On the Use of Long TE and Short TE SV MR Spectroscopy to improve the Automatic Brain Tumor Diagnosis

Juan M. García-Gómez[1], Salvador Tortajada[1], César Vidal[1],
Margarida Julià-Sapé[2], Jan Luts[3], Sabine Van Huffel[3],
Carles Arús[2], Montserrat Robles[1]
[1]IBIME, Itaca, Universidad Politécnica de Valencia,
[2]Universitat Autònoma de Barcelona,
[3]Katholieke Universiteit Leuven, ESAT-SCD (SISTA)

January 31, 2007

Abstract

The statistical improvement of the brain tumour classification by two Echo Times together instead of only Short TE (20-32 ms) or only Long TE (135-136 ms) was measured empirically. A clinically validated dataset of 50 low grade meningiomas (MEN), 105 aggressive tumors (AGG, glioblastoma and metastases) and 30 low-grade glial tumours (LGG, astrocytomas grade II, oligodendrogliomas and oligoastrocytomas) was used to fit and evaluate predictive models based on the combination of features from the Short TE and Long TE spectra. Feature selection by means of the Stepwise algorithm and the Relief-F algorithm, and feature extraction by Principal Components Analysis (PCA) were applied to get relevant features from the original spectra. Least Squares Support Vector Machines and Linear Discriminant Analysis were applied to fit the pairwise and multiclass classifiers respectively. Significant differences among the performance estimations of the classifiers were found when Short TE, Long TE or both spectra were used. For discriminating between MEN from the other two classes, both Short TE and Long TE combined are needed to obtain the best results. On the other hand, for distinguishing between AGG and LGG, acquisition of short TE should suffice. LS-SVMs were useful when combining Short TE and Long TE features and when using full spectra due to its generalization capability based on regularization.

Keywords: MRS; Short TE; Long TE; Pattern recognition; Brain; Cancer

1 Introduction

Brain tumors are the second fastest growing cause of cancer death among people older than 65 years [1], in addition, they are also the second leading cause of cancer death (after leukemia) in children under fifteen years and young adults up to the age of thirty-four.

Information provided by non-invasive radiological techniques are highly relevant for evidence-based medicine application for diagnosis and patient management and for surgery [2, 3]. Nowadays, the diagnosis and treatment of brain tumors is based on

clinical observations, radiological appearance and often a histopathological diagnosis of a biopsy. Magnetic resonance imaging (MRI) is widely used for determining tumor extension in surgical and radiotherapy planning. MRI provides an initial diagnosis with a sensitivity of 14-100% depending on tumor type and precission of the radiological orientation wording used [4]. Magnetic Resonance Spectroscopy (MRS) is slowly becoming an additional accurate non-invasive technique for initial diagnosis exam of brain masses [3, 5], due to its capability to provide useful chemical information of metabolites for characterizing brain tumors and its complementary role to MRI [6].

The complexity of the MRS signals and their underlying mixture of biological compounds makes the Pattern Recognition (PR) discipline, in combination with Biomedical Signal Processing, the technological choice to produce clinically usable knowledge from the acquired biomedical data. The predictive tools should solve routine questions quickly, to reinforce the radiologists diagnosis found by MRI inspection through a direct and intuitive response. On the other hand, when uncommon cases are analyzed, radiologists may use the predictive tools, more as a research support tool for comparing and analyzing the results and relevant features provided by the models of bilateral discrimination questions already solved on available data [7].

Short TE (20-35 ms) MRS data acquisition allows to observe several metabolites and other compounds considered useful for tumor classification. Hence, macromolecules (MM; 5.4 ppm, 2.9 ppm, 2.25 ppm, 2.05 ppm, 1.4 ppm and 0.87 ppm) are only observed in Short TE spectra [3]. mI and mobile lipids are better detected in Short TE [8]. Single Voxel Short TE MRS is fast (typically 5 min) and robust, so it is very useful for clinical studies [3]. The difficulty of the analysis of Short TE signal is due to a large number of overlapping peaks, a strong macromolecule/membrane originated baseline and a certain sensitivity to artefacts [5].

Long TE (about 135 ms) MRS signals are, but on the one hand, less informative than Short TE, on the other hand, easier to extract relevant information for classification than Short TE MRS [5]. Some resonances with shorter T2 relaxation time than those of the main metabolites will be attenuated at Long TE, and as a result, other interesting resonances are better observed. Lipids resonances (1.3 and 0.9 ppm) and macromolecules will not be the dominating components at long echo times, so contributions from lactate (doublet at 1.33 ppm) and alanine (doublet at 1.47 ppm) are observed [9, 3] as inverted peaks.

Many pattern recognition applications to cancer research have been reported in the last two decades (for reviews, see [10]) and more specifically in brain tumor research (for reviews, see [7, 11, 12]). Hagberg in [11] summarizes classification studies based on PR and clustering methods. Eight of those studies were applied to brain tumor discriminations from normal tissue or other CNS diseases. All of them were based on Linear Discriminant Analysis (LDA) or Artificial Neural Networks (ANN) applied to relative metabolites levels or PCA transformations and evaluated by Leave-One-Out cross-validation. Several works described results for classification of brain tumors based on the MR data available within the INTERPRET project [13, 7, 14]. On the one hand, [5] was interested in the classification using Long TE MRS, on the other hand, works as [7, 15, 16, 17] carried out experiments with Short TE MRS. At the same time, some works related to INTERPRET focused in the classification of tumors based on ^{1}H Magnetic Resonance Spectroscopic Imaging [18, 19].

Feature extraction and selection methods from NMR have been of interest because of the special character of the MRS signals [20, 16, 21, 22]. Recently, Menze et al. [23] published a very complete benchmark of quantization and other feature extrac-

tion methods combined with learning methods to discriminate between recurrent and non-recurrent brain tumors using Long TE MRS. Their results performs at least as well as a manually supervised quantification, and yield, for the automated methods 5-10% higher accuracy than for the manual. Based on Least Squares Support Vector Machines (LS-SVMs) [24], Devos, Lukas et al authors in [17, 5] developed different classifiers for in-vivo short echo time spectra and MRSI with good results. Finally, classifiers based on the Relief-F algorithm combined with LS-SVM for MRSI have been reported to produce good performance [25].

This work focuses on the statistical characterization of the improvement in automatic classification of brain tumors when Short TE, Long TE spectra or both kinds of spectra together are used as predictive variables in Pattern Recognition models. Short TE and Long TE offer complementary views of the chemical composition of brain tumours, but nobody has combined in a single classifier for predictive purpose. For our purpose, multiclass and binary classifiers, based on LDA and LS-SVM with different feature selection and extraction techniques, were trained using the three approaches (Short TE only, Long TE only or Long TE + Short TE). An empirical evaluation was carried out to compare the prediction performance of the models and characterize when their combination produces an advantage in the discrimination of tumours. The results of this study may be of use to guide the selection of acquisition conditions and stages to take into account in the development of decission support systems for brain tumour in-vivo diagnosis.

Accordingly, our aim has been to improve the automatic classification of brain tumors by the combination of Long TE and Short TE spectra into a PR model, following the conclusions reported by Majos [26] on separated Long TE and Short TE data analysis. PR methods are expected to take advantage of the complementary information when combining both spectra.

The rest of this paper is organized as follows. Section 2 reviews the origin and protocols used in the study and summarizes the PR methods and evaluation methodology applied in the experiments. Section 3 describes the experiments of the combination of Short TE and Long TE spectra using multiclass and binary classifiers. Section 4 analyzes the features selected during the experiments, their biological signification and an overview of the outliers. Section 5 summarizes the final conclusions and describes further possible work on the combination of Short TE and Long TE using Pattern Recognition.

2 Methods

2.1 Data acquisition

Data were provided by six international centers in the framework of the INTERPRET [13] project. Data were acquired with SV MRS at 1.5 T. In the acquisition of Long TE spectra, the PRESS sequence was used, with a TR between 1500 and 2020 ms, TE of 135 or 136 ms, spectral width of 1000 or 2500 Hz and 512 or 2048 datapoints. Short TE were acquired using PRESS or STEAM sequences, with TR between 1600 and 2020 ms, TE of 20 or 32 ms, spectral width of 1000 or 2500 Hz and 512 or 2048 datapoints. Each spectrum used for training and testing had passed a quality control based on the validation of both Long TE and Short TE spectra by the INTERPRET QC committee [7].

Table 1: Samples per class used in the study. Short TE and Long TE were acquired during INTERPRET according to a consensus protocol [29]

SuperClass	Class	Samples	Total samples (n)
Aggressive (AGG)	Glioblastoma Multiforme (GM)	77	
	Metastasis (ME)	28	105
Meningioma (MEN)	Meningioma (MEN)	50	50
Low-grade glial mixture (LGG)	Astrocytoma grade II (A2)	20	
	Oligodendroglioma (OD)	5	
	Oligoastrocytoma (OA)	5	30
TOTAL		185	185

Table 2: Datasets used in the study

Dataset	TE[range](ppms)	D(features)	C(classes)	n (samples)
Short TE	Short TE[4.1, 0.5]	190	3 (AGG, MEN, LGG)	185
Long TE	Long TE[4.1, 0.5]	190	3 (AGG, MEN, LGG)	185
Long TE + Short TE	Long TE[4.1, 0.5],Short TE[4.1, 0.5]	380	3 (AGG, MEN, LGG)	185

The brain tumors biopsies were histopathologically classified by three consulting pathologists, based on the histological classification of tumors of the central nervous system set up by the World Health Organization [27].

There were 185 cases in the INTERPRET database [28] that had both SV MRS at Short TE and at Long TE. These were grouped into three diagnostic super-classes: Aggressive tumors -glioblastomas and metastases- (AGG) with 105 patients, Low-grade Meningiomas (MEN) with 50 patients and Low Grade Glial mixture (LGG) of Astrocytomas grade II (A2), Oligodendrogliomas (OD) and Oligoastrocytomas (OA), with 30 patients. Table 1 summarizes the composition of the super-classes in the datasets.

Each sample consists of one Long TE spectrum and one Short TE spectrum. Data is analyzed in the frequency domain with 512 points from 7.1. to -2.7 ppm. Spectra were referenced to total Creatine at 3.03 ppm, Choline at 3.21 ppm or mobile lipids at 1.29 ppm in these order of priority depending on SNR and tumour pattern, see [7] and the INTERPRET webpage [13] for further details. The region of interest was restricted to the interval [4.1,0.5] ppm, composed by 190 points. The same approach than the INTERPRET project [15, 7] was followed in this paper, extended to the simultaneous use of two echo times for classifier development. This consists in handling the points of the region of interest as independent variables.

Figure 1 shows a Long TE spectrum and a Short TE spectrum in the first row and a combined view of both spectra in the second row: "one after the other" approach with Short TE spectrum after the Long TE spectrum. Our approach considers the x-points of the "one after the other" view as independent variables.

As a result, three datasets were generated from the available data for the experiments presented in this paper. Table 2 summarizes the Short TE dataset, the Long TE dataset and the Long TE + Short TE dataset. Afterwards, the datasets were split in training, validation and test subsets during the application of the evaluation methodology.

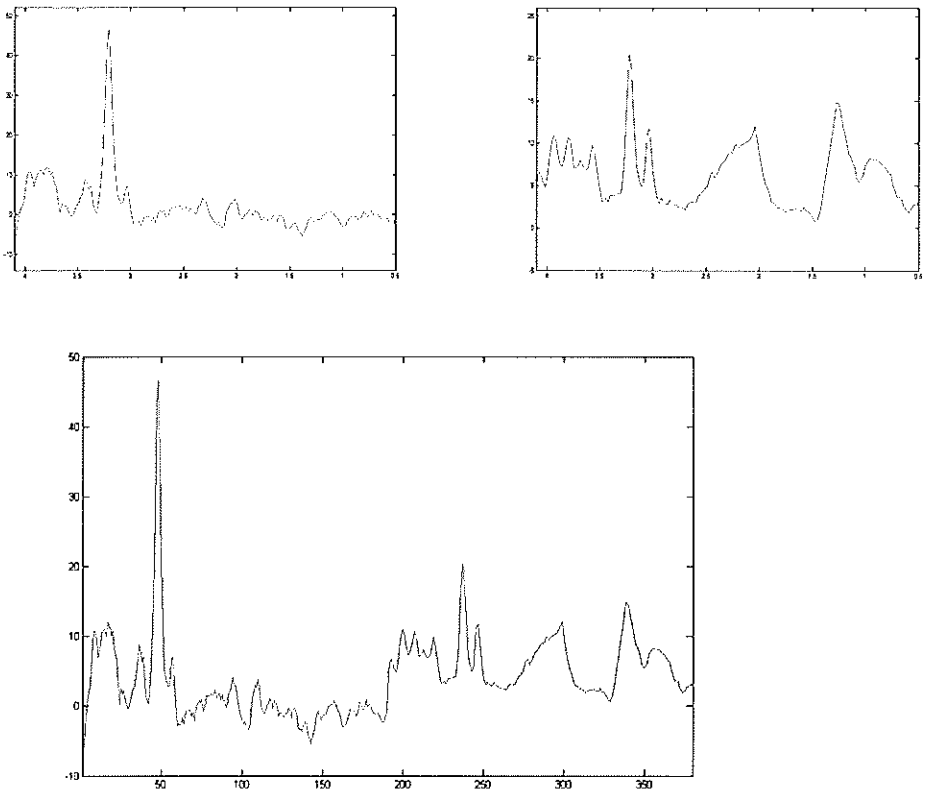

Figure 1: First row shows the Long TE spectrum and the Short TE spectrum from a meningioma, INTERPRET case I0080 [28](x-axis in ppms, y-axis in arbitrary units (a.u.)), the second row the combined view "one after the other" of both spectra (x-axis in relative position of points, y-axis in a.u.)

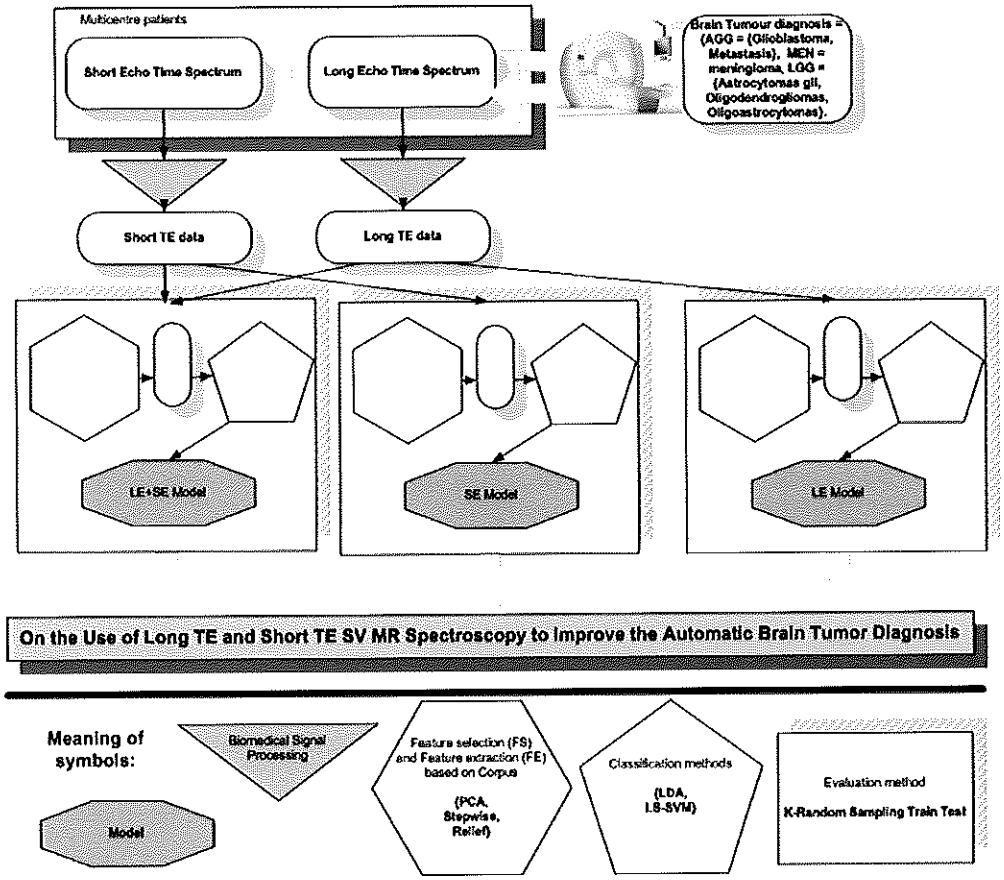

Figure 2: Methodology to study the contribution of Short TE and Long TE spectra to the brain tumor classification.

2.2 Data analysis

Figure 2 summarizes graphically the methodology applied in this study. The main goal was to evaluate the models fitted from Long TE MRS only, models from Short TE MRS only and models using both Long TE and Short TE MRS.

We present the summary of the signal processing carried out to preprocess the spectra and the PR methodology applied to extract features, train the models and estimate their performance.

2.3 Biomedical signal processing and quality control

Long TE and Short TE raw data, acquired in the time domain at six centers, were preprocessed following the well-established MRS data pre-processing protocol and quality control developed during the INTERPRET project [13]. This protocol has been also adopted by the EU eTumour project [30] in order to guarantee the coherence in PR studies and Decision Support System development.

The inputs of the procedure were the water suppressed free induction decay (FID) and water unsuppressed FID signals for each acquisition time. For each spectrum, the expert-spectroscopist evaluation protocol was carried out [7, 29] and after the spectrum was validated as acceptable, the spectrum was included in the database [13].

Processing was carried out by automated modules as in [7] with 1Hz line broadening. In a few instances where problems arose (e.g. unsuppressed water file not available) for which manual processing using MRUI [31] followed by an in-house preprocessing and unit length normalization module was carried out esentially as described [7]. After that, the pre-processing protocol for each spectrum was based on the INTERPRET DMS (Data Manipulation Software) development [32].

2.4 Feature selection and extraction

Feature selection (FS) and feature extraction (FE) are complementary steps commonly used in PR, wherein a relevant subset of the features (FS) or a transformation (FE) of the original features from the data are obtained for application of a learning algorithm. Brief explanation of the FS and FE methods are presented next, for more details extended reviews can be found in [33, 34].

2.4.1 The full spectra

The advantage of using the whole spectrum rather than a set of features is that no assumptions are made as which components are important for discrimination. The main disadvantage of this approach is the high dimensionality the classifier should deal with, and regularization methods are needed in order to avoid the curse of dimensionality [35]. New generation learning methods, such as the Support Vector Machines and its variants, such as the Least Squares Support Vector Machines, are especially designed for dealing with high dimensionality, allowing us to incorporate more predictive variables in the models [36, 24].

2.4.2 Stepwise algorithm for classification

Stepwise selection is one of the most extended algorithms for solving a feature selection task. It applies to regression and classification problems in both statistics and PR disciplines.

Since exhaustive search is impractical when dimensionality grows, the greedy hill climbing approach is applied [37]; then, the subset of features with the highest performance measure will be selected in each step and modified in the next step by the addition or deletion of one variable in the model.

The procedure starts using the initial model, defined by the starting variables provided. For each step, a set of new models are generated by including every single variable that is not in the model (add-one procedure), and by excluding every single variable that is in the model (drop-one procedure) [38]. Stepwise selection compares the performance of the evaluated models in each step by a measure using cross-validation over the training dataset.

Different performance measures exist in the literature, and we adopted the definitions used by Garczarek on [39], where the AS measure (Ability to Separate) measure is defined as the measure of how unlike the classes are after the transformation carried out by the model. The AS measure is calculated by means of equation 1, where the Euclidean distance between the true posterior probability of the sample x for each class c

$(p(c|x))$ and the posterior probability of the model $(p(c|x;\Theta))$ is used.

$$AS = 1 - \frac{\sqrt{C}}{n\sqrt{C-1}} \sum_{i=1}^{n} \sqrt{\sum_{c=1}^{C} (p(c|x_i) - p(c|x_i;\Theta))^2}, \quad (1)$$

where n is the number of samples and C is the number of classes.

If the maximum performance measured in the step improves the performance of the current model plus the tolerance ϵ, then, a new model is made by the inclusion or exclusion of the associated variable, if not, the procedure stops. The procedure stops when performance converges or when the maximum number of variables is reached.

Posterior probabilities obtained by LDA (see section 2.5.1) were used to measure the Ability to Separate in the Stepwise procedure.

2.4.3 Relief-F feature selection

Relief-F is a feature selection method based on how well features distinguish between instances that are near to each other [40]. It has been applied in combination with LS-SVM in MRSI+MRI data with good results [25].

In binary classification problems, the estimation of the quality of each variable is calculated by the accumulation of the distance between randomly selected instances and their k-nearest neighbors of a different class minus the distance to the k neighbors of the same class. The distance is normalized by the range of the feature $(max(x) - min(x))$.

For a selected case of one class and the nearest neighbors of the other class, a large distance for a variable will substantially increase the quality of the variable, whereas, for neighbors of the same class, a large distance for a variable will substantially decrease its quality. Hence, a discriminative variable will have a large distance with samples of other classes and a short distance with samples of the same class.

2.4.4 Principal Components Analysis (PCA)

Principal Components Analysis (PCA) is a well-known projection method commonly used for feature extraction in PR [34].

PCA maps the original D-dimensional data into an orthogonal P-space, where the axes of this new coordinate system lie along the direction of maximum variance of the original data. The more correlated the original variables are, the more variation in the data is explained by the first principal components (or loadings). Hence, feature reduction can be carried out discarting the remaining PCs.

The transformation matrix U(loading vectors $\mathbf{u}_p$) is usually computed by means of the eigenvalue decomposition

$$\hat{S}\mathbf{u}_p = \lambda_p \mathbf{u}_p,$$

where $p \leq min(n, D)$ and using the covariance matrix $\hat{S}$ of the training cases

$$\hat{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T$$

Classifiers are fitted using the highest scores $(\tilde{x}_p = \mathbf{u}_p(\mathbf{x} - \mu))$ of the training cases. In the same way, the loading matrix allows to compute the scores of the test cases to evaluate the performance of the model. In our study, data were normalized to $N(0, 1)$ before applying the PCA transformation.

2.5 Classification methods

2.5.1 Linear discriminant analysis (LDA)

Given a set of samples, we search for a projection that maximizes the separation between the classes projected values. Fisher's proposed solution is a ratio between the difference of the projected means and a measure of dispersion of each class. Ideally, this function should be optimum when the distance between means is maximum and the inside class dispersion is minimum.

2.5.2 Least Squares Support Vector Machines (LS-SVM)

Support Vector Machines (SVM) are classification, nonlinear function estimation and density estimation methodology defined in the context of statistical learning theory and structural risk minimization [36]. SVM defines the optimal separating hyperplane between two classes with the maximal margin. This margin is the minimum distance of patterns of the training set to the hyperplane. Support Vector Machines represent the data in a higher dimensional space where the linear separating hyperplane is built. The explicit construction of the mapping to the higher dimensional space is not necessary by the kernel trick [24].

Least Squares Support Vector Machines (LS-SVM) is a reformulation of the SVMs where a linear system is solved [24]. This characteristic made LS-SVM useful to solve efficiently high dimensional problems and it has been applied extensively for MRS-based classification [5].

Let $\{x_i, y_i\}_{i=1}^n$ be the training set where $x_i \in \mathbb{R}^D$ has the associated label $y_i \in \{-1, 1\}$. The LS-SVM classifier [24] is defined by

$$y(x) = sign[\sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b], \tag{2}$$

where α_i are the support values and b is the independent term corresponding to the solution of a linear system. $K(,)$ is a positive definite kernel, e.g. the RBF kernel as used in this paper ($K(x, x_i) = exp\{-||x - x_i||_2^2/\sigma^2\}$).

2.6 Evaluation procedures

Given the nature of the data used in this study, it is uncommon to have large datasets, and we consider that we are trading in a "small sample" study. In our case we used the following tools to evaluate the performace of our experiments.

Empirical evaluation procedures The goal of empirical evaluation procedures is to obtain the estimation of the true accuracy (or true error) of an inferred classifier by applying it to real data.

There is a consensus in the literature that resampling techniques have to be applied when estimating classifier error (or accuracy) in small-sample datasets [41, 42, 43]. With small samples, random resampling procedures (as cross validation or k-random sampling) are usually applied.

k-Random Sampling Train-Test procedure in external loop of the tuning (model selection, including feature selection) and training procedures [44] was used for obtaining the predictive performance of the models trained with Long TE + Short TE, Long TE and Short TE datasets. External evaluation has shown good estimation of

the performance of the models, avoiding underestimation due to bias produced during model selection. Random Sampling Train-Test evaluations are suggested when a limited number of cases are available.

Receiver operating characteristic (ROC) curves Receiver operating characteristic (ROC) curve is a graphical technique for assessing the performance of a binary classifier system [45]. The sensitivity vs. (1 - specificity) values obtained for different variations of the threshold in the variable under study are represented in a XY plot. The Area Under the ROC Curve (AUC) measures the discrimination capability of a specific classifier depending on its sensitivity and specificity.

ROC was applied in the evaluation of the LS-SVM results varying the threshold of its latent space.

2.6.1 Confidence intervals and significance of differences

We followed the Berrar et al.' caveat presented in [46] for estimating the confidence intervals appropriately. In this review, $(1-\alpha)$%-credibility interval (equivalent to confidence interval in the bayesian framework) for the true error are calculated based on Jeffreys' Beta distribution as:

$$\tau \simeq \epsilon + \frac{2(M-2m)z\sqrt{0.5}}{2M(M+3)} \pm z\sqrt{\frac{\epsilon(1-\epsilon)}{M+2.5}}$$

when $10 \leq M \leq 200$ and $0 \leq m \leq M/2$, being M the test size and m number of errors, ϵ is the observed error and $z = 1.96$ for 95% confidence. In the experiments, equivalent confidence intervals to accuracy were reported; they were calculated based on the accuracy estimated from the k-random sampling train-test evaluation and the test size of one repetition.

Significance of differences among the Short TE, Long TE and Long TE + Short TE combination approach were tested by means of the Friedman's (F) nonparametric two-way analysis of variance test. The test was applied using the best models of Long TE + Short TE and the equivalent models for Long TE only and Short TE for each binary classification.

Comparison of the best models for each binary classification was done by means of the Kruskal-Wallis (KW) nonparametric one-way analysis of variance. When multiple comparisons were carried out, results were analyzed by Tukey's (MCT) honestly significant difference criterion.

3 Experiments

The evaluation procedure for all experiments was 150-random sampling train-test with stratified test sets. The partitions made in the repetitions were random and independent among the experiments. The evaluation was carried out in a nested-loop that covered the feature and model selection in order to avoid underestimations of the true error.

Multiclass and binary classifiers were applied to characterize the behavior of the Short TE models, the Long TE models and the Long TE + Long SE models. Multiclass classifiers discriminate among the three aforementioned super-classes (AGG, MEN and LGG) together. Binary classifiers discriminate between the three pairwise combinations super-classes (AGG vs. MEN, AGG vs. LGG and MEN vs. LGG).

Experiments with Multiclass classifiers. Two procedures were applied to the multiclass classifier, the stepwise procedure based on LDA and the PCA-based feature extraction with a LDA classifier.

For each repetition of the k-random sampling train-test procedure, the stepwise method was used as described in section 2 to extract discriminant subsets of features. Then, the test subset was used to evaluate the performance of the LDA model using the selected features. Three randomly selected features were used in the initialization of the stepwise search. The procedure stopped when no more than 0.1% of improvement was obtained in one step with respect the previous one.

Table 3 shows the k-Random Sampling Train Test evaluation of the Stepwise feature selection based on Ability to Separate (AS) of the predictive LDA classifier applied on Long TE + Short TE, Short TE and Long TE datasets.

Table 3: Results for the classification among Aggressive tumors -Glioblastomas and Metastasis- (AGG), Meningiomas (MEN) and a Low Grade Glial mixture of Astrocytomas grade II, Oligodendrogliomas and Oligoastrocytomas (LGG). 150-Random Sampling Train Test evaluation of the Stepwise feature selection based on Ability to Separate (AS) of the predictive LDA classifier.

Dataset	NVars [min,max]	Accuracy [CI]
Long TE + Short TE	[12,22]	88.71 [83.60,92.67]
Short TE	[5,11]	88.82 [83.74,92.75]
Long TE	[7,14]	82.50 [76.58,87.46]

The behavior of the Long TE + Short TE and Short TE models are similar, because the contribution of the Short TE variables are more relevant than the Long TE ones in Long TE + Short TE models. The accuracy of the Long TE models is considerably lower than the accuracy of the Long TE + Short TE and Short TE models. The number of features in the Long TE + Short TE models varied from 12 to 22. In contrast, Short TE models used a smaller amount of features (from 5 to 11) to produce comparable results. Long TE features are selected in the Long TE + Short TE model but they are not contributing as much as Short TE to the combined model.

The second approach for multiclass discrimination applies PCA to reduce the space dimensionality of the input space and then applies the LDA method to obtain the classification function.

The main goal of the PCA experiment is to study the contribution of the variance of the Short TE and Long TE features in the discrimination of the three-classes by the combined and individual models. For this reason, curves of the accuracy in front of the number of components p were generated for each approach.

The rotation matrix (loading vectors) of the Principal Component Analysis (PCA) is computed using the training cases with normalized variables to $N(0,1)$, then LDA classifier is trained using the p scores of the training cases. The normalization and loading matrix allows us to compute the scores of the test cases to evaluate the performance of the model. The procedure was repeated inside the k-random sampling train-test evaluation loop to avoid the bias in the error estimation.

Figure 3 shows the accuracy estimated by k-Random Sampling Train Test of the PCA feature extraction and LDA classifier applied to Long TE + Short TE, Short TE and Long TE datasets. The number of Principal Components (PCs) used varies from 2 to 15 to show the behavior of the models depending the variables used in the models.

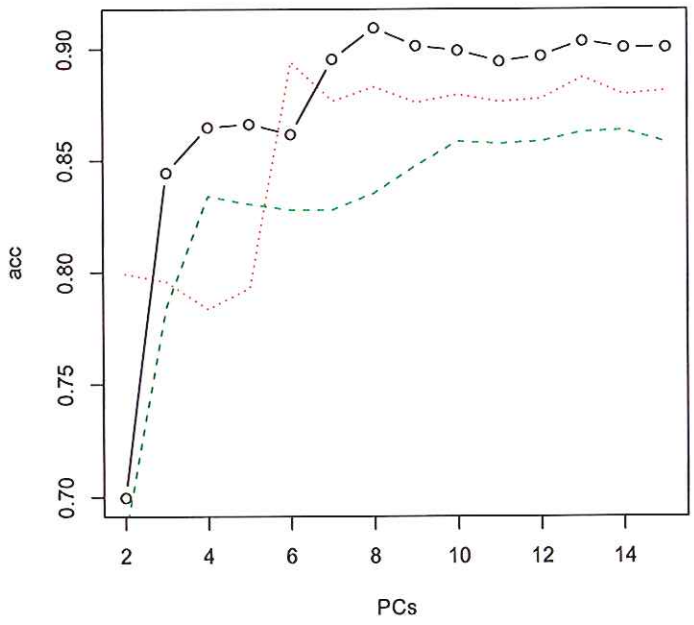

Figure 3: Accuracy vs. Number of Principal Components (PCs) for the Long TE + Short TE (solid line and points) approach follows the Long TE (dasshed line) approach in the 3-6 PCs region and Short TE (dotted line) when more PCs are incorporated in the model.

In general, the expected accuracy using both spectra is higher than the accuracy observed of the Short TE only and Long TE only approaches. Three regions are observed in the accuracy curve of the combined approach: 1) when using two PCs, the accuracy is low, worse than Short TE only approach; 2) when using from 3 to 6 PCs, the behavior seems to follow the Long TE model, but with better accuracy due to the contribution of the Short TE; 3) when using from 7 to 15 PCs, the behavior is increased significatively one more time, just after the increasement of the accuracy of the Short TE approach due to the incorporation of 6th and following components to the model. As a general approximation, we can say that the main discriminative contribution of the Long TE is accumulated in the first principal components, meanwhile the main discriminative contribution of the Short TE is accumulated in posterior principal components.

The models integrating both spectra show the best behavior, this result points out to the more discriminant information is obtained when both spectra are globally analyzed at the same time for classifier development. The inclusion of more features in the models are studied with LS-SVMs in the binary classifiers because of their appropriate treatment of high-dimensional spaces.

Experiments with Binary classifiers. The regularization property of LS-SVM allows to increase the number of features to be used in the classifiers, and then to characterize the results obtained with the multiclass classifiers, a) Which approach (Short TE, Long TE or both spectra) is the best in the pairwise classification of the three superclasses? and b) Confirm that the models with both spectra require the incorporation of enough information from the Long TE spectrum to improve the model.
x1

Table 4 shows the results for the AGG vs. MEN problem using LS-SVM with and without Relief-F feature selection where Long TE + Short TE, Short TE and Long TE datasets were used.

Table 4: LS-SVM for classifying AGG vs. MEN, 150-Random Sampling Train-Test. The percentage of Long TE features with respect to the total is in brackets.

Dataset	Features	Accuracy [CI]	AUC
Long TE + Short TE	10 (Long:0%)	89.97 [84.58,93.96]	0.9541
	50 (Long:17%)	89.22 [83.68,93.38]	0.9597
	100 (Long:20%)	92.18 [87.25,95.63]	0.9794
	380 (Long:50%)	95.27 [91.16,97.79]	0.9921
Short TE	10	90.86 [85.64,94.64]	0.9552
	50	89.78 [84.35,93.81]	0.9646
	100	92.62 [87.79,95.95]	0.9824
	190	92.46 [87.59,95.84]	0.9801
Long TE	10	85.50 [79.38,90.38]	0.9066
	50	91.73 [86.70,95.30]	0.9625
	100	92.15 [87.21,95.61]	0.9735
	190	92.22 [87.30,95.66]	0.9750

In the AGG vs. MEN classifier, the Long TE + Short TE model using 10 features is not using any Long TE features in any of the 150 repetitions. This may be due

to the more individual discrimination power of the Short TE features, since they are selected first by the feature selection algorithms. The accuracy of the model with 50 features decreases slightly compared to 10-features model. Taking into account the behavior of Short TE and Long TE models with respect to the number of features, the decrease must be due to the worse contribution of 50 Short TE features with respect to the 10 first features. In any case, the AUC-ROC of the Long TE + Short TE models increase when the number of features in the models increases. When 100 features are used, the Long TE + Short TE model improvement is considerable with respect to the 10-features and 50-features models, but not better than the 100-Short TE model. The improvement, and best results in accuracy and AUC-ROC are obtained when both full spectra are used. This result is better than the Long TE only model and Short TE only model with full spectra.

Table 5 shows the results for the AGG vs. LGG problem using LS-SVM with and without Relief-F feature selection where Long TE + Short TE, Short TE and Long TE datasets are used.

Table 5: LS-SVM for classifying between AGG vs. LGG, 150-Random Sampling Train-Test. The percentage of Long TE features with respect to the total is in brakets.

Dataset	Features	Accuracy [CI]	AUC
Long TE + Short TE	10 (Long:0%)	92.58 [87.34,96.11]	0.9698
	50 (Long:6%)	91.29 [85.75,95.17]	0.9629
	100 (Long:22%)	91.38 [85.86,95.24]	0.9694
	380 (Long:50%)	91.07 [85.48,95.01]	0.9699
Short TE	10	92.07 [86.71,95.74]	0.9660
	50	91.88 [86.47,95.60]	0.9621
	100	92.87 [87.71,96.31]	0.9722
	190	91.60 [86.13,95.40]	0.9616
Long TE	10	90.49 [84.77,94.58]	0.9541
	50	88.71 [82.64,93.22]	0.9405
	100	88.10 [81.92,92.75]	0.9446
	190	88.41 [82.29,92.98]	0.9472

The best result in the AGG vs. LGG corresponds to the Long TE + Short TE model with 10 features. In fact, this model is not using any feature from the Long TE signal, and is hence equivalent to the Short TE model. When more features are included, accuracy does not increase and the percentage of Long TE features is still low compared to the percentages of Long TE features obtained for the other tasks (AGG vs. MEN and LGG vs. MEN). Long TE models do not improve when more features are included. The behavior of the Long TE + Short TE model is very close to the Short TE model, because the features from Short TE are the main contributors to the combined model.

Table 6 shows the results for the LGG vs. MEN problem using LS-SVM with and without Relief-F feature selection where Long TE + Short TE, Short TE and Long TE datasets were used.

All models for task LGG vs. MEN produced similar good behavior. Long TE features were used in the Long TE + Short TE model in all instances, and with large

Table 6: LS-SVM for classifying between LGG vs. MEN, 150-Random Sampling Train-Test. The percentage of Long TE features with respect to the total is in brackets.

Spectra	Features	Accuracy [CI]	AUC
Long TE + Short TE	10 (Long:19%)	95.90 [90.10,98.64]	0.9753
	50 (Long:42%)	97.53 [92.59,99.29]	0.9960
	100 (Long:48%)	96.81 [91.46,99.03]	0.9975
	380 (Long:50%)	95.75 [89.87,98.58]	0.9964
Short TE	10	95.93 [90.13,98.66]	0.9822
	50	96.02 [90.27,98.70]	0.9927
	100	95.75 [89.87,98.58]	0.9898
	190	94.49 [88.08,97.93]	0.9864
Long TE	10	92.49 [85.38,96.76]	0.9760
	50	93.93 [87.31,97.62]	0.9907
	100	94.54 [88.15,97.96]	0.9925
	190	93.48 [86.70,97.36]	0.9913

percentages in 50- and 100-feature models, where improvements were obtained with respect to the Short TE only and Long TE only models. Both Short TE only and Long TE only models produced good results individually, then the results obtained by the combination classifier were to be expected.

Significant difference results are summarized in Table 7. The general comparison of Long TE + Short TE, Short TE and Long TE approaches reveal significant differences among them. In the pair-wise comparison, the Long TE + Short TE and Long TE are significantly different. Long TE + Short TE and Short TE are not significantly different from each other and Short TE and Long TE models are not either significantly different.

When comparing the best models of the Aggressive vs. Meningioma (AGG vs. MEN) binary classifiers, the Short TE + Long TE model outperforms the Short TE only and Long TE only approaches, however, the differences in performance with respect to the Short TE only and Long TE only approaches are not statistically different.

When comparing the best models of the Agressive vs. Low Grade Glial (AGG vs. LGG) binary classifiers, the Short TE + Long TE model did not outperform the Short TE approach, as we expected by the feature selection analysis. Significant differences in performance between the Short TE only and the Long TE only were obtained.

When comparing the best models of the Low Grade Glial vs. Meningioma (LGG vs. MEN) binary classifiers, the Short TE + Long TE model outperforms the approaches using Short TE only or Long TE spectra only approaches. Significant differences in performance were also obtained when comparing the Short TE only and the Long TE only approaches.

4 Discussion

This study focuses on the automatic discrimination of the most prevalent brain tumors in humans [1]. Metastatic brain tumors have the greatest incidence rate, with breast, lung and melanoma being the most common cancers to metastasize to the brain.

Table 7: Summary of the significance of difference tests among the approaches. Friedman's (F) nonparametric two-way analysis of variance test, Kruskal-Wallis (KW) nonparametric one-way analysis of variance and Tukey's (MCT) honestly significant difference criterion were used.

	Comparison of Long TE + Short TE vs. Short TE vs. Long TE models			
	General	AGG vs. MEN	AGG vs. LGG	LGG vs. MEN
Significance comparison	Yes (F) $p < 0.05$	Yes (KW) $p < 0.05$	Yes (KW) $p < 0.05$	Yes (KW) $p < 0.05$
Long TE + Short TE vs. Short TE	No (MCT)	Yes (MCT)	No (MCT)	Yes (MCT)
Long TE + Short TE vs. Long TE	Yes (MCT)	Yes (MCT)	Yes (MCT)	Yes (MCT)
Short TE vs. Long TE	No (MCT)	No (MCT)	Yes (MCT)	Yes (MCT)

Meningiomas are the most common type of primary brain tumors at 27.4%. Glioblastoma Multiforme is the most common type of Glioma at 23.0 %. Astrocytomas account for 11.3 % and Oligodendrogliomas for 4.0 % of all tumors.

There were significant differences in classifier performance using the different subsets (Long TE, Short TE or both Long TE-Short TE). For discriminating between MEN and the other two super-classes, both Short TE and Long TE are needed to obtain the best result. On the other hand, if we aim to distinguish between AGG and LGG, acquisition of Short TE should be enough. LS-SVMs were useful when combining Short TE and Long TE features; this may be due to its regularized behaviour. These results are consistent with the clinical study described in Majos et al. [26] to evaluate the improvement of the combination of Short TE and Long TE spectra for clinical classification of brain tumors. In their study, tumor classification was slightly better at Short TE (123 [81%] of 151 cases correctly classified) than at Long TE (118 [78%] of 151 cases correctly classified). Meningioma was the only group that showed higher sensitivity and specificity at Long TE. They obtained better results when both TE were considered simultaneously: the suggested diagnosis was correct in 105 (94%) of 112 cases when both TE agreed, whereas the correct diagnosis was suggested by at least one TE in 136 (90%) of 151 cases. Also, a descriptive study was carried out by Tortajada et al [47] where the discrimination of the INTERPRET database using LDA and stepwise feature selection methods was improved using the combined approach.

The rest of this section summarizes the analysis of the features selected during the experiments, their biological interpretation and an overview on the outliers. Additionally, possible result visualization strategies are considered.

The selected features are positioned on the in-vivo average-class spectra to validate their biological origin. Outliers of the Long TE + Short TE, Short TE and Long TE models are also compared in order to study the complementary behavior of the Short TE and Long TE models.

Table 8 lists the most selected features in the 150 repetitions of the stepwise algorithm used to train the multiclass classifier. The number of occurrences are shown in brackets. Most of the selected features in the Long TE + Short TE model were also selected by the Short TE only and Long TE only models.

Figure 4 shows the average spectra per diagnosis and the most selected features

Table 8: Most repeated features in the k-Random Sampling Train Test evaluation. Classification among Aggressive tumors -Glioblastomas and Metastasis- (AGG), Meningiomas (MEN) and a Low Grade Glial mixture of Astrocytomas grade II, Oligodendrogliomas and Oligoastrocytomas (LGG).

Dataset	Top variables (var(n)*)
Long TE + Short TE	STE3.03(49), STE2.32(43), STE2.29(43), STE3.57(41), STE2.31(40), STE1.27(34), STE1.23(32) STE2.34(28), STE1.25(28), STE3.05(25), STE3.02(25), LTE3.03(23), LTE3.19(23), LTE2.42(21), STE3.76(21) LTE2.84(21), LTE2.5(20), LTE1.54(20), LTE2.96(20), STE3.78(19), STE3.11(19), STE3.42(18), LTE2.46(17)
Short TE	STE3.03(56), STE2.32(53), STE3.05(42), STE2.29(38), STE1.23(34), STE1.25(32), STE3.57(28) STE3.02(25), STE3.78(24), STE2.31(23), STE2.34(22), STE1.27(22), STE3.55(18), STE3.76(16)
Long TE	LTE3.03(131), LTE2.13(82), LTE2.31(75), LTE1.19(46), LTE3(40), LTE3.78(39), LTE2.11(35), LTE1.15(34) LTE2.32(31), LTE2.09(30), LTE3.76(29), LTE2.27(29), LTE2.52(27), LTE2.5(26), LTE3.34(25), LTE2.19(23) LTE2.25(22), LTE3.74(21), LTE3.02(20), LTE1.54(20)

corresponding to Table 8 to relate them with the biological origin of the selected features.

The incorporation of the predictor in a Decision Support System (DSS) could be of interest for clinicians to decide whether the tested case is routine or a special one. An interesting visualization of the "universe of cases" is made from the latent space provided by the LDA multiclass classifier. A visual representation of the database for descriptive purposes was generated by means of the 2D projection made by the LDA method. Figure 5 shows the LDA latent space of the best classifiers obtained by the three approaches in the k-random sampling train-test procedure. Therefore, the projection of the best model obtained by the Long TE + Short TE approach provides visually better discrimination than Short TE alone or Long TE alone best models.

Table 9 includes each case considered an outlier by one approach at least. A case was considered outlier when less than 20% of successes were obtained when testing it. The complementary point of view provided by the Short TE models and Long TE models is observed in nine of the thirteen outliers, on the contrary, Long TE + Short TE model is following the Short TE behavior.

Table 9: Outliers in the multi-class classification with stepwise selection.

	Succeses/Occurrences (%)		
Sample code [13, 28]	Long TE + Short TE	Short TE	Long TE
I0009	32.0	25.5	0.0
I0057	100	100	9.1
I0063	6.1	0.0	34.9
I0068	44.7	65.8	8.3
I0080	78.7	85.4	7.8
I0083	97.7	93.0	0.0
I0086	13.04	3.9	63.3
I0332	26.09	4.2	43.5
I0359	100	100	14.6
I0450	0.0	0.0	89.4
I1083	0.0	0.0	0.0
I1275	51.1	48.1	7.9
I1390	0.05	5.6	2.1

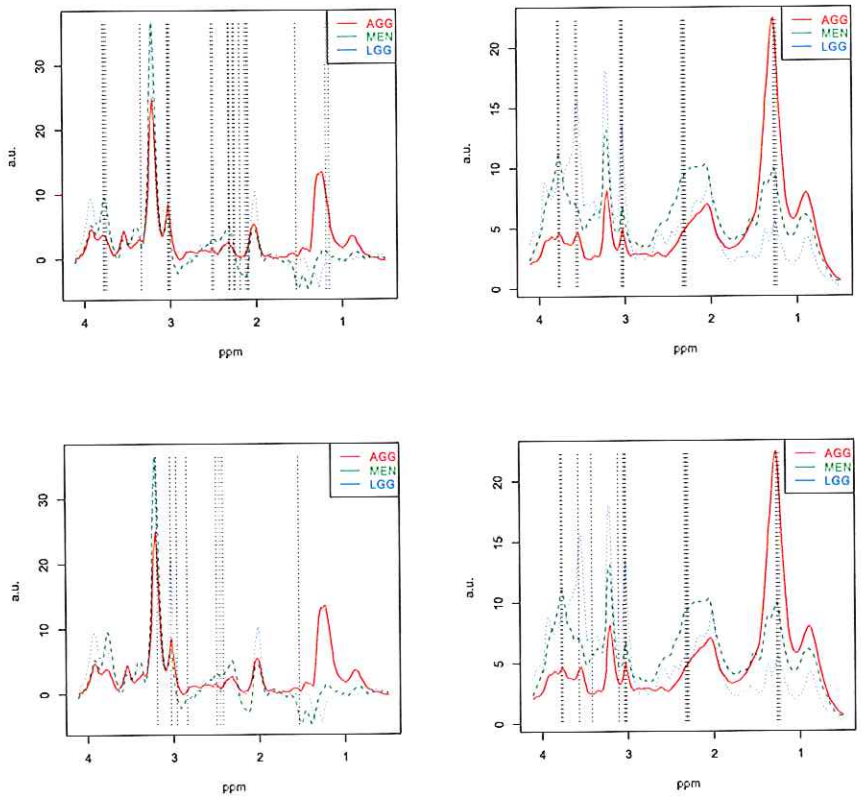

Figure 4: In-vivo Long TE (left) and Short TE (right) average spectra [7] for Aggressive (solid red line), Low-glial (dotted blue line) and Meningioma (dashed green line) diagnosis. The first row shows the selected points by the combined model as vertical black dotted lines and the second row shows, in the same way, the selected points by the Long TE model and Short TE model.

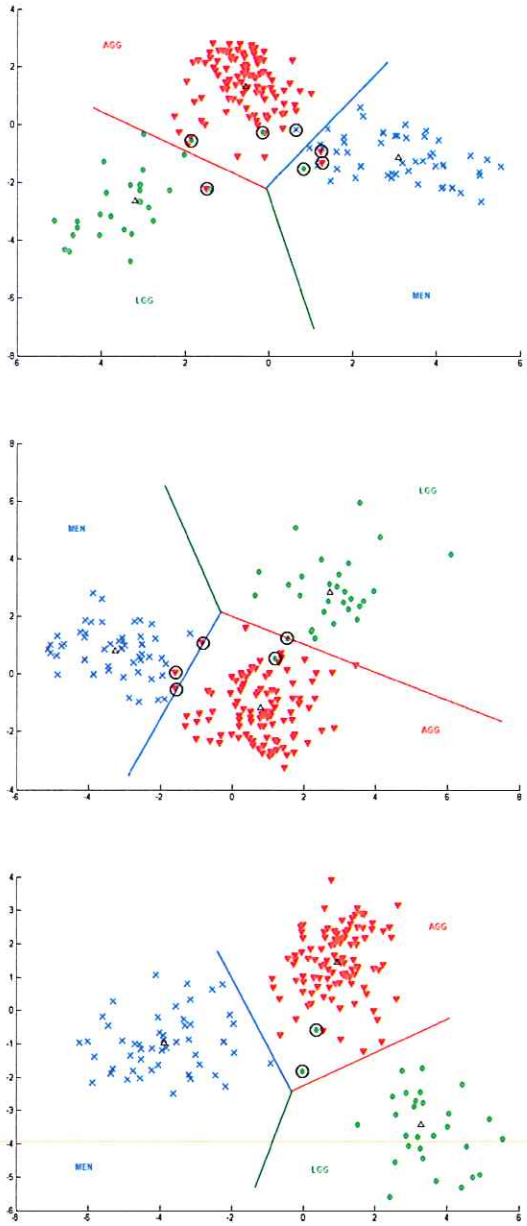

Figure 5: Comparative projections of the LDA latent space (or "universe of cases") for the best model of Short TE (top), Long TE (middle) and the combined approach (bottom). Every symbol is representing a tumor sample projected in the latent space. In each projection, the space is divided by the decision boundary into three different regions, corresponding to each brain tumor class: MEN (blue crosses), AGG (red inverted triangles) and LGG (green circles). Any misclassified sample is marked with a block cicle. Therefore, the projection of the best model obtained by the Long TE + Short TE approach provides visually better discrimination than Short TE alone or Long TE alone best models.

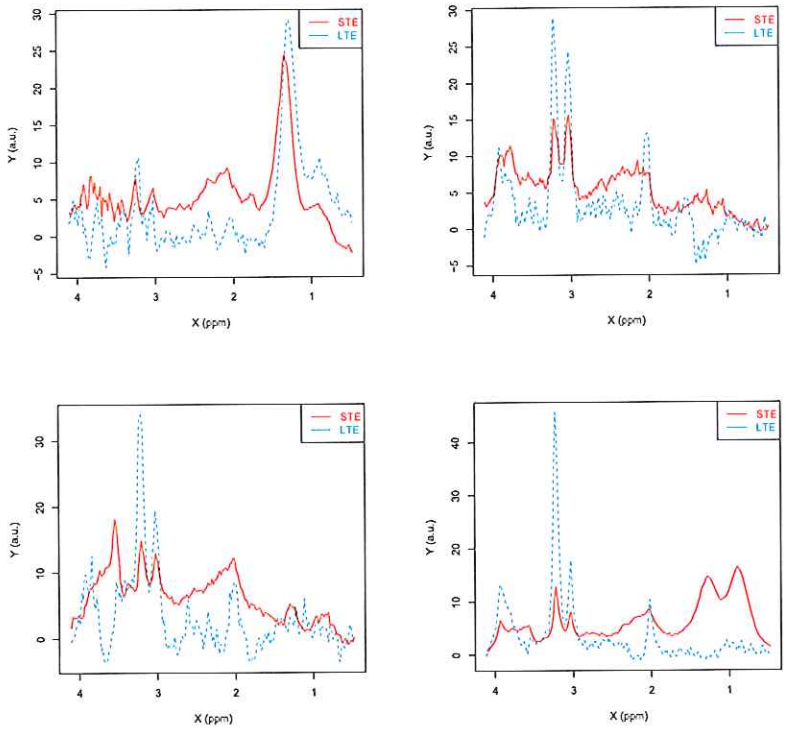

Figure 6: Four outliers analyzed after the experiments. Top-left (I0009) and top-right (I1390) are meningiomas, bottom-left (I0063) is a glioblastoma and bottom-right (I0450) is an oligoastrocytoma. Short TE is presented by solid red lines and Long TE by dash blue lines

Further analysis of the misclassified cases showed that, although validated as usable MEN during INTERPRET, there was one MEN case (see Figure 6, top-left) that showed an abnormal profile for a low-grade meningioma with a clear pattern of lipids at 1.3 ppm and 0.9 ppm. This case was a secretory meningioma and also the representative MRI with the voxel location showed that there was the possibility of contamination by subcutaneous fat.

There was another meningioma that was misclassified (see Figure 6, top-right), where normal tissue contribution can be contaminating the profile, problems appeared in the phasing of the Short TE and an atypical profile in Long TE. There was also one glioblastoma- an AGG tumor (see Figure 6, bottom-left) that did not show the typical necrotic pattern with clear lipid signals at 0.9 and 1.3 ppm that is expected for most of these type of tumours, but instead a distinctive m-Inositol peak at 3.55 ppm constitutes the strongest signal at Short TE and disappears at Long TE confirming the contribution of m-Inositol to this resonance. The spectral pattern of this case closely resembles an average pattern of a LGG tumour and other authors [48] have pointed the possibility of these pattern sub-type belonging to the less common secondary glioblatoma.

There was, also, an oligoastrocytoma (see Figure 6, bottom-right) that shows a clear signal from mobile lipids at Short TE which disappears at Long TE. Note that the 0.9 ppm peak has an equal intensity than the 1.3 ppm peak, which is fairly unusual. This mobile lipid profile observed at Short TE, disappeared at Long TE, Short TE and Long TE + Short TE models miss the cases following the pseudo-necrotic mobile lipid profile observed at some non-aggressive Short TE spectra.

There are still problems in brain tumor classification by MRS. One of the most important ones being the limited number of available spectra per type of tumor [1, 4], especially in tumors of low prevalence, as each tumour type has its own epidemiological distribution [49]. The inclusion of more Low Grade Glial tumors could make it possible to carry out studies in the discrimination among astrocytomas , oligodendrogliomas and oligoastrocytomas. The presence of noise and artifacts in the spectra, even after elimination of the dominating water peak affects the information contain in the spectra. The European projects eTumour [30] and HealthAgents [50] are trying to help in the compilation and analysis of additional brain tumors patterns using in-vivo and ex-vivo data.

5 Conclusions

This study has focused on the use of Short TE and Long TE MRS to improve the automatic brain tumor diagnosis. Complementary information is obtained by both spectra and it can be used to improve the automatic classification of brain tumors following a PR approach. The Combination of Short TE and Long TE in a single classifier performs better in several instances that each echo time data alone.

When models are developed using feature selection from combination of Short TE and Long TE signals, the main source of information in small dimensional models (< 10) comes from Short TE features. The inclusion of Long TE information as PCs or more features in the model improves the accuracy and the AUC of the models with respect to those developed with a lower number of features. These models are better than the models of the same dimensionality that use only Short TE or only Long TE information. There were significant differences in classifier performance using the different subsets (Long TE, Short TE or both Long TE and Short TE). For discriminating MEN from the other two super-classes, both Short TE and Long TE were required to obtain the best results. On the other hand, for distinguishing between AGG and LGG, acquisition of Short TE should be enough. LS-SVMs were useful when combining Short TE and Long TE features; this may be due to its regularized behavior. Results of binary classifiers using LS-SVM suggest that development of multiclass classifiers based on the binary classifiers using Short TE and Long TE features may be of interest.

6 Acknowledgements

This work was partially funded by the European Commission (FP6-2002- LIFESCI-HEALTH 503094 and IST-2004-27214). We thank INTERPRET partners for providing data, in particular C. Majós (IDI-Bellvitge), A. Moreno (Centre Diagnostic Pedralbes), John Griffiths (SGUL), Arend Heerschap (RU), Witold Gajewicz (MUL), and Jorge Calvar (FLENI). We also thank Guillem Mercadal for his feedbak about MRS data processing.

References

[1] BTS: The Brain Tumor Society. (Accessed: 28 January 2007) http://www.tbts.org.

[2] Lin, A.P., Tran, T.T., Ross, B.D.: Impact of evidence-based medicine on magnetic resonance spectroscopy. NMR Biomed **19**(4) (2006) 476–483

[3] Howe, F.A., Opstad, K.S.: 1H MR spectroscopy of brain tumours and masses. NMR Biomed **16**(3) (2003) 123–131

[4] Julia-Sape, M., Acosta, D., Majos, C., Moreno-Torres, A., Wesseling, P., Acebes, J.J., Griffiths, J.R., Arús, C.: Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. J Neurosurg **105**(1) (2006) 6–14 Comparative Study.

[5] Lukas, L., Devos, A., Suykens, J.A.K., Vanhamme, L., Howe, F.A., Majós, C., Moreno-Torres, A., Graaf, M.V.D., Tate, A.R., Arús, C., Huffel, S.V.: Brain tumor classification based on long echo proton MRS signals. Artificial Intelligence in Medicine **31** (2004) 73–89

[6] Galanaud, D., Nicoli, F., Chinot, O., Confort-Gouny, S., Figarella-Branger, D., Roche, P., Fuentes, S., Le Fur, Y., Ranjeva, J.P., Cozzone, P.J.: Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy. Magn Reson Med **55**(6) (2006) 1236–1245

[7] Tate, A.R., Underwood, J., Acosta, D.M., Julia-Sape, M., Majos, C., Moreno-Torres, A., Howe, F.A., van der Graaf, M., Lefournier, V., Murphy, M.M., Loosemore, A., Ladroue, C., Wesseling, P., Luc Bosson, J., Cabanas, M.E., Simonetti, A.W., Gajewicz, W., Calvar, J., Capdevila, A., Wilkins, P.R., Bell, B.A., Remy, C., Heerschap, A., Watson, D., Griffiths, J.R., Arús, C.: Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. NMR Biomed **19**(4) (2006) 411–434

[8] Barba, I., Moreno, A., Martinez-Perez, I., Tate, A.R., Cabanas, M.E., Baquero, M., Capdevila, A., Arús, C.: Magnetic resonance spectroscopy of brain hemangiopericytomas: high myoinositol concentrations and discrimination from meningiomas. J Neurosurg **94**(1) (2001) 55–60 Comparative Study.

[9] Gillies, R.J., Morse, D.L.: In vivo magnetic resonance spectroscopy in cancer. Annu Rev Biomed Eng **7** (2005) 287–326

[10] Lisboa, P.J., Taktak, A.F.G.: The use of artificial neural networks in decision support in cancer: a systematic review. Neural Netw. **19**(4) (2006) 408–415

[11] Hagberg, G.: From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. NMR Biomed **11**(4-5) (1998) 148–156

[12] Ripley, B.D.: Spatial statistics. Hoboken, NJ : Wiley-Interscience (2004)

[13] INTERPRET Consortium: Interpret web site. (Accessed: 27 January 2007) http://azizu.uab.es/INTERPRET/.

[14] Underwood, J., Tate, A.R., Luckin, R., Majós, C., Capdevila, A., Howe, F., Griffiths, J., Arús, C.: A Prototype Decision Support System for MR Spectroscopy-Assisted Diagnosis of Brain Tumours. In et al., V.P., ed.: MEDINFO 2001, IMIA (2001) 561–565

[15] Tate, A.R., Majos, C., Moreno, A., Howe, F.A., Griffiths, J.R., Arús, C.: Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. Magn Reson Med **49**(1) (2003) 29–36

[16] Simonetti, A.W., Melssen, W.J., Szabo de Edelenyi, F., van Asten, J.J.A., Heerschap, A., Buydens, L.M.C.: Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. NMR Biomed **18**(1) (2005) 34–43 Clinical Trial.

[17] Devos, A., Lukas, L., Suykens, J.A.K., Vanhamme, L., Tate, A.R., Howe, F.A., Majos, C., Moreno-Torres, A., van der Graaf, M., Arús, C., Van Huffel, S.: Classification of brain tumours using short echo time 1H MR spectra. J Magn Reson **170**(1) (2004) 164–175

[18] Simonetti, A.W., Melssen, W.J., van der Graaf, M., Postma, G.J., Heerschap, A., Buydens, L.M.C.: A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy. Anal Chem **75**(20) (2003) 5352–5361

[19] Devos, A., Simonetti, A.W., van der Graaf, M., Lukas, L., Suykens, J.A.K., Vanhamme, L., Buydens, L.M.C., Heerschap, A., Van Huffel, S.: The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. J Magn Reson **173**(2) (2005) 218–228

[20] de Edelenyi, F.S., Simonetti, A., Postma, G., Huo, R., Buydens, L.M.C.: (application of independent component analysis to 1h mr spectroscopic imaging exams of brain tumors). Analytica Chimica Acta (2005)

[21] Ma, J., Sun, Z.: Mrs classification based on independent component analysis and support vector machines. his **0** (2005) 509–511

[22] Huang, Y., Lisboa, P.J.G., El-Deredy, W.: Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection. Stat Med **22**(1) (2003) 147–164

[23] Menze, B.H., Lichy, M.P., Bachert, P., Kelm, B.M., Schlemmer, H.P., Hamprecht, F.A.: Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors. NMR Biomed **19**(5) (2006) 599–609

[24] Suykens, J.A.K., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. **9**(3) (1999) 293–300

[25] Luts, J., Heerschap, A., Suykens, J., Huffel, S.V.: A combined mri and mrsi based multiclass system for brain tumour recognition using ls-svms with class probabilities and feature selection. Technical report, Katholieke Universiteit Leuven, Department of Electrical Engineering, ESAT-SCD (SISTA), http://www.esat.kuleuven.ac.be/ (2006)

[26] Majos, C., Julia-Sape, M., Alonso, J., Serrallonga, M., Aguilera, C., Acebes, J.J., Arús, C., Gili, J.: Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE. AJNR Am J Neuroradiol **25**(10) (2004) 1696–1704

[27] Kleihues, P., Burger, P.C., Scheithauer, B.W.: The new WHO classification of brain tumours. Brain Pathol **3**(3) (1993) 255–268

[28] Julia-Sape, M., Acosta, D., Mier, M., Arús, C., Watson, D.: A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. Magn Reson Mater Phy **19**(1) (2006) 22–33

[29] INTERPRET Consortium: Interpret adquisition protocols. (2000) http://azizu.uab.es/INTERPRET/mrsdata/mrsdata.html.

[30] eTumour Consortium: etumour: Web accessible mr decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data. Technical report, FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC, http://www.etumour.net, Accessed: 28 January 2007 (2003)

[31] Naressi, A., Couturier, C., Castang, I., de Beer, R., Graveron-Demilly, D.: Java-based graphical user interface for MRUI, a software package for quantitation of in vivo/medical magnetic resonance spectroscopy signals. Comput Biol Med **31**(4) (2001) 269–286

[32] Buydens, L.: Interpret project deliverable no. 7. d 3.1: Pr strategy. Technical report, Laboratory for Analytical Chemistry. Nijmegen University (2001)

[33] Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artificial Intelligence **97**(1-2) (1997) 273–324

[34] Burges, C.J.: Geometric methods for feature extraction and dimensional reduction: A guided tour. Technical report, Microsoft Research, University of Toronto (2004)

[35] Hastie, T., Tibshirani, R., Friedman, J.H.: The Elements of Statistical Learning. Springer (2001)

[36] Vapnik, V.: The Nature of Statistical Learning Theory. Springer, N.Y. (1995)

[37] Russell, S.J., Norvig, P.: Artificial intelligence: a modern approach. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1995)

[38] Roever, C., Raabe, N., Luebke, K., Ligges, U., Szepannek, G., Zentgraf, M.: The klaR Package. Deparment of Statistics, University of Dortmund. (2006)

[39] Garczarek, U.M.: CLASSIFICATION RULES IN STANDARDIZED PARTITION SPACES. PhD thesis, UNIVERSITY OF DORTMUND, http://deposit.ddb.de (2002)

[40] Robnikikonja, M., Kononenko, I.: Theoretical and empirical analysis of relieff and rrelieff. Mach. Learn. **53**(1-2) (2003) 23–69

[41] Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI. (1995) 1137–1145

[42] Martin, J.K., Hirschberg, D.S.: Small sample statistics for classification error rates I: Error rate measurements. Technical Report ICS-TR-96-22 (1996)

[43] Guyon, I., Makhoul, J., Schwartz, R., Vapnik, V.: What size test set gives good error rate estimates? IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(1) (1998) 52–64

[44] Ambroise, C., McLachlan, G.J.: Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A **99**(10) (2002) 6562–6566

[45] Begg, C.B.: Statistical methods in medical diagnosis. Crit Rev Med Inform **1**(1) (1986) 1–22

[46] Berrar, D., Bradbury, I., Dubitzky, W.: Avoiding model selection bias in small-sample genomic datasets. Bioinformatics (2006)

[47] Tortajada, S., García-Gómez, J.M., Vidal, C., Arús, C., Julià-Sapé, M., Moreno, A., Robles, M.: Improved classification by pattern recognition of brain tumours combining long and short echo time 1h-mr spectra. In ESMRMB, ed.: 23rd annual meeting of European Society of Magnetic Resonance for Medicine and Biology. (2006)

[48] Martinez-Bisbal, M.C., Marti-Bonmati, L., Piquer, J., Revert, A., Ferrer, P., Llacer, J.L., Piotto, M., Assemat, O., Celda, B.: 1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas. NMR Biomed **17**(4) (2004) 191–205 Comparative Study.

[49] Kleihues, P., Cavanee, W., eds.: Pathology and Genetics of Tumours of the Nervous System. International Agency for Research on Cancer (1997)

[50] Arús, C., Celda, B., Dasmahapatra, S., Dupplaw, D., González-Vélez, H., van Huffel, S., Lewis, P., Lluch i Ariet, M., Mier, M., Peet, A., Robles, M.: On the design of a web-based decision support system for brain tumour diagnosis using distributed agents. In: 2006 IEEE/WIC/ACM Int Conf on Web Intelligence & Intelligent Agent Technology (WI-IAT 2006 Workshops), Hong Kong (2006) 208–211