33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Brain Tumour Classification Using Gaussian Decomposition and Neural Networks

Carlos Arizmendi, Daniel A. Sierra, Alfredo Vellido, Enrique Romero

Abstract— **The development, implementation and use of computer-based medical decision support systems (MDSS) based on pattern recognition techniques holds the promise of substantially improving the quality of medical practice in diagnostic and prognostic tasks. In this study, the core of a decision support system for brain tumour classification from magnetic resonance spectroscopy (MRS) data is presented. It combines data pre-processing using Gaussian decomposition, dimensionality reduction using moving window with variance analysis, and classification using artificial neural networks (ANN). This combination of techniques is shown to yield high diagnostic classification accuracy in problems concerning diverse brain tumour pathologies, some of which have received little attention in the literature.**

I. Introduction

Decision making in neuro-oncology is a sensitive undertaking. In this area, in which most diagnostic techniques must be non-invasive, clinicians may benefit from the second opinion provided by automated computer-based MDSS. The availability of such second opinion may reduce the inherent uncertainty in the diagnosis and prognosis of tumours and, thus, facilitate medical practice [1].

This study addresses the problem of human brain tumour diagnosis from the biological signal obtained by MRS. This is a signal in the frequency domain that peaks at specific frequencies or frequency bands most of which are known to correspond to the resonances of specific chemical and biochemical components of the tissue. 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 [2]. We do so using several methodologies that involve signal processing, feature selection and classification, namely and in turn: Gaussian Decomposition (GD) to transform the signal in terms of coefficients of amplitude, standard deviation, and translation [3]; moving window with variance analysis (MWVA) [4] for feature selection; and ANN classifiers with Bayesian regularization [5]–[7].The proposed combination of techniques 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.

This work was partially supported by the Spanish MICINN under grants TIN2006- 08114 and TIN2009-13895-C02-01.

C. Arizmendi, A. Vellido, and E. Romero are with the Department of Computer Languages and Systems at Technical University of Catalonia, Barcelona, 08034, Spain. A. Vellido, corresponding author; phone: +34-93-4137796; fax: +34-93-4137883. avellido@lsi.upc.edu

D.A. Sierra is with the School of Electrical, Electronics and Telecommunications Engineering, Industrial University of Santander, Bucaramanga, 680002, Colombia dasierra@uis.edu.co

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 time of echo (STE: 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 for analysis. 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 spectra are part of this database, and include the classes listed in Table I: nine tumour pathologies, plus abscesses and normal brain tissue.

TABLE I
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

III. Methodology and Results

A. Gaussian Decomposition

Given the nature of the analyzed spectral signals, which are the combination of several sources, it is reasonable to address the problem of their processing through GD (illustrated in Fig. 1). In GD, we assume that a segment of spectral signal (delimited by the $[\lambda_1, \lambda_2]$ interval) can be modeled as:

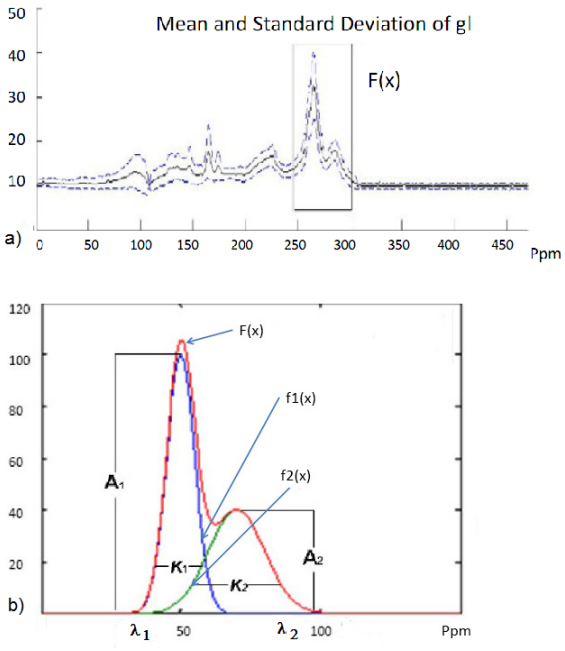

Fig. 1. Top: Illustration of MR spectra: mean $\pm$ standard deviation of *gl* tumour type. Bottom: Illustration of GD of signal $F(x)$ into $f_1(x)$ and $f_2(x)$

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

$$F(x) = 0, \qquad \lambda_1 > x > \lambda_2 \quad (2)$$

where $P = \{A, k, \mu\}$, $k = -1/2\sigma^2$, and $P \in \Re^+$ contains the amplitude (A), standard deviation (σ), and translation (μ) of each one of the n Gaussians.

If we let $Y_j = F(x_j, P)$ be the observed function and $\hat{Y}_j = F(x_j, \hat{P})$, the estimated function, GD attempts to solve the optimization problem

$$\min_{\hat{P}} \left(\varphi(\hat{P}) \right) = \sum_{j=1}^{m} \left(Y_j - \hat{Y}_j(\hat{P}) \right)^2 \quad (3)$$

So, the goal is finding the $\hat{P}$ that minimizes $\varphi(\hat{P})$, evaluated over all the m measured wavelengths x_j. For this, plenty of deterministic optimization methods (e.g. Powell, Flectcher-Reeves, Dandon, Levenberg Marquardt, Nadler, etc.) and non-deterministic ones (e.g. ANN, Genetic Algortihms, Differential Evolution, etc.) are available [8].

To minimize P, we used the Levenberg-Marquardt (LM), Trust-Region Dogleg (TR) and Gauss-Newton (GN) [9] methods, aiming to find the algorithm that best adjusts the spectral signal with Gaussians functions. Different tests, not reported here for the sake of brevity, were carried out to evaluate the performance and accuracy of the fitting. These tests showed TR was the best choice to fit our signals.

B. Selection of the Optimal Number of Gaussians

There is no general consensus about which methodology is the most adequate to decompose and select the optimal number of Gaussians in the signal fitting process [10], due to the many aspects of the problem to be considered (e.g. fitting technique, algorithms, methods to choose the starting points, etc.).

In order to choose a criterion for the choice of this optimal number, an illustrative experiment with 5 artificial Gaussians was carried out. These Gaussians were added, resulting in one single signal, which was then fitted with an increasing number of Gaussians, up to 20 (i.e. the first fitting is done with one Gaussian, the second with two, and so on, until fitting with 20 Gaussians). In each trial i of the reconstruction, the fitting signal $\hat{Y}_i$, formed with i gaussians, was compared with the original signal Y through different quality measures: Mean Square Error (MSE), Power Distortion (PD) and Energy Preservation (EP), with results reported in figure 2.

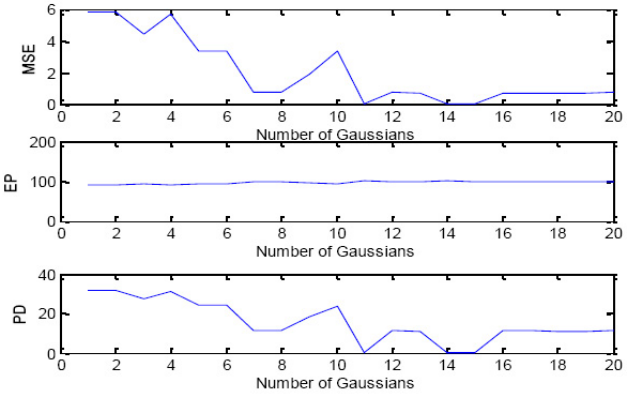

Fig. 2. MSE, EP and PD of the artifitial signal generated, and fitted with a maximun of 20 gaussians

These results indicate that neither MSE nor PD have monotonically decreasing trends (which is plausible because the signal was generated with a given small number of Gaussians). In addition, the selection of starting parameters (initial $\hat{P}$) may lead to convergence to local minima. To avoid this problem, the fitting should be performed with as many trials as possible, and increasing in each trial the number of Gaussians, with different starting points. Figure 2 shows that the PD is 24.12% when 5 Gaussians were used to fit the spectra, whereas a PD of 0.00028% was obtained when 14 Gaussians were used. Therefore, the algorithm requires a redundant number of Gaussians to obtain adequate parameter estimates. Performing the fitting with a redundant number of Gaussians results in a precise estimation of the original parameters, while the less relevant Gaussians are left with very small amplitude, standard deviation and translation values, a situation that has also been observed in previous work [10].

C. Filtering and Baseline Correction

The quality of the decomposition can be improved by adjusting the baseline and filtering the spectra. The TR algorithm is robust enough for fitting a signal with two Gaussians, but the method does not yield adequate results when the signal is potentially composed of a large number of Gaussians. Typical inadequate results include negative magnitudes for the spectra and signal distortion, resulting in the loss of relevant information and misclassification. Because of this, baseline adjustment was performed to those spectra with original negative magnitudes, adding the absolute value of the minimum value found in each one.

As a pre-processing stage, a half-band wavelet filtering was performed using the *Biorthogonal* 3.3 mother wavelet. The adequacy of this mother wavelet has been shown in a previous study, where it achieved a decomposition of the spectra with the minimum number of coefficients, while keeping the MSE to a minimum and yielding the highest signal-to-noise ratio and the lowest PD [11]. Figure 3 exemplifies this process with a brain tumour spectrum, which was fitted using baseline correction and wavelet filtering. The TR algorithm was applied with the following settings: a) the starting values of the translation μ_i, defining the center of each Gaussian, were equally spaced in each trial; b) each initial Gaussian amplitude A_i was computed as the average of those 3 consecutive sample values closer to the associated translation value; and c) the starting value of standard deviation δ_i was set to 1.

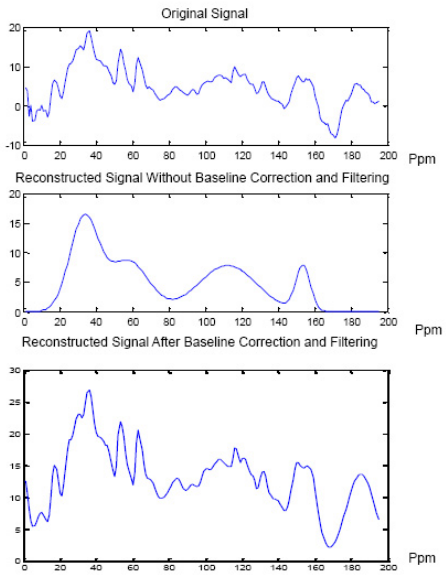

Fig. 3. Top: original example MR spectrum. Middle: reconstructed signal without preprocessing. Bottom: reconstructed signal with the pre-processing procedure described in the main text (filtering and baseline correction).

Since each MRS in the study has a total of 195 clinically relevant frequency intensity values, and there are 3 coefficients (amplitude, standard deviation and translation) to describe each Gaussian, the maximum number of Gaussians allowed by the algorithm, for a given spectrum, is 65. For each trial (that is, for each of the reconstructions with a different number of Gaussians), the statistics MSE, PD and EP were computed, taking the minimum PD of the 65 trials for each spectrum as the representative solution. The mean ($\pm$ standard deviation) of the obtained PD was 8.09 ± 11.95 for fittings of raw data, while a PD of 3.87 ± 1.88 was obtained for fittings of pre-processed data. This result illustrates the better accuracy of the pre-processing procedure followed by the Gaussian Decomposition method to fit the spectra.

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 differentials in MSE between consecutive samples were computed and normalized, so as to produce a maximum value of 100%. The addition of Gaussians stopped when the differential MSE improvement was lower than 1%, and the remaining ones were eliminated. The MSE for the complete MRS data set before this selection of Gaussians resulted in a mean $\pm$ standard deviation of 0.14 ± 0.30; the corresponding values after it were 0.16 ± 0.31, suggesting that the reconstruction of the signal was not significantly affected by this process.

D. Dimensionality Reduction and Classification

Having pre-processed the MR spectra through GD, 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 set to 195 by setting to zero the values corresponding to translations other than those of the Gaussians. This procedure also served to identify repetitions in the translations (by repetition here we understand two translations within the same integer interval). When one such a repetition in translation was identified, the lowest of both translations was moved the closest lower integer value, whereas the highest of both was moved to the closest higher integer value, thus avoiding the overlapping. Having re-arranged the data this way, feature selection was carried out using the MWVA technique [4], [11], from the re-scaled amplitude and standard deviation vectors. A modification of MWVA was implemented, with the dissimilarity index matrix (DIM) obtained in each experiment. This was accomplished by concatenating the amplitude and standard deviation DIM's. Once the new DIM was obtained, feature selection was carried out using the energy criterion described in [4].

E. Classification and Results

Classification problems in the context of this study are binary in nature (one tumour class against another, as multiple-class approaches are hindered by the limited

number of MRS cases available; also, and as remarked in [12], in medical practice, doctors frequently face situations of doubt between two alternative diagnosis, that is, two types of tumour).

Feed-forward ANNs were used in the classification experiments, starting from the features selected and extracted following the procedures described in the previous sections. Different ANN architectures, with between 5 and 40 units in the single 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 [5] as part of a back-propagation process. The adaptive weights and biases were updated according to the Levenberg-Marquardt algorithm [13].

One run of a 5-fold cross-validation was performed for each network, allowing a maximum of 500 epochs. Table II summarizes the results for 20 classification experiments, where G1 (low-grade gliomas) is the union of tumour types a2, oa and od, and G2 (high-grade malignant tumours) is the union of tumour types gl and me. Several quality indicators are reported, including the accuracy (ACC), and the area under the ROC curve (AUC), for the concatenated (CO) DIMs. Most classification problems yield very accurate results, with the worst ones being for diagnostic problems of well-known difficulty, such as me *vs.* gl, or G1 *vs.* G2.

TABLE II

MEAN $\pm$ STANDARD DEVIATION OF AUC AND ACCURACY VALUES FOR ALL CLASSIFICATION EXPERIMENTS.

Experiments	CO Accuracy	CO AUC
$G1\ vs\ G2$	87.35 ± 8.45	0.91 ± 0.08
$G1\ vs\ mm$	89.75 ± 9.78	0.99 ± 0.03
$a2\ vs\ a3$	96 ± 8.94	1.00 ± 0.00
$a2\ vs\ G2$	91.8 ± 0.18	0.96 ± 0.07
$a2\ vs\ ly$	100 ± 0	1.00 ± 0.00
$a2\ vs\ oa$	96 ± 8.94	1.00 ± 0.00
$a3\ vs\ pn$	86.67 ± 18.26	1.00 ± 0.00
$G2\ vs\ mm$	88.13 ± 6.86	0.94 ± 0.04
$gl\ vs\ a3$	94.92 ± 5.26	1.00 ± 0.00
$gl\ vs\ ab$	97.42 ± 3.54	1.00 ± 0.00
$gl\ vs\ ly$	96.25 ± 3.42	0.99 ± 0.03
$gl\ vs\ me$	77.9 ± 2.37	0.87 ± 0.10
$gl\ vs\ no$	96.67 ± 3.04	0.99 ± 0.03
$gl\ vs\ pn$	98.75 ± 2.8	1.00 ± 0.00
$me\ vs\ ly$	90 ± 10.46	1.00 ± 0.00
$me\ vs\ mm$	95 ± 8.15	0.99 ± 0.02
$me\ vs\ no$	100 ± 0	1.00 ± 0.00
$me\ vs\ pn$	100 ± 0	1.00 ± 0.00
$mm\ vs\ ab$	98.18 ± 4.07	1.00 ± 0.00
$od\ vs\ a2$	96 ± 8.94	1.00 ± 0.00

IV. CONCLUSIONS

In this study, the Gaussian Decomposition signal processing technique was used to break down a given MR spectrum into its component coefficients of amplitude, standard deviation and translation of its constituent tones. These coefficients of the constituents tones, which can be associated to specific metabolites, in conjunction with the axis of transformation and the the concatenation of amplitude and standard deviation DIM's, yielded encouraging results in terms of diagnostic discriminatory binary classification. These results are of special relevance for experiments involving tumour types seldom dealt with in the existing literature, such as *ly* or *pn*.

V. ACKNOWLEDMENTS

Authors 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: Drs. Julià-Sapé, Candiota and Olier, Ms. Delgado, Ms. Martín and Mr. Pérez (GABRMN-UAB). Prof. C. Arús, GABRMN coordinator.

REFERENCES

[1] P. Lisboa, A. Vellido, R. Tagliaferri, F. Napolitano, M. Ceccarelli, M. Guerrero, and E. Biganzoli, "Data mining in cancer research," 2010.

[2] M. Julià-Sapé, D. Acosta, M. Mier, C. Arùs, and D. Watson, "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*, vol. 19, no. 1, pp. 22–33, 2006.

[3] U. Haud, "Gaussian decomposition of HI surveys. IV. Galactic intermediate-and high-velocity clouds," *Arxiv preprint arXiv:0802.4163*, 2008.

[4] C. Arizmendi, A. Vellido, and E. Romero, "Frequency selection for the diagnostic characterization of human brain tumours," in *Proceeding of the 2009 conference on Artificial Intelligence Research and Development: Proceedings of the 12th International Conference of the Catalan Association for Artificial Intelligence*. IOS Press, 2009, pp. 391–398.

[5] D. MacKay, "The evidence framework applied to classification networks," *Neural Computation*, vol. 4, no. 5, pp. 720–736, 1992.

[6] v. d. M. O. Bioch, J.C. and R. Potharst, "Classification using Bayesian neural nets," in *Proceedings Benelearn 95*. Universiteit Brussel, Brussel, 1995, pp. 79–90.

[7] R. Neal, *Bayesian Learning from Neural Networks*. Springer-Verlag, New York, 1996.

[8] J. Valdés, A. Barton, and A. Haqqani, "Analysis of mass spectrometry data of cerebral stroke samples: an evolutionary computation approach to resolve and quantify peptide peaks," *Genetic Programming and Evolvable Machines*, vol. 9, no. 3, pp. 257–274, 2008.

[9] K. Madsen, H. Bruun, and O. Imm, "Methods for non-linear least squares problems," 2004.

[10] P. Kalberla, W. Burton, D. Hartmann, E. Arnal, E. Bajaja, R. Morras, and W. Poppel, "The Leiden/Argentine/Bonn (LAB) Survey of Galactic HI: Final data release of the combined LDS and IAR surveys with improved stray-radiation corrections," *Arxiv preprint astro-ph/0504140*, 2005.

[11] C. Arizmendi, J. Hernández-Tamames, E. Romero, A. Vellido, and F. del Pozo, "Diagnosis of brain tumours from magnetic resonance spectroscopy using wavelets and Neural Networks," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 6074–6077.

[12] J. Luts, A. Heerschap, J. Suykens, and S. Van Huffel, "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*, vol. 40, no. 2, pp. 87–102, 2007.

[13] D. Foresee and M. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Neural Networks, 1997., International Conference on*, vol. 3. IEEE, 2002, pp. 1930–1935.