Diagnosis of Brain Tumours from Magnetic Resonance Spectroscopy using Wavelets and Neural Networks

Carlos Arizmendi, Juan Hernández-Tamames, Enrique Romero, Alfredo Vellido, Francisco del Pozo

Abstract—The diagnosis of human brain tumours from non-invasive signal measurements is a sensitive task that requires specialized expertise. In this task, radiology experts are likely to benefit from the support of computer-based systems built around robust classification processes. In this brief paper, a method that combines data pre-processing using wavelets with classification using Artificial Neural Networks is shown to yield high diagnostic classification accuracy for a broad range of brain tumour pathologies.

I. INTRODUCTION

Human brain tumours are complex and often aggressive pathologies of low prevalence but significant social impact. The accurate diagnosis of these tumours is essential in order to provide a prognosis of tumour development: life expectancy largely depends on the accurate estimation of the tumour type and grade.

This study addresses the problem of human brain tumour diagnosis on the basis of biological signal data obtained by Magnetic Resonance Spectroscopy (MRS). MRS has a role in providing biochemical information to aid the radiological diagnosis of brain tumours [1], enabling the quantification of metabolite concentrations non-invasively, and thereby avoiding serious risks of brain damage. Clinicians are often not trained to make sense of the MR spectral signal. Additionally, the natural high dimensionality of the spectra, the presence of noise and artefacts, and the low amount of data usually available for specific brain tumour types, all complicate their diagnostic-oriented classification. As a result, the computer-based, semi-automated processing, analysis and interpretation of the MRS spectra should be valuable as support for medical decision makers.

In this study, we analyze a set of MRS data from the multi-centre, international INTERPRET database [2], using several methodologies that involve signal processing, feature selection and classification. Often, the most determinant step in this computer-based data analysis is data pre-processing. For this, we first use the Discrete Wavelet Transform

(DWT) and a filtering process, together with data compression, for the decomposition of the spectra in terms of approximation and detail coefficients, in a change of representation of the spectra that entails minimum loss of relevant information. This decomposition by itself does not alleviate the high dimensionality of the data. For this reason, dimensionality reduction is implemented using Moving Window with Variance Analysis (MWVA) [3] for feature selection and Principal Components Analysis (PCA) for feature extraction. The processed data are classified 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, some of which have seldom been analyzed in this setting.

II. ANALYZED DATA

This study relies on a database created under the framework of the European project INTERPRET [2], an international collaboration of centers from four different countries. The database includes a set of single-voxel proton MRS (SV ^{1}H-MRS), measured at short echo time (SET: 273 patients). A total of 195 frequency intensity values (measured in parts per million (ppm), an adimensional unit of relative frequency position in the data vector), were considered in this study. Class labeling was performed according to the World Health Organization (WHO) system for diagnosing brain tumours by histopathological analysis of a biopsy sample. The analyzed database includes the classes listed in Table I, which consist of nine tumour pathologies, plus abscesses and normal brain tissue.

III. METHODOLOGY AND RESULTS

A. Wavelet Transform

The Wavelet Transform (WT) is a linear operation that decomposes a signal into components at different scales [4]. The WT of a function $f(t)$ for a wavelet function $\psi(t)$ is given by (1), where a is the scale and τ the position of the wavelet $\{a, \tau: \in R\}$. The inverse transform is given by (2).

$$Wf(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \Psi \left(\frac{t - \tau}{a} \right) dt \qquad (1)$$

$$f(t) = \frac{1}{K_\Psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Wf(a, \tau) \Psi(a, \tau) \frac{da d\tau}{a^2} \quad (2)$$

$$K_\Psi = \int_{-\infty}^{\infty} \frac{|\Psi(w)|^2}{w}\,dw \quad < \infty$$

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

An important development for the application of wavelet theory in Discrete Signal Processing was presented by Mallat [5] using Multirresolution Analysis (MA). In this context, the WT in a discrete domain is implemented via an octave filter bank, as a cascade of low- and high-pass filters, followed by sub-sampling, as illustrated in Figure 1. The reconstruction procedure, except for rounding errors, leads to the restoration of the original signal if no coefficient is altered.
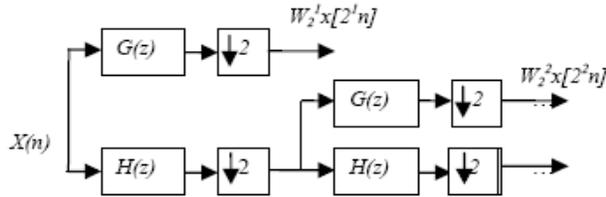

Figure 1. Decomposition algorithm of the DWT with two decomposition levels, the original signal $x(n)$ is passed through the high-pass filters $G(Z)$ and low-pass filters $H(Z)$.

The application of Mallat's model [5] together with Donoho's approach for signal filtering by thresholding [6] and statistical coefficients for data compression permit reducing the noise level, as well as representing the MRS signal without loss of relevant information, while keeping the dimensionality of the system as low as possible.

B. Wavelet Filtering with Threshold or Shrinkage

Frequently, the observed signal $X(t)$ can be considered 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 an estimate of $S(t)$, or $\hat{S}(t)$. The suggested model allows this through the use of WT as described in (3), where

$$Y = W_{(\Psi,j)}(X)$$
$$Z = D(Y,\lambda) \qquad (3)$$
$$\hat{S} = W_{(\Psi,j)}^{-1}(Z)$$

$D(.,\lambda)$ is the filtering operator for threshold λ and $W_{(\psi,\,j)}(.)$ and $W^{-1}_{(\psi,\,j)}(.)$ denote, in turn, the WT and its inverse, with wavelet function ψ and j decomposition levels. The denoising of the available MRS spectra was carried out according to the following three consecutive steps [4], each described in its own sub-section.

B.1. Threshold calculation

Three alternative choices of threshold were considered in the experiments, according to the following statistical estimators developed by Donoho [6]:

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

- *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.

- *Threshold Minimax:* The threshold is selected following the minimax principle, commonly used in statistics to design estimators (see [7]).

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

B.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 are 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.

- *Mln*: As *Sln* but with the calculation of detail coefficients level by level.

B.3. Implementation of the threshold

Once the threshold is calculated and scaled, the thresholding process $D(Y,\lambda)$ is implemented through two alternative methods: Hard thresholding ($D_h(Y,\lambda)$) and Soft thresholding ($D_s(Y,\lambda)$) according to (4).

$$D_h(Y,\lambda) = \begin{cases} Y, & |Y| \geq \lambda \\ 0, & |Y| < \lambda \end{cases}$$
$$\qquad (4)$$
$$D_s(Y,\lambda) = \begin{cases} sgn(Y)(|Y| - \lambda), & |Y| \geq \lambda \\ 0, & |Y| < \lambda \end{cases}$$

C. Wavelet Mother Selection for MRS Data

The DWT was applied to the original signal data, making the decomposition to the maximum allowable level (see Section III.A). It was implemented with different mother wavelets, ranging from different orders of Biorthogonals (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, 8), Coiflet (1 to 5), Daubechies (1 to 43), and Symlet (1 to 25). For every mother wavelet, the absolute values of each decomposition coefficient 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). Finally, *Q1* the index was computed as follows:

$$Q1(r) = \frac{\overline{SNR}_{Re}(r)}{\overline{MSE}_{Re}(r) + \overline{NDC}_{Re}(r)} \qquad (5)$$

where r is the order analyzed, and the *Re* subindex corresponds to the rescaled data between 1 and 3. The maximum values of *Q1* indicate the orders with the best reconstruction error using the minimum NDC. The 2 highest values of *Q1* in each wavelet function are shown in Table II.

TABLE II
PERFORMANCES INDEX FOR OPTIMAL WAVELET SELECTION

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	193.9	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

Once the initial set of wavelet orders were chosen, as shown in Table II, the filtering methodology explained in section B was used to denoise the spectrum signal and to eliminate irrelevant information. In order to determine the appropriated scaling, Donoho [6] recommends the MSE as a measure of performance for each of the experiments. Therefore, the MSE was calculated for each spectrum of the reconstructed signal following the scheme described in (3), implementing all the options allowed by the combination of threshold estimation (*Sqtwolog*, *Rigrsure* and *Minimax*), threshold scaling (*Sln*, *One* and *Mln*), and *Hard* thresholding. The *Hard* function was used because it often yields smaller MSE than the *Soft* one, and also because it preserves the magnitudes of the MRS spectra. The results obtained 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 are compared in Figure 2, showing that the *Rigrsure-Sln-Hard* procedure produces the best results among all combinations.
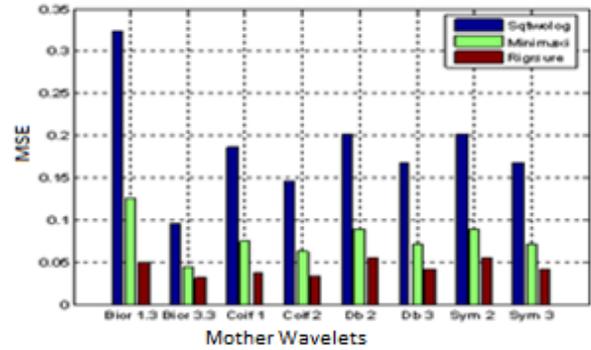

Figure 2. Comparison of the MSE for the three thresholds when Sln is applied.

To determine the final wavelet, the average value of several statistics was computed for the *Rigsure-Sln-Hard* combination. They include: SNR, Preserved Energy (EP), Percentage of Distortion (PRD) and Compression Ratio (CR), expressed as follows:

$$EP = \frac{\sum_{k=1}^{N}[\hat{x}(k)]^2}{\sum_{k=1}^{N}[x(k)]^2} * 100 \qquad (6)$$

$$PRD = \sqrt{\frac{\sum_{k=1}^{N}[x(k) - \hat{x}(k)]^2}{\sum_{k=1}^{N}[x(k)]^2}} * 100 \quad (7)$$

$$CR = L_O/L_C \qquad (8)$$

where $\hat{x}$ is the reconstructed signal, L_o is the cardinality of the decomposition coefficients of the original signal, and L_c is the cardinality of decomposition coefficients different to zero. This set of statistics has been used to choose the optimal wavelet in previous related works concerning ECG signal filtering [8] and classification tasks [9], among others. For a more objective criterion in choosing the optimal wavelet function, the index *Q2* was computed:

$$Q2(r) = \left[\frac{\overline{SNR}_{Re}(r) + \overline{EP}_{Re}(r) + \overline{CR}_{Re}}{\overline{MSE}_{Re}(r) + \overline{PRD}_{Re}(r)} \right] \qquad (9)$$

The *Q2* values for the wavelet functions of Table II are shown in Table III. It can be observed that the maximum value for this index is given by the Biortogonal (3.3). Therefore, this wavelet was chosen as optimal for our study.

D. Dimensionality Reduction and Classification

After processing the MR spectra with wavelet Biortogonal (3.3) and filtering it with the combination *Rigsure-Sln-Hard*, we proceeded to reduce the dimensionality of the data using MWVA and PCA, taking as input variables the decomposition coefficients. The MWVA is a feature selection filter method proposed in [3]. For PCA, principal components were added one at a time until the differential cumulative variance between two consecutive components was less than 1%. An average of 10.15 and 13 variables were obtained for MWVA and PCA respectively

TABLE III
STATISTICS FOR THE EVALUATION OF THE PERFORMANCE OF
MOTHER WAVELETS

Wavelet	MSE	SNR	EP	PRD	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

TABLE IV
MEAN AND STANDARD DEVIATION OF AUC VALUES FOR THE
EXPERIMENTS IN WHICH DIMENSIONALITY WAS REDUCED
USING MWVA AND PCA

Experiments	MWVA	PCA	SET
G1 *vs* G2	0.97±0.03	0.92±0.06	1
G1 *vs* mm	0.98±0.02	0.98±0.02	1
a2 *vs* G2	0.99±0.01	0.94±0,07	1
gl *vs* me	0.90±0.07	0.73±0.09	1
a2 *vs* oa	1.00±0.00	0.95±0.05	1
gl *vs* no	1.00±0.00	1.00±0.00	1
G2 *vs* mm	0.99±0.01	0.97±0.02	1
me *vs* mm	0.98±0.01	0.99±0.00	1
me *vs* no	1.00±0.00	1.00±0.00	1
a2 *vs* a3	0.89±0.11	0.90±0.07	2
a2 *vs* ly	1.00±0.00	0.96±0.04	2
a3 *vs* pn	1.00±0.00	0.70±0.00	2
gl *vs* a3	0.97±0.03	0.92±0.06	2
gl *vs* ab	0.99±0.01	0.75±0.04	2
gl *vs* ly	0.91±0.08	0.95±0.03	2
gl *vs* pn	0.97±0.06	0.93±0.06	2
me *vs* ly	0.96±0.05	0.89±0.08	2
me *vs* pn	1.00±0.00	0.98±0.02	2
mm *vs* ab	1.00±0.00	1.00±0.00	2
od *vs* a2	1.00±0.00	0.95±0.06	2

Feed-forward ANN with one hidden layer were used in the classification experiments. Each network was trained using 5, 10, 30 and 40 units in the hidden layer and one unit in the output layer. The networks were trained with Bayesian regularization and back-propagation, updating the weights and bias according to the Levenberg-Marquardt algorithm [10]. One run of a 5-fold cross-validation was performed for each network, with a maximum of 500 epochs. Table IV shows the best resulting values of the area under the ROC curve (AUC) for each experiment. G1 (low grade gliomas) is the union of classes a2, oa and od. G2 (high-grade malignant tumours) is the union of classes gl and me. In this table, a value of 1 in the SET column indicates experiments which have been reported in previous research with, at best, comparable results, while a value of 2 indicates experiments that, to the best of our knowledge, have not been previously investigated in a similar setting.

The results of a Wilcoxon test show that the differences among the mean and median classification values for MWVA and PCA are statistically significant (*p-value = 0.011*), in favour of MWVA.

IV.   CONCLUSIONS.

In this study, a DWT procedure was applied to the pre-processing of MR spectra corresponding to several brain tumour pathologies. This procedure yielded very encouraging results in terms of diagnostic discriminatory binary classification. In particular, they were quite good in the classification of pathologies for which few results, if any, had been previously reported. The results are of special relevance for the experiments gl *vs* me and a2 *vs* a3, which according to existing literature are specially difficult classification problems.

The proposed methodology for selecting the optimal wavelet developed study concludes that the Biortogonal (3.3) wavelet, implemented with the combination *Rigsure-Sln-Hard*, generates the best MR spectra representation without loss of relevant information. The dimensionality reduction performed using PCA and MWVA achieved an average data compression of 93.29% and 94.76% respectively.

REFERENCES

[1]   M. Murphy, A. Loosemore *et al.* "The contribution of proton magnetic resonance spectroscopy (^{1}H MRS) to clinical brain tumour diagnosis", Br J Neurosurg, vol. 16, pp. 329–334, 2002.

[2]   M. Julià-Sapé, D. Acosta, M. Mier, C. Arús, D. Watson, and the INTERPRET Consortium, "A multi-centre, web-accessible and quality control checked database of in vivo MR spectra of brain tumour patients," Magnetic Resonance Materials in Physics. MAGMA, vol. 19, pp. 22–33, 2006.

[3]   C. Arizmendi, A. Vellido, E. Romero . "Frequency selection for the diagnostic characterization of human brain tumours", Frontiers in Artificial Intelligence and Applications, vol. 202, pp. 391-398, 2009.

[4]   P.D. Agoris *et al.* "Threshold selection for wavelet denoising of partial discharge data," International Symposium on Electrical Insulation, September 2004, USA.

[5]   S. Mallat, "A Wavelet Tour of Signal Processing," Academic Press, 1999, USA.

[6]   D. Donoho, "De-noising by soft-thresholding," IEEE Trans. on Information theory, vol.41, Issue. 3, pp. 613-627, 1995.

[7]   D. Guo -Fei *et al.* "A study of wavelet thresholding denoising," IEEE Proceedings of ICSP, vol.1, pp. 329-332, 2000.

[8]   O. Olarte, D. Sierra. "Determinación de los parámetros asociados al filtro wavelet por umbralización aplicado a filtrado de interferencias electrocardiográficas," Universidad Industrial de Santander UIS Ingenierías, vol. 6,Issue 2, pp 33-44, 2007.

[9]   E. Rivas, J. Burgos, J. García, "Condition assessment of power OLTC by vibration analysis using wavelet transform ". IEEE Transactions On Power Delivery, vol. 24, pp. 687-694, 2009.

[10]   F. Foresee, and M. T. Hagan. "Gauss-Newton approximation to Bayesian regularization," Proceedings of the 1997 International Joint Conference on Neural Networks 1997, pp. 1930-1935.