Possibilistic classification of Brain Tumors by MRS based on Functional Data Analysis and Subpattern discovery

J. García-Gómez[1], I. Epifanio[2], M. Julià-Sapé[3,4], D. Monleón[3,5], J. Vicente[1], S. Tortajada[1], E. Fuster[1], A. Moreno-Torres[6], A. Peet[7,8], F. Howe[9], B. Celda[3,10], C. Arús[3,4], and M. Robles[1]

[1]ITACA-IBIME, Universidad Politécnica de Valencia, Valencia, Valencia, Spain, [2]Departament de Matemàtiques, Universitat Jaume I, Valencia, Valencia, Spain, [3]CIBER de Bioingeniería, Biomateriales y Nanomedicina, Spain, [4]Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain, [5]Fundación de Investigación del Hospital Clínico Universitario de Valencia, Valencia, Spain, [6]Research Department, Centre Diagnòstic Pedralbes, Barcelona, Spain, [7]University of Birmingham, Birmingham, United Kingdom, [8]Birmingham Children's Hospital, Birmingham, United Kingdom, [9]St George's Hospital Medical School, London, United Kingdom, [10]Departamento de Química-Física, Universitat de València, Valencia, Spain

Introduction

MRS is becoming an additional accurate non-invasive technique for initial examination of Brain Tumors (BT) [1]. When designing Clinical Decision Support Systems for Brain Tumors based on MRS, it would be of interest to be able to accept any prospective case. Besides, due to the possible tumor heterogeneity in the acquired voxel, acquisition artifacts or molecular tumor subtypes, the in-vivo MRS pattern may be heterogeneous within each diagnostic class. Previous work has focused on the most frequent brain tumor classes [1,2], which were usually characterized by assuming unimodal distributions [1].

Purpose

Our aim has been to evaluate and interpret, in the largest multicenter database available to us, a possibilistic classification strategy of BT using ^{1}H SV Short TE MRS based on Functional Data Analysis (FDA) [3] and subpattern search [4].

Materials

The dataset used in this study consisted on 967 ^{1}H-MR spectra of histopathologically diagnosed brain tumor cases compiled by ten institutions in the framework of the INTERPRET and eTUMOUR projects [5,6]. The acquisition protocol was SV ^{1}H-MRS at 1.5T at Short TE (TE 20-32ms/ TR 1600- 2020ms/ Sweep width 1000, 2500Hz/ Spectral vector datapoints 512, 1024, and 2048). Signal processing was performed following the protocols defined in [2]. The distribution of cases by pathology was: 27 Astrocytoma I (AS1), 74 Astrocytoma II (AS2), 37 Astrocytoma III (AS3), 252 Glioblastoma (GBM), 16 Lymphoma (LYM), 26 Medulloblastoma (MED), 80 Meningioma (MEN), 118 Metastasis (MET), 40 Oligodendroglioma II (OD2), 12 Oligodendroglioma III (OD3), 26 Oligoastrocytoma II (OA2), 7 Oligoastrocytoma III (OA3). There were 225 cases that belonged to other less frequent tumor types (OTH) and there were 27 Normal brain spectra (NOR).

Methods

A function of frequency defined in a base of cubic splines on 94 knots in the [0.5,4.1 ppm] interval was constructed for each spectrum [3]. In order to fit the spectral shape, first derivative functions were registered to their mean function. Unsupervised learning was applied to the functional PCA (fPCA) decomposition of the spectral functions to detect outliers and subpatterns (SP) within each diagnosis. One-class LDA models [7] were trained for the 14 initial pathology diagnoses to develop the fPCA-LDA classifier. Additionally, the Subpattern-based fPCA-LDA classifier consisted on one-class LDA classification of the subpatterns found during the unsupervised step. The AUC of the ROC curve estimated on a 5x5 Cross Validation re-sampling was calculated for each one-class classifier. The performance of the Subpattern-based fPCA-LDA classifier was calculated as the weighted average of the AUC over the mixture of subpattern classifiers of each diagnosis.

Results

The performances of the fPCA-LDA classifier and of the Subpattern-based fPCA-LDA classifier are shown in Table 1. Two subpatterns were detected in AS3, four in GBM, two in MED, three in MET and six in the OTH class. Besides, 43 samples were considered outliers by the unsupervised learning procedure. As an example, two out of the four subpatterns of GBM are presented in Figure 1. The 1st glioblastoma SP shows high variability in the 1.3 and 3.2-4.1ppm regions and low variability in the 2-2.5ppm one. The 2nd glioblastoma SP is characterized by high peaks at 1.3ppm and low variability in the rest of the regions.

Table 1: Performance estimation (ROC-AUC) by diagnosis for the fPCA-LDA classifier and Subpattern-based fPCA-LDA classifier. Number of subpatterns (SP) is also shown.

Diag.	**Performance of the fPCA-LDA classifier** (AUC)	**Performance of the Subpattern-based fPCA-LDA classifier** (AUC)
AS1	0.70	0.70
AS2	0.79	0.79
AS3	0.66	0.75 (2 SP)
GBM	0.69	0.90 (4 SP)
LYM	0.69	0.75
MED	0.85	0.92 (2 SP)
MEN	0.86	0.89
MET	0.64	0.79 (3 SP)
OD2	0.79	0.78
OD3	0.60	0.66
OA2	0.68	0.70
OA3	0.83	0.79
OTH	0.49	0.81 (6 SP)
NOR	0.94	0.96

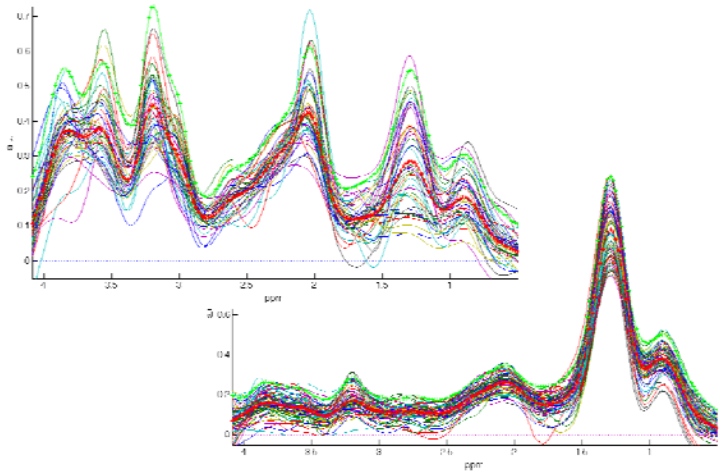

Figure 1: The two more populated subpatterns observed for GBM. Bottom: 1st subpattern, characterized by a high 1.3ppm peak. Top: 2nd subpattern, characterized by high variability in the 1.3 and 3.2-4.1ppm and low variability in the 2-2.5ppm regions.

Discussion & Conclusions

The possibilistic classifier developed in this study has been evaluated on the largest multicenter database of MRS of brain tumors available to us and each subpattern obtained in our study was characterized by its mean and its PCA decomposition functions. The proposed methodology based on FDA and subpattern analysis overperformed the unimodal fPCA-LDA approach and the interpretability of the naturally occurring heterogeneous patterns of the diagnoses. The detected in-vivo MR spectral pattern subtypes seem robust and reproducibly encountered. Correlation of these subtypes with clinical, transcriptomic and metabolomic data is in progress.

Acknowledgements - This work was partially funded by the European Commission (FP6-2002-LIFESCIHEALTH 503094) and (IST-2004-27214). We thank eTUMOUR and INTERPRET partners for providing data, in particular C. Majós (IDI-Bellvitge), J. Griffiths (SGUL), A.Heerschap (RU), W. Gajewicz (MUL), and Jorge Calvar (FLENI), and J Capellades (IDI-Badalona),.

References - [1] NMR Biomed. 2006 Jun; 19(4):411-34 [2] DOI 10.1007/s10334-008-0146-y [3] Ramsay, Functional Data Analysis. Springer. 2005. [4] CM Bishop, Pattern Recognition and Machine Learning. Springer. 2006. [5] INTERPRET: azizu.uab.es [6] eTUMOUR: www.etumour.net [7] D Tax, One-class classification. DUT. 2001