Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel proton MRS ⋆

A. Vellido, [a,*] E. Romero, [a] M. Julià-Sapé, [b,c,d] C. Majós, [e,b] À. Moreno-Torres, [b] J. Pujol, [f] C. Arús [c,b,d]

[a] *Dept. de Llenguatges i Sistemes Informàtics - Universitat Politècnica de Catalunya. C./ Jordi Girona, 1-3. 08034, Barcelona, Spain.*

[b] *Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), 08193, Cerdanyola del Vallès, Spain.*

[c] *Departament de Bioquímica i Biología Molecular (BBM). Unitat de Biociències Universitat Autònoma de Barcelona (UAB), 08193, Cerdanyola del Vallès, Spain.*

[d] *Institut de Biotecnologia i de Biomedicina (IBB), UAB, 08193, Cerdanyola del Vallès, Spain*

[e] *Institut de Diagnòstic per la Imatge (IDI), CSU de Bellvitge, 08907, L'Hospitalet de Llobregat, Barcelona, Spain*

[f] *Institut d'Alta Tecnologia-PRBB, CRC Corporació Sanitària, Hospital del Mar, 08003, Barcelona, Spain*

Abstract

This paper investigates methods for the accurate and robust differentiation of metastases from glioblastomas on the basis of single-voxel proton MRS (SV-^{1}H-MRS) information. SV-^{1}H-MR spectra from a total of 109 patients (78 glioblastomas and 31 metastases) from the multi-center, international INTERPRET database, plus a test set of 40 patients (30 glioblastomas and 10 metastases) from three different centers of the Barcelona (Spain) metropolitan area, were analyzed using a robust method for feature (spectral frequency) selection coupled with a linear-in-the-parameters single-layer perceptron classifier. A parsimonious selection of 5 frequencies yielded, for the test set, an area under the ROC curve (AUC) of 0.86, and an area under the convex hull of the ROC curve (AUCCH) of 0.91. These accurate results for the discrimination between glioblastomas and metastases were, moreover, obtained using a small number of frequencies that are amenable of metabolic interpretation, which should ease their use as diagnostic markers. Importantly, the prediction can be expressed as a simple formula based on a linear combination of these frequencies. As a result, new cases could be straightforwardly predicted by integrating this formula into a computer-based medical decision support system. This work also shows that the combination of spectra acquired at different echo

times (short echo time, 20-32 ms, and long echo time, 135-144 ms) is key to the successful discrimination between glioblastomas and metastases from SV-^{1}H-MRS.

Key words: SV-^{1}H-MRS, feature selection, high-grade malignant tumors, metastases, glioblastomas, pattern recognition, medical decision support system

---

Abbreviations

- AUC: Area Under the (ROC) Curve
- AUCCH: Area Under the (ROC) Curve Convex Hull
- BER: Balanced Error Rate
- Cr: Creatine
- CSI: Chemical Shift Imaging
- FN: False Negative
- FP: False Positive
- FS: Feature Selection
- GABA: γ-Aminobutyric Acid
- Glu/Gln: Glutamate/Glutamine
- LDA: Linear Discriminant Analysis
- LET: Long Echo Time
- MDSS: Medical Decision Support System
- MLP: Multi-Layer Perceptron
- NAA: N-Acetyl Aspartate
- ROC: Receiver Operating Characteristic
- SET: Short Echo Time
- SLP: Single-Layer Perceptron
- STEAM: STimulated Echo Acquisition Mode
- SVM: Support Vector Machine
- TN: True Negative
- TP: True Positive

---

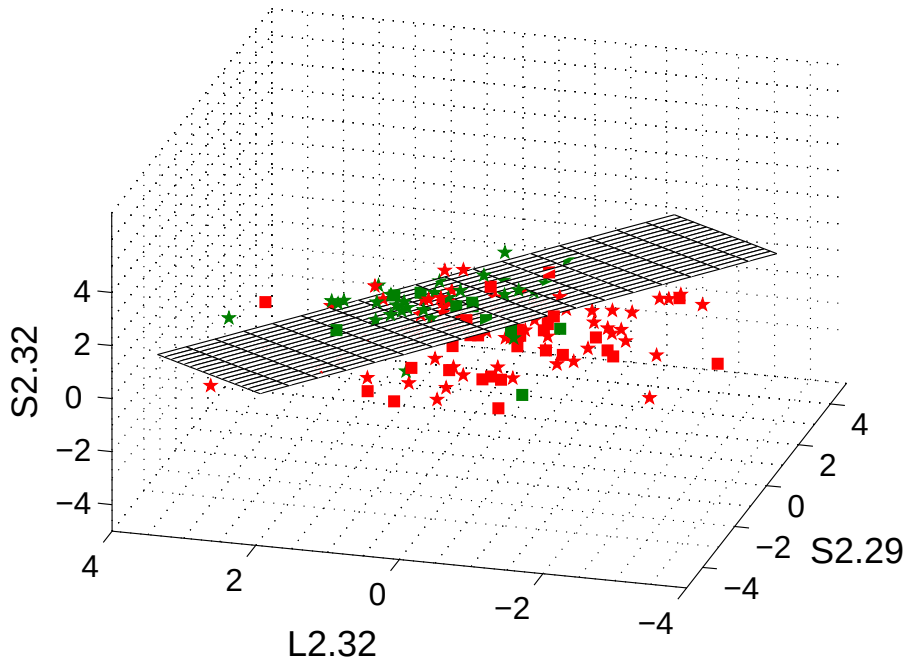
* Short title: *Differentiation of Metastases from Glioblastomas using SV-^{1}H-MRS.*

* Corresponding author; avellido@lsi.upc.edu, tel.: +34 93 4137796, fax: +34 93 4137833.

Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel proton MRS

A. Vellido, E. Romero, M. Julià-Sapé, C. Majós, À. Moreno-Torres, J. Pujol, C. Arús

Graphical abstract. SV-^{1}H-MR spectra from a total of 109 patients (78 glioblastomas and 31 metastases) from the multi-center, international INTERPRET database, plus a test set of 40 patients (30 glioblastomas and 10 metastases) from three different centers, were analyzed using a robust method for feature selection coupled with a linear-in-the-parameters perceptron classifier. A linear combination of five spectral frequencies yielded, for the test set, an area under the ROC curve (AUC) of 0.86, and an area under its convex hull (AUCCH) of 0.91. In the accompanying figure, visual representation of all cases using the spectral amplitudes corresponding to a subset of three selected frequencies from a combination of two echo times, separated by an optimal class-discriminating plane.



1 Introduction

Metastatic brain tumors often arise as multifocal lesions in adults with a history of malignancy. Before treatment is delivered, malignant neoplastic tumors (metastasis, high-grade glioma, and malignant lymphoma) must be differentiated. Amongst these three, metastases and high-grade gliomas are the hardest to differentiate due to their radiological similarity [1].

The discrimination of glioblastoma from solitary metastasis is a challenging problem that arises when a necrotic mass appears within the brain. It is also a highly relevant decision, as maximal surgical resection is the treatment of choice for glioblastomas, while a solitary metastasis is the result of a systemic tumoral process, and its treatment ultimately depends on the origin and the degree of dissemination of the tumor.

Radiology plays an important role in this type of discrimination. The diagnosis of metastasis is quite obvious when multiple brain lesions are found and a primary extracranial tumoral process is known. The situation is different when a solitary mass is found, because radiological findings from conventional Magnetic Resonance Imaging (MRI) may be very similar for both types of tumor. Further diagnostic support can be obtained from the so called *physiological* MR techniques. Most of them use the infiltrative pattern of growth of glioblastomas to accomplish the differentiation. Metastasis is, in fact, an extra-axial process that does not infiltrate the surrounding parenchyma and, accordingly, perfusion MR, diffusion MR, and spectroscopy measurements in the adjacent brain parenchyma should be those of normal parenchyma or edema. On the other hand, glioblastoma is an infiltrative process that should show a tumoral, or at least abnormal, pattern surrounding the tumoral enhancing process when using these techniques. A recent example of this type of study, based on the differences in metabolite ratios in the enhancing tumor and the peritumoral edema, aiming to discriminate between tumor infiltration (glioblastomas) and tumor-free edema (metastases), can be found in the work of Server *et al.* [2].

In this study, we aimed to find if some characteristics of the focal enhancing mass could help to differentiate between glioblastoma and solitary metastasis. Our hypothesis was that tumoral processes originated in the brain should be, at least to some extent, different from those originated elsewhere, and that those differences should be found in the single-voxel proton MRS (SV-^{1}H-MRS) signal.

The existing literature [1,3–5] considers this differentiation problem from SV-^{1}H-MRS as one of great difficulty. As stated by Opstad *et al.* [4], the radiological appearance of intracranial metastases and high-grade gliomas is often similar and dominated in both cases by large peak intensities corresponding

to neutral lipids, a byproduct of necrosis [6]. This problem has often been circumvented by considering both pathologies as part of a more general class of high-grade malignant tumors [7–12].

Most of the aforementioned studies that address the problem of differentiation between glioblastoma and solitary metastasis analyzed small patient samples, often from a single clinical source. Glioblastoma cases usually predominated in all the analyzed databases, but no active compensation procedure for the different prevalences of both pathologies was used. With some exception [5], the cited studies investigated a limited number of peak intensities, or their ratios, at prescribed frequencies, reflecting prior knowledge of which metabolites should be considered as relevant to the differentiation problem. As remarked by Huang *et al.* [13], such approach risks throwing away large parts of the spectrum where useful discriminatory information may be present.

The use of test data sets (that is, independent collections of unseen cases) is a requisite to prove the validity or generalization capability of a proposed model [14]. It is not enough that we define a model that correctly classifies a given data set. The model must also be able to correctly classify unseen, out-of-sample, data cases. In other words, not using a test set entails that results can only reflect accurately the analyzed data, therefore being of questionable use for subsequent out-of-sample predictions. In the cited literature, a test set was only used by García-Gómez *et al.* [5].

Other studies have gone beyond SV-^{1}H-MRS to tackle the differentiation between metastasis and glioblastoma. Alternative techniques include MRI and two-dimensional Turbo Spectroscopic Imaging information [15], diffusion tensor imaging [16,17] and multiple-voxel MRS with 2D chemical-shift imaging (CSI) and peak amplitude ratios [2]. Recent studies have also resorted to Morphometric Analysis of MR Images [18].

In a clinical setting, the discrimination of glioblastomas from metastases becomes a decision making problem. Diagnostic decision making in neuro-oncology is, for rather obvious reasons, an extremely sensitive matter. Taking into account that most diagnostic techniques in this domain must rely upon non-invasive data acquisition methods, clinicians might benefit from at least partially automated computer-based decision support using pattern recognition techniques [19]. There is no technological barrier for the use of SV-^{1}H-MRS information in computer-based medical decision support systems (MDSS), given that this type of data can be acquired and processed automatically (therefore becoming part of routine clinical examination) [4].

In this domain, diagnostic decision support requires methods that are both robust and interpretable by the radiology expert. In diagnostic classification-oriented pattern recognition of MRS data, one way to comply with the inter-

pretability requirement is through data dimensionality reduction and, more specifically, through feature selection (FS). A thorough FS procedure is applied in this paper to the problem of discriminating between metastatic brain tumors and glioblastomas, on the basis of SV-^{1}H-MRS from the multi-center, international INTERPRET database [20].

The proposed FS procedure is seamlessly interwoven with classification using a simple and linear-in-the-parameters machine learning model, namely the Single-Layer Perceptron (SLP) [21]. The FS technique is based on the hypothesis that irrelevant features produce smaller variations in the SLP output prediction than relevant ones. We hypothesize that the combination of a thorough and robust FS procedure and a linear classifier will lead to improved generalization results in the discrimination task for a test set. We also expect (in accordance to some existing studies [22]) to obtain the best discrimination results from the combination of SV-^{1}H-MRS data acquired at different echo times.

2 Materials and methods

2.1 The INTERPRET SV-^{1}H-MRS database and medical DSS

The available data are SV-^{1}H-MR spectra acquired *in vivo* from brain tumor patients. They are part of the multi-center, international web-accessible INTERPRET project database [20]. They were gathered from hardware produced by several manufacturers (GE, Philips and Siemens) and expressed in different formats. A total of eight clinical centres in five countries contributed cases to the database: CDP (Centre Diagnòstic Pedralbes-CETIR, with units at Pedralbes, Barcelona and Esplugues del Llobregat, Spain), IDI Bellvitge (Institut de Diagnòstic per la Imatge - Unitat Bellvitge, L'Hospitalet del Llobregat, Spain), SGUL (St George's University of London, UK), UMCN (Universitair Medisch Centrum Nijmegen, The Netherlands), UJF (Unité mixte Université Joseph Fourier/INSERM U594, Grenoble, France), FLENI (Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia, Buenos Aires, Argentina), MUL (Uniwersytet Medycznyw Lodz, Lodz, Poland). The data from CDP were in turn gathered from six hospitals in the Barcelona metropolitan area, including: Hospital de Bellvitge, L'Hospitalet del Llobregat; Hospital de la Santa Creu i Sant Pau, Barcelona; Hospital Clínic, Barcelona; Hospital Germans Trias i Pujol, Badalona; Hospital Mútua de Terrassa, Terrassa; and Hospital Sant Joan de Déu, Esplugues del Llobregat.

The criteria for the inclusion of cases in this study (in which there are only two tumor types out of the many available from the INTERPRET database)

were: (a) that the case had single voxel, 1.5T spectra acquired both at short and long echo times from a nodular region of the tumour; (b) that the voxel was located in the same region as where a subsequent biopsy was obtained; (c) that the spectra had not been discarded because of acquisition artifacts or other data quality reasons and (d) that a histopathological diagnosis was agreed among a committee of neuropathologists.

The analyzed data were acquired at short and long echo times (SET and LET). They include 78 glioblastomas (WHO 9440/3)) and 31 metastases (WHO 8000/6). Processing was performed with the INTERPRET Data Manipulation Software [1] [23], including UL2 unit length normalization [24]. Data were further scaled to zero mean and unit variance. Clinically-relevant regions of the spectra were sampled to obtain 195 frequency intensity values (data features), spanning approximately from 4.22 to 0.49 ppm (parts per million). These 109 cases were used in the feature selection and classification procedure described in section 2.2.

The echo time is an influential parameter in ^{1}H-MRS data acquisition. In SET spectra (typically acquired at 20-40 ms; in this study: STEAM 20 ms and PRESS 30-32 ms) some metabolites are better detected (e.g. lipids, myo-inositol, glutamine and glutamate). However, there may be numerous overlapping resonances (e.g. glutamate/glutamine at 2.2 ppm) which make the spectra difficult to interpret [25]. The use of LET (in this study: PRESS 135-144 ms) yields fewer metabolites but with more clearly resolved peaks and less baseline distortion, resulting in a more readable spectrum. Existing studies have resorted to either SET or LET MRS to discriminate between different types of high-grade tumors. There is evidence, though, that the combined use of both echo times could be advantageous [22,25]. In this study, the spectra acquired at SET and LET from the same patient (when both echo times were available) were combined through straight concatenation of the spectra, as in [22]. The combined echo times of the aforementioned 109 cases were used to build the diagnostic prediction models.

The INTERPRET database forms the core of a computer-based MDSS [23], designed to assist radiologists in diagnosing and grading brain tumors using *in vivo* SV-^{1}H-MRS. It includes automated pattern recognition techniques (such as Linear Discriminant Analysis -LDA- classifiers), and the results corresponding to some classification problems are displayed in a two-dimensional representation space that can be navigated using an intuitive graphical user interface that links the representation of a case with the corresponding underlying data (spectra and, eventually, image). The final goal of the INTERPRET MDSS is that of facilitating the incorporation of the results of pattern recognition analysis into an overall diagnostic procedure in which the possible

[1] URL: http://gabrmn.uab.es/dms

algorithmic and mathematical intricacies are transparent to the clinician. The latest official release of the MDSS is its version 3.0.2 [23] [2] .

A test data set of 40 cases (30 glioblastomas and 10 metastases) was kept aside to test the generalization capability of the selected model in the classification task. This is, as mentioned in the introduction, the capability of correctly classifying unseen, out-of-sample, data cases. These data came from three different clinical centers in the Barcelona metropolitan area: CETIR-CDP (Centre Diagnòstic Pedralbes, Unitat Esplugues, Esplugues del Llobregat), CRC-Corporació Sanitaria-IAT (Institut d'Alta Tecnología, Barcelona), and IDI-Badalona (Institut de Diagnòstic per la Imatge - Unitat Badalona, Badalona), and were acquired as part of the EU-funded eTUMOUR research project [26] [3] . Ethics committee approval for data accrual was gathered in the context of the INTERPRET and eTUMOUR projects.

2.2 Feature selection and classification methods

As some of the studies briefly reviewed in the introductory section reflect, most of the available spectral frequency range in the SV-^{1}H-MRS data is likely to be of little relevance to the discrimination between high-grade glioblastomas and metastases. This reveals the importance of using an adequate and quantitatively-motivated FS automated procedure. Such procedure should be robust enough to yield a selection of frequencies that was not only relevant to the sample of patients under study, but also able to yield good diagnostic predictions for unseen data in test sets.

In this study, we propose the use of an exhaustive FS procedure associated to a simple pattern recognition classification model: the SLP [21] artificial neural network. In the mathematical specification used in this study (and detailed in the appendix), the SLP is similar to logistic regression, although the adaptive parameters of the model are obtained through standard back-propagation techniques. The SLP is preferred as a partner for FS rather than more complex alternative pattern recognition classifiers such as Multi-Layer Perceptrons (MLP) [27,28] or linear Support Vector Machines (SVM) [29] for several reasons: The former can be computationally too expensive for the number of MRS frequencies analyzed in the available database; MLP parameters are also more difficult to adjust appropriately, and the saliency (relevance) of every feature is also likely be more independent for SLP than for linear SVM. In addition, linear models such as the proposed SLP have performed well with these data in previous studies [7], and their classification can be straightforwardly interpreted in terms of the original features.

[2] URL: http://gabrmn.uab.es/dss
[3] URL: http://www.etumour.net

Two components of the FS procedure must be explicitly specified: the feature subsets evaluation measure and the search procedure through the space of all possible feature subsets. The first is computed as the sum of the individual saliencies of the features, which are a simple function of the adaptive weights (parameters) of the SLP. This method is based on the hypothesis that irrelevant features produce smaller variations in the output values than relevant ones, with smaller output variations being the result of small model weights. The search is implemented as a backward selection procedure with an iterative selection process controlled by the previously defined saliency measure. All the technical details of the proposed SLP-based FS process are explained in the appendix.

All the reported results were obtained with SLP classifiers whose training processes were balanced to account for the different tumor type prevalences, i.e., to compensate for the different number of cases corresponding to each of the two analyzed pathologies. In doing so, it prevented the SLP favoring the accurate classification of only the most prevalent class. In our experiments, the balancing process involved modifying the back-propagated error of the metastases, which was multiplied by the ratio of glioblastomas to metastases. This has a similar effect to over-sampling the least frequent class (metastases).

The FS procedure is decremental and it was repeated, starting from the complete data, a number of times under different initialization conditions to ensure the reliability of the FS outcome (that is, to ensure the consistency of the selected subsets of features). The results reported and discussed in the next sections were the best obtained for the test set.

2.3 Discrimination quality measures

Several quality measures were used to report the classification (discrimination) results. They all use the concepts of true and false predictions (true positives -TP- correspond to correct metastases predictions and true negatives -TN- correspond to correct glioblastoma predictions; likewise, false positives -FP- are false metastasis predictions and false negatives -FN- are false glioblastoma predictions). Measures include the accuracy (percentage of total correctly classified cases, that is, ratio of true cases, TP+TN, to all cases), sensitivity (ratio of TP to all metastases) and specificity (ratio of TN to all glioblastomas), measured at the mid-range threshold.

The ROC curve represents the values of sensitivity with respect to (1-specificity) obtained by varying the values of the discrimination threshold across its range. ROC analysis has its roots precisely in the radiology area [30,31]. The areas under the ROC plot (AUC, [32,33]) and under the convex hull of the ROC

plot (AUCCH, [34]) are routinely used as appropriate measures for qualifying classification results. As stated by Metz [35], "ROC analysis provides the most comprehensive description of diagnostic accuracy available to date, because it estimates and reports all of the combinations of sensitivity and specificity that a diagnostic test is able to provide". In our study, these areas approximate the probability that the SLP will rank a randomly chosen positive case (a metastasis) higher than a randomly chosen negative one (a glioblastoma), and are closely related to the Mann-Whitney U test. Both areas are reported because the AUC may underestimate the quality of the prediction for small data samples (such as the test set in this study), while the AUCCH can, at most, slightly overestimate such quality.

3 Results

Maximum overall accuracies of 85% *in the test set* were obtained for several, extremely parsimonious, selected combinations of LET and SET frequencies. They are summarized in Table 1. These maximum test accuracies corresponded to training accuracies (using the 109 INTERPRET concatenated spectra) of, in turn and following the same order as in Table 1, 79%, 79.8%, 82.6%, and 80.7%. Note that the selections reported in this table are very similar to each other, which is a clear indication of the stability of the FS procedure. Some frequencies (such as those at 2.29, 2.32 and 3.01 ppm, for instance) appear repeatedly in the selection. Interestingly, in one case (2.32) they are selected at both echo times. They are, in any case, extremely consistent; that is, they repeatedly come up as the final result over the battery of performed experiments.

Experiments were also carried out using only LET or SET data separately. Using LET, a maximum accuracy of 77.5% in the test set was achieved. Using only SET, the result decreased to a 75%. Removing the SET features from the selected subsets listed in Table 1 (one frequency in the first three subsets and two in the 4^{th} one) reduced the accuracy from 85% to 77.5% (subset 1), 80% (subset 2), and 82.5% (subsets 3 and 4). The removal of the LET frequencies decreased the performance to values in the range 62.4%-65.1%.

Importantly, the simplicity of the SLP classifiers (combined with the parsimony of the frequency subset selections) allows us to express these predictions as simple formulae. These are expressed as a linear combination of frequencies, as listed in Table 2. These formulae can be used for classification by expressing the SLP output (prediction) y for a given case (spectrum) $\mathbf{x}$ as

$$y(\mathbf{x}) = \tanh(a(\mathbf{x})), \tag{1}$$

Table 1
The four subsets of frequencies from the concatenation of LET and SET data for which an 85% accuracy in the test set was achieved in the experiments with SLP. Frequencies expressed in ppm (rounded to two decimal places). Letter prefixes stand for LET (L) and SET (S) frequencies.

Subset ♯	Features Selected
1	**L2.32-L2.29-S2.32**
2	**L2.32-L2.29-S2.17-L2.02-L3.01**
3	**L2.32-L2.29-S2.32-L3.42-L3.36-L3.01**
4	**L2.32-L2.29-S2.17-L2.02-L3.01-S2.15**

Table 2
Classification formulae for each of the selections listed in Table 1 (presented in the same order). Coefficients rounded to the third decimal place.

Subset ♯	Classification formula
1	a(**x**)=3.473-0.548*L2.32+0.484*L2.29-0.493*S2.32
2	a(**x**)=1.383-0.088*L3.01-0.377*L2.32+ 0.250*L2.29+0.153*L2.02-0.261*S2.17
3	a(**x**)=-0.612+0.349*L3.42-0.214*L3.36-0.049*L3.01-0.436*L2.32+0.453*L2.29-0.160*S2.32
4	a(**x**)=0.671+0.090*L3.01-0.375*L2.32+0.249*L2.29+0.156*L2.02-0.205*S2.17-0.053*S2.15

so that $y(\mathbf{x}) \in [-1, 1]$.

Given a mid-range classification threshold of $y(\mathbf{x}) = 0$, a value of $y(\mathbf{x}) > 0$ would correspond to a metastasis diagnostic prediction; therefore, an output of value 1 would correspond to a fully confident metastasis prediction. Likewise, a value of $y(\mathbf{x}) < 0$ would correspond to a glioblastoma diagnostic prediction, with an output of value -1 indicating a fully confident glioblastoma diagnosis. Moreover, the higher the absolute value of a positive coefficient in the formula, the stronger the influence of the corresponding frequency in a metastasis prediction; likewise, the higher the absolute value of a negative coefficient, the stronger the influence of the corresponding frequency in a glioblastoma prediction. These criteria provide the expert with an explicit quantitative ranking of the relevance of individual frequencies on the diagnostic prediction. To help interested users, a protocol to process and evaluate new cases is provided as supplementary information. A simple spreadsheet that can be used to obtain predictions for new cases is also available online [4].

Detailed test predictions, together with *sensitivity* and *specificity* values (corresponding to confusion matrices at mid-range classification threshold of $y(\mathbf{x}) = 0$), AUC and AUCH results, are listed in Table 3.

According to the quality measures reported in Table 3, the second and fourth subsets of selected frequencies yield the best and most adequately balanced

[4] URL: http://gabrmn.uab.es/GBM-MET-formula.xls

Table 3
Detailed test set classification results for each of the selections listed in Table 1 (presented in the same order). First column: number of subset, as in previous tables. Second column: Total number of correctly classified cases (CCC), out of 40. Third column: True positives (TP) and negatives (TN); false positives (FP) and negatives (FN). Fourth column: corresponding sensitivity (*sen*) and specificity (*spe*), as a percentage. Fifth and sixth columns: AUC and AUCCH results (maximum area possible value: 1).

Subset ♯	CCC	TP, TN, FP, FN	*sen*/*spe*	AUC	AUCCH
1	34/40	7 TP, 27 TN, 3 FN, 3 FP	70/90	0.78	0.86
2	34/40	9 TP, 25 TN, 1 FN, 5 FP	90/83.3	0.86	0.91
3	34/40	6 TP, 28 TN, 4 FN, 2 FP	60/93.3	0.83	0.87
4	34/40	9 TP, 25 TN, 1 FN, 5 FP	90/83.3	0.86	0.91

results in the diagnostic prediction with the test set. The best AUC value of 0.86 and AUCH value of 0.91 must be compared with an AUC of 0.84 reported in [4] using the lipid peak area ratio, but no independent test set. The results are also to be compared with those reported by García-Gómez *et al.* [5]. In that study, for SET spectra acquired at 1.5 T, the best results for a balanced test set were obtained using peak integration techniques and a LDA classifier. An error rate of 0.22 and a corresponding BER of 0.21 were reported. From the results in Table 3, an error rate of 0.15 and a BER of 0.13 were achieved in our experiments.

Note that the fourth solution is almost identical to the second one but for the addition of SET 2.15 (contiguous to the also selected SET 2.17). Therefore, and following a *lex parsimoniae* criterion, the second solution, including five frequencies, was chosen to be the one implemented in the INTERPRET computer-based MDSS described in section 2.1. This new development, which provides predictions for new cases, will shortly be available in version 3.1 of the MDSS. Figure 1 displays the five selected frequencies on top of the mean amplitudes of both tumor types.

The direct 3-D data visualization for this best solution is not possible, as it consists of 5 frequencies. Instead, a visualization of the SLP predictions y can be provided, as in Figure 2. In this figure, the further the case from the threshold $y(\mathbf{x}) = 0$ and the closer to the left limit: $y(\mathbf{x}) = -1$, the more confident the glioblastoma prediction is. Correspondingly, the further the case from the threshold $y(\mathbf{x}) = 0$ and the closer to the right limit: $y(\mathbf{x}) = 1$, the more confident the metastasis prediction is.

Instead, the first selection reported in Table 1, which only includes three frequencies, allows direct 3-D visualization, as illustrated by Figure 3. Importantly for interpretation purposes, the differentiation surface generated by the SLP classifier, separating glioblastomas from metastases, can also be explicitly displayed.

4 Discussion

The importance of using a test set to qualify the results must be again stressed at this point. If only the original INTERPRET data had been used to create the SLP prediction models, there would be no guarantee against the possibility that the models only reflected accurately those data, therefore being of no use for subsequent out-of-sample predictions.

The best FS results are not only parsimonious but also quite consistent. Note that the sign of the coefficients in the formulae reported in Table 2 remains stable when the same frequencies are selected in different experiments. Note also that all selected subsets include frequencies from both echo times, although with a clear preponderance of LET ones. This is a definite indication that the combination of echo times is necessary for the improvement of the differentiation capabilities of the model. The results reported in the previous section show that the combination of echo times yields better results that the separate use of either echo time. Furthermore, the performance has been

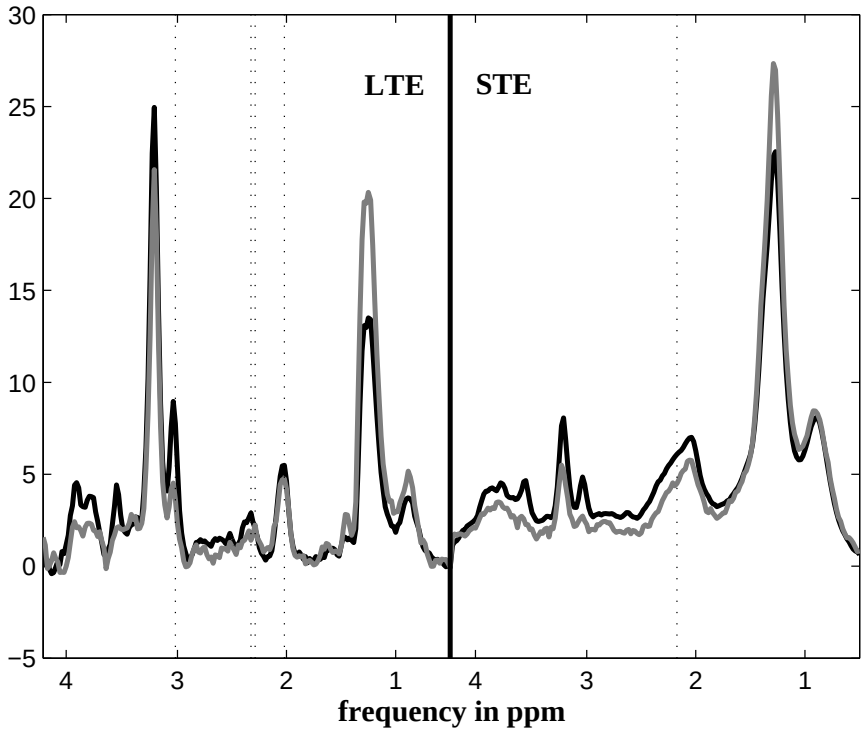


Fig. 1. Location of the frequencies of the best subset selected **L2.32-L2.29-S2.17-L2.02-L3.01** on the LET+SET concatenated data used in the FS procedure (SLP *training* data). LET spectra on the left and SET on the right. Mean metastases amplitudes as a grey solid line. Mean glioblastomas as a black solid line. Selected frequencies as dashed vertical lines. The vertical axis is kept unlabeled, as it corresponds to normalized intensity arbitrary units.

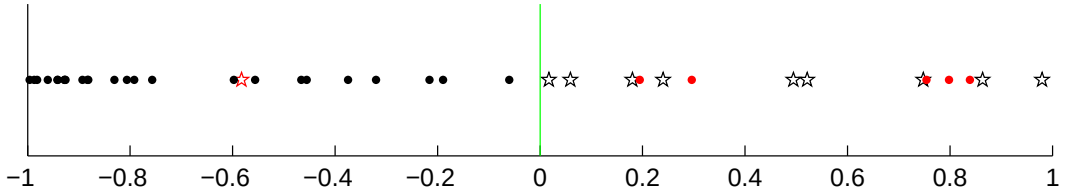


Fig. 2. Visual representation of the SLP predictions $y(\mathbf{x})$ for the test set, corresponding to the best selection of frequencies, as described in the main text. True positives (correctly classified metastases) are represented as black stars; false negatives as red stars (there is only one misclassified metastasis); true negatives as black dots; and false positives (misclassified glioblastomas) as red dots. The decision threshold is set at $y(\mathbf{x}) = 0$. This representation of the data is provided in a forthcoming version of the INTERPRET MDSS.

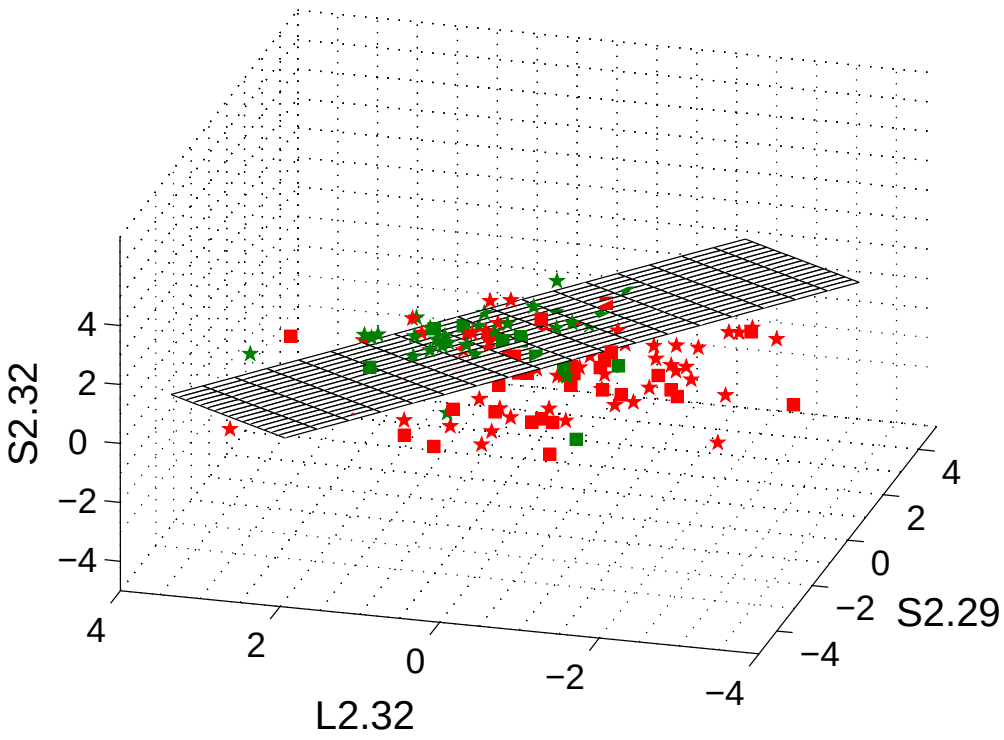


Fig. 3. Visual representation of the data (amplitudes) corresponding to the subset of three selected spectral frequencies **L2.32-L2.29-S2.32** from the LET+SET data set, as described in Table 1. Each of the axes is one of the frequencies, as labeled. Glioblastomas are represented in red and metastases in green. Data used to create the classifier are represented by stars, while data of the test set are represented by squares. The differentiation surface is displayed almost in profile in order to better appreciate the separation of cases. Given that the SLP is a linear classifier, the decision surface is a plane. A case that fell within this surface would correspond to a SLP prediction $y(\mathbf{x}) = 0$. Note that the Euclidean distance from the points to the differentiation plane can be directly used as a measure of our confidence on the diagnostic prediction provided by the classifier.

shown to deteriorate as well if frequencies of either time of echo are removed from the subsets selected when both echo times are used.

The well-balanced test accuracy achieved with the proposed method for both metastases and glioblastomas also validates our approach for the design of SLP classifiers, in which the training procedure included a mechanism to actively compensate for the different tumor type prevalences. This is explicitly reflected by the excellent BER and ROC analysis results.

The frequencies in the best discriminating subset (**L2.32-L2.29-S2.17-L2.02-L3.01**) belong to well-known frequency ranges that are relevant for brain tumor pattern recognition. The LET frequency at 3.01 ppm mostly represents total creatine, while LET at 2.02 ppm, depending on the size, infiltrative nature and degree of necrosis of the tumour, will be mostly contributed by either NAA (due to partial volume effects), or, when mobile lipids resonances are apparent at ca. 1.3 and 0.9 ppm, by the -CH=CH-**CH2**- methylene group of fatty acyl chain of those mobile lipids, due to their low T2. In [1], a qualitative analysis concluded that the presence of intra-tumoral creatine (Cr) is a marker for gliomas, whereas its absence might be an indicator of metastasis. In this respect, [36] reported significantly higher total creatine content in glioblastoma (3.15 +/- 0.30 μmol/g fresh wet weight, n = 59) than in metastases (1.85 +/- 0.28 μmol/g fresh wet weight, n = 18) in quantitative data from hydrosoluble metabolites in extracts of biopsies. This is in qualitative agreement with the average unit length normalized tumor type pattern shown in Figure 1.

On the other hand, quantitation from in vivo MRS data using the LCModel approach [4] failed to find significant differences in creatine content between glioblastoma (n=23) and metastases cases (n=24), although further work from the same institution [37] using proton high-resolution magic angle spinning analysis of brain tumor biopsies did indeed find higher creatine in glioblastomas (2.99$\pm$0.39 mM, n=24) than in metastases (1.24$\pm$0.21 mM, n=8). Nevertheless, it should be taken into account that the current study does not use the absolute metabolite content directly, but unit length UL2 normalized intensities instead. As a result, higher mobile lipids content, specially at LET, will contribute to an apparent metabolite content decrease in metastases versus glioblastoma after normalization.

A definite lipid signal was also concluded in [1] to indicate cellular necrosis in glioblastoma and metastasis, whereas no lipid signal at short echo time should lead to the exclusion of a metastasis diagnosis.

Finally, 2.32 and 2.29 ppm at LTE, together with 2.17 at SET, may have contributions from Glutamate/Glutamine (Glu/Gln) at 2.32 and 2.17 ppm and GABA at 2.29 ppm [38] while 2.02 at LET will have strong contribution of necrotic mobile lipids and/or NAA. In summary, small changes in the normalized intensity of creatine, mobile lipids, NAA and Glu/Gln/GABA resonances seem to provide the best discrimination handle in the problem addressed in this study. It may be of interest to recall here a previous study on pattern

recognition-based discrimination of glioblastoma and metastasis biopsies from proton high-resolution magic angle spinning information [39], for which these spectral ranges were also relevant to the discrimination.

The parsimonious nature of the best frequency subset selections obtained has been shown to ease the visualization of both the data used to create the classifier and those in the test set, as illustrated by Figures 2 and 3. This visualization can be an important element to facilitate the expert interpretation of the results as implemented in the INTERPRET MDSS.

There exist alternative approaches to the use of SV-^{1}H-MRS for discriminating glioblastomas from metastases, based in a multi-voxel approach [2,40]. These may have the advantage of not compelling the radiologist to decide *a priori* which is the best placement location for the sampled volume. They can also use differences in the infiltrative pattern among glioblastoma and metastasis and the corresponding peritumoural MRS pattern differences to discriminate between the two tumour types. In any case, even if a multi-voxel examination can easily detect the presence of infiltration, as shown by Server *et al.* [2], the same study detected differences not only in the peritumoural edema but also in the NAA/Cr ratio of the LTE in the tumoural core (a higher ratio in metastasis, 1.43 ± 1.09, than in glioblastomas, 0.87 ± 0.89), indicating that differences in the spectral pattern also exist. Even if certainly advantageous, this approach is not free from controversy [41,42] and, besides, not all clinical centres are equipped for performing good-quality multi-voxel data acquisition and postprocessing. Therefore, our approach based on SV-^{1}H-MRS should be of practical use, especially in clinical settings where no multi-voxel analysis is available, or in cases in which a single mass is located near the skull. Mixing both approaches with a focus on the peritumoural area may be an interesting goal for future research, aiming to improve the 80% specificity achieved by Server *et al.* [2] by decreasing the number of FP.

4.1 Analysis of the misclassified cases from the test set

All six misclassified cases from the test set (corresponding to the best solution) were further analyzed. In most, the spectra showed an unusual pattern with respect to the tumor class mean spectrum.

One metastasis, namely case et2893, was classified as a glioblastoma. An unusual spectral pattern was found with metabolite signals in addition to necrotic lipids (Figure 4), despite the voxel being properly located into the solid part of the tumor, avoiding normal tissue.

The other five misclassified cases were glioblastomas. Three of them showed a marked lactate signal in the LET spectrum (Figure 5). The reason for their

misclassification seems to be other, though: the three cases show atypically low levels of Glu/Gln at SET; moreover, et2042 shows atypically low values for all the features selected at LET; et3010 shows high levels of NAA at LET; and et2054 shows low levels of total creatine.

A further two cases, et3496 and et3194, showed an unusual pattern: the SET spectrum of case et3496 (Figure 6, left) was judged by the expert spectroscopists of eTumour as having poor, unexpected signals, possible due to scalp lipid contamination and poor phasing. In case et3194 (Figure 6, right), spectra at both echo times would be more characteristic of a lower grade, although the histopathological diagnosis was clear in the diagnosis of glioblastoma, with three consulting pathologists, as well as the originating pathologist, agreeing on the diagnosis. This can be due to heterogeneity within the tumor, which has been described in [43,44]. This is corroborated by the corresponding imaging, which reveals that the tumor consists on a cyst with adjacent solid regions. The voxel was correctly positioned over the solid region of the tumor.

In summary, the analysis of the misclassified cases brings about the fact that, in order to increase the reliability of the computer-based classifier in the classification of new cases, the expert should ensure that the voxel is positioned according to the criteria used to acquire the cases used to develop the classifier. This analysis also exemplifies that if the spectrum is artifactual, or it suffers from deficient water suppression or low signal-to-noise ratio, classification results are likely to be unreliable.

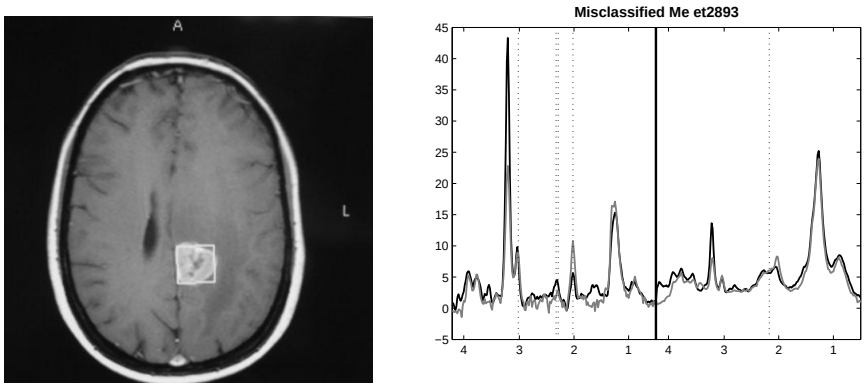


Fig. 4. Single misclassified metastasis: case *et2893*. Left) Voxel correctly placed over hyperintense area. Right) Spectrum as a solid black line. Mean metastases amplitudes as a grey solid line. LET part shows unusually high choline-containing compounds peak. At SET, the choline-containing compounds signal is also the second-highest intense peak after necrotic lipids. The high choline signal at LET will produce an apparent decrease of other sginals upon UL2 normalization. Note though that, given that the classification procedure is based on the 5 selected features, the misclassification of this case seems to be caused mainly by an atypically low level of NAA at LET and high levels of Glu/Gln and GABA at LET. Locations of the frequencies of the best subset selected from the LET+SET concatenated data (**L2.32-L2.29-S2.17-L2.02-L3.01**), represented as dashed vertical lines. LET data on the left and SET on the right of the solid vertical middle black line.

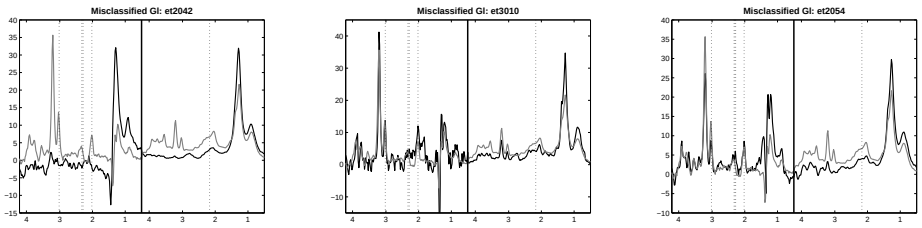


Fig. 5. Three misclassified glioblastomas with marked lactate signal at LET. Representation as in previous figure. Left) At LET, case et2042 shows overlapping lipid-lactate signals. Middle) Also at LET, case et3010 has a large inverted lactate doublet overlapping some contribution from lipid at 1.28ppm. In this case, the voxel reference image was positioned over a heterogeneous, necrotic area. The patient also had large areas of edema surrounding the solid part of the tumour. Right) For case et2054, the voxel was positioned over post-contrast images showing a predominant contribution from necrosis and a small percentage of hyperintense viable tissue, hence the lipid-lactate overlapping peaks and the choline-containing compounds being the most intense signals in the spectrum.

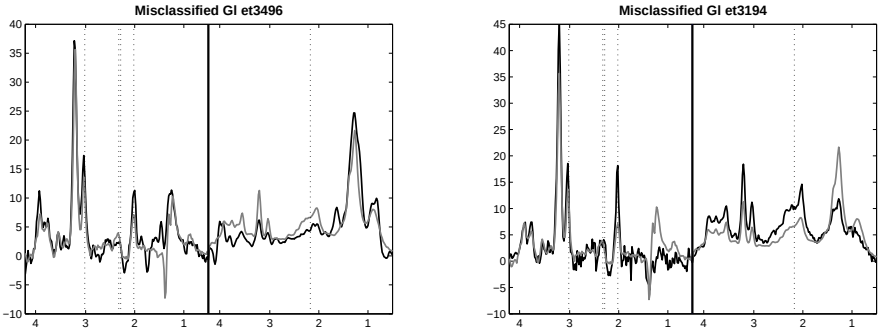


Fig. 6. Two more misclassified glioblastomas: cases *et3496*(left) and *et3194*(right), as described in the text. Representation as in previous figures.

5 Conclusions

In the introduction to this study, it was hypothesized that tumoral processes originated in the brain should be, at least to some extent, different from those originated elsewhere, and that those differences should be found in SV-^{1}H-MRS signal. The reported experimental results confirm this hypothesis and show that a robust feature selection method, coupled with a simple linear-in-the-parameters SLP model can differentiate metastases and glioblastomas to a high degree of accuracy from SV-^{1}H-MRS. The generalizability of these differentiation results is reinforced by the fact that they were obtained for a retrospective and multi-center independent test set of cases. The combination of SV-^{1}H-MRS acquired at different echo times is crucial to this classification success, although LET data predominate in the selected subsets of frequencies.

A differential advantage of the proposed procedure is that it allows us to obtain a simple linear prediction formula for such a difficult problem, based on metabolically interpretable frequencies. Such prediction formula could be

directly applied by interested clinical centers for performance evaluation, yielding predictions for new cases. This could be accomplished manually or through the INTERPRET MDSS.

Acknowledgements

This research was partially funded by Spanish MICINN R+D Plan TIN2009-13895-C02-01 and SAF2008-03323 projects. *Centro de Investigación Biomédica en Red (Bioingeniería, Biomateriales y Nanomedicina*, CIBER-BNN) is an initiative of *Instituto de Salud Carlos III* (ISCIII), Spain, which is co-funded by EU FEDER. Authors gratefully acknowledge former INTERPRET, eTumour and HealthAgents European project partners. Data providers: Dr. J. Capellades (IDI), Dr. F.A. Howe and Prof. J.Griffiths (SGUL), Prof. A. Heerschap (RU), Prof. L Stefanczyk and Dr. J.Fortuniak (MUL), and Dr. J. Calvar (FLENI); data curators: Dr. A.P. Candiota, Dr. I. Olier, Ms. T. Delgado, Ms. J. Martín, Ms. M. Camisón, and Mr. A. Pérez (all from GABRMN-UAB).

Appendix

FS with classification using Single Layer Perceptrons

The FS problem can be defined as follows: given a set of d features, select a subset that performs the best under a certain evaluation measure. From a computational point of view, the definition of FS usually leads to a search problem in a space of 2^d elements. In this case, two components must be specified: the feature subsets evaluation measure and the search procedure through the space of feature subsets. If any of these two components depends on an external model, it must also be specified.

In the remaining of this section, the constituent elements of the proposed FS procedure associated to the SLP model are outlined in some detail.

FS with SLP: The model

SLP artificial neural networks with sigmoidal output units were used in the reported experiments both for the feature subsets evaluation measure (within the FS process) and to obtain the test set accuracy (within the learning and generalization process). The number of output units was set to the number of

classes of the problem. Therefore, the activation y_j of the output unit j for a d-dimensional input vector x is computed as

$$y_j = g\left(\sum_{i=1}^{d} x_i \cdot \omega_{ji} + b_j\right), \tag{2}$$

where ω_{ji} is the weight that connects the input unit i with the output unit j, b_j is the bias of the output unit j, and $g(z)$ is a sigmoidal function. The SLP were trained in this study so as to minimize the sum-of-squares error.

FS with SLP: Feature subsets evaluation measure

The evaluation measure (the relevance) of a feature subset was computed as the sum of the individual saliencies of its features. The saliency s_i of a feature i over O outputs was computed as:

$$s_i = \sum_{j=1}^{O} |\hat{\omega}_{ji}|,$$

where $\hat{\omega}_{ji}$ are the weights of the trained SLP.

This method is based on the hypothesis that irrelevant features produce smaller variations in the output values than relevant ones. Hence, a natural way to compare the relevance of two features is to compare the absolutes values of the derivatives of the output function with respect to their respective input units in the trained model.

Formally, the derivative in the trained model of the output function y_j in Eq.2 with respect to an input feature x_i is

$$\frac{\partial y_j}{\partial x_i} = g'\left(\sum_{i=1}^{d} x_i \cdot \hat{\omega}_{ji} + b_j\right) \cdot \hat{\omega}_{ji},$$

and, for every j,

$$\frac{|\partial y_j / \partial x_{i_1}|}{|\partial y_j / \partial x_{i_2}|} = \frac{|\hat{\omega}_{ji_1}|}{|\hat{\omega}_{ji_2}|}.$$

Therefore, the variation (in absolute value) of the output function is smaller for input features with smaller weights (in absolute value), and they are the main candidates to be eliminated in a FS process. In summary, for linear discriminant functions such as SLP, the magnitude of the weights corresponding to a feature is considered as an indicator of its importance. Similar ideas can be found elsewhere (see, for example [29] or [45]).

FS with SLP: Search procedure

A backward selection procedure was used as an iterative selection process guided by the previously defined saliency measure. Starting from the complete set of available features, a subset of them was removed at every step of the algorithm according to the evaluation measure. Since the evaluation measure of a feature subset is computed as the sum of the saliencies of its features, the features to be removed at every step are those with the smallest saliency. The number of features removed at every step is a parameter of the system that controls the granularity of the selection and the computational cost.

FS with SLP: The algorithm

The FS algorithm applied in this study consists of three general phases:

(1) Perform a backward selection procedure starting with the whole set of features. At every step:
   (a) Train a SLP with the remaining features.
   (b) Compute the saliency of every feature.
   (c) Remove 50% of the remaining features.

   For every feature subset obtained, estimate its generalization performance through 5-fold cross validation. Out of all the results, keep the previous to the best one for the next phase (to avoid missing a possible generalization maximum in intermediate, not analyzed, subsets).
(2) The second phase is similar to the first one, except for:
   (a) The initial feature subset is the one obtained in the first phase.
   (b) At every step, 20% of the remaining features are removed.
(3) The third phase is similar to the second one, except for:
   (a) The initial feature subset is the one obtained in the second phase.
   (b) At every step, one feature is removed.

References

[1] Ishimaru, H, Morikawa, M, Iwanaga, S, Kaminoyo, M, Ochi, M, Hayashi, K. Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. *Eur Radiol.* 2001; **11**:1784-91.

[2] Server, A, Josefsen, R, Kulle, B, Mæhlen, J, Schellhorn, T, Gadmar, Ø, Kumar, T, Haakonsen, M, Langberg, CW, Nakstad, PH. Proton magnetic resonance spectroscopy in the distinction of high-grade cerebral gliomas from single metastatic brain tumors. *Acta Radiol.*, 2010; **51**:316-325.

[3] Fan, G, Sun, B, Wu, Z, Guo, Q, Guo, Y. In vivo single-voxel proton MR spectroscopy in the differentiation of high-grade gliomas and solitary metastases. *Clin Radiol.*, 2004; **59**:77-85.

[4] Opstad, KS, Murphy, MM, Wilkins, PR, Bell, BA, Griffiths, JR, Howe, FA. Differentiation of metastases from high-grade gliomas using short echo time ^{1}H spectroscopy. *J Magn Reson Imaging*, 2004; **20**:187-192.

[5] García-Gómez, JM, Luts, J, Julià-Sapé, M, Krooshof, P, Tortajada, S, Robledo, JV, Melssen, W, Fuster-García, F, Olier, I, Postma, G, Monleón, D, Moreno-Torres, À, Pujol, J, Candiota, AP, Martínez-Bisbal, MC, Suykens, J, Buydens, L, Celda, B, Van Huffel, S, Arús, C, Robles, M. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy, *Magn Reson Mater Phy.* MAGMA, 2009; **22**:5-18.

[6] Auer, DP, Gössl, C, Schirmer, T, Czisch, M. Improved lipid analysis of ^{1}H-MR spectra in the presence of mobile lipids. *Magn Reson Med.*, 2001; **46**:615-618.

[7] Tate, AR, Underwood, J, Acosta, DM, Julià-Sapé, M, Majós, C, Moreno-Torres, À, Howe, FA, van der Graaf, M, Lefournier, V, Murphy, MM, Loosemore, A, Ladroue, C, Wesseling, P, Bosson, JL, Cabañas, ME, Simonetti, AW, Gajewicz, W, Calvar, J, Capdevila, A, Wilkins, PR, Bell, BA, Rémy, C, Heerschap, A, Watson, D, Griffiths, JR, Arús, C. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. *NMR Biomed.*, 2006; **19**:411-434.

[8] Minguillón, J, Tate, AR, Arús, C, Griffiths, JR. Classifier combination for in vivo Magnetic Resonance Spectra of brain tumours. In *Multiple Classifier Systems*, Roli, F, Kittler, J (eds.) *Lecture Notes in Computer Science*, 2364, Springer, 2002; 282-292.

[9] Lukas, L. Brain tumor classification based on long-echo proton MRS signals. *Artif Intell Med.*, 2004; **31**:73-89.

[10] Devos, A, Lukas, L. Classification of brain tumours using short echo time ^{1}H MR spectra. *J Magn Reson*, 2004; **170**:164-175.

[11] Vellido, A, Romero, E, González-Navarro, FF, Belanche-Muñoz, Ll, Julià-Sapé, M, Arús, C. Outlier exploration and diagnostic classification of a multi-centre ^{1}H-MRS brain tumour database. *Neurocomputing*, 2009; **72**:3085-3097.

[12] González-Navarro, FF, Belanche-Muñoz, LlA, Romero, E, Vellido, A, Julià-Sapé, M, Arús, C. Feature and model selection with discriminatory visualization for diagnostic classification of brain tumours. *Neurocomputing*, 2010; **73**:622-632.

[13] Huang, Y, Lisboa, PJG, El-Deredy, W. Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection, *Stat Med.*, 2003; **22**:147-164.

[14] Altman, DG, Royston, P. What do we mean by validating a prognostic model? *Stat Med*, 2000; **19**:453-473.

[15] Luts, J. Classification of Brain Tumors Based on Magnetic Resonance Spectroscopy. PhD Thesis, Katholieke Universiteit Leuven, Belgium, 2010. URL: http://homes.esat.kuleuven.be/ jluts/phd.pdf

[16] Tsuchiya, K, Fujikawa, A, Nakajima, M, Honya, K. Differentiation between solitary brain metastasis and high-grade glioma by diffusion tensor imaging. *Brit J Radiol.*, 2005; **78**:533-537.

[17] Wang, W, Steward, CE, Desmond, PM. Diffusion Tensor Imaging in glioblastoma multiforme and brain metastases: The role of p, q, L, and Fractional Anisotropy. *Am J Neuroradiol.* (AJNR), 2009; **30**:203-208.

[18] Blanchet, L, Krooshof, PWT, Postma, GJ, Idema, AJ, Goraj, B, Heerschap, A, Buydens, LMC. Discrimination between metastasis and glioblastoma multiforme based on Morphometric Analysis of MR Images, *Am J Neuroradiol.* (AJNR), 2011; **32**:67-73.

[19] Lisboa, PJG, Vellido, A, Tagliaferri, R, Napolitano, F, Ceccarelli, M, Martín-Guerrero, JD, Biganzoli, E. Data mining in cancer research, *IEEE Comput Intell M.*, 2010, **5**:14-18.

[20] Julià-Sapé, M, Acosta, D, Mier, M, Arús, C, Watson, D, The INTERPRET Consortium. A multi-centre, web-accessible and quality control checked database of in vivo MR spectra of brain tumour patients. *Magn Reson Mater Phy.* (MAGMA), 2006; **19**:22-33.

[21] Widrow, B, Lehr, MA. 30 years of adaptive neural networks: perceptron, Madaline, and backpropagation, *P IEEE.*, 1990, **78**:1415-1442.

[22] García-Gómez, JM, Tortajada, S, Vidal, C, Julià-Sapé, M, Luts, J, Moreno - Torres, À, Van Huffel, S, Arús, C, Robles, M. The influence of combining two echo times in automatic brain tumor classification by MRS. *NMR Biomed.*, 2008; **21**:1112-1125.

[23] Pérez-Ruiz, A, Julià-Sapé, M, Mercadal, G, Olier, I, Majós, C, Arús, C. The INTERPRET decision-support system version 3.0 for evaluation of Magnetic Resonance Spectroscopy data from human brain tumours and other abnormal brain masses. *BMC Bioinformatics*, 2010; **11**:581.

[24] Tate, AR, Majós, C, Moreno, À, Howe, FA, Griffiths, JR, Arús, C. Automated classification of short echo time in in vivo ^{1}H brain tumor spectra: a multicenter study, *Magn Reson Med.* 2003, **49**: 29-36.

[25] Majós, C, Julià-Sapé, M, Alonso, J, Serrallonga, M, Aguilera, C, Acebes, JJ, Arús, C, Gili, J. Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE, *Am J Neuroradiol.* (AJNR), 2004; **25**:1696-1704.

[26] Julià-Sapé, M, Mier, M, Lurgi, M, Estanyol, F, Rafael, X, Delgado-Goñi, T, Camisón, M, Martínez-Bisbal, M, Celda, B, Arús, C, eTumour Consortium. The eTUMOUR database: a tool for annotation and curation of multidimensional human brain tumor data. *Proceedings of the* 17^{th} *Scientific*

Meeting, International Society for Magnetic Resonance in Medicine ISMRM, Honolulu, USA, 2009: 3475.

[27] Lisboa, PJG, Kirby, SPJ, Vellido, A, Lee, YYB, El-Deredy, W. Assessment of statistical and neural networks methods in NMR spectral classication and metabolite selection. *NMR Biomed.*, 1998; **11**:225-234.

[28] Romero, E, Sopena, JM, Performing feature selection with Multi-Layer Perceptrons. *IEEE T Neural Networ.*, 2008; **19**:431-441.

[29] Guyon, I, Weston, J, Barnhill, S, Vapnik, VN. Gene selection for cancer classification using support vector machines. *Mach Learn.*, 2002; **46**:389-422.

[30] Metz, CE. Basic principles of ROC analysis. *Semin Nucl Med.*, 1978; **8**:283-298.

[31] Sweets, JA. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol.*, 1979; **14**:109-121.

[32] Hanley, JA, McNeil, BJ. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, 1982; **143**:29-36.

[33] Hanley, JA, McNeil, BJ. A method of comparing the areas under Receiver Operating Characteristic curves derived from the same cases. *Radiology*, 1983; **148**:839-843.

[34] Barber, CB, Dobkin, DP, Huhdanpaa, H. The quickhull algorithm for convex hull. *ACM T Math Software*, 1996; **22**:469-483.

[35] Metz, CE. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *J Am Coll Radiol.*, 2006; **3**:413-422.

[36] Candiota, AP. [Contribució a la millora del diagnòstic i de la valoració pronòstica de tumors cerebrals humans]. PhD Thesis. Universitat Autònoma de Barcelona, 2004, URL: http://hdl.handle.net/10803/3525, last accessed 15/07/11.

[37] Wright, A, Fellows, G, Griffiths, J, Wilson, M, Bell, B, Howe, FA. *Ex vivo* HRMAS of adult brain tumours: metabolite quantification and assignment of tumour biomarkers. *Mol Cancer*, 2010; **9**:66.

[38] Hu, J, Yang, S, Xuan, Y, Jiang, Q, Yang, Y, Haacke, EM. Simultaneous detection of resolved glutamate, glutamine, and gamma-aminobutyric acid at 4 T. *J Magn Reson.*, 2007; **185**:204-213.

[39] Poullet, J-B, Martínez-Bisbal, MC, Valverde, D, Monleón, D, Celdá, B, Arús, C, Van Huffel, S. Quantification and classification of high-resolution magic angle spinning data for brain tumor diagnosis. *Proceedings of the* 29^{th} *Annual International Conference of the IEEE EMBS*, Lyon, France, 2007; 5407-5410.

[40] Law, M, Cha, S, Knopp, EA, Johnson, G, Arnett, J, Litt, AW. High-Grade Gliomas and Solitary Metastases: Differentiation by Using Perfusion and Proton Spectroscopic MR Imaging, Radiology, 2002; **222**:715-721.

[41] Sijens, PE. Response to article "Proton magnetic resonance spectroscopy in the distinction of high-grade cerebral gliomas from single metastatic brain tumors". *Acta Radiol.*, 2010; **51**:326-328.

[42] Server, A. Response to a letter by Paul E. Sijens. *Acta Radiol*, 2010; **51**:329-333.

[43] Kunz, M, Thon, N, Eigenbrod, S, Hartmann, C, Egensperger, R, Herms, J, Geisler, J, la Fougere, C, Lutz, J, Linn, J, Kreth, S, von Deimling, A, Tonn, JC, Kretzschmar, HA, Pöpperl, G, Kreth FW. Hot spots in dynamic ^{1}8FET-PET delineate malignant tumor parts within suspected WHO grade II gliomas. *Neuro-Oncology*, 2011, **13**:307-316.

[44] Paulus, W, Peiffer, J. Intratumoral histologic heterogeneity of gliomas. A quantitative study. *Cancer*, 1989; **64**:442-447.

[45] Steppe, JM, Bauer, KW. Feature saliency measures. *Comput Math Appl.*, 1997; **33**:109-126.