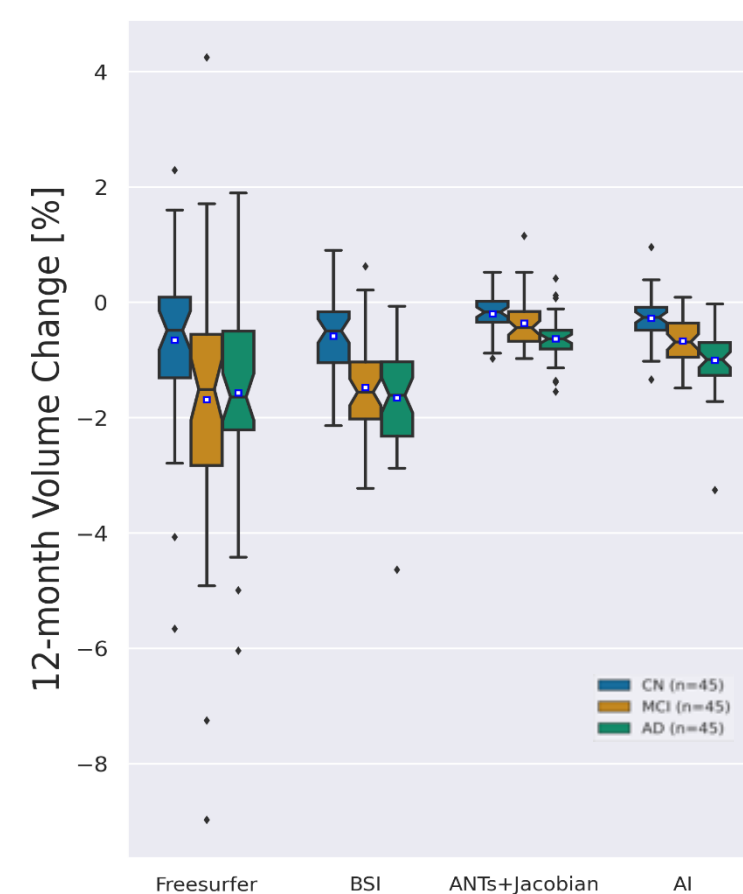


Figure 5: WB volume change (negative change = loss) estimated by IXIQ.Ai (CNN + Jacobian), compared to FreeSurfer⁸, BSI⁹, and ANTS (<https://github.com/ANTsX/ANTs/wiki>) + Jacobian. p-value: Mann-Whitney U-test. CLES: Common Language Effect Size.

CN vs MCI

	FreeSurfer	BSI	ANTs+J	IXIQ.Ai
p-value	0.004	5e-7	0.007	8e-6
CLES	0.67	0.81	0.67	0.78
Median (SD) group difference [%]	1.02 (0.16)	1.06 (0.02)	0.27 (0.01)	0.43 (0.01)

CN vs AD

	FreeSurfer	BSI	ANTs+J	IXIQ.Ai
p-value	0.005	1e-7	6e-8	2e-10
CLES	0.67	0.83	0.84	0.89
Median (SD) group difference [%]	1.16 (0.11)	1.12 (0.03)	0.46 (0.01)	0.73 (0.01)

MCI vs AD

	FreeSurfer	BSI	ANTs+J	IXIQ.Ai
p-value	0.86	0.41	0.009	0.001
CLES	0.48	0.55	0.66	0.70
Median (SD) group difference [%]	0.13 (0.18)	0.05 (0.03)	0.18 (0.01)	0.30 (0.01)

The IXIQ.Ai segmentations (CNN) had high ground-truth overlaps (mean DSC>97). Visual QC found consistently high segmentation quality (see Figure 4): 98% of the whole-brain and 100% of the thalamus segmentations passed QC.

The IXIQ.Ai volume change method (CNN + Jacobian), applied to the WB, detected significant group differences in changes over 12 months, with similar or higher CLES and similar or lower SD of group differences (%) than other methods (Figure 5).

The WB and thalamus volume differences were normally distributed and the distributions zero-centred.

Conclusions

- The WMH workflow shows higher overlap and volume correlation with the ground truth than comparable automated methods.
- The brain region workflow produces high-quality and reliable whole-brain and thalamus regions for volumetric analysis.
- The IXIQ.Ai framework provides a scalable and robust automatic solution for volumetric analysis of brain WMH and regions of interest in MS.

References

- Sastre-Garriga J, Pareto D, Battaglini M, et al. (2020). MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. *Nat Rev Neurol*, 16: 171-182. doi: 10.1038/s41582-020-0314-x
- Miller DH, Barkhof F, Frank JA, Parker GJ, Thompson AJ. (2002). Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance. *Brain*, 125(Pt 8): 1676-1695. doi: 10.1093/brain/awf177
- Valdés Cabrera D, Smyth P, Blevins G, Emery D, Beaulieu C. (2022). Diffusion imaging of fornix and interconnected limbic deep grey matter is linked to cognitive impairment in multiple sclerosis. *Eur J Neurosci*, 55(1): 277-294. doi: 10.1111/ejn.15539
- Weatheritt J, Joles R, Wolz R, Rueckert D. (2020a). Fully automatic AI segmentation of subcortical regions. *Neurother*, 17(SUPPL 1): S21. doi: 10.1007/s13311-020-00945-z
- Weatheritt J, Rueckert D, Wolz R. (2020). Transfer learning for brain segmentation: pre-task selection and data limitations. In: Papiez B, et al. (eds) *Medical Image Understanding and Analysis 2020. Commun Comput Inf Sci*, vol 1248. Springer, Cham. doi: 10.1007/978-3-030-52791-4_10
- Reinwald M, Johnson EB, Scähll RI, Wolz R. (2020). Using artificial intelligence for fast, reliable, and automatic segmentation of the thalamus. *Neurother*, 17(SUPPL 1): S13. doi: 10.1007/s13311-020-00945-z
- Commonwick O, Istace A, Kain M, et al. (2018). Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Sci Rep*, 8(1): 13650. doi: 10.1038/s41598-018-31911-7.
- Reuter M, Schmansky NJ, Rosas HD, Fischl B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage*, 61(4): 1402-1418. doi: 10.1016/j.neuroimage.2012.02.084
- Freeborough PA, Fox NC. (1997). The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans Med Imaging*, 16(5):623-629. doi: 10.1109/42.640753.