

# **Clinical decision support systems for brain tumour diagnosis: classification and evaluation approaches**

JAVIER VICENTE ROBLEDO

EDITORIAL  
UNIVERSITAT POLITÈCNICA DE VALÈNCIA



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

UNIVERSITAT POLITÈCNICA DE VALÈNCIA  
DEPARTAMENTO DE FÍSICA APLICADA  
PROGRAMA DE PROMOCIÓN DEL CONOCIMIENTO

**Clinical Decision Support Systems for Brain  
Tumour Diagnosis: Classification and Evaluation  
Approaches**

DOCTORAL THESIS

**Presented by Javier Vicente Robledo**

Supervised by Dr. Juan Miguel García-Gómez  
and Dr. Montserrat Robles Viejo

**Valencia - Spain  
1st October 2012**



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



This editorial is member of the UNE, which guarantees the diffusion and commercialization of its publications at national and international level..

© Javier Vicente Robledo

First edition, 2012

© of the present edition:

Editorial Universitat Politècnica de València  
[www.editorial.upv.es](http://www.editorial.upv.es)

ISBN: 978-84-8363-962-7 (printed)

Any unauthorized copying, distribution, marketing, editing, and in general any other exploitation, for whatever reason, of this piece of work or any part thereof, is strictly prohibited without the authors' expressed and written permission.

---

## Agradecimientos / Acknowledgement

Quisiera dar mi agradecimiento en primer lugar a la Doctora Montserrat Robles y al Doctor Juan Miguel García-Gómez, quienes han supervisado esta Tesis. Fueron ellos los que me propusieron un lejano día de 2.003 que formara parte de su grupo de investigación y, sin ellos, este trabajo sencillamente no habría sido posible. A Montse por su tesón inagotable y su magia para sacar tiempo de donde parece que no lo hay y a Juanmi por su capacidad de trabajo incansable y el gran entusiasmo que desprende y que me inspira a seguir investigando.

Esta Tesis la comparto también con los compañeros de trabajo que, a base de apoyo y colaboración, se han convertido en buenos amigos: A Salvador Tortajada por el rigor y metodología de sus planteamientos teóricos, a Elies Fuster por sus conocimientos de tratamiento de señales, a Carlos Sáez y Adrián Bresó por su inestimable ayuda en las implementaciones que se describen en este trabajo y a Miguel Esparza por sus malabarismos para hacer que el grupo siga existiendo (y que todos cobremos a fin de mes). A todos ellos por los almuerzos llenos de risas y café malucho de máquina. Un agradecimiento muy especial a Alfredo Navarro y César Vidal, quienes compartieron conmigo los pasos iniciales que sirvieron para que el grupo lo formen hoy personas tan maravillosas como lo son ellos.

Agradezco también a todos los miembros del grupo de investigación IBIME-Itaca por su compañerismo y profesionalidad. A los recién llegados a nuestra línea de investigación: Alfonso Pérez y Javier Juan. A los que pasaron por el grupo mientras este trabajo se gestaba: Patricia Paniagua, Miriam González y Ana Ferrer. A la gente de Sistemas de Información, tanto a los que se quedan: David Moner, José Alberto Maldonado, Luís Marco, Diego Boscá, Vicente Cutanda, Estíbaliz Parceró y Alejandro Mañas, como a los que se fueron: Maria Jesús Portolés, Ernesto Reig y Carlos Angulo. También va mi agradecimiento a la línea de Imagen Médica y la gente que está vinculada a ella: Jose Vicente Manjón, Mariam de la Iglesia, Juanjo Lull, Nuria Lull y Javier García Panach. Sin olvidarme tampoco de José de Paula, que con su buen hacer consigue que todo el grupo de investigación funcione un poquito mejor cada día.

I would also like to specially thank to Doctor Andrew C. Peet and his team from University of Birmingham and Birmingham Children's Hospital NHS Foundation Trust for the unvaluable support given on the pediatric brain tumour research presented in this Thesis. Thank you also to Nigel Davies for his suggestions and Martin Wilson for showing such a patience on solving my doubts about quantitation.

---

My gratitude goes also to Professor Paul Lewis, Professor Srinandan Dasma-hapatra, from the University of Southampton, and Doctor Madalina Croitoru, currently at University Montpellier II, for sharing their expertise, mathematical knowledge, time and friendship during the first design stages of the audit method I present in this Thesis.

This PhD Thesis was partially developed in the context of the European Commission projects eTUMOUR (contract no. FP6-2002-LIFESCIHEALTH 503094) and HEALTHAGENTS (contract no. FP6-2005-IST-027213). Taking part in these projects has been a great opportunity of research in the challenging and thrilling field of brain tumour diagnosis. I feel lucky to have met great researchers and greater people: Prof. Bernardo Celda, Prof. Sabine Van Huffel, Dr. Pieter Wesseling, Prof. Carles Arús, Prof. Lutgarde Buydens, Prof. Geert Postma, Dr. Daniel Monleón, Magí Lluch, Dr. Mariola Mier, Giulia Valsecchi, Dr. Margarida Julià-Sapé, Dr. Ana Paula Candiota, Dr. Franklyn Howe, Dr. Jan Luts and Dr. Anca Croitoru.

None of this research would have been possible without the generosity of patients and relatives who kindly contributed with their biomedical data and information. This PhD Thesis is devoted to all of them. I also thank the following eTUMOUR and INTERPRET partners for providing data, in particular: J. Capellades (IDI-Badalona), C. Majós (IDI-Bellvitge), A. Moreno (Centre Diagnòstic Pedralbes), W. Gajewicz (MUL), J. Calvar (FLENI), A. Capdevila (H Sant Joan de Déu), A. Heerschap (RUNMC), and W. Semmier (DKFZ-Heidelberg).

Mi más cálido agradecimiento a mis padres, José Luís y Marina, y a mi hermana Mónica. Por habérmelo puesto tan fácil. Porque los necesito a mi lado. Porque si me hubiera faltado tan solo una gota de todo el cariño y apoyo que me han dado en cada una de las etapas de esta Tesis, todo habría sido mucho más difícil.

Y mi más sentida dedicatoria a Tatiana, la chica que se coló en mi vida un día que coincidió conmigo en el tranvía de camino al trabajo. Porque no se me ocurre mejor acompañante que ella para el resto de mi viaje.

# Abstract

During the last decades, the availability of great amounts of biomedical information has fostered the development of tools to allow the extraction and inference of knowledge. The increase of biomedical technologies to assist the clinical experts in their decisions has contributed to the incorporation of an evidence-based medicine paradigm centered in the patient.

The contributions of this Thesis are focused in the development of tools to assist in the clinician's workflow decision process in the diagnosis of brain tumours (BTs) using Magnetic Resonance Spectroscopy (MRS).

This Thesis contributes with the development of Pattern Recognition (PR)-based classifiers trained with MRS data and devoted to child and adult patients for tumour type and aggressiveness level assessment. These classifiers take advantage of the biochemical differences found in BT in children and adults in order to provide discrimination.

The development of classification models aimed at the discrimination of the three most prevalent pediatric brain tumour types is one of the major contributions of this Thesis. A common location of these tumours is the cerebellum, where it is difficult to distinguish the tumour type with Magnetic Resonance Imaging alone. Hence, obtaining high accuracy in the discrimination from MRS data of pilocytic astrocytomas, ependymomas and medulloblastomas is crucial to establish a surgical strategy for tumour resection, since each tumour type requires different actions to be taken to obtain good prognosis.

In addition, it is concluded that the combination of single voxel MRS at 1.5T at two different Echo Time (TE), Short-TE and Long-TE, improves the classification of pediatric brain tumours over the use of one TE alone. This finding extends and corroborates similar results achieved with MRS data from adults.

A novel on-line method to audit predictive models using a Bayesian perspective for Decision Support Systems (DSSs) devoted to clinical environment is also presented in this Thesis. This audit method positively affects and improves the clinician's decision workflow in a clinical environment by deciding which is the classifier that best suits each particular case being evaluated and by allowing the detection of possible misbehaviours due to population differences or data shifts in the clinical site. The efficacy of such a method is shown for the problem of diagnosis with a multi-centre database of MRS data of BTs.

This Thesis complements the audit method by contributing with a methodology for prior probability assessment to a set of classifiers. The similarity model, inspired also in the Bayesian approach, allows the DSS to select the most adequate classifier for each test case attending to contextual information, which is information not used in the design of the classifiers but related to the case or its environment.

The results of this Thesis have directly contributed to the eTUMOUR (*Web accessible MR decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data, 2004-2009*), and HEALTHAGENTS (*Agent-based Distributed Decision Support System for Brain Tumour Diagnosis and Prognosis, 2006-2008*) European Union projects of the 6th Framework Programme. As a result to the scientific contributions studied in the Thesis, two translational applications can be emphasized. They consist in the incorporation of practical solutions to improve the clinical decision workflow supplied by CURIAM BT, a clinical DSS for BT diagnosis support: the incorporation of the pediatric classifiers as an effective non-invasive pre-operative tool to define the tumour resection strategy; and the incorporation of the audit method and of the similarity model as tools to select the adequate classifier for each case.

# Resumen

A lo largo de las últimas décadas, la disponibilidad cada vez mayor de grandes cantidades de información biomédica ha potenciado el desarrollo de herramientas que permiten extraer e inferir conocimiento. El aumento de tecnologías biomédicas que asisten a los expertos médicos en sus decisiones ha contribuido a la incorporación de un paradigma de medicina basada en la evidencia centrada en el paciente.

Las contribuciones de esta Tesis se centran en el desarrollo de herramientas que asisten al médico en su proceso de toma de decisiones en el diagnóstico de tumores cerebrales (TC) mediante Espectros de Resonancia Magnética (ERM).

En esta Tesis se contribuye con el desarrollo de clasificadores basados en Reconocimiento de Patrones (RP) entrenados con ERM de pacientes pediátricos y adultos para establecer el tipo y nivel de agresividad del tumor. Estos clasificadores especializados son capaces de aprovechar las diferencias bioquímicas existentes entre los TC infantiles y de adultos para llevar a cabo la discriminación.

Una de las principales contribuciones de esta Tesis consiste en el desarrollo de modelos de clasificación enfocados a discriminar los tres tipos de tumores cerebrales pediátricos más prevalentes. El cerebelo suele ser una localización habitual de estos tumores, resultando muy difícil distinguir el tipo mediante el uso de Imagen de Resonancia Magnética. Por lo tanto, obtener un alto acierto en la discriminación de astrocitomas pilocíticos, ependimomas y meduloblastomas mediante ERM resulta crucial para establecer una estrategia de cirugía, ya que cada tipo de tumor requiere de unas acciones diferentes si se quiere obtener un buen pronóstico del paciente.

Asimismo, se concluye que la combinación de señales de ERM adquiridas en dos



tiempos de eco, tiempo de eco corto y tiempo de eco largo, mejora la clasificación de tumores cerebrales pediátricos frente al hecho de usar únicamente los ERM de un único tiempo de eco. Esta evidencia amplía y corrobora resultados similares publicados para datos ERM de adultos.

También se presenta en esta Tesis un nuevo método ‘on-line’ para auditar modelos predictivos para Sistemas de Ayuda a la Decisión (SADs) médica para entornos clínicos empleando una aproximación Bayesiana. Este método de auditoría interviene positivamente en el proceso de toma de decisiones del médico en un entorno clínico porque decide cuál es el clasificador que mejor se adecua al caso particular que el médico está evaluando y detecta posibles comportamientos no deseados debido a diferencias poblacionales o cambios en los datos del centro. La eficacia de dicho método queda demostrada para el problema de diagnóstico de TC con una base de datos multicéntrica de ERM.

Además se ha desarrollado un modelo de similitud que complementa el anterior método de auditoría contribuyendo con una metodología para establecer la probabilidad a priori en un conjunto de clasificadores. Dicho modelo de similitud, también inspirado en una aproximación Bayesiana, permite que el SAD seleccione el clasificador más adecuado a cada caso a evaluar atendiendo a información contextual, que es información que no se ha usado en el diseño de los clasificadores pero que está relacionada con el caso o su entorno.

Los resultados de esta tesis han contribuido directamente a los proyectos eTUMOUR (*Web accessible MR decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data, 2004-2009*), y HEALTHAGENTS (*Agent-based Distributed Decision Support System for Brain Tumour Diagnosis and Prognosis, 2006-2008*), del 6º Programa Marco de la Unión Europea. Como resultados traslacionales a partir de las contribuciones de esta Tesis, cabe destacar la incorporación de dos soluciones prácticas para mejorar el proceso de toma de decisiones que ofrece CURIAM BT, un SAD médica para apoyo al diagnóstico de TC: La incorporación de clasificadores pediátricos como una herramienta no invasiva y efectiva para definir la estrategia de resección del tumor en el preoperatorio; y la incorporación del método de auditoría y el modelo de similitud como herramientas para seleccionar el clasificador más adecuado para cada caso.

# Resum

Al llarg de les últimes dècades, la disponibilitat cada vegada major de grans quantitats d'informació biomèdica ha potenciat el desenvolupament d'eines que permeten extraure i inferir coneixement. L'augment de tecnologies biomèdiques que assisteixen als experts mèdics en les seues decisions ha contribuït a la incorporació d'un paradigma de medicina basada en l'evidència centrada en el pacient.

Les contribucions d'aquesta Tesi se centren en el desenvolupament d'eines que assisteixen al metge en el seu procés de presa de decisions en el diagnòstic de tumors cerebrals (TC) mitjançant Espectres de Ressonància Magnètica (ERM).

En aquesta Tesi es contribueix amb el desenvolupament de classificadors basats en Reconeixement de Patrons (RP) entrenats amb ERM de pacients pediàtrics i adults per establir el tipus i nivell d'agressivitat del tumor. Aquests classificadors especialitzats són capaços d'aprofitar les diferències bioquímiques existents entre els TC infantils i d'adults per dur a terme la discriminació.

Una de les principals contribucions d'aquesta Tesi consisteix en el desenvolupament de models de classificació enfocats a discriminar els tres tipus de tumors cerebrals pediàtrics més prevalents. El cerebel es la localització habitual d'estos tumors, el que fa difícil distingir el tipus del tumor emprant únicament Imatge de Ressonància Magnètica. Per tant, obtenir un alt encert en la discriminació d'astrocitomes pilocitics, ependimomes i meduloblastomes mitjançant ERM resulta crucial per establir una estratègia de cirurgia, ja que cada tipus de tumor requereix unes accions diferents si es vol obtenir un bon pronòstic del pacient.

Tanmateix, es conclou que la combinació de senyals de ERM adquirides en dos temps d'eco, temps d'eco curt i temps d'eco llarg, millora la classificació de tumors cerebrals pediàtrics enfront del fet d'usar únicament els ERM de un

temps d'eco. Aquesta evidència amplia i corrobora resultats similars publicats per a dades ERM d'adults.

També es presenta en aquesta Tesi un nou mètode 'on-line' per auditar models predictius per a Sistemes d'Ajuda a la Decisió (SADs) mèdica per a entorns clínics emprant una aproximació Bayesiana. Aquest mètode d'auditoria intervé positivament en el procés de presa de decisions del metge en un entorn clínic perquè decideix quin és el classificador que millor s'adequa al cas particular que el metge està avaluant i detecta possibles comportaments no desitjats a causa de diferències poblacionals o canvis en les dades del centre. L'eficàcia d'aquest mètode queda demostrada per al problema de diagnòstic de TC amb una base de dades multicèntrica de ERM.

Amés, s'ha desenvolupat un model de similitud que complementa el mètode d'auditoria, contribuint amb una metodologia per establir la probabilitat a priori en un conjunt de classificadors. El model de similitud, inspirat també en una aproximació Bayesiana, permet que el SAD seleccioni el classificador més adequat a cada cas a avaluar atenent a informació contextual, que és informació que no s'ha usat en el disseny dels classificadors però que està relacionada amb el cas o el seu entorn.

Els resultats d'aquesta Tesi han contribuït directament als projectes eTUMOUR (*Web accessible MR decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data, 2004-2009*), i HEALTHAGENTS (*Agent-based Distributed Decision Support System for Brain Tumour Diagnosis and Prognosis, 2006-2008*), del 6<sup>è</sup> Programa Marc de la Unió Europea. Cal destacar com a resultats traslacionala a partir de les contribucions d'esta Tesis, la incorporació de dues solucions pràctiques per millorar el procés de presa de decisions que ofereix CURIAM BT, un SAD mèdic per a suport al diagnòstic de TC: La incorporació de classificadors pediàtrics com una eina no invasiva i efectiva per definir l'estratègia de resecció del tumor en el preoperatori; i la incorporació del mètode d'auditoria i el model de similitud com a eines per a seleccionar el classificador més adequat per a cada cas.

# Glossary

## Mathematical Notation

The notation used generally follows the standard conventions of mathematics and statistics. If  $X$  is a random variable then  $x$  denotes its value. Bold letters denote vectors, so  $\mathbf{x} = (x_i)$  is a vector with components  $x_i$ ,  $i = 1, \dots, m$  with  $m$  being deduced from the context.

$\mathbf{x}_j$	data vector describing the $j^{\text{th}}$ sample
$t_j$	label associated to the $j^{\text{th}}$ sample
$k_j$	set of values describing the contextual information of the $j^{\text{th}}$ sample
$N$	number of samples in $\mathbf{Z}$
$d_i$	number of free parameters in model $\mathcal{M}_i$
$\mathbf{Z}$	set of $N$ data $\{(k_j, \mathbf{x}_j, t_j)\}_1^N$ . $k_j$ can be null.
$\theta_i$	vector of parameters
$\Omega$	set of possible values of $\theta_i$
$\mathcal{M}_i$	a model
$\mathbf{Z}_i$	set of samples a model $\mathcal{M}_i$ has been trained with
$\mathbf{Z}_{-i}$	set of samples not used in $\mathcal{M}_i$
$T_i$	set of labels of $\mathbf{Z}_i$
$z$	a sample expressed as a tuple $(k, \mathbf{x}, t)$

---

$\{f(\theta_i), \theta_i \in \Omega\}$	functional form of model $\mathcal{M}_i$
$P(\theta_i \mathcal{M}_i)$	‘priori’ distribution of model $\mathcal{M}_i$
$P(\mathbf{Z} \theta_i, \mathcal{M}_i)$	likelihood of model $\mathcal{M}_i$ (information about $\theta_i$ and $\mathcal{M}_i$ derived from the observation of $\mathbf{Z}$ )
$P(\mathcal{M}_i)$	‘subjective’ prior distribution over the model space
$P(\mathbf{Z} \mathcal{M}_i)$	model likelihood or evidence for $\mathbf{Z}$
$P(\mathbf{Z})$	unconditional distribution of $\mathbf{Z}$
$P(\mathcal{M}_i \mathbf{Z})$	posterior distribution of a model $\mathcal{M}_i$
$P(\theta_i \mathbf{Z}, \mathcal{M}_i)$	distribution of $\theta_i$ of a model in the light of the data $\mathbf{Z}$
$\hat{\theta}_i$	mode of the posterior density. Maximum likelihood estimate
$V$	inverse of the Hessian of the $-\log P(\hat{\theta}_i \mathbf{Z})$ $A^{-1}$ , where $A = -\nabla\nabla\log P(\theta \mathbf{Z}) _{\hat{\theta}_i}$
$P(\mathbf{Z} \hat{\theta}_i, \mathcal{M}_i)$	best fit likelihood
$\log P(\mathbf{Z} \mathcal{M}_i)$	Bayesian Information Criterion (BIC) multiplied by $-\frac{1}{2}$
$P(z^{(N+1)} \mathbf{Z}, \mathcal{M}_i)$	predictive distribution for the unknown sample $z^{(N+1)}$
$CT$	Correspondence Table
$m_{sl}(t)$	function that makes reference to the ‘most specific label’ of $t$
$z^*$	sample $z$ labeled with its most specific label $(\mathbf{x}, m_{sl}(t))$
$L$	set of common labels obtained by intersection of all the $m_{sl}(t)$ gathered in each $\mathbf{Z}_i$
$\mathcal{L}(\mathbf{Z}, T_i, CT)$	function which transforms the labels of $\mathbf{Z}$ into one of the labels specified in $T_i$ according to the correspondence table CT

---

$\mathbf{Z}^{T_i}$	result of applying $\mathcal{L}(\mathbf{Z}, T_i, CT)$
$\mathbf{Z}_{DSS}$	dataset introduced in a Decision Support System (DSS) after deployment in a clinical site and composed by a set of samples different to any sample in $\mathbf{Z} = \cup \mathbf{Z}_i$
$z^{(\kappa)}$	current test sample
$k^{(\kappa)}$	contextual information of the current test sample
$w$	set of parameters modeling the contextual information
$P(z^{(\kappa)} \mathbf{Z}_i)$	similarity of a test sample to the training set of the model $\mathcal{M}_i$

## Acronyms and Abbreviations

<b>ACC</b>	Accuracy
<b>Ala</b>	Alanine
<b>AMARES</b>	Advanced Method for Accurate, Robust and Efficient Spectral fitting of MRS data with use of prior knowledge
<b>ANN</b>	Artificial Neural Networks
<b>AQSES</b>	Automated Quantitation of Short Echo time MRS Spectra
<b>Asp</b>	Aspartate
<b>ATRT</b>	Atypical teratoid Rhabdoid Tumour
<b>AUC</b>	Area Under the Curve
<b>BAR</b>	Balanced Accuracy Rate
<b>BF</b>	Bayes Factor
<b>BIC</b>	Bayesian Information Criterion
<b>BT</b>	Brain Tumours
<b>BTloc</b>	Brain Tumour location
<b>CADS</b>	Computer Aided Diagnosis System
<b>CF</b>	Classification Framework
<b>Cho</b>	Choline
<b>CNS</b>	Central Nervous System
<b>Cr</b>	Creatine
<b>CSF</b>	cerebrospinal fluid
<b>CT</b>	Computed Tomography
<b>CURIAM</b>	CURIAM Decision Support System for Clinical Environment
<b>CV</b>	K-fold Cross Validation

---

<b>DASTRO</b>	Diffuse Astrocytoma
<b>dDSS</b>	distributed Decision Support System
<b>DSS</b>	Decision Support System
<b>EPEN</b>	Ependymoma grade II
<b>eTUMOUR</b>	EC project, contract no. FP6-2002-LIFESCIHEALTH 503094
<b>FID</b>	Free induction decay
<b>FP</b>	False Positive
<b>GABA</b>	$\gamma$ Aminobutyric acid
<b>GE</b>	General Electric
<b>Glc</b>	Glucose
<b>Gln</b>	Glutamine
<b>Glu</b>	Glutamate
<b>Glx</b>	Glutamate + Glutamine
<b>Gly</b>	Glycine
<b>GUI</b>	Graphical User Interface
<b>GPC</b>	glycerophosphocholine
<b>Gua</b>	Guanidinoacetate
<b>HEALTHAGENTS</b>	EC project, contract no. FP6-2005-IST 027213
<b>HG</b>	High grade
<b>HLSVD</b>	Hankel-Lanczos Singular Value Decomposition
<b>HR-MAS</b>	High Resolution Magic Angle Spinning Nuclear Magnetic Resonance
<b>IBIME</b>	Informática Biomédica
<b>ICA</b>	Independent Component Analysis
<b>ICT</b>	Information and Communication Technology



<b>i.i.d.</b>	independent and identically distributed
<b>INTERPRET</b>	EC project, contract no. FP5-1999-IST-10310
<b>ITACA</b>	Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas
<b>jMRUI</b>	Java magnetic resonance user interface
<b>KNN</b>	K-Nearest neighbors
<b>kRSTT</b>	k-Random Sampling Train-Test
<b>Lac</b>	Lactate
<b>LDA</b>	Linear Discriminant Analysis
<b>LG</b>	Low grade
<b>LOO</b>	Leave-one-out
<b>LS-SVM</b>	Least-Squares Support Vector Machines
<b>Long-TE</b>	Long Time Echo
<b>MED</b>	Medulloblastoma
<b>MEN</b>	Low-grade meningioma
<b>MET</b>	Metastases
<b>ml</b>	myo-Inositol
<b>ML</b>	Mobile lipids
<b>MLP</b>	Multilayer Perceptron
<b>MMLip09</b>	Macromolecules and lipids components at 0.9 ppm
<b>MMLip13</b>	Macromolecules and lipids components at 1.3 ppm
<b>MMLip20</b>	Macromolecules and lipids components at 2.0 ppm
<b>MR</b>	(Nuclear) Magnetic Resonance
<b>MRI</b>	Magnetic Resonance Imaging

---

<b>MRS</b>	Magnetic Resonance Spectroscopy
<b>MRSI</b>	Magnetic Resonance Spectroscopic Imaging
<b>NAA</b>	N-Acetyl Aspartate
<b>NMF</b>	Non-Negative Matrix Factorization
<b>NMR</b>	Nuclear Magnetic Resonance
<b>NOR</b>	Normal brain spectra
<b>OTH</b>	other less frequent tumour types
<b>PCA</b>	Principal Components Analysis
<b>PCh</b>	Phosphocholine
<b>PCs</b>	principal components
<b>PF</b>	Posterior Fossa
<b>PI</b>	Peak integration
<b>PILOA</b>	Pilocytic Astrocytoma
<b>PINEOB</b>	Pineoblastoma
<b>PNS</b>	Peripheral Nervous System
<b>PNET</b>	primitive neuroectodermal tumour
<b>ppm</b>	parts per million
<b>PR</b>	Pattern Recognition
<b>PRESS</b>	Point-Resolved Spectroscopy
<b>PROBE</b>	Proton Brain Exam
<b>QDA</b>	Quadratic Discriminant Analysis
<b>QUEST</b>	Quantitation based on semi-parametric quantum estimation algorithm
<b>ROC</b>	Receiver operating characteristic curve

<b>SASTRO</b>	Subependymal giant cell astrocytoma
<b>Scyllo</b>	scyllo-inositol
<b>Short-TE</b>	Short Time Echo
<b>SNR</b>	signal-to-noise ratio
<b>STEAM</b>	STimulated Echo Acquisition Mode
<b>SV</b>	Single voxel
<b>SVM</b>	Support Vector Machines
<b>TARQUIN</b>	Totally Automatic Robust Quantitation in NMR
<b>Tau</b>	Taurine
<b>TE</b>	Echo time
<b>TP</b>	True Positive
<b>TR</b>	Recycling time
<b>ULN</b>	Unit length normalization
<b>VOI</b>	Volume of interest
<b>WHO</b>	World Health Organization
<b>XML</b>	Extensible Markup Language

# Contents

<b>Abstract</b>	<b>v</b>
<b>Resumen</b>	<b>vii</b>
<b>Resum</b>	<b>ix</b>
<b>Glossary</b>	<b>xi</b>
<b>Contents</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Hypothesis . . . . .	2
1.3 Goals . . . . .	2
1.4 Contributions . . . . .	3
1.5 Research Projects and Partners . . . . .	4
1.6 Outline of the Thesis . . . . .	5
Bibliography . . . . .	10
<b>2 Theoretical foundations</b>	<b>11</b>
2.1 Brain Tumours . . . . .	11
2.2 Nuclear Magnetic Resonance . . . . .	14
2.3 MRS of the brain from a biochemical perspective . . . . .	14
2.4 Pattern Recognition (PR) . . . . .	27
2.5 PR applied to BT diagnosis using MRS . . . . .	43
2.6 DSSs in Brain Tumours (BT) Research . . . . .	49
Bibliography . . . . .	69

---

<b>3</b>	<b>Brain tumour classification in childhood and adulthood</b>	<b>71</b>
3.1	Introduction . . . . .	72
3.2	Materials . . . . .	73
3.3	Methods . . . . .	76
3.4	Results . . . . .	77
3.5	Discussion . . . . .	83
3.6	Conclusions . . . . .	86
	Bibliography . . . . .	90
<b>4</b>	<b>Automatic classification of pediatric brain tumours</b>	<b>91</b>
4.1	Introduction . . . . .	92
4.2	Material and methods . . . . .	94
4.3	Results . . . . .	97
4.4	Discussion . . . . .	107
4.5	Concluding remarks . . . . .	110
	Bibliography . . . . .	116
<b>5</b>	<b>Audit method suited for DSS in clinical environment</b>	<b>117</b>
5.1	Introduction . . . . .	118
5.2	Methods . . . . .	120
5.3	Evaluation . . . . .	127
5.4	Results . . . . .	129
5.5	Discussion . . . . .	133
5.6	Concluding remarks . . . . .	138
	Bibliography . . . . .	141
<b>6</b>	<b>Similarity model for proper prior assessment</b>	<b>143</b>
6.1	Introduction . . . . .	144
6.2	Methods . . . . .	146
6.3	Results . . . . .	152
6.4	Discussion . . . . .	159
	Bibliography . . . . .	169
<b>7</b>	<b>CURIAM DSS for BT diagnosis support</b>	<b>171</b>
7.1	DSSs for cancer research in clinical environments . . . . .	172
7.2	Inclusion of pediatric classifiers into CURIAM BT . . . . .	173
7.3	Inclusion of audit method into CURIAM BT . . . . .	181
7.4	Discussion . . . . .	185
7.5	Conclusions . . . . .	189
	Bibliography . . . . .	195

---

<b>8</b>	<b>Concluding remarks and future work</b>	<b>197</b>
8.1	Conclusions . . . . .	197
8.2	Future work . . . . .	199
	<b>Appendix A</b>	<b>201</b>
<b>A</b>	<b>Human brain cells, tissues and physiological processes</b>	<b>201</b>
A.1	Brain cells . . . . .	201
A.2	Tissues and regions . . . . .	209
A.3	Physiological processes in brain cells . . . . .	212
	Bibliography . . . . .	218
	<b>List of Figures</b>	<b>224</b>
	<b>List of Tables</b>	<b>224</b>

# Chapter 1

## Introduction

*“A goal is a dream with a deadline.”*

*Napoleon Hill*

### 1.1 Motivation

A trend towards a patient-focused medicine based on the evidence paradigm is becoming a reality. During the last decades, the increase of the biomedical information has fostered the development of tools able to extract and infer knowledge from this information. Brain Tumours (BT) diagnosis with Magnetic Resonance Spectroscopy (MRS) demands such necessities since it is considered a task of high complexity both in volume of information and difficulty of analysis. Complexity in BT diagnosis becomes a major concern when differences in type of tumour and location in the brain are associated to the patient's age. Pattern Recognition (PR) techniques can assist to clinical experts to interpret and analyse the data in the decision workflow of diagnosis. In a clinical Decision Support System (DSS) counting with a varied set of predictive models, the need of a tool that objectively evaluates the behaviour and performance of the predictive models in the clinical environment arises. Furthermore, when the predictive models are specialized according to a patient profile (i.e. age or localization of the lesion), a tool able to select the predictive model/s more suitable to the patient's characteristics it is also desirable.

This Thesis is aimed on two main focuses: the development of predictive models for pediatric brain tumour diagnosis and its incorporation into a clinical DSS; and the development of a tool for PR-based DSS that allows the evaluation of the performance of the predictive models and the selection of the suitable predictive model attending to the patient profile.

The people to whom this Thesis is directed may be interested in the application of Information and Communication Technology (ICT) to support Health Systems and Biomedical Research.

## 1.2 Hypothesis

The line of argument of this Thesis is based on two hypotheses:

- Primary hypothesis: The use of automatic PR techniques applied to MRS data can aid in the diagnosis of the most common pediatric brain tumours.
- Secondary hypothesis: The proposed audit method can provide a framework to determine the most suitable classification model for each test case in the context of a clinical Decision Support System (DSS) with a set of PR-based classifiers for brain tumour classification with MRS data.

## 1.3 Goals

In order to verify these hypothesis, the main goal of this Thesis is the development of computer-assisted support based on the PR discipline for brain tumour diagnosis and suited to clinical environment.

This goal can be subdivided into several goals:

The first goal is to develop predictive models with high accuracy in classification aiming at the diagnosis of pediatric brain tumours with MRS data.

The second goal is to contribute with an evaluation methodology to measure the prediction skills of the classifiers working in a clinical DSS. This goal will be achieved by means of a tool to audit the prediction skills of the classifiers capable to select the most suitable predicting models in a clinical DSS attending to the patient profile.



Finally, the third goal is the implementation and integration of the predictive models and the proposed audit methodology to assess the proper classifier to a patient into a DSS for clinical environment.

## 1.4 Contributions

The scientific results of this Thesis concern the application of Pattern Recognition to brain tumour research. The contributions of this Thesis have been published in scientific journals and proceedings of congresses in the fields of Pattern Recognition, Applied Artificial Intelligence, Medicine, Magnetic Resonance, and Molecular Biology.

The research related to the inference of predictive models for brain tumour diagnosis to establish the aggressiveness in children and adults (Chapter 3) was presented in [1].

Chapter 4 describes the development of classifiers for the three most prevalent tumour types in children. This work has been recently accepted as a journal paper [2] and presented in [3].

The design of an audit model of brain tumour classifiers using a Bayesian approach is described in Chapter 5 and has been presented in [4] and in [5]. These works also described the main aspects regarding the model to establish the more adequate classifier to a given patient profile (Chapter 6).

The implementation of the audit method and the model to assess prior probabilities to the classifiers deployed into a clinical DSS is described in Chapter 7. This Chapter also describes the incorporation of the pediatric classifiers developed in Chapters 3 and 4. Partial sections of Chapter 7 have appeared as scientific contributions in [6, 7, 8, 9, 10, 11].

This Thesis continues the research initiated by Dr. Juan Miguel García-Gómez [12] and follows with the exploration of PR techniques applied to brain tumour diagnosis. Contributions have been made with the preparation of the datasets and the development of PR-based models for a multiproject-multicenter evaluation of brain tumour classifiers based on MRS [13].

Additional contributions of these Thesis have been made to two other lines of research on brain tumour diagnosis led by Elies Fuster-Garcia and Salvador Tortajada regarding the study of compatibility of PR models trained with 1.5T

MRS with samples of 3T MRS and the design of incremental learning algorithms and strategies for brain tumour diagnosis with MRS, respectively. The results of these studies were accepted in journal papers [14, 15]. Collaboration on the preparation of the datasets, the writing and experimental design were carried out in these works.

## 1.5 Research Projects and Partners

Part of this Thesis stem from the research developed related to the EC project, contract no. FP6-2002-LIFESCIHEALTH 503094 (eTUMOUR) and the EC project, contract no. FP6-2005-IST 027213 (HEALTHAGENTS), which took place between 2004 and 2009.

- eTUMOUR (2004-2009) *Web accessible MR decision-support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data* is a European Union Integrated Project of the 6th Framework Programme in the *Molecular imaging for early detection of tumours and monitoring of treatment*. The contributions of work package 3 (*Pattern Recognition*) of the eTUMOUR project, which was led by the IBIME-ITACA group at the Universitat Politècnica de València, were focused on the development of PR-based engines that offered an objective solution to specific tumour discrimination problems using (Nuclear) Magnetic Resonance (MR) in vivo data. The classification engines reported in Chapters 3 and 4 are developed from the multi-centre MRS database acquired in the project. A strong collaboration with pediatric oncologist Dr. Andrew C. Peet and neuropathologist Dr. Pieter Wesseling was established in these studies for the definition of the set of questions to be addressed, for the compilation and validation of the datasets, and for the interpretation of the results under a clinical point of view.
- HEALTHAGENTS (2006-2008) *Agent-based Distributed Decision Support System for Brain Tumour Diagnosis and Prognosis* is another European Union Specific Targeted Research Project of the 6th Framework Programme in the *Information Society of Technologies*. The objective of this project is to create a multi-agent DSS to assist in the early diagnosis of brain tumours and to create a distributed Data Warehouse with the world's largest network of interconnected databases of clinical, histological, and molecular phenotype data of brain tumours. The ranking

model for brain tumour classifiers, a preliminary development of the audit method and prior model assessment described in Chapters 5 and 6 was incorporated in the distributed Decision Support System (dDSS) developed for the classification framework for the HEALTHAGENTS agent-based distributed network. In these studies, Dr. Srinandan Dasmahapatra and Dr. Madalina Croitoru actively contributed to the definition of the mathematical model. A strong collaboration among the clinical partners of the eTUMOUR consortium was established: Neuropathologist Dr. Pieter Wesseling was the clinician in charge of the quality control validation of the dataset used for training the brain tumour classifiers evaluated with the audit model.

## 1.6 Outline of the Thesis

The technical aspects covered in the Thesis include the processing, feature extraction and modeling of MRS brain tumour data; the inference and evaluation of predictive models; the integration of the predicting models into DSSs for the clinical environment; and the follow-up of predictive models within a DSS by means of an audit method which measures and reports their performance. Figure 1.1 illustrates the chapters of the Dissertation and establishes the relations between them.

Chapter 2 introduces several concepts from the fields involved in this Dissertation, e.g., Brain Tumours, Pattern Recognition and Decision Support Systems. It gives an introduction on the biochemical composition observed in *in vivo* Magnetic Resonance Spectroscopy (MRS) data in the brain; it offers a survey of previous research carried out in automatic classification and diagnosis of brain tumour with *in vivo* MR spectroscopy;

Chapter 3 presents some initial results on discrimination and classification of pediatric and adult brain tumours using MRS from a multicenter European database of patients. Biochemical differences in the nature of child and adult BT are analyzed, and classifiers devoted to assess the aggressiveness of BT are developed for children and adults. The incompatibility of using classifiers trained with adults to assess the aggressiveness in children and vice-versa is analyzed. This incompatibility is easily solved with a double-fold predictive model which, according to the age of the patient, decides what predicting model (one trained from pediatric MRS data and other from adult MR spectra) will give a better

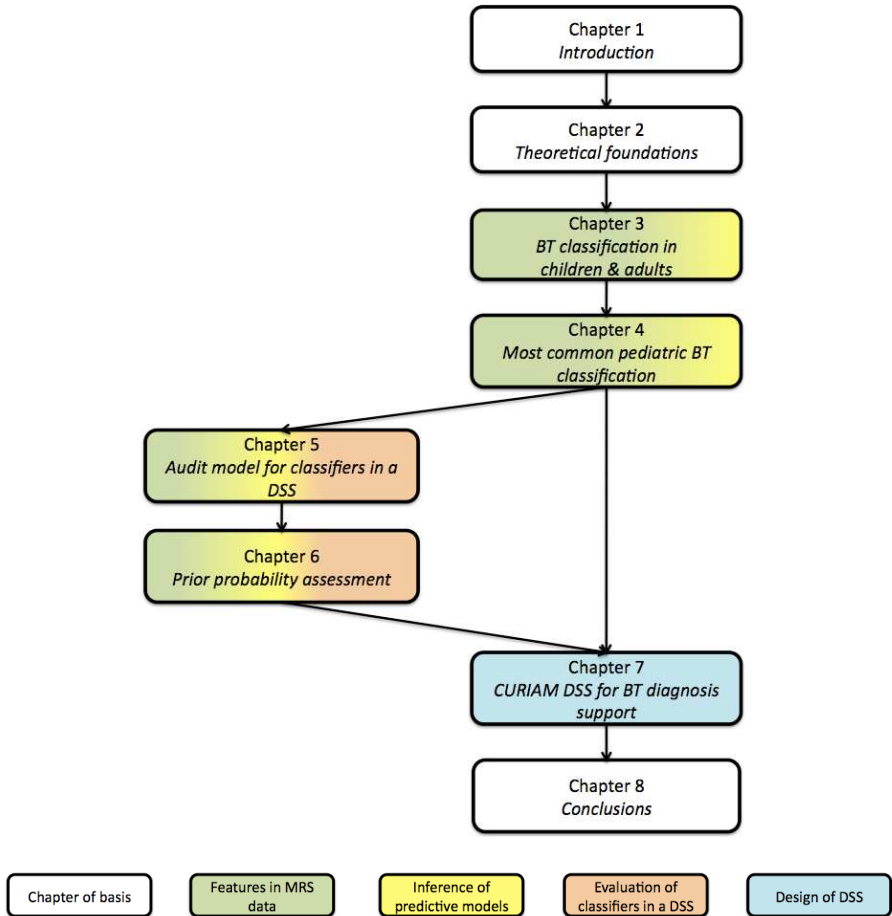


Figure 1.1: Flowchart of the Dissertation.

performance in the task of assessing BT.

The results obtained in Chapter 3 are used to further investigate in two main branches: pediatric BT discrimination (Chapter 4), and the development of an audit method to decide which classifier fits best the answer given to a case under evaluation (Chapters 5 and 6).

Chapter 4 pursues the aim of evaluating the automatic classification of the three most common brain tumour types in children with a large multi-center MRS data collection. The study presented in this chapter contributes also to investigate whether the combination of single voxel  $^1\text{H}$  MRS at 1.5T at two different Echo time (TE), Short Time Echo (Short-TE) and Long Time Echo (Long-TE) improves the classification of pediatric brain tumours over the use of either TE alone. The results obtained with these classifiers are directly applied to the predictive engines of the Clinical DSS for brain tumours (see Chapter 7 for details).

The next two chapters present a novel on-line method to audit predictive models using a Bayesian perspective. This auditing method has been specifically designed for DSSs devoted to clinical or research environments. In Chapter 5 we describe the mathematical aspects of the audit method and focus on how it suits and improves the decision workflow of a clinical environment. The evaluation of the audit model involved a DSS for clinical purpose offering support through several classifiers. The efficacy of such a method is shown for the problem of diagnosis with a multi-centre database of MRS data of brain tumours.

Chapter 6 complements the previous chapter by presenting a methodology for prior probability assessment to a set of classifiers based on a similarity model. The evaluation of this model for proper prior assessment is performed with a multi-center database of MRS brain tumour data.

Chapter 7 describes the process of gathering the results obtained in the previous chapters into CURIAM Decision Support System for Clinical Environment (CURIAM), a DSS for BT diagnosis support. We describe the incorporation of the classifiers developed in Chapters 3 and 4 into CURIAM. This Chapter also shows the incorporation of the audit method and the similarity model presented in Chapters 5 and 6 to the CURIAM DSS in order to be applied in a clinical environment.

Finally, Chapter 8 summarizes the conclusions of the dissertation.

---

## Bibliography

- [1] J. Vicente, J. García-Gómez, S. Tortajada, E. Fuster-Garcia, A. Capdevila, A. Peet, B. Celda, and M. Robles, “Age-Filtered MRS Classifier To Overcome The Differences In Childhood And Adulthood Brain Tumours,” in *Proceedings 17th of the Scientific Meeting, International Society for Magnetic Resonance in Medicine, Honolulu, USA*, p. 981, April 2009.
- [2] J. Vicente, E. Fuster-Garcia, S. Tortajada, J. M. García-Gómez, N. Davies, K. Natarajan, M. Wilson, R. G. Grundy, P. Wesseling, D. Monleón, B. Celda, M. Robles, and A. C. Peet, “Accurate classification of childhood brain tumours by *in vivo*  $^1\text{H}$  MRS – a multi-centre study,” *European Journal of Cancer*, 2012. In Press.
- [3] J. Vicente, E. Fuster-Garcia, S. Tortajada, J. M. García-Gómez, N. Davies, K. Natarajan, M. Wilson, R. G. Grundy, P. Wesseling, D. Monleón, B. Celda, M. Robles, and A. C. Peet, “Accurate classification of childhood brain tumours by *in vivo*  $^1\text{H}$  MRS – a multi-centre study,” in *Proceedings of the 15th International Symposium on Pediatric Neuro-Oncology, Toronto, Canada*, 2012. In press.
- [4] J. Vicente, J. M. García-Gómez, S. Tortajada, A. T. Navarro, F. A. Howe, A. C. Peet, M. Julià-Sapé, B. Celda, P. Wesseling, M. Lluch-Ariet, and M. Robles, *Ranking of Brain Tumour Classifiers using a Bayesian Approach*, vol. 5517 of *Lecture Notes in Computer Science*, pp. 1005–1012. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.
- [5] J. Vicente, S. Tortajada, E. Fuster-Garcia, J. M. García-Gómez, and M. Robles, “An audit method suited for decision support systems for clinical environment,” in *Proceedings of 2012 IEEE 6th International Conference on Intelligent Systems (IS’12), Sofia, Bulgaria*, pp. 281–288, Sept. 2012.
- [6] J. M. García-Gómez, S. Tortajada, J. Vicente, C. Sáez, X. Castells, J. Luts, M. Julià-Sapé, A. Juan-Císcar, S. V. Huffel, A. Barcelo, J. Ariño, C. Arús, and M. Robles, *Genomics and Metabolomics Research for Brain Tumour Diagnosis Based on Machine Learning*, vol. 4507/2007 of *Lecture Notes in Computer Science*, pp. 1012–1019. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.
- [7] M. Lluch-Ariet, F. Estanyol, M. Mier, C. Delgado, H. González-Vélez, T. Dalmas, M. Robles, C. Sáez, J. Vicente, S. Huffel, J. Luts, C. Arús,

- A. P. C. Silveira, M. Julià-Sapé, A. Peet, A. Gibb, Y. Sun, B. Celda, M. C. M. Bisbal, G. Valsecchi, D. Dupplaw, B. Hu, and P. Lewis, "On the implementation of healthagents: Agent-based brain tumour diagnosis," in *Agent Technology and e-Health* (R. Annicchiarico, U. Cortés, C. Urdiales, M. Walliser, S. Brantschen, M. Calisti, and T. Hempfling, eds.), Whitestein Series in Software Agent Technologies and Autonomic Computing, pp. 5–24, Birkhäuser Basel, 2008.
- [8] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, M. Esparza, A. Navarro, E. Fuster-Garcia, M. Robles, L. Martí-Bonmatí, and C. Arús, "A generic decision support system featuring an assembled view of predictive models for magnetic resonance and clinical data," in *Book of abstracts of the ESMRMB 2008: 25th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Valencia, Spain*, p. 483, Springer-Verlag, Oct. 2008.
- [9] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, E. Fuster-Garcia, M. Esparza, A. Navarro, and M. Robles, "Curiam BT 1.0, Decision Support System for Brain Tumour Diagnosis," in *Book of Abstracts of the ESMRMB: 26th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Antalya, Turkey*, vol. 22, p. 538, Springer Berlin / Heidelberg, May 2009.
- [10] C. Sáez, J. M. García-Gómez, J. Vicente, S. Tortajada, J. Luts, D. Dupplaw, S. V. Huffel, and M. Robles, "A generic and extensible automatic classification framework applied to brain tumour diagnosis in HealthAgents," *The Knowledge Engineering Review*, vol. 26, no. Special Issue 03, pp. 283–301, 2011.
- [11] J. Vicente, C. Sáez, S. Tortajada, E. Fuster-Garcia, M. Esparza, M. Robles, and J. García-Gómez, "Curiam BT kids, a Clinical DSS for pediatric brain tumour diagnosis," in *Book of Abstracts of the ESMRMB: 29th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Lisbon, Portugal*, Springer Berlin / Heidelberg, October 2012. Accepted. In press.
- [12] J. M. García-Gómez, *Pattern recognition approaches for biomedical data in computer-assisted cancer research*. PhD thesis, Dpto. de Sistemas Informáticos y Computación. Universidad Politécnica de Valencia, 2008.

- [13] J. M. García-Gómez, J. Luts, M. Julià-Sapé, P. Krooshof, S. Tortajada, J. Vicente, W. Melssen, E. Fuster-Garcia, I. Olier, G. Postma, D. Monleón, A. Moreno-Torres, J. Pujol, A. P. Candiota, M. C. Martínez-Bisbal, J. Suykens, L. Buydens, B. Celda, S. Van Huffel, C. Arús, and M. Robles, “Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy,” *Magnetic Resonance Materials Physics, Biology and Medicine*, vol. 22, no. 1, pp. 5–18, 2009.
- [14] E. Fuster-Garcia, C. Navarro, J. Vicente, S. Tortajada, J. García-Gómez, C. Sáez, J. Calvar, J. Griffiths, M. Julià-Sapé, F. Howe, J. Pujol, A. Peet, A. Heerschap, A. Moreno-Torres, M. Martínez-Bisbal, B. Martínez-Granados, P. Wesseling, W. Semmler, J. Capellades, C. Majós, A. Alberich-Bayarri, A. Capdevila, D. Monleón, L. Martí-Bonmatí, C. Arús, B. Celda, and M. Robles, “Compatibility between 3T 1H SV-MRS data and automatic brain tumour diagnosis support systems based on databases of 1.5T 1H SV-MRS spectra,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 24, no. 1, pp. 35–42, 2011.
- [15] S. Tortajada, E. Fuster-Garcia, J. Vicente, P. Wesseling, F. a. Howe, M. Julià-Sapé, A.-P. Candiota, D. Monleón, A. Moreno-Torres, J. Pujol, J. R. Griffiths, A. Wright, A. C. Peet, M. C. Martínez-Bisbal, B. Celda, C. Arús, M. Robles, and J. M. García-Gómez, “Incremental Gaussian Discriminant Analysis based on Graybill and Deal weighted combination of estimators for brain tumour diagnosis,” *Journal of biomedical informatics*, vol. 44, no. 4, pp. 677–87, 2011.



## Chapter 2

# Theoretical foundations

*“A jug fills drop by drop.”*

*Buddha*

The first part of this chapter (Sections 2.1, 2.2 and 2.3) describes the problem of brain tumour diagnosis with the use of *in vivo* MRS and gives a description of the biochemical composition observed in *in vivo* MRS data in the brain. In Section 2.4 a description of the PR methods used in this Thesis is given. A survey of previous research carried out in automatic classification and diagnosis of brain tumour with *in vivo* MR spectroscopy will be exposed in Section 2.5. Section 2.6 reviews the design of clinical DSSs in BT Research.

### 2.1 Brain Tumours

A brain tumour is an abnormal mass of tissue in which some cells grow and multiply uncontrollably, apparently unregulated by the mechanisms that control normal cells<sup>1</sup>.

---

<sup>1</sup>National Brain Tumor Society. <http://www.brainumor.org/> (Online. Accessed 23-04-2012)

There are more than 120 types of brain tumours. They can be divided into primary and secondary brain tumours, attending to the type and location of the cells that originate from the tumour. Primary brain tumours are these originated in the brain itself (neurons, glial cells, lymphatic tissue, blood vessels, . . .), in the cranial nerves, in the brain envelopes (meninges), in the skull, pituitary, or pineal gland. Secondary brain tumours are created by cells spreading from cancers primarily located in other regions (metastatic tumours).

Nowadays, most medical institutions use the World Health Organization (WHO) classification system to identify brain tumors [1]. The WHO classifies the tumours of the Central Nervous System (CNS) as a taxonomy [2]<sup>2</sup>. attending to cell origin and how the cells behave, from the least aggressive (benign) to the most aggressive (malignant). Some tumour types are assigned a grade, ranging from Grade I (least malignant) to Grade IV (most malignant), which signifies the rate of growth.

The groups of tumours established in this Thesis were defined by medical experts participating in the European Projects INTERPRET [3], HEALTHAGENTS [4] and eTUMOUR [5] and following the WHO Classification.

The most frequent primary brain tumour types are from the meninges (33.8%), the second most common primary tumour types are derived from glial cells (32%) followed by those located in the sellar region (13.5%). Cranial and spinal nerves tumours and Lymphomas account for 8.7% and 2.4% respectively. Glioblastoma is the most frequent tumour of glial origin (53.8%) [6]<sup>3</sup>.

The distributions of the tumour types by age are not uniform. In early ages (0-19), embryonal/primitive/medulloblastoma and pilocytic astrocytoma are the most common types of tumours. In young adults (20-34), pituitary tumours and meningiomas are the most prevalent types. In older adults, the meningioma and, also, the malignant tumour glioblastoma are the most frequent tumours. The most common brain tumour types are summarized in Table 2.1 by age.

According to Pollack [7], the overall distribution of the various pediatric tumour types differs markedly from that of adults. Benign gliomas, primitive neuroectodermal tumour (PNET) and craniopharyngiomas account for a substantially higher percentage of brain tumours in children than in adults. In contrast, malignant gliomas, meningiomas, Schwann cell and pituitary tumours, and metastases, which are the most common brain tumours in adults, are comparatively

---

<sup>2</sup><http://www.brainlife.org/classification/who2007.htm>

<sup>3</sup>Information of brain tumour cells can be found in Section A.1 of Apendix A

Table 2.1: Most common brain tumour and CNS types by age in years (CBTRUS 2010, Statistical Report [6]).

Age	Most Common Histology	Second Most Common Histology
0-4	Embryonal/primitive/medulloblastoma	Pilocytic astrocytoma
5-9	Pilocytic astrocytoma	Malignant Glioma
10-14	Pilocytic astrocytoma	Non-malignant and malignant neuronal/glial
15-19	Pituitary	Pilocytic astrocytoma
20-34	Pituitary	Meningioma
35-44	Meningioma	Pituitary
45-54	Meningioma	Glioblastoma
55-64	Meningioma	Glioblastoma
65-74	Meningioma	Glioblastoma
75-84	Meningioma	Glioblastoma
85+	Meningioma	Glioblastoma

uncommon in children [8, 9, 10]. There are also notable differences between children and adults in the distribution of tumours by location. The vast majority of adult brain tumours arise in and around the cerebral hemispheres whilst approximately 50% of brain tumours in children older than 1 year of age develop infratentorially [11]<sup>4</sup>.

The growth of a tumour takes up space within the skull and interferes with normal brain activity. A tumour can cause damage by increasing pressure in the brain, by shifting the brain or pushing against the skull, and by invading and damaging nerves and healthy brain tissue. The location of a brain tumor influences the type of symptoms that occur: visual disturbance or respiratory arrest, as well as epilepsy, or sudden intracranial hypertension may occur.

When one or more of these symptoms arise, a radiological examination based on Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) is usually performed. Information provided by non-invasive radiological techniques is highly relevant for evidence-based medicine applied to the diagnosis, patient management and to surgery [12, 13]. MRI is widely used for determining tumour extension in surgical and radiotherapy planning, with a sensitivity of 14%-100% depending on tumour type and precision of the wording used by radiologist [14].

Nevertheless, the gold-standard method in the brain tumour diagnosis is the histological examination of tumour tissue samples obtained either by means of brain biopsy or open surgery. Up to date, histopathology provides the main information for deciding the treatment and the prognosis for each patient.

---

<sup>4</sup>Information of brain tissues and regions can be found in Section A.2 of Appendix A

MRS is slowly becoming an additional accurate non-invasive technique for initial examination of brain masses [13, 15, 16, 17], due to its capability to provide useful chemical information of different metabolites for characterizing brain tumours and its complementary role to MRI [18, 19, 20, 21].

## 2.2 Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) spectroscopy is generally known as an analytical method in chemistry to identify molecules and to determine their biophysical characteristics. NMR (or MR) is, thus, defined as the phenomenon where the nuclei of certain atoms absorb and emit energy because of the effect of an oscillating magnetic field when they are immersed in other static magnetic field [22].

The main application of NMR in the clinic is to obtain detailed anatomical images non-invasively throughout the human body by MRI. Magnetic Resonance Imaging (MRI) is the use of the NMR phenomenon to visualize the structure and function of a body. It is largely used in medical imaging because of its contrasts of soft tissues.

However, not only MRI but also NMR spectroscopy has several clinical and biomedical applications. When an NMR technique is used *in vivo*<sup>5</sup> is referred to as Magnetic Resonance Spectroscopy (MRS). MRS is the use of the NMR phenomenon to study the physical, chemical, and biological properties of organic and inorganic molecules in a non-destructive, non-invasive manner. Just as in its application in chemistry, MRS allows the detection of relatively small molecules. The obtained MR spectra supply information on metabolic pathways and changes therein, which makes MRS a very suitable technique to monitor metabolic changes due to disease and to follow treatment.

## 2.3 MRS of the brain from a biochemical perspective

---

<sup>5</sup>An *in vivo* technique means that it is applied directly to the patient, in contrast to *ex vivo*, where a tissue sample is extracted from the patient.

MR spectroscopy has arisen as a non-invasive tool for the investigation of metabolic changes in living tissues that are otherwise inaccessible. Interpretation of MRS data is based on general knowledge of biochemical processes in association with pathological changes revealed as changes in metabolite signals in MR spectra [23].

At the present, TE used in *in vivo*  $^1\text{H}$  MRS by most studies range between 18 and 288 ms. A spectrum acquired with a TE < 45 ms is commonly considered a Short Time Echo (Short-TE) spectrum, Long Time Echo (Long-TE) spectrum otherwise.

Short-TE (20-35 ms)  $^1\text{H}$  MRS allows to observe several metabolites and other compounds considered useful for tumour classification. N-Acetyl Aspartate (NAA) (2.01 parts per million (ppm)), Choline (Cho) (3.21 ppm), Creatine (Cr) (3.02, 3.92 ppm), Glutamate + Glutamine (Glx) (2.04, 2.46 ppm), Alanine (Ala) (1.4, 3.78 ppm), myo-Inositol (mI) (3.26, 3.53 ppm) and Mobile lipids (ML) (0.92 and 1.29 ppm) are mostly observed in Short-TE spectra [24, 13]. MRS allows us to obtain metabolite concentration information in a volume defined by a Single voxel (SV). SV Short-TE  $^1\text{H}$  MRS is fast (typically 5 min) and robust, and considered, thus, adequate for clinical studies [25, 13]. These signals, though, show a large number of overlapping peaks, a strong ML-originated baseline and a certain sensitivity to artifacts [15].

Long-TE (about 135 ms)  $^1\text{H}$  MRS is less informative than Short-TE, because resonances with short T2 may be lost [23]. However, Long-TE signals are easier to analyze than Short-TE signals [15]. ML will not be the dominating components at Long-TE, making possible the study of the contributions of Lactate (Lac) and Ala as inverted peaks (see Table 2.2) [13, 26].

The resonance peaks identified in *in vivo* MR spectra are briefly described in the next sections from the viewpoint of their chemical features emphasizing the characteristics specific to children and adults when possible.

### **2.3.1 N-Acetyl Aspartate (NAA)**

#### **Role and origin of NAA**

The resonance at 2.01 ppm is the highest peak in the spectra of a normal adult brain and it is called NAA peak.

Table 2.2: List of resonance peaks in brain  $^1\text{H}$  MR spectroscopy for Short-TE and Long-TE (indicated by a  $\checkmark$  symbol).

PEAK	Shift (ppm)	Short-TE	Long-TE
Macromolecules and lipids components at 0.9 ppm (MMLip09)	0.92	$\checkmark$	$\checkmark$
Macromolecules and lipids components at 1.3 ppm (MMLip13)	1.29	$\checkmark$	$\checkmark$
Lactate (Lac)	1.31	$\checkmark$	$\checkmark$
Alanine (Ala15)	1.47	$\checkmark$	$\checkmark$
N-Acetyl Aspartate (NAA)	2.01	$\checkmark$	$\checkmark$
Glx: Glutamate (Glu) + Glutamine (Gln)	2.04, 2.46	$\checkmark$	–
Creatine (Cr)	3.02, 3.92	$\checkmark$	$\checkmark$
Choline (Cho)	3.21	$\checkmark$	$\checkmark$
myo-Inositol (mI)	3.26, 3.53	$\checkmark$	–
Taurine (Tau)	3.42	$\checkmark$	–
Glycine (Gly)	3.55	$\checkmark$	$\checkmark$
Alanine (Ala38)	3.78	$\checkmark$	–

N-Acetyl Aspartate is produced in the mitochondria of the neurons and transported into the neuronal cytoplasm [27]. Although the exact role of NAA is presently not known, it continues to be used as a marker of neuronal density and viability. NAA is also found in axons as it is transported along axons from the site of synthesis in the neuronal mitochondria. The utility of NAA as an axonal marker is supported by the loss of NAA in many white matter diseases like multiple sclerosis [17]. NAA is undetectable in astrocytes and oligodendrocytes, although large amounts of NAA (twice that of neurons) are found in immature oligodendrocyte and astrocyte progenitor cells [28]. The concentration of NAA in the brain has anatomical variations. It is 35% higher in grey matter and white matter than in the thalamus [29]. In a similar fashion, we find higher concentrations of NAA in the grey matter than in the white matter at a ratio of 1.5, which reflects a higher neuronal density in grey matter [30, 31].

The level of NAA at birth is only 50% of that in adulthood [32] and increases considerably with brain development, paralleled by a decrease in Cho. The NAA/Cho ratio is reversed in children in the first year, in whom myelination is nearly complete [33]<sup>6</sup>. The NAA concentration does not reach the level of the adult brain until the 16th year [34].

<sup>6</sup>Definition of myelination can be found in Section A.3 of Appendix A

## **NAA levels in abnormalities**

Changes in the signal intensity of NAA generally appear as a reduction. First, a reduction in or disappearance of NAA is a sign of a reduced number or absence of neurons within a region. Second, it is a sign of the dysfunction or death of neurons or axons. Hence, NAA levels appear to be a sensitive but non-disease-specific, marker of neuronal number and viability [35, 36]. NAA in brain white matter could be a biochemical index of axonal integrity [37].

The levels of NAA in brain tumour spectra show large variation. In principle, if the spectra are obtained from tumour cells, the NAA signal might be absent in gliomas and meningiomas since glial and meningeal cells do not contain detectable amounts of NAA. However, as the Volume of interest (VOI) do not always avoid neurons present in brain tissues in clinical measurement, the appearance of NAA signal in tumour spectra may depend on whether the growth pattern of the tumour is infiltrative or circumscribed (see Figure 2.1)<sup>7</sup>. In the later case, like in typical meningiomas, some lymphomas and metastases, tumours have clear margins and might show almost no NAA resonance peak. MR spectra of tumours that tend to grow infiltratively, such as astrocytomas and some lymphomas, show a detectable NAA of various signal intensities (see Figures 2.2 and 2.3) but, even in these cases, the NAA peak is severely decreased or undetectable.

When dealing with children, large reductions in NAA with elevated Cho or Cho/Cr compared with normal-appearing brain tissue generally indicates tumour [13].

### **2.3.2 Choline (Cho)**

#### **Role and origin of Cho**

Cho is a metabolic marker of membrane density and integrity with its peak located at 3.2 ppm. Choline is found in much higher concentrations in glial cells, that is, oligodendrocytes and astrocytes, than in neurons [33, 28].

In neonatal brain, the spectrum shows a high signal intensity of Cho and a relatively low signal intensity of NAA. The absolute Cho content is significantly

---

<sup>7</sup>Definitions of tumoral growth patterns is given in section A.3 of Appendix A

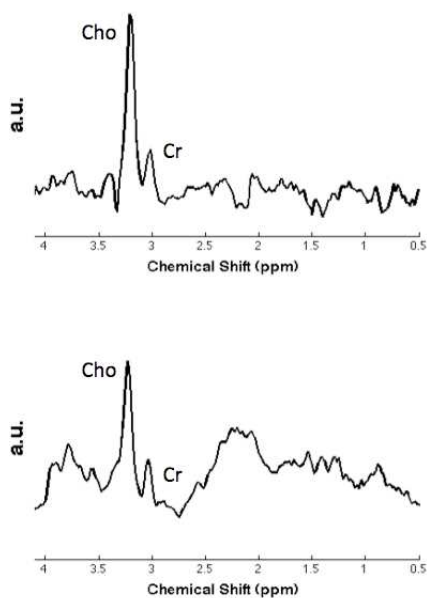


Figure 2.1: Proton spectra of a low grade meningioma (64-year-old, male; case et3028 [5]), showing dominant Cho. Cr and NAA are reduced. Spectra were obtained with Long-TE (top) and Short-TE (bottom).



higher (+ 82%) than the corresponding values in the adult [32]. With maturation of the brain, considerable changes are observed in the signal intensity of Cho as well as that of NAA. Changes are rapid during the first year of life, after which the Cho-to-NAA relationship is reversed, reflecting near completion of myelination [33]. After three years of age, the relationship begins to stabilize [33].

Choline-related molecules are precursors in the synthesis and breakdown products of membrane phospholipids<sup>8</sup>. The signal intensity of Cho, therefore, might be increased by the acceleration of membrane synthesis and breakdown. The elevated concentrations of Cho may be associated with proliferation of both benign and malignant cells and with high cellularity.

### **Cho concentrations in abnormalities**

The relationship between Cho elevation in proton spectra and the cellular proliferative activity has been extensively studied in adults [38]. In tumours, the Cho signal is correlated with malignancy [39]. However, in cerebral infarction and inflammation the Cho signal may also increase [40], which may make difficult a correct differentiation between tumours and other pathologies.

An increase in Cho tends to be detected in high-grade gliomas (see Figures 2.2 and 2.3). Many glioblastomas reveals a high Cho peak, but Cr and NAA are not detectable in a certain number of high-grade gliomas [41, 42]. Necrotic high-grade lesions have a reduced amount of Cho [43].

In children, a significantly shorter survival time has been observed for patients with the highest Cho/NAA ratios, independent of the tumour type [44, 13]. In low-grade pediatric gliomas the relative Cho was found to be higher in the group of patients whose tumours progressed [13].

### **2.3.3 Creatine (Cr)**

#### **Role and origin of Cr**

In simplistic terms, Cr is a marker of ‘energy metabolism’. The central peak at 3.02 ppm on the spectrum represents the sum of creatine and phosphocreatine.

---

<sup>8</sup>Definition of phospholipids membrane can be found in Section A.3 of Appendix A

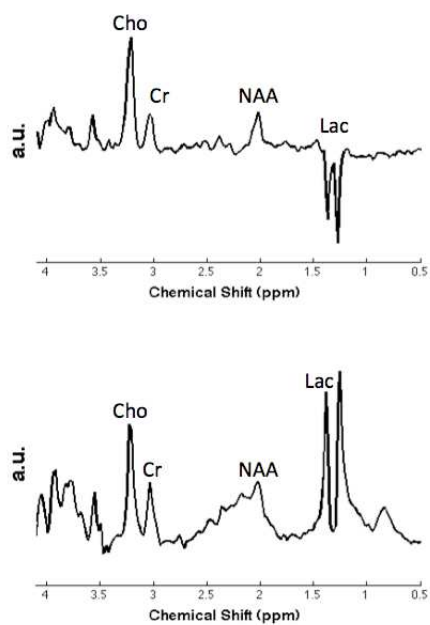


Figure 2.2: Proton spectra of an anaplastic astrocytoma (42-year-old, female; case et3136 [5]), showing intensive Cho and Lac. Spectra were obtained with Long-TE (top) and Short-TE (bottom).

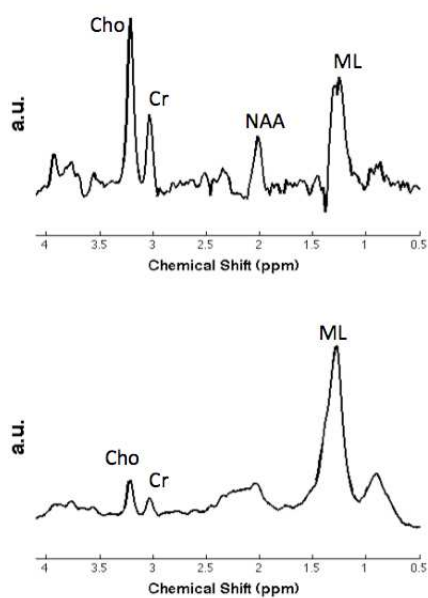


Figure 2.3: Proton spectra of a glioblastoma (73-year-old, male; case et2357 [5]). Spectra were obtained with Long-TE (top) and Short-TE (bottom).

In the clinical setting, Cr is assumed to be stable and is used for calculating metabolite ratios (Cho/Cr and NAA/Cr ratios) [17].

Cr is found in much higher concentrations in oligodendrocytes and astrocytes than in neurons [28]. Since a reduction in Cr signal intensity indicates an energy deficit in glial cells, it can be used as a reliable marker of cellular integrity.

### **Cr concentration in abnormalities**

An absence of Cr in brain lesions is generally encountered when the lesions lack creatine kinase, such as meningiomas, lymphomas and metastatic brain tumours [42]. A reduction in Cr signal intensity occurs also in severely compromised physiologies, such as hypoxic tissues and aggressive tumours. In high-grade gliomas the Cr level tends to decrease, probably reflecting physiological severity within the glial cells (see Figures 2.2 and 2.3). Porto *et al.* [45] suggest that total Cr concentrations combined with lactate could be helpful in the differential diagnosis of pilocytic astrocytomas and diffuse astrocytomas in children.

## **2.3.4 Glutamate + Glutamine (Glx)**

### **Role and origin of Glx**

Glu and Gln are generally inseparable at 1.5T and result in a complex of peaks (Glx) between 2.05 and 2.46 ppm which are most readily observed at Short-TE. In the brain, Glu and Gln are found primarily within the neurons and astrocytes. Glu is an excitatory neurotransmitter, being the most abundant neurotransmitter in the brain [17]. Glu and Gln play a role in detoxification and regulation of neurotransmitters [23].

### **Glx concentration in abnormalities**

Toxicity to brain cells can occur in pathological states during prolonged extracellular exposure to high Glu concentrations. The neurotoxicity of Glu has been reported in a wide range of neurodegenerative diseases, including multiple sclerosis, Alzheimer's disease and schizophrenia [46].

Glx is a prominent feature of meningiomas, possibly reflecting altered energy metabolism involving partial oxidation of glutamine rather than glycolysis [47]<sup>9</sup>, the end product being Ala, which is also elevated in meningiomas [48, 47].

### **2.3.5 Lactate (Lac)**

#### **Role and shape of Lac**

When present, Lac is recognized as a doublet (twin peak) at 1.31 ppm. Lac is characterized by variable projection of the peak at different TEs: On acquisitions using Long-TE, the doublet peak is inverted below the baseline, but at Short-TE, the doublet peak projects above the baseline (see Figure 2.2) [17]. In normal brain tissues where glucose is consumed in an oxygenated metabolism, Lac is present only in small amounts and is not resolved using the normal spectroscopic techniques. The appearance of the Lac peak, thus, is always associated with the abnormal state of the brain.

#### **Lac concentrations in abnormalities**

Lac is raised in a variety of conditions: reduced oxygen supply, accelerated glycolysis, dysfunction of mitochondria, inflammation and fermentation. Anaerobic glycolysis is a major pathway in the production of Lac and, thus, the Lac level could be an index of impairment of oxidative metabolism [49]. Most tumour cells have low respiration and accelerated glycolysis. Thus, the absence of Lac may suggest slow growth. On the contrary, high-grade tumours exhibit a trend toward a higher Lac signal [50]. Lac may reflect ischemic compromise in a large tumour, especially in the central area, leading to anaerobic glycolysis.

### **2.3.6 myo-Inositol (mI)**

#### **Role and origin of mI**

mI is a simple sugar appearing as a single peak at 3.26 and 3.53 ppm which can be observed only at Short-TE. It is absent from neurons and, since mI is synthesized in glial cells, it is considered to be a glial marker [17, 39].

---

<sup>9</sup>Definition of glycolysis can be found in Section A.3 of Appendix A

### **mI concentrations in abnormalities**

An increase in mI content is believed to represent glial proliferation or an increase in glial cell size, both of which may occur in inflammation [51]. mI is believed to be mostly found in astrocytes [52] and is high in low-grade gliomas [48], but low or absent in non-glial tumours such as schwannomas [53] and meningiomas [48, 54]. Higher levels of mI have been reported to distinguish heman-giopericytomas from meningiomas [24].

### **2.3.7 Glycine (Gly)**

#### **Role and origin of Gly**

Gly is an amino acid present at almost the same chemical shift position as mI (3.55 ppm). It has a longer transversal relaxation, T<sub>2</sub>, so that it is also present in MR spectra obtained at Long-TE [17, 39].

#### **Gly concentrations in abnormalities**

A small peak at 3.55 ppm has been observed in astrocytoma spectra at Long-TE and is thought to be Gly with its longer T<sub>2</sub> rather than mI [48, 55]. High Resolution Magic Angle Spinning Nuclear Magnetic Resonance (HR-MAS) studies of tumour biopsies indicate that Gly and mI are both elevated in astrocytomas, decreasing with grade, and are absent in meningiomas [56].

In a recent study [57], the Gly concentration revealed as a promising biomarker of malignancy in pediatric brain tumours.

### **2.3.8 Taurine (Tau)**

#### **Role and origin of Tau**

Tau is an organic acid that resonates at 3.42 ppm which is only observed at Short-TE. Tau is implicated in a wide array of physiological phenomena including inhibitory neurotransmission [58], membrane stabilization [59], protection against Glu excitotoxicity [60] and prevention of epileptic seizures [61].

Tau is generally considered difficult to accurately measure in *in vivo* MRS because of signal overlap from stronger signals of mI at 3.26 and 3.53 ppm and Cho at 3.21 ppm [62].

### **Tau concentration in abnormalities**

Measurement of Tau in gliomas in *in vivo* MRS can be a useful technique for monitoring tumour apoptosis in the clinic [62], specially in glial tumours.

Related with pediatric brain tumours, Tau reveals as a biomarker for medulloblastoma since it is observed in highly elevated concentrations [63, 64, 65].

### **2.3.9 Alanine (Ala)**

#### **Role and origin of Ala**

Ala is an amino acid with uncertain function with a doublet peak at 1.47 ppm, but its presence may be overshadowed by Lac, which resonates at 1.31 ppm. The Ala peak shows a phase inversion in a spectrum acquired with Long-TE. It has a role in the citric acid cycle and is undetectable in normal brain spectra.

#### **Ala concentrations in abnormalities**

Ala may be found in some meningiomas [42] and pyogenic abscesses [66].

### **2.3.10 Mobile lipids (ML)**

#### **Role and origin of ML**

ML are narrow peaks at 1.3 and 0.9 ppm which are a major component of the brain, representing up to 20% [38]. This lipids are so-called ML because they apparently come from the cell membranes during the ongoing metabolic changes associated with programmed cell death (apoptosis)<sup>10</sup> [67]. Although in normal healthy brain these peaks are absent, an inappropriate volume selection may result in contamination of the spectrum with lipid signals from tissue close to the skull [17, 39].

---

<sup>10</sup>Definition of apoptosis is given in Section A.3 of Appendix A

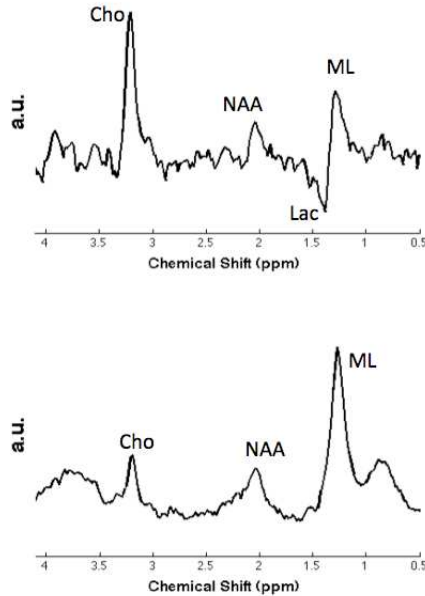


Figure 2.4: Proton spectra of a metastasis (43-year-old, male; case et3181 [5]). Cho and ML peaks are detected in addition to Lac. Spectra were obtained with Long-TE (top) and Short-TE (bottom).

### ML concentrations in abnormalities

ML signals in pathology are generally associated with necrosis<sup>11</sup> such as in high-grade brain tumours or metastases. *Ex vivo* studies have demonstrated a relation between the ML signal and necrosis percentage in high-grade astrocytomas [68]. Hence, the increase in content of MR-visible lipids might be due to hypoxia (intracellular accumulation) and necrosis (extracellular accumulation) [69, 70]. Many metastatic brain tumours show ML, sometimes intensively [41, 71] (see Figure 2.4). Like in high-grade gliomas, the presence of ML in metastases may be related to tissue necrosis [72].

<sup>11</sup>Definition of necrosis can be found in Section A.3 of Appendix A



## 2.4 Pattern Recognition (PR)

The relationship between metabolite concentrations found in tumour lesions and their classification is a complex task that needs to be performed by experts. Traditionally, as an initial research step, a univariate analysis to the MRS data can be carried out in order to assess the levels of individual, specific metabolites. Depending on the tumour type and stage, different combinations of metabolite concentrations can be observed. The univariate analysis may explain the contribution of a metabolite but lacks the ability to explain how two or more metabolites behave together in a tumour lesion. The Pattern Recognition (PR) discipline can offer such inferential multivariate approach. The PR discipline has been successfully applied in diagnosis and characterization of MRS data since the early 90s.

The PR discipline studies how to recognize an object through regularities that indicate that it belongs to a certain class or category. Typical applications of PR emulates easy tasks for humans, such as speech recognition, optical character recognition, face recognition; or other specialized activities, such as language translation, fingerprint identification, or quality control of industrial manufacturing [73].

There are two alternative possibilities of design of a PR system: to implement a set of rules and exceptions based on some heuristic reasoning; or to base the development on a Machine Learning approach, on which a data set is used to fit an adaptive model to solve the problem. Machine Learning provides the mathematical and computational mechanisms to infer knowledge in a formal model from specific data of a given domain [74, 73]. The methodology applied in this Thesis solves the PR in brain tumour classification using the Machine Learning approach.

The life cycle of a PR problem based on Machine Learning can be divided into two main phases: the Training phase and the Recognition phase (see Figure 2.5). During the Training phase, a set of signals from the problem domain (the training corpus) is used to adapt a mathematical function of the output values, e.g diagnosis, treatment, doses or risk. In this phase, the pre-processing and the features extracted from the signals are established. Then, an adaptive model is fitted, selected and evaluated in order to obtain the best generalization for solving new cases (recognition phase).

Once the model is ready, it can be incorporated into a Decision Support System (DSS) to be used for the recognition of new cases.

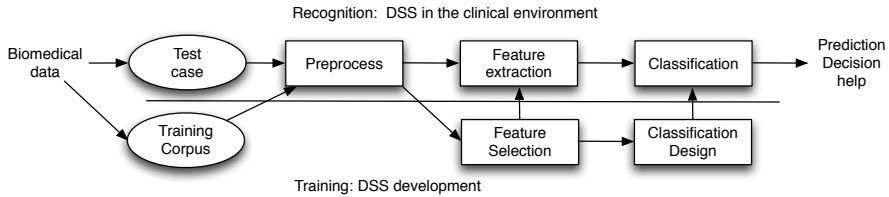


Figure 2.5: PR solution based on the Machine Learning approach.

### 2.4.1 PR methods for classification

Different methods are proposed in the literature to perform classifications which, in most cases, consist on estimating the real probability distributions from the observation of a finite set of cases. From the wide variety of classification methods available, this Thesis proposes the use of standardized and widely accepted methods in BT characterization with MRS data. Hence, based on the benchmark results of previous studies, a set of classification methods are chosen to allow direct comparison with formerly reported achievements. A survey of these studies is provided in Section 2.5. A brief description of the classifiers used in this dissertation is presented below.

#### Gaussian Parametric Models

A Gaussian parametric model is a method used to characterize or separate two or more classes of objects [75]. This separation is performed by estimating the max-likelihood of the Gaussian distribution for the classes under the study. Linear Discriminant Analysis (LDA) is a commonly used method to find linear combination of features based on the assumption that all the classes share a common variance (see Figure 2.6-left). Quadratic Discriminant Analysis (QDA) can also be performed by assuming that the covariances of the classes are independent. In this case, quadratic decision boundaries can be obtained for separating the classes (see Figure 2.6-right).

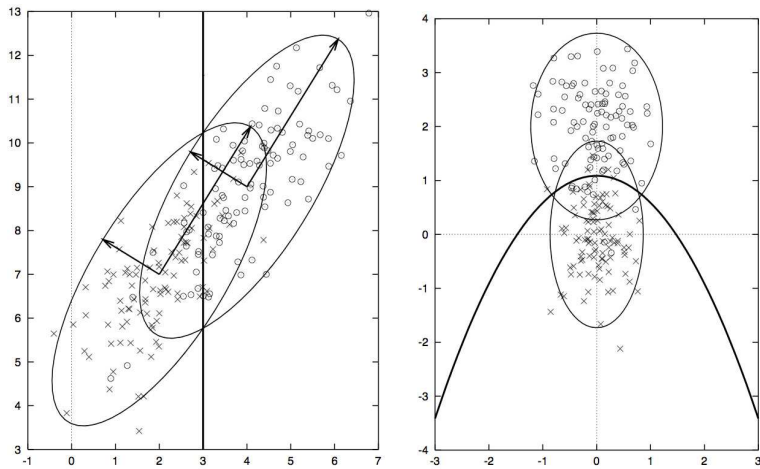


Figure 2.6: The left plot shows the linear decision boundary obtained by LDA for a two-class problem in a 2D space. In addition, the ellipsoids depicting the 95% confidence interval for Mahalanobis distances around the mean of each class are depicted. The right plot shows the quadratic solution obtained by QDA for a similar problem and the 95% confidence interval for Mahalanobis distance of each class.

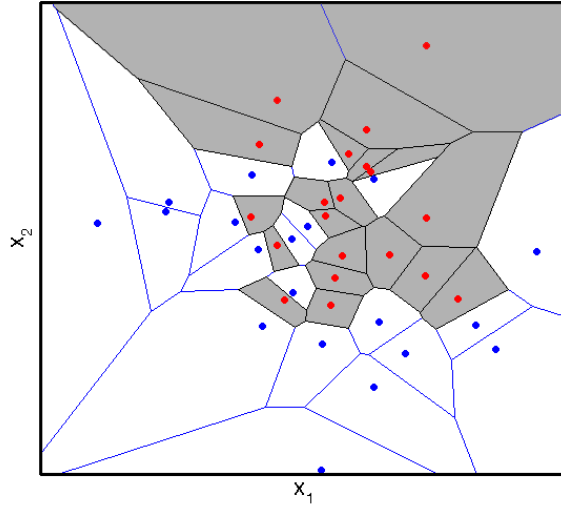


Figure 2.7: In two dimensions, the nearest-neighbor algorithm leads to a partitioning of the input space into cells, each labeled by the category of the training point it contains.

### K-Nearest neighbors (KNN)

The K-Nearest neighbors (KNN) is a non-parametric method of classification based on closest training examples in the feature space. In KNN, a sample is classified according to the majority vote of the  $k$  neighbors. In other words, the class of that sample will be that of the most common class among the  $k$  nearest neighbors. KNN is a type of instance-based learning where the function is approximated locally and all computation is deferred until classification [76, 77] (see Figure 2.7).

## 2.4.2 Feature selection and feature extraction

The feature extraction and feature selection step consists in searching for the optimal (or suboptimal) set of features that better represents the objects for the goal of the PR task. The feature extraction methods used in our experiments were based on multivariate statistical analysis and signal processing.

### Feature extraction with multivariate statistical analysis

Feature extraction can provide dimensionality reduction, eliminate redundant information, filtrate the available information avoiding noise artifacts and non-informative data and, thus, simplify the classification problem. There are several methods described in the literature to perform feature extraction. Some of the techniques most generally used in BT diagnosis and characterization with MRS are described:

**Principal Components Analysis (PCA):** is a projection method commonly used for feature extraction in PR [78, 79]. PCA maps the original  $D$ -dimensional data into an orthogonal  $P$ -space, where the axes of this new coordinate system lie along the direction of maximum variance of the original data. The more correlated the original variables are, the more the data variation is explained by the first principal components (PCs) of the analysis. Hence, feature reduction can be carried out discarding the remaining PCs.

PCA is closely related to LDA in that both techniques look for linear combinations of variables which best explain the data [80]. PCA, though, does not take into account the differences between the classes of data as LDA does.

**Independent Component Analysis (ICA):** a technique that extracts statistically independent components from a set of measured signals. The method is used for blind source separation and can extract a pure signal from a set of mixtures [81]. ICA is an eligible feature extraction technique for MRS spectra due to the presence of partial volume effects. Partial volume effects occur when a signal from a specific voxel contains components of different tissue types. The input for the ICA method is usually the full region of interest of the real spectrum.

**Non-Negative Matrix Factorization (NMF):** is a feature extraction technique useful when there are many attributes and the attributes are ambiguous or have weak predictability. By combining attributes, NMF can produce meaningful patterns. While other feature extraction techniques such as PCA are able to learn holistic, not parts-based, representations, the NMF can learn parts of the signal [82].

NMF decomposes multivariate data by creating a user-defined number of features (parts). Each feature is a linear combination of the original attribute set; the coefficients of these linear combinations are non-negative. These non-negative combination, obtained by using non-negativity constraints, lead to a parts-based representation because they allow only additive, not subtractive, combinations.

NMF decomposes a data matrix  $V$  into the product of two lower rank matrices  $W$  and  $H$  so that  $V \approx WH$ . NMF uses an iterative procedure to modify the initial values of  $W$  and  $H$  so that the product approaches  $V$ . The procedure terminates when the approximation error converges or the specified number of iterations is reached.

PCA was the multivariate analysis technique of choice in the experiments presented in this dissertation. Analogously to the PR methods selected for classification, PCA has been widely accepted in the literature and is a simple, standard and straight multivariate technique that obtains reasonably good results in MRS feature extraction [83, 84].

## **Feature extraction with signal processing**

### **Peak integration (PI)**

Deviations in the MRS spectra are caused by differences in metabolite concentrations. Thus, the spectral ranges circumscribing signals of relevant brain metabolites contain information enough to discriminate between brain tumours whilst the remaining regions include mainly spectral noise. In order to reduce the dimensionality and exclude the noise in the data analysis, an estimation of the levels of metabolites can be performed.

The amplitude of a resonance is proportional to the integral of the corresponding peak in the spectrum. Nevertheless, accurate estimation of these integrals is difficult due to several factors such as spectral noise, nonzero baseline, peak

overlap or the discrete nature of the signal. By using the trapezoidal rule, a Peak integration (PI) can be performed. The procedure is as follows:

Let  $s = (s_1, \dots, s_l)$  be a sample of the function  $s(t)$  from which we want to approximate the integral with respect to  $t = (t_1, \dots, t_l)$ .  $I = \int_{t_1}^{t_l} s(t)dt$ . The trapezoidal integration is computed as:

$$\begin{aligned}\Delta t &= (t_2 - t_1, \dots, t_l - t_{l-1}) \\ s' &= \left( \frac{s_1 + s_2}{2}, \dots, \frac{s_{l-1} + s_l}{2} \right) \\ I &= \Delta t s'\end{aligned}\tag{2.1}$$

The area under the frequency peak in the magnitude spectrum is then calculated for each selected metabolite. The intensities of fifteen spectral ranges which have proportionality to the concentrations of the 11 main metabolites observed in Short-TE are selected and integrated within a window of 0.15 ppm. In the case of Long-TE spectra, nine spectral ranges are integrated, corresponding to 8 main metabolites. Table 2.3 reflects the metabolite concentrations considered by Short-TE and Long-TE PI. This technique has been successfully applied in previous studies with MRS from adults [83, 84].

### **Pediatric PI**

An alternative extraction of estimations devoted to children (pediatric PI) was designed for this Thesis. The pediatric PI extracts only the relevant metabolite concentrations according to Peet *et al.* [85], what results in 12 estimations for Short-TE and 7 estimations for Long-TE. The estimations correspond to 10 and 7 main metabolites in Short-TE and Long-TE respectively. The basic difference between PI and pediatric PI configurations was that the later did not estimate the Gly and the Ala at 3.78 ppm because the concentrations registered from 1.5T Short-TE in-vivo spectra are not usually significant in children. Table 2.3 reflects the resonance peaks of metabolites considered by each PI configuration.

### **Metabolites quantitation**

Quantification of metabolites and components present in  $^1\text{H}$  MRS *in-vivo* data is a vast signal processing discipline that offers a biological interpretation. Although, strictly speaking, the problem of quantifying signals present in MRS data falls beyond the feature extraction scope, quantitation algorithms can be

Table 2.3: Metabolites associated to each peak/region extracted with PI and pediatric PI for Short-TE and Long-TE (indicated by a  $\checkmark$  symbol). Estimations are performed within a window of 0.15 ppm centered in the peak/region.

PEAK	Short-TE		Long-TE	
	PI	Pediatric PI	PI	Pediatric PI
MMLip09	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
MMLip13	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Lac	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Ala15	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
NAA	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Glx: Glu + Gln	$\checkmark$	$\checkmark$	–	–
Cr	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
Cho	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
mI	$\checkmark$	$\checkmark$	–	–
Tau	$\checkmark$	$\checkmark$	–	–
Gly	$\checkmark$	–	$\checkmark$	–
Ala38	$\checkmark$	–	–	–

applied to *in-vivo* data as a feature extraction technique to obtain an estimation of the metabolite contributions in the spectrum.

Several quantitation algorithms have been proposed in the literature [86]. These algorithms can be categorised as follows: black-box, peak fitting and basis set. Black-box methods for *in-vivo* data are based on signal decomposition algorithms and are effective at extracting peak parameters from simple data. Hankel-Lanczos Singular Value Decomposition (HLSVD) is a computationally efficient black-box algorithm [87] successfully applied in MR spectra of human brain. The main drawback of black-box methods is that additional knowledge of spectral features cannot be incorporated. The Advanced Method for Accurate, Robust and Efficient Spectral fitting of MRS data with use of prior knowledge (AMARES) algorithm [88] was developed to address this issue allowing a greater level of prior knowledge in the fitting model.

Black-box and peak fitting methods are effective approaches to quantify sparse spectra such as Long-TE. For complex data, like Short-TE, which includes more metabolic information, methods that incorporate a metabolite basis set have been shown to be more appropriate [89].

LCModel<sup>TM</sup> is a commercial software [90] widely used for the analysis of Short-TE MRS that incorporates a metabolite basis set into the fitting model. The al-



gorithm models data in the frequency domain using a linear combination of metabolite, lipid and macromolecule signals combined with a smoothing splines to account for baseline signals. The Quantitation based on semi-parametric quantum estimation algorithm (QUEST) [91] has been recently developed under an open source license. QUEST also uses a combination of time-domain fitting but, in contrast with LCMoDel<sup>TM</sup>, HLSVD is used to model the baseline signal. Automated Quantitation of Short Echo time MRS Spectra (AQSES) [92] is a recent open source quantitation algorithm that also uses a combination of time-domain fitting and penalized splines to model the baseline. It differs from LCMoDel<sup>TM</sup> in the use of a variable projection method to estimate the amplitudes of the metabolite basis set in a reduction in the number of model parameters. In the same line drawn by AQSES, the Totally Automatic Robust Quantitation in NMR (TARQUIN) open source algorithm [93] performs a time-domain truncation to eliminate baseline interference and models the remaining signal with a parameterised basis set containing metabolites and, in contrast to QUEST and AQSES, lipids and macromolecule signals.

The problem with the use of sophisticated quantification algorithms is that they are in general hard to automate. AQSES requires to set the parameter that controls the smoothness of the baseline. In QUEST a selection of the number of points for baseline construction is required. TARQUIN, though, does not require any parameterization, allowing a non-expert user to process MRS spectra without a great investment of time. TARQUIN algorithm offers comparable results to the commercial algorithm LCMoDel<sup>TM</sup> [93], with two main advantages: It is offered under an open source license, which allows a free use for educational and research purposes, and it is totally automated and no manual intervention is needed to perform the quantitation.

The TARQUIN (version 4.1.1) software tool is used in the experiments of this dissertation. The quantitation of MRS data is performed with the standard metabolite library provided [93] by TARQUIN, which consisted on 16 different metabolites and nine lipids and macromolecular components. Attending to the recommendation in the LCMoDel manual [94], a simulated negative singlet at 3.94 ppm (-CrCh<sub>2</sub>) was included to account for attenuation of the CH<sub>2</sub> peak of Cr. The quantitation of metabolites, lipids and macromolecular components was performed using the spectrum acquired without water suppression as reference. A total of 21 variables were estimated after grouping the macromolecular and lipid components into three main resonances around 0.9 ppm (MMLip09), 1.3 ppm (MMLip13) and 2.0 ppm (MMLip20). Table 2.4 lists the set of variables estimated with TARQUIN.

Table 2.4: Metabolites, lipids and macromolecular components of the basis set provided by TARQUIN.

Elements in the TARQUIN basis set
Ala
Cr
Glc
Gln
Glu
Gly
mI
Lac
NAA
Scyllo
Tau
GPC
PCh
-CrCH <sub>2</sub>
Gua
GABA
Asp
Cho: GPC+PCh
Glx: Glu+Gln
MMLip09
MMLip13
MMLip20

### 2.4.3 Evaluation procedures

Usually, when dealing with small-sample datasets, resampling techniques have to be applied when estimating the classifier performance [95].

K-fold Cross Validation (CV) with stratified blocks was the evaluation procedure used for some experiments of the Thesis. In the CV evaluation, the data set is divided into  $K$  subsets or blocks, and a training-test evaluation is repeated  $K$  times. In each training-test evaluation, one of the  $K$  blocks is used as a test set and the remaining  $K - 1$  blocks are used together as a training set. Once every block has been used as test set, the average error across all  $K$  trials is computed. This procedure guarantees that every data sample is used in a test set exactly once, the other  $K - 1$  times being used as part of the training set. The main drawback of this method is that the training algorithm needs  $K$  runnings in order to estimate the result. On the other hand, the greater  $K$  is, the lower the variance of the resulting estimate is.

The evaluation procedure of k-Random Sampling Train-Test (kRSTT) with stratified test sets with  $k$  repetitions was also applied for some experiments. This technique consists on randomly splitting the dataset into a set of samples to train the classifier (training set) and a set of samples to test the trained classifier (test set) repeating this procedure  $k$  times. This method avoids under-estimation of the true error when the evaluation is carried out in a nested-loop that covers the feature and model selection.

### 2.4.4 Evaluation metrics

The performance of the classifiers was measured in terms of Accuracy (ACC). ACC is defined as the degree of closeness of measurements of a quantity to its actual (true) value, ( $ACC = \frac{\sum_c R_c}{N}$ , where  $C$  is the number of classes,  $R_c$  is the number of right answers to the class  $c$  and  $N$  is the total number of cases).

Another measure used in this Dissertation is the Balanced Accuracy Rate (BAR), which is the average of the success rate obtained for each class [96], ( $BAR = \frac{\sum_c \frac{R_c}{N_c}}{C}$ , where  $N_c$  is the number of cases of the class  $c$ ). BAR is useful to ensure that the sensitivities of the classes are high and balanced when the prevalence of one or more classes is unbalanced with respect to other in the dataset. BAR is a metric easily understandable in the clinical domain since it can be seen as the mean of the sensitivities of each class.

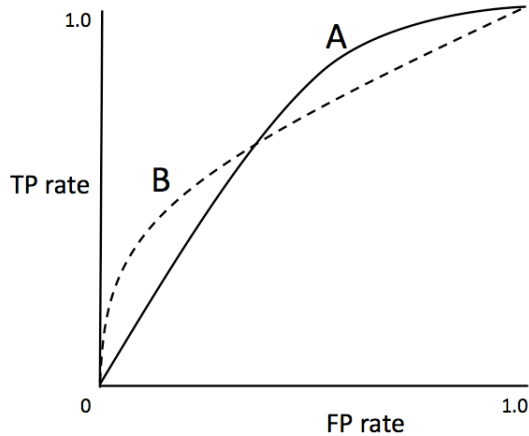


Figure 2.8: Comparison of two classifiers A and B with ROC curve when the AUC is the same for both classifiers.

Although generally known and applied in the medical context, the evaluation metric of the Area Under the Curve (AUC) in the Receiver operating characteristic curve (ROC) (based in the sensitivity and the specificity metrics) was discarded since it is not robust enough when dealing with unbalanced classes: As explained in [97], AUC does not consider cost bias. Figure 2.8 shows the example of two classifiers A and B with the same AUC value. If the classes they discriminate were balanced, there would be no objective way to decide which classifier is better. However, in an imbalanced dataset situation, classifier A should be chosen rather than B, because at higher True Positive (TP) rate region classifier A has smaller False Positive (FP) rate than classifier B. Hence, classifier A is preferred over classifier B even though they have the same AUC value. Therefore, the problem with conventional AUC is that it does not consider cost bias, because this metric sums up the areas with equal weights of 1, which is an affordable assumption only when the classes are balanced.

### 2.4.5 Bayesian inference learning

One of the objectives pursued in this Thesis is to develop a tool to compare the performance of different classifiers. The Bayesian paradigm orbitates around the Bayes' rule or theorem, which relates the conditional and priori probabilities of events  $A$  and  $B$  in the following way:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (2.2)$$

where  $P(A)$  is the prior probability, in the sense that it does not take into account any information about  $B$ .  $P(B|A)$  is the conditional probability of  $B$  given  $A$ ,  $P(B)$  is the prior of  $B$  and acts as a normalizing constant and  $P(A|B)$  is the conditional probability of  $A$  given  $B$ . It is usually called the posterior probability because it depends upon the specified value of  $B$ . Intuitively, Bayes' rule describes the way in which one's beliefs about observing  $A$  are updated by having observed  $B$ .

Let us define  $\mathbf{Z}$ , which represents the set of  $N$  data  $\{(\mathbf{x}_j, t_j)\}_1^N$ , where  $\mathbf{x}_j$  is a data vector describing the  $j^{\text{th}}$  sample and  $t_j$  is its associated label.

Then, the Bayes' rule can be used for two well-differentiated tasks in Machine Learning. In a first level of inference, known as Model Fitting, the Bayes' theorem can be used to fit the parameters of a model for a set of data  $\mathbf{Z}$ . In a second level, called Model Comparison, the Bayes' rule can be applied to compare different models in the light shed by  $\mathbf{Z}$ .

Let us define a model  $\mathcal{M}_i$  with a vector of parameters  $\theta$ . Every model is defined by its functional form,  $\mathcal{M}_i = \{f(\theta), \theta \in \Omega\}$ , where  $\Omega$  is the set of possible values of  $\theta$ , and two probability distributions: a 'priori' distribution  $P(\theta|\mathcal{M}_i)$  which states the plausible values of the parameters of the model; and the likelihood  $P(\mathbf{Z}|\theta, \mathcal{M}_i)$  which contains the information about  $\theta$  and  $\mathcal{M}_i$  derived from observation of  $\mathbf{Z}$ .

In the first level of inference, we assume that our model  $\mathcal{M}_i$  can explain the data  $\mathbf{Z}$  and use the Bayes' rule to fit its parameters  $\theta$  according to that data. That is, we obtain the posterior probability of the parameters  $\theta$ :

$$P(\theta|\mathcal{M}_i, \mathbf{Z}) = \frac{P(\theta|\mathcal{M}_i)P(\mathbf{Z}|\theta, \mathcal{M}_i)}{P(\mathbf{Z}|\mathcal{M}_i)}. \quad (2.3)$$

The normalising constant  $P(\mathbf{Z}|\mathcal{M}_i)$  is the model likelihood for  $\mathbf{Z}$ , also known as the evidence for  $\mathcal{M}_i$  [98]. The evidence is usually ignored in this inference level because it is irrelevant in the choice of  $\theta$ .

Nevertheless, the evidence becomes important in the second level of inference, where we assume that we have several models that can explain the data  $\mathbf{Z}$  and we wish to infer which model is most plausible given that data. The posterior probability of a model  $\mathcal{M}_i$  is:

$$P(\mathcal{M}_i|\mathbf{Z}) = \frac{P(\mathcal{M}_i)P(\mathbf{Z}|\mathcal{M}_i)}{P(\mathbf{Z})}, \quad (2.4)$$

where the term  $P(\mathcal{M}_i)$  is a ‘subjective’ prior distribution function over the model space.  $P(\mathbf{Z})$  is the unconditional distribution of  $\mathbf{Z}$  which is constant with respect to  $\mathcal{M}_i$ . So, if  $\mathbf{Z}$  is fixed, it can be ignored:

$$P(\mathcal{M}_i|\mathbf{Z}) \propto P(\mathcal{M}_i)P(\mathbf{Z}|\mathcal{M}_i) \quad (2.5)$$

If we have no reasons to assign differing priors  $P(\mathcal{M}_i)$  to the alternative models, models  $\mathcal{M}_i$  are ranked by evaluating the evidence [99].

If we include the parameter space  $\theta$  for model  $\mathcal{M}_i$ , the solution of  $P(\mathcal{M}_i|\mathbf{Z})$  in Equation 2.5 can be rewritten by marginalization as:

$$P(\mathcal{M}_i|\mathbf{Z}) \propto P(\mathcal{M}_i) \int P(\mathbf{Z}|\theta, \mathcal{M}_i)P(\theta|\mathcal{M}_i)d\theta \quad (2.6)$$

The calculation of this integral is analytically difficult. We could assume that  $P(\theta|\mathbf{Z}, \mathcal{M}_i)$ , the distribution of  $\theta$  of a model in the light of the data  $\mathbf{Z}$ , is approximately normal,  $N(\hat{\theta}, V)$ , where  $\hat{\theta}$  is the mode of the posterior density and  $V$  is the inverse of the Hessian of the  $-\log P(\hat{\theta}|\mathbf{Z})$  which, for a normal distribution, is the covariance matrix. Then we could apply a Laplace approximation according to the Hessian  $V^{-1}$  evaluated at  $\hat{\theta}$  to calculate the integral in Equation 2.6 [99, 100]:

$$\begin{aligned}
 P(\mathbf{Z}|\mathcal{M}_i) &\simeq \underbrace{P(\mathbf{Z}|\hat{\theta}, \mathcal{M}_i)}_{\text{Best fit likelihood}} \underbrace{P(\hat{\theta}|\mathcal{M}_i)(2\pi)^{\frac{d}{2}}|V|^{-\frac{1}{2}}}_{\text{Occam factor}} \quad (2.7) \\
 \text{Evidence} &\simeq \text{Best fit likelihood} \quad \text{Occam factor}
 \end{aligned}$$

The evidence, thus, is found by taking the best fit likelihood that the model can achieve and multiplying it by an ‘Occam factor’ [101], which is a term with magnitude less than one that penalizes  $\mathcal{M}_i$  for having the parameters  $\theta$ . The more complex the model is, the greater the penalization becomes. Figure 2.9 depicts an example of the evidence of two models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  according to the data  $\mathbf{Z}$ . We can deduce that  $\mathcal{M}_2$  is a more complex model than  $\mathcal{M}_1$  because it is capable of making a greater variety of predictions. In other words, its predictive probability  $P(\mathbf{Z}|\mathcal{M}_2)$  can predict a greater variety of data sets.  $\mathcal{M}_1$ , in contrast, is capable of making only a limited range of predictions,  $P(\mathbf{Z}|\mathcal{M}_1)$ , centered in the region  $C$ . However, in the case where the data are compatible with both models (data falling into the region  $C$ ), the simpler model  $\mathcal{M}_1$  will be a better candidate and a more probable model. The Occam factor, thus, can be defined as the ratio of the posterior accessible volume of the parameter space of  $\mathcal{M}_i$  (that is, the values the model finally chose in the light of  $\mathbf{Z}$ ) to the prior accessible volume (all the possible values of  $\theta$ ) [99].

If we suppose some other simplifications described in [100], page 64, Equation 2.7 can be simplified to

$$\log P(\mathbf{Z}|\mathcal{M}_i) \simeq \log P(\mathbf{Z}|\hat{\theta}, \mathcal{M}_i) - \frac{d}{2} \log N \quad (2.8)$$

where  $N$  is the number of samples in  $\mathbf{Z}$ ,  $\hat{\theta}$  is a maximum likelihood estimate and  $d$  is the number of free parameters in model  $\mathcal{M}_i$  [76]. Notice that Equation 2.8 is the Bayesian Information Criterion (BIC) [102] multiplied by  $-\frac{1}{2}$ . BIC is a statistical measure used as a model selection criterion. This criterion is independent of the prior of  $\theta$  and measures the parameterized model in terms of data prediction and the complexity of the model (e.g., the number of parameters) [103].

The main idea in the Bayesian framework is to assess a preference score to the alternative models  $\mathcal{M}_i$  by the evaluation of the evidence  $P(\mathbf{Z}|\mathcal{M}_i)$ . The evidence

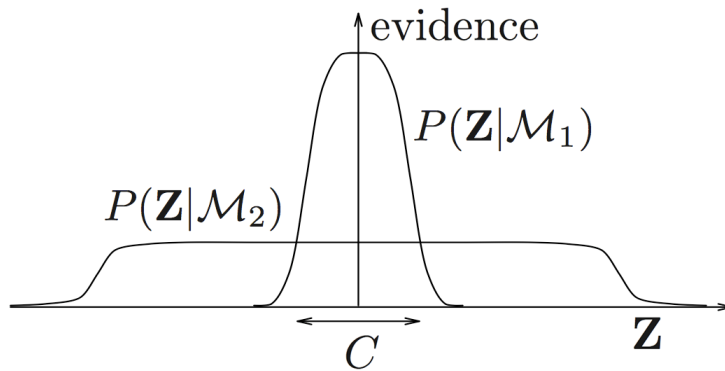


Figure 2.9: This figure, adapted from [99], shows graphically why the Bayesian inference embodies Occam's razor, that is, why complex models can be less probable. The horizontal axis represents the space of possible datasets  $\mathbf{Z}$ . The evidence of two models are shown. The evidence is the probability of the data given the model  $\mathcal{M}_i$ . In other words, how well the model  $\mathcal{M}_i$  is able to explain or predict the data. Model  $\mathcal{M}_1$  is a simple model able to make a limited range of predictions. Model  $\mathcal{M}_2$  is a more powerful model able to make a wider range of predictions. Nevertheless, the powerfulness of  $\mathcal{M}_2$  is lower than that of  $\mathcal{M}_1$  in the region  $C$ . If we suppose that equal prior probabilities have been assigned to both models, then, if the data falls in region  $C$ , the *simpler* model  $\mathcal{M}_1$  will be the *more probable* model.



can be evaluated for different modeling tasks like regression, classification or density estimation. It is also a measure that naturally embodies the principle of Occam's razor.

## 2.5 PR applied to BT diagnosis using MRS

In order to perform a non-invasive diagnosis by means of *in vivo* MRS, the use of automatic techniques based on PR can considerably complement and improve the knowledge of clinical experts.

This section gives a review of previous research on automatic classification and diagnosis of brain tumour with *in vivo* MRS applied to adult and children patients.

### 2.5.1 Survey of studies performed with MRS data from adults

During the last two decades a large amount of studies of cancer research based on PR have been reported [104], as well as brain tumour research [105, 106, 107]. Hagberg [106] reviewed the classification of brain tumour with MRS based on PR and clustering methods. A total of eight studies covered in this review were focused on brain tumour discrimination from normal tissue or other CNS diseases. All of them applied LDA or Artificial Neural Networks (ANN) to relative metabolite concentrations or PCA transformations. Performance of these classification tools was evaluated by Leave-one-out (LOO) CV.

There are several publications describing results for classification of brain tumour based on the MR spectra collected in the framework of the EC project INTERPRET [105]. In [108, 109], linear and kernel-based methods on MRS features extracted by automatic procedures were applied. Some studies [15] were focused on the automatic classification using Long-TE  $^1\text{H}$  MRS, whereas other studies were focused on the use of Short-TE  $^1\text{H}$  MRS [105, 108, 110, 111]. An extensive benchmark study of quantitation and PR-based feature extraction methods in combination with learning strategies was published in [107] to discriminate between recurrent and non-recurrent brain tumours using Long-TE MRS. They reported that PR methods perform at least as well as the ones based on manual quantitation obtaining 5%-10% higher accuracy when automatic techniques were applied. Classifiers for *in vivo* Short-TE spectra and

Magnetic Resonance Spectroscopic Imaging (MRSI) obtained good results using Least-Squares Support Vector Machines (LS-SVM) [112] in [15, 111].

Further investigation with LS-SVM and feature extraction methods for MRS was carried out by Luts *et al.* [84] during a subsequent EC project called eTUMOUR. Luts *et al.* studied the effect of automating the feature extraction step in the preprocessing protocol of Short-TE spectra. They conclude that the use of automated PR methods like ICA, Relief-F or a regularization technique applied to the full spectra produced the highest overall performances in adults since these methods allowed a good handling of high dimensional data. In addition, Postma *et al.* [113] evaluated the automatic feature selection techniques with MR spectroscopy data from brain tumour cases in order to see whether the features selected correspond to what was expected from the literature and clinical experience. They conclude that the automatic feature selection was a useful tool to obtain relevant features but, nevertheless, evaluation of the obtained features is still needed.

A recent study reported results on the application of Discrete Wavelet Transform procedure as a pre-process step to the MRS data [114]. In this work, Arizmendi *et al.* concluded that the use of this procedure, followed by a standard feature selection or feature extraction technique, can yield good results in terms of diagnostic binary classification of brain tumours with Bayesian Neural Networks. In [115], Majòs *et al.* compared the Long-TE and Short-TE discrimination capacity in clinical use. Taking this study as starting point, in [116], García-Gómez *et al.* combined the use of Short-TE and Long-TE MRS to improve the automatic brain tumour classification of the most prevalent brain tumours in adults [14] (low grade glial tumours, aggressive tumours –glioblastoma and metastasis-, and meningiomas). This study revealed that the combination of Short-TE and Long-TE in a classifier performed better in several instances than the use of one TE  $^1\text{H}$  MRS only. Further investigation in the interpretation, visualization and classification of MRS brain tumour data is reported in [117]: several feature selection methods are explored for dimensionality reduction in Short-TE, Long-TE and combination of both TEs, demonstrating that feature selection methods can improve the classification performance over the use of the full set of features. They also confirmed the result reported in [116], showing that the use of combined TE information can help to achieve better classification performance. Finally, they contribute with a linear dimensionality-reduction technique for visualizing the data corresponding to the selected frequencies. Ortega-Martorell *et al.* [118] reported the use of NMF methods as an effective unsupervised feature extraction technique for Short-TE, Long-TE and combina-

tion of both TEs MRS data from adults brain tumours. Vellido *et al.* recently reported in [119] a simple yet powerful method to discriminate glioblastomas from metastatic brain tumour with SV MRS. From 109 patients and an independent test set of 40 patients, they obtained a simple formula based on a linear combination from three frequencies from Short-TE and Long-TE that achieved an AUC of 0.91.

In the framework of the eTUMOUR project, some contributions have also been made with the development of PR-based models for a multiproject-multicenter evaluation of brain tumour classifiers based on MRS. In this work [83], García-Gómez coordinated the efforts of the three universities in charge of developing predicting tools within eTUMOUR. A multicenter multiproject database was used to produce the classifiers. This approach consisted on performing a multiproject evaluation based on training classifiers with the SV  $^1\text{H}$  MRS data of a multicenter project, INTERPRET [105, 3, 120], and evaluating the results with the MRS data acquired in a subsequent multicenter project, eTUMOUR [5]. This study obtained good classification results with Short-TE MR spectra. Besides, the obtained predictive models revealed as a powerful tool to detect outlier or misclassified brain tumour spectra.

Several recent works confirm MRS as a non-invasive useful tool to assess brain tumour diagnosis. Fellows *et al.* demonstrated in [121] that, in patients undergoing stereotactic biopsy, the combination of neuroradiological and spectroscopic evaluation allow good diagnosis of glioblastoma with such an accuracy that could replace the need for biopsy. In [16], Opstad *et al.* reported that, for homogeneous-appearing tumors, significant correlations were found between *in vivo* and *ex vivo*  $^1\text{H}$  MRS concentrations of several metabolites: Cr, mI, total Cho, and the approximately 1.3 and 0.9 ppm lipids). They concluded that *ex vivo* astrocytoma biopsy HR-MAS  $^1\text{H}$  spectra have similar metabolic profiles to that obtained *in vivo* and therefore detailed *ex vivo* characterization of glioma biopsies can directly relate to the original tumour. In [122], Vellido *et al.* developed a method to overcome the problems derived from the high dimensionality of the MRS data. This method encompassed a nonlinear dimensionality reduction, followed by an outlier detection and expert opinion. Afterwards, a feature selection and automatic classification process was applied. The method achieved a classification performance improvement due to the effective outlier removal. Cruz-Barbosa and Vellido explored in [123] the use of semi-supervised variants of Generative Topographic Mapping, a manifold learning model, as a set of learning techniques able to address the diagnostic discrimination of different brain tumour pathologies and establish the outcome of aggressive tumour

types. The performance reported was comparable to the results achieved by Semi-Supervised Support Vector Machines (SVM) and Laplacian Eigenmaps.

PR techniques can be used to predict prognosis aspects rather than diagnosis. In [124], Majós *et al.* assessed the predictive value of MRS regarding survival in high grade astrocytomas. By dividing a set of 187 high grade astrocytomas retrospectively diagnosed into 2 groups according to survival better or worse than the median, they discovered four significant points in Short-TE and Long-TE that allowed the stratification into 2 prognostic groups. This first approach revealed the value of MRS as a predicting tool for the length of survival in patients with high grade astrocytoma that can be used to stratify prognostic groups.

New lines of research related with application of PR applied to MRS are currently being explored. A study to assess whether PR-based classifiers trained with 1.5T MRS spectra are able to correctly classify spectra acquired with 3T was performed by Fuster-Garcia *et al.* [125]. They obtained classifiers trained with 1.5T samples that showed similar accuracy for independent test sets of 1.5T and 3T. In addition, they reported non-significant differences with most metabolite ratios and spectral pattern. Thus, these results encourage the use of existing classifiers based on 1.5T datasets for diagnosis with 3T. This is highly convenient since there exists large 1.5T databases compiled throughout several years and the prediction models based on 1.5T acquisitions can be used for diagnosis of cases acquired with 3T instruments. In addition, the result reported in [125] can also be applied to the work presented by Tortajada *et al.* [126], where they presented an incremental learning algorithm and successfully applied it to several diagnosis classification problems using MR spectra. Tortajada *et al.* realized that in a clinical or research setting, the gathering, pre-process, and validation of samples is expensive and time-consuming. Thus, they proposed an incremental learning approach in order to build an initial classifier with a smaller number of samples and update it incrementally when new data are collected. The incremental learning algorithm for Gaussian Discriminant Analysis (iGDA) they introduced did not require access to the previously used data and was able to include new classes that were not in the original analysis. This can allow the customization of the models to the distribution of data at a particular clinical site.

### 2.5.2 Survey of studies performed with MRS data from children

During the last decades several clinical studies have revealed the use of Short-TE MRS as a useful tool for children with nervous system diseases [33, 127] and, more specifically, in tumour diagnosis and treatment in children [128, 64, 85]; Characterizing a brain mass as neoplasm, following treatment response of primary neoplasm, differentiating residual or recurrent primary neoplasm from postsurgical changes and identifying inactive neoplasms or in remission are some of the applications the use of MRS can improve.

Wang *et al.* [44] quantified the visible metabolites in 35 Long-TE MR spectra of primary pediatric cerebellar tumours. Quantification of the NAA/Cho and Cr/Cho ratios was performed from a semiautomatic pre-processing methodology applied to the spectra. Discriminant analysis was applied to differentiate the three major tumour types in children (low grade astrocytoma, ependymoma and PNET) giving a sensitivity/specificity values of 0.91/0.84, 0.75/0.92, and 0.82/0.89, respectively, based on the data of the study. They obtained an accuracy of 0.85 discriminating the three tumour types using the metabolites ratios of NAA/Cho and Cr/Cho.

The use of ANN has also been applied in pediatric brain tumour research in [129]. In this study, Arle *et al.* obtained data from 33 children presenting Posterior Fossa (PF) tumours. A computer-based Multilayer Perceptron (MLP) was developed to combine MRS data (ratios of NAA, Cho, and Cr) with 10 characteristics of tumour tissue obtained from MR image findings, as well as tumour size and the patient's age and sex. The neural network trained with the three spectroscopy ratios alone had a predictive ability of only 58% when discriminating among three categories (PNET, astrocytoma, or ependymoma/other). Another MLP trained with data including MR imaging characteristics, age, sex, and tumor size, obtained an accuracy improved to 72%, consistent with the predictions of the neuroradiologist who was using the same information. A network that used all of the data (imaging data, spectroscopic data and a limited clinical information) was able to identify 95% of the tumours correctly.

Schneider *et al.* [130] calculated the apparent diffusion coefficient (ADC) and quantified the metabolites signals from diffusion-weighted imaging (DWI) and MRS in differentiating medulloblastoma, ependymoma, pilocytic astrocytoma, and infiltrating glioma in a pediatric population of 17 children. Combination of ADC values and metabolites allowed discrimination between the four tumour

groups with a predictive value of 1. Linear discriminant analysis using DWI and MRS using water as internal reference, fully discriminated the four most frequent posterior fossa tumours in children.

Davies *et al.* [131] performed univariate metabolite comparisons and multivariate classifiers to 35 SV Short-TE spectra from children with untreated cerebellar tumours (18 medulloblastomas, 12 pilocytic astrocytomas and five ependymomas). Spectra were analysed using LCMoDel<sup>TM</sup> to yield metabolite profiles, and key metabolite assignments were verified by comparison with HR-MAS NMR of representative tumour biopsy samples. Successful classification was achieved for glial-cell (astrocytoma + ependymoma) versus non-glial-cell (medulloblastoma) tumours, with an accuracy of 0.95. The accuracy was 0.93 in the discrimination of astrocytoma vs medulloblastoma and also in the discrimination of astrocytoma vs medulloblastoma vs ependymoma.

Hao *et al.* [132] applied ICA in order to determine automatically the metabolite signals which make up MR spectra. The two common techniques for applying ICA, blind source separation (BSS) and feature extraction (FE) were examined in this study using simulated data and the findings confirmed on patient data. Short-TE, low and high field (1.5 and 3 T) *in vivo* brain tumour MR spectra of childhood astrocytoma, ependymoma and medulloblastoma were generated by using a quantum mechanical simulator with ten metabolite and lipid components. ICA of simulated data showed that individual metabolite components can be extracted from a set of MRS data. The BSS method generated independent components with a closer correlation to the original metabolite and lipid components than the FE method when the number of spectra in the dataset was small. When the methods were applied to the patient dataset (117 patients, Short-TE, 1.5T), results were consistent with the synthesized experiments.

Porto *et al.* [45] tried to differentiate between WHO grade I pilocytic astrocytoma and diffuse, fibrillary WHO grade II astrocytoma in children by using normalized concentrations of total Cho and total Cr. To do so, they counted with 16 children with histologically proven astrocytomas (11 children with pilocytic astrocytoma and 5 children with diffuse astrocytoma). They reported that normalized total Cho did not show any statistically significant difference between the two groups. However, they detected a strong trend toward higher values of normalized total Cr in the diffuse astrocytomas. They concluded that choline as a single parameter is not reliable in the differential diagnosis of low-grade astrocytomas in children and suggest that total Cr concentrations combined with lactate could be helpful in the differential diagnosis of pilocytic

astrocytomas and diffuse astrocytomas in children.

To our knowledge, most of the research carried out on MRS of brain tumours in children are based on relatively small sample populations and/or gathered from a unique clinical institution [44, 45, 129, 130, 131, 132]. This is due to low incidence of paediatric brain tumours and the difficulties of coordinating multi-centre studies. In [132], MRS data from 117 patients was available. Nevertheless, all the cases were coming from the same institution. MRS data from 35 patients was applied in [44] as well as in [131]. The study of Arle *et al.* [129] was performed with the data of 33 different patients. Regarding Schneider *et al.* [130], they only counted with 17 patients. In [45], 16 children with astrocytomas were studied. Although these studies showed good classification results when discriminating the three main brain tumour classes in children (ependymoma, medulloblastoma and astrocytoma), it is necessary a multi-centre evaluation to demonstrate the reproducibility of these performances.

## 2.6 DSSs in BT Research

A DSS is a computer-based system that aids in the process of decision making [133].

With increasing quantities of clinical data being collected in hospitals, the practice of evidence-based medicine is becoming more and more feasible. The use of clinical DSSs is based on the goal of making the hospital data repositories, and the information contained therein, available to physicians at the point of care. Clinical DSSs are *active knowledge systems which use two or more items of patient data to generate case-specific advice* [134]. Clinical DSSs facilitate access to high-quality medical practices, which contributes to better health care [135].

The expected benefits of using a clinical DSS are an increase on the quality of treatment by providing automated alerts and consistency checks; an increase in treatment efficacy by providing best-practice guidelines; an increase in knowledge by providing information when and where it is needed, and a cost reduction by eliminating redundant tests [136, 137]. The potential benefits of clinical DSSs are summarized by Coiera and Sintchenko in three major items [138, 139]: improved patient safety, improved quality of care, and improved efficiency in health care delivery.

Clinical DSSs based on Pattern Recognition (PR) have been widely accepted in medical applications due to their capability for flexibility, accuracy and in-

interpretability [104]. Studies on the evaluation of the use of clinical DSSs indicate a positive effect on clinical practice [140, 141, 142] and quality of primary care [143].

Several examples of clinical DSSs related to the diagnosis of brain tumours studied in this Thesis are:

- The SV INTERPRET GUI<sup>12</sup> based on an LDA latent space projection. The SV INTERPRET GUI provides easy access to a database of spectra, images and clinical information from 304 validated cases of human brain tumours. It is designed to allow the display of classification plots which are useful for automating the classification of tumour spectra [105]. Recently, the INTERPRET DSS version 3.0 has been released [144]. Version 3.0 improves the initial analysis capabilities of the first version by incorporating an embedded database, user accounts, more diagnostic discrimination capabilities and the possibility to analyse data acquired under additional data acquisition conditions. Other improvements include a customisable Graphical User Interface (GUI). A picture of the INTERPRET GUI can be seen in Figure 2.10. Most diagnostic problems have been addressed through a pattern-recognition based approach, in which classifiers based on LDA were trained and tested.
- The Computer Aided Diagnosis System (CADs) of eTUMOUR, which offers advanced functionalities for patient management, as well as aided diagnosis on several questions with capability to accept SV Long-TE and Short-TE spectra. It is designed to provide the user with a percentage of confidence associated to the answer it provides when supporting diagnosis. The classifiers are built from the most important repository of brain tumours in Europe.
- The HEALTHAGENTS dDSS. This DSS differs from the previous in its architecture definition. It offered a web accessible DSS distributed along a network of clinical and producer nodes<sup>13</sup>. Conceived as a collaborative network where users (clinical nodes) request services (support in the diagnosis) to knowledge suppliers (producer nodes, constituted by PR research groups), this dDSS is able to address more than 10 different questions, sort the results attending to a ranking model and show latent space projections of the LDA classifiers it contains [145, 146].

---

<sup>12</sup>[http://azizu.uab.es/INTERPRET/sv\\_tutorial/index.php](http://azizu.uab.es/INTERPRET/sv_tutorial/index.php)

<sup>13</sup>Currently not accessible



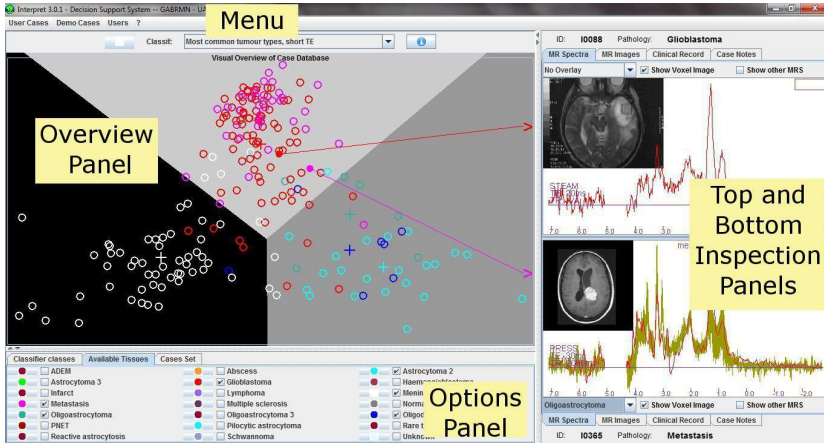


Figure 2.10: INTERPRET DSS version 3.0 GUI. Image obtained from [144].

- **SpectraClassifier.** Is a solution for designing and implementing MRS-based classifiers [147]. It incorporates several feature selection tools and implements the Fisher LDA classification method and several evaluation techniques (see Figure 2.11). It is able to read low resolution in-vivo MRS (single-voxel and multi-voxel) and HR-MAS, processed with existing tools (jMRUI, INTERPRET, 3DiCSI or TopSpin). In addition, to facilitate exchanging data between applications, a standard format capable of storing all the information needed for a dataset was developed. Data from the INTERPRET project was used with the purpose of bug-testing and validating the methods.
- **CURIAM.** Designed as a general purpose DSS for clinical environments, CURIAM BT offers a powerful classification framework able to use most kind of biomedical data and allows a specialized support on the diagnosis of brain tumours based on the grounds of MRS of adults and pediatric patients [148, 149]. It is able to automatically or semi-automatically pre-process MRS from three major manufacturers and implements several PR techniques for brain tumour classification. It has been designed as a tool for offering a tailored decision support, taking into account the patient profile in order to give the user the most adequate answer to each evaluated case (see Figure 2.12).

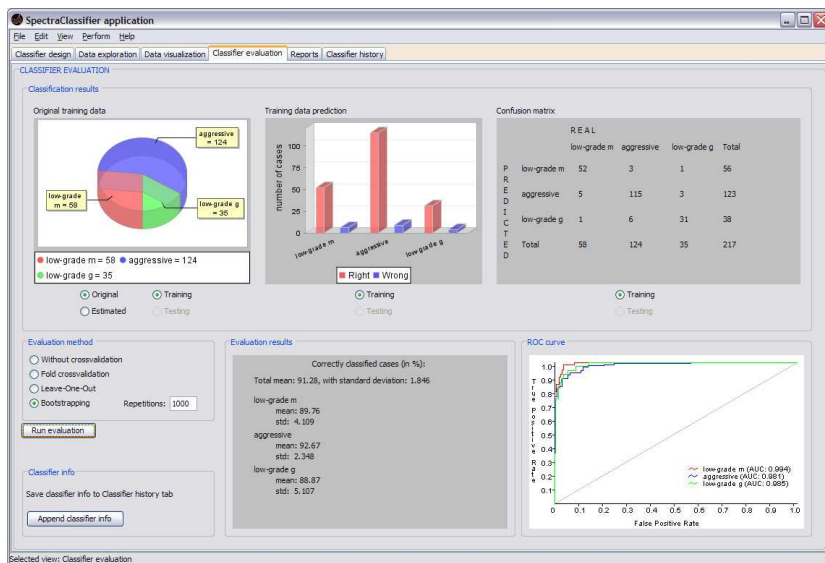


Figure 2.11: SpectraClassifier tool GUI. It is not a DSS in itself, but allows to easily create PR-based classifiers with MRS data.

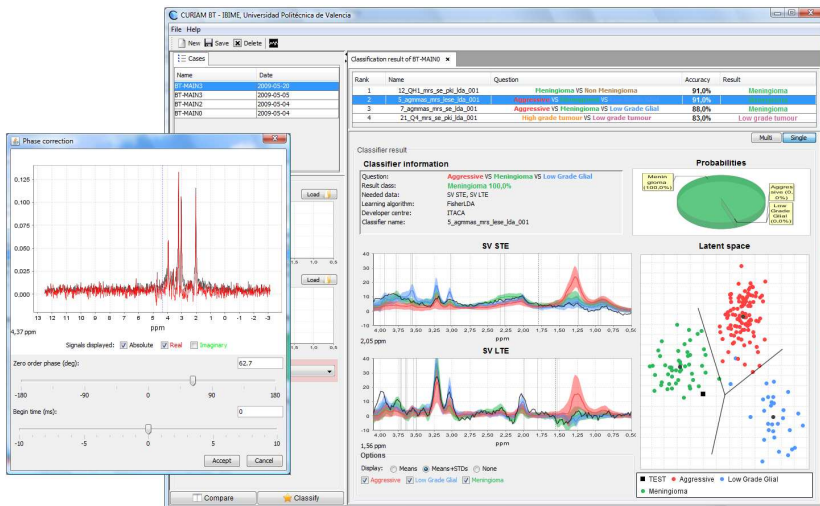


Figure 2.12: CURIAM BT GUI. The image shows the classification result screen of the CURIAM BT and a phase correction screen for MRS processing.

---

## Bibliography

- [1] BTS, *The Brain Tumor Society*, (Accessed: 23rd April 2012). <http://www.braintumor.org>.
- [2] D. N. Louis, H. Ohgaki, O. D. Wiestler, and W. K. Cavenee, *WHO classification of tumours of the central nervous system*. Lyon, France: IARC Press, 2007.
- [3] INTERPRET Consortium, “INTERPRET.” Web site, 1999-2001. IST-1999-10310, EC, <http://azizu.uab.es/INTERPRET> (Accessed: 22nd April 2012).
- [4] M. Mier, “Healthagents, annex 1. agent-based distributed decision support system for brain tumour diagnosis and prognosis,” tech. rep., HealthAgents consortium, IST-2004-27214, VI framework programme, EC, 2005.
- [5] eTUMOUR Consortium, “eTumour: Web accessible MR Decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data.” Web site. FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC, <http://www.etumour.net> (Accessed: 22nd April 2012; at writing time it was temporarily unavailable).
- [6] CBTRUS, “2010 CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2006,” tech. rep., Central Brain Tumor Registry of the United States, 2010.
- [7] I. F. Pollack, “Pediatric brain tumors,” *Seminars in Surgical Oncology*, vol. 16, pp. 73–90, 1999.
- [8] L. Rorke, *Pediatric neurosurgery: surgery of the developing nervous system*, ch. Introductory survey of pediatric brain tumors, pp. 335–337. W.B. Saunders Company, second edition ed., 1989.
- [9] J. Farwell, D. G.J., and J. Flannery, “Central nervous system tumors in children,” *Cancer*, vol. 40, pp. 3123–3132, 1977.
- [10] M. Salzman, “The morbidity and mortality of brain tumors. a perspective on recent advances in therapy,” *Neurologic Clinics*, vol. 3, pp. 229–257, 1985.

- [11] I. F. Pollack, “Brain tumors in children [review],” *New England Journal of Medicine*, vol. 331, pp. 1501–1507, 1994.
- [12] A. P. Lin, T. T. Tran, and B. D. Ross, “Impact of evidence-based medicine on magnetic resonance spectroscopy,” *NMR in Biomedicine*, vol. 19, pp. 476–483, Jun 2006.
- [13] F. A. Howe and K. S. Opstad, “<sup>1</sup>H MR spectroscopy of brain tumours and masses,” *NMR in Biomedicine*, vol. 16, pp. 123–131, May 2003.
- [14] M. Julià-Sapé, D. Acosta, C. Majós, A. Moreno-Torres, P. Wesseling, J. J. Acebes, J. R. Griffiths, and C. Arús, “Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database,” *Journal of Neurosurgery*, vol. 105, pp. 6–14, Jul 2006.
- [15] L. Lukas, A. Devos, J. A. K. Suykens, L. Vanhamme, F. A. Howe, C. Majós, A. Moreno-Torres, M. V. D. Graaf, A. R. Tate, C. Arús, and S. V. Huffel, “Brain tumor classification based on long echo proton MRS signals,” *Artificial Intelligence in Medicine*, vol. 31, pp. 73–89, 2004.
- [16] K. S. Opstad, A. J. Wright, B. A. Bell, J. R. Griffiths, and F. A. Howe, “Correlations between in vivo <sup>1</sup>H MRS and ex vivo <sup>1</sup>H HRMAS metabolite measurements in adult human gliomas,” *Journal of Magnetic Resonance Imaging*, vol. 31, no. 2, pp. 289–297, 2010.
- [17] D. P. Soares and M. Law, “Magnetic resonance spectroscopy of the brain: review of metabolites and clinical applications,” *Clinical Radiology*, vol. 64, pp. 12–21, Jan. 2009.
- [18] D. Galanaud, F. Nicoli, O. Chinot, S. Confort-Gouny, D. Figarella-Branger, P. Roche, S. Fuentes, Y. Le Fur, J.-P. Ranjeva, and P. J. Cozzzone, “Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy,” *Magnetic Resonance in Medicine*, vol. 55, pp. 1236–1245, Jun 2006.
- [19] C. Majós, C. Aguilera, J. Alonso, M. Julià-Sapé, S. Castaner, J. Sánchez, A. Samitier, A. León, A. Rovira, and C. Arús, “Proton MR spectroscopy improves discrimination between tumor and pseudotumoral lesion in solid brain masses,” *American Journal of Neuroradiology*, vol. 30, no. 3, pp. 544–551, 2009.

- 
- [20] P. Georgiadis, S. Kostopoulos, D. Cavouras, D. Glotsos, I. Kalatzis, K. Sifaki, M. Malamas, E. Solomou, and G. Nikiforidis, “Quantitative combination of volumetric MR imaging and MR spectroscopy data for the discrimination of meningiomas from metastatic brain tumors by means of pattern recognition,” *Magnetic Resonance Imaging*, vol. 29, no. 4, pp. 525 – 535, 2011.
- [21] G. J. Postma, J. Luts, A. J. Idema, M. Julià-Sapé, A. Moreno-Torres, W. Gajewicz, J. A. Suykens, A. Heerschap, S. V. Huffel, and L. M. Buydens, “On the relevance of automatically selected single-voxel MRS and multimodal MRI and MRSI features for brain tumour differentiation,” *Computers in Biology and Medicine*, vol. 41, no. 2, pp. 87 – 97, 2011.
- [22] A. Dodd, “An introduction to the theory of nuclear magnetic resonance,” tech. rep., University of Melbourne, 2003.
- [23] E. R. Danielsen, *Magnetic Resonance Spectroscopy Diagnosis of Neurological Diseases. 2nd ed.* London: Informa Healthcar, 2007.
- [24] I. Barba, A. Moreno, I. Martinez-Perez, A. R. Tate, M. E. Cabanas, M. Baquero, A. Capdevila, and C. Arús, “Magnetic resonance spectroscopy of brain hemangiopericytomas: high myoinositol concentrations and discrimination from meningiomas,” *Journal of Neurosurgery*, vol. 94, pp. 55–60, Jan 2001.
- [25] M. Kaminogo, H. Ishimaru, M. Morikawa, M. Ochi, R. Ushijima, M. Tani, Y. Matsuo, J. Kawakubo, and S. Shibata, “Diagnostic potential of short echo time MR spectroscopy of gliomas with single-voxel and point-resolved spatially localised proton spectroscopy of brain,” *Neuroradiology*, vol. 43, pp. 353–363, May 2001.
- [26] R. J. Gillies and D. L. Morse, “In vivo magnetic resonance spectroscopy in cancer,” *Annual Review of Biomedical Engineering*, vol. 7, pp. 287–326, 2005.
- [27] H. Knizley, “The enzymatic synthesis of n-acetyl-l-aspartic acid by a water-insoluble preparation of a cat brain acetone powder,” *Journal of Biological Chemistry*, vol. 242, no. 1, pp. 4619–4622, 1967.
- [28] J. Urenjak, S. Williams, D. Gadian, and M. Noble, “Proton nuclear magnetic resonance spectroscopy unambiguously identifies different neural cell types,” *The Journal of Neuroscience*, vol. 13, pp. 981–989, 1993.

- [29] J. Frahm, H. Bruhn, M. Gyngell, K. Merboldt, W. Haenicke, and R. Sauter, "Localized proton NMR spectroscopy in different regions of the human brain *in vivo*. relaxation times and concentration of cerebral metabolites," *Magnetic Resonance in Medicine*, vol. 11, pp. 47–67, 1989.
- [30] S. Moyher, S. Nelson, L. Wald, R. Henry, J. Kurhanewicz, and D. Vigneron, "High spatial resolution of MRS and segmented MRI to study NAA in cortical grey matter and white matter of the human brain," *Abstracts of the International Society Magnetic Resonance in Medicine*, vol. 1, p. 332, 1995.
- [31] M. A. McLean, F. G. Woermann, G. J. Barker, and J. S. Duncan, "Quantitative analysis of short echo time 1H-MRSI of cerebral gray and white matter," *Magnetic Resonance in Medicine*, vol. 44, no. 3, pp. 401–411, 2000.
- [32] R. Kreis, T. Ernst, and B. D. Ross, "Development of the human brain: *In vivo* quantification of metabolite and water content with proton magnetic resonance spectroscopy," *Magnetic Resonance in Medicine*, vol. 30, pp. 424–437, 1993.
- [33] W. Grodd, I. Kraegelo-Mann, U. Klose, and R. Sauter, "Metabolic and destructive brain disorders in children: Findings with localized proton MR spectroscopy," *Radiology*, vol. 181, pp. 173–181, 1991.
- [34] M. van der Knaap, J. van der Grond, van Rijen P.C., J. Faber, J. Valk, and K. Willemse, "Age-dependent changes in localized proton and phosphorus MR spectroscopy of the brain," *Radiology*, vol. 176, pp. 509–515, 1990.
- [35] O. Henriksen, P. Gideon, B. Sperling, S. Olsen, H. Joergensen, and P. Arlien-Soeborg, "Cerebral lactate production and blood flow in acute stroke," *Journal of Magnetic Resonance Imaging*, vol. 2, pp. 511–517, 1992.
- [36] H. Bruhn, J. Frahm, M. Gyngell, K. Merboldt, W. Haenicke, and R. Sauter, "Cerebral metabolism in man after acute stroke: new observations using localized proton NMR spectroscopy," *Magnetic Resonance in Medicine*, vol. 9, pp. 126–131, 1989.
- [37] N. De Stefano, S. Narayanan, P. Matthews, and et al, "Proton MR spectroscopy to assess axonal damage in multiple sclerosis and other white

- matter disorders,” *Journal of Neurovirology*, vol. 6, no. 2, pp. 121–129, 2000.
- [38] K. Imamura, “Proton MR spectroscopy of the brain with a focus in chemical issues,” *Magnetic Resonance in Medicine*, vol. 2, no. 3, pp. 117–132, 2003.
- [39] M. van der Graaf, “*In vivo* magnetic resonance spectroscopy: basic methodology and clinical applications.,” *European Biophysics Journal*, vol. 39, pp. 527–40, Mar. 2010.
- [40] S. Venkatesh, R. Gupta, L. Pal, N. Husain, and M. Husain, “Spectroscopic increase in choline signal is a nonspecific marker for differentiation of infective/inflammatory from neoplastic lesions of the brain,” *Journal of Magnetic Resonance Imaging*, vol. 14, pp. 8–15, 2001.
- [41] P. Demaerel, K. Johannik, P. Van Hecke, and et al, “Localized Proton MR  $^1\text{H}$  NMR spectroscopy in fifty cases of newly diagnosed intracranial tumors,” *Journal of Computer Assisted Tomography*, vol. 15, no. 1, pp. 67–76, 1991.
- [42] H. Kugel, W. Heindel, R. Ernestus, J. Bunke, R. du Mesnil, and G. Friedman, “Human brain tumors: spectral patterns detected with localized  $^1\text{H}$  MR spectroscopy,” *Radiology*, vol. 183, pp. 701–709, 1992.
- [43] M. Harada, M. Tanouchi, H. Nishitani, H. Miyoshi, K. Bandou, and S. Kannuki, “Non-invasive characterization of brain tumor by *in-vivo* proton magnetic resonance spectroscopy,” *Japanese Journal of Cancer Research*, vol. 86, pp. 329–332, 1995.
- [44] Z. Wang, L. N. Sutton, A. Cnaan, J. C. Haselgrove, L. Rorke, H. Zhao, L. T. Bilaniuk, and R. A. Zimmerman, “Proton MR spectroscopy of pediatric cerebellar tumors,” *American Journal of Neuroradiology*, vol. 16, pp. 1821–1833, 1995.
- [45] L. Porto, M. Kieslich, K. Franz, T. Lehrbecher, U. Pilatus, and E. Hattingen, “Proton magnetic resonance spectroscopic imaging in pediatric low-grade gliomas,” *Brain Tumor Pathology*, vol. 27, pp. 65–70, 2010.
- [46] N. Sailasuta, T. Ernst, and L. Chang, “Regional variations and the effects of age and gender on glutamate concentrations in the human brain,” *Magnetic Resonance Imaging*, vol. 26, pp. 667–675, June 2008.



- [47] D. J. Manton, M. Lowry, S. J. Blackband, and A. Horsman, "Determination of proton metabolite concentrations and relaxation parameters in normal human brain and intracranial tumours," *NMR in Biomedicine*, vol. 8, pp. 104–112, 1995.
- [48] F. A. Howe, S. J. Barton, S. A. Cudlip, M. Stubbs, D. E. Saunders, M. Murphy, P. Wilkins, K. S. Opstad, V. L. Doyle, M. A. McLean, B. A. Bell, and J. R. Griffiths, "Metabolic profiles of human brain tumors using quantitative in vivo  $^1\text{H}$  magnetic resonance spectroscopy," *Magnetic Resonance in Medicine*, vol. 49, pp. 223–232, Feb 2003.
- [49] P. Matthews, F. Andermann, K. Silver, G. Karpati, and D. Arnolg, "Proton MR spectroscopic characterization of differences in regional brain metabolic abnormalities in mitochondrial encephalomyopathies," *Neurology*, vol. 42, pp. 2484–2490, 1993.
- [50] K. Chang, I. Song, S. Kim, M. Han, H. KIM, S. Seong, H. Jung, and M. Han, "*In vivo* single-voxel proton MR spectroscopy in intracranial cystic masses," *Magnetic Resonance in Medicine*, vol. 32, pp. 549–555, 1998.
- [51] Y. Rosen and L. R. E. Soares, "Recent advances in magnetic resonance neurospectroscopy," *Neurotherapeutics*, vol. 4, no. 3, pp. 330–345, 2007.
- [52] B. Ross and T. Michaelis, "Clinical applications of magnetic resonance spectroscopy," *Magnetic Resonance Quality*, vol. 10, pp. 191–247, 1994.
- [53] M. Murphy, A. Loosemore, C. A. G., F. A. Howe, A. T. Tate, S. A. Cudlip, P. R. Wilkins, J. R. Griffiths, and B. A. Bell, "The contribution of  $^1\text{H}$  MRS to clinical brain tumour diagnosis," *Journal of Neurosurgery*, vol. 16, pp. 329–334, 2002.
- [54] M. Castillo, J. K. Smith, and L. Kwock, "Correlation of *myo-inositol* levels and grading of cerebral astrocytomas," *American Journal in Neuroradiology*, vol. 21, pp. 1645–1649, 2000.
- [55] I. Mader, W. Roser, G. Hagberg, M. Schneider, R. Sauter, J. Seelig, E. W. Radue, and W. Steinbrich, "Proton chemical shift imaging, metabolic maps, and single voxel spectroscopy of glial brain tumours," *MAGMA: Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 4, pp. 139–150, 1996.

- [56] L. L. Cheng, C. I. W., D. N. Louis, and R. G. Gonzalez, "Correlation of high-resolution magic angle spinning proton magnetic resonance spectroscopy with histopathology of intact human brain tumor specimens," *Cancer Research*, vol. 58, pp. 1925–1832, 1998.
- [57] N. P. Davies, M. Wilson, K. Natarajan, Y. Sun, L. MacPherson, M.-a. Brundler, T. N. Arvanitis, R. G. Grundy, and a. C. Peet, "Non-invasive detection of glycine as a biomarker of malignancy in childhood brain tumours using in-vivo 1H MRS at 1.5 tesla confirmed by ex-vivo high-resolution magic-angle spinning NMR," *NMR in biomedicine*, vol. 23, pp. 80–7, Jan. 2010.
- [58] M. F. Olive, "Interactions between taurine and ethanol in the central nervous system," *Amino Acids*, vol. 23, no. 4, pp. 345–57, 2002.
- [59] T. C. Birdsall, "Therapeutic applications of taurine," *Alternative medicine review: a journal of clinical therapeutic*, vol. 3, pp. 128–36, Apr. 1998.
- [60] R. Leon, H. Wu, Y. Jin, J. Wei, C. Buddhala, H. Prentice, and W. J-Y, "Protective function of taurine in glutamate-induced apoptosis in cultured neurons," *Journal of Neuroscience Research*, vol. 87, no. 5, pp. 1185–1194, 2009.
- [61] A. El Idrissi, J. Messing, J. Scalia, and E. Trenkner, "Prevention of epileptic seizures by taurine," *Advances in Experimental Medicine and Biology*, no. 526, pp. 515–25, 2003.
- [62] K. Opstad, B. Bell, J. Griffiths, and F. Howe, "Taurine: a potential marker of apoptosis in gliomas," *British journal of cancer*, vol. 100, pp. 789–94, Mar. 2009.
- [63] A. Moreno-Torres, I. Martínez-Pérez, M. Baquero, J. Campistol, A. Capdevila, C. Arús, and J. Pujol, "Taurine detection by proton magnetic resonance spectroscopy in medulloblastoma: contribution to noninvasive differential diagnosis with cerebellar astrocytoma," *Neurosurgery*, vol. 4, no. 55, pp. 824–829, 2004.
- [64] A. Panigrahy, M. D. Krieger, I. González-Gómez, X. Liu, J. G. McComb, J. L. Finlay, M. D. Nelson, F. Gilles, and S. Blüml, "Quantitative short echo time <sup>1</sup>H-MR spectroscopy of untreated pediatric brain tumors : Pre-operative diagnosis and characterization," *American Journal of Neuroradiology*, vol. 27, pp. 560–572, 2006.

- [65] A. C. Peet, N. P. Davies, L. Ridley, M.-A. Brundler, D. Kombogiorgas, S. Lateef, K. Natarajan, S. Sgouros, L. MacPherson, and R. G. Grundy, "Magnetic resonance spectroscopy suggests key differences in the metastatic behaviour of medulloblastoma.," *European journal of cancer (Oxford, England : 1990)*, vol. 43, pp. 1037–44, Apr. 2007.
- [66] D. R., R. Gupta, H. Poptani, R. Roy, S. Sharma, and M. Husain, "Role of *in vivo* proton magnetic resonance spectroscopy in the diagnosis and management of brain abscesses," *Neurosurgery*, vol. 42, pp. 37–43, 1998.
- [67] T. Musacchio, M. Toniutti, R. Kautz, and V. P. Torchilin, "1H NMR Detection of Mobile Lipids as a Marker for Apoptosis: The Case of Anti-cancer Drug-Loaded Liposomes and Polymeric Micelles," *Molecular Pharmaceutics*, vol. 6, no. 6, pp. 1876–1882, 2009. PMID: 19737025.
- [68] A. Kuesel, S. Donnelly, W. Halliday, G. Sutherland, and I. Smith, "Mobile lipids and metabolic heterogeneity of brain tumors as detectable by *ex vivo* 1H MR spectroscopy," *NMR in Biomedicine*, vol. 7, pp. 172–180, 1994.
- [69] C. Rémy, N. Fouilhé, I. Barba, E. Sam-Laï, H. Lahrech, M.-G. Cucurella, M. Izquierdo, A. Moreno, A. Ziegler, R. Massarelli, M. Décorps, and C. Arús, "Evidence that mobile lipids detected in rat brain glioma by 1H nuclear magnetic resonance correspond to lipid droplets," *Cancer Research*, vol. 57, pp. 407–414, 1997.
- [70] I. Martinez-Pérez, A. Moreno, I. Barba, A. Capdevila, and C. Arús, "Large lipid droplets observed by electron microscopy in six human brain tumors with lipid 1H MRS *in vivo* pattern," in *Abstracts of the International Society for Magnetic Resonance in Medicine, New York*, p. 976, 1996.
- [71] H. Poptani, R. Gupta, R. Roy, R. Padney, V. Jain, and D. Chhabra, "Characterization of intracranial mass lesions with *in vivo* proton MR spectroscopy," *American Journal of Neuroradiology*, vol. 16, pp. 1593–1603, 1995.
- [72] A. Kuesel, S. Donnelly, W. Halliday, G. Sutherland, and I. Smith, "Mobile lipids and metabolic heterogeneity of brain tumors as detectable by *ex vivo* 1H MR spectroscopy," *NMR in Biomedicine*, vol. 7, pp. 172–180, 1994.
- [73] C. M. Bishop, *Pattern Recognition and Machine Learning*. Information Science and Statistics, New York: Springer-Verlag, 2006.

- 
- [74] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: John Wiley & Sons, 2nd. ed., 2000.
- [75] B. Vandeginste, D. Massart, L. Buydens, S. de Jong, P. Lewi, and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*. Science Publishers, 1998.
- [76] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. Springer, August 2001.
- [77] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. John Wiley and Sons, inc., 2001.
- [78] C. J. Burges, “Geometric methods for feature extraction and dimensional reduction: A guided tour,” tech. rep., Microsoft Research, University of Toronto, 2004.
- [79] K. Fukunaga, *Introduction to statistical pattern recognition (2nd ed.)*. San Diego, CA, USA: Academic Press Professional, Inc., 1990.
- [80] A. M. Martinez and C. K. Avinash, “PCA versus LDA,” *IEEE Transactions on Pattern Analysis and Machine Learning*, no. 2, pp. 228–233, 2001.
- [81] P. Comon, “Independent component analysis, a new concept?,” *Signal Processing*, vol. 36, no. 3, pp. 287 – 314, 1994.
- [82] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 1, pp. 788 – 791, 1999.
- [83] J. M. García-Gómez, J. Luts, M. Julià-Sapé, P. Krooshof, S. Tortajada, J. Vicente, W. Melssen, E. Fuster-Garcia, I. Olier, G. Postma, D. Monleón, A. Moreno-Torres, J. Pujol, A. P. Candiota, M. C. Martínez-Bisbal, J. Suykens, L. Buydens, B. Celda, S. Van Huffel, C. Arús, and M. Robles, “Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy,” *Magnetic Resonance Materials Physics, Biology and Medicine*, vol. 22, no. 1, pp. 5–18, 2009.
- [84] J. Luts, J. Pouillet, J. M. García-Gómez, A. Heerschap, M. Robles, J. A. K. Suykens, and S. Van Huffel, “Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra,” *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 288–298, 2008.

- [85] A. C. Peet, S. Lateef, L. MacPherson, K. Natarajan, S. Sgouros, and R. G. Grundy, "Short echo time  $^1\text{H}$  magnetic resonance spectroscopy of childhood brain tumours.," *Child's Nervous System*, vol. 23, no. 2, pp. 163–9, 2007.
- [86] J.-B. Poulet, D. M. Sima, and S. V. Huffel, "MRS signal quantitation: A review of time- and frequency-domain methods," *Journal of Magnetic Resonance*, vol. 195, no. 2, pp. 134 – 144, 2008.
- [87] E. Cabanes, S. Confort-Gouny, Y. L. Fur, G. Simond, and P. Cozzone, "Optimization of Residual Water Signal Removal by HLSVD on Simulated Short Echo Time Proton MR Spectra of the Human Brain," *Journal of Magnetic Resonance*, vol. 150, no. 2, pp. 116 – 125, 2001.
- [88] L. Vanhamme, A. van den Boogaart, and S. V. Huffel, "Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge," *Journal of Magnetic Resonance*, vol. 129, no. 1, pp. 35 – 43, 1997.
- [89] M. Kanowski, J. Kaufmann, J. Braun, J. Bernarding, and C. Tempelmann, "Quantitation of simulated short echo time  $^1\text{H}$  human brain spectra by LCMoel and AMARES," *Magnetic Resonance in Medicine*, vol. 51, no. 5, pp. 904–912, 2004.
- [90] A. U. Provencher, "Estimation of metabolite concentrations from localized *in vivo* proton NMR spectra," *Magnetic Resonance in Medicine*, vol. 30, pp. 672–679, 1993.
- [91] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. v. Ormondt, and D. Graveron-Demilly, "Time-domain semi-parametric estimation based on a metabolite basis set," *NMR in Biomedicine*, vol. 18, no. 1, pp. 1–13, 2005.
- [92] J.-B. Poulet, D. M. Sima, A. W. Simonetti, B. De Neuter, L. Vanhamme, P. Lemmerling, and S. Van Huffel, "An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES," *NMR in Biomedicine*, vol. 20, no. 5, pp. 493–504, 2007.
- [93] M. Wilson, G. Reynolds, R. a. Kauppinen, T. N. Arvanitis, and A. C. Peet, "A constrained least-squares approach to the automated quantitation of *in vivo*  $^1\text{H}$  magnetic resonance spectroscopy data," *Magnetic Resonance in Medicine*, vol. 65, pp. 1–12, Jan. 2011.

- 
- [94] Stephen Provencher, “LCModel manual.” Website, 2011. <http://s-provencher.com/pages/lcm-manual.shtml> (Accessed: 24th April 2012).
- [95] J. K. Martin and D. S. Hirschberg, “Small sample statistics for classification error rates I: Error rate measurements,” Tech. Rep. ICS-TR-96-22, 1996.
- [96] I. Guyon, A. R. S. A. Alamdari, G. Dror, and J. M. Buhmann, “Performance prediction challenge,” in *IJCNN '06. International Joint Conference on Neural Networks*, pp. 1649–1656, 2006.
- [97] C. G. Weng and J. Poon, “A new evaluation measure for imbalanced datasets,” in *Proceedings of the 7th Australian Data Mining Conference (AusDM 2008). Conferences in Research and Practice in Information Technology, Glenelg, Australia*, pp. 27–32, 2008.
- [98] D. J. C. MacKay, “Bayesian interpolation,” *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.
- [99] D. J. C. Mackay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [100] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge University Press, January 1996.
- [101] S. F. Gull, “Developments in maximum-entropy data analysis,” in *Maximum Entropy and Bayesian Methods* (J. Skilling, ed.), pp. 53–71, Kluwer Academic, Dordrecht, 1989.
- [102] H. Jeffreys, *Theory of Probability*. Oxford University Press, third ed., 1961.
- [103] A. R. Liddle, “Information criteria for astrophysical model selection,” *Monthly Notices of the Royal Astronomical Society*, vol. 377, pp. 74–79, 2007.
- [104] P. J. Lisboa and A. F. G. Taktak, “The use of artificial neural networks in decision support in cancer: a systematic review,” *Neural Networks*, vol. 19, no. 4, pp. 408–415, 2006.
- [105] A. R. Tate, J. Underwood, D. M. Acosta, M. Julià-Sapé, C. Majós, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, M. E.

- Cabañas, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arús, “Development of a decision support system for diagnosis and grading of brain tumours using *in vivo* magnetic resonance single voxel spectra,” *NMR in Biomedicine*, vol. 19, no. 4, pp. 411–434, 2006.
- [106] G. Hagberg, “From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods,” *NMR in Biomedicine*, vol. 11, no. 4-5, pp. 148–156, 1998.
- [107] B. H. Menze, M. P. Lichy, P. Bachert, B. M. Kelm, H.-P. Schlemmer, and F. A. Hamprecht, “Optimal classification of long echo time *in vivo* magnetic resonance spectra in the detection of recurrent brain tumors,” *NMR in Biomedicine*, vol. 19, pp. 599–609, Aug 2006.
- [108] A. W. Simonetti, W. J. Melssen, F. Szabo de Edelenyi, J. J. A. van Asten, A. Heerschap, and L. M. C. Buydens, “Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification,” *NMR in Biomedicine*, vol. 18, pp. 34–43, Feb 2005.
- [109] Y. Huang, P. J. G. Lisboa, and W. El-Deredy, “Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection,” *Statistics in Medicine*, vol. 22, pp. 147–164, Jan 2003.
- [110] A. R. Tate, C. Majos, A. Moreno, F. A. Howe, J. R. Griffiths, and C. Arús, “Automated classification of short echo time *in vivo* 1H brain tumor spectra: a multicenter study,” *Magnetic Resonance in Medicine*, vol. 49, pp. 29–36, Jan 2003.
- [111] A. Devos, L. Lukas, J. A. K. Suykens, L. Vanhamme, A. R. Tate, F. A. Howe, C. Majos, A. Moreno-Torres, M. van der Graaf, C. Arús, and S. Van Huffel, “Classification of brain tumours using short echo time 1H MR spectra,” *Journal of Magnetic Resonance*, vol. 170, pp. 164–175, Sep 2004.
- [112] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [113] G. J. Postma, J. Luts, A. J. Idema, M. Julià-Sapé, A. A. Moreno-Torres, W. Gajewicz, J. A. K. Suykens, A. Heerschap, S. Van Huffel, and L. M. C. Buydens, “On the relevance of automatically selected single-voxel MRS

- and multimodal MRI and MRSI features for brain tumour differentiation,” *Computers in Biology and Medicine*, vol. 41, no. 2, pp. 87–97, 2011.
- [114] C. Arizmendi, A. Vellido, and E. Romero, “Classification of human brain tumours from MRS data using Discrete Wavelet Transform and Bayesian Neural Networks,” *Expert Systems with Applications*, vol. 39, no. 1, pp. 5223–5232, 2012.
- [115] C. Majos, M. Julià-Sapé, J. Alonso, M. Serrallonga, C. Aguilera, J. J. Acebes, C. Arús, and J. Gili, “Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE,” *American Journal of Neuroradiology*, vol. 25, pp. 1696–1704, Nov 2004.
- [116] J. M. García-Gómez, S. Tortajada, C. Vidal, M. Julià-Sapé, J. Luts, À. Moreno-Torres, S. Van Huffel, C. Arús, and M. Robles, “The effect of combining two echo times in automatic brain tumor classification by MRS,” *NMR in Biomedicine*, vol. 21, no. 10, pp. 1112–1125, 2008.
- [117] F. F. González-Navarro, L. A. B.-M. noz, E. Romero, A. Vellido, M. Julià-Sapé, and C. Arús, “Feature and model selection with discriminatory visualization for diagnostic classification of brain tumors,” *Neurocomputing*, vol. 73, no. 4-6, pp. 622 – 632, 2010.
- [118] S. Ortega-Martorell, P. J. G. Lisboa, A. Vellido, M. Julià-Sapé, and C. Arús, “Non-negative matrix factorisation methods for the spectral decomposition of MRS data from human brain tumours,” *BMC Bioinformatics*, vol. 13, no. 1, p. 38, 2012.
- [119] A. Vellido, E. Romero, M. Julià-Sapé, C. Majós, A. Moreno-Torres, J. Pujol, and C. Arús, “Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel  $^1\text{H}$  MRS,” *NMR in Biomedicine*, vol. 25, no. 1, pp. 819–828, 2012.
- [120] M. Julià-Sapé, D. Acosta, M. Mier, C. Arús, and D. Watson, “A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 19, pp. 22–33, Feb 2006.
- [121] G. A. Fellows, A. J. Wright, N. A. Sibtain, P. Rich, K. S. Opstad, D. J. McIntyre, B. A. Bell, J. R. Griffiths, and F. A. Howe, “Combined use of neuroradiology and  $^1\text{H}$ -MR spectroscopy may provide an intervention



- limiting diagnosis of glioblastoma multiforme,” *Journal of Magnetic Resonance Imaging*, vol. 32, no. 5, pp. 1038–1044, 2010.
- [122] A. Vellido, E. Romero, F. F. González-Navarro, L. A. Belanche-Muñoz, M. Julià-Sapé, and C. Arús, “Outlier exploration and diagnostic classification of a multi-centre  $^1\text{H}$ -MRS brain tumour database,” *Neurocomputing*, vol. 72, no. 13–15, pp. 3085 – 3097, 2009.
- [123] R. Cruz-Barbosa and A. Vellido, “Semi-supervised analysis of human brain tumours from partially labeled MRS information, using manifold learning models,” *International Journal of Neural Systems*, vol. 21, no. 1, pp. 17–29, 2011.
- [124] C. Majós, J. Bruna, M. Julià-Sapé, M. Cos, A. Camins, M. Gil, J. Acebes, C. Aguilera, and C. Arús, “Proton MR Spectroscopy Provides Relevant Prognostic Information in High-Grade Astrocytomas,” *American Journal of Neuroradiology*, vol. 32, no. 1, pp. 74–80, 2011.
- [125] E. Fuster-Garcia, C. Navarro, J. Vicente, S. Tortajada, J. García-Gómez, C. Sáez, J. Calvar, J. Griffiths, M. Julià-Sapé, F. Howe, J. Pujol, A. Peet, A. Heerschap, A. Moreno-Torres, M. Martínez-Bisbal, B. Martínez-Granados, P. Wesseling, W. Semmler, J. Capellades, C. Majós, A. Alberich-Bayarri, A. Capdevila, D. Monleón, L. Martí-Bonmatí, C. Arús, B. Celda, and M. Robles, “Compatibility between 3T  $^1\text{H}$  SV-MRS data and automatic brain tumour diagnosis support systems based on databases of 1.5T  $^1\text{H}$  SV-MRS spectra,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 24, no. 1, pp. 35–42, 2011.
- [126] S. Tortajada, E. Fuster-Garcia, J. Vicente, P. Wesseling, F. a. Howe, M. Julià-Sapé, A.-P. Candiota, D. Monleón, A. Moreno-Torres, J. Pujol, J. R. Griffiths, A. Wright, A. C. Peet, M. C. Martínez-Bisbal, B. Celda, C. Arús, M. Robles, and J. M. García-Gómez, “Incremental Gaussian Discriminant Analysis based on Graybill and Deal weighted combination of estimators for brain tumour diagnosis,” *Journal of biomedical informatics*, vol. 44, no. 4, pp. 677–87, 2011.
- [127] M. Shevell, S. Ashwal, and E. Novotny, *Seminars in Pediatric Neurology*, vol. 6, ch. Proton Magnetic Resonance Spectroscopy: Clinical applications in children with nervous system diseases, pp. 68–77. W.B. Saunders Company, 1999.

- 
- [128] S. E. Byrd, T. Tomita, P. S. Palka, C. F. Darling, J. P. Norfray, and J. Fan, "Magnetic resonance spectroscopy (MRS) in the evaluation of pediatric brain tumors, Part II: clinical analysis," *Journal of the National Medical Association*, vol. 88, pp. 717–723, 1996.
- [129] J. E. Arle, C. Morriss, Z. J. Wang, R. A. Zimmerman, P. G. Phillips, and L. N. Sutton, "Prediction of posterior fossa tumor type in children by means of magnetic resonance image properties, spectroscopy, and neural networks," *Journal of Neurosurgery*, vol. 86, pp. 755–761, 1997.
- [130] J. F. Schneider, S. Confort-Gouny, A. Viola, Y. Le Fur, P. Viout, M. Ben-nathan, F. Chapon, D. Figarella-Branger, P. Cozzone, and N. Girard, "Multiparametric differentiation of posterior fossa tumors in children using diffusion-weighted imaging and short echo-time  $^1\text{H}$ -MR Spectroscopy," *Journal of Magnetic Resonance Imaging*, vol. 26, no. 1, pp. 1390–1398, 2007.
- [131] N. P. Davies, M. Wilson, L. M. Harris, K. Natarajan, S. Lateef, L. Macpherson, S. Sgouros, R. G. Grundy, T. N. Arvanitis, and A. C. Peet, "Identification and characterisation of childhood cerebellar tumours by in vivo proton MRS," *NMR in Biomedicine*, vol. 21, no. 8, pp. 908–918, 2008.
- [132] J. Hao, X. Zou, M. P. Wilson, N. P. Davies, Y. Sun, A. C. Peet, and T. N. Arvanitis, "A comparative study of feature extraction and blind source separation of independent component analysis (ICA) on childhood brain tumour  $^1\text{H}$  magnetic resonance spectra," *NMR in Biomedicine*, vol. 22, no. 8, pp. 809–818, 2009.
- [133] P. Finlay, "Introducing decision support systems (2nd edition)," *The Journal of the Operational Research Society*, vol. 46, no. 10, pp. 1282–1283, 1995.
- [134] J. Wyatt and D. Spiegelhalter, "Field trials of medical decision-aids: potential problems and solutions," *Proceedings of Annual Symposium on Computer Application in Medical Care*, pp. 3–7, 1991.
- [135] D. Brand, "Electronic decision support for Australia's health sector," tech. rep., National Electronic Support Taskforce, 2002.
- [136] I. Sim, P. Gorman, R. a. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann, and P. C. Tang, "Clinical decision support systems for the practice of

- evidence-based medicine,” *Journal of the American Medical Informatics Association*, vol. 8, no. 6, pp. 527–34, 2001.
- [137] E. S. Berner, ed., *Clinical Decision Support Systems: Theory and Practice*. Springer, 1998.
- [138] E. Coiera, *Guide to Health Informatics*, ch. Clinical Decision Support Systems. Hodder Arnold, 2003.
- [139] V. Sintchenko, “Improving patient safety and quality of health care through pathology informatics,” in *Transforming Australia’s Health System/Proceedings of the National Health Information Summit*, pp. 156–158, 2004.
- [140] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, “Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success,” *BMJ*, vol. 330, no. 7494, p. 765, 2005.
- [141] A. X. Garg, N. K. J. Adhikari, H. McDonald, M. P. Rosas-Arellano, P. J. Devereaux, J. Beyene, J. Sam, and R. B. Haynes, “Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review,” *Journal of the American Medical Association*, vol. 293, no. 10, pp. 1223–1238, 2005.
- [142] B. Kaplan, “Evaluating informatics applications—clinical decision support systems literature review,” *International journal of medical informatics*, vol. 64, no. 1, pp. 15–37, 2001.
- [143] B. C. Delaney, D. A. Fitzmaurice, A. Riaz, and F. D. Hobbs, “Can computerised decision support systems deliver improved quality in primary care?. Interview by Abi Berger,” *BMJ*, vol. 319, no. 7220, p. 1281, 1999.
- [144] A. Pérez-Ruiz, M. Julià-Sapé, G. Mercadal, I. Olier, C. Majós, and C. Arús, “The INTERPRET Decision-Support System version 3.0 for evaluation of Magnetic Resonance Spectroscopy data from human brain tumours and other abnormal brain masses.,” *BMC Bioinformatics*, vol. 11, no. 1, pp. 581–581, 2010.
- [145] H. González-Vélez, M. Mier, M. Julià-Sapé, T. N. Arvanitis, J. M. García-Gómez, M. Robles, P. H. Lewis, S. Dasmahapatra, D. Dupplaw, A. Peet, C. Arús, B. Celda, S. V. Huffel, and M. Lluch i Ariet, “HealthAgents:

- Distributed multi-agent brain tumour diagnosis and prognosis,” *Journal of Applied Intelligence*, p. 191, 2007.
- [146] C. Sáez, J. M. García-Gómez, J. Vicente, S. Tortajada, J. Luts, D. Dupplaw, S. V. Huffel, and M. Robles, “A generic and extensible automatic classification framework applied to brain tumour diagnosis in HealthAgents,” *The Knowledge Engineering Review*, vol. 26, no. Special Issue 03, pp. 283–301, 2011.
- [147] S. Ortega-Martorell, I. Olier, M. Julia-Sape, and C. Arus, “SpectraClassifier 1.0: a user friendly, automated MRS-based classifier-development system,” *BMC Bioinformatics*, vol. 11, no. 1, p. 106, 2010.
- [148] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, M. Esparza, A. Navarro, E. Fuster-Garcia, M. Robles, L. Martí-Bonmatí, and C. Arús, “A generic decision support system featuring an assembled view of predictive models for magnetic resonance and clinical data,” in *Book of abstracts of the ESMRMB 2008: 25th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Valencia, Spain*, p. 483, Springer-Verlag, Oct. 2008.
- [149] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, E. Fuster-Garcia, M. Esparza, A. Navarro, and M. Robles, “Curiam BT 1.0, Decision Support System for Brain Tumour Diagnosis,” in *Book of Abstracts of the ESMRMB: 26th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Antalya, Turkey*, vol. 22, p. 538, Springer Berlin / Heidelberg, May 2009.

## Chapter 3

# Brain tumour classification in childhood and adulthood

*“We do not choose to be born. We do not choose our parents. We do not choose our historical epoch, the country of our birth, or the immediate circumstances of our upbringing. We do not, most of us, choose to die; nor do we choose the time and conditions of our death. But within this realm of choicelessness, we do choose how we live.”*

*Somebody*

Several studies confirm that the nature of child BT may be totally different from adults. This chapter focuses on developing pediatric classifiers with high accuracy that overcome the biochemical differences found in children and adult BT. Automatic BT classification of MRS data from two multi-centre projects (INTERPRET and eTUMOUR) was performed to develop classifiers specialized in child patients and adult patients. Our evaluation demonstrates that children classifiers tested with an adult test set show a poor performance and vice-versa. A filter based on the normal probability density function of the training dataset’s age can successfully overcome these differences and obtain a classifier that globally behaves as predicted by the training performance.

*The core of this chapter has been published in the proceedings of the 17th Scientific Meeting of the International Society for Magnetic Resonance in Medicine Conference [1]. The pre-processing protocol applied to the spectra is described in Section 3.2 and was published in [2]. Sections 3.4.1 and 3.4.2 include an analysis of spectral features highlighting the metabolic differences observed in child and adult brain tumours described in Chapter 1.*

### 3.1 Introduction

Although limited studies have been reported in children, they confirm that childhood BT differ from those arising in adulthood in their relative incidences, histological features, sites of origin and responsiveness to therapy [3, 4, 5, 6, 7]. Section 2.5.2 of Chapter 2 showed that PR applied to *in-vivo*  $^1\text{H}$  MRS is becoming an important tool for additional accurate non-invasive technique for classification and characterization of pediatric tumours. However, the results reported were based on relatively small sample populations and/or gathered from a unique clinical institution.

In the context of the multi-centre projects eTUMOUR and INTERPRET, 489 (93 children, 396 adults) SV  $^1\text{H}$  MRS at 1.5T histopathologically diagnosed brain tumour cases from 10 different centers in Europe and South America have been gathered. This amount of multi-centre data offered us the opportunity to demonstrate the reproducibility of the studies reporting the metabolic differences in the nature of children and adult BT.

In this chapter we applied PR techniques to pediatric and adults BT with a multi-centre MRS database. With the obtained classifiers, a study of the compatibility was performed. This study consisted of determining whether a classifier aimed to adult BT diagnosis could correctly predict pediatric cases and vice-versa. Finally, a design of a filter was proposed. This filter was based on the patient age to internally decide when the classifier specialized with child patients should be used.

Table 3.1: Number of samples divided by aggressiveness (bold letters) and associated tumour types. Notice that there are 2 adult cases labeled as OTH and considered aggressive (corresponding to one anaplastic meningioma and one germinoma). The remaining OTH cases, both in children and adults, belonged to non-aggressive tumour types.

Aggressiveness	Tumour type	Child	Adult
<b>Aggressive</b>		<b>36</b>	<b>225</b>
	HG glial	12	156
	MED	24	2
	MET	0	65
	OTH	0	2
<b>Non-aggressive</b>		<b>57</b>	<b>171</b>
	LG glial	48	77
	MEN	3	69
	OTH	6	20
	NOR	0	5
TOTAL CASES		93	396

## 3.2 Materials

### 3.2.1 Cases

489 SV  $^1\text{H}$  MRS at 1.5T (TE of 20-32 ms, Recycling time (TR) of 1600-2020 ms) histopathologically diagnosed brain tumor cases acquired by ten institutions in the framework of the European projects INTERPRET (2000-2002) [8, 9, 10] and eTUMOUR (2004-2009) [11] has been used for this study. Signal quality and histopathological diagnoses of the cases were validated by the INTERPRET Clinical Data Validation Committee [8], the eTUMOUR Clinical Validation Committee, and expert spectroscopists [12].

The 489 spectra belonged to 93 children and 396 adults. The classes considered in this study were based on the histological classification of the CNS tumours as described by the WHO Classification [13]. Pathology distribution in children was: 60 glials (12 High grade (HG) and 48 Low grade (LG)), 24 Medulloblastoma (MED), 3 Low-grade meningioma (MEN) and 6 cases belonging to other less frequent tumour types (OTH). Distribution of pathology in adults was: 233 glials (156 HG, 77 LG), 2 MED, 69 MEN, 65 Metastases (MET), 5 Normal brain spectra (NOR) and 22 OTH. Mean age in children and adults was  $11.5 \pm 8$  and  $54 \pm 13$  respectively. Summary of the data available is gathered in Table 3.1.

## Data acquisition

In the INTERPRET project, a protocol was defined to guarantee the compatibility of the signals acquired at different hospitals [10, 12]. The subsequent multicenter project eTUMOUR has benefited from the data and expertise gathered by INTERPRET. The INTERPRET acquisition protocols for clinical, radiological, and histopathological data were extended to *ex-vivo* transcriptomic (DNA microarrays) and metabolomic (HR-MAS) data acquisition in eTUMOUR [14].

276 SV  $^1\text{H}$  MR at 1.5T spectra from the INTERPRET database [12] were included. These signals were acquired with Siemens, General Electric (GE), and Philips instruments by six international centres. Data were acquired avoiding areas of necrosis or cysts and with minimum contamination from the surrounding non-tumoral tissue. The acquisition protocols included PRESS or STEAM sequences, with TR between 1600 and 2020 ms, TE of 20 or 30-32 ms (Short-TE), spectral width of 1000 - 2500Hz, and 512, 1024 or 2048 data-points.

Slightly differing were the spectral parameters of the 213 SV  $^1\text{H}$  MR spectra at 1.5T acquired from the 10 international centres participating in eTUMOUR: PRESS, PROBE or STEAM sequences, with TR between 1500 and 2000 ms, TE of 20 or 30 ms, spectral width of 500 - 2500Hz, and 512, 1024 or 2048 data-points.

## Pre-processing

Each signal was pre-processed according to the INTERPRET protocol. A fully automatic pre-processing pipeline was available for the INTERPRET signals. Besides, a semi-automated pipeline was defined for some new file formats of the eTUMOUR spectra from GE and Siemens manufacturers. The semi-automatic pipeline was designed to ensure compatibility of its output with the automatic one.

### *Automatic pipeline*

The steps of the automatic pre-processing pipeline were:

1. The Klose algorithm was applied to the water-suppressed Free induction decay (FID) of each case for eddy current correction: The quality of volume-localized MRS is affected by distortions, called eddy currents, in the gradient directions. Eddy currents distortions can be corrected using



affine registration to a reference volume. The Klose algorithm performs such correction by using the water signal FID of a case as a reference for water-suppressed FID of that case [15].

2. Suppression of the residual water signal is a prerequisite to an accurate quantification of cerebral metabolites. The HLSVD time-domain selective filtering using 10 singular values and a water region of [4.33, 5.07] ppm was applied for residual water resonance suppression. This algorithm allows complete elimination of the residual water signal [16].
3. Apodization was applied to the resulting signal in order to improve the signal-to-noise ratio (SNR) of the spectrum. Apodization consists in a point-by-point multiplication of the time domain signal by a decaying exponential function. By doing so, the initial part of the FID, where the SNR is higher, gains more weight compared to the latter part. The apodization was performed with a Lorentzian function of 1Hz of damping.
4. A zero filling was applied to the frequency domain spectrum in order to interpolate the number of points to the maximum number used in the acquisition protocols (2048).
5. Afterwards, the baseline offset is estimated as the mean value of the region [11, 9] and [-2, -1] ppm and subsequently subtracted from the spectrum.
6. The region [4.33, 5.07] ppm was set to zero to avoid the influence of possible residuals of the water peak.
7. A Unit length normalization (ULN) is applied to the spectra using the region of interest [7.1, -2.7] ppm. The ULN allows a normalization of all the spectra under pre-processing to the Euclidean norm.
8. A frequency alignment check was then performed by referencing the ppm-axis to the total Cr at 3.03 ppm; the Cho at 3.21 ppm or the ML at 1.29 ppm, depending on the tumour pattern and on the SNR. Each peak is searched in a range close to the expected position and, if none of them are found, the spectrum no shift correction is applied.
9. Finally, a last step of reduction of the number of points of the spectra to 190 points was performed for the defined region of [0.5, 4.1] ppm. This region corresponds to the range where the main metabolites are found in the brain.

### *Semiautomatic pipeline*

Some of the eTUMOUR samples were pre-processed by a semi-automatic pipeline that was partially based on the Java magnetic resonance user interface (jMRUI) software package [17]. The modifications of the semiautomatic pipeline with respect to the automatic one were: (1) The phase of the water-suppressed FID was mainly corrected with the reference water. (2) Additional manual phase correction of zero order (non-frequency dependent) and first order (proportional to the frequency) was performed after the zero-filling step. (3) The Hankel singular value decomposition algorithm of the jMRUI implementation was used to remove the residual water. The filter was parameterized as in the automatic pre-processing.

As a result, a pre-processing pipeline based on different software implementations but compatible with the automatic one was set up, and comparable signals for developing the PR models were obtained.

## **3.3 Methods**

### **3.3.1 Feature extraction**

Two different PR automatic feature extraction methods were applied in this study: PI and PCA. Both techniques were chosen because of the good results reported in [2]. PI technique was applied after the (semi)automatic processing pipeline to estimate the relative concentration of metabolites integrating the frequency signal around the resonance frequencies. This technique has been successfully applied in previous studies with MRS from adults [2, 18]. Additionally, Principal Components Analysis (PCA) [19, 20] was applied in the region of [4.1, 0.5] ppm interval, taken as input for the classification models the principal components that explained the 85% of the total variance. PCA applied to the PI was also applied as dimensionality reduction technique to the inputs of the classifiers.

### **3.3.2 Classification and evaluation**

The classification problem of discriminating between two levels of aggressiveness in BT (aggressive or non-aggressive) is addressed for child and adult cases.

LDA and KNN techniques were applied to generate the predictive models. Classifiers were evaluated following this methodology: A corpus of 60 children cases and 300 adult cases was available for training purposes. An initial evaluation estimated on a 10-Folded CV was calculated for each classifier. Then, children classifiers were tested with two independent test sets: one containing 33 children cases and other with 396 adult cases. Analogously for the adult classifiers, an independent test set of 96 adult cases and a second with the total set of 93 childhood cases were applied.

The performance of the classifiers was measured with the ACC and the BAR both in the training stage and when evaluated with the independent test sets.

### **3.3.3 Age-filtered classifier**

The pediatric classifier that showed best performance according to the CV evaluation was selected. Analogously, the adult classifier with best CV performance was also selected. With them, a filter based on the normal probability density of the ages of their respective training cases was applied in order to decide what classifier should be used for each case to be evaluated. This ‘filtered classifier’ was tested with both independent test set of 33 children cases and 96 adult cases. ACC and BAR were calculated for this combined independent test set.

## **3.4 Results**

### **3.4.1 Spectral features**

A detailed view of the mean profile of the different tumour types is shown in Figure 3.1. Visual discrimination of the different tumour types can be done attending to the metabolite concentrations. The mean spectra of every tumour type differs visually to the mean spectra of normal adult brain tumour (NOR), with its characteristically prominent NAA peak, a neuronal marker located at 2.00 ppm [6, 7]. In LG glials, NAA and ML at 0.9 ppm and 1.3 ppm shows more variability in children than in adults. HG glials in children shows a great elevation of Choline (Cho) peak at 3.2 ppm with respect to the Cr peak at 3.0 ppm, whereas in adults the elevation of Mobile lipids (ML) is the most prominent feature. MEN shows the characteristic plateau around the NAA resonance both in children and adults although the concentration of NAA may

be slightly greater in children. MED mean spectra is also similar in children and adults, with two relevant aspects: the great elevation of Cho compared to the Cr intensity; and a lower amount of NAA compared to the LG glials. The adult MET mean spectra shows a similar profile to that of HG glials in adults, with a prominent ML peak.

When joining the tumour types attending to their aggressiveness, several visual key features prevail in the mean spectra obtained (Figure 3.2): aggressive tumours in children display an increase in Cho peak with respect to the Cr peak. In contrast, in the mean spectra of aggressive tumours in adults the main feature is the presence of an elevated ML peak. Mean spectra of non-aggressive tumours shows a similar profile in children and adults with an slightly superior concentration of NAA observed in children.

### 3.4.2 Metabolite comparison

The estimated metabolite concentrations in aggressive and non-aggressive BT in children and adults can be seen in Table 3.2. The Kruskal-Wallis test for the analysis of the variance ( $\alpha = 0.05$ ) has been applied to determine the significant differences in metabolite concentrations in pediatric and adult aggressive and non-aggressive tumours. Bonferroni test correction (with  $\alpha = 0.01$ ) was applied to avoid bias in the multiple comparison.

Statistical significant differences were observed between child and adult aggressive BT (significance of  $p < 0.01$  in every concentration peak but the Ala peak), whilst non-aggressive BT in children and adults did not reflect remarkable differences in metabolite concentration, with exception of the estimated concentration of the Cho peak ( $p < 0.05$ ) and the peaks related with ML (Macromolecules and lipids components at 0.9 ppm (MMLip09), with  $p < 0.05$  and Macromolecules and lipids components at 1.3 ppm (MMLip13)+Lac, with  $p < 0.01$ ).

In adults, all the metabolite concentrations had significant differences between aggressive and non-aggressive tumours ( $p < 0.01$ , Table 3.2-Bottom). In contrast, only the differences of five metabolite concentrations were statistically significant ( $p < 0.01$ ) in pediatric aggressive and non-aggressive tumours: mI, Glx, Cho, NAA and Ala.

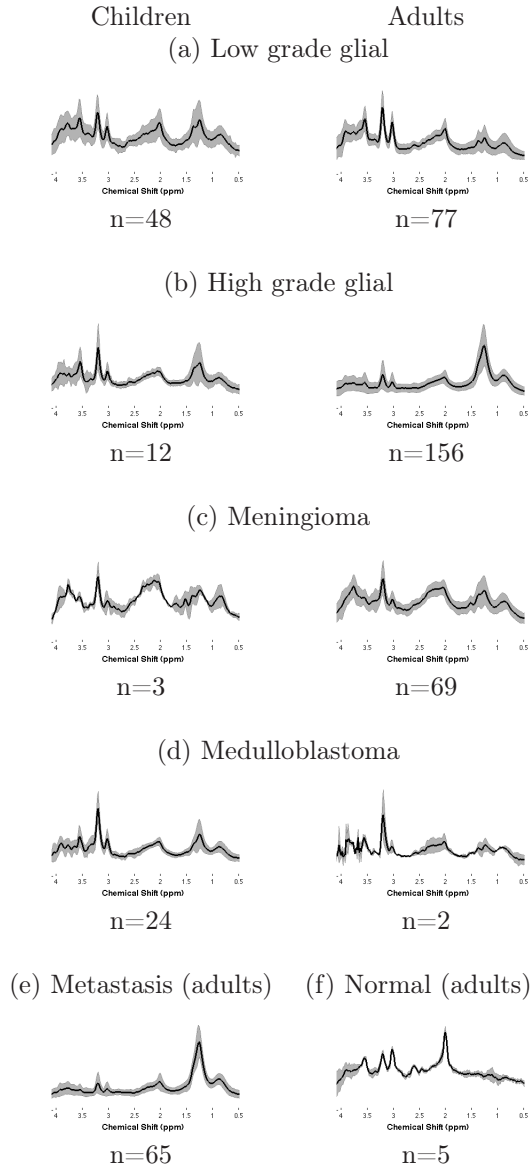


Figure 3.1: Mean spectra of children (left) and adults (right) of several tumour types with standard deviation indicated by the shaded region. Number of patients for each tumour type is indicated beneath each graph. X-axis is in ppm. Y-axis is in arbitrary units.

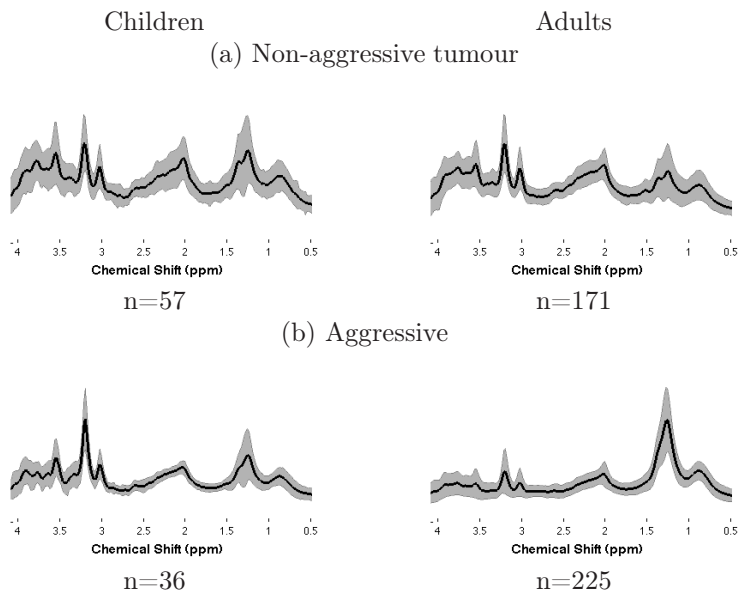


Figure 3.2: Mean spectra of pediatric (left) and adult (right) aggressive and non-aggressive tumours with standard deviation indicated by the shaded region. Number of patients for each tumour type is indicated beneath each graph.

Table 3.2: Estimated metabolite concentrations in aggressive and non-aggressive brain tumours according to the PI method. The total contribution of each metabolite when reasoning at several ppm is calculated. The standard deviation of the concentrations is given in brackets. Analysis of the variance (Kruskal-Wallis test, with  $\alpha=0.05$ , corrected with Bonferroni test) is shown for comparison of aggressiveness in children and adults. On the top, differences in children and adults for the same level of aggressiveness. On the bottom, differences of aggressive and non-aggressive tumours grouped by age (in children and in adults).

Differences in the aggressiveness attending to age						
Metabolite Concentrations	Children	Aggressive Adults	Signif.	Children	Non-aggressive Adults	Signif.
MMLip09	0.89 (0.32)	1.16 (0.36)	<0.01	0.93 (0.41)	0.81 (0.41)	<0.05
MMLip13+Lac	1.59 (0.90)	2.9 (1.11)	<0.01	1.52 (0.88)	1.18 (0.75)	<0.01
Total Ala	1.83 (0.41)	1.87 (0.53)	–	2.21 (0.54)	2.11 (0.56)	–
NAA	1.17 (0.25)	0.97 (0.39)	<0.01	1.27 (0.50)	1.33 (0.38)	–
Total Cr	2.07 (0.72)	1.16 (0.77)	<0.01	1.97 (0.70)	2.13 (0.72)	–
Cho	2.10 (0.66)	0.86 (0.59)	<0.01	1.37 (0.49)	1.56 (0.54)	<0.05
Gly	1.31 (0.52)	0.59 (0.49)	<0.01	1.33 (0.63)	1.28 (0.55)	–
Glx	1.87 (0.39)	1.45 (0.60)	<0.01	2.05 (0.74)	2.19 (0.56)	–
Total ml	3.11 (0.89)	1.31 (0.93)	<0.01	2.58 (1.02)	2.68 (0.91)	–
Tau	0.84 (0.40)	0.40 (0.34)	<0.01	0.93 (0.40)	0.91 (0.36)	–
Discrimination by aggressiveness in children and in adults						
Metabolite Concentrations	Children		Signif.	Adults		Signif.
	Aggressive	Non-aggressive		Aggressive	Non-aggressive	
MMLip09	0.89 (0.32)	0.93 (0.41)	–	1.16 (0.36)	0.81 (0.41)	<0.01
MMLip13+Lac	1.59 (0.90)	1.52 (0.88)	–	2.9 (1.11)	1.18 (0.75)	<0.01
Total Ala	1.83 (0.41)	2.21 (0.54)	<0.01	1.87 (0.53)	2.11 (0.56)	<0.01
NAA	1.17 (0.25)	1.27 (0.50)	<0.01	0.97 (0.39)	1.33 (0.38)	<0.01
Total Cr	2.07 (0.72)	1.97 (0.70)	–	1.16 (0.77)	2.13 (0.72)	<0.01
Cho	2.10 (0.66)	1.37 (0.49)	<0.01	0.86 (0.59)	1.56 (0.54)	<0.01
Gly	1.31 (0.52)	1.33 (0.63)	–	0.59 (0.49)	1.28 (0.55)	<0.01
Glx	1.87 (0.39)	2.05 (0.74)	<0.01	1.45 (0.60)	2.19 (0.56)	<0.01
Total ml	3.11 (0.89)	2.58 (1.02)	<0.01	1.31 (0.93)	2.68 (0.91)	<0.01
Tau	0.84 (0.40)	0.93 (0.40)	–	0.40 (0.34)	0.91 (0.36)	<0.01

Table 3.3: Training and independent test set performances of adult classifiers trained with 300 samples and children classifiers trained with 60 samples. These classifiers discriminate aggressive and non-aggressive BT. Average performance of the 6 classifiers (PI+LDA; PI+KNN; PCA+LDA; PCA+KNN; PI+PCA+LDA; PI+PCA+KNN) is shown as well as the performance of the best one for each category, according to the CV evaluation.

Adults Classifiers	Training Performance		Adults Test Set (96 cases)		Children Test Set (93 cases)	
	ACC	BAR	ACC	BAR	ACC	BAR
Average performance	0.84	0.83	0.78	0.78	0.53	0.50
Best classifier (PI+LDA)	0.85	0.85	0.85	0.85	0.62	0.60

Children Classifiers	Training Performance		Adults Test Set (396 cases)		Children Test Set (33 cases)	
	ACC	BAR	ACC	BAR	ACC	BAR
Average performance	0.77	0.80	0.51	0.52	0.76	0.75
Best classifier (PI+PCA+LDA)	0.84	0.87	0.55	0.50	0.88	0.88

### 3.4.3 Classification

Mean training performance of adult classifiers was ACC=0.84; BAR=0.84 (see Top Table 3.3). On the other hand, mean training performance of childhood classifiers was slightly inferior with ACC=0.77; BAR=0.80 (see Bottom Table 3.3).

When applying the adult independent test set to the adult classifiers, similar results to the training performance were obtained. On the contrary, when the children test set was applied to the adult classifiers, a low mean performance was observed (ACC=0.52; BAR=0.50). Analogously, the performance obtained with the childhood independent test set in children classifiers is similar to the training performance. When the adult test set was applied, the performance became significantly worse (mean performance of ACC=0.51; BAR=0.52).

The best classifier of each age category, according to the CV evaluation, was selected and a filter based on the normal probability density of the ages of their respective training cases was applied in order to decide the convenient classifier for each case to be evaluated. Figure 3.3 depicts a diagram of the decision process



Table 3.4: Independent test set performance of an age-filtered classifier combination of the best children and adult classifiers. Mixed Test Set of 33 children samples and 96 adults samples.

	Mixed Test Set	
	ACC	BAR
Age-Filtered Classifier	0.84	0.85

to decide which classifier should perform the classification attending to the age of a test case. This ‘filtered classifier’ was tested with an independent test set of 33 children cases and 96 adult cases. The performance obtained was similar to the performances obtained by each classifier individually when tested with a test set in accordance to the age of their training samples (ACC=0.84: BAR=0.85; see Table 3.4).

## 3.5 Discussion

### 3.5.1 Tumour characterization from metabolite concentration

The estimation of metabolite concentrations was carried out with the PI technique. Relevant differences were detected between pediatric and adult aggressive tumours, being statistically significant the differences between the concentrations of every peak but the Ala. These significant differences agree with the works of [3, 4, 5, 6, 7], which reported notable differences between children and adults in the distribution of tumours by location, and therefore, with regard to the metabolite concentrations found in the lesions.

In contrast, most of the metabolite concentration did not present relevant differences in the non-aggressive tumours. Only the Cho peak ( $p < 0.05$ ) and the peaks related with the ML (MMLip09, with  $p < 0.05$  and MMLip13+Lac, with  $p < 0.01$ ) showed significant differences.

The significance in Cho concentrations agrees with the fact that children present higher concentrations compared to the amount of NAA due to the brain cell formation during the childhood [21, 22]. Regarding the ML peaks, elevated ML is usually associated to hypoxia and necrosis [23, 24, 25]. These phenomena occur in HG glials and metastasis, where a great peak of lipids predominates.

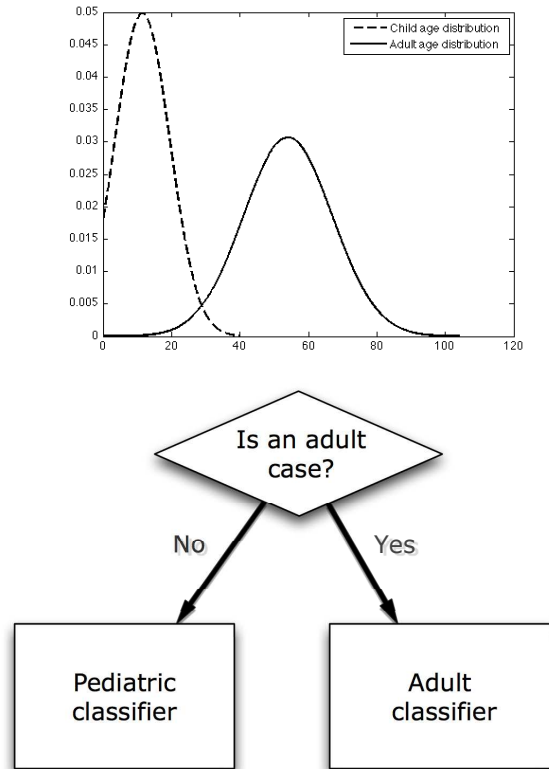


Figure 3.3: Diagram of the ‘filtered classifier’ that decides which specialized classifier should give the answer attending to the age of the test case with respect to the normal probability density of the ages of the cases used to train each classifier.

The appearance of the Lac peak is always associated with the abnormal state of the brain. Low concentration of Lac may suggest low growth. On the contrary, medulloblastomas exhibit a trend toward a higher Lac signal [26].

The Kruskal-Wallis test was also applied to analyze the significant differences between metabolite concentration of aggressive and non-aggressive tumours in adults. It revealed that every metabolite concentration obtained with PI had a significant different concentration for this discrimination problem ( $p < 0.01$ ). The visual examination of both mean signals (Figure 3.2-Right) gave an intuition of how different both levels of aggressiveness are. This was confirmed by the univariate analysis. Since all the relevant metabolite concentrations were significantly different, it might be plausible that a multivariate analysis of these data will obtain a high accuracy rate.

Analogously, the differences in the aggressiveness level in pediatric BT was also analyzed. In contrast with the results with adults, only five metabolite concentration revealed as significantly different. From this result we can infer that this diagnostic discrimination has more inherent complexity in children compared to adults.

The higher presence of Glx and its end product, Ala, in non-aggressive tumours compared to aggressive tumours in children ( $p < 0.01$ ) may be due to the contribution of low grade meningiomas. Important concentrations of these metabolites are usually found in low grade meningiomas [27, 28]. This behaviour is also observed in adult tumours. Higher mI concentrations are found in pediatric aggressive tumours compared to non-aggressive. This might be due to the contribution of high grade glials, since mI is considered a glial marker and its elevation associated to inflammatory processes and increase and proliferation of glial cells [29]. A significant reduction of NAA is observed in the pediatric aggressive tumours compared with the non-aggressive ones ( $p < 0.01$ ). NAA is a marker of neuron activity, both in adults and children. A reduction of NAA indicates reduction or absence of neurons in the region [30, 31]. In children, large reductions in NAA with elevated Cho generally indicates tumour [32]. Cho is also significantly higher in the pediatric aggressive tumours compared to the non-aggressive. In children, a significantly shorter survival time has been observed for patients with the highest Cho/NAA ratios, independent of the tumour type [33, 32].

### **3.5.2 Classification of aggressiveness**

LDA and KNN classifiers obtained similar performances both in the training and test evaluation with children and adults. Although non-statistically significant, lower performance was achieved when discriminating aggressiveness levels in children. This might be explained by the complexity of the problem: univariate analysis revealed that every PI input was statistically different in adults aggressive tumours compared to non-aggressive. In children, though, only five metabolite concentrations were reported to be statistically different. This would explain why the best pediatric classifier was trained with PI+PCA; the firsts principal components of PCA applied after PI made that only the metabolite concentration with significantly relevant information were used to train the classifier. In comparison, the data used to train the adult classifiers (PCA, PI or PCA+PI) did not reveal as a critical factor to improve the performance since both KNN and LDA achieved similar results.

Both types of classifiers performed as predicted by their respective training performance when an independent test set of samples with age accordance was applied. On the contrary, when children classifiers were tested with an adult test set and vice-versa (adult classifiers with a children test set), performance lowered drastically. This result reinforces the idea suggested in [3] that the nature of childhood and adulthood BT may be totally different.

A simple filter based on the normal probability density function of the age estimated in the training dataset can successfully overcome these differences and obtain a classifier that globally behaves as predicted by the training performance.

## 3.6 Conclusions

The results obtained with this experiment shed light to the need of developing a tool that decides which classifier is more adequate to a specific patient. From our point of view, in a clinical DSS oriented to BT diagnosis, and specially if it gives support to both adult and children patients, a tool for auditing and selecting the most adequate predictive model is of the utmost importance.

Chapters 5 and 6 will describe a tool that tries to pursue this aim. Next chapter focuses on the development of classifiers for children.

## Bibliography

- [1] J. Vicente, J. García-Gómez, S. Tortajada, E. Fuster-Garcia, A. Capdevila, A. Peet, B. Celda, and M. Robles. Age-Filtered MRS Classifier To Overcome The Differences In Childhood And Adulthood Brain Tumours. In *Proceedings 17th of the Scientific Meeting, International Society for Magnetic Resonance in Medicine, Honolulu, USA*, page 981, April 2009.
- [2] J. M. García-Gómez, J. Luts, M. Julià-Sapé, P. Krooshof, S. Tortajada, J. Vicente, W. Melssen, E. Fuster-Garcia, I. Olier, G. Postma, D. Monleón, A. Moreno-Torres, J. Pujol, A. P. Candiota, M C. Martínez-Bisbal, J. Suykens, L. Buydens, B. Celda, S. Van Huffel, C. Arús, and M. Robles. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *Magnetic Resonance Materials Physics, Biology and Medicine*, 22(1):5–18, 2009.
- [3] CBTRUS. 2010 CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2006. Technical report, Central Brain Tumor Registry of the United States, 2010.
- [4] I. F. Pollack. Brain tumors in children [review]. *New England Journal of Medicine*, 331:1501–1507, 1994.
- [5] I. F. Pollack. Pediatric brain tumors. *Seminars in Surgical Oncology*, 16: 73–90, 1999.
- [6] K. Imamura. Proton MR spectroscopy of the brain with a focus in chemical issues. *Magnetic Resonance in Medicine*, 2(3):117–132, 2003.
- [7] A. Rossi, C. Gandolfo, G. Morana, M. Severino, M. Garrè, and A. Cama. New MR sequences (diffusion, perfusion, spectroscopy) in brain tumours. *Pediatric Radiology*, 40(6):999–1009, June 2010.
- [8] A. R. Tate, J. Underwood, D. M. Acosta, M. Julià-Sapé, C. Majós, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, M. E. Cabañas, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arús. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. *NMR in Biomedicine*, 19(4):411–434, 2006.

- 
- [9] INTERPRET Consortium. INTERPRET. Web site, 1999-2001. IST-1999-10310, EC, <http://azizu.uab.es/INTERPRET> (Accessed: 22nd April 2012).
- [10] M. Julià-Sapé, D. Acosta, M. Mier, C. Arús, and D. Watson. A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 19(1):22–33, Feb 2006.
- [11] eTUMOUR Consortium. eTumour: Web accessible MR Decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data. Web site. FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC, <http://www.etumour.net> (Accessed: 22nd April 2012; at writing time it was temporarily unavailable).
- [12] M. van der Graaf, M. Julià-Sapé, F. A. Howe, A. Ziegler, C. Majós, A. Moreno-Torres, M. Rijpkema, D. Acosta, K. S. Opstad, Y. M. van der Meulen, C. Arús, and A. Heerschap. MRS quality assessment in a multicentre study on MRS-based classification of brain tumours. *NMR in Biomedicine*, 21(2):148–158, 2008.
- [13] D. N. Louis, H. Ohgaki, O. D. Wiestler, and W. K. Cavenee. *WHO classification of tumours of the central nervous system*. IARC Press, Lyon, France, 2007.
- [14] A. J. Wright, C. Arús, J. P. Wijnen, A. Moreno-Torres, J. R. Griffiths, B. Celda, and F. A. Howe. Automated quality control protocol for MR spectra of brain tumors. *Magnetic Resonance in Medicine*, 59(6):1274–1281, 2008.
- [15] U. Klose. In vivo proton spectroscopy in presence of eddy currents. *Magnetic Resonance in Medicine*, 14(1):26–30, 1990.
- [16] E. Cabanes, S. Confort-Gouny, Y. Le Fur, G. Simond, and P.J. Cozzone. Optimization of Residual Water Signal Removal by HLSVD on Simulated Short Echo Time Proton MR Spectra of the Human Brain. *Journal of Magnetic Resonance*, 150(2):116 – 125, 2001. ISSN 1090-7807. doi: 10.1006/jmre.2001.2318.

- [17] A. Naressi, C. Couturier, I. Castang, R. de Beer, and D. Graveron-Demilly. Java-based graphical user interface for MRUI, a software package for quantitation of in vivo/medical magnetic resonance spectroscopy signals. *Computers in Biology and Medicine*, 31(4):269–286, 2001.
- [18] J. Luts, J.B. Poulet, J. M. García-Gómez, A. Heerschap, M. Robles, J. A. K. Suykens, and S. Van Huffel. Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. *Magnetic Resonance in Medicine*, 60(2):288–298, 2008.
- [19] K. Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
- [20] C. J.C. Burges. Geometric methods for feature extraction and dimensional reduction: A guided tour. Technical report, Microsoft Research, University of Toronto, 2004.
- [21] R. Kreis, T. Ernst, and B. D. Ross. Development of the human brain: *In vivo* quantification of metabolite and water content with proton magnetic resonance spectroscopy. *Magnetic Resonance in Medicine*, 30:424–437, 1993.
- [22] W. Grodd, I. Kraegelo-Mann, U. Klose, and R. Sauter. Metabolic and destructive brain disorders in children: Findings with localized proton MR spectroscopy. *Radiology*, 181:173–181, 1991.
- [23] C. Rémy, N. Foulhé, I. Barba, E. Sam-Laï, H. Lahrech, M-G. Cucurella, M. Izquierdo, A. Moreno, A. Ziegler, R. Massarelli, M. Décorps, and C. Arús. Evidence that mobile lipids detected in rat brain glioma by 1H nuclear magnetic resonance correspond to lipid droplets. *Cancer Research*, 57:407–414, 1997.
- [24] I. Martinez-Pérez, A. Moreno, I. Barba, A. Capdevila, and C. Arús. Large lipid droplets observed by electron microscopy in six human brain tumors with lipid 1H MRS *in vivo* pattern. In *Abstracts of the International Society for Magnetic Resonance in Medicine, New York*, page 976, 1996.
- [25] A.C. Kuesel, S.M. Donnelly, W. Halliday, G.R. Sutherland, and I.C.P. Smith. Mobile lipids and metabolic heterogeneity of brain tumors as detectable by *ex vivo* 1H MR spectroscopy. *NMR in Biomedicine*, 7:172–180, 1994.

- [26] K.H. Chang, I.C. Song, S.H. Kim, M.H. Han, H.D. KIM, S.O. Seong, H.W. Jung, and M.C. Han. *In vivo* single-voxel proton MR spectroscopy in intracranial cystic masses. *Magnetic Resonance in Medicine*, 32:549–555, 1998.
- [27] F A Howe, S J Barton, S A Cudlip, M Stubbs, D E Saunders, M Murphy, P Wilkins, K S Opstad, V L Doyle, M A McLean, B A Bell, and J R Griffiths. Metabolic profiles of human brain tumors using quantitative *in vivo* <sup>1</sup>H magnetic resonance spectroscopy. *Magnetic Resonance in Medicine*, 49(2):223–232, Feb 2003.
- [28] D. J. Manton, M. Lowry, S. J. Blackband, and A. Horsman. Determination of proton metabolite concentrations and relaxation parameters in normal human brain and intracranial tumours. *NMR in Biomedicine*, 8:104–112, 1995.
- [29] Y Rosen and Lenkinski R E Soares. Recent advances in magnetic resonance neurospectroscopy. *Neurotherapeutics*, 4(3):330–345, 2007. ISSN 1933-7213.
- [30] O. Henriksen, P. Gideon, B. Sperling, S. Olsen, H.S. Joergensen, and P. Arlien-Soeborg. Cerebral lactate production and blood flow in acute stroke. *Journal of Magnetic Resonance Imaging*, 2:511–517, 1992.
- [31] H. Bruhn, J. Frahm, M.L. Gyngell, K.D. Merboldt, W. Haenicke, and R. Sauter. Cerebral metabolism in man after acute stroke: new observations using localized proton NMR spectroscopy. *Magnetic Resonance in Medicine*, 9:126–131, 1989.
- [32] F. A. Howe and K. S. Opstad. <sup>1</sup>H MR spectroscopy of brain tumours and masses. *NMR in Biomedicine*, 16(3):123–131, May 2003.
- [33] Z. Wang, L. N. Sutton, A. Cnaan, J. C. Haselgrove, L.B. Rorke, H. Zhao, L. T. Bilaniuk, and R. A. Zimmerman. Proton MR spectroscopy of pediatric cerebellar tumors. *American Journal of Neuroradiology*, 16:1821–1833, 1995.



## Chapter 4

# Automatic classification of childhood brain tumours from multi-centre *in vivo* $^1\text{H}$ MRS

*“I hope our wisdom will grow with our power, and teach us, that the less we use our power the greater it will be.”*

*Thomas Jefferson*

The aim of the current chapter is to evaluate the automatic classification of pediatric brain tumours by combining single-voxel  $^1\text{H}$  MRS at two different echo times (TE). Our hypotheses are (1) that the classification provides an accurate non-invasive diagnosis in multi-centre datasets and (2) the combination of TE improves the accuracy of classification compared with the use of one TE alone.

78 patients under 16 years old with histologically proven brain tumours from 10 international centres were investigated. Discrimination of 29 medulloblastomas, 11 ependymomas and 38 pilocytic astrocytomas was evaluated. Single voxel MRS was undertaken prior to diagnosis (1.5T PRESS, PROBE or STEAM, TE 20-32 ms, and PRESS, 135-136 ms). MRS data was processed using three strategies, determination of metabolite concentrations using TARQUIN software and

automatic feature extraction with Peak integration and Principal Components Analysis. Linear Discriminant Analysis was applied to this data to produce diagnostic classifiers. An evaluation of the diagnostic accuracy based on resampling to measure the BAR was performed.

The accuracy of the diagnostic classifiers for discriminating the three tumour types was found to be high (BAR 0.98) when a combination of TE was used. The combination of both TE significantly improved the classification performance ( $p < 0.01$ , Tukey's test) compared with the use of one TE alone. Other tumour types were classified accurately as glial or primitive neuroectodermal (BAR 1.00).

<sup>1</sup>H-MRS with combined Short and Long TE has very good accuracy for the non-invasive diagnosis of common childhood brain tumours and should become part of routine clinical assessment for these children.

*The core of this chapter has been recently accepted as a scientific paper in the European Journal of Cancer [1] and published in the proceedings of the 15th International Symposium of Pediatric Neuro-Oncology [2].*

## 4.1 Introduction

The conventional MR images are inaccurate in discriminating between most childhood brain tumours and a diagnosis is usually made from biopsy samples taken at operation with the histology of the tumour being used to formulate the treatment plan. Whilst histopathology provides a definitive diagnosis there would be several advantages to obtaining an accurate non-invasive diagnosis.

For tumours where surgical resection is not the initial therapeutic option, an accurate non-invasive diagnosis would avoid an invasive procedure. For tumours where surgery is undertaken at diagnosis, accurate diagnostic information on the tumour type prior to initial surgery would help surgical decision-making, allow timely adjuvant therapy planning and aid discussions with the family. A common site for childhood tumours is the cerebellum and surgical resection is usually the initial therapeutic intervention. However, the importance of a complete resection varies between the tumour types. A complete macroscopic resection is highly prognostic for ependymomas [3, 4], whereas small amounts of residual medulloblastoma (up to 1.5cm<sup>2</sup>) are not of prognostic significance if treated with radiotherapy and chemotherapy [5] and small residual masses of pilocytic

astrocytoma may be observed without further treatment [6]. Histopathology is usually not available for several days after the operation and intraoperative histopathology is commonly used to inform the surgeons of the likely tumour type but the techniques available are not accurate [7] and this strategy does not allow patient-specific clinical management planning prior to surgery. With improved adjuvant treatment, therapeutic strategies may evolve to ones in which surgery is undertaken at a later point, which is already common for childhood tumours outside the brain.

Histopathology remains the ‘gold standard’ for classifying childhood brain tumours and is the basis for treatment planning in the majority of cases. However, patients with identical histopathological diagnosis can respond in different ways to treatment and there is increasing evidence that additional information from tumour biology can improve the classification [8]. Advances in imaging have allowed tissue properties to be probed non-invasively giving important insights into *in vivo* tumour biology [9]. The aims of modern imaging are therefore not just to give a non-invasive histological diagnosis but rather to improve the classification of tumours.

Multivariate analysis of automated MRS processing is a powerful technique that can yield rapid and robust results and promises to translate into routine clinical practice.

However, a multi-centre evaluation of these techniques is required. Although a large number of multi-centre studies on automatic classification of brain tumours has been reported in adults [10, 11, 12, 13, 14, 15, 16, 17], these results cannot be extrapolated to children since the overall distribution of the tumour types, locations and etiology differs markedly from that of adults [18, 19, 20, 21]. Also, combining Short-TE and Long-TE MRS can potentially give accurate quantification of more metabolites than either technique alone [22] but this has currently not been reported in pediatric brain tumours.

The present study investigates the accuracy of tumour metabolite profiles measured by  $^1\text{H}$  MRS as a diagnostic aid for common childhood brain tumours. The aim of this work is twofold: first, to evaluate the automatic classification of pediatric brain tumours using a large multi-centre  $^1\text{H}$  MRS dataset; and second, to test whether maximizing the metabolite information available by combining Short-TE and Long-TE MRS improves the automatic classification of pediatric brain tumours compared with the use of one TE alone.

---

## 4.2 Material and methods

### 4.2.1 Cases and data acquisition

The study includes 97 patients under 16 y.o. (mean age  $7.3 \pm 4.7$ ) with histologically proven brain tumour collected from 10 international centres in the framework of the eTUMOUR project (2004-2009) [23]. Histopathological diagnoses were validated in the context of clinical setting and radiological images and reviewed by the multidisciplinary Clinical Validation Committee. MRS data was reviewed for quality control by expert spectroscopists of the eTUMOUR project [10, 24].

The classes considered in this study were defined according to the histological classification of the CNS tumours established by the WHO [25]. The 97 cases were distributed as follows: 38 Pilocytic Astrocytoma (PILOA), 20 of them in the PF; 11 Ependymoma grade II (EPEN), 7 in the PF; 29 MED, all of them in the PF; Additionally we included 10 Diffuse Astrocytoma (DASTRO), 3 in the PF; 3 Subependymal giant cell astrocytoma (SASTRO), 2 in the ventricular atrium and 1 in the frontal lobe; one primitive neuroectodermal tumour located in the frontal lobe; 3 Atypical teratoid Rhabdoid Tumour (ATRT), 2 in the PF; and 2 Pineoblastoma (PINEOB) in the pineal region. Table 4.1 documents the available cases.

Acquisition protocols for clinical, radiological and histopathological data were defined to ensure the compatibility of the data acquired at the different hospitals [26, 24].

Single voxel  $^1\text{H}$  MRS at 1.5T from 90 patients were acquired at Short-TE and 61 spectra at Long-TE, from which 54 patients had both echo times available from scanners of three major manufacturers (Philips, Siemens and General Electric). Voxels were placed avoiding as much as possible areas of necrosis or cysts and with minimum contamination from the surrounding non-tumoral tissue. The acquisition protocols for Short-TE included PRESS, PROBE or STEAM sequences, with TR of 1500-2000ms, TE of 20 or 30ms, spectral width of 500 - 2500Hz, and 512, 1024 or 2048 data-points. Long-TE spectra were acquired with the PRESS sequence with TR of 1500-2020ms, TE of 135 or 136ms, spectral width of 1000-2500Hz and 512 or 2048 data-points.

Table 4.1: Number of cases.

Label	Short-TE	Long-TE	Short-TE+Long-TE	Total different cases
PILOA	37	27	26	38
EPEN	9	7	5	11
MED	28	15	14	29
SASTRO	2	3	2	3
DASTRO	8	8	6	10
ATRT	3	0	0	3
PNET	1	1	1	1
PINEOB	2	0	0	2
TOTAL	90	61	54	97

## 4.2.2 MRS processing

Two different MRS data processing methods were compared: quantitation of metabolite concentrations using the time-domain fitting algorithm TARQUIN (version 4.1.1) [27]; and the automatic feature extraction technique of PI [17, 28].

### Quantitation with TARQUIN

The MRS data were processed and fitted with TARQUIN (version 4.1.1) software tool [27]. The quantitation of metabolites, lipids and macromolecular components was performed using the spectrum acquired without water suppression as reference as defined in Section 2.4.2 of Chapter 2.

Mean metabolite estimations were compared using non-parametric analysis of variance and multiple comparison tests to detect statistically significant differences between tumour types.

Two quality control criteria were taken into account in order to include the samples in the analysis: (1) A quality fit ( $Q$ ) lower or equal to 2, where  $Q$  is supplied by TARQUIN and defined as the standard deviation of the frequency domain residual between 0.2 and 4.0 ppm divided by the standard deviation of the spectral noise; and (2) a SNR of the spectra greater than 5. These criteria were established based on the minimum quality standard that the commercial software LCModel<sup>TM</sup> recommends [27, 29]. The inclusion of data that not accomplish these two quality criteria would negatively affect to the performance of the classifiers due to the incorporation of spurious information like noise artifacts and distorted baseline signals in the spectra.

### Semiautomatic processing pipeline

Each spectrum was also semiautomatically processed following the MRS processing pipeline defined in Chapter 3, Section 3.2.1. This processing was performed as a previous step to the application of PR feature extraction approaches.

#### 4.2.3 Feature extraction

Two different PR automatic feature extraction methods were explored in order to compare them with the quantitation: PI and PCA. Both techniques were chosen because of the good results reported in [17]. PI technique was applied after the semiautomatic processing pipeline to estimate the relative concentration of metabolites integrating the frequency signal around the resonance frequencies. This technique has been successfully applied in previous studies with MRS from adults [17, 28]. The pediatric PI, as described in Section 2.4.2 of Chapter 2, was also used in this study.

Additionally, PCA [30, 31] was applied in two different ways: (1) to the region of [4.1, 0.5] ppm interval on the spectra obtained after the semiautomatic processing pipeline, and (2) to the variables estimated after applying quantitation and after applying PI. In both cases, we took as input for the classification models the minimum principal components that explained at least the 85% of the total variance.

#### 4.2.4 Classification methods and evaluation technique

A set of three classification problems is addressed in this study: a binary discrimination of glial-cell tumours (EPEN and PILOA) from MED; a binary discrimination of PILOA and MED; and a three-class classification for discriminating EPEN from PILOA and from MED. Obtaining high diagnostic accuracy from PILOA, EPEN and MED is clinically relevant to define a surgical strategy for tumour resection. Since a common location of these tumours is the PF, where it is difficult to distinguish the tumour type with MRI alone, we obtained the classifiers trained only with the tumour cases located in the PF. Afterwards, new classifiers trained with tumours in any brain location were also obtained. For each discrimination problem, classifiers were designed and evaluated using features from Short-TE and Long-TE alone and combining both time of echo

(Short-TE+Long-TE). Our results were compared with the results reported in previous works [32, 33, 34, 35].

Based on the benchmark results of previous studies [17, 22, 28], we choose LDA [36] as the classification technique to apply to the pediatric brain tumour types.

Classifiers were evaluated with the kRSTT strategy with  $k$  set to 150 and the performance of the classifiers was measured with the BAR. The BAR is useful to ensure that the sensitivity of every class is high and balanced when the prevalence of one or more classes is unbalanced with respect to other in the dataset. This situation occurs in our study, when discriminating EPEN from MED and from PILOA, because the amount of EPEN cases was low compared to the number of the other two tumour types (see Table 4.1).

In order to check the significance of differences between the use of combination of both time echo and the use of Short-TE or Long-TE alone, the Friedman's (F) non-parametric two-way analysis of variance test with  $\alpha=0.05$  [37] was applied, for each classification problem, to the classifiers trained with each feature set from Short-TE, Long-TE and the combination of both time echo. When applying multiple comparisons, the results were analyzed with the Tukey's (T) honestly significant difference criterion with  $\alpha=0.01$  [38].

The feature sets compared in our experiments were: (1) the principal components after applying the semiautomatic processing, (2) the variables obtained applying PI, (3) the variables obtained applying pediatric PI, (4) the principal components after applying PI, (5) the quantitation variables estimated with TARQUIN, (6) the subset of significant quantitation variables estimated with TARQUIN, and (7) the principal components of the TARQUIN quantitation.

## 4.3 Results

### 4.3.1 Spectral features

Several key features allow visual discrimination of PILOA, EPEN and MED. Figures 4.1 and 4.2 show the Short-TE and Long-TE mean spectra of the tumour types. Minimum differences are found between the mean spectra of the tumours in the PF and those of tumours in any location. All tumour spectra display an increase in Cho peak (3.2ppm) with respect to Cr peak (3.0ppm). This

is most notable in Long-TE MRS. NAA (2.0ppm) presents a less prominent peak in MED and EPEN compared with PILOA. Elevation of macromolecules and lipids at 0.9ppm and 1.3ppm (MMLip09, and MMLip13) is observed in Short-TE. Regarding Long-TE, the inverted peak at 1.3ppm, corresponding to Lactate can be seen in PILOA and EPEN but not in MED.

### 4.3.2 Univariate metabolite comparison

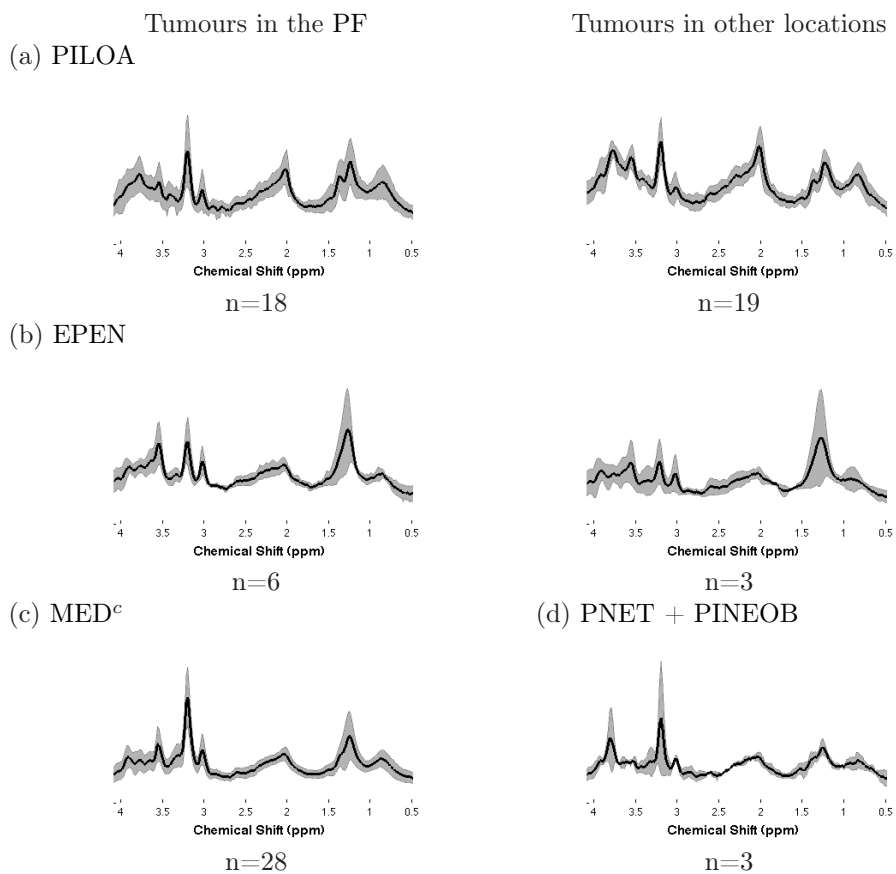
In Table 4.2, the metabolite concentrations estimated with TARQUIN software in Short-TE and Long-TE for the three tumour types found in any brain location can be seen. The Kruskal-Wallis non-parametric test for the analysis of the variance ( $\alpha = 0.05$ ) has been applied to determine the significant differences in metabolite concentrations for the three questions addressed for classification. For all questions, both Cho components, glycerophosphocholine (GPC) ( $p \leq 0.01$ ) and Phosphocholine (PCh) ( $p \leq 0.01$  for Short-TE;  $p \leq 0.05$  for Long-TE) showed significant differences. Cr ( $p \leq 0.01$ ) and Tau ( $p \leq 0.01$ ) concentrations revealed as significant in both time of echo. The mI metabolite peak ( $p \leq 0.01$ ) was significant in Short-TE for discrimination of PILOA and MED and discrimination of the three tumour types. Although mI revealed as significant in Long-TE, this may be due to the contribution of Gly, which was not included in the TARQUIN basis set and, therefore, no attempt was made by the algorithm to separate it from mI. Macromolecules and lipids at 0.9, 1.3 and 2.0 ppm ( $p \leq 0.05$ ,  $p \leq 0.01$  and  $p \leq 0.01$ , respectively) showed statistical differences in Short-TE.

### 4.3.3 Classification

Tables 4.3, 4.4 and 4.5 summarize the classification results of the different questions addressed in this study for the cases in the PF and the cases in any brain location and shows the performance when using Short-TE, Long-TE and a combination of both time of echo. Each discrimination was tackled with the quantitation estimated with TARQUIN, with PI and pediatric PI and with PCA applied after performing the semiautomatic processing, applied to the variables estimated with TARQUIN and to the PI variables.

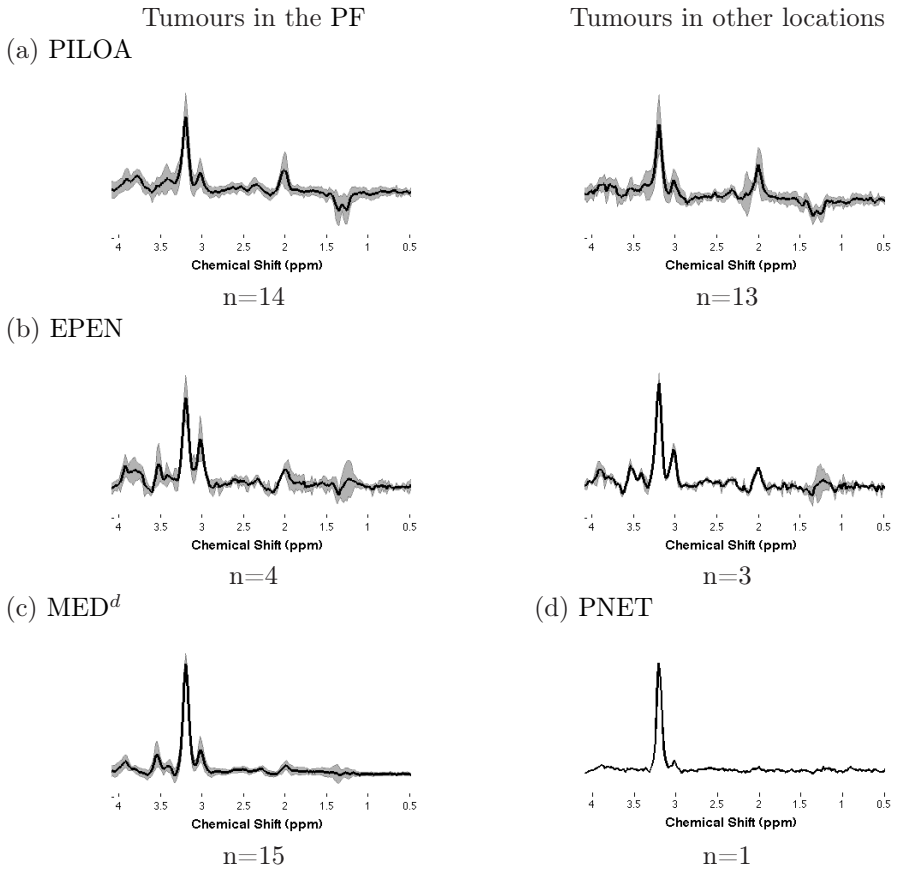
The BAR obtained with the models trained with tumour cases from any location was slightly higher than the BAR of the models trained only with the PF cases. Comparable performances were obtained with the different approaches





<sup>c</sup> MED tumours by definition originate in the PF.

Figure 4.1: Short-TE mean spectra of tumours located in the PF (left) and in any other location than PF (right) with standard deviation indicated by the shaded region. Number of patients is indicated beneath each graph.



<sup>d</sup> MED tumours by definition originate in the PF.

Figure 4.2: Long-TE mean spectra of tumours located in the PF (left) and in any other location than PF (right) with standard deviation indicated by the shaded region. Number of patients is indicated beneath each graph.

Table 4.2: Estimated metabolite concentrations (mM) at several ppm calculated with TARQUIN relative to total water from Short-TE spectra (top) and Long-TE spectra (bottom). The standard deviation of the concentrations is given in brackets. The  $p$ -value of the analysis of the variance (Kruskal-Wallis test,  $\alpha = 0.05$ ) is shown for different discriminating questions when significant differences are observed.

Metabolite	Short-TE			Kruskal-Wallis test ( $p$ value)		
	Tumour Type			(PILOA, EPEN) vs MED	PILOA vs MED	PILOA vs EPEN vs MED
	PILOA (mM)	EPEN (mM)	MED (mM)			
Ala	0.9 (0.9)	2.5 (2.3)	0.6 (0.5)	–	–	–
Cr	1.3 (1.1)	3.4 (1.5)	3.6 (1.9)	<0.01	<0.01	<0.01
Glc	2.7 (1.7)	2.3 (1.8)	1.9 (1.2)	<0.01	<0.01	<0.01
Gln	3.7 (1.9)	6.7 (3.0)	4.1 (3.1)	–	–	–
Glu	2.0 (1.6)	3.1 (3.2)	2.5 (1.5)	–	–	–
mI	2.1 (1.9)	9.0 (5.3)	5.3 (3.2)	–	<0.01	<0.01
Lac	2.4 (1.9)	3.2 (4.3)	2.9 (3.0)	–	–	–
NAA	1.6 (1.4)	1.1 (0.3)	1.5 (1.0)	–	–	–
Scyllo	0.7 (0.8)	0.6 (0.4)	0.9 (0.5)	<0.05	<0.01	<0.05
Tau	1.8 (1.3)	1.8 (1.6)	4.5 (3.6)	<0.01	<0.01	<0.01
GPC	0.9 (0.4)	1.6 (0.5)	2.2 (1.2)	<0.01	<0.01	<0.01
PCh	1.6 (2.4)	1.3 (1.0)	2.2 (1.6)	<0.01	<0.01	<0.01
-CrCH <sub>2</sub>	1.7 (1.8)	3.2 (1.9)	1.3 (1.1)	–	–	–
Gua	1.8 (1.7)	1.2 (1.2)	2.1 (2.0)	–	–	–
GABA	3.8 (7.9)	2.2 (1.7)	2.9 (3.5)	–	–	–
Asp	9.1 (24.2)	4.5 (3.6)	3.6 (2.7)	–	–	–
Cho: GPC+PCh	1.4 (1.6)	2.4 (1.1)	4.1 (1.6)	<0.01	<0.01	<0.01
Glx: Glu+Gln	4.9 (2.2)	8.9 (4.2)	5.5 (3.1)	–	–	<0.05
MMLip09	4.9 (2.9)	8.1 (5.5)	8.3 (5.9)	<0.05	<0.05	<0.05
MMLip13	7.7 (7.0)	27.3 (17.6)	20.5 (18.9)	<0.01	<0.01	<0.01
MMLip20	6.2 (2.9)	10.8 (3.3)	11.3 (5.2)	<0.01	<0.01	<0.01

Metabolite	Long-TE			Kruskal-Wallis test ( $p$ value)		
	Tumour Type			(PILOA, EPEN) vs MED	PILOA vs MED	PILOA vs EPEN vs MED
	PILOA (mM)	EPEN (mM)	MED (mM)			
Ala	0.9 (0.9)	0.9 (1.1)	1.4 (0.8)	<0.05	<0.05	<0.05
Cr	1.3 (1.1)	5.8 (2.2)	5.6 (2.1)	<0.01	<0.01	<0.01
Glc	2.7 (1.7)	7.2 (5.4)	3.0 (1.8)	–	–	–
Gln	3.7 (1.9)	7.2 (3.5)	3.4 (2.1)	–	–	–
Glu	2.0 (1.7)	5.3 (1.9)	5.8 (2.3)	–	–	–
mI <sup>b</sup>	2.1 (1.9)	24.1 (14.8)	31.1 (14.5)	<0.01	<0.01	<0.01
Lac	2.4 (1.9)	2.2 (1.0)	1.9 (1.7)	–	–	–
NAA	1.6 (1.4)	2.4 (1.3)	2.3 (1.6)	–	–	–
Scyllo	0.7 (0.9)	0.9 (0.6)	0.8 (0.6)	–	–	–
Tau	1.8 (1.3)	4.3 (4.5)	6.9 (6.4)	<0.01	<0.01	<0.01
GPC	0.9 (0.4)	2.1 (1.3)	4.1 (4.2)	<0.01	<0.01	<0.01
PCh	1.6 (2.4)	3.0 (0.9)	5.7 (5.6)	<0.05	<0.05	<0.05
-CrCH <sub>2</sub>	1.7 (1.8)	3.9 (3.2)	3.4 (1.7)	–	–	–
Gua	1.8 (1.7)	1.5 (0.1)	0.7 (0.4)	<0.05	<0.05	–
GABA	3.9 (7.9)	–	1.3 (0.8)	–	–	–
Asp	9.1 (24.2)	2.4 (1.3)	3.4 (3.1)	–	<0.05	–
Cho: GPC+PCh	1.4 (1.6)	4.3 (1.9)	9.5 (5.9)	<0.01	<0.01	<0.01
Glx: Glu+Gln	4.9 (2.2)	10.5 (4.9)	7.4 (3.5)	–	–	–
MMLip09	4.9 (2.9)	1.8 (1.1)	0.5 (0.2)	–	–	<0.05
MMLip13	7.7 (7.0)	9.2 (8.8)	3.6 (2.0)	–	<0.05	<0.05
MMLip20	6.2 (2.9)	1.6 (1.3)	1.2 (0.9)	<0.05	<0.05	–

<sup>b</sup> Concentration of mI may contain Gly contribution. Gly was not included in the TARQUIN basis-set.

followed to calculate the classification models for each discrimination question and input data (Short-TE, Long-TE and combination). The (PILOA + EPEN) vs MED discrimination (Table 4.3) obtained a best BAR of 0.91 for Short-TE and Long-TE alone and 0.98 for the combination of both time of echo. Discrimination of PILOA vs MED (Table 4.4) obtained a best BAR of 0.92, 0.95 and 0.96 respectively for Short-TE, Long-TE and combination of both time of echo. The discrimination of the three classes (Table 4.5) obtained the best BAR performance of 0.81 for Short-TE, 0.88 for Long-TE and 0.98 for the combination of both time of echo. In the three addressed questions, the best performance obtained with the combination of both time of echo showed a significant improvement ( $p < 0.01$ , Tukey's test,  $\alpha = 0.01$ ) compared to the best performance obtained with one TE alone.

Figure 4.3 shows the LDA latent spaces from Short-TE, Long-TE and the combination of both echo times obtained for the discrimination of PILOA, EPEN and MED located in the PF. In addition, Figure 4.3-d shows the latent space from Short-TE for the three tumour types in any brain location.

Other tumour types were classified as glial or primitive neuroectodermal according to the result with the classifier developed for PILOA, EPEN and MED. A BAR of 0.91 was obtained with Short-TE, 0.67 for Long-TE and 1.00 was achieved with the combination of both echo times. Figure 4.4 shows the latent space from Figure 4.3-d with other tumour types projected. The ATRT and PNET cases fall close to the boundaries of the MED area, the only exception being the ATRT case located in the frontal lobe. The DASTRO and SASTRO cases are spread all over the PILOA and EPEN area, never within the MED area.

Table 4.3: Balanced Accuracy Rate (BAR) of the classifiers trained with Short-TE, Long-TE and combination of both time of echo (Short-TE+Long-TE) for the discrimination problem of (PILOA, EPEN) vs MED.

Cases from tumours located at the PF			
(PILOA, EPEN) vs MED			
	Short-TE	Long-TE	Short-TE+Long-TE
full spectra PCA	0.74	<b>0.94</b>	<b>0.94</b>
PI	<b>0.83</b>	0.89	0.86
Pediatric PI	0.82	0.91	0.86
PI + PCA	<b>0.83</b>	0.87	0.90
TARQUIN	0.77	0.81	0.81
TARQUIN signif	0.71	0.76	0.83
TARQUIN + PCA	0.74	0.86	0.84
Cases from any brain tumour location			
(PILOA, EPEN) vs MED			
	Short-TE	Long-TE	Short-TE+Long-TE
full spectra PCA	0.86	<b>0.91</b>	0.94
PI	0.88	0.85	0.89
Pediatric PI	0.87	0.85	0.87
PI + PCA	<b>0.91</b>	0.85	0.87
TARQUIN	0.82	<b>0.83</b>	<b>0.98</b>
TARQUIN signif	0.81	0.87	0.86
TARQUIN + PCA	0.81	0.85	0.91

Table 4.4: Balanced Accuracy Rate (BAR) of the classifiers trained with Short-TE, Long-TE and combination of both time of echo (Short-TE+Long-TE) for the discrimination problem of PILOA vs MED.

Cases from tumours located at the PF			
PILOA vs MED			
	Short-TE	Long-TE	Short-TE+Long-TE
full spectra PCA	0.81	<b>0.94</b>	<b>0.95</b>
PI	0.86	0.93	0.94
Pediatric PI	0.83	0.93	0.91
PI + PCA	<b>0.89</b>	0.92	<b>0.95</b>
TARQUIN	0.87	0.91	0.80
TARQUIN signif	0.78	0.91	0.89
TARQUIN + PCA	0.82	0.92	0.89
Cases from any brain tumour location			
PILOA vs MED			
	Short-TE	Long-TE	Short-TE+Long-TE
full spectra PCA	0.88	0.93	<b>0.96</b>
PI	<b>0.92</b>	0.94	0.95
Pediatric PI	0.90	0.91	0.94
PI + PCA	0.91	0.93	<b>0.96</b>
TARQUIN	0.89	0.94	0.94
TARQUIN signif	0.89	<b>0.95</b>	0.93
TARQUIN + PCA	0.89	<b>0.95</b>	0.94

Table 4.5: Balanced Accuracy Rate (BAR) of the classifiers trained with Short-TE, Long-TE and combination of both time of echo (Short-TE+Long-TE) for the discrimination problem of PILOA vs EPEN vs MED.

Cases from tumours located at the PF			
	PILOA vs EPEN vs MED		
	Short-TE	Long-TE	Short-TE+Long-TE
full spectra PCA	0.65	0.78	<b>0.98</b>
PI	0.65	0.62	0.90
Pediatric PI	0.65	0.64	0.90
PI + PCA	0.71	0.60	0.93
TARQUIN	0.67	<b>0.94</b>	0.96
TARQUIN signif	0.70	0.75	0.96
TARQUIN + PCA	<b>0.72</b>	0.92	0.97
Cases from any brain tumour location			
	PILOA vs EPEN vs MED		
	Short-TE	Long-TE	Short-TE+Long-TE
full spectra PCA	0.80	0.76	0.96
PI	0.76	0.69	0.92
Pediatric PI	0.77	0.63	0.93
PI + PCA	0.78	0.67	0.92
TARQUIN	0.79	0.83	0.98
TARQUIN signif	<b>0.81</b>	<b>0.88</b>	<b>0.99</b>
TARQUIN + PCA	0.77	0.81	<b>0.99</b>

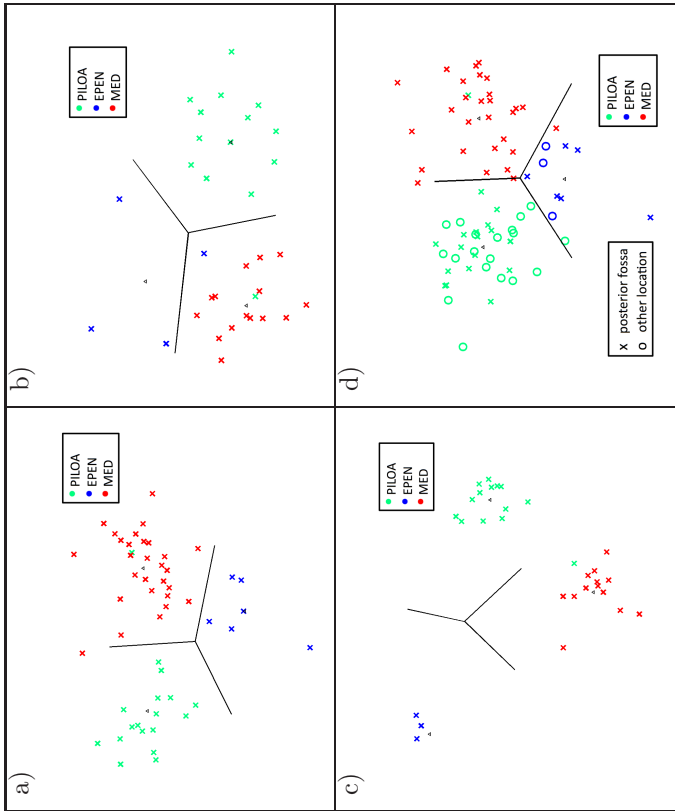


Figure 4.3: LDA latent space for the discrimination of PILOA, EPEN and MED located in the PF using PI applied to: a) Short-TE; b) Long-TE; and c) to both time of echo. Each triangle represents the centroid of each cloud of samples. Figure d) represents the LDA latent space for discrimination of PILOA, EPEN and MED using PI applied to Short-TE for tumours located in the PF (represented with 'x') and other locations (represented with 'o').

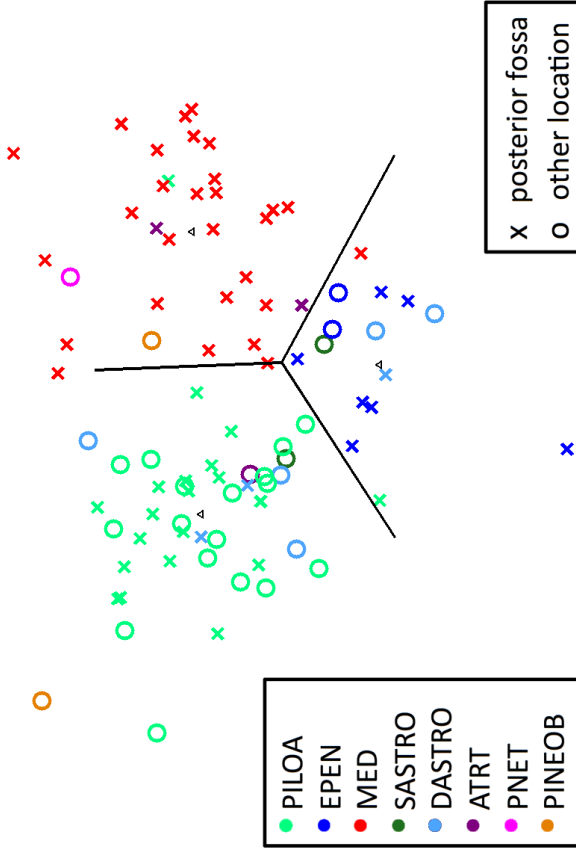


Figure 4.4: LDA latent space for the discrimination of Medulloblastoma (MED), Pilocytic Astrocytoma (PILOA) and Ependymoma grade II (EPEN) using PI applied to Short-TE. The projection of Subependymal giant cell astrocytoma (SASTRO), Diffuse Astrocytoma (DASTRO), Atypical teratoid Rhabdoid Tumour (ATRT), primitive neuroectodermal tumour (PNET) and Pineoblastoma (PINEOB) is shown.



## 4.4 Discussion

Combination of  $^1\text{H}$  MRS with a conventional MRI investigation can yield a metabolite profile of the tumour, showing impressive results in brain tumour diagnosis. In 1996, Pruel *et al.* demonstrated that high diagnostic rates could be achieved for brain tumours in adults by combining MRS with pattern recognition techniques [39]. However, generalization of this approach to multiple centres has faced considerable challenges and accurate diagnosis with MRS has only recently been confirmed in large multi-centre studies of adults [17].

Limited single-centre studies have shown the promise of MRS in the non-invasive diagnosis and characterization of childhood brain tumours. These studies focused on the three most prevalent tumour types: EPEN, MED and PILOA. For example, Wang *et al.* [32] collected data from 26 patients using a Long-TE MRS technique, which detected a small number of metabolites, and presented these as ratios. They obtained an accuracy of 0.85 discriminating the three tumour types using the metabolites ratios NAA:Cho and Cr:Cho. Arle *et al.* [33] obtained an accuracy of 0.88 with a neural network using metabolites ratios of NAA, Cho and Cr from MRS data of 33 patients.

Improvements in technology allowed a larger number of metabolites to be quantified using Short-TE MRS. Schneider *et al.* [34] combined Short-TE MRS and diffusion-weighted imaging data from 17 patients. They obtained a predictive value of 1 in all the cases when applying an LDA with seven variables from the diffusion-weighted image and six metabolites. Davies *et al.* [35] used an automated method for fitting MRS data of 35 patients to quantify 25 metabolite, lipid and macromolecule concentrations and used this as an input to an LDA. They reported an accuracy of 0.93 when discriminating the three tumour types.

To our knowledge, our study is the first to report key differences in pediatric brain tumours in a large multi-centre  $^1\text{H}$  MRS investigation. Diagnostic classifiers were developed for discrimination of MED, EPEN and PILOA and high levels of accuracy achieved. Different methods of MRS data processing and classifier building were compared, and similar results obtained. A significant improvement in the diagnosis rates was obtained with the combination of Long-TE and Short-TE compared to the performance when using either TE alone. The performances of classifiers using one TE were similar to those reported previously [32, 33, 34, 35]. More sophisticated classification techniques could have been used to address this experiment, nevertheless, linear classifiers were applied for the sake of comparison with previous works and to state clear that the

high performance is achievable with simple and standard methods.

The pediatric PI omitted the information of Gly and Ala at 3.78 ppm since these concentrations are usually not significant in children when registered from 1.5T MRS. No significant differences were achieved between the performance of the classifiers trained with PI and those trained with pediatric PI. Hence, the pediatric PI feature extraction might be a more convenient technique to be applied to MRS from children since it reduces the input to the classifiers respect to the conventional PI, initially designed for MRS from adults.

Estimation of metabolite concentrations was performed with the TARQUIN software (Table 4.2). Results are in broad accordance with other studies of pediatric brain tumours [40, 41, 35] where the metabolite concentrations were estimated using the LCModel<sup>TM</sup> software [42]. Significant differences in metabolite concentrations of PILOA, EPEN and MED were found with the Kruskal-Wallis non-parametric test for the analysis of the variance with  $\alpha = 0.05$ . GPC and PCh are higher ( $p \leq 0.01$ ) in MED (grade IV) and EPEN (grade II) compared to PILOA (grade I). This reveals Cho as an important marker for distinguishing between these tumours. Other studies have observed this behaviour, considering high peaks of Cho as an indicator of cell proliferation and tumour malignancy [19, 20]. As previously reported [43, 40, 41], Tau concentration is significantly higher in MED than glial tumours ( $p \leq 0.01$ ). This is a useful biomarker but little is known currently about its role in these tumours. Differences in Cr and mI are also significant ( $p \leq 0.01$ ). Greater concentrations of both metabolites are detected in EPEN and MED compared to PILOA. mI has been associated with grade II tumours in adults, being particularly high in grade II EPEN [35]. Lipids and macromolecules are significantly different between the three tumour types MMLip09 and MMLip13 in both echo times ( $p \leq 0.05$ ), MMLip20 in Short-TE ( $p \leq 0.01$ ). PILOA has lower concentrations of lipids and macromolecules, whereas EPEN shows significantly more elevated concentrations of MMLip13 compared to PILOA and MED (see also Figure 4.1). Elevated mobile lipids usually suggest hypoxia, apoptosis and necrosis and are associated with high malignancy and poor survival [44, 45].

García-Gómez *et al.* [22] found that the combination of both TEs improves the results when performing automatic classification of brain tumours in adults with <sup>1</sup>H MRS. Our study confirms this finding in pediatric brain tumours, obtaining significant differences ( $p < 0.01$ , Tukey's test,  $\alpha = 0.01$ ) between the performance of the models trained with both TEs and the performance of the models trained with Short-TE or Long-TE alone. This behaviour is observed with either MRS

analysis technique applied (quantitation with TARQUIN or PI).

Discrimination of PILOA, MED and EPEN using one TE achieved accuracies similar to the performances reported by equivalent single-centre studies with smaller datasets [32, 33, 34, 35]. In most studies, overall accuracy is reported rather than BAR. Nevertheless, BAR is a more realistic measure of accuracy, especially when the classifier accuracies vary between diagnostic groups and the groups vary greatly in size. In our study, the diagnostic accuracy of the Short-TE classifier was lower for EPEN than the other two tumour types. This led to a lower BAR (0.79 with TARQUIN; 0.76 with PI) than overall accuracy (0.84 with TARQUIN; 0.80 with PI) since there was a smaller number of EPEN. The diversity of their metabolite profiles might have contributed to the lower classification of EPEN. Short-TE determines a larger number of metabolites than Long-TE and, hence, is prone to greater variability in quality which might be accentuated in a multi-centre study.

Figure 4.4 illustrates the potential for generalization of our classifiers to other tumour types. Although these classifiers have been trained with PILOA, EPEN and MED, they effectively separate other low-grade glial tumour types (SASTRO and DASTRO) from embryonal tumour types (ATRT and PNET). The combination of both TEs achieved a BAR of 1.00 for this diagnostic classification.

Although PI is a signal pre-process feature extraction that benefits from prior knowledge of the metabolite distribution in the spectra and TARQUIN is a complex algorithm for quantitation of components in the MRS data, the general-purpose feature selection technique of PCA achieved similar classification performance. Classifiers trained with PCA after TARQUIN performed worse than those with TARQUIN alone. This might happen due to the fact that only the components explaining the 85% are taken into account with PCA and, applied after TARQUIN, might omit some relevant information. When treating simple signals like Long-TE, where the variability and amount of information is lower than in Short-TE, PCA classifiers worked slightly better. Analogously, the performance achieved with TARQUIN was similar to the performance obtained with PI and PCA. These results agree with the ones reported by Luts *et al.* [28]. In this study, several multivariate feature extraction techniques were applied to MRS. In addition, two quantitation algorithms, AQSES and QUEST, were applied. The performance of classifiers trained with traditional PR feature extraction techniques was slightly better to the performance achieved when the classifiers were trained with the quantitation information. They conclude that, for classification purposes, the use of simpler feature extraction techniques,

which can be easily automated and used as a black-box pre-process method, are preferable to using sophisticated metabolite quantitation algorithms which require computational availability and are difficult to implement and automate.

Nevertheless, metabolite quantitation might be useful if a biological explanation is desired. Processing MRS data to yield a metabolite profile prior to diagnostic classification provides the concentrations of specific metabolites for additional analysis. Expertise in interpreting these values can be gained by the radiologist and thereby add to the confidence in the result from the diagnostic classifier. The TARQUIN software is a highly automated and stable method for determining metabolite concentrations from MRS data and allows a non-expert user to process MRS spectra at various echo times without difficulty.

The use of TARQUIN quantitation in automatic DSSs provides a powerful clinical tool. The CURIAM DSS incorporates the classifiers developed in this work, offering the possibility of giving advice both to adult and children cases. CURIAM [46, 47, 48] has been evaluated at several hospitals in Europe providing decision support for brain tumour diagnosis based on  $^1\text{H}$  MRS.

## 4.5 Concluding remarks

$^1\text{H}$  MRS data was collected at diagnosis from children with brain tumours in 10 international centres in Europe and South America and was used to test the ability of MRS to discriminate between different tumour types. Our results show that particularly high diagnostic accuracies are achieved when MRS is collected at two TEs and that this accuracy can be achieved with data collected from multiple centres. MRS with automated processing and pattern recognition provides a useful technique for accurate, non-invasive diagnosis and classification of childhood brain tumours and thereby a powerful diagnostic tool for clinical practice.

## Bibliography

- [1] J. Vicente, E. Fuster-Garcia, S. Tortajada, J. M. García-Gómez, N. Davies, K. Natarajan, M. Wilson, R. G. Grundy, P. Wesseling, D. Monleón, B. Celda, M. Robles, and A. C. Peet, “Accurate classification of childhood brain tumours by *in vivo*  $^1\text{H}$  MRS – a multi-centre study,” *European Journal of Cancer*, 2012. In Press.

- [2] J. Vicente, E. Fuster-Garcia, S. Tortajada, J. M. García-Gómez, N. Davies, K. Natarajan, M. Wilson, R. G. Grundy, P. Wesseling, D. Monleón, B. Celda, M. Robles, and A. C. Peet, “Accurate classification of childhood brain tumours by *in vivo*  $^1\text{H}$  MRS – a multi-centre study,” in *Proceedings of the 15th International Symposium on Pediatric Neuro-Oncology, Toronto, Canada*, 2012. In press.
- [3] R. Venkatramani, G. Dhall, M. Patel, J. Grimm, C. Hawkins, G. McComb, M. Krieger, K. Wong, S. O’Neil, and J. L. Finlay, “Supratentorial ependymoma in children: To observe or to treat following gross total resection?,” *Pediatric Blood & Cancer*, 2011. on-line.
- [4] B. Horn, R. Heideman, R. Geyer, I. Pollack, R. Packer, J. Goldwein, T. Tomita, P. Schomberg, J. Ater, L. Luchtman-Jones, K. Rivlin, K. Lamborn, M. Prados, A. Bollen, M. Berger, G. Dahl, E. McNeil, K. Patterson, D. Shaw, M. Kubalik, and C. Russo, “A multi-institutional retrospective study of intracranial ependymoma in children: Identification of risk factors,” *Journal of Pediatric Hematology/Oncology*, vol. 21, no. 3, pp. 203–211, 1999.
- [5] A. L. Albright, J. H. Wisoff, P. M. Zeltzer, J. M. Boyett, L. B. Rorke, and P. Stanley, “Effects of medulloblastoma resections on outcome in children: A report from the children’s cancer group,” *Neurosurgery*, vol. 38, no. 2, pp. 265–271, 1996.
- [6] H. Ogiwara, R. M. Bowman, and T. Tomita, “Long-term follow-up of pediatric benign cerebellar astrocytomas,” *Neurosurgery*, vol. 70, no. 1, pp. 40–8, 2012.
- [7] P. L. Kubben, K. J. ter Meulen, O. E. Schijns, M. P. ter Laak-Poort, J. J. van Overbeeke, and H. van Santbrink, “Intraoperative MRI-guided resection of glioblastoma multiforme: a systematic review,” *The Lancet Oncology*, vol. 12, no. 11, pp. 1062–1070, 2011.
- [8] D. Ellison, M. Kocak, J. Dalton, and et al., “Definition of disease-risk stratification groups in childhood medulloblastoma using combined clinical, pathologic, and molecular variables,” *Journal of Clinical Oncology*, vol. 29, no. 11, pp. 1400–1407, 2010.
- [9] A. Peet, N. Davies, L. Ridley, and et al., “Magnetic resonance spectroscopy suggests key differences in the metastatic behaviour of medulloblastoma,” *European Journal of Cancer*, vol. 6, no. 43, pp. 1037–1044, 2007.

- 
- [10] A. R. Tate, J. Underwood, D. M. Acosta, M. Julià-Sapé, C. Majós, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, M. E. Cabañas, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arús, “Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra,” *NMR in Biomedicine*, vol. 19, no. 4, pp. 411–434, 2006.
- [11] G. Hagberg, “From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods,” *NMR in Biomedicine*, vol. 11, no. 4-5, pp. 148–156, 1998.
- [12] B. H. Menze, M. P. Lichy, P. Bachert, B. M. Kelm, H.-P. Schlemmer, and F. A. Hamprecht, “Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors,” *NMR in Biomedicine*, vol. 19, pp. 599–609, Aug 2006.
- [13] L. Lukas, A. Devos, J. A. K. Suykens, L. Vanhamme, F. A. Howe, C. Majós, A. Moreno-Torres, M. V. D. Graaf, A. R. Tate, C. Arús, and S. V. Huffel, “Brain tumor classification based on long echo proton MRS signals,” *Artificial Intelligence in Medicine*, vol. 31, pp. 73–89, 2004.
- [14] A. W. Simonetti, W. J. Melssen, F. Szabo de Edelenyi, J. J. A. van Asten, A. Heerschap, and L. M. C. Buydens, “Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification,” *NMR in Biomedicine*, vol. 18, pp. 34–43, Feb 2005.
- [15] A. Devos, L. Lukas, J. A. K. Suykens, L. Vanhamme, A. R. Tate, F. A. Howe, C. Majos, A. Moreno-Torres, M. van der Graaf, C. Arús, and S. Van Huffel, “Classification of brain tumours using short echo time 1H MR spectra,” *Journal of Magnetic Resonance*, vol. 170, pp. 164–175, Sep 2004.
- [16] Y. Huang, P. J. G. Lisboa, and W. El-Deredy, “Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection,” *Statistics in Medicine*, vol. 22, pp. 147–164, Jan 2003.
- [17] J. M. García-Gómez, J. Luts, M. Julià-Sapé, P. Krooshof, S. Tortajada, J. Vicente, W. Melssen, E. Fuster-Garcia, I. Olier, G. Postma, D. Monleón, A. Moreno-Torres, J. Pujol, A. P. Candiota, M. C. Martínez-Bisbal,

- J. Suykens, L. Buydens, B. Celda, S. Van Huffel, C. Arús, and M. Robles, "Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy," *Magnetic Resonance Materials Physics, Biology and Medicine*, vol. 22, no. 1, pp. 5–18, 2009.
- [18] I. F. Pollack, "Brain tumors in children [review]," *New England Journal of Medicine*, vol. 331, pp. 1501–1507, 1994.
- [19] I. F. Pollack, "Pediatric brain tumors," *Seminars in Surgical Oncology*, vol. 16, pp. 73–90, 1999.
- [20] K. Imamura, "Proton MR spectroscopy of the brain with a focus in chemical issues," *Magnetic Resonance in Medicine*, vol. 2, no. 3, pp. 117–132, 2003.
- [21] A. Rossi, C. Gandolfo, G. Morana, M. Severino, M. Garrè, and A. Cama, "New MR sequences (diffusion, perfusion, spectroscopy) in brain tumours," *Pediatric Radiology*, vol. 40, pp. 999–1009, June 2010.
- [22] J. M. García-Gómez, S. Tortajada, C. Vidal, M. Julià-Sapé, J. Luts, À. Moreno-Torres, S. Van Huffel, C. Arús, and M. Robles, "The effect of combining two echo times in automatic brain tumor classification by MRS," *NMR in Biomedicine*, vol. 21, no. 10, pp. 1112–1125, 2008.
- [23] eTUMOUR Consortium, "eTumour: Web accessible MR Decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data." Web site. FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC, <http://www.etumour.net> (Accessed: 22nd April 2012; at writing time it was temporarily unavailable).
- [24] M. van der Graaf, M. Julià-Sapé, F. A. Howe, A. Ziegler, C. Majós, A. Moreno-Torres, M. Rijpkema, D. Acosta, K. S. Opstad, Y. M. van der Meulen, C. Arús, and A. Heerschap, "MRS quality assessment in a multicentre study on MRS-based classification of brain tumours," *NMR in Biomedicine*, vol. 21, no. 2, pp. 148–158, 2008.
- [25] D. N. Louis, H. Ohgaki, O. D. Wiestler, and W. K. Cavenee, *WHO classification of tumours of the central nervous system*. Lyon, France: IARC Press, 2007.
- [26] M. Julià-Sapé, D. Acosta, M. Mier, C. Arús, and D. Watson, "A multicentre, web-accessible and quality control-checked database of in vivo

- 
- MR spectra of brain tumour patients,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 19, pp. 22–33, Feb 2006.
- [27] M. Wilson, G. Reynolds, R. a. Kauppinen, T. N. Arvanitis, and A. C. Peet, “A constrained least-squares approach to the automated quantitation of in vivo  $^1\text{H}$  magnetic resonance spectroscopy data,” *Magnetic Resonance in Medicine*, vol. 65, pp. 1–12, Jan. 2011.
- [28] J. Luts, J. Poulet, J. M. García-Gómez, A. Heerschap, M. Robles, J. A. K. Suykens, and S. Van Huffel, “Effect of feature extraction for brain tumor classification based on short echo time  $^1\text{H}$  MR spectra,” *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 288–298, 2008.
- [29] J. Slotboom, A. Nirko, C. Brekenfeld, and D. van Ormondt, “Reliability testing of in vivo magnetic resonance spectroscopy (MRS) signals and signal artifact reduction by order statistic filtering,” *Measurement Science and Technology*, vol. 20, no. 10, pp. 1040–30, 2009.
- [30] K. Fukunaga, *Introduction to statistical pattern recognition (2nd ed.)*. San Diego, CA, USA: Academic Press Professional, Inc., 1990.
- [31] C. J. Burges, “Geometric methods for feature extraction and dimensional reduction: A guided tour,” tech. rep., Microsoft Research, University of Toronto, 2004.
- [32] Z. Wang, L. N. Sutton, A. Cnaan, J. C. Haselgrove, L. Rorke, H. Zhao, L. T. Bilaniuk, and R. A. Zimmerman, “Proton MR spectroscopy of pediatric cerebellar tumors,” *American Journal of Neuroradiology*, vol. 16, pp. 1821–1833, 1995.
- [33] J. E. Arle, C. Morriss, Z. J. Wang, R. A. Zimmerman, P. G. Phillips, and L. N. Sutton, “Prediction of posterior fossa tumor type in children by means of magnetic resonance image properties, spectroscopy, and neural networks,” *Journal of Neurosurgery*, vol. 86, pp. 755–761, 1997.
- [34] J. F. Schneider, S. Confort-Gouny, A. Viola, Y. Le Fur, P. Viout, M. Ben-nathan, F. Chapon, D. Figarella-Branger, P. Cozzone, and N. Girard, “Multiparametric differentiation of posterior fossa tumors in children using diffusion-weighted imaging and short echo-time  $^1\text{H}$ -MR Spectroscopy,” *Journal of Magnetic Resonance Imaging*, vol. 26, no. 1, pp. 1390–1398, 2007.



- [35] N. P. Davies, M. Wilson, L. M. Harris, K. Natarajan, S. Lateef, L. Macpherson, S. Sgouros, R. G. Grundy, T. N. Arvanitis, and A. C. Peet, "Identification and characterisation of childhood cerebellar tumours by in vivo proton MRS," *NMR in Biomedicine*, vol. 21, no. 8, pp. 908–918, 2008.
- [36] R. A. Fisher, *Statistical methods for research workers*. Edinburgh, UK: Oliver and Boyd, 1925.
- [37] J. H. Friedman, "Regularized Discriminant Analysis," *Journal of American Statistical Association*, no. 405, pp. 165–175, 1989.
- [38] Y. Hochberg and A. C. Tamhane, *Multiple Comparison Procedures (Wiley Series in Probability and Statistics)*. Wiley, 1987.
- [39] M. C. Preul, Z. Caramanos, D. L. Collins, J. G. Villemure, R. Leblanc, A. Olivier, R. Pokrupa, and D. L. Arnold, "Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy," *Nature Medicine*, vol. 2, no. 3, pp. 323–325, 1996.
- [40] A. Panigrahy, M. D. Krieger, I. González-Gómez, X. Liu, J. G. McComb, J. L. Finlay, M. D. Nelson, F. Gilles, and S. Blüml, "Quantitative short echo time  $^1\text{H}$ -MR spectroscopy of untreated pediatric brain tumors : Preoperative diagnosis and characterization," *American Journal of Neuroradiology*, vol. 27, pp. 560–572, 2006.
- [41] A. C. Peet, S. Lateef, L. MacPherson, K. Natarajan, S. Sgouros, and R. G. Grundy, "Short echo time  $^1\text{H}$  magnetic resonance spectroscopy of childhood brain tumours," *Child's Nervous System*, vol. 23, no. 2, pp. 163–9, 2007.
- [42] A. U. Provencher, "Estimation of metabolite concentrations from localized *in vivo* proton NMR spectra," *Magnetic Resonance in Medicine*, vol. 30, pp. 672–679, 1993.
- [43] A. Moreno-Torres, I. Martínez-Pérez, M. Baquero, J. Campistol, A. Capdevila, C. Arús, and J. Pujol, "Taurine detection by proton magnetic resonance spectroscopy in medulloblastoma: contribution to noninvasive differential diagnosis with cerebellar astrocytoma," *Neurosurgery*, vol. 4, no. 55, pp. 824–829, 2004.
- [44] C. Rémy, N. Fouilhé, I. Barba, E. Sam-Laï, H. Lahrech, M.-G. Cucurella, M. Izquierdo, A. Moreno, A. Ziegler, R. Massarelli, M. Décorps, and C. Arús, "Evidence that mobile lipids detected in rat brain glioma by  $^1\text{H}$

- 
- nuclear magnetic resonance correspond to lipid droplets,” *Cancer Research*, vol. 57, pp. 407–414, 1997.
- [45] I. Martínez-Pérez, A. Moreno, I. Barba, A. Capdevila, and C. Arús, “Large lipid droplets observed by electron microscopy in six human brain tumors with lipid 1H MRS *in vivo* pattern,” in *Abstracts of the International Society for Magnetic Resonance in Medicine, New York*, p. 976, 1996.
- [46] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, M. Esparza, A. Navarro, E. Fuster-Garcia, M. Robles, L. Martí-Bonmatí, and C. Arús, “A generic decision support system featuring an assembled view of predictive models for magnetic resonance and clinical data,” in *Book of abstracts of the ESMRMB 2008: 25th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Valencia, Spain*, p. 483, Springer-Verlag, Oct. 2008.
- [47] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, E. Fuster-Garcia, M. Esparza, A. Navarro, and M. Robles, “Curiam BT 1.0, Decision Support System for Brain Tumour Diagnosis,” in *Book of Abstracts of the ESMRMB: 26th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Antalya, Turkey*, vol. 22, p. 538, Springer Berlin / Heidelberg, May 2009.
- [48] C. Sáez, J. M. García-Gómez, J. Vicente, S. Tortajada, J. Luts, D. Dupplaw, S. V. Huffel, and M. Robles, “A generic and extensible automatic classification framework applied to brain tumour diagnosis in HealthAgents,” *The Knowledge Engineering Review*, vol. 26, no. Special Issue 03, pp. 283–301, 2011.

## Chapter 5

# Audit method suited for DSS in clinical environment

*“Don’t judge each day by the harvest you reap but by the seeds that you plant.”*

*Robert Louis Stevenson*

This chapter presents a novel on-line method to audit predictive models using a Bayesian perspective. The auditing model has been specifically designed for DSSs suited for clinical or research environments. Taking as starting point the working diagnosis supplied by the clinician, this method compares and evaluates the predictive skills of those models able to answer to that diagnosis. The approach consists in calculating the posterior odds of a model through the composition of a prior odds, a static odds and a dynamic odds. To do so, this method estimates the posterior odds from the cases that the comparing models had in common during the design stage and from the cases already viewed by the DSS after deployment in the clinical site. In addition, if an ontology of the classes is available, this method can audit models answering related questions, which offers a reinforcement to the decisions the user already took and gives orientation on further diagnostic steps. The main technical novelty of this approach lies in the design of an audit model adapted to suit the decision workflow of a clinical environment. The audit model allows deciding which is the classifier that best

suits each particular case under evaluation and allows the detection of possible misbehaviours due to population differences or data shifts in the clinical site. We show the efficacy of our method for the problem of brain tumour diagnosis with MRS.

*The core of this chapter has been published as a chapter of proceedings in [1] and in [2]. Section 5.5 depicts how the audit method can effectively influence the user's decision process. It also gives a description of how it can be combined with the incremental learning algorithms described in [3].*

## 5.1 Introduction

Early studies focused on evaluation of predictive models in a DSS expected the models to be able to predict ‘correct’ diagnosis by examining the diagnostic accuracy of the DSS functioning in isolation [4, 5]. Recent evaluations, though, balance the value of testing the system and the impact of the DSS on the user’s diagnostic plans [6, 7]. This means that the suggestions made by the DSS should positively influence the user’s diagnostic reasoning.

In order to make a DSS useful for routine clinical use, a trustworthiness feeling needs to be created in the clinician. Thus, for a clinical diagnosis DSS based on predictive models obtained with inference methods, showing the performance evaluated in laboratory might not suffice.

Let us assume that a DSS contains  $M$  models of classification and that every model has been trained with the same data  $\mathbf{Z} = \{(\mathbf{x}_j, t_j)\}_1^N$ , a set of  $N$  samples where  $\mathbf{x}_j$  is a data vector describing the  $j^{th}$  sample and  $t_j$  is its associated label. The posterior probability of a model  $\mathcal{M}_i$  can be expressed as:

$$P(\mathcal{M}_i|\mathbf{Z}) = \frac{P(\mathcal{M}_i)P(\mathbf{Z}|\mathcal{M}_i)}{P(\mathbf{Z})}, \quad (5.1)$$

where  $P(\mathbf{Z}|\mathcal{M}_i)$  is the model likelihood or evidence for  $\mathbf{Z}$ . The term  $P(\mathcal{M}_i)$  is a ‘subjective’ prior over the model space which expresses our prior believe on the basis of experience. This term is typically overwhelmed by the objective term, the evidence [8].  $P(\mathbf{Z})$  is usually ignored since it is assumed that models are compared for the same  $\mathbf{Z}$ .

Typically, when a trained and evaluated classifier is introduced in a DSS, it is assumed that its predictive performance will remain in the course of time. Such assumption, though, may be unrealistic, especially in biomedical domains where dynamic conditions of the environment may change the assumed conditions in the models: modification of the data distribution,  $P(\mathbf{Z})$  (covariate shift [9, 10]), inclusion of new classes through time, which modifies the prior  $P(\mathcal{M}_i)$ , (prior probability shift [10]) or a change in the definition of the classes itself,  $P(\mathbf{Z}|\mathcal{M}_i)$ , (concept shift [10, 11, 12]) might take place.

In order to give guidance in the user's diagnostic workflow, we propose a method that, taking as starting point the diagnosis supplied by the clinician, compares and evaluates the predictive skills of those models able to answer to such diagnosis. In addition, this auditing process should also be capable of comparing those predictive models able to answer more general diagnosis (superclasses) since they could serve as a mechanism to reinforce the decisions already taken by the clinician. Analogously, audit of predictive models discriminating subclasses of the initial diagnosis is also desirable since these predictive models might give guidance on the next steps to take in order to refine his/her diagnostic process. Such comparisons can be performed by our method if an ontology describing the relationships among the different diagnosis labels is available.

The Bayesian paradigm offers a model comparison framework that allows it to objectively assess the predictive skills of two or more classifiers by comparing the posterior odds. This approach has been typically followed for model selection under the assumption that all the models are trained with the same data. Nevertheless, this is not the case when deployed classifiers addressing similar and related discriminations have to be compared. The proposed method overcomes the limitation of the model comparison under the Bayesian paradigm and allows the comparison of predictive models trained with different datasets and answering related questions.

The capabilities of this method are shown for the problem of brain tumour diagnosis with MRS. The results obtained reveal the proposed method as able to objectively compare predictive models answering related problems but also to take part in the physicians' diagnostic decisions, contributing to assess the role and potential benefits of the DSS in real clinical setting.

## 5.2 Methods

### 5.2.1 Bayesian approach

To compare two models  $\mathcal{M}_m$  and  $\mathcal{M}_l$  we form the posterior odds

$$\frac{P(\mathcal{M}_m|\mathbf{Z})}{P(\mathcal{M}_l|\mathbf{Z})} = \frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} \frac{P(\mathbf{Z}|\mathcal{M}_m)}{P(\mathbf{Z}|\mathcal{M}_l)}. \quad (5.2)$$

If the odds are greater than one we choose model  $m$ , otherwise we choose model  $l$ . If we assume a uniform distribution of the prior probabilities  $P(\mathcal{M}_i)$ , models  $\mathcal{M}_i$  are ranked by evaluating the evidence [13].

$\frac{P(\mathbf{Z}|\mathcal{M}_m)}{P(\mathbf{Z}|\mathcal{M}_l)}$  is a ratio of the evidences and is called the Bayes Factor (BF), the contribution of the data toward the posterior odds [14].

Several techniques are available for computing BF. An exhaustive review can be found in [13]. The Bayesian Information Criterion (BIC) gives a rough approximation to the logarithm of  $P(\mathbf{Z}|\mathcal{M}_i)$  and can be used for calculating Equation 5.2 [15]. We consider BIC in this study to calculate an approximation to BF.

### 5.2.2 Comparison of models adapted to a clinical environment

In order to take part in the decision workflow of a clinician, a DSS for diagnosis based on inference models deployed in a clinical environment should inform the user, according to his/her proposed diagnosis, about which of the predictive models available are going to give a useful advise. Let us call  $L$  to the set of labels supplied by the clinician as working diagnosis, which is the preliminary diagnosis given by the clinician and is based on experience, clinical epidemiology, and early confirmatory evidence provided by ancillary studies. A sensible mechanism to decide which predictive models are audited is selecting those able to discriminate  $L$  or, at least, some of the labels  $l_j$  in  $L$ .

Additionally, the predictive models may be trained from different sets of data acquired from different patients and centres. Let us call  $\mathbf{Z}_i$  to the arbitrary set of samples each model  $\mathcal{M}_i$  has been estimated with.

In order to apply the Bayesian framework, the  $\mathcal{M}_i$  models are required to be compared with a common dataset  $\mathbf{Z}$ . We propose to obtain  $\mathbf{Z}$  from the samples of each  $\mathbf{Z}_i$  labeled with  $L$ .

Furthermore, most of biomedical problems show equivalences and polymorphisms in the classes involved in the discrimination process. This is the case, for example, of brain tumour diagnosis where, depending on the detail level of the addressed question, a tumour can be named with different terms. Thus, depending on the detail in the diagnosis, a glioblastoma can be labeled as aggressive tumour or high grade glial tumour [16]. This variety of labels for the same concept can be depicted in an ontology. An ontology is a specification of a conceptualisation that consists of a poset (partially ordered set) of concept types, a poset of relations between these concepts and, sometimes, a set of instances of the concepts [17]. Let us call correspondence table,  $CT$ , to a tabular structure that reflects the hierarchy between classes where each column indicates an ‘is a’ relationship. An example of  $CT$  for brain tumour types is given in Table 5.1.

We propose a method capable of auditing models related through the hierarchy in the classes they discriminate. If we define the function  $m_{sl}(t)$  that makes reference to the ‘most specific label’ of each set of input variables  $\mathbf{x}$ , we can define  $z^* = (\mathbf{x}, m_{sl}(t))$  as the sample  $z$  labeled with its most specific label (the left-most column of the  $CT$ ). Figure 5.1 shows an example of the labels  $T_i$  and  $m_{sl}(T_i)$  of four corpora for brain tumour discrimination.

To apply our method for comparing  $\mathcal{M}_i$  models estimated from  $\mathbf{Z}_i$ , we need to know the set of labels  $T_i$  each model can discriminate among. Besides, it is assumed that we can express any sample  $z$  into its most specific label,  $z^*$ . Then, we can produce  $\mathbf{Z}$  as the union of the samples in  $\mathbf{Z}_i$  having a most specific label equivalent to the most specific label of any element of  $L$ . Formally:

$$\begin{aligned} \mathbf{Z} &= \{z^{*(k)} | z^{*(k)} = (\mathbf{x}^{(k)}, m_{sl}(t^{(k)})), m_{sl}(t^{(k)}) \in \{m_{sl}(l_j)\}\}, \\ & \quad k = 1, \dots, |\mathbf{Z}^\cup|, \\ \mathbf{Z}^\cup &= \bigcup_{i=1}^M \mathbf{Z}_i, \\ & \quad j = 1, \dots, |L|. \end{aligned} \tag{5.3}$$

In order to calculate the evidence from  $\mathbf{Z}$ , we need a mechanism to transform the labels in  $\mathbf{Z}$  into the labels  $T_i$  that a model  $\mathcal{M}_i$  understands. Let us define

Table 5.1: Correspondence table (CT) based on WHO classification of tumours of the central nervous system.

WHO Label	is an aggressive <sup>a</sup>	is a gII glial <sup>b</sup>	is a grade I-II <sup>c</sup>	is a grade III-IV <sup>d</sup>	is a men <sup>e</sup>
GLIOBLASTOMA	✓	—	—	✓	—
METASTASIS	✓	—	—	✓	—
ANAPLASTIC ASTROCYTOMA	✓	—	—	✓	—
ANAPLASTIC OLIGOASTROCYTOMA	✓	—	—	✓	—
ANAPLASTIC OLIGODENDROGLIOMA	✓	—	—	✓	—
DIFFUSE ASTROCYTOMA	—	✓	✓	—	—
OLIGOASTROCYTOMA	—	✓	✓	—	—
OLIGODENDROGLIOMA	—	✓	✓	—	—
PILOCYTIC ASTROCYTOMA	—	—	✓	—	—
FIBROUS MENINGIOMA	—	—	✓	—	✓
MENINGIOMA	—	—	✓	—	✓
MENINGOTHELIAL MENINGIOMA	—	—	✓	—	✓

<sup>a</sup> aggressive tumour

<sup>b</sup> glial tumour grade II

<sup>c</sup> grade I or II tumour type

<sup>d</sup> grade III or IV tumour type

<sup>e</sup> meningioma grade II



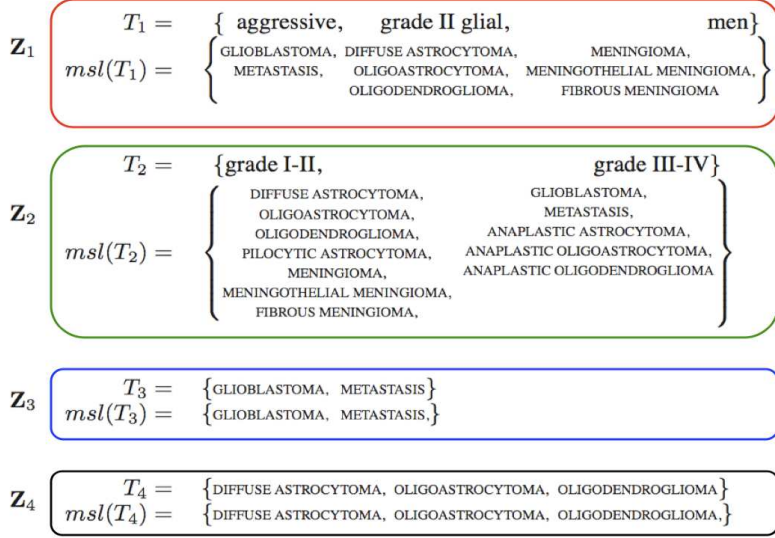


Figure 5.1: Example of corpora for brain tumour diagnosis. Each corpus  $\mathbf{Z}_i$  has samples labeled as  $T_i$ , grouped from the tumour types defined in  $msl(T_i)$ , according to the  $CT$  in Table 5.1.

$\mathbf{Z}^{T_i} = \mathcal{L}(\mathbf{Z}, T_i, CT)$  to be the result of applying the function  $\mathcal{L}$  which transforms the labels of  $\mathbf{Z}$  into one of the labels specified in  $T_i$  according to  $CT$ . This mechanism will allow the models to explain the data according to the labels they discriminate. Formally:

$$P(\mathcal{M}_i|\mathbf{Z}) = P(\mathcal{M}_i|\mathbf{Z}^{T_i}). \quad (5.4)$$

Figure 5.2 illustrates the process for obtaining the common  $\mathbf{Z}$  and shows how the function  $\mathcal{L}$  works in two different scenarios.

Once  $\mathbf{Z}^{T_i}$  has been calculated for each model, we can perform the comparison

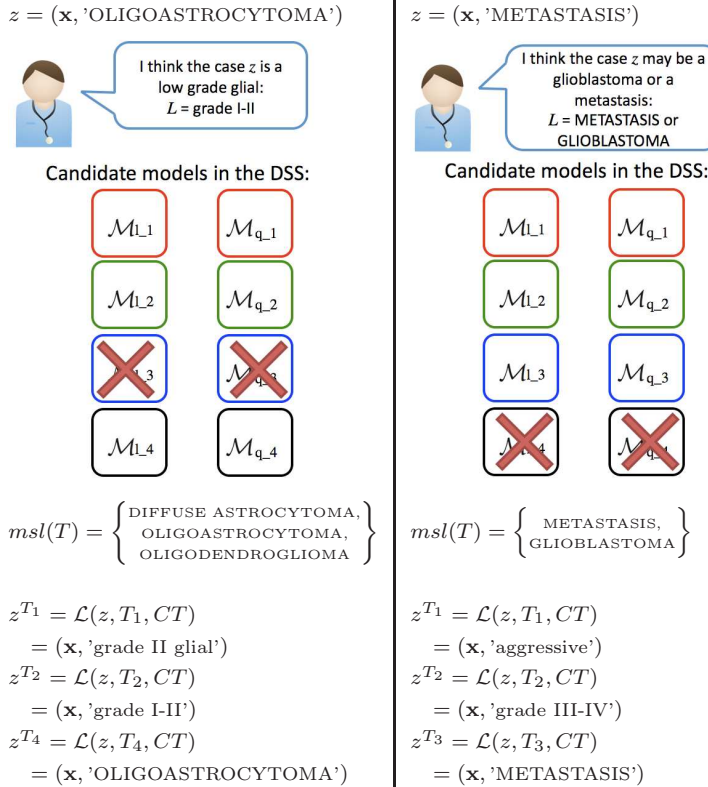


Figure 5.2: Example of two scenarios: A clinician, who does not know the diagnostic label of the sample  $z$ , introduces  $z$  in the DSS and proposes a working diagnosis. Then, a selection of the models in the DSS is performed attending to that working diagnosis. Models  $\mathcal{M}_{l_i}$  and  $\mathcal{M}_{q_i}$  are linear and quadratic Gaussian modeled from the  $\mathbf{Z}_i$  defined in Figure 5.1. Once the models are selected, the common set of cases labels,  $msl(T)$  is obtained to build  $\mathbf{Z}$  to calculate the static odds. Examples of the function  $\mathcal{L}$  that maps the label of a sample  $z$  into one of the labels  $T_i$  (according to the  $CT$ ) are also given.

of two models  $\mathcal{M}_m$  and  $\mathcal{M}_l$  in the light of  $\mathbf{Z}$  forming the posterior odds

$$\frac{P(\mathcal{M}_m|\mathbf{Z})}{P(\mathcal{M}_l|\mathbf{Z})} = \frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} \frac{P(\mathbf{Z}^{T_m}|\mathcal{M}_m)}{P(\mathbf{Z}^{T_l}|\mathcal{M}_l)}. \quad (5.5)$$

### 5.2.3 Audit of dynamic performances

If the DSS is able to store the data introduced by the users, an audit according to the predictive performances for these data is possible. This dynamic auditing will give a real vision of the predictive skills of the classifiers in the environment of the DSS.

Let us suppose that  $\mathbf{Z}_{DSS}$  is a dataset of new samples introduced into the DSS by the users and that it is composed by a set of samples different to any sample in  $\mathbf{Z}$ . If we assume that samples in  $\mathbf{Z}$  and  $\mathbf{Z}_{DSS}$  are independent and identically distributed (i.i.d.), we can calculate the next posterior odds for comparing  $\mathcal{M}_m$  and  $\mathcal{M}_l$ :

$$\begin{aligned} \frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} &= \frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} \frac{P(\mathbf{Z}, \mathbf{Z}_{DSS}|\mathcal{M}_m)}{P(\mathbf{Z}, \mathbf{Z}_{DSS}|\mathcal{M}_l)} = \\ &= \frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} \frac{P(\mathbf{Z}|\mathcal{M}_m)}{P(\mathbf{Z}|\mathcal{M}_l)} \frac{P(\mathbf{Z}_{DSS}|\mathbf{Z}, \mathcal{M}_m)}{P(\mathbf{Z}_{DSS}|\mathbf{Z}, \mathcal{M}_l)}. \end{aligned} \quad (5.6)$$

We have assumed that the samples in  $\mathbf{Z}$  are i.i.d.. Therefore, we can split  $\mathbf{Z} = \mathbf{Z}_i \cup \mathbf{Z}_{-i}$ , where  $\mathbf{Z}_i$  is the data used to design a model  $\mathcal{M}_i$ , and  $\mathbf{Z}_{-i} = (\mathbf{Z} \setminus \mathbf{Z}_i)$  is the set of samples not used in  $\mathcal{M}_i$  estimation. Notice that the set of samples  $\mathbf{Z}_{-i}$  is independent to  $\mathbf{Z}_i$  because they are i.i.d. and it is also independent to  $\mathcal{M}_i$  because its samples has not been used to design  $\mathcal{M}_i$ . Then, Equation 5.6 can be rewritten as:

$$\frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} = \frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} \frac{P(\mathbf{Z}|\mathcal{M}_m)}{P(\mathbf{Z}|\mathcal{M}_l)} \frac{P(\mathbf{Z}_{DSS}|\mathbf{Z}_{-m}, \mathbf{Z}_m, \mathcal{M}_m)}{P(\mathbf{Z}_{DSS}|\mathbf{Z}_{-l}, \mathbf{Z}_l, \mathcal{M}_l)}. \quad (5.7)$$

Hence, Equation 5.7 can be simplified to

$$\frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} = \frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} \frac{P(\mathbf{Z}|\mathcal{M}_m)}{P(\mathbf{Z}|\mathcal{M}_l)} \frac{P(\mathbf{Z}_{DSS}|\mathbf{Z}_m, \mathcal{M}_m)}{P(\mathbf{Z}_{DSS}|\mathbf{Z}_l, \mathcal{M}_l)}. \quad (5.8)$$

To perform the calculation of the posterior of each model according to  $\mathbf{Z}_{DSS}$ , we need to apply our mechanism to translate the labels in  $\mathbf{Z}_{DSS}$  into the labels,  $T_i$ , that a model  $\mathcal{M}_i$  understands as in Equation 5.4:

$$P(\mathcal{M}_i|\mathbf{Z}_{DSS}) = P(\mathcal{M}_i|\mathbf{Z}_{DSS}^{T_i}). \quad (5.9)$$

Where  $\mathbf{Z}_{DSS}^{T_i} = \mathcal{L}(\mathbf{Z}_{DSS}, T_i, CT)$ , the result of applying the function  $\mathcal{L}$  which transforms the labels of  $\mathbf{Z}_{DSS}$  into one of the labels specified in  $T_i$  according to  $CT$ .

Thus, applying Equation 5.9 to Equation 5.8 we obtain:

$$\frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} = \underbrace{\frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)}}_{\text{Posterior odds}} \underbrace{\frac{P(\mathbf{Z}^{T_m}|\mathcal{M}_m)}{P(\mathbf{Z}^{T_l}|\mathcal{M}_l)}}_{\text{Static odds}} \underbrace{\frac{P(\mathbf{Z}_{DSS}^{T_m}|\mathbf{Z}_m, \mathcal{M}_m)}{P(\mathbf{Z}_{DSS}^{T_l}|\mathbf{Z}_l, \mathcal{M}_l)}}_{\text{Dynamic odds}}, \quad (5.10)$$

where  $\mathbf{Z}^{T_i} = \mathbf{Z}_i^{T_i} \cup \mathbf{Z}_{\neg i}^{T_i}$ . Since  $\mathbf{Z}_i^{T_i} = \mathcal{L}(\mathbf{Z}_i, T_i, CT) = \mathbf{Z}_i$ , the relation between  $\mathbf{Z}^{T_i}$  and  $\mathbf{Z}_i$  is:  $\mathbf{Z}^{T_i} = \mathbf{Z}_i \cup \mathbf{Z}_{\neg i}^{T_i}$ .

We call static odds to  $\frac{P(\mathbf{Z}^{T_m}|\mathcal{M}_m)}{P(\mathbf{Z}^{T_l}|\mathcal{M}_l)}$  because it compares the prediction abilities of both models with  $\mathbf{Z}$ , which has been produced from the  $\mathbf{Z}_i$  used to tune the parameters of each model  $\mathcal{M}_i$ .

Analogously, we call  $\frac{P(\mathbf{Z}_{DSS}^{T_m}|\mathbf{Z}_m, \mathcal{M}_m)}{P(\mathbf{Z}_{DSS}^{T_l}|\mathbf{Z}_l, \mathcal{M}_l)}$  the dynamic odds because it measures the predictive ratio of the two models with respect to a set of new samples  $\mathbf{Z}_{DSS}$  introduced in the DSS and not previously used during the design of the models.

Finally, to compare  $M$  models, Equation 5.10 can be generalized:

$$\frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{\sum_{l=1}^M P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} = \frac{P(\mathcal{M}_m)P(\mathbf{Z}^{T_m}|\mathcal{M}_m)P(\mathbf{Z}_{DSS}^{T_m}|\mathbf{Z}_m, \mathcal{M}_m)}{\sum_{l=1}^M P(\mathcal{M}_l)P(\mathbf{Z}^{T_l}|\mathcal{M}_l)P(\mathbf{Z}_{DSS}^{T_l}|\mathbf{Z}_l, \mathcal{M}_l)}. \quad (5.11)$$

## 5.3 Evaluation

The evaluation of the audit method was performed with a multi-center database of MRS data of brain tumours. Gaussian discriminants were trained and audited according to the formulae described above, simulating the real scenario of a DSS working in a clinical environment.

### 5.3.1 Procedure

A database from a multi-center project was divided into two datasets. One dataset was used to create the predictive models that give support in a DSS. To do so, several interesting questions  $T_i$  were defined and  $\mathbf{Z}_i$  datasets were obtained according to each  $T_i$ . Linear and quadratic Gaussian discriminant models ( $\mathcal{M}_{l\_i}$  and  $\mathcal{M}_{q\_i}$ , respectively) were fed with each  $\mathbf{Z}_i$ . The other dataset,  $\mathbf{Z}_{DSS}$ , was used as an independent test set. Each sample in  $\mathbf{Z}_{DSS}$  represented a case introduced into the DSS by a clinician in order to obtain support. Each sample in  $\mathbf{Z}_{DSS}$  had an associated proposed diagnosis supplied by the clinician.

To simulate a DSS running in a real clinical setting, we assume that there is an internal order in the  $N$  samples of  $\mathbf{Z}_{DSS}$ . The procedure is as follows: the clinician introduces into the DSS the  $n$ -th sample from  $\mathbf{Z}_{DSS}$  for diagnosis support along with his/her proposed diagnosis  $L_n$ . Then, the static odds of each predictive model are calculated from  $\mathbf{Z}$ , a dataset obtained from the cases in the different  $\mathbf{Z}_i$  matching  $L_n$ . The dynamic odds are also calculated for each predictive model the set of cases previously introduced in the DSS,  $z_{DSS}^1 \dots z_{DSS}^{n-1}$ , that also match  $L_n$ . Finally, the posterior odds are calculated by combination of the static and dynamic odds as described in Equation 5.11. For this evaluation we assume equal priors for each model. Next chapter will address an approach to assess prior probabilities.

A randomized algorithm to set the internal order of the samples from  $\mathbf{Z}_{DSS}$  was used, repeating the procedure  $k$  times. This repetition procedure prevents from obtaining variance in our results when  $k$  is big enough. In this work,  $k$  was set to 100.

### 5.3.2 Database

The database used for this evaluation consisted of 682 Single Voxel  $^1\text{H}$  MRS signals at 1.5T at Short-TE (TE, 20-32 ms) from the European project eTUMOUR [18, 19]. Several studies applying PR-based feature extraction methods in combination with learning strategies to the eTUMOUR spectroscopy database have been previously reported [20, 21, 22].

The PI selection feature technique was applied to the spectroscopy data. This method has been described in detail in Section 3.3 of this Thesis. PI is a method described and successfully applied in [20] that allows a reduction from the whole spectra to 15 parameters. PI has proportionality to the concentration of the main metabolites in each spectra.

All the classes considered in this study are described in Table 5.1 and were based on the histological classification of the central nervous system tumours as described by the WHO Classification [23].

The eTUMOUR spectroscopy database was divided into two datasets. One dataset from where 4 corpora  $\mathbf{Z}_i$  were defined as described in Figure 5.1. From the other dataset,  $\mathbf{Z}_{DSS}$  was obtained containing only the samples labeled as ‘GLIOBLASTOMA’, ‘METASTASIS’, ‘DIFFUSE ASTROCYTOMA’, ‘OLIGOASTROCYTOMA’ or ‘OLIGODENDROGLIOMA’, that corresponds to the elements in  $T_3$  and  $T_4$  described in Figure 5.1. With such a  $\mathbf{Z}_{DSS}$ , we were able to simulate the two scenarios depicted in Figure 5.2: One scenario where the clinician would provide a general working diagnosis  $L$  of ‘grade I-II’ tumour when dealing with samples labeled as any of the elements in  $T_4$ ; and a second scenario where the clinician would express the differential diagnosis  $L$  of ‘METASTASIS or GLIOBLASTOMA’ when trying to diagnose samples labeled as any of the elements in  $T_3$ . Table 5.2 shows the number of samples available.

### 5.3.3 Classifiers

Gaussian discriminant have been selected to train the predictive models and evaluate the audit method. Parametric Gaussian discriminant functions can describe linear boundaries when the covariance matrices of all the classes are equal and quadratic decision boundaries if a covariance matrix is calculated per class [24].

Gaussian discriminant were chosen because calculating the complexity of each

Table 5.2: Number of brain tumour samples available for  $\mathbf{Z}_i$  and  $\mathbf{Z}_{DSS}$ . Each  $\mathbf{Z}_i$  discriminates the labels  $T_i$  described in Figure 5.1.

Label	$\mathbf{Z}_i$	$\mathbf{Z}_{DSS}$		
GLIOBLASTOMA	182	126		
METASTASIS	67	49		
ANAPLASTIC ASTROCYTOMA	17	0		
ANAPLASTIC OLIGOASTROCYTOMA	4	0		
ANAPLASTIC OLIGODENDROGLIOMA	9	0		
DIFFUSE ASTROCYTOMA	52	19		
OLIGOASTROCYTOMA	17	7		
OLIGODENDROGLIOMA	23	17		
PILOCYTIC ASTROCYTOMA	17	0		
FIBROUS MENINGIOMA	13	0		
MENINGIOMA	40	0		
MENINGOTHELIAL MENINGIOMA	23	0		
aggressive	279	175		
gII glial	92	43		
grade I-II	185	43		
grade III-IV	279	175		
men	76	0		

Dataset	Number of samples
$\mathbf{Z}_1$	447
$\mathbf{Z}_2$	464
$\mathbf{Z}_3$	249
$\mathbf{Z}_4$	92
$\mathbf{Z}_{DSS}$	218

Table 5.3: Number of parameters associated to each predictive model.

Complexity	$GL$ (Gauss linear)	$GQ$ (Gauss quadratic)
(in parameters)		
(2 outputs)	152	272
(3 outputs)	168	408

predictive model, which is required for the BIC criterion, is straightforward. Complexity in Gaussian discriminant is measured in terms of the cardinality of the mean vector, the covariance matrix and the prior probabilities associated to each class. Linear and quadratic Gaussian discriminants were calculated for each corpus. Table 5.3 shows the complexities associated to each model.

## 5.4 Results

According to the diagnosis proposed to each sample in  $\mathbf{Z}_{DSS}$ , two scenarios can be identified. A set of models is selected as auditing candidates (see Figure 5.2) attending to the proposed diagnosis. Then, the audit model calculates the performance of these relevant candidates. Figure 5.3 shows the average likelihood

calculated for both scenarios. The likelihood was averaged by  $k = 100$ , the times that the experiment was repeated to avoid variances in the results, and was calculated from all the samples in  $\mathbf{Z}_{DSS}$  whose proposed diagnosis  $L$  was ‘grade I-II’ (top) or ‘METASTASIS or GLIOBLASTOMA’ (bottom) and from the common  $\mathbf{Z}$  obtained from the  $\mathbf{Z}_i$  of the eligible predictive models according to  $L$ . In general, when the likelihood of the models is measured from  $\mathbf{Z}$ , better results are reported compared to the likelihood measured from  $\mathbf{Z}_{DSS}$ . This behaviour is expected because the samples in  $\mathbf{Z}$  have been used to tune the parameters of the predictive models, whilst the samples in  $\mathbf{Z}_{DSS}$  remain totally independent. Nevertheless, the differences in the average likelihood are of no importance for each predictive model except in the quadratic Gaussian model that discriminates  $T_4$ , where the likelihood from  $\mathbf{Z}_{DSS}$  drops dramatically compared to the likelihood from  $\mathbf{Z}$ . This may be due to the complexity inherent to the discrimination of the three classes in  $T_4$ .

Another interesting view of the usefulness of our audit method for clinical DSSs is depicted in Figure 5.4, which shows the evolution of the posterior odds. The X axis represents the samples of  $\mathbf{Z}_{DSS}$  introduced in the DSS in the course of time. When the clinician introduces the  $n$ -th sample from  $\mathbf{Z}_{DSS}$  along with  $L$ , the posterior odds of each eligible model are calculated from:  $\mathbf{Z}$ , a dataset obtained from the cases in the different  $\mathbf{Z}_i$  matching  $L$  (static odds) and from those cases previously introduced into the DSS,  $z_{DSS}^1 \dots z_{DSS}^{n-1}$ , which also match  $L$  (dynamic odds). Thus, the posterior odds in Figure 5.4 is calculated using the Equation 5.11 and covering the  $\mathbf{Z}_{DSS}$  so that the first point in the figure corresponds to the static posterior, (since no samples in  $\mathbf{Z}_{DSS}$  are evaluated yet), and the last point corresponds to the scenario where all the samples of  $\mathbf{Z}_{DSS}$  have been introduced in the DSS. Although, in the limit, all the models showed similar posterior odds, those models answering more general questions ( $T_1$  and  $T_2$ ) obtained a slightly better posterior than the models solving more specific questions ( $T_3$  and  $T_4$ ) for the two scenarios. These results are in agreement with previous studies [20], where classifiers discriminating superclasses of tumour types obtained better performances than classifiers discriminating more specific tumour groups. All the predictive models show stationary behaviour in the limit.



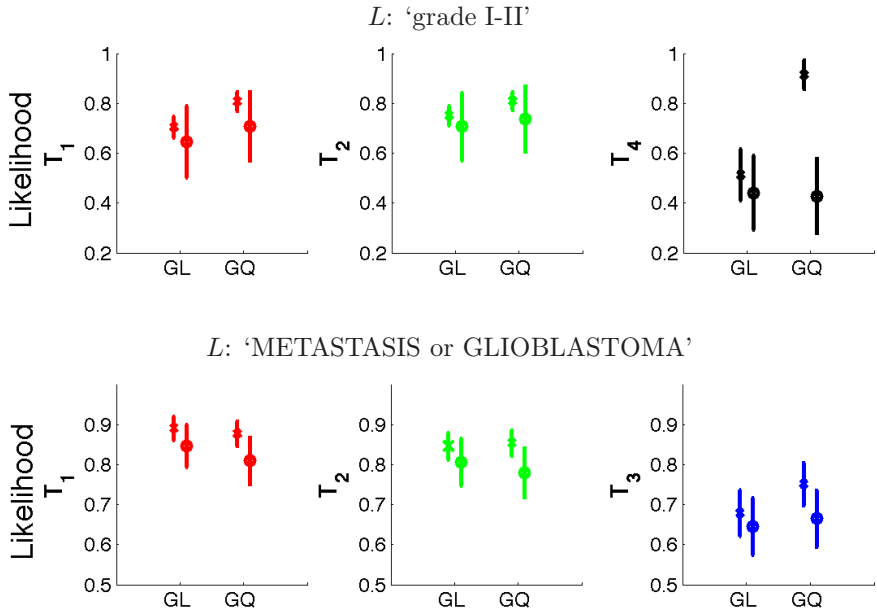


Figure 5.3: Average likelihood of the models according to the scenario where the clinician proposes  $L$  as 'grade I-II' (top) and the scenario where the working diagnosis  $L$  supplied is 'METASTASIS or GLIOBLASTOMA' (bottom), as described by Figure 5.2. Notice that  $T_1, \dots, T_4$  correspond to the discrimination labels described in Figure 5.1. The likelihood of the models are depicted with the error bars for  $Z$  (crosses) and for  $Z_{DSS}$  (circles) after repeating the experiment  $k = 100$  times.

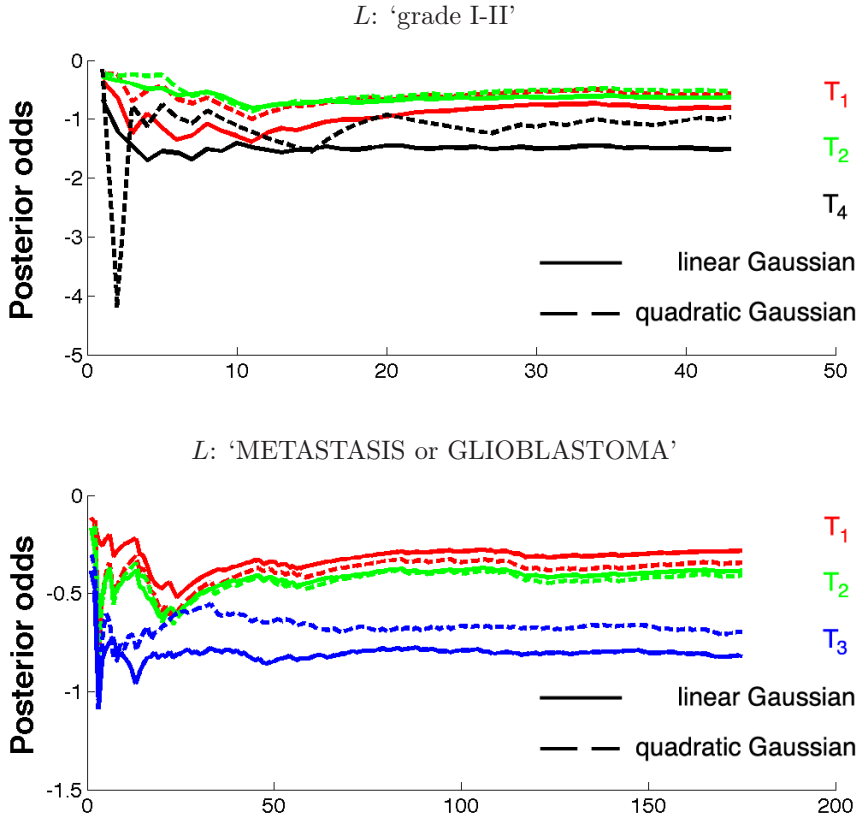


Figure 5.4: Evolution of the posterior odds (in logarithmic scale) of the predictive models as the samples of  $\mathbf{Z}_{DSS}$  are introduced into the DSS in the course of time.

## 5.5 Discussion

In this section we discuss considerations about the audit method's design as well as how the proposed method can improve the decision support experience.

### 5.5.1 Design concerns

The proposed audit method relies on the Bayesian approach, which allows it to objectively select the more adequated model among two or more. When performing model comparison inference under a Bayesian paradigm, the models need to be compared in the light of the same data. Furthermore, the audit method is designed to work on-line, that is, the performance of each eligible model needs to be calculated for each case introduced in the DSS for decision support.

The audit method has been designed to work each time the clinician introduces a case in the DSS. Thus, it is required to have a good time response and optimizations to improve the performance are needed. Precalculation of several operations during low activity periods of the DSS (at nights, sundays, holydays, ...) could positively improve the response time of the audit method. Eligible operations to be precalculated are the calculation of the common  $\mathbf{Z}$  from the  $\mathbf{Z}_i$  of each model, the complexity of each model in terms of its free parameters, or the calculation of the dynamic performance attending to  $\mathbf{Z}_{DSS}$ .

Considering that the scenario of comparison of general and specific models is common in clinical environments, we proposed a mechanism to transform the labels of  $\mathbf{Z}$  into the labels  $T_i$  each model  $\mathcal{M}_i$  can understand. By doing so, we can express  $P(\mathcal{M}_i|\mathbf{Z}) = P(\mathcal{M}_i|\mathbf{Z}^{T_i})$ , and calculation of the posterior of each model is properly performed. In addition, we restrict the posterior calculation to only the set of samples  $\mathbf{Z}$  that any comparing model can deal with. This might be seen as sub-optimal because the models discriminating subclasses are imposing that the samples in  $\mathbf{Z}$  should belong to such subclasses. Nevertheless, we consider this design as a compromise solution: if  $\mathbf{Z}$  was obtained from the union of all the samples in each  $\mathbf{Z}_i$ , the situation where a model would try to predict a sample belonging to a class it cannot discriminate upon would arise and misclassifications would occur. Consequently, this would lead to an undesired side-effect of punishing those classifiers discriminating very specific labels, since its evidence would dramatically drop.

Taking into account the previous concerns, the audit method can perform a consistent comparison between models in a non-painfully time allowing a ‘fair’ comparison of classifiers discriminating super-classes, which typically will give right answers to a general discrimination problem, and classifiers devoted to sub-classes that might address more challenging questions but be trained with lower number of samples.

The design of the audit method is compatible with the multiple inheritance in an ontology. In the *CT* described in Table 5.1, each row with more than one ‘is a’ relationship can be seen as a term that inherits from several super-terms. GLIOBLASTOMA, for example, inherits from the group of tumours that can be considered as aggressive but also from those with high grading stage (grade III-IV). With the mechanism defined to obtain the candidates for comparison attending to the working diagnosis, the audit method can effectively select models where multiple inheritance in some of their labels might occur: in case that the user defined his/her working diagnosis as ‘Glioblastoma’, classifiers discriminating this specific tumour type would be selected but, in addition, also those discriminating the label of ‘grade III-IV’ or ‘aggressive’, if they are available in the system.

One possible limitation of the audit method, though, is that it calculates the evidence by using the BIC. The BIC is a rough approximation to the evidence and limits the use of this audit method to the classification techniques where the calculation of the parameters is possible. Implementation of alternatives to the calculation of the evidence using regularization methods should be studied in order to allow the auditing of other classification techniques. Special attention should be taken on the performance of such regularization methods since they usually are computationally expensive and we want them to work in an on-line scenario where the models will be dynamically audited for each case introduced in the DSS for diagnostic support.

### **5.5.2 Decision support for diagnostic confirmation or further research**

The main purpose of the audit method is to positively influence the user’s diagnostic reasoning. Starting from an initial diagnosis, all the models able to discriminate one or more of the label diagnosis supplied by the clinician are selected. Once the models are evaluated, the DSS interface can be designed

in order to (a) just show the answer given by the model with the best audit performance, which will be the best suited for the case under evaluation, or (b) to show all the classifiers' answers.

By showing only the classifier with best performance able to deal with the proposed diagnosis, we omit information that can distract the user and focus his/her attention on the specific question he/she wants to be answered. Thus, the adoption of (a) is the most recommended for diagnostic confirmation, where the clinician will not invest more than 30-60 seconds to obtain reinforcement to his/her decision [18, 19].

Alternatively, the option of showing all the answers, (b), is best suited for clinical research purposes or diagnosis support for a non-trivial case that requires further time investment. The fact of showing to the user the classifiers able to answer super-classes of the proposed diagnosis can give him/her arguments to decide whether the previous reasoning decision steps were right. In addition, allowing the user to observe the diagnosis support given by classifiers dealing with sub-classes, can help on refining the user's reasoning process.

By asking the user at the moment of the data insertion whether he/she considers the case under evaluation as rutinary, a DSS running the proposed audit method would choose to show just the answer of the best audited model. Otherwise, all the answers, sorted by granularity of the question and by the performance obtained, will be shown.

### **5.5.3 Complementary use to incremental learning algorithms**

In a clinical or research setting, the gathering, pre-process, and validation of samples is expensive and time-consuming. Usually, the PR-based DSSs rely on classification models obtained from a unique training set and the learning stops once this set has been processed [3]. If the DSS has the ability to store the data introduced by the users, this new data can be used to retrain the models. An incremental learning approach allow us to build an initial classifier with a smaller number of samples and update it incrementally when new data are collected.

Our proposed audit method can help on the use of incremental learning algorithms in two different ways: On one hand, by deciding the moment when a retraining should be performed. A stationary behaviour in the evolution of

the posterior odds (Figure 5.4) usually indicates that the predictive capabilities of the classifiers are stable and no further classification improvement will be reached unless new training samples are introduced. This phenomenon could be used as an indicator to retrain or evolve the predictive models by using the data  $\mathbf{Z}_{DSS}$  already introduced into the DSS.

On the other hand, it can also help on detecting when a learning technique or discrimination problem has reached its ‘learning roof’, that is, when the performance does not improve although new cases are used for retraining. Keeping both the last retrained version of a classifier and the one previous to the last retraining would allow to detect if no performance improvement is achieved in the course of time.

#### 5.5.4 Detector of misbehaving models and data shift

Measuring both static and dynamic odds is useful to detect problems in the clinical setting: If a bias between the static and the dynamic odds is detected, it might be due to possible overtraining in the predictive models. The performance evaluated in laboratory might be optimistic for a number of reasons. Here we briefly describe some of them:

- **Low number of training samples:** The number of samples available to train a classifier is a key factor in the success (accuracy) of the results. If a classifier obtains an elevated training performance from few data, an overtraining might be happening: the classifier has learned the characteristics of each sample and is able to perfectly discriminate them but, due to this excessive training, the classifier has a poor generalization power and limited abilities in discriminating other cases different to the ones used for training. Classifiers trained with a low amount of data might probably obtain a good performance when tested in lab, but chances are high that this discrimination ability might be spurious. That is, it is highly probable that there exists a combination of variables able to perfectly separate the low amount of data. This happens due to a phenomenon called curse of dimensionality [25]: when the amount of cases available is smaller than the number of variables used from each case, the available data becomes sparse and this sparsity makes difficult that any learning method can achieve a good result with statistical significance. In the specific domain of brain tumour diagnosis from MRS signals, if the number of cases available is not

significantly greater than the number of variables extracted from the MR variables to train the classifier, the curse of dimensionality might occur.

- **Use of a poor evaluation strategy:** Since having a great amount of data to train classifiers is not always possible, a good design of the training and evaluation stage can help to overcome or, at least, reduce the aforementioned problems. Usually, when dealing with small-sample datasets, resampling techniques can be applied when estimating the classifier performance [26]. These techniques try to optimize the use of the available samples in order to give a non-optimistic performance measure and avoiding the overtraining. In addition, the set of training samples might contain unbalanced classes. If so, a proper evaluation metric, like BAR, that avoids overoptimistic measures, should to be taken into account for the design of a evaluation strategy.
- **Lack of an independent test set to evaluate the classifier:** The use of a test set (independent collection of data not previously seen by the classifier) to prove the generalization capability of the model is a requisite [27]. Although a good evaluation technique can take profit of the available training data and avoid as much as possible any optimistic bias in the evaluation, the use of an independent test is the best practise to establish an objective evaluation of a classifier.

If a classifier suffering from some of these conditions is deployed in a real environment for decision support, its performance will low dramatically. That is, the dynamic odds will be biased with respect to the static odds.

Furthermore, a relevant worsening of the dynamic odds with respect to the static odds might also indicate data shift in the clinical centre when acquiring the biomedical data: classifiers are very sensitive to the input they use to perform a classification. If the clinical center does not follow the MRS acquisition protocol, chances are high that the classifiers will be unable to give a proper answer. A similar result is to be expected if the test spectra would not be preprocessed following the same steps used with the training samples. In addition, a great bias may also be an indicator of different patient populations if the training cases were gathered from clinical centre(s) different to the clinical site where the DSS is deployed.

## 5.6 Concluding remarks

The proposed audit method, suited for DSSs in clinical or research environment, is able to compare predictive models from the initial diagnosis made by the clinician. Comparison is performed using a Bayesian framework and focusing on auditing only those models that are relevant to help on the diagnosis of the current patient.

If an ontology or hierarchy of the labels related to the biomedical domain is available, our audit model compares not only the predictive models able to answer the diagnosis proposed by the clinician, but also those models that can deal with more general labels and those dealing with sub-classes of the given diagnosis. Auditing of the models dealing with super-classes can reinforce the decisions already taken by the clinician. Meanwhile, auditing of models focused on sub-classes can guide to the user on which further step should be taken to address a definitive diagnosis to the current case.

This audit method evaluates both the static and dynamic odds. Static odds tell us how well the predictive models deal with data they were trained with, similarly to an evaluation performed on laboratory, where the models were designed. Dynamic odds evaluation allows to study the effectiveness of the models in a real clinical environment. Measuring both static and dynamic odds is useful to detect problems in the clinical setting: If a bias between the static and the dynamic odds is detected, it might be due to possible overtraining in the predictive models, data shift in the clinical centre when acquiring the biomedical data or even an indicator of different patient populations.

Experiments on real datasets of brain tumour diagnosis with MRS have been performed. We emphasize that this audit method is an effective tool to transmit trustworthiness to the final user and potential benefit of deploying a clinical DSS in real clinical setting.

## Bibliography

- [1] J. Vicente, J. M. García-Gómez, S. Tortajada, A. T. Navarro, F. A. Howe, A. C. Peet, M. Julià-Sapé, B. Celda, P. Wesseling, M. Lluch-Ariet, and M. Robles, *Ranking of Brain Tumour Classifiers using a Bayesian Approach*, vol. 5517 of *Lecture Notes in Computer Science*, pp. 1005–1012. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.



- [2] J. Vicente, S. Tortajada, E. Fuster-Garcia, J. M. García-Gómez, and M. Robles, "An audit method suited for decision support systems for clinical environment," in *Proceedings of 2012 IEEE 6th International Conference on Intelligent Systems (IS'12)*, Sofia, Bulgaria, pp. 281–288, Sept. 2012.
- [3] S. Tortajada, E. Fuster-Garcia, J. Vicente, P. Wesseling, F. a. Howe, M. Julià-Sapé, A.-P. Candiota, D. Monleón, A. Moreno-Torres, J. Pujol, J. R. Griffiths, A. Wright, A. C. Peet, M. C. Martínez-Bisbal, B. Celda, C. Arús, M. Robles, and J. M. García-Gómez, "Incremental Gaussian Discriminant Analysis based on Graybill and Deal weighted combination of estimators for brain tumour diagnosis," *Journal of biomedical informatics*, vol. 44, no. 4, pp. 677–87, 2011.
- [4] H. R. Warner, A. F. Toronto, L. G. Veasey, and R. Stephenson, "A mathematical approach to medical diagnosis: application to congenital heart disease," *Journal of the American Medical Association*, vol. 177, pp. 75–81, 1961.
- [5] R. A. Miller and F. E. Masarie, "The demise of the "Greek Oracle" model for medical diagnostic systems," *Methods of Information in Medicine*, vol. 29, pp. 1–2, 1990.
- [6] P. Ramnarayan, R. R. Kapoor, M. Coren, V. Nanduri, A. L. Tomlinson, P. M. Taylor, J. C. Wyatt, and J. F. Britto, "Measuring the impact of diagnostic decision support on the quality of clinical decision making: development of a reliable and valid composite score," *Journal of the American Medical Informatics Association*, vol. 10, pp. 563–572, 2003.
- [7] S. Dreiseitl and M. Binder, "Do physicians value decision support? A look at the effect of decision support systems on physician opinion," *Artificial Intelligence in Medicine*, vol. 33, pp. 25–30, 2005.
- [8] D. J. C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.
- [9] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *Journal of Statistical Planning and Inference*, vol. 90, pp. 227–244, Oct. 2000.
- [10] J. G. Moreno-Torres, T. Raeder, R. Alaíz-Rodríguez, N. V. Chawla, and F. Herrera, "A unifying view on dataset shift in classification," *Pattern Recognition*, vol. 45, no. 1, pp. 521–530, 2012.

- 
- [11] N. W. Street and Y. Kim, “A streaming ensemble algorithm (SEA) for large-scale classification,” in *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 377–382, ACM, 2001.
- [12] M. A. Maloof and R. S. Michalski, “Incremental learning with partial instance memory,” *Artificial Intelligence*, vol. 154, no. 1-2, pp. 95–126, 2004.
- [13] D. J. C. Mackay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [14] R. E. Kass and A. E. Raftery, “Bayes factors,” *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 773–795, 1995.
- [15] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. Springer, August 2001.
- [16] M. Julià-Sapé, D. Acosta, C. Majós, A. Moreno-Torres, P. Wesseling, J. J. Acebes, J. R. Griffiths, and C. Arús, “Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database,” *Journal of Neurosurgery*, vol. 105, pp. 6–14, Jul 2006.
- [17] T. Gruber, “Ontology (computer science),” in *Encyclopedia of Database Systems* (L. Liu and T. M. Özsu, eds.), Springer-Verlag, 2008.
- [18] A. R. Tate, J. Underwood, D. M. Acosta, M. Julià-Sapé, C. Majós, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, M. E. Cabañas, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arús, “Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra,” *NMR in Biomedicine*, vol. 19, no. 4, pp. 411–434, 2006.
- [19] eTUMOUR Consortium, “eTumour: Web accessible MR Decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data.” Web site. FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC, <http://www.etumour.net> (Accessed: 22nd April 2012; at writing time it was temporarily unavailable).

- [20] J. M. García-Gómez, J. Luts, M. Julià-Sapé, P. Krooshof, S. Tortajada, J. Vicente, W. Melssen, E. Fuster-Garcia, I. Olier, G. Postma, D. Monleón, A. Moreno-Torres, J. Pujol, A. P. Candiota, M. C. Martínez-Bisbal, J. Suykens, L. Buydens, B. Celda, S. Van Huffel, C. Arús, and M. Robles, “Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy,” *Magnetic Resonance Materials Physics, Biology and Medicine*, vol. 22, no. 1, pp. 5–18, 2009.
- [21] J. M. García-Gómez, S. Tortajada, C. Vidal, M. Julià-Sapé, J. Luts, À. Moreno-Torres, S. Van Huffel, C. Arús, and M. Robles, “The effect of combining two echo times in automatic brain tumor classification by MRS,” *NMR in Biomedicine*, vol. 21, no. 10, pp. 1112–1125, 2008.
- [22] J. Luts, J. Pouillet, J. M. García-Gómez, A. Heerschap, M. Robles, J. A. K. Suykens, and S. Van Huffel, “Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra,” *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 288–298, 2008.
- [23] D. N. Louis, H. Ohgaki, O. D. Wiestler, and W. K. Cavenee, *WHO classification of tumours of the central nervous system*. Lyon, France: IARC Press, 2007.
- [24] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.
- [25] R. Bellman, *Dynamic Programming*. Courier Dover Publication, 2003.
- [26] J. K. Martin and D. S. Hirschberg, “Small sample statistics for classification error rates I: Error rate measurements,” Tech. Rep. ICS-TR-96-22, 1996.
- [27] D. G. Altman and P. Royston, “What do we mean by validating a prognostic model?,” *Statistics in Medicine*, vol. 19, no. 4, pp. 453–473, 2000.

## Chapter 6

# Similarity model for proper prior assessment

*“I define nothing. Not beauty, not patriotism. I take each thing as it is, without prior rules about what it should be.”*

*Bob Dylan*

A DSS based on Machine Learning can offer support to experts by means of a set of relevant predictive models. In a dynamic scenario, a DSS should also be dynamic by being able to cope with changes in the environment and conditions. Biomedical domains are a clear example of a dynamic scenario since dynamic conditions of the environments may change the assumed conditions in the models. The definition of a mechanism that calculates the suitability of the classifiers attending to each particular test case can improve the performance of the predictions in such dynamic systems. This chapter contributes with a methodology for prior probability assessment to a set of classifiers based on a similarity model. This similarity model, inspired in a Bayesian approach, is a natural extension of the audit method presented in the previous chapter. Such mechanism will allow the DSS to select on-line the most adequate classifier for each test case attending to contextual information, which is information not used in the design stage of the classifiers, but related to the case or its environment.

The evaluation of this prior assessment based on a similarity model has been carried out for brain tumour grade diagnosis from MR spectra data. For this scenario, the age, gender and location of the tumour in the brain have been defined as contextual information. Specialized classifiers aimed to adult and children patients were designed with MRS data. Non-specialized classifiers designed with the spectra and the contextual information were also developed. A similarity model was defined for each specialized classifier attending to the contextual information. The overall performance of the DSS when using the prior assessment calculated from the similarity model was comparable to the performance of the non-specialized classifiers. In addition, the discrimination of set of particularly difficult cases like those of children or young adults improved when using the similarity model for prior probability assessment with the specialized models. The evaluation demonstrates the viability of integration of this mechanism into a real DSS for dynamic environments.

*This chapter complements the audit method presented in Chapter 5 by focusing in assessing prior probabilities to each candidate model. The core of this chapter has been published in the proceedings of [1] and in [2]. The adult and children classifiers developed in this work are based on the question addressed in Chapter 3. Section 6.4.4 describes how the similarity model can be combined with the information related to compatibility of classifiers trained with 1.5T MR spectra with 3T datasets presented in [3].*

## 6.1 Introduction

The conception of a DSS aimed to clinical diagnosis and prognosis as an ‘evidence adaptive’ system is becoming a new trend design [4]. Such systems incorporate automated and continuous updating to reflect the most recent advances in clinical science and local practice knowledge. These systems, likewise, provide solutions dealing with the incorporation of information from diverse sources and adaptability to varied practice settings.

In that way, a dynamic DSS based on PR classifiers should include mechanisms to guide on which classifier is adequated to give support on the decision related to a specific test case.

This work contributes with a methodology for calculating the prior probabilities of a set of classifiers according to a specific test case. The definition of this mechanism can contribute to improve the performance of the predictions in the

dynamic systems. The approach presented in this chapter is a natural extension of the development described in Chapter 5. Hence, the Bayesian paradigm is our basis to obtain a proper prior assessment of two or more comparing models.

In the audit method presented in Chapter 5, efforts were focused in evaluating the predictive abilities of classification models by comparing its posterior odds. The method, though, treated the prior probabilities associated to each model in a simple fashion: those models unable to deal with the working diagnosis proposed by the clinician were discarded as candidates. In other words, their prior probabilities were set to 0. Then, the audit method assigned an arbitrary prior, based on a uniform distribution probability, to the  $M$  models able to answer to the working diagnosis. This treatment of the prior probabilities, though, might be oversimplistic, specially in the environment of a set of classifiers giving diagnosis support in a clinical DSS. Chapter 3 depicted such a plausible scenario: two classifiers performing the same discrimination (aggressive brain tumours VS non-aggressive brain tumours) but aimed at different patient profiles: one devoted to adult and other to pediatric cases. A filter mechanism based on the patient's age was proposed as a solution and allowed an adequated selection for classification.

The purpose of this chapter is to unify the solution proposed in Chapter 3 into the audit method of Chapter 5. More specifically, we propose the calculation of the prior probabilities of the candidate models attending to a similarity model based on extra information of the data from which each classification model has been designed. This extra information about the samples, that we will call *contextual information*, is not necessarily used for design purposes but is related to the environment of the problem.

In that way, given a test sample, the prior probability assigned to each classification model will be calculated according to the similarity between the contextual information of the test sample and the contextual information of the design dataset of each model.

The sections below are structured as follows: Section 6.2 depicts how our proposal is linked with the audit method described in Chapter 5, our proposed mechanism to calculate the prior probabilities using the similarity model, and the procedure followed to its evaluation. Section 6.3 shows comparative results of the mechanism in a dynamic scenario of prediction of the aggressiveness of BT. Section 6.4 describes some theoretical issues related to the calculation of the prior probabilities with the similarity model, recites medical significance to the diagnosis of the aggressiveness of BTs and concludes that the proposed

mechanism can contribute to improve the overall performance in dynamic environments.

## 6.2 Methods

### 6.2.1 Prior probabilities in the audit method

A sample  $z$  is defined as a pair  $(\mathbf{x}, t)$ , where  $\mathbf{x}$  is a data vector describing the sample and  $t$  is its associated label. The objective of the audit method is to compare among the classification models available in a DSS in order to assist about which of them gives a better classification support. Let us call  $z^\kappa$  to the test sample a clinician wants to introduce in the DSS. Notice that at the moment the user introduces  $z^\kappa$ , the label  $t^\kappa$  of the test case remains unknown both to the clinician (who is asking to the DSS for confirmation or support) and to the classifiers in the DSS. Hence, the audit method takes as starting point the initial diagnosis,  $L$ , proposed by the user and selects the  $M$  models able to answer any element  $l_j$  of  $L$ . That is, calculates the prior probability of each model in the DSS according to this rule:

$$P(\mathcal{M}_i) = \begin{cases} 0 & \text{if } msl(T_i) \cap msl(L) = \emptyset \\ 1/M & \text{otherwise,} \end{cases} \quad (6.1)$$

where  $T_i$  is the set of labels that  $\mathcal{M}_i$  is able to discriminate,  $msl(t)$  is a function that makes reference to the ‘most specific label’ as defined in Section 5.2 of Chapter 5, and  $M$  is the total amount of models with  $msl(T_i) \cap msl(L) \neq \emptyset$ .

In other words, the prior probabilities are assessed previously to the comparison of the  $M$  candidates according to the prior beliefs of the clinician. This prior assessment allows to discard those models unable to answer the working diagnosis and assigns an arbitrary uniform probability to each selected candidate. Then, the comparison of the  $M$  selected candidates is performed as described in Section 5.2 of Chapter 5:

$$\frac{P(\mathcal{M}_m | \mathbf{Z}, \mathbf{Z}_{DSS})}{P(\mathcal{M}_l | \mathbf{Z}, \mathbf{Z}_{DSS})} = \underbrace{\frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)}}_{\text{Prior odds}} \underbrace{\frac{P(\mathbf{Z}^{T_m} | \mathcal{M}_m)}{P(\mathbf{Z}^{T_l} | \mathcal{M}_l)}}_{\text{Static odds}} \underbrace{\frac{P(\mathbf{Z}_{DSS}^{T_m} | \mathbf{Z}_m, \mathcal{M}_m)}{P(\mathbf{Z}_{DSS}^{T_l} | \mathbf{Z}_l, \mathcal{M}_l)}}_{\text{Dynamic odds}} \quad (6.2)$$

where each classification model  $\mathcal{M}_i$  is estimated with a set of samples  $\mathbf{Z}_i$ , where  $\mathbf{Z}_{DSS}$  represents the samples already introduced in the DSS (and with a known label associated) and where  $\mathbf{Z}$  was obtained as a combination of all the  $\mathbf{Z}_i$  from the comparing models.

## 6.2.2 Prior assessment based on the similarity model

The audit model, so far, assumes that we have no reason to assign strongly differing priors  $P(\mathcal{M}_i)$  to the  $M$  comparing models. Thus, the prior odds of Equation 6.2 can be ignored since both comparing models have the same prior probability  $1/M$ .

However, given a dynamic environment where change in the samples may occur, it could be useful to define a mechanism that gives a tailored recommendation on which of the  $M$  models are more suited to correctly predict the label of  $z^\kappa$ . Our proposal is to assess the prior probabilities of the models of the DSS attending not only to the working diagnosis,  $L$ , but also to a measure of the similarity of  $z^\kappa$  to  $\mathbf{Z}_i$  with respect to extra information not used during the design stage of  $\mathcal{M}_i$ .

Let us redefine  $\mathbf{Z} = \{(\mathbf{k}_j, \mathbf{x}_j, t_j)\}_1^N$  as a set of  $N$  samples where  $\mathbf{x}_j$  is the data vector describing the  $j^{th}$  sample,  $t_j$  is its associated label and  $\mathbf{k}_j$  is a set of values describing the contextual information. Contextual information is defined as extra information about the samples not used in the input space of the classification models.

So far, for each  $\mathbf{Z}_i$ , we had a  $\mathcal{M}_i$ , a learning model with parameters  $\theta$  that tried to assign to each set of variables  $\mathbf{x}_j$  of  $\mathbf{Z}_i$  its corresponding label  $t_j$ . Let us now define  $\mathcal{K}_i$ , which we will call similarity model of  $\mathbf{Z}_i$ , as a model with a set of parameters  $\omega$  modeling the  $\mathbf{k}_j$  values, the contextual information of  $\mathbf{Z}_i$ . Then, the prior probability of  $\mathcal{M}_i$  can be expressed as:

$$P(\mathcal{M}_i) = \begin{cases} 0 & \text{if } msl(T_i) \cap msl(L) = \emptyset \\ P(\mathcal{K}_i|z^\kappa) & \text{otherwise,} \end{cases} \quad (6.3)$$



where the calculation of the posterior  $P(\mathcal{K}_i|z^\kappa)$  is:

$$P(\mathcal{K}_i|z^\kappa) = P(\mathcal{K}_i|\mathbf{k}^\kappa) = \frac{P(\mathcal{K}_i)P(\mathbf{k}^\kappa|\mathcal{K}_i)}{P(\mathbf{k}^\kappa)} = \frac{P(\mathcal{K}_i)}{P(\mathbf{k}^\kappa)} \int P(\mathbf{k}^\kappa|\omega, \mathcal{K}_i)P(\omega|\mathcal{K}_i)d\omega. \quad (6.4)$$

The  $\mathbf{k}^\kappa$  is the contextual information associated to the current sample  $z^\kappa$ .  $P(\mathcal{K}_i)$  is the prior probability of the contextual model, a 'subjective' prior distribution function over the model space.  $P(\mathcal{K}_i|z^\kappa)$  is the posterior of the similarity model for  $\mathbf{Z}_i$  where, by definition of Equation 6.3,  $\sum_i P(\mathcal{K}_i|z^\kappa) = 1$ . The posterior of the similarity model is a way of objectively determine the similarity of a test sample to the dataset  $\mathbf{Z}_i$  used for  $\mathcal{M}_i$  in the design stage.

The calculation of the integral of Equation 6.4 can be locally approximated as a Gaussian by maximization of the posterior probability:

$$\log P(\mathbf{k}^\kappa|\mathcal{K}_i) \simeq \log P(\mathbf{k}^\kappa|\hat{\omega}, \mathcal{K}_i) \quad (6.5)$$

where  $\hat{\omega}$  is a maximum likelihood estimate.

Now that we have a procedure to calculate the posterior of the similarity model, we can use it to estimate the prior odds of two comparing models:

$$\frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} = \frac{\frac{P(\mathcal{K}_m)}{P(\mathbf{k}^\kappa)}P(\mathbf{k}^\kappa|\mathcal{K}_m)}{\frac{P(\mathcal{K}_l)}{P(\mathbf{k}^\kappa)}P(\mathbf{k}^\kappa|\mathcal{K}_l)} = \frac{P(\mathcal{K}_m)P(\mathbf{k}^\kappa|\mathcal{K}_m)}{P(\mathcal{K}_l)P(\mathbf{k}^\kappa|\mathcal{K}_l)} \quad (6.6)$$

If we assumed that we have no reasons to assign strongly differing priors to each similarity model, Equation 6.6 can be simplified to:

$$\frac{P(\mathcal{M}_m)}{P(\mathcal{M}_l)} = \frac{P(\mathbf{k}^\kappa|\mathcal{K}_m)}{P(\mathbf{k}^\kappa|\mathcal{K}_l)} \quad (6.7)$$

And hence, Equation 6.2 can be rewritten:

$$\frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} = \frac{P(\mathbf{k}^\kappa|\mathcal{K}_m)}{P(\mathbf{k}^\kappa|\mathcal{K}_l)} \frac{P(\mathbf{Z}^{T_m}|\mathcal{M}_m)}{P(\mathbf{Z}^{T_l}|\mathcal{M}_l)} \frac{P(\mathbf{Z}_{DSS}^{T_m}|\mathbf{Z}_m, \mathcal{M}_m)}{P(\mathbf{Z}_{DSS}^{T_l}|\mathbf{Z}_l, \mathcal{M}_l)} \quad (6.8)$$

Finally, to compare  $M$  models, equation 6.8 can be generalized:

$$\frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{\sum_{l=1}^M P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} = \frac{P(\mathbf{k}^\kappa|\mathcal{K}_m)P(\mathbf{Z}^{T_m}|\mathcal{M}_m)P(\mathbf{Z}_{DSS}^{T_m}|\mathbf{Z}_m, \mathcal{M}_m)}{\sum_{l=1}^M P(\mathbf{k}^\kappa|\mathcal{K}_l)P(\mathbf{Z}^{T_l}|\mathcal{M}_l)P(\mathbf{Z}_{DSS}^{T_l}|\mathbf{Z}_l, \mathcal{M}_l)} \quad (6.9)$$

### 6.2.3 Evaluation of the prior assessment with the similarity model

The goal of the prior assessment based on the similarity model is to select the adequate classification model for each test case evaluated in the DSS in order to improve the overall performance of the system. To measure the performance of the similarity model, we propose to evaluate how it affects the overall performance of a DSS running classifiers devoted to diagnosis aggressiveness of BTs with MRS. Age, gender and location of the tumour in the brain are defined as contextual information for this problem. We have divided the dataset into different ‘specialized’ sets, attending to some criteria. These specialized sets are divided into a set for classification model design and a set for independent test.

The classification models are obtained with the LDA technique. LDA is a Gaussian Parametric model designed to find boundaries between classes [5] based on the maximum likelihood estimation of the parameters of the Gaussian distributions for the classes considered under the study.

In order to assess the prior probabilities to each classifier (or corpora  $\mathbf{Z}_i$ ), we define a similarity model  $\mathcal{K}_i$  based on the estimation of the probability density mixture of the contextual information. By using this estimation, we can establish a mixture of binary, discrete and continuous variables assuming that they follow a Bernoulli, multinomial and Gaussian distribution given the nature of the contextual information variables. Each contextual model  $\mathcal{K}_i$  calculates, attending to the contextual information of a test sample, the probability of belonging to its design dataset,  $\mathbf{Z}_i$ . In other words, how similar is the test sample to the samples used in the design of each model. The highest probability determines the winner candidate.

Taking the contextual information of each design dataset, a similarity model is set up. A simulation of the scenario of a DSS using the similarity to apply the

prior probability assessment is then performed. That is, for each test sample, the similarity model as defined in Equation 6.3, decides which of the classifier candidates is more adequate (attending to the contextual information). The prediction for this sample is carried out by the classifier selected by the similarity model.

A set of ‘non-specialized’ LDA classifiers is also designed, for comparison purposes, using together the MRS data and the contextual information. Then, combining every independent test set from each specialized dataset, we will compare the performance of the non-specialized classifiers with the specialized ones using our proposed similarity model.

## 6.2.4 Dataset for evaluation

To evaluate the prior assessment based on the similarity model, an MRS database for BT aggressiveness is used.

The dataset consisted on 1042 MRS at 1.5T histopathologically diagnosed brain tumour cases acquired by ten institutions in the framework of the eTUMOUR [6] project. The classes considered in this study were based on the histological classification of the CNS tumours as described by the WHO Classification [7]. Signal processing was performed following the protocols defined in [8, 9]. The cases were divided into adults (889 cases; mean age  $52 \pm 15$ ) and children, who are patients with age under 16 years old (153 cases; mean age  $7 \pm 5$ ). Division of the cases into these groups responds to the differences reported in the nature of brain tumour in children and adults [10]. Brain tumours were categorized into aggressive and non-aggressive attending to this distribution: In adults, 357 samples of non-aggressive tumours and 532 samples of aggressive tumours. In children, 89 samples of non-aggressive and 64 labeled as aggressive tumour. Table 6.1 shows the cases distributed in design and independent test attending to the patient age and the tumour classes. Figure 6.1 shows the mean MRS data of each tumour type for children and adults.

A total of 68 cases of young adults ( $17 \leq \text{age} \leq 25$  years old), with 40 cases of non-aggressive tumours and 28 aggressive tumours, was available. A set of 28 cases (17 non-aggressive; 11 aggressive) was defined as independent test set of young adults for evaluation purposes. These 28 cases were those not used in the design stage of the classifiers.

A PI around mean frequency positions corresponding to particular metabolites

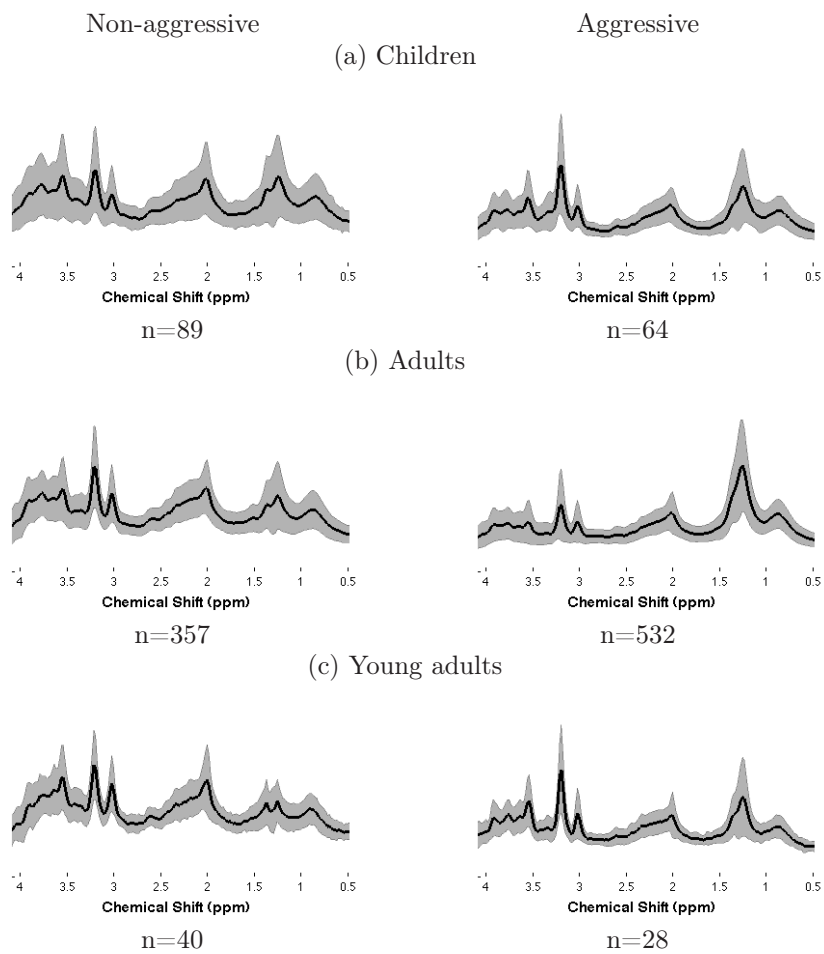


Figure 6.1: Mean spectra of pediatric (top), adult (center) and young adult (bottom, age between 17 and 25 years old) non-aggressive and aggressive tumours with standard deviation indicated by the shaded region. Number of patients for each tumour type is indicated beneath each graph.

Table 6.1: Number of samples for the brain tumour grade diagnosis problem divided into design and test.

Classes	Design stage ( $\mathbf{Z}_i$ )			Independent test ( $\mathbf{Z}_{DSS}$ )		
	Children	Adults	Both	Children	Adult	Both
Non-aggressive	58	246	304	31	111	142
Aggressive	44	347	391	20	185	205
TOTAL	102	593	695	51	296	347

in the region of [0.5, 4.1] ppm interval is carried out. PI is a method successfully applied in [11] that allows a reduction from the whole spectra to 15 parameters. PI has proportionality to the concentration of the main metabolites in each spectra.

Three different variables are used as contextual information: the age of the patient (continuous), gender (binary) and the tumour location in brain (categorical). Distributions of age, gender and brain tumour location are given in Figures 6.2, 6.3 and 6.4, respectively.

## 6.3 Results

Table 6.2 summarizes the performances obtained by two specialized classifiers trained with PI extracted from MRS data. One is tuned exclusively from adult cases; the other, with pediatric cases. The classifiers showed an optimistic bias when comparing the likelihood from  $\mathbf{Z}_i$  (design stage) with the likelihood measured from all the samples of the independent test set ( $\mathbf{Z}_{DSS}$ ). Nevertheless, this bias is not as relevant if we take a closer examination: these specialized classifiers obtain a likelihood similar to the one in design stage when are evaluated with an independent dataset with age accordance. In contrast, when the children classifier is evaluated with an adult test set, and vice-versa, the performance dramatically decreases. Having an age accordance seems to be an important factor for showing a proper performance and Figure 6.1 can help to explain why this happens: Although the mean spectra of aggressive BT cases of children and adult are visually different, the non-aggressive mean spectra of pediatric and adult cases looks similar, the main difference being that the pediatric mean spectra presents greater variability. Thus, a classifier devoted to

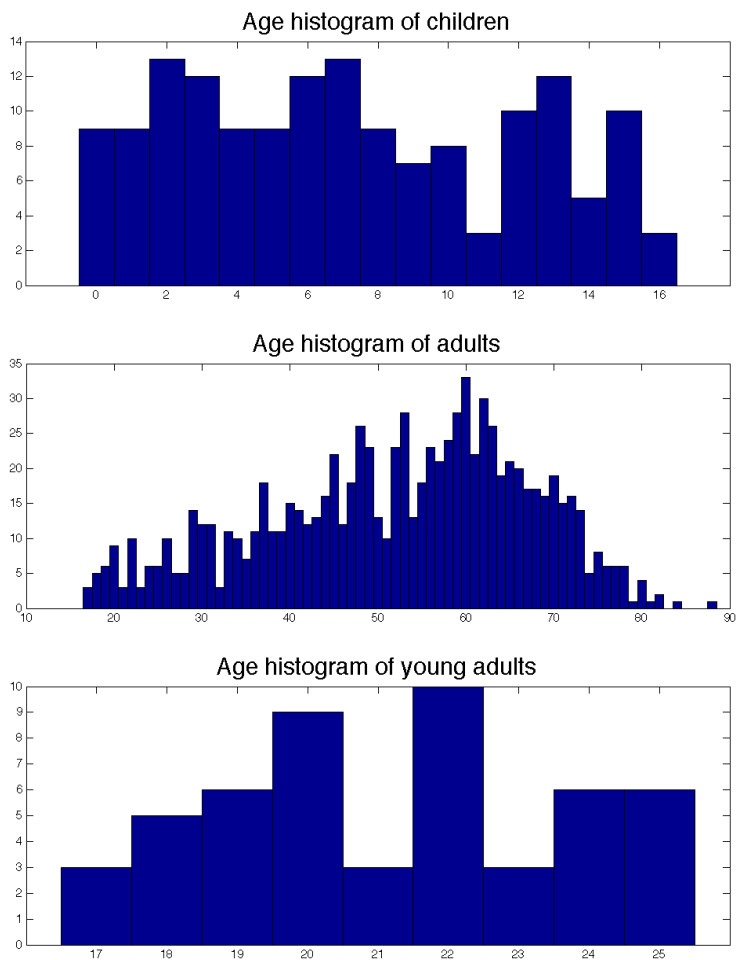


Figure 6.2: Histograms of the age in children, adults and young adults (age between 17 and 25 years old) with brain tumour.



Figure 6.3: Histogram of gender in children, adults and young adults (age between 17 and 25 years old) with brain tumour.

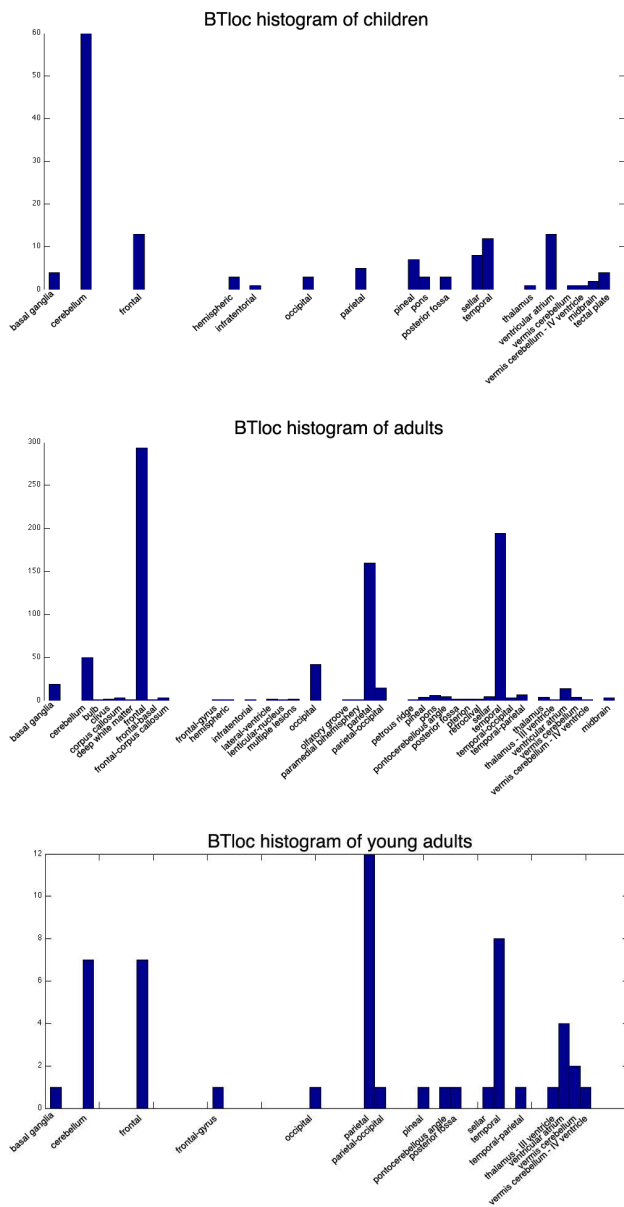


Figure 6.4: Histogram of BTloc in children, adults and young adults (age between 17 and 25 years old) with brain tumour.



Table 6.2: Likelihood achieved by BT aggressiveness classification from PI extracted from the MRS data specialized in ages: Children (age under 16 years old) and adults (age over 16 years old).

Classifier	Specialized classifiers (by age)				
	Number of training samples	Design evaluation ( $Z_i$ )	Independent test ( $Z_{DSS}$ )		
			Children (51 cases)	Adults (296 cases)	All (347 cases)
Adults	593	0.76	0.53	0.78	0.75
Children	102	0.71	0.77	0.53	0.53

discriminate children cases needs to be able to deal with such variability but with less training samples to learn, compared to a classifier devoted to adults.

To measure the overall performance of a DSS, we use the similarity model in order to choose the most adequate specialized classifier for each sample to be tested. The overall performance of the DSS measured as the likelihood obtained according to the different combinations of contextual information is depicted in Table 6.3. Different combinations of contextual information are tested to model this proper prior assessment. The overall performance when using the contextual information to decide which specialized classifier should be used shows an improvement when the age and the BTloc are considered as contextual information (Table 6.2). Gender information, on the contrary helps to discriminate better the pediatric cases but the overall performances does not improve when compare to the performance of the classifier specialized in adults.

In order to evaluate the use of the contextual information applied to specialized classifiers, we compare the performances of Table 6.3 with those of non-specialized classifiers. Non-specialized classifiers are trained with all the samples (adult and children) plus the contextual information. Table 6.4 summarizes the performance of the non-specialized classification models, which combine the PI applied to MRS data with the contextual information. These classification models can be seen as a reference for the estimation of the performance achievable for this problem. Table 6.4 shows the evaluation of independent test sets according to the age: children (patients under 16 years old) and adults (age > 16 years old). It also shows the result of discriminating all the samples.

Analysis of Table 6.4 reveals similar results for the different combinations of PI

Table 6.3: Overall likelihood of a DSS with specialized classifiers when using a similarity model to determine which classifiers is more adequate to each test sample.

Contextual information (similarity model)	Overall Performance		
	Independent test		
	Children (51 cases)	Adults (296 cases)	All (347 cases)
Age	0.77	0.78	0.78
Gender	0.76	0.64	0.65
BTloc	0.80	0.76	0.77
Age + gender	0.77	0.64	0.65
Age + BTloc	0.80	0.77	0.78
Age + gender + BTloc	0.78	0.76	0.75

and contextual information. The best performance when evaluating with all the test cases is 0.74, almost identical to the likelihood of 0.75 measured with the data of the design stage. The performance achieved when testing child cases (around 0.65) is inferior to the global. In contrast, the likelihood for adult cases is slightly better (0.77) than that of the global performance (measured with adult and child test cases).

Attending to the results of Tables 6.3 and 6.4, the overall likelihood of the DSS obtained when applying the similarity model is comparable to the likelihood of the non-specialized classifiers although slightly better. Nevertheless, these improvement is marginal since no statistical significant differences are found between the performances of Tables 6.3 and 6.4 ( $p < 0.01$ , Tukey's test,  $\alpha = 0.05$ ).

Finally, Table 6.5 shows the result of testing the different classifiers to an independent dataset of 28 young adults. A likelihood of 0.70 is achieved by the classifier specialized in adult cases. The likelihood achieved by the pediatric classifier is slightly inferior, with a likelihood of 0.65. If we use the similarity model to select the adequate specialized classifier according to the contextual information, the overall performance of the DSS can improve up to 0.80 if using the age or the BTloc. In contrast, all the non-specialized classifiers obtained a likelihood of 0.60 or 0.67. The difference in performance achieved with the use of the contextual information is statistically significant ( $p < 0.01$ , Tukey's test,

Table 6.4: Likelihood of the non-specialized models for BT aggressiveness classification (trained with 695 cases). These models are trained with both adults and children cases and combine the PI extracted from the MRS data with the contextual information: age, gender and Brain Tumour location (BTloc).

Non-specialized classifiers				
Classifier designed with:	Likelihood for $\mathbf{Z}$	Likelihood for $\mathbf{Z}_{DSS}$		
		Children (51 cases)	Adults (296 cases)	All (347 cases)
PI	0.74	0.65	0.74	0.72
PI + age	0.74	0.63	0.77	0.74
PI + gender	0.74	0.65	0.74	0.72
PI + BTloc	0.75	0.65	0.72	0.72
PI + age + gender	0.75	0.63	0.77	0.74
PI + age + BTloc	0.75	0.63	0.74	0.74
PI + age + gender + BTloc	0.75	0.63	0.77	0.74

Table 6.5: Likelihood of specialized classifiers (designed with PI), specialized classifiers using the similarity model and non-specialized classifiers (designed with PI and contextual information) for an independent test set of young adult patients.

	Classifier	Young Adults test set (28 cases)
Specialized Classifiers	Adults (593 cases)	0.70
	Children (102 cases)	0.65
Overall performance with similarity model	Age	0.80
	Gender	0.65
	BTloc	0.80
	Age + gender	0.65
	Age + BTloc	0.80
	Age + gender + BTloc	0.70
Non-specialized Classifiers (695 cases)	PI	0.67
	PI + Age	0.60
	PI + gender	0.67
	PI + BTloc	0.67
	PI + age + gender	0.60
	PI + age + BTloc	0.60
	PI + age + gender + BTloc	0.60

$\alpha = 0.05$ ) compared to performance obtained with non-specialized classifiers. It seems that, for this discrimination problem, the use of the variables of age, BTloc or gender might be more convenient if used as contextual information rather than using them as input for the classifiers.

## 6.4 Discussion

A proposal for prior probabilities assessment is designed to select the most adequate classification model in a DSS. This prior assessment is founded on the Bayesian paradigm. It calculates the prior assigned to each classification model attending to a similarity model. The evaluation of this prior probabilities assessment model has been performed by simulating a DSS with a set of specialized classifiers for BT aggressiveness diagnosis that selected the more adequate classifier attending to contextual information.

### 6.4.1 Dealing with the Simpson's paradox

The Simpson's paradox [12, 13] is a phenomenon whereby an event A increases the probability of other event B in a given population and, at the same time, decreases the probability of B in every subpopulation. This phenomenon might occur with the audit method and, particularly, when using the contextual information and the similarity model to assess proper priors.

An example of such effect is formally described: given a population of adults and children with BT that can be categorized as aggressive or non-aggressive, the rate of aggressive cases is greater in adults than in children. Nevertheless, if such population is divided attending to the center from each case was acquired (subpopulations), the probability is reversed:

$$P(\text{Aggressive}|\text{Adult}) > P(\text{Aggressive}|\text{Children}), \quad (6.10)$$

$$P(\text{Aggressive}|\text{Adult, Centre A}) < P(\text{Aggressive}|\text{Children, Centre A}), \quad (6.11)$$

$$P(\text{Aggressive}|\text{Adult, Centre B}) < P(\text{Aggressive}|\text{Children, Centre B}). \quad (6.12)$$

Table 6.6 depicts Simpson’s paradox numerically (numbers are given for pedagogical purposes). Looking at equations (6.11) and (6.12), the prevalence of aggressive tumours in children is higher than that of adults in both centres. Paradoxically, if we look at the whole population (Equation 6.10), aggressive tumours are more prevalent in adults. How is this possible?

Table 6.6: Rate of aggressive tumours in children and adults in two centres A and B and both combined.

All	Aggressive	Non-aggressive	Aggressiveness rate	
Adults	20	20	40	50%
Children	16	24	40	40%
	36	44	80	
Centre A	Aggressive	Non-aggressive	Aggressiveness rate	
Adults	18	12	30	40%
Children	7	3	10	70%
	25	15	40	
Centre B	Aggressive	Non-aggressive	Aggressiveness rate	
Adults	2	8	10	20%
Children	9	21	30	30%
	11	29	40	

According to global statistical studies like that of CBTRUS [10], aggressive tumours are less prominent in children. However, a naive classifier answering attending to the frequencies observed would say otherwise if it was trained only with the data in Centre A, for example.

Lindley and Novick in [14] explained the reason behind this paradox using the concept of exchangeability and, more recently, Pearl solved this contradiction using the paradigm of causality [15]. Although using somewhat different approximations, both works stated that this paradox occurs due to distorted causal interpretations. In our example, if the fact that a case is acquired from Centre A is perceived to be the cause for both the aggressiveness level and the age of the patient, then the aggressiveness needs to be evaluated separately for each center (as in (6.11) (6.12)) and then averaged accordingly. Hence, if the fact of coming from a particular centre is the only confounding factor, (6.11) and (6.12) just represent the aggressiveness rates in the respective populations while (6.10) represents merely the evidential weight in the absence of centre information, and the paradox dissolves.

In other words, the correct interpretation of the data available and the understanding of the acquisition circumstances makes the paradox disappear. Numbers in Table 6.6 make sense if we knew that Centre A was a hospital specialized in treatment of aggressive BT cases both in adults and children and that Centre B was a children hospital where most of the cases were pediatric and the remaining were from young adults.

The use of the similarity model might arise Simpson's paradoxes depending on the variables taken under consideration as contextual information. Nevertheless, a proper interpretation of the subpopulations obtained should effectively dissolve any contradictory explanation. Since the contextual information is extra information not mandatory for classification purposes, a convention can be taken in order to avoid possible Simpson's paradoxes: whenever the contextual information is available in a test set (in our example, the centre from where it comes), the similarity model can be calculated (according to the data information of each centre A or B). Otherwise, equal priors will be assigned to each candidate model and the final decision of choosing the best classification model will be left to the user of the DSS or performed arbitrarily by taking into account the evidences of the candidates.

### **6.4.2 Theoretical issues**

The proposed model to assess the prior distribution of the classification models relies on the contextual information, which is extra information about the samples not used for design purposes but related to the samples or their environment. In that way, contextual knowledge about the design samples can be incorporated and applied for determining the prior of each candidate model.

Several proposals have been made in the literature to model the prior probabilities of comparing models. Some authors claimed that prior probabilities can be inspired from trust models in multi-agent systems, which promote the exchange of information on the reliability of individual agents through feedback systems to provide reputation information. This feedback mechanism induces a substantial improvement in transaction efficiency and benefits the trust and trustworthy behavior for the whole community [16] (for review see [17]). However, prior assessment based on trust models may rely only on subjective perception of the experience of an interaction partner (user), but ignoring the 'objective' information related to the data from where each classifier was build-up.

A mechanism that indicates the similarity of a test sample to the data used to

design a model can compensate the natural bias that a model can present when its performance is evaluated attending to cases to which is not aimed. According to the statistics [10], in a clinical site assisting both adults and children with BT, the incidence of pediatric BT cases represents the 15-20% of the total seen cases. The dynamic performance of a pediatric classifier in a DSS can be biased if such classifier has been selected as candidate for adult cases. With the proposed similarity model, when the sample under evaluation is that of a children, the prior probabilities assigned would reduce this possible bias by suggesting to the clinician an appropriate model according not only to the performance but also to the contextual information.

In order to obtain a prior assessment as objective as possible, each similarity model  $\mathcal{K}_i$  must take into account the same contextual information: if it is established that the contextual information is the patient's age, the  $\mathcal{K}_i$  associated to each  $\mathcal{M}_i$  should consider only that variable. By following this premise, the comparison of a test sample among the different candidate models will be performed attending always to the same set of variables, allowing a 'fair' prior assessment.

### 6.4.3 BT - case study

BT aggressiveness offers a good framework to test this proposal for proper prior assessment. The classification task is performed with MRS data but other contextual information is also available. The importance of this contextual information has been reported in several studies: the age is a relevant factor for determining the grade of tumours because the nature of child brain tumours may be totally different from adults [10]. The data used in this experiments clearly follows a normal distribution in adults (Figure 6.2). In the case of young adults and children patients, the distribution does not seem to be Gaussian and this might be due to the difference in the amount of data available from these populations. Analogously to the age, the differences in tumour nature makes also reference to the location in the brain, which varies with the brain development and aging. These anatomical differences can be seen in the distributions gathered in Figure 6.4. BTloc in young adults reveals that these patients can be considered as a mid-term between adults and children: several tumours are located in frontal, parietal and temporal locations where brain tumours are typically found in adults. Nevertheless, a significant amount of tumours are also found in the cerebellum and the ventricular atrium, which are locations where pediatric tumours might appear. Hence, although anatomical location of brain

tumours is not sufficient to make classifications based on this variable alone, tumour location could be used as an informative prior [18]. Similarly, information about the gender distribution [10] might be an interesting factor. Figure 6.3 reveals that, with independence of the age, brain tumours are more prevalent in men. This is specially observed in children, where the number of males doubles that of females.

The information regarding the use of age, gender and BTloc as contextual data has been consensuated and recommended by oncologists, spectroscopists and radiologists of the consortium of the multicenter project HEALTHAGENTS (2005-2008) [19].

The nature of the BT aggressiveness diagnosis fosters the specialization of classification models attending to an age scope. These specialized classifiers improve the performance of non-specialized classifiers, even when the non-specialized models are designed using also the contextual information. This improvement, though, is only observed when the specialized classifiers are evaluated with test samples similar to those used in the design stage.

The scenario proposed in this evaluation retakes the question addressed in Chapter 3 but making use of more tumour cases. The prevalence of pediatric cases is of 14.7% with respect to the total. This is in broad accordance with the literature [10] and also depicts an scenario similar to the one in Chapter 3, where the incidence of pediatric brain tumours was of 19%.

The results described in Table 6.2 correspond to two specialized classifiers attending to the patient's age. These results are similar to the ones described in Chapter 3. The overall performance of the pediatric classifier (likelihood of 0.53) is closer to that obtained when tested with adult cases. This is explained by the great amount of adult cases in the  $\mathbf{Z}_{DSS}$  (296 adult cases and only 51 pediatric cases). The overall performance of the adult classifier (likelihood of 0.75) was also lowered by the presence of the pediatric cases in  $\mathbf{Z}_{DSS}$  but, analogously, it remained closer to the performance obtained when evaluated only with adult cases. The results of these specialized classifiers can correspond to an scenario of a DSS using an audit method that assigns equal priors to each classifier.

Table 6.3 summarizes the overall performance of the DSS when our similarity model is used to decide which specialized classifier is to be used for each test case according to the contextual information. Compared to the overall performances described in Table 6.2, the DSS is able to globally achieve a slight performance improvement. These results are in agreement with the age-filtered classifier



---

described in Chapter 3.

Another interesting aspect of this evaluation is the performance comparison with classifiers that make use of both adult and children cases and, in addition, make use of the variables defined as contextual information (Table 6.4). The first row of Table 6.4 describes the performance of a classifier trained only with the PI. The achieved performance, though, is slightly inferior than the performance that the specialized classifiers can achieve with some combinations of contextual information (Table 6.3). The other rows of Table 6.4 describe a similar performance. This implies that the age, gender and BTloc have a minimal weight compared to the information given by the PI in the discrimination abilities of the classifiers obtained when using contextual information as input. Nevertheless, these variables, if used as contextual information, can improve the overall performance of the DSS if used with specialized classifiers. In addition, a further analysis reveals that the use of the similarity model improved, in some cases, the performance of the pediatric cases.

A final performance comparison is described in Table 6.5, which shows the predictive power of the classifiers when discriminating an independent dataset of 28 young adults. The performance achieved by any non-specialized classifier (likelihood of 0.67 at the best) is comparable to that of the classifier specialized in children (likelihood of 0.65). The use of the similarity model, though, slightly improves the performance up to a likelihood of 0.80, a statistically significant improvement ( $p < 0.01$ , Tukey's test,  $\alpha = 0.05$ ). This result demonstrates that the similarity model for assigning priors can help to achieve a better global classification for particularly difficult cases like those of young adults.

The similarity model based on the use of the contextual information to establish a proper prior to the specialized models for a test set allows the measurement of the overall performance of a DSS. In our results, the overall performance is slightly better than the one obtained with the non-specialized classifiers and, hence, the use of the similarity model can help to improve the diagnostic performance of specific populations like the pediatric cases or young adult cases, whose brain tumours are considered in the mid-way between children and adults nature [10].

The proposed approach of using the similarity model to establish the appropriateness of a classifier to each test case resembles a meta-learning approach where an exploitation and understanding of the learning itself is done in order to increase in efficiency through experience [20].

Other biomedical domains where the classification task should be specialized attending to one or several criteria might take benefit from using this approach of proper prior assessment based on a similarity model.

#### **6.4.4 Limitations and future work**

The prior probability assessment based on the similarity model has been calculated with the estimation of the probability density mixture of the contextual information. This estimation establishes mixtures of binary, discrete and continuous variables assuming independence between these variables. Any other classification technique able to estimate a distribution model and to calculate a probability to each class considered is susceptible to be used instead. Alternative methods like the web metrics for auditing the importance of web pages may be studied. Some of these algorithms take into account several evolving prominent features of the web like its hierarchical structure [21] or use features that are independent of the link structure of the web, like the frequency with which users visit web pages [22]. Such features, applied to the contextual information, may be useful to face the covariate and concept drift phenomena underlying the biomedical domains.

In order to assess the prior probabilities to each classifier (or corpora  $\mathbf{Z}_i$ ), we define a similarity model  $\mathcal{K}_i$  based on the estimation of the probability density mixture of the contextual information. By using this estimation, we can establish a mixture of binary, discrete and continuous variables assuming that they follow a Bernoulli, multinomial and Gaussian distribution given the nature of the contextual information variables. Each contextual model  $\mathcal{K}_i$  calculates, attending to the contextual information of a test sample, the probability of belonging to its design dataset,  $\mathbf{Z}_i$ . In other words, how similar is the test sample to the samples used in the design of each model. The highest probability determines the winner candidate.

The proposed similarity model needs contextual information in order to assess proper priors. This contextual information should not be used for design purposes in order to keep the independence between the prior distribution and the evidence. Hence, the proposed model requires some extra features that effectively describe the profile of each sample but allowing, at the same time, that its omission in the design stage does not drastically alter the prediction skills of the classifiers. Biomedical domains, fortunately, offer a good framework where

the classification task should be specialized attending to one or several criteria (which might be used as contextual information).

In the specific biomedical domain of brain tumour diagnosis with MRS data, another variable that may be useful as contextual information is the magnetic field of the MR scanner. The data used for this study was acquired from 1.5T MR scanners. Nevertheless, 3T scanners are becoming widely available in the clinical environment because of an improvement in the SNR and spectral resolution, which enhances the diagnostic abilities of MRS. Gathering a database of 3T with the amount of data currently available at 1.5T would take many years and can imply the risk that, by the time it will be achieved, the technology might have evolved again and scanners with greater magnetic fields will be available. In [3], we demonstrated that 3T MRS data can be used with the currently available brain tumour diagnostic classifiers trained on databases of 1.5T spectra. To do so, we evaluated a set of classifiers trained with 1.5T with two independent test sets of 1.5T and 3T. Although no significant differences were detected in the evaluation, there could be specific classification problems where the metabolites in the [3.5–4] ppm region are of paramount relevance for proper discrimination [3, 23]. Including the magnetic field of each of the training samples as contextual information might help on the selection of an appropriate candidate in case of co-existence of 1.5T and 3T MRS trained classifiers.

The evaluation of the prior assessment based on a similarity model presented in this chapter demonstrates the viability of integration of the prior assessment model into the clinical DSS CURIAM [24, 25, 26].

## Bibliography

- [1] J. Vicente, J. M. García-Gómez, S. Tortajada, A. T. Navarro, F. A. Howe, A. C. Peet, M. Julià-Sapé, B. Celda, P. Wesseling, M. Lluch-Ariet, and M. Robles, *Ranking of Brain Tumour Classifiers using a Bayesian Approach*, vol. 5517 of *Lecture Notes in Computer Science*, pp. 1005–1012. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.
- [2] J. Vicente, S. Tortajada, E. Fuster-Garcia, J. M. García-Gómez, and M. Robles, “An audit method suited for decision support systems for clinical environment,” in *Proceedings of 2012 IEEE 6th International Conference on Intelligent Systems (IS’12), Sofia, Bulgaria*, pp. 281–288, Sept. 2012.

- [3] E. Fuster-Garcia, C. Navarro, J. Vicente, S. Tortajada, J. García-Gómez, C. Sáez, J. Calvar, J. Griffiths, M. Julià-Sapé, F. Howe, J. Pujol, A. Peet, A. Heerschap, A. Moreno-Torres, M. Martínez-Bisbal, B. Martínez-Granados, P. Wesseling, W. Semmler, J. Capellades, C. Majós, A. Alberich-Bayarri, A. Capdevila, D. Monleón, L. Martí-Bonmatí, C. Arús, B. Celda, and M. Robles, “Compatibility between 3T 1H SV-MRS data and automatic brain tumour diagnosis support systems based on databases of 1.5T 1H SV-MRS spectra,” *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 24, no. 1, pp. 35–42, 2011.
- [4] G. Purcell, *What makes a good clinical decision support system*. British Medical Association, 2005.
- [5] R. A. Fisher, *Statistical methods for research workers*. Edinburgh, UK: Oliver and Boyd, 1925.
- [6] eTUMOUR Consortium, “eTumour: Web accessible MR Decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data.” Web site. FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC, <http://www.etumour.net> (Accessed: 22nd April 2012; at writing time it was temporarily unavailable).
- [7] D. N. Louis, H. Ohgaki, O. D. Wiestler, and W. K. Cavenee, *WHO classification of tumours of the central nervous system*. Lyon, France: IARC Press, 2007.
- [8] A. R. Tate, J. Underwood, D. M. Acosta, M. Julià-Sapé, C. Majós, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, M. E. Cabañas, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arús, “Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra,” *NMR in Biomedicine*, vol. 19, no. 4, pp. 411–434, 2006.
- [9] M. van der Graaf, M. Julià-Sapé, F. A. Howe, A. Ziegler, C. Majós, A. Moreno-Torres, M. Rijpkema, D. Acosta, K. S. Opstad, Y. M. van der Meulen, C. Arús, and A. Heerschap, “MRS quality assessment in a multicentre study on MRS-based classification of brain tumours,” *NMR in Biomedicine*, vol. 21, no. 2, pp. 148–158, 2008.

- 
- [10] CBTRUS, “2010 CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004-2006,” tech. rep., Central Brain Tumor Registry of the United States, 2010.
- [11] J. M. García-Gómez, J. Luts, M. Julià-Sapé, P. Krooshof, S. Tortajada, J. Vicente, W. Melssen, E. Fuster-Garcia, I. Olier, G. Postma, D. Monleón, A. Moreno-Torres, J. Pujol, A. P. Candiota, M. C. Martínez-Bisbal, J. Suykens, L. Buydens, B. Celda, S. Van Huffel, C. Arús, and M. Robles, “Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy,” *Magnetic Resonance Materials Physics, Biology and Medicine*, vol. 22, no. 1, pp. 5–18, 2009.
- [12] E. H. Simpson, “The interpretation of interaction in contingency tables,” *Journal of the Royal Statistical Society*, vol. Series B, no. 13, pp. 238–241, 1951.
- [13] C. R. Blyth, “On Simpson’s paradox and the sure-thing principle,” *Journal of the American Statistical Association*, vol. 67, no. 1, pp. 364–366, 1972.
- [14] D. V. Lindley and M. R. Novick, “The role of exchangeability in inference,” *The Annals of Statistics*, vol. 9, no. 1, pp. 45–58, 1981.
- [15] J. Pearl, *Simpson’s Paradox: An Anatomy*, ch. 6. Cambridge, UK: Cambridge University Press, 2009.
- [16] G. Bolton, E. Katok, and A. Ockenfels, “How effective are electronic reputation mechanisms? an experimental investigation,” *Management Science*, vol. 50, no. 1, pp. 1587–1602, 2004.
- [17] S. D. Ramchurn, D. Huynh, and N. R. Jennings, “Trust in multi-agent systems,” *Knowledge Engineering Review*, vol. 19, no. 1, pp. 1–25, 2004.
- [18] G. M. Reynolds, A. C. Peet, and T. N. Arvanitis, “Generating prior probabilities for classifiers of brain tumours using belief networks,” *BMC medical informatics and decision making*, vol. 7, p. 27, 2007.
- [19] H. González-Vélez, M. Mier, M. Julià-Sapé, T. N. Arvanitis, J. M. García-Gómez, M. Robles, P. H. Lewis, S. Dasmahapatra, D. Dupplaw, A. Peet, C. Arús, B. Celda, S. V. Huffel, and M. Lluch i Ariet, “HealthAgents: Distributed multi-agent brain tumour diagnosis and prognosis,” *Journal of Applied Intelligence*, p. 191, 2007.

- [20] R. Vilalta and Y. Drissi, “A perspective view and survey of meta-learning,” *Artificial Intelligence Review*, vol. 18, no. 2, pp. 77–95, 2002.
- [21] N. Eiron, K. S. McCurley, and J. A. Tomlin, “Ranking the web frontier,” in *WWW '04: Proceedings of the 13th international conference on World Wide Web*, (New York, NY, USA), pp. 309–318, ACM Press, 2004.
- [22] M. Richardson, A. Prakash, and E. Brill, “Beyond pagerank: machine learning for static ranking,” in *WWW '06: Proceedings of the 15th international conference on World Wide Web*, (New York, NY, USA), pp. 707–715, ACM Press, 2006.
- [23] P. B. Barker, D. O. Hearshen, and M. D. Boska, “Single-voxel proton mrs of the human brain at 1.5t and 3.0t,” *Magnetic Resonance in Medicine*, vol. 45, no. 5, pp. 765–769, 2001.
- [24] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, M. Esparza, A. Navarro, E. Fuster-García, M. Robles, L. Martí-Bonmatí, and C. Arús, “A generic decision support system featuring an assembled view of predictive models for magnetic resonance and clinical data,” in *Book of abstracts of the ESMRMB 2008: 25th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Valencia, Spain*, p. 483, Springer-Verlag, Oct. 2008.
- [25] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, E. Fuster-García, M. Esparza, A. Navarro, and M. Robles, “Curiam BT 1.0, Decision Support System for Brain Tumour Diagnosis,” in *Book of Abstracts of the ESMRMB: 26th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Antalya, Turkey*, vol. 22, p. 538, Springer Berlin / Heidelberg, May 2009.
- [26] C. Sáez, J. M. García-Gómez, J. Vicente, S. Tortajada, J. Luts, D. Dupplaw, S. V. Huffel, and M. Robles, “A generic and extensible automatic classification framework applied to brain tumour diagnosis in HealthAgents,” *The Knowledge Engineering Review*, vol. 26, no. Special Issue 03, pp. 283–301, 2011.

## Chapter 7

# CURIAM DSS for BT diagnosis support

*“I do not fear computers. I fear the lack of them.”*

*Isaac Asimov*

The inclusion of DSSs into a clinical site may improve the clinicians’ decision routine workflows. The final purpose of the experiments reported in previous chapters is the improvement of brain tumour diagnosis in clinical or research environments throughout the integration of the PR developments in clinical DSSs.

In this chapter, after a brief review of previous clinical DSSs in the literature, we focus on CURIAM BT, a DSS for brain tumour diagnosis support, which includes the specific contributions described in this dissertation: the incorporation of the pediatric classifiers as an effective non-invasive pre-operative tool to define the tumour resection strategy; and the incorporation of the audit method and the similarity model as tools that will positively influence the decision process followed by the clinicians.

*CURIAM BT diagnosis support is a DSS that borrows in the multidisciplinary research line Biomedical-Mining from the IBIME group of the research institute ITACA in the Universitat Politècnica de València. The development of*

*CURIAM has been paralleled and complementary to those of the eTUMOUR, and HEALTHAGENTS European Union projects of the 6th Framework Programme. Partial sections of this Chapter have appeared as scientific contributions in [1, 2, 3, 4, 5, 6].*

## 7.1 Decision support systems in clinical environments for cancer research

Since the first clinical DSS designed in the 1970s by Leaper et al. [7] for the support of diagnosis and surgery of acute abdominal pain based on a naive Bayesian approach, there have been numerous attempts to automate and improve the decision support in the clinical practice. Related to cancer diagnosis and treatment, the first efforts were materialized in 1981 with a clinical DSS designed by Shortliffe et al. [8] to assist physicians in the treatment of cancer patients receiving chemotherapy. A continuous series of specific clinical DSSs aimed to cancer detection and diagnosis have been published in clinical and technical journals in the last three decades. Most studies have focused on the solution of questions related to specific medical problems, such as breast [9, 10], gastrointestinal [11, 12, 13], hematologic [14, 15, 16], oral [17], lung [18, 19], bladder [20], or prostate [21, 22, 23] cancer detection or diagnosis. Nevertheless, few systems have invested effort in providing generic distributed solutions to standardize the incorporation of predictive models in clinical DSSs [24].

In order to obtain a translational clinical DSS, several aspects need to be accomplished [25, 26, 27]: (1) the accuracy in solving the specific problem should be demonstrated; (2) the adequate evaluation of the capability to generalize the performance on new cases of the problem domain should be given; (3) the clinical community will accept easier the DSS if it is transparent in relation to clinical knowledge; (4) the DSS should provide an easy access to the users; and (5) the DSS needs to be integrated with the clinical routine workflow.

These are key points in the deployment of the system in a clinical environment and have been taken into account in the design of CURIAM BT, which also provides a generic solution to standardize the incorporation of predictive models.

Sections 7.2 and 7.3 describe, respectively, the incorporation of the pediatric classifiers reported in this dissertation into CURIAM BT DSS for brain tumour diagnosis support and how the audit method, along with the similarity model,



are included into CURIAM BT. Section 7.4 will discuss how the incorporation of these contributions into CURIAM BT addresses the key points described as mayor issues in the design of a translational DSS for clinical environment. Finally, this chapter finishes with some conclusions.

## **7.2 Inclusion of pediatric classifiers into CURIAM BT**

CURIAM BT is a DSS able to handle several predictive models which can deal with categorical and numerical data or signals. In order to extend the system capabilities, CURIAM BT offers the feature of incorporating new classifiers to the system just by simple ‘drag & drop’. This functionality enables both to improve the available classifiers and the incorporation of new ones.

### **7.2.1 Classification Framework (CF)**

CURIAM BT architecture is modular. It contains an independent and automatic Classification Framework (CF) that allows:

- (1) To ensure — due to the CF capability to be easily extended with new techniques — that new improvements in PR and brain tumour classification can be rapidly integrated from validated research results into clinical environments.
- (2) To incorporate new classifiers into CURIAM by including their specification files. These specification files are created following the provided Extensible Markup Language (XML)<sup>1</sup> specification, as shown in [24]. The use of the XML standard enables new classifier specifications to be extended when the framework is improved with new methods. It also enables the inclusion of additional information about the field to which a classifier is related. This information can be displayed to the user to enhance the decision support task. In addition, XML-based documents can be interpreted by many computer tools and applications, which enables the management of classifiers as well as its sharing with other systems or tools.

---

<sup>1</sup>Extensible Markup Language, <http://www.w3.org/XML/>

## 7.2.2 PR techniques in CURIAM BT

In the context of brain tumour diagnosis, procedures such as data pre-processing, feature selection, and extraction are crucial and can drastically affect the general accuracy of the system [28, 29]. In this way, one PR technique may work better for a specific problem than another. Our experiments follow evaluation protocols to ensure the reliability of the obtained classifiers and their accuracies [30]. The CF mainly focuses on the classification process, leaving the training process to the laboratory tools that individual PR research groups are accustomed to using. Therefore, the PR techniques that can be included in the CF do not need to implement the training algorithm process but just the algorithm related to the classification process. The PR techniques that are currently available in the CF for brain tumour classification are LDA, KNN, LS-SVM, PCA and ICA.

## 7.2.3 A Classifier in CURIAM BT

The CF separates the specification of a classifier from the implementation of the PR techniques it uses. A classifier added to the system is defined by a set of specification files in order to make use of any of the techniques included in the CF. This approach enables technique implementations that are already part of the system to be reused.

PR classifiers, as defined in this framework, are structured by different actions for making predictions, which can include various methods for feature selection and extraction, data transformation, and classification. Conceptually, these methods can be considered as classifier actions, where each method takes an input and provides an output. A classifier can also pipeline a set of these actions before obtaining its final result.

In statistical PR, a data sample consists of an array of features, and most types of data can be expressed in this way for data mining, e.g. medical and biological data. Even if classifiers use combinations of different data, this array could be formed as a concatenation of the two original arrays. The CF can receive these arrays, and select from them the set of features a classifier needs to make a classification. In this way, the input of the framework becomes standard. An example of such functionality can be seen in Figure 7.1. Since the CF is offered as a generic tool, the classifiers are able to deal with most types of data and solve the questions that can be expressed by this data.

By allowing an open data input and the possibility of using any classifica-

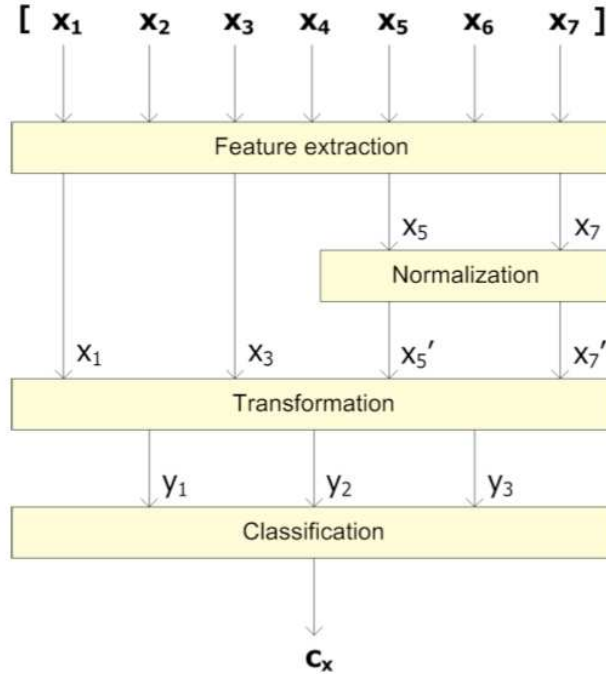


Figure 7.1: Example of a set of classifier actions pipelined to obtain a classification of an array of features. From the whole set of features (from  $x_1$  to  $x_7$ ), four are selected by the feature extraction filter ( $x_1, x_3, x_5$  and  $x_7$ ). Normalization is applied to  $x_5$  and  $x_7$ . A transformation is then applied to  $x_1, x_3,$  and the normalized features  $x_5'$  and  $x_7'$ . Finally, the output of the transformation is sent to the classification module in order to obtain the posterior probabilities of each class.

tion technique, or even a combination of techniques, the CF provides a flexible method of including classifiers in the system. An XML specification template has been defined in the CF. Therefore, to add a classifier to the system, it is only necessary to write the corresponding XML specification and add it to the framework.

The XML classifier specification is divided into three separate XML files (Figure 7.2):

**description.xml**: contains information related to the description of the classifier, such as what kind of input data is used, the question it solves, where it was developed, and what cases have been used for its training. Based on this file, the classifier gives information to the users so that they know whether it is suitable to their problem or not.

**profile.xml**: contains information related to the classifier's profile: classifier accuracy figures, statistical information on the training data, and dynamic values about the use of the classifier — such as a measure of the correctness of the classifier answers.

**method.xml**: contains the specification of the classification process. This file enables the inclusion of the actions required by the classifier, where the set of parameters used for its execution and the selection of features for its input are defined for each classifier action. The **method.xml** file describes which outputs will be returned by the classifier, and these can be any combination of output from any of the actions. In addition, extra elements such as plots or data visualizations can be included in this file.

A simple classification scenario using the CF can be described as follows: the client module (a user) requests the description of the available classifiers to check which fit its needs. Once a valid classifier is selected, it is instantiated by indicating its name. The classifier parses the content of its **method.xml** specification and creates an array of classifier actions. Parameters specified in the XML are loaded by these actions. When the classifier is ready to perform a classification, the data sample to be classified is received from the client. Data passes through the pipeline of classifier actions, and generates the corresponding outputs at each step. Once the final action is complete, the classifier generates the corresponding output which is sent back to the client module. This procedure is illustrated in Figure 7.3.

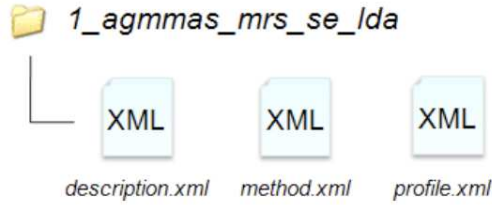


Figure 7.2: A classifier is represented by three XML specification files.

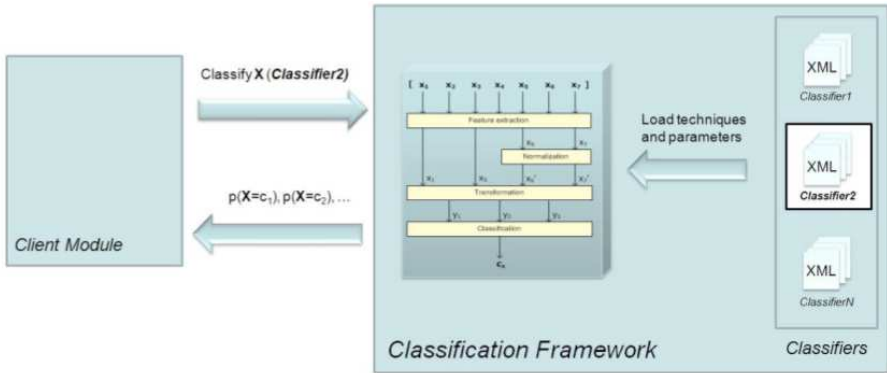


Figure 7.3: A client module requests the classification of X using the classifier Classifier2. The classifier actions specified in the `method.xml` file of Classifier2 are loaded with their corresponding parameters to create the pipeline. Probabilities are sent back to the client module once the classification of X is complete.

Table 7.1: Performance of the classifiers included into CURIAM BT for discrimination of aggressive and non-aggressive BT, extracted from Table 3.3. Balanced Accuracy Rate (BAR) of Adult Classifiers trained with 300 Short-TE samples and Children Classifiers trained with 60 Short-TE samples.

Classifier	Training Performance	Adults Test Set (96 cases)	Children Test Set (93 cases)
Adult classifier (PI+LDA)	0.85	0.85	0.60
Pediatric classifier (PI+PCA+LDA)	0.87	0.50	0.88

## 7.2.4 Pediatric classifiers

In CURIAM BT, the inclusion of new classifiers is straightforward and consists in creating the associated XML files describing the actions each classifier needs to perform in order to obtain a classification.

Hence, from the results reported in Chapter 3, two classifiers for assessment of tumour aggressiveness are incorporated into CURIAM BT. One aimed to adult patients and other aimed to children (patients under 16 years old). Both dichotomic classifiers are able to discriminate between non-aggressive tumours — generally low tumour grade (grades I and II) with a slow growing tendency — and aggressive tumours (high tumour grades, that is, stage III and IV of the tumour). The best adult classifier applies LDA after a feature extraction with PI. In contrast, the classifier devoted to children with best performance applies PI followed by PCA and then applies an LDA. These classifiers have been trained from SV MRS acquired at Short-TE. The adult classifier has been trained from 396 samples and the classifier devoted to children has been trained with 93 samples. The classification performance of both classifiers is described in Table 7.1, extracted from Table 3.3. In addition, the mean spectra of each class is also included into each classifier. This allows easy comparison of the case under evaluation with respect to the mean spectra of aggressive and non-aggressive tumours.

Regarding the classification of the three most prevalent brain tumour in children, nine classifiers are incorporated into CURIAM BT. They are able to answer three different discriminations: a binary discrimination of glial-cell tumours (EPEN and PILOA) from MED; a binary discrimination of PILOA and

Table 7.2: Performance of the nine pediatric classifiers included into CURIAM BT, extracted from Table ?? . Balanced Accuracy Rate (BAR) of the classifiers trained with PI applied to Short-TE, Long-TE and combination of both time of echo (Long-TE+Short-TE) for each discrimination problem.

Classifier	Short-TE	Long-TE	Short-TE+Long-TE
(PILOA, EPEN) vs MED	0.88	0.85	0.89
PILOA vs MED	0.92	0.94	0.95
PILOA vs EPEN vs MED	0.76	0.69	0.92

MED; and a three-class classification for discriminating EPEN from PILOA and from MED. Each discrimination is performed with Short-TE spectra alone, with Long-TE spectra alone and with a combination of both times of echo, which gives a total of nine classifiers (3 discriminations  $\times$  3 MRS combinations). All nine classifiers apply PI to perform an LDA afterwards. In Chapter 4, no significant differences were reported between the classifiers trained using only brain tumour cases in the PF compared to classifiers trained from cases at any brain location. Thus, in order to offer discrimination tools trained with the maximum available number of samples, the classifiers introduced into CURIAM BT were trained from cases in any brain location. The performance of these classifiers is summarized in Table 7.2, extracted from Table ?? . The mean spectra of each class is also included into each classifier to allow easy comparison of the case under evaluation with respect to the mean spectra of the three most prevalent pediatric tumours.

Figure 7.4 shows two screenshots of CURIAM BT with pediatric classifiers. Classifiers in CURIAM BT do not just give an answer but show the posterior probability they assign to each class they discriminate. They also show the mean spectra and standard deviation for each class in the training dataset. Figure 7.4, top, shows the children classifier for assessing the tumour aggressiveness. Below, the pediatric classifier for discrimination of EPEN, PILOA and MED can be seen. Since it is an LDA discriminating three classes, a projection of the latent space is also shown. Each coloured dot in the latent space represents a labeled case from the training set of the classifier. Notice that the test case is projected as a squared dot in order to show where it belongs, according to the classifier.



Figure 7.4: Screenshots of CURIAM BT DSS for tumour diagnosis. Top, the classification result screen of the children classifier discriminating among aggressive and non-aggressive tumour types from Short-TE spectra. Bottom, the classification result screen of an LDA classifier discriminating PILOA, EPEN and MED from a combination of Short-TE and Long-TE MR spectra.



## **7.3 Inclusion of the audit method and the similarity model into CURIAM BT**

Since several classifiers could coexist in CURIAM BT, users may want the system to identify which classifiers are more suitable for a particular case to diagnose. The audit method, along with the similarity model to assess prior probabilities, pursues this aim. These tools can also be used to rank the obtained results from a set of classifiers, so the higher the position in the ranking, the more reliable the result. In addition, it can also be useful to solve possible conflicts between classifiers giving contradictory answers. This can happen when a test case is close to a decision boundary in one or more classifiers.

### **7.3.1 Storage of feedback diagnosis**

The audit method is based on a statistical method to compute a score for classifiers. Two types of classifier performance are taken into account: static and dynamic. Static performance is the evidence of the classifier obtained after training, during the evaluation stage of the classifier. It is an initial approach in assessing how good the classifier is according to the evaluation stage — and is usually based on an independent test set, or other techniques such as kRSTT or CV. Dynamic performance is a measure that represents the performance of the classifier over time, based on user validations of previous classifications carried out in the system. To work properly, feedback of the diagnosis from the test cases introduced into CURIAM BT is needed from the clinicians. The CF stores the results that classifiers provided for a case. When the user updates the label of a case with the diagnosis obtained from the histopathological analysis, the CF is notified and, depending on the correctness of the answers given by the classifiers, the dynamic performance values will increase or decrease. Figure 7.5 shows a diagram depicting this process.

### **7.3.2 Contextual information for proper prior assessment**

The similarity model works with the audit method by giving a measure of the suitability of a case for the classifier by using the contextual information. The classifier specification defines the distributions of the contextual information of the cases in the training set. In a clinical case, for example, values such as

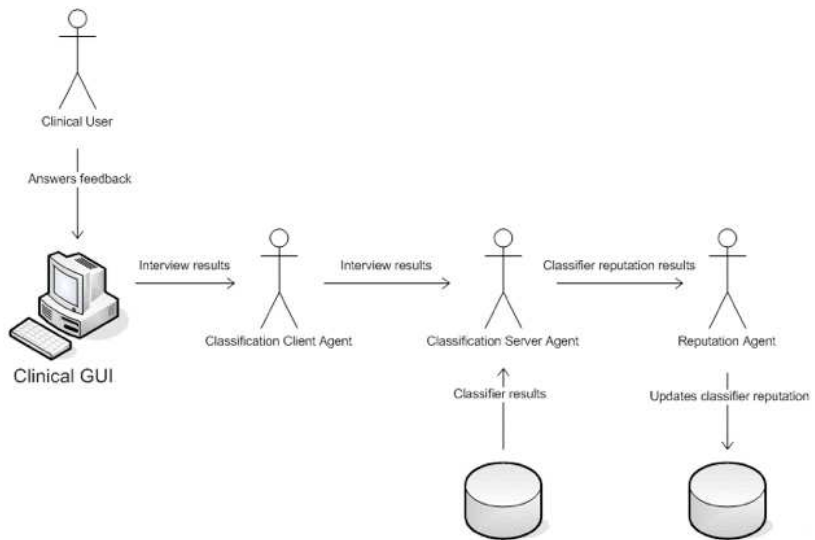


Figure 7.5: The Clinical User answers the feedback interview through the CURIAM BT Graphical User Interface. When finished, the Classification Client Agent sends the interview result to the Classification Server Agent. When the interview results are received, the Classification Server Agent joins the previously stored result of the Classifier Agent with the interview results, creating a request. This request is sent to the Reputation Agent which updates the classifier reputation according to the audit method.

patient age, gender or the location of the tumour in the brain could be used. By combining the prior odds of the similarity model and the evidence odds provided by the audit method, a score is calculated for each classifier, which is then used for sorting the classifiers.

This contextual information is asked to the user at the moment he/she introduces the MRS case into CURIAM BT. Then, when the user wants to classify the case with CURIAM BT, the DSS selects the candidate classifiers by asking the user which is (are) the label(s) he/she thinks the case should belong to (Figure 7.6). Afterwards, according to this proposed diagnosis and the data supplied with the current case (a Short-TE spectrum alone, a Long-TE spectrum or both spectra), CURIAM BT selects the classifier candidates that can give answer using that input. While all the candidates are running in background, the audit method calculates the scores. If contextual information has been given, then, the audit method will call the similarity model in order to assess the prior probability of each candidate. If, on the contrary, no contextual information is available for the current case, the audit method will assign equal priors to each candidate classifier.

Equation 7.1 represents the formula of the audit method used when comparing two classifiers  $\mathcal{M}_m$  and  $\mathcal{M}_l$ .

$$\frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} = \frac{P(\mathbf{k}^\kappa|\mathcal{K}_m)}{P(\mathbf{k}^\kappa|\mathcal{K}_l)} \frac{P(\mathbf{Z}^{T_m}|\mathcal{M}_m)}{P(\mathbf{Z}^{T_l}|\mathcal{M}_l)} \frac{P(\mathbf{Z}_{DSS}^{T_m}|\mathbf{Z}_m, \mathcal{M}_m)}{P(\mathbf{Z}_{DSS}^{T_l}|\mathbf{Z}_l, \mathcal{M}_l)} \quad (7.1)$$

This formula takes into account the data with which each classifier has been trained ( $\mathbf{Z}_m$  and  $\mathbf{Z}_l$ , respectively). The first term of the ratio,  $\frac{P(\mathbf{k}^\kappa|\mathcal{K}_m)}{P(\mathbf{k}^\kappa|\mathcal{K}_l)}$ , is calculated by the similarity model and refers to the contextual score, where  $\mathcal{K}_m$  and  $\mathcal{K}_l$  represent the contextual information associated with each classifier (i.e. the mean age of the training cases); and  $\mathbf{k}^\kappa$  is the contextual information of the current test case (namely, the age of the patient being tested); mid term,  $\frac{P(\mathbf{Z}^{T_m}|\mathcal{M}_m)}{P(\mathbf{Z}^{T_l}|\mathcal{M}_l)}$ , refers to the evaluation of the classifier while ‘in lab’ and before releasing it into the system. Finally, the rightmost term represents the dynamic performance, that is, the measurement of how well each classifier is predicting over time when working in the system. The cases being predicted are represented in the formula with  $\mathbf{Z}_{DSS}$ .

Equation 7.2 is the formula implemented in CURIAM BT, which is a generalization of equation 7.1 for the case when the system needs to compare  $M$  models:

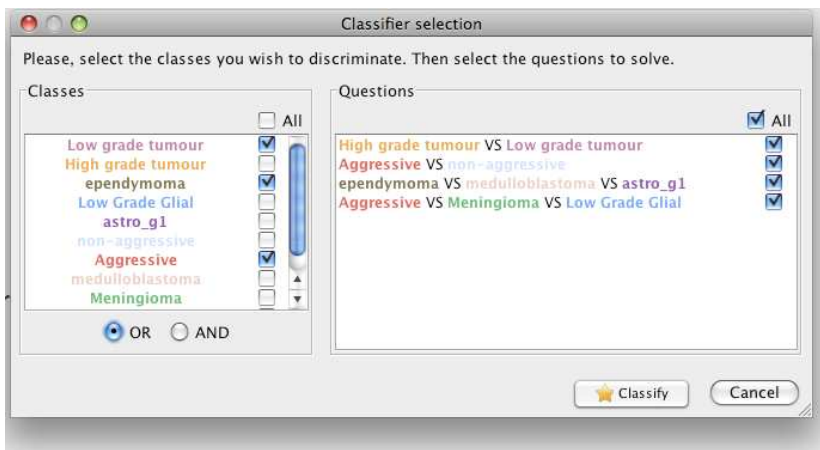


Figure 7.6: Screenshot of CURIAM BT. The user is asked to provide his/her initial diagnosis by choosing, on the left side of the screen, one or more of the labels the classifiers are able to discriminate. As the user selects labels for the initial diagnosis, CURIAM BT shows on the right side of the screen the classifiers that can deal with one or more of these labels.

$$\frac{P(\mathcal{M}_m|\mathbf{Z}, \mathbf{Z}_{DSS})}{\sum_{l=1}^M P(\mathcal{M}_l|\mathbf{Z}, \mathbf{Z}_{DSS})} = \frac{P(\mathbf{k}^\kappa|\mathcal{K}_m)P(\mathbf{Z}^{T_m}|\mathcal{M}_m)P(\mathbf{Z}_{DSS}^{T_m}|\mathbf{Z}_m, \mathcal{M}_m)}{\sum_{l=1}^M P(\mathbf{k}^\kappa|\mathcal{K}_l)P(\mathbf{Z}^{T_l}|\mathcal{M}_l)P(\mathbf{Z}_{DSS}^{T_l}|\mathbf{Z}_l, \mathcal{M}_l)} \quad (7.2)$$

## 7.4 Discussion

The combined use of the CF with the audit method and similarity model enables up-to-date management of sets of classifiers. Due to several factors, such as a poor design, overtraining, or an unrepresentative training set, the level of classifier trust may vary over time. The dynamic component of the audit method penalizes, or rewards, classifiers depending on their accuracy over time. The more reliable classifiers earn better scores. Therefore, in a large set of classifiers, even for the same question, the best classifiers will be identified by leaving to the CF the task of selecting which classifiers to show to the final user. This approach enables the use of multiple classifiers for the same question, each built from different PR methods or trained from different datasets. Moreover, the contextual component of the similarity model will emphasize those classifiers most suitable for a case of particular contextual information. Therefore, the audit method can make intelligent recommendations over a large set of classifiers.

In addition, a brief discussion on how the pediatric classifiers and the audit method contribute to accomplish the main aspects suggested in [25, 26, 27] for the design of a translational clinical DSS is given in the next subsections.

### 7.4.1 Demonstrate the accuracy in solving a specific problem

CURIAM BT is a clinical DSS for brain tumour diagnosis based on SV MRS data. It is specialized in the classification of brain tumour cases by solving different questions regarding diagnosis or grading. Different questions can be selected by the user according to the tumour classes or groups from which to obtain a prediction.

With the incorporation of the diagnostic classifiers presented in chapters 3 and 4, CURIAM BT is able to solve the specific problem of determining the aggressiveness of a brain tumour in adults and in children and to solve the problem of

discriminating among the three most common pediatric brain tumours. These classifiers have been properly evaluated ‘in lab’ and their performances and results are supported by scientific publications. The performance of each classifier in CURIAM BT is available to the user in order to let him/her know how reliable the classifier is.

The accuracy in solving a specific problem is not only supported by scientific publications but also is shown as a classification result: the answer given by a classifier in CURIAM BT is accompanied by a piechart that shows the posterior probability associated to each discriminating class (Figure 7.6). In that way, the diagnostic orientation given is supported by the *belief degree* of the classifier. This allows the user to have more confidence, for example, in a classifier that discriminates two classes and assigns a 0.90 posterior probability to one of the labels than in another dichotomic classifier assigning 0.60 of posterior probability to the winner label.

#### **7.4.2 Provide an adequate evaluation of the capability to generalize the performance on new cases of the problem domain**

The audit method described in Chapter 5 provides an appropriate framework to measure the generalization capabilities of a set of classifiers in CURIAM BT. The audit method calculates a score to each comparing classifier based on the cases used in the design stage (static odds) and the cases already entered in CURIAM BT for diagnosis support (dynamic odds). This method, as discussed in Section 5.5, is able to detect misbehaving models and data shift in the clinical site: it may occur that in some cases like that where a clinical site applies a different acquisition protocol to the one followed by our classifiers, the ‘in lab’ performance is not enough to demonstrate the predictive abilities of a classifier in solving a specific problem. In such scenario, the audit method contributes to detect classifiers optimistically biased that behaved correctly when trained but show poor generalization ability once deployed in the system and, hence, can give guidance on why this bias is happening.

### **7.4.3 Acceptance from the clinical community by offering clinical knowledge**

The GUI of CURIAM BT has been designed taking into account the input provided by expert spectroscopists and radiologists from a series of interviews performed for obtaining the system's requirements previous to the development of the DSS. One of the mayor premises was that the system did not provided just an answer in a 'black box' fashion, but that this answer was supported for clinical knowledge that would help the clinician in their diagnosis. Several aspects contribute in offering this required clinical knowledge:

- A classifier in CURIAM BT shows the answer accompanied with the posterior probability they assign to each class they discriminate. Such information provides a belief degree in the answer given by the classifier and helps the clinician to decide the confidence that such answer can provided to formulate his/her final diagnostic decision.
- A classifier in CURIAM BT displays a plot with the mean spectra and standard deviation of each class in the training dataset. This allows the user to contrast the mean spectra with the signal of the case under evaluation. Amplification of a specific region of the plot is allowed (zoom). Several checkboxes in the GUI give detailed management of this plot by allowing to display the mean spectra of each class, the mean spectra and the standard deviations, and allowing the visualization of the mean spectra from all the classes the classifier discriminates among or just of some specific classes. In order to provide an easy detection of the metabolites involved in the spectra, vertical lines indicating the main metabolites in the plot (and, optionally, their names) can also be displayed.
- An LDA classifier discriminating three classes displays in CURIAM BT a plot with the projection of the latent space of the samples used to train the classifier. The test case under evaluation is projected and displayed in the latent space plot. This provides a visual interpretation of the classifier, that no longer remains as a 'black box' but shows how its decision boundaries and the samples from each class are distributed. This plot provides zoom in and out functionalities. In addition, the user can click on any training sample. By doing so, CURIAM BT shows in the mean spectra plot the signal of the case selected by the user. With this functionality, CURIAM BT provides the user with a tool to compare the case

under evaluation not only with respect to the mean spectra, but also with specific cases used in the design of a classifier.

#### 7.4.4 Provide an easy access to the users

CURIAM BT has recently been evaluated in three different clinical centers. This evaluation has been carried out under the Spanish project *CURIAM-Sistema de ayuda a la decisión medica. Plan de explotación y prueba de concepto en el Hospital Universitario Doctor Peset*. One of the aspects covered by this project was the evaluation of the usability of the DSS with a quantitative study and a qualitative study. The qualitative study evaluated the usefulness and ease of use of the system perceived by the users. The tool used to measure the system's usability was the TAM questionnaire (Technology Acceptance Model) adapted to the specific problem of the brain tumour diagnosis with MRS. The quantitative study was designed as a randomized controlled trial to measure the added value and effectiveness of CURIAM BT as an additional tool to the usual diagnostic techniques in novel clinicians.

Users gave a mean value of 3.6 over 5 to the usefulness of CURIAM BT. The ease of use was scored with a mean value of 4.3 over 5. With regard to the quantitative study, a total of 46 cases were assigned to control (diagnosis using usual techniques) and 41 to experimental (usual techniques and CURIAM BT). 57% of diagnostic accuracy was achieved in the control group, whereas 70% was obtained in the experimental group. The results in the quantitative study showed a clear tendency of improvement in the diagnosis success with the use of CURIAM BT as an additional tool, specially in novel clinicians.

This evaluation demonstrate that CURIAM BT can be accessible to clinical users and reveals that this DSS is seen as a useful tool for diagnostic support both by junior and senior clinicians.

#### 7.4.5 Integration with the clinical routine workflow

The pediatric classifiers presented in Chapter 4 are designed to assist a multi-disciplinary team on the neurosurgical strategy for tumour resection in children. Depending on the tumour type, a different action protocol should be taken. Hence, a macroscopic total resection is recommended for EPEN; a resection and posterior application of irradiation and chemotherapy is recommended for



MED; and partial resection can be performed in case of PILOA since these lesions show good prognosis although small residuals are observed after resection.

CURIAM supplies tools for automatic pre-processing of MRS raw data from different manufacturers (Siemens, GE, Philips) by means of automatic calls to external jMRUI [31] and jDMS [32] applications. Additionally, jMRUI or DMS pre-processed signals can also be loaded. A manual phase correction tool is also provided. These capabilities allow an easier integration of CURIAM BT into a research environment, where the use of jMRUI and jDMS applications is common and extended; and into clinical environment by enabling its compatibility with raw data from the three main manufacturers of MRS.

In addition, CURIAM offers two views conceived for confirmation diagnosis of routine cases and for research purposes for cases needing further investigation. For routine cases, the results of the available models are assembled in a common view specialized for the task domain. This assemble view offers a faster and wider criterion for clinical diagnosis by means of several classifiers for different features, questions or learning algorithms (Figure 7.7). In this view, the results are sorted according to the scores assigned by the audit method, so the first answers correspond to the classifiers more suited for the case under evaluation. Besides, an advanced view for specific classifier can also be displayed for research purposes on specific interesting cases (Figure 7.4). In this view, the classifier shows the answer accompanied with the posterior probability they assign to each class they discriminate. It also shows the mean spectra and standard deviation for each class in the training dataset. In some cases (LDA discriminating three classes) a projection of the latent space is also shown.

## 7.5 Conclusions

This Chapter has described the inclusion of the scientific contributions studied in this Thesis into CURIAM BT, a DSS for BT diagnosis support: The incorporation of the pediatric classifiers as an effective non-invasive pre-operative tool to define the tumour resection strategy; and the incorporation of the audit method and the similarity model. Besides, a discussion on how these contributions positively influence the decision process followed by the clinicians is also given, demonstrating that such practical solutions are suited for deployment in clinical or research environments.

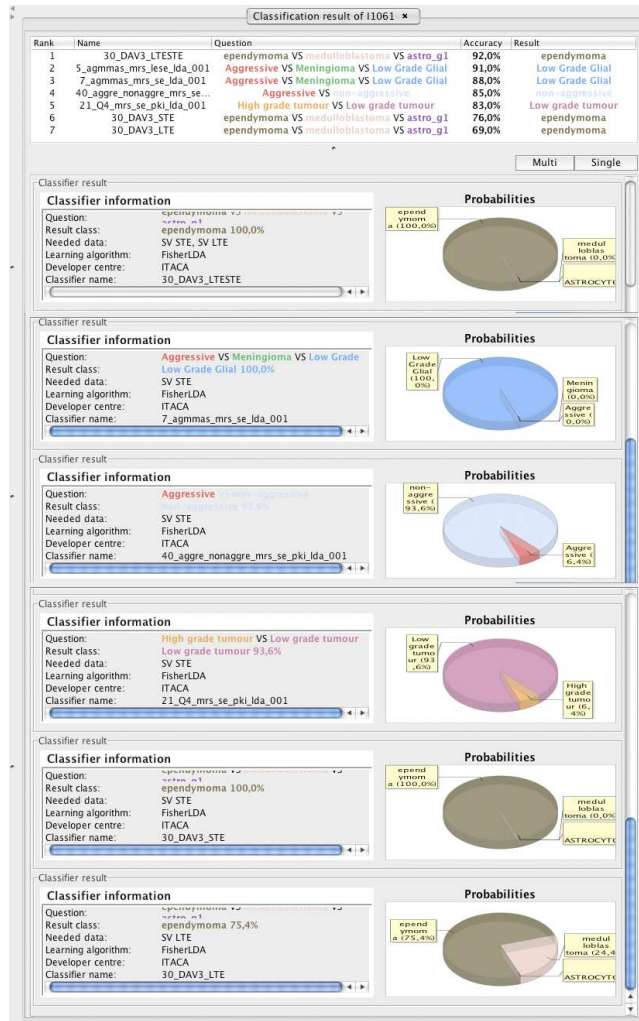


Figure 7.7: Screenshot of the multiple classifiers view in CURIAM BT. The classifiers are sorted according to the audit method. Each classifier shows its answer along with a piechart with the posterior probabilities for each discriminating class.

## Bibliography

- [1] J. M. García-Gómez, S. Tortajada, J. Vicente, C. Sáez, X. Castells, J. Luts, M. Julià-Sapé, A. Juan-Císcar, S. V. Huffel, A. Barcelo, J. Ariño, C. Arús, and M. Robles, *Genomics and Metabolomics Research for Brain Tumour Diagnosis Based on Machine Learning*, vol. 4507/2007 of *Lecture Notes in Computer Science*, pp. 1012–1019. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.
- [2] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, M. Esparza, A. Navarro, E. Fuster-Garcia, M. Robles, L. Martí-Bonmatí, and C. Arús, “A generic decision support system featuring an assembled view of predictive models for magnetic resonance and clinical data,” in *Book of abstracts of the ESMRMB 2008: 25th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Valencia, Spain*, p. 483, Springer-Verlag, Oct. 2008.
- [3] M. Lluch-Ariet, F. Estanyol, M. Mier, C. Delgado, H. González-Vélez, T. Dalmas, M. Robles, C. Sáez, J. Vicente, S. Huffel, J. Luts, C. Arús, A. P. C. Silveira, M. Julià-Sapé, A. Peet, A. Gibb, Y. Sun, B. Celda, M. C. M. Bisbal, G. Valsecchi, D. Dupplaw, B. Hu, and P. Lewis, “On the implementation of healthagents: Agent-based brain tumour diagnosis,” in *Agent Technology and e-Health* (R. Annicchiarico, U. Cortés, C. Urdiales, M. Walliser, S. Brantschen, M. Calisti, and T. Hempfling, eds.), Whitestein Series in Software Agent Technologies and Autonomic Computing, pp. 5–24, Birkhäuser Basel, 2008.
- [4] C. Sáez, J. García-Gómez, J. Vicente, S. Tortajada, E. Fuster-Garcia, M. Esparza, A. Navarro, and M. Robles, “Curiam BT 1.0, Decision Support System for Brain Tumour Diagnosis,” in *Book of Abstracts of the ESMRMB: 26th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Antalya, Turkey*, vol. 22, p. 538, Springer Berlin / Heidelberg, May 2009.
- [5] C. Sáez, J. M. García-Gómez, J. Vicente, S. Tortajada, J. Luts, D. Dupplaw, S. V. Huffel, and M. Robles, “A generic and extensible automatic classification framework applied to brain tumour diagnosis in HealthAgents,” *The Knowledge Engineering Review*, vol. 26, no. Special Issue 03, pp. 283–301, 2011.

- 
- [6] J. Vicente, C. Sáez, S. Tortajada, E. Fuster-Garcia, M. Esparza, M. Robles, and J. García-Gómez, “Curiam BT kids, a Clinical DSS for pediatric brain tumour diagnosis,” in *Book of Abstracts of the ESMRMB: 29th Annual Scientific Meeting of the European Society for Magnetic Resonance in Medicine and Biology, Lisbon, Portugal*, Springer Berlin / Heidelberg, October 2012. Accepted. In press.
- [7] D. J. Leaper, F. T. De Dombal, J. C. Horrocks, and J. R. Staniland, “Computer-assisted diagnosis of abdominal pain using estimates provided by clinicians.,” *The British journal of surgery*, vol. 59, no. 11, pp. 897–898, 1972.
- [8] E. H. Shortliffe, A. C. Scott, M. B. Bischoff, A. B. Campbell, W. va. Melle, and C. D. Jacobs, “Oncocin: An expert system for oncology protocol management,” in *Seventh International Joint Conference on Artificial Intelligence*, (Vancouver), 1981.
- [9] P. J. G. Lisboa, H. Wong, P. Harris, and R. Swindell, “A bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer.,” *Artificial intelligence in medicine*, vol. 28, no. 1, pp. 1–25, 2003.
- [10] J. L. Y. Liu and J. C. Wyatt, “The case for randomized controlled trials to assess the impact of clinical information systems,” *Journal of the American Medical Informatics Association*, vol. 18, no. 2, pp. 173–180, 2011.
- [11] P. J. Lucas, H. Boot, and B. G. Taal, “Computer-based decision support in the management of primary gastric non-hodgkin lymphoma.,” *Methods of information in medicine*, vol. 37, no. 3, pp. 206–219, 1998.
- [12] M. M. Zheng, S. M. Krishnan, and M. P. Tjoa, “A fusion-based clinical decision support for disease diagnosis from endoscopic images.,” *Computers in biology and medicine*, vol. 35, no. 3, pp. 259–274, 2005.
- [13] J. Saleem, L. Militello, N. Arbuckle, M. Flanagan, D. Haggstrom, J. Linder, and B. Doebbeling, “Provider perceptions of colorectal cancer screening clinical decision support at three benchmark institutions,” in *American Medical Informatics Association (AMIA) Annual Symposium Proceedings, 2009*, pp. 558–562, AMIA, 2009. PMC2815413.
- [14] D. J. Foran, D. Comaniciu, P. Meer, and L. A. Goodell, “Computer-assisted discrimination among malignant lymphomas and leukemia using

- immunophenotyping, intelligent image repositories, and telemicroscopy.” *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 4, no. 4, pp. 265–273, 2000.
- [15] W. L. Tung and C. Quek, “Genso-fdss: a neural-fuzzy decision support system for pediatric all cancer subtype identification using gene expression data.” *Artificial intelligence in medicine*, vol. 33, no. 1, pp. 61–88, 2005.
- [16] R. Ratei, L. Karawajew, F. Lacombe, K. Jagoda, G. Del Poeta, J. Kraan, M. De Santiago, J. Kappelmayer, E. Bjorklund, W.-D. Ludwig, J. W. Gratama, and A. Orfao, “Discriminant function analysis as decision support system for the diagnosis of acute leukemia with a minimal four color screening panel and multiparameter flow cytometry immunophenotyping.” *Leukemia*, vol. 21, no. 6, pp. 1204–1211, 2007.
- [17] G. S. Nayak, S. Kamath, K. M. Pai, A. Sarkar, S. Ray, J. Kurien, L. D’Almeida, B. R. Krishnanand, C. Santhosh, V. B. Kartha, and K. K. Mahato, “Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra: classification of normal premalignant and malignant pathological conditions.” *Biopolymers*, vol. 82, no. 2, pp. 152–166, 2006.
- [18] G. Coppini, S. Diciotti, M. Falchini, N. Villari, and G. Valli, “Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms.” *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 7, no. 4, pp. 344–357, 2003.
- [19] P.-M. Roy, P. Durieux, F. Gillaizeau, C. Legall, A. Armand-Perroux, L. Martino, M. Hachelaf, A.-E. Dubart, J. Schmidt, M. Cristiano, J.-M. Chretien, A. Perrier, and G. Meyer, “A computerized handheld decision-support system to improve pulmonary embolism diagnosis,” *Annals of Internal Medicine*, vol. 151, no. 10, pp. 677–686, 2009.
- [20] P. Spyridonos, D. Cavouras, P. Ravazoula, and G. Nikiforidis, “A computer-based diagnostic and prognostic system for assessing urinary bladder tumour grade and predicting cancer recurrence.” *Medical informatics and the Internet in medicine*, vol. 27, no. 2, pp. 111–122, 2002.
- [21] B. M. Kelm, B. H. Menze, C. M. Zechmann, K. T. Baudendistel, and F. A. Hamprecht, “Automated estimation of tumor probability in prostate

- magnetic resonance spectroscopic imaging: pattern recognition vs quantification.,” *Magnetic Resonance in Medicine*, vol. 57, no. 1, pp. 150–159, 2007.
- [22] R. Montironi, L. Cheng, A. Lopez-Beltran, R. Mazzucchelli, M. Scarpelli, and P. H. Bartels, “Decision support systems for morphology-based diagnosis and prognosis of prostate neoplasms,” *Cancer*, vol. 115, no. S13, pp. 3068–3077, 2009.
- [23] D. Berry, B. Halpenny, S. Wolpin, D. B. Joyce, W. Ellis, B. Lober, M. Justin, and J. Wulff, “Development and evaluation of the Personal Patient Profile-Prostate (P3P), a Web-based Decision Support System for men newly diagnosed with localized prostate cancer,” in *AMIA Annual Symposium Proceedings, 2010*, p. e67, AMIA.
- [24] J. M. Garcia-Gomez, C. Vidal, J. Vicente, L. Martí-Bonmatí, and M. Robles, “Medical decision support system for diagnosis of soft tissue tumors based on distributed architecture,” in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, vol. 5, pp. 3225–8, Jan. 2004.
- [25] B. Kaplan, “Evaluating informatics applications—clinical decision support systems literature review,” *International journal of medical informatics*, vol. 64, no. 1, pp. 15–37, 2001.
- [26] P. J. Lisboa and A. F. G. Taktak, “The use of artificial neural networks in decision support in cancer: A systematic review,” *Neural Networks*, vol. 19, pp. 408–415, May 2006.
- [27] C. Bryan and S. A. Boren, “The use and effectiveness of electronic clinical decision support tools in the ambulatory/primary care setting: a systematic review of the literature,” *Informatics in Primary Care*, vol. 16, no. 2, pp. 79–91, 2008.
- [28] J. Luts, J. Poulet, J. M. García-Gómez, A. Heerschap, M. Robles, J. A. K. Suykens, and S. Van Huffel, “Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra,” *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 288–298, 2008.
- [29] J. M. García-Gómez, S. Tortajada, C. Vidal, M. Julià-Sapé, J. Luts, À. Moreno-Torres, S. Van Huffel, C. Arús, and M. Robles, “The effect

- of combining two echo times in automatic brain tumor classification by MRS,” *NMR in Biomedicine*, vol. 21, no. 10, pp. 1112–1125, 2008.
- [30] J. M. García-Gómez, J. Luts, M. Julià-Sapé, P. Krooshof, S. Tortajada, J. Vicente, W. Melssen, E. Fuster-Garcia, I. Olier, G. Postma, D. Monleón, A. Moreno-Torres, J. Pujol, A. P. Candiota, M. C. Martínez-Bisbal, J. Suykens, L. Buydens, B. Celda, S. Van Huffel, C. Arús, and M. Robles, “Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy,” *Magnetic Resonance Materials Physics, Biology and Medicine*, vol. 22, no. 1, pp. 5–18, 2009.
- [31] Magnetic Resonance User Interface, “MRUI: Magnetic Resonance User Interface.” Web site. [http://sermn02.uab.es/mrui/mrui\\_Overview.shtml](http://sermn02.uab.es/mrui/mrui_Overview.shtml) (Accessed:24th April 2012).
- [32] A. R. Tate, J. Underwood, D. M. Acosta, M. Julià-Sapé, C. Majós, A. Moreno-Torres, F. A. Howe, M. van der Graaf, V. Lefournier, M. M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. L. Bosson, M. E. Cabañas, A. W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P. R. Wilkins, B. A. Bell, C. Rémy, A. Heerschap, D. Watson, J. R. Griffiths, and C. Arús, “Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra,” *NMR in Biomedicine*, vol. 19, no. 4, pp. 411–434, 2006.

## Chapter 8

# Concluding remarks and future work

*“Success is getting what you want. Happiness is wanting what you get.”*

*Dale Carnegie*

### 8.1 Conclusions

This Thesis compiles some applications of the PR discipline for the analysis of biomedical data. More specifically, the aim of this Thesis is to develop computer-assisted support of brain tumour diagnosis for clinical environments based on the PR discipline.

This aim can be particularly described in several aims: to develop predictive models with high accuracy in classification aiming at the diagnosis of pediatric brain tumours with MRS data; to develop an evaluation methodology to measure the prediction skills of the classifiers working in a DSS; and to implement and integrate these developments into a DSS for clinical environment.

The technical aspects covered in the Thesis include the processing, feature extraction and modeling of MRS brain tumour data; the inference and evaluation



of predictive models; the integration of the predicting models into DSSs for the clinical environment; and the follow-up of predictive models within a DSS by means of an audit method which measures and reports their performance.

The conclusions extracted from this Thesis are:

- There exist biochemical differences in children BT compared to those arising in adults. The spectral features and metabolite comparison carried out with MRS data from two multi-centre European datasets confirms this finding.
- Classifiers aiming to assess aggressiveness of a brain tumour have been developed devoted to adult patients and child patients. They show good performance, but it dramatically lowered when children classifiers were tested with an adult test set and vice-versa. A filter based on the normal probability density function of the training dataset's age allowed to obtain a classifier able to assess the aggressiveness of a brain tumour with high accuracy, improving the overall classification performance.
- This experiment sheds light to the need of developing a tool that decides which classifier is more adequate to a specific patient. In a clinical DSS oriented to BT diagnosis and giving support to both adult and children patients, the need of a tool for auditing and selecting the most adequate predictive model is of paramount importance.
- Classification models aimed at the discrimination of the three most common brain tumour types in children have been developed. These classifiers were trained and evaluated with a multi-centre European dataset of MRS data of pilocytic astrocytomas, ependymomas and medulloblastomas.
- The combination of single voxel MRS at 1.5T at two different TE, Short-TE and Long-TE, improves the classification of pediatric brain tumours over the use of one TE alone in the discrimination of pilocytic astrocytomas, ependymomas and medulloblastomas.
- MRS with automated processing and pattern recognition provides a useful technique for accurate, non-invasive diagnosis and classification of childhood brain tumours and thereby a powerful diagnostic tool for clinical practice.
- An audit method suited for PR-based DSSs devoted to clinical or research environment has been designed and implemented. Taking as starting point

the initial diagnosis made by the clinician, it allows to objectively audit, based on the Bayesian framework, the models relevant to help on the diagnosis of the current patient.

- The designed audit method improves the user's decision workflow by comparing not just the predictive models able to answer the diagnosis proposed by the clinician, but also those models that can deal with more general labels, offering reinforcement to the decision steps already taken, and those dealing with sub-classes of the given diagnosis, which can help to further refine the diagnosis process.
- The audit method contributes to detect potential problems in the clinical setting such as possible overtraining of the predictive models, data shift in the centre when acquiring the biomedical data or the existence of different patient population.
- A prior probability assessment inspired in the Bayesian approach has been designed and implemented. This approach is defined as a similarity model that allows the DSS to select the most adequate classifiers for each test case attending to contextual information, which is information not used in the design of the classifiers but related to the case or its environment. This prior probability assessment is a natural extension of the audit method.
- Two practical solutions that apply directly to the results of the eTUMOUR and HEALTHAGENTS projects have been developed to improve the clinical decision workflow supplied by CURIAM BT, a DSS for BT diagnosis support: The incorporation of the pediatric classifiers as an effective non-invasive pre-operative tool to define the tumour resection strategy; and the incorporation of the audit method and of the similarity model as tools that will positively influence the decision process followed by the clinicians.

## 8.2 Future work

Some of the future lines of investigation directly related to the results of this Thesis are:

- Development of predictive models to assess prognosis in children patients undergoing a brain tumour resection.

- Evaluation of the impact of incorporating CURIAM BT with the pediatric classifiers in a clinical environment as an effective non-invasive pre-operative tool to define the tumour resection strategy.
- Integration of metabolite quantitation tools like TARQUIN, developed at University of Birmingham and Birmingham Children's Hospital NHS Foundation Trust, into CURIAM BT in order to wider its pre-processing skills.
- Combination of the audit method with the use of incremental learning algorithms in order to: decide the moment when a retraining should be performed; and help on detecting when a learning technique or discrimination problem has reached its 'learning roof'.
- Application of the audit method and similarity model to other challenging biomedical domains such as diagnosis of cardiovascular pathologies from electrocardiography.

# Appendix A

## Human brain cells, tissues and physiological processes

*“The chief function of the body is to carry the brain around.”*

*Thomas A. Edison*

This appendix gathers the definitions of concepts and terms related to brain cell biology and brain tumours. Although description of every brain cell, tissue and physiological process is beyond the scope of this dissertation, this appendix tries to compile all (or almost all) the concepts cited along the present Thesis to give a richer vision to the reader non-familiarized with the brain biology. The definitions have been gathered from [1].

### A.1 Brain cells

#### Neurons

A **neuron** is an electrically excitable cell that processes and transmits information by electrical and chemical signaling. Chemical signaling occurs via

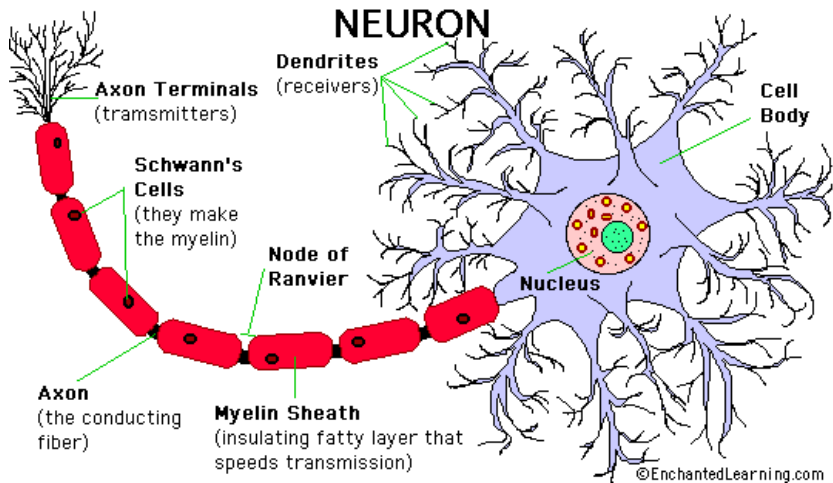


Figure A.1: Structure of a typical neuron.

From <http://www.enchantedlearning.com/subjects/anatomy/brain/Neuron.shtml>

synapses, specialized connections with other cells. Neurons connect to each other to form networks. Neurons are the core components of the nervous system, which includes the brain, spinal cord, and peripheral ganglia.

A typical neuron possesses a cell body (often called the soma), dendrites and an axon (see Figure A.1). **Dendrites** are filaments that arise from the cell body, often extending for hundreds of microns and branching multiple times, giving rise to a complex 'dendritic tree'. An **Axon** is a special cellular filament that arises from the cell body at a site called the axon hillrock and travels for a distance, as far as 1 m in humans or even more in other species. The cell body of a neuron frequently gives rise to multiple dendrites, but never to more than one axon, although the axon may branch hundreds of times before it terminates. At the majority of synapses, signals are sent from the axon of one neuron to a dendrite of another.

All neurons are electrically excitable, maintaining voltage gradients across their membranes by means of metabolically driven ion pumps, which combine with ion channels embedded in the membrane to generate intracellular-versus-extracellular concentration differences of ions such as sodium, potassium, chloride, and calcium. Changes in the cross-membrane voltage can alter the function of voltage-

dependent ion channels. If the voltage changes by a large enough amount, an all-or-none electrochemical pulse called an action potential is generated, which travels rapidly along the cell's axon, and activates synaptic connections with other cells when it arrives.

Neurons of the adult brain do not generally undergo cell division, and usually cannot be replaced after being lost, although there are a few known exceptions. In most cases they are generated by special types of stem cells, although astrocytes (a type of glial cell) have been observed to turn into neurons as they are sometimes pluripotent.

## **Glial Cells**

**Glial cells**, commonly called neuroglia or simply glia (Greek for "glue"), are non-neuronal cells that maintain homeostasis, form myelin, and provide support and protection for the brain's neurons. In the human brain, there is roughly one glia for every neuron with a ratio of about two neurons for every three glia in the cerebral grey matter.

As the Greek name implies, glia are commonly known as the glue or "bricks and mortar" of the nervous system; this, however, is not fully accurate. The four main functions of glial cells are to surround neurons and hold them in place, to supply nutrients and oxygen to neurons, to insulate one neuron from another, to destroy pathogens, and remove dead neurons. They also modulate neurotransmission.

Some glial cells function primarily as the physical support for neurons. Others regulate the internal environment of the brain, especially the fluid surrounding neurons and their synapses, and nutritify neurons. During early embryon development, glial cells direct the migration of neurons and produce molecules that modify the growth of axons and dendrites.

Glial cells can have chemical synapses and release neurotransmitters. For example, astrocytes are crucial in clearance of neurotransmitter from within the synaptic cleft, which provides distinction between arrival of action potentials and prevents toxic build-up of certain neurotransmitters such as Glu (excitotoxicity).

Glia are also crucial in the development of the nervous system and in processes such as synaptic plasticity and generation of synapsis. Futhermore, these cells have a role in the regulation of repair of neurons after injury. In the CNS, the

astrocytes are capable to enlarge and proliferate to form a scar and produce inhibitory molecules that inhibit regrowth of a damaged or severed axon. In the Peripheral Nervous System (PNS), glial cells known as Schwann cells promote repair; After axonal injury, Schwann cells regress to an earlier developmental state to encourage regrowth of the axon. This difference between PNS and CNS raises hopes for the regeneration of nervous tissue in the CNS. For example, the spinal cord may be able to be repaired following injury or severance.

Some types of glial cells are described in deeper detail:

- **Astrocytes**

Located at the CNS, are the most abundant glial cell. Astrocytes (also called astroglia) have numerous projections that anchor neurons to their blood supply (see Figure A.2). They regulate the external chemical environment of neurons by removing excess ions, specially potassium, and recycling neurotransmitters released during synaptic transmission. The current theory suggests that astrocytes may be the predominant ‘building blocks’ of the blood-brain barrier. Astrocytes may regulate vasoconstriction and vasodilation by producing substances such as arachidonic acid.

It has recently been shown that astrocyte activity is linked to blood flow in the brain, and that this is what is actually being measured in fMRI.

- **Oligodendrocytes**

Oligodendrocytes are cells that coat axons in the CNS with their cell membrane forming a specialized membrane differentiation called myelin, producing the so-called myelin sheath (see Figure A.3). The myelin sheath provides insulation to the axon that allows electrical signals to propagate more efficiently.

**Oligodendrocyte precursor cells** in nervous tissue cells precede oligodendrocytes, and may also be able to generate neurons and astrocytes.

- **Ependymal cells**

Ependymal cells, also named ependymocytes, line the cavities of the CNS and make up the walls of the ventricles. These cells create and secrete cerebrospinal fluid (CSF) and beat their cilia to help circulate that CSF and make up the Blood-CSF barrier (see Figure A.4). They are also thought to act as neural stem cells.

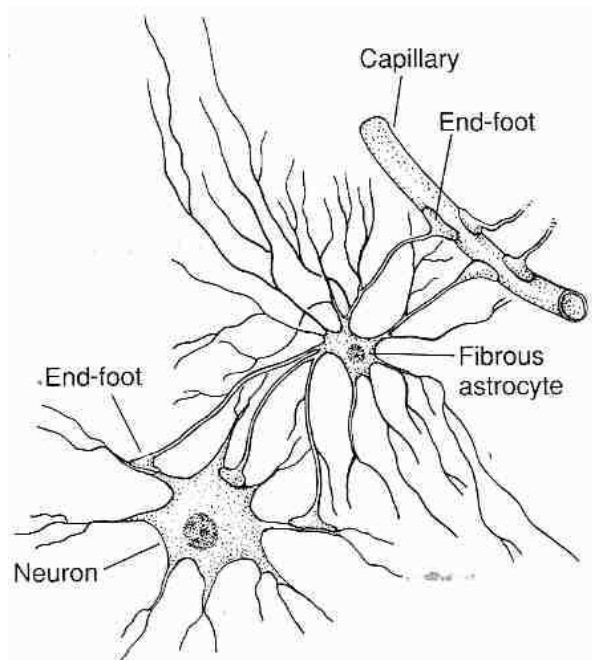


Figure A.2: Astrocyte with end-feet on capillary and neuron.  
From <http://www.benbest.com/cryonics/protocol.html#BBB>



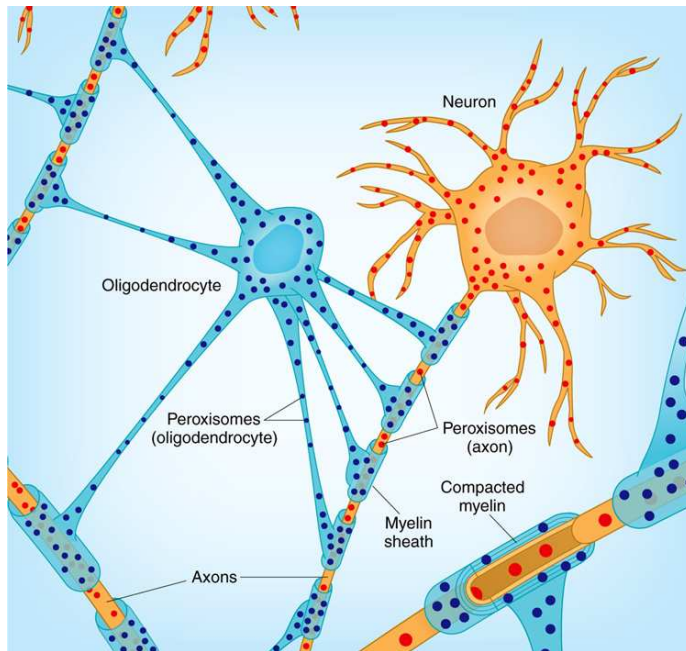


Figure A.3: Oligodendrocyte simultaneously wrapping multiple axons with a myelin sheath. Nodes of Ranvier, which are small unmyelinated axonal regions, can be observed. Peroxisomes are present in myelin sheaths surrounding axons and are also present in the axons. From [2].

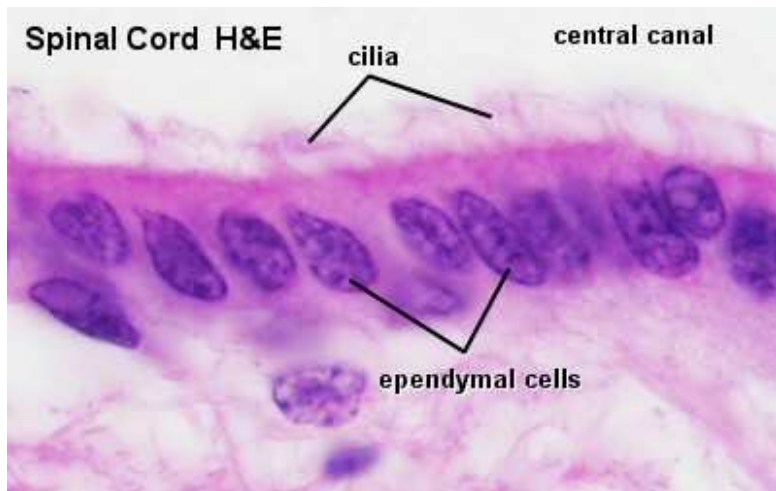


Figure A.4: Ependymal cells along the spinal cord. Image obtained with an HE stain (hematoxylin and eosin stain). From School of Anatomy and Human Biology, The University of Western Australia.

<http://www.lab.anhb.uwa.edu.au/mb140/CorePages/Nervous/Nervous.htm>

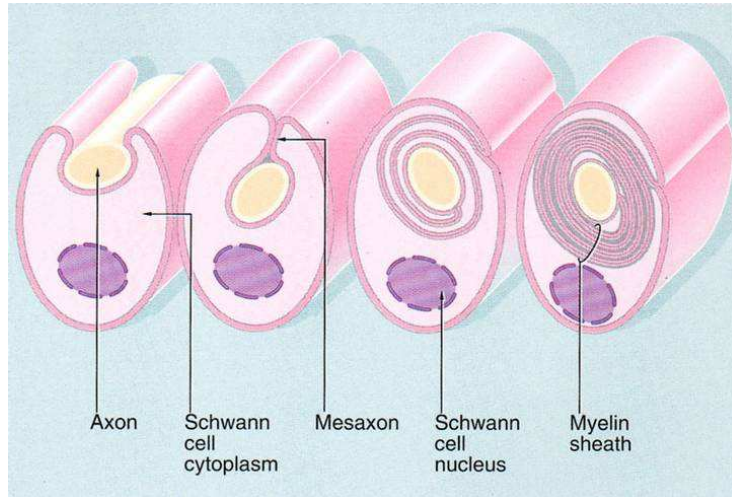


Figure A.5: Diagram of the process of myelination of an axon. Myelination begins with the invaginations of a single nerve axon into a Schwann cell; a mesoaxon is then formed. As myelination proceeds, the mesoaxon rotates around the axon enveloping it in concentric layers of Schwann cell cytoplasm and plasma membrane.

From Wheater's.

[http://www.mc.vanderbilt.edu/histology/images/histology/nervous\\_tissue/display/schwann3.jpg](http://www.mc.vanderbilt.edu/histology/images/histology/nervous_tissue/display/schwann3.jpg)

- **Schwann cells**

Similar in function to oligodendrocytes, Schwann cells provide myelination to axons in the PNS (see Figure A.5). They also have phagocytotic activity and clear cellular debris that allows for regrowth of PNS neurons.

### A.1.1 Arachnoidal cells

**Arachnoidal cells** are located in the arachnoid villi in the meninges. Meningiomas arise from arachnoidal cells.

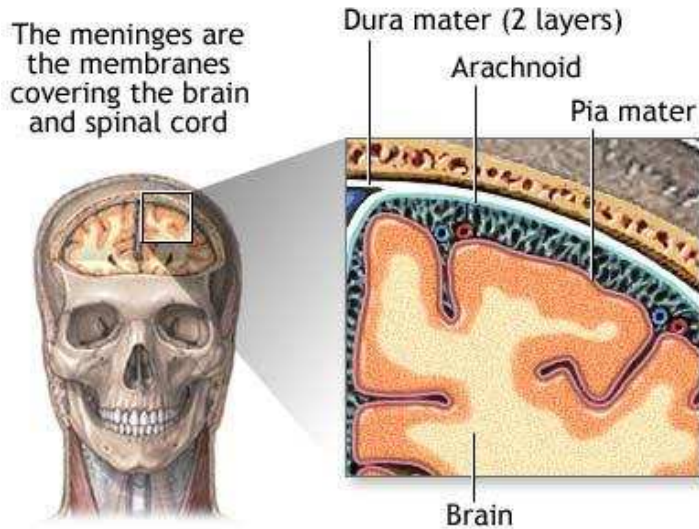


Figure A.6: Diagram of the meninges.

From <http://www.nlm.nih.gov/medlineplus/ency/imagepages/19080.htm>

## A.2 Tissues and regions

### A.2.1 Meninges

The **meninges** is the system of membranes which envelops the CNS. The meninges consist of three layers: the dura mater, the arachnoid mater, and the pia mater (see Figure A.6). The primary function of the meninges and of the CSF is to protect the CNS.

#### **Dura mater**

The dura mater is a thick, durable membrane, closest to the skull. It contains larger blood vessels which split into the capillaries in the pia mater. It is composed of dense fibrous tissue, and its inner surface is covered by flattened cells

like those present on the surfaces of the pia mater and arachnoid. The dura mater is a sac which envelops the arachnoid and has been modified to serve several functions. The dura mater surrounds and supports the large venous channels (dural sinuses) carrying blood from the brain toward the heart.

### **Arachnoid mater**

The middle element of the meninges is the arachnoid mater, so named because of its spider web-like appearance. It provides a cushioning effect for the CNS. The arachnoid mater exists as a thin, transparent membrane. It is composed of fibrous tissue and, like the pia mater, is covered by flat cells also thought to be impermeable to fluid. The arachnoid does not follow the convolutions of the surface of the brain and so looks like a loosely fitting sac. In the region of the brain, particularly, a large number of fine filaments called arachnoid trabeculae pass from the arachnoid through the subarachnoid space to blend with the tissue of the pia mater.

### **Pia mater**

The pia or pia mater is a very delicate membrane. It is the meningeal envelope which firmly adheres to the surface of the brain and spinal cord. As such it follows all the minor contours of the brain (gyri and sulci). It is a very thin membrane composed of fibrous tissue covered on its outer surface by a sheet of flat cells thought to be impermeable to fluid. The pia mater is pierced by blood vessels which travel to the brain and spinal cord, and its capillaries are responsible for nourishing the brain.

### **Spaces**

The subarachnoid space is the space which normally exists between the arachnoid and the pia mater, which is filled with CSF.

Normally, the dura mater is attached to the skull, or to the bones of the vertebral canal in the spinal cord. The arachnoid is not attached to the dura mater, while the pia mater is attached to the CNS tissue. When the dura mater and the arachnoid separate through injury or illness, the space between them is the subdural space.

## **A.2.2 White matter**

**White matter** is one of the two components of the CNS and consists mostly of myelinated axons. White matter is the tissue through which messages pass between different areas of grey matter within the nervous system (see Figure A.7). Using a computer network as an analogy, the grey matter can be thought of as the actual computers themselves, whereas the white matter represents the network cables connecting the computers together. The white matter is white because of the fatty substance (myelin) that surrounds the nerve fibers (axons). This myelin is found in almost all long nerve fibers, and acts as an electrical insulation. This is important because it allows the messages to pass quickly from place to place.

The brain in general (and especially a child's brain) can adapt to white-matter damage by finding alternative routes that bypass the damaged white-matter areas, and can therefore maintain good connections between the various areas of grey matter.

Unlike grey matter, which peaks in development in a person's twenties, the white matter continues to develop, and peaks in middle age.

## **A.2.3 Grey matter**

Grey matter is a major component of the CNS, consisting of neuronal cell bodies, neuropil (dendrites and both unmyelinated axons and myelinated axons), glial cells (astrocytes and oligodendrocytes) and capillaries. Grey matter contains neural cell bodies, in contrast to white matter, which does not and mostly contains myelinated axon tracts.

The function of grey matter is to route sensory or motor stimulus to interneurons of the CNS in order to create a response to the stimulus through chemical synapse activity. Grey matter structures (cortex, deep nuclei) process information originating in the sensory organs or in other grey matter regions. This information is conveyed via specialized nerve cell extensions (long axons), which form the bulk of the cerebral, cerebellar, and spinal white matter.

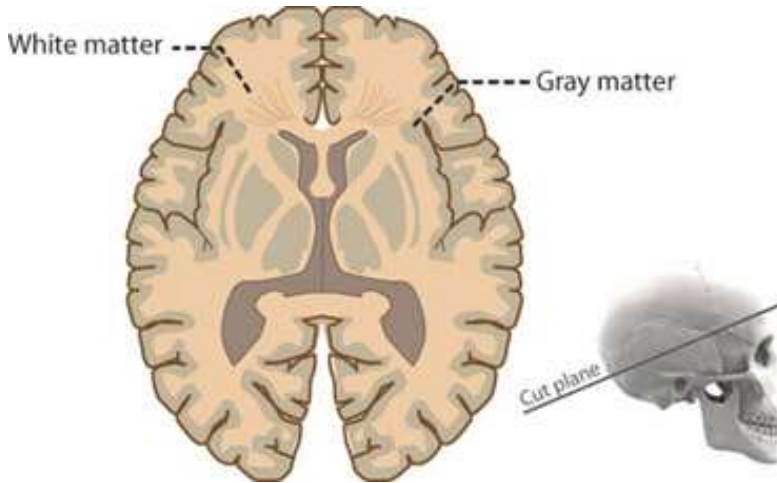


Figure A.7: White and grey matter.

From <http://science-naturalphenomena.blogspot.com/2009/04/white-matter.html>

## A.3 Physiological processes in brain cells

### A.3.1 Myelination

**Myelination** is the process of supplying myelin to the neurons. **Myelin** is a dielectric (electrically insulating) material that forms a layer, the myelin sheath, usually around only the axon of a neuron. It is essential for the proper functioning of the nervous system. Myelin is an outgrowth of a glial cell. Schwann cells supply the myelin for peripheral neurons (see Figure A.5), whereas oligodendrocytes myelinate the axons of the central nervous system (see Figure A.3). In healthy humans, the myelination process is nearly complete at the first year of life [3].

The main purpose of a **myelin sheath** is to increase the speed at which impulses propagate along the myelinated fiber. Along unmyelinated fibers, impulses move continuously as waves, but, in myelinated fibers, they hop or ‘propagate by saltation’. Thus, myelination helps prevent the electrical current from leaving the axon.

When a peripheral fiber is severed, the myelin sheath provides a track along which regrowth can occur. Unmyelinated fibers and myelinated axons of the mammalian central nervous system do not regenerate.

## **Demyelination and Dysmyelination**

**Demyelination** is the loss of the myelin sheath insulating the nerves, and is the hallmark of some neurodegenerative autoimmune diseases, including multiple sclerosis, acute disseminated encephalomyelitis or transverse myelitis, among others.

Research to repair damaged myelin sheaths is ongoing. Techniques include surgically implanting oligodendrocyte precursor cells in the CNS and inducing myelin repair with certain antibodies. While there have been some encouraging results in mice (via stem cell transplantation), it is still unknown whether this technique can be effective in replacing myelin loss in humans.

**Dysmyelination** is characterized by a defective structure and function of myelin sheaths; unlike demyelination, it does not produce lesions. Such defective sheaths often arise from genetic mutations affecting the biosynthesis and formation of myelin. Human diseases where dysmyelination has been implicated include leukodystrophies (Pelizaeus-Merzbacher disease, Canavan disease, phenylketonuria) and schizophrenia.

### **A.3.2 Phospholipids Membrane**

Cells are limited by a continuous thin barrier called membrane made of lipid molecules. This membrane, known as lipid bilayer or phospholipid membrane, surround the cells of almost all living organisms and many viruses. The lipid bilayer is the barrier that keeps ions, proteins and other molecules where they are needed and prevents them from diffusing into areas where they should not be.

Lipid bilayers are usually made of phospholipids, which have a hydrophilic head and two hydrophobic tails. When phospholipids are exposed to water, they arrange themselves into a two-layered sheet (a bilayer) with all of their tails pointing toward the center of the sheet (see Figure A.8).



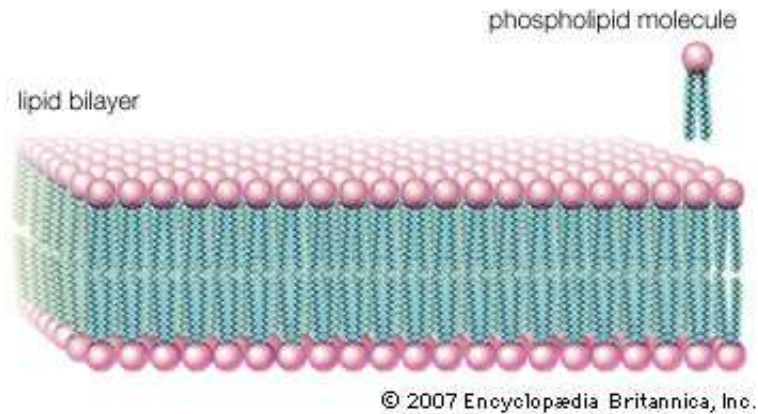


Figure A.8: Lipid bilayer.  
From Encyclopaedia Britannica Inc.  
<http://www.britannica.com/EBchecked/media/92240/Phospholipid-molecules>

### A.3.3 Apoptosis

**Apoptosis** is the process of programmed cell death (PCD) that may occur in multicellular organisms. Biochemical events lead to characteristic cell changes (morphology) and death.

In contrast to necrosis, which is a form of traumatic cell death that results from acute cellular injury, apoptosis, in general, confers advantages during an organism's life cycle. For example, the differentiation of fingers and toes in a developing human embryo occurs because cells between the fingers apoptose; the result is that the digits are separate. Between 50 and 70 billion cells die each day due to apoptosis in the average human adult. For an average child between the ages of 8 and 14, approximately 20 billion to 30 billion cells die a day.

Excessive apoptosis causes atrophy, such as in ischemic damage, whereas an insufficient amount results in uncontrolled cell proliferation, such as cancer.

## **Process**

The process of apoptosis is controlled by a diverse range of cell signals, which may originate either extracellularly (extrinsic inducers) or intracellularly (intrinsic inducers).

**Extracellular signals** may include toxins, hormones, growth factors, nitric oxide or cytokines, and therefore must either cross the plasma membrane or transduce to effect a response. These signals may positively (i.e., trigger) or negatively (i.e., repress, inhibit, or dampen) affect apoptosis.

A cell initiates **intracellular apoptotic signalling** in response to a stress, which may bring about cell suicide. The binding of nuclear receptors by glucocorticoids, heat, radiation, nutrient deprivation, viral infection, hypoxia and increased intracellular calcium concentration, for example, by damage to the membrane, can all trigger the release of intracellular apoptotic signals by a damaged cell.

Before the actual process of cell death is precipitated by enzymes, apoptotic signals must cause regulatory proteins to initiate the apoptosis pathway. This step allows apoptotic signals to cause cell death, or the process to be stopped, should the cell no longer need to die. Several proteins are involved, but two main methods of regulation have been identified: **targeting mitochondria functionality**, or **directly transducing the signal** via adaptor proteins to the apoptotic mechanisms. Another extrinsic pathway for initiation identified in several toxin studies is an increase in calcium concentration within a cell caused by drug activity, which also can cause apoptosis via a calcium binding protease calpain.

- **Execution**

A cell undergoing apoptosis shows a characteristic morphology (see Figure A.9):

1. Cell shrinkage and rounding are shown because of the breakdown of the proteinaceous cytoskeleton by caspases.
2. The cytoplasm appears dense, and the organelles appear tightly packed.
3. Chromatin undergoes condensation into compact patches against the nuclear envelope in a process known as pyknosis, a hallmark of apoptosis.
4. The nuclear envelope becomes discontinuous and the DNA inside it is fragmented in a process referred to as karyorrhexis. The nucleus breaks

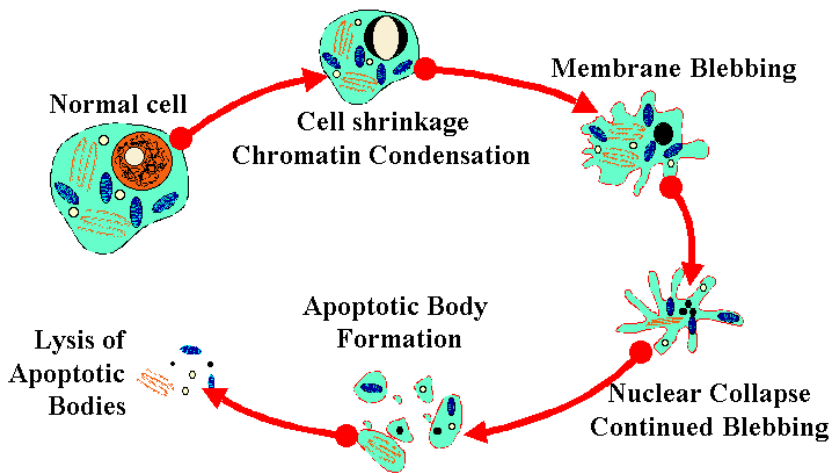


Figure A.9: Diagram of the process of apoptosis. From Kalkmakoff & Ward. University of Otago, Dunedin, New Zealand.  
<http://www.microbiologybytes.com/virology/kalkmakoff/baculo/baculohostinteract.html>

into several discrete chromatin bodies or nucleosomal units due to the degradation of DNA.

5. The cell membrane shows irregular buds known as blebs.

6. The cell breaks apart into several vesicles called apoptotic bodies, which are then phagocytosed.

Apoptosis progresses quickly and its products are quickly removed, making it difficult to detect or visualize.

- **Removal of dead cells**

The removal of dead cells by neighboring phagocytic cells has been termed efferocytosis. Dying cells that undergo the final stages of apoptosis display phagocytotic molecules, such as phosphatidylserine, on their cell surface. Upon recognition, the phagocyte reorganizes its cytoskeleton for engulfment of the cell. The removal of dying cells by phagocytes occurs in an orderly manner without eliciting an inflammatory response.

### **A.3.4 Necrosis**

**Necrosis** is the premature death of cells and living tissue. Necrosis is caused by factors external to the cell or tissue, such as infection, toxins, or trauma. This is in contrast to apoptosis, which is a naturally occurring cause of cellular death. While apoptosis often provides beneficial effects to the organism, necrosis is almost always detrimental and can be fatal.

Cells that die due to necrosis do not usually send the same chemical signals to the immune system that cells undergoing apoptosis do. This prevents nearby phagocytes from locating and engulfing the dead cells, leading to a build-up of dead tissue and cell debris at or near the site of the cell death. For this reason, it is often necessary to remove necrotic tissue surgically, a process known as debridement.

### **A.3.5 Glycolysis**

Glycolysis (from glycolysis, an older term for glucose + -lysis degradation) is the metabolic pathway that converts glucose,  $C_6H_{12}O_6$ , into pyruvate,  $CH_3COCOO^- + H^+$ . The free energy released in this process is used to form the high-energy

compounds, ATP (adenosine triphosphate) and NADH (reduced nicotinamide adenine dinucleotide).

### A.3.6 Growth pattern of brain tumours

There exists two growth pattern in brain tumour:

- **Infiltrative growth:** A process whereby inflammatory or other types of disease spread throughout an organ.
- **Circumscribed growth:** Bounded or limited growth; confined to a limited space.

In circumscribed tumours, like in typical meningiomas, some lymphomas and metastases, tumours have clear margins and might show almost no NAA resonance peak. MR spectra of tumours that tend to grow infiltratively, such as astrocytomas and some lymphomas, show a detectable NAA of various signal intensities but, even in these cases, the NAA peak is severely decreased or undetectable.

## Bibliography

- [1] J. Nolte, *The Human Brain: an introduction to its functional anatomy*. Mosby Elsevier, 2009.
- [2] P. Aubourg, “Axons need glial peroxisomes,” *Nature Genetics*, vol. 39, no. 8, pp. 936–938, 2008.
- [3] W. Grodd, I. Kraegelo-Mann, U. Klose, and R. Sauter, “Metabolic and destructive brain disorders in children: Findings with localized proton MR spectroscopy,” *Radiology*, vol. 181, pp. 173–181, 1991.

# List of Figures

1.1	Flowchart of the Dissertation. . . . .	6
2.1	Spectra of a low grade meningioma . . . . .	18
2.2	Spectra of an anaplastic astrocytoma . . . . .	20
2.3	Spectra of a glioblastoma . . . . .	21
2.4	Spectra of a metastasis . . . . .	26
2.5	PR solution based on the Machine Learning approach. . . . .	28
2.6	LDA and QDA decision boundaries . . . . .	29
2.7	Nearest-neighbor partitioning of a 2-dimensional space. . . . .	30
2.8	ROC curve when AUC is the same . . . . .	38
2.9	Evidence embodies Occam’s razor . . . . .	42
2.10	INTERPRET DSS Version 3.0 GUI . . . . .	50
2.11	SpectraClassifier tool GUI . . . . .	52
2.12	CURIAM DSS GUI . . . . .	53
3.1	Mean spectra of children and adult BT types . . . . .	79
3.2	Aggressive and non-aggressive tumours mean spectra . . . . .	80
3.3	Diagram of the ‘filtered classifier’ attending to age . . . . .	84
4.1	Short-TE mean spectra of pediatric tumours. . . . .	99

4.2 Long-TE mean spectra of pediatric tumours. . . . .	100
4.3 LDA latent space from Short-TE, Long-TE and combination. . . . .	105
4.4 LDA latent space with all tumour types projected. . . . .	106
5.1 Corpora of brain tumours to apply the audit method. . . . .	123
5.2 Example of scenarios for the audit method. . . . .	124
5.3 Average likelihoods. . . . .	131
5.4 Evolution of the posterior odds. . . . .	132
6.1 Mean spectra of aggressive and non-aggressive tumours . . . . .	151
6.2 Histograms of age in BT patients . . . . .	153
6.3 Histograms of gender in BT patients . . . . .	154
6.4 Histograms of brain tumour location in patients . . . . .	155
7.1 Example of a set of classifier actions. . . . .	175
7.2 A classifier is represented by three XML specification files. . . . .	177
7.3 Diagram of classification request. . . . .	177
7.4 CURIAM BT screenshots with children classifiers. . . . .	180
7.5 Diagram of diagnosis feedback in CURIAM BT. . . . .	182
7.6 Selection of classifier candidates in CURIAM BT. . . . .	184
7.7 View of Multiple classifiers in CURIAM BT. . . . .	190
A.1 Structure of a Neuron . . . . .	202
A.2 Astrocyte . . . . .	205
A.3 Oligodendrocyte . . . . .	206
A.4 Ependymal cells . . . . .	207
A.5 Schwann cell . . . . .	208
A.6 Diagram of the Meninges . . . . .	209
A.7 White and grey matter . . . . .	212

LIST OF FIGURES

---

A.8 Lipid bilayer . . . . . 214  
A.9 Cell apoptosis . . . . . 216



# List of Tables

2.1	Most common brain tumour and CNS types by age . . . . .	13
2.2	List of resonance peaks in brain $^1\text{H}$ MR spectroscopy . . . . .	16
2.3	Metabolites measured with PI and pediatric PI . . . . .	34
2.4	Basis set of TARQUIN. . . . .	36
3.1	Number of cases of children and adults . . . . .	73
3.2	Metabolite quantitation in children and adult brain tumours . . .	81
3.3	Performance of ‘aggressive VS non-aggressive’ classifiers . . . . .	82
3.4	Performance of age-filtered ‘aggressive VS non-aggressive’ classifier	83
4.1	Number of cases. . . . .	95
4.2	Estimated metabolite concentrations in pediatric tumours. . . . .	101
4.3	BAR of the pediatric classifiers. . . . .	103
4.4	BAR of the pediatric classifiers. . . . .	103
4.5	BAR of the pediatric classifiers. . . . .	104
5.1	<i>CT</i> based on the WHO classification of brain tumours. . . . .	122
5.2	Number of samples for audit model evaluation. . . . .	129
5.3	Complexity of the models. . . . .	129

6.1	Datasets for prior assessment. . . . .	152
6.2	Likelihood of classifiers specialized by age. . . . .	156
6.3	Likelihood of specialized classifiers with prior assessment. . . . .	157
6.4	Likelihood of non-specialized models. . . . .	158
6.5	Likelihood of classifiers with young adult dataset. . . . .	158
6.6	Example of the Simpson’s Paradox. . . . .	160
7.1	‘aggressive VS non-aggressive’ classifiers in CURIAM BT . . . . .	178
7.2	Pediatric classifiers incorporated into CURIAM BT. . . . .	179