Expert Systems with Applications 39 (2012) 5223–5232

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Classification of human brain tumours from MRS data using Discrete Wavelet Transform and Bayesian Neural Networks

Carlos Arizmendi *, Alfredo Vellido, Enrique Romero

Dept. de Llenguatges i Sistemes Informàtics (LSI), Universitat Politècnica de Catalunya (UPC), 08034 Barcelona, Spain

ARTICLE INFO

Keywords:
Magnetic Resonance Spectroscopy
Brain tumours
Medical Decision Support
Bayesian Neural Networks

ABSTRACT

The diagnosis of brain tumours is an extremely sensitive and complex clinical task that must rely upon information gathered through non-invasive techniques. One such technique is Magnetic Resonance Spectroscopy. In this task, radiology experts are likely to benefit from the support of computer-based systems built around robust classification processes. In this paper, a Discrete Wavelet Transform procedure was applied to the pre-processing of spectra corresponding to several brain tumour pathologies. This procedure does not alleviate the high dimensionality of the data by itself. For this reason, dimensionality reduction was subsequently implemented using Moving Window with Variance Analysis for feature selection or Principal Component Analysis for feature extraction. The combined method yielded very encouraging results in terms of diagnostic discriminatory binary classification using Bayesian Neural Networks. In most cases, the classification accuracy improved on previously reported results.

1. Introduction

The commodification of healthcare is leading to a rapidly increasing demand for personalization of patients' treatments. This demand is crucial in the clinical management of life-threatening pathologies such as cancer. Meeting this demand requires abundant resources, but also a sophisticated management of information systems. One of the reasons for this is that the amount of medical data available for analysis and knowledge extraction is also increasing exponentially. The surge in novel techniques for the non-invasive measurement and acquisition of medically-relevant data, in various forms including signals and image, is behind this situation. The resulting vast amount of information should be understood as a valuable asset for the Pattern Recognition and Computational Intelligence communities (Vellido & Lisboa, 2009).

This paper deals with the problem of diagnostic decision making in clinical neuro-oncology. Decision making in neuro-oncology is an obviously sensitive undertaking. Oncology in general is becoming a data-intensive field, in which new data acquisition techniques appear at a staggering pace. *Nature* recently devoted the cover of one of its issues (Nature, 2009) to advances in cancer research. Even within the very narrow context of just this journal issue and research field, several next-generation sequencing approaches were introduced with the specific target of monitoring genetic changes in tumour cells. The increasing reliance on microarrays data in genomics, and on protein chips and tissue arrays in proteomics, add to the wealth of information about active metabolic pathways that is available from other non-invasive data-acquisition techniques such as functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), or Magnetic Resonance Spectroscopy (MRS), just to name a few. Neuro-oncology is indeed an area in which the use of non-invasive measurement techniques is almost compulsory, unless invasive action becomes unavoidable. This means that plenty of data are often available to the decision maker. For this reason, neuro-oncology clinicians should benefit from the use of computer-based Medical Decision Support System (MDSS) tools.

The tumours of the central nervous system (CNS) represent around the 2% of the total of cancers diagnosed around the world. Annually, about 175,000 people are diagnosed with tumours that affect the CNS (Steward & Kleihues, 2003), out of which 29,000 occur in Europe (Comprehensive Cancer Monitoring Programme in Europe[1]). The incidence ratio of this pathology is of 7 persons per 100,000. The current gold standard for classification of brain tumours is class labeling according to the World Health Organization (WHO) system for diagnosing brain tumours by histopathological analysis of a biopsy sample. The WHO system lists dozens of different brain tumour pathologies, of an extremely diverse etiology and prognosis.

This paper offers evidence that CI-based tools can be successfully used to assist medical decision making in the diagnosis of a broad palette of brain tumour pathologies, on the basis of MRS information. MRS resorts to the use of strong magnetic fields for the generation of energy exchanges between the external magnetic

* Corresponding author.
E-mail address: carlos.julio.arizmendi@upc.edu (C. Arizmendi).

[1] http://www-dep.iarc.fr/hmp/camon.htm.

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.11.017

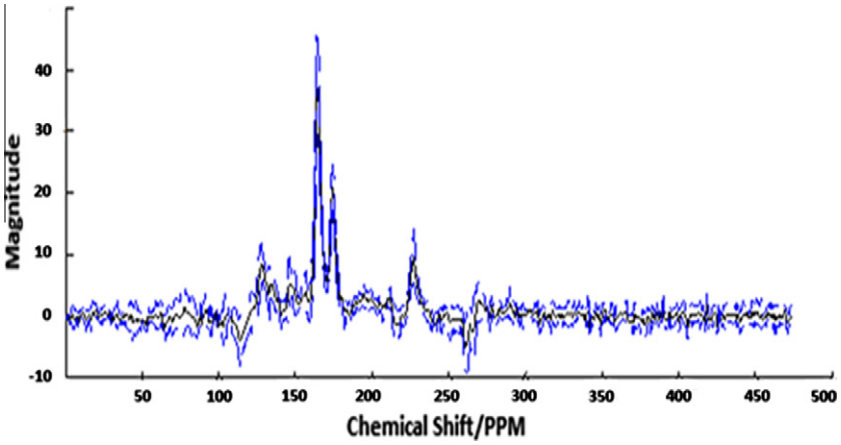

Fig. 1. Example of MRS signal: mean ± standard deviation of the spectra acquired through MRS corresponding to the oligoastrocytomas grade II type of brain tumour. The chemical shift, measured in ppm as described in Section 2, is expressed in the frequency domain.

field and the protons that are present in abundance in all living tissue. A radio-frequency machine picks up the energy exchanged, which is coded using 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 resonances of specific chemical and biochemical components of the tissue, as illustrated in Fig. 1. The wave profile is an indication of the quantities in which the components are present. Therefore, those substances that are present in big quantities in the tissue will have higher peaks associated than those present in lower concentrations.

In this study, we analyze a set of MRS data from the multi-centre, international INTERPRET database (Julià-Sapé, Acosta, Mier, Arús, & Watson, 2006). Even being one of the few available databases of this kind, it consists of only several hundred MR spectra corresponding to a number of pathologies (listed in Table 1). Each spectral frequency is taken to be a data variable, taking us to a difficult data analysis setting that involves a small sample of high-dimensional cases. Such setting requires the use of data pre-processing techniques. Often, the most determinant step in computer-based diagnostic classification analysis is precisely data pre-processing. Previous studies have resorted to different feature extraction and selection approaches (see, for instance, (Huang, Lisboa, & El-Deredy, 2003; Ladroue, Howe, Griffiths, & Tate, 2003; Leardi, 2000). In this study, we first use the Discrete Wavelet Transform (DWT) and a filtering process for the decomposition of the spectra in terms of approximation and detail coefficients, in a change of data representation that entails minimum loss of relevant information. This decomposition process by itself does not reduce the high dimensionality of the data. For this reason, the DWT process is followed by dimensionality reduction using Moving Window with Variance Analysis (MWVA: Arizmendi, Vellido, & Romero, 2009) for feature selection or Principal Component Analysis (PCA) for feature extraction. Diagnostic classification is then accomplished using Artificial Neural Networks (ANN) with Bayesian regularization. The proposed combination of methodologies is shown to yield high diagnostic classification accuracy for a broad range of brain tumour pathologies, for some of which this type of computer-based automated classification has, to the best of the authors' knowledge, never previously been reported.

Table 1
Different tumour types from the INTERPRET database analyzed in this paper, together with the corresponding number of cases available.

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	22
oa: Oligoastrocytomas grade II	6
od: Oligodendrogliomas grade II	7
pn: Primitive neuroectodermal tumours and medulloblastomas	9

2. Analyzed data

This paper investigates a multi-centre, international database of single-voxel, proton MRS (SV-^{1}H-MRS) data corresponding to several brain tumour pathologies (see Table 1: nine tumour pathologies were analyzed, as well as abscesses and normal brain tissue). It was created under the framework of the European project INTERPRET, an international collaboration of centers from 4 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). For detailed information on data acquisition and processing, and on database characteristics, see, for instance, INTERPRET Project: Data Protocols[2] and (Julià-Sapé et al., 2006). Class labeling was performed according to the WHO system for diagnosing brain tumours by histopathological analysis of a biopsy sample. 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 the experiments reported in this paper, a set of SV-^{1}H-MRS measured at short echo time (SET: 273 patients) were used. A total of 195 clinically relevant frequency intensity values (measured in parts per million (ppm), an adimensional unit of relative frequency position in the data vector), were used in this study.

3. Methodology and results

3.1. Wavelet Transform and its application to MRS data

The extraction of relevant information from *in vivo* proton MRS may be a difficult task. The presence of noise, artifacts (Vellido,

[2] http://azizu.uab.es/INTERPRET/cdap.html.

Lisboa, & Vicente, 2006; Vellido et al., 2009) and severe overlaps among spectral peaks (de Graaf & Bovée, 1990) can contribute to such difficulty. In current clinical settings, these limitations are emphasized by the use of SET data acquired at low intensity magnetic fields.

The use of wavelet techniques has been proposed to circumvent these problems (Luca, Mainardi, Pietro, Giuseppe, & Cerutti, 2002). In (Serrai, Nadal-Desbarats, Poptani, Glickson, & Senhadji, 2000) and (Serrai, Senhadji, Wang, Akoka, & Stroman, 2003), for instance, filtering processes based on the continuous Wavelet Transform (CWT) method were proposed to isolate the lactate signal from overlapping lipid resonances. An investigation on the use of wavelet techniques in the specific area of MRS signal-based brain tumour diagnostic analysis, with a similar setting to the one used in the current study, can be found in (García-Gómez et al., 2009).

The CWT of a signal $x(t)$ with mother wavelet $\psi(\cdot)$ is defined as:

$$W(\tau,s)=\frac{1}{\sqrt{|s|}}\int_{-\infty}^{\infty}x(t)\psi\left(\frac{t-\tau}{s}\right)dt. \tag{1}$$

The transformed signal $W(\tau)$ is a function of the translation parameter τ and the scale s. The signal energy is normalized at every scale by dividing the wavelet coefficients by $1/\sqrt{|s|}$. This ensures that the wavelets have the same energy at every scale. The original signal can be reconstructed with the inverse CWT, defined by:

$$x(t)=\frac{1}{C_\psi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}W(\tau,s)\frac{1}{s^2}\psi\left(\frac{t-\tau}{s}\right)d\tau ds. \tag{2}$$

$$C_\psi=\int_o^\infty\frac{|\psi(f)|^2}{f}df<\infty.$$

Many real applications, though, need to be defined in a discrete domain. This is the case of the MRS data analyzed in this study, described in a discrete frequency domain. An important development for the application of wavelet theory in discrete signal processing was presented in (Mallat, 1999), using Multirresolution Analysis (MA). The DWT is implemented via an octave filter bank, as a cascade of low-pass L(z) and high-pass H(z) filters, followed by subsampling, as illustrated in Fig. 2. Every pair of filters represents a decomposition level. The reconstruction of the original signal is possible using the synthesis filter bank where the signals are upsampled and passed through the filters L′(z) and H′(z). The reconstruction procedure, except for rounding errors, leads to the restoration of the original signal if no coefficient is altered.

The joint application of Mallat's model, Donoho's approach for signal filtering by thresholding (Donoho, 1995) and the computation of statistical coefficients for data compression, will allow to reduce the noise level and to represent the MRS signal without loss of relevant information, while keeping the dimensionality of the system as low as possible.

3.2. Wavelet Filtering

Frequently, the observed signal $x(t)$ can be assumed to consist of a real signal $s(t)$ plus additive white noise $n(t)$. Shrinkage filtering aims to denoise the observed signal $x(t)$ and recover $\hat{x}(t)$, an estimate of $s(t)$. The suggested model allows this through the use of WT by computing

$$\begin{aligned} y &= W_{(\psi,j)}(x)\\ z &= D(y,\lambda)\\ \hat{x} &= W_{(\Psi,j)^{-1}}(z)\end{aligned} \tag{3}$$

where $D(\cdot,\lambda)$ is the filtering operator for threshold λ and $W_{(\psi,j)}(\cdot)$ and $W^{-1}_{(\psi,j)}(\cdot)$ denote, in turn, the WT and its inverse, with wavelet function ψ and j decomposition levels; y is the Wavelet Transform of $x(t)$, z the result of the filtering operator and $\hat{x}$ the estimated signal after filtering. The computation of the filtering operator $D(\cdot,\lambda)$ for the denoising of the available MRS spectra was carried out according to the following three consecutive steps (Guo, 2000), each described in its own sub-section.

3.2.1. Threshold calculation

Three alternative choices of threshold were considered in the experiments, according to the following statistical estimators developed by Donoho (1995):

- Universal threshold (*Sqtwolog*): the threshold is chosen to be $\lambda=2\times\log(n)$. where n represents the length of the signal.

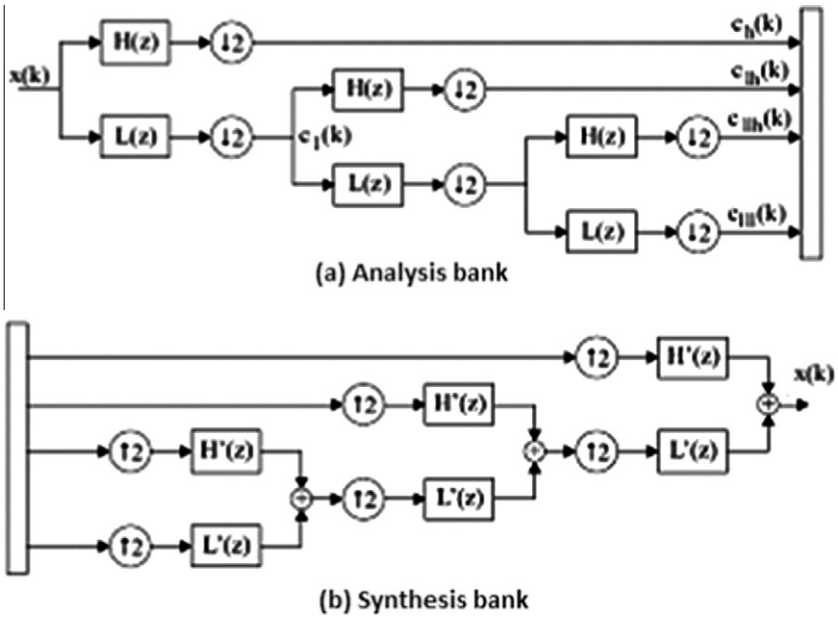

Fig. 2. Graphical illustration of a DWT decomposition algorithm with three decomposition and reconstruction levels, represented by, in turn, analysis and synthesis banks.

Table 2
Results of the different quality measures, as described in the main text, for the selection of the optimal mother wavelet.

Wavelet	MSE	SNR	NDC	Q1
Symlet (2)	0.65	191.43	213	0.91
Symlet (3)	0.70	193.58	216	0.95
Coiflet (1)	0.66	192.75	216	0.50
Coiflet (2)	0.75	193.90	238	0.37
Daubichie (2)	0.62	191.58	213	0.93
Daubichie (3)	0.66	193.97	216	0.98
Biortogonal (1.3)	0.77	228.38	216	0.62
Biortogonal (3.3)	0.64	217.38	221	0.65

- Threshold applying the principle of Stein's Unbiased Risk *(Rigrsure):* the procedure requires obtaining a new vector *NV* (*k*), rearranging data from minimum to maximum and taking the square root (Donoho, 1995), where *k* is the index in the risk algorithm (Guo et al., 2000).
- Threshold *Minimax:* the threshold is selected following the minimax principle, commonly used in statistics to design estimators (Donoho, 1995).

Minimax, *Rigrsure* and *Sqtwolog*, are function names taken from Matlab® wavelet toolbox.

3.2.2. Threshold scaling

The thresholds are usually weighted by a factor σ, a scaling of the mean absolute deviation based on the wavelet decomposition level. Three types of weighting were considered:

- *One*: the weighting term is scalar (e.g., $\sigma = 1$).
- *Sln*: the weighting is computed by averaging the detail coefficients of the first level of decomposition, divided by 0.6745 (Misiti, Misiti, Oppenheim, & Poggi, 2002).
- *Mln*: as *Sln* but with the calculation of the detail coefficients level by level.

3.2.3. Implementation of the threshold

Once the threshold is calculated and scaled, the thresholding process $D(y,\lambda)$ can be implemented through two alternative methods: *Hard* thresholding $D_h(y,\lambda)$ and *Soft* thresholding $D_s(y,\lambda)$ according to:

$$D_h(y,\lambda) = \begin{cases} y, & |y| \geq \lambda \\ 0, & |y| < \lambda \end{cases}$$
$$D_s(y,\lambda) = \begin{cases} Sgn(y)(|y| - \lambda), & |y| \geq \lambda \\ 0, & |y| < \lambda \end{cases} \quad (4)$$

3.3. Mother wavelet selection for MRS data

3.3.1. Selection of the mother wavelet orders

The DWT was applied to the original SV-^{1}H-MRS data, taking the decomposition to the maximum allowable level. Different mother wavelets and, for each, different orders were used to implement the DWT: Biorthogonals (orders: 1.1, 1.3, 1.5, 2.2, 2.4, 2.6, 2.8, 3.1, 3.3, 3.5, 3.7, 3.9, 4.4, 5.5, 6, and 8), Coiflet (orders: 1–5), Daubechies (1–43), and Symlet (1–25). For every mother wavelet, the absolute values of the decomposition coefficients

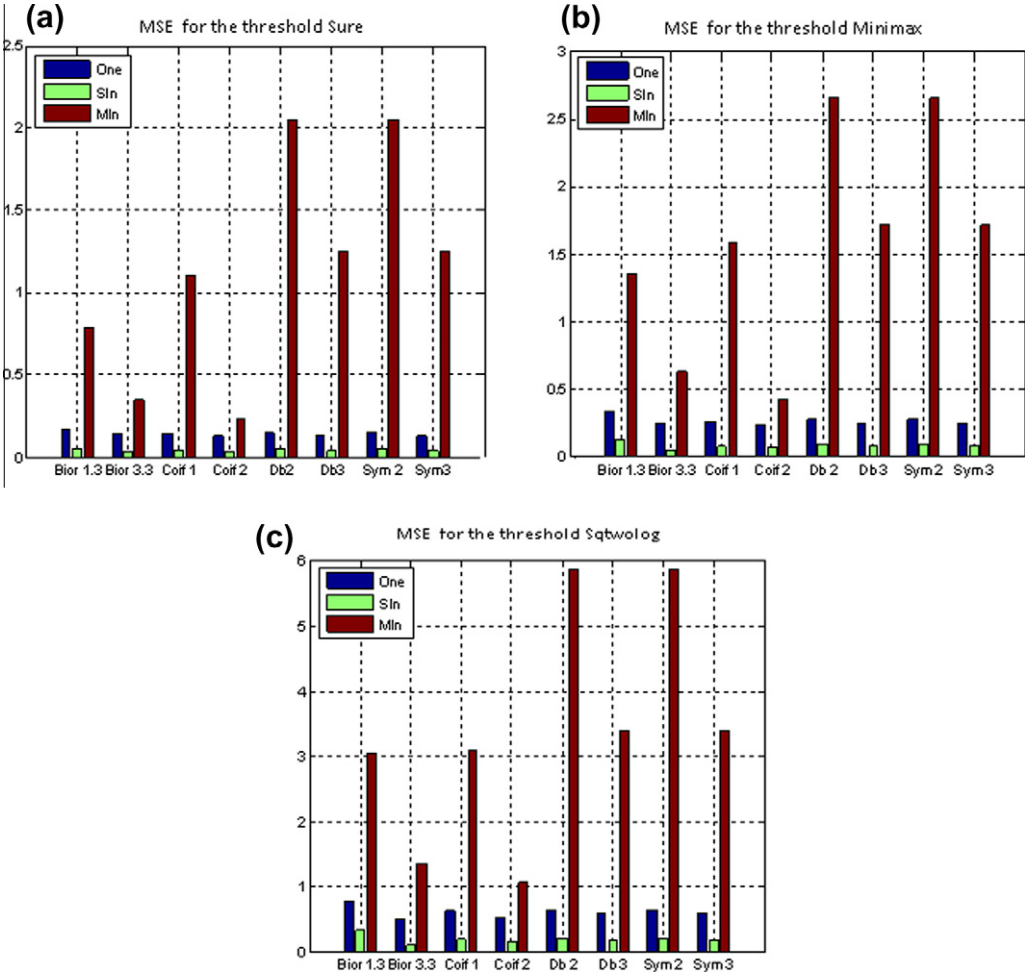

Fig. 3. MSE corresponding to the signals reconstructed by threshold *Rigrsure* (a) *Minimax* (b) and *Sqtwolog* (c) for different mother wavelets and weighting schemes.

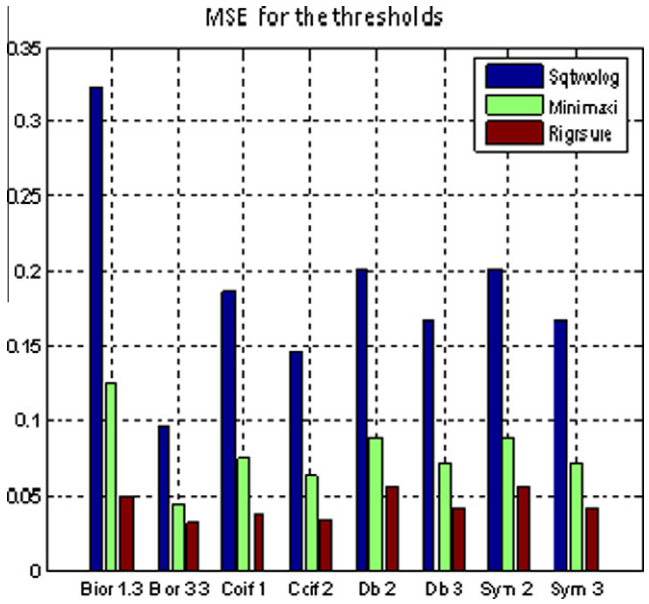

Fig. 4. Comparison of the MSE obtained with the best three thresholds, for several mother wavelets, when the *Sln* weighting scheme is applied.

Table 3
Statistics for the final comparison of the performance of the best mother wavelets once the *Rigrsure-Sln-Hard* procedure has been selected.

Wavelet	MSE	SNR	EP	PD	CR	Q2
Coiflet (1)	0.037	190.81	99.86	3.30	1.61	2.51
Coiflet (2)	0.033	192.86	99.86	3.08	1.43	2.78
Symlet (2)	0.056	188.23	99.78	3.78	1.68	1.18
Symlet (3)	0.042	191.94	99.83	3.28	1.63	2.10
Daubichie (2)	0.056	188.23	99.78	3.78	1.68	1.18
Daubichie (3)	0.042	191.94	99.83	3.29	1.63	2.10
Biortogonal (1.3)	0.050	182.73	99.92	4.47	1.68	1.25
Biortogonal (3.3)	0.032	193.62	99.81	3.02	1.55	3.20

were sorted in descending order, and the signal of each spectrum was reconstructed by adding consecutive coefficients. The average mean square error (MSE) and signal-to-noise ratio (SNR) were calculated over the whole set of patients for each wavelet order r, together with the number of decomposition coefficients (NDC):

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}[x(i) - \hat{x}(i)]^2 \tag{5}$$

$$\mathrm{SNR} = 10\log\left[\sum_{i=1}^{n}\frac{[x(i)]^2}{\mathrm{MSE} * n}\right] \tag{6}$$

where $\hat{x}$ is the reconstructed signal. Finally the $Q1$ index for the order r was computed with the mean of the aforementioned statistics, as follows:

$$Q1(r) = \frac{\overline{\mathrm{SNR}}_{Re}(r)}{\overline{\mathrm{MSE}}_{Re}(r) + \overline{\mathrm{NDC}}_{Re}(r)} \tag{7}$$

The *Re* subindex corresponds to a rescaling of the data between 1 and 3. The maximum values of $Q1$ indicate the orders with the best reconstruction error using the minimum NDC. The two highest values of $Q1$ for each wavelet function are reported in Table 2.

3.3.2. Selection of the final mother wavelet

Once the initial set of wavelet orders was chosen, as shown in Table 2, the filtering methodology described in Section 3.2 was used to denoise the spectrum signal. In order to determine the appropriate scaling, we followed Donoho (1995), who establishes the MSE as a measure in the design of estimators by threshold, with the condition that the filtered signal be as smooth as the original one. The MSE was thus calculated for each spectrum of the reconstructed signal, following the scheme described in 3.2: All the combinations of threshold estimation (*Sqtwolog*, *Rigrsure* and *Minimax*), threshold scaling (*Sln*, *One* and *Mln*), and *Hard* thresholding were implemented. The *Hard* function was used because it often yields smaller MSE than the *Soft* one and, furthermore, our

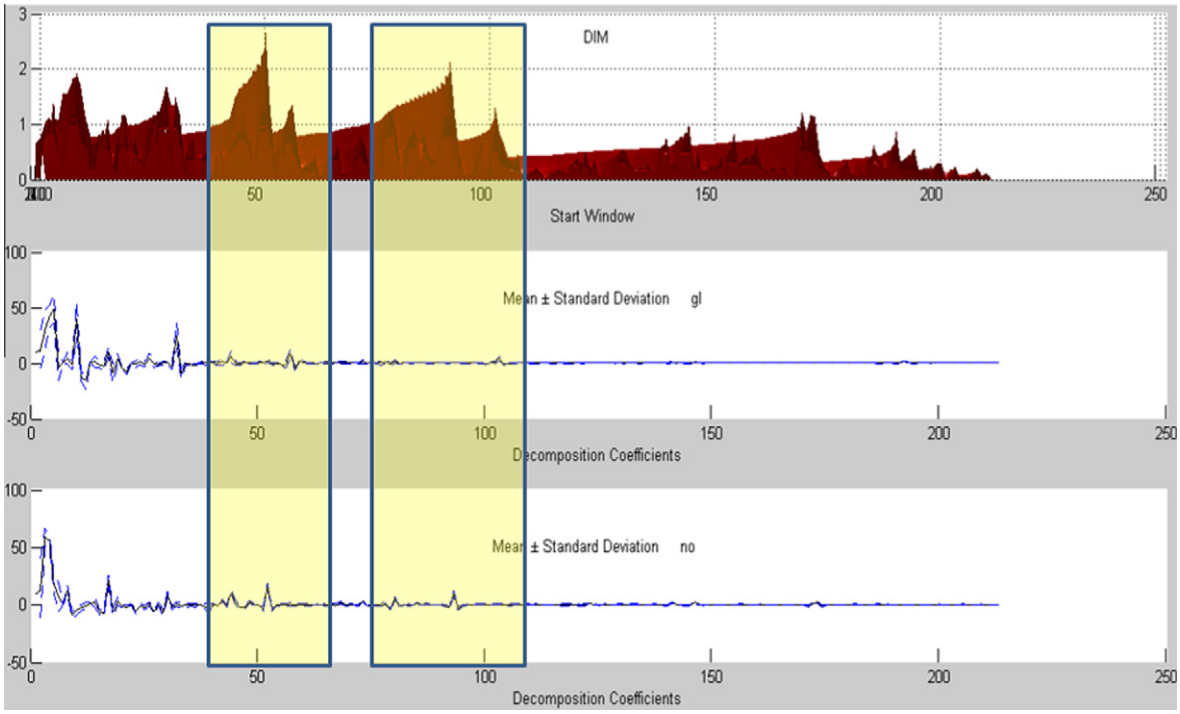

Fig. 5. The top display is a graphic illustration of the DIM corresponding to the experiment gl *vs.* no. The centre and bottom displays correspond, in turn, to the decomposition coefficients of gl and no. The stripes mark the areas that generate the largest differences between the coefficients of gl and no, which correspond to the highest DIM values.

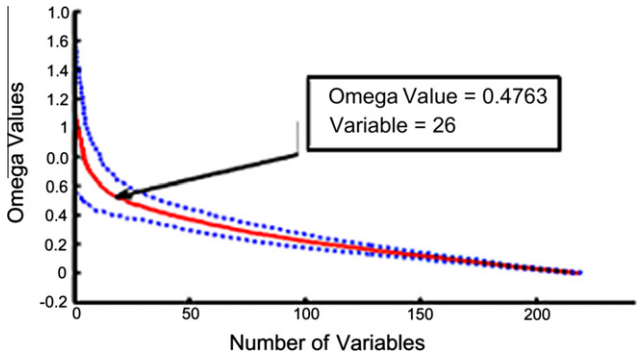

Fig. 6. Mean ± standard deviation of Ω, in descending order, as a function of the number of variables. The Ω ratio and the number of variables corresponding to a total of energy of 11% are highlighted.

results indicated that Soft thresholding could result in a decrease in the height of the resonance frequencies with a high degree of skewness, decreasing the importance of certain metabolites that are useful for classification. The results displayed in Fig. 3 show that, for all wavelets, the lower MSE is achieved when applying the *Sln* weighting scheme, regardless of the threshold calculation. The MSE of the three types of thresholds when *Sln* scaling is applied can be compared in Fig. 4, which shows that the *Rigrsure-Sln-Hard* procedure yields the best results among all combinations.

In order to determine the final wavelet, the average values of several statistics were computed for the *Rigsure-Sln-Hard* combination. They include: SNR, Energy Preserved (EP), Percentage of Distortion (PD) and Compression Ratio (CR) defined as:

$$\mathrm{EP} = \frac{\sum_{i=1}^{n}[\hat{x}(i)]^2}{\sum_{i=1}^{n}[x(i)]^2} * 100 \tag{8}$$

$$\mathrm{PD} = \sqrt{\frac{\sum_{i=1}^{n}[x(i)-\hat{x}(i)]^2}{\sum_{i=1}^{n}[x(i)]^2}} * 100 \tag{9}$$

$$\mathrm{CR} = L_0/L_c \tag{10}$$

where L_o is the cardinality of the decomposition coefficients of the original signal, and *Lc* the cardinality of decomposition coefficients that are different from zero. This set of statistics has been used to choose the optimal wavelet in previous related works concerning ECG signal filtering (Olarte & Sierra, 2007) and classification tasks (Rivas, Burgos, & García-Prada, 2009), among others. For a further objective criterion in choosing the optimal wavelet function, the *Q2* index was computed:

$$Q2(r) = \frac{\overline{\mathrm{SNR}}_{Re}(r) + \overline{\mathrm{EP}}_{Re}(r) + \overline{\mathrm{CR}}_{Re}(r)}{\overline{\mathrm{MSE}}_{Re}(r) + \overline{\mathrm{PD}}_{Re}(r)}. \tag{11}$$

Again, the *Re* subindex corresponds to a rescaling of the data between 1 and 3. The *Q2* values for the wavelet functions of Table 2 are shown in Table 3. The maximum value for this index was obtained for the Biortogonal (3.3) wavelet. Therefore, this wavelet was chosen for further experimentation.

3.4. Dimensionality reduction and classification

After processing the MR spectra with the Biortogonal (3.3) wavelet and filtering them with the combination *Rigsure-Sln-Hard*, the next necessary step of the analysis entails using the obtained decomposition coefficients as input in a dimensionality reduction process, using MWVA or PCA. This should help the subsequent classification process by removing redundant and unnecessary information from the input.

Table 4
Number of variables selected by the MWVA procedure (Balanced and unbalanced procedures in 2nd and 3rd columns, respectively) and features extracted by PCA (4th column) for the problems shown in the 1st column. G1 (low-grade gliomas) is the union of tumour types a2, oa and od. Meanwhile, G2 (high-grade malignant tumours) is the union of tumour types gl and me.These groupings are clinically relevant and have often been used in previous research (Devos & Lukas, 2004; Minguillón, Tate, Arús, & Griffiths, 2002; Tate, & Underwood. et al., 2006).

Experiments	MWVA.B	MWVA.UB	PCA
G1 *vs* G2	9	9	14
G1 *vs* mm	9	9	13
a2 *vs* a3	19	13	12
a2 vs G2	21	21	11
a2 *vs* ly	9	17	13
a2 *vs* oa	7	19	12
a3 *vs* pn	29	3	13
G2 *vs* mm	26	3	12
gl *vs* a3	3	13	18
gl *vs* ab	3	7	11
gl *vs* ly	5	6	11
gl *vs* me	12	15	15
gl *vs* no	17	17	17
gl *vs* pn	11	3	13
me *vs* ly	3	3	12
me *vs* mm	23	19	15
me *vs* no	10	11	14
me *vs* pn	3	7	10
mm *vs* ab	5	5	12
od *vs* a2	7	3	14
Average	10.15	11.15	12.52

3.4.1. Dimensionality reduction with MWVA and PCA

The MWVA is 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:

$$\mathrm{WGV}(w,i) = \sum_{a=1}^{n1}\frac{\|c1_{a,w,i} - \mu c1_{w,i}\|^2}{n1\sqrt{w}} + \sum_{a=1}^{n2}\frac{\|c2_{a,w,i} - \mu c2_{w,i}\|^2}{n2\sqrt{w}} \tag{12}$$

$$\mathrm{BGV}(w,i) = \frac{\|\mu c1_{w,i} - \mu c2_{w,i}\|^2}{\sqrt{w}} \tag{13}$$

$$\Omega(w,i) = \frac{\mathrm{BGV}(w,i)}{\mathrm{WGV}(w,i)}. \tag{14}$$

Table 5
Mean ± standard deviation of AUC and accuracy values for all balanced classification experiments.

Experiments	MWVA AUC	PCA AUC	MWVA ACCURACY	PCA ACCURACY
G1 *vs* G2	0.97 ± 0.04	0.91 ± 0.07	93.19 ± 5.04	88.74 ± 2.52
G1 vs mm	0.98 ± 0.01	0.97 ± 0.03	95.00 ± 8.14	92.25 ± 3.12
a2 vs a3	0.90 ± 0.16	0.95 ± 0.05	68.00 ± 17.9	68.00 ± 17.90
a2 vs G2	0.98 ± 0.02	0.95 ± 0.06	96.70 ± 3.47	88.53 ± 7.38
a2 vs ly	1.00 ± 0.00	0.95 ± 0.04	92.00 ± 10.9	84.00 ± 21.90
a2 vs oa	1.00 ± 0.00	1.00 ± 0.00	100.00 ± 0.00	74.00 ± 13.40
a3 vs pn	1.00 ± 0.00	0.70 ± 0.00	93.33 ± 14.9	53.33 ± 29.80
G2 vs mm	0.98 ± 0.01	0.96 ± 0.00	96.73 ± 2.28	92.12 ± 4.95
gl vs a3	0.99 ± 0.01	0.80 ± 0.10	94.91 ± 5.25	80.66 ± 8.02
gl vs ab	1.00 ± 0.00	0.80 ± 0.12	97.50 ± 3.42	84.66 ± 7.72
gl vs ly	1.00 ± 0.00	0.87 ± 0.08	90.00 ± 7.12	85.00 ± 8.38
gl vs me	0.90 ± 0.05	0.67 ± 0.12	71.23 ± 8.66	50 ± 16.66
gl vs no	1.00 ± 0.00	1.00 ± 0.00	100.00 ± 0.00	93.33 ± 12.00
gl vs pn	0.98 ± 0.02	0.93 ± 0.04	93.75 ± 6.25	82.50 ± 15.60
me vs ly	0.95 ± 0.06	0.90 ± 0.06	90.00 ± 5.59	72.50 ± 13.70
me vs mm	0.99 ± 0.01	0.99 ± 0.01	95.00 ± 6.84	92.50 ± 5.22
me vs no	1.00 ± 0.00	1.00 ± 0.00	100.00 ± 0.00	96.00 ± 5.47
me vs pn	1.00 ± 0.00	1.00 ± 0.00	100.00 ± 0.00	85.00 ± 16.30
mm vs ab	1.00 ± 0.00	1.00 ± 0.00	100.00 ± 0.00	78.18 ± 15.21
od vs a2	1.00 ± 0.00	1.00 ± 0.00	84.00 ± 16.73	64.00 ± 32.86

Table 6
Mean ± standard deviation of AUC and (100-BER) values for all unbalanced classification experiments.

Experiments	MWVA.AUC	PCA.AUC	MWVA 100-BER	PCA 100-BER
G1 VS G2	0.97 ± 0.03	0.93 ± 0.06	90.22 ± 5.10	84.69 ± 10.76
G1 vs mm	0.98 ± 0.02	0.96 ± 0.00	96.23 ± 3.67	92.12 ± 4.52
a2 vs a3	0.89 ± 0.11	1.00 ± 0.00	55.00 ± 10.55	66.67 ± 24.90
a2 vs G2	0.99 ± 0.00	0.90 ± 0.00	94.05 ± 1.18	91.58 ± 28.56
a2 vs ly	1.00 ± 0.00	1.00 ± 0.00	86.67 ± 21.26	95.00 ± 21.38
a2 vs oa	1.00 ± 0.00	0.95 ± 0.10	100.00 ± 15.42	70.84 ± 6.84
a3 vs pn	1.00 ± 0.00	0.70 ± 0.00	85.00 ± 23.73	50.00 ± 9.77
G2 vs mm	0.99 ± 0.00	0.98 ± 0.00	93.46 ± 22.36	93.06 ± 0.00
gl vs a3	0.97 ± 0.00	0.93 ± 0.00	90.05 ± 0.00	83.01 ± 1.82
gl vs ab	0.99 ± 0.00	0.82 ± 0.14	93.62 ± 13.74	65.86 ± 21.54
gl vs ly	0.91 ± 0.00	0.88 ± 0.00	70.15 ± 18.18	75.24 ± 6.89
gl vs me	0.90 ± 0.00	0.70 ± 0.10	77.24 ± 7.18	56.96 ± 6.33
gl vs no	1.00 ± 0.00	1.00 ± 0.00	100.00 ± 0.00	98.67 ± 20.62
gl vs pn	0.97 ± 0.00	0.97 ± 0.00	87.86 ± 5.46	76.62 ± 5.69
me vs ly	0.96 ± 0.00	0.92 ± 0.00	85.24 ± 0.00	76.19 ± 4.37
me vs mm	0.98 ± 0.00	0.98 ± 0.00	95.23 ± 3.36	93.89 ± 5.13
me vs no	1.00 ± 0.00	1.00 ± 0.00	100.00 ± 12.24	95.24 ± 26.14
me vs pn	1.00 ± 0.00	0.98 ± 0.00	100.00 ± 0.00	87.39 ± 14.52
mm vs ab	1.00 ± 0.00	1.00 ± 0.00	99.00 ± 2.23	90.78 ± 7.47
od vs a2	1.00 ± 0.00	0.85 ± 0.13	81.67 ± 20.75	60.84 ± 10.86

For every group, $c1_{a,w,i}$ and $c2_{a,w,i}$ are w-dimensional vectors that represent the window of width w of the element a that starts at position i. Here, the elements in the two groups (tumour types in this study) are represented in numerical matrices X1 (of dimension $n1 \times n$) and X2 (of dimension $n2 \times n$), where $n1$ and $n2$ are the number of elements in groups 1 and 2, respectively, and n is the input dimension. Therefore, $c1_{a,w,i}$ and $c2_{a,w,i}$ are the w-dimensional vectors in row a starting at column i of X1 and X2, respectively. Vectors $\mu c1_{w,i}$ and $\mu c2_{w,i}$ are the mean w-dimensional vectors of $c1_{a,w,i}$ and $c2_{a,w,i}$ over a (the centroids of every group for a fixed width and starting point).

In order to determine the final width of the window, the values of $\Omega(w,t)$ were computed for increasing values of w (from 1 to n) and i (from 1 to $n - w + 1$) and stored in a triangular matrix with zeros on its upper diagonal called Dissimilarity Index Matrix (DIM). As an example, Fig. 5 displays a graphic illustration of the DIM matrix for the specific experiment gl *vs.* no (see tumour labels in Table 1), where good values for the width of the window can be visually inspected by looking for large values of $\Omega(w,t)$. In our study, and following the procedure described in (Arizmendi, et al. 2009), the value of the optimal width w was found to be 1. Therefore, every window corresponds to a single variable.

The initial feature selection was carried out by obtaining a ranking (in descending order) of the values of Ω. The variables were divided in groups corresponding to the set of variables whose energy gradually provided 1% of the total energy. Fig. 6 provides the average of the Ω ratio for the 20 tumour classification experiments investigated. The value of Ω can be seen to decrease exponentially as new variables are added, starting its linear trend with the variable 26 (11% of the total energy). Therefore, variables added from this point are not likely to contribute significantly to the classification task.

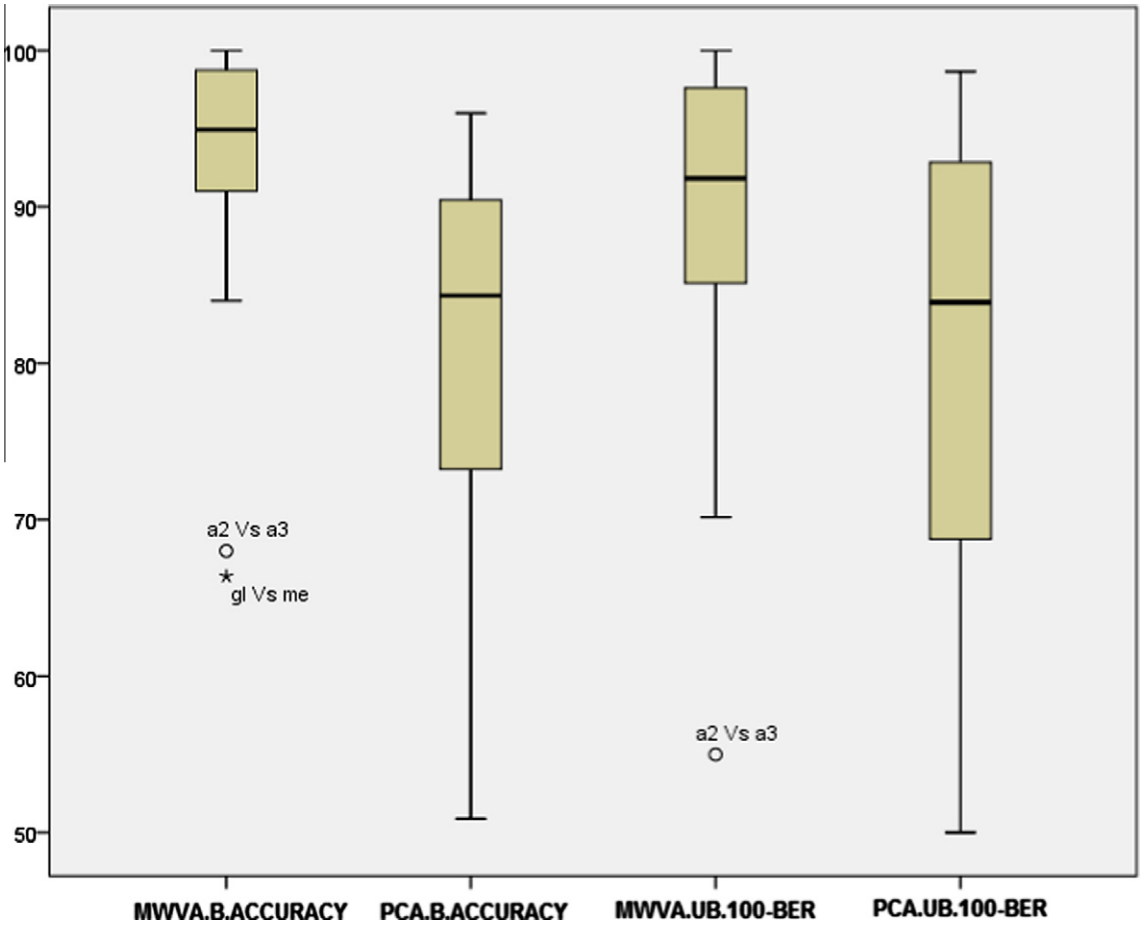

Fig. 7. Boxplot of the accuracy and 100-BER values corresponding to the unbalanced and balanced experiments of Tables 5 and 6. Each box represents the lower quartile (bottom line), median (line in the middle), and upper quartile (top line) values. The whiskers are lines extending from each end of the box to represent the extent of the rest of the data. Classification problems with atypical results are left outside the limits of the boxplots.

In the feature extraction experiments with PCA, principal components were added one at a time until the differential cumulative variance between two consecutive components was less than 1%.

A number of dimensionality reduction experiments were carried out and their results are summarized in Table 4. The number of selected (MWVA) and extracted (PCA) features varies widely depending on the problem at hand, as it would be reasonable to expect. These numbers vary from as little as three selected (for gl *vs.* a3, gl *vs.* ab, me *vs.* ly, me *vs.* pn) and extracted (for G2 *vs.* mm, a3 *vs.* pn, gl *vs.* pn, me *vs.* ly, od *vs.* a2) features, to as much as 29 selected features (a3 *vs.*c pn, which is a particularly complicated classification problem). These mostly parsimonious data representations yielded by the dimensionality reduction techniques should also ease the biomedical interpretation of the associated classification results.

3.4.2. Classification

Classification problems in this context are binary in nature (one tumour class against another, as multiple-class approaches are hindered by the limited number of MRS cases available; also, as remarked in (Luts et al., 2007), in medical practice, doctors frequently face situations of doubt between two different diagnosis, (i.e. types of tumour).

Feed-forward ANNs were used in the classification experiments starting from the features selected and extracted through dimensionality reduction, as reported in Table 4. Different network architectures between 5 and 40 units in the only hidden layer were employed. Given that all classifications are binary, one unit in the output layer does 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).

One run of a five-fold cross-validation was performed for each network, 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, as reported in Table 1), the original datasets were re-sampled, by over-sampling the minority class and under-sampling the majority class (Japkowicz, 2000).

Tables 5 and 6 show the detailed best results for all the analyzed problems. They include several quality indicators, including the balanced error rate (BER), the area under the ROC curve (AUC), and the accuracy for the balanced (B) and unbalanced (UB) groups.

3.4.3. Comparison with other studies using the INTERPRET database

The binary classification of low grade gliomas (G1) *vs.* meningiomas (mm), glioblastomas (gl) *vs* metastases (me), and me *vs.* mm was addressed in (García-Gómez et al., 2009). This study used a variation of the INTERPRET SET database analyzed here, which makes the comparison especially relevant. The authors reported a BER result of 0.91 for G1 *vs.* mm, with cross-validation, feature extraction with Independent Component Analysis (ICA) and a Least-Squares Support Vector Machine (LS-SVM) classifier (to be compared with a BER of 0.96 reported in Table 6); a BER result of 0.60 was reported for gl *vs.* me, with spectral Peak Integration (PI) and a Linear Discriminant Analysis (LDA) classifier (to be compared with a BER of 0.77 reported in Table 6); finally, a BER result of 0.95 was reported for me *vs.* mm, with PCA and a MLP classifier (identical to our result reported in Table 6).

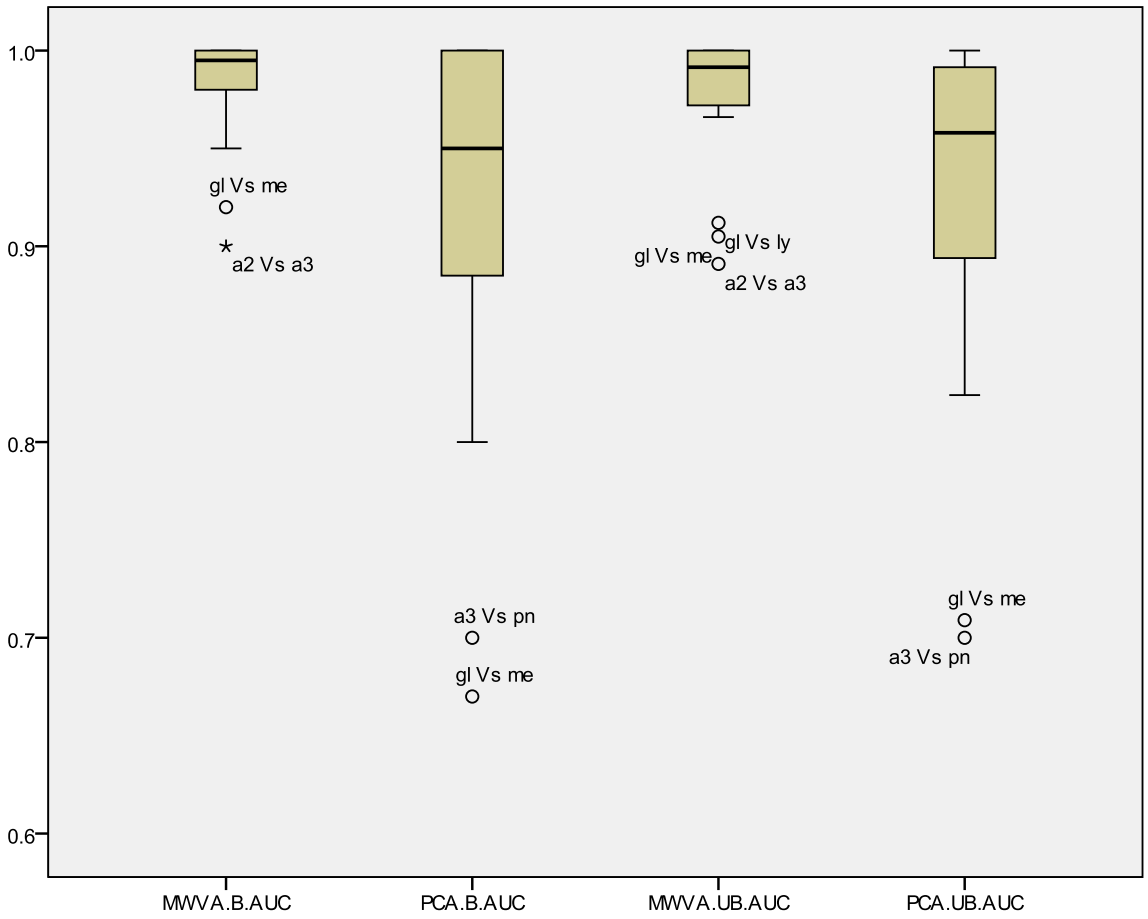

Fig. 8. Boxplots, as in Fig. 7, of the AUC values corresponding to the unbalanced and balanced experiments of Tables 5 and 6.

Some similar diagnostic classification problems were explored in (Lukas, 2004). They involved data from the same general INTERPRET database, but acquired at long echo time (LET). Experiments were reported for gl, me, mm, and low-grade astrocytomas (a2). More specifically, the following binary problems were considered: gl *vs.* mm, gl *vs.* me, gl *vs.* a2, mm *vs.* me, mm *vs.* a2, and me *vs.* a2. Data were classified using linear and nonlinear methods, namely, LDA, SVM and 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 0.90 with MWVA, as reported in Table 6). For mm *vs.* me, the best mean AUC in (Lukas, 2004) was 0.97 for LS-SVM without dimensionality reduction (corresponding to our 0.99 with both MWVA and PCA, in Table 6).

The work by Luts and colleagues (Luts et al. 2007) goes beyond our approach as they combined imaging and spectroscopy information, also from the same database, by employing magnetic resonance spectroscopic imaging (MRSI). All spectra were pre-processed through PI. Unfortunately, results from this study cannot be directly compared with ours, as multiple voxels from a very limited number of patients are used, instead of single voxels from multiple patients.

The difficult problem of discriminating between different grades of astrocytomas a2 and a3 is 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. The author reports a mean test accuracy of just under 70% for 20 PCs and LS-SVM. This can be compared with our result of around 68% (although with a very high standard deviation of 17.9%), in Table 5. Another classically difficult problem: me *vs.* gl is also dealt with in (Ladroue, 2003). A maximum accuracy of only 55% is reported. This compares to our result of 71.23%, using MWVA for dimensionality reduction.

A rather easier classification problem, that of discriminating between low-grade gliomas (G1) and high-grade malignant tumours (G2), has been far more commonly 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 0.97, for both balanced and unbalanced experiments with MWVA, reported in Tables 5 and 6. The same method was used in (Devos, 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 results of 0.98 and 0.99 for balanced and unbalanced experiments with MWVA, respectively reported in Tables 5 and 6.

All in all, these comparisons with previous results reported in similarly-oriented literature hint at the possibility that the differential advantage provided by our proposed method lies mainly in the wavelet-based spectra pre-processing stage of data analysis as well as in the use of MWVA for dimensionality reduction.

3.4.4. Discussion

Figs. 7 and 8 provide a telling summary of the results for all problems. All indicators consistently show that dimensionality reduction using MWVA tends to achieve better and more homogeneous results than PCA throughout the experiments. This is an encouraging outcome, given that feature selection provides solutions that are easier to interpret clinically than those obtained with feature extraction (González-Navarro et al., 2010). A Wilcoxon test was carried out to statistically compare the classification results corresponding to MWVA and PCA, looking for evidence of significant differences between the results obtained through both methods. Results of this test are reported in Table 7 and sustain the previous comments.

Table 7
p-Values corresponding to the Wilcoxon test of the comparison between classification results obtained with the feature selection provided by MWVA and the feature extraction provided by PCA.

PCA	MWVA	P-values
PCA.B.AUC	MWVA.B.AUC	0.005
PCA.B.ACCURACY	MWVA.B.ACCURACY	0.001
PCA.UB.AUC	MWVA.UB.AUC	0.004
PCA.B.BER	MWVA.UB.BER	0.007

The few atypically poor results highlighted in these figures alert us of the special difficulty of some classification experiments. In some cases, such difficulty is well reported in the existing literature, as for astrocytomas of similar grade: a2 *vs.* a3 (see for instance, (Ladroue, 2003), or high-grade malignant tumours: gl *vs.* me (see Romero, Vellido, Julià-Sapé, & Arús, 2009). In a few other cases (gl *vs.* ly, a3 *vs.* pn), such difficulty has not been reported.

Finally, the results obtained with balanced datasets are consistently better than those obtained with the original unbalanced datasets, although not in a statistically significant way, as evaluated through a Wilcoxon test. This holds both for feature selection with MWVA and extraction with PCA. These results can only justify to a certain extent the use of the class-balancing strategy.

4. Conclusion

The diagnosis of neuro-oncology pathologies is a critical task for medical experts in hospital environments. Most decisions in this context bound to be made on the basis of a combination of doctors' experience and background knowledge, and information gathered through non-invasive measurement techniques.

Human brain tumours are a very diverse family of pathologies, and require an accurate diagnosis of both type and grade of malignity, in order to select those tailored therapies that could maximize the chances of survival. The availability of tumour information in the form of signal and image makes the use of computer-based diagnostic assistance advisable.

A sensitive stage in the computer-based analysis of tumour information is that of data pre-processing. In this paper, we have outlined and evaluated a method in which MRS signal information is first processed using DWT techniques, a data transformation that is followed by feature selection or extraction for dimensionality reduction. All this is a preliminary step for diagnostic-oriented binary classification using Bayesian ANNs.

The comparison of the results obtained in our experiments with those previously reported in the recent scientific literature provide evidence of the adequacy of the proposed data pre-processing approach. To the best of the authors' knowledge, some of the classification experiments in which encouraging results have been obtained had never previously been investigated as a problem of binary classification. Such experiments should therefore provide medical experts in neuro-oncology with some preliminary knowledge related to the discriminability of the concerned tumour pathologies on the basis of MRS information.

In this study, spectroscopy data acquired at short echo time were analyzed. Future work should involve similar experiments with data acquired at long echo times or even with data obtained by combination of different acquisition echo times.

Acknowledgments

This research was partially funded by Spanish MICINN R+D projects TIN2006-08114 and TIN2009-13895-C02-01. Authors gratefully acknowledge the former INTERPRET European project partners. Data providers: Dr. C. Majós (IDI), Dr. À. Moreno-Torres (CDP), Dr. F.A. Howe and Prof. J.Griffiths (SGUL), Prof. A. Heerschap

(RU), Prof. L Stefanczyk and Dr J.Fortuniak (MUL) and Dr. J. Calvar (FLENI); data curators: Dr. M.Julià-Sapé, Dr. A.P. Candiota, Dr. I. Olier, Ms. T. Delgado, Ms. J. Martín and Mr. A. Pérez (all from GABRMN-UAB). GABRMN coordinator: Prof. C. Arús.

References

Arizmendi, C., Vellido, A., & Romero, E. (2009). Frequency selection for the diagnostic characterization of human brain tumours. *Frontiers in Artificial Intelligence and Applications, 202*, 391–398.

de Graaf, A. A., & Bovée, W. M. (1990). Improved quantification of *in vivo* ^{1}H NMR spectra by optimization of signal acquisition and processing and by incorporation of prior knowledge into 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., et al. (2004). Classification of brain tumours using short echo time ^{1}H MR spectra. *Journal of Magnetic Resonance, 170*(1), 164–175.

Donoho, D. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory, 41*(3), 613–627.

Foresee, F., & Hagan, M. (1997). Gauss-Newton approximation to Bayesian regularization. In *Proceedings of the IEEE international conference on neural networks (ICNN)* pp. 1930–1935.

García-Gómez, J. 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., 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.

Guo, D et al. (2000). A study of wavelet thresholding denoising. In *Proceedings of the IEEE 5th international conference on signal processing (WCCC-ICSP)* pp. 329–332.

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.

Japkowicz, N. (2000). The class imbalance problem: significance and strategies. In *Proceedings of the 2000 international conference on artificial intelligence (IC-AI'2000)* pp. 111–117.

Julià-Sapé, M., Acosta, D., Mier, M., Arús, C., & Watson, D.the INTERPRET Consortium. (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.

Ladroue, C.L.C. (2003). Pattern recognition techniques for the study of Magnetic Resonance Spectra of brain tumours, PhD thesis, St. George's Hospital Medical School, UK.

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

Leardi, R. (2000). Application of genetic algorithm–PLS for feature selection in spectral data sets. *Journal of Chemometrics, 14*(5-6), 643–655.

Luca, T., Mainardi, D., Pietro, L., Giuseppe, S., & Cerutti, S. (2002). A wavelet packets decomposition algorithm for quantification of *in vivo* ^{1}H-MRS parameters. *Medical Engineering & Physics, 24*(3), 201–208.

Lukas, L. (2004). Brain tumor classification based on long-echo proton MRS signals. *Artificial Intelligence in Medicine, 31*(1), 73–89.

Luts et al. (2007). A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection. *Artificial Intelligence in Medicine, 40*(2), 87–102.

MacKay, D. (1992). The evidence framework applied to classification networks. *Neural Computation, 4*(5), 720–736.

Mallat, S. (1999). *A wavelet tour of signal processing*. San Diego (CA), USA: Academic Press.

Minguillón, J., Tate, A.R., Arús, C., & Griffiths, J.R. (2002). Classifier combination for *in vivo* Magnetic Resonance Spectra of brain tumours. In Roli, F., and Kittler, J. (Eds.), *Multiple Classifier Systems. Lecture Notes in Computer Science, 2364* pp. 282–292.

Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J-M. (2002). *Wavelet toolbox for use with MatLab*. The Math Works, Inc.

Nature (2009), 461 (7265), pp. 697–836.

Olarte, O., & Sierra, D. (2007). Determinación de los parámetros asociados al filtro wavelet por umbralización aplicado a filtrado de interferencias electrocardiográficas. *UIS Ingenierías, 6*(2), 33–44.

Rivas, E., Burgos, J. C., & García-Prada, J. C. (2009). Condition assessment of power OLTC by vibration analysis using Wavelet Transform. *IEEE Transactions on Power Delivery, 24*(2), 687–694.

Romero, E., Vellido, A., Julià-Sapé, M., & Arús, C. (2009). Discriminating glioblastomas from metastases in a SV ^{1} H-MRS brain tumour database. In *Proceedings of the 26th annual meeting of the european society for magnetic resonance in medicine and biology (ESMRMB)* p. 18.

Serrai, H., Nadal-Desbarats, L., Poptani, H., Glickson, J., & Senhadji, L. (2000). Lactate editing and lipid suppression by continuous Wavelet Transform analysis: Application to simulated and ^{1} H-MRS brain tumor time-domain data. *Magnetic Resonance in Medicine, 43*, 649–656.

Serrai, H., Senhadji, L., Wang, G., Akoka, S., & Stroman, P. (2003). Lactate double quantification and lipid signal suppression using a new biexponential decay filter: Application to simulated and ^{1}H MRS brain tumor time-domain data. *Magnetic Resonance in Medicine, 50*(3), 623–626.

Steward, B., & Kleihues, P. (2003). *World health organization*. IARC Press: World Cancer Report.

Tate, R., Underwood, et al. (2006). Development of a decision support system for diagnosis and grading of brain tumours using *in vivo* Magnetic Resonance single voxel spectra. *NMR in Biomedicine, 19*(4), 411–434.

Vellido, A., & Lisboa, P. J. G. (2009). *Investigating human cancer with Computational Intelligence techniques. KES recent research results series*. East Sussex, UK: Future Technology Press.

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