Spectral Prototype Extraction for the discrimination of glioblastomas from metastases in a SV 1H-MRS brain tumour database

S. Ortega-Martorell[1,2], I. Olier[3], A. Vellido[4], M. Julià-Sapé[1,2], and C. Arús[1,2]

[1]Grup d'Aplicacions Biomèdiques de la RMN (GABRMN), Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Barcelona, Spain, [2]Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Zaragoza, Spain, [3]Institut de Neurociències (INc), Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Barcelona, Spain, [4]Departament de Llenguatges i Sistemes Informàtics (LSI), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

Introduction: Feature Extraction (FE) is often performed in MRS datasets prior to classification for diagnostic purposes. Principal Components Analysis (PCA), the most commonly used FE technique in MRS has some limitations in this scenario, including: lack of robustness in the presence of noise; the lack of a standard criterion for the selection of the most adequate number of extracted components; also, each extracted component is a linear combination of all the spectral frequencies, making their metabolic description difficult and hence limiting their interpretability as diagnostic predictors. Moreover, PCA bypasses the fact that MRS does not comply with the independent and identically-distributed (i.i.d.) condition. We propose a novel FE technique, Spectral Prototype Extraction (SPE) that overcomes these limitations while providing competitive results in the difficult problem of discriminating metastases from glioblastomas on the basis of SV ^{1}H-MR information. It is based on a manifold-constrained Hidden Markov Model (HMM). Its Bayesian formulation imbues it with regularization properties that minimize the negative effect of the presence of noise in the data. This model, Variational Bayesian Generative Topographic Mapping Through Time (VB-GTM-TT: [1]) segments the MRS in an interpretable way. Each of the resulting spectral prototypes (SP) represents a complete interval of frequencies or a collection of such intervals. Importantly, each frequency is represented in only one SP, while all frequencies are represented in at least one SP.

Subjects and methods: The available data were SV ^{1}H-MR spectra at 1.5T, acquired from brain tumour patients at short and long echo times. They included 78 glioblastomas (WHO 9440/3) and 31 metastases (WHO 8000/6) from a multicentre study. These were extracted from the web-accessible INTERPRET project database [2]. Two spectra were available from the same VOI, for each case: 1) Short TE (SET) PRESS or STEAM (1600-2000/20-32) (TR/TE in ms); and 2) Long TE (LET) PRESS (1600-2000/135-144). Spectra were automatically processed with the INTERPRET data manipulation software and if considered necessary (zero and first order phase), previously adjusted with jMRUI. Clinically-relevant regions of the spectra were sampled to obtain 195 frequency intensity values (observed data features from the pattern recognition point of view) in the 4.24 – 0.50 ppm range. For their analysis, PCA and SPE were applied to SET and LET. Relevant SP for further classification were selected using a sequential forward (greedy stepwise) method and validated using a correlation-based criterion. A subsequent classification step was applied with Linear Discriminant Analysis (LDA) to both PCA and SPE using SpectraClassifier [3]. Classifier results were validated through bootstrap with 1000 repetitions, and averaged accuracy (AA) and standard deviation (SD) values as well as Area Under the ROC Curve (AUC) were obtained.

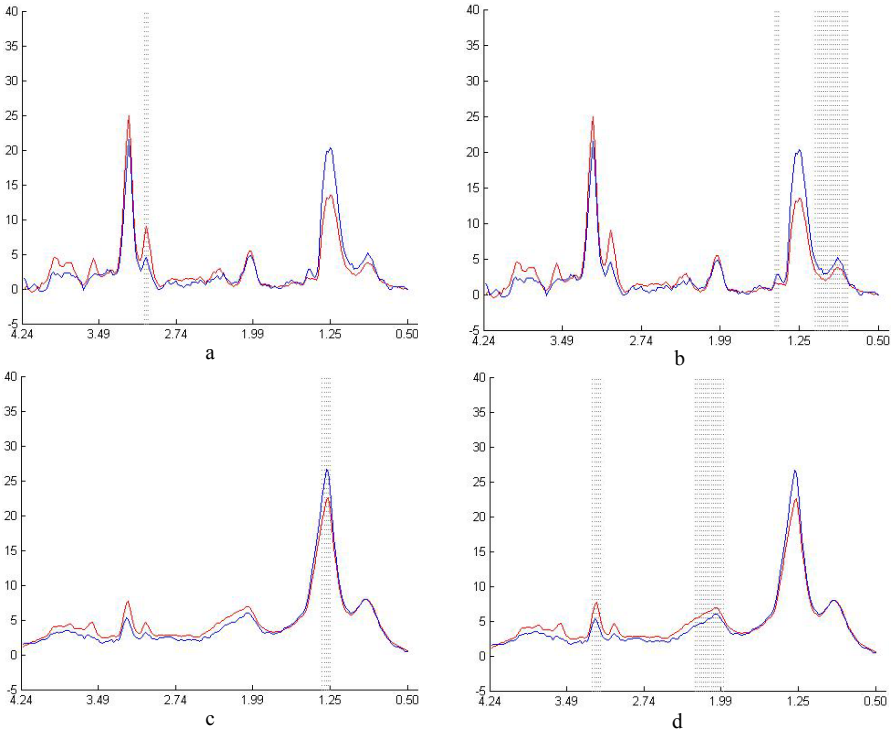

Figure 1: Example of the SPE results. Red: mean spectrum of glioblastomas. Blue: mean spectrum of metastases. **Top)** Two of the 20 SP describing LET. a) It contains 3 frequencies (from 3.05 to 3.01ppm) corresponding to Creatine. b) It contains 20 frequencies (from 1.48 to 1.44ppm and 1.09 to 0.79ppm) corresponding to Alanine and Lipids. **Bottom)** Two of the 23 SP describing SET. c) It contains 5 frequencies (from 1.32 to 1.25ppm) that correspond to mobile Lipids. d) It contains 20 frequencies (from 3.24 to 3.16ppm and 2.24 to 1.97ppm) that correspond to total Choline and mobile Lipids/Glx/Macromolecules, respectively.

Experimental results and discussion: An illustrative example of the SPE results is shown and discussed in Fig.1. All the obtained SP turned out to be readily interpretable frequency intervals or concatenations of intervals, as expected. The dimensionality of the data was radically reduced (20 SP for LET and 22 for SET). This, together with the frequency interval-like nature of the resulting SP should make the subsequent diagnostic classification easier to explain for radiology experts. A sample of comparative classification results of the application of PCA and SPE together with LDA are compiled in Table 1. Both models yield similar classification performance for LET and SET, with few statistically significant differences (according to a t-Student test with $p \leq 0.05$ significance threshold), in favour of SPE in LET and in favour of PCA in SET. Further research should investigate the performance of these FE techniques with nonlinear classifiers and their robustness with independent test sets.

	PCA				SPE			
	PC	**AA ± SD**	**AA per class ± SD**	**AUC**	**SP**	**AA ± SD**	**AA per class ± SD**	**AUC**
LET	17	76.85 ± 4.082	Gl: 76.61 ± 4.814 Me: 77.52 ± 7.370	0.812	18	78.03 ± 3.983	Gl: 74.39 ± 4.928 Me: 87.18 ± 5.995	0.847
SET	17	70.05 ± 4.111	Gl: 69.57 ± 4.979 Me: 71.17 ±7.595	0.769	22	67.85 ± 4.320	Gl: 67.63 ± 5.077 Me: 68.24 ± 7.558	0.771

Table 1: Classification results for PCA-based LDA and SPE-based LDA. The results chosen for comparison are the best obtained for each method.

Conclusions: SPE offers the capability of creating spectral prototypes which correspond to known metabolites or groups of metabolites that can be used as readily interpretable input features in classifiers, yielding results that are comparable to those of PCA-based classification.

References: [1] Procs. of IJCNN, 2008, pp.517-522. [2] MAGMA, 2006, 19(1), pp.22-33. [3] Procs. of ISMRM, 2009, pp.3477.