Automated classification of brain tumours from short echo time *in vivo* MRS data using Gaussian Decomposition and Bayesian Neural Networks

CrossMark

Carlos Arizmendi [a], Daniel A. Sierra [b], Alfredo Vellido [c,d], Enrique Romero [c,*]

[a] Universidad Autònoma de Bucaramanga, Colombia
[b] Electrical, Electronics and Telecommunications Engineering School, Universidad Industrial de Santander, Colombia
[c] Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Spain
[d] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Cerdanyola del Vallès, Spain

ABSTRACT

Neuro-oncologists must ultimately rely on their acquired knowledge and accumulated experience to undertake the sensitive task of brain tumour diagnosis. This task strongly depends on indirect, non-invasive measurements, which are the source of valuable data in the form of signals and images. Expert radiologists should benefit from their use as part of an at least partially automated computer-based medical decision support system. This paper focuses on Magnetic Resonance Spectroscopy signal analysis and illustrates a method that combines Gaussian Decomposition, dimensionality reduction by Moving Window with Variance Analysis and classification using adaptively regularized Artificial Neural Networks. The method yields encouraging results in the task of binary classification of human brain tumours, even for tumour types that have seldom been analyzed from this viewpoint.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Magnetic Resonance Spectroscopy (MRS) is a non-invasive measurement technique that can help diagnosing human brain tumours that remain ambiguous after clinical investigation. Being the brain the subject of study, this non-invasiveness limits the risks of functional and cognitive damage associated to surgical intervention.

MRS resorts to the use of strong magnetic fields for the generation of energy exchanges between an external magnetic field and the protons that are present in abundance in all living tissue (Liang & Lauterbur, 1999). A radio-frequency machine detects these energy exchanges, which are then coded using sophisticated mathematical software. The result is a signal in the frequency domain that peaks at specific frequencies or frequency bands that are known to correspond to the the sites of resonance for specific chemical and biochemical components of the tissue. The signal profile is a quantitative indicator of the presence of these components: those substances that are present in big quantities in the tissue will yield higher peaks than those present in lower concentrations. As a result, *in vivo* MRS enables the quantification of metabolite concentrations in the brain tissue, providing experts with a detailed signature of localized tissue metabolism that should shed light on tumour characterization.

This technique has evolved rapidly over the past decades, helping to discover informative correlations between brain tumour types and spectral patterns. Although MRS has a great potential for clinical decision making support, there are several factors that make MRS signal processing difficult (Cousins, 1995) and thus hamper its application: signal degradation related to the sensitivity of the technique, thermal noise from the sample, noise from the electronic components, technical limitations when measuring the *in vitro* tissue, as well as time limitation during measurement. In general, *in vivo* MRS signals are characterized by a low signal-to-noise ratio (SNR), strongly overlapping spectral components (Tate et al., 1996) and the presence of the residual water peak in the case of proton MRS (^{1}H-MRS), which even after pre-saturation, dominates the proton free induction decay, causing baseline distortions in the frequency domain (particularly for resonances closer to the water peak). Furthermore, radiologists are more commonly trained in the use and interpretation of imaging techniques than in MRS interpretation, limitation compounded by the fact that this task requires considerable experience. The automated processing, analysis and interpretation of the MRS spectra is therefore highly valuable. Additionally, the intrinsically high dimensionality of the spectra, the presence of noise and artifacts, and the low amount of data available for specific pathologies (i.e., for specific brain tumour

* Corresponding author. Tel.: +34 934137796.
  E-mail addresses: carizmendi@unab.edu.co (C. Arizmendi), dasierra@uis.edu.co (D.A. Sierra), avellido@lsi.upc.edu (A. Vellido), eromero@lsi.upc.edu (E. Romero).

types) often complicate their diagnostic-oriented classification (Arizmendi, Vellido, & Romero, 2009).

This paper aims to contribute new tools for the automated classification analysis of brain tumours from MRS data. The proposed method is a combination of signal processing, variable selection, and machine learning classification techniques for tumour diagnosis. It is tested using data from the international, multi-center INTERPRET database (Julià-Sapé, Acosta, Mier, Arùs, & Watson, 2006).

The MR spectrum, as expressed over a frequency range, is the result of computing the Fast Fourier Transform (FFT) of the FID signal, and we can therefore safely assume that is the sum of a set of tones at the corresponding resonance frequencies. The first step of the proposed method consists in breaking down the spectra in individual components using the Gaussian Decomposition (GD) technique. Each tone can be approximated with a Gaussian function, so that the spectrum as a whole is approximated through a sum of Gaussians. From each of them, the coefficients of amplitude, standard deviation and translation parameters are used. These can be associated to specific metabolites to improve the interpretability of the results, which is a key goal in medical practice (Vellido, Martín-Guerrero, & Lisboa, 2012). The second step of the method involves dimensionality reduction in the form of a feature selection filter method known as Moving Window with Variance Analysis (MWVA) (Arizmendi et al., 2009). In the third and final step, the remaining relevant information is fed to an Artificial Neural Network (ANN) classifier with Bayesian regularization (MacKay, 1992).

The main technical contribution of the current study is the combination of a feature extraction technique such as GD, never used for MRS analysis and rarely in medical imaging studies, with a subsequent feature selection step of the analysis that should ease the interpretation of the results. The proposed combination of techniques is shown in the reported experiments to yield high diagnostic classification accuracy for a broad range of brain tumour pathology discrimination problems, some of which have seldom been analyzed from this viewpoint. These results come to justify the adequacy of the proposed combination of methods.

2. Materials and methods

2.1. Database

This study used MRS available from a database hosted at the *Grup d'Applications Biomèdiques de la Ressonància Magnètica Nuclear* (GABRMN) at *Universitat Autònoma de Barcelona* (UAB), in Spain. This database was created under the framework of the European project INTERPRET, an international collaboration of centers from four different countries. More specifically, the data were collected by CDP (Centre Diagnòstic Pedralbes, Barcelona, Spain), IDI (Institut de Diagnòstic per la Imatge, Barcelona, Spain), SGHMS (St. George's Hospital Medical School, London, UK) and UMCN (University Nijmegen Medical Center, Nijmegen, Netherlands).

The original criteria for the selection of cases to be included in the database were: (a) that each considered case had a single voxel short TE, 1.5 T spectrum acquired from a nodular region of the tumour; (b) that the voxel was located in the same region as where subsequent biopsy was obtained; (c) that the short TE spectrum had not been discarded because of acquisition artefacts or other reasons; and (d) that a histopathological diagnosis was agreed among a committee of neuropathologists. In those cases in which the spectra were obtained from normal volunteers without the pathology, or corresponded to abscesses or clinically proven metastases, biopsy was not required. For further details on data acquisition and processing, and on database characteristics, we

Table 1
Analyzed classes from the INTERPRET database.

	Tumour class	Number of cases
a2:	Astrocytomas, grade II	22
a3:	Astrocytomas, grade III	7
ab:	Brain abscesses	8
gl:	Glioblastomas	86
ly:	Lymphomas	10
me:	Metastases	38
mm:	Meningiomas grade I	58
no:	Normal cerebral tissue, white matter	22
oa:	Oligoastrocytomas grade II	6
od:	Oligodendrogliomas grade II	7
pn:	Primitive neuroectodermal tumours and medulloblastomas	9

refer readers to INTERPRET project (2014) and Julià-Sapé et al. (2006). Class labelling was performed according to the World Health Organization (WHO) system for diagnosing brain tumours by histopathological analysis of a biopsy sample (Kleihues et al., 2002). The database includes, amongst others, the types of tumours (as well as normal tissue and brain abscesses) listed in Table 1.

The database includes proton MRS (^{1}H-MRS) with removal of water obtained using PRESS and STEAM sequences. The spectra were acquired at different times of echo: long time of echo (LTE, 135–136 ms, 266 patients) and short time of echo (STE, 30–32 ms, 304 patients). The time repetition (TR) was set between 1.50 and 2.02 ms; the spectral bandwidth, from 1000 Hz to 2500 Hz, and the total number of spectral frequencies was set to 512.

2.2. Basics of Gaussian decomposition

Several studies have investigated the problem of the extraction of relevant information for the task of diagnostic classification from MRS. The presence of noise and artifacts in the spectra (Vellido, Lisboa, & Vicente, 2006; Vellido et al., 2009) and the strong overlapping between spectral peaks (De Graaf & Bovee, 1990), amongst other causes, are known to make the extraction of discriminant information between classes difficult. This is especially true for techniques used for brain tumour discrimination such as Principal Component Analysis (PCA) (Devos et al., 2004) or the Discrete Wavelet Transform (DWT) (Arizmendi, Vellido, & Romero, 2012; Tate et al., 1996), but also for source extraction techniques such as Independent Component Analysis (ICA) and Non-negative Matrix Factorization (NMF), which tend to obtain sources that reflect tissue types instead of tissue metabolites (Huang, Lisboa, & El-Deredy, 2003; Vilamala, Lisboa, Ortega-Martorell, & Vellido, 2013).

The technique of Peak Integration (PI) has also been suggested as a candidate to overcome the problem of peak overlapping (García-Gómez et al., 2009; Hoch & Stern, 1996). However, the precise estimation of the peak integrals is difficult due the existence of a non-zero baseline, overlapping between peaks, and the discrete nature of the spectra (Devos et al., 2004). Because of the strong overlapping, the use of prior knowledge for peak selection becomes mandatory (Lukas et al., 2004).

The MR spectra is the summary of a set of tones at the corresponding resonance frequencies for the metabolites present in the sample (Elguero, Alkorta, Pardo, Claramunt, & Santa María, 2004). In signal preprocessing, the approximation of each tone by a Gaussian is a plausible procedure. This way, the spectrum can be approximated as the sum of a set of Gaussian functions. The method proposed in the current study includes the use of the GD signal processing technique to break down a given MR spectrum into its

Gaussian components. Despite having been used in a variety of fields, such as astronomy (Haud, 2010; Haud & Kalberla, 2007) and laser instrumentation (Wagner, Ullrich, Ducic, Melzer, & Studnicka, 2006), to name a few, GD has rarely been applied in the medical domain in general and in MR in particular. In Suzuki, Sakai, and Jara (2006), GD was used to fit the pixel histograms of relaxation times for MRI. A similar approach can be found in Oros-Peusquens, Laurila, and Shah (2008). To the best of the authors' knowledge, this technique has never been used to analyze MRS data.

We describe each of the constituent tones obtained by GD through its coefficients of amplitude, standard deviation and translation. These descriptive coefficients can then be associated to specific metabolites, thus increasing the interpretability of the results. Unlike PI, GD does not require prior knowledge to extract the individual information of the metabolites. It also overcomes the problem of the separation of overlapping spectral components. For all the reasons stated above, we consider GD an adequate tool for the investigation of MRS.

Consider a typical MRS section as that illustrated in Fig. 1, with two resonance peaks associated with specific metabolites. Assume that the interval of spectral signal delimited by the $[\lambda_1, \lambda_2]$ interval is modeled by superposition of n Gaussians, as:

$$F(x,P) = \sum_{i=1}^{n} A_i e^{-k_i(x-\mu_i)^2}, \quad \lambda_1 \leqslant x \leqslant \lambda_2 \tag{1}$$

$$F(x) = 0, \quad x < \lambda_1 \quad \text{or} \quad x > \lambda_2 \tag{2}$$

where $k_i = -1/2\sigma_i^2$, and $P_i = \{A_i, k_i, \mu_i\} \in \Re^{3+}$ contains information of the amplitude (A_i), standard deviation (σ_i), and translation (μ_i) of each Gaussian. Therefore, the set $P = \{P_1, P_2, \ldots P_n\}$ contains all the parameters corresponding to the n Gaussians used to fit the MRS signal.

Let $Y = \{Y_1, Y_2, \ldots, Y_m\}$ be the observed function (observed MRS) for a data set $X = \{x_1, x_2, \ldots, x_m\}$ and $\widehat{Y} = \{\widehat{Y}_1, \widehat{Y}_2, \ldots, \widehat{Y}_m\} = \{F(x_1, \widehat{P}), F(x_2, \widehat{P}), \ldots, F(x_m, \widehat{P})\}$ the estimated approximated function for a set of parameters $\widehat{P}$. GD aims to solve the optimization problem:

$$\min_{\widehat{P}} \left(\varphi(x, Y, \widehat{P}) \right) = \sum_{i=1}^{m} (Y_i - \widehat{Y}_i)^2 \tag{3}$$

As the x wavelenght points are the same in Y and $\widehat{Y}$, the function to minimize becomes:

$$\min_{\widehat{P}} \left(\varphi(\widehat{P}) \right) = \sum_{i=1}^{m} \left(Y_i - \widehat{Y}_i(\widehat{P}) \right)^2 \tag{4}$$

Therefore, the optimization objective is to find the $\widehat{P}$ that minimizes $\varphi(\widehat{P})$. This can be seen as a non-linear least squares problem, for which plenty of deterministic and nondeterministic optimization methods exist (Barton & Valdés, 2008). To minimize $\varphi(\widehat{P})$, we used the Trust-Region Dogleg (TR) method, because it yields better accuracy than the Levenberg–Marquardt (LM). Fig. 2 exemplifies the results of the LM and TR algorithms in solving the problem of fitting the sum of two artificially generated Gaussians with added Gaussian noise. While LM does not fit the data properly, TR produces an excellent fit of the signal.

In this study, the Matlab® Curve Fitting Toolbox was used to decompose the spectrum into its constituent tones (i.e., to obtain the coefficients of amplitude, standard deviation and translation), by minimizing the mean squared error (MSE) between the original spectra and the fitted ones.

2.3. Moving Window with Variance Analysis

One of the main characteristics of the analyzed MRS is their high dimensionality, due to the fact that each measured frequency is treated as a data feature (or attribute/variable). It is well-known that only a few of these frequencies (or short intervals of frequencies) are associated to identifiable metabolites that may be present in the tumoural tissue (Govindaraju, Young, & Maudsley, 2000). On the other hand, it is also well investigated that some of these metabolites are especially relevant as tumour markers. Nevertheless, the relative relevance of individual metabolites or groups of them for the automatic discrimination of MRS spectra according to tumour types and grades is still an open multivariate data analysis problem (González-Navarro et al., 2010). For this reason, the
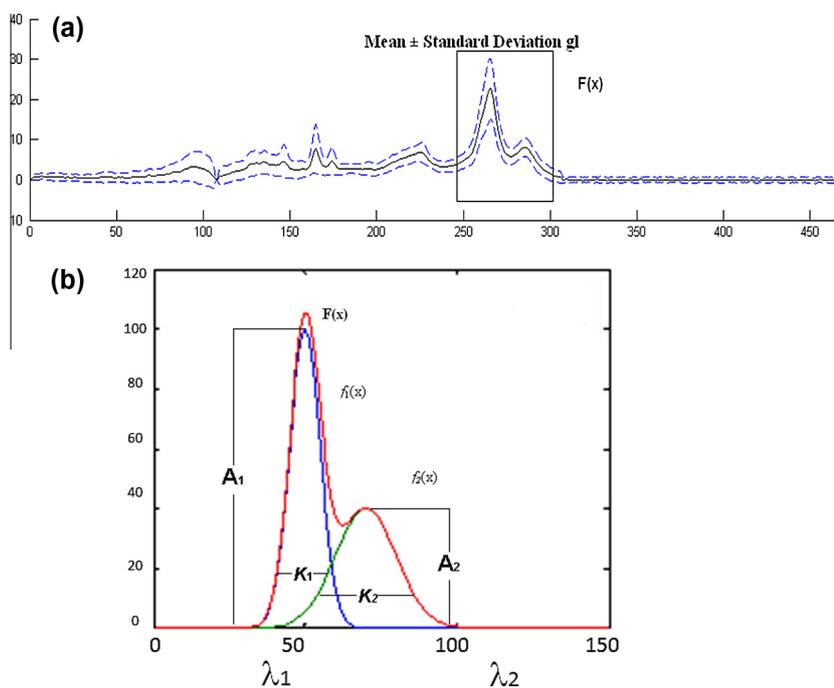
Fig. 1. *Left*: Mean ± standard deviation of the brain tumour spectrum $F(x)$ acquired by MRS for a Glioblastoma case. *Right*: Gaussian Dissection of $F(x)$ in $f_1(x)$ and $f_2(x)$.
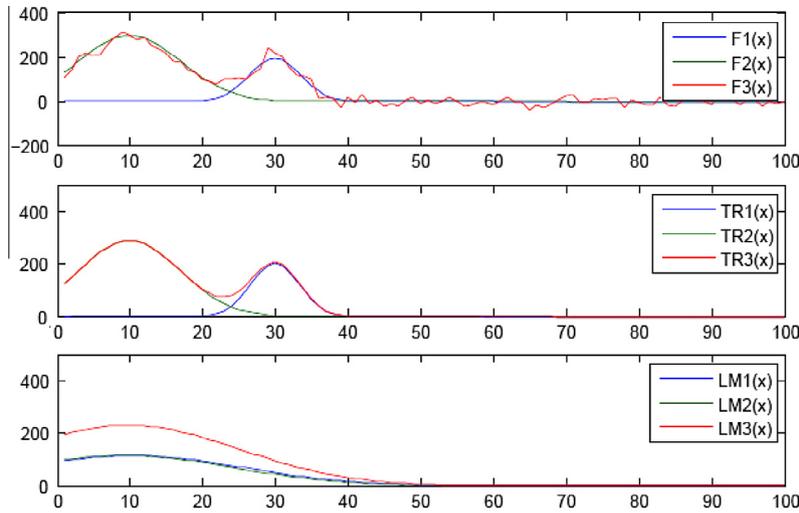
Fig. 2. $F1(x)$ and $F2(x)$ are artificially generated Gaussians with Gaussian noise added where $F3(x) = F1(x) + F2(x) +$ Gaussian noise (the Gaussian noise has mean 0 and standard deviation 1). In turn, $LM1(x), LM2(x), TR1(x)$ and $TR2(x)$ are the adjusted Gaussians with the LM and TR algorithms, respectively; (i.e., $LM3(x) = LM1(x) + LM2(x)$ and $TR3(x) = TR1(x) + TR2(x)$).

development of robust and efficient feature selection techniques for this domain is of special importance.

In this paper, we use MWVA: a feature selection filter method proposed in Arizmendi et al. (2009), which consists of the combination of the Moving Window technique in conjunction with the calculation of a standard ratio Ω, defined as the quotient between the between-groups variance (BGV) and the within-groups variance (WGV) for a particular width w of the window. The reader is referred to Arizmendi et al. (2009) for more details on this method.

2.4. Aspects of the use of GD for the analysis of MRS

2.4.1. Selection of the optimal number of Gaussians

There is no general consensus about which methodology is the most adequate for the selection of the optimal number of Gaussians in the signal fitting process (Kalberla et al., 2005), due to the many aspects of the problem that must be considered (e.g., fitting technique, algorithms, methods to choose the starting points, etc.).

The following experiment with artificial data was devised: Five artificial Gaussians were generated and added together, resulting in one single smooth signal. Table 2 summarizes the coefficients of the five generated Gaussians in the experiment. This test signal was then reconstructed in different trials with a varying number of Gaussians (in trial 1, it was reconstructed with one Gaussian, in trial 2 with two Gaussians, and so on until completing 20 trials). In each reconstruction trial, three figures of merit were computed: the MSE, Preserved Energy (PE) and Power Distortion (PD). The results for these figures of merit are shown in Fig. 3 as a function of the number of Gaussians used to recover the signal. The PE is used as an index to gauge if the algorithm is fitting the spectra in an appropriate manner, comparing the area under the curve of the adjusted signal and the original one. If the PE is low, the algorithm is correctly fitting the signal.

Table 2
Coefficients of the 5 artificially generated Gaussians.

Gaussians generated	Amplitude	Std deviation	Translation
Gaussian 1	10	15	30
Gaussian 2	10	25	60
Gaussian 3	10	10	90
Gaussian 4	10	15	120
Gaussian 5	10	5	150

These results with a toy example indicate that neither MSE nor PD have monotonically decreasing trends. In addition, the selection of starting parameters (initial $\widehat{P}$) may lead to convergence towards local minima. In consequence, the fitting should be performed with as many trials as possible, increasing the number of Gaussians in each trial, and with different starting points.

Fig. 3 also shows that the PD is 24.12% when 5 Gaussians were used to fit the spectra, whereas a PD as small as 0.00028% was obtained when 14 Gaussians were used. The two cases are portrayed in Fig. 4, which is self-explanatory in terms of the difference in goodness-of-fit.

Table 3 summarizes the coefficients (amplitude (A), standard deviation (SD) and translation (T)) of the trials with 5 (the original number), 11, 14, and 15 Gaussians. Figures in bold indicate those parameters that best match the original ones. For the fitting with 14 Gaussians, a zero MSE is obtained. As reported in previous research (Haud, 2000; Kalberla et al., 2005), the algorithm uses more Gaussians than necessary to fit the signal properly, setting to zero those Gaussians that fit noise or that are irrelevant for the reconstruction.

In conclusion, both the PD and MSE are good criteria to choose the number of Gaussians of a fitting process. In contrast, PE is useful to observe the behaviour of the algorithm: If PE shows too many oscillations or is too low, this will be an indication that the algorithm is not fitting the signal properly. Regarding the number of Gaussians needed to fit the spectra, the algorithm tends to use more than necessary, although it does not necessarily yield a bad quality reconstruction.

2.4.2. Filtering and Baseline correction

The GD method separates an MR spectrum into its constituent tones, expressed by coefficients of amplitude, standard deviation, and translation of the corresponding Gaussian functions. These parameters must not be negative when the TR method is used, in order to obtain a good fitting. Therefore, a baseline correction is required to ensure positive values for all the components of the spectrum. In this study, the baseline correction consisted in adding the absolute value of the minimum negative amplitude to the spectrum.

As an additional preprocessing procedure, a half band wavelet filtering was carried out using the Biorthogonal 3.3 mother wavelet. The half-band filtering corresponds to picking up only the approximation coefficients of the first decomposition level (Mallat,
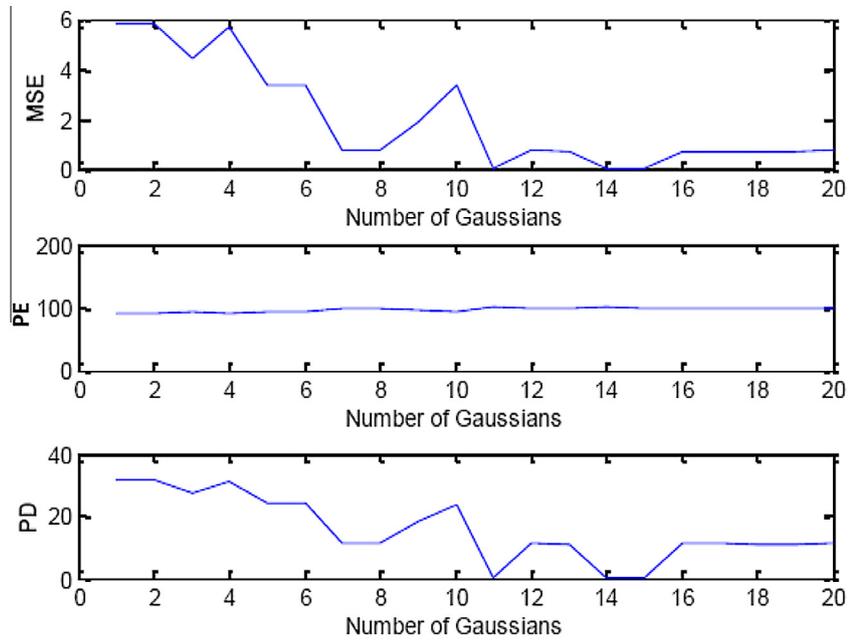
Fig. 3. MSE, PE and PD corresponding to the fitting of the 5 artificially generated Gaussians. The fit was carried out with a maximum of 20 Gaussians.
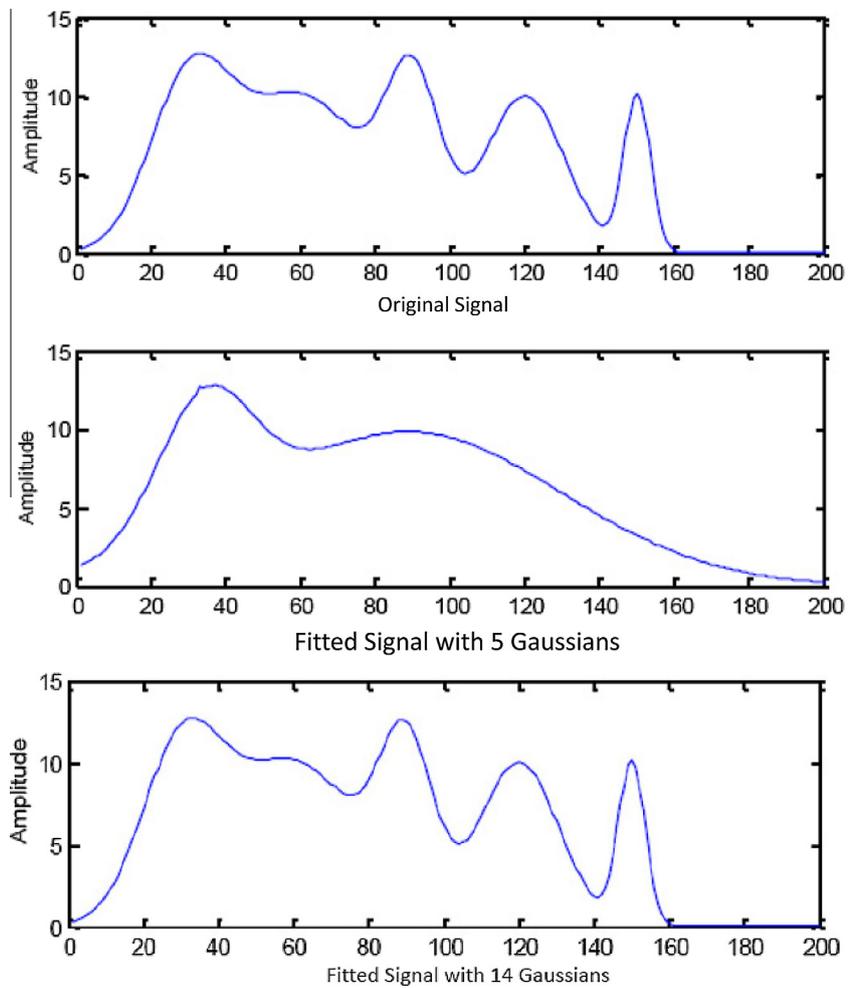
Fig. 4. The top figure corresponds to the original signal, whereas the middle figure corresponds to the signal fitted with 5 Gaussians and the bottom one to the signal fitted with 14 Gaussians.

Table 3
Coefficients for trials with different numbers of Gaussians corresponding to the lowest values of PD. The trials are ordered, from left to right, according to their PD, so that the trial with 14 Gaussians has the best PD, and the trial with 11, the worst one (out of the best trials considered). For comparison purposes, the trial with 5 Gaussians is also included (rightmost). Results in boldface indicate the Gaussians that are closer to the original ones.

14 Gaussians			15 Gaussians			11 Gaussians			5 Gaussians		
A14	SD14	T14	A15	SD15	T15	A11	SD11	T11	A5	D5	T5
0.00	0.60	12.77	0.08	0.18	9.53	**10.00**	**15.00**	**30.00**	**12.74**	**0.02**	**32.95**
10.00	**25.00**	**60.00**	10.00	15.00	30.00	12.15	0.03	32.55	**9.54**	**2.22**	**66.12**
10.00	**15.00**	**30.00**	10.15	0.10	59.67	3.02	0.14	48.42	**8.53**	**17.88**	**34.17**
10.18	0.07	52.55	7.94	0.00	69.30	9.76	0.15	63.51	**9.82**	**58.03**	**88.64**
9.63	0.10	65.32	**10.00**	**25.00**	**60.00**	**10.00**	**5.00**	**150.00**	**0.00**	**0.00**	**165.00**
2.85	0.06	75.73	**10.00**	**10.00**	**90.00**	**10.00**	**10.00**	**90.00**			
10.00	**10.00**	**90.00**	0.64	14.70	123.62	**10.00**	**24.99**	**60.00**			
4.67	0.04	104.42	10.01	0.02	120.06	**6.51**	**14.69**	**118.32**			
9.64	0.15	116.50	**9.40**	**14.96**	**119.76**	3.73	14.58	122.96			
10.00	**15.00**	**120.00**	2.20	4.89	148.98	0.12	0.07	165.25			
0.00	18.22	169.84	**7.92**	**4.96**	**150.28**	0.00	0.00	176.00			
10.00	**5.00**	**150.00**	0.00	0.00	168.00						
0.00	0.00	169.00	0.00	0.00	180.01						
0.00	0.00	0.00	12.45	0.13	13.50						
			10.24	0.02	48.08						

Spectrum of an astrocytoma 2

Spectrum of an oligodendodrioma
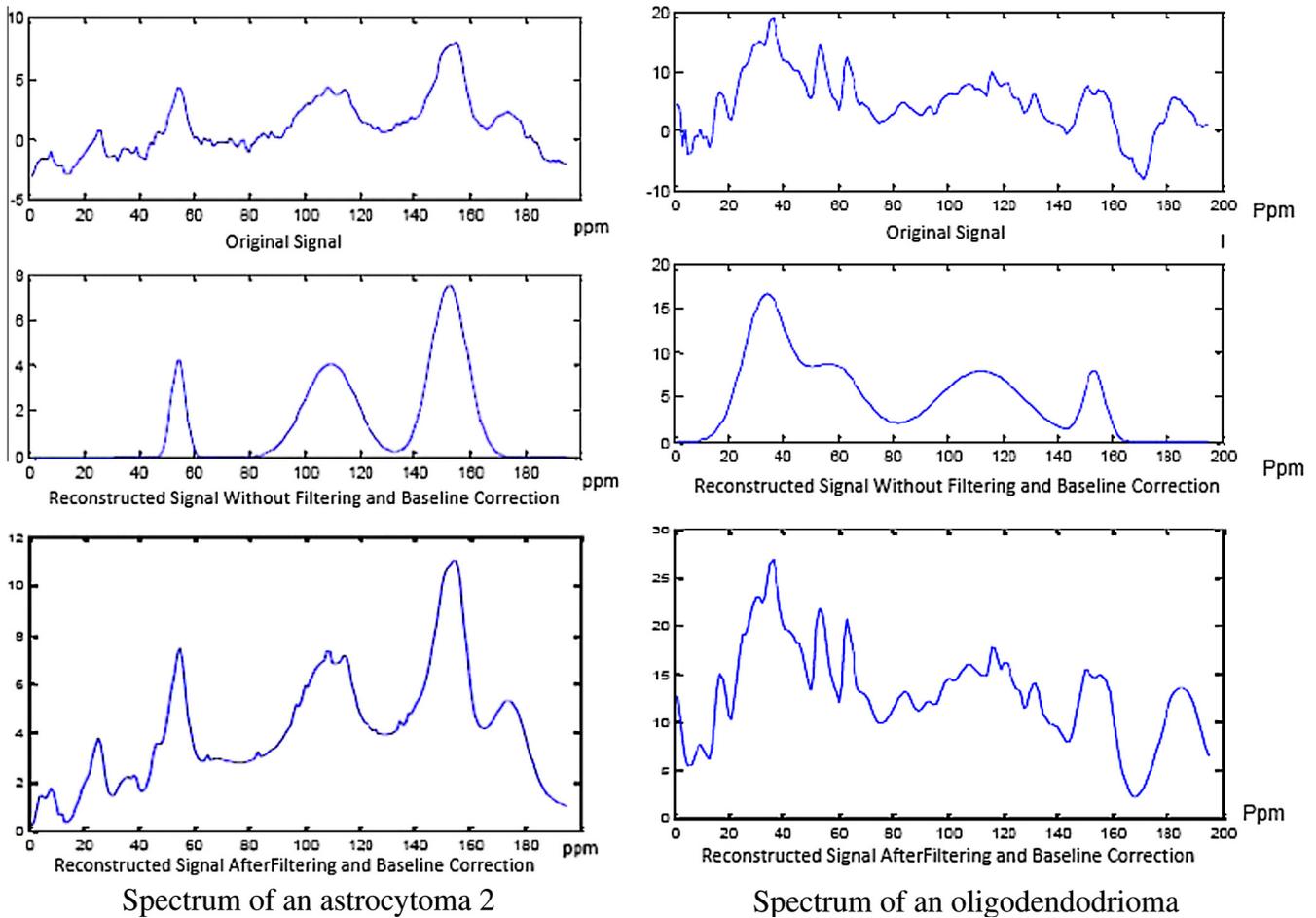
Fig. 5. *Top*: Original spectra; *Middle*: Reconstructed signal without preprocessing; *Bottom*: Reconstructed signal with filtering and baseline correction.

1999). In previous studies, the Biorthogonal 3.3 wavelet decomposed the MR spectra with the minimum number of coefficients, while keeping MSE and PD low and SNR high (Arizmendi, Hernandez-Tamames, Romero, Vellido, & Del Pozo, 2010).

Fig. 5 illustrates the signal processing procedure with two cases from the database, corresponding to different types of brain tumour. They were fitted with and without baseline correction and wavelet filtering. It is clear from these examples that the

Table 4
Fit convergence criteria for GD.

Parameter	Description	Default value
MaxFunEvals	Max. number of function (model) evaluations allowed	1000
MaxIter	Max. number of iterations allowed	1500
TolFun	Termination tolerance wrt the function (model) value	10^{-8}
TolX	Termination tolerance wrt the coefficients	10^{-6}
DiffMinChange	Min. change coefficients for finite difference Jacobians	10^{-4}
DiffMaxChange	Max. change coefficients for finite difference Jacobians	1

reconstruction is much better when wavelet filtering and baseline correction is applied.

2.4.3. Construction of the final GD models

The following initial values were set for the parameters of the algorithm:

- The initial translation values μ_i, defining the center of each Gaussian f_i, were equally spaced in each trial.
- The initial amplitude A_i for each Gaussian f_i was computed as the average of the 3 consecutive sample values in the spectrum that were closer to the associated translation value.
- The initial standard deviation δ_i of each Gaussian f_i was set to one.

Since the coefficients of the Gaussians were restricted to be positive, the restrictions in the TR algorithm were set to fit positive coefficients. The convergence criteria were set to the values compiled in Table 4, which resulted in a good fitting performance.

2.5. Feature selection of the GD models using MWVA

Having pre-processed the MR spectra and after reconstructing the signal using an adequate number of Gaussians, two vectors with the values of amplitude and standard deviation of each Gaussian were created. These values were positioned at the coordinates of the translations of their corresponding Gaussian. The cardinality of these vectors (lower than the 195 of the original spectra) was then artificially set to 195 by assigning zero values to translations other than those of the Gaussians themselves. This procedure also served to identify repetitions in the translations (by repetition, we understand two translations within the same integer interval). When such a repetition in translation was identified, the lower translation value was truncated to the closest integer value, while the higher translation value was rounded to the next integer value, thus avoiding the overlapping. When more than two repetitions happened in the same interval, the highest of the translations were moved to the closest higher integer, while the other translations were positioned, according to their value (ranked from the highest to the lowest, following the aforementioned procedure), behind the highest integer initially estimated. This process finished when no further overlapping was found between translations.

Feature selection was performed with MWVA (Arizmendi et al., 2009, 2010) (see section 2.3), applying it to the re-scaled vectors of amplitude and standard deviation (the data were re-scaled to zero mean and unit variance). Each value of Ω in the Dissimilarity Index Matrix (DIM) was labeled with the coordinates k and l, where k indicates the position of the spectrum where the window starts and l indicates the window width (w) used.

Once the DIM of the amplitude and standard deviation of each experiment were obtained, the selection of the relevant variables was carried out with MWVA in two different data sets: The first one corresponded to the concatenation of amplitude and standard deviation. The second one corresponded to the average of the amplitude and standard deviation.

2.6. MRS classification

Feed-forward ANNs were used in the classification experiments starting from the features selected and extracted through dimensionality reduction and variable selection. Different network architectures between 5 and 40 units in the only hidden layer were investigated. Given that all classification problems were binary, one unit in the output layer did suffice. In order to avoid data overfitting, the networks were trained with Bayesian regularization (MacKay, 1992) as part of a back-propagation process. The adaptive weights and biases were updated according to the Levenberg–Marquardt algorithm (Foresee & Hagan, 1997).

Obviously, there is a wide choice of available classifiers to perform this task. Given that the focus of this study is in the impact of feature extraction and selection of the signal processing method and not in the classification method itself, we restrict our choice to a technique that combines the flexibility of nonlinear models with the penalization of model over-complexity conferred by the Bayesian regularization framework in which it is developed. To some extent, the choice was also justified by the results obtained in our experiments and reported in Section 3.2, as they were shown to outperform results presented in other studies that analyzed similar data using Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and unregularized ANNs, amongst other techniques.

One run of a 5-fold cross-validation was performed for each ANN, allowing a maximum of 500 epochs. To address the issue of class imbalance (the number of cases available from each tumour type is always small, but widely varying), the original datasets were optionally re-sampled, by over-sampling the minority class and under-sampling the majority class (Japkowicz, 2000).

3. Results

3.1. Final Gaussian dissection models

In this study, and following the recommendations of previous research (Vellido et al., 2009, 2012), each MR spectrum was represented by 195 clinically-relevant frequency intensity values (data features). Considering as well that each Gaussian is defined by three parameters, the maximum allowed number of Gaussians was limited to 65. For each trial, the MSE and PD figures of merit were computed, taking the minimum PD out of the 65 trials for each spectrum as the representative solution.

Table 5 shows the MSE, PD and the statistics of the fitting of the spectral signal in two settings: (1) when decomposing the spectrum without preprocessing and (2) when decomposing the spectrum after preprocessing (i.e., preprocessing using wavelet filtering with baseline correction). These results fully confirm the adequacy of the pre-processing procedure.

In order to remove the less relevant amongst the obtained Gaussians, retaining only the most important ones, the individual area of each Gaussian in the trial with minimum PD was computed. These areas were subsequently ranked in descending order. The ranked Gaussians were then added sequentially, reconstructing the spectra and calculating the MSE after each addition. The differences in MSE between consecutive additions were computed and normalized, so as to produce the maximum value of 100%. The

Table 5
Mean ± Std of the MSE and PD corresponding to the MR Spectrum preprocessed and without preprocessing. The last line presents the Mean ± SD of the number of Gaussians used in the reconstruction.

	With preprocessing	Without preprocessing
Mean ± SD of MSE	0.1419 ± 0.3045	0.7218 ± 3.0375
Mean ± SD of PD	3.8754 ± 1.8893	8.0960 ± 11.9532
Mean ± SD of Gaussians used	55.1217 ± 10.1610	42.7303 ± 19.3334

Table 6
Best accuracy results and their corresponding methods.

Experiments	Clasification	Methods
G1 vs G2	87.35 ± 8.45	UB.CO
G1 vs mm	89.83 ± 5.69	B.CO
a2 vs a3	96.00 ± 8.94	B.CO, UB.CO, UB.AV
a2 vs G2	91.90 ± 8.55	B.CO
a2 vs ly	100.00 ± 0.00	UB.CO
a2 vs oa	100.00 ± 0.00	B.CO
a3 vs pn	93.33 ± 14.91	B.AV
G2 vs mm	88.80 ± 1.95	B.CO
gl vs a3	96.17 ± 5.64	B.CO
gl vs ab	97.42 ± 3.54	UB.CO
gl vs ly	96.25 ± 3.42	UB.CO
gl vs me	77.90 ± 2.37	UB.CO
gl vs no	96.67 ± 3.04	UB.CO, B.CO
gl vs pn	98.75 ± 2.80	UB.CO
me vs ly	95.00 ± 6.85	B.CO, B.AV
me vs mm	95.00 ± 8.15	UB.CO
me vs no	100.00 ± 0.00	UB.CO
me vs pn	100.00 ± 0.00	UB.CO, UB.AV
mm vs ab	100.00 ± 0.00	B.CO
od vs a2	96.00 ± 8.94	B.CO, B.AV, UB.CO, UB.AV

Table 7
Best BER results and their corresponding methods.

Experiments	Classification	Methods
G1 vs G2	16.45 ± 10.92	B.CO, UB.CO
G1 vs mm	9.77 ± 8.05	B.CO, UB.CO
a2 vs a3	2.50 ± 5.59	B.CO
a2 vs G2	6.88 ± 5.91	B.CO
a2 vs ly	0.00 ± 0.00	UB.CO
a2 vs oa	0.00 ± 0.00	B.CO
a3 vs pn	5.00 ± 9.12	B.AV
G2 vs mm	13.33 ± 3.26	UB.CO
gl vs a3	2.05 ± 3.01	B.CO
gl vs ab	2.81 ± 2.98	UB.AV
gl vs ly	8.38 ± 3.75	B.CO
gl vs me	17.33 ± 4.14	UB.AV
gl vs no	2.10 ± 1.92	B.CO
gl vs pn	0.67 ± 1.49	UB.CO
me vs ly	10.00 ± 13.69	B.AV
me vs mm	4.00 ± 6.52	UB.CO
me vs no	0.00 ± 0.00	B.CO, UB.CO
me vs pn	0.00 ± 0.00	B.CO, UB.CO
mm vs ab	0.00 ± 0.00	B.CO
od vs a2	2.50 ± 5.59	B.CO, UB.CO, B.AV, UB.AV

Table 8
Best AUC results and their corresponding methods.

Experiments	Classification	Methods
G1 vs G2	0.93 ± 0.07	B.CO
G1 vs mm	0.99 ± 0.01	B.CO
a2 vs a3	1.00 ± 0.00	B.CO, UB.CO, UB.AV, B.AV
a2 vs G2	0.96 ± 0.04	B.CO
a2 vs ly	1.00 ± 0.00	UB.CO, B.CO, B.AV, UB.AV
a2 vs oa	1.00 ± 0.00	UB.CO, B.CO, B.AV, UB.AV
a3 vs pn	1.00 ± 0.00	UB.CO, B.CO, B.AV, UB.AV
G2 vs mm	0.95 ± 0.04	B.CO
gl vs a3	1.00 ± 0.00	B.CO, UB.CO, UB.AV
gl vs ab	1.00 ± 0.00	B.CO, B.AV, UB.CO, UB.AV
gl vs ly	0.99 ± 0.03	UB.CO
gl vs me	0.88 ± 0.09	B.CO
gl vs no	0.99 ± 0.01	B.CO
gl vs pn	1.00 ± 0.00	B.CO, UB.CO
me vs ly	1.00 ± 0.00	UB.CO, B.CO, B.AV, UB.AV
me vs mm	1.00 ± 0.00	B.CO, B.AV, UB.AV
me vs no	1.00 ± 0.00	UB.CO, B.CO, B.AV, UB.AV
me vs pn	1.00 ± 0.00	UB.CO, B.CO, B.AV, UB.AV
mm vs ab	1.00 ± 0.00	UB.CO, B.CO, B.AV, UB.AV
od vs a2	1.00 ± 0.00	UB.CO, B.CO, B.AV, UB.AV

3.2. Classification results

The best results for each of the performance evaluation metrics (which are the Area Under the ROC Curve (AUC), the accuracy, and the Balanced Error Rate (BER)), together with their corresponding methods, are summarized in Tables 6–8. Figs. 6–8 display the box-plots of the global results of AUC, BER and accuracy, for all the analyzed problems in balanced (B) and unbalanced (UB) groups, and for the concatenated (CO) and averaged (AV) DIM.

The labels in the box-plot figures are built by joining the acronym of the method (i.e., CO and AV) and the performance indicator (i.e., Accuracy, BER and AUC). The descriptive statistics (i.e., mean, standard deviation, median, and 25/50/75 quartiles) for the AUC, $100 - BER$ and Accuracy results are summarized, in turn, in Tables 9–11.

4. Discussion

The box-plots in Figs. 6–8 provide evidence that the concatenated method yields the best scores. A Wilcoxon test was applied to assess the existence of significant differences between the scores of averaged and concatenated DIMs. The results of the statistical test are compiled in Table 12, showing that the concatenated method is significantly better than the averaged one (p-value $\leqslant 0.05$). In turn, the results obtained with balanced datasets are consistently better than those obtained with the original unbalanced ones, although not in a statistically significant way, as evaluated again through a Wilcoxon test (see Table 13). This last result supports the importance of using class-balancing algorithmic strategies in order to limit the bias introduced by heterogeneous class sample size in the classification process, but also indicates that, in the tumour classification problems analyzed in this study, the negative impact of class unbalance is not too significant.

Quite a few of the classification problems addressed in this study have previously been investigated using similar datasets extracted from the INTERPRET project database. Tumour type discrimination was carried out in García-Gómez et al. (2009) using a slightly different subset of MRS data obtained at SET from the INTERPRET database analyzed in our experiments. This makes the comparison with this study especially relevant.

The authors reported in García-Gómez et al. (2009) a BER result of 40% in the difficult problem of discriminating between two types

addition of Gaussians stopped when the MSE improvement was lower than 1%, and the remaining ones were eliminated.

The mean (± standard deviation) MSE for the complete MRS data set before this reduction in the number of constituting Gaussians was 0.14 (±0.30), with an average of 55.12 (±10.16) Gaussians used; the corresponding values after reduction were 0.16 (±0.31) with an average of 25.55 (±8.35) Gaussians. These results confirm that the elimination process does not affect the reconstruction in any significant way, while managing a reduction of 46.36% in the number of Gaussians initially considered for the reconstruction of the spectra.

Fig. 6. Boxplot of the AUC values. Each box represents the lower quartile (bottom line), median (line in the middle), and upper quartile (top line) values. The whiskers are lines extending from each end of the box to represent the extent of the rest of the data. Classification problems with atypical results are left outside the limits of the boxplots.
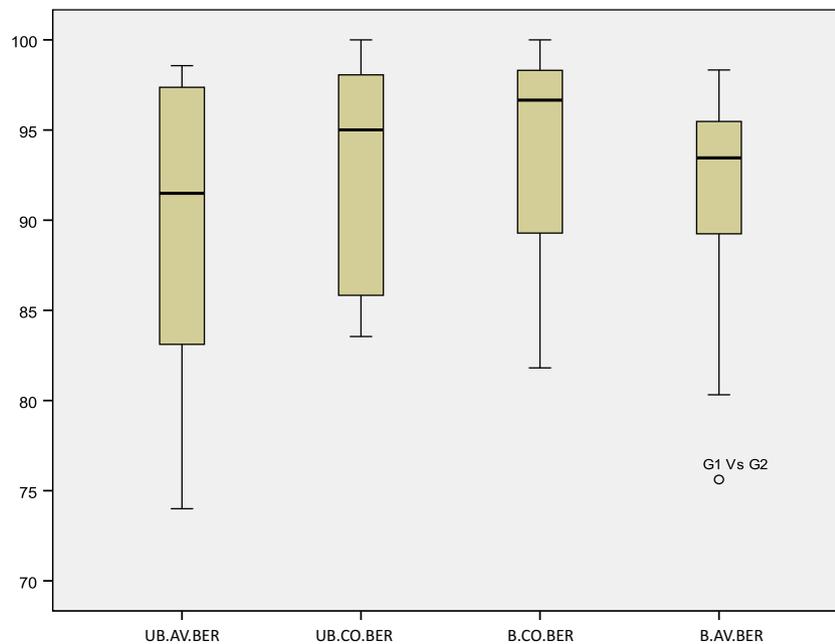
Fig. 7. Boxplots of the BER values, as in Fig. 6.

of aggressive tumours, namely *gl* and *me*, using spectral PI and a Linear Discriminant Analysis (LDA) classifier. This result is to be compared with a BER as low as 17.33%, reported in Table 7. The best results to date in this problem using the INTERPRET database were reported in Vellido et al. (2012), where a test AUC of 0.86 was obtained using a concatenation of LET and SET spectra for the same patients. Also in García-Gómez et al. (2009), a BER result of 5% was reported for the easier problem of discrimination between *me* and *mm*, this time using PCA and a MLP ANN classifier; this result is not too different from the result of 4% reported in Table 7.

Some similar diagnostic classification problems were explored in Lukas et al. (2004). They involved data from the same general INTERPRET database, but acquired at LET. Experiments were reported for *gl*, *me*, *mm*, and low-grade astrocytomas (*a*2). More

specifically, the following binary problems were considered: *gl vs. mm*, *gl vs. me*, *gl vs. a*2, *mm vs. me*, *mm vs. a*2, and *me vs. a*2. Data were classified using linear and nonlinear methods, namely, LDA, SVM and Least-Squares SVM (LS-SVM). Experiments with and without dimensionality reduction (with PI and heuristic spectral sub-regions selection) were performed. Results were qualified using the AUC measure. Only two of their experiments correspond to our setting: The best reported mean AUC for the *gl vs. me* problem is 0.64 with PI and LS-SVM. Although not directly comparable, due to the use of different echo times, our corresponding result is 0.88 with MWVA, as reported in Table 8. For *mm vs. me*, the best mean AUC in Lukas et al. (2004) was 0.97 for LS-SVM without dimensionality reduction. This is to be compared with the AUC 1.00 reported in Table 8.
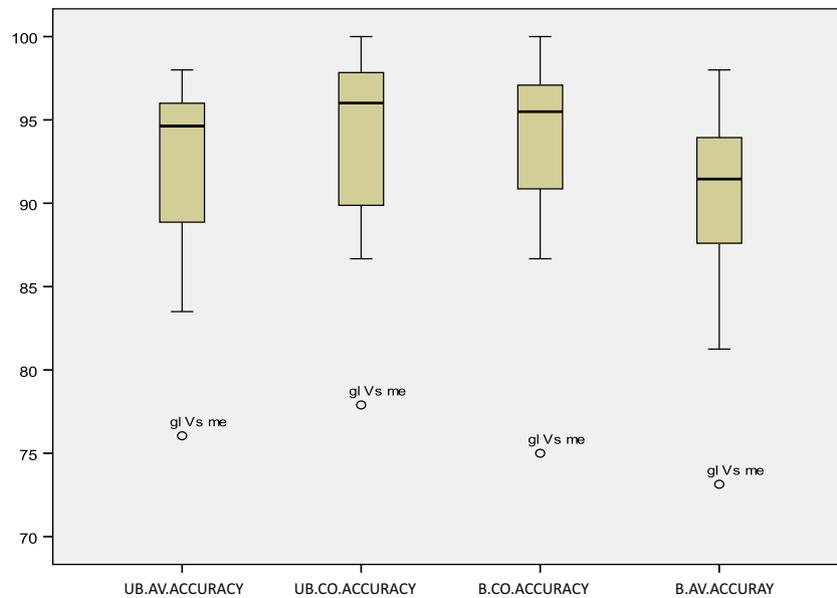
Fig. 8. Boxplots of the ACCURACY values, as in Fig. 7.

Table 9
AUC: Descriptive statistics for each of the tests performed.

Performance indicator	N	Mean	Std deviation	Min	Max	Percent		
						25	50	75
B.CO.AUC	20	0.985	0.0318	0.88	1.00	0.9925	1.0000	1.0000
UB.CO.AUC	20	0.983	0.0359	0.87	1.00	0.9900	1.0000	1.0000
B.AV.AUC	20	0.969	0.0599	0.82	1.00	0.9825	1.0000	1.0000
UB.AV.AUC	20	0.971	0.0523	0.83	1.00	0.9650	1.0000	1.0000

Table 10
BER: Descriptive statistics for each of the tests performed.

Performance indicator	N	Mean	Std deviation	Min	Max	Percent		
						25	50	75
UB.CO.BER	20	92.732	6.222	83.55	100.00	85.42	95.00	98.34
B.AV.BER	20	91.054	6.623	75.62	98.33	89.18	93.47	95.64
B.CO.BER	20	94.035	5.863	81.81	100.00	88.93	96.66	98.47
UB.AV.BER	20	89.821	8.784	74.00	98.57	82.89	91.48	97.44

Table 11
Accuracy: Descriptive statistics for each of the tests performed.

Performance indicator	N	Mean	Std deviation	Min	Max	Percent		
						25	50	75
UB.AV.ACCURACY	20	91.995	5.708	76.05	98.00	88.30	94.64	96.00
B.CO.ACCURACY	20	93.671	6.041	75.00	100.00	90.35	95.50	97.30
UB.CO.ACCURACY	20	93.847	5.692	77.90	100.00	89.81	96.00	98.01
B.AV.ACCURACY	20	89.881	5.929	73.14	98.00	87.55	91.46	94.25

The difficult problem of discriminating between different grades of astrocytomas $a2$ and $a3$ was addressed in Ladroue (2003). Here, PCA was used for dimensionality reduction and LDA, LS-SVM and K-Nearest Neighbor (K-NN) were used as classifiers. A mean test accuracy of just under 70% for 20 PC's and LS-SVM was reported. This can be compared with our result of around 96.00%, reported in Table 6.

Another classically difficult problem: me vs. gl was also investigated in Ladroue, Howe, Griffiths, and Tate (2003). A maximum accuracy of only 55% was reported, compared to our result of 77.90%, using MWVA for dimensionality reduction.

A different tumour classification problem, that of discriminating between low-grade gliomas ($G1$) and high-grade malignant tumours ($G2$), has far more commonly been analyzed in the literature. Using PCA followed by LDA to distinguish between $G2$ and mm, a mean AUC of 0.94, with 6 principal components was reported in Devos (2005); this must be compared with our results of 0.95 reported in Table 8. The same method was used in Devos

Table 12
p-values between the concatenated and averaged methodologies in the Wilcoxon test.

Methodologies	*p*-value
UB.CO.AUC-UB.AV.AUC	0.018
UB.CO.BER-UB.AV.BER	0.019
UB.CO.ACCURACY-UB.AV.ACCURACY	0.006
B.CO.ACCURACY- B.AV.ACCURACY	0.003
B.CO.BER- B.AV.BER	0.003
B.CO.AUC-B.AV.AUC	0.007

Table 13
p-values between the balanced and unbalanced methodologies in the Wilcoxon test.

Methodologies	*p*-values
B.AV.AUC-UB.AV.AUC	0.096
B.CO.BER-UB.CO.BER	0.102
B.CO.ACCURACY-UB.CO.ACCURACY	0.777
B.AV.ACCURACY-UB.AV.ACCURACY	0.609
B.AV.BER-UB.AV.BER	0.542
B.CO.AUC-B.CO.AUC	0.028

(2005) to distinguish between $G2$ and $a2$, obtaining a mean AUC of 0.92, also using 6 principal components; this must be compared with our result of 0.96 reported in Table 8.

Importantly, the proposed method has achieved very encouraging classification results in problems concerning tumour types that have received little attention from the point of view of pattern recognition-based diagnostic classification from MRS. They include lymphomas (ly) and primitive neuroectodermal tumours and medulloblastomas (pn). The former have been discriminated from low-grade astrocytomas ($a2$) and high-grade malignant tumours (me and gl) with average accuracies ranging from 95 to 100% (see Table 6). Note that these tumours are difficult to discriminate from glioblastomas and metastases even with state-of-the-art data acquisition methods (Wang et al., 2011). The latter were classified within a similar range of accuracy from $a3$, gl and me, which neatly improves on previous results presented in Majós et al. (2002).

All in all, these comparisons with previous results reported in similarly-oriented studies of the INTERPRET database MRS data provide some evidence to support that the differential advantage provided by our proposed method lies mainly in the GD method and the pre-processing of the data using the wavelet filtering stage, combined with the use of MWVA for dimensionality reduction. Having said that, the proposed method has several limitations to take into account. Firstly, because it is a combination of several data pre-processing stages (GD, filtering, feature selection) with classification; there is no guarantee that alternative combinations of pre-processing methods (using, for instance, different filtering techniques or different feature selection approaches) could not yield better results. Secondly, there is no guarantee that the proposed pairing of proposed pre-processing strategy and classifier is, again, optimal in any sense.

5. Conclusions

In this study, the GD signal processing technique was used to break down a given MR spectrum into its constituent tones, represented by coefficients of amplitude, standard deviation and translation of their corresponding Gaussians. In doing so, GD provides information about the conformation of the metabolite peaks. Simultaneously, it alleviates the problem of metabolite-associated broad amplitude peak overlapping between metabolites resonating at similar frequencies, overcomes the noise and baseline problems, and eases the subsequent feature selection and the classification processes.

These coefficients obtained by GD, in conjunction with the axis of transformation and the the concatenation of amplitude and standard deviation DIM's, were used for the discrimination between different types of brain tumours on the basis of their MR spectra.

The obtained diagnostic classification results were very encouraging and rank among the best obtained to date using alternative methods to analyze similar data. An exception to this pattern are the experiments that involve some super-class group (G1 or G2). This could be due to the calculation of the exact variations of the translations of each Gaussian in each tumour class of the experiments, so that, when several tumour types are merged in a group, the difficulty of calculating the exact translation of the metabolite increases. As a result, the errors in each pattern are accumulated, making the process of feature selection difficult and, subsequently, affecting classification negatively. This problem could be overcome by making a transformation of the amplitude and standard deviation vectors, so as to expand the area of each one translation component, smoothing the effect of variation in translations.

These results are of special relevance for experiments involving tumour types that have rarely been studied using pattern recognition methods from MRS data, in particular those involving the discrimination of lymphomas and primitive neuroectodermal tumours from astrocytomas and high-grade brain tumours.

The practical implication of our proposed method and the results of its application to a number of brain tumour classification problems is straightforward. As stated in the introduction, expert radiologists face a very sensitive diagnostic task, in which they might be assisted using decision support systems that include automated classification tools based on metabolic tissue signature provided by MRS (Ortega-Martorell, Olier, Julià-Sapé, & Arús, 2010; Pérez-Ruiz et al., 2010). Our method has shown to provide consistently strong results in many tasks of brain tumour discrimination, which would make it a good candidate for such medical decision support system.

Future research will be devoted to the investigation of the association between the most relevant variables (amplitude, translation and standard deviation) and their correspondence with the known metabolites in the spectrum. This should result in a more clinically-interpretable outcome that could be used in decision-making. Furthermore, alternative and more exhaustive feature selection and classification methods to be used in combination with GD should be investigated.

In this study, spectroscopy data acquired only at STE were analyzed. Future work should also involve similar experiments with data acquired at long echo times or even with data obtained by combination of different acquisition echo times, an approach that has recently been shown to be fruitful (García-Gómez et al., 2009; Vellido et al., 2012). Despite the fact that the analyzed database is among the best in the world of its kind, future work should aim to validate the reported results using alternative MRS databases.

Spanish "Ley Orgánica de Protección de Datos de Carácter Personal" (LOPD), "Ley Orgánica 15/1999" and the "95/46/EU directive on data protection, December 13th, 1999". All patients or their legal representatives signed informed consent forms, agreeing to the study and to the use of their deidentified (anonymised) data for research.

References

Arizmendi, C., Hernandez-Tamames, J., Romero, E., Vellido, A., & Del Pozo, F. (2010). Diagnosis of brain tumours from magnetic resonance spectroscopy using wavelets and Neural Networks. In *Proceedings of the annual international conference of the ieee engineering in medicine and biology society* (pp. 6074–6077).

Arizmendi, C., Vellido, A., & Romero, E. (2009). Frequency selection for the diagnostic characterization of human brain tumours. In *Proceedings of the 12th international conference of the catalan association for artificial intelligence* (pp. 391–398). IOS Press.

Arizmendi, C., Vellido, A., & Romero, E. (2012). Classification of human brain tumours from MRS data using discrete wavelet transform and bayesian neural networks. *Expert Systems with Applications, 39*(5), 5223–5232.

Barton, A. J., & Valdés, J. J. (2008). Computational intelligence techniques applied to magnetic resonance spectroscopy data of human brain cancers. In *Proceedings of the 6th international conference on rough sets and current trends in computing. Lecture notes in computer science* (Vol. 5306, pp. 485–494). Springer.

Cousins, J. P. (1995). Clinical MR spectroscopy: Fundamentals, current applications, and future potential. *American Journal of Roentgenology, 164*(6), 1337–1347.

De Graaf, A. A., & Bovee, W. (1990). Improved quantification of in vivo NMR spectra by optimization of signal acquisition and processing and by incorporation of prior knowledge into the spectral fitting. *Magnetic Resonance in Medicine, 15*(2), 305–319.

Devos, A. (2005). *Quantification and classification of MRS data and applications to brain tumour recognition* (Ph.D. thesis), Katholieke University, Leuven, Belgium.

Devos, A., Lukas, L., Suykens, J. A. K., Vanhamme, L., Tate, A. R., Howe, F. A., et al. (2004). Classification of brain tumours using short echo time ^{1}H MR spectra. *Journal of Magnetic Resonance, 170*(1), 164–175.

Elguero, J., Alkorta, I., Pardo, C., Claramunt, R. M., & Santa María, M. D. (2004). Resonancia Magnética Nuclear de Protón. Aplicaciones en Química Orgánica. http://www.uned.es/dpto-quim-org-bio [Online; accessed 13th of February, 2014].

Foresee, F. D., & Hagan, M. T. (1997). Gauss–Newton approximation to Bayesian regularization. In *Proceedings of the international joint conference on neural networks, Houston, Texas, USA* (pp. 1930–1935).

García-Gómez, J. M., Luts, J., Julià-Sapé, M., Krooshof, P., Tortajada, S., Robledo, J. V., et al. (2009). Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *Magnetic Resonance Materials in Physics, Biology and Medicine (MAGMA), 22*(1), 5–18.

González-Navarro, F. F., Belanche-Muñoz, L. A., Romero, E., Vellido, A., Julià-Sapé, M., & Arús, C. (2010). Feature and model selection with discriminatory visualization for diagnostic classification of brain tumors. *Neurocomputing, 73*(4–6), 622–632.

Govindaraju, V., Young, K., & Maudsley, A. A. (2000). Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine, 13*(3), 129–153.

Haud, U. (2000). Gaussian decomposition of the Leiden/Dwingeloo survey. I. Decomposition algorithm. *Astronomy and Astrophysics, 364*, 83–101.

Haud, U. (2010). Gaussian decomposition of HI surveys. V. Search for very cold clouds. *Astronomy and Astrophysics, 514*, A27arXiv:1001.4155 .

Haud, U., & Kalberla, P. M. W. (2007). Gaussian decomposition of HI surveys. 3. Local HI. *Astronomy and Astrophysics, 466*, 555–564arXiv:astro-ph/0702149 .

Hoch, J. C., & Stern, A. S. (1996). *NMR Data Processing.* Blackwell: Wiley.

Huang, Y., Lisboa, P. J. G., & El-Deredy, W. (2003). Tumour grading from magnetic resonance spectroscopy: A comparison of feature extraction with variable selection. *Statistics in Medicine, 22*(1), 147–164.

INTERPRET project, Data Protocols, [Online; accessed 13th of February, 2014. <http://azizu.uab.es/INTERPRET/cdap.html>.

Japkowicz, N. (2000). The class imbalance problem: Significance and strategies. In *Proceedings of the international conference on artificial intelligence* (Vol. 1, pp. 111–117).

Julià-Sapé, M., Acosta, D., Mier, M., Arùs, C., & Watson, D. (2006). A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. *Magnetic Resonance Materials in Physics, Biology and Medicine (MAGMA), 19*(1), 22–33.

Kalberla, P. M. W., Burton, W. B., Hartmann, D., Arnal, E. M., Bajaja, E., Morras, R., et al. (2005). The Leiden/Argentine/Bonn (LAB) survey of galactic HI. *Astronomy and Astrophysics, 440*(2), 775–782.

Kleihues, P., Louis, D. N., Scheithauer, B. W., Rorke, L. B., Reifenberger, G., Burger, P. C., et al. (2002). The WHO classification of tumors of the nervous system. *Journal of Neuropathology and Experimental Neurology, 61*(3), 215–225.

Ladroue, C. L. C. (2003). *Pattern recognition techniques for the study of magnetic resonance spectra of brain tumours* (Ph.D. thesis). St. George's Hospital Medical School, University of London, United Kingdom.

Ladroue, C., Howe, F. A., Griffiths, J. R., & Tate, A. R. (2003). Independent component analysis for automated decomposition of in vivo magnetic resonance spectra. *Magnetic Resonance in Medicine, 50*(4), 697–703.

Liang, Z. P., & Lauterbur, P. C. (1999). *Principles of magnetic resonance imaging: A signal processing perspective. IEEE press series in biomedical engineering.* Wiley-Blackwell.

Lukas, L., Devos, A., Suykens, J. A. K., Vanhamme, L., Howe, F. A., Majos, C., et al. (2004). Brain tumor classification based on long echo proton MRS signals. *Artificial Intelligence in Medicine, 31*(1), 73–89.

MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Computation, 4*(3), 415–447.

Majós, C., Alonso, J., Aguilera, C., Serrallonga, M., Acebes, J. J., Arús, C., et al. (2002). Adult primitive neuroectodermal tumor: Proton MR spectroscopic findings with possible application for differential diagnosis. *Radiology, 225*(2), 556–566.

Mallat, S. G. (1999). *A wavelet tour of signal processing.* New York: Academic Press.

Oros-Peusquens, A. M., Laurila, M., & Shah, N. J. (2008). Magnetic field dependence of the distribution of NMR relaxation times in the living human brain. *Magnetic Resonance Materials in Physics, Biology and Medicine (MAGMA)* (21), 131–147.

Ortega-Martorell, S., Olier, I., Julià-Sapé, M., & Arús, C. (2010). Spectraclassifier 1.0: A user friendly, automated mrs-based classifier-development system. *BMC Bioinformatics, 11*, 106.

Pérez-Ruiz, A., Julià-Sapé, M., Mercadal, G., Olier, I., Majós, C., & Arús, C. (2010). The INTERPRET decision-support system version 3.0 for evaluation of magnetic resonance spectroscopy data from human brain tumors and other abnormal brain masses. *BMC Bioinformatics, 11*, 581.

Suzuki, S., Sakai, O., & Jara, H. (2006). Combined volumetric T_1, T_2 and secular-T_2 quantitative MRI of the brain: Age-related global changes (preliminary results). *Magnetic Resonance Imaging* (24), 878–887.

Tate, A. R., Watson, D., Eglen, S., Arvanitis, T. N., Thomas, E. L., & Bell, J. D. (1996). Automated feature extraction for the classification of human *in vivo* ^{13}C NMR spectra using statistical pattern recognition and wavelets. *Magnetic Resonance in Medicine, 35*(6), 834–840.

Vellido, A., Lisboa, P. J. G., & Vicente, D. (2006). Robust analysis of MRS brain tumour data using t-GTM. *Neurocomputing, 69*(7–9), 754–768.

Vellido, A., Martín-Guerrero, J., & Lisboa, P. (2012). Making machine learning models interpretable. In *Proceedings of the european symposium on artificial neural networks (ESANN)* (pp. 163–172). Belgium: Bruges.

Vellido, A., Romero, E., González-Navarro, F. F., Belanche-Muñoz, L. A., Julià-Sapé, M., & Arús, C. (2009). Outlier exploration and diagnostic classification of a multi-centre 1H-MRS brain tumour database. *Neurocomputing, 72*(13-15), 3085–3097.

Vellido, A., Romero, E., Julià-Sapé, M., Majós, C., Moreno-Torres, Á., Pujol, J., et al. (2012). Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel ^{1}H MRS. *NMR in Biomedicine, 25*(6), 819–828.

Vilamala, A., Lisboa, P. J. G., Ortega-Martorell, S., & Vellido, A. (2013). Discriminant convex non-negative matrix factorization for the classification of human brain tumours. *Pattern Recognition Letters, 34*(14), 1734–1747.

Wagner, W., Ullrich, W., Ducic, V., Melzer, T., & Studnicka, N. (2006). Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing, 60*(2), 100–112.

Wang, S., Kim, S., Chawla, S., Wolf, R. L., Knipp, D. E., Vossough, A., et al. (2011). Differentiation between glioblastomas, solitary brain metastases, and primary cerebral lymphomas using diffusion tensor and dynamic susceptibility contrast-enhanced mr imaging. *American Journal of Neuroradiology, 39*(3), 507–514.