

<https://helda.helsinki.fi>

---

## Machine learning-based prediction of motor status in glioma patients using diffusion MRI metrics along the corticospinal tract

Shams, Boshra

2022-05-02

---

Shams , B , Wang , Z , Roine , T , Aydogan , D B , Vajkoczy , P , Lippert , C , Picht , T & Fekonja , L S 2022 , ' Machine learning-based prediction of motor status in glioma patients using diffusion MRI metrics along the corticospinal tract ' , Brain communications , vol. 4 , no. 3 , 141 . <https://doi.org/10.1093/braincomms/fcac141>

---

<http://hdl.handle.net/10138/346329>

<https://doi.org/10.1093/braincomms/fcac141>

---

publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*









*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# BRAIN COMMUNICATIONS

## Machine learning-based prediction of motor status in glioma patients using diffusion MRI metrics along the corticospinal tract

 Boshra Shams,<sup>1,2</sup>  Ziqian Wang,<sup>1</sup>  Timo Roine,<sup>3,4</sup>  Dogu Baran Aydogan,<sup>3,5,6</sup>  Peter Vajkoczy,<sup>1</sup>  Christoph Lippert,<sup>7,8</sup>  Thomas Picht<sup>1,2</sup> and  Lucius S. Fekonja<sup>1,2</sup>

Along tract statistics enables white matter characterization using various diffusion MRI metrics. These diffusion models reveal detailed insights into white matter microstructural changes with development, pathology and function. Here, we aim at assessing the clinical utility of diffusion MRI metrics along the corticospinal tract, investigating whether motor glioma patients can be classified with respect to their motor status. We retrospectively included 116 brain tumour patients suffering from either left or right supratentorial, unilateral World Health Organization Grades II, III and IV gliomas with a mean age of  $53.51 \pm 16.32$  years. Around 37% of patients presented with preoperative motor function deficits according to the Medical Research Council scale. At group level comparison, the highest non-overlapping diffusion MRI differences were detected in the superior portion of the tracts' profiles. Fractional anisotropy and fibre density decrease, apparent diffusion coefficient axial diffusivity and radial diffusivity increase. To predict motor deficits, we developed a method based on a support vector machine using histogram-based features of diffusion MRI tract profiles (e.g. mean, standard deviation, kurtosis and skewness), following a recursive feature elimination method. Our model achieved high performance (74% sensitivity, 75% specificity, 74% overall accuracy and 77% area under the curve). We found that apparent diffusion coefficient, fractional anisotropy and radial diffusivity contributed more than other features to the model. Incorporating the patient demographics and clinical features such as age, tumour World Health Organization grade, tumour location, gender and resting motor threshold did not affect the model's performance, revealing that these features were not as effective as microstructural measures. These results shed light on the potential patterns of tumour-related microstructural white matter changes in the prediction of functional deficits.

- 1 Department of Neurosurgery, Charité - Universitätsmedizin Berlin, Klinik für Neurochirurgie mit Arbeitsbereich Pädiatrische Neurochirurgie, Campus Charité Mitte, Charitéplatz 1, 10117 Berlin, Germany
- 2 Cluster of Excellence: 'Matters of Activity. Image Space Material', Humboldt University Berlin, Berlin, Germany
- 3 Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland
- 4 Turku Brain and Mind Center, University of Turku, Turku, Finland
- 5 Department of Psychiatry, Helsinki University and Helsinki University Hospital, Helsinki, Finland
- 6 A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland
- 7 Digital Health - Machine Learning, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany
- 8 Hasso Plattner Institute for Digital Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Correspondence to: Lucius S. Fekonja  
Charité - Universitätsmedizin Berlin, Klinik für  
Neurochirurgie mit Arbeitsbereich Pädiatrische Neurochirurgie  
Campus Charité Mitte, Charitéplatz 1, 10117 Berlin, Germany  
E-mail: [lucius.fekonja@charite.de](mailto:lucius.fekonja@charite.de)

**Keywords:** machine learning; support vector machine; tractography; diffusion MRI; corticospinal tract

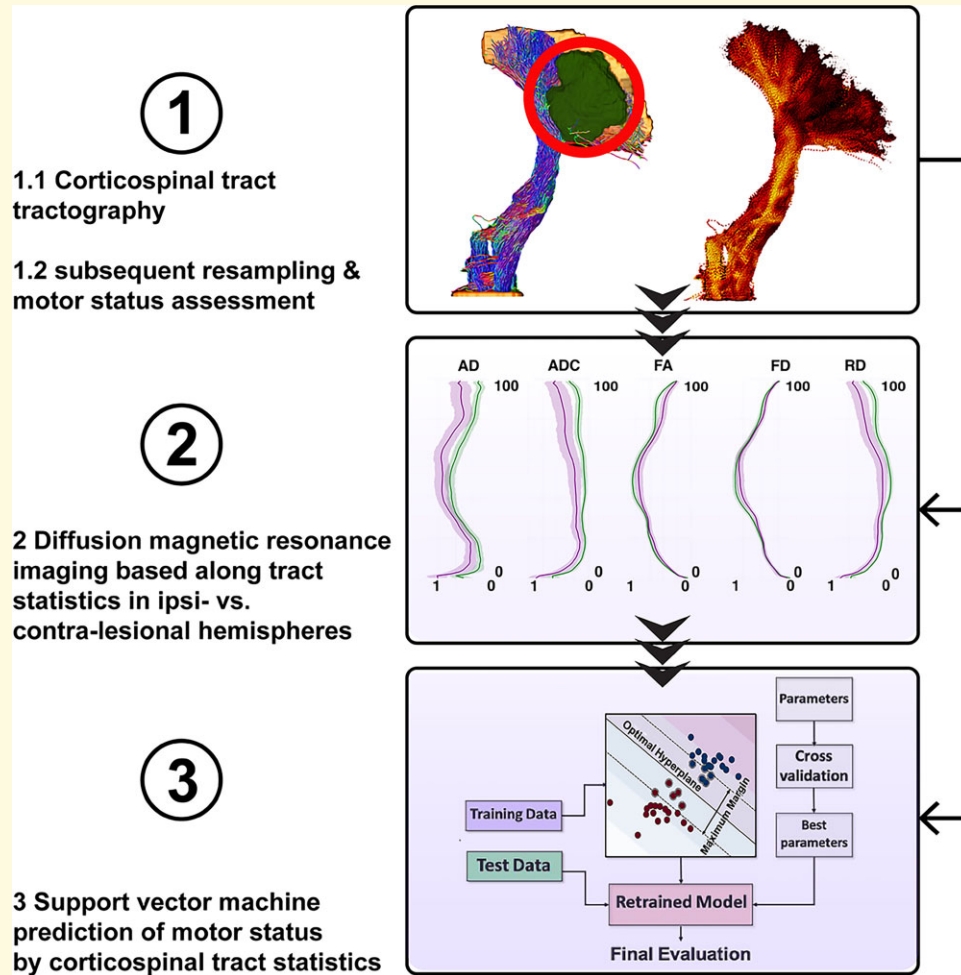
Received October 16, 2021. Revised March 01, 2022. Accepted May 24, 2022. Advance access publication May 27, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abbreviations:** AD = axial diffusivity; ADC = apparent diffusion coefficient; BMRC = British Medical Research Council; CCA = canonical correlation analysis; CSD = constrained spherical deconvolution; CST = corticospinal tract; CV = cross validation; dMRI = diffusion magnetic resonance imaging; DTI = diffusion tensor imaging; FA = fractional anisotropy; FBA = fixel-based analysis; FD = fibre density; FDR = false discovery rate; FOD = fibre orientation distribution; GBM = glioblastoma; KNN = k-nearest neighbour; KU = kurtosis; M = mean; Mdn = median; ML = machine learning; MRC = medical research council; PCA = principal component analysis; RD = radial diffusivity; RFE = recursive feature elimination; RMT = resting motor threshold; ROI = Region of interest; SK = skewness; STD = standard deviation; SVM = support vector machine; TMS = transcranial magnetic stimulation; WM = white matter

## Graphical Abstract



## Introduction

Gliomas are known as the most frequent and malignant human brain tumours, characterized by poor prognosis and high morbidity.<sup>1</sup> Gliomas infiltrating the motor system potentially cause various degrees of damage to the white matter (WM) architecture and might lead to substantial motor function impairments.<sup>1,2</sup> Diffusion MRI (dMRI)<sup>3,4</sup> has shown potential by enabling non-invasive delineation of the WM pathways known as tractography<sup>5-9</sup>. Tractography has been frequently used for preoperative planning or analysing the effects of the tumour on WM and structural connectivity, for example, to investigate tumour infiltration and its impact on surrounding tissues.<sup>10,11</sup>

Along tract statistics enables WM characterization using various dMRI metrics.<sup>12-14</sup> These measures have gained great interest since they reveal insights into WM development, function and disease.<sup>15</sup> Along tract diffusion tensor imaging (DTI)-derived metrics such as apparent diffusion coefficient (ADC; a measure of the overall diffusivity), axial diffusivity (AD; the diffusion rate along the main axis of diffusion), fractional anisotropy (FA; the directional preference of diffusion) or radial diffusivity (RD; rate of diffusion in the transverse direction) and a more complex dMRI-based metric, namely fibre density (FD) have been previously used to study tumour-induced local microstructural changes.<sup>16,17</sup> Here, we investigate whether along tract metrics can be used as predictive features to detect motor

function impairments. Recently, we have shown that the segmental DTI-derived metrics, such as ADC and FA are associated with motor deterioration in patients with brain tumours.<sup>18</sup> Multiple fibre populations are found in up to 90% of the WM voxels and 30–40% of these WM voxels contain more than three fibre populations.<sup>19–22</sup> Moreover, non-WM contamination is found in more than a third of the WM voxels<sup>23</sup> and multi-tissue constrained spherical deconvolution (CSD) methods<sup>24–27</sup> have been used to account for it. As a result, CSD-based metrics in addition to DTI metrics (such as AD, ADC, FA or RD) are critical. By estimating fibre orientation distributions (FODs) in each voxel based on the expected signal from a single collinearly oriented fibre population, CSD can discriminate complex fibre populations.<sup>28</sup> Probabilistic tractography algorithms, such as the iFOD2, have been proposed to overcome the limitations of tensor-based tractography methods by using the rich information in FODs.<sup>29</sup> A complete picture of the underlying WM architecture is critical for risk assessment, neurosurgical planning and as well for prediction models.<sup>30</sup> To that end, modern CSD-based FD and fixel-based analysis (FBA) approaches, in addition to traditional DTI methods, provide promising opportunities because they are related to the intra-axonal restricted compartment that is limited to a given fibre orientation within a voxel.<sup>31,32</sup> More recently, we used FD for fibre orientation-specific study of dMRI properties along the tract in relation to infiltrating tumours,<sup>16</sup> which was previously focused on group-based analyses.<sup>31</sup> Yet, research lacks the individual and tract-specific characterization of WM microstructure investigating the association between tumour impact on structural connectivity and clinical assessment.

Machine learning (ML) methods have recently gained remarkable success in clinical applications such as diagnostic, prognostic and predictive analytics using various modalities of MRI scans.<sup>33–36</sup>

Here, we employed ML methods using along corticospinal tract (CST)-related dMRI metrics (e.g. AD, ADC, FA, FD and RD) to predict motor deficits in patients with motor-related glioma, focusing on individual diagnosis rather than GroupWise comparisons. We used ML methods based on support vector machines (SVMs) which is a powerful method, easy to interpret and well suitable method to handle large dimensional data sets.<sup>37–39</sup> SVM has been used in various clinical applications such as tumour segmentation and classification, e.g. to distinguish low-grade gliomas from high-grade gliomas.<sup>40–43</sup> In previous studies, glioma grading has been performed using resting-state functional MRI<sup>44</sup> and radiomic.<sup>45</sup> In addition, it has been used as a tool for non-invasive prediction of tumour consistency to classify the tumour as soft or firm for preoperative planning.<sup>46</sup> Furthermore, it is known as an analysis tool for predicting the prognosis and survival time of tumour patients using multi-modal imaging.<sup>35,47</sup> Considering our tumour patients' cohort and multi-modal quantitative assessment of WM along CST based on dMRI metrics, we investigated our hypotheses using SVM-based analysis.

We designed our SVM models with an embedded feature selection method called recursive feature elimination (RFE)<sup>48,49</sup> using histogram-based features of dMRI-based tract profiles. The histogram-based analysis is one of the most useful methods in many neuroimaging applications, e.g. classification and clustering tasks by which we can summarize and preserve more information from first-order statistics of the original data than simple averaging of the data values.<sup>50,51</sup> A histogram model can be specified for a specific image feature type independently of any real image content.<sup>50,51</sup>

In addition, we developed an SVM model using principal component analysis (PCA)-derived<sup>52,53</sup> components, regardless of histogram-based features. PCA transforms data from high-dimensional space (all segmental information of dMRI-based tract profiles) into a low-dimensional space.

Furthermore, patient demographics and clinical variables including the resting motor threshold (RMT), a transcranial magnetic stimulation (TMS)-derived neurophysiological marker, were incorporated and fed into our designed models. We then assessed the impact of all these features, e.g. demographics, clinical and microstructural features, on the performance of the model.

## Materials and methods

### Patient cohort

We included 116 left- and right-handed adult patients in this retrospective study (43 females, 73 males, average age =  $53.51 \pm 16.32$ , age range = 20–87). Only patients with an initial diagnosis of supratentorial, unilateral World Health Organization (WHO) Grades II, III and IV gliomas (16 WHO Grade II, 23 WHO Grade III and 77 WHO Grade IV) were included. All tumours were infiltrating or immediately adjacent to M1 and/or the CST either in the left or right hemisphere. Patients with recurrent tumours, previous radiochemotherapy or multilocular tumours were not included. The motor status was graded preoperatively according to the medical research council (MRC) scale for muscle power. Grade, 0 means no muscle power, and 5 means full muscle strength. All patients with MRC < 5 were assigned to the group with motor deficits (Class 1), and others (MRC = 5) were assigned to the group without motor deficits (Class 0).

### Image acquisition

Clinical MRI data were acquired preoperatively at Charité University Hospital, Berlin, Department of Neuroradiology over the past years, and data acquisition started before HARDI techniques were commonly used. The centre performed scans on 3 T Siemens Skyra scanner with dedicated 32-channel head/neck coil. The protocol included whole-brain high-resolution structural data, contrast-enhanced T1-weighted images, with TR/TE/TI 2300/2.32/900 ms flip

angle = 9°, field of view = 256 × 256, 192 sagittal slices, 1 mm isotropic resolution, acquisition time: 5 min as well as a single shell diffusion-weighted volume with TR/TE 7500/95 ms, 2 × 2 × 2 mm<sup>3</sup> voxels, 128 × 128 matrix, 60 axial slices, with 40 equally distributed orientations for diffusion-sensitizing gradients at *b*-value of 1000 s/mm<sup>2</sup>, for a total acquisition time of 12 min.

## Transcranial magnetic stimulation

Non-invasive functional motor mapping of both ipsilesional and contralesional hemispheres was performed in each patient using navigated TMS (nTMS) with NeXimia Navigated Brain Stimulation (Nexstim Oy, Helsinki, Finland). Each patient's head was registered to the structural MRI and the composite muscle action potentials were captured by the integrated electromyography unit (sampling rate 3 kHz, resolution 0.3 mV; Neuroline 720, Ambu). The muscle activity (motor evoked potential, MEP amplitude ≥ 50 μV) was recorded by surface electrodes on the abductor pollicis brevis and first dorsal interosseous. Initially, the first dorsal interosseous hotspot, defined as the stimulation area that evoked the strongest MEP, was determined. Subsequently, the resting motor threshold, defined as the lowest stimulation intensity that repeatedly elicits MEPs, was defined using a threshold-hunting algorithm within the Nexstim eximia software. Mapping was performed at 105% resting motor threshold and 0.25 Hz. All MEP amplitudes > 50 μV (peak to peak) were considered as motor positive responses and exported in the definitive mapping.<sup>54</sup> The subject-specific positive responses of the first dorsal interosseous were exported as binary 3 mm<sup>3</sup> voxel masks per response in the T1 image space.

## Preprocessing and processing of MRI data

Preprocessing and processing of MRI data were performed as described earlier.<sup>16</sup> Briefly, all T1 images were linearly (affine) registered to the dMRI data sets using advanced normalization tools (ANTs).<sup>55,56</sup> Furthermore, we registered the human motor area template (HMAT) atlas to subject space with ANTs using the Symmetric normalization transformation model<sup>55,57</sup> to obtain M1 seeding region of interests (ROIs).<sup>55,57</sup> The preprocessing of dMRI data included the following and was performed within MRtrix3<sup>58</sup> in sequential order: denoising,<sup>59</sup> removal of Gibbs ringing artefacts,<sup>60</sup> correction of subject motion,<sup>61</sup> eddy currents<sup>62</sup> and susceptibility-induced distortions<sup>63</sup> in FSL,<sup>64</sup> and subsequent bias field correction with ANTs N4.<sup>65</sup> Each dMRI data set and processing step was visually inspected for outliers and artefacts. Scans with excessive motion were initially excluded based on a predefined threshold (if >10% outlier slices; however, this was not the case in the current cohort). We upsampled the dMRI data to a 1.3 mm isotropic voxel size before computing FODs to increase anatomical contrast and improve downstream tractography results and

statistics.<sup>66</sup> To obtain AD, ADC, FA and RD scalar maps, we first used diffusion tensor estimation using an iteratively reweighted linear least squares estimator, resulting in scalar maps of tensor-derived parameters.<sup>3,67</sup> For voxel-wise modelling, we used a robust and fully automated and unsupervised method. This method allowed us to obtain three-tissue response functions for white and grey matter and cerebrospinal fluid from our data with the use of spherical deconvolution for subsequent usage in multi-tissue CSD-based tractography.<sup>24,27,68</sup>

## Tractography

Probabilistic multi-tissue tractography was performed based on the WM FODs with the iFOD2 algorithm<sup>69</sup> as described earlier,<sup>16</sup> with the slight modification of using the above-mentioned HMAT atlas-derived M1 seeding ROI.<sup>70</sup> In brief, an inclusion ROI was defined in the medulla oblongata, tracking parameters were set to default with an FOD amplitude cutoff value of 0.1, a streamline minimum length of 5 × voxel size, and a maximum streamline length of 100 × voxel size. For each CST tractogram, we computed default *n* = 5000 streamlines per hemisphere. Each streamline per tractogram was resampled along its length to 100 equidistant points. Subsequently, we mapped AD, ADC, FA, FD and RD scalar metrics along the derived 100 equidistant points per streamline.

## Data preparation

We generated dMRI-based CST profiles, by which AD, ADC, FA, FD and RD were quantified at 100 segments along the CST using the values of the 100 points per streamline. To create a tract profile that is robust to outliers, we used two different methods and compared the results. In the first method, we calculated the median (Mdn) values across the 5000 streamlines per tractogram along its 100 segments. In the second method, we computed the segment-wise weighted mean (M) of the dMRI measures across streamlines. The streamline-wise contribution was weighted by the inverse Mahalanobis distance of the streamlines from the tract core (M). Streamlines that were more distant from M were considered less important.<sup>15</sup> We used both ipsi- and contralesional CST profiles as input features (predictor variables); thus, the dimension of the imaging-based feature space was 1000 (5 metrics × 2 hemispheres × 100 segments = 1000 features).

## Statistical analysis

Statistical analysis and data visualization were carried out using Python 3.8.6. The main packages used were Scipy,<sup>71</sup> Seaborn,<sup>72</sup> Statsmodels<sup>73</sup> and Matplotlib.<sup>74</sup> To compare categorical variables, Fisher's exact test (when the expected frequency was less than five per category) or Pearson's  $\chi^2$  test (for larger values) were employed. Two-tailed Student's *t*-tests or Mann-Whitney U tests (Wilcoxon rank-sum test) were performed to compare continuous variables. Effect

size ( $r$ ) for Mann–Whitney U statistics was calculated as the Z-statistic divided by the square root of the number of samples. A significance level of  $P < 0.05$  was considered as cut-off. With respect to multiple comparison analyses, statistically, significant  $P$ -values were false discovery rate (FDR) corrected using the Benjamini–Hochberg procedure.<sup>75</sup> For all univariate statistical analyses, Mdn-based tract profiles were used. Canonical correlation analysis (CCA) was performed as a multivariate correlation analysis to identify and measure the association among all dMRI-based extracted features (both in ipsi- and contralesional hemispheres) and age.<sup>76</sup> This analysis extracts meaningful information from a pair of data sets, dMRI-based features and age, by seeking pairs of linear combinations from two sets of variables with the maximum pairwise correlation. CCA was performed both on the patient’s cohort and on each patient’s group (Class 0; Class 1) separately. A more detailed analysis investigating the associations between each metric and age was also performed.

## SVM classification

SVM has gained a widespread application in the neuroimaging context as either classification or regression method.<sup>77,78</sup> An SVM classifier aims to find hyperplanes with maximal margins between classes. As a supervised ML method, SVM can be extended to complex instances that are not linearly separable using so-called kernel tricks.<sup>79,80</sup> Kernel techniques map input features from one space to a higher dimensional feature space in which different classes can be distinguished by a separating hyperplane. All ML analysis methods should be balanced between their predictive accuracy and descriptive power.<sup>81</sup> Accordingly, in the present study, we developed different models based SVM method using dMRI-based features, demographic and clinical variables to predict the motor status (Class 0; Class1) preoperatively (cf. Fig. 1).

Some segments along tract profiles were missed (not-a-number value) when the tract profiles were generated. These missing values were imputed using two different interpolation methods<sup>82</sup>: (i) Mdn and (ii) k-nearest neighbour (KNN).<sup>83</sup> Before fitting a model to our data, imputation of missing values and feature standardization were performed. To enable our classifier to learn from low and high variance metrics, we removed each feature’s M and scaled it to a unit variance ( $z$ -score). Training and test sets within each cross validation (CV) were standardized separately by M and standard deviation (STD) derived from the training set to prevent information leakage between testing and training data sets.

Four different SVM models were trained and tested using Mdn-based and Mahalanobis-based weighted M tract profiles with the above-mentioned interpolation methods for imputation of missing values (SVM\_1: Mdn-based imputation method and Mdn-based tract profile; SVM\_2: KNN-based imputation method and Mdn-based tract profile; SVM\_3: Mdn-based imputation methods and Mahalanobis-based weighted M tract profile; SVM\_4: KNN-based imputation method and Mahalanobis-based weighted M tract profile).

As a preprocessing step, to reduce the high dimensional imaging-based feature space, a set of statistical features was calculated as a high-level representation to measure different properties of dMRI-based tract profiles’ distributions. Descriptive statistics such as M as a central tendency, STD as a measure of variability and kurtosis (KU), and skewness (SK) as measures of shape were extracted as histogram-based features (Fig. 2). The tract profile statistics were calculated for ipsi- and contralesional tractograms (4 measures  $\times$  5 metrics  $\times$  2 hemispheres = 40 features) and were fed into the models. We further incorporated patient demographics and clinical data such as age, gender, tumour grade, tumour location and RMT ratio, fed them into the aforementioned models, and compared the results.

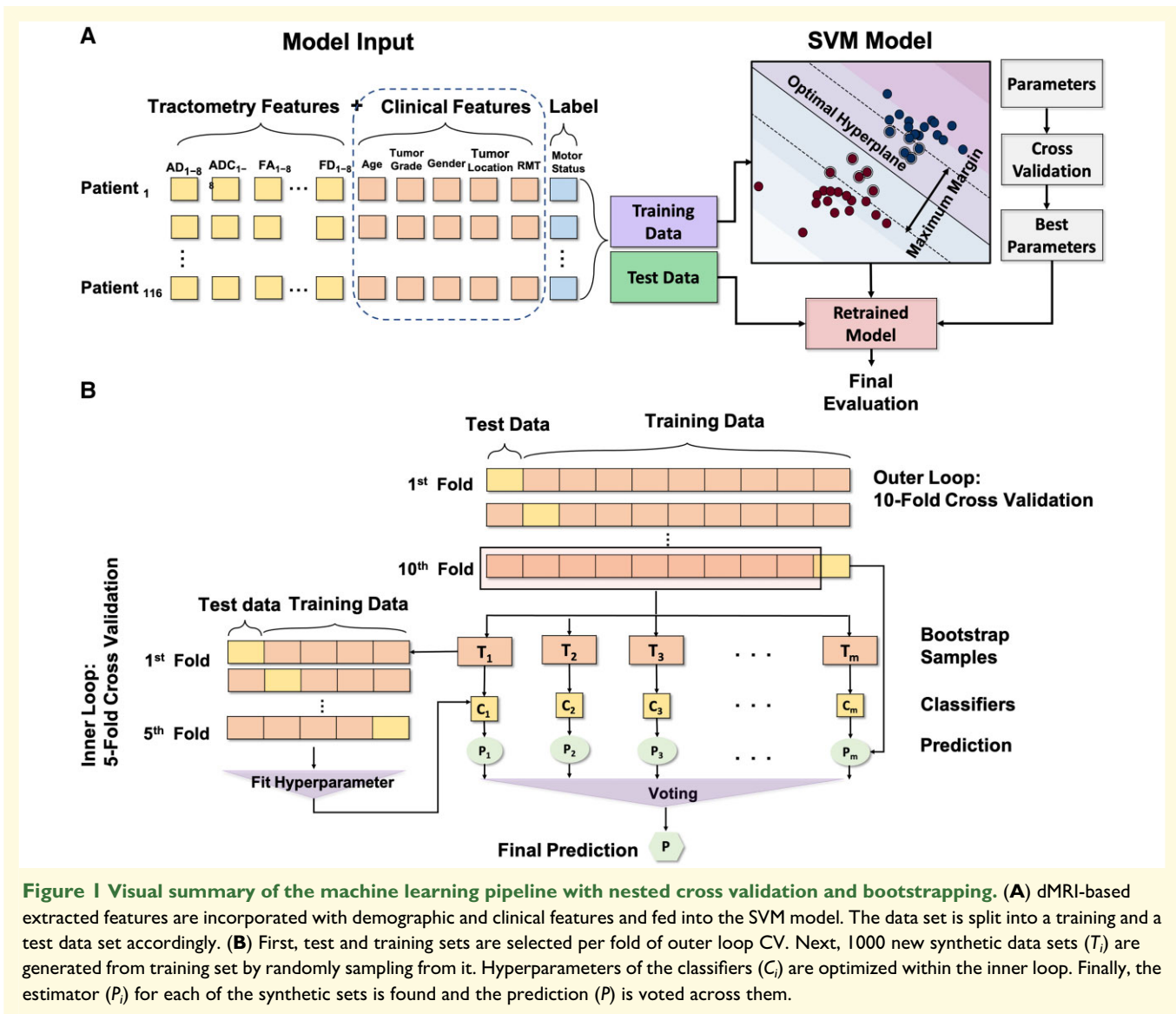
The most relevant features were selected using the RFE method based on SVM (SVM-RFE).<sup>48,49</sup> This method recursively removes features that contribute least to the prediction based on the linear SVM classifier weight coefficients before the actual learning phase. Subsequently, selected features were used to train and validate the SVM model with a linear kernel.

To investigate how well each dMRI metric (e.g. AD, ADC, FA, FD, RD) performed in classifying the patients with respect to their motor status, different SVM models were trained and tested, e.g. SVM\_AD, SVM\_ADC, SVM\_FA, SVM\_FD and SVM\_RD, with KNN-based imputation method for missing values and Mdn-based tract profile (cf. Table 1). To assess the predictive power of patient demographics and clinical variables regardless of imaging-based features (when ignoring the neuroimaging analysis pipeline), an SVM model (SVM\_clinical) was developed using only patients’ age, tumour WHO grade, tumour location, gender and RMT.

Additionally, a model was developed using all values of ipsi- and contralesional dMRI-based tract profiles without performing above-mentioned feature extraction method. To reduce the high-dimensional imaging-based feature space (1000), PCA<sup>52,53</sup> was performed on Mdn-based tract profiles, and the first four components were fed into an SVM model with the linear kernel (SVM\_5) using KNN-based imputation method (cf. Table 1).

We evaluated our models (SVM\_1-5; SVM\_AD-RD; SVM\_clinical) using nested CV with a 10-fold outer loop and a 5-fold inner loop. Our model key hyperparameter C for penalty<sup>84</sup> was optimized in the inner CV loop and the best performing model was applied to the outer CV loop test set to evaluate the model selected by the inner loop. C was tested from 0 to 10 with a 0.1 step size.

Our data set was imbalanced because the proportion of patients with motor deficits to patients without motor deficits was nearly 1 to 2 (cf. Table 2). We used a stratified 10-fold CV to ensure that class distributions in each data split matched the distribution in the complete training data set. We additionally assigned the class weights,  $w_j$  (Class 1: MRC < 0; Class 0: MRC = 5) inversely proportional to their respective frequencies as  $w_j = \frac{n_{\text{samples}}}{n_{\text{classes}} * n_{\text{samples}_j}}$ , where  $n_{\text{samples}}$  is the number of samples,  $n_{\text{classes}}$  is the number of classes and  $n_{\text{samples}_j}$  is the number of samples per class.



Bootstrap aggregating (bagging) has been introduced as a method to reduce the variance of a given estimator.<sup>85</sup> Bagging involves applying an estimator to multiple bootstrap samples and voting the results across them. These estimators can use CV themselves to select fine-tuning parameters trading off bias and variance of the bootstrap sample-specific candidate estimators. We used this approach in our models (SVM\_1-4; SVM\_AD-RD) with 1000 resampled training sets per fold of outer loop CV and lastly voted among all 1000 generated models.

We evaluated the performance of our models using the overall accuracy, the ratio of correctly predicted samples over the entire cohort, sensitivity, specificity and the area under the receiver operating curve (AUC).

We used various SVM models to predict tumour-related motor deficits. To this end, four SVM models (SVM\_1-4) were trained using histogram-based features of dMRI-based CST profiles, and one SVM model (SVM\_5) was trained using four PCA components of segmental

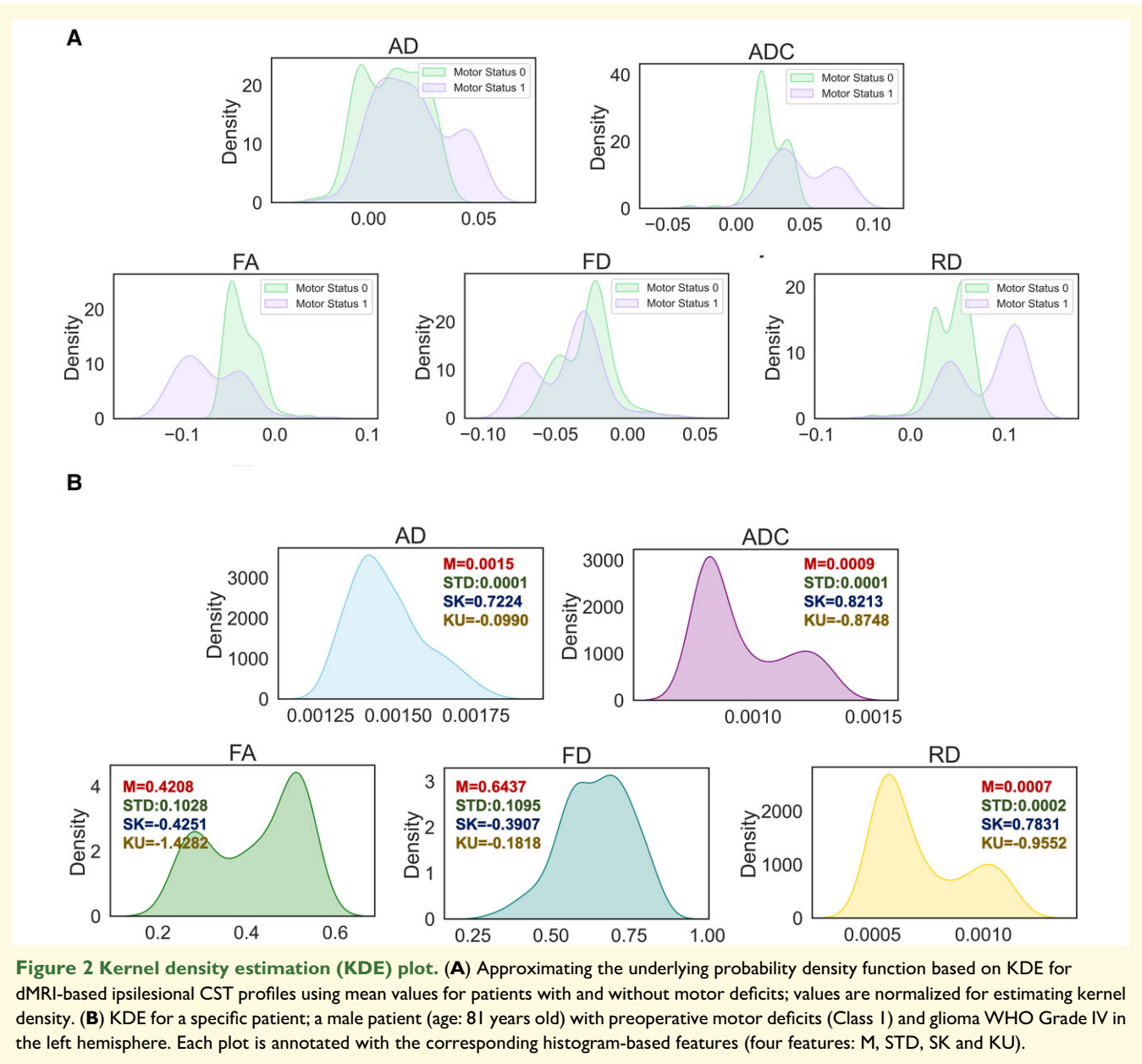
information of dMRI-based CST profiles. Finally, an additional SVM was trained (SVM\_Clinical) using clinical and demographic features.

## Data availability

Raw data that support the findings of this study are not publicly available due to information that could compromise the privacy of the research patients. However, the code we have used is openly available on <https://github.com/CUB-IGL/Machine-learning-based-prediction-of-motor-status-in-glioma-patients-using-dMRI-metrics-along-CST.git> and is referred to at the corresponding passage in the article.

## Results

Forty-five (37.9%) of the recruited 116 patients presented with preoperative motor deficits (MRC < 5; cf. Table 2).



There were no significant differences in gender ( $\chi^2[1, N = 116] = 0.51, p = 6.3e - 1$ ) or hemispheric pathology position ( $\chi^2[1, N = 116] = 0.08, p = 7.7e - 1$ ) in relation to motor deficits. Patients with motor deficits (Class 1) were older ( $58.64 \pm 15.45$ ) than patients without motor deficits (Class 0) ( $50.25 \pm 15.85$ ), with a highly significant difference between them ( $t[114] = 2.83, p = 5e - 3$ ), although a medium effect was found ( $g = 5.3e - 1$  95% CI = [0.155 - 0.914]). There were no significant differences in tumour locations and RMT ratio in both ipsi- and contralesional hemispheres in relation to motor deficits (cf. Table 2). We also found significant difference between the glioma WHO Grades III and IV between patients with and without motor deficits ( $\chi^2[1, N = 116] = 0.86, p = 3e - 2$ ).

Moreover, considering CCA, we found a strong positive correlation between all dMRI-based extracted features and

age, which was statistically significant ( $r_s(114) = 0.7, p = 3.3e - 18$ ). The correlation coefficients for the CCA are provided in Supplementary Table 1. This correlation was stronger in the group with motor deficits (Class 1;  $r_s(45) = 0.93, p = 6.12e - 21$ ). Considering each metrics separately, the correlation was weaker (AD:  $r_s(45) = -0.59, p = 1.64e - 5$ ; ADC:  $r_s(45) = 0.64, p = 2.32e - 6$ ; FA:  $r_s(45) = -0.42, p = 3.2e - 4$ ; FD:  $r_s(45) = -0.24, p = 0.1$ ; RD:  $r_s(45) = 0.51, p = 3.54e - 4$ ). In the group without motor deficits (Class 0), the dMRI-based extracted measures were negatively correlated to age ( $r_s(71) = -0.89, p = 1.0e - 25$ ). Considering each metrics separately, the correlation was weaker (AD:  $r_s(71) = 0.33, p = 4.0e - 4$ ; ADC:  $r_s(71) = 0.53, p = 2.07e - 6$ ; FA:  $r_s(71) = -0.49, p = 1.64e - 5$ ; FD:  $r_s(71) = -0.55, p = 4.87e - 7$ ; RD:  $r_s(71) = 0.58, p = 1.2e - 7$ ).



**Table 1** The overview of all SVM models

Model	Input data	Imputation method for missing values	Method for generating the tract profile	PCA components	Histogram-based features
SVM_1	dMRI metrics	Mdn	Mdn	No	Yes
SVM_2	dMRI metrics	KNN	Mdn	No	Yes
SVM_3	dMRI metrics	Mdn	Weighted-mean	No	Yes
SVM_4	dMRI metrics	KNN	Weighted-mean	No	Yes
SVM_5	dMRI metrics	KNN	Mdn	Yes	No
SVM_AD	dMRI (AD metric)	KNN	Mdn	No	Yes
SVM_ADC	dMRI (ADC metric)	KNN	Mdn	No	Yes
SVM_FA	dMRI (FA metric)	KNN	Mdn	No	Yes
SVM_FD	dMRI (FD metric)	KNN	Mdn	No	Yes
SVM_RD	dMRI (RD metric)	KNN	Mdn	No	Yes
SVM_clinical	Clinical + demographic	No	No	No	No

**Table 2** Demographic and neuropathological overview of the patient cohort

	Patients without motor deficits Class 0	Patients with motor deficits Class 1	P-value
<i>Demographics</i>			
Cohort size	71 (62%)	45 (37%)	–
Age	50.25 ± 15.85	58.64 ± 15.45	0.005
Female	25 (35%)	18 (41%)	0.63
Male	47 (65%)	26 (60%)	0.63
<i>Ipsilesional hemisphere</i>			
Left	30 (42%)	17 (39%)	0.77
Right	41 (57%)	28 (64%)	0.77
<i>Tumour location</i>			
Frontal	35 (49%)	25 (55%)	0.55
Parietal	19 (27%)	12 (27%)	0.91
Insular	10 (14%)	5 (11%)	0.77
Temporal	7 (10%)	3 (7%)	0.73
<i>Glioma WHO grade</i>			
II	13 (18%)	3 (7%)	0.8 <sup>a</sup> , 0.05 <sup>b</sup>
III	18 (25%)	5 (11%)	0.8 <sup>a</sup> , 0.03 <sup>c</sup>
IV	40 (55%)	37 (84%)	0.05 <sup>b</sup> , 0.03 <sup>c</sup>
<i>RMT* (V/m)</i>			
Ipsilesional hemisphere	33.72 ± 7.2	34.64 ± 9.15	0.57
Contralesional hemisphere	34.3 ± 6.05	35.62 ± 8.64	0.37

\*Resting Motor Threshold (RMT), TMS-derived neurophysiological marker, indicating cortical excitability

<sup>a</sup>WHO Grade II versus III.

<sup>b</sup>WHO Grade II versus IV.

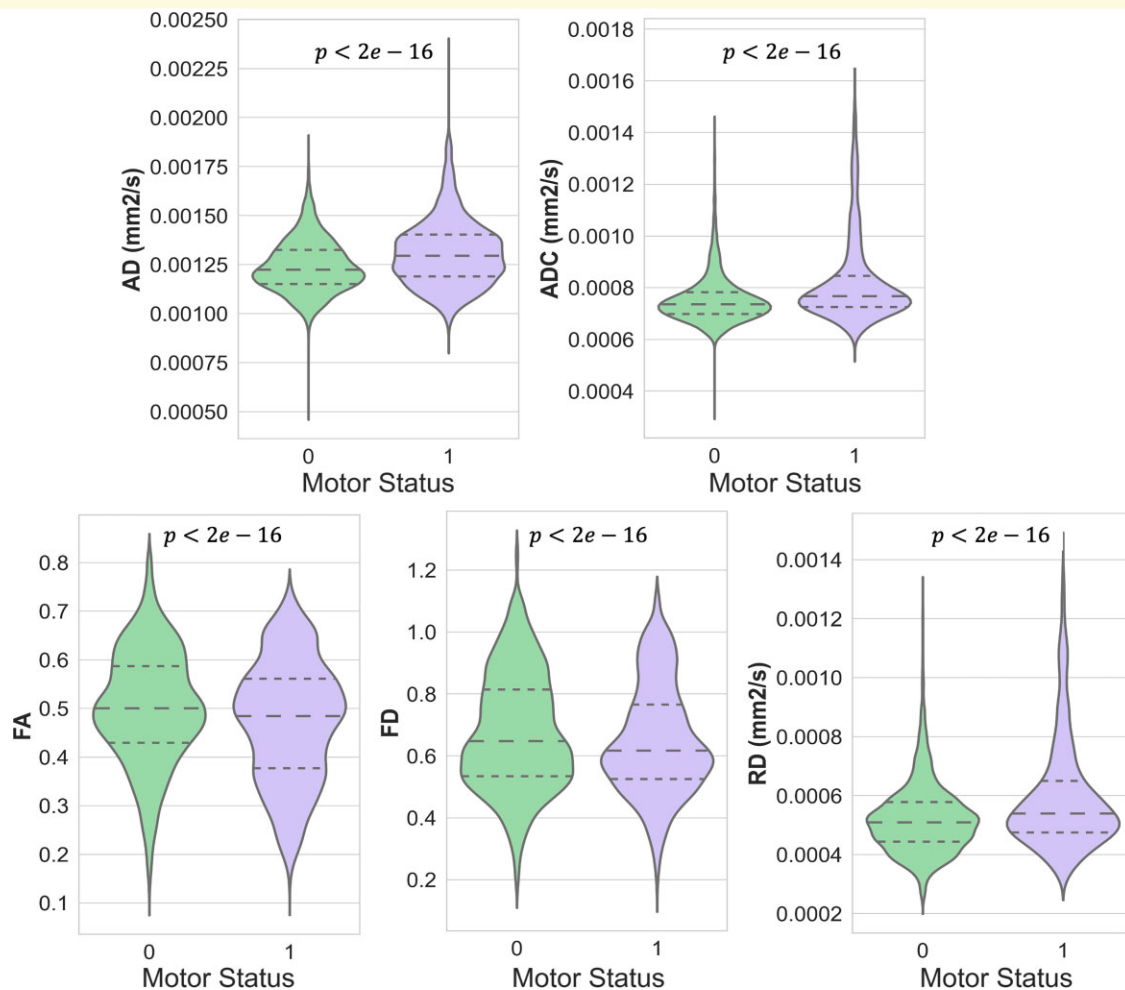
<sup>c</sup>WHO Grade III versus IV.

## Group-wise statistical analysis

A comprehensive group-wise analysis was performed over the entire CST and segment-wise along the 100 segments to compare the differences in dMRI metrics between the two patient groups with (Class 1) and without (Class 0) motor deficits. The violin plots in Fig. 3 show significant differences in dMRI-based measures in the ipsilesional CST between the two patient groups.

Figure 4 shows a segment-wise comparison of dMRI metrics between the two groups of patients (Class 0; Class 1) in ipsilesional CST. We found significant segment-wise differences, surviving FDR correction, between the two groups in the ipsilesional CST profiles in relation to ADC, AD, FA and RD metrics. However, no significant segment-wise

differences were found with respect to FD (cf. Supplementary Table 2). These differences were larger in ADC and RD, especially in the tracts' middle and peritumoural areas. The ipsilesional tract profiles led to larger significant differences compared with the differences between ipsi- and contralesional CSTs in group analyses (cf. Supplementary Table 3). In the latter case, only 6 nodes (93–98th) in the ADC, 9 nodes (88–96) in FA and 11 nodes (88–98) in RD metrics of the ipsilesional tracts showed significant differences between the two groups (class 0; class 1) mainly at the superior portion (area between the cortex and the internal capsule) of the CST. Moreover, the significant segments in contralesional tract profiles were only seen in a few segments in AD and ADC tract profiles (cf. Supplementary Table 4).



**Figure 3 Violin plots.** Violin plots are illustrating the frequency distribution of all segmental values related to AD, ADC, FA, FD and RD metrics over the ipsilesional CST. *P*-values were calculated using a two-tailed Student's *t*-test ( $n = 100 \times 71$ , Class 0, without motor deficit, MRC = 5;  $n = 100 \times 45$  Class 1, with the motor deficit, MRC < 5); horizontal lines indicate median and quartiles.

Additionally, we used M, STD, KU and SK as histogram-based measures of tract profiles and performed group-wise analyses on each measure for ipsi- and contralesional tract profiles (cf. [Supplementary Table 5](#)). As shown in [Fig. 5A–E](#), the M measure over the ipsilesional CST profiles of AD, ADC, FA and RD was significantly different between the two patient groups with and without motor deficits (AD:  $U = 1000$ ,  $p = 3.6e-4$ ,  $r = 0.37$ ; ADC:  $U = 974$ ,  $p = 2.07e-4$ ,  $r = 0.4$ ; FA:  $U = 1214$ ,  $p = 1.5e-2$ ,  $r = 0.24$ ; RD:  $U = 1003$ ,  $p = 3.8e-4$ ,  $r = 0.42$ ) as well as the KU measure of ipsilesional CST profile of FA value ( $U = 1034$ ,  $p = 7.1e-4$ ,  $r = 0.35$ ). Further in [Fig. 5B](#) and [E](#), STD of ADC and RD profiles showed a highly significant increase in the patient group with motor deficit (ADC:  $U = 1032$ ,  $p = 6.8e-4$ ,  $r = 0.35$ ; RD:  $U = 1065$ ,  $p = 1.2e-3$ ,  $r = 0.33$ ).

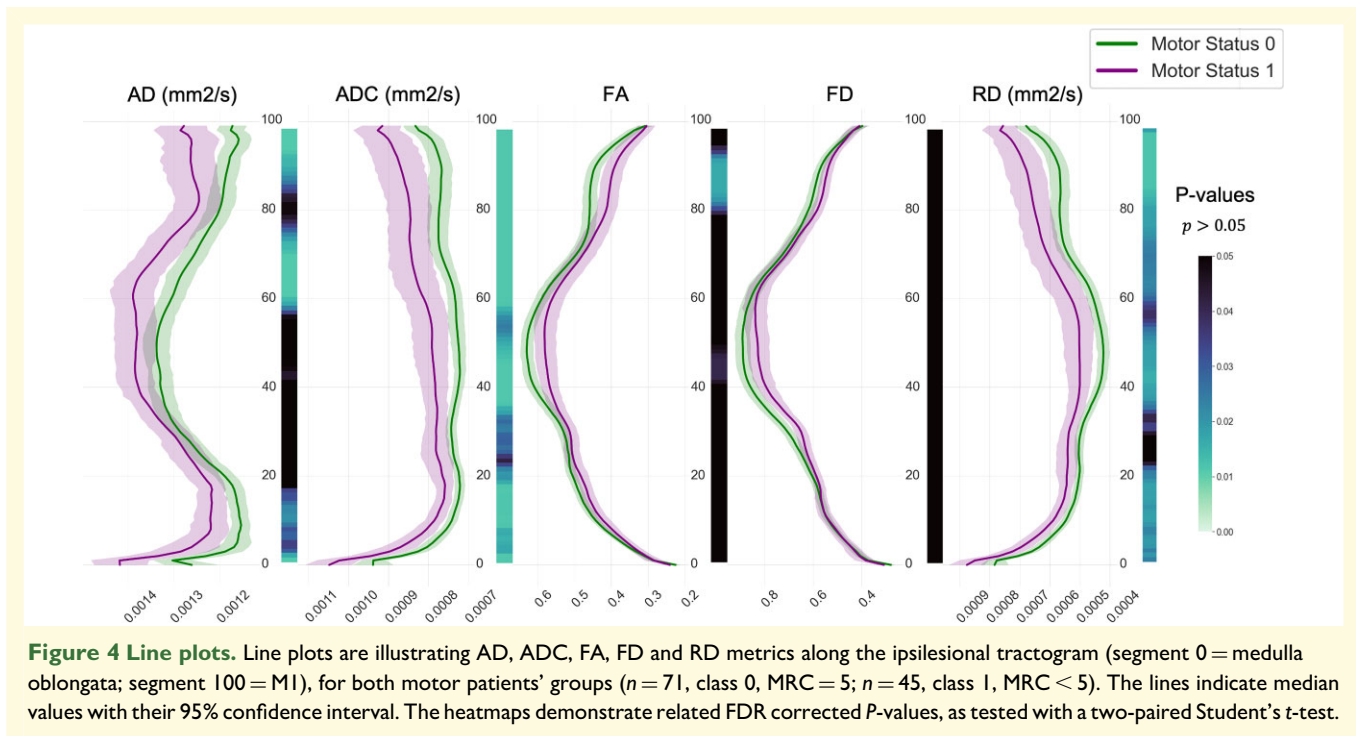
Ipsilesional tract profile measures led to larger differences compared with the contralesional tract profile measures. Nine ipsilesional tract profile measures (extracted features) were significantly different between the two patient groups (see [Fig. 5](#)), while only three measures of contralesional tract

profiles were significantly different (cf. [Supplementary Table 5](#)).

## SVM classification

SVM\_clinical, using all patient demographics and clinical variables as input features, received a low performance score (58% accuracy, 82% sensitivity, 43% specificity and 62% AUC). Among all SVM\_AD-RD models, when using microstructural measures in relation to each metric separately, SVM\_FA reached the highest accuracy (67%), sensitivity (60%) and AUC (68%), and SVM\_ADC and SVM\_RD reached the highest specificity (80%). The AUC in SVM\_ADC and SVM\_AD reached nearly the same score (67%, [Fig. 4](#)).

Considering SVM\_1-4 models, using all measures of dMRI-based tract profiles, the best classifier performance was achieved with SVM\_2, which reached 74% accuracy, 74% sensitivity, 75% specificity and 77% AUC ([Table 3](#)). With the KNN interpolation method, the number of nearest



neighbours when  $K = 10$  yielded the best model performance.

The five most effective features that were selected by SVM-RFE are FA\_KU, FA\_SK, RD\_KU, ADC\_STD and ADC\_M; effect sizes and FDR-corrected  $P$ -values were also calculated (FA\_KU:  $U = 1034$ ,  $p = 1.1e - 3$ ,  $r = 0.35$ ; FA\_SK:  $U = 1413$ ,  $p = 1.5e - 1$ ,  $r = 0.11$ ; RD\_KU:  $U = 1377$ ,  $p = 1.3e - 1$ ,  $r = 0.13$ ; ADC\_STD:  $U = 1032$ ,  $p = 1.1e - 3$ ,  $r = 0.35$ ; ADC\_M:  $U = 974$ ,  $p = 1.0e - 3$ ,  $r = 0.4$ ). Table 4 shows these selected features with their respective learned weights using SVM\_2.

In the last step, patient demographics and clinical features were integrated into our models (SVM\_1-4). The feature selection method did not select them in any of the trained models (SVM\_1-4) and the models' performances remained unchanged. The receiver operating characteristic curves for SVM\_1-4, SVM\_AD-RD and SVM\_clinical are presented in Fig. 6.

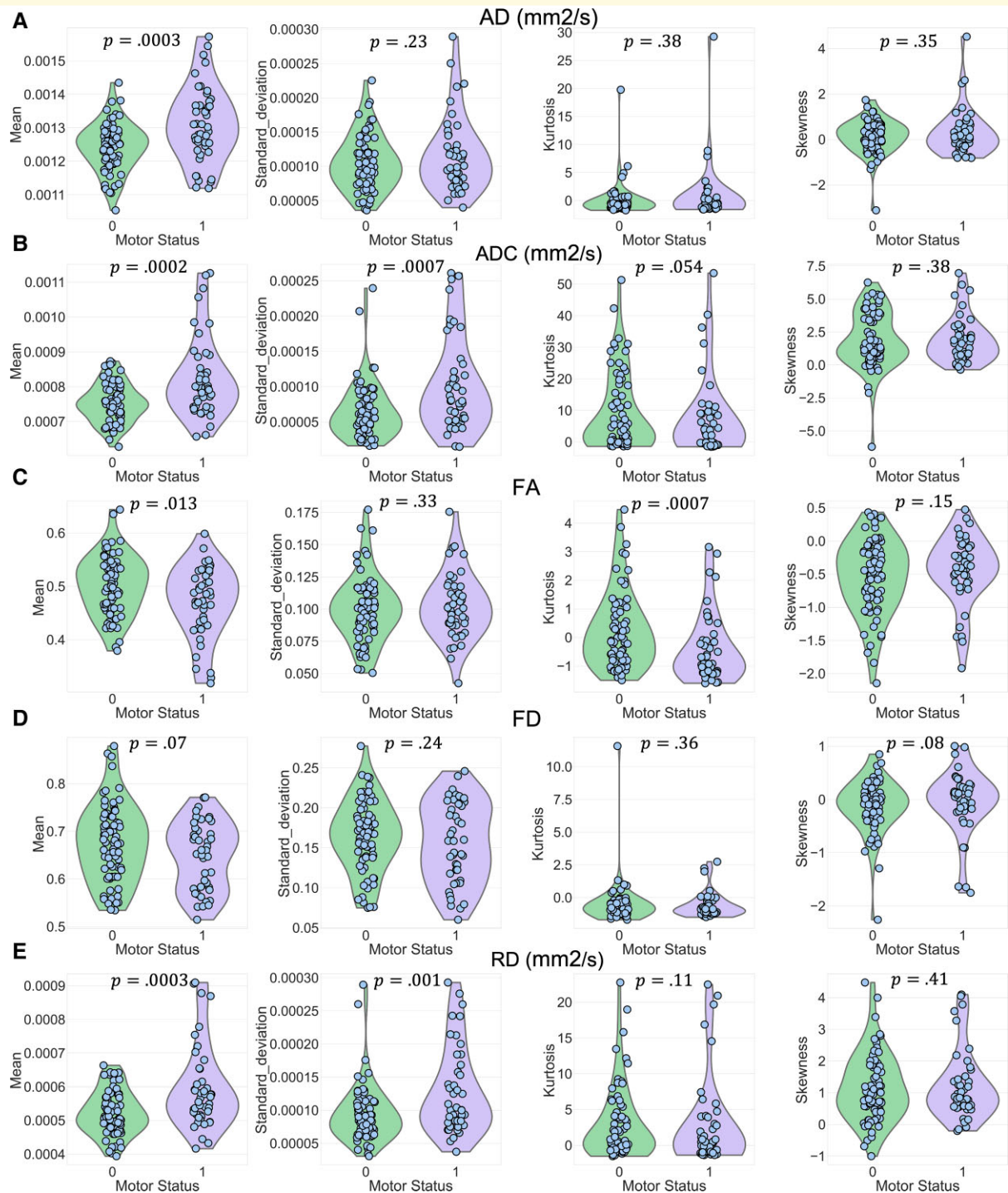
Moreover, we developed a model using all segment-wise information along ipsi- and contralesional tract profiles for all dMRI metrics (ADC, RD, AD, FA and FD) with PCA (SVM\_5). This model reached 63% accuracy, 64% sensitivity, 63% specificity and 70% AUC. Summarizing the SVM results, we showed that the best model performance was achieved when we generated a tract profile based on median values and imputed the missing values with the KNN method (SVM\_2). Moreover, our model that included only clinical and demographic features (SVM\_clinical) was less accurate than the models with microstructural measurements. All models were trained separately using each modality (SVM\_AD-RD) and were also less accurate than other

models that included imaging features from all modalities (SVM\_1-4). Additionally, our SVM model, which used PCA components, performed poorly compared with the SVM\_1-4 models.

## Discussion

We investigated how glioma-induced microstructural alterations to WM were associated with functional motor deficits by mapping dMRI metrics along the CST. In segment-wise group comparison, we found significant differences between the ipsilesional tract profiles in the two patient groups (Class 0; Class 1) in relation to all dMRI metrics except for FD. The ipsilateral differences were mainly seen at the level of the glioma (superior portion of the tract), which showed the direct influence of the tumour area on motor function. However, tumour impact on tract metrics could also be detected in areas relatively distant from the tumour and peritumoural oedema especially in ADC and RD, demonstrating the spreading of the local tumour effect with respect to the motor status.

FA and ADC have been largely used to evaluate the WM, and both lower FA values and higher ADC values are associated with WM impairment. FA has been shown to be sensitive to detecting changes in WM water diffusion in cases of neuropathology and tends to decrease in areas where tumour cells have invaded WM.<sup>86-88</sup> ADC indicates the M diffusivity of water molecules and an increase in ADC has been observed in pathologies accompanied by, e.g. oedema or necrosis<sup>29,89</sup> and higher values are expected in voxels with low



**Figure 5 Violin plots.** Violin plots are illustrating different histogram-based measures of ipsilesional CST profiles based on (A) AD (mm<sup>2</sup>/s), (B) ADC (mm<sup>2</sup>/s), (C) FA, (D) FD and (E) RD (mm<sup>2</sup>/s) metrics in two patient groups ( $n = 71$ , Class 0, MRC = 5;  $n = 45$ , class 1, MRC < 5). The analyses were done using Mann–Whitney U-tests.

anisotropy. These metrics have been used to describe specific types of WM impairments, e.g. demyelination, axonal injury, inflammation or necrosis.<sup>90,91</sup> However, these metrics may be affected by several factors<sup>90,91</sup> and they might be less sensitive to distinguish specific types of WM impairments.<sup>90,91</sup> The majority of our cohort with glioblastomas

(GBMs; 77 patients) are characterized by diffuse infiltration into the normal brain. We observed the greatest elevation in values along the CST in relation to ADC, which could be explained by GBMs' frequently developing within the WM and spreading throughout the brain along fibres,<sup>92</sup> and limited diffusion because of their tissue properties. However, this

**Table 3 SVM\_I-4 model performances using Mdn versus KNN interpolation methods and Mdn versus Mahalanobis-based weighted mean tract profile**

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)
<i>Mdn-based profile</i>				
Mdn interpolation (SVM_1)	73	73	73	76
KNN interpolation (k = 10) (SVM_2)	<b>74</b>	<b>74</b>	<b>75</b>	<b>77</b>
<i>Weighted M profile</i>				
Mdn interpolation (SVM_3)	68	66	70	70
KNN interpolation (k = 10) (SVM_4)	69	66	70	70

**Table 4 SVM2 selected features with their respective learned weights using SVM-RFE**

Features	Weights
FA_KU	1.29
FA_SK	1.17
RD_KU	1.07
ADC_STD	1.01
ADC_M	0.61

explanation should be taken with caution. These results confirmed that the ipsilesional CST fibres were affected and correlated with the patient's motor deficits.

Furthermore, AD and RD are measures of diffusion in perpendicular and parallel directions to the principal axis of diffusion, respectively. The incorporation of these metrics has been shown to lead to better differentiation between axonal injury or degeneration (AD) and pathological demyelination (RD).<sup>29,89,93–95</sup> Myelin fragmentation results in an increase in RD since the myelin sheaths block water diffusion out of the axon.<sup>96</sup> Demyelination could be more of a chronic/slow process, while axonal degeneration could be a more acute and potentially clinically more detrimental process.

Several studies in brain tumour patients investigated DTI metrics as a diagnostic and prognostic biomarker for motor function.<sup>87,97–101</sup> Decreases in FA values were associated with preoperative motor deficits.<sup>87,97,98</sup> Other studies showed that lower average FA values within the affected CST as well as higher average ADC values are significantly associated with postoperative motor deficits.<sup>18</sup>

Here, we did not see a strong improvement in our SVM model performance on basis of AD. However, the KU of tract profile based on RD was highly effective (Table 4). Besides, we even found differences in segments/regions along contralesional tract profiles with respect to ADC and AD, supporting the assumption that contralesional CST may play a role in the postoperative motor outcome and recovery of motor function. These results may have implications for the compensatory strategy of the brain.

## SVM analysis

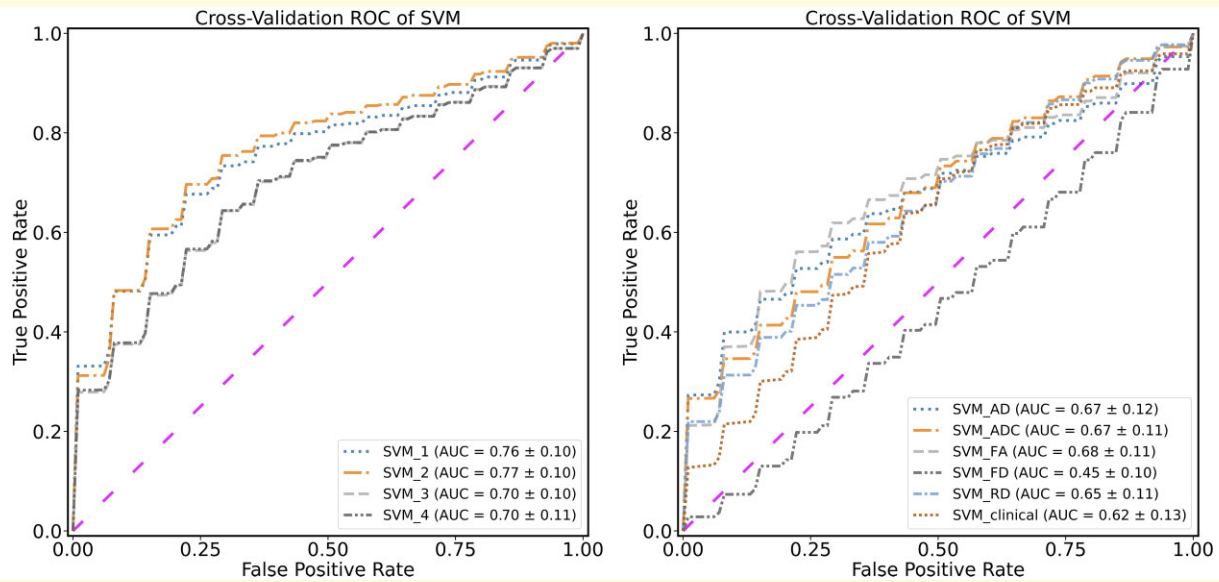
As explained earlier, SVM has been used in various clinical applications such as tumour segmentation and classification,<sup>40–43</sup> prediction of tumour consistency,<sup>46</sup> prediction of prognosis and survival time of tumour patients using multi-modal imaging.<sup>35,47</sup>

## Microstructural measures

Considering dMRI-based features within the ipsi- and contralesional CST, we successfully developed several SVM-based models to classify patients with respect to their motor deficits (Class 0; Class 1). The best model performance achieved an AUC of 76.6% (SVM\_2). The most effective features were FA\_KU, ADC\_M, RD\_KU, FA\_SK and ADC\_STD. Accordingly, ADC and FA were identified as the most relevant metrics which better accounted for the detection of motor function impairment.<sup>102</sup> These findings are also in line with our previous studies.<sup>16,18</sup> In Fekonja *et al.*,<sup>16</sup> we found FA and ADC metrics to be the most relevant metrics for detecting CST impairments, and in Rosenstock *et al.*,<sup>18</sup> we found that peritumoural ADC and FA were strongly associated with post-operative motor deficits (motor outcome).

Here, FD was used to better account for the fibre orientation-specific microstructural properties in relation to infiltrating tumours. CSD can distinguish complex fibre populations in the brain. CSD estimates FODs within each voxel, based on the expected signal from a single collinearly oriented fibre population.<sup>28</sup> Modern, CSD-based FD and FBA methods offer promising opportunities since they are related to the intra-axonal restricted compartment that is specific to a certain fibre orientation within a voxel.<sup>16,31</sup> However, the stronger contribution of FA and ADC to our SVM models showed that these metrics are more robust for detecting motor impairment than FD in typical, clinical single-shell  $b = 1000$  dMRI sequences, but might perform better when using multi-shell dMRI data. This finding is consistent with our previous results where ADC and FA presented higher sensitivity to detect CST impairment in tumour patients than FD.<sup>16</sup>

The univariate analysis confirmed our SVM model results to some extent, showing the predictive power of each tract profile's measure (input feature) separately. All models which were trained on each AD, ADC, FA and AD metric separately, reached a relatively high specificity though they dealt with low sensitivity. Among them, SVM\_FA provides a higher sensitivity but lower specificity, and SVM\_ADC and SVM\_RD provided higher specificity. Combining these metrics (SVM\_2) led to both high sensitivity and specificity. FA\_SK and RD\_KU were not significantly different between the two patient groups, although they were selected as two of the most effective features. Features identified as significantly relevant and/or predictively relevant can agree or diverge, and numerous studies have been conducted demonstrating the differences between highly predictive and highly significant variable sets.<sup>103,104</sup> Sometimes a strong predictivity fails



**Figure 6 Receiver operating characteristic (ROC) curves.** ROC of the discriminative performance of the SVM\_1-4 and SVM\_ADC-RD models show the true positive rate against the false-positive rate for different thresholds in comparison to classifiers with a random performance level (diagonal, dashed line).

to be significant, as it only provides supplementary information and could increase the predictive power of ML models just in combination with other features. The SVM\_2 model reached the best performance among all trained models which showed that the Mdn profile corresponded to better prediction accuracy in comparison with the Mahalanobis-based weighted M tract profiles—a method described in Yeatman *et al.*<sup>15</sup> and Richie-Halford *et al.*, that implies the robustness of the Mdn method in the specific case since outlier segments with extreme values do not bias the Mdn. The superiority of the Mdn profile has been previously shown in,<sup>106</sup> where microstructural models were used to understand the role of WM in relation to cognitive development. Our SVM\_1-4 models, which were based on the extracted features, were more efficient and robust compared with SVM\_5 since these models were less complex in their design. The high dimensional dMRI data (100 values per tract profile) and low available number of patient data resulted in lower performance in SVM\_5. The variety in glioma location in relation to the CST as well as the low sample size restricted our analysis when SVM\_5 tried to capture all patterns of microstructural variations. Therefore, we were not able to detect all segment-wise variations as efficiently as possible in SVM\_5. In the SVM\_1-4 models, we were able to summarize the segment-wise information of dMRI-based CST profiles as different statistical measures to detect informative glioma-induced microstructural alterations to WM to predict functional motor deficits.

## Demographic and clinical features

The SVM model with only demographics and clinical variables (SVM\_clinical), such as age, gender, glioma location,

glioma WHO grade and RMT ratio, showed a poor performance (56% accuracy and 62% AUC) while the models with microstructural measures as input, e.g. SVM\_2 (74% accuracy and 77% AUC), reached to relatively high performance. We further assessed the performance of the SVM\_1-5 models integrating patient demographics and clinical features and saw that none of them affected the SVM models' performances. This could indicate the lower effectiveness of these features compared with tract profile-based characteristics (microstructural measures). In addition, the dMRI-based measures could be associated with different variables. Since the patient's age was significantly different between the two patient groups, it was expected to improve our predictive models' performances in combination with the microstructural measures. However, the integration of patients' age did not improve the models' performances. The performed CCA showed a strong and significant correlation between dMRI extracted features and age. Interestingly, we found a strong and significant correlation in the reverse direction for both motor groups (Class 0; Class 1). This confirms previous findings<sup>107–109</sup> in which WM changes in relation to age and its variation as a function of age were investigated. In a recent study, age has been accurately predicted by FA and ADC metrics.<sup>105</sup> These results justify as well that if taking into account the microstructural measures, age is of critical importance in distinguishing between the two motor groups.

According to CCA analysis and our SVM results, we could conclude that the information provided with age was apparently sufficiently covered by dMRI-based extracted features and thus no additional information was found considering patients' age.

In conclusion, SVM\_2 was the best model, which showed that the median-based tract profiles were more accurate than

the M-weighted tract profile. This might be due to the median being more robust to outliers. The inclusion of demographic and clinical features did not affect model performance. This means that the dMRI-based measures are more powerful than these features in this context. The SVM\_5 model performed poorly compared with SVM\_2, indicating that the patterns of segmental information could not be perfectly detected with PCA.

## Translational aspect

The body of evidence that preservation of the WM connectivity is a key to preserving function is steadily growing. Therefore, not only the presurgical assessment of the spatial relation of the tumour and the tracts but also a detailed analysis of the impact the tumour already exerts on the WM is of great importance. As demonstrated in this study, ML shows a promising potential to address the microstructural effects of brain tumours on the WM, which is not accessible with traditional statistical methods, since it allows for discovering patterns in dMRI data and well-approximating complex relationships. Future studies need to further correlate ML finding with functional outcomes to establish new biomarkers for WM resilience to surgical manipulation with the promise to become a powerful prognostic tool in future neurosurgery.

## Limitations

Tractography suffers from a wide range of limitations that make its routine use problematic.<sup>16,110</sup> Tractograms contain both false-positive<sup>111</sup> and false-negative<sup>112</sup> streamlines. In addition, tractography cannot distinguish between afferent and efferent connections, and streamlines may terminate improperly,<sup>58</sup> especially in the case of tumours and oedema. Furthermore, our results are atlas and tractography algorithm dependent, since other tractography methods or atlas choices would possibly result in different tractograms.<sup>16</sup> The dMRI data used for this study consists of a typical clinical single-shell acquisition and is thus limited for FD measurements due to incomplete attenuation of apparent extra-axonal signal.<sup>16,113</sup> Furthermore, all patients received preoperative steroids to reduce oedema, which may cause a confounding effect. However, there is evidence that oedema has no strong influence on tractography results.<sup>18</sup> In addition, the British Medical Research Council (BMRC) motor status does not necessarily detect subtle or apractic motor deficits which might correlate with early tumour effects on the WM. Indeed, the main limitation of our study is the relatively small sample size we could include to perform the ML analysis. Moreover, we binarized motor deficits due to a low number of samples per class. This led to inaccuracy and affected the performance of our ML models. To develop models performing multiclass classification which could consider differences in the degree of motor power (MRC = 1, 2, 3, 4, 5), larger samples would be needed per class.

A reliable way to evaluate the final ML model is to split the data into training, validation and test sets. The training set is

used for learning and fitting the model's parameters, a validation set is used to tune the model's hyperparameters, while the test set is kept as an unseen data set to assess the performance of the final tuned ML model. This procedure offers an unbiased robust estimate of real model performance.<sup>114</sup> However, this approach typically requires a large number of subjects, which is difficult to recruit in a clinical cohort, because they usually consist of small sample sizes for many ML methods. Moreover, to overcome the problem of the curse of dimensionality, where increasing the number of features requires larger training data to define a generalizable model, a reasonable sample size for training the ML models is necessary. CV<sup>84,115</sup> is a common solution to estimate the model's performance in case of small sample size or when validation with a separate data set is not feasible. However, model selection without nested CV uses the same test data to tune the model's hyperparameters and evaluate the model,<sup>84</sup> which is known to yield overly optimistic scores.<sup>80</sup> To this end, here we performed nested CV and bootstrapping methods to enhance the generalizability of the ML models as well as to prevent an over-optimistic performance estimate. Nested CV uses a series of training, validation and test splits and fits the model iteratively using a pair of nested loops. In the inner loop, an optimal set of model's hyperparameters is found using methods such as grid search on each training set, and each set of hyperparameters is evaluated using *k*-fold CV. In the outer loop, generalization error is estimated by averaging test set scores over several data set splits. Moreover, bootstrap aggregating (bagging)<sup>116</sup> is applied to reduce the variance of a given estimator which uses CV<sup>85</sup> itself to select fine-tuning parameters, trading off bias and variance of the bootstrap sample-specific candidate estimators. Overall, in the nested CV, the model is trained only using the training data (with five-fold inner CV). But the reported model accuracies are obtained using separate test data sets (with 10-fold CV). Therefore, the final test scores are computed on a completely independent set of samples than the training data as shown in Fig. 1.

## Conclusion

In this study, we analysed dMRI-based metrics to assess microstructural WM changes in correlation with the motor status of patients with gliomas in the motor system. We successfully developed SVM models to predict motor deficits in a heterogeneous multivariate data set. ADC, FA and RD were highly predictive dMRI metrics. Additionally, we showed that dMRI metrics are better predictors than demographic and clinical variables, such as age, glioma grade and RMT ratio. Careful selection and testing of ML modelling are mandatory to prevent over- or underfitting and misinterpretation of data.

## Acknowledgements

B.S., L.S.F. and T.P. acknowledge the support of the Cluster of Excellence Matters of Activity. Image Space Material is

funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC 2025–390648296. T.R. received support from the Finnish Cultural Foundation.

## Funding

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2025 - 390648296.

## Competing interests

The authors report no competing interests.

## Supplementary material

Supplementary material is available at *Brain Communications* online

## References

- Weller M, Wick W, Aldape K, *et al*. Glioma. *Nat Rev Dis Prim*. 2015;1(1):15017.
- Giesseks A, Westphal M. Glioma invasion in the central nervous system. *Neurosurgery*. 1996;39(2):235–252.
- Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J*. 1994;66(1):259–267.
- Le Bihan D, Johansen-Berg H. Diffusion MRI at 25: exploring brain tissue structure and function. *Neuroimage*. 2012;61(2):324–341.
- Bartsch AJ, Biller A, Homola GA. Chapter 23 - presurgical tractography applications. In: Johansen-Berg H, ed. *Behrens TEJBT-DMRI (Second E)*. Academic Press; 2014:531–567.
- Mori S. Chapter 9 - Three-dimensional tract reconstruction. *Mori SBT-I to DTI*. Elsevier Science BV; 2007:93–123.
- Pujol S. Chapter 4 - Imaging white matter anatomy for brain tumor surgery. *Golby AJBT-I-GN*. Academic Press; 2015:91–121.
- Dell'Acqua F, Tournier J-D. Modelling white matter with spherical deconvolution: How and why? *NMR Biomed*. 2019;32(4):e3945.
- Jeurissen B, Descoteaux M, Mori S, Leemans A. Diffusion MRI fiber tractography of the brain. *NMR Biomed*. 2019;32(4):e3785.
- Essayed WI, Zhang F, Unadkat P, Cosgrove GR, Golby AJ, O'Donnell LJ. White matter tractography for neurosurgical planning: A topography-based review of the current state of the art. *NeuroImage Clin*. 2017;15:659–672.
- Romano A, D'andrea G, Minniti G, *et al*. Pre-surgical planning and MR-tractography utility in brain tumour resection. *Eur Radiol*. 2009;19(12):2798–2808.
- Bells S, Cercignani M, Deoni S, *et al*. Tractometry-comprehensive multi-modal quantitative assessment of white matter along specific tracts. *Proc ISMRM*. 2011;19:678.
- Colby JB, Soderberg L, Lebel C, Dinov ID, Thompson PM, Sowell ER. Along-tract statistics allow for enhanced tractography analysis. *Neuroimage*. 2012;59(4):3227–3242.
- O'Donnell LJ, Westin C-F, Golby AJ. Tract-based morphometry for white matter group analysis. *Neuroimage*. 2009;45(3):832–844.
- Yeatman JD, Dougherty RF, Myall NJ, Wandell BA, Feldman HM. Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS One*. 2012;7(11):e49790.
- Fekonja LS, Wang Z, Aydogan DB, *et al*. Detecting corticospinal tract impairment in tumor patients with fiber density and tensor-cased metrics. *Front Oncol*. 2021;10:3256.
- D'Souza S, Ormond DR, Costabile J, Thompson JA. Fiber-tract localized diffusion coefficients highlight patterns of white matter disruption induced by proximity to glioma. *PLoS One*. 2019;14(11):e0225323.
- Rosenstock T, Giampiccolo D, Schneider H, *et al*. Specific DTI seeding and diffusivity-analysis improve the quality and prognostic value of TMS-based deterministic DTI of the pyramidal tract. *NeuroImage Clin*. 2017;16:276–285.
- Jeurissen B, Leemans A, Tournier J-D, Jones DK, Sijbers J. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Hum Brain Mapp*. 2013;34(11):2747–2766.
- Riffert TW, Schreiber J, Anwander A, Knösche TR. Beyond fractional anisotropy: extraction of bundle-specific structural metrics from crossing fiber models. *Neuroimage*. 2014;100:176–191.
- Vos SB, Jones DK, Jeurissen B, Viergever MA, Leemans A. The influence of complex white matter architecture on the mean diffusivity in diffusion tensor MRI of the human brain. *Neuroimage*. 2012;59(3):2208–2216.
- Tournier JD. Diffusion MRI in the brain—Theory and concepts. *Prog Nucl Magn Reson Spectrosc*. 2019;112-113:1–16.
- Roine T, Jeurissen B, Perrone D, *et al*. Isotropic non-white matter partial volume effects in constrained spherical deconvolution. *Front Neuroinform*. 2014;8:28.
- Dhollander T, Raffelt D, Connelly A. Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image. *ISMRM Workshop on Breaking the Barriers of Diffusion MRI*. 2016:5.
- Roine T, Jeurissen B, Perrone D, *et al*. Informed constrained spherical deconvolution (iCSD). *Med Image Anal*. 2014;8:28.
- Rasmussen PM, Madsen KH, Lund TE, Hansen LK. Visualization of nonlinear kernel models in neuroimaging by sensitivity maps. *Neuroimage*. 2014;8:28.
- Jeurissen B, Tournier JD, Dhollander T, Connelly A, Sijbers J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *Neuroimage*. 2014;103:411–426.
- Tournier J-D, Calamante F, Gadian DG, Connelly A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage*. 2004;23(3):1176–1185.
- Tournier J-D, Mori S, Leemans A. Diffusion tensor imaging and beyond. *Magn Reson Med*. 2011;65(6):1532–1556.
- Mormina E, Longo M, Arrigo A, *et al*. MRI tractography of corticospinal tract and arcuate fasciculus in high-grade gliomas performed by constrained spherical deconvolution: Qualitative and quantitative analysis. *Am J Neuroradiol*. 2015;36(10):1853–1858.
- Raffelt DA, Tournier J-D, Smith RE, *et al*. Investigating white matter fibre density and morphology using fixel-based analysis. *Neuroimage*. 2017;144:58–73.
- Dhollander T, Clemente A, Singh M, *et al*. Fixel-based analysis of diffusion MRI: Methods, applications, challenges and opportunities. *Neuroimage*. 2021;241:118417.
- Zhu W, Xie L, Han J, Guo X. The application of deep learning in cancer prognosis prediction. *Cancers (Basel)*. 2020;12(3):603.
- Chandrashekar G, Sahin F. A survey on feature selection methods. *Comput Electr Eng*. 2014;40(1):16–28.
- Molina-García D, Vera-Ramírez L, Pérez-Beteta J, Arana E, Pérez-García VM. Prognostic models based on imaging findings in glioblastoma: Human versus Machine. *Sci Rep*. 2019;9(1):5982.
- Senders JT, Arnaout O, Karhade A V, *et al*. Natural and artificial intelligence in neurosurgery: A systematic review. *Neurosurgery*. 2018;83(2):181–192.
- Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–297.



38. Noble WS. What is a support vector machine? *Nat Biotechnol.* 2006;24(12):1565–1567.
39. Bzdok D, Krzywinski M, Altman N. Machine learning: Supervised methods. *Nat Methods.* 2018;15(1):5–6.
40. Polly FP, Shil SK, Hossain MA, Ayman A, Jang YM. Detection and classification of HGG and LGG brain tumor using machine learning. In: 2018 International Conference on Information Networking (ICOIN); 2018:813–817.
41. Gurbină M, Lascu M, Lascu D. Tumor detection and classification of MRI brain image using different wavelet transforms and support vector machines. In: 2019 42nd International Conference on Telecommunications and Signal Processing (TSP); 2019:505–508.
42. Vamvakas A, Williams SC, Theodorou K, et al. Imaging biomarker analysis of advanced multiparametric MRI for glioma grading. *Phys Medica.* 2019;60:188–198.
43. Kumari N, Saxena S. Review of brain tumor segmentation and classification. In: 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT); 2018:1–6.
44. Wu J, Qian Z, Tao L, et al. Resting state fMRI feature-based cerebral glioma grading by support vector machine. *Int J Comput Assist Radiol Surg* 2015;10(7):1167–1174.
45. Tian Q, Yan L-F, Zhang X, et al. Radiomics strategy for glioma grading using texture features from multiparametric MRI. *J Magn Reson Imaging.* 2018;48(6):1518–1528.
46. Fan Y, Hua M, Mou A, et al. Preoperative noninvasive radiomics approach predicts tumor consistency in patients with acromegaly: development and multicenter prospective validation. *Front Endocrinol (Lausanne).* 2019;10:403.
47. Nie D, Zhang H, Adeli E, Liu L, Shen D. *3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients BT - medical image computing and computer-assisted intervention – MICCAI 2016.* In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. Springer International Publishing; 2016:212–220.
48. Chen X, Jeong JC. Enhanced recursive feature elimination. In: Sixth International Conference on Machine Learning and Applications (ICMLA 2007); 2007:429–435.
49. Jain A, Zongker D. Feature selection: evaluation, application, and small sample performance. *IEEE Trans Pattern Anal Mach Intell.* 1997;19(2):153–158.
50. Garali I, Adel M, Bourennane S, Guedj E. Histogram-based features selection and volume of interest ranking for brain PET image classification. *IEEE J Transl Eng Heal Med.* 2018;6:2100212.
51. Blachnik M, Laaksonen J. *Image classification by histogram features created with learning vector quantization BT - artificial neural networks - ICANN 2008.* In: Kůrková V, Neruda R, Koutník J, eds. Springer Berlin Heidelberg; 2008:827–836.
52. Jolliffe I. *Principal component analysis BT - international encyclopedia of statistical science.* In: Lovric M, ed. Springer Berlin Heidelberg; 2011:1094–1096.
53. Hotelling H. Analysis of a complex of statistical variables into principal components. *J Educ Psychol.* 1933;24(6):417–441.
54. Picht T, Frey D, Thieme S, Kliesch S, Vajkoczy P. Presurgical navigated TMS motor cortex mapping improves outcome in glioblastoma surgery: a controlled observational study. *J Neurooncol.* 2016;126(3):535–543.
55. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage.* 2011;54(3):2033–2044.
56. Avants BB, Tustison NJ, Stauffer M, Song G, Wu B, Gee JC. The Insight ToolKit image registration framework. *Front Neuroinform.* 2014;8:44.
57. Grabner G, Janke AL, Budge MM, Smith D, Pruessner J, Collins DL. Symmetric atlas and model based segmentation: An application to the hippocampus in older adults. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2006:58–66.
58. Tournier J-D, Smith R, Raffelt D, et al. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage.* 2019;202:116137.
59. Veraart J, Novikov DS, Christiaens D, Ades-Aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. *Neuroimage.* 2016;142:394–406.
60. Kellner E, Dhital B, Kiselev VG, Reisert M. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magn Reson Med.* 2016;76(5):1574–1581.
61. Leemans A, Jones DK. The B-matrix must be rotated when correcting for subject motion in DTI data. *Magn Reson Med.* 2009;61(6):1336–1349.
62. Andersson JLR, Graham MS, Drobnjak I, Zhang H, Filippini N, Bastiani M. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *Neuroimage.* 2017;152:450–466.
63. Andersson JLR, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *Neuroimage.* 2003;20(2):870–888.
64. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. *Fsl. Neuroimage.* 2012;62(2):782–790.
65. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging.* 2010;29(6):1310–1320.
66. Dyrby TB, Lundell H, Burke MW, et al. Interpolation of diffusion weighted imaging datasets. *Neuroimage.* 2014;103:202–213.
67. Veraart J, Sijbers J, Sunaert S, Leemans A, Jeurissen B. Weighted linear least squares estimation of diffusion MRI parameters: Strengths, limitations, and pitfalls. *Neuroimage.* 2013;81:335–346.
68. Tournier J-D, Calamante F, Connelly A. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage.* 2007;35(4):1459–1472.
69. Tournier J-D, Calamante F, Connelly A. *Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions.* 2010.
70. Mayka MA, Corcos DM, Leurgans SE, Vaillancourt DE. Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: A meta-analysis. *Neuroimage.* 2006;31(4):1453–1474.
71. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat Methods.* 2020;17(3):261–272.
72. Waskom M, Gelbart M, Botvinnik O, et al. *Mwaskom/seaborn: v0.11.1 (December 2020).* Published online December 20, 2020. doi:10.5281/ZENODO.4379347
73. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with Python. In: Proceedings of the 9th Python Science Conference; 2010.
74. Hunter JD. Matplotlib: A 2D graphics environment. *Comput Sci Eng.* 2007;9(3):90–95.
75. Haynes W. *Benjamini–Hochberg method BT - encyclopedia of systems biology.* In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, eds. Springer New York; 2013:78.
76. Haroon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: an overview with application to learning methods. *Neural Comput.* 2004;16(12):2639–2664.
77. Wang Z, Dreyer F, Pulvermüller F, et al. Support vector machine based aphasia classification of transcranial magnetic stimulation language mapping in brain tumor patients. *NeuroImage Clin.* 2021;29:102536.
78. Gaonkar B, Shinohara R T, Davatzikos C. Interpreting support vector machine models for multivariate group wise analysis in neuroimaging. *Med Image Anal.* 2015;24(1):190–204.
79. Schölkopf B, Smola AJ, Bach F. *Learning with kernels: Support vector machines, regularization, optimization, and beyond.* MIT Press; 2002.
80. Shawe-Taylor J, Cristianini N. *Kernel methods for pattern analysis.* Cambridge University Press; 2004.

81. Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci*. 2019;116(44):22071.
82. Stewart TG, Zeng D, Wu MC. Constructing support vector machines with missing data. *WIREs Comput Stat*. 2018;10(4):e1430.
83. Batista GE, Monard MC. A study of K-nearest neighbour as an imputation method. *HIS*. 2002;87:48.
84. Lever J, Krzywinski M, Altman N. Points of significance: model selection and overfitting. *Nat Med*. 2016;3:703–704.
85. Petersen ML, Molinaro AM, Sinisi SE, van der Laan MJ. Cross-Validated Bagged Learning. *J Multivar Anal*. 2008;25(2):260–266.
86. Stadlbauer A, Gruber S, Nimsky C, et al. Preoperative grading of gliomas by using metabolite quantification with high-spatial-resolution proton MR spectroscopic imaging. *Radiology*. 2006;238(3):958–969.
87. Stadlbauer A, Nimsky C, Gruber S, et al. Changes in fiber integrity, diffusivity, and metabolism of the pyramidal tract adjacent to gliomas: A quantitative diffusion tensor fiber tracking and MR spectroscopic imaging study. *AJNR Am J Neuroradiol*. 2007;28(3):462–469.
88. Kinoshita M, Hashimoto N, Goto T, et al. Fractional anisotropy and tumor cell density of the tumor core show positive correlation in diffusion tensor magnetic resonance imaging of malignant brain tumors. *Neuroimage*. 2008;43(1):29–35.
89. Alexander AL, Lee JE, Lazar M, Field AS. Diffusion tensor imaging of the brain. *Neurotherapeutics*. 2007;4(3):316–329.
90. Yeh F-C, Irimia A, Bastos DC de A, Golby AJ. Tractography methods and findings in brain tumors and traumatic brain injury. *Neuroimage*. 2021;245:118651.
91. Hui ES, Cheung MM, Chan KC, Wu EX. B-value dependence of DTI quantitation and sensitivity in detecting neural tissue changes. *Neuroimage*. 2010;49(3):2366–2374.
92. Brooks LJ, Clements MP, Burden JJ, et al. The white matter is a pro-differentiative niche for glioblastoma. *Nat Commun*. 2021;12(1):2184.
93. Harsan LA, Poulet P, Guignard B, et al. Brain dysmyelination and recovery assessment by noninvasive in vivo diffusion tensor magnetic resonance imaging. *J Neurosci Res*. 2006;83(3):392–402.
94. Sun S-W, Liang H-F, Trinkaus K, Cross AH, Armstrong RC, Song SK. Noninvasive detection of cuprizone induced axonal damage and demyelination in the mouse corpus callosum. *Magn Reson Med*. 2006;55(2):302–308.
95. Concha L. A macroscopic view of microstructure: Using diffusion-weighted images to infer damage, repair, and plasticity of white matter. *Neuroscience*. 2014;276:14–28.
96. Weber AM, Pukropski A, Kames C, et al. Pathological insights from quantitative susceptibility mapping and diffusion tensor imaging in ice hockey players pre and post-concussion. *Front Neurol*. 2018;9:575.
97. Morita N, Wang S, Kadakia P, Chawla S, Poptani H, Melhem E. Diffusion tensor imaging of the corticospinal tract in patients with brain neoplasms. *Magn Reson Med Sci*. 2011;10(4):239–243.
98. Kim CH, Chung C-K, Kim J, Jahng T, Lee JH, Song I. Use of diffusion tensor imaging to evaluate weakness. *J Neurosurg*. 2007;106(1):111–118.
99. Laundre BJ, Jellison BJ, Badie B, Alexander AL, Field AS. Diffusion tensor imaging of the corticospinal tract before and after mass resection as correlated with clinical motor findings: Preliminary data. *Am J Neuroradiol*. 2005;26(4):791.
100. Yuanzheng H, Lichao M, Xiaolei C, Bainan X. Functional outcome of surgery for glioma directly adjacent to pyramidal tract depicted by diffusion-tensor based fiber tracking. *Turk Neurosurg*. 2015;25(3):438–445.
101. Kovanlikaya I, Firat Z, Kovanlikaya A, et al. Assessment of the corticospinal tract alterations before and after resection of brainstem lesions using Diffusion Tensor Imaging (DTI) and tractography at 3 T. *Eur J Radiol*. 2011;77(3):383–391.
102. Liu D, Liu Y, Hu X, et al. Alterations of white matter integrity associated with cognitive deficits in patients with glioma. *Brain Behav*. 2020;10(7):e01639.
103. Lo A, Chernoff H, Zheng T, Lo SH. Why significant variables aren't automatically good predictors. *Proc Natl Acad Sci U S A*. 2015;112(45):13892–13897.
104. Bzdok D, Engemann D, Thirion B. Inference and prediction diverge in biomedicine. *Patterns*. 2020;1(8):100119.
105. Richie-Halford A, Yeatman JD, Simon N, Rokem A. Multidimensional analysis and detection of informative features in human brain white matter. *PLOS Comput Biol*. 2021;17(6):e1009136.
106. Huber E, Henriques RN, Owen JP, Rokem A, Yeatman JD. Applying microstructural models to understand the role of white matter in cognitive development. *Dev Cogn Neurosci*. 2019;36:100624.
107. Tian L, Ma L. Microstructural changes of the human brain from early to mid-adulthood. *Front Hum Neurosci*. 2017;11:393.
108. Lövdén M, Köhncke Y, Laukka EJ, et al. Changes in perceptual speed and white matter microstructure in the corticospinal tract are associated in very old age. *Neuroimage*. 2014;102:520–530.
109. Salat DH, Tuch DS, Greve DN, et al. Age-related alterations in white matter microstructure measured by diffusion tensor imaging. *Neurobiol Aging*. 2005;26(8):1215–1227.
110. Schilling KG, Nath V, Hansen C, et al. Limits to anatomical accuracy of diffusion tractography using modern approaches. *Neuroimage*. 2019;185:1–11.
111. Maier-Hein KH, Neher PF, Houde J-C, et al. The challenge of mapping the human connectome based on diffusion tractography. *Nat Commun*. 2017;8(1):1349.
112. Aydogan DB, Jacobs R, Dulawa S, et al. When tractography meets tracer injections: a systematic study of trends and variation sources of diffusion-based connectivity. *Brain Struct Funct*. 2018;223(6):2841–2858.
113. Raffelt D, Tournier J-D, Rose S, et al. Apparent fibre density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. *Neuroimage*. 2012;59(4):3976–3994.
114. Kuhn M, Johnson K. *Applied predictive modeling*. Springer; 2013.
115. Refaeilzadeh P, Tang L, Liu H. Cross-validation. *Encycl Database Syst*. 2009;5:532–538.
116. Hillis DM, Bull JJ. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol*. 1993;42(2):182–192.