



Brain tumor classification based on long echo proton MRS signals

L. Lukas^a, A. Devos^{a,*}, J.A.K. Suykens^a, L. Vanhamme^a, F.A. Howe^b,
C. Majós^c, A. Moreno-Torres^d, M. Van Der Graaf^e, A.R. Tate^b,
C. Arús^f, S. Van Huffel^a

^aSCD-SISTA, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Heverlee (Leuven), Belgium

^bCRC Biomedical Magnetic Resonance Research Group, Department of Biochemistry and Immunology, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK

^cInstitut de Diagnòstic per la Imatge (IDI), CSU de Bellvitge, Autovia de Castelldefels km 2.7, L'Hospitalet de Llobregat, 08907 Barcelona, Spain

^dCentre Diagnòstic Pedralbes, Unitat Esplugues, C/Josep Anselm Clavé 100, 08950 Esplugues de Llobregat, Spain

^eDepartment of Radiology, University Medical Center Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands

^fDepartament de Bioquímica i Biologia Molecular, Unitat de Ciències, Edifici Cs, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain

Received 28 April 2003; received in revised form 7 August 2003; accepted 17 January 2004

KEYWORDS

Brain tumors;
Classification;
Magnetic resonance spectroscopy (MRS);
Linear discriminant analysis (LDA);
Support vector machine (SVM);
Least squares support vector machine (LS-SVM)

Summary There has been a growing research interest in brain tumor classification based on proton magnetic resonance spectroscopy (¹H MRS) signals. Four research centers within the EU funded INTERPRET project have acquired a significant number of long echo ¹H MRS signals for brain tumor classification. In this paper, we present an objective comparison of several classification techniques applied to the discrimination of four types of brain tumors: *meningiomas*, *glioblastomas*, *astrocytomas grade II* and *metastases*. Linear and non-linear classifiers are compared: linear discriminant analysis (LDA), support vector machines (SVM) and least squares SVM (LS-SVM) with a linear kernel as linear techniques and LS-SVM with a radial basis function (RBF) kernel as a non-linear technique. Kernel-based methods can perform well in processing high dimensional data. This motivates the inclusion of SVM and LS-SVM in this study. The analysis includes optimal input variable selection, (hyper-) parameter estimation, followed by performance evaluation. The classification performance is evaluated over 200 stratified random samplings of the dataset into training and test sets. Receiver operating characteristic (ROC) curve analysis measures the performance of binary classification, while for multiclass classification, we consider the accuracy as performance measure. Based on the complete magnitude spectra, automated binary classifiers are able to reach an area under the ROC curve (AUC) of more than 0.9 except for the hard case glioblastomas versus metastases. Although, based on the available long

* Corresponding author. Tel.: +32-16-321-926; fax: +32-16-321-970.
E-mail address: andy.devos@esat.kuleuven.ac.be (A. Devos).

echo ^1H MRS data, we did not find any statistically significant difference between the performances of LDA and the kernel-based methods, the latter have the strength that no dimensionality reduction is required to obtain such a high performance.

© 2004 Elsevier B.V. All rights reserved.

1. Introduction

Brain tumors are the second leading cause of cancer death in children under 15 years and young adults up to the age of 34. These tumors are also the second fastest growing cause of cancer death among humans older than 65 years [1]. Early detection and correct treatment based on accurate diagnosis are important steps to improve disease outcome.

Currently, magnetic resonance spectroscopy (MRS) in combination with magnetic resonance imaging (MRI) are important tools to identify the location, size and type of brain tumors. So far, MRS has been proven to be an accurate non-invasive technique which can give detailed chemical information of metabolites present in the suspected brain tumors [2,3]. Under physiological conditions, several important metabolites are observed: *NAA* (*N*-acetyl aspartate) as a neuronal marker; *Cho* (choline-containing compounds) as membrane precursors and degradation products; *Cr* (total creatine) as a measure of the energy status; *glucose*; and *mi* (*myo*-inositol). Under pathological conditions, the presence of some resonances can be indicative: a doublet of *Lac* (lactate); *lipids* and/or *some low molecular weight proteins* which might occur even under normal conditions; *Ace* (acetate) and certain amino acids, such as *Ala* (alanine), *Gln* (glutamine), *Glu* (glutamate) and *Gly* (glycine).

In comparison to in vitro spectroscopy, in vivo spectroscopy signals are more difficult to analyze because of their broader resonances, strongly overlapping peaks, lower signal-to-noise ratio and higher number of artifacts. Cousins [4] discusses the influence of the echo time TE on the spectral pattern of an MRS signal. The above-mentioned metabolites can be detected in short echo ^1H MRS signals. However, short echo ^1H MRS signals are more difficult to analyze than long echo ^1H MRS signals due to a higher number of overlapping peaks, a stronger baseline and a higher sensitivity to artifacts. In comparison, long echo ^1H MRS signals are poorer in information but they allow a more reliable analysis and testing of classification methods.

Many studies have been performed to classify MRS signals. Lindon et al. [5] overviewed pattern recognition methods and their applications in

biomedical magnetic resonance. Several studies [6–11] also show some progress in automated pattern recognition for brain tumor classification based on MR data. These studies are either based on MRI (e.g. [11]), MRI combined with MR spectroscopic imaging (MRSI) (e.g. [7]), long echo (e.g. [6,8,9]) or short echo ^1H MRS (e.g. [10]), but most of the papers investigate only one classification method and restrict data collection to one center only. As performance measure either the training performance is considered or test performance on a specifically selected set. In our study we measure the binary classification performance based on the receiver operating characteristic (ROC) curve analysis over 200 stratified random samplings of training and test set. ROC analysis is commonly used in medicine [12] 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 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 like the accuracy, e.g. used in [11], which is only based on a single cut-off point (e.g. for one specific value of the false-positive rate). Various clinical studies focus on the prediction of the malignancy of tumors, more specifically for brain gliomas (e.g. [6,11]). Thereby, they consider only two classes: low-grade and high-grade gliomas. In our study, astrocytomas of grade II and glioblastomas (also called astrocytomas of grade IV) are included, which are large subtypes of, respectively, low-grade and high-grade gliomas. Additionally, we consider two other common brain tumor types, namely metastases and meningiomas.

Moreover, this paper reports the results of a comparative study on a *multicenter* dataset of MRS signals. This dataset was developed in the framework of the EU funded INTERPRET project [13]. Several INTERPRET partners [7,9,10,14–19] have already published results for classification of brain tumors based on MR data available within the project. The papers [7,15,18] focus on the use of ^1H MRSI data, while others consider the use of short or long echo ^1H MRS. Nevertheless, most of these studies are based on a previous version of the dataset or focus on a specific technique. For example, in [10], 144 short echo ^1H MRS spectra from three contributing centers were used, originating

from three groups of brain tumors; meningiomas, low-grade astrocytomas and aggressive tumors. The latter group includes glioblastomas and metastases. Note that these groups correspond to the same four tumor groups as considered in this paper. But Tate et al. selects a specific training and test set; the data from two centers formed the training set (94 spectra) and the data from the third center were used for testing (50 spectra). Based on this specific test set an accuracy of 96% was obtained using LDA.

In this study several methods are applied on all histopathologically validated long echo ^1H MRS data from four common brain tumor types as available in the final status of the database development. We mention three additional points differing with previous classification studies within the framework of the INTERPRET project. First of all, we investigate what can be obtained as typical performance on a representative test set. Therefore, we construct 200 different combinations of training and independent test set. Second of all, the discrimination ability was judged by the AUC, which is, in contradiction to the accuracy, a global measure. Only in one other INTERPRET study [15] ROC analysis was also applied to compare two diagnostic methods for classification based on ^1H MRSI. Third of all, four different techniques are applied for classification; linear as well as non-linear techniques. We investigate binary as well as multiclass classification. Moreover, this analysis includes optimal input variable selection and (hyper-) parameter estimation.

Several classification techniques are compared in this paper. We evaluate the performance of linear discriminant analysis (LDA), support vector machines (SVMs) and the least squares version of support vector machines (LS-SVMs) in classifying brain tumors based on long echo ^1H MRS spectra. The support vector machine [20,21] is a training algorithm for learning classification and regression rules from data. It applies the idea of kernel representation from mathematical analysis, for example, using either linear, polynomial, radial basis functions (RBF) or multi-layer perceptrons (MLP) as its learning kernel. SVMs were first introduced by Vapnik in the 1960s for classification and have recently become an area of intense research owing to developments in the techniques and theory coupled with extensions to density estimation and regression. SVMs arose from statistical learning theory; the aim being to solve only the problem of interest without solving a more difficult problem as an intermediate step. SVMs are based on the structural risk minimization principle, closely related to regularization theory. This principle incorporates capacity

control to prevent overfitting and is thus a partial solution to the bias-variance trade-off dilemma.

Least squares SVM [22] uses equality constraints and solves a set of linear equations in the dual space instead of solving a quadratic programming problem as for the standard SVM. This simplifies the computations and enhances the speed considerably. There exists a link between the LS-SVM classifier formulation with the well-known Fisher discriminant analysis, namely by extending it to a high-dimensional feature space. Some parameters have to be tuned to achieve a high level performance of the (LS-)SVM, including the regularization parameter and the kernel parameter corresponding to the kernel type.

The paper is organized as follows. Section 2 explains the material and methods used for classification; description of the data and short explanation of the kernel based methods SVM and LS-SVM. Section 3 summarizes the results of binary classification using complete spectra, selected frequency regions and peak integrated values, consecutively. Afterwards, results of the multiclass classification approach are also mentioned. In Section 4, we discuss the classification performance of the classifiers, the limitations of the dataset and the influence of dimensionality reduction. Finally, Section 5 presents the conclusions.

2. Material and methods

2.1. Material

The data were provided by CDP (Centre Diagnòstic Pedralbes, Barcelona, Spain), IDI (Institut de Diagnòstic per la Imatge, Barcelona, Spain), SGHMS (St. George's Hospital Medical School, London, UK) and UMCN (University Medical Center Nijmegen, Nijmegen, The Netherlands) in the framework of the INTERPRET project. It concerns long echo ^1H MRS data, acquired both with and without water suppression using a PRESS sequence (the repetition time TR is between 1500 and 2020 ms, the echo time TE = 135 or 136 ms, the spectral width SW = 1000 or 2500 Hz, the number of datapoints is 512 or 2048) (Table 1).

Four main classes are considered, corresponding to four brain tumor types, i.e. glioblastomas, meningiomas, metastases and astrocytomas (grade II). They are labeled as class 1 (glio), class 2 (meni), class 3 (meta) and class 4 (astroll), respectively. All data have passed a quality control and validation process, which was regulated by strict rules agreed on by all INTERPRET partners. After thorough examinations, the brain tumors were histopathologically classified by three pathologists. These class

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

Center (acquisition scheme)	1	2	3	4	Total
CDP (PRESS, TE = 135 ms)	38	16	5	6	65
IDI (PRESS, TE = 136 ms)	28	27	16	6	77
SGHMS (PRESS, TE = 136 ms)	10	9	11	7	37
UMCN (PRESS, TE = 136 ms)	1	1	0	2	4
Total	77	53	32	21	183

The rows correspond to the acquisition center, while the columns mention the type of brain tumor. The acquisition scheme is a PRESS sequence and TE denotes the echo time.

assignments were based on the histological classification of tumors of the central nervous system (CNS) set up by the World Health Organization (WHO).

The raw data are acquired in the time domain at the aforementioned centers. A few preprocessing steps are carried out: frequency alignment and phase correction with Klose's method [23] and filtering of the dominating residual water peak using HSVD [24]. The initial point of the time domain signal was removed, because it was often affected by artifacts. The resulting signal is transformed to the frequency domain by a FFT. For each signal the

L2-normalized magnitude spectrum (of unit length) is considered only in the frequency region of interest (4.17–0.94 ppm), corresponding to 108 input variables. Fig. 1 depicts the mean magnitude frequency spectra of the four considered classes.

2.2. Methods

Several classification techniques can be applied to separate the given MR spectra. The techniques we apply in this paper are chosen so that we consider linear as well as non-linear methods: LDA, SVM and LS-SVM.

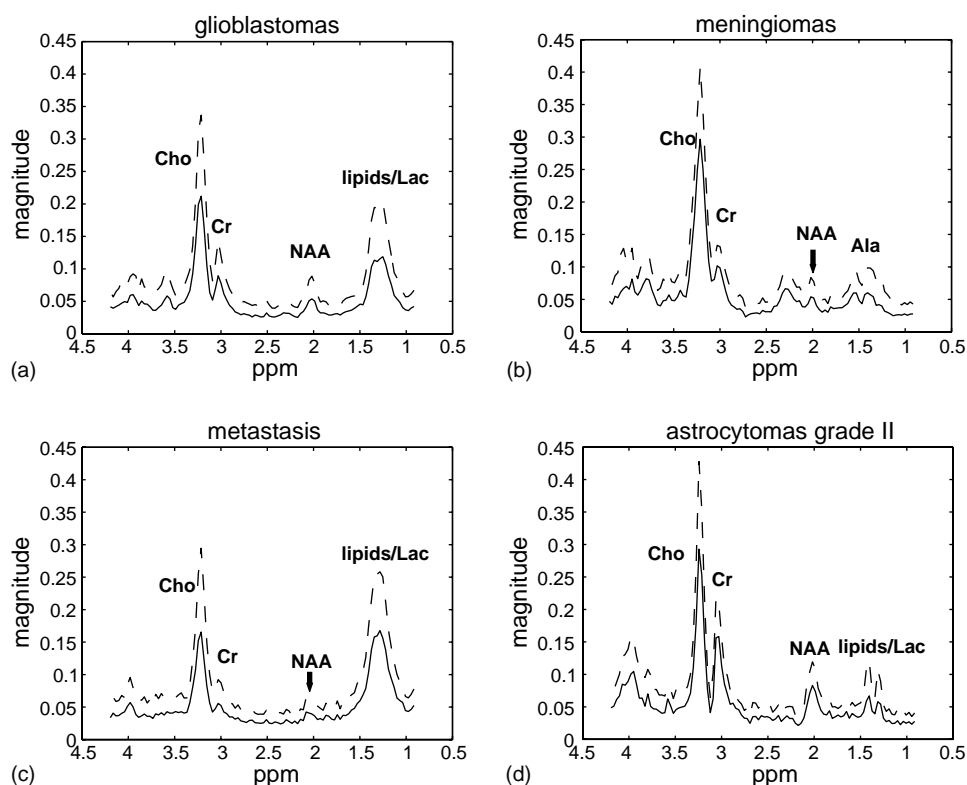


Figure 1 Mean L2-normalized magnitude frequency spectra of the four considered classes: class 1 (top-left), class 2 (top-right), class 3 (bottom-left) and class 4 (bottom-right) correspond to the glioblastomas, meningiomas, metastases and astrocytomas (grade II), respectively. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

Linear discriminant analysis [25,26] basically projects the data $\mathbf{x}_k \in \mathbb{R}^n$ from the original input space into a one-dimensional variable $z_k \in \mathbb{R}$ and makes a discrimination using this projected variable. This approach tries to maximize between-class variances and minimize the within-class variances for two given classes.

Linear principal component analysis (PCA) is applied to select the input variables. It reduces the 108 given spectral variables to a minimal set of variables which cover 75% variance of the data.

Quite often, different classes do not have equally distributed datapoints and their distributions are also overlapping among classes, which causes the problem to be linearly non-separable. Here, two kernel-based classifiers SVM and LS-SVM (briefly explained below) are assessed. SVM and LS-SVM with linear kernel can be regarded as regularized linear classifiers, while LS-SVM with RBF kernel is regarded as a regularized non-linear classifier.

A *support vector machine* [20,21] is a universal learning machine, which has become more established and performs well in many classification problems. The principles of SVM are as follows:

- (1) Consider the training samples $\{\mathbf{x}_k, \mathbf{y}_k\}_{k=1}^N$, $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{y}_k \in \{-1, +1\}$. The classifier in the primal space is defined by: $y(x) = \text{sign}[w^T \varphi(x) + b]$, $k = 1, \dots, N$, in which w is a weighting function.
- (2) The SVM performs a non-linear mapping φ of the input vectors $\mathbf{x}_k \in \mathbb{R}^n$ from the input space into a high dimensional feature space. Some kernel functions can be used for this mapping, e.g. linear, polynomial, RBF kernels.
- (3) In the feature space, an optimal linear decision rule is constructed by calculating a separating hyperplane which has the largest margin:

$$\min_{w, e_k} J(w, e_k) = \frac{1}{2} w^T w + C \sum_{k=1}^N e_k$$

$$\text{s.t. } y_k [w^T \varphi(x_k) + b] \geq 1 - e_k, \quad e_k \geq 0, \quad k = 1, \dots, N$$

in which C is a regularization constant.

- (4) This hyperplane is the solution of the following quadratic programming (QP) problem:

$$\max_{\alpha} J(\alpha) = \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k=1}^N \sum_{l=1}^N \alpha_k \alpha_l y_k y_l K(\mathbf{x}_k, \mathbf{x}_l)$$

satisfying the constraints $\sum_{k=1}^N \alpha_k y_k = 0$ and $0 \leq \alpha_k \leq C$ for $k = 1, \dots, N$ where $\{\mathbf{x}_k \in \mathbb{R}^n | k = 1, \dots, N\}$ is the training sample set, and $\{y_k \in \{-1, +1\} | k = 1, \dots, N\}$ the corresponding class labels. $K(x, x_k)$ is a symmetric kernel function in the input space which satisfies Mercer's theorem: $K(x, x_k) = \varphi(x)^T \varphi(x_k)$.

- (5) Those input vectors $\mathbf{x}_k \in \mathbb{R}^n$ with corresponding non-zero α_k are called support vectors. They are located in the boundary margin and contribute to the construction of the separating hyperplane.
- (6) Classification in the input space is calculated by mapping the separating hyperplane back into the input space (SV, set of support vectors):

$$y(x) = \text{sign} \left[\sum_{\mathbf{x}_k \in \text{SV}} \alpha_k y_k K(x, \mathbf{x}_k) + b \right].$$

Recently, a *least squares version (LS-SVM)* has been proposed [22,27], incorporating equality instead of inequality constraints as in the SVM case. This simplifies the computation of the solution, namely by solving a set of linear equations. The modifications are:

- (1) The constrained optimization problem in the primal space is reformulated as

$$\min_{w, b, e} J(w, b, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2$$

$$\text{s.t. } y_k [w^T \varphi(x_k) + b] = 1 - e_k, \quad k = 1, \dots, N$$

The conditions for optimality are $y_k [w^T \varphi(x_k) + b] - 1 + e_k = 0$, $\alpha_k = \gamma e_k$, $\sum_{k=1}^N \alpha_k y_k = 0$ and $w = \sum_{k=1}^N \alpha_k y_k \varphi(x_k)$, $k = 1, \dots, N$.

- (2) Here, non-zero support values α_k are spread over all datapoints. Each α_k value is proportional to the error of the corresponding datapoint. No sparseness property raises as in the standard SVM case. But, interestingly, in the LS-SVM case one can relate a high support value to a high contribution of the datapoint on the decision line.
- (3) Elimination of w and e from the previous equations gives

$$\left[\begin{array}{c|c} 0 & Y^T \\ \hline Y & \Omega + \gamma^{-1} I \end{array} \right] \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 1_v \end{bmatrix} \quad (1)$$

with $Y = [y_1 \dots y_N]^T$, $1_v = [1 \dots 1]^T$, $e = [e_1 \dots e_N]^T$, $\alpha = [\alpha_1 \dots \alpha_N]^T$, $(\Omega)_{kl} = y_k y_l K(\mathbf{x}_k, \mathbf{x}_l)$. This set of linear equations is easier to solve rather than the QP problem as in the standard SVM.

In certain problems, non-linear techniques could improve classification performance, especially when data are linearly non-separable. Therefore, in addition to the use of linear kernels in SVM and LS-SVM classifiers, we also apply LS-SVM classifiers with RBF kernels.

The MRS spectra were classified using Steve Gunn's MATLAB Support Vector Machines toolbox [28,29] and KULeuven's MATLAB/C LS-SVMlab toolbox [27,30,31] for LS-SVM classification with both linear and RBF kernels.

2.3. Selected frequency regions

It is well known that characteristic peaks at certain frequencies correspond to important metabolites in the brain [2,3,32–35]. These peaks might be used as discriminatory features to distinguish tumor types. In particular, when their appearance clearly differ in size and shape in between spectra of different tumor types. Instead of using complete spectra as input variables to the classifier, selection of the most explanatory input features can be used. One approach is based on selected frequency regions: therefore, the input variables within certain regions of the magnitude spectrum which are assumed to contain most of the information as input features are selected. Hence, the redundancy produced by spectral noise and artefacts in the spectrum is reduced. Characteristic metabolites can be observed in the following regions of the magnitude MRS spectrum: *Cho* and *Cr* (2.95–3.3 ppm); *NAc* (1.95–2.1 ppm); *Lac*, *Ala* and *lipid1* (1.15–1.55 ppm); *lipid2* (0.9–1.0 ppm). Note that these selected regions are based on the metabolites that are assumed to be most characteristic according to prior knowledge available from field experts participating in this study. Nevertheless, this selection is still subjective as the size of the regions could be altered or some other resonances (e.g. from metabolites with a typically lower intensity at a long echo time; ml, Gln, Gly, etc.) could also have been included.

2.4. Peak integration

Another approach to select the most explanatory input is based on *peak integration*. The amplitude of a resonance is proportional to the integral of the corresponding peak in the spectrum. However, precise estimation of the peak integrals is difficult due to several factors, including non-zero baseline, peak overlap, noise and also the discrete nature of the spectrum. Peak integration is performed here by using the trapezoidal rule. For each selected metabolite the area under the frequency peak in the magnitude spectrum is calculated. These regions cover: *Cho* (3.1–3.3 ppm); *Cr* (2.95–3.05 ppm); *NAc* (1.95–2.1 ppm); *Lac* and *Ala* (1.25–1.55 ppm); and *lipid1* (1.1–1.25 ppm).

2.5. Training and test data

2.5.1. Binary classification

Binary classification can be used to distinguish two different tumor types. Instead of using a one-against-all scheme, the classes are pairwise compared by means of a binary classifier. Consider four types of brain tumors, then six binary classifiers can be constructed to separate the following pairs:

- glioblastomas versus meningiomas,
- glioblastomas versus metastases,
- glioblastomas versus astrocytomas grade II,
- meningiomas versus metastases,
- meningiomas versus astrocytomas grade II, and
- metastases versus astrocytomas grade II.

By classifying in pairs, we obtain more information about:

- (1) the distribution of two classes and their overlap,
- (2) the balance of the data distribution of the classes, and
- (3) the performance of the classifier which can be measured using ROC analysis.

The dimension of the input features to LDA is reduced by PCA. The number of principal components is determined by the number of components that account for 75% of the total variance of the given data. Note that PCA is not used when peak integrated values are taken as input features, as peak integration already significantly reduces the dimension.

To achieve a high level of performance in SVMs, some hyperparameters must be tuned. These adjustable hyperparameters include: a regularization parameter, which determines the tradeoff between minimizing the training errors and minimizing the model complexity. In case of a RBF kernel, also a kernel parameter (the width σ) must be selected. We choose the value of hyperparameters C for SVM, γ for LS-SVM with a linear kernel and (σ, γ) for LS-SVM with a RBF kernel through leave-one-out (LOO) cross-validation, while bounding the search to avoid overfitting.

The experiment consists of the following steps:

- (1) the data are divided in a training set (2/3 of the data) and a test set (remainder) using stratified random sampling,
- (2) train the classifiers and use the test set to evaluate the performance,
- (3) the index of the misclassified spectra is noted.

This randomization is repeated 200 times to avoid bias possibly introduced by selection of a specific training and test set. In this way we try to obtain a

representative performance on the test set. ROC [12] analysis is used to evaluate the binary classifiers. The performance is then measured by the mean AUC and its pooled standard error calculated from 200 randomizations.

2.5.2. Multiclass classification

In the framework of binary classification we assume that a new MRS spectrum belongs to one of the two considered classes. Nevertheless, in medical practice, the number of possible tumor types is mostly not restricted to two types. This motivates the development of multiclass classifiers, that handle all classes in one construction, which extends the classifiers mentioned in the previous section. With this setup, the classifier is expected to classify a certain spectrum as one of the four tumor types.

Various pattern recognition techniques have been tried to distinguish MRS spectra of class 1 (glio) and class 3 (meta), but none gives satisfactory results [9,36]. Alternatively, as was suggested by Tate et al. in [10], we can merge these two classes, obtaining a new group called class 5, containing only aggressive (aggr) tumors. This scheme is depicted as step 1 shown on the left part of Fig. 2. A voting scheme is applied to decide which class is chosen based on the three outputs of the contributing binary classes. With a minimum two-out-of-three vote, a certain class is taken if two or three of the binary classifiers give the same output, otherwise the classifier considers the output as *undecided*. Step 2 is carried out, as illustrated in the right part: if the output of step 1 is class 5, then further classify the spectrum either into class 1 (glio) or 3 (meta) using the binary classifier 13. If the output is class 2 or 4, then the output of step 2 is the same as the output of step 1. Four binary classifiers are the building blocks of this multiclass classifier: binary classifiers 24 and 13 are available from the

previous section, additionally two binary classifiers are required:

- meningiomas versus aggressive tumors (class 2 versus class 5),
- astrocytomas grade II versus aggressive tumors (class 4 versus class 5).

2.6. Statistical analysis

From 200 runs, the mean AUCs (\overline{AUC}) is listed in the tables, as well as the standard error (SE) on the AUC. For each binary classifier C , the mean and standard error of the AUC is calculated.

Consider two classifiers C_1 and C_2 that handle the same input data; e.g. C_1 is PCA/LDA and C_2 is LS-SVM with a linear kernel applied to the complete spectra of classes 1 and 2. Let the AUC of each classifier $C_i, i = 1, 2$ be $A_{i,l}$ with standard error $SE_{i,l}, i = 1, 2; l = 1, \dots, M$, with M the number of stratified randomizations ($M = 200$). The pooled statistics are then given by ($i = 1, 2$), where n_l is the amount of samples for the stratified randomization $l = 1, \dots, M$:

$$\bar{A}_i = \frac{1}{n} \sum_{l=1}^M A_{i,l}, \quad (2)$$

$$SE_i = \sqrt{\frac{1}{N-M} \sum_{l=1}^M (n_l - 1) SE_{i,l}^2}, \quad (3)$$

$$SE_i = \sqrt{\frac{1}{M} \sum_{l=1}^M SE_{i,l}^2}. \quad (4)$$

The last line is satisfied, since the test set contains an equal amount of samples for each stratified randomization, i.e. $\forall l n_l = n, N = \sum_{l=1}^M n_l$.

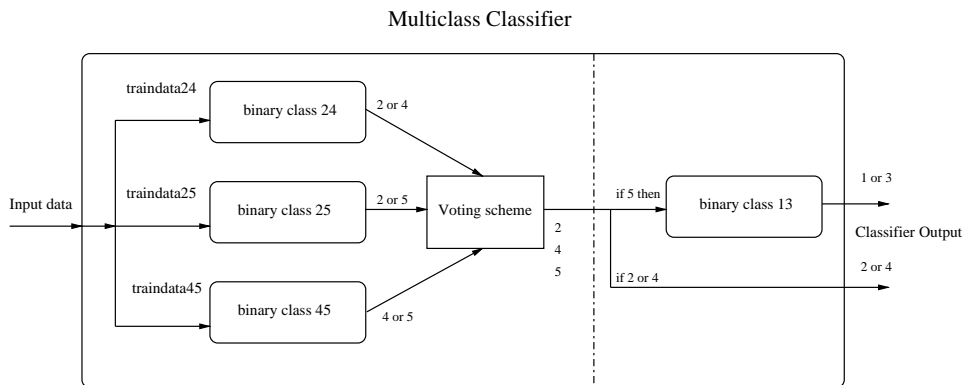


Figure 2 Two-steps classification. The left part shows step 1, classification of three tumor classes: (2) meni, (4) astroll, and (5) aggressive tumors. The right part, or step 2, further refines the classification if the output is class 5 and assigns the spectra of this class either to class 1 (glio) or 3 (meta).

A general approach to statistically test whether the areas under two ROC curves derived from the same samples differ significantly from each other is then given by the critical ratio z , defined as [37]:

$$z = \frac{\bar{A}_1 - \bar{A}_2}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}}$$

in which r is a quantity representing the correlation introduced between the two areas by studying the same samples. In our study we calculate the z -value based on the pooled statistics $\bar{A}_i, SE_i, i = 1, 2$ from 200 runs as calculated in Eqs. (2)–(4). If the resulting z -value satisfies $z \geq 1.96$, then \bar{A}_1 and \bar{A}_2 are statistically different. The cut-off value 1.96 is taken as the quantity for which, under the hypothesis of equal AUCs ($\bar{A}_1 = \bar{A}_2$), $z \geq 1.96$ occurs with a probability of $\alpha = 0.05$ under a normal distribution.

This ROC analysis is performed for binary classification. Although ROC analysis has been extended to multiclass classification [38], the result is generally non-intuitive and computationally expensive. This motivates the use of the correct classification rate as performance measure for multiclass classification.

3. Results

3.1. Classification using complete spectra

In the following the classification performance of LDA, SVM, and LS-SVM (using linear and RBF kernels) are reported. The result using the complete spectra are summarized in Table 2, while Fig. 3 shows the boxplots corresponding to the same cases. Note that the boxplots display the median of the AUC values and the Interquartile Range (IQR), while the tables display the mean and standard error of the AUC values. The latter can be used to calculate the z -value (Section 2.6).

3.2. Classification using selected frequency regions

By selecting the values within specific frequency ranges in the spectra, the number of datapoints is reduced from 108 to 30. For the LDA classifier, PCA is applied to further reduce this input dimension, covering at least 75% of the variance. These input variables are different from those obtained for the complete spectrum, due to the higher degree of freedom in the latter case. The results of 200 runs of stratified random samplings of the L2-normalized magnitude MRS spectra are shown in Table 3 and Fig. 4.

3.3. Classification using peak integration

Table 4 and Fig. 5 show the results of the ROC analysis for classification using peak integration. Five peak integrated values are used as input of the classifiers. The linear classifier LDA is used without applying PCA.

3.4. Multiclass approach

As mentioned above, two additional binary classifiers are constructed by merging glioblastomas and metastases into one class of aggressive tumors. Table 5 shows the performance of these classifiers using the complete spectra as input.

3.4.1. Training performance

One way to train the multiclass classifier is by feeding all the spectra to the classifier and train each binary classifier with the corresponding classes. For example, use the spectra of class 2 and class 4 to train the binary classifier 24, and similarly for the others.

Table 6 shows a comparison of the multiclass classifier performance. The first row shows the percentage of correctly classified spectra in the

Table 2 Classification using *complete spectra*

Classes	PCA/LDA	SVM lin	LS-SVM lin	LS-SVM RBF
glio-meni	0.9528 ± 0.0306 ₍₈₎	0.9519 ± 0.0335	0.9506 ± 0.0338	0.9560 ± 0.0304
glio-meta	0.5926 ± 0.1036 ₍₆₎	0.6323 ± 0.0942	0.6431 ± 0.0983	0.5851 ± 0.1037
glio-astroll	0.9180 ± 0.0627 ₍₇₎	0.9159 ± 0.0565	0.9351 ± 0.0524	0.9385 ± 0.0486
meni-meta	0.9605 ± 0.0375 ₍₅₎	0.9642 ± 0.0337	0.9711 ± 0.0307	0.9701 ± 0.0306
meni-astroll	0.9313 ± 0.0725 ₍₁₀₎	0.9661 ± 0.0390	0.9581 ± 0.0482	0.9595 ± 0.0456
meta-astroll	0.9612 ± 0.0533 ₍₄₎	0.9695 ± 0.0418	0.9740 ± 0.0393	0.9721 ± 0.0377

Average performance on the test set from 200 runs of stratified random samplings of the L2-normalized magnitude MRS spectra. As performance measure we use the mean AUC and its pooled standard error (SE). The number between the brackets mentions the number of principal components used.

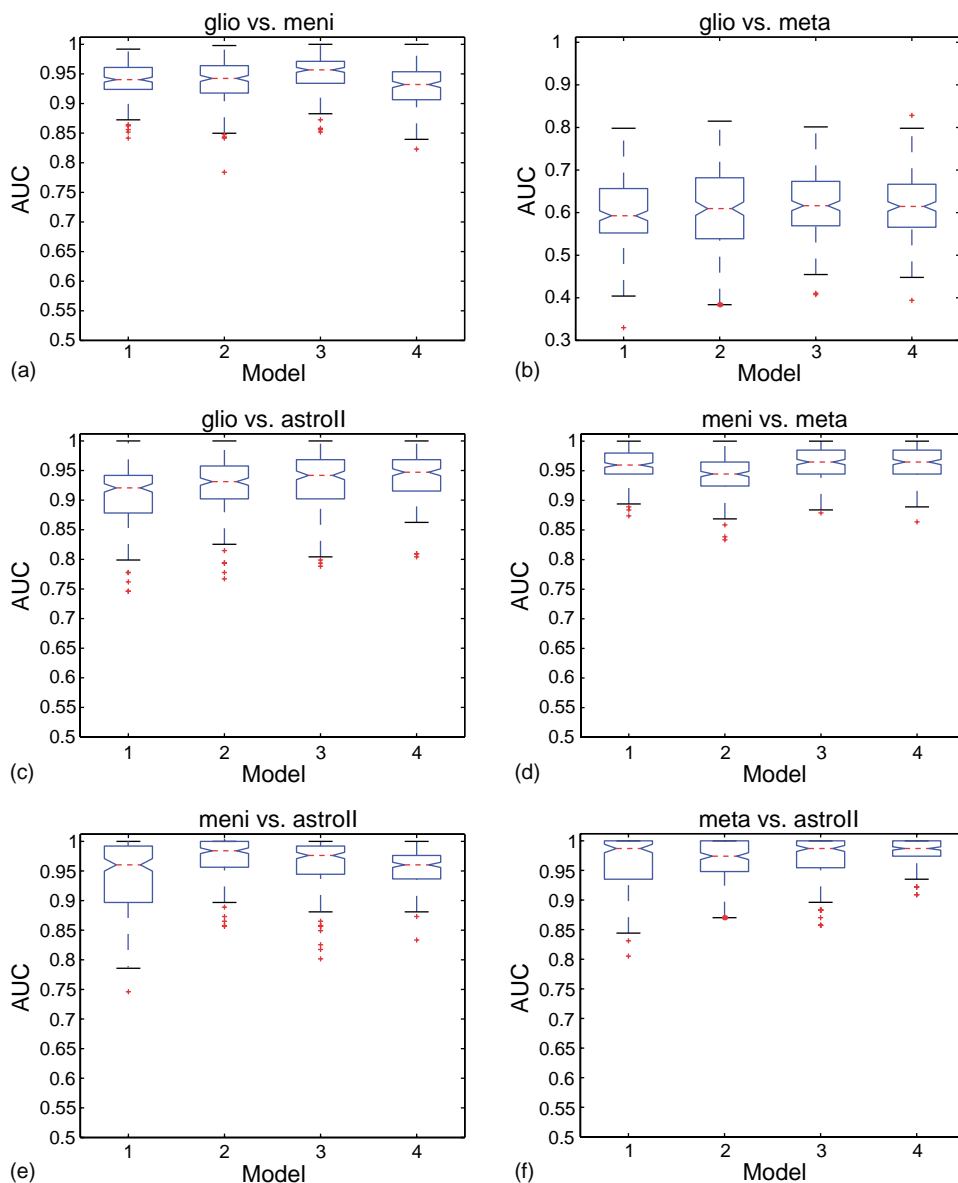


Figure 3 Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models: (1) PCA-LDA, (2) SVM with linear kernel, (3) LS-SVM with linear kernel, (4) LS-SVM with RBF kernel. Six figures correspond to the binary classifiers using *complete spectra*: (a) glio vs. meni, (b) glio vs. meta, (c) glio vs. astroll, (d) meni vs. meta, (e) meni vs. astroll and (f) meta vs. astroll.

Table 3 Classification using *selected frequency regions*

Classes	PCA/LDA	SVM lin	LS-SVM lin	LS-SVM RBF
glio-meni	$0.7643 \pm 0.0722_{(2)}$	0.8532 ± 0.0575	0.8922 ± 0.0494	0.9187 ± 0.0413
glio-meta	$0.6381 \pm 0.1004_{(2)}$	0.5081 ± 0.1044	0.6368 ± 0.0998	0.5576 ± 0.1030
glio-astroll	$0.8319 \pm 0.0776_{(2)}$	0.8692 ± 0.0713	0.8849 ± 0.0660	0.9012 ± 0.0594
meni-meta	$0.9212 \pm 0.0525_{(2)}$	0.9098 ± 0.0594	0.9339 ± 0.0475	0.9534 ± 0.0374
meni-astroll	$0.9079 \pm 0.0645_{(3)}$	0.9592 ± 0.0410	0.9619 ± 0.0422	0.9617 ± 0.0411
meta-astroll	$0.9173 \pm 0.0689_{(2)}$	0.9459 ± 0.0549	0.9698 ± 0.0389	0.9642 ± 0.0429

Average performance on the test set from 200 runs of stratified random samplings of the L2-normalized magnitude MRS spectra. As performance measure we use the mean AUC and its pooled standard error (SE). The number between the brackets mentions the number of principal components used.

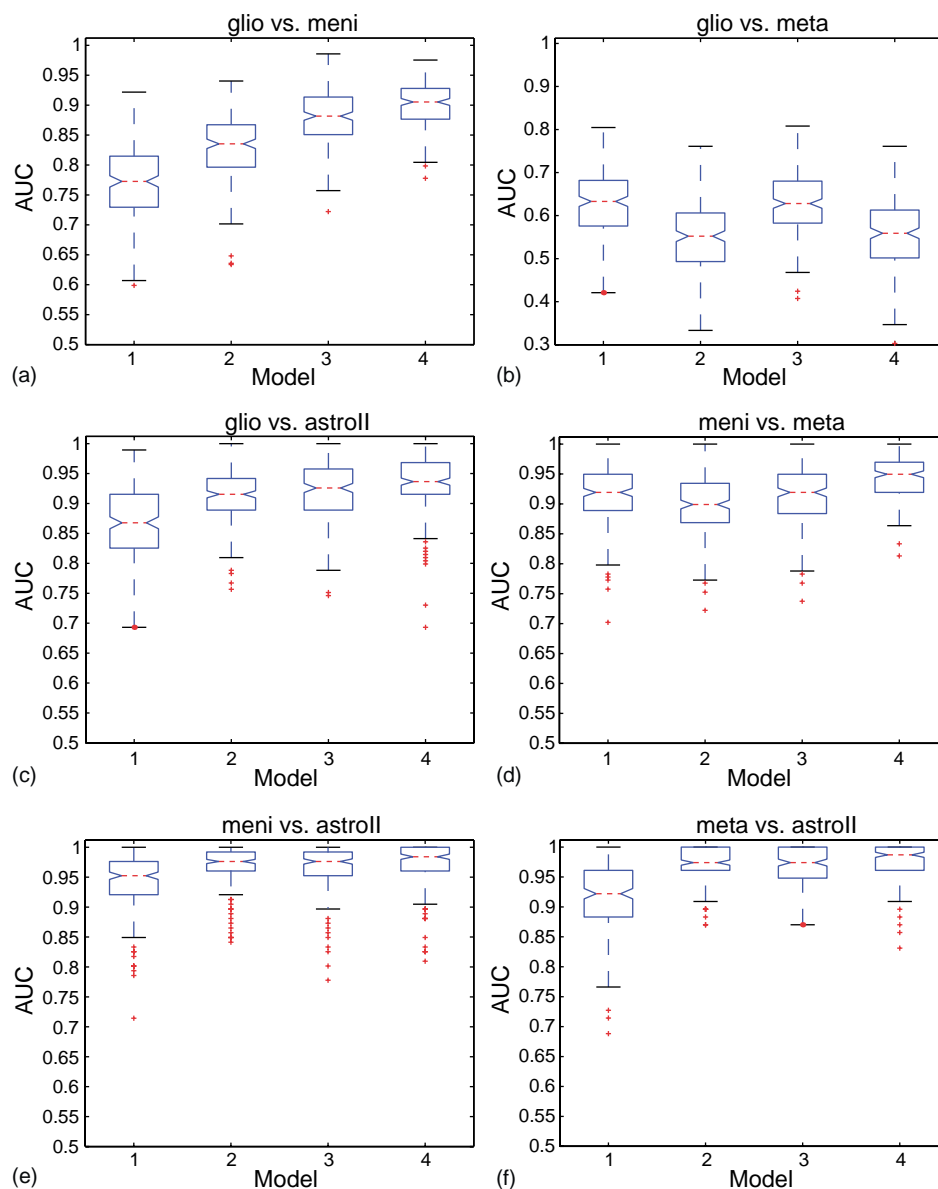


Figure 4 Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models: (1) PCA-LDA, (2) SVM with linear kernel, (3) LS-SVM with linear kernel, (4) LS-SVM with RBF kernel. Six figures correspond to the binary classifiers using *selected frequency regions*: (a) glio vs. meni, (b) glio vs. meta, (c) glio vs. astroll, (d) meni vs. meta, (e) meni vs. astroll and (f) meta vs. astroll.

Table 4 Classification using *peak integrated values*

Classes	LDA	SVM lin	LS-SVM lin	LS-SVM RBF
glio-meni	0.8504 ± 0.0586	0.8561 ± 0.0577	0.8448 ± 0.0593	0.8677 ± 0.0550
glio-meta	0.6252 ± 0.1007	0.6236 ± 0.1005	0.6434 ± 0.1006	0.6264 ± 0.0988
glio-astroll	0.8773 ± 0.0635	0.8916 ± 0.0571	0.8787 ± 0.0628	0.8818 ± 0.0631
meni-meta	0.9103 ± 0.0628	0.9113 ± 0.0618	0.9191 ± 0.0585	0.9357 ± 0.0473
meni-astroll	0.8441 ± 0.0858	0.8297 ± 0.0926	0.8485 ± 0.0851	0.8281 ± 0.0921
meta-astroll	0.9592 ± 0.0461	0.9727 ± 0.0376	0.9597 ± 0.0453	0.9521 ± 0.0528

Average performance on the test set from 200 runs of stratified random samplings of the L2-normalized magnitude MRS spectra. As performance measure we use the mean AUC and its pooled standard error (SE).

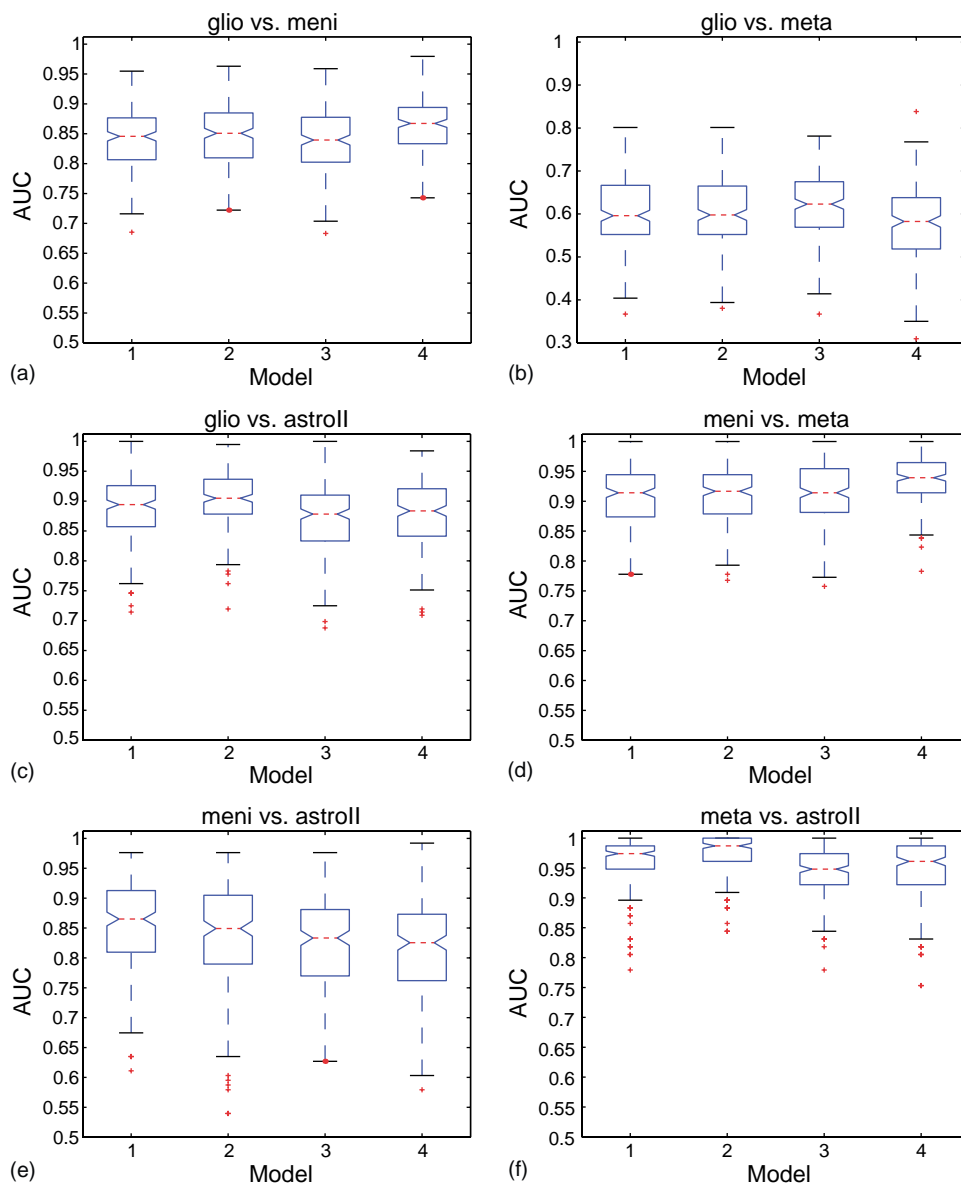


Figure 5 Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models: (1) LDA, (2) SVM with linear kernel, (3) LS-SVM with linear kernel, (4) LS-SVM with RBF kernel. Six figures correspond to the binary classifiers using *peak integration*: (a) glio vs. meni, (b) glio vs. meta, (c) glio vs. astroll, (d) meni vs. meta, (e) meni vs. astroll and (f) meta vs. astroll.

class of *meningiomas*, *astrocytomas grade II* and *aggressive tumors*. One undecided case arose when using PCA/LDA with complete spectra classification, 15 when using PCA/LDA and one when using

LS-SVM classification both with the selected frequency regions as input variables.

In the second step, we use classifier 13 to further subclassify the *aggressive class*. Using this subclassification

Table 5 Classification using *complete spectra*

Classes	LDA	SVM lin	LS-SVM lin	LS-SVM RBF
meni-aggr	$0.9433 \pm 0.0306_{(6)}$	0.9409 ± 0.0343	0.9620 ± 0.0279	0.9110 ± 0.1121
astroll-aggr	$0.9230 \pm 0.0674_{(6)}$	0.9343 ± 0.0458	0.9416 ± 0.0502	0.9129 ± 0.1137

Average performance on the test set from 200 runs of stratified random samplings of the L2-normalized magnitude MRS spectra. As performance measure we use the mean AUC and its pooled standard error (SE). The number between the brackets mentions the number of principal components used.

Table 6 One-step classification using *complete spectra*

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
Compl. spec.	84.6995	93.9891	97.8142
Disc. feat.	65.0273	84.6995	90.1639
Peak integ.	75.9563	77.0492	80.8743

Percentage of correctly classified spectra using all L2-normalized magnitude MRS spectra to assess the training performance.

Table 7 Two-steps classification using *complete spectra*

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
Compl. spec.	71.0383	78.1421	83.6066
Disc. feat.	50.2732	68.8525	74.8634
Peak integ.	61.7486	62.2951	67.7596

Percentage of correctly classified spectra using all L2-normalized magnitude MRS spectra to assess the training performance.

sification, the multiclass classifier's performance is shown in [Table 7](#).

3.4.2. Test performance

Besides using all the spectra to choose the hyper-parameters and to train the classifiers, one can also select 2/3 of the dataset as training set and use the remainder as test set. This stratified random sampling is repeated for 200 runs. The results are shown in [Table 8](#) for one-step classification, which assigns

the spectra to one of the three following classes: 2, 4 or 5. [Table 9](#) shows the classifier performance after two-steps classification, which assigns the spectra to 1 of the 4 following classes: 1, 2, 3 or 4. Each spectrum of class 5 in step 1, is either assigned to class 1 or class 3 in step 2. In [Tables 8 and 9](#) we mention the mean correct classification rate, the mean misclassification rate and the mean percentage of undecided cases and their standard deviation. The correct classification rate is defined as the

Table 8 One-step classification using *complete spectra*

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
Correct	80.1855 ± 4.2853	82.7823 ± 3.3449	83.5726 ± 3.5058
Misclass	14.0887 ± 4.0665	13.6532 ± 3.1140	12.5565 ± 3.3290
Undecided	05.7258 ± 2.6110	03.5645 ± 2.0870	03.8710 ± 2.1144

Average performance on the test set from 200 runs of stratified random samplings (2/3 of the data used for training, 1/3 for testing). 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 9 Two-steps classification using *complete spectra*

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
Correct	63.1532 ± 4.7255	65.7984 ± 3.3449	66.8145 ± 3.5058
Misclass	31.1210 ± 4.6858	30.6371 ± 3.1706	29.3145 ± 3.5954
Undecided	05.7258 ± 2.6110	03.5645 ± 2.0870	03.8710 ± 2.1144

Average performance on the test set from 200 runs of stratified random samplings (2/3 of the data used for training, 1/3 for testing). 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. Note that the number of undecided cases is equal to that for the one-step classifier.

percentage of correctly classified spectra, while the misclassification rate is the percentage of misclassified cases.

4. Discussion

In this section we discuss various issues concerning the results we obtained using the available long echo ^1H MRS data. We do not necessarily claim that these remarks generally hold for similar analyses on other data.

4.1. Limitations

MRS signals of brain tumors contain chemical information about metabolites characteristic for the type of tumor. Nevertheless, there are still some factors making it hard to construct a classifier which is able to discriminate between different brain tumors using MRS signals:

- (1) The limited number of available spectra per type of tumor (see Table 1). Especially the amount of available metastases and astrocytomas grade II is low. This makes it difficult to construct a classifier with a high generalization capacity.
- (2) The presence of noise and artefacts in the spectra. Even after elimination of the dominating water peak, remaining artefacts might affect important peaks in the spectra.
- (3) The large variances within each class and the overlap between spectra of different brain tumor types (see Fig. 1). For example, the mean spectra of glio and meta show a very similar characteristic pattern, which makes the discrimination between glio and meta a very hard problem. This problem is also observed in [9,36]. Further discussion about this is addressed below.

4.2. Glioblastomas versus metastases

Although we obtained a low performance for distinguishing glioblastomas (glio) and metastases (meta), there are indications that these tumor types might be separable based on MR. In [8] Szabo De Edelenyi et al. introduced the so-called nosologic images, which is an approach to analyze ^1H MRSI data of brain tumors. It is a tool that assigns the spectroscopic data of each voxel in the spectroscopic image to a histopathological class. Classification was carried out by LDA applied on six metabolite values obtained from long echo ^1H MRSI spectra (TE = 272 ms), together with the unsuppressed

water area. Their study included 77 images, of which 24 high-grade gliomas and 10 metastases, for which they obtained a training performance of 87% following a leave-one-out (LOO) procedure. For the high-grade gliomas and metastases, respectively, 19 and 6 spectra were correctly assigned.

Researchers [39–41] have found a few metabolite peaks or ratios which might contribute to the discrimination of high-grade gliomas and metastases. Law et al. concluded out of a study based on MRSI that, despite the small size of their dataset (11 high-grade gliomas, 6 metastases), the *Cho/Cr* ratio was significantly higher in high-grade gliomas than in metastases; this was the case for the tumoral region as well as the peritumoral region. Also based on perfusion-weighted MRI they have found different characteristics. Opstad et al. [41] have considered short echo ^1H MRS spectra (TE = 30 ms) from 25 glioblastomas and 34 metastases. Based on these data, they were able to find a significant difference in the ratio of the 1.3 ppm and the 0.9 ppm lipid/macromolecule peaks between these two tumor groups. This lipid peak area (LPA) ratio was 2.6 ± 0.6 for glioblastomas and 3.8 ± 1.4 for metastases ($P \leq 0.0001$). Based on ^1H MRS, Ishimaru et al. [39] have shown that the absence of *Cr* might indicate a diagnosis of metastasis, while in short echo the absence of lipids may exclude metastasis. In Fig. 1 we do indeed notice a large mean lipid peak in metastases, but also in glioblastomas. This latter might be due to the occurrence of necrotic tissue in part of the glioblastomas. This partially explains the large variation we especially observe within this class and the similarity with the class of metastases.

4.3. Classification techniques

In general, LDA as linear classifier, preceded by PCA (except for peak integration) performs quite well in solving the brain tumor classification problem. This is in correspondence with [10]. However, due to its linear boundary, overlapping classes are very difficult to handle.

As stated above, the small dataset available also forms a limitation for training. Therefore, the discrimination boundary will strongly correlate with the training set. Especially LDA requires a significant amount of datapoints to be able to draw a linear separating line between overlapping classes. In addition, it is possible that the separating line is very dependent on the selected training set.

Kernel-based classifiers are less sensitive to the amount of datapoints; although the dimension is larger than the number of datapoints, these classifiers could draw an optimal separating boundary,

without applying any dimensionality reduction (such as PCA). Kernel-based classifiers, SVM and LS-SVM, feature the advantage of detecting automatically important characteristics independently of the input pattern.

Based on the statistical analysis, described in Section 2.6, no statistically significant difference was found between the AUC values for any of the classification techniques applied to the available long echo ^1H MRS data. The highest z-value (1.72) was obtained, when comparing PCA/LDA with LS-SVM with a linear kernel based on the frequency selected regions; this is still lower than the cut-off value (1.96).

Also from visual inspection, we cannot conclude that there is a clear difference in between the considered classification techniques. In particular, the best performing technique depends on the considered classes and the type of input. To be more specific, when comparing the classification techniques we can group them in two ways (e.g. consider only the results with the complete spectra as input):

- linear (LDA, SVM lin, LS-SVM lin) versus non-linear techniques (LS-SVM RBF): in the cases glio-meni, glio-astroll the mean AUC values are slightly higher for the non-linear technique, while in the other cases the AUC values are in the same range or slightly lower.
- LDA versus kernel-based techniques (SVM lin, LS-SVM lin, LS-SVM RBF): in the cases meni-meta, meni-astroll and meta-astroll the kernel-based techniques perform slightly better.

Additionally, we can still consider the comparison of SVM versus LS-SVM. Out of this we can only conclude that the best performing technique is also quite dependent on the case.

4.4. Influence of dimensionality reduction

For classification using selected frequency regions (Section 3.2) and peak integrated values (Section 3.3) the input dimension is reduced by selecting only spectral regions which contain resonances of important metabolites. The underlying idea for this dimensionality reduction is to remove any redundant input features and reduce the influence of noise and artefacts. Hence, we try to enhance the discriminatory chemical information present in the spectra.

In contradiction to our expectations, we observe that the results in Sections 3.2 and 3.3 on average are worse than in Section 3.1. This seems to imply that this approach to dimensionality reduction also reduces part of the valuable information, that is present in the excluded frequency regions, which

was important to explain the variance between the brain tumor classes. More specifically, here we are not considering resonances which typically have a low intensity at long echo times (because of a small T2-value or cancellation due to J-modulation) [42,43]: *ml* (e.g. with triplets and multiplet at 3.26 and 3.57 ppm); *Glu* (e.g. multiplets at 2.33 and 3.74 ppm); *Gln* (e.g. multiplets at 2.43 and 3.75 ppm); *Gly* (singlet at 3.55 ppm). In Fig. 1 we indeed notice a few small peaks around the specified resonances.

4.5. Multiclass classification

Multiclass classifiers handle all classes in one construction. We reduce this problem to a set of four binary classification problems, as explained in Section 2.5.2. Hence, we obtain four separating functions instead of one (one for each binary problem). Multiclass classifiers with the proposed scheme show a high learning capability. This is illustrated by the correct classification rates for the first step using the complete spectra as input: 84.7% (PCA/LDA), 93.9% (LS-SVM lin) and 97.8% (LS-SVM RBF). Given an independent test set as input, the classifiers on average give a quite good generalization performance: 80.2% (PCA/LDA), 82.8% (LS-SVM lin) and 83.6% (LS-SVM RBF).

In the second step of the multiclass classifier we combine the output of the first step with the binary classifier that separates glio and meta. The test performances reduce to 63.1% correct classification (LDA/PCA), 65.8% (LS-SVM lin), 66.8% (LS-SVM RBF). This can be explained by the hard binary problem glio versus meta. As observed in the binary classification, these two classes are very similar. Therefore, separating them from one single class 5 into class 1 and class 3 deteriorates the total performance of the classifier.

Although, no ROC analysis for multiclass classification was performed in order to test for significant differences in between the classification techniques, the following indications can be noted, without drawing a general conclusion. The results, after step 1 as well as step 2, yield a clearly higher training performance for the kernel-based methods than for LDA. Moreover, the kernel-based methods on average perform slightly better on an independent test set than LDA. This is clear from the mean percentage of correctly classified spectra which is a few percentages higher for LS-SVM (see Tables 8 and 9 and Fig. 6). Also the mean percentage of undecided cases differs slightly in favor of LS-SVM. This indicates that kernel-based methods can generalize at least as well as LDA based on a small dataset.

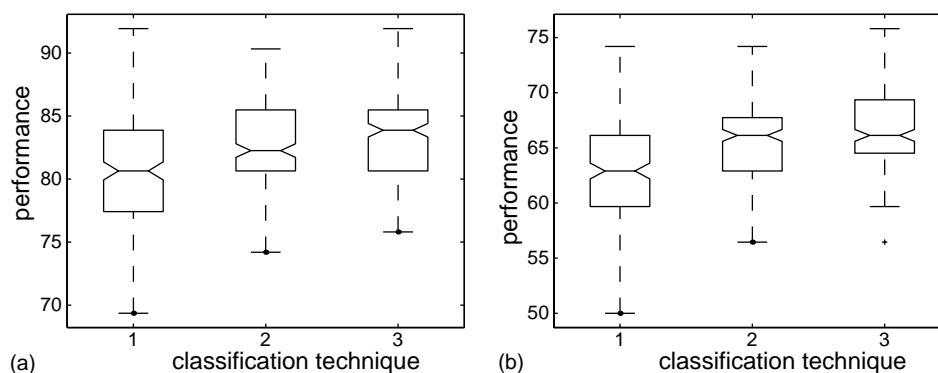


Figure 6 Boxplots of the correct classification rate on the test set from 200 runs of stratified random samplings (2/3 of the data for training, 1/3 for testing) of the three models: (1) LDA, (2) LS-SVM with linear kernel, (3) LS-SVM with RBF kernel. Two figures correspond to the multiclass classifiers using *complete spectra*: (a) output of step 1 classifier, (b) output of step 2 classifier.

5. Conclusions

This paper shows a comparative study of brain tumor classification based on long echo ^1H MRS signals. Linear as well as non-linear classifiers are compared. All techniques are applied automatically, including (hyper-) parameter selection, training and testing. Also for use in clinical practice, all techniques are easy to automate for analysis of independent data.

Binary classification gives more insight on the distributions of each class and their overlap. Except for the hard case glioblastomas versus metastases, all classifiers based on the complete magnitude spectra are able to reach an AUC of more than 0.9. Based on the available data, we were not able to statistically prove any difference in performances between the classification techniques, for binary as well as multiclass classification. This indicates that kernel-based methods and LDA statistically perform as well for classification of brain tumors based on a small set of long echo ^1H MRS data.

However, each of the applied techniques has its characteristics. LDA requires a prior dimensionality reduction of input variables (e.g. by applying PCA), while dimensionality reduction is done automatically in kernel-based methods.

We expected that dimensionality reduction, by selecting frequency regions or peak integration, would reduce the disturbing noise and artefacts in the spectra. However, the described approach for selecting resonance peaks of long echo ^1H MRS spectra resulted in a lower performance. It might be necessary to include additional spectral information to increase classification performance. This also motivates further research in learning the peak pattern of short echo ^1H MRS, for which data are also provided within the INTERPRET project.

By using magnitude spectra, phasing problems are avoided. Nevertheless, with respect to real spectra, in magnitude spectra there occurs more peak overlap. Also, in real spectra at long echo time TE (TE = 135, 136 ms) the peaks of Ala and Lac are inverted. This might reduce the ability to distinguish tumor types based on subtle differences in the spectral pattern. In order to test for this effect, in a future study also real spectra could be included as input features.

Discriminating aggressive tumor types, glioblastomas and metastases, using long echo ^1H MRS spectra clearly is a very hard problem due to the highly similar pattern of the spectra from both classes, possibly due to the presence of some necrotic tissue. In order to address this problem, the discriminatory information present in the ^1H MRS spectrum should be enhanced, potentially by improvements in the acquisition of ^1H MRS signals. In particular, improvements are expected when processing short echo ^1H MRS signals, since more metabolites are visible in these spectra. Moreover, this spectral information is spread out over a larger amount of peaks, thereby enlarging the number of possible discriminatory features. This is part of future research.

Acknowledgements

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 project. 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.

References

- [1] The Brain Tumor Society. <http://www.tbts.org>.
- [2] Mukherji SK, editor. Clinical applications of magnetic resonance spectroscopy. Wiley-Liss, 1998.
- [3] Smith ICP, Stewart LC. Magnetic resonance spectroscopy in medicine: clinical impact. *Prog Nucl Mag Res Sp* 2002; 40:1–34.
- [4] Cousins JP. Clinical MR spectroscopy: fundamentals, current applications, and future potential. *AJR Am J Roentgenol* 1995;164:1337–47.
- [5] Lindon JC, Holmes E, Nicholson JK. Pattern recognition methods and applications in biomedical magnetic resonance. *Prog Nucl Mag Res Sp* 2001;39:1–40.
- [6] Herminghaus S, Dierks T, Pilatus U, Möller-Hartmann W, Wittsack J, Marquardt G, et al. Determination of histopathological tumor grade in neuroepithelial brain tumors by using spectral pattern analysis of in vivo spectroscopic data. *J Neurosurg* 2003;98:74–81.
- [7] Simonetti AW, Melssen WJ, van der Graaf M, Heerschap A, Buydens LMC. Brain tumor classification and probability maps using MRI and MRSI data. *Anal Chem* 2003;75(20): 5352–61.
- [8] Szabo De Edelenyi F, Rubin C, Estève F, Grand S, Décorps M, Lefournier V, et al. *Nature Med.* 2000;6:1287–9.
- [9] Tate AR, Griffiths JR, Martínez-Pérez I, Moreno A, Barba I, Cabañas ME, et al. Towards a method for automated classification of 1H MRS spectra from brain tumours. *NMR Biomed* 1998;11:177–91.
- [10] Tate AR, Majós C, Moreno A, Howe FA, Griffiths JR, Arús C. Automated classification of short echo time in vivo ¹H brain tumor spectra: a multicenter study. *Magn Reson Med* 2003;49:29–36.
- [11] Ye C-Z, Yang J, Geng D-Y, Zhou Y, Chen N-Y. Fuzzy rules to predict degree of malignancy in brain glioma. *Med Biol Eng Comput* 2002;40:145–52.
- [12] Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol* 1979;14(2):109–21.
- [13] International network for pattern recognition of tumours using magnetic resonance. <http://carbon.uab.es/INTERPRET/>.
- [14] Ladroue C, Tate AR, Howe FA, Griffiths JR. 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. p. 147–8.
- [15] Lefournier V, Szabo De Edelenyi F, Estève F, Grand S, Bessou P, Boubagra K, et al. Nosologic images for classification of brain tumors with ¹H MRSI: clinical performance. 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. 91–2.
- [16] Lukas L, Devos A, Suykens JAK, Vanhamme L, Van Huffel S, Tate AR, et al. The use of LS-SVM in the classification of brain tumors based on magnetic resonance spectroscopy signals. In: Proceedings of the European Symposium for Artificial Neural Networks (ESANN), Bruges, Belgium, April 24–26, 2002. p. 131–5.
- [17] Lukas L, Devos A, Suykens JAK, Vanhamme L, Van Huffel S, Tate AR, et al. The use of LS-SVM in the classification of brain tumors based on ¹H-MR spectroscopy signals. In: Proceedings of the IEE Symposium on Medical Applications of Signal Processing, Savoy Place, London, UK, October 7, 2002. p. 15/1–5.
- [18] Szabo De Edelenyi F, Estève F, Rémy C, Buydens L. 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.
- [19] Tate AR, Griffiths JR, Howe FA, Pujol J, Arús C. 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.
- [20] Vapnik V. The nature of statistical learning theory. New York: Springer, 1995.
- [21] Vapnik V. Statistical learning theory. New York: Wiley, 1998.
- [22] Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. *Neur Proc Lett* 1999;9(3):293–300.
- [23] Klose U. In vivo proton spectroscopy in presence of eddy currents. *Magn Reson Med* 1990;14:26–30.
- [24] Barkhuijsen H, De Beer R, Van Ormondt D. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *J Magn Reson* 1987;73:553–7.
- [25] Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. New York: Wiley, 2001.
- [26] Ripley BD. Pattern recognition and neural networks. Cambridge: Cambridge University Press, 1996.
- [27] Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J. Least squares support vector machines. Singapore: World Scientific, 2002.
- [28] Gunn SR. Support vector machines for classification and regression. Technical Report. Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.
- [29] MATLAB support vector machines toolbox. <http://www.isis.ecs.soton.ac.uk/isystems/kernel>.
- [30] MATLAB/C LS-SVMlab toolbox. <http://www.esat.kuleuven.ac.be/sista/lssvmlab>.
- [31] Pelckmans K, Suykens JAK, Van Gestel T, De Brabanter J, Lukas L, Hamers B, et al. LS-SVMlab Toolbox User's Guide. Internal Report 02-145. ESAT-SISTA, K. U. Leuven, Leuven, Belgium, 2002.
- [32] Howe FA, Barton SJ, Cudlip SA, Stubbs M, Saunders DE, Murphy M, et al. Metabolic profiles of human brain tumors using quantitative in vivo ¹H magnetic resonance spectroscopy. *Magn Reson Med* 2003;49:223–32.
- [33] Leclerc X, Huisman TAGM, Sorensen AG. The potential of proton magnetic resonance spectroscopy (¹H) in the diagnosis and management of patients with brain tumors. *Curr Opin Oncol* 2002;14:292–8.
- [34] Majós C, Alonso J, Aguilera C, Serrallonga M, Acebes JJ, Arús C, et al. Adult primitive neuroectodermal tumor: proton MR spectroscopic findings with possible application for differential diagnosis. *Radiology* 2002;225:556–66.
- [35] Murphy M, Loosemore A, Clifton AG, Howe FA, Tate AR, Cudlip SA, et al. The contribution of proton magnetic resonance spectroscopy (¹H MRS) to clinical brain tumour diagnosis. *Br J Neurosurg* 2002;16(4):329–34.

- [36] Poptani H, Kaartinen J, Gupta RK, Niemitz M, Hiltunen Y, Kauppinen RA. Diagnostic assessment of brain tumours and non-neoplastic brain disorders in vivo using proton nuclear magnetic resonance spectroscopy and artificial neural networks. *J Cancer Res Clin Oncol* 1999;125:343–9.
- [37] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;148:839–43.
- [38] Srinivasan A. Note on the location of optimal classifiers in n -dimensional ROC space. Technical Report PRG-TR-2-99. Oxford University Computing Laboratory, Oxford, England, 1999.
- [39] Ishimaru H, Morikawa M, Iwanaga S, Kaminogo M, Ochi M, Hayashi K. Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. *Eur Radiol* 2001;11:1784–91.
- [40] Law M, Cha S, Knopp EA, Johnson G, Arnett J, Litt AW. High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. *Radiology* 2002;222:715–21.
- [41] Opstad KS, Griffiths JR, Bell BA, Howe FA. In vivo lipid T2 relaxation time measurements in high-grade tumors: differentiation of glioblastomas and metastases. In: Proceedings of the 11th Scientific Meeting and Exhibition (ISMRM 03), Toronto, Canada, July 10–16, 2003, p. 754.
- [42] Ernst T, Hennig J. Coupling effects in volume selective 1H spectroscopy of major brain metabolites. *Magn Reson Med* 1991;21:82–96.
- [43] Govindaraju V, Young K, Maudsley AA. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR Biomed* 2000;13:129–53.