Prediction of Time Between CIS Onset and Clinical Conversion to MS using Random Forests

Viktor Wottschel[1,2], Daniel C. Alexander[2], Declan T. Chard[3], Christian Enzinger[4], Massimo Filippi[5], Jette Frederiksen[6], Claudio Gasperini[7], Antonio Giorgio[8], Maria A. Rocca[5], Alex Rovira[9], Nicola De Stefano[8], Mar Tintoré[9], David H. Miller[3], and Olga Ciccarelli[1]

[1]NMR Research Unit, Department of Brain Repair and Rehabilitation, Queen Square MS Centre, UCL Institute of Neurology, London, London, United Kingdom, [2]Microstructure Imaging Group, Centre for Medical Image Computing, Department for Computer Science, UCL, London, London, United Kingdom, [3]NMR Research Unit, Department of Neuroimflammation, Queen Square MS Centre, UCL Institute of Neurology, London, United Kingdom, [4]Department of Neurology and Section of Neuroradiology, Medical Unversity of Graz, Graz, Graz, Austria, [5]Neuroimaging Research Unit, Vita-Salute San Raffaele University, Milan, Milan, Italy, [6]Department of Neurology, Glostrup Hospital and University of Copenhagen, Copenhagen, Copenhagen, Denmark, [7]Neurology Unit, San Camillo-Forlanini Hospital, Rome, Rome, Italy, [8]Department of Neurological and Behavioral Sciences, University of Siena, Siena, Siena, Italy, [9]Department of Radiology and Neuroimmunology Unit, Hospital Vall d'Hebron, Barcelona, Barcelona, Spain

Target audience Researchers with interest in multiple sclerosis (MS) or machine learning

Purpose To create a fully automated machine-learning-based model that uses MRI data to predict the conversion to MS in patients with clinically isolated syndrome (CIS).

Introduction

CIS is diagnosed after a first neurological attack and can be considered an early stage of MS as ~80% of all CIS patients will have a second relapse within 20 years[1]. The prediction of this second clinical relapse which marks the clinical conversion to MS (i.e., clinically-definite MS, CDMS) is very challenging, and many clinical and radiological predictors of CDMS have been identified[1]. Machine learning techniques such as support vector machines (SVMs) have been widely applied to neuroimaging data in order to associate MRI features with binary clinical outcomes. A single-centre study has shown that it is possible to predict short-time conversion after 1 and 3 years with an accuracy of ~75 % using a priori defined features from baseline MRI measures and clinical characteristics, which were applied to support vector machines (SVMs)[2].

Random forests are another type of machine learning techniques that can easily be applied to regression problems, and consist of an ensemble of decision trees for regression where each tree is created from independent bootstraps from the input data.

The present study shows the feasibility of using random forests with European multi-centre MRI data (obtained at CIS onset) to predict the actual date of conversion to CDMS rather than just a binary outcome at a fixed time point.

Methods

Data: This is a retrospective, multi-centre study, organised and developed within the Magnetic Resonance in Multiple Sclerosis (MAGNIMS) network (www.magnims.eu). Five different European MS centres (Barcelona, Spain; Copenhagen, Denmark; London, United Kingdom; Milan, Italy; Siena, Italy) contributed to this study providing 419 T1 weighted (T1w) images acquired at onset of CIS, as well as both date of CIS onset and date of conversion based on a second clinical relapse. Patients were scanned in each centre following their local MRI protocols. The study received approval from the local ethics committees and written informed consent was obtained prior to the study. For the present study, only 203 patients who converted to CDMS within 3000 days were included.

Preprocessing: Intensity non-uniformities in the T1w MRI scans were removed using N3[3], the brain regions were extracted using BET[4], and finally all scans were registered nonlinearly to the MNI152 space using NiftyReg[5]. All brain voxels were concatenated to a vector for each patient resulting in a matrix of size 203x261701 for the whole cohort.

Random Forest Regression: The patients were subdivided 100 times in semi-random training and testing sets such that every set contains half of the patients from each individual centre. For each of the 100 training/testing sets a random forest was created by using a training set and the associated time-to-conversion (in days) as input to the TreeBagger function from the Statistics Toolbox in MATLAB 2013a. Each forest consists of 100 independent trees with a minimum of 5 observations per tree leaf. We present the average results over all 100 training/testing sets.

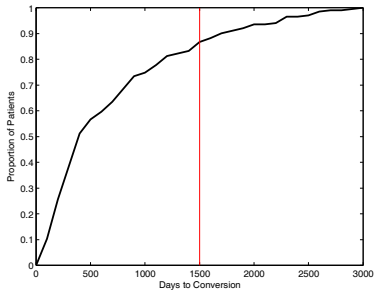

Figure 1: Proportion of patients who converted to CDMS over time.

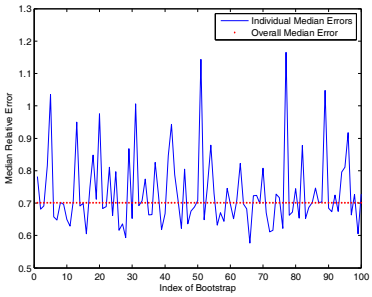

Figure 2: Median errors of the individual bootstraps.

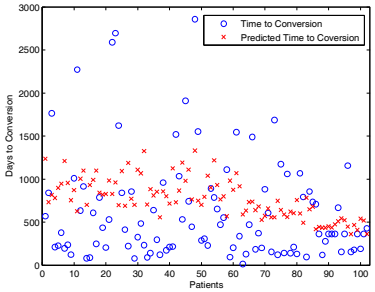

Figure 3: Overview of the actual time to conversion (blue circles) and the predicted values (red crosses) for one example bootstrap.

Results

We gathered MRI data from 203 patients with CIS who converted to CDMS within 8.2 years: 42 % of patients converted to CDMS in one year, 65 % in two years, and 91 % in 5 years as shown in Figure 1. The median relative error for the prediction of the time to conversion is 0.7 (range 0.57-1.16) with respect to the individual patients' conversion times. The median errors of the each bootstrap can be seen in Figure 2. A comparison of the actual time to conversion and the predicted values is shown in Figure 3 for one example bootstrap.

Discussion

The performance of the regression model we construct demonstrates useful predictive power, especially for patients who convert between 1 and 5 years, whereas the errors are quite large for long-term converters. This error results partially from a non-uniform distribution of conversion times as 88 % of the patients convert within the first 1500 days and the remaining time is covered by training samples from only up to 24 patients as shown in Figure 1. However, this reflects the real life situation as 30 % of all CIS patients (including non-converters) convert within 1 year and 80 % within 20 years[1]. The regression problem is particularly challenging, because all patients are at a very early stage of the disease and show similar characteristics in terms of lesion distribution and T1 signal changes[6].

Conclusion

We have presented a fully automated, purely image-based approach to estimate the time to clinical conversion from CIS onset to MS using a heterogeneous MRI data set previously acquired in different European centres. Future work will improve the accuracy by making use of additional MRI modalities such as T2 and diffusion weighted images as they reflect microstructural changes, which may be relevant for clinical conversion to CDMS.

References

[1] Miller et al., Clinically Isolated Syndromes, Lancet Neurol., 2013, 11, 157-169 [2] Ciccarelli et al., Predicting Conversion to MS in Patients with CIS using Machine Learning Techniques, ECTRIMS, No. 113, 2012 [3] Sled et al., A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data, IEEE TMI, 1998, 17(1), pp. 87-97 [4] Smith, Fast Robust Automated Brain Extraction, HBM, 2002, 17(3), 143-155 [5] Modat et al., Fast Free-form Deformations using Graphics Processing Units, Comput. Methods Programs Biomed., 2010, 98, 278-284 [6] Giorgio et al., Location of Brain Lesions Predicts Conversion of CIS to Multiple Sclerosis, Neurology, 2013, 80(3), 234-241

Acknowledgements Thanks to the MS Society in UK, the NIHR UCL-UCLH BRC, and the UCL GCS