Identification of significant Metabolic Markers from MRSI data for Brain Cancer Classification

M. G. Kounelakis, M. E. Zervakis, M. E. Blazadonakis, G. J. Postma, L. M. C. Buydens, A. Heerschap and X. Kotsiakis

***Abstract*— Investigation of the significance of metabolites' peak area ratios derived from brain Magnetic Resonance Spectroscopic Imaging (MRSI) spectra, in brain tumors classification, has been applied. Results have shown that in most binary classifications using SVM and LSSVM classifiers, the accuracy achieved was greater than 0.90 AUC except the case of Gliomas grade 2 vs Gliomas grade 3 where 0.84 AUC was recorded due to the great heterogeneity of these two types of tumor. The minimum but also biologically significant set of features (markers), where maximum AUCs recorded, was derived. Ratios of N-acetyl-aspartate, Choline, Creatine and Lipids metabolites found to play the most crucial role in brain tumors discrimination. The biological importance of these markers was also verified by literature. Finally the influence of four magnetic resonance image (MRI) intensities on the classification process was also measured. It was found that MRI data do not improve significantly the classification accuracies.**

I. Introduction

Accurate diagnosis is essential for optimum clinical management of patients with intracranial tumors. When accessible, most tumors are surgically removed, but there is a balance between removing as much tumor tissue as possible whilst maintaining vital brain functions. Therefore a noninvasive and accurate prediction of tumor type can reduce unnecessary surgical biopsy for non-cancerous lesions and less accessible tumors, which could be treated by radio- or chemo-therapy rather than surgical resection.

1H magnetic resonance spectroscopy (or proton MRS) can provide information on the biochemical profile of tissue and is increasingly being used as a noninvasive method for classifying brain lesions. Earlier MRS studies show clear differences between the 1H spectra of brain tumors and normal brain tissue [1], [2]. There are two methods of proton magnetic resonance spectroscopy, namely single voxel and multivoxel, with or without spectroscopic imaging. MRS or single voxel spectroscopy acquires one signal from a certain volume element (voxel), while MRSI or multi voxel spectroscopy acquires simultaneously signals from a two dimensional grid of voxels. MRSI can facilitate the identification of heterogeneity of a tumorous region, since spatial variations of the tissue characteristics can be assessed at metabolite level. For each voxel, the intensity of the biochemically relevant metabolites can be determined as shown in Fig. 1. MRSI has large potential for clinical applications as it provides spectroscopic as well as spatial information of the brain tissue.

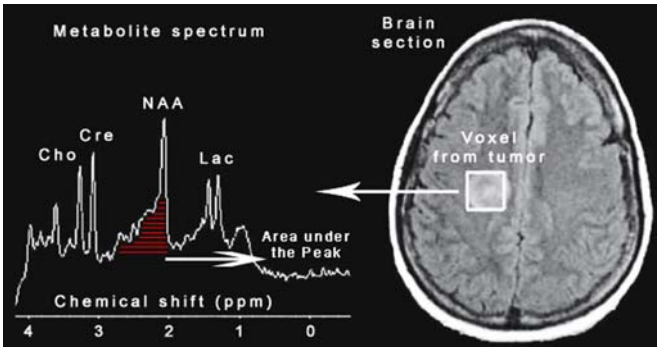


Fig. 1. Interpretation of a voxel to its spectrum. Y axis: peak heights (metabolites concentration). X axis: frequency (position) in parts per million. Cho = Choline, Cre = Creatine, NAA = N-acetyl-aspartate, Lac = Lactate.

The acquisition of the patient's metabolic spectra enables the determination of several types of features that can be used for brain tumor classification purposes. These may be the metabolites amplitudes, the area under the metabolites peaks (red area in Fig. 1) or even the ratios of the metabolites peak areas.

The main goal of this work is to rigorously study the potential of MRSI information and identify the smallest possible set of biologically significant features (markers) from brain metabolites' spectra in order to accurately classify different types of labeled brain tumors. Towards this direction, ratios of metabolite peak areas were used, which are extracted from twelve well known brain metabolites. Ratio type features have been successfully used in studies over the last decade concerning the discrimination of brain tumors type, grade and heterogeneity [3]-[7]. Measuring peak area ratios of metabolites has the advantage of canceling out the effects of general reduction in measured metabolite concentrations, which are due to variations in cellular density. Most of these studies though, involve only ratios of two or three known metabolites, like NAA (N-

Manuscript submitted July 3, 2008.

M. G. Kounelakis, M. E Zervakis and M. E. Blazadonakis are with the Technical University of Crete, Department of Electronic and Computer Engineering.

G. J. Postma and L. M. C. Buydens are with the Radboud University, Department of Analytical Chemistry.

A. Heerschap, is with the University Medical Center of Nijmegen, Department of Radiology.

X. Kotsiakis, is with the General Hospital of Chania, Department of Neurosurgery.

acetyl aspartate), CHO (Choline) and CRE (Creatine) and none of them combines MRI data in the classification process.

The novelty of our approach is in the way we exploit multiple ratio features within a multimodal framework. First, we extended the idea of ratio type features, by constructing a feature set that includes 31 unique peak area ratios stemming from the biological significance of metabolites. Using then combinations of these 31 features, 5 more feature sets are also built in order to reveal any possible intrinsic metabolite variations that could assist more accurate brain tumor discrimination. Furthermore, MRI features that come from 4 imaging intensities are combined with the above ratio features in order to measure their influence in the classification process.

In the sequel we test pattern recognition methods for feature selection and classification purposes. The dataset and methods used are discussed in Section II, while the results are presented in Section III. Section IV elaborates on the findings and Section V concludes this paper.

II. Materials and Methods

A. Materials

The dataset consists of magnetic resonance imaging (MRI) data and short echo magnetic resonance spectroscopy imaging (MRSI) data from 24 patients and 4 control volunteers. From these patients, 21 have been diagnosed with a glial brain tumor of specific grade and 3 with meningiomas.

The MR data was collected by the Radboud University as part of INTERPRET (http://carbon.uab.es/INTERPRET) project and contains T1-weighted, T2-weighted, proton density (PD) weighted images and gadolinium-enhanced (GD) T1-weighted images, as well as water suppressed and unsuppressed 1H-MRSI spectral images (two-dimensional MRSI data). The core dataset [8] consists of 669 pre-processed spectral sets containing data (MRI as well as MRS) from 669 voxels of 24 brain tumour patients and 4 healthy persons. Each patient case passed strict quality control and validation procedures, including consensus histopathologic determination. Per tissue type voxels are taken from homogeneous regions.

More specifically, the set contains the following information.

(1) Healthy tissue from volunteers: normal tissue from healthy persons (4 patients),

(2) Healthy tissue from patients: apparently normal tissue from the contralateral half of the brain of patients (4 patients),

(3) Cerebro spinal fluid (CSF): CSF from patients, where the ventricles were clearly visible and the voxels were located as far from the tumour as possible (10 patients).

(4) Grade 2 gliomas: diffuse astrocytomas (5 patients), oligodendrogliomas (2 patients), and mixtures (3 patients),

(5) Grade 3 gliomas: anaplastic astrocytomas (1 patient), oligodendrogliomas (2 patients), and undefined (1 patient),

(6) Grade 4 gliomas: glioblastomas (7 patients),

(7) Meningiomas (3 patients).

Design of binary classifications

For the needs of this study, nine binary classification schemes were constructed as shown in Table I. These classification schemes were selected owing to their clinical importance.

CSF voxels were removed from the classification process since they can be easily discriminated from other tissues, so they have no clinical interest.

TABLE I
BINARY CLASSIFICATION SCHEMES

Classes	Patients	Voxels
Healthy vs Tumor	8 vs 24	218 vs 351
Healthy vs Glio	8 vs 21	218 vs 303
Healthy vs MNG	8 vs 3	218 vs 48
GR2 vs GR3	10 vs 4	176 vs 57
GR2 vs GR4	10 vs 7	176 vs 70
GR3 vs GR4	4 vs 7	57 vs 70
GR2 vs MNG	10 vs 3	176 vs 48
GR3 vs MNG	4 vs 3	57 vs 48
GR4 vs MNG	7 vs 3	70 vs 48

Glio = Gliomas (sum of GR2, GR3 and GR4), MNG = Meningiomas, GR2 = Gliomas Grade 2, GR3 = Gliomas Grade 3, GR4 = Gliomas Grade 4.

Construction of feature sets

In this study, two input feature ensembles were used, (a) six sets of metabolites' peak areas ratios (Table II) and (b) the set of 4 imaging intensities mentioned above. The imaging intensities were also combined with the six ratio sets. Ratios of metabolites' peak areas were measured within each voxel independently, not in relation to healthy tissue.

The metabolites' peak areas from the MR spectrum were obtained by peak integration [8], from the following spectral regions: TCreatine (Cre1 at 3.95 ppm), Glutamate (Glu1 at 3.75 ppm), Myo-inositol (Mi at 3.56 ppm), Glutamate/Glutamine (Glx at 3.44 ppm), Choline (Cho at 3.20 ppm), Creatine (Cre at 3.02 ppm), Glutamine (Glu2 at 2.20 ppm), N-acetyl aspartate (NAA at 2.02 ppm), Alanine (Ala at 1.48 ppm), Lactate (Lac at 1.33 ppm), Lipids1 (L1 at 1.30 ppm) and Lipids2 (L2 at 0.90 ppm). Only the spectral values in the region of interest (0.5–4.0 ppm) were used as input features. In this region the metabolites' peaks can be easily measured.

TABLE II
RATIOS' SETS

Ratios' Sets	Number of features
Ratios set 1	10
Ratios set 2	10
Ratios set 3	13
Ratios set 4	15
Ratios set 5	21
Ratios set 6 (union of Ratios' sets 1 to 5)	31

B. Methods

Towards the identification of a minimum number of input features that produce the best classification of available samples, an RFE-SVM strategy was employed, which has

been applied to microarray genomic data with remarkable results [9]. In particular, we use a variant of this scheme that embeds a Fisher's filter criterion within the operation of the SVM classifier [10]. This criterion shown in equation 1, focuses only on the support vectors by ranking their input features as:

$$f(x_i) = \frac{|\mu_+(x_i) - \mu_-(x_i)|}{\sigma_+(x_i) + \sigma_-(x_i)} \quad (1)$$

where $\mu_+(x_i) - \mu_-(x_i)$ and $\sigma_+(x_i) + \sigma_-(x_i)$ are the mean and standard deviation values of feature x_i in the positive and negative classes, respectively. The features with the smallest Fisher's value are eliminated and the most significant ones are kept for classification purposes. A brief presentation of the pseudocode of this method is presented below.

Feature selection and classification pseudocode
1. Let n be the initial number of input features
Loop:
2. While ($n \geq 0$)
3. Create and train the SVM classifier using any type of kernel.
4. Locate the Support Vectors (SVs).
5. Based on the Support Vectors only rank the features according to the value returned by Fisher criterion.
6. Remove the feature with the smallest Fisher metric, in absolute value. More than one feature can be removed in each iteration.
7. Estimate the classification accuracy of the *n* surviving features using a linear SVM classifier.
8. End While *(end of Loop)*
9. Output as significant features the set of surviving features achieving the best classification accuracy.

For classification purposes, both the SVM classifier and its least squares variant (LS-SVM) [11] were tested. In both cases the Radial Basis Function (RBF) kernel was selected, since it is considered a good choice when multidimensional and heterogeneous data are under scrutiny, like the case of Brain tumors we investigate. The parameters of RBF kernels (gamma, sigma) were tuned in such a way that the smallest possible number of support vectors would be retained for training purposes, as to avoid overtraining.

Classifier performance evaluation

The 10-Ford cross validation (CV) and Leave-Patients-Out CV strategies were tested. In 10-Fold CV, 100 stratified random splits of each dataset were created. Train sets contained 90% of the voxels of each of the two classes (binary classification) and the remaining 10% of these two classes were contained in the test sets. In Leave Patients Out CV, the test sets were constructed based on the principle that every patient from each class would be contained at least once in the test sets, so that the entire dataset could be tested at least once.

III. Experimental results

Accuracy measurements were obtained using the global metric representing the Area under the Receiver Operating Characteristic curve (AUROC). Confidence intervals (CI) were also estimated for statistical purposes. The results obtained using the Ratios' sets as inputs into the feature selection and classification process are presented in Table III below.

TABLE III
MAXIMUM AUC MEASURED USING RATIOS SETS

Feature Set	Ratios of Metabolites Peak Areas				
Classifier evaluation	10-Fold CV		Leave Patients Out CV		
Feature selection	SVM	LS-SVM	SVM	LS-SVM	CI
Healthy vs Tumor	**0.99 / 6**	0.98 / 9	0.97 / 5	0.98 / 8	0.004
Healthy vs Glio	**0.99 / 5**	0.99 / 9	0.98 / 8	0.98 / 6	0.005
Healthy vs MNG	**0.98 / 2**	0.98 / 4	0.90 / 9	0.97 / 13	0.009
GR2 vs GR3	**0.84 / 8**	0.84 / 9	0.78 / 7	0.63 / 12	0.028
GR2 vs GR4	**0.99 / 4**	0.99 / 5	0.99 / 4	0.97 / 4	0.004
GR3 vs GR4	**0.98 / 2**	0.98 / 2	0.97 / 3	0.97 / 1	0.013
GR2 vs MNG	**0.92 / 7**	0.92 / 8	0.89 / 10	0.84 / 10	0.022
GR3 vs MNG	0.92 / 2	**0.94 / 7**	0.81 / 8	0.83 / 8	0.015
GR4 vs MNG	**0.90 / 5**	0.90 / 6	0.86 / 2	0.89 / 2	0.018

Numbers in bold present the maximum AUC measured. The number at the right side of the backslash symbol corresponds to the number of features where the maximum AUC was recorded, i.e. the most significant features.

The identities of the most significant features (markers) are shown in Table IV. These markers were also ranked according to their frequency of appearance in the feature selection and classification process. Feature ranking is based on the number of times in a 100 runs process (100 stratified random splits) a feature is weighted as significant and it is also depicted in Table IV.

Ratios of the metabolites' peak areas provide quite satisfactory results in most binary classifications. The 10-Fold CV method gives better AUROC results compared to the Leave Patients Out method. The SVM prediction model appears to be the best choice for such binary classifications.

More specifically, in Healthy vs Tumor, Gliomas and MNG the accuracy measures achieved are very high (greater than 0.98). It is also observed that in Healthy vs Tumor case the number of features needed to achieve 0.99 AUROC is 6, which is a bit larger than the other two tests. This is reasonable since Tumor class contains both Gliomas and Meningiomas. It is also remarkable the fact that only 2 features can discriminate Healthy vs MNG.

TABLE IV
MOST SIGNIFICANT FEATURES FOR EACH BINARY CLASSIFICATION SCHEME

Binary schemes	Most frequent features at max AUC								Max AUC
Healthy vs Tumor	NAA/ CHO	CHO/ CRE	CHO/ S	NAA/ CRE	LIPS / CHO	NAA / S			0.99
Healthy vs Glio	NAA / CHO	CHO / CRE	CHO / S	LAC / CRE	ALA / CRE	MI / S			0.99
Healthy vs MNG	MI / S	NAA / S							0.98
GR2 vs GR3	LIPS / CRE	CRE / S	LIPS / CHO	LAC / CRE	ALA / CRE	MI / S	ALA / S	NAA/ CRE	0.84
GR2 vs GR4	CRE / S	CHO / S	MI / S	ALA / S					0.99
GR3 vs GR4	CHO / S	MI / S							0.98
GR2 vs MNG	NAA / CHO	ALA / CRE	GLU1 / GLU2	L1 / CRE	NAA / CRE	MI / CRE	GLX / CRE		0.92
GR3 vs MNG	ALA / S	NAA / CHO	MI / S	LAC / CHO	MI / CHO	NAA / S	CHO / S		0.94
GR4 vs MNG	NAA / CHO	LIPS / CHO	NAA / S	LAC / CRE	ALA / S				0.90
Ranking	1st	2nd	3rd	4rth	5th	6th	7th	8th	

Rank most frequent to less frequent features (from left to right). Frequency of appearance was measured at each run (100 runs)

LIPS = L1+L2, S = sum of the 12 metabolites peak areas.

In the cases of Gliomas, the accuracy measures again take high values with a small number of features, except for the case of GR2 vs GR3 where AUROC is 0.84. This is explained due to the fact that these two classes present great heterogeneity as mentioned above. Finally, in Gliomas vs MNG cases, the accuracy measures are quite high but the number of features needed is increased. This is also expected since the MNG class involves only 3 patients, which makes the training of the classifier and the feature selection process more complex.

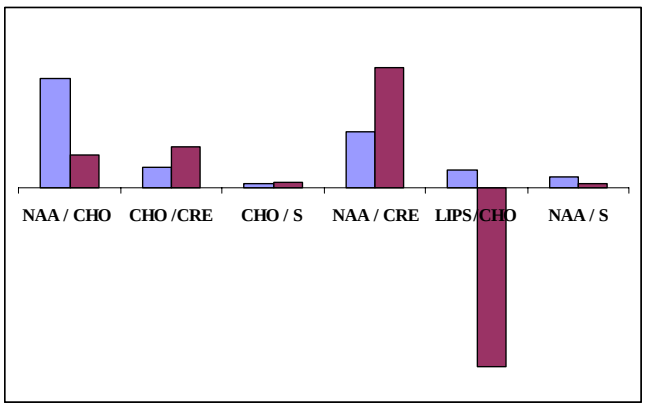
Adding the image features into the process of feature selection and classification, the maximum AUROC's are lower compared to those obtained without image features, as indicated in Table V. We now apply the SVM and LS-SVM classifiers, but with input features set to the most significant ratio features already derived (and shown in Table V) together with the image features from T1, T2, PD and GD imaging intensities. In most binary classification schemes there is either a decrease in the AUROC value or an increase in the number of significant features. This fact shows that image features can not improve the discrimination of brain tissues.

Another important issue concerns the statistical significance of the features selected as most important. Towards this direction, the mean value of each significant ratio feature in each class was estimated. The results are illustrated in Fig. 2 below. Blue bars represent the mean value of each feature in one class and the red ones represent the mean value in the other class. Statistical analysis was applied using SPSS version 16.0 software tool.

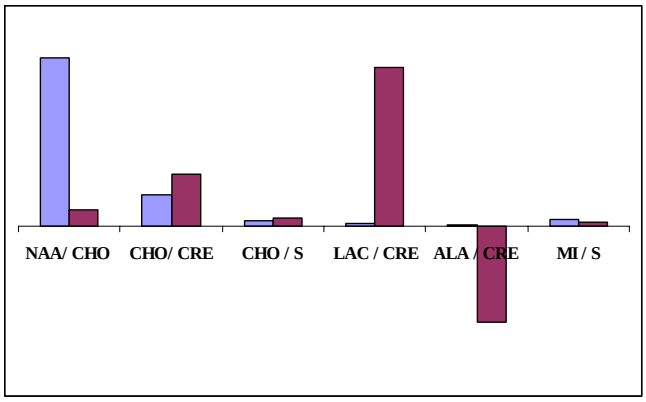
As we can see from this analysis the ratios NAA/CHO, CHO/CRE and LIPS/CHO contribute in the discrimination of Healthy from all Tumors (Gliomas, MNG) and LAC/CRE, ALA/CRE ratios in Healthy from Gliomas. MI/S and NAA/S ratios are able to discriminate Healthy from MNG. LIPS/CRE, LIPS/CHO, LAC/CRE and NAA/CRE clearly differentiate in GR2 and GR3 classes and CRE/S, CHO/S and MI/S assist the classification between GR2 and GR4 but also GR3 and GR4. Finally ALA/S, ALA/CRE, NAA/CHO and MI/S ratios are important in Gliomas vs MNG classification.

TABLE V
MAXIMUM AUC USING MOST SIGNIFICANT FEATURES FROM THE RATIOS SETS WITH IMAGE FEATURES

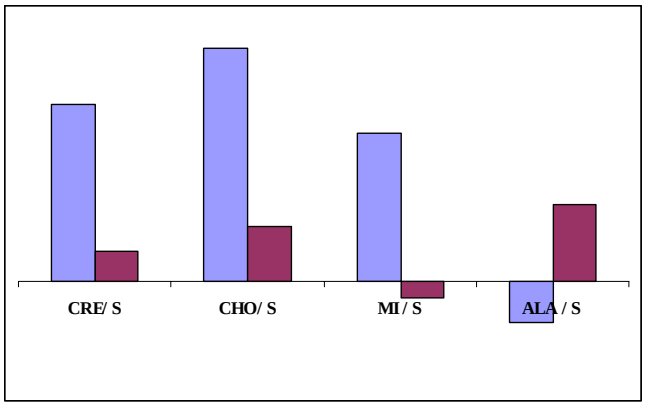
Feature Set	Most significant features with Image features (T1, T2, PD, GD)	
Classes	Max AUC	CI
Healthy vs Tumor	0.99 / 8	0.004
Healthy vs Glio	0.97 / 5	0.008
Healthy vs MNG	0.98 / 2	0.012
GR2 vs GR3	0.82 / 8	0.027
GR2 vs GR4	0.99 / 4	0.004
GR3 vs GR4	0.97 / 2	0.016
GR2 vs MNG	0.86 / 9	0.025
GR3 vs MNG	0.91 / 5	0.019
GR4 vs MNG	0.85 / 9	0.029



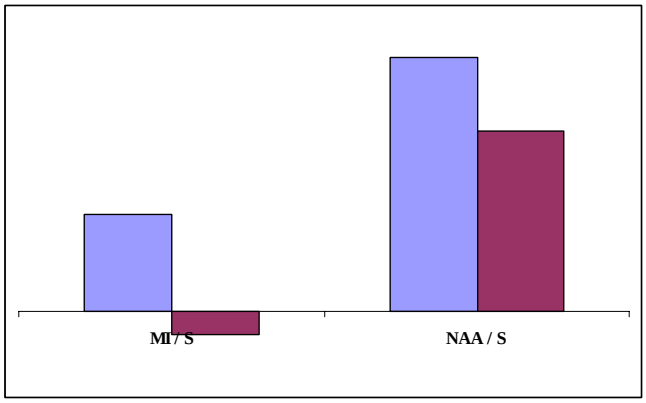
(a) Comparison of the means in Healthy vs Tumor



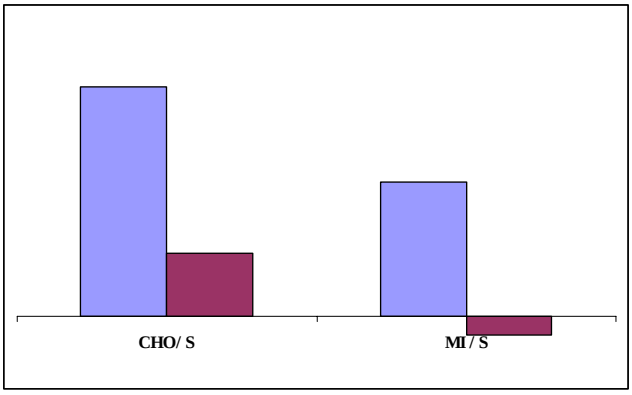
(b) Comparison of the means in Healthy vs Gliomas



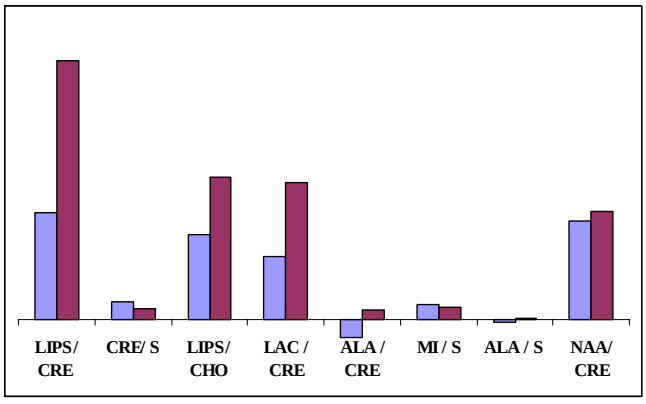
(e) Comparison of the means in GR2 vs GR4



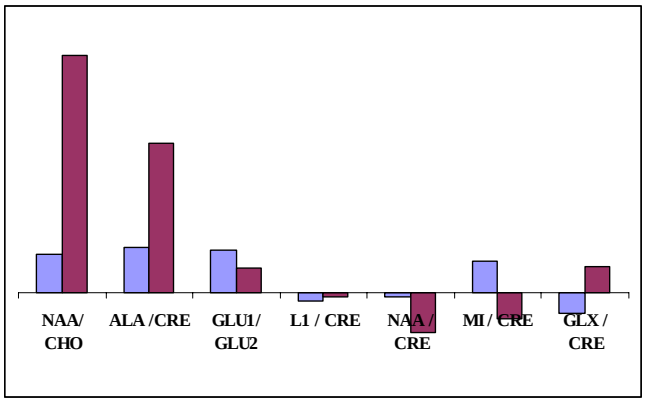
(c) Comparison of the means in Healthy vs Meningiomas



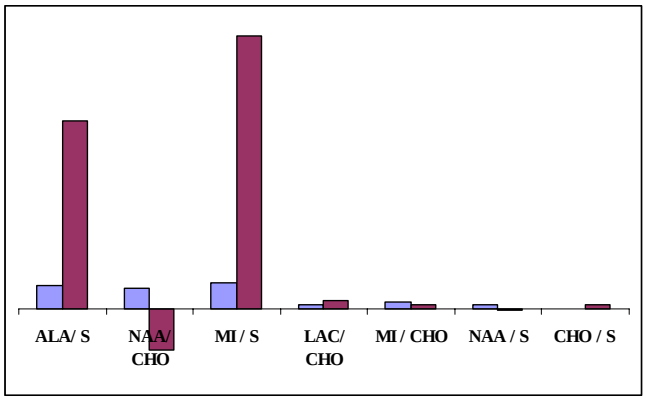
(f) Comparison of the means in GR3 vs GR4



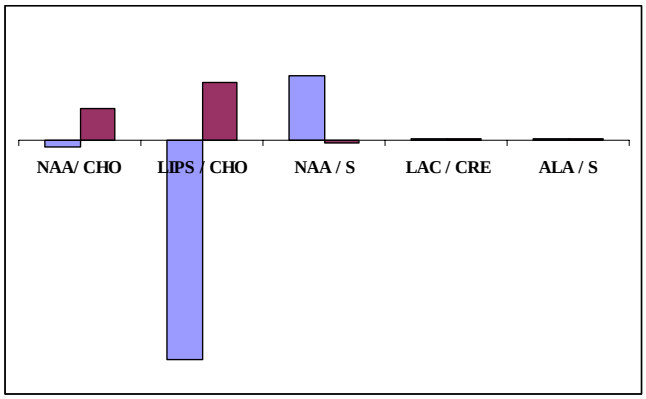
(d) Comparison of the means in GR2 vs GR3



(g) Comparison of the means in GR2 vs Meningiomas



(h) Comparison of the means in GR3 vs Meningiomas



(i) Comparison of the means in GR4 vs Meningiomas

Fig. 2. Comparison of the means of the most significant ratio features in each classification scheme. Blue bars correspond to the first type of tumor and red to the second one.

IV. DISCUSSION

Support Vector Machines using 10-Fold CV provide the best classification accuracy results in most binary classification schemes. This is expected due to the fact that Leave Patients Out CV approach excludes the data from at least one patient from the training phase, rendering the testing process more difficult. This is more obvious for the GR2 vs GR3 scheme, were the number of patients of GR3 is very small and there is great heterogeneity between these tumor types.

In Healthy vs Tumor, Healthy vs Gliomas and Healthy vs MNG classifications it is observed that NAA, CHO and CRE play a very important role. Even more, LIPS, LAC, CRE and CHO have significant contribution in gliomas discrimination. Finally, ALA, NAA and CHO and CRE assist the discrimination of gliomas versus meningiomas. The behavior of these metabolites is also verified in the literature [12], [13].

Comparing the means of these markers it is also obvious that NAA and CRE decrease in Tumors, while CHO increases. Lipids and LAC show a significant increase in Gliomas. Also ALA Lipids increase in Meningiomas.

Adding image features derived from T1, T2, PD and GD imaging modalities did not improve the classification results. Furthermore, it was observed that even when an increase in the accuracy was reported with the use of imaging features, the number of total features required for maximizing AUROC was significantly increased.

V. CONCLUSIONS

The present study reveals that 1H MRSI is an important adjunct to the clinical imaging modalities for non-invasive diagnosis of viable tumours. For most cases, binary classification based on metabolic information from MRSI provides good diagnostic results. Furthermore, the ratios of metabolites' peak areas can assist the diagnosis of Brain tumor types and reveal intrinsic characteristics of this complex disease.

ACKNOWLEDGMENT

Present work was supported by Biopattern, IST EU funded project, Proposal/Contract no.: 508803.

REFERENCES

[1] B. Ross and T. Michaelis, "Clinical applications of magnetic resonance spectroscopy", *Magnetic Resonance Quarterly*, vol .10, pp. 191-247 (1994).

[2] F. A. Howe, "Magnetic resonance spectroscopy in vivo. In Principles and Practice of MRI: Selected Topics", Markisz JA, Whalen JP (eds), *Appleton & Lange*, pp. 17-107 (1998).

[3] W. Roser et al., "Assignment of glial brain tumors in humans by in vivo H-magnetic resonance spectroscopy and multidimensional metabolic classification" *MAGMA*, vol. 5, pp. 179-183, (1997).

[4] A. Kumar, S. Kaushik, R. P. Tripathi, P. Kaur and S. Khushu, "Role of in vivo proton MR spectroscopy in the evaluation of adult brain lesions: Our preliminary experience", *Neurology India*, vol. 51, pp. 474-478, (2003).

[5] G. Fun, B. Suna, Z. Wua, Q. Guoa and Y. Guob, "In vivo single-voxel proton MR spectroscopy in the differentiation of high-grade gliomas and solitary metastases", *Clinical Radiology*, vol. 59, pp. 77-85, (2004).

[6] F. G. Lehnhardt, C. Bock, G. Rohn, R. I. Ernestus and M. Hoehn1, "Metabolic differences between primary and recurrent human brain tumors: a 1H NMR a 1H NMR spectroscopic investigation", *NMR in Biomedicine*, vol. 18, pp. 371-382, (2005).

[7] L. M. Harris et al., "The use of short-echo-time 1H MRS for childhood cerebellar tumours prior to histopathological diagnosis", *Pediatric Radiology*, vol. 37, pp. 1101-1109, (2007).

[8] A.W. Simonetti et al., "A Chemometric Approach for Brain Tumor Classification Using Magnetic Resonance Imaging and Spectroscopy", *Analytical Chemistry*, vol. 75, pp. 5352-5361 (2003).

[9] M. E. Blazadonakis and M. Zervakis, "Polynomial and RBF Kernels as Marker Selection Tools – A Breast Cancer Case Study", *Proceedings of the 6th International Conference on Machine Learning and Applications*, Cincinnati Ohio, 2007, pp. 488-493.

[10] N. V. Vapnik, "The Nature of Statistical Learning Theory" (Springer–Verlag New York), 1999.

[11] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers", *Neural Processing Letters*, vol. 9, pp. 293-300 (1999).

[12] F. A. Howe and K. S. Opstad, "1H MR spectroscopy of brain tumours and masses", *Nuclear Magnetic Resonance in Biomedicine*, vol. 16, pp. 123-131 (2003).

[13] K. S. Opstad et al., "Differentiation of metastases from high-grade gliomas using short echo time 1H spectroscopy", *Journal of Magnetic Resonance Imaging*, vol. 20, pp. 187-192 (2004).