Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 170 (2004) 164–175
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Classification of brain tumours using short echo time ^{1}H MR spectra

A. Devos[a,*], L. Lukas[a], J.A.K. Suykens[a], L. Vanhamme[a], A.R. Tate[b], F.A. Howe[c], C. Majós[d], A. Moreno-Torres[e], M. van der Graaf[f], C. Arús[g], S. Van Huffel[a]

[a] *SCD-SISTA, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Heverlee (Leuven), Belgium*
[b] *Department of Paediatric Epidemiology and Biostatistics, Institute of Child Health, London WC1N 1EH, UK*
[c] *CRUK Biomedical Magnetic Resonance Research Group, Department of Biochemistry and Immunology, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK*
[d] *Institut de Diagnòstic per la Imatge (IDI), CSU de Bellvitge, Autovia de Castelldefels km 2.7, L'Hospitalet de Llobregat, 08907 Barcelona, Spain*
[e] *Centre Diagnòstic Pedralbes, Unitat Esplugues, C/ Josep Anselm Clavé, 100, 08950 Esplugues de Llobregat, Spain*
[f] *Department of Radiology, University Medical Center Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands*
[g] *Departament de Bioquímica i Biologia Molecular, Unitat de Ciències, Edifici Cs, Universitat Autonòma de Barcelona, 08193 Cerdanyola del Vallès, Spain*

Received 1 December 2003; revised 26 May 2004
Available online 20 July 2004

Abstract

The purpose was to objectively compare the application of several techniques and the use of several input features for brain tumour classification using Magnetic Resonance Spectroscopy (MRS). Short echo time ^{1}H MRS signals from patients with glioblastomas ($n = 87$), meningiomas ($n = 57$), metastases ($n = 39$), and astrocytomas grade II ($n = 22$) were provided by six centres in the European Union funded INTERPRET project. Linear discriminant analysis, least squares support vector machines (LS-SVM) with a linear kernel and LS-SVM with radial basis function kernel were applied and evaluated over 100 stratified random splittings of the dataset into training and test sets. The area under the receiver operating characteristic curve (AUC) was used to measure the performance of binary classifiers, while the percentage of correct classifications was used to evaluate the multiclass classifiers. The influence of several factors on the classification performance has been tested: L2- vs. water normalization, magnitude vs. real spectra and baseline correction. The effect of input feature reduction was also investigated by using only the selected frequency regions containing the most discriminatory information, and peak integrated values. Using L2-normalized complete spectra the automated binary classifiers reached a mean test AUC of more than 0.95, except for glioblastomas vs. metastases. Similar results were obtained for all classification techniques and input features except for water normalized spectra, where classification performance was lower. This indicates that data acquisition and processing can be simplified for classification purposes, excluding the need for separate water signal acquisition, baseline correction or phasing.

Keywords: Brain tumour classification; Short echo time MRS; Linear discriminant analysis; Least squares support vector machines

1. Introduction

In vivo magnetic resonance spectroscopy (MRS) is a noninvasive technique which provides chemical information of metabolites present in living tissue and can be used to help characterize human brain tumours [1–3]. A histopathological analysis of a biopsy is the present gold standard for diagnosis of an abnormal brain mass suspected of being a brain tumour. A biopsy is not without risk of morbidity and mortality and cannot be carried out in all instances (e.g., brain stem tumours, paediatric tumours). Additionally, there are inherent inaccuracies in the gold standard [4] which can lead to

* Corresponding author. Fax: +32-16-321970.
E-mail address: adevos@esat.kuleuven.ac.be (A. Devos).

1090-7807/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2004.06.010

misclassification or imprecision in establishing the final diagnosis. MRS has the potential to improve the diagnosis of brain tumours, with no additional risk to the patient.

Several studies [5–17] have already shown progress in automated pattern recognition for brain tumour classification based on MRS data. Several partners from the EU funded INTERPRET project (IST-1999-10310) [18], who provided the data for this study, have already published promising results for classification of brain tumours based on MRS data available within the project [7,12,15,16,19–24]. In [16], for example, Tate et al. used linear discriminant analysis to classify 144 short echo time spectra from three contributing centres, considering three groups of brain tumours: meningiomas, low-grade astrocytomas, and aggressive tumours (glioblastomas and metastases combined). Tate et al. selected specific training and test sets; forming the training set from two centres (94 spectra) and taking the third centre as test set (50 spectra). Based on this setting a classification accuracy of 96% was achieved.

However, most of the INTERPRET studies were based on a previous version of the dataset, did not use receiver operating characteristic (ROC) analysis or focused on a specific technique. The present work follows a distinct approach in several aspects with respect to previous studies. Several well-established classification and feature selection techniques are applied to the multicentre dataset of short echo time ^{1}H MRS signals. The spectra were acquired on scanners from GE, Philips, and Siemens, the three leading manufacturers of these MR scanners, using the major acquisition sequences PRESS and STEAM.

Linear and nonlinear classification techniques are objectively compared for binary and multiclass classification. The algorithms are designed to automatically extract the most important features which are then used to classify each spectrum according to its corresponding tumour type. By applying the techniques to several input features, the influence of normalization methods, baseline correction, phasing, and dimensionality reduction of the input data is tested.

Binary classification performance is measured based on the receiver operating characteristic (ROC) curve analysis over 100 stratified random samplings of training and test set. ROC analysis is commonly used in medicine [25–27] to objectively judge the discrimination ability of various statistical methods for predictive purposes, which can be measured by the area under the ROC curve (AUC). The AUC gives then a global measure of the clinical efficiency over a range of test cut-off points on the ROC curve. This is in contrast to performance measures such as the accuracy, which is only based on a single cut-off point (e.g., for one specific value of the false-positive rate).

In our previous study on long echo time magnitude spectra [7], automated binary classifiers reached a mean AUC of more than 0.90 except for the most difficult binary classification to discriminate glioblastomas from metastases. In comparison with long echo time spectra, short echo time spectra are more difficult to analyze due to a higher number of overlapping peaks, a more prominent baseline and a higher sensitivity to artefacts. However, short echo time spectra are richer in information than long echo time spectra, as several resonances diminish in apparent intensity at longer echo times [28,29], because of a small T_2 value or signal cancellation due to *J*-modulation [30,31]. For example, resonances of *mI* (myo-Inositol,triplets and multiplet at 3.26 and 3.57 ppm), *Glu* (glutamate, multiplets at 2.33 and 3.74 ppm), *Gln* (glutamine, multiplets at 2.43 and 3.75 ppm) are much less pronounced in spectra at longer echo times. Here, we report our results of an extended study based on short echo time spectra, similar to [7]. We investigate whether the applied pattern recognition techniques are able to exploit the large amount of information available in short echo time MR spectra, despite the problems in analyzing these data, to discriminate several types of brain tumours.

2. Materials

The short echo time INTERPRET database contains single voxel ^{1}H MR spectra from six centres: Centre Diagnòstic Pedralbes (CDP), Barcelona (Spain); Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia (FLENI), Buenos Aires (Argentina); Institut de Diagnòstic per la Imatge (IDI), Barcelona (Spain); Uniwersytet Medyczny w Łodzi (MUL), Łodzi (Poland); St. George's Hospital Medical School (SGHMS), London (United Kingdom), and Universitair Medisch Centrum Nijmegen (UMCN), Nijmegen (The Netherlands). Table 1 reports the most important acquisition parameters. Quality control criteria have been applied to MR system performance and data from all centres [32]. Data were acquired at 1.5 T either by a STEAM- or PRESS-sequence, at an echo time (TE) between 20 and 32 ms. Note that TE can differ somewhat from one centre to another, or even between data from the same centre, due to variations in the optimized acquisition protocols supplied by the manufacturers. This range of TE will mostly affect the appearance of signals from metabolites with coupled spins (e.g., glutamate/glutamine (*Glx*), lactate (*Lac*), and alanine (*Ala*)) and the intensity of signals from macromolecules which have a short T_2. The resonance pattern of signals from coupled spins is also affected by the choice of the acquisition sequence used (PRESS or STEAM). Nevertheless, Tate et al. [16] concluded that pattern recognition is less sensitive to the effects of differences in acquisition parameters than had been expected.

Four common brain tumour types are considered in our study: glioblastomas, meningiomas, metastases,

Table 1
Overview of the acquisition parameters of the short echo time ^{1}H MRS data from all centres

Centre	Manufacturer	Sequence	TE (ms)	TR (ms)	SW (Hz)	N
CDP	GE	STEAM	20	1600, 2000	2500	2048
SGHMS	GE	STEAM	30	2000, 2018, 2020	2500	2048
		PRESS	30	2000, 2018, 2020	2500	2048
FLENI	GE	PRESS	30	2000	2500	2048
IDI	Philips	PRESS	30,31,32	2000	1000	512
UMCN	Siemens	STEAM	20	2000	1000	512
MUL	Siemens	STEAM	20	2000	1000	512

The columns respectively correspond to the name of the centre, the manufacturer of the scanner, the sequence used, the echo time, the repetition time, the spectral width, and the number of points in the original FID.

and astrocytomas grade II, which for the purpose of grouping are labeled as class 1, 2, 3, and 4, respectively (Table 2). Although the INTERPRET database does contain data from other types of brain tumours there are much fewer spectra of these. To allow statistically valid comparisons and to prevent unwanted bias due to large differences in the numbers of spectra within each group, we focused the analysis on the brain tumour types with more than 20 spectra available. Moreover, the selected tumour types are the major tumour groups occurring in the brain and include both benign/low grade (meningiomas and astrocytomas grade II) and malignant (glioblastomas and metastases) tumour types, and tumours of the same original cell type (glioblastomas vs. astrocytomas grade II). All data have passed a quality control and validation process, which was regulated by strict rules agreed on by all INTERPRET partners [33]. Tumour assignment was based on the histological classification of tumours of the central nervous system (CNS) set up by the World Health Organization (WHO) [34].

It is necessary to preprocess the raw MRS data prior to classification. First, frequency alignment and zero order phase correction was performed with Klose's method [35]. Second, the residual water peak was filtered out using HSVD [36], with 15 singular values over the water region of 4.09–5.31 ppm. The initial point of the time-domain signal was removed, because it was often affected by instrumental artefacts. The FIDs acquired with a GE-scanner had a different spectral width and number of points than the Siemens- and Philips-data (Table 1). To match these parameters, the GE-data were interpolated using piecewise cubic splines in the time domain. Hence, after preprocessing, all FIDs had 511 points (512 minus the deleted first point) which provided spectra with a final spectral width of 1000 Hz. The interpolation of the GE-data actually resulted in a four times lower input dimension than the original GE-data, which is more appropriate for classification. No additional referencing of the spectra was made, apart from the frequency alignment with Klose's method, leaving a slight misalignment (<0.08 ppm).

The resulting signal was then transformed to the frequency domain by fast Fourier transform (FFT). For classification, only the data points in the frequency region of interest (4.17–0 ppm), corresponding to 138 input variables, were used. Fig. 1 depicts the mean L2-normalized magnitude spectra without baseline correction of the four classes.

3. Methods

The standard linear discriminant analysis (LDA) [37–39] technique is compared to the kernel-based least squares support vector machine (LS-SVM) [39,40] technique with linear and RBF kernels. One of the interesting advantages of LS-SVM is its ability to learn and generalize processing of high-dimensional data but without dimensionality reduction, which is important for the present application.

Classification can be applied directly to the spectra, in which all spectral points are regarded as features. However, reduction of these features to a smaller number can be performed to further improve the perfor-

Table 2
Number of short echo time ^{1}H MRS data of glioblastomas (class 1), meningiomas (2), metastases (3), and astrocytomas grade II (4)

Centre	Acquisition scheme	Class 1	Class 2	Class 3	Class 4	Total
CDP	STEAM	30	12	5	4	51
SGHMS	STEAM	12	4	5	6	27
	PRESS	6	8	11	4	29
FLENI	PRESS	2	0	0	1	3
IDI	PRESS	31	30	17	5	83
UMCN	STEAM	2	1	0	1	4
MUL	STEAM	4	2	1	1	8
	Total	87	57	39	22	205

The rows correspond to the acquisition centre, while the columns show the type of brain tumour.

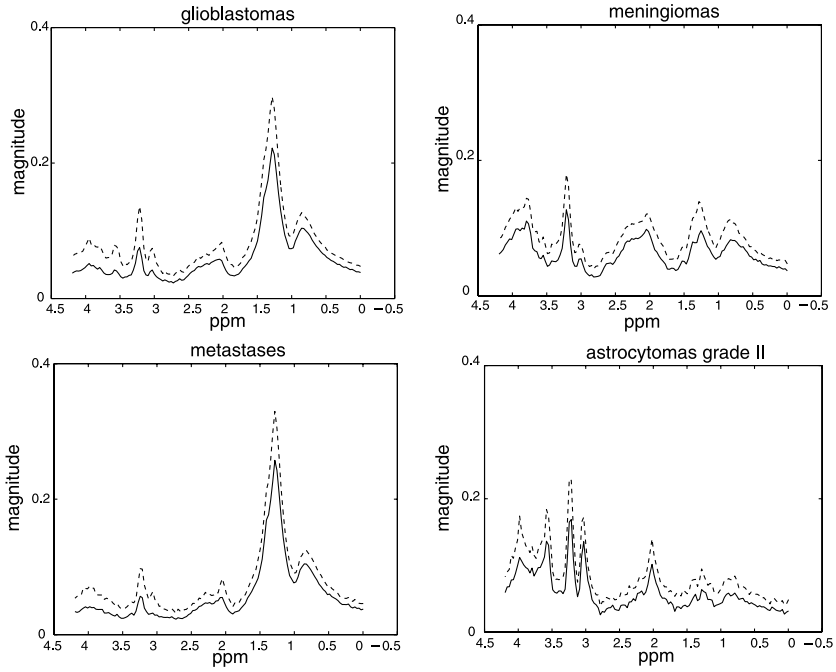

Fig. 1. Mean L2-normalized magnitude complete MRS spectra without baseline correction of the four classes: class 1 (top-left), class 2 (top-right), class 3 (bottom-left), and class 4 (bottom-right) correspond to the glioblastomas, meningiomas, metastases, and astrocytomas grade II, respectively. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

mance, also decreasing the amount of complexity and simplifying the calculation. Dimensionality reduction techniques try to extract the most characteristic input features, to minimize redundancy and to exclude noise and artefacts in the spectrum. This approach implicitly assumes that the discarded variables only contain a minor amount of discriminative information.

One approach to dimensionality reduction is to select the input features that are assumed to be most characteristic according to some prior knowledge (Sections 3.2.2 and 3.2.3). This approach is taken prior to using each of the classification techniques and is compared to the direct use of the complete spectra. A second approach is to select the input variables based on some criterion; principal component analysis (PCA) uses a minimal number of PCs that explain a certain amount of variability in the data. This technique is applied prior to the LDA classification technique, since LDA requires dimensionality reduction so that the number of input variables is not too high with respect to the number of training data.

3.1. Input features

3.1.1. Normalization

A normalization procedure should be applied that tries to compensate for signal intensity differences between spectra due to effects independent of the type of tumour tissue. Two normalization approaches are compared, which were elaborated in the following way:

- L2-normalization: the spectrum (the full spectral range of 1000 Hz was used) is considered as a vector, and the L2-norm of that vector is defined as the square root of the sum of the squares of the elements in that vector. In this study, we divided each spectral value by the L2-norm of the spectrum. Hence, a L2-normalized spectrum has unity norm.
- Water normalization: for each metabolite signal, a corresponding water unsuppressed signal was available, acquired with the same acquisition parameters as the metabolite signal and originating from the same voxel. The amplitude of the water unsuppressed signal was estimated with KULeuven ESAT-SCD software developed by Leentje Vanhamme (personal communication). The water unsuppressed signal was modeled in the time domain by a Voigt model [41], with an additional first order term that corrects for eddy currents. HSVD was used to obtain initial parameter estimates, which were subsequently used as starting values in a nonlinear least squares algorithm to obtain the final parameter estimates. Each spectral value in the metabolite spectrum was then divided by the resulting estimate of the intensity of the water peak.

Within the INTERPRET project, data were acquired using standard protocols, (i.e., short echo time PRESS or STEAM), to obtain MRS data measured under acquisition parameters that are as uniform as possible (Table 1). Water normalization compensates for differences in voxel size and receiver coil sensitivity variations,

assuming that different brain regions and tissues have the same water content. However, this assumption is not completely valid [42]. Absolute quantitation [30,43,44] might solve this problem, if used with an external reference signal, but is more time consuming and less practical for routine clinical situations.

3.1.2. Magnitude vs. real spectra

An MRS signal is typically represented by its real and imaginary part, and the process of ''phasing'' is required to correctly produce the real part of the spectrum for classification. Although automated phasing techniques exist, they are less suitable for in vivo data with large background signals. An alternative to using phased spectra is to calculate the magnitude spectrum, which makes the signal phase independent. In this paper, we compare the use of real and magnitude spectra for classification. This extends our previous study on long echo time signals, in which only magnitude spectra were considered [7].

3.1.3. Baseline correction

Short echo time ^{1}H MRS signals are characterized by the presence of an unknown broad baseline underlying the sharper resonances of the metabolites of interest, that hinders the assessment of the intensity (i.e., quantitation) of low weight metabolites. Several parametric and nonparametric approaches to model the macromolecular baseline are developed [45–50], but most of these techniques are integrated as part of a quantitation approach to estimate the MRS parameters. Independently of quantitation, the baseline can be modeled, e.g., by polynomial or spline functions, but this requires the selection of several spectral points to define the fitted baseline, a process which is heavily user-dependent.

In this paper, another approach was used, which requires the selection of only one parameter. The preprocessed FID was multiplied by an exponentially decreasing apodization function [12,51], given by $g_n = e^{-\beta n \Delta t}$, $n = 0,\ldots,N-1$, with N the number of points in the FID and a parameter β. The resulting FID is a model for the broad baseline. The value $\beta = 0.15$ was chosen to avoid, on the one hand a too flat baseline and hence overestimation of several metabolite resonances because of baseline contributions to the metabolite peaks (β too high), and on the other hand the underestimation of several metabolite resonances (β too low). The chosen baseline model was subtracted from the original signal to obtain the baseline corrected signal. The performances of classification using the baseline and nonbaseline corrected spectra are compared.

3.2. Dimensionality reduction by prior knowledge

3.2.1. Complete spectra

The spectra, obtained after preprocessing (Section 2) were used as input features.

3.2.2. Selected frequency regions

It is well known that characteristic resonance peaks correspond to important brain metabolites [1,3,42,52–54]. It seems reasonable then, that these peaks might be used as discriminatory features to distinguish tumour types, in particular for those regions of the ^{1}H spectrum which are clearly different between spectra of different tumour types. Thus, as an alternative to using complete spectra, selected frequency regions were used which are assumed to contain most of the information. Hence, the redundancy generated by spectral noise and artefacts in the spectrum was reduced. By taking a range of ±0.075 ppm around the resonance frequencies [31] of several characteristic metabolites, the number of variables was reduced to 71. The following regions of the spectrum were selected: *L2* (lipids at ≈0.9 ppm; 0.825–0.975 ppm); *L1* (lipids at ≈1.2 ppm; 1.125–1.275 ppm); *Lac* (^{3}CH$_3$-group; 1.235–1.385 ppm); *Ala* (^{1}CH$_3$-group; 1.395–1.545 ppm); *NAA* (*N*-acetyl aspartate, ^{2}CH$_3$-group; 1.935–2.085 ppm); *Glx* (^{4}CH$_2$-group; 2.270–2.510 ppm); *Cr* (total creatine, N(CH$_3$)-group; 2.955–3.105 ppm); *Cho* (choline containing compounds, N(CH$_3$)$_3$-group; 3.115–3.265 ppm); *mI* (^{5}CH-group) + *Tau* (taurine, ^{2}CH$_2$-group) (3.195–3.345 ppm); *mI*/*Gly*(glycine) (^{1}CH-, ^{3}CH-, ^{4}CH-, and ^{6}CH-groups; 3.450–3.680 ppm); *Tau* (^{1}CH$_2$-group; 3.345–3.495 ppm); *Glx* + *Ala* (^{2}CH-groups; 3.665–3.815 ppm); and *Cr* (^{2}CH$_2$-group; 3.845–3.995 ppm). This set of metabolite peaks is an extension of the one used in our study on long echo time MRS [7], because several additional metabolites (e.g., *Glx*, *mI*, and *Tau*) are more visible in short echo time spectra.

3.2.3. Peak integration

Another approach to select the most discriminatory input is based on peak integration. However, precise estimation of the peak integrals is difficult due to several factors, including nonzero baseline, peak overlap, noise, and the discrete nature of the spectrum. Peak integration in the current study was performed using the trapezoidal rule [55]. For each selected metabolite the area under the frequency peak in the spectrum was calculated. We used the same frequency ranges as for the selected frequency regions above, and so reduced the number of input variables to 13.

3.3. Experimental setting

3.3.1. Binary classification

Given four types of brain tumours, six binary classifiers can be constructed to separate the following pairs: glioblastomas vs. meningiomas (class 1 vs. 2), glioblastomas vs. metastases (1 vs. 3), glioblastomas vs. astrocytomas grade II (1 vs. 4), meningiomas vs. metastases (2 vs. 3), meningiomas vs. astrocytomas grade II (2 vs. 4), and metastases vs. astrocytomas grade II (3 vs. 4).

We applied the aforementioned techniques in the following way:

(1) LDA and LS-SVM with a linear kernel were used as linear techniques, while LS-SVM with an RBF kernel was used as a nonlinear technique.
(2) LS-SVM techniques were applied using the complete spectra, using the selected frequency regions as well as using peak integration.
(3) Feature selection using PCA prior to LDA was applied except for the case of peak integration where a specific dimensionality reduction was already applied. PCA reduces the original number of spectral variables to a minimal set of variables that accounts for 75% of the variance of the data. Increasing the number of selected PCs, for example, by taking 80 or 85% of the variance, causes rank deficiency when applying LDA. Rank deficiency typically occurs when the number of input variables is too high with respect to the number of training data. An alternative criterion could be the selection of the number of PCs by inspecting a scree plot [56], but this was excluded as it is difficult to automate.

LS-SVM classification was applied using KULeuven's MATLAB/C LS-SVMlab toolbox [39,57,58] using the same approach as described in [7]. The experiment consisted of the following steps:

(1) division of the dataset into a training (2/3 of the data) and a test set (remainder),
(2) training the classifiers using the training set,
(3) evaluation of the performance using the test set.

Stratified random sampling was used while dividing the dataset to preserve the proportion of the classes. The procedure was repeated 100 times to avoid bias possibly introduced by selection of a specific training and test set. In this way, we tried to obtain a representative test performance. ROC [25–27] analysis was applied to measure the discrimination ability of the binary classifiers. The classification performance was then measured by the mean AUC and its pooled standard error (SE) calculated from 100 randomizations, as described in [7]. The z ratio [59] was applied to statistically test whether the areas under two ROC curves derived from the same samples differ significantly from each other at a significance level of 0.05. We note that when interpreting the reported results the mean AUC as well as its pooled SE should be taken into account. The mean AUC, as an average performance, gives an indication of a typical AUC obtained using the given input data, while the pooled SE indicates how reliably the mean AUC is estimated.

3.3.2. Multiclass classification

Using a binary classifier, a new spectrum of unknown tumour type may be assigned to one of the two considered classes. However, in medical practice, the number of possible tumour types is mostly not restricted to two types. Therefore, a multiclass scheme has been developed to handle all classes in one construction, thereby extending the restricted use of the binary classifiers mentioned in Section 3.3.1.

MR spectra of class 1 (glioblastomas) and class 3 (metastases) show a very similar pattern (Fig. 1), hence both classes are merged, resulting into a new group of aggressive tumours, labeled as class 5 (cf. [7,16]). A scheme to discriminate classes 2, 4, and 5 is depicted as step one on the left part in Fig. 2. A voting scheme has been applied to decide which class is chosen based on the three outputs of the contributing binary classifiers. A certain class is taken if two of the binary classifiers give the same output, otherwise the classifier considers the output as *undecided*. Step two is carried out as illustrated in the right part: if the output of step one is class 5, then the spectrum is further classified either into class 1 or 3 using the binary classifier 1 vs. 3. The output of step two does not change if step one gives class 2 or 4 as output. Four binary classifiers are the building blocks of this multiclass classifier: classifiers 1 vs. 3 and 2 vs. 4, and meningiomas vs. aggressive tumours (class 2 vs. 5),

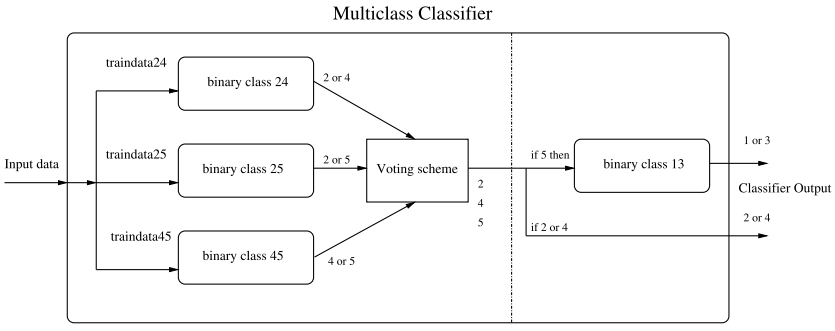

Fig. 2. Two-step classification. The left part shows step one, classification of three tumour classes: meningiomas (class 2), astrocytomas grade II (4), and aggressive tumours (5). The right part, or step two, further refines the classification if the output is class 5 and assigns the spectra of this class either to glioblastomas (class 1) or metastases (3).

and astrocytomas grade II vs. aggressive tumours (class 4 vs. 5) as two additional classifiers.

Similarly as for binary classification, stratified random splitting was applied to select 2/3 of the dataset as training and the remainder as test set. The multiclass classifier was trained by feeding all training data to the classifier and each binary classifier was trained with the corresponding classes. For example, the training data of classes 2 and 4 were used to train the binary classifier 2 vs. 4, and similarly for the others. This fully determined the multiclass classifier constructed by four binary classifiers (Fig. 2). The resulting classifier was applied to the independent test set. This procedure was repeated for 100 runs. The test performance is given by the mean correct classification rate, the mean misclassification rate, the mean percentage of undecided cases, and their standard deviation. The correct classification rate is defined as the percentage of correctly classified spectra, while the misclassification rate is the percentage of misclassified cases.

4. Results

4.1. Binary classification

The test performance of the classifiers using the L2-normalized complete magnitude spectra without baseline correction is shown in Table 3. Tables 4–8 each differ from Table 3 in one single aspect. By comparing Table 3 with Tables 4–8, the individual effect on the classification performance of normalization, baseline correction, real vs. magnitude spectra, dimensionality reduction by selected frequency regions, and peak integration, respectively, can be determined. In order not

Table 3
Binary classification using L2-normalized complete magnitude spectra without baseline correction

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
1 vs. 2	$0.9556 \pm 0.0278_{(4)}$	0.9725 ± 0.0210	0.9730 ± 0.0204
1 vs. 3	$0.5907 \pm 0.0967_{(4)}$	0.5917 ± 0.0939	0.5873 ± 0.0942
1 vs. 4	$0.9660 \pm 0.0292_{(3)}$	0.9635 ± 0.0305	0.9615 ± 0.0314
2 vs. 3	$0.9540 \pm 0.0441_{(4)}$	0.9770 ± 0.0263	0.9781 ± 0.0254
2 vs. 4	$0.9965 \pm 0.0093_{(11)}$	0.9959 ± 0.0100	0.9902 ± 0.0168
3 vs. 4	$0.9856 \pm 0.0248_{(2)}$	0.9916 ± 0.0178	0.9903 ± 0.0194

As performance measure we use the mean test AUC and its pooled standard error SE from 100 runs of stratified random samplings. The number between the brackets mentions the number of principal components used (column PCA/LDA).

Table 4
Binary classification using water normalized complete magnitude spectra without baseline correction

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
1 vs. 2	$0.8399 \pm 0.0607_{(1)}$	0.9294 ± 0.0397	0.9581 ± 0.0281
1 vs. 3	$0.5492 \pm 0.0935_{(1)}$	0.5917 ± 0.0928	0.5950 ± 0.0935
1 vs. 4	$0.9485 \pm 0.0403_{(1)}$	0.9560 ± 0.0339	0.9598 ± 0.0326
2 vs. 3	$0.8924 \pm 0.0618_{(1)}$	0.9616 ± 0.0470	0.9704 ± 0.0309
2 vs. 4	$0.7427 \pm 0.1199_{(1)}$	0.9930 ± 0.0139	0.9793 ± 0.0303
3 vs. 4	$0.9784 \pm 0.0297_{(1)}$	0.9757 ± 0.0386	0.9837 ± 0.0284

For further explanation we refer to Table 3.

Table 5
Binary classification using L2-normalized complete magnitude spectra with baseline correction

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
1 vs. 2	$0.9591 \pm 0.0265_{(5)}$	0.9625 ± 0.0271	0.9616 ± 0.0269
1 vs. 3	$0.5762 \pm 0.0939_{(5)}$	0.5914 ± 0.0931	0.6002 ± 0.0930
1 vs. 4	$0.9650 \pm 0.0295_{(4)}$	0.9584 ± 0.0335	0.9587 ± 0.0334
2 vs. 3	$0.9607 \pm 0.0388_{(5)}$	0.9685 ± 0.0308	0.9736 ± 0.0281
2 vs. 4	$0.9965 \pm 0.0099_{(13)}$	0.9973 ± 0.0082	0.9960 ± 0.0100
3 vs. 4	$0.9725 \pm 0.0411_{(3)}$	0.9813 ± 0.0309	0.9836 ± 0.0282

For further explanation we refer to Table 3.

Table 6
Binary classification using L2-normalized complete real spectra without baseline correction

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
1 vs. 2	$0.9543 \pm 0.0279_{(6)}$	0.9743 ± 0.0196	0.9742 ± 0.0199
1 vs. 3	$0.5628 \pm 0.0984_{(4)}$	0.5655 ± 0.0970	0.5753 ± 0.0973
1 vs. 4	$0.9755 \pm 0.0240_{(5)}$	0.9717 ± 0.0264	0.9691 ± 0.0278
2 vs. 3	$0.9562 \pm 0.0425_{(6)}$	0.9702 ± 0.0319	0.9734 ± 0.0288
2 vs. 4	$0.9959 \pm 0.0102_{(13)}$	0.9947 ± 0.0119	0.9917 ± 0.0154
3 vs. 4	$0.9924 \pm 0.0167_{(4)}$	0.9893 ± 0.0210	0.9875 ± 0.0233

For further explanation we refer to Table 3.

Table 7
Binary classification using selected frequency regions of the L2-normalized magnitude spectra without baseline correction

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
1 vs. 2	$0.9060 \pm 0.0430_{(2)}$	0.9748 ± 0.0202	0.9733 ± 0.0200
1 vs. 3	$0.5959 \pm 0.0932_{(2)}$	0.5898 ± 0.0945	0.6085 ± 0.0930
1 vs. 4	$0.9500 \pm 0.0358_{(1)}$	0.9636 ± 0.0318	0.9620 ± 0.0315
2 vs. 3	$0.9495 \pm 0.0468_{(2)}$	0.9712 ± 0.0316	0.9683 ± 0.0336
2 vs. 4	$0.9958 \pm 0.0103_{(7)}$	0.9956 ± 0.0105	0.9971 ± 0.0086
3 vs. 4	$0.9947 \pm 0.0133_{(1)}$	0.9924 ± 0.0172	0.9929 ± 0.0165

For further explanation we refer to Table 3.

Table 8
Binary classification using peak integrated values of the L2-normalized magnitude MR spectra without baseline correction

Classes	LDA	LS-SVM lin	LS-SVM RBF
1 vs. 2	0.9675 ± 0.0243	0.9704 ± 0.0221	0.9713 ± 0.0210
1 vs. 3	0.5054 ± 0.0940	0.5762 ± 0.0956	0.6073 ± 0.0935
1 vs. 4	0.9580 ± 0.0335	0.9616 ± 0.0328	0.9539 ± 0.0355
2 vs. 3	0.9641 ± 0.0364	0.9640 ± 0.0378	0.9632 ± 0.0379
2 vs. 4	0.9968 ± 0.0090	0.9965 ± 0.0092	0.9875 ± 0.0201
3 vs. 4	0.9892 ± 0.0219	0.9938 ± 0.0147	0.9938 ± 0.0146

Note that in this case no PCA prior to LDA is applied, because peak integration already involves dimensionality reduction. For further explanation we refer to Table 3.

to overload the paper, the results for the other cases are not tabulated, but can be obtained by correspondence with the authors. We will further discuss only the tabulated results included in this paper in detail. However, the classification results for the other input features all showed a similar behaviour.

In general, all techniques achieved a good performance with mean AUCs up to 0.9971, with only slightly and nonsignificantly different results when mutually comparing classification techniques. However, the discrimination of glioblastomas and metastases remains a very difficult binary problem, resulting in a very poor performance (with a mean AUC in between 0.5054 and 0.6085 and a high pooled SE of up to 0.0984) for all the classification techniques investigated. A slightly lower performance was observed for the discrimination of glioblastomas and metastases when applying PCA/LDA as compared to LS-SVMs (Tables 4, 5, and 8), but for all other binary classification problems a high performance was reached with a mean AUC of more than 0.95. An exception occurred when using water normalized spectra and selected frequency regions, for which almost all results are worse with respect to using L2-normalized complete spectra. The discrimination of classes 1 and 2 and of classes 2 and 4 yielded a significantly lower performance for water normalization, using PCA/LDA (classes 1 vs. 2 and 2 vs. 4) and using LS-SVM with a linear kernel (classes 1 vs. 2). Although less striking, when using selected frequency regions a significantly lower performance was obtained for the classification of classes 1 vs. 2 when applying PCA/LDA as compared to using the complete spectra.

4.2. Multiclass approach

After merging of glioblastomas and metastases into one class of aggressive tumours, two additional binary classifiers were constructed in addition to those previously discussed. Table 9 shows the performance of these classifiers using the complete spectra as input. Tables 10 (first step) and 11 (second step) report the multiclass test results using complete spectra. A mean correct test classification rate of 85.1% (PCA/LDA), 87.3% (LS-SVM lin), and 86.6% (LS-SVM RBF) has been reached after the first step for all classification techniques. After the second step, the test performance has been reduced to 67.6% (PCA/LDA), 69.4% (LS-SVM lin), and 68.5% (LS-SVM RBF) because of the difficulty in discriminating classes 1 and 3.

Table 9
Binary classification using complete spectra for two additional binary problems, needed to implement the multiclass classification

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
2 vs. 5	$0.9570 \pm 0.0245_{(4)}$	0.9803 ± 0.0148	0.9801 ± 0.0148
4 vs. 5	$0.9701 \pm 0.0232_{(3)}$	0.9728 ± 0.0225	0.9725 ± 0.0226

Average performance on the test set from 100 runs of stratified random samplings of the L2-normalized magnitude MR spectra without baseline correction. As performance measure we use the mean AUC and its pooled standard error SE.

Table 10
One-step multiclass classification using complete spectra

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
Correct	85.0882 ± 4.2939	87.2794 ± 3.6466	86.6324 ± 3.4857
Misclass	9.3676 ± 3.6722	8.7941 ± 3.5409	9.8971 ± 3.4336
Undecided	5.5441 ± 2.4183	3.9265 ± 2.0592	3.4706 ± 1.9102

Average test performance from 100 runs of stratified random samplings of the L2-normalized magnitude MR spectra without baseline correction. The first, second, and third rows give, respectively, the mean correct classification rate, the mean misclassification rate and the mean percentage of undecided cases, each with their standard deviation.

Table 11
Two-step multiclass classification using complete spectra

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
Correct	67.6176 ± 4.9418	69.4118 ± 3.6082	68.5441 ± 4.0417
Misclass	26.8382 ± 4.5043	26.6618 ± 3.6122	27.9853 ± 4.0984
Undecided	5.5441 ± 2.4183	3.9265 ± 2.0592	3.4706 ± 1.9102

Average test performance from 100 runs of stratified random samplings of the L2-normalized magnitude MR spectra without baseline correction.

5. Discussion

We next discuss various issues concerning the results we obtained using the short echo time ^{1}H MRS data available. We do not necessarily claim that these remarks generally hold for similar analyses on other data.

5.1. Classification techniques

LDA has been widely used and has obtained good results for many applications including brain tumour classification [7,16,60]. The method has the strength of being simple and computationally fast, and works well for data that are linearly separable. Nevertheless, LDA also has some disadvantages. These include the nonuniqueness of the derived discrimination function [61] and the potential poor performance in the case of severe nonlinearity. Moreover, in the case of a high number of variables relative to the number of training data (e.g., in the paper presented), LDA requires the data to be reduced in dimensionality, as by PCA for example. In contrast to classical techniques, LS-SVM as a kernel-based technique is also suited for linearly nonseparable problems and is able to find a global solution [39]. Kernel-based methods are also less sensitive to the number of data. Although the dimension is larger than the number of data, these classifiers have the advantage to

robustly learn the peak pattern and draw an optimal separating boundary, even without applying any dimensionality reduction. Linear as well as nonlinear classifiers can be selected by choosing an appropriate kernel. However, in contrast to classical techniques, LS-SVM is a more complex method that requires to select a set of hyperparameters in order to obtain a high performance.

LDA, as a linear classifier, performed quite well in discriminating brain tumours using complete spectra. Except for the most difficult binary classification to discriminate glioblastomas and metastases, classification of L2-normalized non-baseline corrected complete magnitude spectra reached a mean AUC of at least 0.95. Using the same input features as LDA, LS-SVMs with a linear and an RBF kernel both reached a mean AUC of at least 0.96, which is similar to the performance of LDA. Although the tables indicate a few differences in the classification performance, no statistical significance was found between the AUC values from different classification techniques. Due to the limited number of data available (i.e., limited in terms of classification), nonlinear techniques, in this case LS-SVMs with an RBF kernel, cannot fully exploit the advantage of a nonlinear method.

These findings are in agreement with our previous study on long echo time spectra reported in [7] and results described in [16]. Direct comparison of the performances even reveals overall slightly higher mean AUC values with respect to the use of long echo time spectra (confront Tables 3, 7, and 8 with [7]), but this was not statistically tested because different datasets were used.

5.2. Influence of input features

The purpose of normalization is to obtain consistent scaling of the measurement data. Using the complete magnitude spectra without baseline correction, a lower performance was found for water normalization compared to L2-normalization. This includes a few significant differences when using PCA/LDA and LS-SVM with a linear kernel. Only slightly different results were observed due to the normalization method when using nonlinear techniques.

Although Klose's method was used to phase the spectra, there still exists residual phase variation in the real spectra, which may reduce the accuracy of classification. Phasing problems can be avoided by using magnitude spectra, but this yields a significant broadening of the peaks [55]. Hence, magnitude calculation results in a larger overlap of the individual peaks, which potentially distorts the information contained in the spectrum. However, no significant differences in classification performance were found between the use of complete real and magnitude spectra without baseline correction.

In short echo time spectra the appearance of the metabolite peaks is heavily influenced by peak overlap and the presence of a macromolecular baseline, hence a baseline correction was applied in an attempt to improve resolution of the metabolite peaks. Spectra with baseline correction were expected to have clearer peak patterns and to result in a more robust classification, but no improvement was found. Although our baseline correction may simplify the identification of metabolite peaks, macromolecule signals included in the baseline may contribute to the biochemical fingerprint of the tumours. Hence, baseline correction may result in loss of some valuable information, which may explain why classification does not improve after baseline correction.

5.3. Influence of dimensionality reduction

In classifying MR spectra using selected frequency regions and peak integration, all techniques achieved at least a mean AUC of 0.90 except for the binary classification problem to discriminate glioblastomas and metastases. In comparison to classification using complete spectra, classification using selected frequency regions or peak integrated values reached a similar performance. This suggests that the selected metabolite regions or integrated signal areas include most of the information that correctly distinguishes between types of brain tumour. Only for the discrimination of glioblastomas and meningiomas PCA/LDA performed significantly worse using selected frequency regions than using the complete spectra.

5.4. Multiclass classification

Although no statistical analysis has been carried out for multiclass classification, we can make some observations from the results. All classification techniques reached a high test performance of 85–87% for distinguishing classes 2, 4, and 5 in the first step. In the second step, the test performances reduced with about 18% because of the difficulties in separating glioblastomas from metastases, which share very similar peak patterns. Therefore, separating class 5 (aggressive tumours) into class 1 (glioblastomas) and 3 (metastases) deteriorates the total performance of the classifier.

5.5. Clinical practice

We have performed an objective comparison of several classification techniques, based on several types of input features. This requires the selection of a training set to construct a classifier and an independent test set to evaluate that classifier. Therefore we implemented an experimental setting of 100 stratified randomizations, which resulted each time in another classifier.

The authors are aware of the fact that, from a clinical point of view, a practical classification system requires

only one specific classifier. The classification system is then used to assign new MRS data to a certain class, corresponding to the type of the tumour and its grade. However, our objective comparison makes it possible to obtain an overview of the performance of several techniques, with their advantages and disadvantages, and to evaluate the influence on classification of the type of input feature used. From such results it should be possible to make an informed choice of classification technique, as well as of the most appropriate input features and implement this in a practical classification system. For example, this selection could depend on the computational complexity and on the types of brain tumours the system typically should be able to discriminate. It should be noted that all techniques are applied automatically, including (hyper-)parameter selection, training and testing. Hence, for use in clinical practice, all techniques are easy to automate for analysis of independent data.

6. Conclusions

In general, no significant differences were found with respect to the use of classification techniques. Linear techniques found the best separating linear boundary for most of the cases. Kernel-based methods have the advantage of robustly learning the peak pattern and were able to reach a high mean AUC (except for glioblastomas vs. metastases), even without the use of any feature reduction. Kernel-based methods are also more robust against different types of normalization than LDA, which may be a result of data information loss due to PCA.

From our analysis, we found that the best classification could be obtained using L2-normalised magnitude spectra without baseline correction and simply using peak integration or PCA for dimensionality reduction. This is an important result if it would be generally true as it means that neither the difficult task of accurate spectral phasing, or acquisition of a water reference for lineshape correction and zero order phasing would be required. Thus, data acquisition times can be reduced and preprocessing protocols can be simplified. This would be particularly useful for spectroscopic imaging data, for which there is a significant time penalty in acquiring a full water reference data set.

Additionally, although peak integration may not accurately quantify overlapping metabolite signals [62] and PCA only expresses up to 75% of the data, automated dimensionality reduction by peak integration or PCA speeds up the computation and does not appear to compromise the classification efficiency for the comparison made in this study. Further studies are still needed to assess whether other approaches to dimensionality reduction, such as more accurate model-based quantification methods [48,62], yield any significant improvement in classification performance.

Acknowledgments

This research work was carried out at the ESAT laboratory and the Interdisciplinary Center of Neural Networks ICNN of the Katholieke Universiteit Leuven, in the framework of the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture (IUAP Phase V-22), the Concerted Action Project MEFISTO of the Flemish Community, the FWO projects G.0407.02 and G.0269.02 and the IDO/99/03 and IDO/02/009 projects. AD research financed by IWT grant of the Flemish Institute for the promotion of scientific-technological research in the industry. LVH is a postdoctoral researcher with the National Fund for Scientific Research FWO—Flanders. Use of the data provided by the EU funded INTERPRET project (IST-1999-10310; http://carbon.uab.es/INTERPRET/) is gratefully acknowledged. Margarida Julià-Sapé is gratefully acknowledged for all data management tasks. The clinical partners represented by Dr. Witold Gajewicz (MUL, Łodzi, Poland) and Jorge Andrés Calvar (FLENI, Ciudad de Buenos Aires, Argentina) are gratefully acknowledged for providing data.

References

[1] S.K. Mukherji (Ed.), Clinical Applications of Magnetic Resonance Spectroscopy, Wiley-Liss, New York, 1998.
[2] S.J. Nelson, Multivoxel magnetic resonance spectroscopy of brain tumors, Mol. Cancer Ther. 2 (2003) 497–507.
[3] I.C.P. Smith, L.C. Stewart, Magnetic resonance spectroscopy in medicine: clinical impact, Prog. Nucl. Magn. Reson. Spectrosc. 40 (2002) 1–34.
[4] M.A. Mittler, B.C. Walters, E.G. Stopa, Observer reliability in histological grading of astrocytoma stereotactic biopsies, J. Neurosurg. 85 (6) (1996) 1091–1094.
[5] D.L. Arnold, E.A. Shoubridge, J.G. Villemure, W. Feindel, Proton and phosphorus magnetic resonance spectroscopy of human astrocytomas in vivo. Preliminary observations on tumor grading, NMR Biomed. 3 (4) (1990) 184–189.
[6] S. Herminghaus, T. Dierks, U. Pilatus, W. Möller-Hartmann, J. Wittsack, G. Marquardt, C. Labish, H. Lanfermann, W. Schlote, F.E. Zanella, Determination of histopathological tumor grade in neuroepithelial brain tumors by using spectral pattern analysis of in vivo spectroscopic data, J. Neurosurg. 98 (2003) 74–81.
[7] L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, F.A. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, A.R. Tate, C. Aruś, S. Van Huffel, Brain tumour classification based on long echo time proton MRS signals, Artif. Intell. Med. 31 (2004) 73–89.
[8] M.E. Meyerand, J.M. Pipas, A. Mamourian, T.D. Tosteson, J.F. Dunn, Classification of biopsy-confirmed brain tumors using single-voxel MR spectroscopy, Am. J. Neuroradiol. 20 (1999) 117–123.

[9] T.R. McKnight, S.M. Noworolski, D. Vigneron, S. Nelson, An automated technique for the quantitative assessment of 3D-MRSI data from patients with glioma, J. Magn. Reson. Imag. 13 (2001) 167–177.
[10] M.C. Preul, Z. Caramanos, D.L. Collins, J.-G. Villemure, R. Leblanc, A. Olivier, R. Pokrupa, D.L. Arnold, Accurate, noninvasive diagnosis of human brain tumors by using magnetic resonance spectroscopy, Nat. Med. 2 (3) (1996) 323–325.
[11] A. Rutter, H. Hugenholtz, J.K. Saunders, I.C.P. Smith, Classification of brain tumors by ex vivo ^{1}H NMR spectroscopy, J. Neurochem. 64 (4) (1995) 1655–1661.
[12] A.W. Simonetti, W.J. Melssen, M. van der Graaf, A. Heerschap, L.M.C. Buydens, A new chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy, Anal. Chem. 75 (20) (2003) 5352–5361.
[13] R.L. Somorjai, B. Dolenko, A.K. Nikulin, N. Pizzi, G. Scarth, P. Zhilkin, W. Halliday, D. Fewer, N. Hill, I. Ross, M. West, I.C.P. Smith, S.M. Donnelly, A.C. Kuesel, K.M. Briere, Classification of 1H MR spectra of human brain neoplasms: the influence of preprocessing and computerized consensus diagnosis on classification accuracy, Magn. Reson. Imag. 6 (3) (1996) 437–444.
[14] F. Szabo De Edelenyi, C. Rubin, F. Estève, S. Grand, M. Décorps, V. Lefournier, J.-F. Le Bas, C. Rémy, A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images, Nat. Med. 6 (2000) 1287–1289.
[15] A.R. Tate, J.R. Griffiths, I. Martínez-Pérez, A. Moreno, I. Barba, M.E. Cabañas, D. Watson, J. Alonso, F. Bartumeus, F. Isamat, I. Ferrer, F. Vila, E. Ferrer, A. Capdevila, C. Arús, Towards a method for automated classification of ^{1}H MRS spectra from brain tumours, NMR Biomed. 11 (1998) 177–191.
[16] A.R. Tate, C. Majós, A. Moreno, F.A. Howe, J.R. Griffiths, C. Arús, Automated classification of short echo time in in vivo ^{1}H brain tumor spectra: a multicenter study, Magn. Reson. Med. 49 (2003) 29–36.
[17] C.-Z. Ye, J. Yang, D.-Y. Geng, Y. Zhou, N.-Y. Chen, Fuzzy rules to predict degree of malignancy in brain glioma, Med. Biol. Eng. Comput. 40 (2002) 145–152.
[18] International network for Pattern Recognition of Tumours Using Magnetic Resonance. Available from<http://carbon.uab.es/INTERPRET/>.
[19] C. Ladroue, F.A. Howe, J.R. Griffiths, A.R. Tate, Independent component analysis for automated decomposition of in vivo magnetic resonance spectra, Magn. Reson. Med. 50 (2003) 697–703.
[20] C. Ladroue, A.R. Tate, F.A. Howe, J.R. Griffiths, Exploring magnetic resonance data with independent component analysis, in: Proceedings of the 19th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB02), Cannes, France, August 22–25, 2002, pp. 147–148.
[21] L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, S. Van Huffel, A.R. Tate, C. Majós, C. Arús, The use of LS-SVM in the classification of brain tumors based on magnetic resonance spectroscopy signals, in: Proceedings of the European Symposium for Artifical Neural Networks (ESANN02), Bruges, Belgium, April 24–26, 2002, pp. 131–135.
[22] L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, S. Van Huffel, A.R. Tate, C. Majós, C. Arús, The use of LS-SVM in the classification of brain tumors based on ^{1}H-MR Spectroscopy signals, in: Proceedings of the IEE Symposium on Medical applications of signal processing, Savoy Place, London, UK, October 7, 2002, pp. 15/1–5.
[23] F. Szabo De Edelenyi, F. Estève, C. Rémy, L. Buydens, Application of independent component analysis to ^{1}H MR spectroscopic imaging exams of brain tumors, in: Proceedings of the 19th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB02), Cannes, France, August 22–25, 2002, p. 91.
[24] A.R. Tate, J.R. Griffiths, F.A. Howe, J. Pujol, C. Arús, Differentiating types of human brain tumours by MRS. A comparison of pre-processing methods and echo times, in: Proceedings of the Ninth Scientific Meeting & Exhibition (ISMRM01), Glasgow, Scotland, April 21–27, 2001, p. 2284.
[25] J.A. Hanley, B.J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1982) 29–36.
[26] N.A. Obuchowski, Receiver operating characteristic curves and their use in radiology, Radiology 229 (2003) 3–8.
[27] J.A. Swets, ROC analysis applied to the evaluation of medical imaging techniques, Invest. Radiol. 14 (2) (1979) 109–121.
[28] I. Barba, A. Moreno, I. Martinez-Perez, A.R. Tate, M.E. Cabanas, M. Baquero, A. Capdevila, C. Arús, Magnetic resonance spectroscopy of brain hemangiopericytomas: high myoinositol concentrations and discrimination from meningiomas, J. Neurosurg. 94 (1) (2001) 55–60.
[29] T. Ernst, J. Hennig, Coupling effects in volume selective 1H spectroscopy of major brain metabolites, Magn. Reson. Med. 21 (1) (1991) 82–96.
[30] T. Ernst, R. Kreis, B.D. Ross, Absolute quantitation of water and metabolites in the human brain. I. Compartments and water, J. Magn. Reson. B 102 (1993) 1–8.
[31] V. Govindaraju, K. Young, A.A. Maudsley, Proton NMR chemical shifts and coupling constants for brain metabolites, NMR Biomed. 13 (2000) 129–153.
[32] M. van der Graaf, M. Rijpkema, Y. vander Meulen, C. Majós, A. Moreno, A. Ziegler, F.A. Howe, K. Opstad, A. Heerschap, System quality assurance in INTERPRET, a multicenter study on brain tumor MR spectroscopy, MAGMA 15 (Suppl.1) (2002) 245–246.
[33] Quality Control of MR spectra in the INTERPRET-validated database. Available from<http://carbon.uab.es/INTERPRET/mrs_data/docs/MRS_QC.pdf>.
[34] P. Kleihues, W.K. Cavenee, World Health Organization Classification of Tumours: Pathology & Genetics, Tumours of the nervous system, International Agency for Research on Cancer (IARC), Lyon, 2000.
[35] U. Klose, In vivo proton spectroscopy in presence of eddy currents, Magn. Reson. Med. 14 (1990) 26–30.
[36] H. Barkhuijsen, R. De Beer, D. Van Ormondt, Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals, J. Magn. Reson. 73 (1987) 553–557.
[37] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley, New York, 2001.
[38] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, Cambridge, 1996.
[39] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least Squares Support Vector Machines, World Scientific, Singapore, 2002.
[40] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neur. Proc. Lett. 9 (3) (1999) 293–300.
[41] I. Marshall, J. Higinbotham, S. Bruce, A. Freise, Use of Voigt lineshape for quantification of in vivo ^{1}H spectra, Magn. Reson. Med. 37 (1997) 651–657.
[42] F.A. Howe, S.J. Barton, S.A. Cudlip, M. Stubbs, D.E. Saunders, M. Murphy, P. Wilkins, K.S. Opstad, V.L. Doyle, M.A. McLean, B.A. Bell, J.R. Griffiths, Metabolic profiles of human brain tumors using quantitative in vivo ^{1}H Magnetic resonance spectroscopy, Magn. Reson. Med. 49 (2003) 223–232.
[43] O. Henriksen, In vivo quantitation of metabolite of concentrations in the brain by means of proton MRS, NMR Biomed. 8 (1995) 139–148.

[44] R. Kreis, T. Ernst, B.D. Ross, Absolute quantitation of water and metabolites in the human brain. II. Metabolite concentrations, J. Magn. Reson. B 102 (1993) 9–19.
[45] R. Bartha, D.J. Drost, P.C. Williamson, Factors affecting the quantification of short echo in-vivo ^{1}H MR spectra: prior knowledge, peak elimination, and filtering, NMR Biomed. 12 (1999) 205–216.
[46] L. Hofmann, J. Slotboom, C. Boesch, R. Kreis, Model fitting of 1H-MR spectra of the human brain: incorporation of short-T1 components and evaluation of parametrized vs. non-parametrized models, in: Proceedings of the Ninth Scientific Meeting of the International Society of Magnetic Resonance in Medicine (ISMRM), Philadelphia, USA, May 22–28, 1999, p. 586.
[47] L. Hofmann, J. Slotboom, B. Jung, P. Maloca, C. Boesch, R. Kreis, Quantitative ^{1}H-magnetic resonance spectroscopy of human brain: influence of composition and parameterization of the basis set in linear combination model-fitting, Magn. Reson. Med. 48 (2002) 440–453.
[48] S. Provencher, Estimation of metabolite concentrations from localized in-vivo proton NMR spectra, Magn. Reson. Med. 30 (1993) 672–679.
[49] U. Seeger, U. Klose, I. Mader, W. Grodd, T. Nägele, Parametrized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases, Magn. Reson. Med. 49 (2003) 19–28.
[50] B.J. Soher, K. Young, A.A. Maudsley, Representation of strong baseline contribution in ^{1}H MR spectra, Magn. Reson. Med. 45 (2001) 966–972.
[51] I.D. Campbell, C.M. Dobson, R.J.P. Williams, A.V. Xavier, Resolution enhancement of protein PMR spectra using the difference between a broadened and a normal spectrum, J. Magn. Reson. 11 (1973) 172–181.
[52] X. Lecrerc, T.A.G.M. Huisman, A.G. Sorensen, The potential of proton magnetic resonance spectroscopy (^{1}H) in the diagnosis and management of patients with brain tumors, Curr. Opin. Oncol. 14 (2002) 292–298.
[53] C. Majós, J. Alonso, C. Aguilera, M. Serrallonga, J.J. Acebes, C. Arús, J. Gili, Adult primitive neuroectodermal tumor: proton MR spectroscopic findings with possible application for differential diagnosis, Radiology 225 (2002) 556–566.
[54] M. Murphy, A. Loosemore, A.G. Clifton, F.A. Howe, A.R. Tate, S.A. Cudlip, P.R. Wilkins, J.R. Griffiths, B.A. Bell, The contribution of proton magnetic resonance spectroscopy (^{1}H MRS) to clinical brain tumour diagnosis, Br. J. Neurosurg. 16 (4) (2002) 329–334.
[55] J.C. Hoch, A.S. Stern, NMR data processing, Wiley, New York, 1996.
[56] A.C. Rencher, Methods of multivariate analysis, in: Wiley Series in Probability and Mathematical Statistics (1995).
[57] MATLAB/C LS-SVMlab toolbox. Available from<http://www.esat.kuleuven.ac.be/sista/lssvmlab>.
[58] K. Pelckmans, J.A.K. Suykens, T. Van Gestel, J. De Brabanter, L. Lukas, B. Hamers, B. De Moor, J. Vandewalle, LS-SVMlab Toolbox User's Guide, Internal Report 02–145, ESAT-SISTA, K.U. Leuven, Leuven, Belgium, 2002.
[59] J.A. Hanley, B.J. McNeil, A method of comparing the areas under receiver operating characteristic curves derived from the same cases, Radiology 148 (1983) 839–843.
[60] G. Hagberg, A.P. Burlina, I. Mader, W. Roser, E.W. Radue, J. Seelig, In vivo proton MR spectroscopy of human gliomas: definition of metabolic coordinates for multi-dimensional classification, Magn. Reson. Med. 34 (1995) 242–252.
[61] J.C. Lindon, E. Holmes, J.K. Nicholson, Pattern recognition methods and applications in biomedical magnetic resonance, Prog. Nucl. Magn. Reson. Spectrosc. 39 (2001) 1–40.
[62] L. Vanhamme, T. Sundin, P. Van Hecke, S. Van Huffel, MR spectrocsopy quantitation: a review of time-domain methods, NMR Biomed. 14 (2001) 233–246.