Classification of MS Patients from the Geometry and Texture of White Matter Lesions

Bernd Taschler[1], Kerstin Bendfeldt[2], Nicole Müller-Lenke[2], Till Sprenger[2], Ernst-Wilhelm Radü[2], and Thomas E. Nichols[3]

[1]Centre for Complexity Science, University of Warwick, Coventry, UK,
[2]Medical Image Analysis Center, University Hospital Basel, Basel, Switzerland,
[3]Department of Statistics, University of Warwick, Coventry, UK

Introduction

In the assessment of MS, MRI data is to a large extent only used in a qualitative way, to assess the dissemination of lesions in space and time [4,5]. Studies have shown that conventional MRI measures have rather low predictive value and are therefore poor indicators for determining clinical outcomes in MS [3].
We propose an objective classification of MS disease subtype (CIS, RLRM, PRP, SCP, PRL) using support vector machine (SVM). In addition to traditional demographic and clinical measures, our features include detailed aspects of lesion geometry (as measured by Minkowski functionals) and statistics of image intensities within lesions.

Methods

Minkowski functionals [1] are used to characterize the connectivity and shape of lesions. In 3D space there are four functionals, corresponding to volume, surface area, mean breadth and Euler-Poincaré characteristic, which provide pose-independent summaries of lesion geometry. Furthermore, the original MRI images (normalized to whole-brain median of 100) are used to compute various 'texture' statistics (see Tab.1). These features are combined into summary measures over the whole brain or 13 ROI's delineating white matter track regions. In addition to demographic data and clinical scores (EDSS, PASAT), the fraction of gray matter volume to whole brain volume is also included as a feature.

SVM is a binary classification scheme based on finding a separating hyperplane that seeks to split the data set into two groups. As non-linear kernel we use radial basis functions, $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left[-||\mathbf{x}_i - \mathbf{x}_j||^2/2\sigma^2\right]$, and adopt an one-vs-one approach based on pairwise classifiers and a majority voting scheme to make predictions.
To estimate prediction accuracies, stratified k-fold cross-validation is carried out, where k is given by the number of elements in the smallest class (here k=10). Additionally, nested cross-validation is used to optimize the model parameters and ensure unbiased estimates of out-of-sample accuracy.

Data. 250 subjects were scanned on a 1.5T scanner at the University Hospital Basel, Switzerland, collecting T1, T2 & T1-Gd-enhanced images. White matter lesion masks were created by a semi-automatic procedure and each scan was affine registered to MNI space using trilinear interpolation [2]. Number of subjects per subtype: 11 CIS, 173 RLRM, 13 PRP, 43 SCP, 10 PRL.

contact: b.taschler@warwick.ac.uk

Tab.1: Features used for classification.

demographic info	sex, age, disease duration
clinical scores	EDSS (& subscores), PASAT
gray matter	GM-volume ratio to brain volume
standard measures†	total lesion count, total lesion load
lesion geometry†	Euler-Poincaré characteristic; volume, surface area, mean breadth: sum total, mean, median, max., min., standard dev.
intra-lesion intensity†	sum total, mean, median, std. dev.

† from T1, T2, T1-Gd MRI respectively; whole brain summaries or split according to 13 WM ROI's.

Tab.2: Confusion matrix for best feature set‡; overall & average accuracy: 0.560 & **0.478**.

	CIS	RLRM	PRP	SCP	PRL
CIS	**0.818**	0.182	0.000	0.000	0.000
RLRM	0.162	**0.584**	0.058	0.081	0.116
PRP	0.000	0.231	**0.308**	0.231	0.231
SCP	0.023	0.093	0.116	**0.581**	0.186
PRL	0.000	0.400	0.200	0.300	**0.100**

‡incl. GM volume, T2 median volume by WM ROI's, whole brain summaries for T1 mean-breadth standard deviation, T2 mean-breadth median, T1 & T1-Gd total intra-lesion intensities, alongside demographic and clinical covariates;

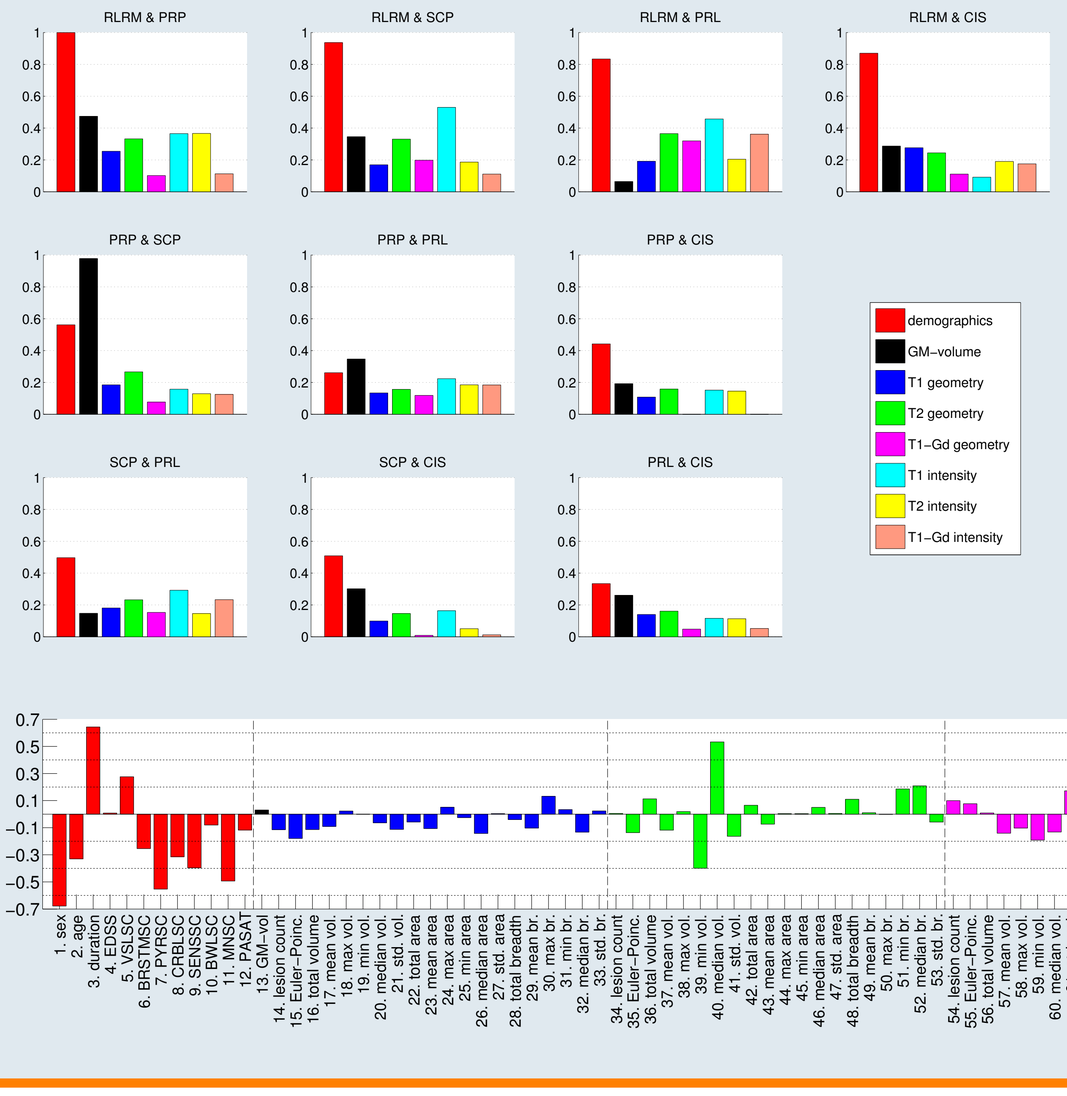

Fig.1 (*left*): Normalized root mean square errors of SVM weights across all classifiers, showing the relative significance of different kinds of features during classification.

Fig.2 (*below*): Example of standardized SVM weights for one classifier (RLRM vs. PRL). Features (whole-brain summaries) with positive weights correlate with RLRM, negative weights with PRL.

Results

We considered a number of about 50 different subsets of features as guided by scientific considerations. Tab.2 shows the confusion matrix for the feature set with the highest average prediction accuracy of 47.8% (overall 56.0%). In comparison, using only demographic and clinical covariates yields considerably lower accuracies (42.0% overall and 39.8% average accuracy).
An example of normalized support vector weights for the classifier involving RLRM and PRL is given in Fig.2. The relevance of different features varies depending on which groups are involved in the classification. For instance, median T2w lesion volume is important in RLRM vs. PRL, but less so for other groups.
The quadratic means of SVM-weights shown in Fig.1 give a comparison between different sorts of features and their variability.

In general, a comparison across classifiers indicates that the median is in many cases a better measure than the mean, that the maximum lesion volume, area or mean breadth of lesions is more meaningful than the respective minimum, and that the Euler characteristic is more useful than a simple lesion count.

Conclusions

Geometry and intra-lesion intensity improve objective classification of MS subtype. While this shows the value of detailed quantitative MRI lesion features, total accuracy remains modest ($\sim 50\%$) and thus more work is needed to improve classifier accuracy.

References
[1] Arns CH, et al. (2001), *Phys Rev E* 63(3), 031112.
[2] Bendfeldt K, et al. (2009), *NeuroImage* 45: 60–67.
[3] Lövblad KO, et al. (2010), *AJNR* 31: 983–989.
[4] MacKay-Altman R, et al. (2011), *Stat Med* 31: 449–469.
[5] Morgan C, et al. (2010), *Mult Scl* 16: 926–934.