Research Article

NMR IN BIOMEDICINE

Received: 4 April 2011, Revised: 1 August 2011, Accepted: 13 September 2011, Published online in Wiley Online Library: 13 November 2011

(wileyonlinelibrary.com) DOI: 10.1002/nbm.1797

Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel ^{1}H MRS

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

This article investigates methods for the accurate and robust differentiation of metastases from glioblastomas on the basis of single-voxel ^{1}H MRS information. Single-voxel ^{1}H MR spectra from a total of 109 patients (78 glioblastomas and 31 metastases) from the multicenter, international INTERPRET database, plus a test set of 40 patients (30 glioblastomas and 10 metastases) from three different centers in 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. For the test set, a parsimonious selection of five frequencies yielded an area under the receiver operating characteristic curve of 0.86, and an area under the convex hull of the receiver operating characteristic curve of 0.91. Moreover, these accurate results for the discrimination between glioblastomas and metastases were obtained using a small number of frequencies that are amenable to 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 TEs (short TE, 20–32 ms; long TE, 135–144 ms) is key to the successful discrimination between glioblastomas and metastases from single-voxel ^{1}H MRS. Copyright © 2011 John Wiley & Sons, Ltd.
Supporting information may be found in the online version of this article.

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

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 because of their radiological similarity (1).

The discrimination between glioblastoma and 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, whereas a solitary metastasis is the result of a systemic tumoral process, and its treatment ultimately depends on the origin and 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 MRI may be very similar for both types of tumor. Further diagnostic support can be obtained from so-called physiological MR techniques. Most use the infiltrative pattern of growth of glioblastomas

* *Correspondence to: A. Vellido, Departamento de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, C/Jordi Girona, 1–3, 08034, Barcelona, Spain.*
E-mail: avellido@lsi.upc.edu

[a] *A. Vellido, E. Romero*
Departamento de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Barcelona, Spain

[b] *M. Julià-Sapé, C. Majós, À. Moreno-Torres, C. Arús*
Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Cerdanyola del Vallès, Spain

[c] *M. Julià-Sapé, C. Arús*
Departament de Bioquímica i Biología Molecular (BBM), Unitat de Biociències Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Spain

[d] *M. Julià-Sapé, C. Arús*
Institut de Biotecnologia i de Biomedicina (IBB), Unitat de Biociències Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Spain

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

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

Abbreviations used: *AUC, area under the (ROC) curve; AUCCH, area under the (ROC) curve convex hull; FN, false negative; FP, false positive; FS, feature selection; GABA, γ-aminobutyric acid; Glu/Gln, glutamate/glutamine; MDSS, medical decision support system; NAA, N-acetylaspartate; ROC, receiver operating characteristic; SLP, single-layer perceptron; TN, true negative; TP, true positive; WHO, World Health Organization.*

to accomplish the differentiation. Metastasis is, in fact, an extra-axial process that does not infiltrate into 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. However, 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 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 determine whether certain characteristics of the focal enhancing mass could aid in the differentiation between glioblastoma and solitary metastasis. Our hypothesis was that tumoral processes originating in the brain should be, at least to some extent, different from those originating elsewhere, and that these differences should be found in the single-voxel (SV) ^{1}H MRS signal.

The existing literature (1,3–5) considers this differentiation problem by SV ^{1}H MRS to be 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 have addressed the problem of differentiation between glioblastoma and solitary metastasis have analyzed small patient samples, often from a single clinical source. Glioblastoma cases usually predominate in all the analyzed databases, but no active compensation procedure for the different prevalences of the two pathologies has been used. With some exceptions (5), the cited studies have 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 an approach risks throwing away large parts of the spectrum where useful discriminatory information may be present.

The use of test datasets (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 dataset. 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 the 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 two-dimensional chemical shift imaging 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. For rather obvious reasons, diagnostic decision making in neuro-oncology is an extremely sensitive matter. Taking into account that most diagnostic techniques in this domain must rely on noninvasive 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 (MDSSs), given that this type of data can be acquired and processed automatically (therefore becoming part of the 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 interpretability requirements is through data dimensionality reduction and, more specifically, through feature selection (FS). A thorough FS procedure is applied in this article to the problem of discriminating between metastatic brain tumors and glioblastomas on the basis of SV ^{1}H MRS from the multicenter, 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 than relevant ones in the SLP output prediction. 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 with some existing studies (22)] to obtain the best discrimination results from the combination of SV ^{1}H MRS data acquired at different TEs.

MATERIALS AND METHODS

The INTERPRET SV ^{1}H MRS database and MDSS

The available data are SV ^{1}H MR spectra acquired *in vivo* from patients with brain tumors. They are part of the multicenter, 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 centers 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 of the many available from the INTERPRET database) were as follows: (i) that the case had SV 1.5-T spectra acquired at both short and long TEs from a nodular region of the tumor; (ii) that the voxel was located in the same region in which a subsequent biopsy was obtained; (iii) that the spectra had not been discarded because of acquisition artifacts or other data quality reasons; and (iv) that a histopathological diagnosis was agreed among a committee of neuropathologists.

The analyzed data were acquired at short and long TEs. They included 78 glioblastomas [World Health Organization (WHO) 9440/3)] and 31 metastases (WHO 8000/6). Processing was performed with the INTERPRET Data Manipulation Software (http://gabrmn.uab.es/dms) (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 4.22 to 0.49 ppm. These 109 cases were used in the FS and classification procedure described in the next section.

TE is an influential parameter in ^{1}H MRS data acquisition. In short-TE spectra (typically acquired at 20–40 ms; in this study: stimulated echo acquisition mode, 20 ms; point-resolved spectroscopy, 30–32 ms), some metabolites are better detected [e.g. lipids, myo-inositol, glutamine (Gln) and glutamate (Glu)]. However, there may be numerous overlapping resonances (e.g. Glu/Gln at 2.2 ppm) which make the spectra difficult to interpret (25). The use of long TE (in this study: point-resolved spectroscopy, 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 short- or long-TE MRS to discriminate between different types of high-grade tumor. There is evidence, however, that the combined use of both TEs could be advantageous (22,25). In this study, the spectra acquired at short and long TE from the same patient (when both TEs were available) were combined through straight concatenation of the spectra, as in ref. (22). The combined TEs 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 the diagnosis and grading of brain tumors using *in vivo* SV ^{1}H MRS. It includes automated pattern recognition techniques (such as linear discriminant analysis classifiers), and the results corresponding to certain 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 to facilitate the incorporation of the results of pattern recognition analysis into an overall diagnostic procedure in which the possible algorithmic and mathematical intricacies are transparent to the clinician. The latest official release of the MDSS is version 3.0.2 (http://gabrmn.uab.es/dss) (23).

A test dataset of 40 cases (30 glioblastomas and 10 metastases) was kept aside to test the generalization capability of the selected model in the classification task. As mentioned in the Introduction, this is 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 (http://www.etumour.net) (26). Ethics committee approval for data accrual was gathered in the context of the INTERPRET and eTUMOUR projects.

FS and classification methods

As some of the studies briefly reviewed in the Introduction 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 a procedure should be sufficiently robust to yield a selection of frequencies that is not only relevant to the sample of patients under study, but is 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 with 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 multilayer perceptrons (27,28) or linear support vector machines (29) for several reasons: the former may be computationally too expensive for the number of MRS frequencies analyzed in the available database; multilayer perceptron parameters are also more difficult to adjust appropriately, and the saliency (relevance) of every feature is also likely to be more independent for SLP than for linear support vector machines. 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.

Two components of the FS procedure must be specified explicitly: 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 do relevant features, 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, that is to compensate for the different numbers of cases corresponding to each of the two analyzed pathologies. In doing so, this prevented the SLP favoring the accurate classification of only the most prevalent class. In our experiments, the balancing process involved modification of the back-propagated error of the metastases, which was multiplied by the ratio of glioblastomas to metastases. This has a similar effect to oversampling the least frequent class (metastases).

The FS procedure is decremental and 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 following sections were the best obtained for the test set.

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 (TPs) correspond to correct metastasis predictions and true negatives (TNs) correspond to correct glioblastoma predictions; likewise, false positives (FPs) correspond to false metastasis predictions and false negatives (FNs) correspond to 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 receiver operating characteristic (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 the qualification of 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), whereas the AUCCH can, at most, slightly overestimate such quality.

RESULTS

Maximum overall accuracies of 85% in the test set were obtained for several, extremely parsimonious, selected combinations of long- and short-TE 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%. It should be noted 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 ppm), they are selected at both TEs. 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 long- or short-TE data separately. Using long TE, a maximum accuracy of 77.5% in the test set was achieved. Using only short TE, the result decreased to 75%. Removing the short-TE features from the selected subsets listed in Table 1 (one frequency in the first three subsets and two in the fourth) reduced the accuracy from 85% to 77.5% (subset 1), 80% (subset 2) and 82.5% (subsets 3 and 4). The removal of the long-TE 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:

$$y(\mathbf{x}) = \tanh[a(\mathbf{x})] \qquad [1]$$

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 unity 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 –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 Supporting information. A simple spreadsheet that can be used to obtain predictions for new cases is also available online (http://gabrmn.uab.es/GBM-MET-formula.xls).

Detailed test predictions, together with sensitivity and specificity values [corresponding to confusion matrices at the mid-range classification threshold of $y(\mathbf{x}) = 0$], AUC and AUCCH 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 results in the diagnostic prediction with the test set. The best AUC value of 0.86 and AUCCH value of 0.91 must be compared with an AUC value of 0.84 reported in ref. (4) using the lipid peak area ratio, but no independent test set. The results should also be compared with those reported by García-Gómez *et al*. (5). In that study, for short-TE spectra acquired at 1.5 T, the best results for a balanced test

Table 1. The four subsets of frequencies from the concatenation of long-TE and short-TE data for which an accuracy of 85% in the test set was achieved in the experiments with the single-layer perceptron (SLP). Frequencies expressed in parts per million (ppm) (rounded to two decimal places). Letter prefixes stand for long-TE (L) and short-TE (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	$a(\mathbf{x}) = 3.473 - 0.548\text{L}2.32 + 0.484\text{L}2.29 - 0.493\text{S}2.32$
2	$a(\mathbf{x}) = 1.383 - 0.088\text{L}3.01 - 0.377\text{L}2.32 + 0.250\text{L}2.29 + 0.153\text{L}2.02 - 0.261\text{S}2.17$
3	$a(\mathbf{x}) = -0.612 + 0.349\text{L}3.42 - 0.214\text{L}3.36 - 0.049\text{L}3.01 - 0.436\text{L}2.32 + 0.453\text{L}2.29 - 0.160\text{S}2.32$
4	$a(\mathbf{x}) = 0.671 + 0.090\text{L}3.01 - 0.375\text{L}2.32 + 0.249\text{L}2.29 + 0.156\text{L}2.02 - 0.205\text{S}2.17 - 0.053\text{S}2.15$

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 subsets as in Tables 1 and 2. Second column: total number of correctly classified cases (CCCs) out of 40. Third column: true positives (TPs) and negatives (TNs), and false positives (FPs) and negatives (FNs). Fourth column: corresponding sensitivity (Sen) and specificity (Spe) as a percentage. Fifth and sixth columns: area under the receiver operating characteristic curve (AUC) and area under the receiver operating characteristic curve convex hull (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

set were obtained using peak integration techniques and a linear discriminant analysis classifier. An error rate of 0.22 and a corresponding balanced error rate of 0.21 were reported. From the results in Table 3, an error rate of 0.15 and a balanced error rate of 0.13 were achieved in our experiments.

It should be noted that the fourth solution is almost identical to the second, but for the addition of short-TE 2.15 (contiguous to the also selected short-TE 2.17). Therefore, and following a *lex parsimoniae* criterion, the second solution, including five frequencies, was chosen for implementation in the INTERPRET computer-based MDSS described in The INTERPRET SV ^{1}H MRS database and MDSS section. 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 three-dimensional data visualization for this best solution is not possible, as it consists of five frequencies. Instead, a visualization of the SLP predictions y can be provided, as in Fig. 2. In this figure, the further is the case from the threshold $y(\mathbf{x}) = 0$ and the closer to the left limit $y(\mathbf{x}) = -1$, the more confident is the glioblastoma prediction. Correspondingly, the further is the case from the threshold $y(\mathbf{x}) = 0$ and the closer to the right limit $y(\mathbf{x}) = 1$, the more confident is the metastasis prediction.

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

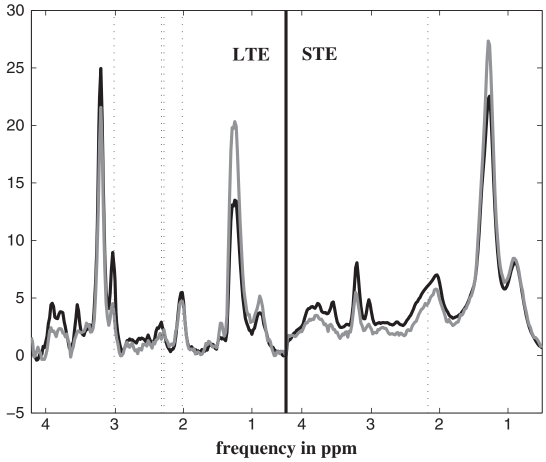

Figure 1. Location of the frequencies of the best subset selected (L2.32–L2.29–S2.17–L2.02–L3.01) on the long-TE + short-TE concatenated dataset used in the feature selection (FS) procedure [single-layer perceptron (SLP) training data]. Long-TE spectra on the left (LTE) and short-TE spectra on the right (STE). Mean metastasis amplitudes are shown as a full gray line and mean glioblastoma amplitudes as a full black line. Selected frequencies are shown as broken vertical lines. The vertical axis is unlabeled, as it corresponds to normalized intensity arbitrary units.

DISCUSSION

The importance of using a test set to qualify the results must again be stressed at this point. If only the original INTERPRET data had been used to create the SLP prediction models, there

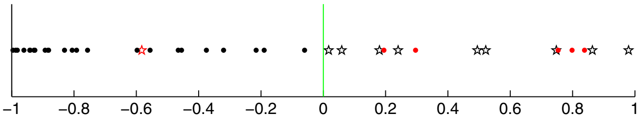

Figure 2. Visual representation of the single-layer perceptron (SLP) predictions $y(\mathbf{x})$ for the test set, corresponding to the best selection of frequencies, as described in the text. True positives (correctly classified metastases) are represented as black stars, false negatives as red stars (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 medical decision support system (MDSS).

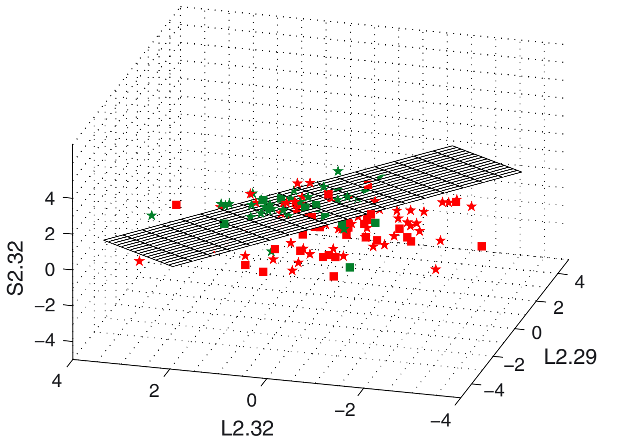

Figure 3. Visual representation of the data (amplitudes) corresponding to the subset of three selected spectral frequencies (L2.32–L2.29–S2.32) from the long-TE + short-TE dataset, as described in Table 1. Each of the axes is one of the frequencies, as labeled. Glioblastomas are represented in red and metastases in green. The data used to create the classifier are represented by stars, and the 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 single-layer perceptron (SLP) is a linear classifier, the decision surface is a plane. A case that falls within this surface would correspond to an SLP prediction $y(\mathbf{x}) = 0$. It should be noted that the Euclidean distance from the points to the differentiation plane can be used directly as a measure of our confidence on the diagnostic prediction provided by the classifier.

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. It should be noted that the signs of the coefficients in the formulae reported in Table 2 remain stable when the same frequencies are selected in different experiments. It should also be noted that all selected subsets include frequencies from both TEs, although with a clear preponderance of long-TE frequencies. This is a definite indication that the combination of TEs is necessary for the improvement of the differentiation capabilities of the model. The results reported in the previous section show that the combination of TEs yields better results that the separate use of either TE. Furthermore, the performance has been shown to deteriorate if frequencies of either TE are removed from the subsets selected when both TEs 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 includes a mechanism to actively compensate for the different tumor type prevalences. This is explicitly reflected by the excellent balanced error rate 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 long-TE frequency at 3.01 ppm mostly represents total creatine, whereas the long-TE frequency at 2.02 ppm, depending on the size, infiltrative nature and degree of necrosis of the tumor, will mostly be contributed by either *N*-acetylaspartate (NAA) (as a result of partial volume effects) or, when mobile lipid resonances are apparent at *c.* 1.3 and 0.9 ppm, by the –CH = CH–**CH$_2$**– methylene group of the fatty acyl chain of these mobile lipids, because of their low *T*2. In ref. (1), a qualitative analysis concluded that the presence of intratumoral creatine is a marker for gliomas, whereas its absence might be an indicator of metastasis. In this respect, ref. (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 Fig. 1.

However, quantification from *in vivo* MRS data using the LCModel approach (4) failed to find significant differences in creatine content between cases of glioblastoma ($n = 23$) and metastasis ($n = 24$), although further work from the same institution (37) using ^{1}H high-resolution magic angle spinning analysis of brain tumor biopsies did indeed find higher creatine in glioblastomas (2.99 ± 0.39 mM, $n = 24$) than in metastases (1.24 ± 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 lipid contents, especially at long TE, will contribute to an apparent metabolite content decrease in metastases *versus* glioblastoma after normalization.

A definite lipid signal was also concluded in ref. (1) to indicate cellular necrosis in glioblastoma and metastasis, whereas no lipid signal at short TE should lead to the exclusion of a diagnosis of metastasis.

Finally, 2.32 and 2.29 ppm at long TE, together with 2.17 ppm at short TE, may show contributions from Glu/Gln at 2.32 and 2.17 ppm and γ-aminobutyric acid (GABA) at 2.29 ppm (38), whereas 2.02 ppm at long TE will have a strong contribution from 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 for 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 ^{1}H 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 Figs 2 and 3. This visualization can be an important element in facilitating 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 discrimination between glioblastomas and metastases, based on a multivoxel approach (2,40). These have the advantage of not compelling the radiologist to decide *a priori* which is the best placement location for the sampled volume. They also use differences in the infiltrative pattern among glioblastomas and metastases, and the corresponding peritumoral MRS pattern differences, to discriminate between the two tumor types. Moreover, even though a multivoxel examination can easily detect the presence of infiltration, as shown by Server *et al.* (2), the same study can also detect differences not only in the peritumoral edema but also in the NAA/creatine ratio of the long-TE results in the tumoral core [a higher ratio in metastasis (1.43 ± 1.09) than in glioblastoma (0.87 ± 0.89)], indicating that differences in the spectral pattern also exist. Although certainly advantageous, this approach is not free from controversy (41,42) and, moreover, not all clinical centers are equipped to perform good-quality multivoxel data acquisition and post-processing. Therefore, our approach based on SV ^{1}H MRS should be of practical use, especially in clinical settings in which no multivoxel analysis is available, or in cases in which a single mass is located near the skull. Mixing both approaches with a focus on the peritumoral 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 FPs.

Analysis of the misclassified cases from the test set

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

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

The other five misclassified cases were glioblastomas. Three showed a marked lactate signal in the long-TE spectrum (Fig. 5). However, there were other reasons for their misclassification: the three cases showed atypically low levels of Glu/Gln at short TE; moreover, et2042 showed atypically low values for all the features selected at long TE, et3010 showed high levels of NAA at long TE and et2054 showed low levels of total creatine.

A further two cases, et3496 and et3194, showed an unusual pattern: the short-TE spectrum of case et3496 (Fig. 6, left) was

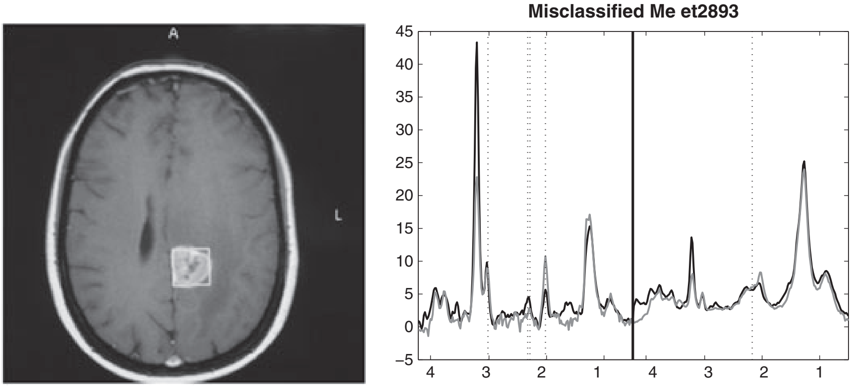

Figure 4. Single misclassified metastasis: case et2893. Left: voxel correctly placed over hyperintense area. Right: spectrum as a full black line. Mean metastasis amplitudes are shown as a full gray line. The long-TE part shows an unusually high choline-containing compound peak. At short TE, the choline-containing compound signal is also the second highest intense peak after necrotic lipids. The high choline signal at long TE will produce an apparent decrease in other signals on UL2 normalization. It should be noted, however, that, given that the classification procedure is based on the five selected features, the misclassification of this case seems to be caused mainly by an atypically low level of *N*-acetylaspartate (NAA) at long TE and high levels of glutamate/glutamine (Glu/Gln) and γ-aminobutyric acid (GABA) at long TE. The locations of the frequencies of the best subset selected from the long-TE + short-TE concatenated data (L2.32–L2.29–S2.17–L2.02–L3.01) are represented as broken vertical lines. Long-TE data on the left and short-TE data on the right of the middle vertical full black line.

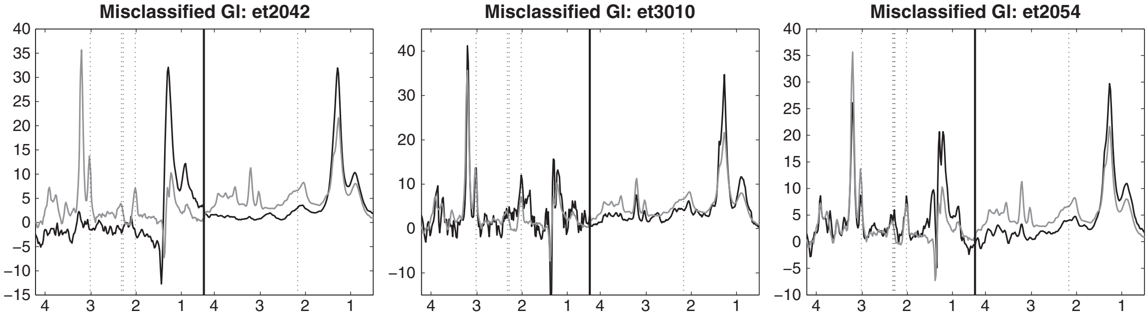

Figure 5. Three misclassified glioblastomas with marked lactate signal at long TE. Representation as in Fig. 4. Left: at long TE, case et2042 shows overlapping lipid–lactate signals. Middle: also at long TE, case et3010 shows a large inverted lactate doublet overlapping some contribution from lipid at 1.28 ppm. 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 tumor. 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 overlapping lipid–lactate peaks and the choline-containing compounds are the most intense signals in the spectrum.

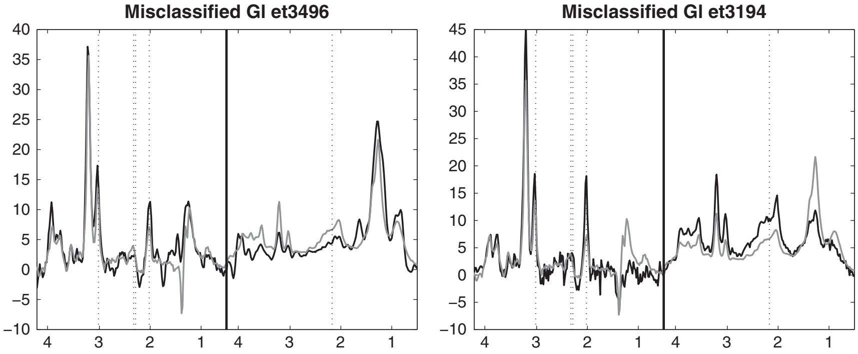

Figure 6. Two further misclassified glioblastomas: cases et3496 (left) and et3194 (right), as described in the text. Representation as in Figs 4 and 5.

judged by the expert spectroscopists of eTUMOUR as having poor, unexpected signals, possibly because of scalp lipid contamination and poor phasing. In case et3194 (Fig. 6, right), spectra at both TEs were more characteristic of a lower grade, although the histopathological diagnosis was clearly glioblastoma, with three consulting pathologists, as well as the originating pathologist, agreeing on the diagnosis. This may be a result of heterogeneity within the tumor, which has been described in refs. (43,44). This is corroborated by the corresponding imaging, which reveals that the tumor consists of 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 indicates 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 suffers from deficient water suppression or low signal-to-noise ratio, the classification results are likely to be unreliable.

CONCLUSIONS

In the Introduction section, it was hypothesized that tumoral processes originating in the brain should be, at least to some extent, different from those originating elsewhere, and that these differences should be detected in the SV ^{1}H MRS signal. The reported experimental results confirm this hypothesis and show that a robust FS 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 multicenter independent test set of cases. The combination of SV ^{1}H MRS acquired at different TEs is crucial to this classification success, although long-TE 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 a prediction formula could be applied directly 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 TIN200913895-C02-01 and SAF2008-03323 projects. Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN) is an initiative of Instituto de Salud Carlos III (ISCIII), Spain, which is co-funded by EU FEDER. The authors gratefully acknowledge former INTERPRET, eTUMOUR and HealthAgents European project partners. Data providers: Dr J. Capellades (Institut de Diagnòstic per la Imatge, Spain); Dr F. A. Howe and Professor J. Griffiths (St George's University of London, UK); Professor A. Heerschap (Radboud University (RU), Nijmegen, The Netherlands.); Professor L. Stefanczyk and Dr J. Fortuniak (Uniwersytet Medycznyw Lodz, Lodz, Poland); Dr J. Calvar (Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia, Buenos Aires, Argentina). Data curators: Dr A. P. Candiota, Dr I. Olier, Ms T. Delgado, Ms J. Martín, Ms M. Camisón and Mr A. Pérez (all from Grup d'Aplicacions Biomèdiques de la Ressonància Magnètica Nuclear (GABRMN), Universitat Autònoma de Barcelona (UAB) (GABRMN-UAB)).

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: 1784–1791.
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. Phys. 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: Roli F, Kittler J (eds). *Multiple Classifier Systems*, *Lecture Notes in Computer Science*. Springer: Berlin; 2002, 2364: 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, 2010. Available at: http://homes.esat.kuleuven.be~jluts/phd.pdf. [Accessed on 21 July 2011].
16. Tsuchiya K, Fujikawa A, Nakajima M, Honya K. Differentiation between solitary brain metastasis and high-grade glioma by diffusion tensor imaging. Br. 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. 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. 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. Phys. 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. 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 17th Scientific Meeting ISMRM*, Honolulu, HI, 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 classification 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. Available at: http://hdl.handle.net/10803/3525 [Accessed on 21 July 2011].
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 29th Annual International Conference 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 Oncol. 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.

APPENDIX

Feature selection (FS) with classification using single-layer perceptrons (SLPs)

The FS problem can be defined as follows: given a set of d features, let us select a subset that performs 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 remainder of the Appendix, the constituent elements of the proposed FS procedure associated with 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:A1

$$y_j = g\left(\sum_{i=1}^{d} x_i\cdot\omega_{ji} + b_j\right) \qquad \text{(A1)}$$

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 SLPs 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:

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

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 do relevant features. 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 Equation [A1] 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:

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

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 an 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, refs. (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 was removed at every step of the algorithm according to the evaluation measure. As 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:

   (i) train an SLP with the remaining features;
   (ii) compute the saliency of every feature;
   (iii) remove 50% of the remaining features.

For every feature subset obtained, estimate its generalization performance through five-fold cross validation. From all the results, keep the previous to the best result 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, except for:

   (i) the initial feature subset is that obtained in the first phase;
   (ii) at every step, 20% of the remaining features are removed.

(3) The third phase is similar to the second, except for:

   (i) the initial feature subset is that obtained in the second phase;
   (ii) at every step, one feature is removed.