



Deep Learning Corpus Callosum Segmentation as a Neurodegenerative Marker in Multiple Sclerosis

Michael Platten , Irene Brusini, Olle Andersson, Russell Ouellette , Fredrik Piehl, Chunliang Wang, and Tobias Granberg 

From the Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden (MP, RO, FP, TG); School of Engineering Sciences in Chemistry, Biotechnology and Health, Royal Institute of Technology, Stockholm, Sweden (MP, IB, OA, CW); Department of Neuroradiology, Karolinska University Hospital, Stockholm, Sweden (MP, RO, TG); Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden (IB); Department of Neurology, Karolinska University Hospital, Stockholm, Sweden (FP); and Center for Neurology, Academic Specialist Center, Stockholm Health Services, Stockholm, Sweden (FP)

ABSTRACT

Background and Purpose: Corpus callosum atrophy is a sensitive biomarker of multiple sclerosis (MS) neurodegeneration but typically requires manual 2D or volumetric 3D-based segmentations. We developed a supervised machine learning algorithm, DeepnCCA, for corpus callosum segmentation and relate callosal morphology to clinical disability using conventional MRI scans collected in clinical routine.

Methods: In a prospective study of 553 MS patients with 704 acquisitions, 200 unique 2D T₂-weighted MRI scans were delineated to develop, train, and validate DeepnCCA. Comparative FreeSurfer segmentations were obtained in 504 3D T₁-weighted scans. Both FreeSurfer and DeepnCCA outputs were correlated with clinical disability. Using principal component analysis of the DeepnCCA output, the morphological changes were explored in relation to clinical disease burden.

Results: DeepnCCA and manual segmentations had high similarity (Dice coefficients 98.1±.11%, 89.3±.76%, for intracranial and corpus callosum area, respectively through 10-fold cross-validation). DeepnCCA had numerically stronger correlations with cognitive and physical disability as compared to FreeSurfer: Expanded disability status scale (EDSS) ±6 months ($r = -.22$, $P = .002$; $r = -.17$, $P = .013$), future EDSS ($r = -.26$, $P < .001$; $r = -.17$, $P = .012$), and future symbol digit modalities test ($r = .26$, $P = .001$; $r = .24$, $P = .003$). The corpus callosum became thinner with increasing cognitive and physical disability. Increasing physical disability, additionally, significantly correlated with a more angled corpus callosum.

Conclusions: DeepnCCA (<https://github.com/plattenmichael/DeepnCCA/>) is an openly available tool that can provide fast and accurate corpus callosum measurements applicable to large MS cohorts, potentially suitable for monitoring disease progression and therapy response.

Keywords: Multiple sclerosis, magnetic resonance imaging, artificial intelligence, deep learning, corpus callosum.

Acceptance: Received September 12, 2020, and in revised form January 14, 2021. Accepted for publication January 14, 2021.

Correspondence: Address correspondence to Michael Platten, Department of Clinical Neuroscience, Karolinska Institutet, 17177, Stockholm, Sweden. E-mail: michael.platten@ki.se.

Acknowledgments and Disclosures: We would like to thank all the healthcare professionals at the Karolinska University Hospital involved in acquiring data for this study. We would also like to thank all the patients and their families for making this study possible.

This project was funded by the Stockholm Region (ALF grants: 20120213, 20150166, and 20180660) and Karolinska Institutet (PhD grants: Clinical Scientist Training Program and Researching Intern Grants). Tobias Granberg was supported by the Swedish Society for Medical Research, MERCK Grant for Multiple Sclerosis Innovation, as well as the Christer Lindgrens and Eva Fredholm foundation.

M.P. declares no conflicts of interest. I.B. declares no conflicts of interest. O.A. declares no conflicts of interest. R.O. declares no conflicts of interest. F.P. declares no conflicts of interest. C.W. declares no conflicts of interest. T.G. declares no conflicts of interest.

J Neuroimaging 2021;31:493-500.

DOI: 10.1111/jon.12838

Introduction

Multiple sclerosis (MS) is an immune-mediated neuroinflammatory and neurodegenerative disorder, which represents a leading cause of neurological disability among young and middle-aged individuals.¹ Magnetic resonance imaging (MRI) facilitates early diagnosis of MS,² and is a key part of the MS diagnostic criteria.³ MRI offers a unique ability to monitor the inflammatory and neurodegenerative aspects of MS,⁴ which is clinically valuable because both components are coupled to disease progression and worsened clinical outcomes.⁵

The corpus callosum is a dense white matter structure connecting the cerebral hemispheres. Callosal atrophy is signif-

icantly associated with cognitive and physical impairment in MS.⁶ While the corpus callosum is relatively stable in the normal aging population, it exhibits significant atrophy in MS,⁷ with a mean annual atrophy rate of 2.5%.⁸ Corpus callosum atrophy is, therefore, a strategic biomarker to monitor MS-related neurodegeneration.

There are several methods for segmenting the corpus callosum based on MRI, including manual segmentations,⁹ atlas-based software, such as FreeSurfer,¹⁰ and machine learning algorithms.¹¹ FreeSurfer can render 3D segmentations of the corpus callosum with high precision but lower accuracy, as compared to manual delineations.¹² In recent years, artificial

intelligence has become more widely applied in radiology and convolutional neural networks, a subclass of deep learning inspired by human layered neuronal connectivity, are particularly useful in segmenting radiological images.¹³ Convolutional neural networks have previously been trained and applied to MS data for lesion detection,¹⁴ and segmentation of large tissue compartments, such as gray and white matter,^{15,16} but it has not been trained and applied for corpus callosum segmentations in MS.

While measuring corpus callosal atrophy as a biomarker for disease progression and evaluation of the potential neuroprotective effects of therapies is clinically very useful, the morphological attributes of atrophy may elucidate further features specific to MS neurodegeneration. It has been shown that corpus callosum morphological parameters vary between the sexes,¹⁷ as well as in neurodegenerative diseases, such as Alzheimer's disease¹⁸ and MS.¹⁹

In this study, we developed an openly available MS-tailored deep learning algorithm, DeepnCCA (<https://github.com/plattenmichael/DeepnCCA/>), implemented in Python and using TensorFlow and Keras for normalized corpus callosum area measurements. This produces a quick and accurate measurement that can be applied as an MS biomarker of neurodegeneration in large MRI datasets. Additionally, we compare the output with FreeSurfer corpus callosum segmentations and correlate this with clinical disability. Furthermore, utilizing the output from DeepnCCA, shape analysis was explored to evaluate which morphological changes of the corpus callosum are associated with worse cognitive and physical disability.

Methods

Study Design

This study is based on a population-based prospective cohort study in the County of Stockholm, Stockholm Prospective Multiple Sclerosis (STOP-MS), initiated in January 2001, with the objective of identifying prognostic factors for long-term outcomes in newly diagnosed MS patients, as previously described.²⁰ The study was approved by the Regional Ethics

Table 1. Cohort Demographics at the First Included Brain MRI

Patients (<i>n</i> = 353)	
Age at first scan (years)	39 ± 11
Sex (% female)	77
Mean disease duration at first scan (years) ^a	1.1 ± 4.1
Subtype: RRMS/SPMS/PPMS/NA (%)	55/26/5/14
Number of T ₂ and T ₁ -weighted MRIs	504
Normalized corpus callosum area (%)	3.4 ± .61
FreeSurfer corpus callosum volume (mm ³)	4003 ± 624
Median EDSS ± 6 months (<i>n</i> = 204), IQR	2.0 (1.0-3.0)
Median EDSS future (<i>n</i> = 227), IQR ^b	2.5 (1.0-5.0)
Median SDMT future z-score (<i>n</i> = 151) IQR ^c	-1.3 (-2.4 to -.5)

^a Mean disease duration from diagnosis to first scan.

^b Average number of years between scan and EDSS was 9.7 ± 3.9 years.

^c Average number of years between scan and SDMT was 8.1 ± 4.0 years.

All the data represent mean ± standard deviation unless otherwise indicated.

RRMS, relapsing-remitting MS; SPMS, secondary progressive MS; PPMS, primary progressive MS; NA, not available; *n*, sample size; EDSS, Expanded Disability Status Scale; SDMT, Symbol Digit Modalities Test; IQR, interquartile range.

Review Board in Stockholm (reg. no. 02-548, 2009/2107-31/2, 2009/2107-31/2, 2018/2711-32, 2020-01954, 2020-02794) and written informed consent was obtained from all participants. For this study, only patients who received the diagnosis of MS and underwent a brain MRI at the Karolinska University Hospital in Huddinge, Stockholm, Sweden, were included.

Clinical Data

Clinical data variables were derived from the Swedish MS registry (SMSreg). Table 1 presents the demographics of the participants. The MS subtype²¹ and diagnosis were set by a neurologist according to concurrent diagnostic criteria.²²⁻²⁴ We analyzed two of the most established clinical outcome scales used in MS; the Expanded Disability Status Scale (EDSS) reflecting physical impairment in different neurological domains and the Symbol Digit Modalities Test (SDMT) for deficits in cognitive processing speed, transformed into a z-score.²⁵ Figure 1 summarizes the available clinical data and corresponding MRIs.

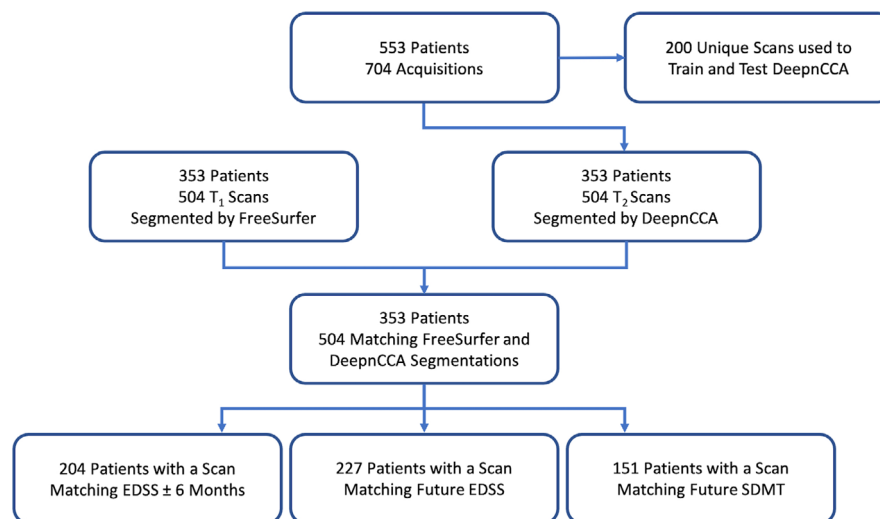


Fig 1. Flowchart of data analysis.
Flowchart of available MRI and clinical data.

Table 2. MRI Acquisition Parameters

Sequence Scanner, T	T ₂ -weighted Siemens Vision 1.5 T	T ₁ -weighted
Sequence	Sagittal 2D TSE	Axial 3D GRE (MPRAGE)
Voxel size, mm	.9 × .9 × 4.0	1.0 × 1.0 × 1.5
Echo time, milliseconds	96	7.0
Repetition time, milliseconds	3,500	13.5
Inversion time, milliseconds	N/A	300
Flip angle, °	180	15

T, Tesla; TSE, Turbo Spin Echo; GRE, Gradient Echo; MPRAGE, Magnetization-Prepared Rapid Gradient Echo; N/A, not applicable.

Magnetic Resonance Imaging

All brain MRIs were performed according to a standardized protocol in clinical routine at the Karolinska University Hospital in Huddinge, Stockholm, Sweden. A total of 704 brain MRI scans were available. In order to select the mid-sagittal slice for analysis, a script was created that extracted the middle slice. Since not all 2D T₂-weighted scans were aligned appropriately with an anatomically located mid-sagittal middle slice, a step for quality control of data was introduced. A trained rater and physician (M.P.) went through the MRIs where the middle slice was misaligned and manually extracted the anatomically located mid-sagittal slice. This control was done prior to introducing any of the data to DeepnCCA. No data were discarded from analysis. One scanner was used, and the technical details are presented in Table 2. The data are representative of a real-life clinical scenario rather than a controlled treatment trial.

Deep Learning Algorithm

Overview

An in-house de novo algorithm was implemented in Python using TensorFlow and Keras. The network architecture is based on a 2D U-net architecture, applied on independent slices, and relies on image augmentation to increase sample size.²⁶ We modified the U-net by performing batch normalization and adding a dropout layer, which randomly dropped 25% of its connections. Figure 2 presents the processing structure. The Adam optimizer was utilized, which has demonstrated empirically to work well in practice.²⁷ Two networks were trained, one for the corpus callosum area and another for the intracranial area, analogously to how manual segmentations are performed. The corpus callosum area is then divided by the intracranial area to account for scaling effects, resulting in the normalized corpus callosum area. The training dataset consisted of 200 manually delineated corpus callosum and intracranial areas by a trained rater and physician (M.P.), quality controlled by a radiologist (T.G.). The dataset was selected at random from the pool of available MRI, while simultaneously ensuring that individual patients were only represented once. The patients from the training set were not included in any other analyses, as this would have potentially inflated the DeepnCCA results. The performance of the algorithm was evaluated through a 10-fold cross-validation.

Preprocessing and Image Augmentation

Volumetric 3D images are typically recommended for MS evaluation.² Historically, however, 2D acquisitions have been the clinical norm. Due to the shape of the corpus callosum, it lends itself especially well to being segmented on sagittal 2D images and 3D images can be reconstructed as sagittal 2D images. In the current study, we, therefore, focused on 2D-based segmentation in order to provide an alternative segmentation

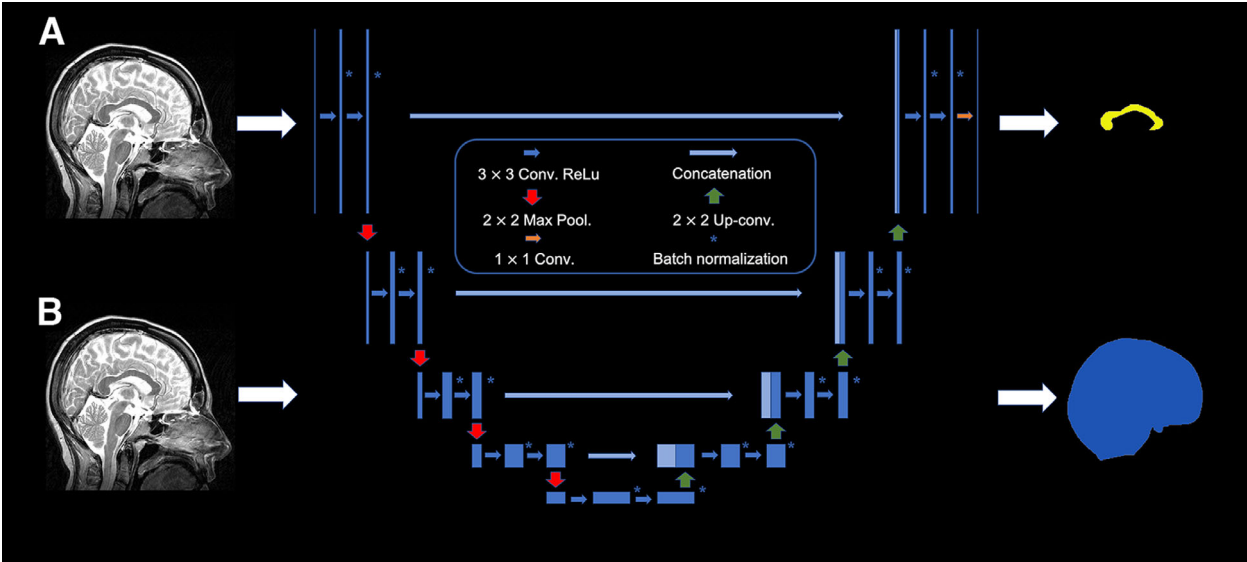


Fig 2. Architecture of the DeepnCCA U-nets. Two separate algorithms were trained for the corpus callosum area (A) and the intracranial area (B), using U-net architectures. This is similar to the original U-net model by Ronneberger et al,²⁶ but with addition of dropout and batch normalization. The pathway has an encoding and decoding path with concatenation in order to retain information from more original upstream images. Conv, convolution; ReLU, rectified linear unit; Max Pool, max pooling; Up-Conv, up-convolution.

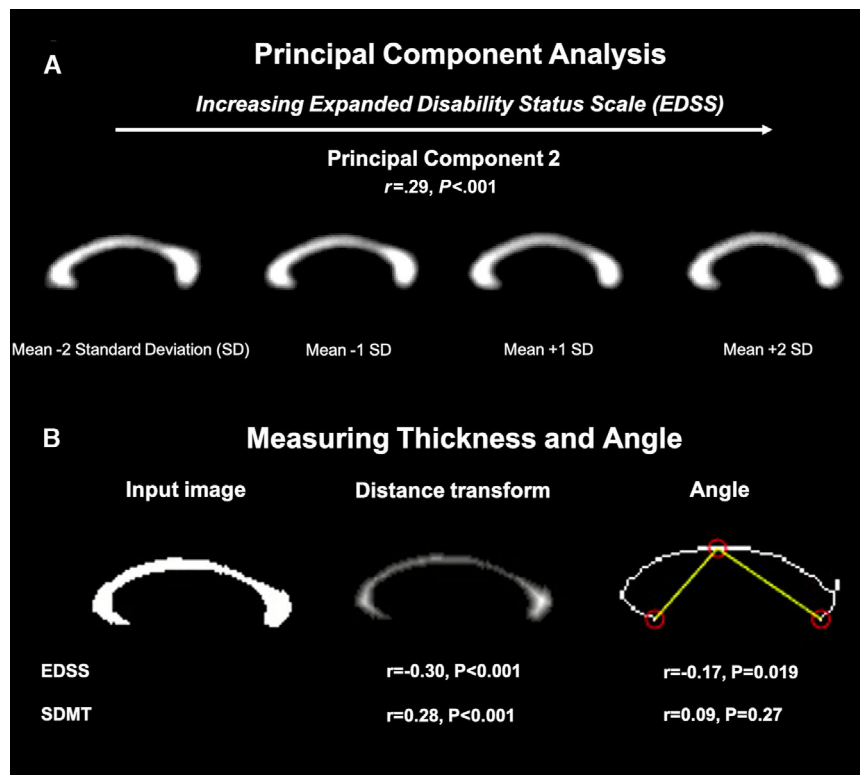


Fig 3. Shape changes associated with physical and cognitive disability.

(A) The principal component of shape variation that significantly correlated with the EDSS physical disability scores. It is a visual representation of how the shape changes with increasing disability. (B) Data representations used for the computation of the bending angle and average thickness. The distance transform (middle) is computed from the input segmentation (left) and the average value along the midline between the two end points defines the average thickness of the corpus callosum. From the input segmentation mask, its skeleton is identified (right). From the skeleton, it is possible to automatically detect its end points as well as the midpoint between them. These three points can be connected by two vectors (in yellow) and the angle between such vectors represents the callosal bending angle. SDMT, symbol digit modalities test.

method to manual segmentations in cohorts where only 2D images are available.

The first step of the preprocessing was intensity normalization, since pixel intensities in MRI are arbitrarily scaled, and affects the weights of convolutional filters.²⁸ This was done by mapping the 5th and 95th percentiles of the pixel intensity of each slice to between 0 and 1. We also resized the slices to a resolution of 256×256 , in order to have a uniform dataset. We then increased our training sample size by randomly applying five methods of image augmentation: rotation (-10 to $+10^\circ$), displacement (x- and y-axis translation, -10 to $+10$ pixels), scaling (-6% to $+6\%$ of the original size), horizontal mirroring and adding Gaussian background noise that causes minor distortions and a grainier appearance of the image. At each epoch of training, the randomization started over, thus creating newly augmented images for the next training batch. Thereby, the number of training data was only limited by the number of epochs being performed.

Shape Analysis

Using the output from the DeepnCCA algorithm, we created shape variation models for the corpus callosum. The aim was to see how the morphology of the corpus callosum correlated with the clinical disability as measured by EDSS and SDMT. We, furthermore, extracted morphological parameters, corpus callosum thickness, and angle to evaluate how these changed with disease burden.

Principal Component Analysis

To perform the principal component analysis (PCA), distance transforms were first extracted from all available segmentations. The approach of performing PCA using distance transforms was inspired by Leventon et al.²⁹ All analyses were performed using Python with PCA tools from the Scikit-learn library. PCA was run on all distance transforms to compute the principal components that preserve 99% of the total variance present in the dataset. For each of the identified principal components, the projections of every image on that principal component were extracted and correlated with the corresponding disability scores.

Bending Angle and Mean Thickness

The outputs of the aforementioned model lent inspiration to measure both the bending angle and mean thickness of the corpus callosum. A corpus callosum skeleton was identified based on the midline of the corpus callosum, and the bending angle was defined as the angle between the two vectors joining each end point of the corpus callosum to its midpoint. The thickness was computed as the average distance transform of the segmentation along its midline. Figure 3 contains a visualization of the methodology while simultaneously presenting the results. All computations were performed on MATLAB (MATLAB R2019a, The Mathworks Inc., Natick, MA, USA).

Table 3. 10-Fold Cross-Validation of the Deep Learning Training

K-Fold number	Dice coefficient of test set intracranial area, %	Dice coefficient of test set corpus callosum area, %
Fold 1	98.13	89.41
Fold 2	97.91	89.63
Fold 3	98.02	89.96
Fold 4	98.14	88.77
Fold 5	97.89	89.75
Fold 6	98.20	90.32
Fold 7	98.08	89.06
Fold 8	97.98	87.54
Fold 9	98.01	89.25
Fold 10	98.16	89.22
Mean \pm SD	98.05% \pm .11%	89.29% \pm .76%

SD, standard deviation.

Statistical Analysis

Clinical Data

Normal distribution was assessed through histograms and Shapiro-Wilk's test of normality. Pearson and Spearman's rho were applied to analyze the correlation between FreeSurfer and DeepnCCA segmentations to the clinical variables EDSS and SDMT. *P*-values of <.05 were considered statistically significant.

DeepnCCA

We applied two main statistical tools for monitoring model efficiency: a loss function and a calculation of the Dice coefficient. The equation for calculating the loss function (Equation 1) and the equation for calculating the Dice coefficient (Equation 2) are shown below. The model was evaluated using 10-fold cross-validation coded in python. For the PCA analysis, a Bonferroni correction was applied due to multiple comparisons.

Binary Cross – Entropy Loss Function

$$= -\frac{1}{N} \sum_{i=1}^N y_i \times \log(p(y_i)) + (1 - y_i) \times \log(1 - p(y_i))$$

N is the number of points/pixels, y_i is label, and $p(y_i)$ is predicted label.

$$\text{Dice Coefficient} = \frac{2TP}{2TP + FP + FN} \quad (2)$$

TP is True Positive, FP is False Positive, and FN is False Negative.

Results

DeepnCCA Performance

Intracranial Segmentations

The DeepnCCA intracranial segmentations had high similarity compared to manual segmentations (mean Dice score 98.1 \pm .11% using a 10-fold cross-validation, see Table 3). The model performance over the first fold, as measured by the Dice score and loss function, began to flatten at around 50 epochs. Figure 4 provides examples of the Dice coefficient corresponding to the DeepnCCA segmentations of two patients. Figure 5

Table 4. Clinical Correlates of FreeSurfer and DeepnCCA

Correlation between FreeSurfer and DeepnCCA outputs (<i>n</i> = 504 scans): .45, <i>P</i> < .001		
Clinical assessment	FreeSurfer (volume)	DeepnCCA (normalized area)
EDSS \pm 6 months (<i>n</i> = 204)	-.17, <i>P</i> = .013	-.22, <i>P</i> = .002
EDSS future (<i>n</i> = 227) ^a	-.17, <i>P</i> = .012	-.26, <i>P</i> < .001
SDMT future (<i>n</i> = 151) ^b	.24, <i>P</i> = .003	.26, <i>P</i> = .001

^a Average number of years between scan and EDSS was 9.7 \pm 3.9 years.

^b Average number of years between scan and SDMT was 8.1 \pm 4.0 years.

^{a,b} Data represent mean \pm standard deviation.

EDSS, Expanded Disability Status Scale; SDMT, Symbol Digit Modalities Test; *n*, sample size.

shows the output from the algorithm from four representative patients with MS.

Corpus Callosum Segmentations

The corpus callosum segmentations had a slightly lower similarity (mean Dice score 89.3 \pm .76% using a 10-fold cross-validation, see Table 3). The model performance as measured by the loss function had a similar curve to that of the intracranial area which began flattening at around 50 epochs.

Clinical Correlates

FreeSurfer versus DeepnCCA

There were 504 MRI acquisitions containing both 2D T₂-weighted and 3D T₁-weighted scans, allowing a qualitative comparison between DeepnCCA and FreeSurfer. The DeepnCCA processing time was 2.5 seconds per segmentation compared to just over 10 hours per CPU core for FreeSurfer on a MacBook Pro with 3.3 GHz Dual-Core Intel Core i7 and 8 GB 2133 MHz LPDDR3 RAM (Apple Inc., Cupertino, CA, USA). No graphics processing unit is required to use either DeepnCCA or FreeSurfer. The correlation between FreeSurfer's corpus callosum volume and DeepnCCA's normalized corpus callosum area was .45 (*P* < .001) as shown in Table 4. Furthermore, to evaluate the performance relative to the clinical evaluation of patients, both EDSS and SDMT scores were correlated to the output of each respective tool (Table 4). For correlations with future disability, the last available time point was chosen, providing a clinical follow-up on average 8-10 years after the MRI scan. SDMT scores were only available in the Swedish MS registry for timepoints after the scans had been performed, as such only a correlation with future SDMT was possible.

Shape Analysis

Principal Component Analysis

The PCA revealed 13 separate components that accounted for 99% of the variability in corpus callosum morphology in the patient population. After Bonferroni correction, only EDSS, representing physical disability, correlated significantly with a principal component. Figure 3A visually presents the shape variation for the significant principal component, reflecting a thinner and more angled corpus callosum with higher disability.

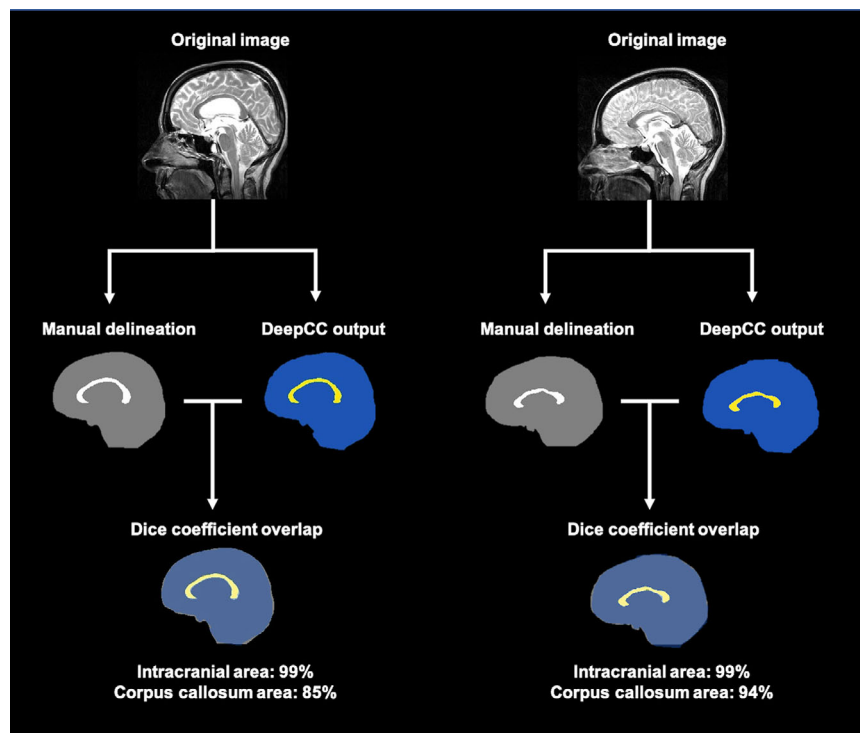


Fig 4. Examples of DeepnCCA segmentations compared to manual segmentations. Comparison of manual versus DeepnCCA segmentations in two representative patients. The Dice coefficient is used as a metric of similarity for both intracranial area and corpus callosum area.

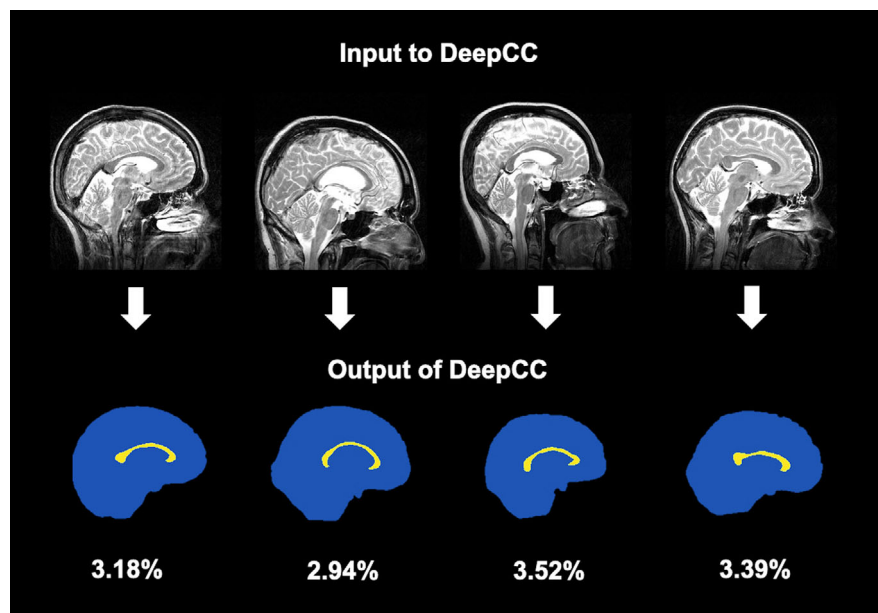


Fig 5. Examples of DeepnCCA algorithm output. Examples of the DeepnCCA output from four representative patients. The normalized corpus callosum value is calculated by dividing the corpus callosum area by the intracranial area.

Bending Angle and Mean Thickness

The PCA results inspired the measurement of both the mean thickness and bending angle of the corpus callosum. Figure 3B shows that the EDSS scores significantly correlated with both the thickness ($r = -.30$; $P < .001$) and the bending angle ($r = -.17$; $P = .043$). While the normalized SDMT z-score only

significantly correlated with the thickness of the corpus callosum ($r = .28$, $P < .001$).

Discussion

In this study, we developed an openly available deep learning algorithm, DeepnCCA, specifically adapted to normalized cor-

pus callosum segmentations in MS patients. Its output showed high similarity with gold standard manual segmentations. DeepnCCA segmentations had numerically higher correlations with disability as compared to the computationally more demanding FreeSurfer segmentations. Shape analysis further found that a thinner and more angled corpus callosum correlates with a higher degree of clinical disability.

For neurodegenerative biomarkers, it is important to consider the pathophysiology of interest. In MS, it has been shown that the corpus callosum atrophies more than other brain structures and therefore may represent a more sensitive marker than global or other regional brain atrophy metrics.⁸ It has also been shown that corpus callosum atrophy is a strong predictor of clinical disability as assessed by EDSS.³⁰ In this study, future cognitive and physical disability scores were available, making it possible to measure an association between baseline atrophy and cognitive and physical disability an average of 8 and 10 years into the future, respectively. It was found that DeepnCCA produced numerically higher correlations to these metrics as compared to FreeSurfer's volumetric output. The novelty of DeepnCCA is due to several aspects; disease specificity, potential clinical value, accuracy, and computational speed, but also that it is applicable to conventional 2D MRI scans. The last aspect was of particular importance here, as it allows for recording of a relevant proxy for neurodegeneration in MS using historic non-3D MRI data.

Previous studies employing corpus callosum segmentations have reported varying levels of success. Manual segmentations have been shown useful and are often used as gold standard when comparing other methods.^{9,12} However, manual measurements are tedious and have an inherent inter- and intrarater variability. In an article by Park et al applying Bayesian inference consisting of sparse representation error and multi-atlas voting, a Dice coefficient was found varying between 87% and 96% for corpus callosum segmentations.³¹ Furthermore, we have recently shown that FreeSurfer, relative to manual segmentations, has a corpus callosum Dice score of 76-81% for the cross-sectional stream and 80-85% for the longitudinal stream.¹² In a study by Van Schependorn et al using an active shape model, the mean Dice score for MS corpus callosum segmentation was 90%, but the algorithm failed to segment one of the patients due to MS-related atrophy.³² A recent review comparing model-based, region-based, thresholding, and machine learning techniques for segmenting the corpus callosum in other diseases has identified that the results tended to be better for the machine learning techniques. However, there are various inherent challenges when comparing methods, such as limitations due to variance in the type of MRI scanner, sequence, and field strength used as input.³³

Many machine learning algorithms in medicine are limited by the amount of data available for training, in that they become limited in their performance to the variance inherent in patient data. It is a fundamental concept of convolutional neural networks that performance increases with larger sample sizes.³⁴ This highlights the importance of data augmentation to increase the sample size, which we established through rotation, scaling, translation, horizontal flip, and a Gaussian noise filter. It is equally important for an algorithm to cater to the disease at hand, in order to handle the common variations. A phenotypically heterogeneous disease, such as MS, where midsagittal structures, such as the corpus callosum often contains lesions,

may complicate segmentation. DeepnCCA, however, has been trained to handle these common variations.

We applied a few modifications to DeepnCCA that deviate slightly from the original U-Net architecture by Ronneberger et al,²⁶ in order to strengthen the algorithm. We performed batch normalization, which allows for much higher learning rates and speeds of training through the network.³⁵ We also added a dropout layer, minimizing coadaptations in the network and thus minimizing overfitting and providing a regularization effect.³⁶ Similar to the original U-net, we applied a rectified linear unit, the most commonly used activator in deep learning networks.³⁷ These modifications allowed us to make an algorithm that is updated with improvements that result in more accurate segmentations.

We found that the shape of the corpus callosum varies with the physical and cognitive disability. With more severe physical disability, we found both a thinner, that is, more atrophied corpus callosum, and increased curvature. For cognitive disability, we found a significant correlation with a thinning of the corpus callosum. It has previously been shown that the shape of the corpus callosum significantly differs between healthy controls and MS patients.¹⁹ As the corpus callosum gets thinner, other structures simultaneously atrophy,³⁸ leading to ventriculomegaly, whereby the corpus callosum theoretically gets displaced upward, which could explain the more angled appearance. In a study by Van Schependorn et al, it was shown that the corpus callosum midbody was thinner in mildly disabled MS patients versus controls.³⁸ Our results indicate that this thinning seems to continue and that it relates significantly to cognitive and physical disability. Moreover, discovering features of MS atrophy may allow for further future machine learning modeling where features can be used to predict prognosis and progression.

There are certain limitations to this study. Due to the clinical measures and MRIs being acquired in a clinical setting in this prospective cohort study, there was a discrepancy in temporal alignment between MRI scans and clinical measurements, which may impact the accuracy of the described correlations. In addition, as the measure of cognitive processing speed, SDMT, was added to the Swedish MS registry at a later stage, correlations between change from baseline to follow up were not possible. From a technical point, it is important to note that 2D scans acquired axially may not allow extraction of a mid-sagittal slice, potentially limiting the application of DeepnCCA. Another limitation is the use of only one scanner with only one set of scan parameters, thus reducing the external validity and generalizability of the results. Furthermore, the manual selection of the anatomically located middle slice introduces an inherent variability between users, which may increase the variability in normalized corpus callosum outputs from the same patient. Future work should, therefore, focus on expanding the training to a larger variety of scans alongside an automatized process for slice selection. For this study, DeepnCCA was trained on 2D T₂-weighted images to allow for a fast automatic neurodegenerative biomarker in datasets without more recent 3D imaging. Future developments will include applications to other image weightings and 3D images, allowing for more direct head-to-head comparisons with volumetric software.

In conclusion, we present a de novo deep learning algorithm that can automatically extract the normalized cor-

pus callosum area, a robust MRI-biomarker of neurodegeneration in MS, in a quick and reliable manner. The algorithm is well adapted to handle the type of neurodegeneration that underlies progression of disability accumulation in MS and the specific morphological changes that occur in this important brain region. Importantly, this tool makes it possible to evaluate neurodegeneration using historic 2D data, in turn allowing for capturing the long-term natural course of MS and the effects of various interventions in large MS cohorts.

References

- Filippi M, Bar-Or A, Piehl F, et al. Multiple sclerosis. *Nat Rev Dis Primers* 2018;4:43.
- Filippi M, Rocca MA, Ciccarelli O, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. *Lancet Neurol* 2016;15:292-303.
- Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol* 2018;17:162-73.
- Inglese M, Petracca M. MRI in multiple sclerosis: clinical and research update. *Curr Opin Neurol* 2018;31:249-55.
- Dekker I, Eijlers AJC, Popescu V, et al. Predicting clinical progression in multiple sclerosis after six and twelve years. *Eur J Neurol* 2019;26:893-902.
- Granberg T, Martola J, Bergendal G, et al. Corpus callosum atrophy is strongly associated with cognitive impairment in multiple sclerosis: results of a 17-year longitudinal study. *Mult Scler* 2015;21:1151-8.
- Martola J, Stawiarz L, Fredrikson S, et al. Progression of non-age-related callosal brain atrophy in multiple sclerosis: a 9-year longitudinal MRI study representing four decades of disease development. *J Neurol Neurosurg Psychiatr* 2007;78:375-80.
- Ouellette R, Bergendal Å, Shams S, et al. Lesion accumulation is predictive of long-term cognitive decline in multiple sclerosis. *Mult Scler Relat Disord* 2018;21:110-6.
- Granberg T, Bergendal G, Shams S, et al. MRI-defined corpus callosal atrophy in multiple sclerosis: a comparison of volumetric measurements, corpus callosum area and index. *J Neuroimaging* 2015;25:996-1001.
- Claesson T-B, Putaala J, Shams S, et al. Comparison of manual cross-sectional measurements and automatic volumetry of the corpus callosum, and their clinical impact: a study on type 1 diabetes and healthy controls. *Front Neurol* 2020;11:27.
- Magnotta VA, Heckel D, Andreasen NC, et al. Measurement of brain structures with artificial neural networks: two- and three-dimensional applications. *Radiology* 1999;211:781-90.
- Platten M, Martola J, Fink K, et al. MRI-based manual versus automated corpus callosum volumetric measurements in multiple sclerosis. *J Neuroimaging* 2019;30:1-7.
- Yasaka K, Akai H, Kunimatsu A, et al. Deep learning with convolutional neural network in radiology. *Jpn J Radiol* 2018;36:257-72.
- Salem M, Valverde S, Cabezas M, et al. A fully convolutional neural network for new T2-w lesion detection in multiple sclerosis. *Neuroimage Clin* 2020;25:102149.
- Narayana PA, Coronado I, Sujit SJ, et al. Deep-learning-based neural tissue segmentation of MRI in multiple sclerosis: effect of training set size. *J Magn Reson Imaging* 2019;51:1487-96.
- Gabr RE, Coronado I, Robinson M, et al. Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: a large-scale study. *Mult Scler* 2020;26:1217-26.
- Prendergast DM, Ardekani B, Ikuta T, et al. Age and sex effects on corpus callosum morphology across the lifespan. *Hum Brain Mapp* 2015;36:2691-702.
- Ardekani BA, Bachman AH, Figarsky K, et al. Corpus callosum shape changes in early Alzheimer's disease: an MRI study using the OASIS brain database. *Brain Struct Funct* 2014;219:343-52.
- Sigirli D, Ercan I, Ozdemir ST, et al. Shape analysis of the corpus callosum and cerebellum in female MS patients with different clinical phenotypes. *Anat Rec (Hoboken)* 2012;295:1202-11.
- Kavaliunas A, Manouchehrinia A, Stawiarz L, et al. Importance of early treatment initiation in the clinical course of multiple sclerosis. *Mult Scler* 2017;23:1233-40.
- Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis. *Neurology* 2014;83:278-86.
- McDonald WI, Compston A, Edan G, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Ann Neurol* 2001;50:121-7.
- Polman CH, Reingold SC, Edan G, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the 'McDonald Criteria'. *Ann Neurol* 2005;58:840-6.
- Polman CH, Reingold SC, Banwell B, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology* 2011;69:292.
- Brochet B, Deloire MSA, Bonnet M, et al. Should SDMT substitute for PASAT in MSFC? A 5-year longitudinal study. *Mult Scler* 2008;14:1242-49.
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *arXiv e-prints* 2015;1505:arXiv.org/abs/1505.04597. 2020.
- Kingma DP, Ba J. Adam: a method for stochastic optimization. <http://arxiv.org/abs/1412.6980>. 2020.
- Shah M, Xiao Y, Subbanna N, et al. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. *Med Image Anal* 2011;15:267-82.
- Leventon ME, Grimson WEL, Faugeras O. Statistical shape influence in geodesic active contours. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2000;1:316-23.
- Vaneckova M, Kalincik T, Krasensky J, et al. Corpus callosum atrophy – a simple predictor of multiple sclerosis progression: a longitudinal 9-year study. *ENE* 2012;68:23-7.
- Park G, Kwak K, Seo SW, et al. Automatic segmentation of corpus callosum in midsagittal based on Bayesian inference consisting of sparse representation error and multi-atlas voting. *Front Neurosci* 2018;12:629.
- Van Schependom J, Jain S, Cambron M, et al. Reliability of measuring regional callosal atrophy in neurodegenerative diseases. *Neuroimage Clin* 2016;12:825-31.
- Cover GS, Herrera WG, Bento MP, et al. Computational methods for corpus callosum segmentation on MRI: a systematic literature review. *Comput Methods Programs Biomed* 2018;154:25-35.
- Akkus Z, Galimzianova A, Hoogi A, et al. Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging* 2017;30:449-59.
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. <http://arxiv.org/abs/1502.03167>. 2020.
- Strivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15:1929-58.
- Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks. *Pattern Recognit* 2018;77:354-77.
- Van Schependom J, Gielen J, Laton J, et al. The effect of morphological and microstructural integrity of the corpus callosum on cognition, fatigue and depression in mildly disabled MS patients. *Magn Reson Imaging* 2017;40:109-14.