Applied Intelligence manuscript No.
(will be inserted by the editor)

HealthAgents: Distributed Multi-Agent Brain Tumour Diagnosis and Prognosis

Horacio González–Vélez · Mariola Mier · Margarida Julià–Sapé · Theodoros N. Arvanitis · Juan M. García–Gómez · Montserrat Robles · Paul H. Lewis · Srinandan Dasmahapatra · David Dupplaw · Andrew Peet · Carles Arús · Bernardo Celda · Sabine Van Huffel · Magí Lluch i Ariet

Received: date / Revised version: date

Abstract We present an agent-based distributed decision support system for the diagnosis and prognosis of brain tumours developed by the HealthAgents project. HealthAgents is a European Union funded research project, which aims to enhance the classification of brain tumours using such a decision support system based on intelligent agents to securely connect a network of clinical centres. The HealthAgents system is implementing novel pattern recognition discrimination methods, in order to analyse in vivo Magnetic Resonance Spectroscopy (MRS) and ex vivo/in vitro High Resolution Magic Angle Spinning Nuclear Magnetic Resonance (HR-MAS) and DNA micro-array data. HealthAgents intends not only to apply forefront agent technology to the biomedical field, but also develop the HealthAgents network, a globally distributed information and knowledge repository for brain tumour diagnosis and prognosis.

Keywords Machine Learning; Decision Support Systems; Computational Intelligence; Agents; Pattern Recognition; Medical Ontologies; Medical Informatics; Magnetic Resonance

Correspondence author: Magí Lluch i Ariet, HealthAgents project coordinator, MicroArt S.L. Parc Cientific de Barcelona, Baldiri Reixac, 4-6 Torre D, 08028 Barcelona Spain E-mail: mlluch@microart.cat

Horacio González–Vélez · Mariola Mier
University of Edinburgh, UK

Margarida Julià–Sapé
Universitat Autònoma de Barcelona, Spain

Theodoros N. Arvanitis
University of Birmingham, UK

Juan M. García–Gómez · Montserrat Robles
Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas, Spain

Paul H. Lewis · Srinandan Dasmahapatra · David Dupplaw
University of Southampton, UK

Andrew Peet
University of Birmingham and Birmingham Children's Hospital, UK

Carles Arús
Universitat Autònoma de Barcelona, Spain

Bernardo Celda
Universitat de València and Instituto de Salud Carlos III, Spain

Sabine Van Huffel
Katholieke Universiteit Leuven, Belgium

Magí Lluch i Ariet
MicroArt S.L., Spain

1 Introduction

Brain tumours remain an important cause of morbidity and mortality in Europe with a crude incidence rate of 8 per 100,000 inhabitants [9]. Even though it is not the most common type of cancer overall, brain tumours account for a greater proportion of tumours in younger age groups. This leads to them being an important cause of cancer in young adults and children. Indeed, brain tumours are the most common solid malignancies in children.

While medical treatment relies on the accurate classification of a tumour, diverse studies document the difficulties faced by radiologists and oncologists in making a non-invasive diagnosis based on traditional cranial imaging only [2,13,19]. The inclusion of innovative techniques, such as Magnetic Resonance Spectroscopy (MRS), gives the opportunity to increase the information available from imaging and potentially improve the accuracy of non-invasive diagnosis. Furthermore, there is emerging evidence that these techniques may provide novel biomarkers of prognosis. The use of histopathology to classify tumours is now being augmented by other investigations on tissue, such as molecular genetics and gene expression, to improve the characterisation of tumours and stratify them into groups of varying prognosis.

The metabolite profiling of tissue by High Resolution Magic Angle Spinning Nuclear Magnetic Resonance (HR-MAS) may further improve this characterisation by probing the downstream consequences of these genetic alterations. The use of ex vivo magnetic resonance spectroscopy in the investigation of tumours gives the potential to link these studies to in-vivo MRS and hence the non-invasive determination of tumour tissue properties [42]. Moreover, we argue that if advanced magnetic resonance data can be made widely available along with clinical data, in a secure and easily accessible way, this will significantly improve the ability of clinicians to determine non-invasively the diagnosis and prognosis of brain tumours.

The HealthAgents project [3,47] is engaged in the development of a distributed, agent-based Decision Support System (DSS), which implements a series of automated classifiers based on pattern recognition methodologies for the diagnosis and prognosis of brain tumours.

Our approach builds upon previous experiences in biomedical informatics, particularly in image processing and computer-aided diagnosis, where physiologic and molecular level tumour discrimination are becoming increasingly used for the early detection of tumours [24]; in machine learning for brain tumour classification using MRS [20], where high classification accuracies have been achieved by various methodologies; and in agents, where meaningfully codified descriptions of service capabilities have facilitated the development of protocols for pipelining them in dynamic ways for genome analysis and medical decision support systems [11,38].

This work documents the first prototype of the DSS, which is comprised of an agent-based architecture, with an associated ontology, data mining techniques, and the protocols for clinical data exchange. It is designed to allow users to preserve their local centre policies for sharing information, whilst allowing them to benefit from the use of a distributed data warehouse. Moreover, it will permit the design of local classifiers targeting specific patient populations.

While the DSS provides a clinical environment using MRS, the machine learning techniques will also be applied to ex-vivo chemometrics, micro-arrays and text mining to correlate the transcriptomic and metabolomic information. The use of multiple complementary data sources will enrich the classification of brain tumours and aid the discovery of novel prognostic markers.

All data is stored anonymously and securely through the HealthAgents network of data marts in order to create a distributed data warehouse. This data warehouse contains the collection of such clinical data, that has been properly anonymised from the original clinical data and information acquired and stored at the various participating European clinical centres. This incipient network grants bona-fide access to any qualified organisation in return for its contribution of clinical data to the DSS. No personal patient information leaves the local centres.

The rest of this paper is structured as follows. First, we provide some background on the underlying techniques for this project: determination of tumour properties, machine learning, and agents. Then we provide the architectural specification, followed by a description of the implementation. Finally, we conclude by reviewing some related work and providing guidelines on our future work.

2 Background

Nowadays the diagnosis and treatment of brain tumours is typically based on clinical symptoms, radiological appearance and, often, a histopathological diagnosis of a biopsy. However, treatment response of histologically or radiologically-similar tumours can vary widely, particularly in children. MRS is a non-invasive technique for determining the tissue biochemical composition of a tumour (metabolic profile) [26]. Additionally, the genomic profile, determined using DNA micro-arrays, facilitates the classification of tumour grades and sub-types, which may not be distinguished by morphologic appearance.

HealthAgents builds upon three areas of expertise:

1. Determination of Tumour Properties
2. Machine Learning
3. Agents and Ontologies

2.1 Determination of Tumour Properties

Diagnosis using Magnetic Resonance Imaging (MRI) is non-invasive, but only achieves variable accuracy depending on the tumour type and grade [28]. In addition to its intrinsic healthcare costs and stress to patients, the stereotactic brain biopsy exhibits significant risks, with an estimated morbidity of 2.4-3.5% [14,21] and a death rate of 0.2-0.8% [14,15]. For tumours whose grade may evolve over time, repeated biopsies would be needed to establish the current status and these may not be clinically advisable or practical. Furthermore, tumour histopathology does not reliably predict response to treatment or outcome for all tumours and there is an increasing emphasis on the discovery of novel biomarkers of tumour behaviour [23].

Hence, there is a need to improve brain tumour classification and non-invasive methods for brain tumour diagnosis and prognosis in order to aid patient management and treatment. In the HealthAgents, three techniques are made available to address the aforementioned requirements:

1. MRS, coupled with conventional MRI, provides metabolite profiles of either a single-voxel of tumour tissue or a grid of voxels, where a molecular image of particular tumour metabolites can be additionally produced Magnetic Resonance Spectroscopic Imaging (MRSI) [26,45].
2. HR-MAS is applied to biopsies in vitro in order to provide metabolomic characterisation which is more detailed than that available from in vivo MRS [4, 36].
3. DNA microarray analysis of biopsies can determine tumour phenotype from gene expression profiles and predict survival more accurately than classical histology [39,41].

2.2 Machine Learning

Defined as the study of computer algorithms which improve automatically through experience, machine learning can be thought as the intersection of computer science and statistics [7]. It uses example data or past experience to optimise a given set of performance criteria. As a field of study in computer science it is sometimes referred to as "data mining," "knowledge discovery from databases," or "advanced data analysis" [40], and entails the solution of a series of sub-problems such as: association, supervised learning (e.g. classification and regression), unsupervised learning, or reinforcement learning [1]. Hence, pattern recognition is often described as a sub-domain of machine learning since its main focus is on supervised and unsupervised learning. Brain tumour research provides several biological domains where both pattern recognition and machine learning techniques can be applied: chemometrics [45], metabolomics, microarrays, genomics, proteomics, and text mining [30].

HealthAgents employs machine learning methods to provide the mathematical and computational mechanisms to infer knowledge in a formal model from specific brain tumour data. HealthAgents samples brain tumour data from a training set (x_i,y_i), where x_i is an input pattern —a metabolic profile— and y_i indicates the class membership —a known pre-diagnosed brain tumour—, with the goal of learning general models from the particular samples. Such models will minimise classification error on future unseen data and, eventually, suggest a brain tumour diagnosis more accurately. In order to address the solution of such classification problems, HealthAgents is developing linear and non-linear classifiers for brain tumours employing Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and Least-Squares Support Vector Machines (LS-SVM) in combination with feature selection and feature extraction methodologies.

LDA maximises the ratio between the difference of the projected means and the dispersion within the classes. Ideally, this function should be optimum when the distance between means is maximum and the inside-class dispersions are minimum. SVM are classification, nonlinear function estimation and density estimation methodologies defined in the context of statistical learning theory, kernel methods and structural risk minimisation [50]. While SVM define the optimal separating hyperplane between two classes with the maximal margin in a high dimensional space by means of the kernel trick, LS-SVM provide a reformulation of the SVM, where a linear system is solved [44].

2.3 Agents and Ontologies

Several modern complex distributed systems are composed of customisable building blocks, known as software agents or, simply, agents. The literature enumerates four important characteristics of agents [10]. First, agents possess an internal knowledge-based state that can be dynamically altered. Second, they have dynamic reasoning capabilities that determine their internal behaviour through constraints or goals. Third, they sustain a communication status that enables them to interact with agents or human entities. Last, they feature a unique identity that provides roaming and service advertising capabilities.

Software agent technology offers an increasingly popular paradigm for the design and development of certain types of software system. This is particularly the case for complex distributed systems, in which components need to communicate and reason about the information they exchange. Other approaches, such as those based on web services, offer similar solutions but the clean and high level software abstraction inherent in the agent approach makes agent technology an appropriate choice for this application.

The HealthAgents scenario is one in which distributed datamarts are being built by widespread hospital groups in various countries throughout Europe. Individual hospitals do not typically encounter sufficient cases of particular tumour types to be able to constitute a sizeable training set to develop robust software-based tumour classifiers capable of providing reliable diagnoses and prognoses when presented with non invasive imaging and spectroscopic data. A key aim of the HealthAgents system is that, through data sharing between hospitals across Europe, more powerful diagnostic and prognostic support can be facilitated. Not only will it be possible to build local classifiers based solely on a hospital's own cases, but also global classifiers based on aggregated appropriate cases in the distributed system. Hospitals and countries vary in their approach to restricting the mobility of data and the system design anticipates this variability. To build global classifiers, classifier builder agents will typically gather appropriate cases from across the network, but will be able to work locally within a hospital node if the hospital restricts the movement of data.

Another relatively new set of technologies, on which we draw, is the set called "semantic web technologies", in which ontologies are used firstly to structure the knowledge implicit in the data of the application, secondly as a vehicle for interoperability between software components such as agents and finally to provide a platform for reasoning over that knowledge [25].

Although there are moves towards standardisation, different hospitals often use different schema for their tumour case data and, in order to support interoperability between the data from different hospitals and between agents utilising that data, we have developed several ontologies in a modular fashion. These cover the brain tumour domain and include the relevant medical imaging modalities, clinical information and histopathological classes involved in tumour diagnosis and prognosis. We are including in the ontology relevant knowledge from medical experts, such as any established relationships between anatomical location and tumour type and between clinical data and tumour type. Using this knowledge from the ontology and information from Yellow Pages (YP) agents about classifier agents available in the system, their functionality, performance characteristics and reputation, agents will be able to reason about and recommend appropriate classifiers to be used for a particular case.

In addition to the domain ontology, which describes what sorts of objects are referred to by components of the system, we have developed a separate ontology, which defines the terms to be included in the communication language used by the agents. This means the messaging vocabulary used by agents can be expanded without modification to the individual agents in the system.

The use of agent technology and ontologies is not new. Several authors have described systems in the medical and bio-informatics domains and elsewhere [12,22, 31,32,33,34,38,51]. However, there are several novel aspects to the approach taken here, including the use of the communication language ontology, the implementation of agent functionality through the use of a Lightweight Coordination Calculus (LCC) [43], innovations we are making to handle classifier agent performance and reputation ranking and the integration of an evidence-based search service. Some of these issues are discussed more fully in the following sections.

3 Architectural Specification

Before describing the architectural specification it is instructive to consider a simplified high level overview of the functionality to be achieved.

First, to begin the process, hospitals need patient cases for which the tumour diagnosis is known from biopsy analysis (histopathology, etc) and for which potentially predictive MRI and MRS data is available. These cases are link anonomysed and copied to the hospital's local HealthAgents datamart.

The MRS data is typically in a format dependent on the MRS machine manufacturer and is first preprocessed to a canonical form. At the request of a medical user, and when sufficient cases are available within the datamart, classifiers are developed to answer specific diagnostic questions. Once trained and tested using the appropriate cases from the distributed datamarts, the classifier is added to the system and its existence, its initial performance and reputation, and the profile of its training and test data are published in the HealthAgents YP. The ontologies for the system encompass these descriptive labels.

A medical user, attempting to diagnose a patient for whom MRS data is available, uses a local web based Graphical User Interface (GUI) to initiate entry of the case information into HealthAgents, once again in link anonymised and canonical form. The system, via the GUI, may be able to suggest appropriate classifiers based on the clinical data, tumour location etc or the user may ask, via the GUI, whether appropriate classifiers are available. The GUI consults nearby YP to establish the availability of appropriate classifiers. This is not a straightforward process. Classifiers may be appropriate on the basis of the tumour types between which they can discriminate but may be less obviously suitable when comparing the patient profile of the case to be classified with the profile of the training set used to build a particular classifier.

When performance and reputation of a classifier are taken into account the problem of classifier selection may become a substantial reasoning and negotiation process and several classifiers may be capable of satisfying the request of the user.

In HealthAgents all potentially suitable classifiers are invoked to classify the current case and the various factors influencing classifier choice are used to rank the results unless the user makes a specific overriding choice. The classifiers may be located at classifier nodes anywhere on the HealthAgents network, in which case the data to be classified may be sent from the hospital to the remote classifier nodes for classification. If the hospital does not allow data to leave the local node, classifiers may be run locally. Results from the different classifiers are gathered, ranked and returned to the user via the GUI to support the user's decision making processes. Classifier results are also recorded in the system so that, if and when a confirmed diagnosis is available for a case, an estimate of the "dynamic" performance and reputation of classifiers can be updated.

In addition to the classification processes described above, the HealthAgents system provides an Evidence-based Search Service (EbSS) which seeks, in a context sensitive way, papers from the medical literature to assist the medical user in the current task. The search service has a manual mode in which the users indicate the topics for which supporting material is required but an automatic search mode may also be triggered by the classification processes being undertaken and the resulting literature made available to the user if desired.

This simplified overview of functionality suggests the need for at least the following agents:

– Database agents to handle input and output of cases to and from the hospital datamarts
– Preprocessing agents to convert imaging data to canonical form
– GUI interface agents to handle interaction with medical users at hospital nodes
– YP agents to keep track of resources in the system including the location of case data, classifiers and their profiles, performance and reputations.
– Classifier builder agents to (help to) gather appropriate cases and build, train and test classifier agents
– Classifier agents to provide tumour classifications based on case data
– Petitioner agents to invoke appropriate classifiers and gather and rank results
– EbSS agents to provide the context sensitive information searching

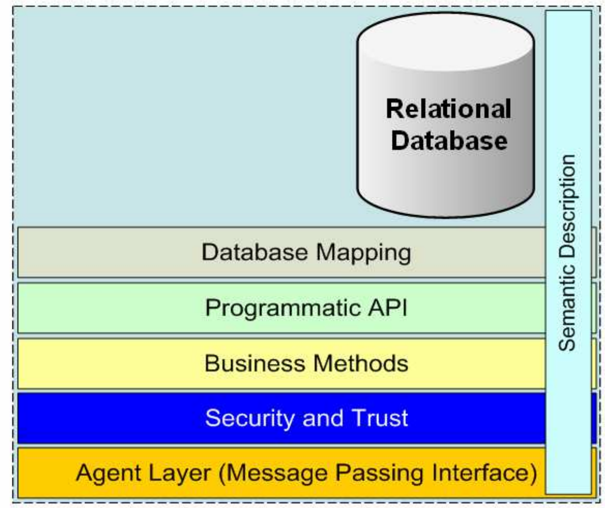


Fig. 1 The HealthAgents Multi-layer Framework

A multi-layer agent framework has been built in order to provide seamless integration of the new functionalities into HealthAgents with minimum programming effort, as well as to support information extraction and analysis in a timely fashion. By deliberately abstracting all specific agent functionality from the interface, this framework enables platform independence. The framework, as depicted in Figure 1, is composed of the following layers:

Database Mapping The database-mapping layer is used to map relational database schemas to the HealthAgents ontological schema.

Application Programming Interface (API) The programmatic API layer abstracts the underlying database interaction from the agent architecture.

Business Methods The business methods layer contributes the main functionality of the agent such as new case classification, data retrieval from a datamart etc.

Security and Trust The security and trust layer is a crucial system component due to the sensitivity of the data. Its functionalities are access control, data marshalling, tracking of on-going data, and the evaluation of reputation and trust of agents.

Agent The agent layer is in charge of all the communications and allows their abstraction from the rest of the system to allow flexibility in the underlying framework. Thus, we can use any agent development platform by modifying this layer only.

Semantic The semantic description contains the description of what the agent holds and what it is able to do.

Listing 1 The Interaction Model for the YP Agent

```
// Here YPID is a yellowpages identifier
// and AID is an agent identifier

a(yellowpages,YPID) ::
(
 (
  // Check if someone is registering with us
  registerRequest(Abilities) <= a(registrant,AID)
  <- register(AID,Abilities) // then
  )
  or
 (
  // Check if someone is searching us
  searchRequest(Abilities) <= a(searcher,AID)
  <- search(Abilities,Results) then
  searchResponse(Results) => a(searcher,AID)
  )
 // ...
)
```

The API at the agent layer consists of the basic messaging interface that queues incoming messages and currently takes them off the queue one-by-one to process them. The messages are automatically tagged with conversation identifiers to relate outgoing messages with their responses. What constitutes a conflicting message very much depends on the agent's functionality and such situations are not explicitly handled in the messaging interface.

That said, formal agent messaging definitions can be used to specify precisely what messages an agent should be expecting in the course of its execution. By providing an executable workflow definition we can simply invoke a workflow and the agent will behave in a determined way, allowing the agent's behaviour to be easily altered or updated by those with the necessary authorisation. Listing 1 shows part of the interaction model for the YP agent, encoded in LCC:

Communications within the HealthAgents network are governed by two complementary ontologies:

1. The *communication ontology* defines an agent language, the HealthAgents Language (HAL), containing message primitives that support the HealthAgents architecture; for example, there are definitions for registration and deregistration messages received by YP agents that specify what data is required in that message. This language has been defined using the Protégé ontology editor [17] as a Web Ontology Language (OWL) [37] ontology. In the agents, a Turtle [5] representation is used for conciseness.
   The ontology has been mapped to Foundation of Intelligent Physical Agents (FIPA) performatives [27] should the underlying agent layer support such messages.

Listing 2 YP Registration for a Classifier Agent

```
@prefix hal: <http://www.healthagents.net/
    HAAgentCommunicationLanguage.owl#>.

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-
    syntax-ns#>.

hal:messageContent
   rdf:type hal:YellowPages_Register_Request;
   hal:has-agent-to-register hal:object1;
   hal:has-abilities hal:object2.

hal:object2
   hal:has-class-name "net.healthagents.agent.
       RDFCollection";
   hal:has-collection-item hal:object1455484972;
   hal:has-collection-item hal:object1638383633.

hal:object1638383633
   hal:has-ability hal:has-name;
   hal:has-class-name "net.healthagents.agent.
       SpecificAgentAbility";
   hal:has-ability-specification "5
       _agmmas_mrs_lese_lda_001".

hal:object1455484972
   hal:has-class-name "net.healthagents.agent.
       SpecificAgentAbility";
   hal:has-ability-specification hal:Classifier;
   hal:has-ability hal:has-type.

hal:object1
   hal:has-class-name "net.healthagents.agent.
       jade.JadeAgentIdentifier";
   hal:has-jade-agent-platform-address <http://
       pasiphae:1633/acc>;
   hal:has-jade-agent-id-name "Classifier@192
       .168.2.11:1099/JADE";
   hal:has-jade-agent-platform-address <http://
       pasiphae:7778/acc>;
   hal:has-jade-agent-platform-address <http://
       pasiphae:1632/acc>.
```

2. The *domain ontology* defines concepts and relations relating to brain tumour diagnosis. The ontology is used to facilitate interoperability between agents and disparate data resources, and also to provide support for agent based learning and reasoning processes.

Listing 2 shows an example of the use of HAL for the process of YP Registration for a classifier agent.

In summary, whilst focusing on a specific knowledge domain –brain tumour diagnosis and prognosis–, HealthAgents is creating a generic intelligent agent communication architecture to securely connect user sites with a distributed database and provide appropriate support for applications built thereon.

Moreover, the architecture specification is intended to support the building of a completely distributed repository of local databases. An overview of the data flow is shown in Fig 2.

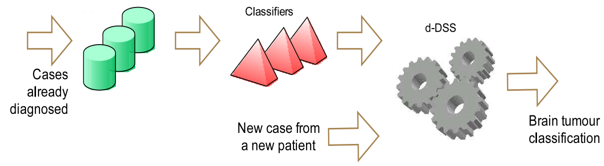


Fig. 2 Overview of the Data Flow in HealthAgents

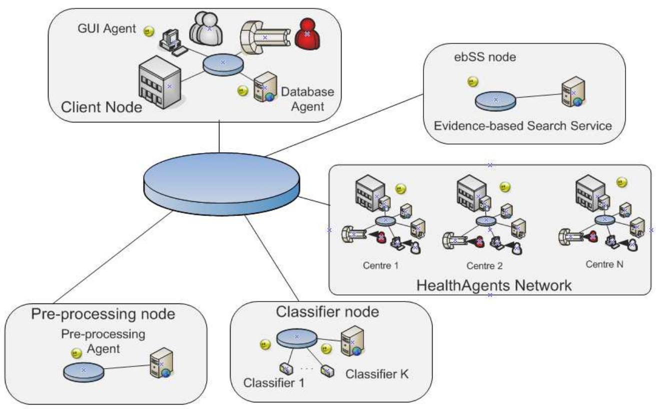


Fig. 3 The Multi-Node HealthAgents Architectural Implementation

4 Implementation

Conceived as an open-source platform, the HealthAgents DSS is implemented using the Jade agent development environment [6], Java, Ant and D2RQ [8], and supported under Windows and Linux platforms, and intended to be distributed into four different types of computing nodes with at least one active agent, as depicted in Figure 3.

Pre-processing node. Involves not only the conversion of time-domain MRS data into frequency-domain data but also the increase of its signal/noise ratio. It requires the application of both a Lorentzian apodisation and a Fast Fourier Transform on the metabolic profile.

Classifier node. Implements classification functions and data projection based on the LDA latent space, implemented as classification agents in the HealthAgents network. These agents provide not only support to the decision-making process during the diagnosis of new patients, but also seamless access to the results at the GUI to the classification model.

Database node. Includes an ontological mapping between the relational database and the HealthAgents ontology. By designing the agent system to utilise semantic web querying mechanisms via D2RQ, we assure the maximum flexibility for integration of different functionality as the network gets larger, as well as providing the ability to run more advanced reasoning over the data.

EbSS node. Provides contextualised searches, classification oriented searches, and generates an on-line literature search.

Client node. Furnishes the HealthAgents GUI to upload the raw data and visualise the result of the classification. This agent manages the user interaction with the MRS raw data, and is crucial to assure patient confidentiality through its anonimisation capabilities.

Currently, the HealthAgents DSS furnishes classifiers for aggressive tumours (glioblastomas and metastasis), benign meningiomas and a low-glial mixture (astrocytomas grade II, oligodendrogliomas and oligoastrocytoma), and its functionality is primarily based on the Interpret DSS system [49].

The system will also ensure that new versions of the classifiers and their models are made available. These updates are based on any newly validated data entering the system, used to adapt and improve the behaviour of the classifiers with respect to the constantly changing evidence in the field. In addition, updates incorporate any feedback from clinical users of the system. This type of feedback is considered the most useful information for the improvement of the DSS. A reputation subsystem using contextual evidence such as user choice of classifiers, clinician feedback and background evaluations of the classifiers will eventually provide quality information and statistics on the classifiers.

4.1 DSS Operation

Firstly, in the HealthAgents DSS operation to upload the MRS raw data to the system, is in situ anonymisation employing the HealthAgents GUI. This is a critical process, because HealthAgents ensures that no personal patient information leaves a clinical centre. This applies to both clinical data records and data files such as MRI and MRS signals (raw data). Patient identifiable information is removed from these data files within the collecting hospital by a process of anonymisation. This user interaction is illustrated in Figure 4 (a).

Secondly, once the data is completely anonymised, the MRS raw data is sent from the client to the preprocessing agent. The pre-processing agent transforms the raw MRS data file from the scanner into the HealthAgents data format, invokes the classifiers and sends the results back to the client.

The classification of a case is done in a specialised node (or nodes) where the trained classifiers reside. This classification is undoubtedly the cornerstone in the HealthAgents functionality as the basic database agent manages the interactions with the HealthAgents network which groups the different clinical centres (data marts) where classifiers are trained. This is depicted in Figure 4 (b). It is important to note that the accuracy of the classifiers depends heavily on the number of cases and therefore on the size of the HealthAgents network. The prototype currently contains a few hundred cases, and new cases will be acquired and incorporated from clinical centres in Spain and the United Kingdom within the following months.

Thirdly, the GUI agent collects the result of the preprocessing module in order to plot the spectra, the MRI (if any) and the classification of the current case in the latent space as shown in Figure 4 (c). While the visualisation is essentially based on previous experiences [45], the agent and web-based capabilities will enable the GUI a seamless operation across networks.

Fourthly, the suggested case classification along with the MRI raw data is presented as local visualisation queries, since personal data never goes out of a clinical centre in order to meticulously preserve the patient identity.

The overall functioning is presented in Figure 5.

5 Evaluation

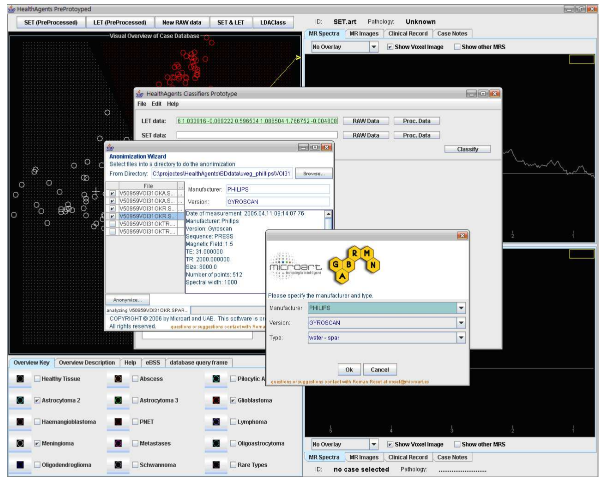
The goal of this initial empirical evaluation is twofold: to evaluate the distributed agent infrastructure and to obtain the estimation of the true accuracy (or true error) of an inferred classifier by applying it to real data.

In order to evaluate HealthAgents and its aforementioned functionalities, we have been deployed the system using the following software:

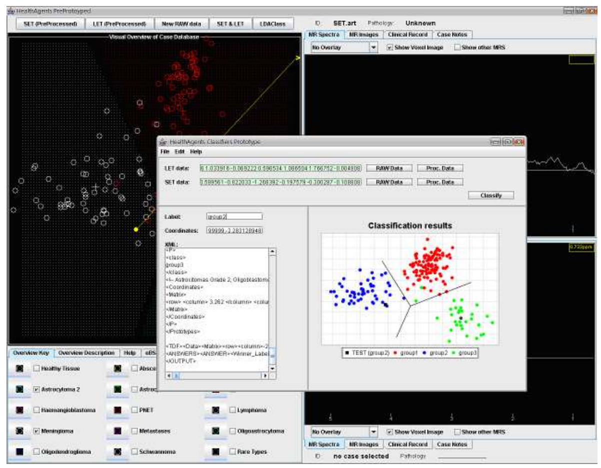
- Java 1.4.2
  - Java 2 Runtime Environment, Standard Edition (build 1.4.2_06-b03)
  - Java HotSpot Client VM (build 1.4.2_06-b03, mixed mode)
- Ant 1.7.0
- Jade 3.4
- D2RQ 0.5

The agent architecture has been deployed into the following nodes:

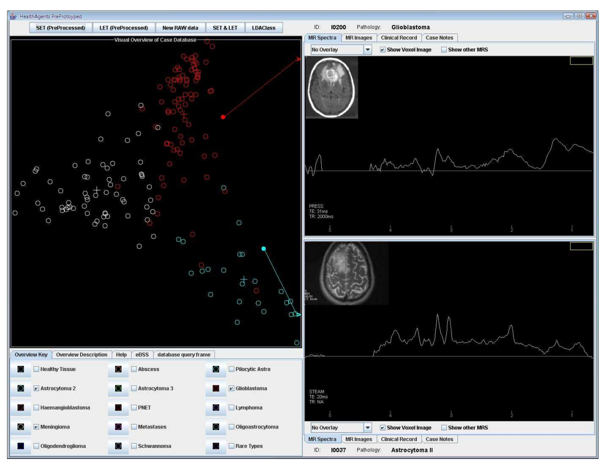
Pre-processing node
Server Dell SC1425
Xeon 3.2Ghz/2MB 300 Mhz FSB processor
1GB Single Rank DDR2 Memory (2x512MB)
OS: Red Hat Enterprise Linux, x86_64 GNU/Linux
Classifier node
Server Dell PowerEdge 1850



(a) Anonymisation



(b) Classification



(c) Visualisation

Fig. 4 The HealthAgents Graphical User Interface (GUI). (a) Anonymisation process of MRS data. (b) Classification. The classification results and their communication messages are combined into a single view. (c) Visualisation. The case classification within the latent space is presented in the upper left part of the screen. Different types of tumours are presented as tick-in boxes in the bottom left portion of the screen. The MRS is presented on the right hand panel along with the MRI of the case. General informative fields on the cases are enlisted in the bottom and top left lines.

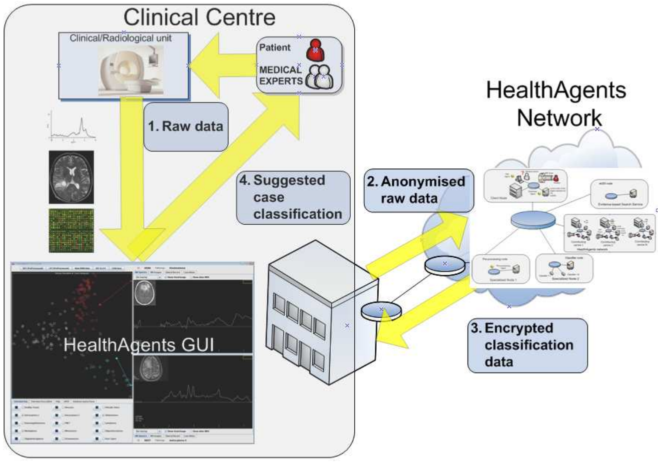


Fig. 5 The Operation of the HealthAgents DSS (functional view)

Listing 3 Log for the HealthAgents Platform Initialisation (28-may-2007)

```
[java] 28-may-2007 12:46:32 jade.core.Runtime
    beginContainer
[java] INFO: ----------------------------------------
[java]     This is JADE3.4 - revision 5874 of
    2006/03/09 14:13:11
[java]     downloaded in Open Source, under LGPL
    restrictions,
[java]     at http://jade.tilab.com/
[java] ----------------------------------------------
[java] 28-may-2007 12:46:33 jade.core.BaseService
    init
[java] INFO: Service jade.core.management.
    AgentManagement initialized
[java] 28-may-2007 12:46:33 jade.core.BaseService
    init
[java] INFO: Service jade.core.messaging.
    Messaging initialized
[java] 28-may-2007 12:46:33 jade.core.BaseService
    init
[java] INFO: Service jade.core.mobility.
    AgentMobility initialized
[java] 28-may-2007 12:46:33 jade.core.BaseService
    init
[java] INFO: Service jade.core.event.Notification
    initialized
[java] 28-may-2007 12:46:33 jade.core.messaging.
    MessagingService boot
[java] INFO: MTP addresses:
[java] http://devel:7778/acc
[java] 28-may-2007 12:46:33 jade.core.
    AgentContainerImpl joinPlatform
[java] INFO: ----------------------------------------
[java] Agent container Main-Container@JADE-IMTP:
    //devel is ready.
[java] ----------------------------------------------
[java] - Welcome to HealthAgents
[java] - Initial Action called.
[java] - Register with Directory Facilitator ...
[java] - Register with Directory Facilitator ...
[java] - Messaging Service Initialised with agent
   rpHAJMSA
[java] - Agent set to net.healthagents.agent.jade
    .JadeMessagingServiceAgent@7c138c63
```

Superclass	*Classes*	*Nsamples*	*Total samples*
gmme	*gm*	74	102
	me	28	
mm	*mm*	51	51
a2odoa	*a2*	20	29
	od	5	
	oa	4	
TOTAL		182	182

Table 1 Brain Tumour Sampling

2x Xeon 3.2Ghz/2MB 300 Mhz FSB processors
2GB Single Rank DDR2 Memory (2x1GB)
OS: Red Hat Enterprise Linux, x86_64 GNU/Linux

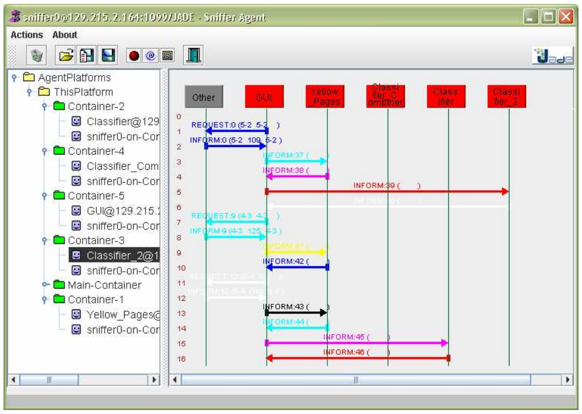
GUI Agent
Workstation, Dell Latitude D610
Intel pentium M 2Ghz processor
1GB Single Rank memory
OS: Microsoft Windows XP Professional [v 5.1]

From a systems infrastructure standpoint, Figure 6 illustrates the system monitoring in HealthAgents. Figure 6 (a) presents a time chart with the message sequence for the connection between the GUI agent and two classifier agents, using a series of service requests through YP agent. Figure 6 (b) shows the monitoring of the system using the HealthAgents Process Manager. Finally, the Listing 3 presents the HealthAgents initialisation log for the overall platform.

As far as the classification is concerned, we have deployed a LDA classifier to perform a high-level discrimination comprising three tumour superclasses: *gmme* containing the glioblastoma multiforme (*gm*) and metastasis (*me*) aggressive tumour classes; *mm* for meningiomas; and *a2odoa* comprising a low-glial mixture of astrocytomas grade II (*a*2), oligodendrogliomas (*od*), and oligoastrocytoma (*oa*).

We have employed single-voxel MRS data on the Interpret database [29], executed on a single node instance, to perform descriptive discrimination of the aforementioned three tumour types. We have employed a discriminative model adjusted using terms from Short Time Echo (STE) and Long Time Echo (LTE) MRS data, and the terms in the three types were matched to single spectral points in the $[0.5..4.1]ppm$ range. A stepwise procedure based on the leaving-one-out evaluation of an LDA classifier was used to obtain the subset of points more discriminant for the multi-class task. A summary of the sampling is presented in Table 1 with further details described in [48].

We have observed that the combined model, LTE and STE, has obtained a good accuracy ($> 90\%$) in the leaving-one-out evaluation, and a marginal improvement compared with models based on STE or LTE alone.



(a) GUI - Classifier Connection

(b) The HealthAgents Process Manager

Fig. 6 System Monitoring in HealthAgents. (a) Sequence of service requests for the connection between the GUI and the classifier agents using the YP agents (b) The operation of the system can be monitored with the HealthAgents Process Manager

6 Related Work

Machine learning surveys have summarised tumour classification techniques based on pattern recognition and clustering methods [20]. Eight of these studies were applied to brain tumour discrimination from normal tissue and other central nervous system diseases. All of them were based on LDA or artificial neural networks and were applied over relative metabolites and principal component transformations. Furthermore, automatic brain tumour grading and image segmentation techniques, based on computational intelligence techniques, have successfully been applied to different case sets in the past five years [16,18,35].

There are a handful of projects which implement computer-assisted evidence-based brain tumour diagnosis using MRS. The Interpret project produced a centralised decision support system for single centres with classification based on histopathological diagnosis [45]. Interpret has successfully been used to discriminate among low-grade meningiomas, high-grade tumours (glioblastomas and metastases), and low-grade glial tumours. The eTUMOUR project incorporates MRS biochemical profiles from single voxel and metabolic spatial distribution by chemical shift imaging [46].

While the functionality of the first prototype is based on the single-voxel version of Interpret [45], HealthAgents expands the original Interpret capabilities with a distributed multi-centre agent architecture, an in-vivo classification method with negotiation, an additional number of cases located in different centres across Europe, and a web-based user interface.

7 Concluding Remarks

In vivo MRS combined with ex vivo/in vitro HR-MAS and gene expression promises to improve the classification of brain tumours and yield novel biomarkers for prognosis. Considerable amounts of highly complex data are required to build reliable specific tumour classifiers and it is a challenge to collect and manage this data. HealthAgents has started to address this problem by building a distributed system of databases centred on the users and managed by agents. As a result, HealthAgents proposes a unique blend of state-of-the-art technologies to develop novel clinical tools for the diagnosis, management and understanding of brain tumours.

HealthAgents extends the traditional scope of machine learning classification by providing a distributed agent-based approach, which enables the system to be re-trained using aggregated sources while preserving security and patient privacy. Future work will include the application of LS-SVM to improve the combined approach and to characterise its behaviour in pairwise classifications. Indeed, HealthAgents is also developing probabilistic mixture models and hierarchical agglomerative clustering for density estimation of heterogeneous brain tumour types and gene co-expression profiles.

The most promising and ambitious development in machine learning for the project is to provide a retraining system for the classifiers deployed in the network. It is expected to enhance the accuracy of the classifiers; to assist wisely in the compilation of additional biomedical data from affiliated clinical centres; and, above all, to improve the data sets leading to a more comprehensive and accurate tumour discrimination.

We argue that the HealthAgents DSS furnishes a completely new approach to brain tumour diagnosis. Since inferences from local predictions are based on limited amounts of data, they may well conflict with one another. Reasoned argument among intelligent agents, in a multi-agent system, will produce a consensus based on data available from a large range of databases hence improving reliability and accuracy. Additionally, HealthAgents aims to provide new concepts relating to the brain tumour domain, while introducing additional elements relating to analytic techniques, such as MRS, in the context of the project.

HealthAgents intends not only to apply agent technology to the biomedical field in a multi-disciplinary fashion, but also to develop the first distributed repository for brain tumour diagnosis, leading eventually to the formation of a special interest data grid, the HealthAgents network.

In this work we have presented the first release of the HealthAgents decision support system. Although still in development, the experience gained from production of an initial prototype strongly suggests that a system based on distributed intelligent agents can produce an innovative software system to help in the fight against one of the most pernicious diseases of our time: cancer.

List of Acronyms

API Application Programming Interface
DSS Decision Support System
EbSS Evidence-based Search Service
FIPA Foundation of Intelligent Physical Agents
GUI Graphical User Interface
HAL HealthAgents Language
HR-MAS High Resolution Magic Angle Spinning Nuclear Magnetic Resonance
LCC Lightweight Coordination Calculus
LDA Linear Discriminant Analysis
LS-SVM Least-Squares Support Vector Machines
LTE Long Time Echo
MRI Magnetic Resonance Imaging
MRS Magnetic Resonance Spectroscopy
MRSI Magnetic Resonance Spectroscopic Imaging
OWL Web Ontology Language
STE Short Time Echo
SVM Support Vector Machines
YP Yellow Pages

Acknowledgements First and foremost, we profoundly thank the HealthAgents Consortium who are ultimately the people in charge of this research endeavour. Without their help and consideration, this article would certainly not have been possible. Second, we thank Francesc Estanyol, Xavier Rafael Palou and Roman Roset for their crucial contribution in the development of the prototype of HealthAgents, Tiphaine Dalmas for the development of the EbSS, and Jan Luts and Javier Vicente for their comments to the machine learning section. Third, we express our gratitude to the anonymous reviewers who have provided us with feedback to improve the overall quality of the final manuscript.

Access to the source code for the Interpret DSS and GUI and for some preprocessing modules is gratefully acknowledged to the Interpret partners [49].

This research has been carried out under the HealthAgents research grant, funded by the Information Society Technologies priority of the European Union Sixth Framework Programme as an Specific Targeted Research Project with contract no.: IST-2004-27214 (2006–2008).

References

1. Alpaydin, E.: Introduction to Machine Learning. Adaptive Computation and Machine Learning. The MIT Press, Cambridge (2004)
2. Armstrong, T.S., Cohen, M.Z., Weinberg, J., Gilbert, M.R.: Imaging techniques in neuro-oncology. Semin. Oncol. Nurs. **20**(4), 231–239 (2004)
3. Arús, C., Celda, B., Dasmahapatra, S., Dupplaw, D., González-Vélez, H., van Huffel, S., Lewis, P., Lluch i Ariet, M., Mier, M., Peet, A., Robles, M.: On the design of a web-based decision support system for brain tumour diagnosis using distributed agents. In: WI-IAT 2006, pp. 208–211. IEEE, Hong Kong (2006)
4. Barton, S., Howe, F., Tomlins, A., Cudlip, S., Nicholson, J., Bell, B., Griffiths, J.: Comparison of in vivo 1H MRS of human brain tumours with 1H HR-MAS spectroscopy of intact biopsy samples in vitro. Magn. Reson. Mater. Phy. **8**(2), 121–128 (1999)
5. Beckett, D.: Turtle - Terse RDF Triple Language. ILRT University of Bristol (2007). `http://www.ilrt.bris.ac.uk/discovery/2004/01/turtle/` (Last accessed: 13 Feb 2007)
6. Bellifemine, F., Poggi, A., Rimassa, G.: JADE: a FIPA2000 compliant agent development environment. In: AGENTS'01, pp. 216–217. ACM Press, Montreal (2001)
7. Bishop, C.M.: Pattern Recognition and Machine Learning. Information Science and Statistics. Springer-Verlag, New York (2006)
8. Bizer, C., Cyganiak, R., Garbers, J., Maresch, O.: D2RQ-Treating Non-RDF Relational Databases as Virtual RDF Graphs. Freie Universitat Berlin, v0.5 edn. (2006)
9. Bray, F., Sankila, R., Ferlay, J., Parkin, D.M.: Estimates of cancer incidence and mortality in Europe in 1995. Eur. J. Cancer **38**(1), 99–166 (2002)
10. Brugali, D., Sycara., K.: Towards agent oriented application frameworks. ACM Computing Surv. **32**(1), 21–27 (2000)
11. Dasmahapatra, S., Dupplaw, D., Hu, B., Lewis, P.H., Shadbolt, N.: Ontology-mediated distributed decision support for breast cancer. In: AIME 2005, *Lect. Notes Comput. Sc.*, vol. 3581, pp. 221–225. Springer-Verlag, Aberdeen (2005)
12. De Turck, F., Decruyenaere, J., Thysebaert, P., Van Hoecke, S., Volckaert, B., Danneels, C., Colpaert, K., De Moor, G.: Design of a flexible platform for execution of medical decision support agents in the intensive care unit. Comput. Biol. Med. **37**(1), 97–112 (2007)
13. DeAngelis, L.M.: Brain tumors. N. Engl. J. Med. **344**(2), 114–123 (2001)
14. Favre, J., Taha, J.M., Burchiel, K.J.: An analysis of the respective risks of hematoma formation in 361 consecutive morphological and functional stereotactic procedures. Neurosurgery **50**(1), 48–57 (2002)
15. Field, M., Witham, T.F., Flickinger, J.C., Kondziolka, D., Lunsford, L.D.: Comprehensive assessment of hemorrhage risks and outcomes after stereotactic brain biopsy. J. Neurosurg. **94**(4), 545–551 (2001)
16. Fletcher-Heath, L.M., Hall, L.O., Goldgof, D.B., Murtagh, F.R.: Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. Artif. Intell. Med. **21**(1-3), 43–63 (2001)
17. Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubézy, M., Eriksson, H., Noy, N.F., Tu, S.W.: The evolution of Protégé: an environment for knowledge-based systems development. Int. J. Hum.-Comput. Stud. **58**(1), 89–123 (2003)

18. Glotsos, D., Tohka, J., Ravazoula, P., Cavouras, D., Nikiforidis, G.: Automated diagnosis of brain tumours astrocytomas using probabilistic neural network clustering and support vector machines. Int. J. Neural Syst. **15**(1-2), 1–11 (2005)
19. González-Vélez, V., Flores-Rodríguez, T., Flores-Avalos, B., González-Vélez, H.: A statistical brain-mapping system for the evaluation of communication disorders. In: CBMS 1997, pp. 167–172. IEEE, Maribor (1997)
20. Hagberg, G.: From magnetic resonance spectroscopy to classification of tumors. a review of pattern recognition methods. NMR Biomed. **11**(4–5), 148–156 (1998)
21. Hall, W.: The safety and efficacy of stereotactic biopsy for intracranial lesions. Cancer **82**(9), 1749–1755 (1998)
22. Hamdi, M.S.: MASACAD: A multiagent-based approach to information customization. IEEE Intell. Syst. **21**(1), 60–67 (2006)
23. Hanahan, D., Weinberg, R.A.: The hallmarks of cancer. Cell **100**(1), 57–70 (2000)
24. Haque, S., Mital, D., Srinivasan, S.: Advances in biomedical informatics for the management of cancer. Ann. N.Y. Acad. Sci. **980**, 287–297 (2002)
25. Hendler, J.: Agents and the semantic web. IEEE Intell. Syst. **16**(2), 30–37 (2001)
26. Howe, F.A., Opstad, K.S.: 1H MR spectroscopy of brain tumours and masses. NMR Biomed. **16**(3), 123–131 (2003)
27. IEEE Computer Society: The Foundation of Intelligent Physical Agents. web site (2007). `http://www.fipa.org/` (Last accessed 30 May 2007)
28. Julià-Sapé, M., Acosta, D., Majós, C., Moreno-Torres, A., Wesseling, P., Acebes, J.J., Griffiths, J.R., Arús, C.: Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. J. Neurosurg. **105**(1), 6–14 (2006)
29. Julià-Sapé, M., Acosta, D., Mier, M., Arús, C., Watson, D., The INTERPRET consortium: A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. Magn. Reson. Mater. Phy. **19**(1), 22–33 (2006)
30. Larrañaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., Lozano, J.A., Armañanzas, R., Santafé, G., Perez, A., Robles, V.: Machine learning in bioinformatics. Brief Bioinform. **7**(1), 86–112 (2006)
31. Lee, C.S., Jiang, C.C., Hsieh, T.C.: A genetic fuzzy agent using ontology model for meeting scheduling system. Inf. Sci. **176**(9), 1131–1155 (2006)
32. Lee, C.S., Pan, C.Y.: An intelligent fuzzy agent for meeting scheduling decision support system. Fuzzy Sets Syst. **142**(3), 467–488 (2004)
33. Lee, C.S., Wang, M.H.: Ontology-based intelligent healthcare agent and its application to respiratory waveform recognition. Expert Syst. Appl. **33**(3), 606–619 (2007)
34. Luck, M., Merelli, E.: Agents in bioinformatics. Knowl. Eng. Rev. **20**(2), 117–125 (2005)
35. Lukas, L., Devos, A., Suykens, J.A.K., Vanhamme, L., Howe, F.A., Majós, C., Moreno-Torres, A., Graaf, M.V.D., Tate, A.R., Arús, C., Van Huffel, S.: Brain tumor classification based on long echo proton MRS signals. Artif. Intell. Med. **31**(1), 73–89 (2004)
36. Martínez-Bisbal, M.C., Martí-Bonmatí, L., Piquer, J., Revert, A., Ferrer, P., Llácer, J.L., Piotto, M., Assemat, O., Celda, B.: 1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo. NMR Biomed. **17**(4), 191–205 (2004)
37. McGuinness, D.L., van Harmelen, F.: OWL Web Ontology Language overview. Standard W3C Recommendation 10 February 2004, World Wide Web Consortium (W3C) (2004). `http://www.w3.org/TR/owl-features/`(Last accessed 13 January 2007)
38. Merelli, E., Armano, G., Cannata, N., Corradini, F., d´Inverno, M., Doms, A., Lord, P., Martin, A., Milanesi, L., Möller, S., Schroeder, M., Luck, M.: Agents in bioinformatics, computational and systems biology. Brief. Bioinform. **8**(1), 45–59 (2007)
39. Mischel, P., Cloughesy, T., Nelson, S.: DNA-microarray analysis of brain cancer: molecular classification for therapy. Nature Rev. Neuroscience **5**, 782–792 (2004)
40. Mitchell, T.M.: Machine learning and data mining. Commun. ACM **42**(11), 30–36 (1999)
41. Nutt, C.L., Mani, D.R., Betensky, R.A., Tamayo, P., Cairncross, J.G., Ladd, C., Pohl, U., Hartmann, C., McLaughlin, M.E., Batchelor, T.T., Black, P.M., von Deimling, A., Pomeroy, S.L., Golub, T.R., Louis, D.N.: Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. Cancer Res. **63**, 1602–1607 (2003)
42. Peet, A.C., Leach, M.O., Pinkerton, C.R., Price, P., Williams, S.R., Grundy, R.G.: The development of functional imaging in the diagnosis, management and understanding of childhood brain tumours. Pediatr. Blood Cancer **44**(2), 103–113 (2005)
43. Robertson, D.: A lightweight coordination calculus for agent systems. In: DALT 2004, *Lect. Notes Comput. Sc.*, vol. 3476, pp. 183–197. Springer-Verlag, New York (2004)
44. Suykens, J.A.K., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. **9**(3), 293–300 (1999)
45. Tate, A.R., Underwood, J., Acosta, D.M., Julià-Sapé, M., Majós, C., Moreno-Torres, A., Howe, F.A., van der Graaf, M., Lefournier, V., Murphy, M.M., Loosemore, A., Ladroue, C., Wesseling, P., Bosson, J.L., Cabañas, M.E., Simonetti, A.W., Gajewicz, W., Calvar, J., Capdevila, A., Wilkins, P.R., Bell, B.A., Rémy, C., Heerschap, A., Watson, D., Griffiths, J.R., Arús, C.: Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. NMR Biomed. **19**(4), 411–434 (2006)
46. The eTUMOUR Consortium: eTUMOUR. web site (2004-2008). `http://www.etumour.net` (Last accessed: 5 Jan 2007)
47. The HealthAgents Consortium: HealthAgents. web site (2006-2008). `http://www.healthagents.net` (Last accessed: 5 Jan 2007)
48. Tortajada, S., García-Gómez, J.M., Vidal, C., Arús, C., Julià-Sapé, M., Moreno, A., Robles, M.: Improved classification by pattern recognition of brain tumours combining long and short echo time 1H-MR spectra. In: ESMRMB 2006: 23rd Annual Scientific Meeting, *Magn. Reson. Mater. Phy.*, vol. 19, chap. Suppl. 1, pp. 168–169. Springer-Verlag (2006)
49. Universitat Autònoma de Barcelona: INTERPRET project. web site (2000–2002). `http://azizu.uab.es/INTERPRET/` (Last accessed: 5 Jan 2007)
50. Vapnik, V.N.: The Nature of Statistical Learning Theory, second edn. Statistics for Engineering and Information Science. Springer-Verlag, New York (1999)
51. Yan, H., Jiang, Y., Zheng, J., Peng, C., Li, Q.: A multilayer perceptron-based medical decision support system for heart disease diagnosis. Expert Syst. Appl. **30**(2), 272–281 (2006)