Classification of Brain Tumours from MR Spectra: The INTERPRET Collaboration and its Outcomes

Short Title: Robust classification of brain tumours from MR spectra

Margarida Julià-Sapé*, John R. Griffiths*#, A. Rosemary Tate, Franklyn Howe, Dionisio Acosta, Geert Postma, Joshua Underwood, Carles Majós , Carles Arús#

*Equal contribution

#Corresponding author

Affiliations:

Margarida Julià-Sapé: 1: Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Cerdanyola del Vallès, Spain. 2: Departament de Bioquímica i Biologia Molecular, Unitat de Bioquímica de Biociències, Edifici Cs, Universitat Autònoma de Barcelona (UAB), 08193, Cerdanyola del Vallès, Spain. 3: Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), 08193, Cerdanyola del Vallès, Spain.

John R. Griffiths: 4: Cancer Reasearch UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK

Anne Rosemary Tate: xxx

Franklyn Howe: xxx

Dionisio Acosta: xxx

Geert Postma: xxx

Joshua Underwood: xxx

Carles Majós: Institut de Diagnòstic per la Imatge (IDI), CSU de Bellvitge, Autovia de Castelldefels km 2.7, 08907, L'Hospitalet de Llobregat, Barcelona, Spain

Carles Arús: 2: Departament de Bioquímica i Biologia Molecular, Unitat de Bioquímica de Biociències, Edifici Cs, Universitat Autònoma de Barcelona (UAB), 08193, Cerdanyola del Vallès, Spain. 1: Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Cerdanyola

del Vallès, Spain. 3: Institut de Biotecnologia i de Biomedicina (IBB), Universitat Autònoma de Barcelona (UAB), 08193, Cerdanyola del Vallès, Spain.

KEYWORDS

human, interpret, brain, tumour, magnetic resonance spectroscopy, classifier, database, decision-support system

ABBREVIATIONS

A2: Astrocytoma WHO grade II.
A3: Astrocytoma WHO grade III.
AB: Abscess.
AGG: Aggressive tumour.
ANN: Artificial Neural Network.
AUC: Area Under the Curve.
CDVC: Clinical Data Validation Committee.
CSF: Cerebrospinal fluid.
DMS: Data manipulation software.
DSS: Decision-support system.
eTumour: Web Accessible MR Decision Support System for Brain Tumour Diagnosis and Prognosis, Incorporating in vivo and ex vivo Genomic and Metabolomic Data.
GB: Glioblastoma.
iDB: INTERPRET database.
INTERPRET: International network for Pattern Recognition of Tumours Using Magnetic Resonance.
jDMS: java data manipulation software.
jMRUI: Java MRUI.
KNN: K-nearest neigbour algorithm.
LDA: Linear discriminant analysis.
LGG: Low grade glial tumour.
LY: Lymphoma.
ME: Metastasis.
mI/Gly: m-Inositol glycine ratio.
MN: Meningioma.
MRUI: Magnetic resonance user interface.
NO: Normal brain.
OA: Oligoastrocytoma WHO grade II.
OD: Oligodendroglioma WHO grade II.
PCA: Principal component analysis.
PN: Primitive neuroectodermal tumour.
QC: Quality control.
viDB: validated INTERPRET database.
WHO: World Health Organisation.

ABSTRACT

The INTERPRET project was a multicentre European collaboration, carried out from 2000 to 2002, which developed a decision-support system (DSS) for helping neuroradiologists with no experience of MRS to utilise spectroscopic data for the diagnosis and grading of human brain tumours. Its development involved gathering a large collection of MR spectra of brain tumours and pseudo-tumoural lesions from seven centres. Consensus acquisition protocols, a standard processing pipeline and strict methods for quality control of the aquired data were put in place. Particular emphasis was put on ensuring diagnostic certainty of each case, for which all cases were evaluated by a clinical data validation committee. One outcome of the project is a database of 304 fully-validated spectra from brain tumours, pseudotumoural lesions and normal brains, along with their associated images and clinical data, which is available to the scientific and medical community and has been the basis of twenty-one papers. The second is the INTERPRET DSS, which has continued to be developed and clinically evaluated after the project ended.

We also review here the results of the post-INTERPRET period. We evaluate the results of the studies that have been performed with the INTERPRET database after sharing its data with other consortia or research groups. A summary of the clinical evaluations that have been performed on the post-INTERPRET DSS versions is also presented. Several have shown that diagnostic certainty can be improved for certain tumour types when the INTERPRET DSS is used in conjunction with conventional radiological image interpretation. About 30 papers concerned with the INTERPRET single voxel dataset have so far been published.

262 words, max 300.

THE ORIGINS OF INTERPRET

Magnetic Resonance Spectroscopy (MRS) of cancers in the human body, which has been possible for 30 years (1), provides a unique and entirely non-invasive method for detecting and quantifying metabolites within tumours. It can nowadays be performed on standard clinical MRI instruments; obtaining a matrix of MR spectra from a suspicious cerebral mass adds about 20 minutes to a routine diagnostic evaluation of a patient. However, despite many years of successful research, MRS is rarely used as a routine clinical method. Likely issues that have inhibited its use include (i) the unfamiliarity of the spectra it produces, which are unlike any data normally used in medical diagnosis; (ii) a perceived requirement for an understanding of biochemistry and metabolism on the part of the end-user; and (iii) the need for time-consuming and expert data manipulation, especially if quantitative metabolite concentrations are required.

During the early 1990s Sian Howells and colleagues demonstrated that subjective expert interpretation of MR spectra was not always necessary, since computer-based pattern recognition methods (sometimes termed “chemometrics”) could classify spectra from animal tumour biopsies (2) and from animal tumours scanned non-invasively in vivo by ^{31}P MRS (3). Several studies over the next few years (4,5) demonstrated that statistical methods could be used to classify ^{1}H MR spectra of human brain tumours, indicating both the type of brain tumour (glioma, meningioma etc.) and in some cases its grade of malignancy. A highly malignant glioblastoma multiforme, for instance, has a completely different spectrum from a low-grade glioma (Figure 1).

Although the statistical methods used largely eliminated issue (ii) - the need for understanding the underlying biochemistry -, expert intervention was still required to assign peaks and to quantify the corresponding metabolites, before data could be entered into the classification algorithm. Consequently, these early statistical methods were not widely adopted for routine clinical use.

In contrast, the chemometric classification methods that had originally been developed by Howells (3,6) had used "raw" spectra, with no quantification or even assignment of the peaks. If such an approach could be applied to in vivo ^{1}H spectra of human tumours, there would be no need for any understanding of the underlying biochemistry or for expert data manipulation, addressing issues (i) and (iii) above. The way would then be open for the development of a completely automated method for spectral classification that could be used by radiologists or other medical practitioners who had no specialist knowledge of MRS. To achieve this aim, it would be necessary to gather a sufficiently large dataset of spectra from tumours with known diagnoses, and then to design an algorithm that would compare a spectrum from a patient undergoing diagnosis with the standard spectra in the database.

With this aim in mind, the initial grant application for what became the INTERPRET project was developed in the late 1990s by Rosemary Tate and John Griffiths, who were then at St George's Hospital Medical School, London, responding to a call from the European Union Framework 5. However, several issues had to be addressed before the application could be completed (7).

1.1. Design of the INTERPRET collaboration

First, it was clear that no single hospital would see sufficient patients to be able to create a database of the size required for automated spectral analysis, so a collaborative project would be required. Second, there was at that time no consensus as to the best algorithm for

classifying the spectra, or indeed the most useful features of the spectra that could be used by the algorithm. Thus the collaboration would have to include data analysis experts. It would also be necessary to provide data early in the project for the data analysts to start their work, which was accomplished by creating a preliminary database of retrospective cases, dating as far back as 1994. Retrospective spectra were accepted if they had been obtained using protocols compatible with those used to obtain the prospective spectra, and if their clinical data passed the same validation protocols as prospective cases.

A third issue was the way in which the results would be made available to the user. Bearing in mind the large variety of cancer types, some of them very rare, it was unlikely that a prototype system could function as a "black box" that gave a simple diagnosis, even within one class of tumors. Building on preliminary work at Sussex University (8) it was considered that the best approach was to implement a decision support system (DSS) that would display the unknown case in a feature-space that allowed the user to see how closely it corresponded to the spectra of cases with known diagnoses. Since it was expected that most users would be radiologists, it would be desirable to store images alongside their spectra, together with a limited amount of clinical data, so that the user would be able to call up the spectra, images and clinical details of individual cases. All this information from each standard case would need to be uploaded into the developing database over the web and then be subject to quality control and curating. Experts in all these matters were therefore recruited to the collaboration.

Brain tumours were the obvious exemplar for several reasons. Lesions in the brain give better spectra than those from other parts of the body, because (i) the normal brain lacks the lipid deposits present in most normal tissues (although scalp lipids can be a problem for some peripheral brain tumours), which give confounding peaks in tumour spectra; (ii) the brain is not subject to significant respiratory motion; (iii) it was already known that brain tumours could be classified from their ^{1}H spectra(4,5); (iv) most brain tumours are currently diagnosed by a

histopathologist, using a specimen obtained by stereotactic needle biopsy, which is a very unpleasant experience for the patient and incurs morbidity and even mortality(9-11), so a method that could improve the certainty of radiological diagnosis and reduce the need for biopsy would be welcome(12); (v) several European groups were already working on brain tumour MRS, which facilitated the recruitment of collaborators.

Several issues had to be addressed concerning histological verification of the diagnoses. The majority of cases would be tumours that had been diagnosed and graded by histopathology, either from needle biopsies or surgical specimens. These diagnoses were assigned by a group of pathologists who checked each other's diagnoses for consistency. Cases diagnosed without histological examination (e.g. lymphomas, and "pseudotumoural" neurological lesions such as multiple sclerosis lesions or abscesses which can occasionally be mistaken for tumours) were checked by the Clinical Data Validation Committee (CDVC), consisting of practising neurosurgeons, neuroradiologists and cancer clinicians within the collaboration, which also reviewed and validated the clinical, MRI and surgery data associated with all the spectra.

Another problem was MR system heterogeneity. The collaborating institutions used 1.5T MRI instruments from the three major manufacturers (General Electric, Philips and Siemens) but these instruments and their associated software were of different models and generations. Furthermore, both the instruments and the MRS methods would be constantly updated or even replaced during the period of accrual of spectra. The classification method would therefore have to be able to deal with differences in the spectra due to the different instruments, pulse sequences and other technical matters and to focus only on the type and grade of the tumours. In any case, a classification system that could not cope with spectra obtained on different instruments and with different pulse sequences would be unlikely to be of much practical value.

The EU call required that the collaboration should include commercial partners to ensure that the diagnostic tool developed by INTERPRET would be commercially marketed, so a software company (PRAXIM) and an instrument company (Siemens) were invited to join in.

The project took the form of an EU-funded collaboration (*IST*-1999-10310), from January 1st, 2000, to December 31st, 2002, and was led by Carles Arús, from the Universitat Autònoma de Barcelona in Spain. Seven clinical magnetic resonance centres in six countries participated in the prospective data acquisition (Table 1).

MR methods

In 2000, 1.5T was state of the art in clinical magnets and the participating centres used the following 1.5T machines: GE Signa, Signa Advantage, and LX CV/i 1.5T, Philips NT and ACS NT 1.5T and Siemens Vision 1.5T. Consensus acquisition parameters for Single Voxel spectroscopy are summarised in Table 2. Before the MRS acquisition, a basic set of MR images was acquired to aid voxel placement, ensuring that it was entirely located within the lesion and avoiding contamination from normal adjacent brain tissue or oedema. When contrast was administered, MRS was performed after contrast. Two centres also performed multivoxel MRS imaging (MRSI), one using the PRESS long TE protocol and the other with the STEAM short TE protocol (http://gabrmn.uab.es/interpret/mrs_data/mrs_data.html).

MRS processing pipeline

Consensus acquisition protocols were developed to minimise differences in the data format and the post-processing algorithms and methodologies were standardised (13,14). In 2000, the DICOM standard for spectroscopy was still under development, and retrospective data came with a rich variety of spectral widths and numbers of data points (Table 2). Therefore, an automated processing pipeline, the data manipulation software (DMS), was developed for converting spectra into an “INTERPRET Canonical Format”, a 512-point spectrum covering the

[-2.7, 7.1] ppm range (Table 3). The DMS was an adaptation of the Matlab scripts that constituted the "*pre*-java" Magnetic Resonance User Interface (MRUI) software package (http://sermn02.uab.es/mrui/mrui_versions/mrui_versions.shtml).

The canonical spectra had three different uses: (i) classifier development, (ii) display of the spectrum for quality control, (iii) allowing radiologists to input spectra into the DSS, without the need of cumbersome manual processing.

Databases

Three databases were developed during the project. The INTERPRET preliminary database which was used for the initial pattern recognition experiments in the early stages of the project. The INTERPRET comprehensive database (iDB) which contained all the cases accrued during the project; and the validated INTERPRET database (viDB) (15) which contained the subset of iDB cases that fulfilled all quality control criteria (Table 4).

Quality control of MRS data

The quality assessment system ensured traceability of records and quality control (QC) records for both MRS and clinical data (16). Each instrument's performance was checked bimonthly and, more rigorously, annually by measurements on a specially designed phantom and by spectra obtained from five healthy volunteers per centre. QC of MR spectra uploaded to the database was performed automatically on two indices calculated by the DMS: the signal-to-noise ratio of the processed spectrum and the linewidth of the water peak in the unsuppressed water file (Table 4). A final manual check by a committee of expert spectroscopists looked for other artefacts such as large baseline distortions, insufficient removal of the water peak or large phasing errors. All this QC information is stored in the iDB as metadata.

Quality control of clinical data

The CDVC (http://gabrmn.uab.es/interpret/clinical_data/clinical_data.html) verified each case and tagged those suitable for classifier development. The main criteria were a consensus diagnosis and appropriate clinical information (i.e. age, sex, tumour location). Histopathological assessment was performed on a sample obtained from the surgical specimen or, in some cases, from a pre-operative stereotactic biopsy. Two expert pathologists had to agree the diagnosis (15,17). Additionally, CDVC meetings checked that voxel positioning had been performed according to the protocols and that the diagnostic biopsy had been taken from the place where the voxel had been positioned.

Classification of MR spectra

The project eventually accrued 775 cases with MR data (mostly single voxel, partly multi voxel). The database inevitably contained more cases from common cancers and rarer types were represented by too few cases to form an adequate training set. It was therefore necessary to aggregate together certain cancer types in order to have large enough groups for classifier development.

A key issue, as previously mentioned, was to devise software classifiers that would categorise the spectra purely according to the metabolic profiles of the different classes, and would not be influenced by issues such as the scanner brand, the pulse sequence or the echo time. A preliminary study (18) developed classifiers using spectra obtained with different instruments, using STEAM or PRESS sequences, and with TE values that varied from 20 to 32 ms. None of those factors affected classification performance, provided that all the spectra had been processed in the same way and interpolated to the same number of points and sweep width.

The definitive INTERPRET classifier, a short TE classifier for most common brain tumour types, was developed using features selected by correlation analysis(18). The classifier was trained to

distinguish groups of tumour classes: (i) low-grade glial tumours (LGG) (comprising A2, OA and OD); (ii) aggressive tumours (AGG), (comprising GB and ME); and (iii) low-grade meningiomas (MN) which included meningiomas of WHO grades I and II. The number of features was restricted according to the size of the training set (generally n/3, where n is the number in the smallest group) to avoid overfitting (19,20). The "LGG vs. AG vs. MN" classifier gave excellent results with the independent test set (89% accuracy) and the short TE classifier was chosen over the long TE one for its slightly better performance. Development of a "GL vs. ME" classifier was also attempted; it showed reasonable results (70-80% correct classifications) when data from a single centre were used, but when training data from one centre were tested with data from the other centres (18) accuracy dropped to about 62% (short TE) and 48.7% (long TE).

These INTERPRET classifiers, developed at SGUL by Rosemary Tate, used a very simple technique, linear discriminant analysis, run on commercial programs (SAS, SPSS). In another study, Christophe Ladroue and the SGUL group explored the potential of non-linear methods (21) to automatically decompose a given spectrum into "independent components", corresponding to the different metabolites. The group of Sabine van Huffel from the University of Leuven, also compared the performance of linear and non-linear classifiers in two different studies, one devoted to short TE (22) and the other to long TE (23) and they found that linear and non linear methods performed similarly.

There were also attempts to build classifiers using the multivoxel data. A first approach was the use of the information contained in the variously acquired MRI images (T1, T2, Proton Density and Gadolinium enhanced T1). Unsupervised classifiers were constructed in order to segment an image and identify the possible tumorous area in an objective way. A supervised classifier (linear discriminant analysis) allowed discrimination between healthy and tumor regions. In addition, it allowed discrimination between oligodendrogliomas and astrocytomas

for a limited set of patients (24). A next step was the combined use of MRI and MRSI data. After post-processing, several MRSI peaks were quantified and the MRSI images were aligned with the MRI image data. The MRI intensities of each MRI image modality corresponding to each voxel were then quantified. Using an unsupervised classifier it was possible to segment the voxels in the volume of interest, to identify the deviating and possibly tumorous area and to detect possible heterogeneity (25). Supervised classifiers were also built, based upon these data (26) . This system was capable of segmenting and identifying the volume of interest into voxels belonging to healthy tissue, cerebrospinal fluid, WHO grade II, II and IV glial tumours, unknown areas and voxels on which no decision could be made. The voxels could be coloured according to these classes and an associated probability of each class was provided. This resulted in a prototypical Decision Support System for multivoxel data (27) (Figure 2).

The Decision Support System (DSS)

The DSS was developed by the Human-Computer Interaction team at the University of Sussex (8). Key users (radiologists and spectroscopists, both experts and beginners in MRS) , helped to set the functional requirements of the system, and later to evaluate versions of the system containing preliminary datasets and classifiers (28). Usability aspects evaluated included whether a 2D or a 3D display of cases was more useful, and how the system was used by different user types, with a camera recording the user's actions for further analysis. The final DSS prototype (28) was designed to be both a visual interface, displaying the cases in the database, and an automated classifier for new cases. A 2D scatterplot of the cases in the database was shown in the left-hand panel, with each case represented by a symbol indicating the tumour type (Figure 3). The classifier places similar cases close to each other in the 2D scatterplot. Users can enter a spectrum from an undiagnosed case, which is then automatically positioned in the scatterplot, and they can compare their unknown spectrum to those of the cases in the database using the two right-hand spectral visualisation panels. Individual spectra

can also be compared to averaged plots of the different tumour classes, showing their mean spectra and standard deviations. The user can also create overviews of cases, as 2D scatterplots of ppm heights or ppm height ratios.

INTERPRET AFTER INTERPRET

At the end of its 3-year funding period, in December 2002, the INTERPRET project had successfully developed a prototype DSS and its associated database of spectra, and had performed initial tests which verified that use of the prototype DSS by radiologists could significantly improve their diagnosis and grading of tumour spectra. Work on many aspects of INTERPRET has continued, however.

From the INTERPRET DSS prototype to the industrial INTERPRET DSS

To turn the prototype DSS into a commercially marketable system, PRAXIM, one of the participating industrial partners, developed the "Industrial INTERPRET DSS". In 2003, PRAXIM passed its rights to SCITO, a related company, which continued that work.

In the first step towards CE accreditation, the UAB team verified the traceability of all the cases in the database, giving a final list of 304 cases. SCITO then re-engineered the system to make it automated and robust. The outcome was a client-server application called RADIONET, which was intended to provide a unified solution for the radiological examination process, with the spectroscopy as one of the modules of the RADIONET system (http://www.scito.com/produit_radionet_en.html). In 2004, UAB began further developments of the INTERPRET DSS independently of RADIONET, in part motivated by a need to facilitate bilateral clinical collaborations, especially with Carles Majós at the IDI and with Franklyn Howe at SGUL. The DMS was modified so that data could be entered on-line, and the data associated with some cases was corrected and re-labeled, which meant that the java code had to be manipulated and classifiers retrained. Miquel Cabañas recompiled the DMS, Guillem Mercadal

worked on the DSS code and Margarida Julià retrained the classifiers with the SPSS program (http://gabrmn.uab.es/dss).

In parallel to these developments, the preparation phase of another, larger project, eTumour (ftp://ftp.cordis.europa.eu/pub/lifescihealth/docs/canpr315_en.pdf) led by Bernardo Celda at the University of Valencia, took place during 2003. eTumour expanded the INTERPRET approach to permit a multimodal analysis of MRI, MRSI, ex-vivo transcriptomics (RNA microarrays) and metabolomics (HR-MAS) obtained from brain tumour biopsies from the same patients. The idea was to complement the exact categorisation of brain tumours from in vivo MRS with the ex-vivo information, and to explore the potential of all this information taken together in predicting response to therapy. Since eTumour involved all the original INTERPRET data provider partners as well as new ones (29), it was able to benefit from the lessons learned during INTERPRET, particularly with respect to the acquisition protocols, quality control strategies, database, pattern recognition methods and DSS.

Database preservation

The final version of the INTERPRET iDB contains MRS data (SV or MV spectra) for 775 patients (Table 1). Since paraffin biopsy sections could not be obtained from all retrospective cases, a consensus histopathological diagnosis is available for 477 cases. The number of cases with at least one good quality SV short echo spectrum, acquired from the solid part of the mass in the region where the biopsy or tumor resection took place is 282. These cases, together with data from 22 normal volunteers, form the viDB (15), which continues to be the project's publicly available database (http://gabrmn.uab.es/interpret/int_Disc_FrozenDB.shtml). The viDB can be accessed by two modalities. Anyone requesting access for scientific or medical purposes can be granted "view permission", i.e. being able to look at the data but not being able to download either the processed or the unprocessed spectra; it has resulted in about 50

requests. The second modality provides permission to download the raw data. The download facility is available upon request to the coordinator, who then has to obtain the necessary permission from the original data-contributing partners of INTERPRET. This type of access has been granted to four research groups and two consortia (eTumour and HealthAgents (30)). The data sharing schema has been very successful, allowing numerous pattern recognition studies to be performed on the existing data (31-51), although it should be noted that only those accesses that have involved a scientific collaboration have led to publications.

The complete iDB has been a relatively unexploited resource, particularly those cases with different degrees of validation and completeness. Both INTERPRET databases are administered and maintained by the UAB team.

Using the DMS with the new MRS data formats

Further developments in the DMS have taken place during the post-INTERPRET period. The DMS pipeline has also undergone some minor adjustments: Instead of shell scripts it now uses the C language, and line broadening is now set to 1Hz instead of the 0.8 Hz used in the original study. At UAB, Guillem Mercadal developed the jDMS (52), a java based MRS format conversion software, which translates any spectrum processed with jMRUI into the 512-point INTERPRET canonical format. Using jDMS, manually processed spectra can be automatically converted to the canonical format (52). jDMS eliminated the need for constantly updating the DMS MRS format conversion routines to cope with the numerous MRS formats and updates that have appeared in the 12 years since the original software was developed, keeping the DSS alive and usable. It also permits the use of jMRUI for manual correction of the phase of those spectra in which the Klose algorithm does not work perfectly. The jDMS was, *de facto*, adopted as the general solution by the eTumour (29) and HealthAgents (53) consortia, both to provide database display of processed spectra in any format after 2002, for quality control purposes,

and also to obtain data matrices for classifier development using these multicenter datasets (31,33-36,38), and for the CURIAM DSS (54), which derived from these projects. The jDMS also greatly facilitated scientific collaborations with artificial intelligence research groups who could obtain a consistent, clean data matrix without having to learn how to phase a spectrum or process MRS data . Currently, the DMS pipeline is being integrated into the MRUI software as a plug-in (http://www.transact-itn.eu/). The idea is to automate the data processing of any SV or MV spectral format, ready for exporting the processed files into pattern recognition algorithms or into the INTERPRET DSS.

Pattern recognition: does the particular method matter?

Sharing the SV INTERPRET data resulted in numerous published studies (31-44) which mainly developed classifiers or feature extraction methods. The INTERPRET data was used either as the training set for developing new classifiers or as a test set for the classifiers that had been developed. Both approaches were used in a study performed by eTumour (33), in which 253 pairwise classifiers for GB, MN, ME, and LGG were obtained for 211 short TE spectra from INTERPRET (the training set) and 97 spectra from eight different centres of the eTumour consortium (the test set).

What were the lessons learned from the pattern-recognition studies?

Single voxel

1. Classifiers developed in this way are robust, and perform well on independent data from several centres and manufacturers, acquired at different times by different operators.
2. Most studies addressed relatively easy problems that gave results of around 90% (whether it in terms of accuracy, AUC, or any other measure of efficacy)(33), no matter what feature extraction or classification method were employed. The basic paradigm continued to be the “most common tumour types” problem, first attempted for “LGG vs. AGG vs. MN” in

the original INTERPRET study (55). Later authors (31,36,37,40,52) developed variations and simplifications of these bilateral combinations. In (48), for example, all bilateral combinations of MN, OD, OA, A2, GL, ME, NO, LY, PN, A3 (astrocytomas WHO grade III) and AB were tested. In the main INTERPRET-eTumour paper (33), the seven bilateral combinations were between the GB, ME, MN classes and the AGG superclass. In the latest study using the INTERPRET data, published in 2013 (49), the classes were: "LGG vs. ME", "MN vs. LGG", "LGG vs. GL", "MN vs. GL", "MN vs. NO (normal tissue)" and "GB vs. ME".

3. Difficult problems have more rarely been attempted, the paradigm being "GB vs. ME". Despite having very different (and in the case of ME, very heterogeneous) origins, these two groups of tumours have remarkably similar spectra, so GB and ME were originally joined into a single group for analysis of multiclass (i.e. more than two classes) problems (22,23). Discrimination of GB from ME is a case in which the importance of using an independent test set can be observed. Studies that attempted this bilateral discrimination without using a test set, for example (44,48,50), have claimed results in the 90% range. In contrast, in most studies that used an independent test set to evaluate the robustness of the their classifiers performance was no better than random (23,33,49). A successful classification was achieved in one study (51) that had three distinctive characteristics: first, training the classifier with INTERPRET data and testing it with an independent test set of 40 cases from three different hospitals and two manufacturers' instruments. Second, the use of a simple and well-known classifier (a single-layer perceptron) but an exhaustive feature selection method in which 5 runs of a 5-fold cross-validation of each feature were performed to select an initial set of features. The feature selection process was repeated with the top 50% of the features and again with the top 20% of the features. The third characteristic of this study was the use of both short and long TE concatenated spectra as input. The best classifier used 4 features of the long TE and one from the short TE

spectrum and is available in the latest version of the INTERPRET DSS (http://gabrmn.uab.es/dss).

4. There are also outlier cases that are consistently misclassified in studies using different pattern recognition techniques. Four such outliers (cases I0009, I1390, I0063 and I0450) were first recognized in a study (31) that used an already curated dataset of INTERPRET cases. The same problem was found in the INTERPRET-eTumour study (33). Another study (41) used Sammon's mapping for visualization of cases, coupled to generative topographic mapping to automatically identify outliers. They were categorized into artifact-related (outliers caused by artifacts) and class-related (spectra that are outliers with respect to their class).

5. Short and long TE spectra can be concatenated together and then provide help in some bilateral discriminations, e.g. the "GB vs. ME" problem (51), or to distinguish meningiomas from other tumour types (31).

6. 3T data and 1.5T data, processed with the DMS and converted to the INTERPRET canonical format, can be compatible in terms of classification (35). This study trained a classifier with short TE INTERPRET spectra and tested it with sets of short TE eTumour spectra obtained at 1.5T and 3T. The classification problem was "AGG vs. LGG", either using peak heights or integrated peak areas, and simple classifiers (LDA, KNN and ANN). The results on the test set were similar for the 1.5T and the 3T sets, despite the training set consisting mainly of 1.5T spectra.

In a further development of the INTERPRET approach, the Barcelona group at UAB developed SpectraClassifier (56), a user-friendly program that enables biochemists and other users with no expertise in pattern recognition to make their own MRS classifiers for clinical data and preclinical models. It includes simple tools for feature extraction (PCA), selection (greedy

stepwise (57)) and classification (fisher LDA). The system has been successfully used in classification of MRSI data of preclinical models (58) and human tumours (52).

Multivoxel

The studies performed on MV data after INTERPRET follow the path set by Arjan Simonetti, in which the original dataset accrued and analysed during the project was later used in different follow-up studies by the groups of Lutgarde Buydens and Sabine van Huffel. The dataset comprised 4 volunteers and 20 (26) or 24 patients (the rest of papers cited later in this section) with MN, OD WHO grade II, III and IV tumours (GB); spectra were acquired at 1.5T (25). In contrast to the SV studies, the main technique for dimensionality reduction was peak area integration for 5 (25) ,7 (26), 8 (59) or 10 (60,61) of the main metabolites, although principal component analysis followed by LDA was also used (26). The co-registered T1 weighted pre and post Gadolinium, T2 weighted and proton density images (26) were also used, either for confirming the result of the MV classification with the anatomy (25), or to improve classification (62,63). The MV data were submitted to a variety of techniques that had previously been used with SV data, mostly unsupervised: clustering by mixtures of multivariate normal distributions (25), principal component analysis (26,62) and independent component analysis (62,64), Kohonen networks (60), canonical correlation analysis (59), support vector machines (65), or the latter combined (61). Some of the studies performed recognises the drawback of using supervised methods – particularly to deal with unexpected tumour types. Tumour heterogeneity (66), was tackled in different ways: either by assigning voxels of "unknown" pathology to those voxels that did not reach a certain threshold of Mahalanobis distance to the centroid of the class (26) or by a two-step process in which tumour typing is followed by a segmentation in which the model of "mixed tissue" is introduced (59).

Decision support systems

While the MV interface was not further developed, the SV INTERPRET DSS developed by UAB has evolved into the current 3.1 (52), keeping the look and feel of the original. It currently has classifiers for three different problems: “most common tumor types”, “tumour vs. pseudotumoural” disease (67), and “GB vs. ME” (51). It keeps the “make your own overview” feature, allowing users to make peak height or ratio plots. It also offers the possibility of using short, long or short and long TE spectra of the same case. In fact, the “GB vs. ME” classifier requires the input of short and long TE spectra when entering a new case for analysis (51). The multiplicity of classifiers now means that the user must have a clear idea of the question to be answered and whether to attempt the analysis with short TE, long TE or combined TE spectra.

Versions 1 and 2 of the DSS have been clinically tested. When the final version of the DSS was released at the end of 2002, a retrospective multicentre evaluation of the added value of MRS for diagnosing brain tumours was carried out. It compared diagnostic accuracy using conventional MRI alone with diagnostic accuracy using the DSS in addition to MRI. The experiment involved 20 radiologists (only 4 of whom had been concerned with INTERPRET) from 4 European countries, who reviewed a test sample of 16 cases of different brain tumour types. Radiologists had access to the usual set of films showing T1 weighted images prior to and after contrast enhancement, and T2 weighted images before contrast. They were later presented with a list of possible diagnoses and had to choose one of them and rate its likelihood on a six-point scale. They did so first with the MRI and clinical information and then with additional information from the ^{1}H spectrum analysed and displayed by the DSS. The statistical analysis was done by pooling all cases together and then class by class. The pooled analysis showed that adding the MRS analysis from the INTERPRET DSS significantly improved on diagnoses made using MRI alone (AUC_{MRI} =0.88, $AUC_{MRI+MRS}$= 0.92, n=834 readings). The individual class analysis showed higher AUC after using the INTERPRET DSS in MN, GB and ME,

but only in PN were the AUC significantly different (AUC= 0.50 vs. 0.83 without or with INTERPRET respectively)(55).

Version 1 was also evaluated, together with 3 more systems for spectral classification in a single-centre, prospective study of 40 patients, to see if the MRS information added to MRI analysis was useful in the preoperative diagnosis and grading of brain tumours(68). MRI and MRS data were prospectively acquired at IDI. First, radiologists evaluated MRI and spectroscopists evaluated MRS independently – spectroscopists without any added information related to the patient, radiologists with the usual clinical data. After making a prediction on the tumour type using a 5-point scale, they exchanged their predictions and re-evaluated the information. MRS was found to provide added value in the preoperative radiological evaluation of some abnormal brain masses, depending on the tumour type. For example, MRS did not help MRI in low-grade meningiomas, as they were almost perfectly recognised by both techniques (AUC_{MRI}=0.98, AUC_{MRS}=0.98, n=8). However, there are some classes of brain tumours in which MRS significantly helped MRI. These were: glioblastomas and metastases (AUC_{MRI}= 0.83, $AUC_{MRI+MRS}$= 0.93, n= 12), glial tumours WHO grade III (AUC_{MRI}= 0.70 , $AUC_{MRI+MRS}$= 0.84, n= 12), glial tumours WHO grades II-III (AUC_{MRI}= 0.81, $AUC_{MRI+MRS}$= 0.93, n= 13) and tumours of WHO grade IV (AUC_{MRI}= 0.85, $AUC_{MRI+MRS}$= 0.93 , n= 14). Four different systems for evaluating the spectra were tested, the consensus opinion of the expert spectroscopists, the result from the INTERPRET DSS, area classifiers based on long- and short-TE spectra (69), and the MI/Gly ratio (70). The INTERPRET DSS was the most successful of the four systems, and it significantly out-performed the radiologists in diagnosis of astrocytomas of WHO grade III (AUC_{MRI}= 0.66 , AUC_{DSS}= 0.87, n= 9 cases).

A lesson learned from (68) was that, radiologists from public health system centres would not have the time to embark on formal evaluations of a DSS. Therefore when it was necessary to check whether a new version of the system, offering the choice of an added long TE classifier

would give results consistent with those from the prospective study, 6 radiologists with no particular knowledge of spectroscopy and 3 expert MR spectroscopists were left free to use the system to categorise these 40 cases. In most classes and superclasses of brain tumours the AUC were not significantly different from the first study (68) except for astrocytomas WHO grade III. For this class, radiologists had an AUC_{DSS} =0.59 (n=54 cases/ 238 readings) whereas the expert spectroscopists reached an AUC_{DSS} =0.71 (n=27 cases /116 readings). Radiologists and spectroscopists performed similarly on average, but evaluation of each case required on average 5.4 minutes by neuroradiologists and 9 minutes by spectroscopists (71,72). This study demonstrated that radiologists with no expertise in spectroscopy can use the system, by adhering to basic guidelines. The ability to choose an echo time other than short TE did not affect results, which does not seem to agree with previous literature on classifiers (31,69), however the exact use of the system was not recorded.

The DSS version 2 was also evaluated by the SGUL team as an intervention-limiting tool for patients with a grade IV tumour (73). During the period 2002–2007, 89 brain tumor patients referred to the neurosurgical unit had a short TE (30 ms) single voxel spectrum at 1.5 T. The DSS was used purely as a high or low-grade tumour classifier. Taking the whole cohort, prediction of the grade (high/low) with MRI and clinical findings was 82% accurate with MRI, whereas with MRS and the DSS it was 89% and 84% respectively. All 14 biopsied GB patients were diagnosed as GB using a single neuroradiologist and as high-grade by the DSS. On the basis of their 100% success rate with these biopsied patients the authors suggested that the DSS could be used as a confirmatory test in the quite large group of patients for whom brain biopsy would be contraindicated (about 25% in the Barcelona study (68), such as those with poor functional status or with an unfavourable lesion location.

Taking together, these three evaluations show that the system behaves robustly when used in different ways by different classes of users. It can be used either as for visualising the effects of

classifiers or for decision-suppport, and importantly, to analyse spectra from pathologies for which there is not a specific classifier. In the latter case there will be differences in performance depending on the experience of the person using it and the protocol, which can be critical for certain tumour types, such in as astrocytomas of WHO grade III.

The DSS has been distributed to about 200 users and it is currently in use at Uppsala University Hospital in Sweden (74). A survey was performed among registered users of the DSS in the summer of 2011 and again in the summer of 2012. There were 14 responses: 63% of the respondents used the DSS at between once a month and once every six months. Half used it to evaluate real cases, the remainder as a learning tool or for teaching; 64% of the users worked in hospitals and 84% worked in academia (77% worked in academia in collaboration with public clinical centres). All worked with SV spectroscopy and 62% also with MV. A much larger and more extensive survey is ongoing in the context of the TRANSACT project (http://www.transact-itn.eu/), on the use of jMRUI, the INTERPRET DSS and SpectraClassifier.

ACHIEVEMENTS OF INTERPRET AND POSSIBLE FUTURE DEVELOPMENTS

The INTERPRET collaboration set out to develop a non-invasive method for diagnosing and grading brain tumours using ^{1}H MRS, one that would eventually substitute the subjective analysis of a biopsy in cases where such biopsy is not available or advisable, and would require no understanding of MRS or biochemistry on the part of the user. To do that it accrued several hundred quality-controlled spectra, along with their associated MRI images, histopathology and clinical details; eventually the uploading and quality control procedures were automated. INTERPRET also developed and tested numerous mathematical classifiers and created a decision support system – the DSS – for presenting the data to the user; finally, the overall prototype system was tested several times. How well did INTERPRET succeed in its aims?

Successes

A major technical problem that was solved by the INTERPRET consortium was development of a data input procedure that could cope with the proliferation of incompatible and constantly changing MRS data formats used by the three major manufacturers. This, however, required frequent modification of the DMS data input section, which could not be continued once the project's funding ceased. The development of the jDMS obviated the need for this updating, but introduced weaknesses that will be discussed in the next section.

Additional early successes for INTERPRET were the production of classifier algorithms that worked with raw spectra rather than the metabolite concentrations or integrated peak areas required by previous classification programs, and the demonstration that a useful database could be created by combining spectra acquired on instruments from different manufacturers, using fundamentally different acquisition protocols and numerous generations of software. The software classifiers could be "taught" to ignore all these irrelevant issues and focus only on the differences between the tumour classes. It was even possible to classify 3T spectra using a database composed of 1.5T spectra (35). These achievements are of value in their own right, as they suggest that other pattern recognition programmes of this type could be created.

Several clinical studies have tested the practical usefulness of the INTERPRET prototype. In view of the limited numbers of tumour classes that were adequately represented in its database, it was not expected that the prototype would be able to make definitive diagnoses. Despite those initial misgivings, it has performed successfully in several prospective clinical tests, and it is clear that for certain classes of tumour, notably low-grade glial tumours, it significantly improves the diagnostic ability of experienced neuroradiologists.

The INTERPRET prototype is still in regular use more than a decade after the project ended. Its database has provided a resource for many studies, particularly those that developed new

classifiers, and it is even used by some clinicians to assist diagnosis. It also has a role in teaching. Furthermore, two subsequent EU collaborative projects, eTumour and HealthAgents were built on the foundations laid by INTERPRET.

Weaknesses

There were a number of obvious weaknesses of the INTERPRET project outcome. Because of the limited size of its database it was necessary to aggregate some tumour types into unfamiliar groupings (e.g. “AGG” or “LGG”), while many rarer tumour types were not represented at all. This limits the routine use of the DSS for clinical diagnosis. The same problem meant that the DSS was not configured to offer a proposed diagnosis with a percentage probability, a feature that some users find desirable, or even essential.

Another structural weakness (overcome by brute force during the project itself) was the need for constant updating of the format conversion routines due to changes in the various manufacturers’ data formats. Now that regular updating can no longer be supported the jDMS input program has permitted the on-going use of the INTERPRET system by many researchers and clinicians. However, before uploading a spectrum for classification it is now necessary to process it both with jMRUI and jDMS, so the original aim of developing a system that required the user to have no knowledge of MRS has been compromised.

A more fundamental problem is that the classifiers developed so far are not able to make clinically important distinctions, e.g. beween lymphomas from other grade IV brain tumours, or between glial subclasses or grades.

This could be partly due to the fact that conventional histopathological grading of brain tumours can be difficult (and subjective), particularly for distinguishing intermediate grades (75-78). In INTERPRET, 31.9% of the cases did not have an agreed histopathologic diagnosis (either missing or not agreed) (15). Thus pathologists are less likely to agree on the diagnoses

of intermediate grade. Preliminary investigation of some of these spectra had shown a different profile indicative of an intermediate grade. Indeed, some evidence suggests that spectra may be a better indicator of prognosis in terms of survival than the pathology-assigned WHO grade (79,80).

An objection to the whole idea of developing a non-invasive method for brain tumour diagnosis is that "They are all going to be operated on anyway so why not just biopsy them?". As has been mentioned, however, a significant number of patients with suspicious intracerebral massess are not, or should not be biopsied. In addition, efforts are being made to develop more anticancer drugs that will cross the blood-brain barrier. If non-surgical treatment of brain tumours by chemotherapy (and possibly also radiotherapy) becomes possible then a non-invasive diagnostic test would have great value.

Can the INTERPRET concept be further developed?

Even though MRS has been around as long as MRI, and even though it can be performed on most MRI instruments with minimal upgrading, it has never entered routine clinical use. The original impulse to develop INTERPRET came from a desire to find a "killer app" for MRS – that is an application that would be so compelling that it would cause users to purchase the necessary equipment and software. From a health economics perspective, bringing MRS into routine practice requires that the costs associated with the additional imaging and radiologist reporting time, are balanced against a significant increase in diagnostic performance and the resulting costs of the downstream investigations and treatments. Although successful in many ways, INTERPRET did not pass the "killer app" threshold. What would a future version need in order to bring MRS into routine use for brain tumour diagnosis?

1. Clinical needs:

1.1. Diagnosis prediction. Ideally, the system should diagnose and grade all the cancer types and grades that routinely present at a brain cancer centre although there is no certainty, of course, that all tumour types and grades will have MRS-detectable differential features. The system should also be able to advise a user about which tumour types can not be distinguished with MRS at each particular field or acquisition conditionsHowever, to make routine diagnoses , many more classifiers would also have to be developed including:

1.1.1.1. Grade IV tumours (LY vs. PN vs. GB vs. ME).

1.1.1.2. Grades of glioma (WHO grades II vs. III vs. IV).

1.1.1.3. Subtypes of glioma, both the classical ones defined by conventional histopathology (OD vs. A2 vs. OA) and the emerging ones based on genetic markers (81,82) .

1.1.1.4. Less common tumour types, e.g. hemangiopericytoma (83).

1.1.1.5. Children brain tumours (84-86).

1.1.2. A new INTERPRET database should include 3T spectra and contain adequate numbers (~20) of cases of all the major cancer types and grades (although there will always be some cancers so rare that they cannot be scanned in adequate numbers). This database should also include other types of patients, i.e. children. Given the constant technical improvement in spectroscopic techniques (81,87,88), the database should be an open-ended project.

1.1.3.The system should be particularly optimized to give diagnoses for brain masses in patients who cannot (e.g. because they are too old, too young, too sick etc.) or should not (because their masses will not be surgically removed – e.g. lymphomas, abscesses etc.) be biopsied.

1.2. Prognosis prediction: Ideally the new system should be able to tell whether a tumour type or molecular sub-type may respond or not to a certain therapy. For instance, in glioblastomas, the system should be able to differentiate true progression from pseudoprogression or true response from pseudoresponse (89). That too would require the development of a new database of post-therapy cases. This feature would be particularly enhance user acceptance, as the system would then provide action recommendations, rather than mere assessments (90).

1.3. Regional determination of heterogeneity: Another possibility would be to develop a tool for assigning tumour type and grade to each voxel in a spectroscopic image – in effect using the DSS to create a "nosologic image"(26,63,91), either for diagnosis or for prognosis prediction .

2. Workflow needs:

2.1. In order to integrate seamlessly with radiologist workflow and working practices (90) the DSS should fully automate MRS preprocessing and classification, and ideally should be built in into the standard software of the MRI instrument.

2.1.1. It should give percentage likelihoods for each prediction.

2.1.2. It must give diagnoses in terms that are familiar to clinicians – e.g."oligodendroglioma", or "low grade glial tumour" rather than "LGG".

2.2. The system should be able to compare each new case with a database of cases , replicating the standard radiologist professional practice of consulting peers in intricate and difficult cases. In this way, the system could also tutor users in making use of and understanding the significance of MRS data.

2.3. It should motivate users through game-like diagnostic challenges that would demonstrate the added value of MRS data (92).

Two methodological problems must be overcome if the original vision of INTERPRET is to be realized. The first is to find a way to outsource the integration of spectra from all major manufacturers, MRS methods, software generations etc., a problem that required repeated intervention during the original INTERPRET project and that has not been overcome thereafter. The development of DICOM standards for MRS (Supplement 49 in the standard definition ftp://medical.nema.org/medical/dicom/final/sup49_ft.pdf and ftp://medical.nema.org/medical/dicom/multiframe/Presentations/SCAR-2005/Clunie-Erickson-SCAR2005.pdf, (93)) and the SIVIC software framework (94) are steps in the right direction.

Planning a new DSS and larger database leads to a third problem: who would own and maintain them? One solution would be to develop a commercial DSS that could be sold to users or to MRI instrument manufacturers. The ownership of the database necessary for the use of that DSS (and also for the use of the INTERPRET prototype DSS) is a more difficult question. The current iDB and viDB consist of de-identified spectra obtained from patients in many different countries. All these patients gave informed consent, but numerous national legal frameworks were involved. These databases are owned by the original INTERPRET partners who allow them to be used by the INTERPRET DSS. If a larger database is to be developed it would probably be simplest if it is also owned and maintained by a non-commercial entity, as it would be challenging (and probably expensive) to develop a legal framework that allowed its commercial ownership. One possible solution would be to integrate it in a large data infrastructure such as Elixir, which is currently under construction (http://www.elixir-europe.org/about/rationale). An alternative approach to these technological and medico-legal challenges, would be not to have a centralised database but rather a cloud database of cases, whereby each contributing centre owns and manages its own local database, sharing only de-identified spectral features that are relevant for classification.

Patient consent would then only be required if, in the course of a prospective evaluation of the DSS, patient management would be affected by the system recommendations. This alternative approach poses the technological challenge of redesigning the classification algorithms, so that they can learn and classify based on distributed datasets, which is already an active area of research (95,96). This concept, using agents (30) had already been explored during the EU project HealthAgents (37,97-99) but, unfortunately, with no practical continuity or commercial exploitation after the funding period ended.

All the studies performed so far have been designed to demonstrate that use of MRS data by the prototype DSS improves diagnostic power. When we are developing a DSS for practical clinical use we should consider developing classifiers that use both MRS data and clinical or MRI data. Simple parameters such as the patient's age and sex are already available in the database and could significantly refine certain diagnoses. Another source of information that could be exploited in a future system would be the MRI images (or indeed any other images) of the tumour. Texture analysis software could produce quantitative parameters that could be used by a classifier (100,101) . All these data would be easily accessible to a radiologist.

One final speculation: it has always been tacitly assumed that the aim of INTERPRET and its subsequent developments was to reproduce the diagnosis made by a histopathologist. In reality, it is not impossible that MRS-based classification of cancers might *improve* on histopathology. MRS detects changes in the spectral pattern caused by small-molecule metabolites, lipids and even macromolecules in the living tumour, all of which contain a different type of information from that available on the classical microscope slides used for histopathology, or even from the genomic data that are nowadays coming into use for tumour classification (82,102-105).

Perhaps, for instance, MRS-based classifiers could be devised that will provide a more accurate prognosis for some brain tumour sub-types (79,106,107), or define sub-categories that would help to personalize the treatment of patients. Studies of that type will require another level of information gathering. Long-term data on disease progression and treatment response will have to be collected for the patients whose spectra are in the database and then classifiers constructed that will act as prognostic (i.e. course of disease) and predictive (i.e. response to treatment) biomarkers. Finally, it will be necessary to make prospective predictions by using those classifiers on a new cohort of patients and then to wait several years and see whether the predictions were accurate. The big challenge here would be to develop strategies that may work, at least partially, in a non-supervised or semi-supervised way (46). MRS is already more than 30 years old: let us hope that its future is long enough for these possibilities to be realized.

8564 words

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Drs. Marinette van de Graaf, Arend Heerschap and Des Watson for their comments which helped to improve this manuscript.

Funding from project MARESCAN (SAF2011-23870) from Ministerio de Economia y Competitividad in Spain. This work was also partially funded by CIBER-BBN, which is an initiative of the VI National R&D&i Plan *2008-2011, CIBER Actions* and financed by the *Instituto de Salud Carlos III* with assistance from the European Regional Development Fund. JRG acknowledges support from Cancer Research UK, the University of Cambridge and Hutchison Whampoa Ltd.

REFERENCES

1. Griffiths JR, Cady E, Edwards RH, McCready VR, Wilkie DR, Wiltshaw E. 31P-NMR studies of a human tumour in situ. Lancet 1983;1(8339):1435-1436.
2. Howells SL, Maxwell RJ, Peet AC, Griffiths JR. An investigation of tumor 1H nuclear magnetic resonance spectra by the application of chemometric techniques. Magn Reson Med 1992;28(2):214-236.
3. Howells SL, Maxwell RJ, Howe FA, Peet AC, Stubbs M, Rodrigues LM, Robinson SP, Baluch S, Griffiths JR. Pattern recognition of 31P magnetic resonance spectroscopy tumour spectra obtained in vivo. NMR Biomed 1993;6(4):237-241.
4. Preul MC, Caramanos Z, Collins DL, Villemure JG, Leblanc R, Olivier A, Pokrupa R, Arnold DL. Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. Nat Med 1996;2(3):323-325.
5. Usenius JP, Tuohimetsa S, Vainio P, Ala-Korpela M, Hiltunen Y, Kauppinen RA. Automated classification of human brain tumours by neural network analysis using in vivo 1H magnetic resonance spectroscopic metabolite phenotypes. Neuroreport 1996;7(10):1597-1600.
6. Howells SL, Maxwell RJ, Griffiths JR. Classification of tumour 1H NMR spectra by pattern recognition. NMR Biomed 1992;5(2):59-64.
7. Tate AR, Griffiths JR, Martinez-Perez I, Moreno A, Barba I, Cabanas ME, Watson D, Alonso J, Bartumeus F, Isamat F, Ferrer I, Vila F, Ferrer E, Capdevila A, Arus C. Towards a method for automated classification of 1H MRS spectra from brain tumours. NMR Biomed 1998;11(4-5):177-191.
8. Underwood J, Luckin R, Cox R, Watson D, Tate R. Focussing User Studies: Requirements Capture for a Decision Support Tool. 2000. Citeseer. p 88-92.
9. Favre J, Taha JM, Burchiel KJ. An analysis of the respective risks of hematoma formation in 361 consecutive morphological and functional stereotactic procedures. Neurosurgery 2002;50(1):48-56; discussion 56-47.
10. Hall WA. The safety and efficacy of stereotactic biopsy for intracranial lesions. Cancer 1998;82(9):1749-1755.
11. Field M, Witham TF, Flickinger JC, Kondziolka D, Lunsford LD. Comprehensive assessment of hemorrhage risks and outcomes after stereotactic brain biopsy. J Neurosurg 2001;94(4):545-551.
12. Julia-Sape M, Acosta D, Majos C, Moreno-Torres A, Wesseling P, Acebes JJ, Griffiths JR, Arus C. Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. J Neurosurg 2006;105(1):6-14.
13. Witjes H, Melssen WJ, in 't Zandt HJ, van der Graaf M, Heerschap A, Buydens LM. Automatic correction for phase shifts, frequency shifts, and lineshape distortions across a series of single resonance lines in large spectral data sets. J Magn Reson 2000;144(1):35-44.
14. Simonetti AW, Melssen WJ, van der Graaf M, Heerschap A, Buydens LM. Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction. J Magn Reson 2002;159(2):151-157.
15. Julia-Sape M, Acosta D, Mier M, Arus C, Watson D. 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.
16. van der Graaf M, Julia-Sape M, Howe FA, Ziegler A, Majos C, Moreno-Torres A, Rijpkema M, Acosta D, Opstad KS, van der Meulen YM, Arus C, Heerschap A. MRS

quality assessment in a multicentre study on MRS-based classification of brain tumours. NMR Biomed 2008;21(2):148-158.
17. Murphy M, Loosemore A, Ferrer I, Wesseling P, Wilkins PR, Bell BA. Neuropathological diagnostic accuracy. Br J Neurosurg 2002;16(5):461-464.
18. Tate AR, Majos C, Moreno A, Howe FA, Griffiths JR, Arus C. Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. Magn Reson Med 2003;49(1):29-36.
19. Altman DG, Royston P. What do we mean by validating a prognostic model? Statistics in Medicine 2000;19(4):453-473.
20. Broadhurst D, Kell D. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2006;2(4):171-196.
21. Ladroue C, Howe FA, Griffiths JR, Tate AR. Independent component analysis for automated decomposition of in vivo magnetic resonance spectra. Magn Reson Med 2003;50(4):697-703.
22. Devos A, Lukas L, Suykens JA, Vanhamme L, Tate AR, Howe FA, Majos C, Moreno-Torres A, van der Graaf M, Arus C, Van Huffel S. Classification of brain tumours using short echo time 1H MR spectra. J Magn Reson 2004;170(1):164-175.
23. Lukas L, Devos A, Suykens JA, Vanhamme L, Howe FA, Majos C, Moreno-Torres A, Van der Graaf M, Tate AR, Arus C, Van Huffel S. Brain tumor classification based on long echo proton MRS signals. Artif Intell Med 2004;31(1):73-89.
24. Witjes H, Rijpkema M, van der Graaf M, Melssen W, Heerschap A, Buydens L. Multispectral magnetic resonance image analysis using principal component and linear discriminant analysis. J Magn Reson Imaging 2003;17(2):261-269.
25. Wehrens R, Simonetti AW, Buydens LMC. Mixture modelling of medical magnetic resonance data. Journal of Chemometrics 2002;16(6):274-282.
26. Simonetti AW, Melssen WJ, van der Graaf M, Postma GJ, Heerschap A, Buydens LM. A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy. Anal Chem 2003;75(20):5352-5361.
27. Simonetti A. Investigation of brain tumor classification and its reliability using chemometrics on MR spectroscopy and MR imaging data. Nijmegen, the Netherlands: Radboud University Nijmegen; 2004.
28. Underwood J, Tate AR, Luckin R, Majos C, Capdevila A, Howe F, Griffiths J, Arus C. A prototype decision support system for MR spectroscopy-assisted diagnosis of brain tumours. Stud Health Technol Inform 2001;84(Pt 1):561-565.
29. Julià-Sapé M, Lurgi M, Mier M, Estanyol F, Rafael X, Candiota A, Barceló A, García A, Martínez-Bisbal M, Ferrer-Luna R, Moreno-Torres Á, Celda B, Arús C. Strategies for annotation and curation of translational databases: the eTUMOUR project. Database (Oxford) 2012;doi: 10.1093/database/bas035.
30. González-Vélez H, Mier M, Julià-Sapé M, Arvanitis T, García-Gómez J, Robles M, Lewis P, Dasmahapatra S, Dupplaw D, Peet A, Arús C, Celda B, Van Huffel S, Lluch-Ariet M. HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis. Applied Intelligence 2009;30(3):191-202.
31. Garcia-Gomez JM, Tortajada S, Vidal C, Julia-Sape M, Luts J, Moreno-Torres A, Van Huffel S, Arus C, Robles M. The effect of combining two echo times in automatic brain tumor classification by MRS. NMR Biomed 2008;21(10):1112-1125.
32. Luts J, Poullet JB, Garcia-Gomez JM, Heerschap A, Robles M, Suykens JA, Van Huffel S. Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. Magn Reson Med 2008;60(2):288-298.
33. Garcia-Gomez JM, Luts J, Julia-Sape M, Krooshof P, Tortajada S, Robledo JV, Melssen W, Fuster-Garcia E, Olier I, Postma G, Monleon D, Moreno-Torres A, Pujol J, Candiota

AP, Martinez-Bisbal MC, Suykens J, Buydens L, Celda B, Van Huffel S, Arus C, Robles M. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. Magn Reson Mater Phy 2009;22(1):5-18.
34. Vicente J, García-Gómez J, Tortajada S, Navarro A, Howe F, Peet A, Julià-Sapé M, Celda B, Wesseling P, Lluch-Ariet M, Robles M. Ranking of Brain Tumour Classifiers Using a Bayesian Approach. In: Cabestany J, Sandoval F, Prieto A, Corchado J, editors. Bio-Inspired Systems: Computational and Ambient Intelligence. Volume 5517, Lecture Notes in Computer Science: Springer Berlin / Heidelberg; 2009. p 1005-1012.
35. Fuster-Garcia E, Navarro C, Vicente J, Tortajada S, García-Gómez J, Sáez C, Calvar J, Griffiths J, Julià-Sapé M, Howe F, Pujol J, Peet A, Heerschap A, Moreno-Torres À, Martínez-Bisbal M, Martínez-Granados B, Wesseling P, Semmler W, Capellades J, Majós C, Alberich-Bayarri À, Capdevila A, Monleón D, Martí-Bonmatí L, Arús C, Celda B, Robles M. Compatibility between 3T 1H SV-MRS data and automatic brain tumour diagnosis support systems based on databases of 1.5T 1H SV-MRS spectra. Magn Reson Mater Phy 2011:1-8.
36. Tortajada S, Fuster-Garcia E, Vicente J, Wesseling P, Howe FA, Julia-Sape M, Candiota AP, Monleon D, Moreno-Torres A, Pujol J, Griffiths JR, Wright A, Peet AC, Martinez-Bisbal MC, Celda B, Arus C, Robles M, Garcia-Gomez JM. Incremental Gaussian Discriminant Analysis based on Graybill and Deal weighted combination of estimators for brain tumour diagnosis. J Biomed Inform 2011;44(4):677-687.
37. Saez C, Garcia-Gomez JM, Vicente J, Tortajada S, Luts J, Dupplaw D, Van Huffel S, Robles M. A generic and extensible automatic classification framework applied to brain tumour diagnosis in HealthAgents. The Knowledge Engineering Review 2011;26(Special Issue 03):283-301.
38. Vicente J, Fuster-Garcia E, Tortajada S, Garcia-Gomez JM, Davies N, Natarajan K, Wilson M, Grundy RG, Wesseling P, Monleon D, Celda B, Robles M, Peet AC. Accurate classification of childhood brain tumours by in vivo(1)H MRS - A multi-centre study. Eur J Cancer 2012.
39. Vellido A, Julià-Sapé M, Romero E, Arús C. Exploratory Characterization of Outliers in a Multi-centre 1H-MRS Brain Tumour Dataset. In: Lovrek I, Howlett R, Jain L, editors. Knowledge-Based Intelligent Information and Engineering Systems. Volume 5178, Lecture Notes in Computer Science: Springer Berlin / Heidelberg; 2008. p 189-196.
40. Nebot À, Castro F, Vellido A, Julià-Sapé M, Arús C. Rule-Based Assistance to Brain Tumour Diagnosis Using LR-FIR. In: Lovrek I, Howlett R, Jain L, editors. Knowledge-Based Intelligent Information and Engineering Systems. Volume 5178, Lecture Notes in Computer Science: Springer Berlin / Heidelberg; 2008. p 173-180.
41. Vellido A, Romero E, González-Navarro FF, Belanche-Muñoz LA, Julià-Sapé M, Arús C. Outlier exploration and diagnostic classification of a multi-centre 1H-MRS brain tumour database. Neurocomputing 2009;72(13-15):3085-3097.
42. González-Navarro FF, Belanche-Muñoz LA, Romero E, Vellido A, Julià-Sapé M, Arús C. Feature and model selection with discriminatory visualization for diagnostic classification of brain tumors. Neurocomputing 2010;73(4-6):622-632.
43. Cruz-Barbosa R, Vellido A. Semi-supervised analysis of human brain tumours from partially labeled MRS information, using manifold learning models. Int J Neural Syst 2010;21(1):17-29.
44. Colas F, Kok JN, Vellido A. Finding discriminative subtypes of aggressive brain tumours using magnetic resonance spectroscopy. Conf Proc IEEE Eng Med Biol Soc 2010;2010:1065-1068.

45. 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):5296-5307.
46. Ortega-Martorell S, Ruiz H, Vellido A, Olier I, Romero E, Julia-Sape M, Martin JD, Jarman IH, Arus C, Lisboa PJ. A Novel Semi-Supervised Methodology for Extracting Tumor Type-Specific MRS Sources in Human Brain Data. PLoS One 2013;8(12):e83773.
47. Vilamala A, Lisboa PJG, Ortega-Martorell S, Vellido A. Discriminant Convex Non-negative Matrix Factorization for the classification of human brain tumours. Pattern Recognition Letters 2013;34(14):1734-1747.
48. Arizmendi C, Hernandez-Tamames J, Romero E, Vellido A, Del Pozo F. Diagnosis of brain tumours from magnetic resonance spectroscopy using wavelets and Neural Networks. Conf Proc IEEE Eng Med Biol Soc 2010;2010:6074-6077.
49. Fuster-Garcia E, Tortajada S, Vicente J, Robles M, Garcia-Gomez JM. Extracting MRS discriminant functional features of brain tumors. NMR Biomed 2013;26(5):578-592.
50. Arizmendi C, Sierra DA, Vellido A, Romero E. Brain tumour classification using Gaussian decomposition and neural networks. Conf Proc IEEE Eng Med Biol Soc 2012;2011:5645-5648.
51. Vellido A, Romero E, Julia-Sape M, Majos C, Moreno-Torres A, Pujol J, Arus C. Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel (1)H MRS. NMR Biomed 2012;25(6):819-828.
52. Perez-Ruiz A, Julia-Sape M, Mercadal G, Olier I, Majos C, Arus C. The INTERPRET Decision-Support System version 3.0 for evaluation of Magnetic Resonance Spectroscopy data from human brain tumours and other abnormal brain masses. BMC Bioinformatics 2010;11(1):581.
53. Estanyol F, Rafael X, Roset R, Lurgi M, Mier M, Lluch-Ariet M. A Web-accessible distributed data warehouse for brain tumour diagnosis. The Knowledge Engineering Review 2011;26(Special Issue 03):329-351.
54. Sáez C, Martí-Bonmatí L, Alberich-Bayarri Á, Robles M, García-Gómez JM. Randomized pilot study and qualitative evaluation of a clinical decision support system for brain tumour diagnosis based on SV 1H MRS: Evaluation as an additional information procedure for novice radiologists. Computers in Biology and Medicine 2014;45(0):26-33.
55. Tate AR, Underwood J, Acosta DM, Julia-Sape M, Majos C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabanas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Remy C, Heerschap A, Watson D, Griffiths JR, Arus C. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. NMR Biomed 2006;19(4):411-434.
56. Ortega-Martorell S, Olier I, Julia-Sape M, Arus C. SpectraClassifier 1.0: a user friendly, automated MRS-based classifier-development system. BMC Bioinformatics 2010;11.
57. Hall MA. Correlation-based feature subset selection for machine learning. Hamilton, New Zealand: Waikato; 1999. 198 p.
58. Simoes RV, Ortega-Martorell S, Delgado-Goni T, Fur YL, Pumarola M, Candiota AP, Martín J, Stoyanova R, Cozzone PJ, Julià Sapé M, Arus C. Improving the classification of brain tumors in mice with perturbation enhanced (PE)-MRSI. Integr Biol (Camb) 2012;4(2):183-191.
59. De Vos M, Laudadio T, Simonetti AW, Heerschap A, Van Huffel S. Fast nosologic imaging of the brain. Journal of Magnetic Resonance 2007;184(2):292-301.
60. Melssen W, Wehrens R, Buydens L. Supervised Kohonen networks for classification problems. Chemometrics and Intelligent Laboratory Systems 2006;83(2):99-113.

61. Luts J, Laudadio T, Idema AJ, Simonetti AW, Heerschap A, Vandermeulen D, Suykens JAK, Van Huffel S. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. NMR in Biomedicine 2009;22(4):374-390.
62. Simonetti AW, Melssen WJ, Szabo de Edelenyi F, van Asten JJ, Heerschap A, Buydens LM. Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. NMR Biomed 2005;18(1):34-43.
63. Luts J, Laudadio T, Idema AJ, Simonetti AW, Heerschap A, Vandermeulen D, Suykens JA, Van Huffel S. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. NMR Biomed 2009;22(4):374-390.
64. Szabo de Edelenyi F, Simonetti AW, Postma G, Huo R, Buydens LMC. Application of independent component analysis to 1H MR spectroscopic imaging exams of brain tumours. Analytica Chimica Acta 2005;544(1–2):36-46.
65. Krooshof PWT, Üstün Bl, Postma GJ, Buydens LMC. Visualization and Recovery of the (Bio)chemical Interesting Variables in Data Analysis with Support Vector Machine Classification. Analytical Chemistry 2010;82(16):7000-7007.
66. Segebarth CM, Balériaux DF, Luyten PR, Den Hollander JA. Detection of metabolic heterogeneity of human intracranial tumors in vivo by 1h nmr spectroscopic imaging. Magnetic Resonance in Medicine 1990;13(1):62-76.
67. Majos C, Aguilera C, Alonso J, Julia-Sape M, Castaner S, Sanchez JJ, Samitier A, Leon A, Rovira A, Arus C. Proton MR spectroscopy improves discrimination between tumor and pseudotumoral lesion in solid brain masses. AJNR Am J Neuroradiol 2009;30(3):544-551.
68. Julia-Sape M, Coronel I, Majos C, Candiota AP, Serrallonga M, Cos M, Aguilera C, Acebes JJ, Griffiths JR, Arus C. Prospective diagnostic performance evaluation of single-voxel 1H MRS for typing and grading of brain tumours. NMR Biomed 2012;25(4):661-673.
69. Majos C, Julia-Sape M, Alonso J, Serrallonga M, Aguilera C, Acebes JJ, Arus C, Gili J. Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE. AJNR Am J Neuroradiol 2004;25(10):1696-1704.
70. Candiota AP, Majos C, Julia-Sape M, Cabanas M, Acebes JJ, Moreno-Torres A, Griffiths JR, Arus C. Non-invasive grading of astrocytic tumours from the relative contents of myo-inositol and glycine measured by in vivo MRS. JBR-BTR 2011;94(6):319-329.
71. Julià-Sapé M, Pérez-Ruiz A, Olier I, Camins À, Samitier A, Baquero M, Serrallonga M, Domènech S, Grivé E, Majós C, Aguilera C, Arús C. Multicentre clinical evaluation of the INTERPRET DSS 2.0. Magn Reson Mater Phy. Volume 21. Valencia, Spain; 2008. p 522.
72. Julià-Sapé M, Majós C, Camins A, Samitier A, Baquero M, Serrallonga M, Domènech S, Grivé E, Howe FA, Opstad K, Calvar J, Aguilera C, Arús C. Multicentre evaluation of the INTERPRET decision-support system 2.0 for brain tumour classification. NMR Biomed 2014;In Press.
73. Fellows GA, Wright AJ, Sibtain NA, Rich P, Opstad KS, McIntyre DJ, Bell BA, Griffiths JR, Howe FA. Combined use of neuroradiology and 1H-MR spectroscopy may provide an intervention limiting diagnosis of glioblastoma multiforme. J Magn Reson Imaging 2010;32(5):1038-1044.
74. Weis J, Ring P, Olofsson T, Ortiz-Nieto F, Wikström J. Short echo time MR spectroscopy of brain tumors: Grading of cerebral gliomas by correlation analysis of normalized spectral amplitudes. Journal of Magnetic Resonance Imaging 2010;31(1):39-45.
75. Coons SW, Johnson PC. Regional heterogeneity in the proliferative activity of human gliomas as measured by the Ki-67 labeling index. J Neuropathol Exp Neurol 1993;52(6):609-618.

76. 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 Oncol 2001;3(3):193-200.
77. Prayson RA, Agamanolis DP, Cohen ML, Estes ML, Kleinschmidt-DeMasters BK, Abdul-Karim F, McClure SP, Sebek BA, Vinay R. Interobserver reproducibility among neuropathologists and surgical pathologists in fibrillary astrocytoma grading. J Neurol Sci 2000;175(1):33-39.
78. Kros JM, Gorlia T, Kouwenhoven MC, Zheng PP, Collins VP, Figarella-Branger D, Giangaspero F, Giannini C, Mokhtari K, Mork SJ, Paetau A, Reifenberger G, van den Bent MJ. Panel review of anaplastic oligodendroglioma from European Organization For Research and Treatment of Cancer Trial 26951: assessment of consensus in diagnosis, influence of 1p/19q loss, and correlations with outcome. J Neuropathol Exp Neurol 2007;66(6):545-551.
79. Majos C, Bruna J, Julia-Sape M, Cos M, Camins A, Gil M, Acebes JJ, Aguilera C, Arus C. Proton MR spectroscopy provides relevant prognostic information in high-grade astrocytomas. AJNR Am J Neuroradiol 2011;32(1):74-80.
80. Herminghaus S, Dierks T, Pilatus U, Moller-Hartmann W, Wittsack J, Marquardt G, Labisch C, Lanfermann H, Schlote W, Zanella FE. Determination of histopathological tumor grade in neuroepithelial brain tumors by using spectral pattern analysis of in vivo spectroscopic data. J Neurosurg 2003;98(1):74-81.
81. Andronesi OC, Kim GS, Gerstner E, Batchelor T, Tzika AA, Fantin VR, Vander Heiden MG, Sorensen AG. Detection of 2-hydroxyglutarate in IDH-mutated glioma patients by in vivo spectral-editing and 2D correlation magnetic resonance spectroscopy. Sci Transl Med 2012;4(116):116ra114.
82. Andronesi OC, Rapalino O, Gerstner E, Chi A, Batchelor TT, Cahill DP, Sorensen AG, Rosen BR. Detection of oncogenic IDH1 mutations using magnetic resonance spectroscopy of 2-hydroxyglutarate. The Journal of Clinical Investigation 2013;123(9):3659-3663.
83. Barba I, Moreno A, Martinez-Perez I, Tate AR, Cabanas ME, Baquero M, Capdevila A, Arus C. Magnetic resonance spectroscopy of brain hemangiopericytomas: high myoinositol concentrations and discrimination from meningiomas. J Neurosurg 2001;94(1):55-60.
84. Vicente J, Fuster-Garcia E, Tortajada S, Garcia-Gomez JM, Davies N, Natarajan K, Wilson M, Grundy RG, Wesseling P, Monleon D, Celda B, Robles M, Peet AC. Accurate classification of childhood brain tumours by in vivo (1)H MRS - a multi-centre study. Eur J Cancer 2013;49(3):658-667.
85. Orphanidou-Vlachou E, Auer D, Brundler MA, Davies NP, Jaspan T, MacPherson L, Natarajan K, Sun Y, Arvanitis TN, Grundy RG, Peet AC. (1)H magnetic resonance spectroscopy in the diagnosis of paediatric low grade brain tumours. European journal of radiology 2013;82(6):e295-301.
86. Gill SK, Wilson M, Davies NP, Macpherson L, English M, Arvanitis TN, Peet AC. Diagnosing relapse in children's brain tumors using metabolite profiles. Neuro Oncol 2014;16(1):156-164.
87. Nelson SJ, Kurhanewicz J, Vigneron DB, Larson PEZ, Harzstark AL, Ferrone M, van Criekinge M, Chang JW, Bok R, Park I, Reed G, Carvajal L, Small EJ, Munster P, Weinberg VK, Ardenkjaer-Larsen JH, Chen AP, Hurd RE, Odegardstuen L-I, Robb FJ, Tropp J, Murray JA. Metabolic Imaging of Patients with Prostate Cancer Using Hyperpolarized [1-13C]Pyruvate. Science Translational Medicine 2013;5(198):198ra108.

88. Scheenen TW, Heerschap A, Klomp DW. Towards 1H-MRSI of the human brain at 7T with slice-selective adiabatic refocusing pulses. MAGMA 2008;21(1-2):95-101.
89. Wen PY, Macdonald DR, Reardon DA, Cloughesy TF, Sorensen AG, Galanis E, DeGroot J, Wick W, Gilbert MR, Lassman AB, Tsien C, Mikkelsen T, Wong ET, Chamberlain MC, Stupp R, Lamborn KR, Vogelbaum MA, van den Bent MJ, Chang SM. Updated Response Assessment Criteria for High-Grade Gliomas: Response Assessment in Neuro-Oncology Working Group. Journal of Clinical Oncology 2010;28(11):1963-1972.
90. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005;330(7494):765.
91. De Edelenyi FS, Rubin C, Esteve F, Grand S, Decorps M, Lefournier V, Le Bas JF, Remy C. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. Nat Med 2000;6(11):1287-1289.
92. Underwood J, Luckin R. Adequate Decision Support Systems Must Also Be Good Learning Environments. In: Cerri S, Gouardères G, Paraguaçu F, editors. Intelligent Tutoring Systems. Volume 2363, Lecture Notes in Computer Science: Springer Berlin Heidelberg; 2002. p 1012-1012.
93. DA C, B. E. The New CT and MR DICOM Objects: Why All the Fuss? ; 2005; Orlando, Florida.
94. Crane JC, Olson MP, Nelson SJ. SIVIC: Open-Source, Standards-Based Software for DICOM MR Spectroscopy Workflows. International journal of biomedical imaging 2013;2013:169526.
95. Kargupta H. Next generation of data mining. Boca Raton: CRC Press; 2009. xxiv, 605 p., 608 p. of plates p.
96. Samanthula BK, Elmehdwi Y, Jiang W. k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data. CoRR. Volume abs/1403.5001. http://arxiv.org/abs/1403.5001; 2014.
97. Xiao L, Dasmahapatra S, Lewis P, Hu B, Peet A, Gibb A, Dupplaw D, Croitoru M, Estanyol F, Martinez-Miranda J, Gonzalez-Velez H, Lluch-Ariet M. The design and implementation of a novel security model for HealthAgents. The Knowledge Engineering Review 2011;26(Special Issue 03):261-282.
98. Hu B, Croitoru M, Roset R, Dupplaw D, Lurgi M, Dasmahapatra S, Lewis P, Martinez-Miranda J, Saez C. The HealthAgents ontology: knowledge representation in a distributed decision support system for brain tumours. The Knowledge Engineering Review 2011;26(Special Issue 03):303-328.
99. Hu B, Croitoru M, Roset R, Dupplaw D, Lurgi M, Dasmahapatra S, Lewis P, Martínez-Miranda J, SÃ¡ez C. The HealthAgents ontology: knowledge representation in a distributed decision support system for brain tumours. The Knowledge Engineering Review 2011;26(Special Issue 03):303-328.
100. Herlidou-Meme S, Constans JM, Carsin B, Olivie D, Eliat PA, Nadal-Desbarats L, Gondry C, Le Rumeur E, Idy-Peretti I, de Certaines JD. MRI texture analysis on texture test objects, normal brain and intracranial tumors. Magn Reson Imaging 2003;21(9):989-993.
101. Georgiadis P, Kostopoulos S, Cavouras D, Glotsos D, Kalatzis I, Sifaki K, Malamas M, Solomou E, Nikiforidis G. Quantitative combination of volumetric MR imaging and MR spectroscopy data for the discrimination of meningiomas from metastatic brain tumors by means of pattern recognition. Magnetic resonance imaging 2011;29(4):525-535.
102. Brandes AA, Franceschi E, Tosoni A, Blatt V, Pession A, Tallini G, Bertorelle R, Bartolini S, Calbucci F, Andreoli A, Frezza G, Leonardi M, Spagnolli F, Ermani M. MGMT

promoter methylation status can predict the incidence and outcome of pseudoprogression after concomitant radiochemotherapy in newly diagnosed glioblastoma patients. J Clin Oncol 2008;26(13):2192-2197.
103. Reifenberger G, Weber RG, Riehmer V, Kaulich K, Willscher E, Wirth H, Gietzelt J, Hentschel B, Westphal M, Simon M, Schackert G, Schramm J, Matschke J, Sabel MC, Gramatzki D, Felsberg J, Hartmann C, Steinbach JP, Schlegel U, Wick W, Radlwimmer B, Pietsch T, Tonn JC, von Deimling A, Binder H, Weller M, Loeffler M, for the German Glioma N. Molecular characterization of long-term survivors of glioblastoma using genome- and transcriptome-wide profiling. International journal of cancer Journal international du cancer 2014.
104. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, Campos C, Fabius AW, Lu C, Ward PS, Thompson CB, Kaufman A, Guryanova O, Levine R, Heguy A, Viale A, Morris LG, Huse JT, I.K. M, Chan TA. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature 2012;483:479-483.
105. Brennan Cameron W, Verhaak Roel GW, McKenna A, Campos B, Noushmehr H, Salama Sofie R, Zheng S, Chakravarty D, Sanborn JZ, Berman Samuel H, Beroukhim R, Bernard B, Wu C-J, Genovese G, Shmulevich I, Barnholtz-Sloan J, Zou L, Vegesna R, Shukla Sachet A, Ciriello G, Yung WK, Zhang W, Sougnez C, Mikkelsen T, Aldape K, Bigner Darell D, Van Meir Erwin G, Prados M, Sloan A, Black Keith L, Eschbacher J, Finocchiaro G, Friedman W, Andrews David W, Guha A, Iacocca M, O'Neill Brian P, Foltz G, Myers J, Weisenberger Daniel J, Penny R, Kucherlapati R, Perou Charles M, Hayes DN, Gibbs R, Marra M, Mills Gordon B, Lander E, Spellman P, Wilson R, Sander C, Weinstein J, Meyerson M, Gabriel S, Laird Peter W, Haussler D, Getz G, Chin L. The Somatic Genomic Landscape of Glioblastoma. Cell;157(3):753.
106. Hattingen E, Raab P, Franz K, Lanfermann H, Setzer M, Gerlach R, Zanella F, Pilatus U. Prognostic value of choline and creatine in WHO grade II gliomas. Neuroradiology 2008;50(9):759-767.
107. Pope WB, Prins RM, Albert Thomas M, Nagarajan R, Yen KE, Bittinger MA, Salamon N, Chou AP, Yong WH, Soto H, Wilson N, Driggers E, Jang HG, Su SM, Schenkein DP, Lai A, Cloughesy TF, Kornblum HI, Wu H, Fantin VR, Liau LM. Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy. J Neurooncol 2011;107(1):197-205.

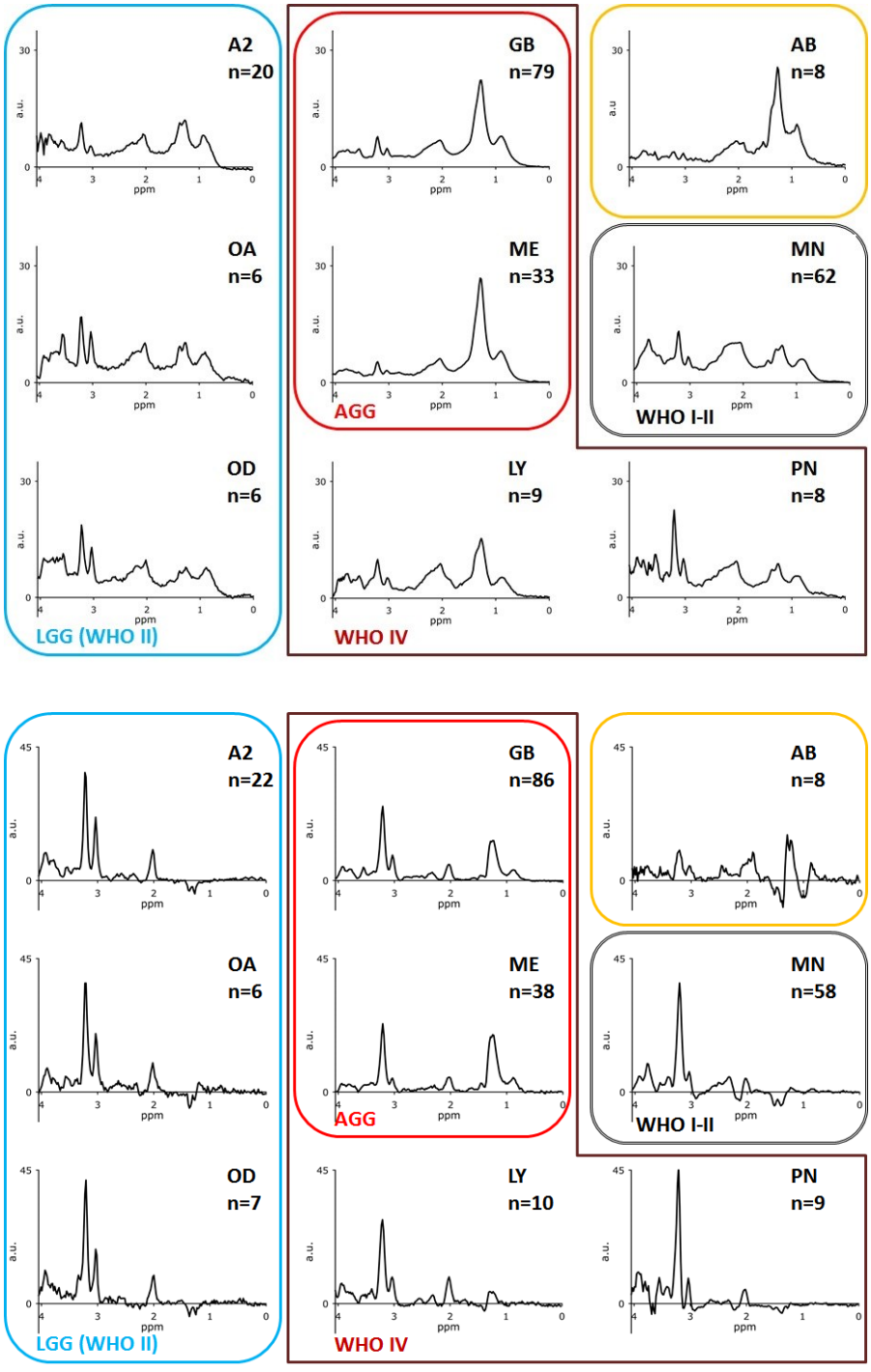

Figure 1. Top, mean (i.e. average) short TE (20-32 msec) mean spectra; bottom, mean long TE (135-144 msec) spectra, both from the INTERPRET validated database (15). A2, astrocytomas of WHO grade II; OA, oligoastrocytomas of WHO grade II; OD, oligodendrogliomas of WHO grade II; GB, glioblastoma multiforme; ME, metastasis; LY, lymphoma; AB, abscess; MN, meningioma; PN, primitive neuroectodermal tumour; LGG, low-grade glial tumours; AGG, aggressive tumours; WHO, World Health Organization; "n", number of cases used to calculate the mean. Roman numbers stand for the WHO grades of the tumours according to the WHO classification. Colour legend: blue, LGG; red, AGG; brown: WHO grade IV; yellow, AB; grey, MN.

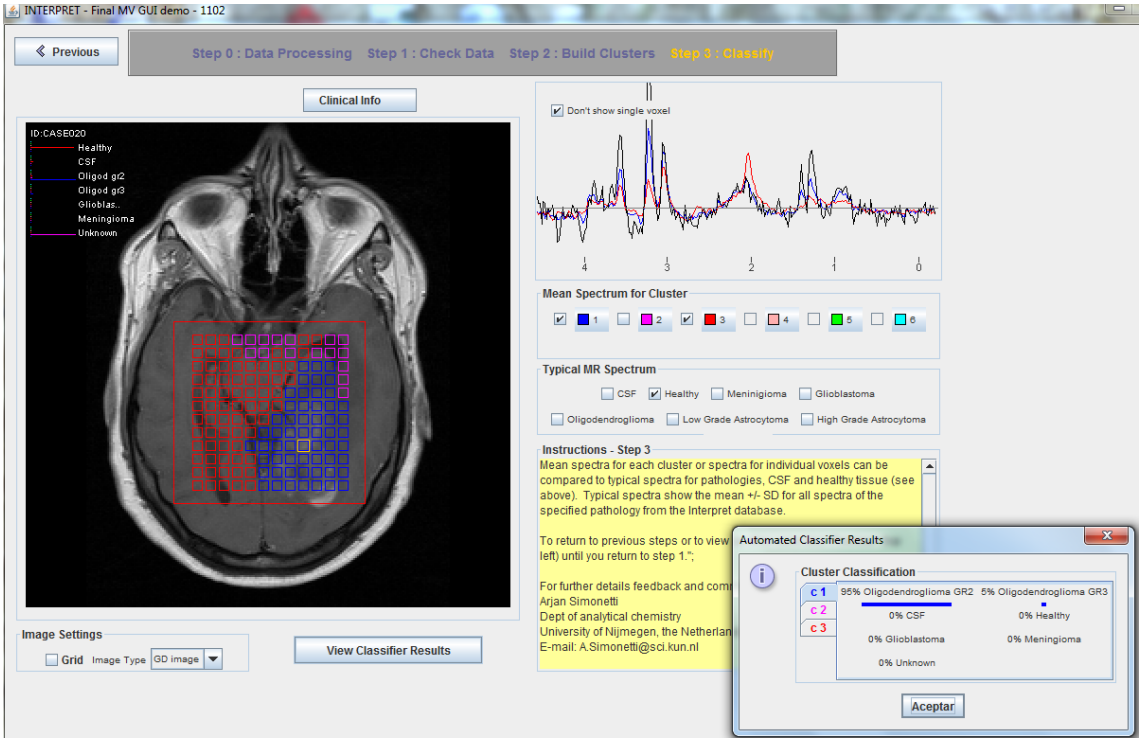

Figure 2. The prototype decision-support system for multivoxel data from Patient I1260, with a diagnosis of OA. The image shows a T1 weighted post gadolinium image on which the results of the clustering algorithm are overlaid. Blue, OD WHO grade II; red, healthy tissue; pink, regions with unspecified diagnosis.

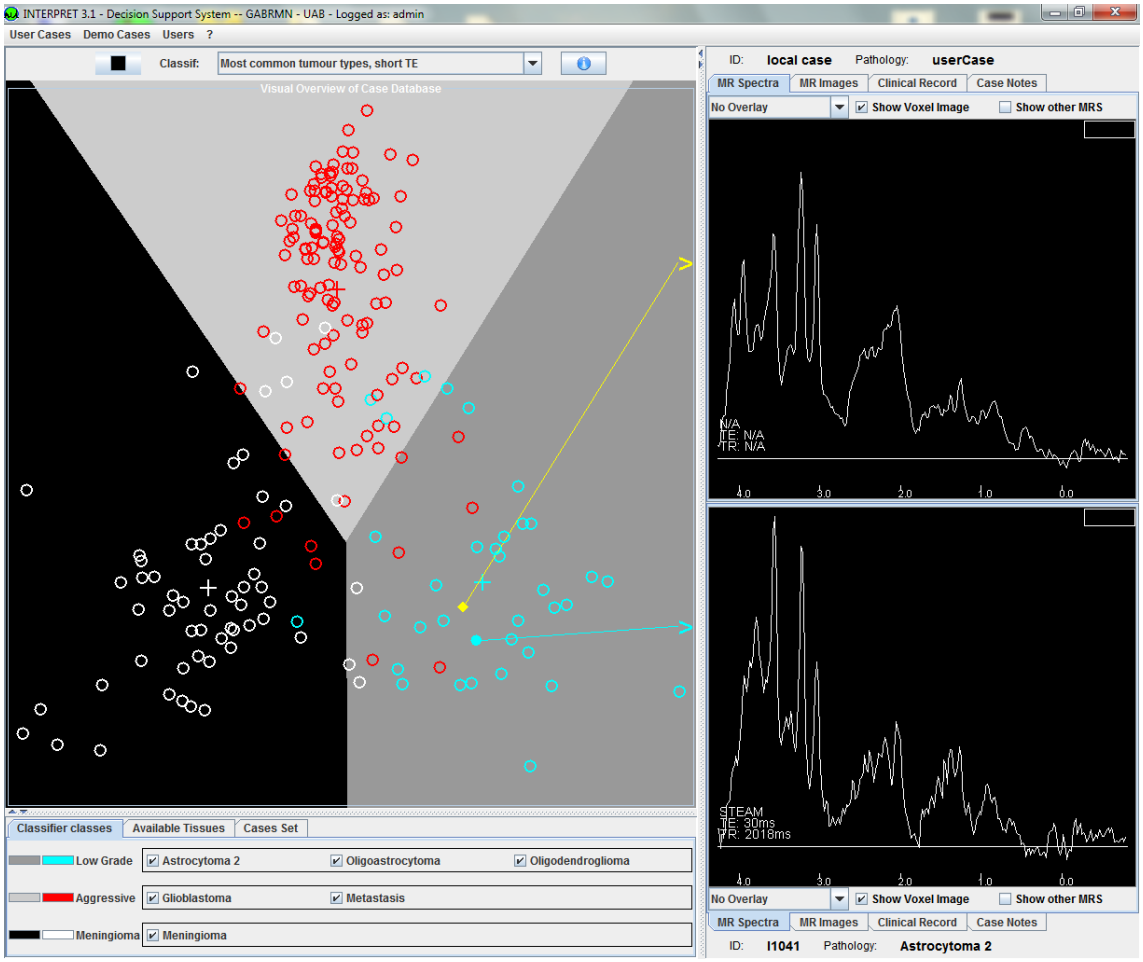

Figure 3. The INTERPRET DSS version 3.1 screen for the "LGG vs. AGG. vs MN" short TE classifier. The screen is divided in two main parts, left and right. The overview space of cases in the database is displayed on the left side, where each case is a coloured circle (see legend on the bottom left). The right part has two panels (top and bottom) for visual inspection of the MRS of individual spectra. In this example, the top right panel displays the short TE spectrum of an astrocytoma of WHO grade II from the study (68), which is displayed as a yellow symbol in the overview space). The bottom right panel displays the short TE spectrum of a case from the INTERPRET validated database: I1041 (blue dot in the overview space).

			Cases contributed			
			iDB			viDB
Contributing centre	**Place**	**Main role**	**Total**	**SV**	**MV**	**SV**
CDP CETIR, Centre Diagnòstic Pedralbes-	Barcelona and Esplugues del Llobregat, Spain	Data contributor	230	230	-	82
IDI, Institut de Diagnòstic per la Imatge-Unitat Bellvitge	L'Hospitalet del Llobregat, Spain	Data contributor	204	204	-	118
SGUL, St George's University of London	London, United Kingdom	Data contributor	159	159	-	75
UMCN, Universitair Medisch Centrum Nijmegen	Nijmegen, The Netherlands	Data contributor	60	46	50	13
UJF, Unité mixte Université Joseph Fourier/INSERM U594	Grenoble, France	Data contributor	70	5	70	-
FLENI, Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia	Buenos Aires, Argentina	Data contributor (Associated partner)	37	37	2	6
MUL, Uniwersytet Medyczny w Lodzi	Lodzi, Poland	Data contributor (Associated partner)	15	15	-	10
UOS, University of Sussex	Brighton, United Kingdom	Database and DSS development, data analysis	-	-	-	-
KUN, Radboud University Nijmegen	Nijmegen, the Netherlands	Data analysis	-	-	-	-
PRAXIM, SARL	Grenoble, France	Commercialisation	-	-	-	-
Siemens AG, Medizinische Technik	Erlangen, Germany	Advisory	-	-	-	-
UAB, Universitat Autònoma de Barcelona	Cerdanyola del Vallès, Spain	Coordination, data managing	-	-	-	-

Table 1. Participating clinical centres as well as roles and number of cases contributed to the databases. Patients scanned at CDP were referred from 6 hospitals in the Barcelona (Spain) area, which contributed clinical data and histology slides. INTERPRET database (iDB), INTERPRET validated database (viDB).

PARAMETER	**STEAM (SHORT TE)**	**PRESS (SHORT TE)**	**PRESS (LONG TE)**
TE	20 ms (20 – 32 ms)	30 -32 ms (30-32 ms)	136 ms (135 – 144 ms)
TR	2000 ms (1600 – 2000 ms)	2000 ms (1600 – 2000 ms)	2000 ms (1600 – 2000 ms)
Volume	4 – 8 cm^3	4 – 8 cm^3	4 – 8 cm^3
Number of averages metabolites	256	192 - 128	192 - 128
Number of averages water	8 to 32	8 to 16	8 to 16
Number of points	512 [Philips] 1024 [Siemens] 2048 [GE]	512 [Philips] 1024 [Siemens] 2048 [GE]	512 [Philips] 1024 [Siemens] 2048 [GE]
Bandwidth	1000 Hz [Philips] 1000 Hz [Siemens] 2500 Hz [GE]	1000 Hz [Philips] 1000 Hz [Siemens] 2500 Hz [GE]	1000 Hz [Philips] 1000 Hz [Siemens] 2500 Hz [GE]
Dummy scans	4	4	4

Table 2. Consensus acquisition protocols for new data with ranges used for retrospective data accepted into the database. TE and TR ranges used for retrospective data are given in parentheses.

ORDER IN WHICH IT WAS PERFORMED	PROCEDURE
1st	Lineshape correction and zero order phasing using water reference with the Klose method
2nd	0.8 Hz exponential line broadening
3rd	Processing by FFT
4th	Water removal by HLSVD: 5 components removed within ±0.37 ppm of water resonance
5th	Residual water suppression: points at 4.2 to 5.1 ppm set to zero
6th	Linear interpolation to 512 points over 1000 Hz of Siemens and Philips data
7th	Spectrum alignment: maximum of choline peak shifted to 3.21ppm
8th	Normalisation of spectrum to Euclidian norm of peak heights

Table 3. Consensus data processing into the INTERPRET canonical format for spectrum display and analysis.

PROCESS	PARAMETERS	ACCEPT IF
Automatic	Water linewidth (WBW)	WBW < 8Hz
Automatic	S = Maximum metabolite signal in range 0 – 3.4ppm; N = standard deviation noise in range 9 – 11ppm. SNR = S/N	SNR > 10
Manual	Visual inspection by expert spectroscopists. Possible artefacts that cause rejection: high scalp lipids; poor phasing; large baseline artefacts; metabolite peaks of suspect origin	2 experts agree

Table 4. Consensus QC of all spectral data.