Multiproject-Multicenter evaluation of automatic brain tumor classification by Magnetic Resonance Spectroscopy

Magma Editorial Office, CRMBM-CNRS, FacultØ de MØdecine, 27 Boulevard Jean Moulin, 13005 Marseille, France

Juan M. García-Gómez[1,*], Jan Luts[2], Margarida Julià-Sapé[3,4],

Patrick Krooshof[5], Salvador Tortajada[1], Javier Vicente Robledo[1],

Willem Melssen[5], Elies Fuster-García[1], Iván Olier[4],

Geert Postma[5], Daniel Monleón[4,6], Àngel Moreno-Torres[7],

Jesús Pujol[8], Ana-Paula Candiota[3,4], M.Carmen Martínez-Bisbal[3,9],

Johan Suykens[2], Lutgarde Buydens[5], Bernardo Celda[9,3],

Sabine Van Huffel[2], Carles Arús[4,3], Montserrat Robles[1]

[1]IBIME-Itaca, Universidad Politécnica de Valencia, Spain

[2]Department of Electrical Engineering (ESAT), Research Division SCD,

Katholieke Universiteit Leuven, Leuven, Belgium

[3]CIBER de Bioingeniería, Biomateriales y Nanomedicina, Spain

[4]Departament de Bioquímica i Biologia Molecular,

Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

[5]Institute for Molecules and Materials. Radboud University.

The Netherlands (Gelderland), Nijmegen

[6]Fundación de Investigación del Hospital Clínico

Universitario de Valencia, Valencia, Spain

[7]Research Department, Centre Diagnòstic Pedralbes,

Esplugues de Llobregat, Spain

[8]Institut d'Alta Tecnologia-PRBB,

CRC Corporació Sanitària. Barcelona

[9]Departamento de Química-Física,

Universitat de Valencia, Spain

September 7, 2008

* Corresponding author: juanmig@upv.es, http://www.ibime.upv.es/; Telf: +34 963877000, ext. 75278. Universidad Politécnica de Valencia, Facultad de Informática. Camino de Vera, s/n, 46022, Valencia, Spain

- Word count of abstract: 200
- Word count of text: 4500
- Number of figures: 7
- Number of tables: 4
- Number of references: 69

Abstract

Automatic brain tumor classification by MRS has been under development for more than a decade. Nonetheless, to our knowledge, there are no published evaluations of predictive models with unseen cases that are subsequently acquired in different centers. The multicenter eTUMOUR project (2004-2009), which builds upon previous expertise from the INTERPRET project (2000-2002) has allowed such an evaluation to take place.

253 pairwise classifiers for Glioblastoma, Meningioma, Metastasis, and Low-Grade Glial diagnosis were inferred based on 211 SV Short TE INTERPRET MR spectra obtained at 1.5T (PRESS or STEAM, $20 - 32ms$) and automatically pre-processed. Afterwards, the classifiers were tested with 97 spectra, which were subsequently compiled during eTUMOUR. In our results based on subsequently acquired spectra, accuracies of around 90% were achieved for most of the pairwise discrimination problems. The exception was for the Glioblastoma versus Metastasis discrimination, which was below 78%. **A more clear definition of metastases may be obtained by other approaches, such as MRSI+MRI.**

Therefore, the prediction of the tumor type of in-vivo MRS is possible using classifiers developed from previously acquired data, in different hospitals with different instrumentation under the same acquisition protocols. This methodology may find application for assisting in the diagnosis of new brain tumor cases and for the quality control of multicenter MRS databases.

Keywords: Magnetic Resonance Spectroscopy, Pattern Classification, Brain tumors, Decision Support Systems, multicenter evaluation study.

1 Introduction

Magnetic Resonance Spectroscopy (MRS) is slowly becoming an accurate non-invasive complement to Magnetic Resonance Imaging (MRI) for initial diagnosis exam of brain masses [1], since it provides useful chemical information about metabolites for characterizing brain tumors [2]. To achieve this status, clinical and Pattern Recognition (PR)-based classification of brain tumors using MRS data has been thoroughly investigated for more than fifteen years [1, 3–13].

The Clinical Decision-Support Systems (CDSSs) based on PR should be developed in such a way so as to obtain high accuracy in classification, interpretability by means of clinical knowledge and the generalization of the performance to new samples obtained subsequently in different clinical centers [14–17]. Standardization of acquisition conditions and protocols should make data from different hospitals compatible and allow the development and evaluation of joint CDSSs. This standardization prevents possible bias from single-center or single-machine studies and, additionally, increases the number of available cases for classifier development and test purposes.

During the INTERPRET project (INTERPRET) [8, 18], a protocol was defined to guarantee the compatibility of the signals acquired at different hospitals [19, 20]. As a result, studies on automated brain tumor classification were carried out using these data. Hence, in previous studies [7, 8, 10, 21], the ability of automatic classifiers based on Short Echo Time (TE) MRS to discriminate among different brain tumor diagnoses was demonstrated. In addition, in [11, 13, 21], the automated classification by means of Long TE MRS was also studied and demonstrated. Other studies evaluated the extension of the classifiers towards ^{1}H Magnetic Resonance Spectroscopic Imaging

(MRSI) [12, 21–24]. Every study reported above was developed and evaluated using data acquired during the same period of time. Besides, other automated classification studies, such as [2, 13, 25–28], have been reported on single-center MRS datasets of brain masses.

In order to provide the clinical community with robust results of automatic classification, the extension of the evaluation in time is advisable. Hence, the validation of classifiers through subsequent cases can consolidate the confidence of clinicians in the potential applicability of these classifiers. The multicenter the eTUMOUR project (eTUMOUR) [29] (2004-2009) has benefited from the data and expertise gathered by INTERPRET. The INTERPRET acquisition protocols for clinical, radiological, and histopathological data were extended to ex-vivo transcriptomic (DNA microarrays) and metabolomic (HR-MAS) data acquisition in eTUMOUR. Furthermore, the raw MRS data acquired during INTERPRET were incorporated into the eTUMOUR dataset for classifier development. This provides a unique opportunity to evaluate INTERPRET-based models by means of cases from a later date from partly different hospitals with different instrumentation, but obtained using the same or compatible acquisition protocols. The multiproject-multicenter evaluation proposed in this study gives a close-up perspective of the conditions that predictive models may face under different real clinical environments.

In this study, six pairwise classifiers for Glioblastoma (GBM), Low-grade Meningioma (MEN), Metastasis (MET), and Low-Grade Glial (LGG) diagnoses were developed and tested on Single-voxel (SV) Short TE MRS signals. Short TE MRS is fast (typically $5min$) and robust, so it is considered to be appropriate for routine clinical studies [1].

Most major hospitals currently use this acquisition protocol for the MRS evaluation of brain tumors. Short TE spectral pattern has been reported to contain a larger amount of information than Long TE spectra, e.g. metabolites and other compounds that are considered useful for classification purposes [1, 8, 11]. Hence, Creatine (Cr) (3.02, $3.92ppm$), Choline (Cho) ($3.21ppm$), N-Acetyl Aspartate (NAA) ($2.01ppm$), myo-Inositol (mI) and Glycine (Gly) ($3.55ppm$), mI/Taurine (Tau) ($3.26ppm$), Glutamate/Glutamine (Glx) ($2.04, 2.46, 3.78ppm$), Lactate (Lac) ($1.31ppm$), and Alanine (Ala) ($1.47ppm$) are observed at Short TE. Furthermore, Macromolecules (MM) ($5.4ppm$, $2.9ppm$, $2.25ppm$, $2.05ppm$, $1.4ppm$ and $0.87ppm$) and Mobile Lipids (ML) are also well detected at Short TE [1, 8]. Comparative studies on the use of Short TE vs. Long TE have shown the benefit of using Short TE or the combination of both echo times for automatic classification purposes [30].

Based on previous results from [10, 11, 18, 21], good performance of the PR models could be expected for most of the classification problems, except for the discrimination of Glioblastoma and Metastasis [10]. Our performance estimations of models trained with INTERPRET data and tested over eTUMOUR cases confirmed this behaviour. We observed that pairwise discrimination between Glioblastoma, Meningioma, Metastasis, and Low-Grade Glial achieved an accuracy of around 90%. The exception was for the discrimination between Glioblastoma and Metastasis that did not perform better than 78%. This study consolidates the results obtained by previous studies in automatic brain tumor classification using MRS. These results may also increase the confidence of the clinical community in the use of CDSSs that incorporate this kind of classifiers for the interpretation of MRS biomedical signals and the

diagnosis of brain tumors.

2 Materials and Methods

2.1 Data acquisition

The training data used for classifier development were SV MRS signals at $1.5T$ at Short TE (Point-Resolved Spectroscopic Sequence (PRESS) or Stimulated Echo Acquisition Mode sequence (STEAM), 20-32ms) that were acquired by international centers in the framework of INTERPRET [18]. The classes considered for inclusion in this study were based on the histological classification of the Central Nervous System (CNS) tumors set up by the World Health Organization (WHO) [31]: GBM, MEN, MET, and LGG (Astrocytoma gII, Oligoastrocytoma gII, or Oligodendroglioma gII). The number of cases by class is summarized in Table 1.

211 SV ^{1}H (Nuclear) Magnetic Resonance (MR) spectra from the INTERPRET database [19] were included. These signals were acquired with Siemens, General Electric (GE), and Philips instruments by six international centers. The acquisition protocols included PRESS or STEAM sequences, with spectral parameters: Repetition Time (TR) between 1600 and 2020ms, TE of 20 or 30-32ms, spectral width of $1000 - 2500Hz$, and 512, 1024, or 2048 data-points, as described in previous studies [19]. Every training spectrum and diagnosis was validated by the INTERPRET Clinical Data Validation Committee (CDVC) and expert spectroscopists [8].

The test data were provided by eight international institutions in the framework of eTUMOUR [29]. The cases with the SV Short TE (STEAM 20 ms, PRESS 30-

32 ms) MRS at 1.5T signal validated by the expert spectroscopist of eTUMOUR and with the original histopathology available before February 28th, 2007) were included. Therefore, 97 cases from eTUMOUR were considered for testing in this study. The test cases used to evaluate the performance of the classifiers were acquired from partly different hospitals in later dates than the training cases and using instruments of the three main manufacturers. Table 2 shows that the percentages of cases by manufacturer included in the test data are similar to the percentages in the training data. Table 3 shows the percentage of cases by center included in the training and test datasets. Forty per cent of training cases belong to one center that afterwards did not provide test data. Besides, 35% of test cases belong to three new centers that were not providers of training data.

2.2 Pre-processing

Each signal was pre-processed according to the INTERPRET protocol. A fully automatic pre-processing pipeline was available for the training data. Besides, a semi-automatic pipeline was defined for some new file formats of the test cases from GE and Siemens manufacturers. The semi-automatic pipeline was designed to ensure compatibility of its output with the automatic one.

2.2.1 Automatic pipeline

The steps of the automatic pre-processing pipeline were: 1) Eddy current correction was applied to the water-suppressed Free Induction Decay (FID) of each case using the Klose algorithm [32]. 2) The residual water resonance was removed using the Hankel-

Lanczos Singular Value Decomposition (HLSVD) time-domain selective filtering using 10 singular values and a water region of $[4.33, 5.07]ppm$. 3) An apodization with a Lorentzian function of $1Hz$ of damping was applied. 4) Before transforming the signal to the frequency domain using the Fast Fourier Transform (FFT), an interpolation was needed in order to increase the frequency resolution of the low resolution spectra to the maximum frequency resolution used in the acquisition protocols (see [8] for details in the acquisition conditions and resolutions). This was carried out with the zero-filling procedure. 5) Afterwards, the baseline offset, which was estimated as the mean value of the region $[11, 9] \cup [-2, -1]ppm$, was subtracted from the spectrum. 6) The normalization of the spectral data vector to the L2-norm was performed based on the data-points in the region $[-2.7, 4.33] \cup [5.07, 7.1]ppm$. 7) Depending on the signal-to-noise ratio (SNR) and the tumor pattern, an additional frequency alignment check of the spectrum was performed by referencing the ppm-axis to (in order of priority) the total Cr at $3.03ppm$ or to the Cho containing compounds at $3.21ppm$ or the ML at $1.29ppm$. 8) Finally, the region of interest was restricted to $[0.5, 4.1]ppm$, obtaining a vector of 190 points for each spectrum where, after the pre-processing filters, the resonances of the main metabolites arise and where the contribution of the residual water is expected to be minimal. In summary, 211 INTERPRET cases and 47 cases of the eTUMOUR test dataset (32 from Philips and 15 from GE) were pre-processed with the automatic pipeline.

2.2.2 Semi-automatic pipeline

Due to limitations of the automatic pre-processing software, 50 test samples were pre-processed by a semi-automatic pipeline that was partially based on the java Magnetic Resonance User Interface (jMRUI) [33]. Some modifications of the semi-automatic pipeline with respect to the automatic pipeline were in the following steps: 1) The phase of the water-suppressed FID was mainly corrected with the reference water. Additional manual zero-order and first-order phase correction was performed when needed. 2) Residual water was removed by means of the jMRUI-implementation of the Hankel Singular Value Decomposition (HSVD) algorithm [34]. The filter was parametrized as in the automatic pipeline. Steps 3-8 remained equivalent to the automatic pre-processing. As a result, a pre-processing pipeline based on different software implementations but compatible with the automatic one was set up, and comparable signals for testing the PR models were obtained.

2.3 Feature extraction

Several feature extraction methods based on PR were applied to the real part of the spectra prior to any classification approach. These methods included direct spectral Peak Integration (PI) on selected metabolite resonance regions [35], peak height of typical resonances (PPM) [36], Principal Component Analysis (PCA) [37, 38], Independent Component Analysis (ICA) [39, 40], and Wavelet transform (WAV) [41, 42]. Finally, some classification approaches were applied to the full region of interest represented by a data vector of 190 points (190). The selected features for the classifiers were derived from previous studies [10, 30] or from model validation based on the

training dataset. In some approaches, Standard Normal Variate (SNV) scaling was applied to the obtained features. The wavelet basis used in the experiments was coiflet 3 with 9 levels [41]. Further information and experimental details about the methods used can be found in Appendix A of the Supplementary Material[1].

2.4 Classification methods

Ten methods were applied to address the pairwise classifications. These methods included parametric discriminant analysis [43]: Linear Discriminant Analysis (LDA), Fisher's rank-reduced version of LDA (FLDA) [44]), Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis with diagonal covariance matrix (dLDA) and Quadratic Discriminant Analysis with diagonal covariance matrix (dQDA). Kernel-based models (Support Vector Machines (SVM) [45] and Least-Squares Support Vector Machine (LSSVM) [46]) were also applied. Additionally, Artificial Neural Networks (Multilayer Perceptron (MLP) [47] and Bi-directional Kohonen Networks (BDK) [25, 48]) and single and ensemble [49] classifiers using K-nearest neighbours and local feature reduced by PCA (PCA-KNN) [50, 51]) were used.

Bayesian strategies for regularization were also applied in some of the classifiers based on LSSVM [52] and MLP [53]. Further information about these methods can be found in Appendix B of the Supplementary Material.

[1] Available from: http://bmg.webs.upv.es/joomla_rpboys/articulos/mmeval_mrs08.pdf

2.5 A measure to evaluate unbalanced classifiers: the Balanced Error Rate (BER).

The performance was measured by means of the Error Rate (ERR) and the Balanced Error Rate (BER). In a binary classifier A vs. B, BER is the average of the error rate on the A and B classes [54]. Let n_A be the number of cases of the class A, and e_A the number of misclassified cases. Let n_B be the number of cases of the class B, and e_B the number of misclassified cases. While the ERR is defined as $\frac{e_A+e_B}{n_A+n_B}$, the BER is defined as $\frac{1}{2}(\frac{e_A}{n_A}+\frac{e_B}{n_B})$. BER is useful when one class is underrepresented compared to the other class, e.g. GBM vs. LGG and GBM vs. MET in the INTERPRET dataset and MEN vs. GBM and MEN vs. MET in the eTUMOUR dataset.

3 Results and discussion

For each task, different combinations of feature extraction and classification methods were applied in the study. An estimation of the ERR and BER for the INTERPRET dataset using a 10-fold Cross Validation (CV) was carried out for each model. Afterwards, the estimations of the ERR and BER were obtained on the Independent Test (IT) dataset of eTUMOUR. Table 4 illustrates the results with the best pairwise classifiers based on the IT estimations. A detailed list of the results is available in Section 1 of the Supplementary Material.

3.1 The classification problems

Most of the discrimination problems among the four classes were solved with high accuracy in the eTUMOUR dataset. Table 4 shows that most of the best classifiers among GBM, MEN, MET, and LGG achieved an accuracy ($1 - ERR$) of around 90%. Such decision support methodologies with these ratios of accuracy may be useful to be incorporated in integrated CDSSs for clinical purposes. Besides, for GBM vs. MET, the best result was an accuracy of 78% of the independent test, which is far from the accuracy obtained for the other discrimination problems. The Glioblastoma vs. Metastasis discrimination by means of the MRS is difficult with the use of SV spectroscopy alone [7, 8, 55–58]. Other approaches, such as MRSI coupled with MRI or the acquisition of an additional adjacent voxel to the brain mass should provide relevant additional information for distinguishing between these two types of tumors [57–59].

Figure 1 shows the box-whisker plot of the performance (BER based on IT) for each problem based on the detailed list of the results (Section 1 of the Supplementary Material). Note the high deviation of the distribution for the GBM vs. MET with respect to the others. In a multiple comparison at a 0.05 *alpha*-level based on the Tukey's honestly significance difference criterion for Kruskal-Wallis nonparametric one-way analysis of variance [60], each problem had a mean rank that was significantly different from the GBM vs. MET problem. The distributions of the other five discrimination problems overlapped among them. Nevertheless, the smallest non-outlier observation of the GBM vs. LGG problem was higher than the smallest non-outlier observation of the remaining problems. This may indicate that the GBM vs. LGG discrimination is more difficult to solve by SV Short TE MRS than the other four discrimination

problems.

The different approaches obtained good results for the discrimination of the GBM and MEN classes. A multilayer perceptron with the full spectra achieved a BER of 0.09. The mode of the distribution of BER was below 0.20 for the GBM vs. MEN problem.

The difficulty of the GBM vs. MET discrimination was clearly observed in both CV-and IT-estimations (see Figure 2). In the distribution of the IT results for this problem, the BER mode was 0.5, and the main distribution of the results ranged from 0.4 to 0.55. Some methods achieved a BER of 0.2; nevertheless, the main mass of the distribution was far from this value, which makes it difficult to ensure reproducibility of these performances. These results agree with those already published in previous studies [8, 10]. This is most probably due to the similar necrotic profile (high lipid peaks mask the rest of the metabolic information) of the Metastasis cases and of most of the Glioblastoma cases.

The mode of the BER for the GBM vs. LGG problem was 0.2. Nevertheless, there was a set of regularized classifiers that obtained a BER of around 0.09. To be more precise, the best BER corresponded to the Bayesian framework for LS-SVM using PI values. Devos et al. [10] obtained comparable performances for this problem using LDA and standard LS-SVMs. In studies [26, 61], significant statistical differences between GBM and LGG and between GBM and astrocytoma grade-III were also found for different metabolite ratios with respect to Cr and/or water. In Long TE, Menze et al. [13] observed a better performance with regularized methods than with the standard ones when classifying normal, non-progressive tumors (with radiation injury and stable

disease) and brain tumors.

As expected, our results confirm that MEN can be easily discriminated from MET no matter what method is used. Most of the BER probability mass of the results was in the interval from 0.1 to 0.2. The best result achieved a BER of 0.07, which was based on PCA and a neural network with Bayesian regularization. These results are consistent with [10].

LSSVM and LDA with different feature extraction methods achieved BER of 0.08 and 0.11 for the Meningioma vs. Low-Grade Glial problem. Most of the results for this problem were in the interval from 0.15 to 0.25, and the mode of the distribution was under 0.2. The low error in MEN vs. LGG was also predicted by the CV results on the INTERPRET data. This result is consistent with the performances reported in Tate et al. in [7] on a three-class discrimination problem: MEN vs. Astrocytomas grade II (A2) vs. Aggressive tumors (AGG) (which is composed of GBM and MET). In that study, the confusion submatrix of MEN vs. A2 indicates no missclassifications between them. Identical results were obtained by Tate et al. in [8] when extending the three-class classifier to MEN vs. LGG vs. AGG.

The distribution of BER for MET vs. LGG had a clear trend towards the lower values (BER of 0.1), showing good performance for all the methods studied in this problem. PI combined with LDA, FLDA, MLP, or LSSVM classification methods obtained the best performance for the IT set. The CV estimations of the errors also indicated good performance by the classifiers. These results are also consistent with [10].

3.2 The pre-processing techniques

Eight out of 50 semi-automatically preprocessed test cases were misclassified at least once by the pairwise BDK classifiers (GBM vs. MET excluded). Also, ten out of 47 of the automatically preprocessed test cases were misclassified at least once by the same classifiers. Based on these rates, no differences were observed in the classification of automatic and semi-automatic pre-processed signals. The semi-automatic pre-processing pipeline applied to the larger part of the test dataset was consistent with the automatic pipeline applied on the training set. This is an important practical conclusion because it suggests the compatibility of different pre-processing software tools, either in an automatic or a semi-automatic fashion for automatic classification in CDSSs.

3.3 The Feature Extraction methods

All the feature extraction methods applied in this study were based on PR. Therefore, we could not make any comparison between PR and metabolite quantification approaches. Approaches that take advantage of the combination of different TE [26, 27, 30, 62–64] were not considered in order to ensure that results could be compared with previous analyses of this type of data [7, 8, 10, 12, 25, 28, 65–67]. Furthermore, although a feature extraction evaluation is not the aim of the present study and the setup of this study is not designed specifically for it, some effects of the different feature extraction methods are reported.

Figure 3 shows the box-whisker plot of the performance (BER) for each Feature Extraction (FE) method. GBM vs. MET classifiers are not included because of their large difference in performance with respect to the other classification problems. The

distributions of the results for all FE methods overlap, and no statistical differences were observed. Nevertheless, a noteworthy fact is the trend toward low values of the Peak Integration method compared to other methods. The study of Devos et al. [10] about the same four classes obtained similar performances when comparing full region of interest, peak regions and PI. In [12], Simonetti et al. compared, PCA, ICA, LCModel [67] and PI for feature extraction on Short TE MRSI data and they also obtained the best results with PI. In a single-center study, Opstad et al. [28] reported that the LCModel quantification obtained better results than PCA for two-step LDA classification. In Long TE spectra, Lukas et al. [11] observed a better performance using the full region of interest rather than using PI or peak region extraction. Finally, Menze et al. [13] and Luts et al. [68] obtained an improvement when PR approaches (e.g. ICA, PCA, binned peak region and WAV) were used in Short or Long TE instead of quantification approaches.

3.4 The classification methods

The diversity of methods used for classification is broad enough to have a good overview of the effect that this selection has on the performance of the classifiers. Figure 4 shows the box-whisker plot of the performance (BER) for each classification method. Analogously to the analysis of FE methods, GBM vs. MET classifiers are not included in the distributions because of their large differences in performance with respect to the other classification methods. As observed in Figure 4, the distributions overlap, but in general, lower results of BER were obtained using a BDK. In [25], BDK was used in PI values to discriminate over tumor grades and other tissues in the INTERPRET

multi-voxel dataset. The study of Devos et al. [10] observed similar performances of their LDA and LSSVM classifiers based on PI and evaluated by the area under the ROC curves. Tate et al. [7, 8] based their three-class classifiers on the LDA due to the ability of this method for projecting the results in a 2-dimensional space for visualization. Note that FLDA shows similar results when compared with the other methods in average; however, other methods like LSSVM and BDK might be preferable for some discrimination problems (e.g. GBM vs. LGG).

Finally, in Figure 2, we summarize and compare the BER estimation obtained by the CV for the INTERPRET training dataset and the IT consisting of the new eTUMOUR cases. Most of the results are in the $(BER(CV) < 0.2, BER(IT) < 0.3)$ region, except for the GBM vs. MET problem, which had a sparse distribution. The general trend in this region is indicated by the black-dashed line. This indicates an underestimation of the BER by the CV evaluation. The underestimation is typically observed in the PR challenges [54], and it is usually produced by the overfitting of the models on the training dataset and the estimation of the error with non-fully independent samples [69]. A noteworthy feature of our study is the evaluation of the predictive models using the new subsequently acquired multicenter test, that ensures the independence of the training and test sets. With respect to the GBM vs. MET results, they are scattered in regions of larger error. For this problem, some overestimations of the CV error are also observed. This may show the difficulty of the problem and the randomness in the results. The results obtained for the rest of the discrimination problems confirm the expected behaviour of the predictive models.

3.5 Use of the study for automatic validation of MRS entries in brain tumour datasets

An intuitive method to compare datasets of signals is the visual inspection of their prototypical patterns. Figure 5 shows plots of the unimodal prototypes of the Short TE spectra for the four tumour groups of the training and test datasets. Each prototype is represented by the unsmoothed mean function and the mean function $\pm$ the standard deviation function. The view is zoomed in the $[0.5, 4.1]ppm$ region used in our experiments. The observed resonances correspond to the main compounds reported in page 8 of Section 1. In general, the training and test prototype patterns for GBM, MET and LGG are close to each other, whereas the MEN prototype differs visually more. This may be because of a higher standard deviation on the test dataset around the $3.21ppm$ peak with respect to the training dataset. Besides, the variation around the $2.2ppm$ is higher in the test-set mean than in the training one.

A practical result of this study is that cases that are repeatedly misclassified by the different techniques can be flagged as being susceptible of revision for possible problems in voxel positioning, acquisition artifact, normal-tissue contamination, or limitation in the classification methodology (e.g. patterns replicated in non-tumoral diseases, atypical MRS patterns and underrepresented tumor subtypes). In this way, even in the absence of biopsy, PR techniques can contribute to the automatic validation of cases, assisting the specialists on the detection of potential source of errors in the biomedical data acquired from patients.

Figures 6 and 7 show some eTUMOUR misclassified cases which may be interest-

ing to review. The eTUMOUR case et2274 was diagnosed by the original pathologist as oligodendroglioma 9450/3 (grade II, WHO), although a comment was added to the free text section of the eTUMOUR database (eTDB) making reference to the presence of areas of anaplastic oligodendroglioma (grade III, WHO). Still, the final diagnosis proposed was grade II oligodendroglioma. The voxel allocation was carried out following the eTUMOUR acquisition protocol. The ML pattern is uncommon, as the high 0.9 and 1.3*ppm* resonances show. The disappearance of these resonances at Long TE (136 ms) discards a significant necrotic contribution (results not shown, but see [30]). This pattern has been observed before [30], for example in the INTERPRET cases I0450 (oligoastrocytoma) and I0179 (oligodendroglioma), which are also misplaced in the Short TE latent space of the INTERPRET Decision-Support System (DSS) 2.0 (`http://azizu.uab.es/INTERPRET`). In summary, et2274 seems to behave as a class outlier and its consistent missclassification in our analysis may be sampling precisely that. The eTUMOUR case et2206 was originally diagnosed as oligoastrocytoma 9382/3 (grade II, WHO), but there were some discrepancies regarding the glial subtype on the validation done by the pathological committee. It was misclassified by every MET vs. LGG classifier, and also by some GBM vs. LGG and MEN vs. LGG classifiers. Its ML pattern at Short TE is also uncommon, having relatively large 0.9, 1.3 and 2.8*ppm* peaks that are reduced at Long TE (results not shown), which suggests, as well, a non-necrotic origin. The eTUMOUR case et2349 is a GBM without clear visible ML, which was misclassified in every classification problem. The review of the experts did not indicate problems in the location of the voxel, being this mainly positioned in the highly cellular part of the tumour. The eTUMOUR case et2197 is a MET

with possible MRS pattern contribution from normal brain parenchyma, as it could be deduced by the relative difference of size between the voxel used for acquisition and the small brain lesion. Its pattern shows similar Cho and Cr peak heights and relatively high NAA at $2ppm$). However, the appearance of high Lac/ML at $1.3ppm$ at the same time suggests abnormality. Nonetheless, it is clearly an uncommon spectral pattern for a MET.

4 Conclusions

This study describes a multiproject-multicenter evaluation of automated brain tumor classifiers using single-voxel Short TE MR spectra. To our knowledge, there is no previous work that evaluates predictive models trained with data acquired from a multicenter project using a new independent test set subsequently acquired from partly different centers. Classifiers were trained with cases acquired by six centers during the 2000-2002 period. They were tested with posterior cases acquired by eight institutions during the 2004-2007 period. This strategy provides a view that is close to a real environment where similar classifiers, integrated in a Clinical Decision-Support System (CDSS), may be used in multiple hospitals to assist in the diagnosis of new cases.

Our major conclusion is that accurate classification of those new cases is feasible using data acquired in different hospitals, different instrumentation, but similar acquisition protocols. Specifically, in our experiments, classifiers developed from the INTERPRET dataset seem to be robust enough for predictive classification of prospec-

tive cases from eTUMOUR.

The pairwise discrimination between Glioblastoma, Meningioma, Metastasis, and Low-grade Glial achieved accuracies of around 90%. However, the discrimination of Glioblastoma and Metastasis did not achieve a result better than 78% accuracy. Our results consolidate the conclusions of previous studies on automatic brain tumor classification using MRS but with multiproject-multicenter data for training and subsequent test.

A well-defined protocol for the acquisition of MRS (e.g. spectral parameters and voxel localization), and the application of quality controls to MRS spectra should allow the reproducibility of such classification rules and the successful use of Decision-Support Systems (DSSs) in clinical environments.

The methodology provided in the present study may also be of use as "automatic flaggers" to help in the quality control of cases during the eTUMOUR multicenter project and beyond. **The approach used in this work could be of use for pediatric brain tumour related studies [70] aimed at providing predictive information to pediatric neurosurgeons.**

Hence, the conclusions obtained in this study are directly applicable to several of the tasks associated to a CDSS development for brain tumor diagnosis and prognosis and its deployment in clinical environments.

Tables and Figures

Table 1: Number of Training (INTERPRET) and Test (eTUMOUR) cases per class used in the study. Short TE ^{1}H MRS data were acquired according to a consensus protocol during the INTERPRET (2000-2002) and eTUMOUR (2004-2009) projects.

Class	INTERPRET	eTUMOUR
GBM	84	28
MEN	57	17
MET	37	32
LGG	33	20
	211	97

Table 2: Breakdown of cases per manufacturer included in the Training (INTERPRET) and Test (eTUMOUR) datasets.

Manufacturer	INTERPRET (%)	eTumour (%)
GE	53.1	54.6
Siemens	6.6	12.4
Philips	40.3	33.0

Table 3: Percentage of cases per acquisition center included in the Training (INTERPRET) and Test (eTUMOUR) datasets. Last row indicates the percentage of training cases that belong to centers that did not produce eTUMOUR cases, and the percentage of test cases that belong to centers that did not acquired training data for INTERPRET.

CENTERS	Training from INTERPRET (%)	Test from eTUMOUR (%)
UMC NIJMEGEN	2.8	1.0
ST. GEORGE'S HOSPITAL	27.0	18.6
MEDICAL UNIV. OF LODZ	3.8	10.3
FLENI	1.9	6.2
IDI-BELLVITGE	40.3	
CENTRE DE DIAG. PEDRALBES	24.2	
CENTRE DE DIAG. PEDRALBES + IAT		28.9
IDI-BADALONA		17.5
UNIV. DE VALENCIA		16.5
HOSPITAL SANT JOAN DE DEU		1.0
% OF CASES OF PROJECT EXCLUSIVE CENTERS	40.3	35.1

Table 4: Best results obtained for the six pairwise classification problems. The ERR and BER estimation based on CV over the INTERPRET data and based on the eTUMOUR IT set are shown. The columns of the table are: id, identification of the classifier; Task: classification problem defined by the classes to discriminate by the classifiers; Features: acronym of the feature extraction method, Classif: acronym of the classification method, CV: results estimated by means of a 10-fold CV in the INTERPRET database, IT: results estimated by means of the independent test, with the INTERPRET database as training and the eTUMOUR dataset as test, ERR: error rate, and BER: Balanced Error Rate. []: interval within every result falls.

				CV		IT	
Task	id	Features	Classif	ERR	BER	ERR	BER
GBM vs. MEN	1.6	190	MLP	0.06	0.07	0.07	0.09
GBM vs. MET	2.13	PI	LDA	0.33	0.40	0.22	0.21
GBM vs. LGG	3.16	PI	LSSVM	0.12	0.18	0.08	0.09
MEN vs. MET	4.21	PCA	MLP	0.05	0.05	0.06	0.07
MEN vs. LGG	5.10	ICA	LSSVM	0.08	0.09	0.08	0.08
MET vs. LGG	6.13,21,25-26	PI	LDA/FLDA/MLP/LSSVM	[0.01,0.04]	[0.01,0.04]	0.06	0.07

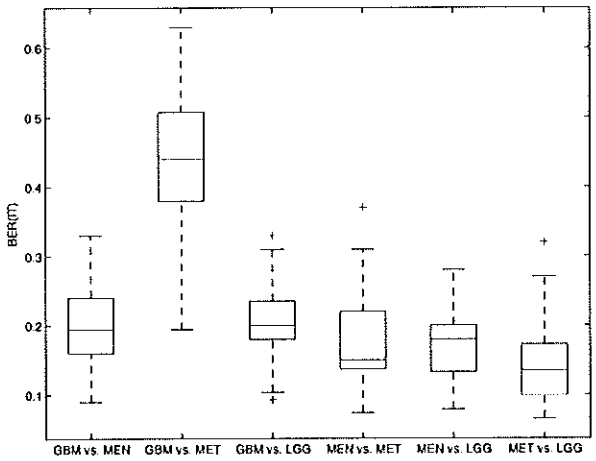

BER(IT)
0.6
0.5
0.4
0.3
0.2
0.1
GBM vs. MEN
GBM vs. MET
GBM vs. LGG
MEN vs. MET
MEN vs. LGG
MET vs. LGG

Figure 1:

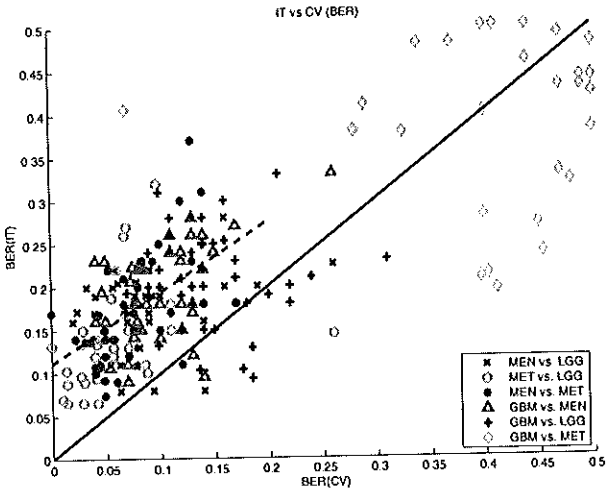

iT vs CV (BER)
BER(iT)
BER(CV)
MEN vs. LGG
MET vs. LGG
MEN vs. MET
GBM vs. MEN
GBM vs. LGG
GBM vs. MET

Figure 2:

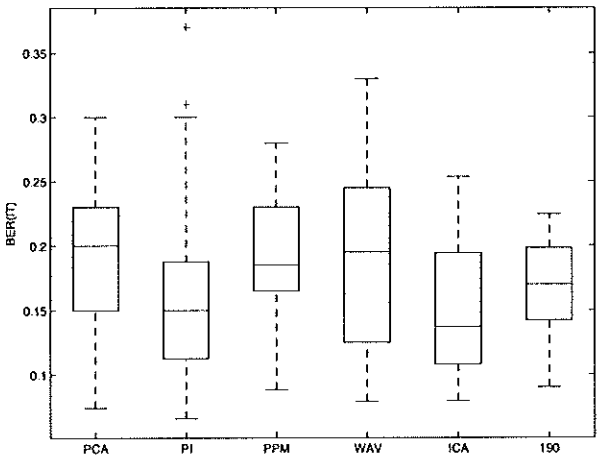


Figure 3:

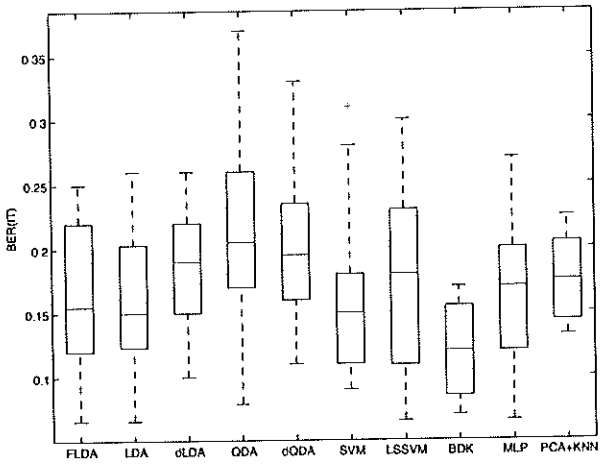

BER(rr)
0.35
0.3
0.25
0.2
0.15
0.1
FLDA
LDA
dLDA
QDA
dQDA
SVM
LSSVM
BDK
MLP
PCA+KNN

Figure 4:

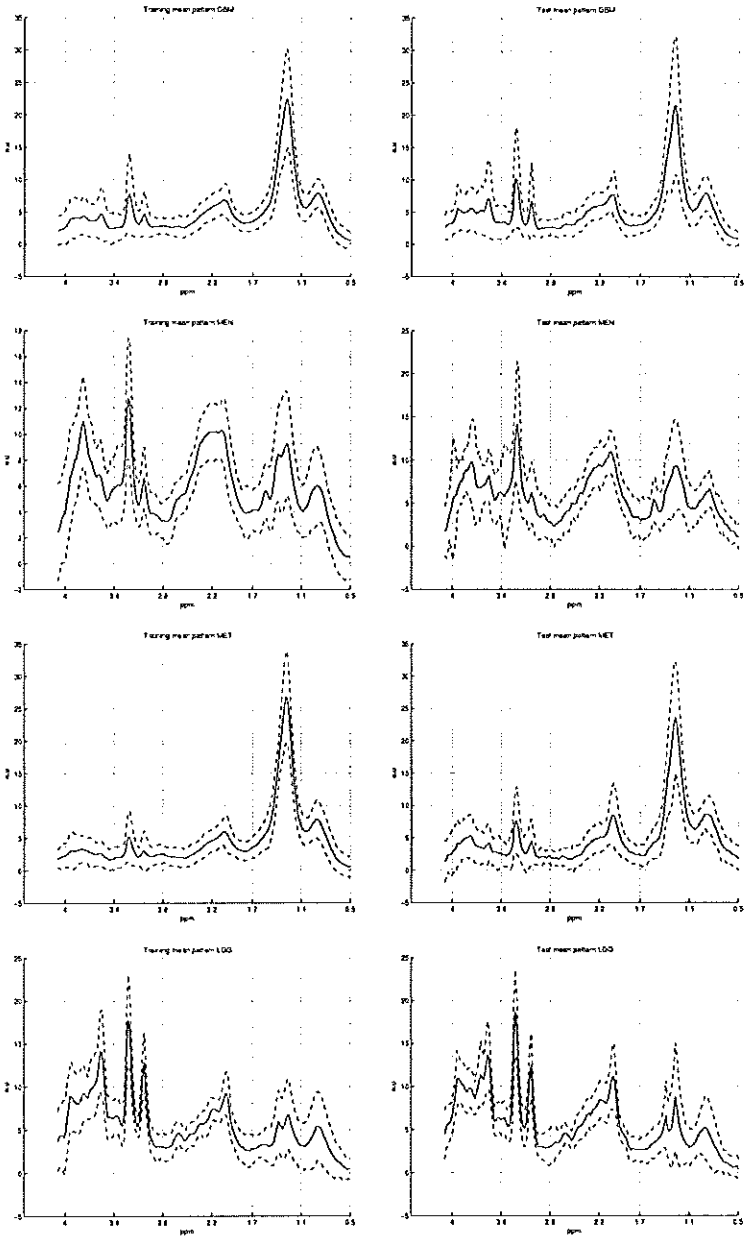
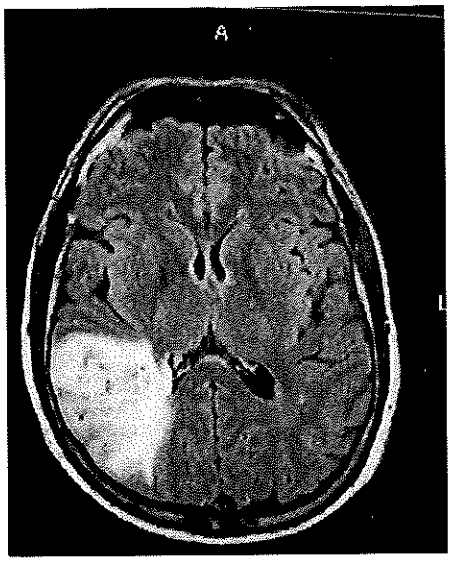
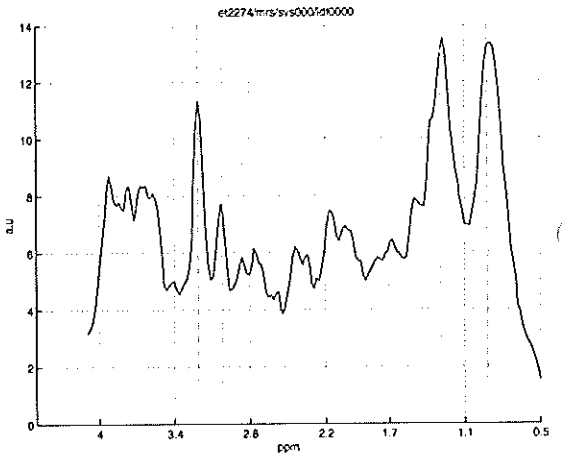


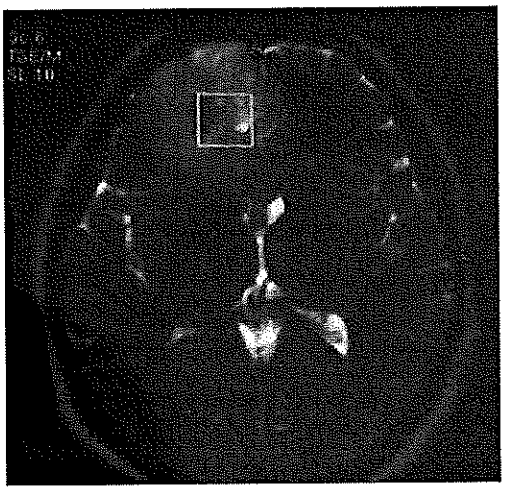
Figure 5:

et2274 (T1-weighted), LGG (OD)





et2206 (T2-weighted), LGG (OA or OD)



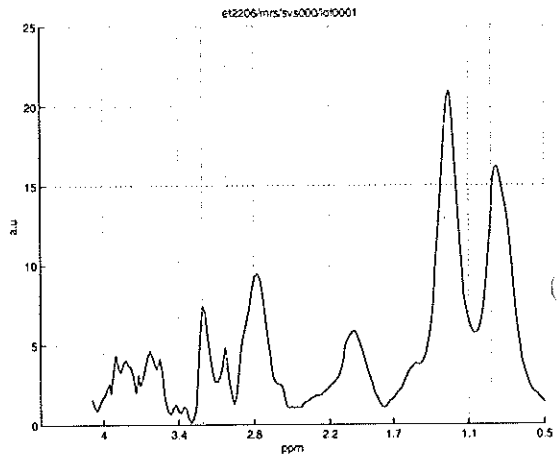
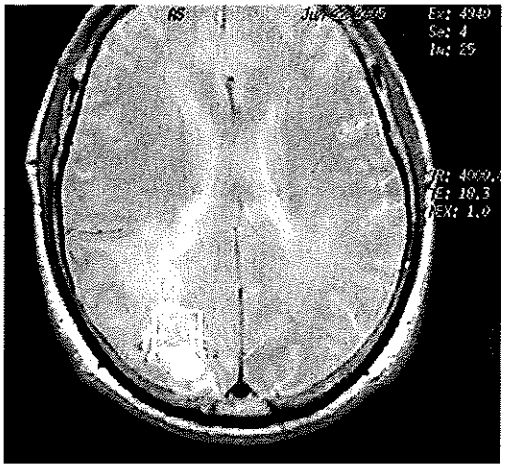
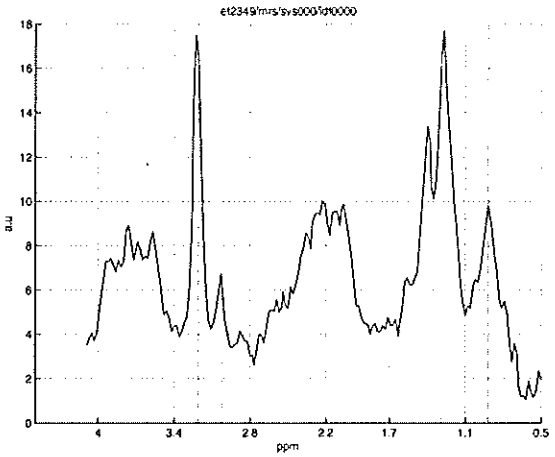


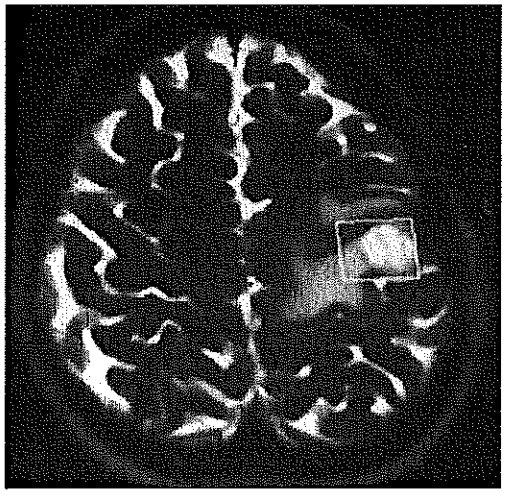
Figure 6:

et2349 (PD-weighted), GBM





et2197 (T2-weighted), MET



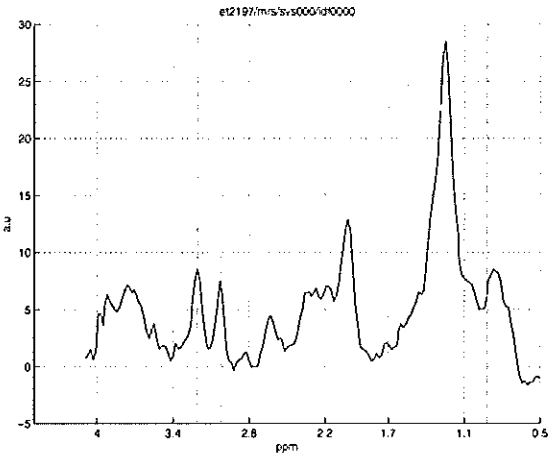


Figure 7:

Figure Legends

- Figure 1: Box-whisker plots of the performance for each problem in the eTUMOUR dataset (based on the detailed list of results included in Section 1 of the Supplementary Material). Performance is measured in BER. The box indicates the region between the lower ($X_{0.25}$) and the upper ($X_{0.75}$) quartiles. The horizontal line inside the box indicates the median of the distribution, and the vertical lines (the "whiskers") extend to at most 1.5 times the box width. Any outlier of the distribution is displayed with a cross (+).

- Figure 2: Scatter plot of the performance measured in BER estimated by the IT set consisting of new eTUMOUR cases and the BER estimated by the CV using the INTERPRET cases. $BER(IT) = BER(CV)$ is represented by the solid-blue line and the trend of the ($BER(CV) < 0.2$, $BER(IT) < 0.3$) region is indicated by the black-dashed line.

- Figure 3: Box-whisker plots of the performance for each Feature Extraction method in the eTUMOUR dataset. Performance is measured in BER and the box-whisker characteristics are the same as in Figure 1.

- Figure 4: Box-whisker plots of the performance for each classification method in the eTUMOUR dataset. Performance is measured in BER and the box-whisker characteristics are the same as in Figure 1.

- Figure 5: Unimodal prototypes of the Short TE spectra for the four tumour groups of the training and test datasets. Each prototype is represented by the unsmoothed mean function and the mean $\pm$ the standard deviation function. The

view is zoomed in the $[0.5, 4.1]ppm$ region used in our experiments.

- Figure 6: Potential outliers (1/2) detected as a consequence of this study. Case numbering corresponds to eTUMOUR database (www.etumour.net) entries). For each case, the reference image and voxel location is shown on the left, and the region of interest of the real part of the Short TE spectrum is shown on the right. For an easier visualization of the spectrum, vertical dashed lines indicates the position of the main resonances: Cho ($3.21ppm$), Cr (3.02), NAA ($2.01ppm$), L1 ($1.29ppm$), L2 ($0.92ppm$).

- Figure 7: Potential outliers (2/2) detected as a consequence of this study. Figure characteristics are the same as in Figure 6.

List of abbreviations

A2	Astrocytomas grade II
AGG	Aggressive tumors
Ala	Alanine
BER	Balanced Error Rate
BDK	Bi-directional Kohonen Networks
CDSS	Clinical Decision-Support System
CDSSs	Clinical Decision-Support Systems
CDVC	Clinical Data Validation Committee
Cho	Choline
CNS	Central Nervous System

Cr	Creatine
CV	Cross Validation
dLDA	Linear Discriminant Analysis with diagonal covariance matrix
dQDA	Quadratic Discriminant Analysis with diagonal covariance matrix
DSS	Decision-Support System
DSSs	Decision-Support Systems
ERR	Error Rate
eTDB	eTUMOUR database
eTUMOUR	the eTUMOUR project
FE	Feature Extraction
FID	Free Induction Decay
FLDA	Fisher's rank-reduced version of LDA
FFT	Fast Fourier Transform
GBM	Glioblastoma
GE	General Electric
Gly	Glycine
Glx	Glutamate/Glutamine
HEALTHAGENTS	the HEALTHAGENTS EC project
HLSVD	Hankel-Lanczos Singular Value Decomposition
HSVD	Hankel Singular Value Decomposition
ICA	Independent Component Analysis
INTERPRET	the INTERPRET project

IT	Independent Test
jMRUI	java Magnetic Resonance User Interface
Lac	Lactate
LDA	Linear Discriminant Analysis
LGG	Low-Grade Glial
LSSVM	Least-Squares Support Vector Machine
MEN	Low-grade Meningioma
MET	Metastasis
mI	myo-Inositol
ML	Mobile Lipids
MLP	Multilayer Perceptron
MM	Macromolecules
MR	(Nuclear) Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MRSI	Magnetic Resonance Spectroscopic Imaging
NAA	N-Acetyl Aspartate
PCA	Principal Component Analysis
PCA-KNN	K-nearest neighbours and local feature reduced by PCA
PI	Peak Integration
PPM	peak height of typical resonances
PR	Pattern Recognition

PRESS	Point-Resolved Spectroscopic Sequence
QDA	Quadratic Discriminant Analysis
SNR	signal-to-noise ratio
SNV	Standard Normal Variate
STEAM	Stimulated Echo Acquisition Mode sequence
SV	Single-voxel
SVM	Support Vector Machines
Tau	Taurine
TE	Echo Time
TR	Repetition Time
WAV	Wavelet transform
WHO	World Health Organization

Acknowledgments

We would like to thank the INTERPRET and eTUMOUR partners for providing data, particularly, Carles Majós (IDI-Bellvitge), John Griffiths and Franklyn Howe (SGUL), Arend Heerschap (RU), Witold Gajewicz (MUL), Jorge Calvar (FLENI), and Antoni Capdevila (H. de Sant Joan de Déu). This work was partially funded by the European Commission: eTUMOUR (contract no. FP6-2002-LIFESCIHEALTH 503094), the HEALTHAGENTS EC project (HEALTHAGENTS) (contract no. FP6-2005-IST 027213), BIOPATTERN (contract no. FP6-2002-IST 508803). The authors appreciate the suggestions from the reviewers that have improved the discussion presented in

this work. We also thank the following for their contributions: Programa de Apoyo a la Investigación y Desarrollo, PAID-00-06 UPV; Research Council KUL: GOA-AMBioRICS, Centers-of-excellence optimisation; Belgian Federal Government: DWTC, IUAPV P6/04 (DYSCO 2007-2011); The following participants acknowledge the following: JVR acknowledges to Programa Torres Quevedo from Ministerio de Educación y Ciencia, co-founded by the European Social Fund (PTQ05-02-03386). JL is a PhD student supported by an IWT grant. DM is supported by the Ministerio de Educación y Ciencia del Gobierno de España for a Ramon y Cajal 2006 Contract. BC and CA gratefully acknowledge the Ministerio de Educación y Ciencia del Gobierno de España (BC: SAF2004-06297 and SAF2007-6547; CA: SAF2005-03650). CIBER-BBN is an initiative of the "Instituto de Salud Carlos III" (ISCIII), Spain.

References

[1] Howe FA, Opstad KS (2003) 1H MR spectroscopy of brain tumours and masses. NMR Biomed 16(3): 123–131

[2] Galanaud D, Nicoli F, Chinot O, Confort-Gouny S, Figarella-Branger D, Roche P, Fuentes S, Le Fur Y, Ranjeva JP, Cozzone PJ (2006) Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy. Magn Reson Med 55(6): 1236–1245

[3] Arnold DL, De Stefano N (1997) Magnetic resonance spectroscopy in vivo: applications in neurological disorders. Ital J Neurol Sci 18(6): 321–329

[4] Poptani H, Kaartinen J, Gupta RK, Niemitz M, Hiltunen Y, Kauppinen RA (1999)

Diagnostic assessment of brain tumours and non-neoplastic brain disorders in vivo using proton nuclear magnetic resonance spectroscopy and artificial neural networks. J Cancer Res Clin Oncol 125(6): 343–349

[5] Moller-Hartmann W, Herminghaus S, Krings T, Marquardt G, Lanfermann H, Pilatus U, Zanella FE (2002) Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. Neuroradiology 44(5): 371–381

[6] Hagberg G (1998) From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. NMR Biomed 11(4-5): 148–156

[7] Tate AR, Majos C, Moreno A, Howe FA, Griffiths JR, Arús C (2003) Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. Magn Reson Med 49(1): 29–36

[8] Tate AR, Underwood J, Acosta DM, Julia-Sape M, Majos C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabanas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Remy C, Heerschap A, Watson D, Griffiths JR, Arús C (2006) Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. NMR Biomed 19(4): 411–434

[9] González-Vélez H, Mier M, Julià-Sapé M, Arvanitis T, García-Gómez J, Robles M, Lewis P, Dasmahapatra S, Dupplaw D, Peet A, Arús C, Celda B, Van Huf-

fel S, Lluch-Ariet M (2007) HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis. Applied Intelligence [Epub ahead of print]

[10] Devos A, Lukas L, Suykens JAK, Vanhamme L, Tate AR, Howe FA, Majos C, Moreno-Torres A, van der Graaf M, Arús C, Van Huffel S (2004) Classification of brain tumours using short echo time 1H MR spectra. J Magn Reson 170(1): 164–175

[11] Lukas L, Devos A, Suykens JAK, Vanhamme L, Howe FA, Majós C, Moreno-Torres A, Graaf MVD, Tate AR, Arús C, Huffel SV (2004) Brain tumor classification based on long echo proton MRS signals. Artif Intell Med 31: 73–89

[12] Simonetti AW, Melssen WJ, Szabo de Edelenyi F, van Asten JJA, Heerschap A, Buydens LMC (2005) Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. NMR Biomed 18(1): 34–43

[13] Menze BH, Lichy MP, Bachert P, Kelm BM, Schlemmer HP, Hamprecht FA (2006) Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors. NMR Biomed 19(5): 599–609

[14] Potts HWW, Wyatt JC, Altman DG (2001) Challenges in Evaluating Complex Decision Support Systems: Lessons from Design-a-Trial. In AIME '01: Proceedings of the 8th Conference on AI in Medicine in Europe, pp. 453–456. Springer-Verlag, London, UK

[15] Lisboa PJ, Taktak AFG (2006) The use of artificial neural networks in decision support in cancer: a systematic review. Neural Netw 19(4): 408–415

[16] Anagnostou T, Remzi M, Djavan B (2003) Artificial neural networks for decision-making in urologic oncology. European Urology 43(6): 596–603

[17] Perner P (2006) Intelligent data analysis in medicine-recent advances. Artif Intell Med 37(1): 1–5

[18] INTERPRET Consortium (Accessed: 28 April 2008) Interpret web site. http://azizu.uab.es/INTERPRET

[19] Julia-Sape M, Acosta D, Mier M, Arús C, Watson D (2006) A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. Magn Reson Mater Phy 19(1): 22–33

[20] van der Graaf M, Julia-Sape M, Howe FA, Ziegler A, Majos C, Moreno-Torres A, Rijpkema M, Acosta D, Opstad KS, van der Meulen YM, Arus C, Heerschap A (2008) MRS quality assessment in a multicentre study on MRS-based classification of brain tumours. NMR Biomed 21(2): 148–158

[21] Devos A (2005) Quantification and classification of Magnetic Resonance Spectroscopy data and applications to brain tumour recognition. Ph.D. thesis, Faculty of Engineering, K.U.Leuven

[22] Simonetti AW, Melssen WJ, van der Graaf M, Postma GJ, Heerschap A, Buydens LMC (2003) A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy. Anal Chem 75(20): 5352–5361

[23] Devos A, Simonetti AW, van der Graaf M, Lukas L, Suykens JAK, Vanhamme L, Buydens LMC, Heerschap A, Van Huffel S (2005) The use of multivariate

MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. J Magn Reson 173(2): 218–228

[24] Laudadio T, Martinez-Bisbal M, Celda B, Van Huffel S (2007) Fast nosological imaging using canonical correlation analysis of brain data obtained by two-dimensional turbo spectroscopic imaging. NMR Biomed 21(4): 311–321

[25] Melssen W, Wehrens R, Buydens L (2006) Supervised Kohonen networks for classification problems. Chemometrics and Intelligent Laboratory Systems 83(2): 99–113

[26] Martinez-Bisbal MC, Celda B, Marti-Bonmati L, Ferrer P, Revert-Ventura AJ, Piquer J, Molla E, Arana R, Dosda-Munoz R (2002) The contribution of Magnetic Resonance Spectroscopy for the classification of high grade glial tumours. The predictive value of macromolecules. Revista de Neurología 34(309-313)

[27] Martinez-Bisbal MC, Ferrer-Luna R, Martinez-Granados B, Monleón D, Esteve V, Piquer J, Revert AJ, Mollá E, Martí-Bonmatí L, Celda B (2005) Glial tumours grading by a combination of (1)H MR short and medium echo time single voxel located by spectroscopic imaging. Magnetic Resonance Materials in Physics 18(S1): S68

[28] Opstad KS, Ladroue C, Bell BA, Griffiths JR, Howe FA (2007) Linear discriminant analysis of brain tumour (1)H MR spectra: a comparison of classification using whole spectra versus metabolite quantification. NMR Biomed 20(8): 763–770

[29] eTumour Consortium (2003) eTumour: Web accessible MR Decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data. Technical report, FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC, http://www.etumour.net, Accessed: 28 April 2008

[30] García-Gómez JM, Tortajada S, Vidal C, Julia-Sape M, Luts J, Moreno-Torres À, Van Huffel S, Arús C, Robles M (2008) The effect of combining two echo times in automatic brain tumor classification by MRS. NMR in Biomedicine 21: Accepted for publication

[31] Kleihues P, Burger PC, Scheithauer BW (1993) The new WHO classification of brain tumours. Brain Pathol 3(3): 255–268

[32] Klose U (1990) In vivo proton spectroscopy in presence of eddy currents. Magn Reson Med 14(1): 26–30

[33] Naressi A, Couturier C, Castang I, de Beer R, Graveron-Demilly D (2001) Java-based graphical user interface for MRUI, a software package for quantitation of in vivo/medical magnetic resonance spectroscopy signals. Comput Biol Med 31(4): 269–286

[34] Cabanes E, Confort-Gouny S, Le Fur Y, Simond G, Cozzone PJ (2001) Optimization of residual water signal removal by HLSVD on simulated short echo time proton MR spectra of the human brain. J Magn Reson 150(2): 116–125

[35] Hoch JC, Stern AS (1996) NMR Data Processing. John Wiley and Sons, Inc., New York, NY

[36] Preul MC, Caramanos Z, Collins DL, Villemure JG, Leblanc R, Olivier A, Pokrupa R, Arnold DL (1996) Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. Nat Med 2(3): 323–325

[37] Burges CJ (2004) Geometric Methods for Feature Extraction and Dimensional Reduction: A Guided Tour. Technical report, Microsoft Research, University of Toronto

[38] Fukunaga K (1990) Introduction to statistical pattern recognition (2nd ed.). Academic Press Professional, Inc., San Diego, CA, USA

[39] Comon P (1994) Independent component analysis, a new concept? Signal Processing 36(3): 287–314

[40] J-F Cardoso, A Souloumiac (1993) Blind beamforming for non Gaussian signals. IEE Proceedings-F 140(6): 362–370

[41] Daubechies I (1992) Ten Lectures on Wavelets (CBMS - NSF Regional Conference Series in Applied Mathematics). Soc for Industrial & Applied Math

[42] Panagiotacopulos N, Lertsuntivit S, Savidge L, Lin A, Shic F, Ross B (2000) Wavelet Analysis of Brain Tumors in Clinical MRS. In Advances in Physics, Electronics and Signal Processing Applications, pp. 290–296

[43] Krzanowski WJ, editor (1988) Principles of multivariate analysis: a user's perspective. Oxford University Press, Inc., New York, NY, USA

[44] Fisher RA (1925) Statistical methods for research workers. Oliver and Boyd, Edinburgh, UK

[45] Vapnik V (1995) The Nature of Statistical Learning Theory. Springer, N.Y.

[46] Suykens JAK, Vandewalle J (1999) Least Squares Support Vector Machine Classifiers. Neural Process Lett 9(3): 293–300

[47] Rosenblatt F (1958) The Perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review 65(6): 386–408

[48] Melssen W, Ustun B, Buydens L (2007) SOMPLS: A supervised self-organising map–partial least squares algorithm for multivariate regression problems. Chemometrics and Intelligent Laboratory Systems 86(1): 102–120

[49] Valentini G, Dietterich TG (2004) Bias-Variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods. Journal of Machine Learning Research 5: 725–775

[50] Hastie T, Tibshirani R, Friedman JH (2001) The Elements of Statistical Learning. Springer

[51] Duda R, Hart P, Stork D (2001) Pattern Classification. John Wiley and Sons, inc.

[52] Van Gestel T, JAK Suykens, Lanckriet G, Lambrechts A, De Moor B, Vandewalle J (2002) Bayesian Framework for Least Squares Support Vector Machine Classifiers, Gaussian Processes and Kernel Fisher Discriminant Analysis. Neural Computation 14: 1115–1147

[53] MacKay DJC (1992) Bayesian Interpolation. Neural Computation 4(3): 415–447

[54] Guyon I, Alamdari ARSA, Dror G, Buhmann JM (2006) Performance Prediction Challenge. In IJCNN '06. International Joint Conference on Neural Networks, pp. 1649–1656

[55] Ishimaru H, Morikawa M, Iwanaga S, Kaminogo M, Ochi M, Hayashi K (2001) Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. Eur Radiol 11(9): 1784–1791

[56] Opstad KS, Murphy MM, Wilkins PR, Bell BA, Griffiths JR, Howe FA (2004) Differentiation of metastases from high-grade gliomas using short echo time 1H spectroscopy. J Magn Reson Imaging 20(2): 187–192

[57] Law M, Cha S, Knopp EA, Johnson G, Arnett J, Litt AW (2002) High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. Radiology 222(3): 715–721

[58] Burtscher IM, Skagerberg G, Geijer B, Englund E, Stahlberg F, Holtas S (2000) Proton MR spectroscopy and preoperative diagnostic accuracy: an evaluation of intracranial mass lesions characterized by stereotactic biopsy findings. AJNR Am J Neuroradiol 21(1): 84–93

[59] Laudadio T, Luts J, Martinez-Bisbal M, Celda B, Huffel SV (2008) Differentiation between brain metastasis and glioblastoma using MRI and two-dimensional turbo spectroscopic imaging data. In Proc. of the 4th European Medical and Biomedical Engineering congress, p. Accepted for publication.

[60] Hochberg Y, Tamhane AC (1987) Multiple comparison procedures. John Wiley & Sons, Inc., New York, NY, USA

[61] Celda B, Monleon D, Martinez-Bisbal MC, Esteve V, Martinez-Granados B, Pinero E, Ferrer R, Piquer J, Marti-Bonmati L, Cervera J (2006) MRS as endogenous molecular imaging for brain and prostate tumors: FP6 project "eTUMOR". Adv Exp Med Biol 587: 285–302

[62] Tortajada S, García-Gómez JM, Vidal C, Arús C, Julià-Sapé M, Moreno A, Robles M (2006) Improved classification by pattern recognition of brain tumours combining long and short echo time 1H-MR spectra. In SpringerLink, editor, Book of Abstracts ESMRMB 2006 - Supplement 1, Journal Magnetic Resonance Materials in Physics, Biology and Medicine, volume 19, pp. 168–169

[63] García-Gómez JM, Tortajada S, Vicente J, Sáez C, Castells X, Luts J, Julià-Sapé M, Juan-Císcar A, Van Huffel S, Barcelo A, Ariño J, Arús C, Robles M (2007) Genomics and Metabolomics Research for Brain Tumour Diagnosis Based on Machine Learning. In IWANN: Lecture Notes in Computer Sicences, volume 4507/2007, pp. 1012–1019

[64] McIntyre DJO, Charlton RA, Markus HS, Howe FA (2007) Long and short echo time proton magnetic resonance spectroscopic imaging of the healthy aging brain. J Magn Reson Imaging 26(6): 1596–1606

[65] Majos C, Julia-Sape M, Alonso J, Serrallonga M, Aguilera C, Acebes JJ, Arús C, Gili J (2004) Brain tumor classification by proton MR spectroscopy: comparison

of diagnostic accuracy at short and long TE. AJNR Am J Neuroradiol 25(10): 1696–1704

[66] Julia-Sape M, Acosta D, Majos C, Moreno-Torres A, Wesseling P, Acebes JJ, Griffiths JR, Arús C (2006) Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. J Neurosurg 105(1): 6–14

[67] Provencher SW (2001) Automatic quantitation of localized in vivo 1H spectra with LCModel. NMR Biomed 14(4): 260–264

[68] Luts J, Poullet JB, Garcia-Gomez JM, Heerschap A, Robles M, Suykens JAK, Van Huffel S (2008) Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. Magn Reson Med 60(2): 288–298

[69] Bishop CM (2006) Pattern Recognition and Machine Learning (Information Science and Statistics). Springer

[70] Davies N, Wilson M, Harris L, Natarajan K, Lateef S, Macpherson L, Sgouros S, Grundy R, Arvanitis T, Peet A (2008) Identification and characterisation of childhood cerebellar tumours by in vivo proton MRS. NMR Biomed 21(8): 908–918