

# The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification

A. Devos<sup>a,\*</sup>, A.W. Simonetti<sup>b</sup>, M. van der Graaf<sup>c</sup>, L. Lukas<sup>a</sup>, J.A.K. Suykens<sup>a</sup>,  
L. Vanhamme<sup>a</sup>, L.M.C. Buydens<sup>b</sup>, A. Heerschap<sup>c</sup>, S. Van Huffel<sup>a</sup>

<sup>a</sup> *K.U. Leuven, ESAT-SCD (SISTA), Leuven, Belgium*

<sup>b</sup> *Laboratory for Analytical Chemistry, University of Nijmegen, Nijmegen, The Netherlands*

<sup>c</sup> *Department of Radiology, University Medical Center Nijmegen, Nijmegen, The Netherlands*

Received 21 September 2004; revised 20 December 2004

Available online 22 January 2005

## Abstract

This study investigated the value of information from both magnetic resonance imaging and magnetic resonance spectroscopic imaging (MRSI) to automated discrimination of brain tumours. The influence of imaging intensities and metabolic data was tested by comparing the use of MR spectra from MRSI, MR imaging intensities, peak integration values obtained from the MR spectra and a combination of the latter two. Three classification techniques were objectively compared: linear discriminant analysis, least squares support vector machines (LS-SVM) with a linear kernel as linear techniques and LS-SVM with radial basis function kernel as a nonlinear technique. Classifiers were evaluated over 100 stratified random splittings of the dataset into training and test sets. The area under the receiver operating characteristic (ROC) curve (AUC) was used as a global performance measure on test data. In general, all techniques obtained a high performance when using peak integration values with or without MR imaging intensities. For example for low- versus high-grade tumours, low- versus high-grade gliomas and gliomas versus meningiomas, the mean test AUC was higher than 0.91, 0.94, and 0.99, respectively, when both MR imaging intensities and peak integration values were used. The use of metabolic data from MRSI significantly improved automated classification of brain tumour types compared to the use of MR imaging intensities solely.

© 2004 Elsevier Inc. All rights reserved.

**Keywords:** Brain tumours; Classification; Magnetic resonance imaging; Magnetic resonance spectroscopic imaging; Linear discriminant analysis; Least squares support vector machines

## 1. Introduction

Magnetic resonance imaging (MRI) is an important noninvasive tool for identifying the location and size of brain tumours, because it yields morphological and anatomical information about the brain tissue. However, conventional MRI has a limited specificity is rather non-specific in determining the underlying type of brain tumour and grade [1,2]. More recently developed MR techniques like diffusion-weighted MRI, perfusion-

weighted MRI, and magnetic resonance spectroscopic imaging (MRSI) are promising new techniques in the characterization of brain tumours [3,4]. Diffusion-weighted MRI visualizes the tissue structure and is useful for assessing tumour cellularity, while perfusion-weighted MRI provides measurements that reflect changes in tumour vasculature and tumour grading. MRSI or multi-voxel magnetic resonance spectroscopy (MRS) provides chemical information about metabolites present in normal and abnormal tissue [5–8]. Therefore, the differentiation of abnormal brain tissues, including brain tumours, from normal brain forms a potentially major clinical application of these new techniques. In general, diagnosis

\* Corresponding author. Fax: +32 16 321970.

E-mail address: [adevos@esat.kuleuven.ac.be](mailto:adevos@esat.kuleuven.ac.be) (A. Devos).

of brain tumours is based on the microscopic examination of tissue obtained by a biopsy, which includes risks associated with anesthesia and surgery. It would be very beneficial to the patient if the invasive biopsy could be guided or even avoided by the use of noninvasive techniques like diffusion-weighted MRI, perfusion-weighted MRI, and MRS(I). In this study, we combined the use of conventional MRI intensities and one of the new techniques, more specifically MRSI.

Several studies [9–18] have shown progress in automated pattern recognition for brain tumour classification using MRI or MRS(I). However, currently only few studies (e.g., [14,15]) have used a combination of MRI and MRSI features for classification of brain tumours. To enhance the diagnostic capabilities in clinical practice, we investigated whether the combined use of MR imaging intensities and metabolic data from MRSI could improve the discrimination between several brain tumour and normal brain tissue types. Although the radiologist also uses spatial and morphological information present in the MR images, these features were not taken into account in this study, as they are difficult to quantify. By comparing the results obtained, we evaluated the strength of both, MR imaging intensities and metabolic data from MRSI, in discriminating brain tissue types.

We considered linear as well as nonlinear classification techniques applied to several input features, such as short echo time magnitude spectra, imaging intensities, peak integration values obtained from the spectra and a combination of the latter two. The algorithms were designed to extract the most important features which were then used to classify each spectrum into the corresponding tumour type. As classification is required to be objective and user-friendly, all techniques were automated. The purpose of this paper was twofold:

- To investigate the discriminatory value of MRI intensities and metabolic data extracted from MRSI for automated brain tumour diagnosis. This analysis also provides the typical AUC values achievable for several relevant diagnostic problems of brain tumours.
- To apply and compare several classification techniques, including the investigation of the influence of the input features used.

## 2. Materials

### 2.1. Data

Data from 25 patients with a brain tumour and 4 volunteers were selected from the database developed in the framework of the EU funded INTERPRET project (IST-1999-10310) [19]. All data were provided by the acquisition center UMCN (University Medical Center

Nijmegen), Nijmegen (The Netherlands). Each case was clinically validated. The patients' tumour type was determined by a central consensus histopathological validation. For one of the 25 patients no consensus was reached. Therefore, the data from the tumour region of this patient were not used.

The dataset contained MR images as well as MR spectra, acquired and preprocessed as described in [14]. For each subject, stacked MR images of cross-sections of the whole brain at four contrasts were acquired:  $T_1$ - and  $T_2$ -weighted images, a proton density weighted image and a gadolinium enhanced (Gd-DTPA)  $T_1$ -weighted image ( $256 \times 256$ , FOV = 200 mm, slice thickness = 5 mm). The image values will further be labeled as  $T_1$ ,  $T_2$ , PD, and GD. No Gd-DTPA administration was applied to the healthy volunteers. Besides MR images, also  $^1\text{H}$  MRSI data were acquired for each subject, both with and without water suppression using a  $16 \times 16$  2D STEAM  $^1\text{H}$  MRSI sequence with acquisition parameters TR = 2000 or 2500 ms, TE = 20 ms, slice thickness 12.5 or 15 mm, FOV = 200 mm, SW = 1000 Hz, 1024 data points. The position of the MRSI slice was chosen according to the slice position of the GD image which showed the largest GD enhancement.

To ensure that image pixels from subsequent images originate from the same spatial location, the images were co-aligned [14]. All MRSI data were semi-automatically preprocessed (cf. [14]), which involved:

- Filtering of the k-space data by a Hanning filter of 50% using the LUISE software package (Siemens, Erlangen, Germany).
- Zero filling to  $32 \times 32$ , which involved an increase of the apparent spatial resolution with a factor of 2.
- Spatial 2D Fourier transformation to obtain time domain signals for each voxel.
- Correction for eddy current effects in the MR spectra using a method which prevents the occasional occurrence of eddy current correction induced artefacts [20]. This process resulted in a frequency alignment and zero order phasing of the MR spectra.
- Removal of the dominating residual water using HLSVD [21], with 12 singular values and 4.0–6.0 ppm as residual water region.
- Frequency alignment was performed semi-automatically. First, the position of the *NAA* peak (*N*-acetyl-aspartate,  $^2\text{CH}_3$ -group) in the mean spectrum of an MRSI dataset was set to 2.02 ppm. The obtained shift was used to reset each spectrum of the dataset in the time domain automatically.
- First order phase correction was also manually performed on the mean spectrum of a dataset. The obtained first order time instant was used to automatically correct each spectrum in the dataset.
- Fourier transformation was applied to the time domain data to obtain frequency spectra.

Table 1  
Number of data for each type of brain tissue (brain tumour or healthy tissue)

Label	Pathology	Number of data	Number of subjects
1	Normal from volunteers	142	4
2	Normal from patients	76	4
3	CSF	100	8
4	Gliomas, grade II	176	10
5	Gliomas, grade III	57	4
6	Gliomas, grade IV	70	7
7	Meningiomas	48	3
	Total	669	29

The first and second columns give the label of the classes (from 1 till 7) and the pathology. The third and fourth columns display the number of data and subjects for each class. Note that the total number of subjects was not simply the summation of the number of subjects per class, because for several patients data were available from brain tumour as well as from healthy tissue.

The set of validated data consisted of 10 grade II, 4 grade III, and 7 grade IV gliomas and 3 meningiomas, which gave us 4 classes of brain tumours. From each patient, data were selected from several voxels, identified as lying in the tumour area. Besides brain tumour data, also data from cerebro spinal fluid (CSF) from patients and normal brain tissue from volunteers and patients were selected. Data from all subjects with the same pathology were combined into one class, which resulted in the following 7 classes of pathologies (Table 1):

- (1) normal tissue from volunteers: normal tissue from healthy persons,
- (2) normal tissue from patients: apparently normal tissue from the contralateral half of the brain of patients,
- (3) cerebro spinal fluid (CSF): CSF from patients, where the ventricles were clearly visible and the voxels were located as far from the tumour as possible. Unfortunately, no CSF voxels could be selected from volunteers, since the MRSI slices of volunteers did not include the ventricles.
- (4) grade II gliomas: diffuse astrocytomas (90 voxels from 5 patients), oligodendrogliomas (22 voxels from 2 patients), and mixtures (64 voxels from 3 patients),
- (5) grade III gliomas: anaplastic astrocytomas (4 voxels from 1 patient), oligodendrogliomas (25 voxels from 2 patients), and mixtures (28 voxels from 1 patient),
- (6) grade IV gliomas: glioblastomas,
- (7) meningiomas.

### 3. Methods

#### 3.1. Input features

In this paper, an input pattern was either an MR spectrum, a set of quantified values from the MR spec-

trum, a set of imaging intensities or a combination of the latter two input types. This enabled us to investigate whether the combination of imaging and spectroscopic information can improve the performance for pattern recognition of brain tumours. The following input features were considered:

- Water normalized magnitude spectra (see Figs. 1 and 2). The amplitude of the water unsuppressed signal was estimated as described in [9]. Then each spectral value in the preprocessed water suppressed spectrum was divided by the resulting estimate of the intensity of the water peak. Only the spectral values in the region of interest (0.5–4.0 ppm) were used as input features.
- Metabolite amplitudes obtained by peak integration. Short echo time  $^1\text{H}$  MRSI signals are characterized by the presence of a partially unknown broad baseline underlying the resonances of the metabolites of interest, that hinders the assessment of the intensity (e.g., by peak integration) of low weight molecules. To remove this broad baseline a simple baseline correction [9,14,22] was applied as additional preprocessing step prior to the first order phase correction. This was performed as described in [14]. Amplitude estimates were then obtained from the baseline corrected frequency spectra using peak integration within a spectral range of 0.13 ppm. These selected frequency regions correspond to resonances from metabolites and lipids that are assumed to be characteristic to distinguish between tumour types [6,7,23–26]. As short echo time  $^1\text{H}$  MR in vivo spectra are characterized by substantial peak overlap and a relatively low spectral resolution at the clinical field strength of 1.5 T, a particular region might cover resonances of more than one metabolite. Such regions are [27]: *L2* (lipids at  $\approx 0.9$  ppm; 0.835–0.965), *L1* (lipids at  $\approx 1.2$ ) + *Lac* (lactate,  $^3\text{CH}_3$ -group) + *Ala* (alanine,  $^1\text{CH}_3$ -group) (1.265–1.395 ppm), *NAA* ( $^2\text{CH}_3$ -group; 1.955–2.085 ppm), *Glx* (glutamate/glutamine,  $^3\text{CH}_2$ -group; 2.135–2.265), *Cr* (creatine,  $\text{N}(\text{CH}_3)$ -group; 2.955–

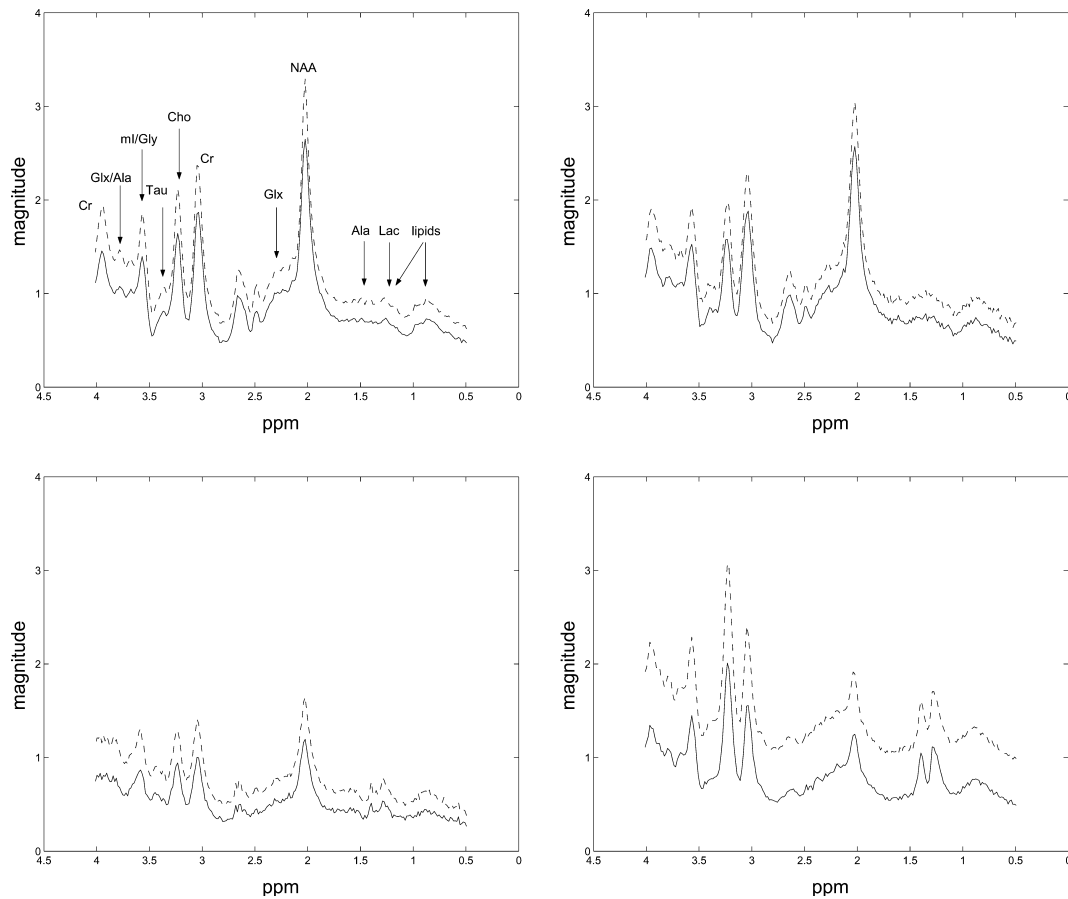


Fig. 1. Mean water normalized magnitude  $^1\text{H}$  MR spectra (TE = 20 ms) of the considered classes: class 1 (top-left), class 2 (top-right), class 3 (bottom-left), and class 4 (bottom-right) correspond to the normal tissue of volunteers, normal tissue of patients, CSF, and gliomas of grade II. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

3.095), *Cho* (choline,  $\text{N}(\text{CH}_3)_3$ -group; 3.135–3.265 ppm), *Tau* (taurine,  $^1\text{CH}_2$ -group; 3.375–3.505 ppm), *mI* (myo-Inositol,  $^1\text{CH}$ -,  $^3\text{CH}$ -,  $^4\text{CH}$ - and  $^6\text{CH}$ -group) + *Gly* (glycine,  $^2\text{CH}_2$ -group) (3.495–3.625 ppm), *Glx* + *Ala* ( $^2\text{CH}$ -groups; 3.685–3.815 ppm), and *Cr* ( $^2\text{CH}_2$ -group; 3.885–4.015 ppm). The resulting peak integration values were then water normalized as described above.

- Imaging intensities (see Figs. 3 and 4). For each of the images ( $T_1$ ,  $T_2$ , PD, and GD), an image value was extracted, producing four additional variables. To obtain the same spatial resolution of both MRI and MRSI data, the resolution of the MRI data was lowered to that of the MRSI grid by averaging the image pixels within each spectroscopic voxel. Each intensity value was divided by the highest intensity in the corresponding downsampled image and scaled to the same range as the spectral data. No GD information was available for the volunteers. Instead the  $T_1$ -image was used, under the assumption that no GD enhancement would occur in the brain tissue of volunteers. However, this is an approximation as Gd-DTPA typically causes an increase in intensity in the blood vessels.

Selection of the data used was based on the visual inspection of the low and high resolution images and MR spectra. To obtain data for a specific histopathological class, the following procedure was used for each patient within this class. The four MR images were plotted together with a segmented image in which voxels were clustered using a model-based clustering algorithm [28] as described in [29]. The clustering provides an objective segmentation based on similarities obtained from the MR images as well as MR spectroscopic features, and thus is considered to be helpful in the selection of voxels. For each pathology, only voxels were included in the dataset if their corresponding MR spectra were found to be typical for that pathology by an expert in MR spectroscopy. Since tumours are known to be heterogeneous, this approach was considered better than for example taking all spectra from one segment. Neighbouring voxels were not selected to avoid too much mutual correlation and as such to ensure that samples were as independent as possible. Although the method of class selection is subjective, we think it is appropriate since in tumour diagnosis a “ground truth” is not available and the number of patients for each specific class is low.

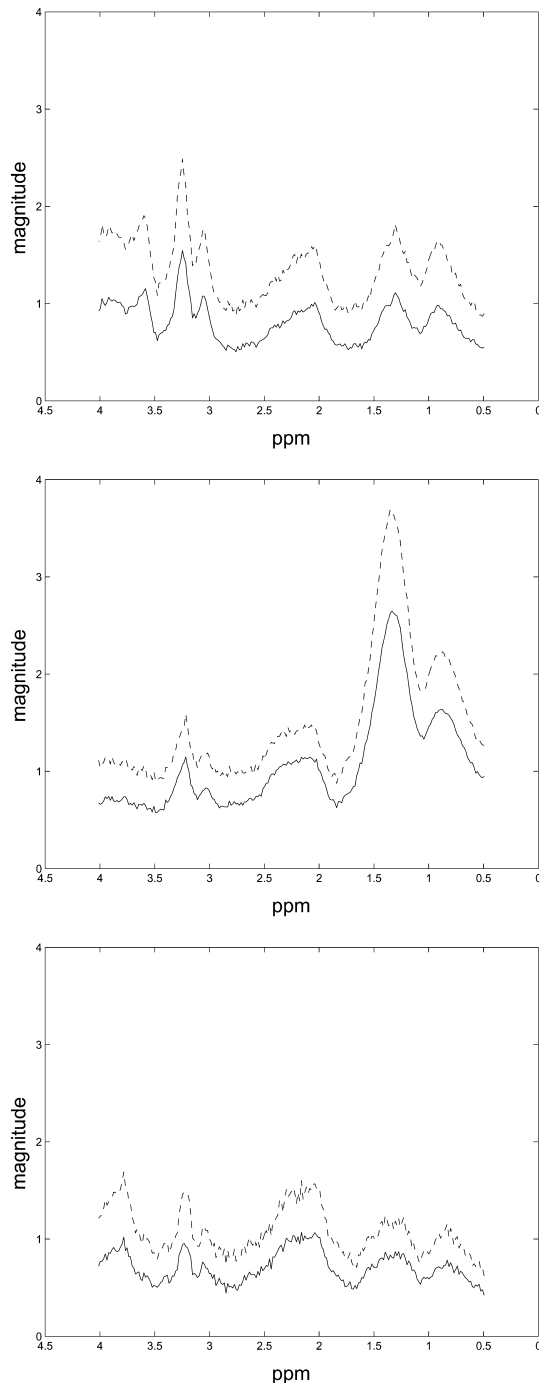


Fig. 2. Mean water normalized magnitude  $^1\text{H}$  MR spectra (TE = 20 ms) of the considered classes: class 5 (top), class 6 (middle), and class 7 (bottom) correspond to gliomas of grade III, grade IV, and meningiomas. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

### 3.2. Experimental approach

Binary classification was performed by linear discriminant analysis (LDA) [30,31], least squares support vector machines (LS-SVM) [32,33] with linear or radial basis function (RBF) kernel, which were also applied

in two previous extensive studies for classification based on long [12] and short echo time  $^1\text{H}$  MRS [9].

To perform meaningful data-analysis in a high dimensional space, a sufficiently large amount of training data is required. This limitation can be overcome, e.g., by dimensionality reduction, which decreases the amount of complexity and risk of overfitting and also simplifies the calculation. In fact, peak integration (Section 3.1) is a feature extraction method that reduces the input dimension and therefore uses prior knowledge about the most discriminatory features in the spectrum. As the input dimension is already small when using imaging intensities (dimension 4), peak integration values (dimension 10) or the combination of both (dimension 14), dimensionality reduction is only required when using magnitude spectra (dimension 231).

Especially LDA suffers from this dimensionality problem, while LS-SVM is able to work without any prior dimensionality reduction, thanks to the primal-dual aspects of the model, even in case of a relatively low number of input data. Therefore, the use of principal component analysis (PCA) as a feature extraction technique was only needed prior to LDA. The 231 given spectral variables were reduced by PCA that retained those components that account for a larger variance than the average over all individual components [34]. This strategy was different to the one taken in [9,12], in which only the largest components were selected that explain a 75% of total variance. In [9,12] it was unfeasible to retain more components due to rank deficiency problems related to a too small number of training samples. Nevertheless, if feasible, it is more appropriate to retain a certain number of components, as performed by the strategy taken in this study.

In certain problems, nonlinear techniques can improve the classification performance [33]. Therefore, in addition to the use of linear kernels in LS-SVM classifiers, we also applied LS-SVM classifiers with RBF kernels. All input patterns were classified using KULeuven's LS-SVMlab MATLAB/C toolbox [33,35,36] for LS-SVM classification with both linear and RBF kernels. Linear as well as nonlinear classifiers were applied automatically, including feature selection, (hyper-) parameter estimation, training and testing.

LS-SVM classifiers require the tuning of a set of hyperparameters to achieve a high level of performance. This tuning was performed in the same way as described in [12]. The experiments consisted of the following steps, similar as in [9,12]:

- (1) divide the dataset in a training set (2/3 of the data) and a test set (remainder),
- (2) train the classifiers using the training set,
- (3) evaluate the performance using the test set.

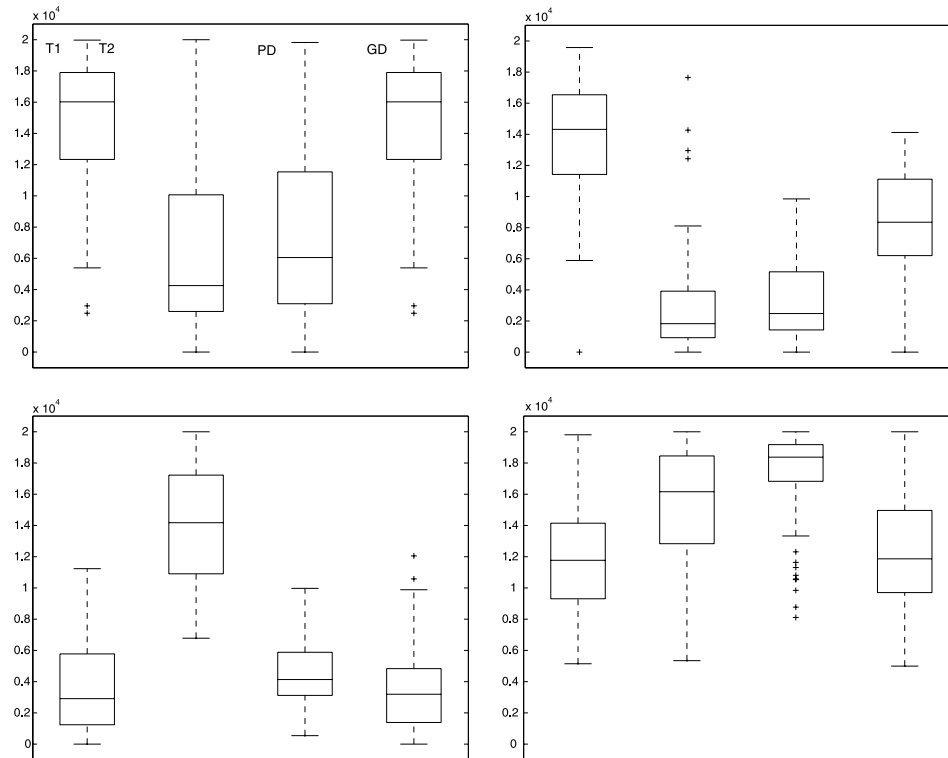


Fig. 3. Boxplots of the imaging intensities of the considered classes: class 1 (top-left), class 2 (top-right), class 3 (bottom-left), and class 4 (bottom-right) correspond to the normal tissue of volunteers, normal tissue of patients, CSF, and gliomas of grade II.

Stratified random sampling was used while dividing the dataset to preserve the proportion of the classes. The aforementioned 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. The test performance was measured by the mean AUC and its pooled standard error calculated from 100 randomizations. As discussed by Obuchowski [37], the area under the receiver operating characteristic (ROC) curve is a good summary measure of the test accuracy. The results were tested for significant differences by the  $z$  test [38], applied as in [12].

#### 4. Results

We evaluated the following binary classifications of brain tissue types:

- Healthy versus tumour tissue (classes 1, 2 and 3; 318 data versus 4, 5, 6, and 7; 351 data; Table 2). The resulting mean AUC was for all techniques and input patterns higher than 0.95. The performance based on imaging intensities alone was significantly lower with respect to using peak integration values (e.g., when using LDA,  $p < 0.01$ ) or the combination of imaging intensities and peak integration values (e.g., when using LDA,  $p < 0.001$ ). Also using magnitude spectra

was significantly worse than using the combination of imaging intensities and peak integration values for LDA and LS-SVM with a linear kernel ( $p < 0.05$ ). In addition, using solely imaging intensities, LDA and LS-SVM with a linear kernel reached a significantly lower performance than LS-SVM with an RBF kernel (both  $p < 0.05$ ).

- Low- versus high-grade tumours (classes 4 and 7; 224 data versus 5 and 6; 127 data; Table 3). Classification based on imaging intensities alone was poor, while discrimination based on peak integration values was significantly better, compared to classification using imaging intensities alone (with,  $p < 10^{-8}$ ,  $< 10^{-8}$ , and  $< 0.001$ , respectively, in case of LDA, LS-SVM lin, and LS-SVM RBF). Also the performances obtained when using magnitude spectra and the combination of imaging intensities and peak integration values were significantly higher than those obtained using imaging intensities alone. LS-SVM RBF achieved a significantly higher performance with respect to LDA and LS-SVM lin, when using imaging intensities ( $p < 0.001$ ), peak integration ( $p < 0.05$ ) and the combination of imaging intensities and peak integration values ( $p < 0.001$ ). Based on magnitude spectra, LS-SVM RBF was significantly better than LDA using 9 PCs ( $p < 0.05$ ). Based on imaging intensities, peak integration, and their combination, LS-SVM RBF also achieved a significantly higher performance than LDA and LS-SVM with a linear kernel.

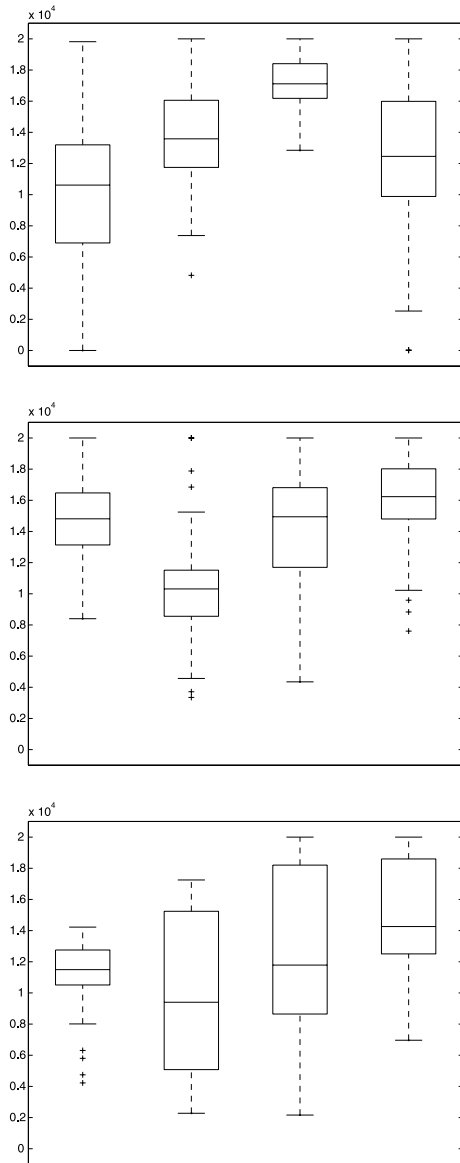


Fig. 4. Boxplots of the imaging intensities of the considered classes: class 5 (top), class 6 (middle), and class 7 (bottom) correspond to gliomas of grade III, grade IV, and meningiomas.

- Low- versus high-grade gliomas (class 4; 176 data versus 5 and 6; 127 data; Table 4). The classification problem resembles that of low- versus high-grade tumours, but reached in general a slightly higher performance.

This might be partially due to the lower heterogeneity in the two compared classes, while in low- versus high-grade tumours also nongliomas (namely meningiomas) were included. Once more, peak integration and the combination of imaging intensities and peak integration provided a significantly better performance than using imaging intensities alone. For example, when comparing the use of peak integration with imaging intensities we obtained  $p < 0.001$  (LDA, LS-SVM lin) and  $p < 0.01$  (LS-SVM RBF) and the significance was even stronger for the combination of imaging intensities and peak integration values. For this problem, classification with LS-SVMs using magnitude spectra also reached a very high mean AUC, for which  $p < 0.001$  (LDA),  $p < 0.00001$  (LS-SVM lin), and  $< 0.01$  (LS-SVM RBF) with respect to imaging intensities. From Figs. 3 and 4 we remark that the imaging intensities for class 4 indeed highly resemble those of classes 5 and 6, which explains the low performance based on the imaging intensities. LS-SVMs achieved a significantly better result than LDA when using magnitude spectra, while LS-SVM RBF provided a significantly higher AUC than LDA and LS-SVM lin when using imaging intensities.

- Gliomas versus meningiomas (classes 4, 5, and 6; 303 data versus class 7; 48 data; Table 5). Note that the same data were used as in low- versus high-grade tumours, but now gliomas were differentiated from meningiomas. At least a mean AUC of 0.99 was reached when based on peak integration or the combination of imaging intensities and peak integration (AUC  $> 0.85$  when using imaging intensities alone). For LDA and LS-SVM lin, using peak integration, combined with imaging intensities or not, gave a significantly better result than using imaging intensities ( $p < 0.001$  for all cases). However, no significant differences were found for LS-SVM RBF with respect to the input type used.
- Grade II versus grade III gliomas (class 4; 176 data versus class 5; 57 data; Table 6). With respect to the significant influence of the input features, we observed the same as for low- versus high-grade gliomas, except when using LS-SVM RBF. In case of LS-SVM RBF, only the combination of imaging intensities and peak integration values gave a signifi-

Table 2

Average test performance for classification of healthy versus tumour tissue from 100 runs of stratified random splittings

Classifier	MRI	Peak integration	MRI/peak integration	Magnitude spectra
LDA	$0.9569 \pm 0.0128$	$0.9912 \pm 0.0042$	$0.9991 \pm 0.0009$	$0.9828 \pm 0.0066_{(10;89.7\%)}$
LS-SVM lin	$0.9570 \pm 0.0128$	$0.9921 \pm 0.0039$	$0.9991 \pm 0.0008$	$0.9755 \pm 0.0097$
LS-SVM RBF	$0.9858 \pm 0.0065$	$0.9974 \pm 0.0036$	$0.9998 \pm 0.0003$	$0.9851 \pm 0.0076$

MRI, imaging intensities; peak integration, peak integration values; MRI/peak integration, imaging intensities and peak integration values; and magnitude spectra, water normalized magnitude spectra. Performance measure for each classifier are the mean AUC and its pooled standard error. When using magnitude spectra, PCA was applied prior to LDA. The number of principal components that was given as input to LDA and the amount of variance explained by those components is mentioned between brackets (row LDA, column magnitude spectra).

Table 3  
Classification of low- versus high-grade tumours

Classifier	MRI	Peak integration	MRI/peak integration	Magnitude spectra
LDA	0.6239 ± 0.0543	0.9210 ± 0.0243	0.9195 ± 0.0255	0.9193 ± 0.0265 <sub>(9.92.8%)</sub>
LS-SVM lin	0.6239 ± 0.0543	0.9328 ± 0.0220	0.9260 ± 0.0240	0.9573 ± 0.0189
LS-SVM RBF	0.8469 ± 0.0385	0.9827 ± 0.0124	0.9918 ± 0.0072	0.9797 ± 0.0161

For further explanation we refer to Table 2.

Table 4  
Classification of low- versus high-grade gliomas

Classifier	MRI	Peak integration	MRI/peak integration	Magnitude spectra
LDA	0.7429 ± 0.0517	0.9452 ± 0.0223	0.9563 ± 0.0203	0.9339 ± 0.0240 <sub>(9.93.4%)</sub>
LS-SVM lin	0.7431 ± 0.0517	0.9453 ± 0.0222	0.9589 ± 0.0193	0.9809 ± 0.0114
LS-SVM RBF	0.8774 ± 0.0354	0.9774 ± 0.0158	0.9920 ± 0.0084	0.9896 ± 0.0081

For further explanation we refer to Table 2.

Table 5  
Classification of gliomas versus meningiomas

Classifier	MRI	Peak integration	MRI/peak integration	Magnitude spectra
LDA	0.8593 ± 0.0406	0.9945 ± 0.0056	0.9961 ± 0.0040	0.9890 ± 0.0083 <sub>(9.92.8%)</sub>
LS-SVM lin	0.8590 ± 0.0407	0.9947 ± 0.0055	0.9964 ± 0.0039	0.9889 ± 0.0092
LS-SVM RBF	0.9520 ± 0.0303	0.9985 ± 0.0025	0.9989 ± 0.0019	0.9912 ± 0.0092

For further explanation we refer to Table 2.

Table 6  
Classification of grade II versus grade III gliomas

Classifier	MRI	Peak integration	MRI/peak integration	Magnitude spectra
LDA	0.7799 ± 0.0689	0.9278 ± 0.0310	0.9480 ± 0.0313	0.9339 ± 0.0274 <sub>(9.93.1%)</sub>
LS-SVM lin	0.7799 ± 0.0688	0.9290 ± 0.0306	0.9488 ± 0.0308	0.9669 ± 0.0196
LS-SVM RBF	0.9018 ± 0.0410	0.9706 ± 0.0217	0.9907 ± 0.0098	0.9861 ± 0.0141

For further explanation we refer to Table 2.

cantly higher mean AUC than imaging intensities alone ( $p < 0.05$ ). Linear classification techniques gave a significantly lower performance than LS-SVM RBF when using imaging intensities ( $p < 0.05$ ).

## 5. Discussion

### 5.1. Classification techniques

Tables 2–6 show that all classification techniques performed very well, especially based on the peak integration values and the combination with imaging intensities. In general, LDA is shown to be competitive with linear LS-SVM classifiers, even when based on the first principal components of the magnitude spectra. Using all PCs that explain more variance than the average, LDA based on magnitude spectra reached a similar performance as based on peak integration or the combination of imaging intensities and peak integration values.

Nevertheless, in several cases a significant difference was found, e.g., for the discrimination of low- and high-grade gliomas, based on magnitude spectra both kernel-based techniques reached a significantly higher performance than PCA/LDA. However, not only the classical linear LDA technique, but also the kernel-based linear LS-SVM did not always reach the performance of the nonlinear LS-SVM. This occurred for the problem low- versus high-grade tumours based on peak integration and the combination of imaging intensities and peak integration.

For several problems, the unbalanced situation (e.g., gliomas versus meningiomas) or the relatively small number of data available forms a limitation for training (e.g., grade II versus grade III gliomas). Therefore, the discrimination boundary might strongly correlate with the training set. Especially LDA requires a significant amount of data to be able to draw a linear separating line between overlapping classes. Kernel-based techniques are less sensitive to the amount of data and the input dimension and are able to detect automatically important characteristics independently of the input



pattern. Hence, these techniques are able to obtain a high performance even without any prior dimensionality reduction, although dimensionality reduction may further improve the results.

### 5.2. Imaging intensities versus metabolic data from MRSI

Although MRI is an established technique for the characterization of brain tumours, it has a few limitations that produce uncertainty for an accurate assessment of the presence and extent of the tumour. In practice, the contrast-enhancing lesion on an MR image is often much smaller than the region of abnormal metabolism [39,40]. From our results we observed that, for some problems the use of imaging intensities alone reached a significantly lower performance, e.g., for the discrimination between low- and high-grade tumours. This is in correspondence to the conclusion of [39–41], that the MR imaging intensities are unable to fully explain the metabolic heterogeneity of brain tumours.

Scaling of the MR data is necessary to correct for effects independent of the tissue characteristics. As a result of the scaling procedure, MR imaging intensities from different subjects or acquired under different conditions, should be more comparable. Although the applied scaling method might not be fully appropriate, the procedure yielded good classification results for healthy versus tumour tissue and gliomas versus meningiomas. Scaling with respect to healthy tissue might be an alternative method that possibly could improve results, but—in contrast to the applied method—would involve processing that is difficult to automate.

The obtained test performances based on metabolic data are in agreement with our previous studies on short and long echo time spectra reported in [9,12]. The results were also similar with respect to those of other MRS(I) studies [10,13–15,17], although these authors used other performance measures than the test AUC. This confirms the statement that spectroscopy is able to add valuable information about the metabolic status of brain tumours, which is in agreement with a few clinical combined MRSI/MRI studies [39–41]. As such multivoxel MRS could be very helpful for the diagnosis of brain tumours in combination with conventional MRI.

However, automated discrimination between different tumour types is still difficult, partially due to the fact that only intensity values are used and no anatomical (e.g., the location and homogeneity of the tumour) or clinical information (e.g., complaints of the patient) is included in the input features. To make an accurate diagnosis a neuroradiologist exploits such clinical features—in contradiction to most classification studies—as they could be very specific for certain types of tumour. For example, meningiomas are a type of tumour of the meninges and are not really a type of brain tumour, while gliomas are intracerebral tumours that start

in glial cells. In an MRI classification study [18] also several other diagnostic factors were found to be important for prediction of brain glioma like age, oedema, blood supply, calcification and haemorrhage. Hence, adding important anatomical and clinical information as input features is expected to further improve automated diagnosis of brain tumours.

### 5.3. Several classification problems

If a classification technique is developed for diagnostic purposes, then the technique should be able to distinguish healthy tissue from tumour tissue. Several metabolic differences between normal cells and various tumour types are reflected in MR spectra: the *NAA* and *Cr* levels are lower in tumour spectra, while the *Cho* level is higher. The applied techniques were able to extract and exploit this differing pattern available in imaging intensities and MR spectra. Healthy brain and tumour tissue could be distinguished almost perfectly from each other.

Classification of low- versus high-grade gliomas reached a high mean AUC ( $>0.93$ ) using any input (except imaging intensities or magnitude spectra with a low number of PCs given to LDA). A similar observation could be made for gliomas versus meningiomas (AUC  $> 0.98$ ), low- versus high-grade tumours (AUC  $> 0.91$ ) and grade II versus grade III gliomas (AUC  $> 0.92$ ).

Using the imaging intensities alone for discriminating gliomas of grade II and grade III, a linear classifier performed poorly. This might partially be due to the unbalanced distribution of the data (176 data of class 4, and 57 of class 5). A significantly higher performance was obtained by using peak integration or combining the imaging intensities and peak integration. Also for the discrimination between gliomas and meningiomas, the use of peak integration and the combination of peak integration and imaging intensities gave a significant improvement with respect to using imaging intensities alone. This was not the case when using LS-SVM with an RBF kernel, since the performance was already high.

## 6. Conclusions

This study investigated the use of several classification techniques for classification of brain tumours using MR imaging intensities and metabolic data from MRSI. The nonlinear technique LS-SVM with an RBF kernel reached in several specific problems a significantly better performance than the linear techniques. Although linear classifiers also performed well, based on these data this indicates that a few diagnostic problems seem to have a nonlinear behaviour.

From current studies it is clear that  $^1\text{H}$  MRSI is an important adjunct to the clinical imaging techniques

for noninvasive diagnosis of viable tumours. For most problems, binary classification based on imaging intensities and metabolic information from MRSI is very well possible. The combined use of MR imaging intensities and metabolic information significantly increased the performance with respect to imaging intensities alone. Also with respect to metabolic data alone, the combination of imaging intensities and metabolic data reached a higher, although not significantly, performance. The results of this study strengthen the statement that imaging intensities and metabolic data provide complementary information for the accurate discrimination between several brain tissue types. Therefore, we motivate the integration of MRSI into a standard clinical examination which is performed for the diagnosis of brain tumours. However, to enhance the quality of automated diagnosis, it would be beneficial that classification datasets also include several other relevant anatomical and clinical parameters. This would enable pattern recognition people to develop and test classifiers that use as much information as relevant.

### 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 Federal Science Policy Office (IUAP Phase IV-02 and IUAP Phase V-22), the EU funded projects BIOPATTERN (EU network of excellence; Contract No. 508803) and eTUMOUR (FP6 integrated project; Contract No. 503094), the Concerted Action Project MEFISTO and AMBIORICS 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 funded by a Ph.D grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). LVH was 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.

### References

- [1] B.L. Dean, B.P. Drayer, C.R. Bird, R.A. Flom, J.A. Hodak, S.W. Coons, R.G. Carey, Gliomas: classification with MR imaging, *Radiology* 174 (1990) 411–415.
- [2] F.I.V. Earnest, P.J. Kelly, B.W. Scheithauer, B.A. Kall, T.L. Cascino, R.L. Ehman, G.S. Forbes, P.L. Axley, Cerebral astrocytomas: histopathologic correlation of MR and CT contrast enhancement with stereotactic biopsy, *Radiology* 166 (1988) 823–827.
- [3] S.J. Nelson, S. Cha, Imaging glioblastoma multiforme, *Cancer J.* 9 (2) (2003) 134–145.
- [4] J. Rees, Advances in magnetic resonance imaging of brain tumours, *Curr. Opin. Neurol.* 16 (2003) 643–650.
- [5] S.J. Nelson, Multivoxel magnetic resonance spectroscopy of brain tumors, *Mol. Cancer Ther.* 2 (2003) 497–507.
- [6] X. Leclerc, T.A.G.M. Huisman, A.G. Sorensen, The potential of proton magnetic resonance spectroscopy ( $^1\text{H}$ ) in the diagnosis and management of patients with brain tumors, *Curr. Opin. Oncol.* 14 (2002) 292–298.
- [7] S.K. Mukherji (Ed.), *Clinical Applications of Magnetic Resonance Spectroscopy*, Wiley-Liss, 1998.
- [8] S.J. Nelson, E. Graves, A. Pirzkall, X. Li, A.A. Chan, D.B. Vigneron, T.R. McKnight, In vivo molecular imaging for planning radiation therapy of gliomas: an application of  $^1\text{H}$  MRSI, *J. Magn. Reson. Imaging* 16 (2002) 464–476.
- [9] A. Devos, L. Lukas, J.A.K. Suykens, L. Vanhamme, A.R. Tate, F.A. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, C. Arús, S. Van Huffel, Classification of brain tumours using short echo time  $^1\text{H}$  MR spectra, *J. Magn. Reson.* 170 (1) (2004) 164–175.
- [10] S. Herminghaus, T. Dierks, U. Pilatus, W. Möller-Hartmann, J. Witsack, 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.
- [11] Y. Huang, P.J.G. Lisboa, W. El-Deredy, Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection, *Statist. Med.* 22 (2003) 147–164.
- [12] 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. Arús, S. Van Huffel, Brain tumour classification based on long echo proton MRS signals, *Artif. Intell. Med.* 31 (1) (2004) 73–89.
- [13] 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.
- [14] A.W. Simonetti, W.J. Melssen, M. van der Graaf, G.J. Postma, 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.
- [15] 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.
- [16] 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  $^1\text{H}$  MRS spectra from brain tumours, *NMR Biomed.* 11 (1998) 177–191.
- [17] A.R. Tate, C. Majós, A. Moreno, F.A. Howe, J.R. Griffiths, C. Arús, Automated classification of short echo time in vivo  $^1\text{H}$  brain tumor spectra: a multicenter study, *Magn. Reson. Med.* 49 (2003) 29–36.
- [18] 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.
- [19] International network for Pattern Recognition of Tumours Using Magnetic Resonance. Available from: <http://carbon.uab.es/INTERPRET/>.
- [20] A.W. Simonetti, W.J. Melssen, M. van der Graaf, A. Heerschap, L.M.C. Buydens, Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction, *J. Magn. Reson.* 159 (2002) 151–157.
- [21] W.W.F. Pijnappel, A. van den Boogaart, R. de Beer, D. van Ormondt, SVD-based quantification of magnetic resonance signals, *J. Magn. Reson.* 97 (1992) 122–134.

- [22] 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.
- [23] I.C.P. Smith, L.C. Stewart, Magnetic resonance spectroscopy in medicine: clinical impact, *Prog. Nucleic Mag. Reson. Spectrosc.* 40 (2002) 1–34.
- [24] 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  $^1\text{H}$  magnetic resonance spectroscopy, *Magn. Reson. Med.* 49 (2003) 223–232.
- [25] 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.
- [26] 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 ( $^1\text{H}$  MRS) to clinical brain tumour diagnosis, *Br. J. Neurosurg.* 16 (4) (2002) 329–334.
- [27] V. Govindaraju, K. Young, A.A. Maudsley, Proton NMR chemical shifts and coupling constants for brain metabolites, *NMR Biomed.* 13 (2000) 129–153.
- [28] C. Fraley, A.E. Raftery, MCLUST: software for model-based cluster analysis, *J. Classif.* 16 (2) (1999) 297–306.
- [29] R. Wehrens, A.W. Simonetti, L.M.C. Buydens, Mixture modeling of medical magnetic resonance data, *J. Chemometr.* 16 (6) (2002) 274–282.
- [30] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, second ed., Wiley, New York, 2001.
- [31] B.D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, 1996.
- [32] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, *Neur. Proc. Lett.* 9 (3) (1999) 293–300.
- [33] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific Publishing, Singapore, 2002.
- [34] A.C. Rencher, *Methods of multivariate analysis*, Wiley series in Probability and Mathematical Statistics (1995).
- [35] LS-SVMlab MATLAB/C toolbox. Available from: <<http://www.esat.kuleuven.ac.be/sista/lssvmlab>>.
- [36] 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.
- [37] N.A. Obuchowski, Receiver operating characteristic curves and their use in radiology, *Radiology* 229 (2003) 3–8.
- [38] 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.
- [39] D.J. Manton, M. Lowry, C. Rowland-Hill, D. Crooks, B. Mathew, L.W. Turnbull, Combined proton MR spectroscopy and dynamic contrast enhanced MR imaging of human intracranial tumours in vivo, *NMR Biomed.* 13 (2000) 449–459.
- [40] S.J. Nelson, D.B. Vigneron, W.P. Dillon, Serial evaluation of patients with brain tumors using volume MRI and 3D  $^1\text{H}$  MRSI, *NMR Biomed.* 12 (1999) 123–138.
- [41] D. Vigneron, A. Bollen, M. McDermott, L. Wald, M. Day, S. Moyher-Moworolski, R. Henry, S. Chang, M. Berger, W. Dillon, S. Nelson, Three-dimensional magnetic resonance spectroscopic imaging of histologically confirmed brain tumors, *Mag. Reson. Imaging* 19 (2001) 89–101.