SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of Multiple Sclerosis

Viktor Wottschel , Declan T. Chard , Christian Enzinger , Massimo Filippi , Jette L. Frederiksen , Claudio Gasperini , Antonio Giorgio , Maria A. Rocca , Alex Rovira , Nicola De Stefano , Mar Tintoré , Daniel C. Alexander , Frederik Barkhof , Olga Ciccarelli , for the MAGNIMS study group and the EuroPOND consortium

PII: S2213-1582(19)30361-4
DOI: https://doi.org/10.1016/j.nicl.2019.102011
Reference: YNICL 102011

To appear in: *NeuroImage: Clinical*

Received date: 15 April 2019
Revised date: 6 September 2019
Accepted date: 17 September 2019

Please cite this article as: Viktor Wottschel , Declan T. Chard , Christian Enzinger , Massimo Filippi , Jette L. Frederiksen , Claudio Gasperini , Antonio Giorgio , Maria A. Rocca , Alex Rovira , Nicola De Stefano , Mar Tintoré , Daniel C. Alexander , Frederik Barkhof , Olga Ciccarelli , for the MAGNIMS study group and the EuroPOND consortium, SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of Multiple Sclerosis, *NeuroImage: Clinical* (2019), doi: https://doi.org/10.1016/j.nicl.2019.102011

Highlights:

- RFE-SVMs predict future outcome of CIS patients with conservative accuracy estimates between 64.9% and 88.1%
- Recursive feature selection improves classification performance compared to using all information
- Relevant features include regional WM lesion load and GM density, as well as the type of CIS onset.
- Cross-validation introduces positive bias on accuracy estimate

SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of Multiple Sclerosis

Viktor Wottschel[1,2], Declan T. Chard[2,3], Christian Enzinger[4], Massimo Filippi[5], Jette L. Frederiksen[6], Claudio Gasperini[7], Antonio Giorgio[8], Maria A. Rocca[5], Alex Rovira[9], Nicola De Stefano[8], Mar Tintoré[9], Daniel C. Alexander[10], Frederik Barkhof[1,2,3,11], Olga Ciccarelli[2,3] for the MAGNIMS study group and the EuroPOND consortium

[1] Dept. of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, Location VUmc, Amsterdam, The Netherlands

[2] Queen Square MS Centre, University College London, London, United Kingdom

[3] National Institute of Health Research (NIHR) University College London Hospitals Biomedical Research Centre

[4] Research Unit for Neuronal Repair and Plasticity, Dept. of Neurology, Medical University of Graz, Graz, Austria

[5] Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy

[6] Rigshospitalet-Glostrup and University of Copenhagen, Copenhagen, Denmark

[7] San Camillo-Forlanini Hospital, Rome, Italy

[8] University of Siena, Siena, Italy

[9] Hospital Vall d'Hebron, Barcelona, Spain

[10] Centre for Medical Image Computing, Dept. of Computer Science, University College London, London, United Kingdom

[11] Institute of Neurology and Healthcare Engineering, University College London, London, United Kingdom

Words (Abstract): 398

Words (Article body): 4722

No. References: 33

No. Figures: 5

No. Tables: 5

Corresponding author

Viktor Wottschel
Department of Radiology and Nuclear Medicine
Amsterdam University Medical Centers, Location VUmc
Postbus 7057
1007 MB Amsterdam
The Netherlands
v.wottschel@vumc.nl

Abstract

Machine learning classification is an attractive approach to automatically differentiate patients from healthy subjects, and predict future disease outcomes. A clinically isolated syndrome (CIS) is often the first presentation of multiple sclerosis (MS), but it is difficult at onset to predict who will have a second relapse and hence convert to clinically definite MS. In this study, we thus aimed to distinguish CIS converters from non-converters at onset of a CIS, using recursive feature elimination and weight averaging with support vector machines. We also sought to assess the influence of cohort size and cross-validation methods on the accuracy estimate of the classification.

We retrospectively collected 400 patients with CIS from six European MAGNIMS MS centres. Patients underwent brain MRI at onset of a CIS according to local standard-of-care protocols. The diagnosis of clinically definite MS at one-year follow-up was the standard against which the accuracy of the model was tested. For each patient, we derived MRI-based features, such as grey matter probability, white matter lesion load, cortical thickness, and volume of specific cortical and white matter regions. Features with little contribution to the classification model were removed iteratively through an interleaved sample bootstrapping and feature averaging approach. Classification of CIS outcome at one-year follow-up was performed with 2-fold, 5-fold, 10-fold and leave-one-out cross-validation for each centre cohort independently and in all patients together.

The estimated classification accuracy across centres ranged from 64.9% to 88.1% using 2-fold cross-validation and from 73% to 92.9% using leave-one-out cross-validation. The classification accuracy estimate was higher in single-centre, smaller data sets than in combinations of data sets, being the lowest when all patients were merged together.

Regional MRI features such as WM lesions, grey matter probability in the thalamus and the precuneus or cortical thickness in the cuneus and inferior temporal gyrus predicted the

occurrence of a second relapse in patients at onset of a CIS using support vector machines. The increased accuracy estimate of the classification achieved with smaller and single-centre samples may indicate a model bias (overfitting) when data points were limited, but also more homogeneous. We provide an overview of classifier performance from a range of cross-validation schemes to give insight into the variability across schemes. The proposed recursive feature elimination approach with weight averaging can be used both in single- and multi-centre data sets in order to bridge the gap between group-level comparisons and making predictions for individual patients.

Keywords:
Multiple sclerosis
Machine learning classification
Feature selection

1 Introduction

Multiple sclerosis (MS) is a disease of the central nervous system that is characterised by neuroinflammation, demyelination and neurodegeneration. The first clinical episode of MS is referred to as a clinically isolated syndrome (CIS). A majority of CIS patients (>80%) will eventually develop a second episode over a course of 20 years[1], which then defines clinically definite MS (CDMS). A shorter time to conversion from CIS to CDMS is associated with a faster disease progression and higher disability subsequently[1]. The number of lesions on the MRI scan at onset of CIS is a clinically highly relevant prognostic factor for the development of CDMS and disability[2].

Machine learning offers tools for learning how to distinguish two or more groups based on their features and subsequently assign new, previously unseen, cases to one of the groups. The idea of supervised learning is to identify common characteristics in the individual groups (i.e., patients with a known diagnosis or clinical outcome) that can be generalised to a larger population. This supervised classification has become increasingly popular in neuroimaging over the last decade with a few applications also in MS[3ï5]. However, only few studies have been performed on the prediction of conversion to CDMS in CIS patients[5ï7], and these have often been limited to one centre[5,6].

A common issue is the selection of relevant features to perform a classification. Some studies in MS and Alzheimerôs disease have used voxelwise grey-matter (GM) probability[4,8], which works well when patient groups can be distinguished based on their extent of (regional) brain atrophy. Other studies used hand-picked features that potentially provide predictive information[5,7]. In a previous single-centre study[5], we showed that support vector machine-based classification predicted clinical outcome in CIS patients with an accuracy score of

71.4% using leave-one-out cross validation. We found that a specific subset of features, mostly related to MS lesions, performed better than individual or all available features. However, as we note in [5] leave-one-out cross-validation may overestimate classification performance on unseen test data.

Here, we aimed to identify CIS patients developing CDMS within the first year of their symptoms, using data collected in six European centres. We introduce a recursive feature elimination scheme, based on weight averaging with support vector machines, in a large set of imaging measures, including GM probability, cortical thickness, T2 white matter lesion load, and volume of specific GM and white matter (WM) regions. These features can be easily and robustly extracted from MRI scans, and we investigated whether our model automatically identified informative features with respect to the classification task. We examined the influence of the cross-validation partitioning on the estimated classification accuracy by using 2-fold, 5-fold, 10-fold and leave-one-out cross-validation on all data sets to provide an overview of the bias introduced by the different schemes. The model was run in each centreôs cohort independently and then in combinations of data sets, including all patient data together in order compare different levels of heterogeneity in the data.

2 Methods

2.1 Data

This is a retrospective study performed on data obtained by six European centres, which are members of the MAGNIMS (Magnetic Resonance Imaging in Multiple Sclerosis, www.magnims.eu) network (Barcelona/Spain (B), Copenhagen/Denmark (C), Graz/Austria (G), London/UK (L), Milan/Italy (M) and Siena/Italy (S)). The total number of CIS patients included was 400, and 91 (22.8%) of them converted from CIS to CDMS within one year. All baseline scans were performed within 14 weeks (SD 7 weeks) of CIS onset. We do not have information on treatment in this retrospective cohort. A more detailed overview of patient characteristics is given in Table 1.

This project was approved locally by the ethics committees and patient consent was obtained prior to data collection.

The inclusion criteria were as follows: (1) Patients with a CIS were examined within three months from symptoms onset; (2) T1-weighted MRI sequences of the brain were obtained at onset of a CIS, using standard-of-care local protocols; (3) Demographic (age, sex) and clinical information (e.g. type of CIS) at baseline and the presence/absence of a second relapse at one year follow-up was available; (4) presence of T2-hyperintense WM brain lesions as outlined in each centre on PD/T2-weighted or FLAIR MRI by experienced researchers, resulting in binary lesion masks.

[Table 1]

2.2 Image processing

Due to the heterogeneity of the MRI data, we used derived measures such as GM probability or cortical thickness (CT) which we believe to be more robust to multi-centre variation compared to direct intensity information. To calculate the features used in the classification experiments, a comprehensive image processing pipeline was created as follows.

1. Bias field correction: all MRI scans were initially corrected for bias field inhomogeneities using the N4 algorithm[9].
2. Lesion filling: WM lesions can have intensities similar to GM on T1-weighted MRI, which can cause problems in registration and segmentation. To reduce this effect, we used a patch-based approach[10] to fill the lesion voxels with intensities similar to their neighbourhood.
3. Registration: lesion masks were created from PD/T2- or FLAIR-weighted images whereas most other image processing is performed in T1 space. Therefore, the PD/T2 or FLAIR MRI scans were affinely registered to T1 space using reg_aladin from the NiftyReg toolbox[11]. Lesion masks were subsequently resampled using the obtained transformation parameters.
4. Brain parcellation: we performed a fine-grained brain parcellation of all T1 scans using the GIF (geodesic information flows) algorithm[12]. This tool segments the brain into 143 ROIs based on the Neuromorphometrics atlas[13], of which most are cortical areas as shown in Figure 1.
5. Merging hemispheres: Measurements from the left and right hemisphere are highly correlated, which is undesirable for machine learning analyses[14]. Therefore, corresponding contralateral ROI values were averaged in order to reduce the noise in the data and reduce collinearity of features. (Please note that we show some results

with unmerged contralateral features in the supplementary material section ôUnmerged Hemispheresô.)

6. Grouping: ROIs were merged into nine larger areas according to their anatomical location. Most of these areas correspond to the anatomical brain lobes, and, therefore we refer to all of them as ôlobesô in the context of this study. These ôlobesô were limbic, insular, frontal, parietal, temporal, occipital, cerebellum, GM and WM. Deep grey matter is defined as thalamus, hippocampus, nucleus accumbens, amygdala, caudate nucleus, pallidum, putamen and basal ganglia.
7. Segmentation: In addition to the 143 ROIs, the GIF algorithm also provides a probabilistic segmentation of GM and WM, as well as binary masks of brain tissue and intracranial volume.
8. Cortical thickness: this was calculated using DiReCT, a registration-based algorithm[15]. It has been shown to have the same degree of reproducibility as the more commonly used Freesurfer method[16] but is faster once WM and GM probability maps are available.
9. ROI masking. We used the ROIs from steps 4 and 6 to calculate local information from GM probability maps, CT maps and lesion masks.

[Figure 1]

2.3 Feature definitions

Following the image processing, an extensive list of features has been defined on different ROI scales as follows.

1. Global features: these features describe whole-brain measures such as overall GM volume, WM volume and brain volume as a percentage of the intracranial volume. In addition, we added demographic and clinical measures such as age, sex, CIS type and EDSS.

Figures

Figure 1: Illustration of the Neuromorphometrics atlas used for brain parcellation in this study.

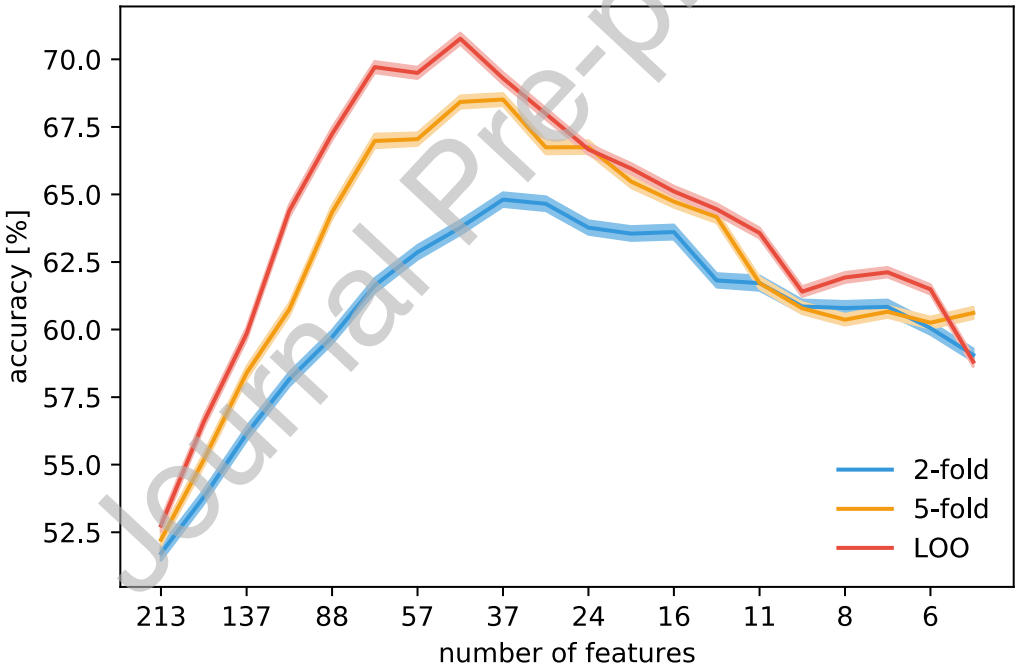

Figure 2: Accuracy estimates achieved at different iterations of the recursive feature elimination. The accuracy estimates increase with the first steps of the RFE, and the accuracy estimates generally increase with the number of folds. The shaded areas indicate 95% confidence intervals over 1000 bootstraps.

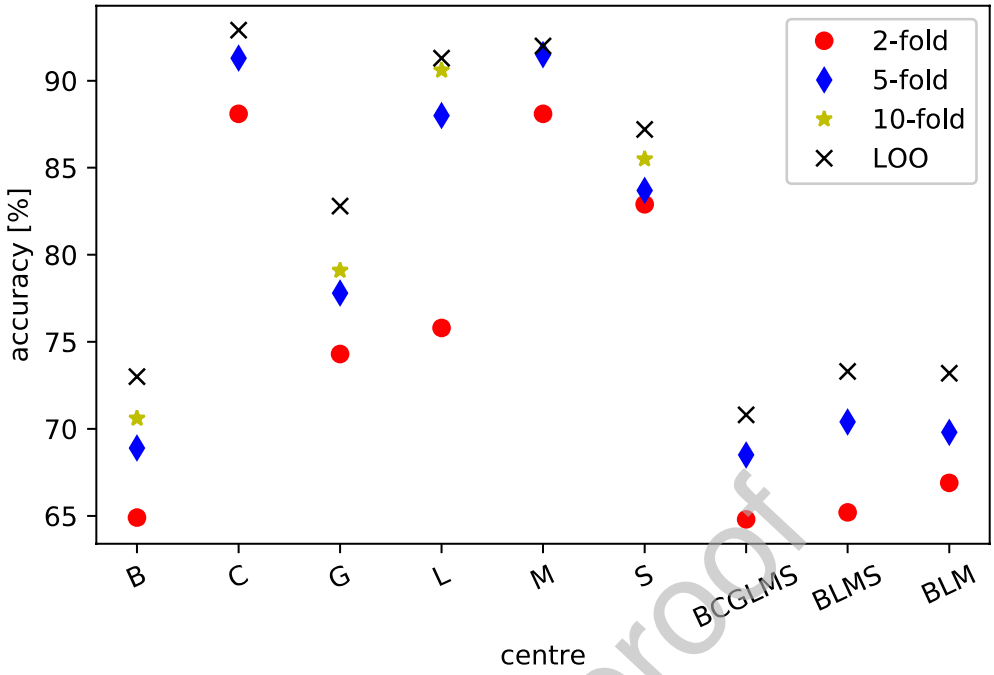

Figure 3: Accuracy estimates per centre or combination of centres for each validation method. Corresponding values for confidence intervals, sensitivity and specificity can be found in tables 2-5.

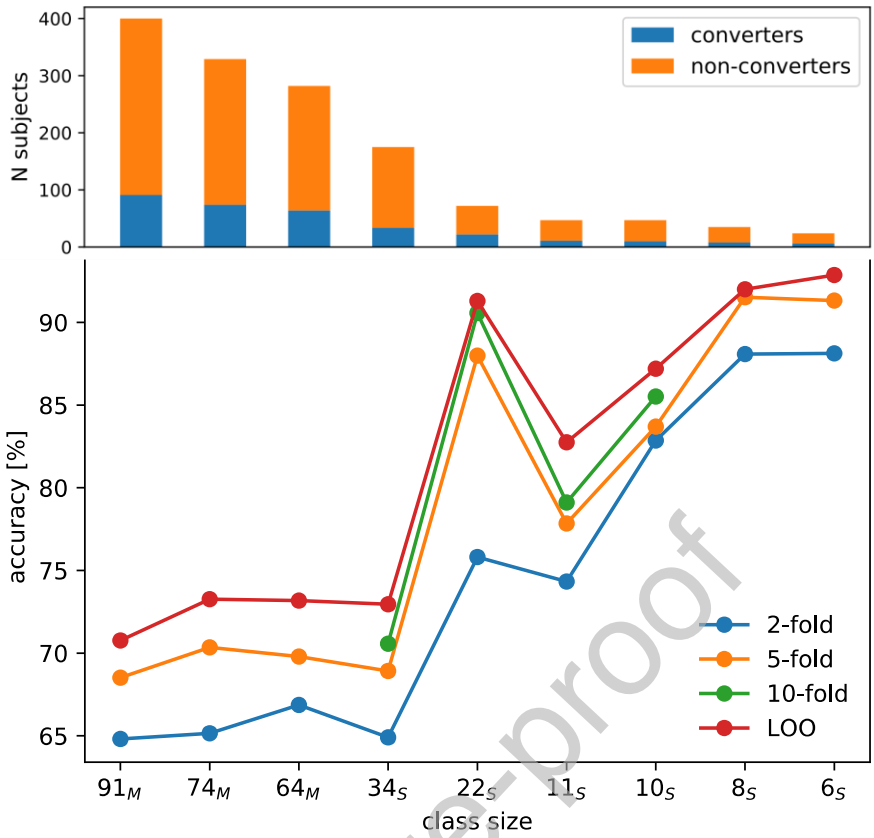

Figure 4: Top: bar chart of the proportion of converters in the cohort. Bottom: estimated classification accuracy relative to the size of the minority class. There is a general increase of estimated accuracy with a decrease in sample size. The subscript M and S indicate multi centre and single centre data sets respectively.

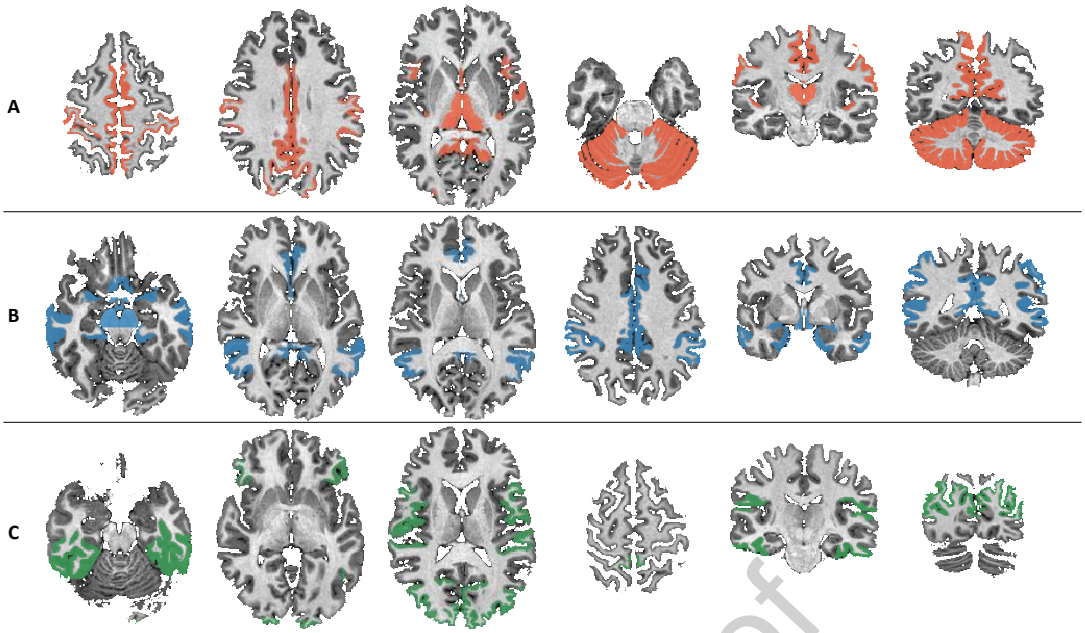

Figure5: Location of features relevant to the prediction of CIS conversion at 1-year follow-up. The colours represent A: GM probability (red), B: regional volume size (blue) and C: cortical thickness (green) respectively. Please note that white matter lesional load across the whole brain was also selected but is not shown here for clarity. Type of CIS onset was selected as the only non-imaging feature. A full list of features can be found in the supplementary material.