Elsevier Editorial System(tm) for Computers in Biology and Medicine
Manuscript Draft

Manuscript Number:

Title: On the relevance of automatically selected single voxel MRS and multimodal MRI and MRSI features for brain tumor differentiation

Article Type: Full Length Article

Keywords: automatic feature selection, brain tumour, differential diagnosis, MRS, MRSI, MRI

Corresponding Author: Dr. Geert Postma,

Corresponding Author's Institution: Radboud University

First Author: Geert Postma

Order of Authors: Geert Postma; Jan Luts, Ph.D.; Albert J Idema, M.D.; Margarida Julià-Sapé, Ph.D.; Ángel Moreno-Torres, M.D.; Witek Gajewicz, M.D.; Johan A Suykens, Ph.D.; Arend Heerschap, Ph.D.; Sabine Van Huffel, Ph.D.; Lutgarde M Buydens, Ph.D.

Abstract: In order to evaluate the relevance of magnetic resonance (MR) features selected by automatic feature selection techniques to build classifiers for differential diagnosis and tissue segmentation two data sets containing MR spectroscopy data from patients with brain tumours were investigated. The automatically selected features were evaluated using literature and clinical experience. It was observed that a significant part of the automatically selected features correspond to what is known from the literature and clinical experience. We conclude that automatic feature selection is a useful tool to obtain relevant and possibly interesting features, but evaluation of the obtained features remains necessary.

Conflict of Interest Statement.

Paper: "On the relevance of automatically selected single voxel MRS and multimodal MRI and MRSI features for brain tumor differentiation."

Authors : Geert J. Postma, Jan Luts, Albert J. Idema, Margarida Julià-Sapé, Ángel Moreno-Torres, Witek Gajewicz, Johan A.K. Suykens, Arend Heerschap, Sabine Van Huffel, Lutgarde M.C. Buydens

Statement: none declared.

On the relevance of automatically selected single voxel MRS and multimodal MRI and MRSI features for brain tumor differentiation.

Authors:

Geert J. Postma[a#*], Jan Luts[b#], Albert J. Idema[c], Margarida Julià-Sapé[d1d2], Ángel Moreno-Torres[e], Witek Gajewicz[f], Johan A.K. Suykens[b], Arend Heerschap[g], Sabine Van Huffel[b], Lutgarde M.C. Buydens[a]

[a] Institute for Molecules and Materials, Radboud University Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, Netherlands, tel.: +31 243652345, fax: +31 243652653

[b] Department of Electrical Engineering (ESAT), Research Division SCD, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

[c] Department of Neurosurgery, Radboud University Nijmegen, University Medical Center, Geert Grooteplein Z18, PO Box 9101, 6500 HB Nijmegen, Netherlands

[d1] Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Cerdanyola del Vallès, España

[d2] Departament de Bioquímica i Biologia Molecular, Unitat de Bioquímica de Biociències, Edifici Cs, Universitat Autònoma de Barcelona, UAB, 08193, Cerdanyola del Vallès, Spain

[e] Research Department, Centre Diagnòstic Pedralbes, Esplugues de Llobregat, Barcelona, Spain. CIBER-BBN, Esplugues de Llobregat, Spain

[f] Department of Radiology, Medical University Lodz, 90-156Lodz, Kopcinskiego str. 22, Poland

[g] Department of Radiology, Radboud University Nijmegen, University Medical Center, Geert Grooteplein Z18, PO Box 9101, 6500 HB Nijmegen, Netherlands

[#] equal contribution.

[*] Corresponding author

Abstract

In order to evaluate the relevance of magnetic resonance (MR) features selected by automatic feature selection techniques to build classifiers for differential diagnosis and tissue segmentation two data sets containing MR spectroscopy data from patients with brain tumours were investigated. The automatically selected features were evaluated using literature and clinical experience. It was observed that a significant part of the automatically selected features correspond to what is known from the literature and clinical experience. We conclude that automatic feature selection is a useful tool to obtain relevant and possibly interesting features, but evaluation of the obtained features remains necessary.

Keywords:

Brain tumour, differential diagnosis, automatic feature selection, MRS, MRSI, MRI.

List of abbreviations

Alanine (Ala); analysis of variance (ANOVA); automatic relevance determination (ARD); choline (Cho), cerebrospinal fluid (CSF); creatine (Cr); gadolinium (Gd); glucose (Glc); glutamine/glutamate (Glx); glycine (Gly); Hankel Lanczos singular value decomposition (HLSVD); lactate (Lac); least squares support vector machine (LS-SVM); linear discriminant analysis (LDA); lipids (Lip); myo-Inositol (mI); magnetic resonance (MR); magnetic resonance imaging (MRI); magnetic resonance spectroscopy (MRS); magnetic resonance spectroscopic imaging (MRSI); N-acetyl-aspartate (NAA); point resolved spectroscopy (PRESS); proton density (PD); quadratic discriminant analysis (QDA); stimulated echo acquisition mode (STEAM); support vector machine (SVM); taurine (Tau); echo time (TE); mixing time (TM); repetition time (TR).

1 Introduction

Automatic feature selection is frequently advised if classifiers are built on data consisting of a high number of variables[1]. This can have several benefits, such as improving the classifier performance and an improved interpretability (the variables that appear to be important for the decision process could lead to hypotheses on the underlying biological process and new insights). During the automatic selection process those features (variables) are selected which are relevant for the classification process. Overviews of the various methods, combined with comparisons of their success are e.g. given in [1-4]. However, the relevance of a selected feature is judged in the context of the classification process (see also [3, 4]). And this is not necessarily similar to whether the selected features represent relevant biological information. An investigation into this matter is described in this article using two data sets consisting of magnetic resonance data for the classification of several types and grades of human brain tumours.

Metabolic information as acquired by single-voxel proton magnetic resonance spectroscopy (MRS) or multivoxel magnetic resonance spectroscopic imaging (MRSI) measurements of human brain tumours, is used to build classifiers for diagnostic purposes [5-7]. The spectral data can be combined with imaging data [6, 7]. As these data consist of an enormous amount of variables, automatic feature selection is normally performed. The peaks in the spectra correspond to various metabolites occurring in the brain. Of most of these metabolites their relation to and occurrence in the various types of brain (tumour) tissue is already known [8-12]. The differentiation between tumour classes or grades can be done in a multivariate way using classification algorithms such as linear discriminant analysis (LDA), neural networks or support vector machines (SVMs), applied to a set of (quantified) peaks or to the complete spectra [7, 8, 13, 14]. Some classification algorithms include feature selection, for others the feature selection must be applied through a separate automatic process. The feature selection can help to separate the noise and non-informative features from those that possibly contain information useful for

the classification. For the interpretability of the results we have restricted ourselves to feature selection on areas where metabolite peaks can be present.

Feature selection procedures can also be used to investigate the influence of specific features on the classification performance and/or their replaceability in the model. Removal of a feature that requires an additional measurement, and/or data processing can result in reduced examination costs, reduced patient burden, or reduced risk for the patient. Also, if some part of the data, e.g. an MRI image, is of bad quality, the effect of its removal can be quantified.

There are several strategies for automatic feature selection [15]. In filter methods the features are selected independently from the classifier, resulting in time efficient algorithms. However, according to [4] the classifier should be involved in the selection process. This is done in wrapper methods, at the cost of increased computational effort. In addition, embedded methods integrate the feature selection or weighting procedure into the learning algorithm.

In a previous paper [16] several feature selection algorithms, belonging to filter (although applied in a wrapper setting) and embedded methods, were applied to multimodal data consisting of MRSI data and MR imaging (MRI) data. The methods were compared based on the performance of the resulting feature sets in binary brain tissue and tumour classifiers. The feature selection was applied to a set of quantified metabolite peaks, extracted from the MR spectra. It was concluded that these methods produced feature sets, including image features, which resulted in accurate classifiers.

In this paper we aim to assess the usefulness of feature selection to obtain relevant and promising features for differential diagnosis of brain tumour tissue based on single voxel data and tissue segmentation based on multivoxel data. Therefore, we extended our previous investigation in several ways. First, by adding a wrapper method as feature selection strategy, second, by testing the feature selection methods also on single-voxel MRS data from patients with 4 types of brain tumours, and third, by inspecting the selected features and comparing them with those known to be relevant from literature for the respective differential diagnosis. Finally we studied the effect of different combinations of image features on the final classification

outcome and investigated whether reduction or replacement of the number and type of measured MR images was possible (the added value of combining MRSI data with MRI data is already proven by e.g. [17, 18]).

The data we investigated encompass a wide range of binary classification problems. Several of these classification problems are clinically important for differential diagnosis (e.g. those involving grade II, grade III and IV gliomas). Others are less important, occur less frequent (i.e. involving atypical meningiomas in the differentiation from glioblastomas or metastases [19]), or are not important for differential diagnosis, but are important for correct tissue segmentation in nosologic imaging (e.g. based on MRSI data, combined with MRI data, [6, 7, 20]). Those classifiers, having no direct clinical significance, served as validation of the obtained conclusions.

2 Material

The feature selection methods are compared on two data sets. The MRS data set contains 211 single-voxel ^{1}H MR short echo time spectra from the INTERPRET database [21]. Data were acquired by 6 contributing centers: Centre Diagnostic Pedralbes (CDP), Fundacion para la Lucha contra las Enfermedades Neurologicas de la Infancia (FLENI), Institut de Diagnostic per la Imatge (IDI), Uniwersytet Medyczny w Lodzi (MUL), St George's Hospital Medical School (SGHMS) and Universitair Medisch Centrum Nijmegen (UMCN). The studies were performed in accordance with the medical ethics regulations of the countries concerned and all patients or their legal representatives had signed informed consent forms, agreeing to the study or the use of anonymized data for research. The data were acquired at 1.5T by stimulated-echo acquisition mode (STEAM) or point resolved spectroscopy sequence (PRESS) using an echo time between 20 and 32 ms. Repetition time varied from 1600 to 2020 ms, spectral width was 1000 or 2500 Hz and number of points equaled 512 or 2048. Each spectrum was subjected to a strict validation process set up by the INTERPRET Quality Control committee [14]. Tumour assignment was performed according to the rules of the World Health Organization. Four types

of brain tumours were included: glioblastomas (84 patients), meningiomas (57 patients), metastases (37 patients) and low grade glial tumours (33 patients). Water suppressed free induction decays were zero order phase corrected by using Klose's method [22] and residual water resonance was removed by Hankel Lanczos singular value decomposition (HLSVD). Signals were baseline corrected and normalization was done based on the L2-norm. Spectra were referenced to total creatine at 3.03 ppm, choline at 3.21 ppm or mobile lipids at 1.29 ppm in this order of priority depending on the signal-to-noise ratio and tumour pattern. The spectra are presented in figure 1. All spectra were quantified using simple peak integration. The following regions were integrated [23, 24]: Cr (Cr1; 3.845-3.995 ppm), Glx (Glx1; 3.665-3.815 ppm), Tau + Glc (3.345-3.495 ppm), mI + Gly (3.450-3.680 ppm), mI + Tau (3.195-3.345 ppm), Cho (3.115-3.265 ppm), Cr (Cr2; 2.955-3.105 ppm), Glx (Glx2; 2.270-2.510 ppm), NAA (1.935-2.085 ppm), Ala (1.395-1.545 ppm), Lac (1.235-1.385 ppm), Lip1 (1.125-1.275 ppm) and Lip2 (0.825-0.975 ppm).

{Figure 1 about here}

The second data set, containing MRSI and MRI data, was acquired in the context of the INTERPRET project. The data were acquired at the UMCN (at 1.5 Tesla) and the study was approved by the ethical committee and followed the rules of the World Health Organization. Data from 25 patients with a brain tumour and 4 volunteers were used. The data set encompassed 6 classes: normal brain tissue from volunteers and apparently normal tissue from the contralateral half of the brain of patients (8 persons), CSF (8 patients), grade II gliomas (10 patients), grade III gliomas (5 patients), meningiomas (two grade II patients and one grade I patient) and glioblastomas (7 patients). A strict quality control procedure was applied to each case and the tumour type was determined by a histopathological consensus. Of each patient water suppressed and unsuppressed proton MRSI data were acquired using a 2D STEAM sequence with the STEAM box positioned totally in the brain (16x16x1024 samples). Signals

arising from fat tissue surrounding the skull were avoided. Additionally, four different MR images were acquired: T1-weighted image, T2-weighted image, PD-weighted image, gadolinium (Gd)-enhanced T1-weighted image. The location of the STEAM box was determined using the Gd-enhanced T1-weighted image showing the largest tumour area. The MRSI slice was centered around an MRI slice. The MR images were registered with respect to the MRSI slice. After processing, among which zero-filling to 32x32 voxels, several voxels, situated in the healthy areas, CSF areas and tumour areas, were selected from the respective patients to obtain a sufficiently large data set (669 voxels in total). The data set consists of 10 quantified areas from the processed short echo MRSI spectra from the selected voxels, and the corresponding MR pixel intensities of the 4 MR images. The MR pixel intensities were averaged for all voxels to lower the resolution of the four MRI images to the one of the MRSI grid. The low resolution images were scaled in agreement to the range of the integrated regions of the spectral data. The image features are referred to as T1, T2, PD and Gd. The ten different quantified spectral areas correspond to [23]: Cr (Cr1; 3.885-4.015 ppm), Glx (Glx1; 3.685-3.815 ppm), mI+Gly (3.495-3.625 ppm), Tau (3.375-3.505 ppm), Cho (3.135-3.265 ppm), Cr (Cr2; 2.955-3.095 ppm), Glx (Glx2; 2.135-2.265 ppm), NAA (1.955-2.085 ppm), Lip1 + Lac + Ala (1.265-1.395 ppm) and Lip2 (0.835-0.965 ppm). Further details on the acquisition parameters, the processing of data and the selection of the voxels can be found in [17].

3 Theory and methods

Three filter methods, one embedded method and one wrapper method were applied to both data sets on all possible differential diagnoses: the Kruskal-Wallis test, the Fisher discriminant criterion, the Relief-F algorithm, the automatic relevance determination (ARD) mechanism, and forward selection respectively. The Kruskal-Wallis test and the Fisher discriminant criterion check the features independently from each other, while the other methods score the features in combination. A brief description of these methods will be provided here. The reader is referred to [16, 29] for a more extensive theoretical and algorithmic description.

3.1 Fisher discriminant criterion

Fisher's criterion is a simple measure that scores each feature by calculating for that feature the (squared) distance between the class means and correcting that distance with the within class variance [30]. The features are ranked based on these scores.

3.2 Kruskal-Wallis test

The Kruskal-Wallis test is a non-parametric alternative to the well-known one factor analysis of variance (ANOVA). In contrast to ANOVA, this test does not assume a normal distribution for the data. The main disadvantage is that the test is less powerful than ANOVA, when the ANOVA assumptions are met. A test statistic is calculated for each feature, comparing the medians of both classes, and the features are ranked according to the outcome.

3.3 Relief-F

Compared to most feature ranking methods, Relief-F does not assume conditional independence of the features. The significance of the features is obtained based on how good they discriminate between samples that are close. It is an iterative procedure in which random samples are picked from the data set, and each time the k nearest neighbors of the same and the opposite class are calculated. Based on these neighboring cases the weights of the features are updated. Again, at the end the features are ranked based on the calculated weights.

3.4 Automatic relevance determination for LS-SVMs

The implementation of this embedded feature selection technique is linked to LS-SVM classifiers in which a kernel (a similarity measure) is applied on the data to take care of non-linearity. ARD introduces a weight for every feature in this kernel. Setting these weights to equal values first the LS-SVM parameters are optimized. Afterwards the weights are iteratively

optimized with a Bayesian probability calculation using the data. The resulting weights are used to select the most relevant features.

3.5 Forward selection

The wrapper method that was applied to the data is forward selection. In forward selection the features are tested with a classifier. First all individual features are tested by building classifiers using only the data of that feature. The feature that provides the best performance with the applied classifier is selected. Then the remaining features are tested in combination with the previously selected feature and the combination with the best classifier performance is selected. This procedure is repeated, each time adding one feature to the already selected set. The procedure is stopped if adding a feature does not improve the classifier performance. The classifier performance can be any measure, e.g. the number of correctly classified samples. There are modifications to this technique, such as allowing changes to the already selected features. This is not applied in this study as additional testing is required which further increases the time of the feature selection process.

3.6 Quadratic discriminant analysis

The applied classifier in the forward selection was quadratic discriminant analysis (QDA). QDA is a variant of LDA which is designed to find boundaries between classes. In general, an unknown object is assigned to the class for which the Mahalanobis distance [17] is minimal. This distance takes the covariance of the data of a class into account, i.e. the multidimensional structure of the data and correlation between the features.

3.7 Experimental setup

For each pair of classes in the separate data sets, the filter methods (Fisher discriminant criterion, the Kruskal-Wallis test, Relief-F) and the embedded method (ARD) were applied 50 times after resampling of the data (so-called bagging analysis). In each of the runs the methods

assign a rank number (in case of ARD a weight) to each feature. These ranks or weights were averaged over the 50 runs, resulting into 4 values for every single combination of feature and classification problem (see table 1 for an example). The use of bagging within the context of feature selection for MR spectroscopy data has already been proposed in [31].

For the Kruskal-Wallis test, the Fisher criterion and Relief-F values between 1 and 13 (or 14, depending on the number of features in the data set) were obtained. The smaller this number, the better the feature is able to distinguish between the two groups. For ARD, weights between 0 and 1 are obtained. A weight that is closer to 1 indicates a more relevant feature. A feature was considered to be relevant if the ARD weight was higher than 0.2 or if at least two out of three filter methods obtained a rank smaller than '6' or if the ARD weight was higher than 0.1 and one filter method obtained a rank smaller than '6'. A feature was labeled as most relevant if the ARD weight was higher than 0.2 and if all filter methods obtained ranks below '6'.

In case of forward selection a double cross-validation was executed using both times a leave-20%-out procedure with QDA. The inner cross-validation was used to test the performance of each feature set in the forward selection procedure. It was repeated 10 times after resampling of the data. The average number of correctly classified samples was used to evaluate the performance of the feature set under investigation. The outer cross-validation procedure (involving 5 consecutive sets of 20% of the data used as test set) was repeated 5 times after a resampling of the data, resulting in 25 feature sets, each with an evaluation score on an independent test set of 20% of the data (as such, 25 times the feature selection process was repeated). For the single-voxel multicenter MRS data set the three most frequently occurring feature sets were evaluated by a leave-20%-out cross-validation using the complete data set (apart from the outer cross-validation), and the best performing one is reported. In case of the MRSI and MRI data set the three most frequently occurring feature sets were evaluated by a leave-patient-out procedure (20 repetitions, again using QDA). These sets were compared with the features found by the other feature selection methods.

In order to investigate whether it was possible to reduce the number and type of measured MR images (MRI-based features), the forward selection procedure was executed on all available peak-integrated MRSI features for all possible combinations of MRI features and without MRI features. The underlying idea is that it could be possible to replace an MRI feature by one or more MRSI features. During the forward selection procedure each time the performance of the features is then evaluated combining the pre-selected MRI features with the MRSI features, selected by the forward selection procedure. This resulted in 16 feature sets for each differential diagnosis. Hereafter, by comparing the obtained performances of the 16 feature sets with the performance of the feature set obtained by applying forward selection on all available features (the 'choice' set in figure 2), it was investigated whether it was possible to reduce the number of MRI features.

4 Results

The selected input features for the various types of differential diagnosis are provided in the order of clinical importance. Note, however, that in case of MRSI (and MRI) data all the differential diagnoses are required for a correct labeling of the different types of tissue in a 2-dimentional image based on these data. The results for the multicenter MRS data set are depicted in Tables 1 and 2. As an example, Table 1 provides the averaged rank or weight values, obtained by the feature selection methods, for all features for the separation of glioblastomas and low grade glial tumour. Weights and ranks that potentially belong to relevant features are printed in boldface. For forward selection the most frequently occurring set is listed. To distinguish glioblastomas from low grade glial tumours mI+Gly, Cr (Cr2, around 3.05ppm), and Lac are clearly relevant, and additionaly Glx1, Tau+Glc, Glx2, Lip1 and Lip2 seem to be useful. In summary, Table 2 lists the features that are selected for all possible types of differential diagnosis.

{Tabel 1 about here}

{Tabel 2 about here}

For the combination of the MRSI and MRI data set a similar analysis is performed (note that the single voxel MRS data set did not contain spectra from healthy tissue, CSF and grade III gliomas, whereas the MRSI and MRI data set did not contain data from metastases, resulting in a partly different set of differential diagnoses). Again, as an example, the results for the differential diagnosis of grade III gliomas and glioblastomas are listed in Table 3. Based on the rank and weight values obtained with the filter methods and embedded method mI+Gly, Lip1+Lac+Ala, and T1 are clearly selected, and additionally Cho is picked out. The forward selection adds Lip2 to this. Table 4 presents the selected features for all possible types of differential diagnosis. The corresponding accuracy scores obtained by applying the filter methods and the embedded method prior to classification already have been reported in [16]. The scores obtained with the forward selection procedure are not reported. These scores are comparable or slightly lower than those reported in [16]. The differences are mainly attributed to the applied classifier.

{Figure 2 about here}

Figure 2 presents an example of the investigation of the possibility of reduction or replacement of the MR image features. The example concerns the differentiation of grade II and grade III gliomas. It illustrates the effect of the preselection of all possible combinations of MR images (image features) on the selected features, together with the resulting final classification performances. From this figure it can be concluded that for the differentiation of grade II from grade III gliomas the best results are obtained if all 4 MR images are available, with the additional use of most MRS features. If one or 2 image features would be available then the best choices are Gd or combinations of T2 with PD or PD with Gd, with a gain of accuracy of about 3%. After evaluating these kinds of figures for all tested binary classifiers it turned out that the

addition of MRI features had a positive effect (statistically significant with p = 0.05, Student T-test) on all investigated differential diagnoses (binary classifiers). A summary of the results is presented in Table 5. The differentiation of CSF and grade III gliomas can be done with PD images alone, so without additional MRS features. The other image features do not have an effect. Using MRS features alone an accuracy of 94% can be obtained.

5 Discussion

In this section the metabolites and image features, found by the different feature selection methods, are evaluated on their relevance based on the knowledge that is available in the MR literature. For this purpose, the studies in [8-12] are used.

Single-voxel data set.

Regarding the results of the multicenter single-voxel MRS data set: in general, it is observed that the results of all methods largely correspond. Sometimes there seemed to be differences in the feature sets selected by the forward selection method compared to those selected by the other four methods. However, for some differential diagnoses the forward selection method produced a large number of possible feature sets with no clear winner with respect to the number of times a certain set was selected. In such situations the feature sets selected by the other four methods appeared to be also present in alternative sets of the forward selection method. Furthermore, a number of metabolites for certain tissue types highly correlate (e.g. Lac, Lip1 and Lip2) so that it hardly matters whether one, two or more of those features are selected (this is also the case for the multivoxel data set). Exceptions to foregoing situations will be discussed below.

Differentiations of glial tumours. In this data set most of the low grade glial tumours are grade II astrocytomas. For the differentiation of low grade glial tumours from glioblastomas, most selected features correspond to what is known from literature: mI is mostly found in

astrocytomas and reduces with grade[12], Cr is reduced in astrocytomas but further reduced in glioblastomas, and Lip1, Lip2 and Lac are characteristics of high grade tumours (and increase with the tumour grade). Glx related peaks are also reduced in glioblastomas (although not observed in [8]). The selection of Ala is probably be due to the effect of the increased neighboring Lac peak and the applied simple peak integration method. The selection of Tau+Glc corresponds to visible differences in the mean tissue levels between both classes (not mentioned in [9]).

Differentiation of metastases and glioblastomas. According to [10] lipids have characteristics that might allow discrimination between metastases and glioblastomas. This corresponds to our findings. Lac is frequently elevated for glioblastomas. The same holds for Cho and mI, whereas Glx sometimes was reduced [11], but slightly elevated in our situation. NAA showed no significant difference in [11], nor in our situation, but was explicitly selected. Investigating the data it appeared that NAA, for a considerable number of glioblastoma patients, showed higher tissue levels compared to those of the metastasis group. The selection of Cr corresponds with a higher average tissue level for glioblastomas compared to metastases, similar to [11], although [10] showed no clear difference. Regarding Tau and Ala: their choice is supported by small differences in average tissue levels (Tau+Glc slightly higher and Ala slightly lower for glioblastomas).

Remaining single voxel MRS based differentiations. Regarding the other clinically less or not interesting differential diagnoses it can be stated that most of the selected features also correspond to what is known from literature (e.g. Glx is a prominent feature of meningiomas). The features labeled as most relevant were mentioned in the literature as showing significant difference. For the differentiation of meningiomas from metastases Cr2 is selected, although it is reported as being low or absent in meningiomas (as well as in metastases). The selection, however, is supported by a clear difference in tissue level (peak area), which is probably due to the applied integration method (which also includes baseline and normalization effects; see fig 2). This is also the reason for the selection of mI-related features and Cr2 for the differentiation

of glioblastomas from meningiomas. For the differentiation of meningiomas from metastases the overlap in integrated ppm ranges is probably the reason for the selection of mI+Tau above the selection of Cho, which is mentioned in the literature. In contrast to the other techniques, the forward selection method did not select the Cr2 peak. The selected features already provide the best differentiation and Cr2 is among the by the other techniques selected features the one with the least difference in mean tissue level between both tumour classes.

{Table 3 about here}

{Table 4 about here}

Multivoxel data set.

For the multivoxel data set MRI information is added. The information from the images seems to be valuable for almost all classification problems (Table 4). This corresponds to e.g. [13, 17, 18]. The comments made at the beginning of the discussion on the features selected from the single-voxel data set are valid here as well: there is considerable correspondence between the feature sets selected by the various techniques, and in case of differences between the feature sets selected by the filter/embedded methods and the forward selection method corresponding features frequently appeared to be present in alternative sets of the forward selection method. Differences which remain present after checking alternative sets of the forward selection method (with similar performance) will be discussed below. The clinically less interesting differentiations are discussed, as well, as they can be part of a segmentation system of brain tissue using this type of data (nosologic imaging) [6, 7, 20]. Moreover, they serve as additional control of the validity of the selected features.

Differentiations of glial tumours. For the differentiation of the various grades of glial tumours, most selected features correspond to what is known from the literature (see also the single voxel section). That in our situation mI(+Gly) is not selected for the differentiation of grade II and III

gliomas corresponds with an absence of a difference in mean tissue level. Cho is selected in the differentiation from glioblastomas as its tissue level is reduced in glioblastomas, but similar to literature the tissue level difference appears to be small between grade II and III gliomas. It is thus not surprising that the filter/embedded methods do not select this feature for the differentiation of those grades. The same story holds for Glx2 and NAA (both its values probably being influenced by the presence of macromolecules). The selection of Glx1 and Tau is due to the chosen peak integration method (the spectra are slightly elevated in that region for grade II and III gliomas) and probably do not represent those metabolites. That the forward selection method still selects peaks/metabolites whose differences are small is due to the effect that even a small difference can contribute to a (small) increase of accuracy. Regarding the selected image features: high grade tumours show frequently an enhancement on Gd-enhanced images whereas they are frequently less intense on T2 images compared to the appearance of grade II gliomas [32]. Grade III gliomas show sometimes contrast enhancement on Gd images compared to grade II gliomas and frequently show less hyperintense tumour borders compared to glioblastomas. The selection of T1 in the various differentiations follows from a decrease in voxel intensity going from grade II to grade III gliomas and an increased intensity (compared to the grade II and III glioma voxels) in glioblastomas. That all image features are selected in the differentiation of grade II and III gliomas is probably due the fact that that differentiation is difficult and in this way as much as possible information (although small differences) is used to improve it. For the differentiation of grade II gliomas from glioblastomas the forward selection provided several combinations of features, all having about the same accuracy and all including image features (e.g. involving T2, see table 5, or with the addition of NAA and Lip1+Lac+Ala).

Differentiations of the various gliomas from healthy tissue. Also in the differentiation of the various gliomas from healthy tissue most selected features correspond to what is known from the literature. NAA, an indicator of viable neurons, is reduced in most tumours, Lac is frequently observed in tumours, Lac and lipids are often most prominent in high grade tumours, and Cr is reduced in astrocytomas compared to normal brain. The selection of mI+Gly in the

differentiation of healthy tissue from grade II gliomas can be dataset dependent and just produces a small increase in performance. It is known that Cho can be increased in lower grade astrocytomas compared to healthy tissue. According to [9] Lip2 is a characteristic feature in the differentiation of healthy tissue from glioblastomas, but it is not selected in our case. This is probably due to the fact that it has hardly an additional value compared to the already selected strong features (such as Lip1+Lac, with which it is correlated). Besides, the feature is present in alternative selections of the forward selection method and is close to the selection threshold for several filter methods. The choices of T1, T2 and PD are to be expected as frequently the tumour area has a lower intensity on T1 and is hyperintense on T2 and PD images.

Differentiations of the various gliomas from CSF. The choices of the metabolites in the CSF - tumour differentiations mainly correspond to those metabolites which have in the specific tumour tissue a high (or elevated) tissue level, as in CSF all metabolite tissue levels are low or absent. The chosen image features correspond to the fact that CSF is only enhanced on T2 images, whereas tumours are enhanced on the other image types compared to CSF on those images. The choice of Tau in CSF-glioblastomas is probably due to the chosen integration method. The feature sets selected by forward selection revealed that for the differentiation of CSF from grade III gliomas and glioblastomas the image features alone can provide an (almost) perfect result. Alternative combinations of image features require spectral features to obtain the same accuracy.

Differentiations of Meningiomas from other tissues. Regarding the meningioma related differentiations a number of general observations from the literature correspond to the selected features (see also the single voxel section). mI is found to be low or absent in non-glial tumours. It is thus not surprising that Glx is selected in most differentiations involving meningiomas, as Glx is low in all other tissue types, and that Cr (Cr2) and mI are selected in all differentiations but the differentiation from glioblastomas, because in the latter the Cr2 and mI+Gly tissue level is also low. Glx is not selected in the differentiations involving healthy tissue or grade II gliomas. However, the ARD mechanism showed elevated values for Glx2 and the feature is

present in alternative selections of the forward selection method. The selection of Lip1+Lac+Ala (area around 1.3 ppm) in the differentiation from grade II gliomas, healthy, and CSF can be explained from the fact that two of the meningiomas are atypical (i.e. grade II): these tumours usually comprise necrotic tissue containing lipids [33, 34]. Checking the average tissue levels it appears that in our situation their tissue level is even higher than those in grade II gliomas (in contradiction to e.g. observations in [8]). A macromolecular contribution to the signal in this ppm range is also reported for meningiomas [8]. The selection of Cho corresponds to the fact that the Cho tissue level is normally low for (benign) meningiomas, although it is reported that they can have intensities corresponding to grade III gliomas [8, 34]. In our data, despite the 2 atypical meningiomas, the latter is not the case. Cho is not selected in the differentiation from grade II gliomas, probably because of the presence of the image features (without image features it is selected and the filter methods, which score the features in isolation, show scores just above the selection threshold). In the differentiation from CSF Cr2 and mI+Gly are selected, as well, because, although these metabolites have a low tissue level in CSF (compared to e.g. healthy tissue) their tissue levels are still higher than in meningiomas. The selection of Tau is probably dataset dependent. The selected image features in the differentiations from gliomas can be explained by comparing the average intensities of both classes, but normally only compared to grade II gliomas meningiomas are hyperintense on Gd images. In the differentiation from healthy tissue and CSF the selected image features correspond to their normal appearance: e.g. hypointense to isointense on T1, isointense to hyperintense on T2 and enhanced on Gd images [35].

The differentiation of healthy tissue from CSF. In the differentiation of healthy tissue and CSF the selected features are to be expected, as in CSF most metabolites have lower tissue levels than in healthy tissue and this effect is more pronounced for those metabolites which have high tissue levels in healthy tissue (e.g. Cr2, NAA), T2 shows an hyperintense CSF region, whereas on e.g. T1 the CSF is more hypointense. On PD images CSF appeared slightly less intense due

to the relative short repetition time applied (3100 ms). The selection of Tau is probably more a baseline effect.

{Table 5 about here}

Possible reduction or alternative choices of MRI features.

The MRI-based features are relevant for all differentiations (binary classification). This can be concluded from table 4. However, the investigation on the relative added value of the image features using QDA (which could help evaluating the trade-off of extra measurement time against gain in classification performance; table 5) revealed that the added value can vary considerably. Especially the differentiations involving grade II gliomas benefit from the availability of image features. In the clinically more interesting differentiation of grade II from grade III gliomas a gain of 6% can be obtained, but that gain is only possible if most or all MRI images (features) are used. In other differentiations the added value is very small and corresponds to 1-2 voxels that are additionally correctly classified. In two of the CSF involving differentiations a 100% correct classification can also be obtained using MRI features only. This is to be expected as especially on the T2 and PD images the CSF region is easy to recognize.

Table 5 provides a relative ordering of the importance of the various image features (alone or in combination). Using this information it can be verified whether the number of MR images that need to be acquired can be reduced or that alternative combinations are possible. T1 and Gd are mostly sufficient, but the extra availability of T2 or PD images is beneficial. The CSF versus grade III gliomas separation requires PD (the other image types have no effect or seem even to be counter productive; without PD the loss in accuracy drops with QDA). The CSF versus grade II gliomas discrimination is less accurate if PD is replaced by T1 + Gd. The grade II versus grade III gliomas separation accuracy drops 4% if T1+Gd are used instead of all image features. The differentiations of gliomas from healthy tissue are only to a small extent influenced by alternative choices of image features (0.2% reduction of the accuracy). If one would like to

replace the Gd image (because of a small NSF risk), this would almost only effect the grade II versus grade III differentiation (2-3% loss of accuracy).

It was observed that the performances of the individual test sets in a leave-20%-out cross-validation could vary a lot. This was investigated and it was observed that this was due to a patient effect: voxels of the same tissue type of a patient are more alike than between patients. This was tested using a leave-patient-out cross-validation. In this cross-validation all voxels of 1-2 patients are taken out as test set (being each time more or less 20% of the data). The classifier performances for the tests of the effect of the different combinations of preselected image features (e.g. Figure 2) were significantly worse (up to 10%), but showed the same tendencies as the normal cross-validation. Therefore, the presented accuracy results are those obtained through the leave-patient-out cross-validation.

Correspondence between the selected features of both data sets.

Three differentiations are in common between the single voxel data set and the multivoxel data set. For these differentiations the selected spectral features mostly correspond. The differences are mostly due to the presence of the image features in the multivoxel data set: if the features can only be selected from the spectra features, most differences disappear. Two differences remain present in the differentiation of glioblastomas and meningiomas: mI+Gly and Cho, which are not selected for the multi voxel data set. This can be explained because for both features the peak area distributions appear to be similar.

6 Conclusions

A significant part of the automatically selected features for the single-voxel data as well as the multivoxel data corresponds to what is known from the literature and clinical experience. However, applying a single feature selection technique does not guarantee that all known biologically or clinically relevant features are found: a combination of several techniques

provides better results. Reasons for this can be due to the assumptions underlying the various feature selection techniques [2], but also due to the fact that not all biologically relevant features for a certain differentiation need to be relevant for the corresponding classifier, and some features can be data set specific. This holds for the clinically more interesting differentiations, but also for the other differentiations which were checked as additional support of our hypothesis and/or are important for proper tissue segmentation in 2-dimensional images based on MRSI (and MRI) data. From this we can conclude that automatic feature selection is a useful tool for obtaining relevant and possibly interesting features for differential diagnosis and brain tissue segmentation, but that careful evaluation of the selected features should be carried out, as well.

There is correspondence between the results obtained with the forward selection method and the filter/embedded methods. Part of the features selected for the corresponding set of binary differentiations, applying single-voxel data and multivoxel data, agree, although this comparison is not completely viable, as both have only part of the features in common.

Some of the selected features, which cannot directly be accounted for from a biological or clinical perspective, can be explained from the data set. They can show data set specific small differences, improving the classification performance or giving hints to potentially interesting additional new features. This deserves further research and clinical validation. The value of MR image information for the classification is already observed in [13, 17, 18].

Our study shows that for our data and the applied classification algorithm the images are relevant in the CSF - glioma differentiations and in the differentiation of grade II and grade III gliomas. For the CSF related problems proton density images show frequently the largest effect. For the discrimination between grade II and grade III glioma, all 4 types of MR images should be used for the best result with as second best option (if less images would be available) the frequently measured T1 and the T1 after gadolinium images.

Conflict of interest statement

None declared.

Acknowledgements

Jan Luts is funded by a PhD grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen); Research supported by Research Council KUL: GOA-AMBioRICS, several PhD/postdoc & fellow grants, Centers-of-excellence optimisation, IDO 05/010 EEG-fMRI; BET-IBM, Universidad Politecnica de Valencia for preprocessing the MRS data; Flemish Government: FWO: PhD/postdoc grants, projects, G.0360.05 (Advanced EEG analysis techniques for epilepsy monitoring), G.0519.06 (Noninvasive brain oxygenation), G.0341.07 (Data fusion), G.0321.06 (Numerical tensor techniques for spectral analysis), G.0302.07 (Support vector machines and kernel methods), research communities (ICCoS, ANMMM); IWT: PhD Grants; Belgian Federal Government: DWTC (IUAP IV-02 (1996-2001), IUAP V-22 (2002-2006): Dynamical Systems and Control: Computation, Identification & Modelling) and Belgian Federal Science Policy Office IUAP P6/04 (Dynamical systems, control and optimization, 2007-2011); EU: use of the data provided by the EU funded INTERPRET project (contract no. IST-1999-10310) is gratefully acknowledged, Marinette van der Graaf is gratefully acknowledged, the clinical partners represented by Jorge Andres Calvar (FLENI, Ciudad de Buenos Aires, Argentina), and dr. Franklin Howe (St George's Hospital Medical School / St George's University of London), are gratefully acknowledged for providing data, BIOPATTERN (contract no. FP6-2002-IST 508803), eTUMOUR (contract no. FP6-2002-LIFESCIHEALTH 503094), HealthAgents (contract no. FP6-2005-IST 027213), FAST (contract no. FP6-019279-2), ESA: Cardiovascular Control (Prodex-8 C90242).

References

[1] I. Guyon, A. Elissee, An Introduction to Variable and Feature Selection, Journal of Machine Learning Research, 3 (2003) 1157.

[2] P. Pudil, P. Somol, Identifying the most Informative Variables for Decision-Making Problems - a Survey of Recent Approaches and Accompanying Problems, Acta Oeconomica Pragensia, 2008 (2008) 37.

[3] A. Blum, P. Langley, Selection of relevant features and examples in machine learning, Artificial Intelligence, 97 (1997) 245.

[4] G.H. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selection problem, International Conference on Machine Learning, 1994, p. 121.

[5] A.R. Tate, J.R. Griffiths, I. Martinez-Pérez, A. Moreno, I. Barba, M.E. Cabañas, D. Watson, J. Alonso, F. Bartumeus, F. Isamat, I. Ferrer, F. Villa, E. Ferrer, A. Capdevilla, C. Arús, Towards a method for automated classification of 1H MRS specra from brain tumours, NMR Biomed., 11 (1998) 177.

[6] F. Szabo de Edelenyi, C. Rubin, F. Estève, S. Grand, M. Décorps, V. Lefournier, J.L. Bas, C. Rémy, A new approach for analyzing proton magnetic resonance pectroscopic images of brain tumors: nosologic images, Nat. Med., 6 (2000) 1287.

[7] A.W. Simonetti, W.J. Melssen, M.v.d. Graaf, A. Heerschap, L.M.C. Buydens, A new chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy, Anal. Chem., 75 (2003) 5352.

[8] K.S. Opstad, C. Ladroue, B.A. Bell, J.R. Griffiths, F.A. Howe, Linear discriminant analysis of brain tumour 1H MR spectra: a comparison of classification using whole spectra versus metabolite quantification, NMR in Biomed, 20 (2007) 763.

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

[10] K.S. Opstad, M.M. Murphy, P.R. Wilkins, B.A. Bell, J.R. Griffiths, F.A. Howe, Differentiation of metastases from high-grade gliomas using short echo time 1H spectroscopy, J. Magn. Reson. Imaging, 20 (2004) 187.

[11] G. Fan, B. Sun, Z. Wu, Q. Guo, Y. Guo, In vivo single-voxel proton MR spectroscopy in the differentiation of high-grade gliomas and solitary metastases, Clin. Radiol., 59 (2004) 77.

[12] M. Castillo, J. Smith, L. Kwock, Correlation of myo-inositol levels and grading of cerebral astrocytomas, American Journal of Neuroradiology, 21 (2000) 1645.

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

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

[15] T.G. Dietterich, Machine learning research: Four current directions, The AI Magazine, 18 (1998) 97.

[16] J. Luts, A. Heerschap, J.A.K. Suykens, S. van Huffel, A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection, Artif. Intell. Med., 40 (2007) 87.

[17] A.W. Simonetti, W.J. Melssen, F. Szabo de Edelenyi, J.J.A. van Asten, A. Heerschap, L.M.C. Buydens, Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification, Nuclear Magnetic Resonance in Biomedicine, 18 (2005) 34.

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

[19] C. Majós, J. Alonso, C. Aguilera, M. Serrallonga, S. Coll, J. Acebes, C. Arús, J. Gili, Utility of proton MR spectroscopy in the diagnosis of radiologically atypical intracranial meningiomas, Neuroradiology, 45 (2003) 129.

[20] J. Luts, T. Laudadio, A.J. Idema, A.W. Simonetti, A. Heerschap, D. Vandermeulen, J.A.K. Suykens, S. van Huffel, Nosologic imaging of the brain: segmentation and classification using MRI and MRSI, NMR Biomed., 22 (2009) 374.

[21] M. Julia-Sape, D. Acosta, M. Mier, C. Arús, D. Watson, A multi-centre, webaccessible and quality control-checked database of in vivo MR spectra of brain tumour patients, Magn Reson Mater Phy, 19 (2006) 22.

[22] U. Klose, In vivo proton spectroscopy in presence of eddy currents, Magn. Reson. Med., 14 (1990) 26.

[23] V. Govindaraju, K. Young, A.A. Maudsley, Proton NMR chemical shifts and coupling constants for brain metabolites, NMR Biomed., 13 (2000) 129.

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

[25] R. Wehrens, A.W. Simonetti, L.M.C. Buydens, Mixture modelling of medical magnetic resonance data, Journal of Chemometrics, 16 (2002) 1.

[26] A.W. Simonetti, W.J. Melssen, M. van der Graaf, A. Heerschap, L.M.C. Buydens, Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction, J. Magn. Reson., 159 (2002) 151.

[27] W.W.F. Pijnappel, A. van den Boogaart, R. de Beer, D. van Ormondt, SVD-based quantification of magnetic resonance signals, J. Magn. Reson., 97 (1992) 122.

[28] Z. Tong, T. Yamaki, K. Harada, K. Houkin, In vivo quantification of the metabolites in normal brain and brain tumors by proton MR spectroscopy using water as an internal standard, Magn. Reson. Imaging, 22 (2004) 735.

[29] M. Robnik-Sikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Machine Learning, 53 (2003) 23.

[30] R.A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics, 7 (1936) 179.

[31] C. Lu, A. Devos, J.A.K. Suykens, C. Arus, S. van Huffel, Bagging linear sparse bayesian learning models for variable selection in cancer diagnosis, IEEE Trans. Inf. Technol. Biomed., 11 (2007) 338.

[32] S.J. Nelson, Multivoxel magnetic resonance spectroscopy of brain tumors, Mol. Cancer Ther., 2 (2003) 497.

[33] A. Shiino, S. Nakasu, M. Matsuda, J. Handa, S. Morikawa, T. Inubushi, Noninvasive evaluation of the malignant potential of intracranial meningiomas performed using proton magnetic resonance spectroscopy, J. Neurosurg., 91 (1999) 928.

[34] C. Majós, J. Alonso, C. Aguilera, M. Serrallonga, S. Coll, J.J. Acebes, C. Arús, J. Gili, Utility of proton MR spectroscopy in the diagnosis of radiologically atypical intracranial meningiomas, Neuroradiology, 45 (2003) 129.

[35] M.P. Buetow, P.C. Buetow, J.G. Smirniotupoulos, Typical, atypical and misleading features in meningiomas, Radiographics, 11 (1991) 1087.

Figure legends

Figure 1: Plot of the mean short echo single-voxel spectrum plus/minus the standard deviation (4-0.5 ppm). (a) low grade gliomas, (b) glioblastomas, (c) metastases, and (d) meningiomas.

Figure 2: The results of the forward feature selection procedure in combination with QDA using fixed preselected combinations of image features for grade II versus grade III gliomas. The selected features, labeled on the right vertical axis (with chemical shift in ppm), are indicated in black on each bar, the preselected set of MRI features are given on the X-axis for each bar and by the dark gray squares (1 means T1, 2 T2, P PD and G means Gd). '-' Means that the feature selection was executed without MRI features, 'choice' means that the feature selection procedure could select from all MRI and MRSI features and is added for comparison. The height of the bars indicates the accuracy ± standard deviation obtained by a repeated leave-patient-out procedure.

Table legends

Table 1: Mean rank and weight values of the features and the selected features of the MRS data set obtained with the various feature selection methods for the differentiation of glioblastomas and low grade glial tumours.

Table 2: The selected features for each differential diagnosis in the single-voxel MRS data set.

Table 3: Mean rank and weight values of the features and the selected features of the MRSI/MRI data set obtained with the various feature selection methods for the differentiation of grade III gliomas and glioblastomas.

Table 4: The selected features for each differential diagnosis in the MRSI/MRI data set.

Table 5: Summary of the results regarding the added value of the MRI features.

Table 1.

	Cr1	Glx1	Tau+Glc	ml+Gly	ml+Tau	Cho	Cr2	Glx2	NAA	Ala	Lac	Lip1	Lip2
ARD[b]	0.19	**0.32**	**0.47**	**0.76**	0.08	0.02	**0.46**	**0.62**	0.01	0.01	**0.70**	0.02	**0.43**
R-F[a]	11.06	7.26	9.78	**4.64**	6.14	**3.72**	**3.60**	10.60	11.04	9.18	**1.42**	**3.34**	9.22
FC[a]	9.90	**5.94**	6.72	**3.24**	**5.86**	7.66	**4.12**	11.88	12.32	8.22	**1.26**	**2.32**	11.56
K-W[a]	9.60	**5.36**	6.26	**2.80**	6.24	8.42	**4.34**	12.08	12.66	8.12	**1.30**	**2.90**	10.92
FW[c]			o	o			o	o		o	o	o	o

a,b: mean rank values (a) and mean weights (b) of the features of the MRS data set obtained with the filter/embedded methods for the differentiation of glioblastomas and low grade glial tumours. The columns refer to the metabolites, the rows depict the feature selection methods (ARD: automatic relevance determination, R_F: Relief-F, FC: Fisher discriminant criterion, K-W: Kruskal-Wallis test, and FW: forward selection). Numbers in boldface belong to features that are potentially relevant.

c: the selected features for the forward selection approach (marked with 'o').

Table 2.

	Cr1	Glx1	Tau+Glc	ml+Gly	ml+Tau	Cho	Cr2	Glx2	NAA	Ala	Lac	Lip1	Lip2
GBM-GrII		x	x o	x̲ o			x̲ o	x o		o	x̲ o	x o	x o
GBM-Met	x	x	x o	x	x o	x̲ o	x o	x̲ o	x o	x	x̲	x o	x o
Men-Met		x̲			x		x	x̲ o		x	x̲ o	x̲	x o
GBM-Men		x̲	x	x	o	x	x o	x̲ o			x̲	x o	x
Met-GrII	o			x̲ o	x o	x	x̲	o	x		x̲ o	x̲	x o
Men-GrII			x	x̲		x	x̲ o	x̲ o		x			

The columns refer to the metabolites, the rows depict the differential diagnosis (GBM: glioblastoma, GrII: glioma grade II, Met: metastasis, Men: meningioma). Features that are selected by the forward procedure are indicated with a 'o', features that are labeled as relevant by the filter/embedded methods are denoted by 'x' and features that are labeled as most relevant are denoted by 'x̲'. A gray cell means that that feature is mentioned in literature as being relevant for the given differentiation. The clinically more relevant differentiations are in bold (left column).

Table 3.

	Cr1	Glx1	ml+Gly	Tau	Cho	Cr2	Glx2	NAA	L1+L+A	Lip2	T1	T2	PD	Gd
ARD[b]	0.006	0.126	**0.475**	0.024	0.136	0.011	0.012	0.001	**0.908**	0.002	**0.878**	0.041	0.012	0.142
R-F[a]	10.560	12.680	**3.880**	9.420	**5.160**	11.600	8.260	13.240	**2.300**	8.280	1.000	8.300	6.180	**4.140**
FC[a]	8.800	7.620	**2.300**	7.820	**3.960**	6.720	12.380	13.900	**2.860**	8.960	1.000	12.540	8.120	8.020
K-W[a]	8.500	7.840	**3.000**	7.120	**4.100**	7.400	12.540	13.820	**1.980**	9.740	1.040	12.540	6.920	8.460
FW[c]					o				o	o	o			

a,b,c: see table 1.

Table 4.

	Cr1	Glx1	mI+Gly	Tau	Cho	Cr2	Glx2	NAA	L1+L+A	Lip2	T1	T2	PD	Gd
GrII-GrIII				o	o	o	o	o	x o	x o	x o	x o	x o	x o
GrIII-GBM		o	x o	o	x	o	o	o	x o	o	x o	o	o	x
GrII-GBM		o	x o	o	o	x o	o		x	x	o	x	o	x o
He-GrII	o		o		x	x o		x o	x			x	x o	
He-GrIII						x		x o	x o		x		x o	
He-GBM			x			x		x o	x o				x o	x o
GBM-Men	x	x		x			x		x o	x	o		x	x
GrIII-Men	o	o	x		x o	x	x				x		x o	x
CSF-GrII					x		x		x o	o	x o	x	x o	x
CSF-GrIII					x				x	x			x o	x
CSF-GBM				x					x	x	x	o	o	x o
He-Men			x		o	x o		x o	x o		x		x o	x
GrII-Men			x o			x			o		x	x o	x o	x
He-CSF				o		x		x o			x	x o	x	x
CSF-Men			o	x		o	x		x o		x	x o	o	x o

Content explanation: see table 2. Additionally: GrIII: glioma grade III, He: healthy tissue, CSF: celebral spinal fluid.

Table 5.

	Gr II vs Gr III	**Gr III vs GBM**	**Gr II vs GBM**	**He vs GrII**	**He vs GrIII**	**He vs GBM**	CSF vs Gr II	CSF vs Gr III	CSF vs GBM	He vs CSF
Max accuracy (max gain)	**86.4% (6.0%)**	**94.7% (2.7%)**	**99.0% (0.8%)**	**99.4% (1.2%)**	**100% (0.4%)**	**99.9% (0.5%)**	99.6% (11.7%)	100% (5.6%)	100% (2.7%)	99.4% (0.8%)
Relevant MRI features	**Gd/T1T2/T2PD/PDGd < T1T2Gd/T2PDGd/T2PDGd**	**PDGd/T1T2 < T2PDGd/T1T2Gd/T2PD/T1T2PD**	**T2/PD Gd/T1T2PD /T1PD/ T1PDGd**	**T1<PD/ T2Gd/PDGd**	**T2/PD**	**PD < T1<Gd**	T1<Gd <T1Gd < T1T2Gd <PD	PD alone	T1/Gd /T1Gd < T1T2/T2PDGd alone	T1 / T2

'Max gain' means the maximal gain in accuracy by adding image features, compared to the situation in which no image features are used. The bottom row lists the image features (or combinations of image features) that have a significant positive effect on the accuracy in the order of

increasing magnitude of the effect. Image features or combinations of image features that have a similar effect are separated with '/', '<' Means 'smaller than' and refers to the comparison of the effect of the listed image features on the accuracy. The clinically more relevant differentiations are in bold

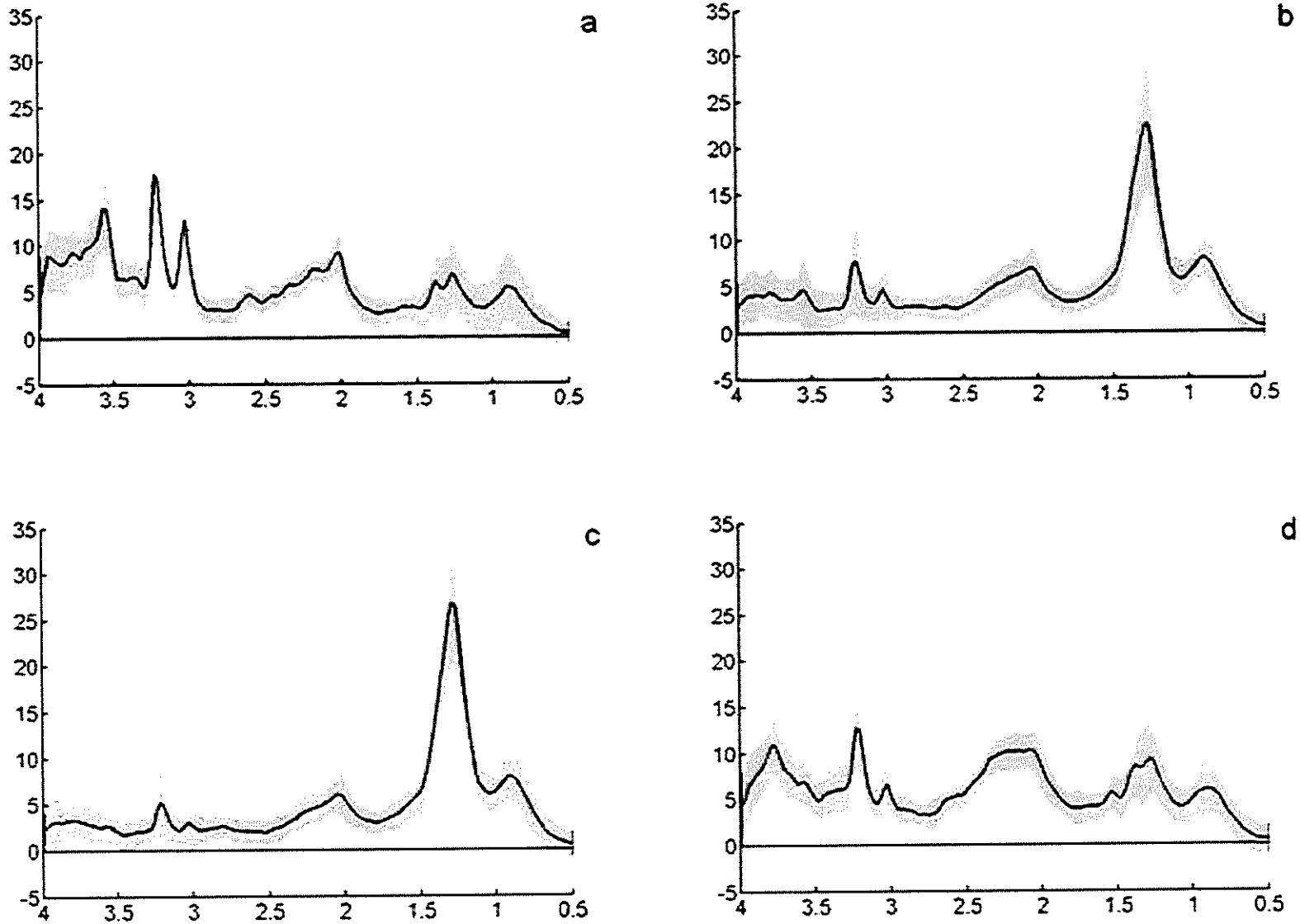
Summary

Automatic feature selection is frequently advised if classifiers are built on data consisting of a high number of variables. This can have several benefits, such as improving the classifier performance and an improved interpretability (the variables that appear to be important for the decision process could lead to hypotheses on the underlying biological process and new insights). During the automatic selection process those features (variables) are selected which are relevant for the classification process. However, the relevance of a selected feature normally is judged in the context of the classification process. And this is not necessarily similar to whether the selected features represent relevant biological information. An investigation into this matter is described in this article using two data sets consisting of magnetic resonance data for the classification of several types and grades of human brain tumours.

One data set consists of single voxel short echo MR spectroscopy data from 211 patients with histologically diagnosed brain tumours (84 glioblastomas, 57 meningiomas, 37 metastases and 33 low grade glial tumours). The second dataset consists of a combination of short echo multivoxel MR spectroscopic imaging (MRSI) data and four types of MR imaging data from selected voxels from 25 patients (10 grade II gliomas, 4 grade III gliomas, 7 grade IV gliomas, and 4 meningiomas) and 4 healthy individuals. These data sets were analyzed with several automatic feature selection routines belonging to three main categories (filter methods, wrapper methods and embedded methods). The automatically selected features were evaluated using literature and clinical experience.

The second multimodal data set was also used to investigate the influence of the MRI features on the classifiers and the possibility to reduce or replace types of MR images. It was observed that a significant part of the automatically selected features for the single-voxel data as well as the multivoxel data correspond to what is known from the literature and clinical experience. The results obtained with the various types of feature selection techniques are very similar. However, applying a single feature selection technique does not guarantee that all known biologically or clinically relevant features are found: a combination of several techniques provides better results. Reasons for this can be due to the assumptions underlying the various feature selection techniques, but also due to the fact that not all biologically relevant features for a certain differentiation need to be relevant for the corresponding classifier, and some features can be data set specific. The latter features can show data set specific small differences, helping in improving the classification performance or can give hints to possible interesting additional new features. We conclude that automatic feature selection is a useful tool to obtain relevant and possibly interesting features for differential diagnosis and brain tissue segmentation and can be used to evaluate alternative combinations of features.

Click here to download high resolution image



Click here to download high resolution image

