Brain Imaging and Behavior
https://doi.org/10.1007/s11682-018-9942-9

ORIGINAL RESEARCH

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

Kerstin Bendfeldt[1] · Bernd Taschler[2,3] · Laura Gaetano[1,4] · Philip Madoerin[1] · Pascal Kuster[1] · Nicole Mueller-Lenke[1] · Michael Amann[1,4] · Hugo Vrenken[5] · Viktor Wottschel[5] · Frederik Barkhof[5,6] · Stefan Borgwardt[1,7,8] · Stefan Klöppel[9] · Eva-Maria Wicklein[10] · Ludwig Kappos[4] · Gilles Edan[11] · Mark S. Freedman[12] · Xavier Montalbán[13] · Hans-Peter Hartung[14] · Christoph Pohl[10,15] · Rupert Sandbrink[10,14] · Till Sprenger[1,4] · Ernst-Wilhelm Radue[1] · Jens Wuerfel[1,15] · Thomas E. Nichols[3]

Abstract

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). To determine whether pattern classification using SVMs facilitates predicting conversion to clinically definite multiple sclerosis (CDMS) from clinically isolated syndrome (CIS). We used baseline MRI data from 364 patients with CIS, randomised to interferon beta-1b or placebo. Non-linear SVMs and 10-fold 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. We applied linear SVM analysis and leave-one-out cross-validation to subgroups of converters ($n = 25$) and non-converters ($n = 44$) based on cortical grey matter segmentations. Highest prediction accuracies of 70.4% ($p = 8e\text{-}5$) were reached with a combination of lesion-specific 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. Alongside standard demographic and clinical measures, both lesion geometry and grey matter based information can aid prediction of conversion to CDMS.

Bernd Taschler contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11682-018-9942-9) contains supplementary material, which is available to authorized users.

✉ Kerstin Bendfeldt
kerstin.bendfeldt@unibas.ch

[1] Medical Image Analysis Center (MIAC AG), Mittlere Str. 83, CH-4031 Basel, Switzerland

[2] German Center for Neurodegenerative Diseases, Bonn, Germany

[3] Department of Statistics, University of Warwick, Coventry, UK

[4] Department of Neurology, University Hospital Basel, Basel, Switzerland

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

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

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

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

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

[10] Bayer Pharma AG, Berlin, Germany

[11] CHU-Hopital Pontchaillou, Rennes, France

[12] University of Ottawa and Ottawa Hospital Research Institute, Ottawa, ON, Canada

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

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

[15] Charité University Medicine Berlin, Berlin, Germany

Published online: 28 August 2018

Keywords Clinically isolated syndrome · Multiple sclerosis · Support vector machine · MRI · Classification · Lesion geometry

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 (Filli et al. 2012; Locatelli et al. 2004) (Bates et al. 2003) (Wei et al. 2004) (Ge et al. 2014).

Clinically isolated syndromes (CIS), such as optic neuritis, brainstem or spinal cord syndromes are the first clinical presentations of MS in about 85% of cases (Scalfari et al. 2010; Confavreux and Vukusic 2006). Over time, most patients with a CIS, with additional clinically silent brain lesions on MRI that suggest disseminated CNS disease, develop relapsing-remitting MS (RRMS) and have a substantial risk for later progression of disability (Brex et al. 2002a; Tintore et al. 2006; Fisniku et al. 2008). The identification of patients with CIS who will develop clinically definite MS (CDMS) in the short-term is particularly relevant from a clinical, therapeutic (Calabrese et al. 2011), and economical point of view (Beer and Kesselring 1994).

A number of clinical features including laboratory investigations and MRI abnormalities have been associated with an increased risk for conversion to CDMS. Among these, MRI is the most informative marker (M. Tintore et al. 2005; M. Tintore et al. 2001; Morrissey et al. 1993; O'Riordan et al. 1998; Brex et al. 2002b) to monitor treatment efficacy. Several studies have confirmed the predictive value of baseline and follow-up MR imaging for conversion to CDMS, depending on the population, follow-up duration, and treatment intervention. To date, however, no reliable method exists to predict who will, and will not, develop MS amongst those with CIS.

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)).

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.

A combination of different MRI-based features has been shown to improve results as compared to using only one measurement (Wottschel et al. 2015). The study 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. A recent study on lesional geometry in MS used 3D-printing of lesions to qualitatively rate their shape and surface characteristics (Newton et al. 2017). Earlier work by Goldberg-Zimring et al. (Goldberg-Zimring et al. 2003) studied longitudinal changes in lesion geometry as captured by spherical harmonics. The study also found that lesional shape was more variable over time than size or volume. Although providing valuable insights, the methods presented in these two studies are not easily applicable to studies with more than a handful of patients. On a larger scale, Gourraud et al. (2013) (Gourraud et al. 2013) used the topological arrangement of lesional clusters (e.g. many small clusters vs. few big clusters) as an additional trait in a GWAS analysis; however, the study did not consider measures derived from individual lesions.

The present study demonstrates that a rich feature set incorporating multiple types of data and in combination with a classifier that is well suited for high-dimensional, multimodal settings, SVMs, provide an alternative approach. 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. By solving the classification problem in a higher dimensional auxiliary space, kernel-based SVMs can more easily adapt to the structure of the data (Schölkopf and Smola 2001). Most importantly, in addition to traditional demographic, clinical and lesional measures, we included aspects of lesion geometry derived from Minkowski functionals (Legland et al. 2007). These functionals provide quantitative measures that are closely related to standard geometric quantities such as volume, surface area and width of individual lesions.

Furthermore, there is increasing evidence that grey matter (GM) atrophy in CIS predicts conversion to CDMS (Calabrese et al. 2011; Raz et al. 2010; Jenkins et al. 2011). Therefore, we also include measures of GM volume and demonstrate how predictive performance of non-linear SVM can be. Finally, in a separate analysis, we used linear SVM to investigate whether SVMs can differentiate CIS from CDMS based on cortical grey matter characteristics.

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) and confirmed by a central committee. 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.

MRI acquisition and pre-processing

Baseline MR images were obtained using an oblique axial acquisition plane and gadolinium diethylenetriamine pentaacetic acid (0.1 mmol/kg of body weight) as contrast agent (T1c). The MR imaging protocol included a T1-weighted (T1w) spin echo after contrast administration with repetition times of 400 to 700 milliseconds and echo times of 5 to 25 milliseconds. A dual echo T2-weighted (T2w) spin echo with repetition times of 2000 to 3000 milliseconds and echo times of 20 to 40 milliseconds (first echo) and 60 to 100 milliseconds (second echo) was acquired as well. The field of view for all examinations was 25 cm with a 256 × 256–pixel matrix, resulting in a pixel size of roughly 1 × 1 mm2. Images were acquired in 2 interleaved sets with a 3-mm gap, resulting in whole-brain coverage with contiguous 3-mm-thick sections.

All MRI scans were performed with 0.1 mmol/kg gadolinium (Gd)-DTPA. The numbers and volumes of hyperintense lesions on T2-weighted images and Gd-enhancing lesions on T1-weighted images were centrally evaluated by the MRI Analysis Centre in Amsterdam, which was kept blinded to treatment allocation.

All of an individual patient's MRI was performed using the same machine. Assessment of the MRIs was performed centrally at the Image Analysis Centre at the Vrije Universiteit Medical Center. The quality assessment included a trial-specific standard operating procedure, training specifications (including a dummy-run procedure for site selection), and internal quality control procedures. During the trial each MRI was subjected to a routine quality control procedure, and additional MRIs were obtained when quality standards were not met.

The numbers and volumes of hyperintense lesions on T2-weighted images (T2 lesions) and Gd-enhancing (Gd lesions) and hypointense lesions on T1-weighted images (T1-hypointense lesions) were determined by evaluators who were blinded to treatment allocation but not to MRI order. Because all the MRIs were performed after Gd administration, T1-hypointense lesions were defined as the subset of new or enlarging lesions on T2-weighted images that were isointense to hypointense relative to gray matter on T1-weighted images. Lesions were identified by trained radiologists who marked the MS lesions on hard copy; their volumes were subsequently determined electronically by trained technicians using in-house–developed software (Show_Images version 3.6.3) based on local thresholding (Barkhof et al. 2007).

All MRI scans used in the present study were performed at 1.5 T and included transaxial contiguous 3-mm dual-echo images. T1-weighted images with sufficient image quality together with their corresponding lesion masks, as well as clinical data at two year follow-up, had to be comparable across scanners. 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.

Statistical analysis of demographic data

Statistical analysis was performed using IBM SPSS version 19.0. Group comparisons were performed using the Student's *t*-test, chi2-test or Wilcoxon rank sum test when appropriate. Values are reportet as mean (SD) or median (IQR). A statistical level of $p < 0.05$ was considered significant.

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)

	Non-linear SVM				Linear SVM			
	Placebo		IFN-b		Placebo		IFN-b	
	conv-	conv+	conv-	conv+	conv-	conv+	conv-	conv+
N	22	39	49	50	25	44	49	49
Mean age, years (SD)	30.6 (7.5)	29.2 (8.0)	31.4 (8.1) *	27.9 (6.8) *	34.1 (6.3) $^{+}$	29.0 (6.7) $^{+}$	31.1 (7.8)	29.8 (7.6)
m/f (ratio)	6/16 (1:2.7)	12/27 (1:2.2)	21/28 (1:1.3)	13/37 (1:2.8)	8/17 (1:2.1)	12/32 (1:2.7)	18/31 (1:1.72)	12/37 (1:3.08)
T_{CDMS}, median (IQR)	1780 (1805–857) $^{\#}$	249 (627–74)	1807 (1820–1794) $^{\#}$	432 (824–111)	1798 (1808–1497)	303 (709–138)	1808 (1822–1791)	423 (1027–204)
EDSS, median (IQR)	1.0 (2.0–0.0)	1.5 (2.0–1.0)	2.0 (2.0–1.0)	2.0 (2.5–1.0)	1.5 (2.0–1.0)	1.5 (2.0–1.0)	1.5 (2.0–1.0)	1.5 (2.0–1.0)
N relapse, median (IQR)	0 (0–0) $^{\#}$	1.0 (1.0–0.75) $^{\#}$	0 (0–0) $^{+}$	1.0 (1.0–0) $^{+}$	0 (0–0) *	1.0 (1.0–0.75) *	0 (0–0) $^{+}$	1.0 (1.0–0) $^{+}$
N T2 lesions at m24, median (IQR)	0 (1.25–0)	0.5 (3.0–0)	0 (0.75–0)	0 (0.75–0)	0 (1.25–0)	0.5 (3.0–0)	0 (0.75–0)	0 (0.75–0)
N Gd-lesions at m24, median (IQR)	0 (0–0)	0 (1.0–0)	0 (1–0)	0 (0.0–0)	0 (0–0)	0 (1.0–0)	0 (1–0)	0 (0–0)
Total (cortical^) GM volume [ml], mean (SD)	554.7 (177.3)	575.6 (234.2)	651.0 (217.1)	587.2 (265.4)	251.52 (102.37)^	284.0 (56.96)^	303.24 (27.88)^	293.64 (12.75)^

*/+/# indicate significant differences between groups ($p < 0.05$)

T treated, NT not treated, SD standard deviation, IQR interquartile range, EDSS expanded disability status scale, bs baseline, y1 year1, Total (*non-linear SVM*) or cortical (*linear SVM*) GM volume - as obtained from the SPM/VBM8 pipeline, PGMVC percentage gray matter volume change, T_{CDMS} time to CDMS (days)

Classification based on lesion-specific geometric measures and feature combinations using non-linear SVMs

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. 2007) 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 Fig. 1) can be found in the supplementary materials.

We also considered a parcellation of the whole brain into 13 regions of interest (ROIs) in order to add some amount of spatial information about lesion location to the feature set. (See supplementary materials for a description of the 13 ROIs.) Lesion dissemination in space is a standard requirement in the criteria for the diagnosis of CDMS (McDonald et al. 2001; C. H. Polman et al. 2005). Ideally, including spatial information will aide classification and prediction accuracy.

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 1.5 T 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 (Fig. 2).

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. Details on the

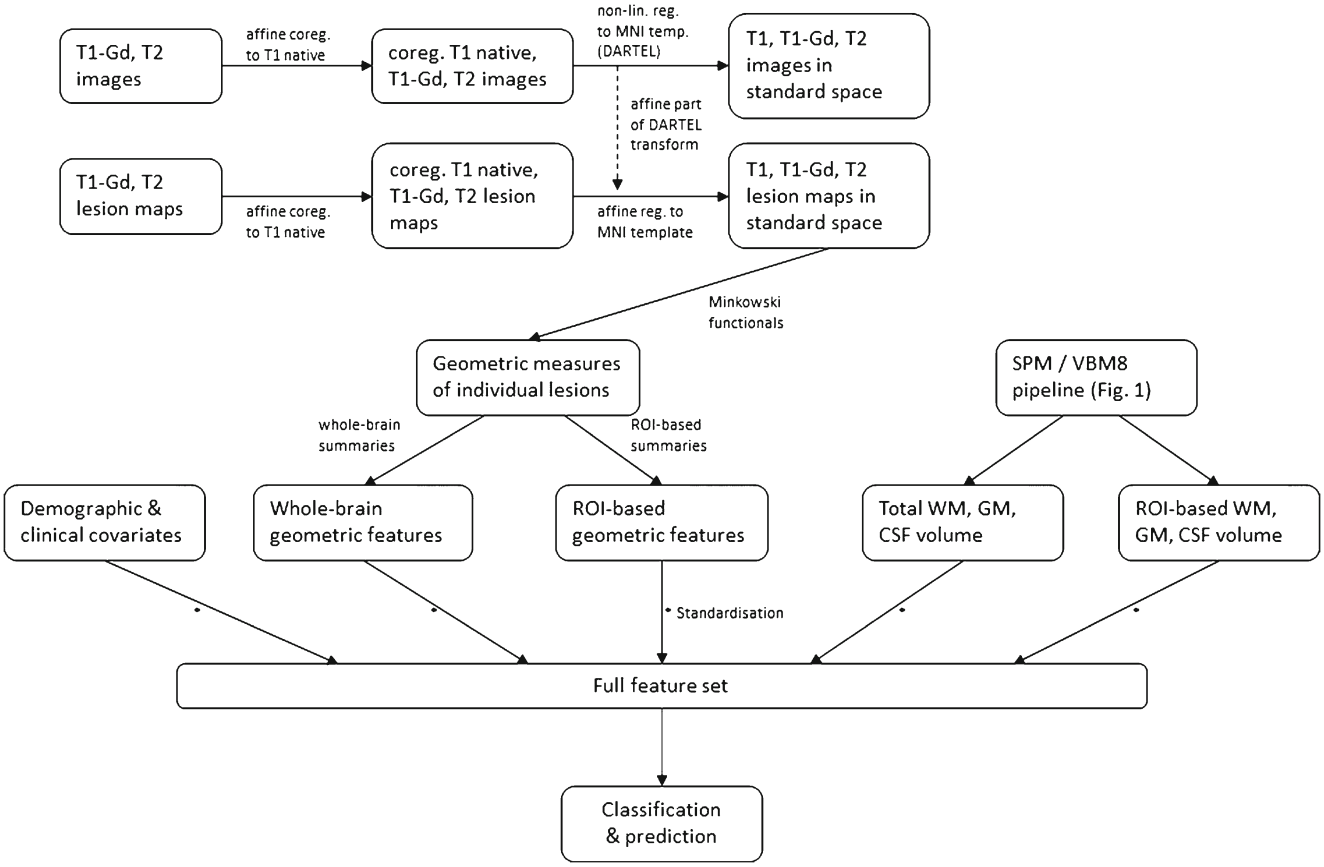

Fig. 1 Preprocessing pipeline for geometry-based lesion measures and subsequent classification with nonlinear SVMs

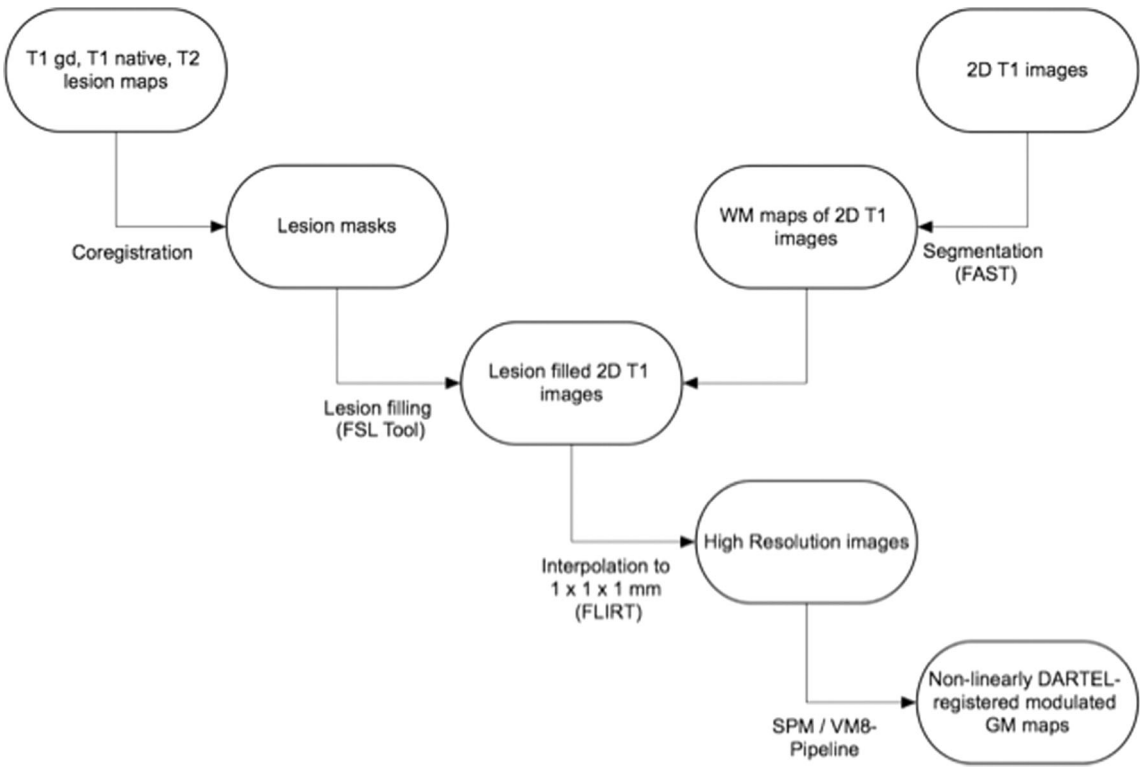

Fig. 2 Preprocessing pipeline

rationale for using linear SVMs in this setting, as well as advantages and limitations of SVMs in general can be found in the supplementary materials.

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.

Classification based on feature combinations

A summary of balanced prediction accuracies of kernel SVMs trained on different combinations of input features is given in Table 2 and visualised in Fig. 3. All values are based on nonlinear SVMs using an RBF kernel. Parameters were optimised through grid search. *P*-values are based on permutation tests of the class labels (1000 permutations per model). 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. In addition, the weights do not indicate whether any interaction between features are relevant during classification. 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.

Table 2 Balanced prediction accuracy of optimised kernel SVM based on different combinations of input features; including 95% confidence intervals and corresponding *p*-values

Model	Input features [resulting number of features]	Placebo group ($N = 61$, conv+: 39, conv-: 22)	IFN-b group ($N = 99$, conv+: 50, conv-: 49)
M1	demographic and clinical covariates (DC): age, sex, EDSS [3 features]	56.7% [48.5–64.9%] ($p = 0.053$) [77:63] *	55.5% [49.0–62.0%] ($p = 0.051$) [98:126] *
M2	DC, lesion count, lesion load (i.e. total lesion volume) [5 f.]	50.1% [37.6–62.6%] ($p = 0.537$)	57.6% [47.9–67.3%] ($p = 0.072$)
M3	DC, GM volume ratio (GM) [4 f.]	54.3% [41.8–66.8%] ($p = 0.261$)	58.6% [48.9–68.3%] ($p = 0.048$)
M4	DC, GM, lesion count, lesion load [6 f.]	60.9% [48.7–73.1%] ($p = 0.046$)	60.6 [51.0–70.2%] ($p = 0.020$)
M5	DC, GM, count, load, Euler-Poincare characteristic (EP) [7 f.]	56.6% [44.2–69.0%] ($p = 0.145$)	54.5% [44.7–64.3%] ($p = 0.193$)
M6	whole-brain summaries* of geometric measures (GEO-brain) [21 f.]	57.9% [45.5–70.3%] ($p = 0.126$)	58.5% [48.8–68.2%] (p = 0.046)
M7	DC, GM, GEO-brain [25 f.]	**67.6%** **[55.9–79.3%]** **($p = 0.0022$)**	64.6% [55.2–74.0%] (p = 0.0017)
M8	DC, GM by ROI, geometric measures by ROI (GEO-ROI) [276 f.]	51.2% [38.7–63.7%] ($p = 0.468$)	59.6% [49.9–69.3%] ($p = 0.035$)
M9	DC, GM, lesion count, subset** of GEO-brain [15 f.]	57.8% [45.4–70.2%] ($p = 0.133$)	**70.4%** **[61.4–79.4%]** **(p = 8e-5)**
M10	DC, GM by ROI, lesion count, subset** of GEO-ROI [159 f.]	50.5% [38.0–63.0%] ($p = 0.491$)	66.6% [57.3–75.9%] ($p = 0.0004$)

* 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

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

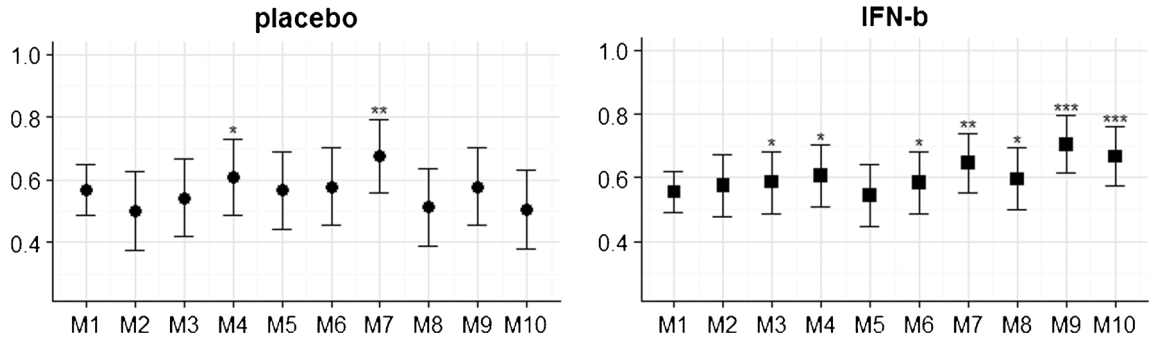

Fig. 3 Prediction accuracies of models as described in Table 2. Bars indicate 95% confidence interval. Significance levels based on permutation testing: * ($p < 0.05$), ** ($p < 0.005$), *** ($p < 0.0005$)

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 4 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% ($p = 8e$-

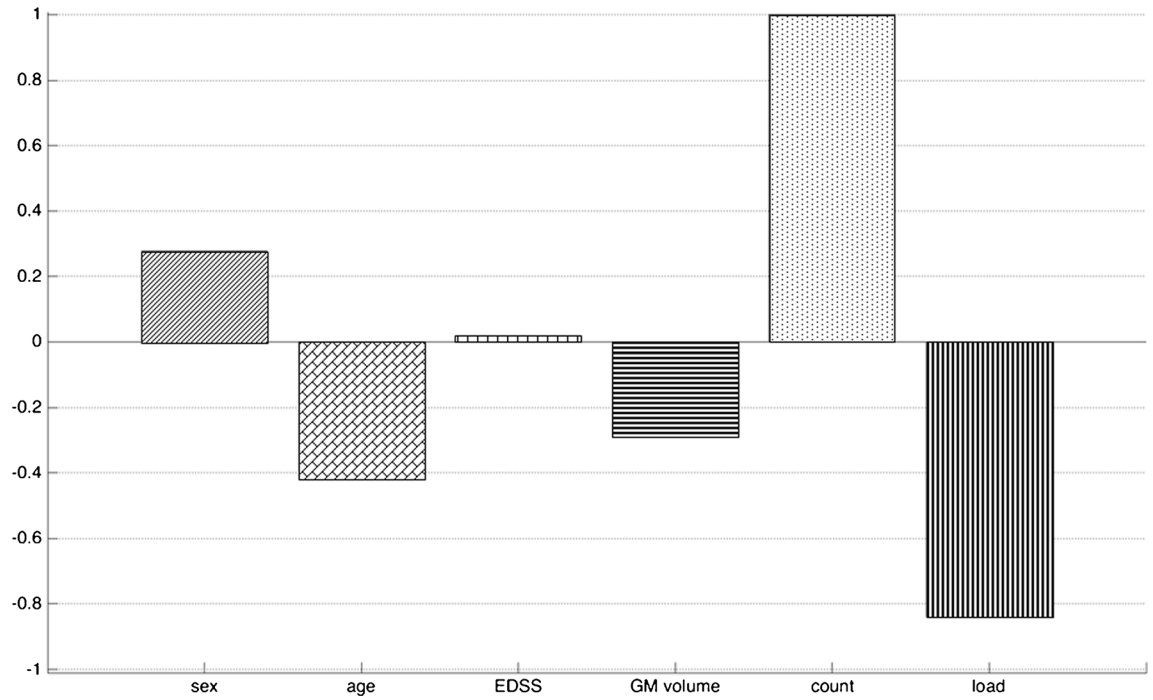

Fig. 4 M4. Individual weights from a linear SVM trained on features including demographic and clinical covariates, grey matter volume, lesion count and total lesion load; based on T1-Gd MRI data of IFN-b treated patients (converters: 50, non-converters: 49). Positive (negative) weights indicate that 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. Note that weights should only be interpreted qualitatively, not quantitatively

5). Note that splitting the same features by ROIs again reduced overall performance (cf. M10).

Figure 5 displays SVM weights for M9 of the corresponding linear classifier. The root mean square values (RMS, or quadratic means) of SVM weights shown in Fig. 6 (accuracy: 60.5%, $p = 0.019$) allow for a comparison between different kinds of features and their variability.

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 (accuracy 64%, $p < 0.0008$). 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.

Classification based on cortical grey matter segmentations

The average accuracy from the 500 bootstraps was 71.2% (95% confidence interval: 70.7–71.6%), 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). 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 (Fig. 7).

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, global 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 clinical praxis. 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. 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. 6). 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. However, it should be noted that imaging data of CIS patients may not necessarily be well suited for splitting into ROIs because most CIS patients have relatively few or even only one lesion.[1] This is due to the fact that the gain in spatial information about lesion locations may be offset by the inclusion of a large number of uninformative, zero-valued features for regions where no lesion is present.

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.

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

[1] 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.

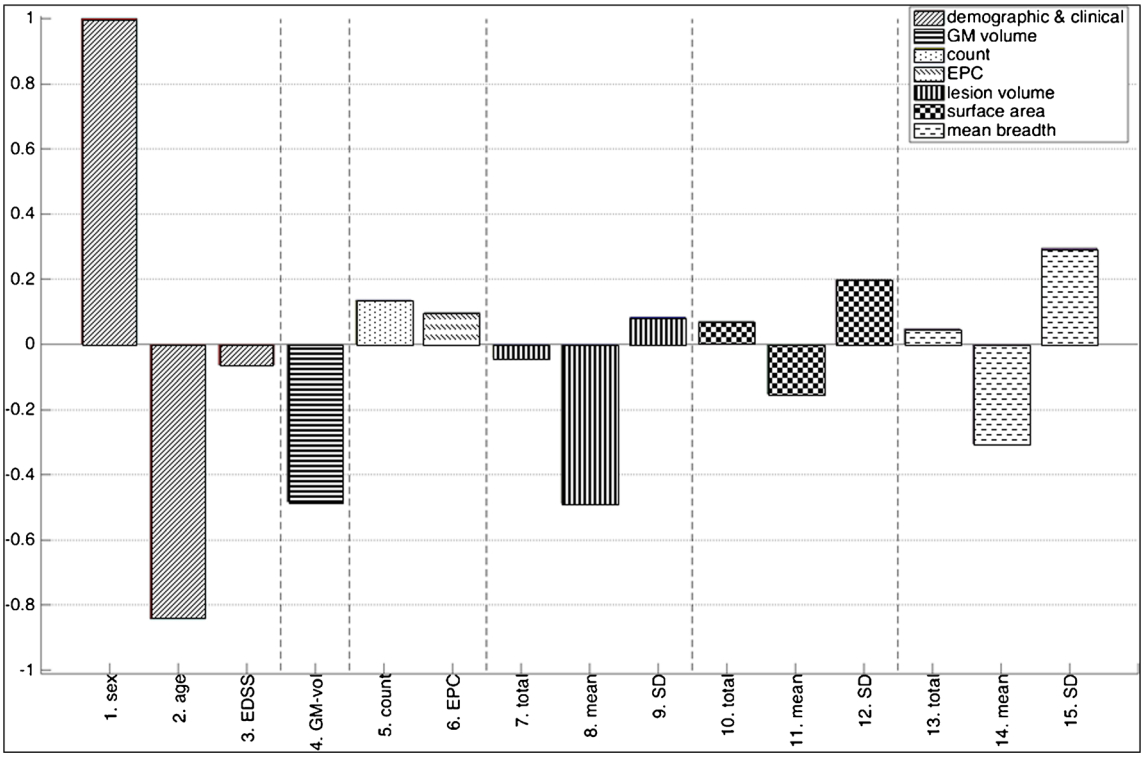

Fig. 5 M9. Individual weights from a linear SVM trained on demographic, clinical, grey matter volume and geometry based features derived from T1-Gd MRI data of IFN-b treated patients (converters: 50, non-converters: 49). Positive (negative) weights indicate that a larger feature value will drive prediction 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. Note that weights should only be interpreted qualitatively, not quantitatively. Abbreviations: EDSS - expanded disability status scale, EPC - Euler-Poincare characteristic, GM - grey matter, SD – standard deviation

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-

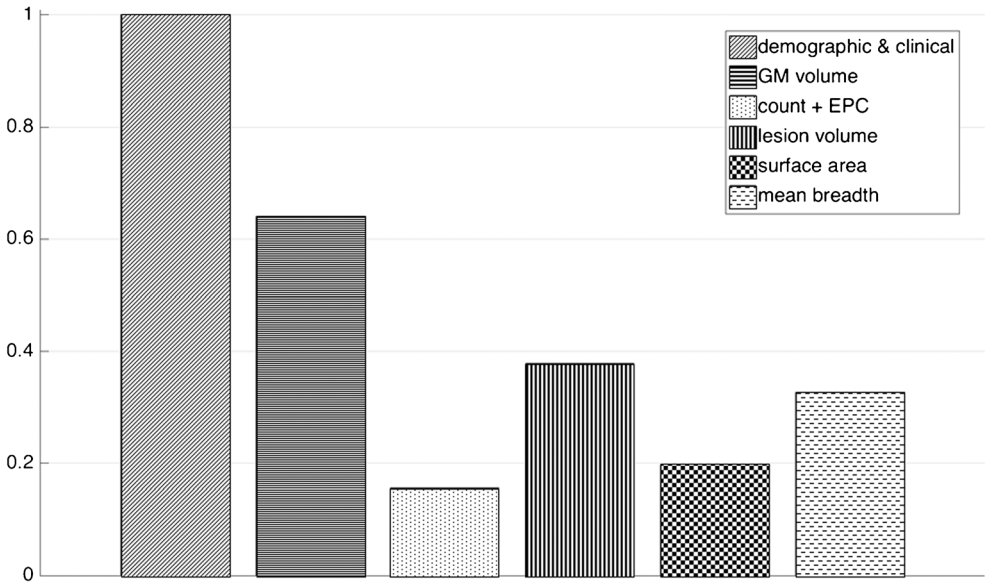

Fig. 6 M9. Root mean squares summaries of the weights in Fig. 4 indicating the relative importance of different kinds of input features during classification. Weights are scaled relatively to the largest individual weight, which is set equal to one. Note that weights should only be interpreted qualitatively, not quantitatively

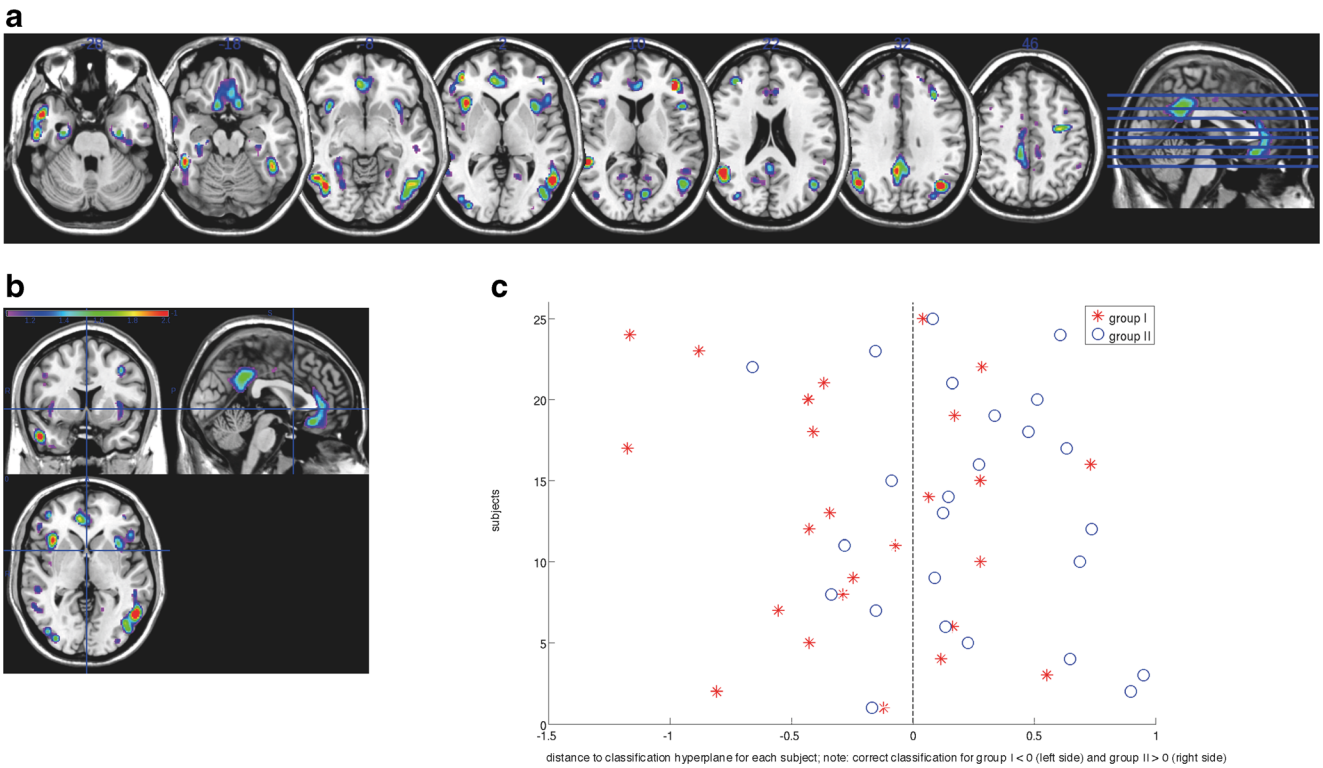

Fig. 7 Weight vector maps showing the most discriminating brain regions between placebo groups for the matched placebo groups (top 10%; accuracy 66%). (A, B) Regions that contributed most to classification accuracy in the matched placebo groups are shown in red, in axial, coronal and sagittal views (z = (−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; $p = 0.6$)

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 Figs. 4 and 5).

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.

It should be noted that a quantitative comparison of prediction accuracies and/or p-values across models is to be avoided. A statistically rigorous method for model comparison would need to be based on, for example, an appropriate permutation method, in order to assess the probability of finding spurious differences between any two given models. We are not aware of the existence of such a method in the context of a cross-validation framework, like the one we use in this work. Although beyond the scope of the current study, the need for the development of an appropriate permutation method for this setting could provide a basis for future work.

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), (M 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 (C. Polman et al. 2008), subclinical structural damage in the brain (i.E. changes in the normal-appearing brain tissue) might also occur early in the 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 which are 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 linear 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 larger, independent test data may further increase classification accuracy. Furthermore, the shift from single predictive models to ensembles of classifiers may produce more generalisable 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 (M Tintore et al. 2008) and genetic factors (Kelly et al. 1993) can improve SVM-based classification accuracy.

The IFN and the placebo group showed different features to have a role in determining conversion from CIS to CDMS. Early initiation of treatment at the stage of CIS has been demonstrated to delay conversion to CDMS, and to positively affect the evolution of clinical and MRI disease aspect (Kappos et al. 2007). Assessment of the balance between burden of treatment and effects on outcome in patients with CIS is clinically important because a considerable proportion of patients will develop CDMS and eventually progress to substantial disability in the following years (Kappos et al. 2009). However, about 15–20% might not develop CDMS, not even after 20 years (Fisniku et al. 2008). In the context of early intervention, multivariate pattern classification may constitute

the framework for MRI based discrimination of those individuals who will benefit from treatment from those who will do well without intervention.

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 MS 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. https://doi.org/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'tReview]. *Journal Neuroimaging, 15*(4 Suppl), 30S–45S. https://doi.org/10.1177/1051228405283901.

Barkhof, F., Polman, C. H., Radue, E.-W., Kappos, L., Freedman, M. S., Edan, G., Hartung, H. P., Miller, D. H., Montalbán, X., Poppe, P., de Vos, M., Lasri, F., Bauer, L., Dahms, S., Wagner, K., Pohl, C., & Sandbrink, R. (2007). Magnetic resonance imaging effects of interferon Beta-1b in the BENEFIT study: Integrated 2-year results. *Archives of Neurology, 64*(9), 1292–1298. https://doi.org/10.1001/archneur.64.9.1292.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion-symptom mapping. *Nature Neuroscience, 6*(5), 448–450. https://doi.org/10.1038/nn1050.

Beer, S., & Kesselring, J. (1994). High prevalence of multiple sclerosis in Switzerland. *Neuroepidemiology, 13*(1–2), 14–18. https://doi.org/10.1159/000110353.

Bergsland, N., Horakova, D., Dwyer, M. G., Dolezal, O., Seidl, Z. K., Vaneckova, M., Krasensky, J., Havrdova, E., & Zivadinov, R. (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, 33*, 1573–1578. https://doi.org/10.3174/ajnr.A3086.

Brex, P. A., Ciccarelli, O., O'Riordan, J. I., Sailer, M., Thompson, A. J., & Miller, D. H. (2002a). A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. [research support, non-U.S. Gov't]. *The New England Journal of Medicine, 346*(3), 158–164. https://doi.org/10.1056/NEJMoa011341.

Brex, P. A., Ciccarelli, O., O'Riordan, J. I., Sailer, M., Thompson, A. J., & Miller, D. H. (2002b). A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *New England Journal of Medicine, 346*(3), 158–164. https://doi.org/10.1056/NEJMoa011341.

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

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

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

Confavreux, C., & Vukusic, S. (2006). Natural history of multiple sclerosis: A unifying concept. *Brain, 129*(3), 606–616. https://doi.org/10.1093/brain/awl007.

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. https://doi.org/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. https://doi.org/10.1016/S1474-4422(12)70003-0.

Filli, L., Hofstetter, L., Kuster, P., Traud, S., Mueller-Lenke, N., Naegelin, Y., Kappos, L., Gass, A., Sprenger, T., Nichols, T. E., Vrenken, H., Barkhof, F., Polman, C., Radue, E. W., Borgwardt, S. J., & Bendfeldt, K. (2012). Spatiotemporal distribution of white matter lesions in relapsing-remitting and secondary progressive multiple sclerosis. *Multiple Sclerosis, 18*(11), 1577–1584. https://doi.org/10.1177/1352458512442756.

Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., Thompson, A. J., & Miller, D. H. (2008). Disability and T2 MRI lesions: A 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain, 131*(Pt 3), 808–817. https://doi.org/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. *Annals of Applied Statistics, 8*(2), 1095–1118.

Goldberg-Zimring, D., Achiron, A., Guttmann, C. R. G., & Azhari, H. (2003). Three-dimensional analysis of the geometry of individual multiple sclerosis lesions: Detection of shape changes over time using spherical harmonics. *Journal of Magnetic Resonance Imaging, 18*(3), 291–301. https://doi.org/10.1002/jmri.10365.

Gourraud, P. A., Sdika, M., Khankhanian, P., Henry, R. G., Beheshtian, A., Matthews, P. M., Hauser, S. L., Oksenberg, J. R., Pelletier, D., & Baranzini, S. E. (2013). A genome-wide association study of brain lesion distribution in multiple sclerosis. *Brain, 136*, 1012–1024. https://doi.org/10.1093/brain/aws363.

Jenkins, T. M., Ciccarelli, O., Atzori, M., Wheeler-Kingshott, C. A. M., Miller, D. H., Thompson, A. J., & Toosy, A. T. (2011). Early pericalcarine atrophy in acute optic neuritis is associated with conversion to multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry, 82*(9), 1017–1021. https://doi.org/10.1136/jnnp.2010.239715.

Kappos, L., Freedman, M. S., Polman, C. H., Edan, G., Hartung, H. P., Miller, D. H., Montalbán, X., Barkhof, F., Radü, E. W., Bauer, L., Dahms, S., Lanius, V., Pohl, C., & Sandbrink, R. (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., Freedman, M. S., Polman, C. H., Edan, G., Hartung, H. P., Miller, D. H., Montalbán, X., Barkhof, F., Radü, E. W., Metzig, C., Bauer, L., Lanius, V., Sandbrink, R., Pohl, C., & Benefit Study Group. (2009). Long-term effect of early treatment with interferon beta-1b after a first clinical event suggestive of multiple sclerosis: 5-year active treatment extension of the phase 3 BENEFIT trial. *Lancet Neurology, 8*(11), 987–997. https://doi.org/10.1016/S1474-4422(09)70237-6.

Kappos, L., Polman, C. H., Freedman, M. S., Edan, G., Hartung, H. P., Miller, D. H., Montalban, X., Barkhof, F., Bauer, L., Jakobs, P., Pohl, C., Sandbrink, R., & for the BENEFIT Study Group. (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., Miller, D. H., Barnett, A. H., & Francis, D. A. (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., Rujescu, D., Giegling, I., Gaser, C., Reiser, M., Möller, H. J., & Meisenzahl, E. M. (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. https://doi.org/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. (2007). 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. *Multiple Sclerosis, 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. https://doi.org/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. https://doi.org/10.1002/sim.4394.

McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H. P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., van den Noort, S., Weinshenker, B. Y., & Wolinsky, J. S. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology, 50*(1), 121–127.

Morrissey, S. P., Miller, D. H., Kendall, B. E., Kingsley, D. P. E., Kelly, M. A., Francis, D. A., et al. (1993). The significance of brain magnetic-resonance-imaging abnormalities at presentation with clinically isolated syndromes suggestive of multiple-sclerosis - a 5-year follow-up-study. *Brain, 116*, 135–146. https://doi.org/10.1093/brain/116.1.135.

Newton, B. D., Wright, K., Winkler, M. D., Bovis, F., Takahashi, M., Dimitrov, I. E., Sormani, M. P., Pinho, M. C., & Okuda, D. T. (2017). Three-dimensional shape and surface features distinguish multiple sclerosis lesions from nonspecific white matter disease. *Journal of Neuroimaging, 27*(6), 613–619. https://doi.org/10.1111/jon.12449.

O'Riordan, J. I., Thompson, A. J., Kingsley, D. P. E., MacManus, D. G., Kendall, B. E., Rudge, P., et al. (1998). The prognostic value of brain MRI in clinically isolated syndromes of the CNS - a 10-year follow-up. *Brain, 121*, 495–503. https://doi.org/10.1093/brain/121.3.495.

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. https://doi.org/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. *Multiple Sclerosis, 19*(14), 1878–1886. https://doi.org/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.

Polman, C. H., Reingold, S. C., Edan, G., Filippi, M., Hartung, H.-P., Kappos, L., Lublin, F. D., Metz, L. M., McFarland, H. F., O'Connor,

P. W., Sandberg-Wollheim, M., Thompson, A. J., Weinshenker, B. G., & Wolinsky, J. S. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the ldquoMcDonald Criteriardquo. *Annals of Neurology, 58*(6), 840–846.

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., Minagar, A., Dwyer, M. G., & Zivadinov, R. (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., & Pantano, P. (2010). Gray- and white-matter changes 1 year after first clinical episode of multiple sclerosis: MR imaging. *Radiology, 257*(2), 448–454. https://doi.org/10.1148/radiol.10100626.

Richert, N. D., Howard, T., Frank, J. A., Stone, R., Ostuni, J., Ohayon, J., Bash, C., & McFarland, H. F. (2006). Relationship between inflammatory lesions and cerebral atrophy in multiple sclerosis. [research support, N.I.H., intramural]. *Neurology, 66*(4), 551–556. https://doi.org/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. https://doi.org/10.1093/brain/awh498.

Scalfari, A., Neuhaus, A., Degenhardt, A., Rice, G. P., Muraro, P. A., Daumer, M., & Ebers, G. C. (2010). The natural history of multiple sclerosis, a geographically based study 10: Relapses and long-term disability. *Brain, 133*(7), 1914–1929. https://doi.org/10.1093/brain/awq118.

Schölkopf, B., Smola, A.J. (2001) Learning with Kernels. MIT Press.

Tintore, M., Rovira, A., Brieva, L., Grive, E., Jardi, R., Borras, C., et al. (2001). Isolated demyelinating syndromes: Comparison of CSF oligoclonal bands and different MR imaging criteria to predict conversion to CDMS. *Multiple Sclerosis, 7*(6), 359–363. https://doi.org/10.1177/135245850100700603.

Tintore, M., Rovira, A., Rio, J., Nos, C., Grive, E., Tellez, N., et al. (2005). Is optic neuritis more benign than other first attacks in multiple sclerosis? *Annals of Neurology, 57*(2), 210–215. https://doi.org/10.1002/ana.20363.

Tintore, M., Rovira, A., Rio, J., Nos, C., Grive, E., Tellez, N., et al. (2006). Baseline MRI predicts future attacks and disability in clinically isolated syndromes. [comparative study Research Support, Non-U.S. Gov't]. *Neurology, 67*(6), 968–972. https://doi.org/10.1212/01.wnl.0000237354.10144.ec.

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. https://doi.org/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. https://doi.org/10.1016/j.nicl.2014.11.021.

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

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