The clinical perspective of large scale projects: A case study of multiparametric MR imaging of pediatric brain tumors

L. ASTRAKAS[1], SONG YE[2], M. ZARIFI[1], F. MAKEDON[2] and A. ARIA TZIKA[1]

[1]NMR Surgical Laboratory, Massachusetts General Hospital, Boston, MA 02114;
[2]Dartmouth College, 6211 Sudikoff Laboratory, Hanover, NH 03755, USA

Abstract. Current developments in medical information technologies provide the clinical researcher with overwhelming amounts of data that need to be retrieved, organized, analyzed, and shared using secure, efficient, and robust protocols. The development of a local research database can provide an infrastructure for improved data management and detailed data analysis. For example, a pediatric brain tumor database of magnetic resonance imaging data, including conventional MRI imaging, hemodynamic MRI, diffusion weighted MRI and MR spectroscopic imaging, combined with neuropathological and neurological evaluation data, will significantly enhance the assessment and treatment of pediatric brain tumor patients. Furthermore, a negotiation system by which different clinical research facilities can share and combine data will permit re-analyses and meta-analyses of large data arrays that are beyond the focus or time constraints of the original researchers. Such a system will greatly enhance the utility of different data sets to a wide array of scientists. At present, efforts to organize medical data locally and between different sites is limited by diversity, interoperability, security, and accountability difficulties.

Introduction

The progress in medical technology coupled with increased computing capabilities have created a new landscape in the hospital environment. Automated diagnostic medical procedures have 'industrialized' the acquisition and storage of medical data, which have created both opportunities and difficulties regarding their management and exploitation. A representative example of this trend is the Picture Archiving and Communications System (PACS), which in combination with the Radiology Information System (RIS), has created filmless radiology departments and revolutionized access to diagnostic images (1). Digital organization schemes in modern hospitals now allow both the clinician and researcher to readily incorporate data from different imaging modalities. This multi-source approach can significantly aid both clinical patient evaluation and the production of new scientific knowledge, as we have established for pediatric brain tumors.

Multi-center cooperative research organizations, such as the Pediatric Brain Tumor Consortium (PBTS; www.pbts.org), which is focused on central nervous system (CNS) childhood tumors, often bring together resources from different institutions and communities. Such large-scale organizations are typically directed at the improvement of treatment strategies, clinical protocols, data analysis, and the under-standing and cure of particular diseases. To date, the utility of such organizations has often been limited by their centralized structures and the absence of complete, secure, and effective methods for sharing data and methodologies among different researchers. Based on our clinical experience with pediatric brain tumors, we present steps to build research-oriented organizations that overcome these limitations. In our scheme, the members of multi-center collaborations place their data in a central archive to create a large-scale database, which is then accessible, under different conditions, to the individual members.

Materials and methods

Our local pediatric brain tumor patient database includes magnetic resonance imaging (MRI) data, combined with neuropathological and neurological evaluation data. While this database is based on more than 200 initial patients, specific inclusion criteria (e.g. type of tumor, treatment protocol, exam protocol) limit its largest dataset to 66 subjects. Magnetic resonance imaging (MRI) is widely used in the diagnosis and follow-up of pediatric brain tumor patients because of its ability to provide anatomical detail in the absence of invasive procedures. Conventional MR imaging is limited as an analytical tool to estimate the extent of active tumors because it fails to provide tissue biochemistry data,

---

Correspondence to: Dr A. Aria Tzika, Department of Surgery, Massachusetts General Hospital, Harvard Medical School, 51 Blossom Street, Room 261, Boston, MA 02114, USA
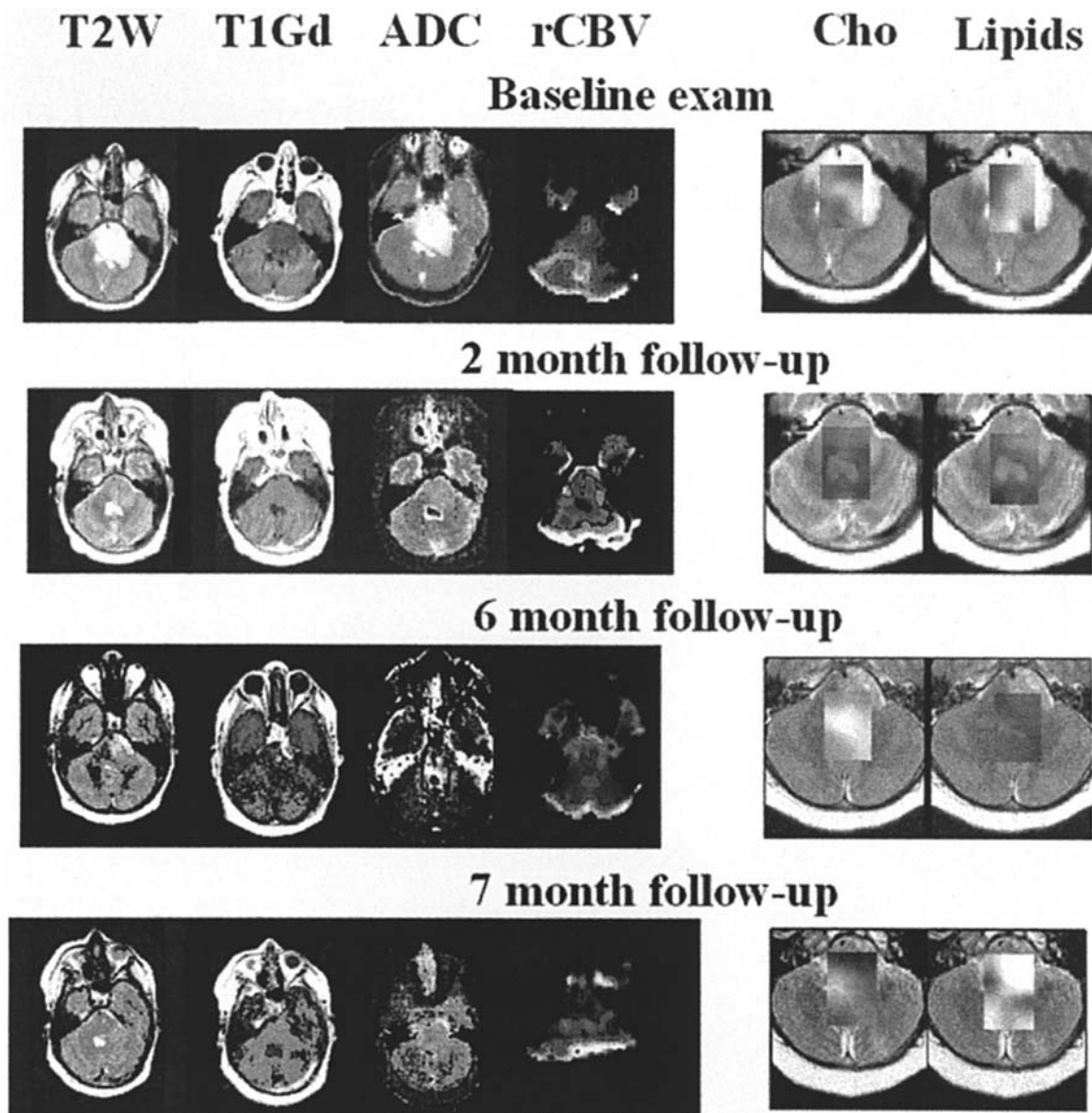E-mail: atzika@partners.org

Figure 1. Conventional T2-weighted MR images (T2W), Gd-enhanced T1-weighted (T1Gd), perfusion MR imaging (rCBV image), diffusion-weighted MR imaging (ADC image) and MR spectroscopic imaging with metabolite mapping, within the posterior fossa of a 7 year-old boy with a brain stem glioma. The baseline MR imaging exam (prior to therapy) showed a large mass in the pons compressing the 4th ventricle of high T2 signal on T2W, with no significant enhancement on T1Gd images, but with high ADC and low perfusion on rCBV. Metabolite images (Cho images) demonstrated increased Cho in the region of the T2 signal abnormality compared to normal surroundings. After 2 months, the resolution of both T2 and ADC hyperintensities was detected and the Cho image showed decreased intensity, while the perfusions remained low on rCBV. Furthermore, the lipid metabolite images demonstrated increased lipids implicating the induction of apoptotic cell death by 7 months. These findings were consistent with stable or regressing disease.

and is unsuitable for tumor classification and interpretation. Also, while MR contrast enhancement can assist in defining tumor borders, it is unreliable in the determination of malignancy (2). We have extended conventional MRI to include hemodynamic MR imaging (HMRI), proton MR spectroscopic imaging (MRSI) and diffusion-weighted MR imaging (DWMRI) to better utilize MRI to study pediatric brain tumors. Furthermore, we have used several data analysis statistical tools, including ANOVA, linear and logistic regressions, correlation analysis, and classification algorithms based on receiver operating characteristic curves.

Proton MR spectroscopy shows the presence and concentration of hydrogen protons attached to different brain compounds. A generated spectrum corresponds to a scale of resonant frequencies versus amplitude (concentration). Molecular compounds identified within cerebral tissue include N-acetyl-aspartate (NAA), a neuronal marker; choline, a cell membrane marker; creatinine, phosphocreatinin and energy metabolites; and lipids and lactate.

Perfusion MR shows the microscopic vascular proliferation, termed neovascularization, associated with tumor growth. Cerebral tissue perfusion can be assessed following a dynamic injection of gadolinium. During its first pass transit through the cerebrovascular bed, which lasts from 5 to 15 sec, gadolinium is restricted to the intravascular space. The restricted presence of highly paramagnetic contrast gadolinium molecules generates microscopic field gradients around the cerebral microvasculature, resulting in shortened T2 relaxation and signal loss. The gadolinium concentration in each pixel
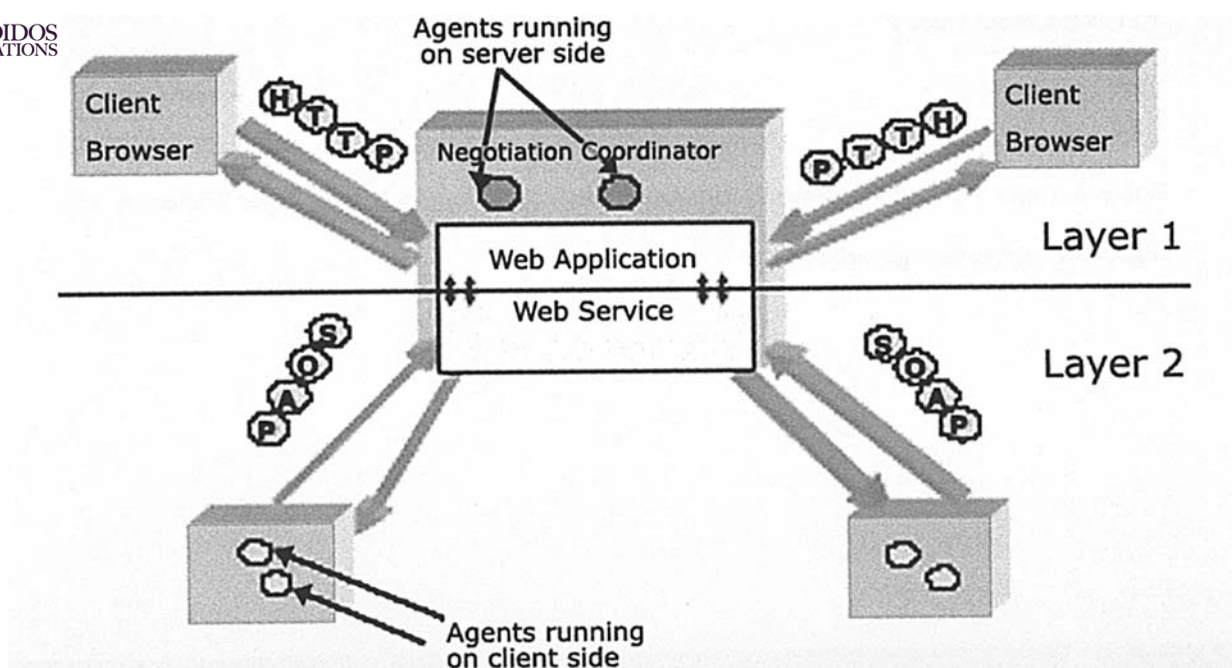
Figure 2. Layer 1 provides web-based negotiation services for human beings, and layer 2 interacts with negotiation agents through web services.

can be calculated from the amount of signal loss, to produce a pixel by pixel relative estimate of blood volume. To this end, maps of cerebral blood volume (CBV) and cerebral blood flow (CBF) can be generated.

DWMRI shows water movement in different tissue areas to produce apparent diffusion coefficient (ADC) maps that describe CNS water molecule mobility. Hyperintense and hypointense ADC areas respectively correspond to free water molecules in cysts, and restricted water molecules in necrotic areas.

Results

We applied multiparametric MRI, including conventional imaging, HMRI, DWMRI, and MRSI, to examine the relationship between MR imaging and MRSI parameters, and correlate these data with histopathology indices and clinical evaluation results. We propose these relationships reflect the functional and biochemical tumor status, as they are influenced by physiological status and correlate with tumor histopathology. Our results show that this multiparametric MR approach: a) predicts histologic findings, b) follows tumor grade serial changes, c) corroborates responses to chemo-therapy and/or radiation, and d) accurately differentiates tumor tissue from radiation necrosis and other structural abnormalities, and normal tissue (3-7). Fig. 1 presents multiparametric data for a 7 year-old patient with a brain stem glioma.

Discussion

The development of a database is essential to improve data collection and management, and achieve greater power and flexibility in detailed data analysis for medical research, such as brain tumor tracking studies. It is further essential that the identification, processing, and retrieval methods be fully automated, especially when considering large amounts of raw data, to minimize human error and uncover parameters that may escape human biased approaches.

Automated retrieval requires the specific tagging of data. In most cases, clinical data headers contain information that is helpful to hospital operational needs, but often irrelevant to research. Clinical data typically offer user-defined fields that can be used as research related tags. For example, fields for disease description or treatment protocol can be useful to automatically identify interesting patients groups. To this end, clinical research departments should adopt research-friendly policies of data entry.

Raw clinical data must often first be processed to be useful. For example, quantification of MRI spectra requires signal processing to create metabolite maps that depict the spatial distribution of MRS detectable brain metabolites. Similarly, raw perfusion data must be processed to calculate various hemodynamic maps (rCBV, rCBF and MTT), and raw diffusion data must be processed to give ADC maps. Unsupervised data processing has become essential, as MRS data analysis has progressed from a single spectrum to 3D grids of spectra. Since processing manipulations cannot be subjectively performed, as they are unknown, robust black-box spectra quantification methods have been developed that use Bayesian estimation or prior knowledge (8). Similar automated algorithms are needed for perfusion imaging when motion correction is involved, and identification of the middle cerebral artery is required to provide the arterial input function (9).

Data analysis aims to answer specific scientific questions, and the full exploitation of all available data is most likely to produce insightful answers. Data organization in a local database can facilitate multiparametric efforts to combine and merge different information types. Such efforts are

limited by difficulties related to data diversity and volume. For example, diversity of the imaging data reflects differences in contrast, resolution, signal-to-noise ratio, geometrical distortions, and movements. Conventional anatomical images have better resolution versus the perfusion or diffusion images and their parametric maps in our data, so the inter- and intra-subject registrations are challenging. Although many robust algorithms exist, including rigid and non-rigid transformations, new image registration protocols continue to be developed (10,11).

The large amounts of data produced by modern medical research instruments pose further difficulties. Although unsupervised data mining techniques can be used to uncover hidden relationships, researcher contribution can reduce the dimensionality of the original dataset. For example, we have used spectroscopic, hemodynamic and diffusion imaging, and less contrast enhanced or T2 imaging to investigate pediatric brain tumors. We chose the rCBV from the many hemodynamic parameters (rCBV, rCBF and MTT), as it better reveals vascularization changes. Furthermore, we chose to assess the highest choline regions in tumors since they exhibit particular importance in tumor monitoring. These examples show how human intervention can be fundamental in medical data design and analysis.

Equally important is that data analysis be clinically oriented. For example, if a new diagnostic index is sought, then the analysis must consider its sensitivity, specificity and accuracy over the entire domain of its values. For instance, where environmental factors are known to increase the incidence of high-grade tumors, a cut-off point with increased sensitivity and decreased specificity is called for, whereas a cut-off point with increased specificity and decreased sensitivity is called for when using a novel therapy that benefits high-grade tumors and significant side effects to low-grade tumors. Also, numerical approaches are preferred in testing the dependence of an index on disease prevalence, or the cost of a false positive, i.e. for a low prevalence condition, high false positive referrals will put an unacceptable burden on the health care system. We have combined spectroscopic indices using a distribution-free approach to meet all of the above requirements (12).

Local hospital databases are often inadequate for research due to limited patient numbers, especially in rare diseases, and limited human/hardware resources preventing robust analyses. These problems can be solved by creating collaborative schemes where knowledge and data are shared between groups and facilities. Indeed, this occurred several years ago for genomics research, which requires the organization and synthesis of great amounts of data. Collaboration allows researchers to pool resources to study open scientific questions and obtain insight that might not otherwise be discovered in isolation.

Although several large disease-oriented research efforts have been instituted in medical research, no universal scheme yet exists to access the resultant data, possibly due to their sensitive nature and extensive non-homogeneity. The Digital Imaging and Communication in Medicine (DICOM) standard is a universal method to transmit medical images and their associated information (13). DICOM has continuously evolved since its inception, and now describes raw images, parametric maps, and medical documents (14). Similar standards are needed in other areas of medical information, such as for histopathology data, to better enable sharing and collaboration.

We propose a flexible front end to medical research databases to increase their utility by translating different local database formats into standard formats and providing access via a single Internet portal to large numbers of different research groups having different levels of access permission. Secure Content Exchange Negotiation System (SCENS) is designed to support the sharing of scientific research by providing incentives specific to a research community (15). SCENS is a trusted third party software infrastructure enabling independent entities to interact and conduct multiple forms of negotiation. With our focus on the flexibility and scalability in design and implementation of SCENS, our implemented SCENS services include user authentication, data quality assessment, web-based negotiation, web service-based negotiation, usage tracking, user evaluation, and other services. We have also developed synthetic tools that simulate negotiation activities to evaluate the whole system.

SCENS has a flexible 3-layer service structure that provides different levels of negotiation services for different users. Layer 1 is a traditional web-based negotiation support system, where users can customize the SCENS negotiation agents via multiple parameters. Layer 2 provides negotiation strategy customization support for users, especially negotiation agents, which are treated as web service consumers through web services. Fig. 2 shows the interrelationship between layers 1 and 2. Layer 3 provides an open automatic negotiation environment. DAML+OIL (16,17), a language with which to create ontologies and mark up information, is used in layer 3 to define a negotiation ontology. This allows agents to acquire knowledge about how to conduct negotiations, and this knowledge includes negotiation protocols and proposals, and conditions. While the negotiation rules are hand-coded into the layer 2 agents, the layer 3 agents can be used in any negotiation activity, given the appropriate ontology.

References

 1. De Backer AI, Mortele KJ and De Keulenaer BL: Picture archiving and communication system - Part one: filmless radiology and distance radiology. J Belge Radiol 87: 234-241, 2004.
 2. Fitz C: Magnetic resonance imaging of pediatric brain tumors. Top Magn Reson Imaging 5: 174-189, 1993.
 3. Tzika AA, Astrakas LG, Zarifi MK, *et al*: Multiparametric MR assessment of pediatric brain tumors. Neuroradiology 45: 1-10, 2003.
 4. Tzika AA, Cheng LL, Goumnerova L, *et al*: Biochemical characterization of pediatric brain tumors by using *in vivo* and *ex vivo* magnetic resonance spectroscopy. J Neurosurg 96: 1023-1031, 2002.
 5. Tzika, AA, Zarifi MK, Goumnerova L, *et al*: Neuroimaging in pediatric brain tumors: Gd-DTPA-enhanced, hemodynamic, and diffusion MR imaging compared with MR spectroscopic imaging. Am J Neuroradiol 23: 322-333, 2002.
 6. Tzika AA, Zurakowski D, Poussaint TY, *et al*: Proton magnetic spectroscopic imaging of the child's brain: the response of tumors to treatment. Neuroradiology 43: 169-177, 2001.
 7. Tzika AA, Astrakas LG, Zarifi MK, *et al*: Spectroscopic and perfusion magnetic resonance imaging predictors of progression in pediatric brain tumors. Cancer 100: 1246-1256, 2004.
 8. Vanhamme L, Sundin T, Hecke T, *et al*: MR spectroscopy quantitation: a review of time-domain methods. NMR Biomed 14: 233-246, 2001.

TJ, Rowley HA and Haughton VM: Automatic calculation arterial input function for cerebral perfusion imaging with MR imaging. Radiology 227: 593-600, 2003.

10. Maintz JB and Viergever MA: A survey of medical image registration. Med Image Anal 2: 1-36, 1998.
11. Crum WR, Hartkens T and Hill DL: Non-rigid image registration: theory and practice. Br J Radiol 77: S140-153, 2004.
12. Astrakas LG, Zurakowski D, Tzika AA, *et al*: Noninvasive magnetic resonance spectroscopic imaging biomarkers to predict the clinical grade of pediatric brain tumors. Clin Cancer Res 10: 8220-8228, 2004.
13. Lynes J and Riha C: Learning the fundamentals of DICOM. Biomed Instrum Technol 38: 35-38, 2004.
14. Hussein R, Engelmann U, Schroeter A, *et al*: DICOM structured reporting: Part 1. Overview and characteristics. Radiographics 24: 891-896, 2004.
15. Ye S, Makedon F, Steinberg T, Shen L, Ford J, Wang Y, Zhao Y and Kapidakis S: SCENS: a system for the mediated sharing of sensitive data. Presented at the Third ACM/IEEE Joint Conference on Digital Libraries, Houston, TX, 2003.
16. Hendler J and McGuinness DL: The DARPA agent markup language. IEEE Intelligent Systems 16: 67-73, 2000.
17. Joint US/EU ad hoc Agent Markup Language Committee (2001). DAML+OIL language specification, http://www.daml.org/2001/03/daml+oil-index.