Pattern Recognition 63 (2017) 593–600

Contents lists available at ScienceDirect

Pattern Recognition

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

BundleMAP: Anatomically localized classification, regression, and hypothesis testing in diffusion MRI

Mohammad Khatami [a], Tobias Schmidt-Wilcke [b], Pia C. Sundgren [c,d], Amin Abbasloo [a], Bernhard Schölkopf [e], Thomas Schultz [a,*]

[a] Department of Computer Science, University of Bonn, Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
[b] Department of Neurology, University Hospital Bergmannsheil, Buerkle-de-la-Camp-Platz 1, 44789 Bochum, Germany
[c] Institute for Clinical Sciences/Diagnostic Radiology, Lund University, SE.221 85 Lund, Sweden
[d] Department of Radiology, University of Michigan Health Systems, 1500 E Medical Center Court, Ann Arbor MI 48109, USA
[e] Empirical Inference Department, Max Planck Institute for Intelligent Systems, Spemannstr. 38, 72076 Tübingen, Germany

ARTICLE INFO

Article history:
Received 31 January 2016
Received in revised form
23 August 2016
Accepted 21 September 2016
Available online 21 September 2016

Keywords:
Disease detection
Manifold learning
Support vector machines
Classification
Regression
Fiber tracking
Diffusion MRI

ABSTRACT

Diffusion MRI (dMRI) provides rich information on the white matter of the human brain, enabling insight into neurological disease, normal aging, and neuroplasticity. We present BundleMAP, an approach to extracting features from dMRI data that can be used for supervised classification, regression, and hypothesis testing. Our features are based on aggregating measurements along nerve fiber bundles, enabling visualization and anatomical interpretation. The main idea behind BundleMAP is to use the ISOMAP manifold learning technique to jointly parametrize nerve fiber bundles. We combine this idea with mechanisms for outlier removal and feature selection to obtain a practical machine learning pipeline. We demonstrate that it increases accuracy of disease detection and estimation of disease activity, and that it improves the power of statistical tests.

1. Introduction

Even though nerve fibers are much too small to be observed with Magnetic Resonance Imaging (MRI) directly, their coherent organization hinders and restricts the natural Brownian heat motion of water in a characteristic manner. This can be probed at a macroscopic level by diffusion MRI (dMRI), making it a unique mechanism for studying white matter organization of the human brain in vivo [1]. However, dMRI sequences involve a large number of whole-brain images, taken with different measurement parameters, and biologically relevant quantities can only be derived by analyzing the relationship between those images.

This makes it challenging to leverage the rich information that diffusion MRI provides on white matter disease for the detection of disease or assessment of disease severity using supervised machine learning [2–5]. In particular, it is desirable to derive features that are not only effective, but that should also be interpretable, indicating which anatomical structures are particularly relevant, and how they might be affected by a given disease.

We present a practical system that provides mechanisms for supervised classification, regression, and hypothesis testing of dMRI data based on features that we derive from diffusion parameters and anatomical structures whose interpretation is familiar to neurologists. The name of our method, BundleMAP, reflects the fact that it combines manifold learning using the ISOMAP method [6] with registration and clustering to achieve a joint parametrization of the fiber bundles in a group of subjects. Significantly extending a previous conference paper [7], we combine this idea with methods for outlier removal and feature selection, and demonstrate that this allows us to detect disease, predict disease activity, and visualize diffusion parameters along major nerve fiber bundles, highlighting specific segments on fiber bundles that differ most between the populations.

The structure of our paper is as follows: After reviewing related work in Section 2, we present the individual steps of our BundleMAP approach in Section 3. In particular, Section 3.3 describes the core idea of using manifold learning for joint parametrization, and Section 3.5 describes a stable method to decide on the number of sections per bundle, and a method for imputing missing features.

* Corresponding author.
E-mail addresses: khatami@cs.uni-bonn.de (M. Khatami), tobias.schmidt-wilcke@bergmannsheil.de (T. Schmidt-Wilcke), pia.sundgren@med.lu.se (P.C. Sundgren), abbasloo@cs.uni-bonn.de (A. Abbasloo), bs@tuebingen.mpg.de (B. Schölkopf), schultz@cs.uni-bonn.de (T. Schultz).

http://dx.doi.org/10.1016/j.patcog.2016.09.020

Section 4 contains results and evaluations. In Section 4.2, we demonstrate that, compared to an earlier work [8], BundleMAP shows superior accuracy in detecting Systemic Lupus Erythematosus, a neuroinflammatory disease that can affect cerebral white matter. In Sections 4.3 and 4.4, we report results of using our system for regression and localized hypothesis testing. In Section 4.6, we find that the fiber bundles that our method highlights as important for disease detection are in excellent agreement with previous findings from the literature [8,9]. In Section 4.7, we discuss advantages of using manifold learning for bundle parametrization over a previously proposed alternative, and perform a direct comparison [10]. Finally, Section 5 concludes the paper.

2. Related work

Localized comparison of white matter structures requires their joint parametrization, i.e., anatomical correspondences between fiber bundles of different subjects. Previous methods for this require manual specification of start and end points [11,12] or manual alignment of a cutting plane [10], whereas BundleMAP works fully automatically. Some more automated methods fit deformable models [13] or match fibers to a tractography atlas [14] which, unlike our tool, makes prior assumptions on the bundle shape.

Among the existing alternatives, tract-based morphometry (TBM) [15] is the approach most similar to ours. However, it has not been used for supervised machine learning. Moreover, we take a novel perspective on the problem, leading to computational techniques that are completely different from those used in TBM. The exact relationship between BundleMAP and TBM is discussed in more detail in Section 3.3.

3. Method: our proposed bundlemap pipeline

The overall pipeline of our BundleMAP method consists of four steps, which are illustrated in Fig. 1. First, the major nerve fiber bundles are automatically extracted. Since this is not fully reliable, a combination of outlier removal techniques is applied at the second stage. In the third and most important step, manifold learning is used to establish a joint parametrization of the remaining fibers by mapping them to the latent fiber bundle core. Finally, this parametrization is used to map some common diffusion parameters as a function of position along the bundles. Adaptive binning, in which feature selection determines the most suitable number of bins, creates our final feature vectors.

3.1. Tracking the fiber bundles

Different diffusion models are available to reconstruct nerve fiber bundles from dMRI data. Among them, Diffusion Tensor Imaging is the most widely used, and is employed in our study [16]. More elaborate methods, such as constrained spherical deconvolution [17] or multi-tensor models [18], significantly increase accuracy especially in fiber crossings, and could easily replace Diffusion Tensor Imaging in our pipeline. Unfortunately, they cannot be used with our current data due to its limited angular resolution.

Even though rules exist on how to place seed points for tracking the major fiber bundles [19], we want to avoid having to follow them manually for each individual subject. Therefore, we define seed regions in a template that represents the average of a large number of healthy normal brains. Since this template is aligned with a widely used brain atlas created at the Montreal Neurological Institute, its coordinates define the so-called MNI space.

For each subject, tractography is performed in its individual coordinate system, which allows us to sidestep the difficult problem of correctly adjusting local fiber directions while spatially transforming dMRI data [20]. Seed points are automatically transferred from the template using a nonlinear transformation obtained from an established algorithm for volumetric registration of Fractional Anisotropy [21]. The resulting fibers are warped back into MNI space using the inverse of that transformation.

3.2. Eliminating erroneous fibers

There are two main sources of error that can lead to the inclusion of erroneous fibers during tracking. The first is automatic placement of seed points using image registration, which is known to suffer from inaccuracies [22]. The second are imperfections in the tractography itself, which are known to occur where tracts run closely together or where, at the imaging resolution, two or more fiber bundles cross [23].

We follow two strategies to remove erroneous fibers. First, anatomical knowledge imposes natural constraints on many bundles, and we filter out fibers that violate them. For example, bundles that are known to connect ipsilateral regions should not cross the mid-sagittal plane. Second, in the set of remaining fibers from all subjects, it is often quite obvious which of them are erroneous, since they follow trajectories that differ substantially from the majority of all reconstructed fibers. We use a one-class support vector machine (SVM) with a radial basis function kernel [24] to separate out those atypical fibers.

A one-class SVM treats its input data as samples from a probability distribution, and estimates the support of that distribution. In other words, it identifies a region in the input space that should

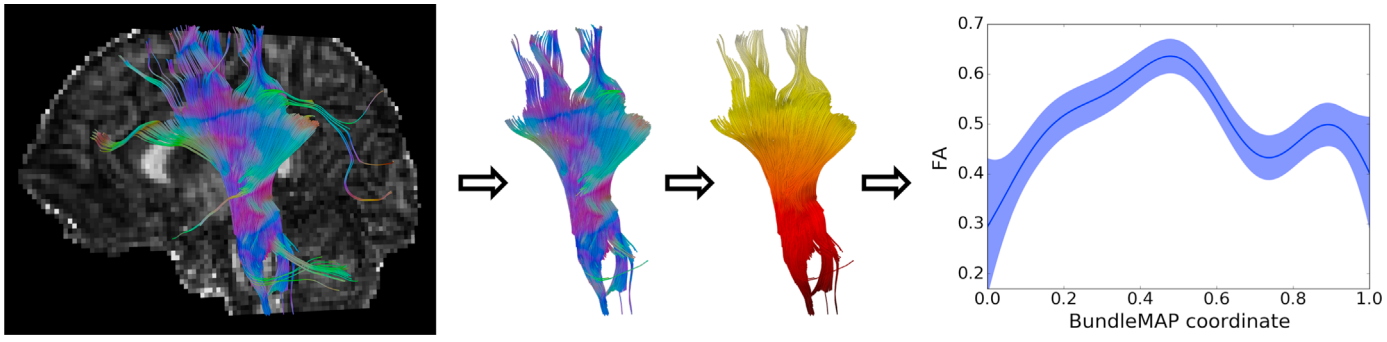

Fig. 1. The four main steps of the BundleMAP pipeline are fiber tractography, outlier removal, joint parametrization of a group of subjects, and derivation of spatially localized features that can be used for supervised learning, hypothesis testing, and visualization.

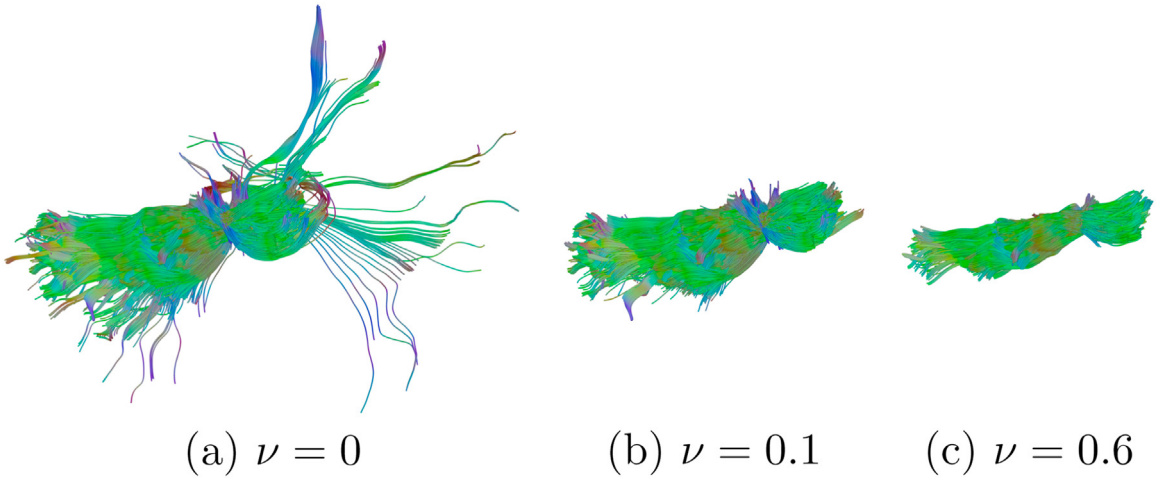

Fig. 2. Fully automated reconstruction of the *anterior thalamic radiation* erroneously includes some streamlines that actually belong to other bundles (a). Our approach uses a one-class SVM to remove them (b). We visually verify the choice of its parameter ν, since overly large values will remove legitimate parts of the bundle (c). (a) $\nu = 0$ (b) $\nu = 0.1$ (c) $\nu = 0.6$.

be as small as possible while containing most, but not necessarily all of the samples. Samples outside this core region can be treated as outliers. The most important parameter in one-class SVM is $\nu \in (0, 1]$, which sets an upper bound on the fraction of samples that will be discarded.

To apply this technique to the identification of atypical fibers, we represent each streamline as a nine-dimensional vector, based on the mean and covariance of its vertices [25]. Values around $\nu = 0.1$ worked well for most of the bundles. In order to ensure that results agree with anatomical knowledge, we allow a human analyst to adjust the exact setting of ν based on visual inspection (cf. Fig. 2).

3.3. Joint parametrization as a manifold learning problem

A key task within our method is to establish a joint parametrization of fiber bundles from different subjects. Finding such correspondences is a common problem in dMRI analysis. Some existing solutions are discussed in Section 2, but each of them has certain drawbacks.

BundleMAP is based on taking a novel perspective on this problem: We view the fiber bundles that have been reconstructed from individual subjects as manifestations of an abstract, idealized fiber bundle core, a latent low-dimensional manifold that is common to all subjects. Most bundles are tubular, and we consider their corresponding fiber bundle core to be a one-dimensional manifold, with the two brain regions connected by the bundle at the far ends.

Our implementation of this idea makes use of ISOMAP [6], a classic nonlinear manifold learning algorithm. We first achieve a rough alignment of the fibers from different subjects by warping them to standard MNI space, as it was described in Section 3.1. We then build a k nearest neighbor graph ($k=12$) on all streamline vertices. The weights of edges that connect points on the same streamline are set to the arc-length distance between them. For edges that connect vertices $\mathbf{v}_{1,2}$ on different streamlines, we average the projected distances onto the respective tangents $\mathbf{t}_{1,2}$, to eliminate the component orthogonal to the underlying manifold:

$$d(\mathbf{v}_1, \mathbf{v}_2) := \frac{1}{2} \sum_{i=1}^{2} \left| \frac{\mathbf{t}_i \cdot (\mathbf{v}_2 - \mathbf{v}_1)}{\| \mathbf{t}_i \|} \right| \tag{1}$$

Performing multidimensional scaling on the distance matrix that results from computing all-pairs shortest path distances on this graph [6] assigns scalars to streamline vertices so that differences between them approximate geodesic distances on the underlying manifold.

However, the results are unique only up to adding constant values and multiplication by -1. To achieve uniqueness and to facilitate interpretation in brains of different sizes, we obtain our final BundleMAP coordinates by normalizing the output of ISOMAP to range $[0, 1]$, so that it specifies a relative position along the bundle. We fix the origin to be on the left for fiber bundles with a primary left-to-right orientation (such as sections of the *corpus callosum*), at the back for those that run back-to-front (such as the *inferior fronto-occipital fasciculus (IFO)*), and at the bottom for those that run bottom-to-top (the *cortico-spinal tract (CST)*).

We require a *joint* parametrization for all subjects, whose $|\mathcal{F}|$ fibers are represented using $|\mathcal{V}|$ vertices overall. Unfortunately, the need to build an $\mathcal{O}(|\mathcal{V}|^2)$ distance matrix makes it prohibitive to simply apply ISOMAP to the union of all fibers. Therefore, we use clustering and interpolation: First, based on ℓ_2 distances between the same feature vectors as in Section 3.2, k means [26] selects a set of representative fibers on which ISOMAP is applied. Note that clustering operates on the streamlines, which are much fewer than the number of all vertices, $|\mathcal{F}| \ll |\mathcal{V}|$.

For each vertex that was not part of the ISOMAP computation, we assign the weighted average of the values at the $k=4$ vertices that lie on representative bundles and are closest in MNI space, with inverse distance weights. Vertices are removed as outliers if their distance to the k-th nearest representative exceeds a threshold ($\theta = 3$ mm). An example that shows the resulting correspondences between five subjects is shown in Fig. 3.

It is well-known that using image registration alone to match different subjects can lead to imperfect correspondences [22]. Visual inspection of the results, and the fact that feature vectors based on BundleMAP achieve higher classification accuracy than two state-of-the-art alternatives (cf. Section 4), suggests that *combining* registration with manifold learning yields automated correspondences with good accuracy.

Our strategy can be described as mapping each point in each individual subject to the corresponding point on an abstract fiber bundle core. This is in contrast to most previous techniques, which work by mapping the individual curves to each other. In particular, tract-based morphometry (TBM) [15] selects a prototype fiber from the input data, and matches other fibers to it. A connection to BundleMAP can be made by interpreting the TBM prototype fiber as an explicit representation of our fiber bundle core. However, in contrast to TBM, BundleMAP treats all input fibers equally, and

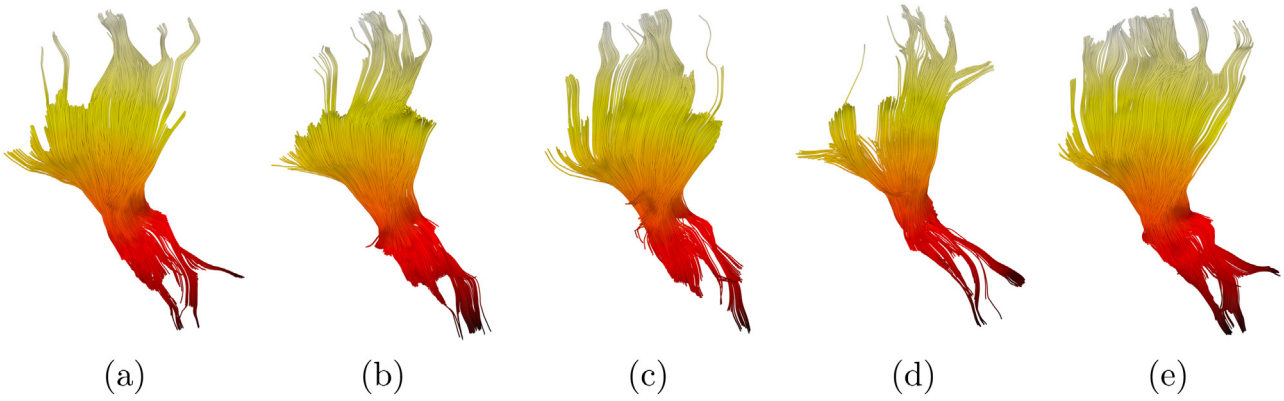

Fig. 3. A joint parametrization of the left *corticospinal tract*, shown from the left, has been created using our method, and is color coded on five example subjects. The success of our method can be seen from the fact that same colors indicate corresponding anatomical locations.

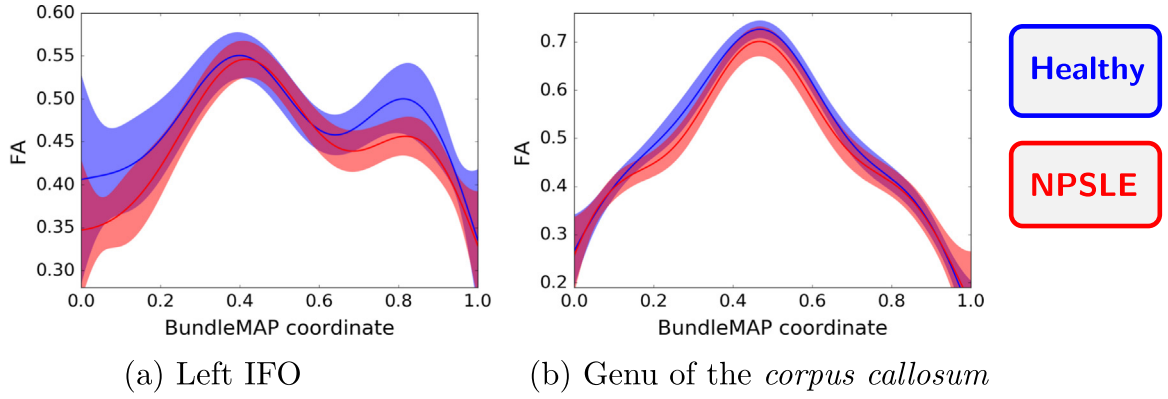

Fig. 4. BundleMAP allows us to plot the mean and a 95% confidence interval of Fractional Anisotropy (FA) along fiber bundles, such as the left *inferior fronto-occipital fasciculus* (a), or the genu of the *corpus callosum* (b). This reveals that, in a comparison between healthy controls (blue) and NPSLE patients (red), some sections of the bundles are more strongly affected than others. (a) Left IFO (b) Genu of the *corpus callosum*.

does not introduce a dependence on one particular representative, whose most suitable choice might not always be obvious.

3.4. Visualizing diffusion parameters as a function of bundlemap coordinate

Quantitative analysis of diffusion MRI is based on parameters such as Fractional Anisotropy (FA), the values of which are often affected by neurodegenerative or neuroinflammatory diseases. Combining our BundleMAP coordinates with Gaussian Process Regression [27] can be used to visualize how FA changes along major nerve fiber bundles, such as the left IFO (Fig. 4 (a)) and the genu of the *corpus callosum* (b). Since FA depends not only on fiber integrity, but also on factors such as neuronal density, orientation dispersion [28], and the crossing of different bundles on the voxel resolution, a dependence on position along the bundle can be seen even in a healthy population.

We performed Gaussian Process Regression (GPR) with a constant linear regressor and the squared exponential covariance function $k(x_p, x_q) = \exp(-\gamma(x_p - x_q)^2)$ with $\gamma = 10$. To create the input of GPR, we subdivided the BundleMAP range $[0, 1]$ into ten uniform bins. In each bin, means and variances of FA were computed in two groups, healthy controls (blue), and patients suffering from systemic lupus erythematosus with neuropsychiatric symptoms (NPSLE, red). We only use GPR for visualizations such as Fig. 4; the derivation of features for prediction will be described next.

3.5. Deriving feature vectors by data-adaptive binning

Fig. 4 illustrates that disease often does not affect all parts of a given nerve fiber bundle equally. Accounting for this helps us derive feature vectors that should be as effective as possible for the detection of disease, while maintaining anatomical interpretability.

Our feature vectors are based on the four widely used diffusion parameters Fractional Anisotropy (FA), mean, axial and radial diffusivity, as well as the number of samples in the reconstructed bundle, which accounts for the fact that disease can make it more difficult to fully sample certain bundles. A simple way to derive a feature vector from a set of n fiber bundles would be to average those five parameters along each of the bundles, leading to a feature vector of size $5 \times n$. BundleMAP allows us to derive stronger features from smaller segments that are more specifically affected by the disease.

We uniformly subdivide the BundleMAP parameter range $[0, 1]$ into b bins, and average data from streamline vertices whose coordinates fall into the same bin. If a fiber bundle could be reconstructed only partially in a subject, its BundleMAP coordinates may span only a subset of the full $[0, 1]$ interval. When creating Fig. 4, we simply excluded subjects without data from the computation of mean and variance in the respective bin.

When using classification and regression techniques that do not support incomplete data, we use the following scheme to fill in missing entries in the feature vectors: For each parameter and bundle, and for each subject i, we compute the average μ_i over all

available points in that bundle. If data is missing in some bin, we impute it from all subjects where it is available. In particular, we take the average of their values in the same bin, and correct it for the global difference in the current subject i by adding $\mu_i - \bar{\mu}$, where the grand mean $\bar{\mu}$ is taken over the μ_j of all subjects from which data was taken. It is important to consider *all* subjects for imputation, since using only those with the same label could lead to overly optimistic estimates of accuracy, as will be discussed in Section 4.6.

Features are strongest if bins are small enough so that some of them are dominated by the disease, but large enough to decrease the variance in the resulting features, and to introduce some tolerance against imperfect correspondences. A good tradeoff is difficult to guess, and can be expected to depend on the particular bundle and disease. Therefore, we follow a data-adaptive strategy in which we precompute features that correspond to different bin sizes, and use feature selection to decide on the number of bins b that is optimal for each task and bundle. In our experiments, we allowed the algorithm to select between $b = \{1, 2, \ldots, 8\}$. We also tried setting the maximum to seven or nine, and found that it had little effect on accuracy.

From the large choice of techniques for feature selection [29], we opt for two methods that are simple and fast to compute: Fisher's criterion for classification, and Pearson correlation for regression. Given a classification task, for each bundle, we select b so that the resulting features maximize the Fisher score

$$F = \frac{\| \mathbf{x}^{(p)} - \mathbf{x}^{(a)} \|^2 + \| \mathbf{x}^{(n)} - \mathbf{x}^{(a)} \|^2}{\frac{1}{|\mathcal{P}|-1} \sum_{\mathbf{x}_i \in \mathcal{P}} \| \mathbf{x}_i - \mathbf{x}^{(p)} \|^2 + \frac{1}{|\mathcal{N}|-1} \sum_{\mathbf{x}_i \in \mathcal{N}} \| \mathbf{x}_i - \mathbf{x}^{(n)} \|^2}, \quad (2)$$

where vectors $\mathbf{x}$ include all features that result from a given bin size, $\mathcal{P}$ and $\mathcal{N}$ denote the sets of positive and negative examples in the training data, $\mathbf{x}^{(p)}$ and $\mathbf{x}^{(n)}$ the respective mean vectors, and $\mathbf{x}^{(a)}$ the grand mean.

This multivariate version of the Fisher score differs from the approach we took in our conference paper [7], which computed Fisher scores independently for each coefficient of $\mathbf{x}$, and took the maximum of the results. We found this new definition to be more robust to the fact that, as the number of bins increases, noise and outliers increase the probability of obtaining a spurious large Fisher score in at least one of them. This is analogous to the well-known problem of performing multiple comparisons in statistical hypothesis testing.

We rate features for regression according to the absolute value of their Pearson correlation with the target variable. This is performed independently for each coefficient of $\mathbf{x}$. To address the problem of multiple comparisons, we consider the mean rather than the maximum of the results. In the final feature vector, we include all features that correspond to the selected number of bins.

It has been demonstrated that, when cross-validation is used, performing feature selection as a pre-process can lead to a drastic overestimation of accuracy [5]. To avoid this, we select the number of bins as part of the cross-validation loop. We found that this choice was made in a stable manner, as shown in Section 4.5.

4. Results

4.1. Experimental setup

We applied BundleMAP to the dMRI data of 56 subjects, each acquired with 15 gradient directions at $b = 800\,\mathrm{s/mm^2}$ and one unweighted measurement at $b=0$. Out of the 56 subjects, 38 were patients with systemic lupus erythematosus (SLE), 19 with neuropsychiatric symptoms (NPSLE), 19 without (non-NPSLE). There were no significant differences between the groups in terms of age, gender, or ethnicity. All diagnoses were made by an experienced rheumatologist according to criteria established by the American College of Rheumatology [30,31]. Details on the cohort, clinical parameters, and criteria for inclusion can be found in [8].

We reconstructed 17 different fiber bundles, including *inferior fronto-occipital fasciculus* (IFO l/r), *inferior longitudinal fasciculus* (ILF l/r), *superior longitudinal fasciculus* (SLF l/r), *cingulum bundle* (CING l/r), *fornix* (FOR l/r), *anterior thalamic radiation* (ATR l/r), and *corticospinal tract* (CST l/r). Since our method provides a one-dimensional parametrization, we deal with the *corpus callosum*, an important bundle that connects the two brain hemispheres and is distinctly two-dimensional, by breaking it down into *splenium, body,* and *genu* (SPLE / BODY / GENU), which are then approximated as being one-dimensional.

Our selection includes bundles from different parts of the brain. If prior knowledge exists on what regions are most affected by a given disease, it can make sense to apply BundleMAP to a smaller set of bundles. Even though several of our bundles are curved or fan out, they were adequately parametrized by our method.

Our current implementation is not optimized for speed. Total processing time on a recent workstation (3.3 GHz, 32 GB of RAM) was less than a day. The cross-validation loop that estimated classification or regression accuracy only took around one minute, while whole-brain tractography, outlier removal, and parametrization each took several hours.

4.2. Benefit for classification

The initial task was to reliably distinguish patients suffering from NPSLE or non-NPSLE from age-matched healthy controls (HC). This amounts to a supervised classification task, for which we used a linear support vector machine [32].

Since the different parameters have vastly different scales, all features were normalized through a z transformation. Adjusting for the fact that subjects had a considerable age range ($18-67$) is a legitimate way to increase classification accuracy [8]. Since the effect of age on diffusion is known to differ between bundles [33], age correction was done separately for each bundle. In each case, we performed a linear regression with subject age as the independent variable and the respective diffusion parameter, z normalized and averaged over the full bundle, as the dependent variable. All features from the same bundle and diffusion parameter were age corrected by subtracting the prediction from the respective linear fit.

Based on the normalized and age-corrected features, our method automatically selected the most suitable number of bins, as described in Section 3.5. Classification accuracy was estimated using leave-one-out cross-validation, with feature selection happening on the training data *within* the cross-validation loop. Classification accuracy was 73% for HC vs. non-NPSLE, and 76% for HC vs. NPSLE. Classification based on averaging all parameters over the full bundles resulted in an accuracy of only 54% (HC vs. non-NPSLE) and 70% (HC vs. NPSLE). This indicates that the anatomical localization enabled by BundleMAP indeed improved accuracy.

The results also improve over our previous work, in which we had applied support vector machines with feature vectors derived from the widely used Tract-Based Spatial Statistics (TBSS) skeleton [22], and achieved an accuracy of 70% for both tasks on the same data [8]. In agreement with this previous work, we are still not able to reliably distinguish non-NPSLE from NPSLE patients based on the dMRI data alone (50% accuracy).

In our previous work [8], feature scaling was required for a useful classification based on the 122,477 TBSS skeleton voxels, many of which are irrelevant to the classification task. We also applied several feature reduction techniques [5,34] to our

BundleMAP features, including the same feature scaling as in [8], removing bundles that correlated only weakly with the target variable in the training data, or recursive feature elimination [35]. However, this did not lead to a significant and reliable improvement of classification rates, suggesting that the ≈200 features from BundleMAP provide, without additional post-processing, a sensible middle ground between the two extremes of "one feature per voxel" and "one feature per bundle".

We note that our conference paper reported an even higher accuracy of 84% for the HC vs. non-NPSLE task [7]. This reflects a programming error that had accidentally caused the use of data that came only from samples in the same class when filling in missing entries in the feature vectors, inflating the estimate of accuracy (cf. Section 3.5).

4.3. Benefit for regression

In a second task, we considered only the SLE patients, and attempted to predict the activity of their disease, as quantified by the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) [36], from the dMRI data. Features were normalized and age corrected as for classification. Regression was performed based on the $k=5$ nearest neighbors, averaged with uniform weights.

SLEDAI values in our sample ranged between 0 (indicating no recent disease activity) and 24. When using the BundleMAP features to predict these scores in a leave-one-out manner, mean and standard deviation of the absolute prediction error was 4.78 ± 3.61. Similar to the classification tasks, BundleMAP improved accuracy compared to using features from averages along the whole bundles (5.13 ± 3.83).

Our overall ability to predict SLEDAI scores from the data is limited by the fact that the index not only counts neurological symptoms, but also many symptoms like rash or hematuria, which are not directly related to the brain and are unlikely to be reflected in the dMRI data. Despite this, both ways of using the dMRI data gave more accurate results than a baseline regression that simply predicted the mean SLEDAI score from the training samples, without making any use of the features (5.43 ± 3.49).

4.4. Benefit for statistical hypothesis testing

In clinical studies, dMRI is most commonly acquired to look for statistically significant differences between populations [5]. We will now demonstrate that BundleMAP also allows us to perform statistical hypothesis testing in an anatomically localized manner. This increases statistical power in cases where populations only differ along some subsection of a bundle.

Consider the bundles and groups shown in Fig. 4. If we average FA along the whole bundle, a two-sided two-sample t test will compute a p-value of $p=0.14$ for the null hypothesis that FAs are the same in healthy controls and NPSLE patients in the left IFO. Therefore, at the common significance level $\alpha = 0.05$, the difference along the bundle as a whole is not significant.

However, using BundleMAP to subdivide the bundle into three equal parts allows for hypothesis tests with more fine-grained localization, leading to p values $p=0.055$, $p=0.527$, and $p=0.003$ for the posterior, central, and anterior part of the left IFO. When interpreting these p values, we have to account for the fact that we now performed three rather than a single test, but even with suitable Bonferroni correction ($p < 0.016$), the difference in the anterior region is clearly strong enough to be counted as significant. In the genu, a whole-tract t test yields $p=0.03$, localized analysis leads to $p=0.028$, $p=0.003$, and $p=0.09$ for the left, central, and right parts. In this case, the p value of the central part, but not the p value from the global analysis, is small enough to survive the Bonferroni correction that would be required if we had investigated the genu together with the body and splenium of the *corpus callosum*.

4.5. Importance and stability of data-adaptive binning

We now report results that confirm that our scheme for data-adaptive selection of an optimal binning resolution, described in Section 3.5, selects a resolution that is specific both to the bundle and to the task at hand.

Fig. 5 shows in what fraction of cross-validation folds each possible number of bins was selected. For each bundle, the left stack represent results from classifying HC vs. non-NPSLE, the right stack HC vs. NPSLE. The number of bins varies between one and eight and, in several cases, depends on the particular classification task. This confirms the importance of making the binning data-adaptive. Interestingly, the numbers for many homologous bundles (left/right) are correlated. If symmetry is broken, this may indicate that disease affects the two hemispheres differently.

Fig. 5 also confirms stability of our adaptive binning. In most cases, one particular number of bins is chosen in most or even all cross-validation folds. This is reassuring, since instability has been linked to poor generalization accuracy [37], and stability has been shown to be a necessary and sufficient condition for learnability [38].

4.6. Using bundlemap to identify relevant bundles

Selecting the optimal binning resolution involves quantifying the importance of features based on their Fisher score (for

Table 1
The top five features for each task (classifying controls vs. NPSLE / non-NPSLE patients or SLEDAI regression) indicate which fiber bundles are particularly affected by each subtype of the disease.

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
HC vs. NPSLE	ATR r	IFO l	GENU	ATR l	ILF r
HC vs. non-NPSLE	FOR r	FOR l	ATR r	BODY	CING l
SLEDAI	ILF l	CST l	SLF l	IFO r	ATR l

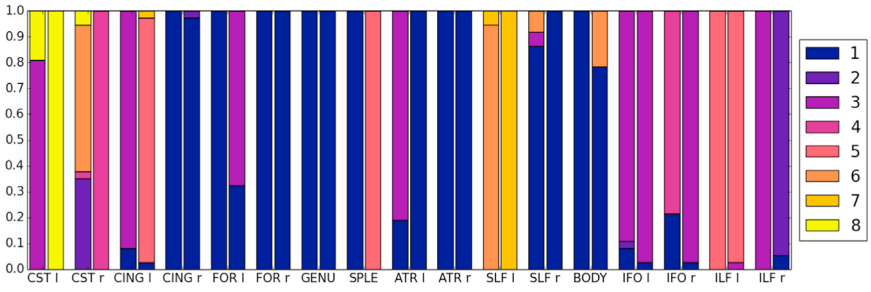

Fig. 5. The frequency with which each possible binning resolution is selected depends on the bundle and task at hand. The two stacks for each bundle represent classification of HC vs. non-NPSLE (left) and HC vs. NPSLE (right).

(a) BundleMAP (b) Arc-length

Fig. 6. Some streamlines in the ILF erroneously bend back into the bundle instead of terminating. Our BundleMAP approach is robust to this (a), while the outliers have a strong impact and are clearly visible as dark curves in an arc-length parametrization (b). (a) BundleMAP (b) Arc-length.

classification tasks) or Pearson correlation (for regression). Averaging these scores over cross-validation folds allows us to rank the fiber bundles according to their relevance to the task at hand.

Table 1 lists the five highest ranking bundles for distinguishing healthy controls from NPSLE or non-NPSLE patients. The results agree well with findings in independent data [9], in which the diffusion properties in the white matter of SLE patients were found to be affected, among others, in the *anterior thalamic radiation,* the *inferior fronto-occipital fasciculus*, and the *fornix*.

Results also largely match our own statistical analysis of the same data with conventional methods [8]. A notable exception is the strong involvement of the *fornix* in non-NPSLE patients, which is suggested by BundleMAP and agrees with work from others [9], but was not revealed by our own TBSS-based analysis of the same data. Interestingly, an independent recent study reported that TBSS suffers from reduced sensitivity in the *fornix*, due to difficulties in achieving a good spatial alignment [39].

The bundles ranked most highly for predicting disease activity in the SLEDAI regression task overlap with those that were rated most useful for detecting the disease in the first place, but also prominently include the *corticospinal tract*, another bundle in which changes of FA in SLE patients were reported previously [9].

4.7. Comparison to arc-length parametrization

Finally, we discuss and demonstrate the benefits of using manifold learning for bundle parametrization over a previously proposed alternative, which defines a cutting plane orthogonal to the bundle and uses the arc-length distance from its intersection with each streamline as a parametrization [10].

Since we found it challenging to reliably orient a cutting plane in a three-dimensional view, we iteratively optimized our initial plane using principal component analysis to align its normal vector with the tangents of the intersected streamlines, then updating the points of intersection. This converged in a few steps and produced a plausible result in all cases.

Reliably sampling the full extent of some bundles, such as *cingulum* or *fornix*, required placement of seed points in more than one region. In these cases, no single cutting plane intersected all streamlines. We address this by finding, for each streamline that is not intersected, the point closest to a streamline that was parametrized, copying the corresponding parameter value, and starting the arc-length measurement from there.

Often, the arc-length parametrization, normalized to $[0, 1]$, was quite similar to our BundleMAP coordinates. However, we found that manifold learning based on a k nearest neighbor graph was more robust to erroneous streamlines that occasionally survive our outlier removal step. For example, some outliers in Fig. 6 bend back into the bundle instead of terminating, which is correctly dealt with by BundleMAP, but not by arc-length.

Due to this, replacing manifold learning with an arc-length parametrization, while leaving everything else unchanged, reduced the classification accuracy from 73%/76% (HC vs. non-NPSLE/NPSLE) to only 68% in both cases.

5. Conclusion

We presented BundleMAP, a system for anatomically localized analysis of diffusion MRI data. Compared to related approaches [10,15], important contributions of our work include the use of manifold learning to achieve a robust joint parametrization of fiber bundles, the use of one-class SVMs for the automated removal of outliers, a data-adaptive mechanism for setting optimal bin sizes, and the application of supervised classification, regression, hypothesis testing, and data visualization in a single framework. We have demonstrated that our system enables improved prediction accuracy, statistical power, and retains anatomical interpretability.

In our future work, we plan to adapt BundleMAP to the processing of modern HARDI and multi-shell dMRI data [40–42], and to integrate it into an even larger system that will fuse information from other modalities, such as fMRI.

Conflict of Interest

None declared.

Acknowledgments

TS received funding through DFG grant SCHU 3040/1-1. PCS has, in part, been supported by funding from MICHR-02075, Alfred Österlunds Research Foundation, Sweden, and Skane University Research Funding, Sweden.

References

[1] D.K. Jones (Ed.), Diffusion MRI: Theory, Method, and Applications, Oxford University Press, 2011.
[2] Y. Rathi, J. Malcolm, O. Michailovich, J. Goldstein, L. Seidman, R.W. McCarley, C.-F. Westin, M.E. Shenton, Biomarkers for identifying first-episode schizophrenia patients using diffusion weighted imaging, in: Proceedings of the Medical Image Analysis and Computer-Assisted Intervention (MICCAI) Part I, Vol. 6361 of LNCS, Springer, 2010, pp. 657–665.
[3] C.-Y. Wee, P.-T. Yap, W. Li, K. Denny, J.N. Browndyke, G.G. Potter, K.A. Welsh-Bohmer, L. Wang, D. Shen, Enriched white matter connectivity networks for accurate identification of MCI patients, NeuroImage 54 (3) (2011) 1812–1822.
[4] Y. Cui, W. Wen, D.M. Lipnicki, M.F. Beg, J.S. Jin, S. Luo, W. Zhu, N.A. Kochan, S. Reppermund, L. Zhuang, P.R. Raamana, T. Liu, J.N. Trollor, L. Wang, H. Brodaty, P. S. Sachdev, Automated detection of amnestic mild cognitive impairment in community-dwelling elderly adults: a combined spatial atrophy and white matter alteration approach, NeuroImage 59 (2) (2012) 1209–1217.
[5] L. O'Donnell, T. Schultz, Statistical and Machine Learning Methods for Neuroimaging: Examples, Challenges, and Extensions to Diffusion Imaging Data, in: I. Hotz, T. Schultz (Eds.), Visualization and Processing of Higher Order Descriptors for Multi-Valued Data, Springer, 2015, pp. 293–313.
[6] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[7] M. Khatami, T. Schmidt-Wilcke, P. Sundgren, A. Abbasloo, B. Schölkopf, T. Schultz, BundleMAP: anatomically localized features from dMRI for detection of disease, in: Proceedings. International'l Workshop on Machine Learning in Medical Imaging (MLMI), Vol. 9352 of LNCS, Springer, 2015, pp. 52–60.
[8] T. Schmidt-Wilcke, P. Cagnoli, P. Wang, T. Schultz, A. Lotz, W.J. Mccune, P. C. Sundgren, Diminished white matter integrity in patients with systemic lupus erythematosus, NeuroImage: Clin. 5 (2014) 291–297.

[9] B.J. Emmer, I.M. Veer, G.M. Steup-Beekman, T.W.J. Huizinga, J. van der Grond, M.A. van Buchem, Tract-based spatial statistics on diffusion tensor imaging in systemic lupus erythematosus reveals localized involvement of white matter tracts, Arthritis Rheum. 62 (12) (2010) 3716–3721.
[10] H. Zhu, M. Styner, N. Tang, Z. Liu, W. Lin, J.H. Gilmore, FRATS: functional regression analysis of DTI tract statistics, IEEE Trans. Med. Imaging 29 (4) (2010) 1039–1049.
[11] I. Corouge, P.T. Fletcher, S. Joshi, S. Gouttard, G. Gerig, Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis, Med. Image Anal. 10 (2006) 786–798.
[12] J.B. Colby, L. Soderberg, C. Lebel, I.D. Dinov, P.M. Thompson, E.R. Sowell, Along-tract statistics allow for enhanced tractography analysis, NeuroImage 59 (2012) 3227–3242.
[13] P.A. Yushkevich, H. Zhang, T.J. Simon, J.C. Gee, Structure-specific statistical mapping of white matter tracts, NeuroImage 41 (2008) 448–461.
[14] D. Wang, Y. Luo, V.C. Mok, W.C. Chu, L. Shi, Tractography atlas-based spatial statistics: Statistical analysis of diffusion tensor image along fiber pathways, NeuroImage 125 (2016) 301–310.
[15] L.J. O'Donnell, C.-F. Westin, A.J. Golby, Tract-based morphometry for white matter group analysis, NeuroImage 45 (2009) 832–844.
[16] P.J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, A. Aldroubi, In vivo fiber tractography using DT-MRI data, Magn. Reson. Med. 44 (2000) 625–632.
[17] J.-D. Tournier, F. Calamante, A. Connelly, Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution, NeuroImage 35 (2007) 1459–1472.
[18] T. Schultz, C.-F. Westin, G. Kindlmann, Multi-diffusion-tensor fitting via spherical deconvolution: A unifying framework, in: T. Jiang, N. Navab, J.P. W. Pluim, M.A. Viergever (Eds.), Proceedings. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Vol. 6361 of LNCS, Springer, 2010, pp. 673–680.
[19] S. Wakana, A. Caprihan, M.M. Panzenboeck, J.H. Fallon, M. Perry, R.L. Gollub, K. Hua, J. Zhang, H. Jiang, P. Dubey, A. Blitz, P. van Zijl, S. Mori, Reproducibility of quantitative tractography methods applied to cerebral white matter, NeuroImage 36 (2007) 630–644.
[20] D.C. Alexander, C. Pierpaoli, P.J. Basser, J.C. Gee, Spatial transformations of diffusion tensor magnetic resonance images, IEEE Trans. Med. Imaging 20 (11) (2001) 1131–1139.
[21] M. Jenkinson, C.F. Beckmann, T.E. Behrens, M.W. Woolrich, S.M. Smith, FSL, NeuroImage 62 (2) (2012) 782–790.
[22] S.M. Smith, M. Jenkinson, H. Johansen-Berg, D. Rueckert, T.E. Nichols, C. E. Mackay, K.E. Watkins, O. Ciccarelli, M.Z. Cader, P.M. Matthews, T.E. J. Behrens, Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data, NeuroImage 31 (4) (2006) 1487–1505.
[23] I.N.C. Lawes, C.A. Clark, Anatomical Validation of DTI and Tractography, in: D. K. Jones (Ed.), Diffusion MRI: Theory, Method, and Applications, Oxford University Press, 2011, pp. 439–447, Ch. 26.
[24] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution, Neural Comput. 13 (2001) 1443–1471.
[25] A. Brun, H. Knutsson, H.J. Park, M.E. Shenton, C.-F. Westin, Clustering fiber tracts using normalized cuts, in: C. Barillot, D. Haynor, P. Hellier (Eds.), Medical Image Computing and Computer-Assisted Intervention (MICCAI), Vol. 3216 of Lecture Notes in Computer Science, Springer, 2004, pp. 368–375.
[26] R. Xu, Clustering, IEEE Series on Computational Intelligence, Wiley, 2009.
[27] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.
[28] H. Zhang, T. Schneider, C.A. Wheeler-Kingshott, D.C. Alexander, NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain, NeuroImage 61 (4) (2012) 1000–1016.
[29] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[30] E.M. Tan, A.S. Cohen, J.F. Fries, A.T. Masi, D.J. Mcshane, N.F. Rothfield, J.G. Schaller, N. Talal, R.J. Winchester, The 1982 revised criteria for the classification of systemic lupus erythematosus, Arthritis Rheum. 25 (11) (1982) 1271–1277.
[31] The american college of rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes, Arthritis & Rheumatism 42 (4) (1999) pp. 599–608.
[32] B. Schölkopf, A.J. Smola, Learning with Kernels, MIT Press, 2002.
[33] E.V. Sullivan, A. Pfefferbaum, Diffusion Tensor Imaging in Aging and Age-Related Neurodegenerative Disorders, in: D.K. Jones (Ed.), Diffusion MRI: Theory, Method, and Applications, Oxford University Press, 2011, pp. 624–643, Ch. 38.
[34] B. Mwangi, T.S. Tian, J.C. Soares, A review of feature reduction techniques in neuroimaging, Neuroinformatics 12 (2014) 229–244.
[35] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2002) 389–422.
[36] C. Bombardier, D.D. Gladman, M.B. Urowitz, D. Caron, C.H. Chang, Derivation of the SLEDAI. a disease activity index for lupus patients. the committee on prognosis studies in SLE, Arthritis Rheum. 35 (6) (1992) 630–640.
[37] J. Honorio, D. Tomasi, R.Z. Goldstein, H.-C. Leung, D. Samaras, Can a single brain region predict a disorder? IEEE Trans. Med. Imaging 31 (11) (2012) 2062–2072.
[38] S. Shalev-Shwartz, O. Shamir, N. Srebro, K. Sridharan, Learnability, stability and uniform convergence, J. Mach. Learn. Res. 11 (2010) 2635–2670.
[39] M. Bach, F.B. Laun, A. Leemans, C.M. Tax, G.J. Biessels, B. Stieltjes, K.H. Maier-Hein, Methodological considerations on tract-based spatial statistics (TBSS), NeuroImage 100 (2014) 358–369.
[40] J.-D. Tournier, S. Mori, A. Leemans, Diffusion tensor imaging and beyond, Magn. Reson. Med. 65 (6) (2011) 1532–1556.
[41] L. Astola, A. Fuster, L. Florack, A riemannian scalar measure for diffusion tensor images, Pattern Recognit. 44 (2011) 1885–1891.
[42] M. Ankele, T. Schultz, Quantifying microstructure in fiber crossings with diffusional kurtosis, in: Proceedings. Medical Image Analysis and Computer-Aided Intervention (MICCAI) Part I, Vol. 9349 of LNCS, 2015, pp. 52–60.

Mohammad Khatami is a member of the Visualization and Medical Image Analysis group at the University of Bonn, Germany. He received a BSc degree in Information Technology Engineering at Islamic Azad University of Iran in 2008, and a MSc degree in Media Informatics at RWTH Aachen University, Germany, in 2016.

Tobias Schmidt-Wilcke is an associate professor for neuroscientific imaging at the Department of Neurology at the Berufsgenossenschaftlichen Universitätsklinikum Bergmannsheil, at the Ruhr-University Bochum (since 2012). Tobias Schmidt-Wilcke is a trained neurologist. From 2009/2011 he worked two years at the Chronic Pain an Fatigue research Center at the University of Michigan. From 2011/2012 he worked at the Department of Neurology in Tübingen.

Pia C. Sundgren is a senior consultant neuroradiologist and full Professor and the academic Head at the Department of Diagnostic Radiology at Lund University, Lund, Sweden and adjunct Professor of Clinical Research at University of Michigan.

Amin Abbasloo is a member of the Visualization and Medical Image Analysis group at the University of Bonn, Germany. He received a BSc degree in physics from the Amirkabir University of Technology, Tehran, Iran in 2008, and a MSc in physics from the University of Bonn in 2013.

Bernhard Schölkopf has researched at AT&T Bell Labs, at GMD FIRST, Berlin, and at Microsoft Research Cambridge, UK, before becoming a Max Planck director in 2001. He received the J.K. Aggarwal Prize of the International Association for Pattern Recognition, the Max Planck Research Award, the Academy Prize of the Berlin-Brandenburg Academy of Sciences and Humanities, and the Royal Society Milner Award.

Thomas Schultz is a junior professor for computer science at the University of Bonn, Germany. After receiving his doctorate in 2009 from Saarland University and MPI Informatics, he spent two years as a DAAD postdoctoral fellow at the University of Chicago, then worked at the MPI for Intelligent Systems.