Brain tumour classification based on long echo time 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]

[a] *SCD-SISTA, Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Heverlee (Leuven), Belgium*

[b] *CRC Biomedical Magnetic Resonance Research Group, Department of Biochemistry and Immunology, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, United Kingdom*

[c] *Institut de Diagnòstic per la Imatge (IDI), CSU de Bellvitge, Autovia de Castelldefels km 2.7, L'Hospitalet de Llobregat, 08907 Barcelona, Spain*

[d] *Centre Diagnòstic Pedralbes, Unitat Esplugues, C/ Josep Anselm Clavé, 100, 08950 Esplugues de Llobregat, Spain*

[e] *Department of Radiology, University Medical Center Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands*

[f] *Departament de Bioquímica i Biologia Molecular, Unitat de Ciències, Edifici Cs, Universitat Autonòma de Barcelona, 08193 Cerdanyola del Vallès, Spain*

Abstract

There has been a growing research interest in brain tumour classification based on proton magnetic resonance spectroscopy (^{1}H MRS) signals. Four research centers within the EU funded INTERPRET project have acquired a significant number of long echo time ^{1}H MRS signals for brain tumour classification. In this paper, we present an objective comparison of several classification techniques applied to the discrimination of four types of brain tumours: *meningiomas*, *glioblastomas*, *astrocytomas grade II* and *metastases*. Linear and nonlinear 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 nonlinear 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 vs. metastases. Although, based on the available long echo time ^{1}H 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.

Key words: brain tumours, classification, Magnetic Resonance Spectroscopy (MRS), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Least Squares Support Vector Machine (LS-SVM)

1 Introduction

Brain tumours are the second leading cause of cancer death in children under 15 years old and young adults up to the age of 34. These tumours are also the second fastest growing cause of cancer death among humans older than 65 years old [41]. 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 tumours. 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 tumours [24,31]. 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 artefacts. Cousins [2] discusses the influence of the echo time TE on the spectral pattern of a MRS signal. The above mentioned metabolites can be detected in short echo ^{1}H MRS signals. However, short echo ^{1}H MRS signals are more difficult to analyze

* Corresponding author. Tel. +32-16-321884, Fax. +32-16-321970.
Email address: lukas.lukas@esat.kuleuven.ac.be (L. Lukas).

than long echo time ^{1}H MRS signals due to a higher number of overlapping peaks, a stronger baseline and a higher sensitivity to artefacts. In comparison, long echo time ^{1}H 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.* [18] overviewed pattern recognition methods and their applications in biomedical magnetic resonance. Several studies [9,30,36,38,40,46] also show some progress in automated pattern recognition for brain tumour classification based on MR data. These studies are either based on MRI (e.g. [46]), MRI combined with MR Spectroscopic Imaging (MRSI) (e.g. [30]), long echo time (e.g. [9,36,38]) or short echo ^{1}H MRS (e.g. [40]), 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 [35] 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 [46], 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 tumours, more specifically for brain gliomas (e.g. [9,46]). 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 tumour types, namely metastases and meningiomas.

Moreover, this paper reports the results of a comparative study on a *multi-center* dataset of MRS signals. This dataset was developed in the framework of the EU funded INTERPRET project [11]. Several INTERPRET partners [14,17,20,30,37–40] have already published results for classification of brain tumours based on MR data available within the project. The papers [17,30,37] focus on the use of ^{1}H MRSI data, while others consider the use of short or long echo time ^{1}H MRS. Nevertheless, most of these studies are based on a previous version of the dataset or focus on a specific technique. E.g. in [40], 144 short echo ^{1}H MRS spectra from 3 contributing centers were used, originating from three groups of brain tumours; meningiomas, low-grade astrocytomas and aggressive tumours. The latter group includes glioblastomas and metastases. Note that these groups correspond to the same four tumour 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 time ^{1}H MRS data from four common brain tumour 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 [17] ROC analysis was also applied to compare two diagnostic methods for classification based on ^{1}H MRSI. Third of all, four different techniques are applied for classification; linear as well as nonlinear 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 (SVM) and the least squares version of support vector machines (LS-SVM) in classifying brain tumours based on long echo ^{1}H MRS spectra. The support vector machine (SVM) [44,45] 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 (LS-SVM) [33] 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 time ^{1}H 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 Hz or 2500 Hz, the number of datapoints is 512 or 2048).

Four main classes are considered, corresponding to four brain tumour 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 (astroII), 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 tumours were histopathologically classified by three pathologists. These class assignments were based on the histological classification of Tumours 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 [13] and filtering of the dominating residual water peak using HSVD [1]. The initial point of the time domain signal was removed, because it was often affected by artefacts. 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 to 0.94 ppm), corresponding to 108 input variables. Figure 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 nonlinear methods: LDA, SVM and LS-SVM.

Linear Discriminant Analysis [4,29] basically projects the data $x_k \in \mathbb{R}^n$ 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 nonseparable. 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 nonlinear classifier.

A **Support Vector Machine (SVM)** [44,45] 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 $\{x_k, y_k\}_{k=1}^N$, $x_k \in \mathbb{R}^n$, $y_k \in \{-1, +1\}$. The classifier in the primal space is defined by: $y(x) = sign[w^T\varphi(x) + b]$, $k = 1, ..., N$, in which w is a weighting function.
(2) The SVM performs a nonlinear mapping φ of the input vectors $x_k \in \mathbb{R}^n$ 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:

$$\begin{aligned} \min_{w,e_k} J(w, e_k) &= \frac{1}{2} w^T w + C \sum_{k=1}^{N} e_k \\ \text{s.t.} \quad y_k[w^T\varphi(x) + b] &\geq 1 - e_k, \\ e_k &\geq 0, \qquad\qquad k = 1, ..., N \end{aligned}$$

in which C is a regularization constant.

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

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

satisfying the constraints $\sum_{k=1}^{N} \alpha_k y_k = 0$ and $0 \leq \alpha_k \leq C$ for $k = 1, ..., N$ where $\{x_k \in \mathbb{R}^n | k = 1, ..., N\}$ is the training sample set, and $\{y_k \in \{-1, +1\} | k = 1, ..., N\}$ 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 $x_k \in \mathbb{R}^n$ with corresponding nonzero 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):

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

Recently, a **least squares version (LS-SVM)** has been proposed [33,34], 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

$$\min_{w,b,e} J(w, b, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2$$
$$\text{s.t.} \quad y_k [w^T \varphi(x_k) + b] = 1 - e_k, \quad k = 1, ..., N$$

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

(2) Here, nonzero 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

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

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

In certain problems, nonlinear techniques could improve classification performance, especially when data are linearly nonseparable. 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 [7,22] and KULeuven's MATLAB/C LS-SVMlab toolbox [23,27,34] 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 [10,16,21,24,25,31]. These peaks might be used as discriminatory features to distinguish tumour types. In particular when their appearance clearly differ in size and shape in between spectra of different tumour 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; mI, Gln, Gly,...) 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 nonzero 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 tumour 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 tumours, then six binary classifiers can be constructed to separate the following pairs:

- glioblastomas vs meningiomas,
- glioblastomas vs metastases,
- glioblastomas vs astrocytomas grade II,
- meningiomas vs metastases,
- meningiomas vs astrocytomas grade II, and
- metastases vs 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 [35] 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 tumour 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 tumour 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 [28,38]. Alternatively, as was suggested by Tate *et al.* in [40], we can merge these two classes, obtaining a new group called class 5, containing only aggresive (aggr) tumours. This scheme is depicted as step one shown on the left part of the Figure 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 two is carried out, as illustrated in the right part: if the output of step one 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 two is the same as the output of step one. 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 vs. aggressive tumours (class 2 vs. class 5),
- astrocytomas grade II vs. aggressive tumours (class 4 vs. class 5).

2.6 Statistical Analysis

From 200 runs, the mean AUCs ($\overline{\text{AUC}}$) 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 $SE_{i,l}, i = 1, 2; l = 1, ..., M$, 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 $l = 1, ..., M$:

$$\overline{A_i} = \frac{1}{n} \sum_{l=1}^{M} A_{i,l}, \tag{2}$$

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

$$= \sqrt{\frac{1}{M} \sum_{l=1}^{M} SE_{i,l}^2} \tag{4}$$

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

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 [8]:

$$z = \frac{\overline{A_1} - \overline{A_2}}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}},$$

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 $\overline{A_i}, SE_i, i = 1, 2$ from 200 runs as calculated in Equations (2)-(4). If the resulting z-value satifies $z \geq 1.96$, then $\overline{A_1}$ and $\overline{A_2}$ are statistically different. The cutoff-value 1.96 is taken as the quantity for which, under the hypothesis of equal AUCs ($\overline{A_1} = \overline{A_2}$), $z \geq 1.96$ 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 [32], 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 Figure 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 into 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 Figure 4.

3.3 Classification using peak integration

Table 4 and Figure 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 tumours. 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 class of *meningiomas*, *astrocytomas grade II* and *aggressive tumours*. One undecided case arose when using PCA/LDA with complete spectra classification, fifteen 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, the multiclass classifier's performance is shown in Table 7.

3.4.2 Test performance

Besides using all the spectra to choose the hyperparameters 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 3 following classes: 2, 4 or 5. Table 9 shows the classifier performance after two-steps classification, which assigns the spectra to one of the 4 following classes: 1, 2, 3 or 4. Each spectrum of class 5 in step one, is either assigned to class 1 or class 3 in step two. In the 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 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 time ^{1}H 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 tumours contain chemical information about metabolites characteristic for the type of tumour. Nevertheless, there are still some factors making it hard to construct a classifier which is able to discriminate between

different brain tumours using MRS signals:

(1) the limited number of available spectra per type of tumour (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 tumour types (see Figure 1). E.g. 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 [28,38]. Further discussion about this is addressed below (section 4.2).

4.2 Glioblastomas versus metastases

Although we obtained a low performance for distinguishing glioblastomas (glio) and metastases (meta), there are indications that these tumour types might be separable based on MR. In [36] Szabo De Edelenyi *et al.* introduced the so called nosologic images, which is an approach to analyze ^{1}H MRSI data of brain tumours. 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 6 metabolite values obtained from long echo time ^{1}H MRSI spectra (TE=272ms), 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 [12,15,26] 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 tumoural region as well as the peritumoural region. Also based on perfusion-weighted MRI they have found different characteristics. Opstad *et al.* [26] have considered short echo ^{1}H MRS spectra (TE=30ms) 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 tumour 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 ^{1}H MRS, Iishimaru *et al.* [12] 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 Figure 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, preceeded by PCA (except for peak integration) performs quite well in solving the brain tumour classification problem. This is in correspondence with [40]. 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 time ^{1}H 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 cutoff value (1.96).

Also from visual inspection, we can not 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) vs nonlinear techniques (LS-SVM RBF):

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

Additionally, we can still consider the comparison of SVM vs 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 tumour 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) [5,6]: *mI* (e.g. with triplets and multiplet at 3.26 and 3.57 ppm), *Glu* (e.g. multiplets at 2.33 ppm and 3.74 ppm), *Gln* (e.g. multiplets at 2.43 ppm and 3.75 ppm), *Gly* (singlet at 3.55 ppm). In Figure 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 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 vs 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 one as well as step two, 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, 9 and Figure 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.

5 Conclusions

This paper shows a comparative study of brain tumour classification based on long echo time ^{1}H MRS signals. Linear as well as nonlinear 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 vs. 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 tumours based on a small set of long echo time ^{1}H 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 time ^{1}H MRS spectra resulted in a lower performance. It migth be necessary to include additional spectral information to increase classification performance. This also motivates further research in learning the peak pattern of short echo ^{1}H 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 tumour 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 tumour types, glioblastomas and metastases, using long echo time ^{1}H 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 ^{1}H MRS spectrum should be enhanced, potentially by improvements in the acquisition of ^{1}H MRS signals. In particular, improvements are expected when processing short echo ^{1}H 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 and JS are 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] 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, 73:553-557.

[2] Cousins JP. Clinical MR spectroscopy: fundamentals, current applications, and future potential. AJR Am J Roentgenol, 1995, 164:1337-1347.

[3] Demaerel P. Clinical applications of localized ^{1}H-NMR spectroscopy in intracranial tumours. Acta Biomedica Lovaniensia 112. Leuven Univ. Press. Belgium, 1995.

[4] Duda RO, Hart PE, Stork DG. Pattern Classification, 2nd ed. John Wiley & Sons, 2001.

[5] Ernst T, Hennig J. Coupling effects in volume selective 1H spectroscopy of major brain metabolites. Magn Reson Med, 1991, 21:82-96.

[6] Govindaraju V, Young K, Maudsley AA. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR Biomed., 2000, 13:129-153.

[7] Gunn SR. Support Vector Machines for Classification and Regression. Technical Report, Image Speech and Intelligent Systems Research Group, University of Southampton, 1997.

[8] Hanley JA, McNeil BJ. A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases. Radiology, 1983, 148:839-843.

[9] Herminghaus S, Dierks T, Pilatus U, Möller-Hartmann W, Wittsack J, Marquardt G, Labish C, Lanfermann H, Schlote W, Zanella FE. 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.

[10] Howe FA, Barton SJ, Cudlip SA, Stubbs M, Saunders DE, Murphy M, Wilkins P, Opstad KS, Doyle VL, McLean MA, Bell BA, Griffith JR. Metabolic Profiles of Human Brain Tumors Using Quantitative In Vivo ^{1}H Magnetic Resonance Spectroscopy. Magn Reson Med, 2003, 49:223-232.

[11] International network for Pattern Recognition of Tumours Using Magnetic Resonance http://carbon.uab.es/INTERPRET/.

[12] Ishimaru H, Morikawa M, Iwanaga S. Kaminogo M, Ochi M, Hayashi K. Differentiation between high-grade glioma and metastastic brain tumor using single-voxel proton MR spectroscopy. Eur. Radiol., 2001, 11:1784-1791.

[13] Klose U. In vivo proton spectroscopy in presence of eddy currents. Magn Reson Med, 1990, 14:26-30.

[14] Ladroue C, Tate AR, Howe FA, Griffiths JR. Exploring magnetic resonance data with independent component analysis. 19th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB02), Cannes, France, August 22-25, 2002, pp. 147-148.

[15] 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, 222:715-721.

[16] Lecrerc X, Huisman TAGM, Sorensen AG. The potential of proton magnetic resonance spectroscopy (^{1}H) in the diagnosis and management of patients with brain tumors. Curr. Opin. Oncol., 2002, 14:292-298.

[17] Lefournier V, Szabo De Edelenyi F, Estève F, Grand S, Bessou P, Boubagra K, Bosson J, Segebarth C, Le Bas J, Rémy C. Nosologic images for classification of brain tumors with ^{1}H MRSI: clinical performance. 19th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB02), Cannes, France, August 22-25, 2002, pp. 91-92.

[18] Lindon JC, Holmes E, Nicholson JK. Pattern recognition methods and applications in biomedical magnetic resonance. Prog. Nucl. Mag. Res. Sp., 2001, 39:1-40.

[19] Lukas L, Devos A, Suykens JAK, Vanhamme L, Van Huffel S, Tate AR, Majós C, Arús C. The use of LS-SVM in the classification of brain tumors based on magnetic resonance spectroscopy signals. Proceedings of the European Symposium for Artifical Neural Networks (ESANN), Bruges, Belgium, April 24-26 2002, pp. 131-135.

[20] Lukas L, Devos A, Suykens JAK, Vanhamme L, Van Huffel S, Tate AR, Majós C, Arús C. The use of LS-SVM in the classification of brain tumors based on ^{1}H-MR Spectroscopy signals. Proceedings IEE Symposium "Medical applications of signal processing", Savoy Place, London, UK, October 7, 2002, p. 15/1-5.

[21] Majós C, Alonso J, Aguilera C, Serrallonga M, Acebes JJ, Arús C, Gili J. Adult primitive neuroectodermal tumor: proton MR spectroscopic findings with possible application for differential diagnosis. Radiology, 2002, 225:556-566.

[22] MATLAB Support Vector Machines toolbox.
http://www.isis.ecs.soton.ac.uk/isystems/kernel.

[23] MATLAB/C LS-SVMlab toolbox.
http://www.esat.kuleuven.ac.be/sista/lssvmlab.

[24] Mukherji SK (editor). Clinical Applications of Magnetic Resonance Spectroscopy. Wiley-Liss, 1998.

[25] Murphy M, Loosemore A, Clifton AG, Howe FA, Tate AR, Cudlip SA, Wilkins PR, Griffiths JR, Bell BA. The contribution of proton magnetic resonance spectroscopy (^{1}HMRS) to clinical brain tumour diagnosis. British J. of Neurosurgery, 2002, 16(4):329-334.

[26] Opstad KS, Griffiths JR, Bell BA, Howe FA. In vivo lipid T2 relaxation time measurements in high-grade tumors: differentiation of glioblastomas and metastases. Proceedings of the Eleventh Scientific Meeting & Exhibition (ISMRM 03), Toronto, Canada, May 10-16 2003, submitted.

[27] Pelckmans K, Suykens JAK, Van Gestel T, De Brabanter J, Lukas L, Hamers B, De Moor B, Vandewalle J. LS-SVMlab Toolbox User's Guide. Internal Report 02-145, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2002.

[28] 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, 125:343-349.

[29] Ripley BD. Pattern Recognition and Neural Networks. Cambridge University Press, 1996.

[30] Simonetti AW, Melssen WJ, van der Graaf M, Heerschap A, and Buydens LMC. Brain tumor classification and probability maps using MRI and MRSI data. Analytical Chemistry, submitted.

[31] Smith ICP, Stewart LC. Magnetic resonance spectroscopy in medicine: clinical impact. Prog. Nucl. Mag. Res. Sp., 2002, 40:1-34.

[32] 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.

[33] Suykens JAK, Vandewalle J. Least Squares Support Vector Machine Classifiers. Neur. Proc. Lett., 1999, 9(3):293-300.

[34] Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J. Least Squares Support Vector Machines. Singapore: World Scientific, 2002.

[35] Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. Invest. Radiol., 1979, 14(2):109-121.

[36] Szabo De Edelenyi F, Rubin C, Estève F, Grand S, Décorps M, Lefournier V, Le Bas J-F, Rémy C. Nature Medicine, 2000, 6:1287-1289.

[37] Szabo De Edelenyi F, Estève F, Rémy C, Buydens L. 19th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB02), Cannes, France, August 22-25, 2002, pp. 91.

[38] Tate AR, Griffiths JR, Martínez-Pérez I, Moreno A, Barba I, Cabañas ME, Watson D, Alonso J, Bartumeus F, Isamat F, Ferrer I, Vila F, Ferrer E, Capdevila A, Arús C. Towards a method for automated classification of 1H MRS spectra from brain tumours. NMR Biomed., 1998, 11:177-191.

[39] 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. Proceedings of the Ninth Scientific Meeting & Exhibition (ISMRM01), Glasgow, Scotland, April 21-27 2001, pp. 2284.

[40] Tate AR, Majós C, Moreno A, Howe FA, Griffiths JR, Arús C. Automated Classification of Short Echo Time in Vivo ^{1}H Brain Tumor Spectra: A Multicenter Study. Magn. Reson. Med., 2003, 49:29-36.

[41] The Brain Tumor Society. http://www.tbts.org.

[42] Tosi MR, Ricci R, Bottura G, Tugnoli V. In vivo and in vitro magnetic resonance spectroscopy investigation of an intracranial mass. Oncology Report 2001, 8:1337-1339.

[43] Van Gestel T, Suykens JAK, Baesens B, Viaene S, Vanthienen J, Dedene G, De Moor B, Vandewalle J. Benchmarking least squares support vector machines. Machine Learning (in press).

[44] Vapnik V. The Nature of Statistical Learning Theory. Springer, New York, 1995.

[45] Vapnik V. Statistical Learning Theory. John Wiley & Sons. 1998.

[46] 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, 40:145-152.

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

Table 1
Number of long echo time ^{1}H MRS data of glioblastomas (class 1), meningiomas (2), metastases (3) and astrocytomas grade II (4). The rows correspond to the acquisition center, while the columns mention the type of brain tumour. The acquisition scheme is a PRESS sequence and TE denotes the echo time.

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

Table 2
Classification using **complete spectra**. 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.

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-astroII	$0.8319 \pm 0.0776_{(2)}$	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-astroII	$0.9079 \pm 0.0645_{(3)}$	0.9592 ± 0.0410	0.9619 ± 0.0422	0.9617 ± 0.0411
meta-astroII	$0.9173 \pm 0.0689_{(2)}$	0.9459 ± 0.0549	0.9698 ± 0.0389	0.9642 ± 0.0429

Table 3
Classification using **selected frequency regions**. 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.

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-astroII	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-astroII	0.8441 ± 0.0858	0.8297 ± 0.0926	0.8485 ± 0.0851	0.8281 ± 0.0921
meta-astroII	0.9592 ± 0.0461	0.9727 ± 0.0376	0.9597 ± 0.0453	0.9521 ± 0.0528

Table 4
Classification using **peak integrated values**. 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.

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
astroII-aggr	$0.9230 \pm 0.0674_{(6)}$	0.9343 ± 0.0458	0.9416 ± 0.0502	0.9129 ± 0.1137

Table 5
Classification using **complete spectra**. 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.

	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

Table 6
One-step classification using **complete spectra**. Percentage of correctly classified spectra using all L2-normalized magnitude MRS spectra to assess the training performance.

	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

Table 7
Two-steps classification using **complete spectra**. Percentage of correctly classified spectra using all L2-normalized magnitude MRS spectra to assess the training performance.

	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	5.7258 ± 2.6110	3.5645 ± 2.0870	3.8710 ± 2.1144

Table 8
One-step classification using **complete spectra**. 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.

	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	5.7258 ± 2.6110	3.5645 ± 2.0870	3.8710 ± 2.1144

Table 9
Two-steps classification using **complete spectra**. 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.

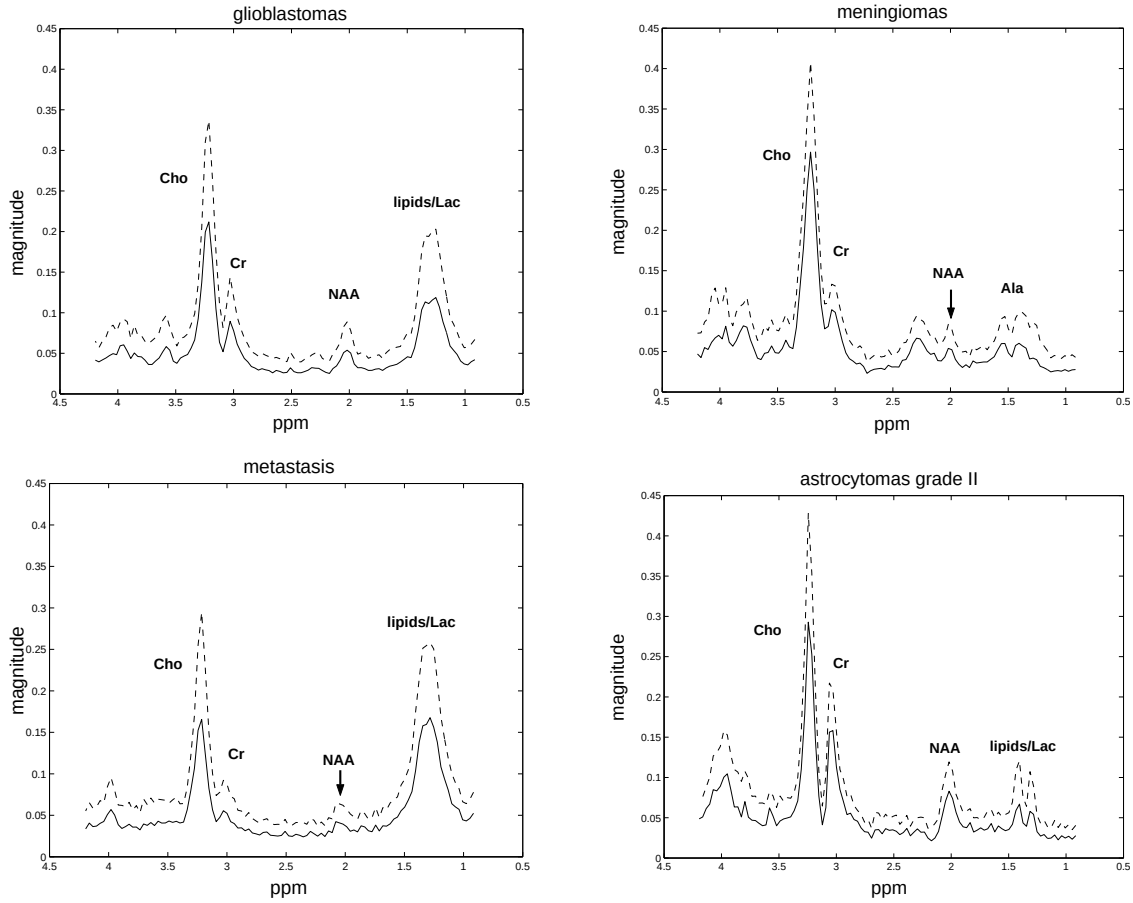

Fig. 1. *Mean L2-normalized magnitude frequency spectra of the four considered classes: class 1 (top-left), class 2 (top-right), class 3 (bottom-left) and class 4 (bottom-right) correspond to the glioblastomas, meningiomas, metastases and astrocytomas (grade II) respectively. The solid lines are the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

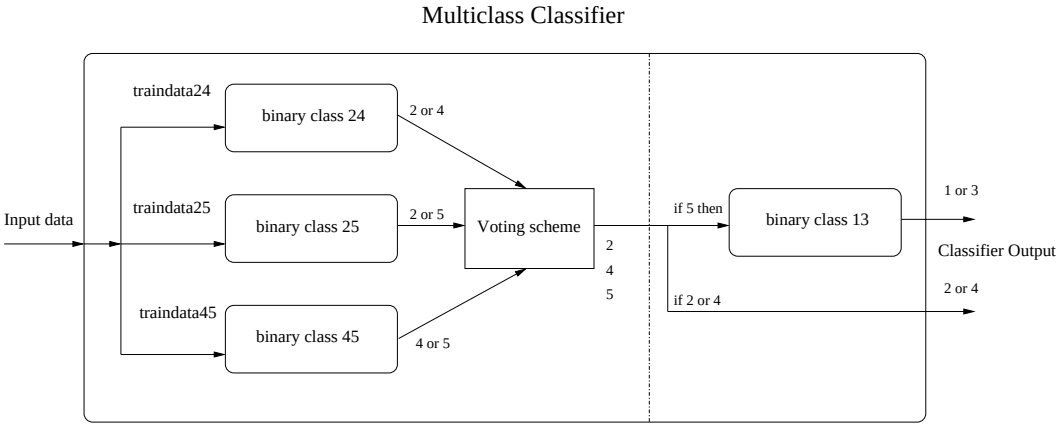

Fig. 2. *Two-steps classification. The left part shows step one, classification of three tumour classes: (2) meni, (4) astroII, and (5) aggressive tumors. The right part, or step two, further refines the classification if the output is class 5 and assigns the spectra of this class either to class 1 (glio)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

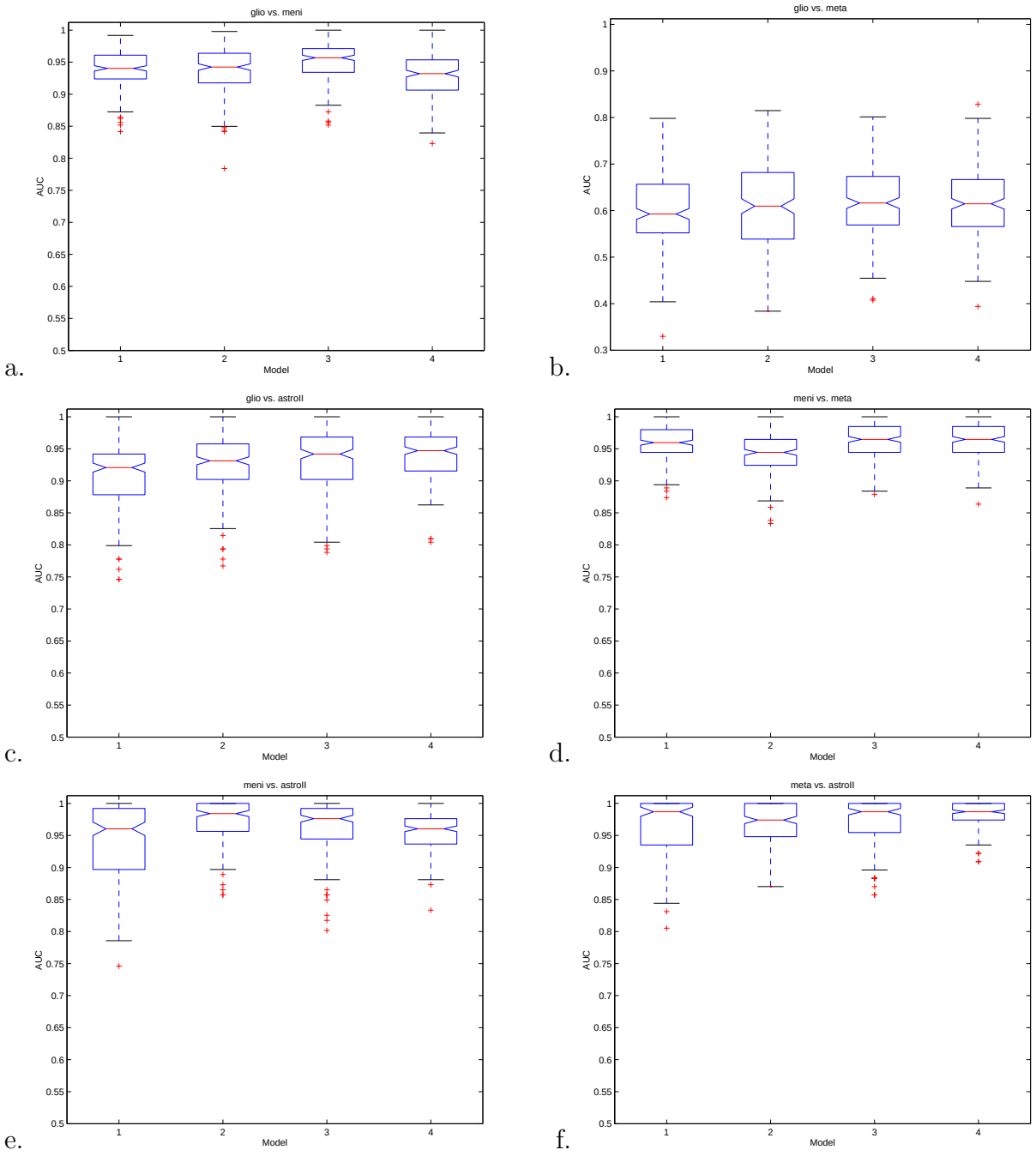

Fig. 3. *Boxplots of the Area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) PCA-LDA, (2) SVM with Linear kernel, (3) LS-SVM with Linear kernel, (4) LS-SVM with RBF kernel. Six figures correspond to the binary classifiers using* **complete spectra***: (a) glio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

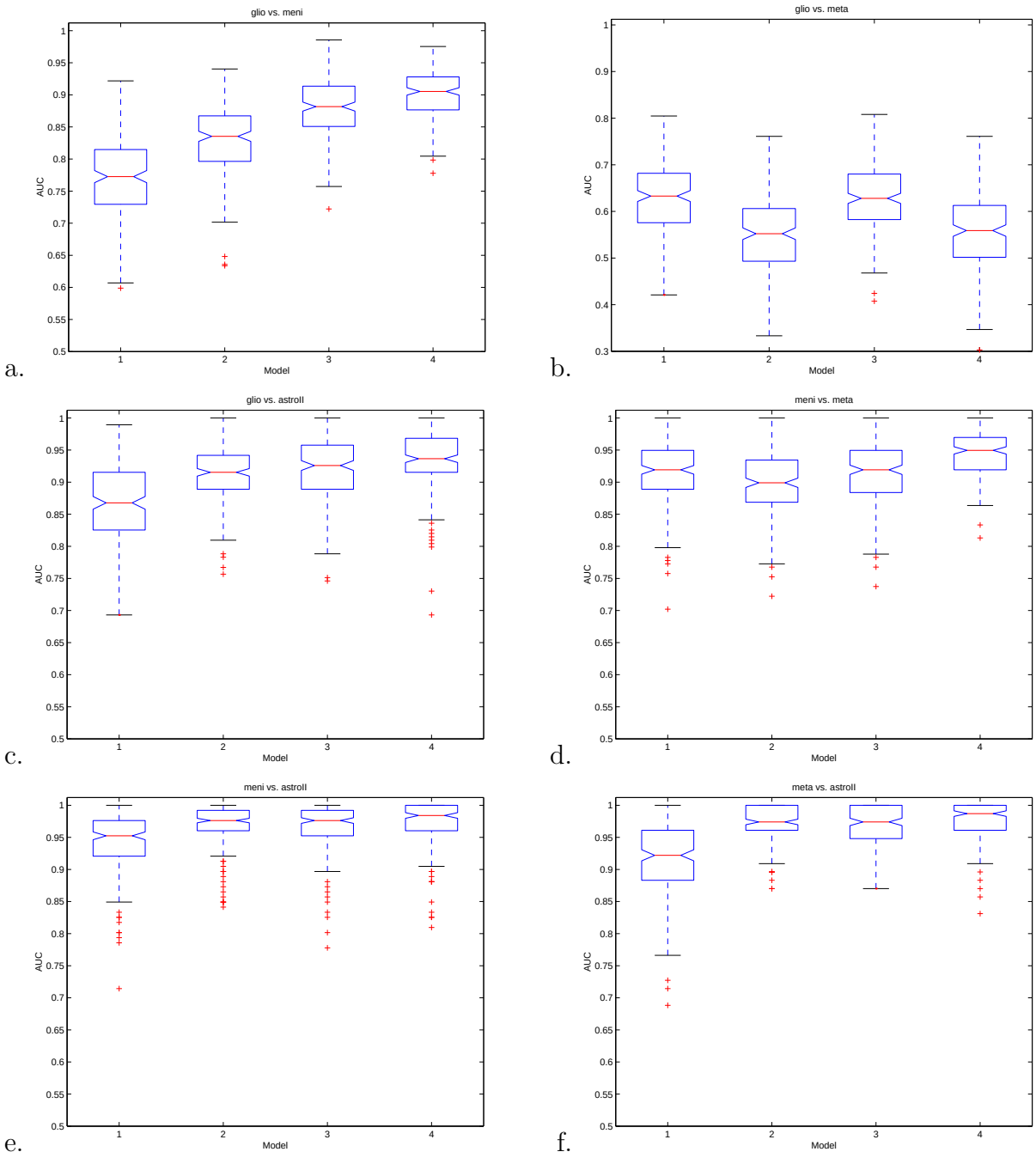

Fig. 4. *Boxplots of the Area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) PCA-LDA, (2) SVM with Linear kernel, (3) LS-SVM with Linear kernel, (4) LS-SVM with RBF kernel. Six figures correspond to the binary classifiers using* **selected frequency regions***: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

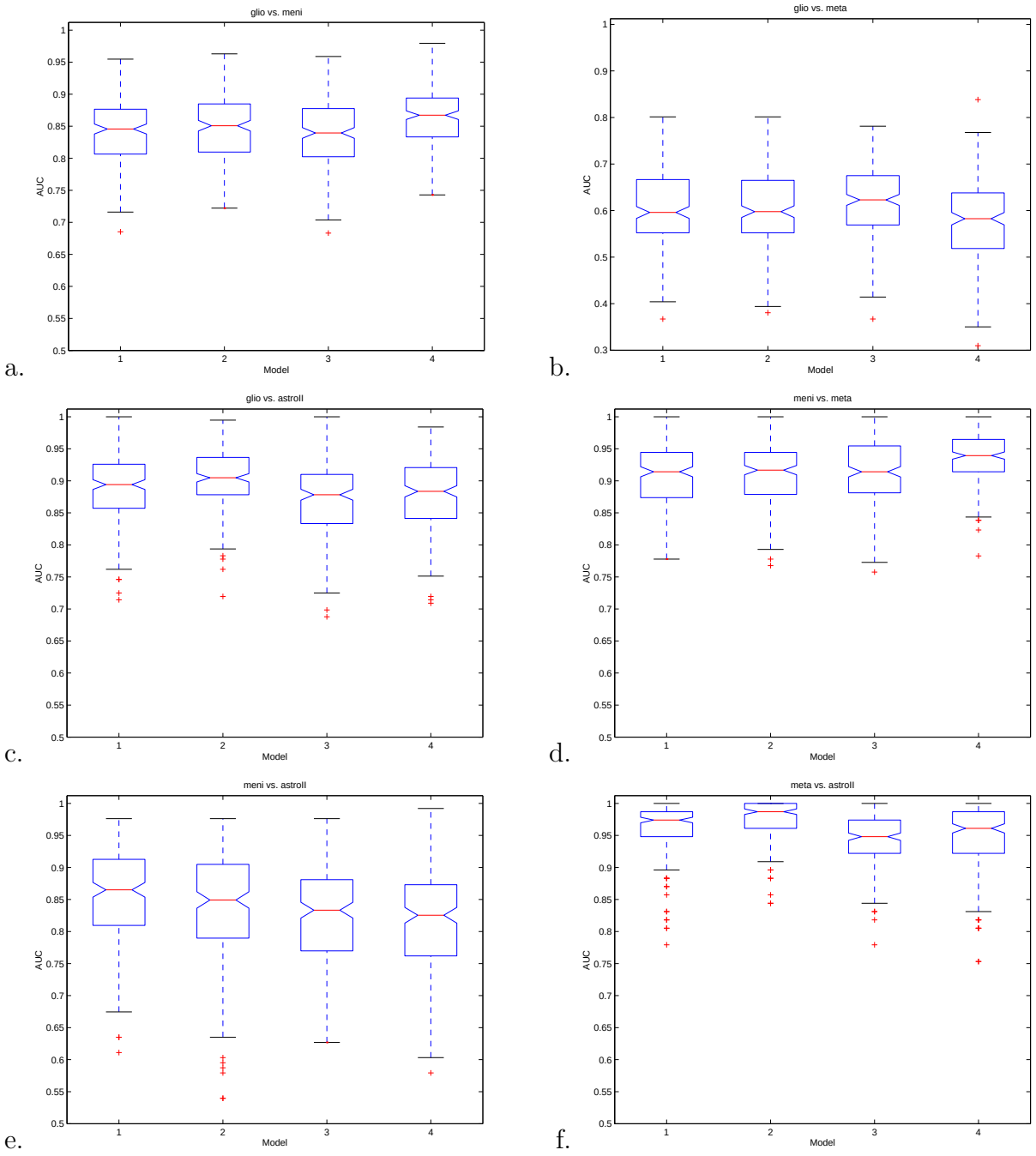

Fig. 5. *Boxplots of the Area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) LDA, (2) SVM with Linear kernel, (3) LS-SVM with Linear kernel, (4) LS-SVM with RBF kernel. Six figures correspond to the binary classifiers using* **peak integration***: (a) glio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

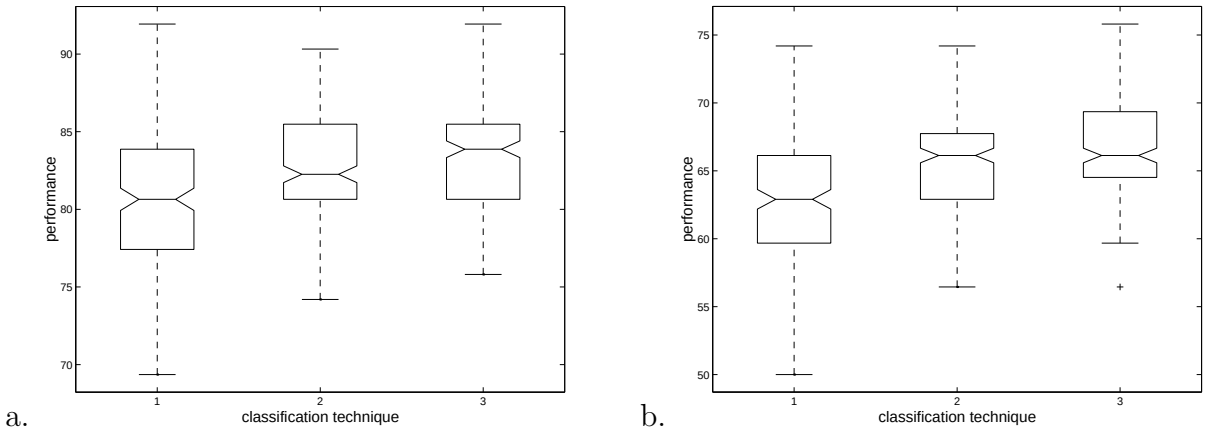

Fig. 6. *Boxplots of the correct classification rate on the test set from 200 runs of stratified random samplings (2/3 of the data for training, 1/3 for testing) of the three models (1) LDA, (2) LS-SVM with Linear kernel, (3) LS-SVM with RBF kernel. Two figures correspond to the multiclass classifiers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