A Random Forest Based Method for classification of White Matter Fiber-Bundles in Multiple Sclerosis

Claudio STAMILE[1], Gabriel KOCEVAR[1], Salem HANNOUN[1], Carole FRINDEL[1],
François COTTON[1,3], Françoise DURAND-DUBIEF[1,2], David ROUSSEAU[1], Dominique SAPPEY-MARINIER[1,4]

[1]CREATIS; CNRS UMR5220; INSERM U1044; Université de Lyon, Université Lyon 1, INSA-Lyon, Villeurbanne, France

[2]Service de Neurologie A, Hôpital Neurologique, Hospices Civils de Lyon, Bron, France

[3]Service de Radiologie, Centre Hospitalier Lyon-Sud, Hospices Civils de Lyon, Pierre-Bénite, France

[4]CERMEP - Imagerie du Vivant, Université de Lyon, Bron, France

{stamile,kocevar,hannoun,frindel,rousseau}@creatis.insa-lyon.fr,
{francoise.durand-dubief,francois.cotton}@chu-lyon.fr,
dominique.sappey-marinier@univ-lyon1.fr

Résumé – Le développement de méthodes avancées d'acquisition et de traitement des images contribue à l'identification de nouveaux biomarqueurs permettant la caractérisation de maladies cérébrales. Dans cette étude, nous nous intéressons à la classification automatique des stades cliniques basés sur les images. Spécifiquement, nous proposons une méthode entièrement automatisée de classification par random forest des différentes formes de scléroses en plaques à partir d'images de tenseur de diffusion. Cette méthode permet d'identifier : *i)* la façon dont chaque faisceau de fibres contribue à la classification et *ii)* quelle métrique de diffusion est la plus descriptive pour analyser la dégradation qui survient dans certains faisceaux de fibres.

Abstract – With the development of advanced image acquisition and processing techniques providing better biomarkers for the characterization of brain diseases, the automatic classification of biomedical imaging constitutes an important field in research. In this work, we describe a new fully automated random forest based method to classify multiple sclerosis clinical forms using information derived from diffusion tensor imaging. This method allows to identify: *i)* how each fiber-bundle contributes to the classification; *ii)* what diffusion metric is more descriptive to analyze the degeneration occurring in certain fiber-bundles.

1 Introduction

Multiple sclerosis (MS) etiology is yet to be understood. MS constitutes a rich source of open problems for image processing. This includes for instance: lesions segmentation algorithms [1], longitudinal statistical analysis [2], new acquisition models [3] and other automatic algorithms, to address specific questions like quantification of well-known brain bio-markers. In this work, we will focus on the classification of MS patients in different clinical groups of the disease progression. Currently, MS patients are classified by the neurologist in different clinical forms based on their clinical history and status. For the first time, we will try to solve this prognostic question using a computer-based method. Due to the unknown etiology of MS, "model based" approaches could be difficult to formalize. This limitation could be easily overcome using a "data-driven" approach based on machine learning algorithms [4]. Therefore, we propose a new fully automated method based on random forest classifier to classify MS clinical forms using diffusion tensor imaging (DTI) derived metrics [5]. Furthermore, we enriched our approach by selecting well-known white matter (WM) fiber-bundles for the analysis. We focalized our analysis to find for each WM region the most sensitive diffusion marker that better discriminate the different MS clinical forms. This work presents the first step for the creation of other "data-driven" methods that could be developed to use data derived from image modalities, longitudinal changes and clinical history of the patient.

2 Materials and Methods

2.1 Subjects and Acquisition

Twenty-five relapsing remitting (RR) patients and 26 secondary progressive (SP) patients with definite MS were included in this study. Clinical history was collected and neurological examinations including the EDSS and the MS functional composite tests, were performed by a board-qualified neurologist for all patients. Patients are diagnosed as definite MS according to the McDonald's criteria [6]. They were then classified by the neurologist in different clinical forms based on their clinical history and status. Twenty-six healthy volunteers with no his-

tory or signs of neurologic disorders served as control subjects for the study. Local ethical committee approval and written informed consent from all participants were obtained.
The DTI protocol was based on a 2D multi-slice spin-echo echo-planar imaging (EPI) sequence (TR/TE=6900/86 ms, acquisition time=7 min). Fifty-one contiguous, 2.5mm thick, axial slices according to the anterior commissure-posterior commissure plane were acquired. Twenty-four diffusion-gradient directions ($b = 1000\frac{s}{mm^2}$) were applied. A nominal isotropic $2.5mm^3$ resolution was obtained by using a matrix size of 96 x 96 over a field of view of 240 x 240 mm. The b0 ($b = 0\frac{s}{mm^2}$) image was acquired four times to increase signal to noise ratio while the other directions were acquired twice.

2.2 Image Pre-Processing and Data Preparation

The entire data processing and classification pipeline is composed of three steps: *i*) registration and pre-processing of DTI data; *ii*) tractography and fiber-bundle extraction; *iii*) feature selection and classification. Diffusion images were processed using the FRMIB software Library (FSL) [7]. First, eddy current correction filter was applied to the 24 diffusion volumes using the b0 as reference. After calculating the tensor model using the FDT module of FSL, fractional anisotropy (FA), mean diffusivity (MD), radial (λr) and axial (λa) diffusivity maps were computed. Finally, all diffusion maps of each subject were co-registered (non-rigid) on the Illinois Institute of Technology Atlas (IIT3) [8]. In order to extract the fiber-bundles, all the 20 regions of interests (ROIs) contained in the JHU fiber-bundles atlas [9] were used as seeds and masks for tractography. To this end, a probabilistic streamline approach was applied on the data of IIT3 atlas using the MRTrix probabilistic tractography algorithm [10]. This process was repeated for each fiber-bundle in the atlas. The suppression of the false positive fibers obtained during the tractography was performed by post-processing tractography outputs. More in detail for each fiber-bundle the Quick Bundle cluster algorithm [11] was applied and only the cluster with the highest mean fiber length was selected. The last process consisted in the automatic extraction of the diffusion metrics from the fiber-bundle. Based on the resampled fibers, each fiber point (x_i, y_i, z_i) was associated with the diffusion metric value of its corresponding voxel (x_i, y_i, z_i). Thus, every point of the fiber-bundle was associated with a set of diffusion metrics values allowing the characterization of the diffusion properties of the entire bundle. For each fiber-bundle four tables were generated, one for each of the four diffusion metrics (FA, MD, λr and λa). Each of those tables has 77 instances (one for each subject) and m features. Each entry of the features vector contains the diffusion value of one of the n voxels belonging to the fiber-bundle. Note that due to the different length and width of the fiber-bundles, the size of the feature vectors is not the same among them.

2.3 RELIEF-F feature selection

In data classification, the feature selection is one of the most important steps during the data preparation. The role of this step is to reduce the cardinality of the feature vector in order to: 1) improve the computational time due to the fact that the classification algorithm has to explore a smallest set of values, and 2) increase the classification performances by selecting the subset of attributes that could better discriminate the different classes in the dataset. Due to its large application in image classification, in this work the RELIEF-F [12] attribute selector was used to perform the feature selection step. Since each fiber-bundle could contain a large number of voxels ($\geq$ 10000) we chose RELIEF-F attribute selection due to its linear complexity and to its high noise tolerance.

The idea behind RELIEF-F is to assign at each feature x_i one value called weight (w_i). The weight value takes into account the relation of the current feature x_i with respect to its neighbor. More in detail, let $Y = \{1 \dots C\}$ be the set with all the classes of the dataset. The weight of each feature is assigned by the following equation:

$$w_i = w_i + \sum_{c \in Y, c \neq Y(x)} \frac{P(c)}{1 - P(c)} (|x_i - NM_c^i(x)| - |x_i - NH_c^i(x)|)$$

where $P(c)$ is the *apriori* to belong to the class c, $NM_c^i(x)$ is set of Near-Miss instances (instances near to x belonging to a different class) and $NH_c^i(x)$ is set of Near-Hit instances (instances near to x belonging to the same class of x).

2.4 Random Forest Classification

In this work, the random forest classifier [13] developed in WEKA [14] was selected due to its robustness and its simple handling. Random forests are part of the ensemble methods for classification that use a collection of z decision trees. The principle of this classification technique is the following: during the training part, for each decision tree bootstrap aggregating (bagging) is used to create sample of the same size as the training data is created. For each node a random subset of features is chosen and each tree is fully grown without pruning. At the end of the training process the random forest generates z decision trees capable to classify new instances. For the testing step, a new instance is given as input to the random forest that will generate z class prediction (one for each tree). The class prediction that will be assigned to the new instance is the most voted class from all the z classifiers.

2.5 Cross Validation and Performances Evaluation

The main requirement of the classifier evaluator is to have two datasets: test and a training test respectively used to train and to test the classifier. If an external test set is not provided, the same dataset should be used to perform both training and test phases. The most common way to perform this split is to use

the so-called K-Fold cross-validation [15]. With this method a given dataset with l instances is randomly split in k different subsets. The instances contained in $k-1$ subsets are then used as training set while the remaining instances are used as test set. The classifier is then evaluated using the classification results on those subsets of data. The test is repeated k-times in order to use all the k subsets as test set. In this work the generalization of the classifier performances was ensured by performing a 10 K-Fold cross-validation. The performance measurements used in this work are based on the analysis of the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) instances classified during the testing phase. Precision, recall and F-Measure were used to measure the classification performances. More in detail, precision reflects the fraction of retrieved instances that are correctly classified, and is defined as $\frac{TP}{TP+FP}$. Recall represents the portion of positive instances that are correctly identified and is defined as $\frac{TP}{TP+FN}$. F-Measure is obtained combining precision and recall and it is defined as $2 * \frac{(precision*recall)}{(precision+recall)}$.

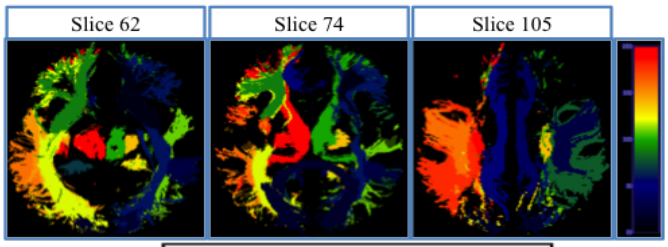
3 Results

The bundle classification pipeline was applied to the 20 fiber-bundles namely: major and minor forceps of the Corpus Callosum (CC), left (L) and right (R), Cortico-Spinal Tract (CST), Inferior Fronto-Occipital Fasciculus (IFOF), Anterior Thalamic Radiation (ATR), and Uncinate Fasciculus (UC), Cingulate Gyrus (CG), Hippocampus (HP), Inferior Longitudinal Fasciculus (ILF), Superior Longitudinal Fasciculus (SLF), Superior Longitudinal Fasciculus Temporal (SLFT). The classification performances were obtained for each fiber-bundle using one diffusion metric as feature vector.

3.1 Classification Results

For each fiber-bundle and each of the diffusion metrics of the 4 tables, the RELIEF- F attribute selection algorithm was applied. Tests were performed (results not showed) to find the best number of features to select and the number of threes of the random forest. Based on those results, the 1000 most relevant features (higher weights) were selected for the classification task performed with 778 trees. The results of the classification performances for each fiber-bundle using the 4 diffusion metrics are shown in Figure 1. The highest classification was achieved with FA where the F-Measure ranged between 71.6% and 85.7% (mean=77.6%). F-Measures were similar in MD and λr ranging between 65.8% and 89.5% (mean=76.8%) and 66.9% and 84.1% (mean=77.0%) respectively. The worst classification performances were obtained for λa, where the F-Measure ranged between 63.7% and 84.5% (mean=75.6%).

The best classification performances were reached for MD in the left ATR, and FA in the left CST. High levels of classification performances were also found in left CST, CC, IFOF and SLF. We ranked each fiber-bundle according to their F-Measure values. This rank position was used to create a map of the fiber-bundles that contributes the most for classification of MS patients (Figure 1).



	F-Measure (%)			
	FA	MD	λr	λa
CC Minor	75,6	74,0	72,9	63,7
CC Major	82,9	84,4	81,9	76,6
CST L	85,7	84,4	80,6	84,4
CST R	81,8	78,1	80,6	72,6
IFOF L	76,5	76,1	81,6	78,8
IFOF R	78,6	75,1	75,3	78,0
ATR L	71,6	89,5	80,4	80,6
ATR R	73,0	77,8	79,2	79,4
UN L	74,3	79,0	77,8	75,1
UN R	72,5	73,7	72,4	69,2
CG L	78,2	73,7	69,1	72,5
CG R	74,1	72,4	77,9	71,2
HP L	74,2	77,0	76,2	78,9
HP R	73,9	67,4	81,7	72,5
ILF L	76,3	77,6	81,8	76,4
ILF R	81,4	76,7	66,9	76,9
SLF L	84,4	80,2	81,4	77,6
SLF R	76,4	75,0	69,1	78,9
SLFT L	84,3	79,0	84,1	76,3
SLFT R	76,4	65,8	68,6	72,5

Figure 1: (Top) Fiber-bundle colored according to their F-Measure; fiber-bundles with low F-Measure value in blue and with high F-Measure in red. (Down) Table with the F-Measure performances obtained from the classification of each fiber-bundle with a particular diffusion metric.

4 Discussion

Our method provided a complete, operator independent and automated processing pipeline applicable in large cohort studies. Such reliability stands on the accuracy and robustness of the pre- and post-processing procedures. By combining the measurements of diffusion metrics on selected fiber-bundles, the random forest classifier demonstrated high degrees of classification performances in terms of F-Measure. The resulting performance values suggest that each fiber-bundle contributes differently to the classification analysis. Indeed, certain fiber-bundles, namely CST, ATR, and IFOF presented a more accurate classification compared to the others. From a clinical point of view, we observed that the fiber-bundles highlighted by the classification are related to the typical MS clinical symptoms like fatigue [16] and motor impairment [17]. Moreover, we showed that the classification results depend also on the sensitivity of each diffusion metrics. In agreement with our previous results, the best classification performances were obtained using FA, MD and λr, while poor classification performances were reached using λa.

5 Conclusion

We describe a first approach to classify clinically the MS patient using only information derived form images. As principal results we have shown how the use of "data-driven" methods like machine learning algorithm are very suitable in environments where build a model is not possible like in MS. Our method present high degree of classification allowing also to rank the WM fiber-bundles according to their capability to discriminate the MS clinical forms. Finally our results are consistent with the clinical studies. In conclusion, this method offers a potential new tool to better characterize the pathological mechanisms occurring along and inside the WM fibers of MS patients. Such improved image biomarker identification could provide a new approach for the classification of different MS clinical forms and to better understand the MS disease evolution, if longitudinal data are available.

6 Acknowledgements

Claudio Stamile is funded by an EU-funded FP7-PEOPLE-2012-ITN project 316679 TRANSACT. This work is supported by the French National Research Agency (ANR) within the national program "Investissements d'Avenir" through the OFSEP project (ANR-10-COHO-002) and the LABEX PRIMES (ANR-11-LABX-0063) of Lyon University (ANR-11-IDEX-0007).

References

[1] X. Lladó, A. Oliver, M. Cabezas, J. Freixenet, J. C. Vilanova, A. Quiles, L. Valls, L. Ramió-Torrentà, and À. Rovira, *Segmentation of multiple sclerosis lesions in brain mri: A review of automated approaches*, Information Sciences, vol. 186, no. 1, pp. 164 – 185, 2012.

[2] A. Grigis, V. Noblet, F. Blanc, F. Heitz, J. de Seze, S. Kremer, and J.-P. Armspach, *Longitudinal change detection: inference on the diffusion tensor along white matter pathways*, Med Image Anal, vol. 17, no. 3, pp. 375–386, 2013.

[3] H. Zhang, T. Schneider, C. A. Wheeler-Kingshott, and D. C. Alexander, *Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain*, Neuroimage vol. 61, no. 4, pp. 1000–1016, 2012.

[4] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

[5] P. B. Kingsley, *Introduction to diffusion tensor imaging mathematics: Part ii. anisotropy, diffusion-weighting factors, and gradient encoding schemes*, Concepts in Magnetic Resonance Part A, vol. 28A, no. 2, pp. 123–154, 2006b.

[6] W.I McDonald, A. Compston, G. Edan, D. Goodkin, H.P. Hartung, F. Lublin, H. F McFarland, D. W. Paty, C. H Polman, S.C Reingold, *Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis*, Annals of neurology, vol. 50, no. 1, pp.121–127, 2001.

[7] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, *Fsl*, Neuroimage, vol. 62, no. 2, pp. 782–790, 2012.

[8] A. Varentsova, S. Zhang, and K. Arfanakis, *Development of a high angular resolution diffusion imaging human brain template*, Neuroimage, vol. 91, pp. 177–186, 2014.

[9] K. Hua,J. Zhang,S. Wakana, J. Setsu, L. Hangyi, R. Xin,D.S. Reich, P.A. Calabresi, J.J Pekar, P.C.M. van Zijl, S. Mori, *Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification*, Neuroimage, vol. 39, pp. 336–347, 2009.

[10] J. Tournier, F. Calamante, and A. Connelly, *Mrtrix: diffusion tractography in crossing fiber regions*, International Journal of Imaging Systems and Technology, vol. 22, no. 1, pp. 53–66, 2012.

[11] E. Garyfallidis, M. Brett, M. M. Correia, G. B. Williams, and I. Nimmo-Smith, *Quickbundles, a method for tractography simplification*, Front Neurosci, vol. 6, p. 175, 2012.

[12] F. Wenbing, W. Quanquan, and Z. Hui, *Feature selection method based on adaptive relief algorithm,* CCEE 2010, vol. 53, no. 2, 2012.

[13] L. Breiman, *Random forests*, Machine learning, vol. 45, no. 1, pp. 5–32, 200.

[14] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, *The WEKA Data Mining Software: An Update*, SIGKDD Explor. Newsl., vol. 11, pp. 10–18, 2009.

[15] R. Kohavi , *A study of cross-validation and bootstrap for accuracy estimation and model selection*, IJCAI, vol. 14, no. 2, 1995, pp. 1137–1145.

[16] M. A. Rocca, L. Parisi, E. Pagani, M. Copetti, M. Rodegher, B. Colombo, G. Comi, A. Falini, and M. Filippi, *Regional but not global brain damage contributes to fatigue in multiple sclerosis*, Radiology, vol. 273, no. 2, pp. 511–520, 2014.

[17] M. Wilson, C. R. Tench, P. S. Morgan, and L. D. Blumhardt, *Pyramidal tract mapping by diffusion tensor magnetic resonance imaging in multiple sclerosis: improving correlations with disability*, J Neurol Neurosurg Psychiatry, vol. 74, no. 2, pp. 203–207, 2003.