RESEARCH ARTICLE

WILEY

Reliable brain morphometry from contrast-enhanced T1w-MRI in patients with multiple sclerosis

Michael Rebsamen[1,2] 🄳 | Richard McKinley[1] | Piotr Radojewski[1,3] |
Maximilian Pistor[4] | Christoph Friedli[4] | Robert Hoepner[4] | Anke Salmen[4] |
Andrew Chan[4] | Mauricio Reyes[5] | Franca Wagner[1] | Roland Wiest[1,3] |
Christian Rummel[1] 🄳

[1]Support Center for Advanced Neuroimaging (SCAN), University Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Inselspital, Bern University Hospital, Bern, Switzerland

[2]Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland

[3]Swiss Institute for Translational and Entrepreneurial Medicine, Bern, Switzerland

[4]Department of Neurology, Inselspital, Bern University Hospital and University of Bern, Bern, Switzerland

[5]ARTORG Center for Biomedical Research, University of Bern, Bern, Switzerland

Correspondence
Michael Rebsamen, Support Center for Advanced Neuroimaging (SCAN), University Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Inselspital, Bern University Hospital, Freiburgstrasse 10, 3010 Bern, Switzerland.
Email: michael.rebsamen@insel.ch

Abstract

Brain morphometry is usually based on non-enhanced (pre-contrast) T1-weighted MRI. However, such dedicated protocols are sometimes missing in clinical examinations. Instead, an image with a contrast agent is often available. Existing tools such as FreeSurfer yield unreliable results when applied to contrast-enhanced (CE) images. Consequently, these acquisitions are excluded from retrospective morphometry studies, which reduces the sample size. We hypothesize that deep learning (DL)-based morphometry methods can extract morphometric measures also from contrast-enhanced MRI. We have extended DL+DiReCT to cope with contrast-enhanced MRI. Training data for our DL-based model were enriched with non-enhanced and CE image pairs from the same session. The segmentations were derived with Free-Surfer from the non-enhanced image and used as ground truth for the coregistered CE image. A longitudinal dataset of patients with multiple sclerosis (MS), comprising relapsing remitting (RRMS) and primary progressive (PPMS) subgroups, was used for the evaluation. Global and regional cortical thickness derived from non-enhanced and CE images were contrasted to results from FreeSurfer. Correlation coefficients of global mean cortical thickness between non-enhanced and CE images were significantly larger with DL+DiReCT ($r = 0.92$) than with FreeSurfer ($r = 0.75$). When comparing the longitudinal atrophy rates between the two MS subgroups, the effect sizes between PPMS and RRMS were higher with DL+DiReCT both for non-enhanced ($d = -0.304$) and CE images ($d = -0.169$) than for FreeSurfer (non-enhanced $d = -0.111$, CE $d = 0.085$). In conclusion, brain morphometry can be derived reliably from contrast-enhanced MRI using DL-based morphometry tools, making additional cases available for analysis and potential future diagnostic morphometry tools.

KEYWORDS
brain morphometry, cortical thickness, deep learning, MRI, post-contrast imaging

1 | INTRODUCTION

Brain morphometry is usually derived from high-resolution, 3D T1-weighted (T1w) MRI, owing to the good gray/white-matter contrast of protocols such as MP-RAGE (van der Kouwe et al., 2008). Morphometry tools such as FreeSurfer (Fischl, 2012) usually expect non-enhanced (pre-contrast) MR images of this type as input data. For diagnostic purposes, T1w images after administration of gadolinium-based contrast agents (GBCAs) are often acquired (Traboulsee et al., 2016), either as a replacement for or in addition to the non-enhanced image. MRI acquired for clinical purposes sometimes include only contrast-enhanced 3D T1w MRI, leaving those data inaccessible for retrospective morphometric analysis. For example, the latest MR imaging guidelines for multiple sclerosis recommend a contrast-enhanced T1w acquisition, but not necessarily a corresponding image without contrast-agent (Wattjes et al., 2021).

Few studies have investigated the influence of contrast agents on morphometry. A significant impact was reported on thalamic volumes (Hannoun et al., 2018) and tissue relaxation times (Warntjes et al., 2014). Good results using FreeSurfer were recently reported for volumes and thickness on a small and pre-selected sample (Lie et al., 2022).

In multiple sclerosis (MS), quantification of the global gray-matter (GM) loss is already used as a secondary outcome measure in clinical trials (Radü et al., 2013; Sastre-Garriga et al., 2020). With the adoption of No evidence of disease activity (NEDA) in its updated version NEDA-4 (Kappos et al., 2016), brain atrophy quantification has become increasingly important for the assessment of individual patients. For research purposes, studying the distribution of MS-typical atrophy patterns over the cortex (like observable in regional GM volume or cortical thickness) in addition to the global GM volume might help improve the understanding of the disease (Eshaghi et al., 2018; Steenwijk et al., 2016). Beyond these univariate approaches, network properties of cortical reorganization due to aging or neurodegenerative processes can be studied by investigating coordinated structural changes of different subregions of the cortex using the concept of structural covariance networks (SCN; Alexander-Bloch et al., 2013; Evans, 2013). In the context of MS, SCNs have been used to study the impact of this reorganization on physical and cognitive impairment (Tewarie et al., 2014), to reveal differences between conversion vs. non-conversion from clinically isolated syndrome to MS (Tur et al., 2018), or to show early longitudinal reorganization processes in patients with relapsing remitting (Fleischer et al., 2019) or primary progressive MS (Tur et al., 2020) as compared to healthy controls. Importantly, none of these findings were noticeable in simpler univariate analyses of global or regional cortical thickness.

Cortical thickness is a frequently used measure in surface-based morphometric analysis (Fischl & Dale, 2000), employed by many of the largest brain morphometry studies (Thompson et al., 2020). Being largely independent of head size (Schwarz et al., 2016) and less correlated to the corresponding volume measures (Winkler et al., 2010) makes it an interesting independent quantitative marker. Its self-explanatory nature is another advantage for potential clinical applications.

The urgent need to make large amounts of additional T1w MRI accessible to morphometric analysis, particularly retrospective analysis of CE images acquired in clinical routine, motivated us to explore the feasibility of deriving global and regional cortical thickness measures from contrast-enhanced MRI. We hypothesized that CE images are a challenge for processing with Freesurfer, resulting in an increased run-time and degraded quality of the reconstructed surface as indicated by the number of surface holes (Monereo-Sánchez et al., 2021).

In this study, we propose an extension for DL+DiReCT (Rebsamen et al., 2020) to cope with contrast-enhanced input data by enriching training data of the deep learning (DL) based segmentation model. For the evaluation, we processed a longitudinal dataset of same-session MR image pairs (non-enhanced and CE T1w) of MS patients with FreeSurfer and DL+DiReCT and assessed the goodness of the cortical thickness measures. Besides correlation coefficients, we analyzed group differences in annual atrophy rates and metrics from structural covariance networks.

2 | MATERIALS AND METHODS

2.1 | MRI dataset

The dataset comprised 454 same-session pairs of non-enhanced and contrast-enhanced (CE), high-resolution T1w-MRI from 75 patients with MS. Among them were nine patients with primary progressive MS (PPMS) and 66 with relapsing remitting MS (RRMS). Detailed demographics are reported in Supplementary Table S1.

Patients with RRMS were all under treatment with Natalizumab (Tysabri, Biogen Corp., Cambridge MA, USA), and therefore underwent regular MRI examinations at an interval of ∼6 months due to their increased risk of progressive multifocal leukoencephalopathy (PML) (Wattjes & Barkhof, 2014). PPMS patients were examined on an annual interval on average. All images were acquired at the Bern University Hospital (Inselspital) on 1.5 T and 3 T scanners from Siemens (Siemens, Erlangen, Germany) with an MP-RAGE protocol (Mugler III & Brookeman, 1990; van der Kouwe et al., 2008). Owing to the retrospective nature of the images acquired in clinical routine, sequence parameters were variable as detailed in Supplementary Figure S1. Contrast-enhanced images were acquired using an MP-RAGE sequence with a water excitation pulse (Norbeck et al., 2020) after administration of a contrast agent. Patients received an intravenous bolus injection of gadobutrol (0.1 ml/kg body weight Gadovist® 1 mmol/ml, Bayer) or, in case of contraindications, gadoterate meglumine (0.2 ml/kg DOTAREM® 0.5 mmol/ml, Guerbet).

2.2 | MRI processing

2.2.1 | DL+DiReCT

DL+DiReCT is an existing morphometry tool to derive global and regional volumes and cortical thickness measures from T1w MRI

(Rebsamen et al., 2020). Its main processing steps include a deep-learning (DL) based segmentation of tissue classes and cortical parcellations followed by a diffeomorphic registration of the GM/WM boundary to the pial surface to derive a voxel-wise cortical thickness map (DiReCT) (Das et al., 2009; Tustison et al., 2013).

A new model for the segmentation was added to DL+DiReCT by enriching the training data with CE images. Initially, DL+DiReCT was trained using a total of 840 non-enhanced T1w MRI, among them 128 images of patients with MS, and auxiliary labels from FreeSurfer, as detailed in Rebsamen et al. (2020). These training data were now enriched with the corresponding 128 same-session CE MRI of these MS patients. Tissue class labels for the CE images were derived from the non-enhanced images by co-registering the image pairs with an affine transformation using *FSL flirt* (Jenkinson et al., 2002; Jenkinson & Smith, 2001) with six degrees of freedom and mutual information as cost function. The resulting transformation matrix was used to resample the CE image into the space of the non-enhanced image using b-spline interpolation. Compared to the initial model with 96 labels, separate labels for WM-hypointensities, left/right inferior horn of the lateral ventricle and choroid plexus were added, resulting in a total of 101 labels. An additional model parcellating the cortex into 74 region per hemisphere according the *Destrieux* atlas (Destrieux et al., 2010) leads to 181 labels. Otherwise, these models were re-trained with identical hyperparameters and network architecture (McKinley et al., 2019).

The tool is publicly available (https://github.com/SCAN-NRAD/DL-DiReCT) including the new models with the "--model v6" option for the *Desikan-Killiany* (Desikan et al., 2006) and "--model v7" for the *Destrieux* atlas (Destrieux et al., 2010). Regional cortical thickness and GM volume estimates for the evaluation were generated from the original T1w images. After brain extraction using HD-BET (Isensee et al., 2019), the new model of DL+DiReCT was applied using default settings (running on one GPU and four CPU cores on Linux).

2.2.2 | FreeSurfer (FS)

All MRI were processed with FreeSurfer 6.0 (Fischl, 2012) using the recon-all pipeline with default settings without manual interventions. Global and regional mean cortical thickness values were extracted from the surface statistics (lh.aparc.stats, rh.aparc.stats) for regions of interest (ROI) as defined by the Desikan-Killiany atlas (Desikan et al., 2006).

2.3 | Evaluation

For the evaluation, we processed all non-enhanced and CE images with FreeSurfer and DL+DiReCT as described above and extracted global and regional mean cortical thickness measures. Pearson correlation coefficients (r) were calculated between image pairs, both across all pairs ($n = 454$) as well as on the subset of pairs with identical sequence parameters (TI/TR, $n = 213$). To allow direct comparison

with previous work (Lie et al., 2022), intra-class correlation coefficients (ICC) were calculated based on mean-rating ($k = 2$), consistency, 2-way mixed-effects model (Koo & Li, 2016) with the R-package *irr* (Gamer et al., 2012).

Leveraging the longitudinal nature of the data, we calculated annual atrophy rates in mm/year by fitting a linear model to the measures of all time points for each patient with age as the covariate. Patients with less than three time points were discarded, resulting in rates from $n = 73$ patients (eight PPMS and 65 RRMS). Group differences in the atrophy rates between the two MS cohorts were quantified with effects sizes using Cohen's d (Torchiano, 2019).

Finally, structural covariance networks (SCN) were constructed (Evans, 2013) using the "brain connectivity toolbox" for Python (Rubinov & Sporns, 2010). ROI-wise cortical thickness measures were corrected for age by fitting linear models to each measure with age as co-variate. The resulting residuals were then used to calculate correlation matrices. Binary undirected graphs were derived by thresholding the correlation matrix at 0.01 intervals. Confidence intervals were determined with random sampling by repeatedly constructing SCNs from 80% of the images 1000 times. Given multiple sclerosis has been associated with SCN disruption (Tewarie et al., 2014), and since this work is not intended as a complete network study, we focused on the example of *global efficiency* (Latora & Marchiori, 2001) as a metric of network integration to compare the two subgroups of MS patients. In contrast to the mean shortest path length, the global efficiency has a finite value also for disconnected graphs. Based on very recent findings in (Fleischer et al., 2019; Tur et al., 2020) we speculated that SCNs could be differently affected in PPMS and RRMS.

Statistical analyses were performed using *R* with the *stats* package version 3.6.2 (R Core Team, 2019). A significance level $\alpha = 0.05$ was used.

3 | RESULTS

The median processing runtime of FreeSurfer was 9.3 h (range: 5.8–38.9) for the non-enhanced images and 15.6 h (7.3–86.3) for the CE images. For DL+DiReCT, the runtimes were 0.3 h (0.2–0.5) for both image types. The median number of surface holes as calculated by FreeSurfer was 74 (range: 15–441) on the non-enhanced images and 332 (72–1091) on the CE images. Detailed runtime statistics are reported in Supplementary Figures S2 and S3. FreeSurfer failed to process 17 CE images. Subsequently reported results are based on the available images (i.e., based on $n = 454$ images from DL+DiReCT and $n = 437$ from FreeSurfer).

Correlation coefficients for the global mean thickness between all non-enhanced and CE image pairs were higher for DL+DiReCT ($r = 0.92$ / ICC $= 0.96$) than for FreeSurfer ($r = 0.75$ / ICC $= 0.85$) as depicted in Figure 1. Equivalent results were observed for the subset of image pairs with identical MR sequence parameters (DL+DiReCT $r = 0.96$ / ICC $= 0.98$, FreeSurfer r $= 0.87$ / ICC $= 0.93$). When calculated from the contrast-enhanced images, cortical thickness values were systematically elevated with both methods. This effect was
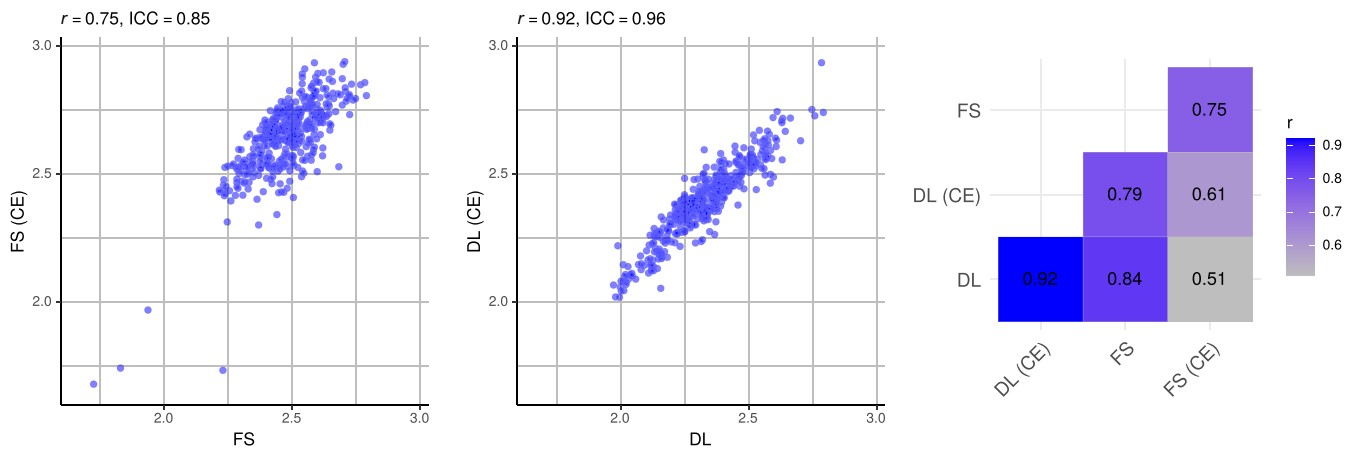
FIGURE 1 Correlations of global mean cortical thickness measures between all pairs of non-enhanced (*x*-axis) and contrast-enhanced (CE, *y*-axis) images. Figures for FreeSurfer (FS, left), DL+DiReCT (DL, middle) and combinations thereof (right matrix)

FIGURE 2 Color-coded Pearson correlation coefficients (*r*) of the ROI-wise average cortical thicknesses between measures derived from all pairs of non-enhanced and CE images
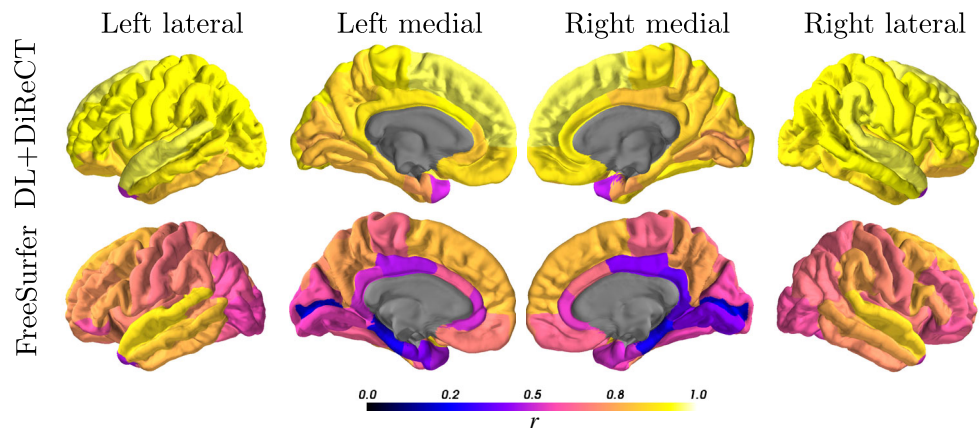
three times more pronounced with FreeSurfer (mean difference = 0.17 mm) than with DL+DiReCT (0.06 mm) as shown in the Bland–Altman plots in Supplementary Figure S4. Regional correlations are depicted in Figure 2, with corresponding values reported in Supplementary Tables S2 and S3 including results for subcortical volumes. Complementary results for the finer-grained parcellations can be found in Supplementary Figure S5.

Subject-wise correlation coefficients (calculated across all 64 regions per subject) were also significantly higher for DL+DiReCT (Supplementary Figure S6). A qualitative example is shown in Figure 3, which was chosen as the subject equal to the 10% quantile from DL+-DiReCT, meaning that 410/454 image pairs showed higher and 44/454 showed lower correlation. Similar examples corresponding to the median correlation coefficient and an analysis of an *outlier* are shown in Supplementary Figures S7 and S8.

3.1 | Annual atrophy rates

Both methods revealed an increased mean global annual atrophy rates for PPMS patients compared to the RRMS group, except FreeSurfer from CE images (see Table 1). The effect sizes were more pronounced with DL+DiReCT (non-enhanced = −0.304, CE = -0.169) than Free-Surfer (non-enhanced = −0.111, CE = 0.085). While these group effects were consistently larger than the differences between the two image types for DL+DiReCT, the group effects from FreeSurfer were similar or smaller in magnitude than between the image types (cf. columns *d* vs. last row in Table 1). Increased atrophy rates were also observed regionally for the PPMS group (Figure 4).

3.2 | Structural covariance networks (SCN)

Analysis using SCN revealed group differences, reflecting a higher global efficiency metric for the PPMS group compared to RRMS (Figure 5). This observation was present for the networks constructed with DL+DiReCT from both image types, as well as with FreeSurfer from the non-enhanced images. In contrast, no such difference could be observed when calculated with FreeSurfer from the CE images. Confidence intervals of the measures from DL+DiReCT were generally smaller than from FreeSurfer, as indicated for the networks constructed at a threshold of 0.5 (see Figure 5 right).
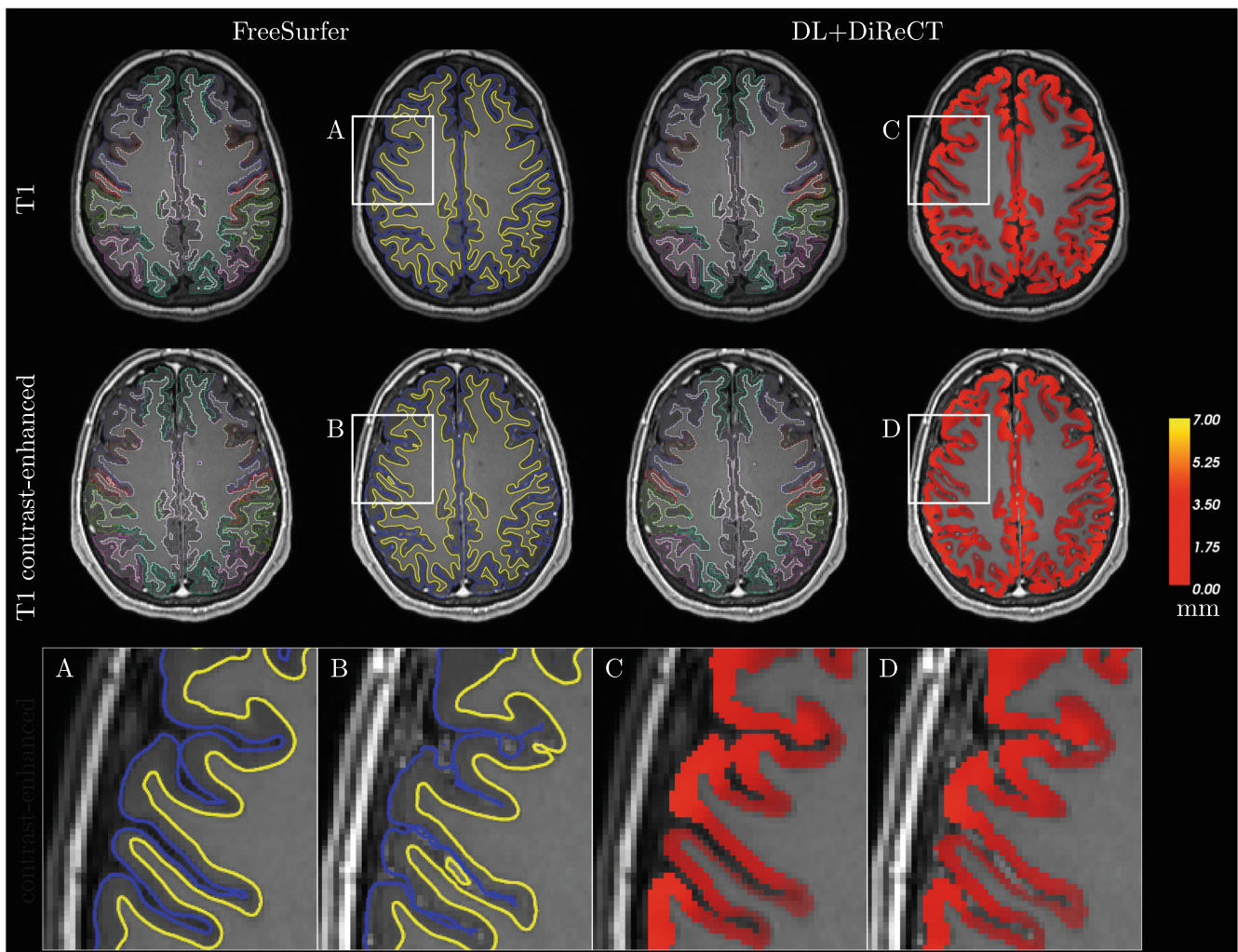
FIGURE 3 Qualitative example (10%-quantile subject-wise correlations across ROIs from DL+DiReCT) with colorized cortical parcellations. To illustrate the cortial thickness, the reconstructed surfaces from FreeSurfer are shown in blue (pial) and yellow (GM/WM) whereas for DL+DiReCT the voxel-wise thickness map is shown. The magnified view shows distortion artifacts in the FreeSurfer surface from CE image (B), particularly in the sulci. Subject-wise correlations over all ROIs of the Desikan-Killiany atlas are $r = 0.65$ for FreeSurfer and $r = 0.95$ for DL+DiReCT (as highlighted in Supplementary Figure S6)

TABLE 1 Mean annual global atrophy rates (mm/year)

	DL+DiReCT			FreeSurfer		
	Non-enhanced	CE	*d*	Non-enhanced	CE	*d*
PPMS [mm/year]	**−0.012**	**−0.009**	**−0.157**	**−0.008**	**−0.009**	**0.084**
RRMS [mm/year]	**−0.003**	**−0.005**	**0.075**	**−0.005**	**−0.011**	**0.224**
Effect size [Cohens'd]	−0.304	−0.169		−0.111	0.085	

Note: Effect size between PPMS and RRMS (last row) and between non-enhanced and CE (column *d*). For better readability, atrophy rates are bold faced.

4 | DISCUSSION

In this study, we investigated the utilization of contrast-enhanced (CE) T1-weighted MRI to derive brain morphometry. A cohort of 75 MS patients with a total of 454 image pairs of non-enhanced and CE images was used for the evaluation. Results from a baseline generated with FreeSurfer were compared to an extended version of

DL+DiReCT by enriching the training data of the deep learning-based segmentation.

After retraining the model with pairs of non-enhanced and CE images of MS patients, DL+DiReCT yielded high correlations between cortical thickness and GM volume values calculated from both image types. This was observed, both globally ($r = 0.91$) and regionally. Although FreeSurfer processed most of the post-contrast images
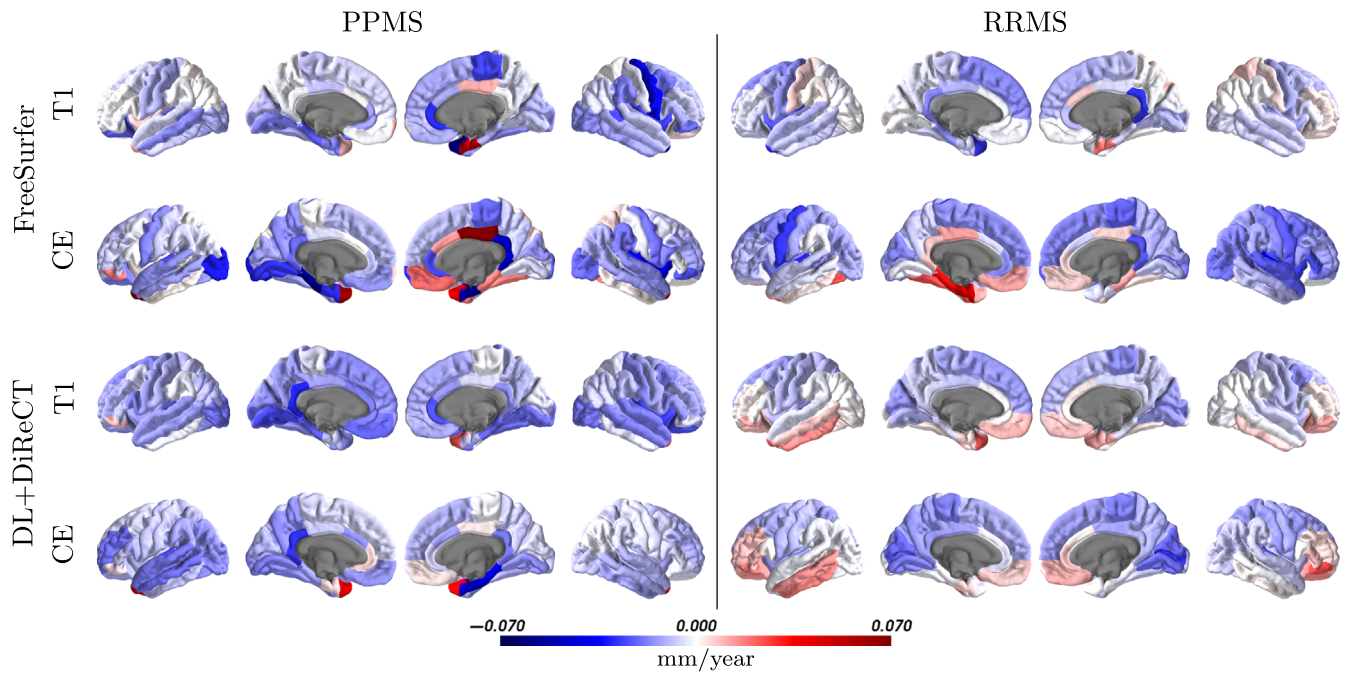
FIGURE 4 Mean annual regional atrophy rates in mm/year for the two MS subroups derived from non-enhanced and contrast-enhanced (CE) images with FreeSurfer and DL+DiReCT
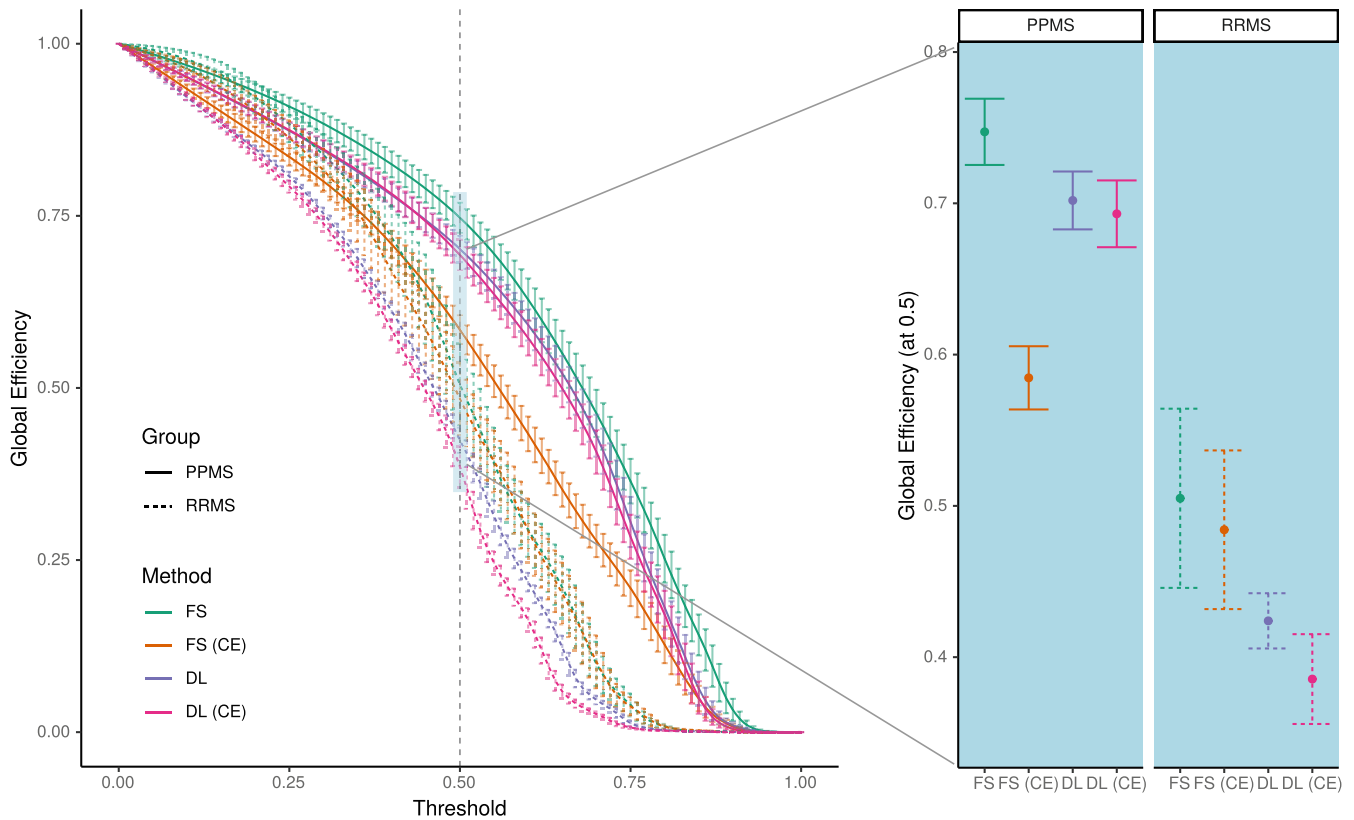

FIGURE 5 Global efficiency metric for the PPMS and RRMS subgroups calculated from structural covariance networks derived from non-enhanced and contrast-enhanced (CE) images with FreeSurfer (FS) and DL+DiReCT (DL). Left: Global efficiency as a function of threshold applied on the correlation matrix. Right: Results corresponding to the threshold at 0.5. Error bars indicate one standard deviation

without premature termination (with the exception of 17/454 images, i.e., 3.7%), we observed a substantially higher number of surface holes (Euler number) than for non-enhanced images. The Euler number can serve as a proxy for the input image quality (Kaufmann et al., 2019) and higher values are generally considered an indicator for a lower quality of the reconstructed surface (Monereo-Sánchez et al., 2021). Consequently, the correlations of cortical thickness estimates between non-enhanced and the corresponding CE images were weaker ($r = 0.75$).

Previous work on morphometry from contrast-enhanced T1w MRI is rare, especially for global and regional cortical thickness measures. A recent dedicated analysis from Lie et al. (2022) reported good to excellent results for global mean thickness (ICC >0.96) using Free-Surfer. The study was, however, based on a much smaller sample size of only 22 patients and included only image pairs with identical MR sequence parameters in the analysis. We were unable to observe such high correlations in our data, even in the subset of image pairs with identical acquisition parameters TI and TR ($r = 0.87$/ICC $= 0.93$ for FreeSurfer). Whether the additional water excitation pulse (Norbeck et al., 2020) in our contrast-enhanced MP-RAGE protocol is causing the remaining difference is difficult to judge. In line with the results from Lie et al. (2022), our results using FreeSurfer yielded an identical mean overestimation of 0.17 mm for the global cortical thickness derived from the CE images compared to the non-enhanced images. This systematic bias was reduced to 0.06 mm using DL+DiReCT.

More pronounced global annual atrophy rates were observed for the PPMS patients compared to the RRMS group under treatment with Natalizumab, in line with expectations (Portaccio et al., 2013; Preziosa et al., 2020). The rates of the RRMS group were similar to previously reported age-related atrophy of -0.004 mm/year in healthy cohorts (Lemaitre et al., 2012). However, only the results from DL+DiReCT consistently yielded larger effect sizes between those two groups than between the image types, likely due to the lower robustness of FreeSurfer on the CE images. Regional atrophy rates revealed a better consistency of the patterns between the two image types for DL+DiReCT than for FreeSurfer (cf. Figure 4). In the RRMS group, the pattern suggested accelerated atrophy rates fronto-parietal as compared to temporal regions. For cortical thickness, less reliable measurement has been reported for the cingulate cortex (Kharabian Masouleh et al., 2020; Rebsamen et al., 2020). We attribute the regions with a putative increase of cortical thickness over time to this kind of uncertainties, especially in the results derived with FreeSurfer from CE images and accentuated in the cingulate cortex and temporo-basal and temporo-polar regions.

For SCNs, sizable uncertainties for edge weights and derived graph measures have recently been demonstrated, particularly when estimated from cortical thickness using FreeSurfer (Carmon et al., 2020). Our findings of smaller error bars when using DL+-DiReCT instead of FreeSurfer together with larger group separation capabilities strongly suggest DL-based input for SCN estimation as an alternative. Enhanced robustness of DL+DiReCT compared to Free-Surfer has been reported before (Rebsamen et al., 2020, 2022; Rusak et al., 2022). SCN analysis revealed clear separation between the two

subgroups of MS, with DL+DiReCT from both non-enhanced and CE images. Whether this finding generalizes remains to be investigated since our PPMS subgroup was small and literature about network efficiency in various phenotypes of MS is sparse (Fleischer et al., 2019; Tur et al., 2018, 2020).

4.1 | Limitations

FreeSurfer failed to process 17 contrast-enhanced MRIs, and we have made no attempts to re-process these cases after manual interventions. No pre-processing of CE images was performed either, and we acknowledge that results from FreeSurfer might yield better results after tuning, e.g., by suppressing high voxel intensities. However, it remains questionable if a globally best threshold or similar parameter could be found and debatable if introducing an additional hyper-parameter is desirable at all. Pre-processing of CE images would likely remain manual labor that needs to be performed on an individual case basis.

All MR images in the analysis were from chronic MS patients showing some degree of white matter lesions. Nevertheless, we refrained from *lesion filling* as cortical thickness measures derived with FreeSurfer have shown to be unaffected (Biberacher et al., 2016; Guo et al., 2019). Consequently, our performance assessment concentrated on global and regional mean cortical thickness, while subcortical GM volumes are only reported for completeness in the Supplementary Materials.

Reflecting the lower prevalence of primary progressive compared to relapsing remitting MS in the population, the analyzed subgroups are unbalanced. A substantially lower number of patients in the PPMS group with fewer follow-up MRIs is a limitation of this study. While group differences for the global atrophy rates are distinct and concordant with expectations from the literature (Portaccio et al., 2013; Preziosa et al., 2020), interpretation of regional atrophy patterns requires a critical appraisal of the results.

Due to the retrospective nature of this study and the fact that images were acquired in clinical routine over a period of 9 years including protocol and scanner upgrades, the data consist of gradient-echo sequences with variations of sequence parameters and sources from different scanners. These variations likely result in lower correlations. However, even when comparing only pairs of MRI with identical parameters (TI/TR), the effect of considerably higher correlations with DL+DiReCT ($r = 0.96$) than for FS ($r = 0.87$) remained.

4.2 | Outlook

The ability to derive brain morphometry reliably from contrast-enhanced T1w MRI will make additional data accessible for quantitative analysis. As a consequence, studies on retrospective clinical data might benefit from a larger sample size. In datasets with both non-enhanced and contrast-enhanced images, one could run the analysis twice with both image types, increasing statistical power. For future

applications of morphometry on individual patients in clinical routine, running an analysis twice might increase confidence in the results.

While correlations between non-enhanced and CE images were excellent, there remains a systematic bias in their absolute values. Whether mixing measures from non-enhanced and CE images in the same analysis is feasible, e.g., by applying a correction factor, remains to be investigated.

5 | CONCLUSIONS

With the proposed deep learning-based morphometry tool (DL+DiReCT), brain morphometry can be derived reliably from contrast-enhanced T1-weighted MRI. The main findings (effect sizes of atrophy rates between groups and network effects from SCN) in the analyzed cohorts were consistent between the non-enhanced and contrast-enhanced images.

DATA AVAILABILITY STATEMENT
The morphometry tool is publicly available, including trained models (https://github.com/SCAN-NRAD/DL-DiReCT). MR images of patients used for the evaluation are not readily available. Derived morphometric measures are available from the authors upon reasonable request.

ETHICS STATEMENT
Patients were identified from the existing neuroimmunological registry and retrospective image analysis of clinical routine MRIs was performed (KEK-BE 2017-01369, KEK-BE 2016-02035).

ORCID
Michael Rebsamen https://orcid.org/0000-0002-8441-1485
Christian Rummel https://orcid.org/0000-0003-2345-7938

REFERENCES
Alexander-Bloch, A., Giedd, J. N., & Bullmore, E. (2013). Imaging structural co-variance between human brain regions. *Nature Reviews Neuroscience*, *14*, 322–336. https://doi.org/10.1038/nrn3465

Biberacher, V., Schmidt, P., Keshavan, A., Boucard, C. C., Righart, R., Sämann, P., Preibisch, C., Fröbel, D., Aly, L., Hemmer, B., Zimmer, C., Henry, R. G., & Mühlau, M. (2016). Intra-and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis. *NeuroImage*, *142*, 188–197. https://doi.org/10.1016/j.neuroimage.2016.07.035

Carmon, J., Heege, J., Necus, J. H., Owen, T. W., Pipa, G., Kaiser, M., Taylor, P. N., & Wang, Y. (2020). Reliability and comparability of human brain structural covariance networks. *NeuroImage*, *220*, 117104. https://doi.org/10.1016/j.neuroimage.2020.117104

Das, S. R., Avants, B. B., Grossman, M., & Gee, J. C. (2009). Registration based cortical thickness measurement. *NeuroImage*, *45*, 867–879. https://doi.org/10.1016/j.neuroimage.2008.12.016

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*, 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Eshaghi, A., Marinescu, R. V., Young, A. L., Firth, N. C., Prados, F., Jorge Cardoso, M., Tur, C., De Angelis, F., Cawley, N., Brownlee, W. J., et al. (2018). Progression of regional grey matter atrophy in multiple sclerosis. *Brain*, *141*, 1665–1677. https://doi.org/10.1093/brain/awy088

Evans, A. C. (2013). Networks of anatomical covariance. *NeuroImage*, *80*, 489–504. https://doi.org/10.1016/j.neuroimage.2013.05.054

Fischl, B. (2012). FreeSurfer. *Neuroimage*, *62*, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, *97*, 11050–11055. https://doi.org/10.1073/pnas.200033797

Fleischer, V., Koirala, N., Droby, A., Gracien, R. M., Deichmann, R., Ziemann, U., Meuth, S. G., Muthuraman, M., Zipp, F., & Groppa, S. (2019). Longitudinal cortical network reorganization in early relapsing–remitting multiple sclerosis. *Therapeutic Advances in Neurological Disorders*, *12*, 1756286419838673. https://doi.org/10.1177/1756286419838673

Gamer, M., Lemon, J., & Singh, I. F. P., (2012). irr: Various Coefficients of Interrater Reliability and Agreement. https://CRAN.R-project.org/package=irr. R package version 0.84.

Guo, C., Ferreira, D., Fink, K., Westman, E., & Granberg, T. (2019). Repeatability and reproducibility of FreeSurfer, FSLSIENAX and SPM brain volumetric measurements and the effect of lesion filling in multiple sclerosis. *European Radiology*, *29*, 1355–1364. https://doi.org/10.1007/s00330-018-5710-x

Hannoun, S., Baalbaki, M., Haddad, R., Saaybi, S., El Ayoubi, N. K., Yamout, B. I., Khoury, S. J., & Hourani, R. (2018). Gadolinium effect on thalamus and whole brain tissue segmentation. *Neuroradiology*, *60*, 1167–1173. https://doi.org/10.1007/s00234-018-2082-5

Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H. P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K. H., & Kickingereder, P. (2019). Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*, *40*, 4952–4964. https://doi.org/10.1002/hbm.24750

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*, 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*, 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Kappos, L., De Stefano, N., Freedman, M. S., Cree, B. A., Radue, E. W., Sprenger, T., Sormani, M. P., Smith, T., Häring, D. A., Piani Meier, D., et al. (2016). Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing–remitting multiple sclerosis. *Multiple Sclerosis Journal*, *22*, 1297–1305. https://doi.org/10.1177/1352458515616701

Kaufmann, T., van der Meer, D., Doan, N. T., Schwarz, E., Lund, M. J., Agartz, I., Alnæs, D., Barch, D. M., Baur-Streubel, R., Bertolino, A., et al. (2019). Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*, 22, 1617–1623. https://doi.org/10.1038/s41593-019-0471-7

Kharabian Masouleh, S., Eickhoff, S. B., Zeighami, Y., Lewis, L. B., Dahnke, R., Gaser, C., Chouinard-Decorte, F., Lepage, C., Scholtens, L. H., Hoffstaedter, F., Glahn, D. C., Blangero, J., Evans, A. C., Genon, S., & Valk, S. L. (2020). Influence of processing pipeline on cortical thickness measurement. *Cerebral Cortex*, 30, 5014–5027. https://doi.org/10.1093/cercor/bhaa097

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

van der Kouwe, A. J., Benner, T., Salat, D. H., & Fischl, B. (2008). Brain morphometry with multiecho MPRAGE. *NeuroImage*, 40, 559–569. https://doi.org/10.1016/j.neuroimage.2007.12.025

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, 87, 198701. https://doi.org/10.1103/PhysRevLett.87.198701

Lemaitre, H., Goldman, A. L., Sambataro, F., Verchinski, B. A., Meyer-Lindenberg, A., Weinberger, D. R., & Mattay, V. S. (2012). Normal age-related brain morphometric changes: Nonuniformity across cortical thickness, surface area and gray matter volume? *Neurobiology of Aging*, 33, 617–e1. https://doi.org/10.1016/j.neurobiolaging.2010.07.013

Lie, I. A., Kerklingh, E., Wesnes, K., van Nederpelt, D. R., Brouwer, I., Torkildsen, Ø., Myhr, K. M., Barkhof, F., Bø, L., & Vrenken, H. (2022). The effect of gadolinium-based contrast-agents on automated brain atrophy measurements by FreeSurfer in patients with multiple sclerosis. *European Radiology*, 32, 3576–3587. https://doi.org/10.1007/s00330-021-08405-8

McKinley, R., Rebsamen, M., Meier, R., Reyes, M., Rummel, C., & Wiest, R. (2019). Few-shot brain segmentation from weakly labeled data with deep heteroscedastic multi-task networks. arXiv preprint arXiv:1904.02436 https://arxiv.org/abs/1904.02436.

Monereo-Sánchez, J., de Jong, J. J., Drenthen, G. S., Beran, M., Backes, W. H., Stehouwer, C. D., Schram, M. T., Linden, D. E., & Jansen, J. F. (2021). Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations-insights from the Maastricht study. *NeuroImage*, 237, 118174. https://doi.org/10.1016/j.neuroimage.2021.118174

Mugler, J. P., III, & Brookeman, J. R. (1990). Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magnetic Resonance in Medicine*, 15, 152–157. https://doi.org/10.1002/mrm.1910150117

Norbeck, O., Sprenger, T., Avventi, E., Rydén, H., Kits, A., Berglund, J., & Skare, S. (2020). Optimizing 3D EPI for rapid t1-weighted imaging. *Magnetic Resonance in Medicine*, 84, 1441–1455. https://doi.org/10.1002/mrm.28222

Portaccio, E., Stromillo, M., Goretti, B., Hakiki, B., Giorgio, A., Rossi, F., De Leucio, A., De Stefano, N., & Amato, M. (2013). Natalizumab may reduce cognitive changes and brain atrophy rate in relapsing–remitting multiple sclerosis: A prospective, non-randomized pilot study. *European Journal of Neurology*, 20, 986–990. https://doi.org/10.1111/j.1468-1331.2012.03882.x

Preziosa, P., Rocca, M. A., Riccitelli, G. C., Moiola, L., Storelli, L., Rodegher, M., Comi, G., Signori, A., Falini, A., & Filippi, M. (2020). Effects of natalizumab and fingolimod on clinical, cognitive, and magnetic resonance imaging measures in multiple sclerosis. *Neurotherapeutics*, 17, 208–217. https://doi.org/10.1007/s13311-019-00781-w

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing https://www.R-project.org/

Radü, E. W., Bendfeldt, K., Mueller-Lenke, N., Magon, S., & Sprenger, T. (2013). Brain atrophy: An in-vivo measure of disease activity in multiple sclerosis. *Swiss Medical Weekly*, 143, w13887. https://doi.org/10.4414/smw.2013.13887

Rebsamen, M., Rummel, C., Reyes, M., Wiest, R., & McKinley, R. (2020). Direct cortical thickness estimation using deep learningbased anatomy segmentation and cortex parcellation. *Human Brain Mapping*, 41, 4804–4814. https://doi.org/10.1002/hbm.25159

Rebsamen, M., Radojewski, P., McKinley, R., Reyes, M., Wiest, R., & Rummel, C. (2022). A quantitative imaging biomarker supporting radiological assessment of hippocampal sclerosis derived from deep learning-based segmentation of T1w-MRI. *Frontiers in Neurology*, 13, 812432. https://doi.org/10.3389/fneur.2022.812432

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, 52, 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Rusak, F., Santa Cruz, R., Lebrat, L., Hlinka, O., Fripp, J., Smith, E., Fookes, C., Bradley, A. P., Bourgeat, P., Initiative, A. D. N., et al. (2022). Quantifiable brain atrophy synthesis for benchmarking of cortical thickness estimation methods. *Medical Image Analysis*, 82, 102576. https://doi.org/10.1016/j.media.2022.102576

Sastre-Garriga, J., Pareto, D., Battaglini, M., Rocca, M. A., Ciccarelli, O., Enzinger, C., Wuerfel, J., Sormani, M. P., Barkhof, F., Yousry, T. A., et al. (2020). MAGNIMS consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. *Nature Reviews Neurology*, 16, 171–182. https://doi.org/10.1038/s41582-020-0314-x

Schwarz, C. G., Gunter, J. L., Wiste, H. J., Przybelski, S. A., Weigand, S. D., Ward, C. P., Senjem, M. L., Vemuri, P., Murray, M. E., Dickson, D. W., Parisi, J. E., Kantarci, K., Weiner, M. W., Petersen, R. C., Jack, C. R., Jr., & Alzheimer's Disease Neuroimaging Initiative. (2016). A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *NeuroImage: Clinical*, 11, 802–812. https://doi.org/10.1016/j.nicl.2016.05.017

Steenwijk, M. D., Geurts, J. J., Daams, M., Tijms, B. M., Wink, A. M., Balk, L. J., Tewarie, P. K., Uitdehaag, B. M., Barkhof, F., Vrenken, H., et al. (2016). Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain*, 139, 115–126. https://doi.org/10.1093/brain/awv337

Tewarie, P., Steenwijk, M. D., Tijms, B. M., Daams, M., Balk, L. J., Stam, C. J., Uitdehaag, B. M., Polman, C. H., Geurts, J. J., Barkhof, F., et al. (2014). Disruption of structural and functional networks in long-standing multiple sclerosis. *Human Brain Mapping*, 35, 5946–5961. https://doi.org/10.1002/hbm.22596

Thompson, P. M., Jahanshad, N., Ching, C. R., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolín, S., Bralten, J., Bruin, W. B., et al. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Translational Psychiatry*, 10, 1–28. https://doi.org/10.1038/s41398-020-0705-1

Torchiano, M. (2019). effsize: Efficient effect size computation. https://doi.org/10.5281/zenodo.1480624. R package version 0.7.6. https://CRAN.R-project.org/package=effsize

Traboulsee, A., Simon, J., Stone, L., Fisher, E., Jones, D., Malhotra, A., Newsome, S., Oh, J., Reich, D., Richert, N., et al. (2016). Revised recommendations of the consortium of MS centers task force for a standardized MRI protocol and clinical guidelines for the diagnosis and follow-up of multiple sclerosis. *American Journal of Neuroradiology*, 37, 394–401. https://doi.org/10.3174/ajnr.A4539

Tur, C., Eshaghi, A., Altmann, D., Jenkins, T., Prados, F., Grussu, F., Charalambous, T., Schmidt, A., Ourselin, S., Clayden, J., et al. (2018). Structural cortical network reorganization associated with early conversion to multiple sclerosis. *Scientific Reports*, 8, 1–15. https://doi.org/10.1038/s41598-018-29017-1

Tur, C., Kanber, B., Eshaghi, A., Altmann, D. R., Khaleeli, Z., Prados, F., Ourselin, S., Thompson, A. J., Gandini Wheeler- Kingshott, C. A., Toosy, A. T., et al. (2020). Clinical relevance of cortical network dynamics in early primary progressive MS. *Multiple Sclerosis Journal*, *26*, 442–456. https://doi.org/10.1177/1352458519831400

Tustison, N. J., Avants, B. B., Cook, P. A., Song, G., Das, S., van Strien, N., Stone, J. R., & Gee, J. C. (2013). The ANTs cortical thickness processing pipeline. In J. B. Weaver & R. C. Molthen (Eds.), *Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging* (p. 86720K). International Society for Optics and Photonics. https://doi.org/10.1117/12.2007128

Warntjes, J., Tisell, A., Landtblom, A. M., & Lundberg, P. (2014). Effects of gadolinium contrast agent administration on automatic brain tissue classification of patients with multiple sclerosis. *American Journal of Neuroradiology*, *35*, 1330–1336. https://doi.org/10.3174/ajnr.A3890

Wattjes, M. P., & Barkhof, F. (2014). Diagnosis of natalizumab-associated progressive multifocal leukoencephalopathy using MRI. *Current Opinion in Neurology*, *27*, 260–270. https://doi.org/10.1097/WCO.0000000000000099

Wattjes, M. P., Ciccarelli, O., Reich, D. S., Banwell, B., de Stefano, N., Enzinger, C., Fazekas, F., Filippi, M., Frederiksen, J., Gasperini, C., Hacohen, Y., Kappos, L., Li, D. K. B., Mankad, K., Montalban, X., Newsome, S. D., Oh, J., Palace, J., Rocca, M. A., … Oh, J. (2021). 2021 MAGNIMS–CMSC–NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. *The Lancet Neurology*, *20*, 653–670. https://doi.org/10.1016/S1474-4422(21)00095-8

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., & Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, *53*, 1135–1146. https://doi.org/10.1016/j.neuroimage.2009.12.028

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.