Classification of Multiple Sclerosis Clinical Forms by 1H Magnetic Ressonance Spectroscopy of Cerebrospinal Fluid

F. X. Aymerich[1,2], J. Alonso[1], M. Comabella[3], M. E. Cabañas[4], M. Tintoré[3], X. Montalban[3], and A. Rovira[1]

[1]Unitat RM Vall Hebron (IDI), Hospital Vall Hebron, Barcelona, Barcelona, Spain, [2]Enginyeria de Sistemes, Automàtica i Informàtica Industrial, Universitat Politècnica de Catalunya, Barcelona, Barcelona, Spain, [3]Unitat Neuroimmunologia Clínica, CEM Cat, Hospital Vall Hebron, Barcelona, Barcelona, Spain, [4]Servei Ressonància Magnètica Nuclear, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain

Introduction

The use of high-resolution 1H nuclear magnetic resonance (1H NMR) spectroscopy in the study of biofluids enables measurement of a large range of biochemical compounds. Previous reports have shown the utility of pattern recognition techniques to differentiate among the metabolic profiles of diseases [1][2][3].

Purpose

To design a fuzzy classifier to differentiate among primary progressive (PP) multiple sclerosis (MS), relapsing remitting (RR) MS, and non-MS conditions by analyzing 1H NMR spectra of cerebrospinal fluid (CSF).

Methods

We studied 22 CSF samples (7 PP, 8 RR and 7 non-MS patients with other neurological diseases) obtained by lumbar puncture. CSF samples of 0.125 mL were adjusted to pH 7.1 and diluted to 0.6 mL with deuterated water. 1H NMR was performed on a Bruker DPX500 spectrometer, equipped with a 1H/13C/BB cryoprobe.

We considered three datasets to study 1H NMR spectra, corresponding to the data located within the aliphatic region (0.5-4.5 ppm), the aromatic region (6.0-9.7 ppm), and the aggregation of the previous two regions. Post-processing tasks to set up the data in order to improve the classification results were performed. These tasks included baseline correction, data spectrum reduction by integration of 0.02 ppm intervals, suppression of low signal to noise ratio regions, normalization, and scaling of spectra data.

The design considered the fusion of classifiers based on decision trees. Then, we considered a two-level classifier. The first level consisted of a series of decision trees designed considering fuzzy techniques. We obtained these decision trees from the study of the more relevant features extracted using the decision tree learner option in RapidMiner [4]. The second level was the fusion of the outputs of the first-level classifiers. These outputs were aggregated by ordered weighted averaging operators [5] considering that the determination of the possibility within the [0,1] interval of belonging to each group (PP; RR; non-MS conditions) required the half more one of first-level classifiers. Samples were then assigned to the group with the highest possibility value. Evaluation of the results was based on correct classification of samples into each group and a robustness index, which was calculated as the difference between the possibility of belonging to the correct group and the highest of the other two possibility values for each sample. A 12-fold cross-validation method was used to estimate the performance of the classifier. Given that we considered three datasets in the spectra, we designed three classifiers, each one optimized to the features of each dataset.

Results

Table 1 shows the classification results for each dataset. According to the cross-validation method, classification levels were very good for the dataset corresponding to aliphatic region and excellent for the other two datasets. Robustness levels were satisfactory for the dataset corresponding to aromatic region and good for the other two datasets. Only the classifier developed for the dataset corresponding to the aggregation aliphatic and aromatic regions achieved the correct classification for all samples

Conclusions

The fuzzy classifier presented allows classification of CSF samples according to the patterns of PP, RR and non-MS groups. Best results were obtained when we considered the information provided by both aliphatic and aromatic regions.

Datasets	Indexes	
	Classification [0,1]	Robustness [-1,1]
Aromatic region	0.96 (0.08)	0.34 (0.11)
Aliphatic region	0.92 (0.11)	0.50 (0.10)
Aromatic + aliphatic region	1.00 (0.00)	0.44 (0.07)

Table 1. Classification results for each dataset. Values shows mean (standard deviation) values obtained for each index.

References

[1] Lutz, NW et al. PLoS One. 2007; 2:e595

[2] t Hart, BA et al.. J Neurol Sci. 2003; 212: 21-30

[3] Tate, AR et al. Magn. Reson. Med., 2003, 49: 29-36

[4] Mierswa, I et al. Proc. of the 12th ACM SIGKDD Int. Conf. on Knowledge and Data Mining,2006; 935-940

[5] Yager, R.R. Fuzzy Sets and Systems. 1993; 59: 125-148.