Pattern recognition of abscesses and brain tumors through MR spectroscopy: Comparison of experimental conditions and radiological findings

Bruno Hebling Vieira[1], Antonio Carlos dos Santos[2], Carlos Ernesto Garrido Salmon[1*]

[1] InBrain Lab, Department of Physics, Faculty of Philosophy, Science and Letters of Ribeirão Preto, University of São Paulo, Ribeirão Preto, SP, Brazil.

[2] Department of Internal Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, SP, Brazil.

Abstract   **Introduction:** The interpretation of brain tumors and abscesses MR spectra is complex and subjective. In clinical practice, different experimental conditions such as field strength or echo time (TE) reveal different metabolite information. Our study aims to show in which scenarios magnetic resonance spectroscopy can differentiate among brain tumors, normal tissue and abscesses using classification algorithms. **Methods:** Pairwise classification between abscesses, brain tumor classes, and healthy subjects tissue spectra was performed, also the multiclass classification between meningiomas, grade I-II-III gliomas, and glioblastomas and metastases, in 1.5T short TE (n = 195), 1.5T long TE (n = 231) and 3.0T long TE (n = 59) point resolved spectroscopy setups, using LCModel metabolite concentration as input to classifiers. **Results:** Areas under the curve of the Receiver Operating Characteristic above 0.9 were obtained for the classification between abscesses and all classes except glioblastomas, reaching 0.947 when classifying against metastases, grade I-II gliomas and glioblastomas (0.980), meningiomas and glioblastomas (0.956), grade I-II gliomas and metastases (0.989), meningiomas and metastases (0.990), and between healthy tissue and all other classes in both conditions except for anaplastic astrocytomas in short TE 1.5T setup. When the multiclass classification agrees with radiological diagnosis the accuracy reaches 96.8% for short TE and 98.9% for long TE. **Conclusions:** The results in the three conditions were similar, highlighting comparable quality, robust quantification and good regularization and flexibility in either algorithm. Multiclass classification provides useful information to the radiologist. These findings show the potential of the development of decision support systems as well as tools for the accompaniment of treatments.

Keywords   Abscesses, Brain, Magnetic resonance spectroscopy, Pattern recognition, Tumors.

Introduction

Magnetic resonance spectroscopy (MRS) is considered a major tool for the biochemical study of tissue in vivo (El-Deredy, 1997; Ladroue, 2004). As a non-invasive technique, MRS becomes attractive for the study of the brain, diagnosis and follows up of various diseases (Gujar et al., 2005). It is particularly useful in the diagnosis of brain tumors, allowing the inference of the

relative or absolute concentration of a variety of substances (El-Deredy, 1997). However, the scope of its use in clinical applications is hampered by the difficulty of direct interpretation of the spectra, which requires training and experience, and direct comparison of different spectra, especially complicated by the large number of metabolites present in a spectrum, often in concentrations small enough to be mistaken for noise (Tate, 1996).

The design of applications able to automatically interpret the exams in order to identify its most relevant information has been one of the alternatives in decision-making tasks (Sharda et al., 1988). Such clinical decision support systems (DSS) are now one of the focuses of the application of artificial intelligence in medicine, with a special emphasis on machine learning and pattern recognition disciplines (Ramesh et al., 2004). A pattern recognition algorithm seeks the best possible categorization by gathering a set of observations by their common nature. Pattern recognition routines can be divided into three main steps: feature extraction, pre-processing, and classification. A feature extraction procedure converts

input data into features, retaining the most informative characteristics of the data. Pre-processing tasks include minimizing the information redundancy of a dataset as well as possible normalization and filtering steps. Lastly, the classification is performed, i.e. rules are set to designate which category new observations belong to through training, which is, adjusting the algorithm to the data set available in order to generalize it to new data. Although it can be summarized in a few steps, a pattern recognition routine may consist of several different algorithms with their particularities and assumptions about the distribution of data.

The application of pattern recognition routines to magnetic resonance spectra has been widely studied (El-Deredy, 1997; Ladroue, 2004; Tate, 1996). Among the first initiatives, Reilly and Kowalski (1971) dealt with spectral analysis in analytical chemistry. In time, the idea was applied to the diagnosis of brain tumors, by classifying spectra obtained in vivo and in vitro (El-Deredy, 1997; Preul et al., 1998). Finally, the INTERPRET project (Tate et al., 2006) facilitated the spread of studies with the establishment of a database containing hundreds of brain tumors and healthy tissue spectra (Julià-Sapé et al., 2006), and further with the design of clinical decision support tools (Julià-Sapé et al., 2015). The INTERPRET project fomented research on pattern recognition of brain tissue magnetic resonance spectra, and it remains the most accessible source for this kind of study, which is carried by diverse research groups worldwide to this date (Julià-Sapé et al., 2015). Several studies explored binary brain tissue classification tasks, which could serve as a DSS aimed at differential diagnosis (Arizmendi et al., 2014; Butzen et al., 2000; García-Gómez et al., 2009; Gray et al., 1998; Luts et al., 2008; Majós et al., 2004; Preul et al., 1998; Server et al., 2010; Vellido et al., 2012), or quality control (Wright et al., 2008), with a few multiclass only studies (Faria et al., 2011; Opstad et al., 2007; Poptani et al., 1999; Tate et al., 1998; 2003), and studies mixing both approaches (García-Gómez et al., 2008; Lukas et al., 2004; Roda et al., 2000). Simpler tasks included the classification between neoplastic and non-neoplastic tissue (Butzen et al., 2000) or between meningioma and non-meningioma (Gray et al., 1998), whereas more complex studies tried to accomplish the classification among many tissue types (Lukas et al., 2004; Tate et al., 2003) or achieve a unified feature extraction, preprocessing and classification framework (Arizmendi et al., 2014). The optimal classification between tumoral and healthy tissue spectroscopy has already been established, as well as the classification between the meningioma, aggressive tumors and low-grade glial tumors superclasses. Abscesses, on the other hand, have rarely been studied. Poptani et al. (1999) report the accuracy, specificity and sensitivity of multiclass classification of high and low-grade

gliomas, tuberculomas, abscesses and normal tissue short TE STEAM single-voxel spectra using artificial neural networks. In general, though, the area under the curve (AUC) of the receiver operating characteristic (ROC) is a more suitable measure, as it infers the discriminative power of a distribution and does not assume an arbitrary threshold for classification.

These tasks are extremely important when imaging diagnosis is not enough. In conventional anatomical images, necrotic or cystic tumors and abscesses may look the same thanks to the usual physiological responses in the brain (Desprechins et al., 1999). Restrictive diffusion within ring enhancement is not enough to differentiate brain abscesses from metastases (Hartmann et al., 2001). These confounding factors may compromise treatments, so a biopsy is usually performed, but it is an invasive procedure and therefore risks are involved. Also, grading of gliomas, a notoriously inhomogeneous type of tumor, has many confounding factors, to the point that approximately 38% of stereotactic biopsy diagnosis may differ from histopathological classification performed in resections from the same patient (Jackson et al., 2001). Brain tumor diagnosis would benefit much from an automated analysis of MR spectra due to these factors, helping to improve the accuracy of the radiological findings. Besides that, a common confounding factor in imaging that play almost no role in MR spectroscopy is the tumor shape. In special, for metastases and glioblastomas, the two most prevalent intra-cerebral tumors among adults which look virtually the same under certain circumstances and may have similar perfusion and contrast enhanced profiles, yet have completely different treatments (Fan et al., 2004).

Considering the data origin, in vivo 1.5 T single voxel spectra analysis is a recurring motif in the literature including data from the INTERPRET (Tate et al., 2006) and eTUMOUR projects (eTumour Consortium, 2008). Few in vitro studies were performed (Faria et al., 2011; Gray et al., 1998). The application of long or short echo-time (TE) spectroscopy has also been reported, as the choice of echo time affects the apparent concentration of every metabolite taking into account its transversal relaxation time (Majós et al., 2004) and also considerations about the dynamic range of the detectors (Ishimaru et al., 2001). In Figure 1, the median and interquartile range of example spectra are shown, highlighting the differences between spectral profiles between two echo times and also between two tissue types, brain abscesses and anaplastic astrocytomas, in both echo times.

In this paper, we propose the training of algorithms in several brain tissue spectra classification tasks in different experimental conditions. We used support vector machines and random forests. Different experimental scenarios were studied: 1.5 T short TE, 1.5 T long TE and 3.0 T long TE spectroscopy. Normal, abscess
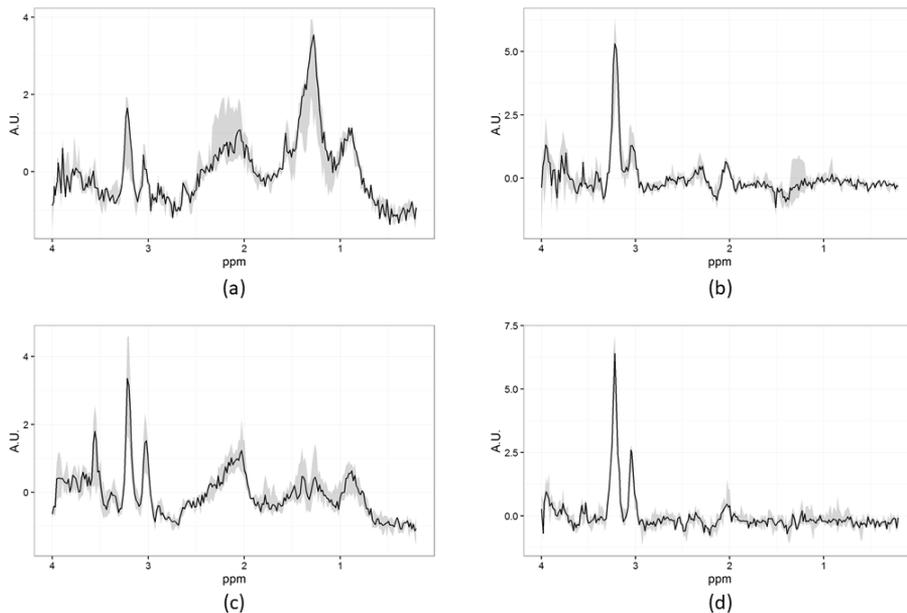
Figure 1. Median 1.5 T single voxel (SV) PRESS normalized spectra with baseline-correction done in LCModel in different echo-times (TE): (a) Abscess short TE (n = 5); (b) Abscess long TE (n = 8); (c) Anaplastic astrocytoma short TE (n = 6); (d) Anaplastic astrocytoma long TE (n = 7). The shadowed areas represent the interquartile ranges.

and tumor tissues were classified with only estimated concentrations of single voxel MRS as input variables. Binary and multiclass classification was performed, and in the case of the latter a novel comparison to radiological findings is provided, which to the best of our knowledge was never reported. Our findings may provide new insights on the application of MRS as a differential diagnosis tool in several tasks, and also on the applicability of 1.5 T and 3.0 T spectroscopy and short and long TE sequences.

Methods

The data were obtained retrospectively from sources that garnered their respective Ethical Review Committee or equivalent approval.

Data and acquisition

Two datasets were assembled from the 1.5 T spectra acquired in CDP (Center Diagnostic Pedralbes at Pedralbes, Barcelona and Esplugues del Llobregat), IDI (Institut de Diagnòstic per la Imatge at Bellvitge), SGUL (St Georges Hospital), UMNC (Universitair Medisch Centrum Nijmegen), UJF (Unité mixte Université Joseph Fourier), FLENI (Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia), MUL (Uniwersytet Medyczny w Lodz) and validated in the INTERPRET project (Tate et al., 2006), obtained in a digital database (Julià-Sapé et al., 2006). The first dataset included 231 long echo time spectra (33 acquired with TE = 135ms, 3 acquired with TE = 144ms, 195 were acquired with TE = 136 ms). The second dataset included 138 short echo time spectra (109 acquired with TE = 30 ms, 29 acquired with TE = 31 ms). The radiological findings annexed to the spectra were also included. In addition, for the third dataset 59 spectra were obtained retrospectively from the server of the HCRP (Hospital das Clínicas de Ribeirão Preto), acquired from a 3.0 T MR scanner, TE = 144 ms, repetition time (TR) 1500-2000 ms, VOI = 8-20 ml. All the spectra in the three datasets were obtained in single voxel mode using a Point Resolved Spectroscopy (PRESS) acquisition including water-scaling. The volume of interest (VOI) was selected in a way to cover the maximum uncontaminated tumor tissue. In the same databases, the final diagnosis, obtained through biopsy or imaging, was acquired for the primary brain tumors. The HCRP dataset lacked both healthy tissue and abscess spectra.

Data processing and metabolite quantitation

Metabolite concentrations were estimated using the LCModel software (Provencher, 2001). The spectra were acquired in many different centers and, ideally, basis spectra should be acquired one for each center.

However, that is impractical. The simulation of the basis spectra ends up being a practical and above all invariant solution in the absence of acquired basis spectra. Simulation was conducted in the Simulation application of the open-source VeSPA suite (Soher et al., 2011). A TE = 31 ms, PRESS, 64MHz, basis was simulated including various metabolites expected to be present in brain tissue. Other basis sets were provided by the software supplier. In total, for long TE spectra, for both 3.0 T and 1.5 T, thirty-one concentrations were estimated, including pure substances and mixtures of associated substances. They are alanine (Ala), creatine (Cr), phosphocreatine (PCr), glutamine (Gln), glutamate (Glu), gliceropphosphocoline (GPC), phosphocoline (PCh), inositol (Ins), lactate (Lac), N-acetylaspartate (NAA), N-acetylaspartylglutamate (NAAG), scyllo-Inositol (Scyllo), creatine singlet correction around 3.94 ppm (-CrCH2), guanidinoacetate or other signals around 3.78 ppm (Gua), and the mixtures GPC+PCh, NAA+NAAG, Cr+PCr and Glu+Gln and lipids (Lip) and macromolecules (MM) denoted following their position in the spectrum in ppm: Lip13a, Lip13b, Lip09, MM09, Lip20, MM20, MM12, MM14, MM17, Lip13a+Lip13b, MM14+Lip13a+Lip13b+MM12, MM09+Lip09 and MM20+Lip20. In addition, for short TE spectra, the concentrations of aspartate (Asp), Gamma-Aminobutyric acid (GABA), glucose (Glc), and taurine (Tau) were also

assessed. These estimated concentrations were used as the independent variables, or features, to the learning algorithms in our study. Logarithmic transformation was applied on the concentration values plus one to reduce the skew and keeping the minimum value of each variable at zero.

Lukas et al. (2004) point out that similar results are obtained when using either the estimated metabolite concentrations, the peak integrals or the constituent points of the spectra as input variables. We opted for using the estimated concentrations estimated through LCModel, as Opstad et al. (2007) discuss certain benefits attained through this method. The whole methodology is summarized in Figure 2. With the whole dataset, several binary classification tasks were assembled with different TE, B_0 and the classes shown in Table 1.

Classification tasks and analysis

Five models were studied in each task: random forests and support vector machines with linear, gaussian, sigmoid and polynomial kernels. Tuning was not performed due to the really limited sample size in most classes, which in turn creates less than optimal models but on the other hand helps to mitigate overfitting. We chose these algorithms because both are two extensively studied algorithms in the literature, due to their robustness to high dimensionality and high feature to observation
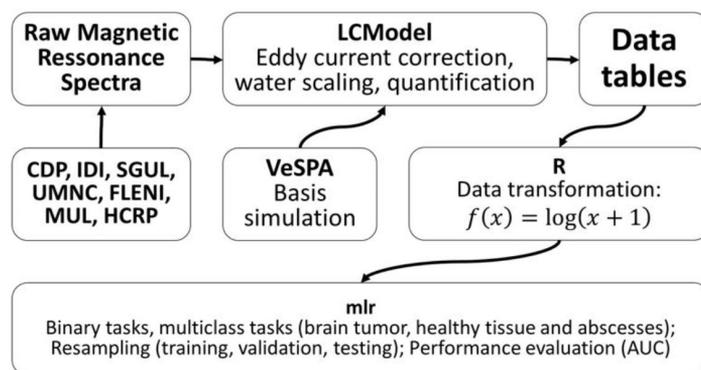
Figure 2. Summary of the methodology for tumor, abscess, and normal tissue classification based on MRS data: single-voxel (SV) spectra were obtained at seven centers and processed in LCModel. Metabolites quantification estimates were concatenated into a dataset. These estimates were log plus one transformed and then used as input to learning algorithms in the mlr package.

Table 1. Classes and sample size distribution for field strength (B_0) and echo-time (TE).

B_0	1.5 T	1.5 T	3.0 T
TE	short	long	long
Abscesses	5	8	-
Healthy tissue	10	15	-
Glioma Grade I-II	16	19	19
Glioma Grade III	6	7	5
Glioma Grade IV	38	61	22
Metastasis	26	28	7
Meningiomas	37	38	6

ratio, such as what is commonly found in microarray analysis (Statnikov et al., 2008), with regularization measures or bootstrap aggregation to avoid overfitting, obtaining highly flexible decision boundaries. The specific parameters used in the implementation of each algorithm used is not important to the scope of our work, where we aim to show in which tasks classification algorithms are effective, even if, unknowingly, less than optimal models are found.

The algorithm that best fit each task was determined using the average performance evaluated through repeated resampling. The validation was conducted through the stratified subsampling method that makes possible to estimate the stability of the obtained models, with a split rate of 2/3 and 200 repetitions. To determine if the algorithms rank differently regarding the evaluation measure, a post-hoc overall Friedman rank sum test (Demšar, 2006) was performed, which null hypothesis states there is no difference among classifiers. If the null hypothesis of this first test is rejected, then a Nemenyi multiple comparison test (Demšar, 2006) with q approximation for unreplicated blocked data can be performed to determine which pairwise differences are significant.

Taking into account the results in the binary classification and similarities between tumor types, another task was created for the classification between meningiomas, low grade gliomas, the concatenation of grade I, II, and III gliomas, and aggressive tumors, the concatenation of metastases and glioblastomas. For this task, repeated stratified nested cross validation was used, with a 10-fold inner cross validation loop and an outer 5-fold repeated 2 times cross validation loop. The SVMs were tuned with a 60 iterations random search strategy for kernel and cost. The results of this multiclass classification for both TE values and 1.5 T spectra were compared to the radiological diagnosis attached to the spectra, which comes as a single line describing the possible inferences from the imaging and spectroscopic studies. The described radiological diagnosis was coerced into one of the three classes, such as the label 'GLIOBLASTOMA/ METASTASIS' being translated into the class 'AGG' (aggressive tumors), or the case was discarded in this analysis as an ambiguous or invalid diagnosis, such as in the case of 'GLIOMA', which is ambiguous as gliomas appear in two classes, or in the case of the label 'ABSCESS', which are not included in this analysis and so are considered invalid.

The mean AUC was measured, and in the multiclass tasks the strategy proposed by Hand and Till (2001) was used. Analyses were carried out in the statistical software R 3.2.1. (R Core Team, 2016), based on the structure developed and made available by Bischl et al. (2016).

Results

The AUC metric was evaluated for all models in all tasks. The highest AUC averaged over resampling repetitions returned for each task are shown in Table 2, in common characters accompanied by the respective learning algorithm denoted as a digit whereas the lowest averages observed are represented in superscript analogously. Even though the comparison between different algorithms is not part of the main aims of our work, a post-hoc overall Friedman rank sum test was performed, showing that, apart from the task studied, the performance measure is not the same for every classifier (p < 0.001). Pairwise comparisons using the Nemenyi multiple comparison test with q approximation for unreplicated blocked data were also performed. Random Forest algorithm ranks significantly better than SVM Linear (p = 0.0081) and SVM Polynomial (p = 0.0015) regarding AUC. No other significant pairwise difference was obtained.

The multiclass classification cross validation results are shown in Table 3. For the long TE task and the short TE task, the classification result and the radiological findings are shown in Table 4. As the radiological diagnosis does not limit itself to the classes presented in the task, the study does not account for instances labelled with uncertainties, ambiguities and diagnostics that do not correspond with the proposed classes. Among the instances, both the classification and the radiologist agreed on the assigned class, even if both were wrong, shown in the main diagonal of each half of Table 4, the accuracy was 98.9% and 96.8% for long TE and short TE, respectively.

Discussion

The classification methodology used here is able to successfully differentiate between tissue types. A lower result was found in the discrimination between healthy subjects tissue and grade III gliomas in short TE data, achieving an AUC of 0.801 in short TE 1.5 T spectra. A reason for this might be due to tissue contamination or low quality of the short TE spectra. Glioma grading takes into account the highest grade observed in the biopsy, and does not mean there are no cells of lower grade glioma in the brain. Also, normal tissue might be present in the spectra thanks to the voxel size. On the other hand, low quality of the spectra might also decrease the differentiation, confounding metabolite peaks and background noise.

Data sources were subject to extensive quantitative quality assessments, at both 3.0 T (Barreto et al., 2014) and 1.5 T (Van der Graaf et al., 2008), and were deemed suitable for quantitative research.

Table 2. Binary classification cross validation results at different field strength (B$_0$) and echo-time (TE). AUCs over 0.9 denoted with boldface.

B0		1.5 T	1.5 T	3.0 T
TE		short	long	long
AA	GBM	**0.797 (4)** *0.674 (1,2)*	0.763 (3) *0.619 (5)*	**0.923 (2)** *0.867 (4)*
AA	LGA	0.703 (2) *0.651 (5)*	0.847 (3) *0.717 (1,2)*	0.733 (3) *0.637 (1,2)*
AA	MEN	0.710 (4) *0.655 (5)*	**0.907 (4)** *0.810 (5)*	0.837 (5) *0.779 (4)*
AA	MET	0.869 (4) *0.804 (1,2)*	0.894 (3) *0.802 (1,2)*	0.896 (3) *0.811 (4)*
AA	ABS	**0.959 (3)** *0.760 (4)*	**0.962 (5)** *0.936 (4)*	-
ABS	GBM	0.838 (3) *0.733 (5)*	0.779 (5) *0.751 (4)*	-
ABS	LGA	**0.983 (3)** *0.953 (4)*	**1.000 (3)** *0.993 (4)*	-
ABS	MEN	**0.940 (3)** *0.822 (2)*	**0.979 (5)** *0.965 (4)*	-
ABS	MET	**0.947 (1,2)** *0.912 (3)*	**0.924 (5)** *0.718 (4)*	-
ABS	HT	**0.996 (3)** *0.966 (4)*	**1.000 (3-5)** *0.998 (1,2)*	-
GBM	LGA	**0.927 (3)** *0.877 (1)*	**0.971 (3)** *0.921 (1,2)*	**0.980 (5)** *0.932 (4)*
GBM	MEN	**0.919 (5)** *0.852 (4)*	**0.956 (3)** *0.900 (1)*	**0.938 (5)** *0.889 (4)*
GBM	MET	0.690 (4) *0.618 (5)*	0.739 (2) *0.620 (4)*	0.691 (4) *0.631 (3)*
LGA	MEN	0.755 (2) *0.722 (5)*	**0.991 (3)** *0.945 (4)*	**0.950 (1,2)** *0.904 (4)*
LGA	MET	**0.984 (5)** *0.949 (4)*	0.989 (5) *0.942 (4)*	0.931 (3) *0.810 (1,2)*
MEN	MET	**0.990 (5)** *0.977 (1)*	0.980 (3) *0.939 (1,2)*	**0.903 (2,5)** *0.818 (4)*
HT	AA	0.801 (4) *0.693 (1,2)*	**0.998 (3)** *0.928 (4)*	-
HT	GBM	**0.999 (5)** *0.986 (3)*	0.997 (3) *0.992 (4)*	-
HT	LGA	**1.000 (3)** *0.883 (4)*	0.988 (3) *0.922 (5)*	-
HT	MEN	**0.988 (3)** *0.950 (4)*	0.998 (3) *0.993 (1,2)*	-
HT	MET	**1.000 (3-5)** *1.000 (1,2)*	1.000 (4) *0.999 (3)*	-

Highest averages are in common characters whereas lowest averages are in italic. The classes are AA = Anaplasic Astrocytoma; ABS = Abscess; GBM = Glioblastoma; LGA = Low Grade Astrocytoma; MEN = Meningioma; MET = Metastases; HT = Healthy Tissue. The digit next to each number refers to the algorithm that achieved that result in that specific task: (1) linear kernel SVM; (2) radial kernel SVM; (3) Random Forest; (4) 3rd degree polynomial kernel SVM; and (5) sigmoid kernel SVM.

Table 3. Multiclass classification cross validation results across field strengths and echo-times (TE).

1.5 T short TE		1.5 T long TE		3.0 T long TE	
AUC	Accuracy	AUC	Accuracy	AUC	Accuracy
0.843	0.752	0.895	0.861	0.755	0.804

The classes are GBM = Glioblastoma; LGA = Low Grade Astrocytoma; MEN = Meningioma; MET = Metastases.

Only the best results are emphasized, as our work is not focused on the comparison between classifiers, but instead on showing in which classification tasks the MRS data alone carries enough information to differentiate between tissue types. AUC over 0.9 was achieved in several binary tasks, but small differences

Res. Biomed. Eng. 2017 September; 33(3): 185-194

Abscesses and brain tumors MRS classification   191

Table 4. Multiclass classification results and radiological findings comparison. Four cells sum to 100% when taken together in each echo-time (TE) evaluation.

| | | MRS classification findings | | | |
| | | 1.5 T long TE | | 1.5 T short TE | |
		Correct	Wrong	Correct	Wrong
Radiological findings	Correct	81.4%	12.4%	71.4%	21.6%
	Wrong	5.3%	0.9%	4.7%	2.3%

in class compositions and resampling strategies make comparisons between studies unreliable. In comparison to the study by Lukas et al. (2004), the most similar to the study presented here, which focused on 1.5 T long TE data, also from the INTERPRET project, using algorithms and resampling strategies similar to the ones used in this study similar results were found in the binary classification tasks. It is important to note that our low-grade glial tumors class consists of grade I-II astrocytomas, whereas the class studied by Lukas et al. consists solely of grade II astrocytomas.

Random Forests ranked high in most tasks while SVMs with linear and polynomial kernel ranked generally lower. We believe this is in part due to the robustness of Random Forests, gained through ensemble averaging and attribute bagging, and in part due to the simpler hypothesis of the linear and polynomial kernel spaces in the case of the SVM. The Gaussian and sigmoid kernel SVMs ranked slightly better, and again this is most probably due to the highly flexible margin provided by these kernels in non-linear space.

The overlap of the classes consisting of metastases and glioblastomas was observed in all experimental conditions. This expected result is in agreement with several previous reports due to the aggressiveness level of these tumor types (Majós et al., 2004; Tate et al., 2003; 2006): despite completely different cell morphology, metastases and glioblastomas have very similar spectra dominated by lipids. NAA is not expected in metastases, but normal tissue contamination is almost unavoidable in the spectroscopy of small single metastases. This led to the union of the two tissue types into one single aggressive tumors category in other studies (Majós et al., 2004; Tate et al., 2003; 2006); an identical strategy was followed in this work. However, the distinction between metastasis and glioblastomas is marginally higher for long TE than it is for short TE spectra. Ishimaru et al. (2001) found that glioblastomas are better resolved from metastases in long TE MRS than in short TE due to overlapping peaks and the dynamic range of the detector being dominated by lipids, diminishing the sensitivity to choline and creatine. This subtle difference suggests that long TE MRS might prove a better option to explore this task in the future with bigger samples.

Our study shows that pattern recognition allied to MRS achieves really high performance in the differential diagnosis of brain abscesses and several tumor types;

one exception was related to glioblastomas. Indeed, Majós et al. (2009) report that MR spectroscopy achieves great levels in the discrimination between tumoral and pseudotumoral tissue. These results indicate that there are differences between abscess and most tumor types considering the spectra, even if not readily visible. Some implications include the possibility of better characterization of post-surgery brain lesions and the contribution to the establishment of differential diagnosis protocols. The diagnostic of brain abscesses is usually made with neuroimaging techniques and, as Lai et al. (2002) point out, there are some advantages regarding the use of diffusion weighted imaging (DWI) instead of 1H-MRS regarding imaging time, despite conflicting findings in the literature regarding DWI.

The difficult classification task between glioblastomas and abscesses shows there is great similarity between both spectral types, as expected due to the presence of lipids in both due to necrosis (Lai et al., 2002). The differential diagnosis can be made through perfusion MRI, as abscesses often demonstrate lower rCBV in the peripheral region than high grade gliomas (Chan et al., 2002).

A well-known finding was confirmed using our methodology. Anaplastic astrocytomas showed overlap with glioblastomas and metastasis and heavy overlap with low grade glial tumors (Majós et al., 2004). This behavior was expected as its metabolic profile has similarities to other tumor types and the diagnosis of anaplastic astrocytomas in the regions of one brain does not preclude the existence of other lower grade gliomas in other regions, augmenting the possibility of contamination. In this work, 3.0 T spectra of anaplastic astrocytomas and glioblastomas achieved a good AUC. One possible explanation is that the use of higher field intensity contributes to a better quantification of J-coupled metabolites somehow related to glioma grading. Alternatively, operator bias could have favored the VOI selection in the 3.0 T dataset.

No differences were obtained for short and long TE data. A previous study showed that short TE achieved just slightly better accuracy than long TE in the binary discrimination between tumor types, and using both TEs improved accuracy (Majós et al., 2004), without considerations about the optimization of thresholds or the separability measured by AUC. In this paper, similar results were obtained for the two magnetic field strengths

data. This might be due to the robustness of magnetic resonance spectroscopy metabolite concentration estimation through LCModel, even for spectra of different quality. 3.0 T spectra underperformed compared to 1.5 T spectra in the multiclass study. Unbalanced classes and smaller sample size are possible causes for this.

The difference in accuracy between the radiological diagnosis and the classification results for the long TE 1.5 T spectra is small, but evident. This is probably due to other information available to the radiology expert, like the shape of the tumor, the occurrence of other primary tumor in other regions, or tumor recurrence. In short TE 1.5 T spectra, the difference in accuracy is bigger in favor of the radiological diagnosis. However, a deeper analysis would be necessary as many instances where the classification was correct were not counted due to ambiguity in the radiological decision, as it does not limit itself to the classes in the present study and has a much broader scope. The complementary character of the application of classification algorithms to radiological differential diagnosis is evident, since in several instances its output was correct when the radiological finding failed. A useful aspect of the use of classification algorithms in these tasks is shown by the overlap of the radiological findings and the classification results. When both the radiologist and the classification agree on the assigned class higher accuracy is guaranteed in either long or short TE spectroscopy when compared to the best individual results. If they disagree, the radiologist is usually right more often. This suggests a strategy that ensures higher confidence to the radiological diagnosis when both diagnoses are the same.

The comparison between classification algorithms is not within the scope of this study. Instead, we aim to show which tasks and how well classification algorithms can perform in the classification of brain tumor magnetic resonance spectra. A drawback in this strategy is we may end with suboptimal models, on the other hand, the effects of overfitting are mitigated.

The number of samples is the biggest limitation regarding the performance of the study. We suggest this favors the dataset with 1.5 T long TE spectra, the biggest by a large margin. The quality of the spectra, specially 1.5 T short TE spectra, may have an impact on the final performance but those are the conditions found in the clinical environment and so the results should still be representative. From combinatorial analysis on our resampled data, the probability of randomly separating sampled metastasis and normal tissue samples in the long TE 1.5 T dataset is approximately 0.1%, and repeating that one hundred times is even less probable. This is also true for other tasks where the mean of resampled AUC evaluations reach unity.

Magnetic resonance spectroscopy also has its limitations, such as inherent variability from the hardware and possible spectral artifacts induced by static field inhomogeneities or crusher gradients amplitudes (Barreto et al., 2014). We believe the earlier acquisitions from the INTERPRET databases might suffer more from noise, since their acquisition were performed with older equipment, as it also depends on spectral resolution and the number of averages.

Using a robust validation technique and five classification algorithms, our work showed the classification of MR spectra of brain masses consistently achieves great performance measured by high AUC and could open new venues in the differential diagnosis of brain tumors and brain abscesses. Our results also provide information on which differential diagnosis are deemed easy to solve using only magnetic resonance spectroscopy information in three experimental conditions: 1.5 T long TE, 1.5 T short TE and 3.0 T long TE point-resolved spectroscopy (PRESS) setups. Previous results from the literature were confirmed as well. We also showed how the accuracy of machine learning methods compare to the radiological diagnosis in both 1.5 T short TE and 1.5 T long TE PRESS single-voxel spectra using only spectral information as input variables. A strategy to improve the accuracy confidence of the radiological diagnosis is provided, showcasing how the use of machine learning algorithms in the differentiation of brain tumors and abscesses can potentially improve the diagnosis.

Acknowledgements

References

Arizmendi C, Sierra DA, Vellido A, Romero E. Automated classification of brain tumours from short echo time in vivo MRS data using Gaussian Decomposition and Bayesian Neural Networks. Expert Systems with Applications. 2014; 41(11):5296-307. http://dx.doi.org/10.1016/j.eswa.2014.02.031.

Barreto FR, Otaduy MCG, Salmon CEG. Evaluation of nuclear magnetic resonance spectroscopy variability. Revista Brasileira de Engenharia Biomédica. 2014; 30(3):242-7. http://dx.doi.org/10.1590/rbeb.2014.023.

Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, et al. mlr: Machine Learning in R. Journal of Machine Learning Research. 2016; 17(170):1-5.

Butzen J, Prost R, Chetty V, Donahue K, Neppl R, Bowen W, Li SJ, Haughton V, Mark L, Kim T, Mueller W, Meyer G, Krouwer H, Rand S. Discrimination between neoplastic and nonneoplastic brain lesions by use of proton MR spectroscopy: The limits of accuracy with a logistic regression model. AJNR. American Journal of Neuroradiology. 2000; 21(7):1213-9. PMid:10954271.

Chan JH, Tsui EY, Chau LF, Chow KY, Chan MS, Yuen MK, Chan TL, Cheng WK, Wong KP. Discrimination of an infected brain tumor from a cerebral abscess by combined MR perfusion and diffusion imaging. Computerized Medical Imaging and Graphics. 2002; 26(1):19-23. PMid:11734370. http://dx.doi.org/10.1016/S0895-6111(01)00023-4.

Demšar J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research. 2006; 7:1-30.

Desprechins B, Stadnik T, Koerts G, Shabana W, Breucq C, Osteaux M. Use of diffusion-weighted MR imaging in differential diagnosis between intracerebral necrotic tumors and cerebral abscesses. AJNR. American Journal of Neuroradiology. 1999; 20(7):1252-7. PMid:10472982.

El-Deredy W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. NMR in Biomedicine. 1997; 10(3):99-124. PMid:9408920. http://dx.doi.org/10.1002/(SICI)1099-1492(199705)10:3<99::AID-NBM461>3.0.CO;2-#.

eTumour Consortium. eTumour: web accessible MR decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data [internet; Technical Report - FP6-2002-LIFESCIHEALTH 503094]. EC; 2008 [cited 2016 Jul 12]. Available from: http://www.etumour.net.

Fan G, Sun B, Wu Z, Guo Q, Guo Y. In vivo single-voxel proton MR spectroscopy in the differentiation of high-grade gliomas and solitary metastases. Clinical Radiology. 2004; 59(1):77-85. PMid:14697379. http://dx.doi.org/10.1016/j.crad.2003.08.006.

Faria AV, Macedo FC Jr, Marsaioli AJ, Ferreira MMC, Cendes F. Classification of brain tumor extracts by high resolution ^{1}H MRS using partial least squares discriminant analysis. Brazilian Journal of Medical and Biological Research. 2011; 44(2):149-64. PMid:21180885. http://dx.doi.org/10.1590/S0100-879X2010007500146.

García-Gómez JM, Luts J, Julià-Sapé M, Krooshof P, Tortajada S, Robledo JV, Melssen W, Fuster-García E, Olier I, Postma G, Monleón D, Moreno-Torres A, Pujol J, Candiota AP, Martínez-

Bisbal MC, Suykens J, Buydens L, Celda B, Van Huffel S, Arús C, Robles M. Multiproject–multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. Magn Reson Mater Phy. 2009; 22(1):5-18. PMid:18989714. http://dx.doi.org/10.1007/s10334-008-0146-y.

García-Gómez JM, Tortajada S, Vidal C, Julià-Sapé M, Luts J, Moreno-Torres À, Van Huffel S, Arús C, Robles M. The effect of combining two echo times in automatic brain tumor classification by MRS. NMR in Biomedicine. 2008; 21(10):1112-25. PMid:18759382. http://dx.doi.org/10.1002/nbm.1288.

Gray HF, Maxwell RJ, Martínez-Pérez I, Arús C, Cerdan S. Genetic programming for classification and feature selection: analysis of 1H nuclear magnetic resonance spectra from human brain tumour biopsies. NMR in Biomedicine. 1998; 11(4-5):217-24. PMid:9719576. http://dx.doi.org/10.1002/(SICI)1099-1492(199806/08)11:4/5<217::AID-NBM512>3.0.CO;2-4.

Gujar SK, Maheshwari S, Björkman-Burtscher I, Sundgren PC. Magnetic resonance spectroscopy. Journal of Neuro-Ophthalmology. 2005; 25(3):217-26. PMid:16148633. http://dx.doi.org/10.1097/01.wno.0000177307.21081.81.

Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. Machine Learning. 2001; 45(2):171-86. http://dx.doi.org/10.1023/A:1010920819831.

Hartmann M, Jansen O, Heiland S, Sommer C, Münkel K, Sartor K. Restricted diffusion within ring enhancement is not pathognomonic for brain abscess. AJNR. American Journal of Neuroradiology. 2001; 22(9):1738-42. PMid:11673170.

Ishimaru H, Morikawa M, Iwanaga S, Kaminogo M, Ochi M, Hayashi K. Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. European Radiology. 2001; 11(9):1784-91. PMid:11511902. http://dx.doi.org/10.1007/s003300000814.

Jackson RJ, Fuller GN, Abi-Said D, Lang FF, Gokaslan ZL, Shi WM, Wildrick DM, Sawaya R. Limitations of stereotactic biopsy in the initial management of gliomas. Neuro-Oncology. 2001; 3(3):193-200. http://dx.doi.org/10.1093/neuonc/3.3.193. PMid:11465400.

Julià-Sapé M, Acosta D, Mier M, Arùs C, Watson D, and the INTERPRET consortium. A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. Magn Reson Mater Phy. 2006; 19(1):22-33. PMid:16477436. http://dx.doi.org/10.1007/s10334-005-0023-x.

Julià-Sapé M, Griffiths JR, Tate RA, Howe FA, Acosta D, Postma G, Underwood J, Majós C, Arús C. Classification of brain tumours from MR spectra: the INTERPRET collaboration and its outcomes. NMR in Biomedicine. 2015; 28(12):1772-87. PMid:26768492. http://dx.doi.org/10.1002/nbm.3439.

Ladroue CLC. Pattern recognition techniques for the study of magnetic resonance spectra of brain tumours [thesis]. London: St George's Hospital Medical School, University of London; 2004.

Lai PH, Ho JT, Chen WL, Hsu SS, Wang JS, Pan HB, Yang CF. Brain abscess and necrotic brain tumor: discrimination with proton MR spectroscopy and diffusion-weighted imaging. AJNR. American Journal of Neuroradiology. 2002; 23(8):1369-77. PMid:12223380.

Lukas L, Devos A, Suykens JAK, Vanhamme L, Howe FA, Majós C, Moreno-Torres A, Van der Graaf M, Tate AR, Arús

C, Van Huffel S. Brain tumor classification based on long echo proton MRS signals. Artificial Intelligence in Medicine. 2004; 31(1):73-89. PMid:15182848. http://dx.doi.org/10.1016/j.artmed.2004.01.001.

Luts J, Poullet J-B, Garcia-Gomez JM, Heerschap A, Robles M, Suykens JAK, Van Huffel S. Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. Magnetic Resonance in Medicine. 2008; 60(2):288-98. PMid:18666120. http://dx.doi.org/10.1002/mrm.21626.

Majós C, Aguilera C, Alonso J, Julià-Sapé M, Castañer S, Sánchez JJ, Samitier A, León A, Rovira A, Arús C. Proton MR spectroscopy improves discrimination between tumor and pseudotumoral lesion in solid brain masses. AJNR. American Journal of Neuroradiology. 2009; 30(3):544-51. PMid:19095788. http://dx.doi.org/10.3174/ajnr.A1392.

Majós C, Julià-Sapé M, Alonso J, Serrallonga M, Aguilera C, Acebes JJ, Arús C, Gili J. Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE. AJNR. American Journal of Neuroradiology. 2004; 25(10):1696-704. PMid:15569733.

Opstad KS, Ladroue CLC, Bell BA, Griffiths JR, Howe FA. Linear discriminant analysis of brain tumour 1H MR spectra: a comparison of classification using whole spectra versus metabolite quantification. NMR in Biomedicine. 2007; 20(8):763-70. PMid:17326043. http://dx.doi.org/10.1002/nbm.1147.

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. Journal of Cancer Research and Clinical Oncology. 1999; 125(6):343-9. PMid:10363566. http://dx.doi.org/10.1007/s004320050284.

Preul MC, Caramanos Z, Leblanc R, Villemure JG, Arnold DL. Using pattern analysis of in vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors. NMR in Biomedicine. 1998; 11(4-5):192-200. PMid:9719573. http://dx.doi.org/10.1002/(SICI)1099-1492(199806/08)11:4/5<192::AID-NBM535>3.0.CO;2-3.

Provencher SQ. Automatic quantitation of localized in vivo 1H spectra with LCModel. NMR in Biomedicine. 2001; 14(4):260-4. PMid:11410943. http://dx.doi.org/10.1002/nbm.698.

R Core Team. R: A language and environment for statistical computing [internet] Vienna: R Foundation for Statistical Computing; 2016 [cited 2016 July 12]. Available from: http://www.r-project.org/.

Ramesh AN, Kambhampati C, Monson JRT, Drew PJ. Artificial intelligence in medicine. Annals of the Royal College of Surgeons of England. 2004; 86(5):334-8. PMid:15333167. http://dx.doi.org/10.1308/147870804290.

Reilly CA, Kowalski BR. Nuclear magnetic resonance spectral interpretation by pattern recognition. Journal of Physical Chemistry. 1971; 75(10):1402-11. http://dx.doi.org/10.1021/j100680a008.

Roda JM, Pascual JM, Carceller F, González-Llanos F, Pérez-Higueras A, Solivera J, Barrios L, Cerdán S. Nonhistological diagnosis of human cerebral tumors by H Magnetic Resonance Spectroscopy and amino acid analysis. Clinical Cancer Research. 2000; 6(10):3983-93. PMid:11051247.

Server A, Josefsen R, Kulle B, Mæhlen J, Schellhorn T, Gadmar Ø, Kumar T, Haakonsen M, Langberg CW, Nakstad PH. Proton magnetic resonance spectroscopy in the distinction of high-grade cerebral gliomas from single metastatic brain tumors. Acta Radiologica (Stockholm, Sweden). 2010; 51(3):316-25. PMid:20092374. http://dx.doi.org/10.3109/02841850903482901.

Sharda R, Barr SH, Mcdonnell JC. Decision support system effectiveness: a review and an empirical test. Management Science. 1988; 34(2):139-59. http://dx.doi.org/10.1287/mnsc.34.2.139.

Soher BJ, Semanchuk P, Young K, Todd D. Vespa-simulation user manual and reference [internet]. 2011 [cited 2016 July 12]. Available from: https://scion.duhs.duke.edu/vespa/project/export/ 2298/trunk/docs/simulation_user_manual.pdf.

Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. BMC Bioinformatics. 2008; 9(1):1. PMid:18647401. http://dx.doi.org/10.1186/1471-2105-9-319.

Tate AR, Griffiths JR, Martínez-Pérez I, Moreno À, 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 in Biomedicine. 1998; 11(4-5):177-91. PMid:9719572. http://dx.doi.org/10.1002/(SICI)1099-1492(199806/08)11:4/5<177::AID-NBM534>3.0.CO;2-U.

Tate AR, Majós C, Moreno A, Howe FA, Griffiths JR, Arús C. Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. Magnetic Resonance in Medicine. 2003; 49(1):29-36. PMid:12509817. http://dx.doi.org/10.1002/mrm.10315.

Tate AR, Underwood J, Acosta DM, Julià-Sapé M, Majós C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabañas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Rémy C, Heerschap A, Watson D, Griffiths JR, Arús C. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. NMR in Biomedicine. 2006; 19(4):411-34. PMid:16763971. http://dx.doi.org/10.1002/nbm.1016.

Tate AR. Pattern recognition analysis of in vivo magnetic resonance spectra [thesis]. Falmer: School of Cognitive and Computing Sciences, University of Sussex; 1996.

Van der Graaf M, Julià-Sapé M, Howe FA, Ziegler A, Majós C, Moreno-Torres A, et al. MRS quality assessment in a multicentre study on MRS-based classification of brain tumours. NMR Biomed. 2008; 21(2):148-58. http://dx.doi.org/10.1002/nbm.1172.

Vellido A, Romero E, Julià-Sapé M, Majós C, Moreno-Torres À, Pujol J, Arús C. Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel 1H MRS. NMR in Biomedicine. 2012; 25(6):819-28. PMid:22081447. http://dx.doi.org/10.1002/nbm.1797.

Wright AJ, Arús C, Wijnen JP, Moreno-Torres A, Griffiths JR, Celda B, Howe FA. Automated quality control protocol for MR spectra of brain tumors. Magnetic Resonance in Medicine. 2008; 59(6):1274-81. PMid:18506793. http://dx.doi.org/10.1002/mrm.21533.