

## A Prototype Decision Support System for MR Spectroscopy-Assisted Diagnosis of Brain Tumours

Joshua Underwood<sup>a</sup>, A. Rosemary Tate<sup>b</sup>, Rosemary Luckin<sup>a</sup>, Carles Majós<sup>c</sup>, Antoni Capdevila<sup>d</sup>,  
Franklyn Howe<sup>b</sup>, John Griffiths<sup>b</sup>, Carles Arús<sup>e</sup>

<sup>a</sup>*School of Cognitive and Computing Sciences, University of Sussex, Falmer, UK.*

<sup>b</sup>*CRC Biomedical MR Research Group, St George's Hospital Medical School, London, UK.*

<sup>c</sup>*Institut de Diagnòstic per la Imatge, Centre Bellvitge (IDI-Bellvitge), Barcelona, Spain.*

<sup>d</sup>*Centre Diagnòstic Pedralbes (CDP), Cetir Grup Mèdic, Barcelona, Spain.*

<sup>e</sup>*Department de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain.*

### Abstract

*Our objective is to develop a decision support system that improves the accuracy of non-invasive brain tumour diagnosis and grading by enabling radiologists to use data from Magnetic Resonance Spectroscopy (MRS). The system, which uses pattern recognition techniques, is trained on a validated database of spectra and associated clinical information to provide automated classification of spectra from brain tumours. An innovative user-interface presents classification results as a two-dimensional overview plot in which points representing cases of different diseases form distinct clusters. Users can inspect any cases in these plots and compare them with the new, unknown spectrum. Hence, the overview plot can both communicate the classification of a case and help provide explanation for that classification in part by supporting human case-based reasoning. This paper describes the development of a prototype system implemented in JAVA.*

### Keywords:

Computer-Assisted Diagnosis; Decision Support; Pattern Recognition; User-Computer Interface; Magnetic Resonance Spectroscopy; Radiology; Neurology;

### Introduction

When planning treatment for a patient with a suspected brain tumour the clinician needs to know exactly what she or he is dealing with. Currently, the gold standard for brain tumour diagnosis is histopathology. This requires surgical extraction of tissue from the area of the tumour and is, needless to say, unpleasant and even hazardous for the patient [1]. Also, this kind of surgery cannot be constantly repeated to follow tumour evolution or response to treatment. Our objective is to improve the accuracy of non-invasive tumour diagnosis and grading consequently reducing the need for biopsy. We aim to supplement clinical and imaging findings with additional information

about a suspect lesion's biochemistry. This is available from proton (<sup>1</sup>H) Magnetic Resonance (MR) spectra, which can be obtained painlessly during the course of a standard MR imaging examination.

An MR spectrum consists of a series of peaks, where the position of the peak along the x-axis (measured in ppm) determines the particular biochemical, and the size of the peak is a measure of its concentration. However, <sup>1</sup>H MR spectra have complex patterns, and few clinicians have the expertise necessary to interpret them and incorporate findings in their diagnostic reasoning. Automating the diagnostic interpretation of tumour spectra might solve this problem; a variety of pattern recognition techniques have been applied with some success [2]. A widely acknowledged drawback of many such techniques is their failure to provide explanation for the diagnosis [3]. This can lead to the rejection of otherwise promising systems by the intended users; understandably, decision-makers generally want to feel they comprehend the reasoning behind their recommendations [4].

Lillehaug and Lajoie [5] argue convincingly that the most appropriate route to improved health care is through empowerment with new skills as opposed to powerful tools and 'somehow' right answers. We address this concern by combining automated pattern recognition with information visualisation techniques. The result is an overview plot similar to that used in the Image Description (ID) Tutor [6], a training system for radiologists. Pattern recognition routines, developed on a training set of spectra from tumours of known type, allow us to generate two-dimensional overview plots. In these plots, points representing spectra from tumours of the same type are clustered together. The classification of a spectrum can be read visually in terms of its membership of these disease clusters. Plots are interactive and allow the user to view and compare MR spectra and other data for any case. Thus a clinician can compare a new spectrum of unknown type with spectra from cases of known pathology. Comparison with these cases can help provide convincing evidence for a

diagnosis. Users can also compare spectra with statistically 'typical' spectra for each disease in the database (figure 1).

To our knowledge the only other system for automated classification of brain tumour spectra is the nosologic imaging system [7]. This is different both in that it uses multi-voxel spectra to produce a classification image and because it currently provides no explanation for classifications given. We are collaborating with the developers of this system in the design of new prototypes.

Provision of explanation is only one of the challenges decision support system (DSS) designers face. Use of DSSs is not commonplace and a principal reason for this is failure to adequately examine the needs of users [8]. In a review of DSSs in radiology Taylor [9] identifies four critical criteria for successful systems. This paper rephrases these as questions. It then describes the principles and methodology guiding our development strategy. The results section lists the system requirements derived so far and describes a current prototype. We conclude by returning to Taylor's criteria and outlining plans for evaluation.

### Criteria for Successful DSSs

**Need:** Is there a clear clinical need for the system? Reducing current reliance on biopsy for diagnosis could reduce suffering, morbidity and even mortality, as well as costs. Furthermore, in the long run we may be able to supply information about 'in vivo' tumour cell biology, proliferation rate, and apoptosis, thus providing a tool for monitoring tumour evolution and response to therapy not presently available via any other technique.

**Veracity:** Are the data decision support is based on available, accurate and complete? We are compiling and validating a large and comprehensive database of spectra and clinical data from brain tumours.

**Practicality:** Does the system respect the practical constraints imposed by the medical domain, approach to work and clinical setting, in which it will be used? We are working closely with clinicians and using 'in-context' interviews [10], observation and iterative evaluation of prototypes in clinical settings to ensure practicality.

**Relevance:** Can the system provide the user with information in a way that is known to improve decision-making? We are planning evaluations to demonstrate the system's potential for improving the accuracy of 'non-invasive' radiological diagnosis of brain tumours.

## Methodology

### Rationale

Focussing dialogue on prototypes helps to establish a shared understanding and a common vocabulary with which to discuss system features. Prototypes also facilitate the active involvement of the intended users in the design process and importantly allow us to investigate the practicality of the system; i.e. will it be usable in the clinical setting? For

these reasons we developed software prototypes at an early stage.

### Prototype Design

In designing prototypes we follow guidelines drawn from our analysis of the medical DSS literature, particularly in the field of radiology. Here, there is much emphasis on the need for medical systems to integrate smoothly with established clinical routines. Szolovits and Pauker [11] state that decision support systems must not interrupt the normal flow of clinical activity. Clearly, any system that requires lengthy interaction will at best be used only as a last resort. Furthermore, we note that clinicians are widely believed to react more positively towards systems which focus on improving their performance rather than offering an independent diagnosis; hence the development of prompting and critiquing systems. Understanding diagnostic reasoning is consequently important in designing for integration of system use with the diagnostic process. Here, we make use of existing work on radiological diagnosis [12]. Radiological image examination can be both data-driven and hypothesis-driven and expert diagnosticians rapidly 'home in' on the small set of possible diagnoses that best explain the available data [6]. Diagnostic indications from our system may be integrated with this process both by prompting radiologists to perform visual search for image features supporting a system-suggested diagnosis and/or by helping to rule in or out candidates from a set of diagnostic hypotheses suggested by inspection of MR images.

This concept of 'small worlds' of confusable diseases is utilised by the ID-Tutor [6]. Concentrating decision support specifically on sets of 'confusable diseases' both ensures focus on real clinical issues and helps to constrain the pattern recognition problem. The ID-Tutor uses an 'overview plot' based on multiple correspondence analysis of image features to show the relative distribution of cases within a disease and the relative distribution across sets of confusable diseases [6]. We employ similar visualisations. To be effective visualisations must be well suited to the current user-task and allow rapid revision to match new tasks [13]. Understanding the tasks is hence essential. For the diagnostic task that is the focus of our system, hierarchical data visualisation as described by Bishop and Tipping is appealing [14]. This uses multiple projections to reveal structure in complex data sets. A top-level projection might reveal tumour and non-tumour clusters, then zooming in on the tumour cluster, reveals malignant and benign groups etc.

In summary, the principles guiding design of our prototypes are as follows:

- Build for flexibility.
- Provide explanations for automated diagnoses.
- Minimise the need for lengthy user interaction.
- Design to integrate with and complement the current diagnostic process.

- Focus support on small sets of confusable diseases.
- Match data presentation to the current user-task.

### User Studies with Prototypes

We are supplementing dialogues focussed on prototypes with interviews and observation at radiological centres [15]. As prototypes evolve these studies continue; as the system becomes more clearly defined, studies become less exploratory with greater focus on evaluation of specific features. Data gathered from these studies are analysed, and trends and consensus summarised as requirements and design suggestions. We have also developed diagrammatic representations of the current radiological examination process and the process as it might be with the decision support system in place. Comments on process models and requirements help validate our interpretation of the data. This process is supported within our development team and to a wider clinical audience through meetings and via a website<sup>1</sup>. System requirements, models and prototypes are continually refined in response to ongoing feedback.

### Results

Our design approach is evolutionary and in reporting results we are trying to hit a moving target. Recent results for automated pattern recognition techniques indicate that it is possible to categorise malignant tumours vs. benign and low grade vs. high-grade gliomas with an accuracy of approx. 90% [16]. However, here we focus on reporting significant themes and system requirements arising from our user studies. We also describe a recent prototype.

### Themes

Many radiologists will be reluctant to use MRS because of unfamiliarity and lack of confidence in its diagnostic utility. However, several radiologist interviewees noted the potential for use of this system in training.

Initial users of our system are likely to be MRS enthusiasts. These users request explanation for classifications in terms of spectral features and are enthusiastic about tools that help identify these.

Visual comparison with disease ‘typical’ spectra (figure 1) and spectra from previous cases of known pathology and outcome is generally felt to be useful for diagnosis.

Radiologists expect to be able to view MR images and possibly clinical information when comparing cases.

Numerical measures of relative confidence in differential diagnoses offered by an automated system are expected.

### Requirements

Discriminate not only tumours but also other potentially confusable lesions, such as abscesses.

Automate spectral acquisition and separate it from decision support. These tasks are performed in different places (at the scanner and in the reporting room) and possibly by different people (radiographer and radiologist).

Provide the ‘overview plot’ most appropriate for solving the current diagnostic problem.

Facilitate comparison of new spectra with ‘statistically typical’ spectra for each disease type in the database.

Facilitate the simultaneous display and comparison of data from two or three cases.

Provide all potentially useful information with each case (MR images, clinical history, outcome).

Facilitate identification of important spectral features and provide advice on probable diagnostic significance.

Highlight the features that lead to automated classification.

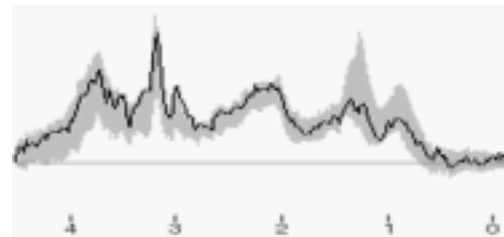


Figure 1. MR spectrum from a meningioma overlaid on the ‘typical spectrum’ (mean +/- S.D. grey) for meningiomas.

### The Prototype

#### Database

The prototype described here uses short echo time (TE=20ms or 30ms) spectra from tumours (Astrocytomas, Glioblastomas, Metastases, Meningiomas, and Oligodendrogliomas) and normal brain (occipital and parietal matter). Data were acquired on clinical Signa 1.5 T MR instruments at two centres<sup>b,d</sup>. Where available for a case, clinical information, selected MR images and additional (TE=136ms) spectra are included in the database.

#### Pattern Recognition and Data Visualisation

To generate ‘overview plots’ we used Linear Discriminant Analysis (LDA) [17], which finds the projection that maximises inter-group distances while minimising intra-group distances. Classifiers were developed using processed and normalised short TE spectra [18]. Principal component (PC) analysis [17] of the datapoint intensities provided discriminatory variables for LDA. The first 5 PC’s were used as input to LDA.

For each pairwise combination of tumour types we calculated a separate overview by performing LDA on the corresponding data subset. LDA provides one less discriminant function than the number of groups to classify; to give two-dimensional plots we always included normal brain as a third group. Overviews plot each case using its discriminant function values. Left of figure 2 shows the

<sup>1</sup> <http://www.cogs.susx.ac.uk/users/joshuau/interpret/>

overview for low-grade Astrocytomas, high-grade Astrocytomas and normal brain. Classifying the entire dataset in three groups, normal brain, malignant tumours and benign tumours provides a top-level overview.

### User Interface

Users select which tumour types are candidate diagnoses and the corresponding overview is displayed. They then introduce the MRS data for a new case; this is automatically processed and the case is positioned in the overview. LDA uses the closest mean case (crosses in the overview) to calculate a case's classification. Boundaries indicating which group mean is closest at any point can be displayed in the overview. The relative distances from any point to any mean case can also be read numerically; bottom left figure 2. The typicality of a case can be assessed in terms of its relation to a disease cluster.

Any case in the overview can be selected and spectra, MR images and clinical information compared with the corresponding data for any other case, including mean cases. The right of figure 2 shows data for the two cases selected in the overview. Spectra can also be compared with the typical spectrum for a disease; see figure 1. Interactive features in the spectrum display areas assist in interpretation of the spectra; ppm value at any position is identified and possible biochemical assignments are suggested. The regions of a spectrum that contribute most to its position in the overview, and hence its classification, can also be highlighted in the spectrum display area.

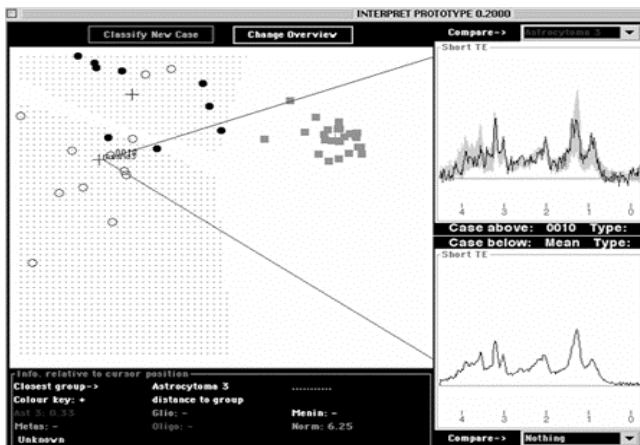


Figure 2. Prototype User Interface

## Discussion

Here we return to the third and fourth of Taylor's criteria for successful DSSs, practicality and relevance [9].

### Practicality

Our system is designed to integrate with the existing clinical context and diagnostic process. On the majority of 1.5T clinical MR systems,  $^1\text{H}$  MR spectra of brain lesions can be routinely acquired during a standard MR examination.

Acquisition of spectra on commercial MR systems is now largely automated, requiring very little extra training and only the decision of where in an abnormality the MRS data will be acquired. This system can provide decision support on-demand at workstations conveniently located both in the reporting room and at the scanner. The requirement for user-interaction is minimal and processing is virtually instantaneous; the user clicks one button, selects the new case data files and views the classification result.

Classifications given by the system can be integrated into the routine clinical diagnostic process either to add weight to a tentative diagnosis suggested by initial inspection of the images, or to prompt consideration of alternative diagnostic hypotheses. Explanation for system-suggested diagnoses is provided primarily via comparison with similar known cases and statistically 'typical' spectra. Users are free to choose how much time they invest in exploring the justification for a classification and this will perhaps vary in relation to the case's diagnostic difficulty.

Our development strategy recognises that ultimately the practicality of a system is determined in use. Introduction of novel systems can change the way people work and consequently system requirements in unpredictable ways that often only become clear in the presence of the device [19]. Iterative validation of requirements, using realistic prototypes with users, can avoid this problem. We are providing clinicians with functional prototypes ready for evaluation in context. We expect this process both to reveal problems that will contribute to our growing understanding of how the system can be used in practice and to drive its evolution through continuous redesign.

### Relevance

Although diagnosis is not made from an MR image in isolation, in many situations the MR image does not always provide the radiologist with enough information to lead to an unequivocal diagnosis. Pattern recognition techniques similar to those we employ are reported to give around 90% accuracy for classification of certain kinds of tumour [2]. However, we are designing a DSS and like others [20] acknowledge that it is the clinician, not the machine, who is ultimately responsible for making decisions. Consequently, rather than measuring automated classifier performance, we plan experimental evaluations comparing the diagnostic performance of radiologists in two conditions, with the DSS and without it.

## Conclusion

We have described a prototype DSS for MRS-assisted brain tumour diagnosis, which presents results from LDA of a tumour spectra database in an 'overview plot'. Future prototypes will incorporate a comprehensive fully validated archive of cases and refined pattern recognition techniques, which may also make use of image features. Navigation through 'overviews' will be restructured to better match specific diagnostic sub-tasks as identified by clinicians.

Experimental evaluations will reveal any potential for improving the accuracy of radiological brain tumour diagnosis. The system continues to evolve in response to data gathered from prototypes in the clinical context.

## Acknowledgements

The EU IST programme (IST-1999-10310: INTERPRET) provided financial support for much of this work. C.A. is also funded by CICYT SAF99-0101 and C.A., A.C. and C.M. by CIRIT, SGR-328. We also thank the Cancer Research Campaign (CRC) and the Royal Society.

## References

- [1] Barnett GH, Miller DW, Weisenberger J. Frameless stereotaxy with scalp-applied fiducial markers for brain biopsy procedures: experiences in 218 cases. *Journal of Neurosurgery* 1999; 91(4):567-576.
- [2] Tate AR, Howells S. (eds) Pattern Recognition Analysis. *NMR Biomed* 1998; 11(4/5):147-256.
- [3] Lavrac N. Machine Learning for Data Mining in Medicine. In: Horn W, Shahar Y, Lindberg G, Adreassen S, Wyatt J, eds. *Proceedings of AIMDM'99*: Springer; 1999; pp. 47-62.
- [4] Swets J, Dawes R, Monahan J. Better Decisions through Science. *Sci Am* 2000; 283(4):82-87.
- [5] Lillehaug S, Lajoie S. AI in medical education - another grand challenge for medical informatics. *Artif Intell Med* 1998; 12(3):197-225.
- [6] du Boulay B, Teather B, du Boulay G, Jeffery N, Teather D, Sharples M, Cuthbert L. From Description to Decision. In: Horn W, Shahar Y, Lindberg G, Adreassen S, Wyatt J, eds. *Proceedings of AIMDM'99*: Springer; 1999; pp. 93-102.
- [7] De Edelenyi FS, Rubin C, Estève F, Grand S, Décorps M, Lefournier V, Le Bas J-F, Rémy C. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. *Nature Medicine* 2000; 6(11):1287-1289.
- [8] Delaney B, Fitzmaurke D, Riaz A, Hobbs R. Can computerised decision support systems deliver improved quality in primary care? *Brit Med Journ* 1999; 319:1-3.
- [9] Taylor P. Computer aids for decision making in radiology. *Brit Journ Radiology* 1995; 68:945-957.
- [10] Beyer H, Holtzblatt K. *Contextual Design*. San Francisco: Morgan Kaufmann; 1998.
- [11] Szolovits P, Pauker SG. Categorical and Probabilistic Reasoning in Medicine Revisited. *Artif Intell* 1993; 59:167-180.
- [12] Cuthbert L, du Boulay B, Teather D, Teather B, Sharples M, du Boulay G. Expert/Novice differences in Diagnostic Medical Cognition - A review of the Literature: University of Sussex; 1999. CSRP 508.
- [13] Shneiderman B. *Designing the User Interface*. 3rd ed: Addison-Wesley; 1998.
- [14] Bishop CM, Tipping ME. Latent Variable Models and Data Visualisation. In: Kay JW, Titterton DM, eds. *Statistics and Neural Networks: Recent Advances at the Interface*. Oxford: OUP; 1999; pp. 147-164.
- [15] Underwood J, Luckin R, Cox R, Watson D, Tate R. Focussing User Studies: Requirements Capture for a Decision Support Tool. In: Singer J, Storey M, Sim S, eds. *Proc ICSE2000 Workshop 5*: 2000; pp. 88-92.
- [16] Tate AR, Griffiths JR, Howe FA, Pujol J, Arús C. Differentiating Types of Human Brain Tumours by MRS. A Comparison of Pre-processing Methods and Echo Times. *Proc ISMRM 2001* 2001.
- [17] Krzanowski WJ. *Principles of Multivariate Analysis*. Oxford: Clarendon Press; 1988.
- [18] Tate AR, Griffiths JR, Martínez Pérez I, Barba I, Cabañas ME, Watson D, Alonso J, Bartemeus F, Isamat F, Ferrer I, Vila F, Ferrer I, Capdevila A, Arús C. Towards a Method for Automated Classification of <sup>1</sup>H MRS Spectra from Brain Tumours. *NMR Biomed* 1998; 11(4-5):177-191.
- [19] Arthur JG, McCarthy AD, Wynn HP, Harley PJ, Baber C. Weak at the Knees? Arthroscopic Surgery Simulation User Requirements. In: *Proceedings of INTERACT'99*: IOS; 1999; pp. 360-365.
- [20] Friedman CP, Elstein AS, Wolf FM, Murphy GCM, Franz TM, Heckerling PS, Fine PL, Miller TM, Abraham V. Enhancement of Clinicians' Diagnostic Reasoning by Computer-Based Consultation. *Journ Am Med Assoc* 1999; 282(19):1851-1856.

## Address for correspondence

Joshua Underwood, School of Cognitive and Computing Sciences, University of Sussex, Falmer, UK.

E-mail: joshuau@cogs.susx.ac.uk.