World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:11, No:08, 2017

Evaluation of Random Forest and Support Vector Machine Classification Performance for the Prediction of Early Multiple Sclerosis from Resting State FMRI Connectivity Data

Authors : V. Saccà, A. Sarica, F. Novellino, S. Barone, T. Tallarico, E. Filippelli, A. Granata, P. Valentino, A. Quattrone

Abstract : The work aim was to evaluate how well Random Forest (RF) and Support Vector Machine (SVM) algorithms could support the early diagnosis of Multiple Sclerosis (MS) from resting-state functional connectivity data. In particular, we wanted to explore the ability in distinguishing between controls and patients of mean signals extracted from ICA components corresponding to 15 well-known networks. Eighteen patients with early-MS (mean-age 37.42±8.11, 9 females) were recruited according to McDonald and Polman, and matched for demographic variables with 19 healthy controls (mean-age 37.55±14.76, 10 females). MRI was acquired by a 3T scanner with 8-channel head coil: (a)whole-brain T1-weighted; (b)conventional T2-weighted; (c)resting-state functional MRI (rsFMRI), 200 volumes. Estimated total lesion load (ml) and number of lesions were calculated using LST-toolbox from the corrected T1 and FLAIR. All rsFMRIs were pre-processed using tools from the FMRIB's Software Library as follows: (1) discarding of the first 5 volumes to remove T1 equilibrium effects, (2) skull-stripping of images, (3) motion and slice-time correction, (4) denoising with high-pass temporal filter (128s), (5) spatial smoothing with a Gaussian kernel of FWHM 8mm. No statistical significant differences (t-test, $p < 0.05$) were found between the two groups in the mean Euclidian distance and the mean Euler angle. WM and CSF signal together with 6 motion parameters were regressed out from the time series. We applied an independent component analysis (ICA) with the GIFT-toolbox using the Infomax approach with number of components=21. Fifteen mean components were visually identified by two experts. The resulting z-score maps were thresholded and binarized to extract the mean signal of the 15 networks for each subject. Statistical and machine learning analysis were then conducted on this dataset composed of 37 rows (subjects) and 15 features (mean signal in the network) with R language. The dataset was randomly splitted into training (75%) and test sets and two different classifiers were trained: RF and RBF-SVM. We used the intrinsic feature selection of RF, based on the Gini index, and recursive feature elimination (rfe) for the SVM, to obtain a rank of the most predictive variables. Thus, we built two new classifiers only on the most important features and we evaluated the accuracies (with and without feature selection) on test-set. The classifiers, trained on all the features, showed very poor accuracies on training (RF:58.62%, SVM:65.52%) and test sets (RF:62.5%, SVM:50%). Interestingly, when feature selection by RF and rfe-SVM were performed, the most important variable was the sensori-motor network I in both cases. Indeed, with only this network, RF and SVM classifiers reached an accuracy of 87.5% on test-set. More interestingly, the only misclassified patient resulted to have the lowest value of lesion volume. We showed that, with two different classification algorithms and feature selection approaches, the best discriminant network between controls and early MS, was the sensori-motor I. Similar importance values were obtained for the sensori-motor II, cerebellum and working memory networks. These findings, in according to the early manifestation of motor/sensorial deficits in MS, could represent an encouraging step toward the translation to the clinical diagnosis and prognosis.

Keywords : feature selection, machine learning, multiple sclerosis, random forest, support vector machine
Conference Title : ICNCN 2017 : International Conference on Neuroinformatics and Computational Neuroscience
Conference Location : Paris, France
Conference Dates : August 28-29, 2017