

Classification of Multiple Sclerosis Patients from the Geometry and Texture of White Matter Lesions

Bernd Taschler¹, Kerstin Bendfeldt², Nicole Müller-Lenke²,
Ernst-Wilhelm Radü², and Thomas E. Nichols³

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



Introduction

Multiple Sclerosis (MS) is a chronic inflammatory-demyelinating disease of the central nervous system. MS is divided into five distinct clinical categories and disease progression differs significantly between these subtypes [3]. Among the five groups, relapsing-remitting (RLRM) is the most common type of MS. The others are primary progressive (PRP), secondary progressive (SCP), progressive relapsing (PRL) and clinically isolated syndrome (CIS). Currently, 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 [5,6]. Other studies have shown that conventional MRI measures have rather low predictive value and are therefore poor indicators for determining clinical outcomes in MS [4].

We propose an objective classification of MS disease subtype using support vector machine (SVM). Unlike previous work, we use a large number of quantitative features derived from three MRI sequences. In addition to traditional demographic and clinical measures, our features include aspects of lesion geometry, measured by Minkowski functionals, and statistics of the image intensities within the 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 images (T1w, T1w-Gd & T2w, normalized to whole-brain median of 100) are used to compute various ‘texture’ statistics (see Tab.1). These features, computed for every lesion, are combined into summary measures over the whole brain. To preserve information about the location of lesions the summary measures are also computed separately for each of 13 ROI’s according to segmentation into white matter track regions. In addition to demographic information and clinical scores (EDSS, PASAT), the fraction of gray matter volume to whole brain volume is also included in the feature list.

SVM is a binary classification scheme based on finding a separating hyperplane that splits the data set into two groups. As a non-linear kernel, radial basis functions are employed: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp[-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2]$. Since this is a multi-class problem, we adopt an one-vs-one approach based on a total set of ten pairwise classifiers for all combinations of the five MS subtypes and combine this with a majority voting scheme [1] 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. The data set used for our analysis consists of 250 subjects scanned at the University Hospital Basel, on a 1.5T scanner, collecting T1w, T2w & T1w-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]. The number of subjects per subtype are: 11 CIS, 173 RLRM, 13 PRP, 43 SCP, 10 PRL.

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
intra-lesion intensity[†]	sum total, mean, median, std. dev.

[†] from T1w, T2w, T1w-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, T2w median volume by WM ROI’s, whole brain summaries for T1w mean-breadth standard deviation, T2w mean-breadth median, T1w & T1w-Gd total intra-lesion intensities, alongside demographic and clinical covariates;

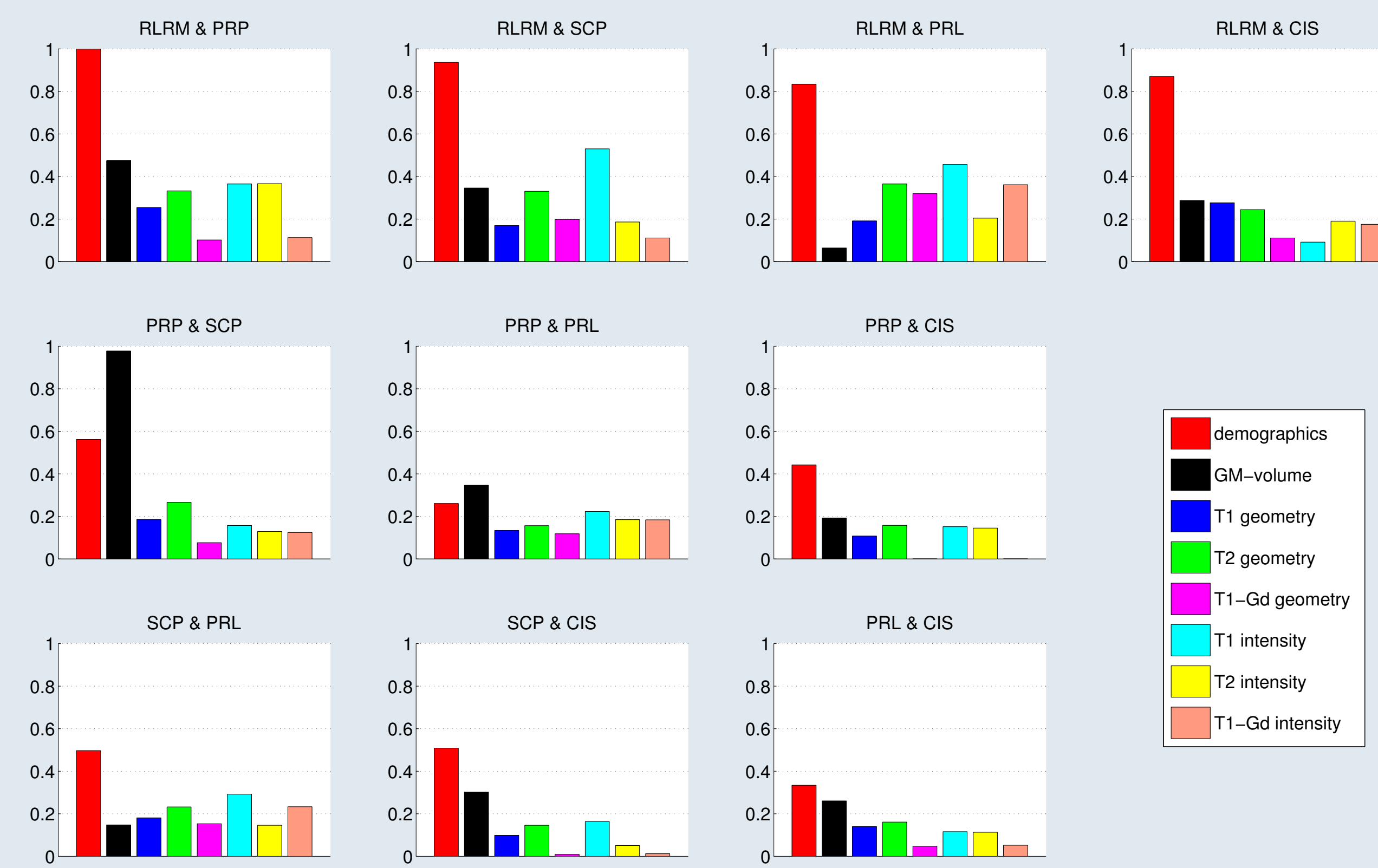
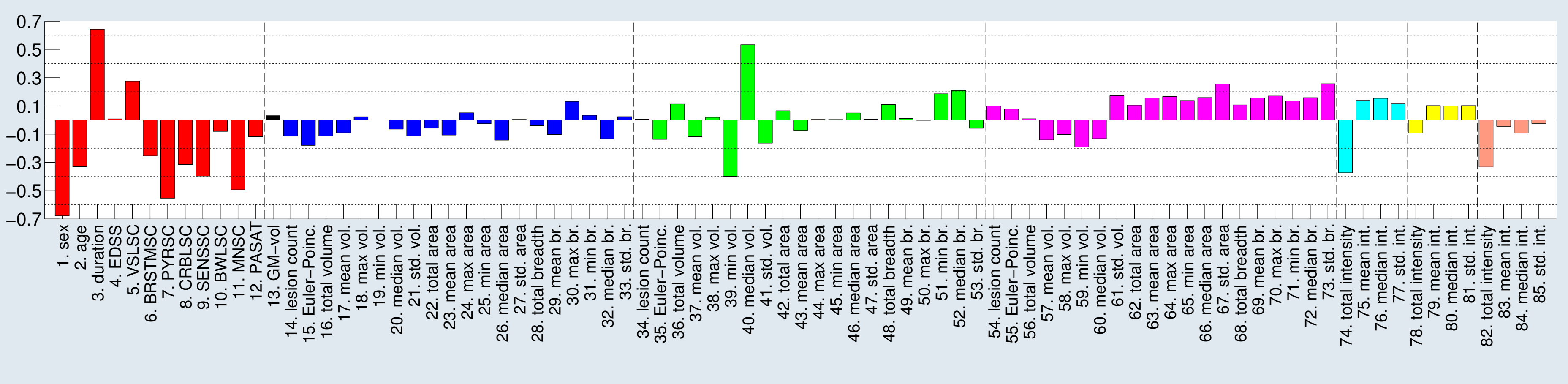


Fig.1 (left): Normalized root mean square errors of SVM weights across all classifiers, showing the relative significance of different kinds of features during classification.

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

An exhaustive combinatorial search across all features is computationally infeasible. Instead we consider a subset of possible feature combinations which are guided by the magnitude of weights of the support vectors. 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 respectively).

An example of normalized support vector weights for the classifier involving the two MS subtypes RLRM and PRL is given in Fig.2. The relevance of different geometry and intensity based 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 in other classifiers; and among the seven EDSS subscores, PYRSC and MNCS seem to be most significant.

The quadratic means of SVM-weights (Fig.1) give a comparison between different sorts of features and their variability. Demographic attributes and clinical scores carry a considerable amount of information about the disease.

In general, with regard to lesion geometry, a comparison across different classifiers indicates that the median is in many cases a better measure than the mean, that the maximum lesion volume, area or mean breadth for a single lesions is more meaningful than the respective minimum, and that the Euler-Poincaré characteristic is more significant than a simple lesion count.

Conclusions

Our work shows that geometry and intra-lesion intensity of T1w hypo-intense and T2w hyper-intense lesions improve objective classification of MS subtype, over and above that obtained by simple demographic, lesion volume and count measures.

References:

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