MRI-based prediction of conversion from clinically isolated syndrome to clinically definite multiple sclerosis using SVM and lesion geometry

Kerstin Bendfeldt[1], Bernd Taschler[2*], Laura Gaetano[1,3], Philip Madoerin[1], Pascal Kuster[1], Nicole Mueller-Lenke[1], Michael Amann[1,3], Hugo Vrenken[5], Viktor Wottschel[5], Frederik Barkhof[5,6], Stefan Borgwardt[1,4,7], Stefan Klöppel[8], Eva-Maria Wicklein[13], Ludwig Kappos[3], Gilles Edan[9], Mark S. Freedman[10], Xavier Montalbán[11], Hans-Peter Hartung[12], Christoph Pohl[13,15], Rupert Sandbrink [12,13], Till Sprenger[1,3], Ernst-Wilhelm Radue[1], Jens Wuerfel[1,15], Thomas E. Nichols[2]

*equally contributing author

[1]Medical Image Analysis Center (MIAC AG) Basel, Switzerland

[2]University of Warwick, Dept. of Statistics, Coventry, United Kingdom

[3]Department of Neurology, University Hospital Basel, Basel

[4]Department of Psychiatry (1), University of Basel, Basel, Switzerland

[5]VU University Medical Center, Amsterdam, The Netherlands

[6] Institutes of Neurology and Healthcare Engineering, UCL, London, UK

[7]King's College London, Department of Psychosis Studies, Institute of Psychiatry, London, UK

[8]Department of Psychiatry and Psychotherapy, Freiburg Brain Imaging, University Medical Center Freiburg, Freiburg, Germany

[9]CHU-Hopital Pontchaillou, Rennes, France

[10]University of Ottawa and Ottawa Hospital Research Institute, Ottawa, Canada

[11]Hospital Universitari Vall d'Hebron, Barcelona, Spain

[12]Department of Neurology, Heinrich-Heine Universität, Düsseldorf, Germany

[13]Bayer Pharma AG, Berlin, Germany

[14]University Hospital of Bonn, Bonn, Germany

[15] Charité University Medicine Berlin, Germany

Running title: Prediction of clinically definite multiple sclerosis

For submission to Human Brain Mapping as *Original Research Paper*

Abstract word count: 200

Correspondence to: PD Dr. Kerstin Bendfeldt,

Medical Image Analysis Centre (MIAC AG),

Mittlere Str. 83

CH-4031 Basel,

Switzerland

E-mail: kerstin.bendfeldt@unibas.ch

Abstract

Background: Neuroanatomical pattern classification using support vector machines (SVMs) has shown promising results in classifying Multiple Sclerosis (MS) patients based on individual structural magnetic resonance images (MRI).

Objectives: To determine whether pattern classification using SVMs facilitates predicting conversion to clinically definite multiple sclerosis (CDMS) from clinically isolated syndrome (CIS).

Methods: We used baseline MRI data from 364 patients with CIS, randomised to interferon beta-1b or placebo. Non-linear SVMs and cross-validation were applied to predict converters/non-converters (175/189) at two years follow-up based on clinical and demographic data, lesion-specific quantitative geometric features and grey-matter-to-whole-brain volume ratios. Furthermore, SVMs were applied to subgroups of n=73 converters/non-converters based on cortical grey matter segmentations.

Results: Highest prediction accuracies of 70.4% (p=8e-5) were reached with a combination of geometric (image-based) and demographic/clinical features. Cortical grey matter was informative for the placebo group (acc.: 64.6%, p=0.002) but not for the interferon group. Classification based on demographic/clinical covariates only resulted in an accuracy of 56% (p=0.05). Overall, lesion geometry was more informative in the interferon group, EDSS and sex were more important for the placebo cohort.

Conclusions: Alongside standard demographic and clinical measures, both lesion geometry and grey matter based information can aid prediction of conversion to CDMS.

Introduction

The advance of non-invasive imaging techniques, such as magnetic resonance imaging (MRI), has made a large impact on the study of Multiple Sclerosis (MS). Measures of abnormalities derived from structural MRI are useful in the context of early diagnosis, treatment planning and monitoring of disease progression. To date, however, MRI data is mainly used in a qualitative way to assess the dissemination of MS lesions in space and time.

The most common quantitative measure is lesion load, i.e. the total lesion volume. Previous work on using lesion load for classification has been inconclusive(Aban et al. 2008), (Zivadinov et al. 2005; MacKay Altman et al. 2012). Other studies have shown that conventional MRI measures have rather low predictive value and are therefore poor indicators for determining the clinical outcomes in MS(Lovblad et al. 2010). Existing quantitative methods that are used for the analysis of MS lesions are to a large extent inapt to fully reflect the given data. These methods include i) comparing lesion probability masks cross-sectionally or longitudinally(Filli et al. 2012; Locatelli et al. 2004) which makes it difficult to associate lesion locations with certain covariates of interest, ii) massive univariate methods such as voxel-based lesion-symptom mapping(Bates et al. 2003) which are ill-suited for the binary nature of lesion data and cannot account for the underlying spatial structure, and iii) smoothing of lesion masks by means of Gaussian kernels(Wei et al. 2004), which does not entirely eliminate the non-Gaussian nature of the data(Ge et al. 2014) and requires an arbitrary choice of smoothing parameter.

One of the reasons for the limited impact of the findings on clinical practice is that neuroimaging studies have typically reported population differences between groups. For neuroimaging to be useful in a clinical setting, however, inferences have to be made at the level of the individual rather than the group. One analytical method that allows such inference is multivariate pattern classification based on support vector machines (SVMs) which are sensitive to spatially distributed and subtle effects in the brain that would be otherwise undetectable using traditional methods which focus on gross differences at group level (for a review on SVMs applied to neurological conditions see(Orru et al. 2012)).

To date, no reliable method exists to predict who will, and will not, develop MS amongst those with CIS. To the best of our knowledge only one study has applied pattern classification to MRI data(Wottschel et al. 2015) in an effort to rectify this fact. The authors showed that a linear SVM correctly predicted CDMS in 71% of patients at 1 year, and in 68% at 3 years of follow-up based on different combinations of lesional and clinical features.

Our present study aims to determine whether multivariate neuroanatomical pattern classification facilitates predicting conversion to CDMS in individuals with a CIS based on structural MRI features and clinical characteristics in a two-year follow-up. Unlike previous work(Wottschel et al. 2015), we employed kernel-based SVMs in a non-linear classification scheme using a large number of quantitative features in a large population of 175 converters and 189 non-converters. Most importantly, in addition to traditional demographic, clinical and lesional measures, we included aspects of lesion geometry derived from Minkowski functionals(Legland et al. 2011). These functionals provide

quantitative measures that are closely related to standard geometric quantities such as volume, surface area and width. We demonstrate how predictive performance of non-linear SVM can be improved by using detailed geometric lesion information as well as grey matter (GM) volume.

Furthermore, we used linear SVM to investigate whether SVMs can differentiate CIS from CDMS based on cortical grey matter network characteristics Data (Wottschel et al. 2015).

Methods

Data and the BENEFIT study

This is a post-hoc analysis of MRI and clinical data from the BENEFIT study(Kappos et al. 2006; Barkhof et al. 2007). Briefly, BENEFIT was a double-blind, placebo-controlled, randomized, parallel-group, multicenter (total of 98 centers), phase 3 study that evaluated the safety, tolerability and efficacy of interferon beta-1b (Betaferon/ Betaseron; Bayer HealthCare Pharmaceuticals Inc) in patients with a monofocal or multifocal presentation of the disease, a first demyelinating event suggestive of MS and at least two clinically silent lesions on a T2-weighted brain magnetic resonance (MRI) scan. Within 60 days of the onset of the first clinical event, and after providing written informed consent, patients (n=468, aged 18-45 years) were randomly assigned, in a 5:3 ratio, to interferon beta-1b 250 μg (n=292) or placebo (n=176) subcutaneously every other day for 2 years or until CDMS was diagnosed by use of the modified Poser criteria(Bakshi et al. 2005). Patients were then eligible to enter the follow-up phase with open-label interferon beta-1b(Kappos et al. 2007). During this double-blind phase, visits were scheduled for collection of Expanded Disability Status Scale (EDSS), MRI, and other efficacy data and for safety data at months 3, 6, 9, 12, 18, and 24.

To be included in the present study, T1-weighted images with sufficient image quality acquired on 1.5T scanners together with their corresponding lesion masks, as well as clinical data at two year follow-up, had to be available. We employed subsets of placebo patients and interferon beta-1b patients who either converted to CDMS (conv+) or did not

convert (conv-) (see Table 1). Definition of the outcome measure was based on the double blind phase.

Classification based on lesion-specific geometric measures

For this approach, we considered a classification and prediction model using kernel-based (radial basis functions) support vector machines. As input to the classifier we used detailed geometric features of individual lesions derived from the available MRI data (T2-, T1-, Gd-lesions).

First, the binary lesion masks were used to identify individual lesions. Minkowski functionals(Legland et al. 2011) were then used to extract and quantify the geometry of each lesion. In standard 3D Euclidean space, Minkowski functionals are directly related to the geometric quantities volume, surface area, mean breadth and the Euler-Poincare (EP) characteristic(Arns et al. 2001; Lang et al. 2001). More details on the computation of geometric features as well as feature reduction and additional preprocessing steps (see also Figure 2) can be found in the supplementary materials.

Classification based on linear SVMs using cortical grey matter segmentations

For linear pattern classification analysis, we used LIBSVM, a library for SVMs (http://www.csie.ntu.edu.tw/-cjlin/libsvm), running under Matlab 7.1 (MathWorks, USA). Two-dimensional T1-weighted images (Placebo: n=44 converters, 25 non-converters; IFN-b: n=49 converters, 49 non-converters) with sufficient image quality aquired on three different 1.5T scanners were interpolated to an isotropic resolution of 1

mm^3,(Richert et al. 2006). Images were processed using FMRIBs software libray FSL for coregistration, segmentation (FAST), lesion filling, and interpolation (FLIRT). Interpolated images were visually checked, and then normalized, segmented and modulated using Statistical Parametric Mapping software (SPM8; Wellcome Department of Imaging Neurosciences, University College London) and Voxel- Based Morphometry (VBM8) toolbox (http://dbm.neuro.uni- jena.de/vbm8/) and DARTEL.

- FIGURE 1 ABOUT HERE

Figure 1. Preprocessing pipeline.

- FIGURE 2 ABOUT HERE

Figure 2: Preprocessing pipeline for geometry-based lesion measures and subsequent classification with nonlinear SVMs.

For the grey matter classification task, the total number of dimensions was determined by the number of voxels within the cortical GM mask. To implement a linear SVM, a kernel matrix was created from the images based on correlation, i.e. the similarity between each pair of subjects.

Predictive performance was evaluated using leave-one-out cross-validation; details are given in the supplementary materials.

Results

The demographic and clinical characteristics of patients are summarised in Table 1.

- TABLE 1 ABOUT HERE

Table 1: Baseline clinical and MRI statistics of subjects used for a) non-linear SVM analysis (T1-Gd data set), and b) linear SVM analysis (cortical grey matter maps). **Abbreviations:** T treated, NT not treated, SD standard deviation, IQR interquartile range, EDSS expanded disability status scale, bs baseline, y1 year1, GMV global grey matter volume, PGMVC percentage grey matter volume change, TCDMS time to CDMS (in days).

Classification based on feature combinations

A summary of prediction accuracies of kernel SVMs trained on different combinations of input features is given in Table 2. All values are based on nonlinear SVMs using an RBF kernel. Parameters were optimised through grid search.

One way of estimating the importance of different features is to visualise the corresponding SVM weights. Since it is not possible to obtain an analytic expression for the weights and thus separate them in the nonlinear case, any illustration has to rely on weights based on linear SVMs. Even in the linear case the plots are meant purely as a qualitative way of visualising the importance of input variables relative to one another. No quantitative assessments should be drawn directly from the magnitude of individual weights.

In general, we found that our trained classifiers were more likely to distinguish correctly between converters and non-converters in the INF-b treated cohort compared to the placebo treated group.

The simplest model (M1) utilised demographic (sex, age) and clinical (EDSS score) covariates only (accuracy: 56%, p=0.05). Including grey matter-to-whole-brain volume ratios (M3) increased prediction accuracy by a few per cent in the INF-b but not in the placebo group.

Model M2 included "traditional" measures such as lesion count and total lesion load (i.e. total lesion volume). These were combined with the grey matter volume feature in M4. Figure 3 shows the relative magnitude of SVM weights (again, resulting from training a linear SVM) for each input feature with classification accuracy equal to 57.6% (p=0.07).

M6 was solely based on whole-brain summary measures of lesion-based geometry and performed on a similar level as using demographic/clinical features only. Combining the two kinds of input information and adding grey matter volume (M7) lead to a significant improvement of classification accuracy, reaching 64.6% (p=0.0017) for the INF-b group.

Model M8 represented the largest possible feature set by splitting the summary measures of lesion geometry into 13 ROIs, plus including all other available covariates. Classifier performance was markedly reduced, most likely due to large amounts of redundant or irrelevant (i.e. equal to zero) input features. Dimensionality reduction via PCA space was only partly able to ameliorate this issue.

In M9 half of the summary statistics were excluded in order to lower the level of redundancy in the data. Only total, mean and standard deviation measures were retained

in the feature set which resulted in a prediction accuracy of 70.4% (p=8e-5). Note that splitting the same features by ROIs again reduced overall performance (cf. M10).

Figure 4 displays SVM weights for M9 of the corresponding linear classifier. The root mean square values (RMS, or quadratic means) of SVM weights shown in Figure 5 (accuracy: 60.5% , p=0.019) allow for a comparison between different kinds of features and their variability.

- TABLE 2 ABOUT HERE

Table 2: Prediction accuracy of optimised kernel SVM based on different combinations of input features; including 95% confidence intervals and corresponding p-values.
* Summaries include total, mean, median, minimum, maximum and standard deviation (SD) of lesion volume, surface area and mean breadth, respectively. Measures are based on T1-Gd MRI data. ** Total, mean and SD of volume, surface area and mean breadth (excluding median, minimum, maximum measures), and EP.
Abbreviations: DC -demographic and clinical covariates, EDSS - expanded disability status scale, EP - Euler-Poincare characteristic, GEO-brain - whole-brain summaries of geometric measures, GM - grey matter, ROI - region of interest.

For reference, training a kernel SVM with input features as specified in model M9 on T2 weighted lesion data resulted in a prediction accuracy of about 60% ($p<0.002$); and similarly for T1 weighted black-hole lesion data of about 64% ($p<0.0008$) correctly classified subjects. We also combined feature sets based on different imaging modalities, at the cost of increasing redundancy, which gave an accuracy of about 63% ($p<0.01$).

Note that for the combined-modality model the size of available data was reduced by roughly 20% as not all imaging modalities were available for all subjects.

- FIGURE 3 ABOUT HERE

Figure 3: M4. Individual weights from a linear SVM trained on features including demographic and clinical covariates, grey matter volume, lesion count and total lesion load; based on T1-Gd MRI data of IFN-b treated patients (converters: 50, non-converters: 49). Positive (negative) weights indicate that a larger feature value will drive prediction towards conversion (non-conversion). For sex, being female (male) corresponds to the positive (negative) axis. Weights are scaled relatively to the largest individual weight, which is set equal to one.

- FIGURE 4 ABOUT HERE

Figure 4: M9. Individual weights from a linear SVM trained on demographic, clinical, grey matter volume and geometry based features derived from T1-Gd MRI data of IFN-b treated patients (converters: 50, non-converters: 49). Positive (negative) weights indicate that a larger feature value will drive prediction towards conversion (non-conversion). For sex, being female (male) corresponds to the positive (negative) axis. Weights are scaled relatively to the largest individual weight, which is set equal to one.
Abbreviations: EDSS - expanded disability status scale, EPC - Euler-Poincare characteristic, GM - grey matter, SD – standard deviation.

- FIGURE 5 ABOUT HERE

Figure 5: M9. Root mean squares summaries of the weights in Fig. 4 indicating the relative importance of different kinds of input features during classification.

Classification based on cortical grey matter segmentations

The average accuracy from the 500 bootstraps was 71.2%, which means that, on average, SVMs correctly predicted CDMS in 71.2% of placebo-treated individuals, with a sensitivity of 64% (converters correctly identified) and specificity of 78.3% (non-converters correctly identified). The accuracy range of the 500 bootstraps was 54-84% with a 95% confidence interval (CI) of 70.7-71.6%. Middle and medial frontal, superior and middle temporal, anterior and posterior cingulate, middle temporal, fusiform, middle occipital, and insular regions contributed most to the classification accuracy as shown here for the matched placebo groups (Figure 6).

- FIGURE 6 ABOUT HERE

Figure 6: Weight vector maps showing the most discriminating brain regions between placebo groups for the matched placebo groups (top 10 %; accuracy 66%). (A, B) Regions that contributed most to classification accuracy in the matched placebo groups are shown in red, in axial, coronal and sagittal views (z = [-28, -18, -8, 2, 10, 22, 31]). (C) Projection of each subject onto the weight vector; with positive patterns (blue circles) discriminating for conv+, and negative patterns (red crosses) discriminating for conv-
CDMS could not be predicted based on cortical segmentations of the treated patients (total accuracy 48%; sens.: 42.9%, spec.: 53.1; p=0.6)

Discussion

Our study aims to determine whether multivariate neuroanatomical pattern classification facilitates predicting conversion to CDMS in individuals with a CIS based on structural MRI features and clinical characteristics in a two-year follow-up. We used geometric measures of individual lesions computed from MRI data in combination with clinical information from patients with CIS to predict conversion to CDMS. We included a large set of quantitative measures on individual lesions to perform classification into one of two groups, converters and non-converters, based on SVMs. In addition to providing an accurate classifier, we aimed to determine which types of features are most important for successful classification.

In the IFN-b group the best performance was achieved with non-linear classification and model M9 utilizing demographic and clinical covariates, grey matter volume, lesion count as well as total, mean and standard deviation measures of lesion geometrics. In the placebo group linear classification using cortical grey matter segmentations yielded the highest accuracy. It should be noted that using prediction accuracies directly for model comparison is not suitable, since differences could arise by chance. The focus should therefore lie on models that reached statistically significant accuracies.

The classification accuracies were only moderate (not exceeding 70%) which limits the applicability of the model for predicting the onset of disease in the brain scans of CIS patients. In general, this application is challenging for classification, as the differences observed in CIS are likely to be much more subtle then those observed in established MS. Not all CIS patients develop MS, and in those who do, disability is highly variable(Fisniku et al. 2008).

Furthermore, it cannot be excluded that some of the patients in the non-converter group may still develop MS in the long-term(Wottschel et al. 2015). Our results, however, are consistent with other studies on prediction of disease onset in MS(Wottschel et al. 2015) and Alzheimer's disease(Young et al. 2013) which reported similar classification accuracies.

Classification based on feature combinations

Consistent with the results from earlier studies(Wottschel et al. 2015)[55], the demographic attributes age and sex as well as the EDSS score carry a considerable amount of information about the disease and its progression (see Fig 5). When considering the full set of all geometric measures, each summarised in a ROI specific value (M10 in Table 2), a large number of these features will contain similar or even redundant information. Consider, for example, the case of a patient having only one lesion, which, in the case of CIS patients, is not uncommon.

Apart from the fact that the measures over 12 out of 13 ROIs will be zero, summary statistics such as mean and maximum lesion volume will be identical. Thus, feature selection or feature reduction becomes a necessity when optimising the classification procedure. A PCA transformation of input features reduces some of this redundancy and, in principle, SVMs can cope with a certain level of inter-dependency between features.

Nevertheless, classification outcomes are still likely to be affected by a lack of distinct information across the feature set. Therefore in this study, a partition of geometric features according to different brain regions turned out to be useful only to some extent.

However, it should be noted that thisparticular data set is not well suited to splitting into ROIs because most CIS patients have relatively few or even only one lesion.[a]

Ideally, as classification outcome varies depending on the combination of features used as input, one would like to find the most informative subset among all available features. This is hindered by the fact that an exhaustive combinatorial search across all features is computationally infeasible; and it would lead to overfitting by selecting a feature set that is specifically suited for the available data set but is unlikely to generalise well to new data. As a way out we considered pre-specified subsets of possible feature combinations as presented in Table 2.

Principally, the examination of weights of individual features for the support vectors across different models can help inform which kinds of features are driving the classification procedure. A comparison of the magnitude of SVM weights provides a qualitative assessment of relative importance of different input features. With regard to our set of geometry-based features, a comparison across different (linear) classifier weights indicated that the EP characteristic is often more significant than a simple lesion count. Among clinical and demographic covariates, age seemed to be the single best predictor with younger patients showing a higher probability to convert to CDMS (cf. negative weight on age in Figures 3 and 4).

[a] With other data sets that comprise of larger numbers of lesions per patients, we found that the segmentation into white matter track regions can substantially increase prediction accuracies compared to those feature sets that rely on whole brain summaries only.

When comparing placebo and IFN-b groups, lesion geometry played a bigger role in the classifier for the IFN-b group. EDSS and sex were more important for the placebo cohort, less so for the IFN-b group. Both features carried predominantly positive weights in all SVM models, indicating that higher EDSS at baseline as well as being female are predictors for conversion. Additionally, we found that the grey matter-to-whole-brain volume ratio in most cases was more informative than either EDSS or sex.

Active plaques typically are associated with Gd- enhancement on MRI and most likely represent the pathological substrate of the attacks in MS(Filippi et al. 2012; Popescu et al. 2013). Therefore, we concentrated here on T1-Gd imaging data. However, a comparison of weights from SVMs trained on T2 or T1 black hole MRI data showed very small variation, especially with respect to the sign of individual weights, across different modalities within each cohort (placebo or IFN-b), and much larger differences between the two cohorts.

Classification based on cortical grey matter segmentations

Using cortical grey matter segmentations, we show that linear SVMs moderately predicted individual conversion/non-conversion to CDMS in 71.2% of placebo-treated CIS patients. Although not exceptionally high, it is worth noting that the accuracy represents the average of 78% specificity and 64% sensitivity. The latter of which might be explained by the possibility that some of the CDMS patients categorized as converters do not show cortical pathology. Notably, patients experiencing a rapid transformation within 1 year show cortical atrophy in contrast to patients with slower disease

progression (Perez-Miralles et al. 2013). We cannot exclude that those converters who were wrongly classified as non-converters are patients with a slower disease progression. In a 2-year follow-up the non-converters are likely to still develop CDMS, some of them maybe even shortly after the last follow-up. Structural grey matter changes, i.e. both, cortical lesions and grey matter atrophy are considered as promising characteristics to track the conversion from CIS to CDMS [20], (Calabrese et al. 2010), (Tintore et al. 2008), (Kelly et al. 1993).

Some studies, however, failed to find a significant cortical atrophy in CIS compared to NC(Dalton et al. 2004; Ceccarelli et al. 2008; Ramasamy et al. 2009) while others have observed a significant cortical atrophy only in CIS(Bergsland et al. 2012) having a dissemination in space (DIS) of the lesions or only in selected brain regions, such as the hippocampus and the deep grey matter(Sastre-Garriga et al. 2005). Multivariate analysis of cortical thickness in patients with CIS who convert early to MS identified atrophy of superior frontal gyrus, thalamus, putamen and cerebellum as independent predictors of conversion to MS[20]. CIS with atrophy of such areas had a double risk of conversion.

These heterogeneous and partly contradictory findings may be explained e.g. by the clinical and paraclinical heterogeneity of patients with CIS, who may have quite different WM lesion load, and, in some cases, will never develop definite MS(Calabrese et al. 2011).

Previously, it was shown that regional GM atrophy is relevant in patients with CIS who convert early to MS(Calabrese et al. 2011; Raz et al. 2010), but about 20 % of CIS patients do not convert to MS after two decades (Fisniku et al. 2008). An accuracy of 70% in our linear SVM analysis reveals that cortical GM patterns played a role for

discrimination of converters and non-converters in the placebo-treated patients. In contrast, converters and non-converters could not be discriminated in the IFN-b groups. In our study, group labeling was based on conversion/non-conversion to CDMS at two year follow-up. However, at follow-up, parts of those patients which would have been expected to convert under placebo changed to non-converters or better “responders” under therapy. The latter might explain why we could not discriminate the treated group accurately using linear SVM.

Limitations

Our study further supports the idea that structural neuroimaging can inform prediction of CDMS. There are some limitations, however, that should be taken into account.

A crucial prerequisite for reliable image classification at the single-case level is the identification of morphometric criteria for distinguishing individuals. Beyond the impact of key demographic, clinical and lesion-associated MRI parameters on the 'natural risk' of CDMS(Polman et al. 2008), subclinical structural damage in the brain (i.e. changes in the normal-appearing brain tissue) might also occur early in disease. However, its functional consequences in CIS patients might be negligible(Kappos et al. 2007). Recently, mapping brain regions with diagnostic information using a “searchlight approach” and leave-one-out cross-validation to identify neuroanatomical patterns relevant for individual classification of patients and controls has been performed successfully in relapsing-remitting MS(Weygandt et al. 2011). Hotspots of MS associated tissue alterations highly informative about the clinical status have been identified in different normal-appearing

areas of the brain. We infer from these results, that the approach could be promising also for classification and prognosis of patients with CIS.

The smaller sample size of the non-converters in the SVM analysis using GM segmentations may have limited the converters vs non-converters classification accuracy, and thus the findings may have been influenced by the heterogeneity of the non-converters subgroup. To account for this, we controlled for the potential effects of covariates, such as age, gender, and scanner.

Generally, it is unclear whether supervised MRI-based pattern recognition can achieve the level of sensitivity and specificity needed in order to be integrated into clinical applications. In the future, feature selection methods and the use of an independent test data set may further increase classification accuracy. Furthermore, the shift from single predictive models to ensembles of classifiers may produce more generalizable diagnostic biomarkers by averaging the diagnostic decisions of numerous predictive models(Koutsouleris et al. 2010). Additionally, it will be interesting to investigate whether other para-clinical markers e.g. synthesis of oligoclonal bands(Tintore et al. 2008) and genetic factors(Kelly et al. 1993) can improve SVM-based classification accuracy.

Conclusions

The main potential of SVM-based classification is that it might be useful for predicting the clinical transition to MS at the individual level. Automatic pattern classification methods have been considered to promote a potentially accessible and objective way to

improve clinical decision making, and may present a measure of the risk of developing MSs in individuals with CIS if sufficiently accurate.

We demonstrated that MRI data of MS lesions contain more information about the disease than is currently utilized in clinical assessments. Both lesion geometry and grey matter based information can aid prediction of conversion to CDMS. Inclusion of other lesional or degenerative MRI features may in the future further improve classification accuracy.

Compliance with Ethical Standards:

Conflict of Interest:

Kerstin Bendfeldt declares that she has no conflict of interest.

Bernd Taschler declares that he has no conflict of interest

Laura Gaetano declares that she has no conflict of interest

Philip Madoerin declares that he has no conflict of interest.

Pascal Kuster declares that he has no conflict of interest

Nicole Mueller-Lenke declares that she has no conflict of interest

Michael Amann declares that he has no conflict of interest.

Hugo Vrenken declares that he has no conflict of interest

Viktor Wottschel declares that he has no conflict of interest

Frederik Barkhof declares that he has no conflict of interest.

Stefan Borgwardt declares that he has no conflict of interest

Stefan Klöppel declares that he has no conflict of interest

Eva-Maria Wicklein declares that she has no conflict of interest.

Ludwig Kappos declares that he has no conflict of interest

Gilles Edan declares that he has no conflict of interest

Mark S. Freedman declares that he has no conflict of interest.

Xavier Montalbán declares that he has no conflict of interest.

Hans-Peter Hartung declares that he has no conflict of interest.

Christoph Pohl † - no conflict of interest.

Rupert Sandbrink declares that he has no conflict of interest

Rupert Sandbrink declares that he has no conflict of interest.

Bernd Taschler declares that he has no conflict of interest

Till Sprenger has received research grants from the Swiss MS Society, Swiss National Research Foundation, EFIC-Grünenthal and Novartis Pharmaceuticals Switzerland. Till Sprengers current and/or previous employers have received compensation for consultation and speaking activities from Mitsubishi Pharma, Eli Lilly, Sanofi Genzyme, Novartis, ATI, Actelion, Electrocore, Biogen Idec, Teva and Allergan.

Ernst-Wilhelm Radue declares that he has no conflict of interest

Jens Wuerfel declares that he has no conflict of interest.

Thomas E. Nichols declares that he has no conflict of interest.

Ethical approval:

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent:

Informed consent was obtained from all individual participants included in the study.

References

Aban, I. B., Cutter, G. R., & Mavinga, N. (2008). Inferences and Power Analysis Concerning Two Negative Binomial Distributions with An Application to MRI Lesion Counts Data. *Computational statistics & data analysis, 53*(3), 820-833, doi:10.1016/j.csda.2008.07.034.

Arns, C. H., Knackstedt, M. A., Pinczewski, W. V., & Mecke, K. R. (2001). Euler-Poincaré characteristics of classes of disordered media. *Physical Review E, 63*(3), 031112.

Bakshi, R., Dandamudi, V. S., Neema, M., De, C., & Bermel, R. A. (2005). Measurement of brain and spinal cord atrophy by magnetic resonance imaging as a tool to monitor multiple sclerosis. [Research Support, Non-U.S. Gov't Review]. *J Neuroimaging, 15*(4 Suppl), 30S-45S, doi:10.1177/1051228405283901.

Barkhof, F., Polman, C. H., Radue, E.-W., Kappos, L., Freedman, M. S., Edan, G., et al. (2007). Magnetic Resonance Imaging Effects of Interferon Beta-1b in the BENEFIT Study: Integrated 2-Year Results. *Arch Neurol, 64*(9), 1292-1298, doi:10.1001/archneur.64.9.1292.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., et al. (2003). Voxel-based lesion-symptom mapping. *Nat Neurosci, 6*(5), 448-450, doi:10.1038/nn1050.

Bergsland, N., Horakova, D., Dwyer, M. G., Dolezal, O., Seidl, Z. K., Vaneckova, M., et al. (2012). Subcortical and Cortical Gray Matter Atrophy in a Large Sample of Patients with Clinically Isolated Syndrome and Early Relapsing-Remitting

Multiple Sclerosis. *American Journal of Neuroradiology*, doi:10.3174/ajnr.A3086.

Calabrese, M., Filippi, M., & Gallo, P. (2010). Cortical lesions in multiple sclerosis. [Review]. *Nature reviews. Neurology, 6*(8), 438-444, doi:10.1038/nrneurol.2010.93.

Calabrese, M., Rinaldi, F., Mattisi, I., Bernardi, V., Favaretto, A., Perini, P., et al. (2011). The predictive value of gray matter atrophy in clinically isolated syndromes. *Neurology, 77*(3), 257-263, doi:10.1212/WNL.0b013e318220abd4.

Ceccarelli, A., Rocca, M. A., Pagani, E., Colombo, B., Martinelli, V., Comi, G., et al. (2008). A voxel-based morphometry study of grey matter loss in MS patients with different clinical phenotypes. *NeuroImage, 42*(1), 315-322.

Dalton, C. M., Chard, D. T., Davies, G. R., Miszkiel, K. A., Altmann, D. R., Fernando, K., et al. (2004). Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. *Brain, 127*(5), 1101-1107, doi:10.1093/brain/awh126.

Filippi, M., Rocca, M. A., Barkhof, F., Bruck, W., Chen, J. T., Comi, G., et al. (2012). Association between pathological and MRI findings in multiple sclerosis. [Comparative Study

Review]. *Lancet neurology, 11*(4), 349-360, doi:10.1016/S1474-4422(12)70003-0.

Filli, L., Hofstetter, L., Kuster, P., Traud, S., Mueller-Lenke, N., Naegelin, Y., et al. (2012). Spatiotemporal distribution of white matter lesions in relapsing-remitting and secondary progressive multiple sclerosis. *Multiple Sclerosis, 18*(11), 1577-1584, doi:10.1177/1352458512442756.

Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., et al. (2008). Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain, 131*(3), 808-817, doi:10.1093/brain/awm329.

Ge, T., Muller-Lenke, N., Bendfeldt, K., Nichols, T. E., & Johnson, T. D. (2014). Analysis of Multiple Sclerosis Lesions Via Spatially Varying Coefficients. *Ann Appl Stat, 8*(2), 1095-1118.

Kappos, L., Freedman, M. S., Polman, C. H., Edan, G., Hartung, H. P., Miller, D. H., et al. (2007). Effect of early versus delayed interferon beta-1b treatment on disability after a first clinical event suggestive of multiple sclerosis: a 3-year follow-up analysis of the BENEFIT study. *Lancet, 370*(9585), 389-397.

Kappos, L., Polman, C. H., Freedman, M. S., Edan, G., Hartung, H. P., Miller, D. H., et al. (2006). Treatment with interferon beta-1b delays conversion to clinically definite and McDonald MS in patients with clinically isolated syndromes. *Neurology, 67*(7), 1242-1249.

Kelly, M. A., Cavan, D. A., Penny, M. A., Mijovic, C. H., Jenkins, D., Morrissey, S., et al. (1993). The influence of HLA-DR and-DQ alleles on progression to multiple sclerosis following a clinically isolated syndrome. *Human immunology, 37*(3), 185-191.

Koutsouleris, N., Patschurek-Kliche, K., Scheuerecker, J., Decker, P., Bottlender, R., Schmitt, G., et al. (2010). Neuroanatomical correlates of executive dysfunction in the at-risk mental state for psychosis. [Research Support, Non-U.S. Gov't]. *Schizophrenia research, 123*(2-3), 160-174, doi:10.1016/j.schres.2010.08.026.

Lang, C., Ohser, J., & Hilfer, R. (2001). On the analysis of spatial binary images. *Journal of microscopy, 203*(3), 303-313.

Legland, D., Kiêu, K., & Devaux, M.-F. (2011). Computation of Minkowski measures on 2D and 3D binary images. *Image Analysis & Stereology, 26*(2), 83-92.

Locatelli, L., Zivadinov, R., Grop, A., & Zorzon, M. (2004). Frontal parenchymal atrophy measures in multiple sclerosis. *Mult Scler, 10*(5), 562-568.

Lovblad, K. O., Anzalone, N., Dorfler, A., Essig, M., Hurwitz, B., Kappos, L., et al. (2010). MR imaging in multiple sclerosis: review and recommendations for current practice. [Research Support, Non-U.S. Gov't

Review]. *AJNR. American journal of neuroradiology, 31*(6), 983-989, doi:10.3174/ajnr.A1906.

MacKay Altman, R., Petkau, A. J., Vrecko, D., & Smith, A. (2012). A longitudinal model for magnetic resonance imaging lesion count data in multiple sclerosis patients. [Research Support, Non-U.S. Gov't]. *Statistics in medicine, 31*(5), 449-469, doi:10.1002/sim.4394.

Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. [Research Support, Non-U.S. Gov't

Review]. *Neuroscience and biobehavioral reviews, 36*(4), 1140-1152, doi:10.1016/j.neubiorev.2012.01.004.

Perez-Miralles, F., Sastre-Garriga, J., Tintore, M., Arrambide, G., Nos, C., Perkal, H., et al. (2013). Clinical impact of early brain atrophy in clinically isolated syndromes. *Mult Scler, 19*(14), 1878-1886, doi:10.1177/1352458513488231.

Polman, C., Kappos, L., Freedman, M., Edan, G., Hartung, H. P., Miller, D., et al. (2008). Subgroups of the BENEFIT study: Risk of developing MS and treatment effect of interferon beta-1b. *Journal of Neurology, 255*(4), 480-487.

Popescu, B. F. G., Pirko, I., & Lucchinetti, C. F. (2013). Pathology of multiple sclerosis: where do we stand? *CONTINUUM: Lifelong Learning in Neurology, 19*(4, Multiple Sclerosis), 901-921.

Ramasamy, D. P., Benedict, R. H. B., Cox, J. L., Fritz, D., Abdelrahman, N., Hussein, S., et al. (2009). Extent of cerebellum, subcortical and cortical atrophy in patients with MS: A case-control study. *Journal of the Neurological Sciences, 282*(1-2), 47-54.

Raz, E., Cercignani, M., Sbardella, E., Totaro, P., Pozzilli, C., Bozzali, M., et al. (2010). Gray- and White-Matter Changes 1 Year after First Clinical Episode of Multiple Sclerosis: MR Imaging. *Radiology, 257*(2), 448-454, doi:10.1148/radiol.10100626.

Richert, N. D., Howard, T., Frank, J. A., Stone, R., Ostuni, J., Ohayon, J., et al. (2006). Relationship between inflammatory lesions and cerebral atrophy in multiple sclerosis. [Research Support, N.I.H., Intramural]. *Neurology, 66*(4), 551-556, doi:10.1212/01.wnl.0000197982.78063.06.

Sastre-Garriga, J., Ingle, G. T., Chard, D. T., Cercignani, M., Ramio-Torrenta, L., Miller, D. H., et al. (2005). Grey and white matter volume changes in early primary progressive multiple sclerosis: a longitudinal study. *Brain, 128*(6), 1454-1460, doi:10.1093/brain/awh498.

Tintore, M., Rovira, A., Rio, J., Tur, C., Pelayo, R., Nos, C., et al. (2008). Do oligoclonal bands add information to MRI in first attacks of multiple sclerosis? *Neurology, 70*(13 Part 2), 1079-1083.

Wei, X., Guttmann, C. R., Warfield, S. K., Eliasziw, M., & Mitchell, J. R. (2004). Has your patient's multiple sclerosis lesion burden or brain atrophy actually changed? [Research Support, Non-U.S. Gov't

Research Support, U.S. Gov't, P.H.S.]. *Mult Scler, 10*(4), 402-406.

Weygandt, M., Hackmack, K., Pfüller, C., Bellmann–Strobl, J., Paul, F., Zipp, F., et al. (2011). MRI Pattern Recognition in Multiple Sclerosis Normal-Appearing Brain Areas. *PloS one, 6*(6), e21138, doi:10.1371/journal.pone.0021138.

Wottschel, V., Alexander, D. C., Kwok, P. P., Chard, D. T., Stromillo, M. L., De Stefano, N., et al. (2015). Predicting outcome in clinically isolated syndrome using machine learning. *Neuroimage Clin, 7*, 281-287, doi:10.1016/j.nicl.2014.11.021.

Young, J., Modat, M., Cardoso, M. J., Mendelson, A., Cash, D., Ourselin, S., et al. (2013). Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *Neuroimage Clin, 2*, 735-745, doi:10.1016/j.nicl.2013.05.004.

Zivadinov, R., Grop, A., Sharma, J., Bratina, A., Tjoa, C. W., Dwyer, M., et al. (2005). Reproducibility and accuracy of quantitative magnetic resonance imaging techniques of whole-brain atrophy measurement in multiple sclerosis. *J Neuroimaging, 15*(1), 27-36, doi:10.1177/1051228404271010.

Legends

Figure 1. Preprocessing pipeline.

Figure 2: Preprocessing pipeline for geometry-based lesion measures and subsequent classification with nonlinear SVMs.

Figure 3: M4. Individual weights from a linear SVM trained on features including demographic and clinical covariates, grey matter volume, lesion count and total lesion load; based on T1-Gd MRI data of IFN-b treated patients (converters: 50, non-converters: 49). Positive (negative) weights indicate that a larger feature value will drive prediction towards conversion (non-conversion). For sex, being female (male) corresponds to the positive (negative) axis. Weights are scaled relatively to the largest individual weight, which is set equal to one.

Figure 4: M9. Individual weights from a linear SVM trained on demographic, clinical, grey matter volume and geometry based features derived from T1-Gd MRI data of IFN-b treated patients (converters: 50, non-converters: 49). Positive (negative) weights indicate that a larger feature value will drive prediction towards conversion (non-conversion). For sex, being female (male) corresponds to the positive (negative) axis. Weights are scaled relatively to the largest individual weight, which is set equal to one.
Abbreviations: EDSS - expanded disability status scale, EPC - Euler-Poincare characteristic, GM - grey matter, SD – standard deviation

Figure 5: M9. Root mean squares summaries of the weights in Fig. 4 indicating the relative importance of different kinds of input features during classification.

Figure 6: Weight vector maps showing the most discriminating brain regions between placebo groups for the matched placebo groups (top 10 %; accuracy 66%). (A, B) Regions that contributed most to classification accuracy in the matched placebo groups are shown in red, in axial, coronal and sagittal views (z = (-28, -18, -8, 2, 10, 22, 31)). (C) Projection of each subject onto the weight vector; with positive patterns (blue circles) discriminating for conv+, and negative patterns (red crosses) discriminating for conv- CDMS could not be predicted based on cortical segmentations of the treated patients (total accuracy 48%; sens.: 42.9%, spec.: 53.1; p=0.6).