



UNIVERSIDAD POLITÉCNICA DE VALENCIA
DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN
INTELIGENCIA ARTIFICIAL, RECONOCIMIENTO DE FORMAS E
IMAGEN DIGITAL

Pattern Recognition Approaches for Biomedical Data in Computer-Assisted Cancer Research

DOCTORAL THESIS

Presented by **Juan Miguel García-Gómez**

Supervised by Dr. Montserrat Robles Viejo
and Dr. Alfons Juan i Císcar

Valencia - Spain
April 2, 2009

Agradecimientos/Agraïments/Acknowledgement

Mi agradecimiento a la Profesora Montserrat Robles y al Profesor Alfons Juan, supervisores de esta Tesis. Las oportunidades y retos que Montse ha producido cada día en el grupo de Informática Biomédica han sido la mayor inspiración para desarrollar este trabajo. La rigurosidad y metodología de Alfons han llevado esta Tesis a su conclusión. El interés de ambos por la ciencia y la tecnología hace avanzar al Reconocimiento de Formas aplicado a la Medicina y la Biología.

A mis compañeros diarios de investigación Miguel Esparza, Javier Vicente, Salvador Tortajada, Alfredo Navarro, Elies Fuster y Carlos Sáez; porque ellos son el futuro; y a César Vidal, porque sin él, esto no sería.

El meu agraïment al Prof. Carlés Arús y a la Dra. Margarida Julià, pels seus ensenyaments en el camp de la Espectroscopia de Resonància Magnètica aplicada als Tumors Cerebrals i mostrar-me com tenir la constància i motivació per fer ciència aplicada a la medicina. Tambien mi agradecimiento al Dr. Luis Martí-Bonmatí que nos ha empujado con tesón al desarrollo de los Sistemas de Ayuda a la Decisión en Medicina.

Agradezco el gran espíritu investigador respirado en el grupo IBIME-Itaca, y que lo crean con su esfuerzo Mariam de la Iglesia, Juanjo Lull, José Alberto Maldonado, Carlos Angulo, José Carbonell, Gracián García, Stefan Goetz, Ana Conesa, David Moner, José Vicente Manjón, Ernesto Reig, Diego Boscá y Luis Marco. Agradezco la ayuda, inestimable en mis comienzos como investigador, del grupo PHRLT-ITI, dirigido por el Prof. Francisco Casacuberta. También fue inestimable la ayuda en los comienzos de mi etapa investigadora del grupo BET, y su director Prof. Antonio Mocholí y del grupo TSB, dirigido por el Prof. Vicente Traver y el Dr. Sergio Guillén. En aquella etapa disfruté al trabajar junto a Carlos Fernández y José Luis Bayo, con los que he compartido gratos momentos de estudio, trabajo y amistad. A su vez, agradezco el apoyo y las oportunidades que ofrecen constantemente los miembros del Departamento de Física Aplicada, y en especial al equipo directivo de la Prof. M^a Carmen Millán y a las unidades docentes de Informática y de la Escuela de Telecomunicaciones. Agradezco el soporte económico e institucional que me ha brindado la Universidad Politécnica de Valencia a través de sus diferentes vicerrectorados.

It was a great opportunity to be at the Katholieke Universiteit Leuven with Prof. Sabine Van Huffel during my PhD mobility period. At KULeuven I have benefited from a close contact with Jan Luts, Diana Sima, PhD and Jean-Baptiste Poulet.

This PhD Thesis was partially developed in the context of the European Commission projects eTUMOUR (contract no. FP6-2002-LIFESCIHEALTH 503094) and HEALTH-AGENTS (contract no. FP6-2005-IST 027213). I would like to thank to the consortia of these projects from I had the opportunity to learn. I feel lucky to have met many people during the development of the eTUMOUR and HEALTHAGENTS projects: Prof. Bernardo Celda, Prof. Lutgarde Buydens, Prof. Geert Postma, Prof. Paul Lewis, Prof. Srinandan Dasmahapatra, Magí Lluch, Mariola Maier, Dr. Daniel Monleón, Xavier Castells, Giulia Valsecchi, Dr. Ana Paula Candiota, Dr. Andrew Peet, Guillem Mercadal, and Patrick Krooshof.

I would like to dedicate the PhD Thesis to the patients that contributed to the science including their hopes on the biomedical data used in this study. I thank the following groups and hospitals for acquiring the biomedical data in the soft tissue tumours and brain

tumours tasks: Carles Majós (IDI-Bellvitge), Jaume Capellades (IDI-Badalona), John Griffiths and Franklyn Howe (SGUL), Arend Heerschap (RU), Witold Gajewicz (MUL), Jorge Calvar (FLENI), Antoni Capdevila (H. de Sant Joan de Déu), Àngel Moreno (CDP), Jesús Pujol (CRC), Joaquin Galant (Hospital Universitario San Juan de Alicante), Hospital Universitario Dr Peset de Valencia, Hospital Cruces de Baracaldo, Hospital Juan Canalejo de La Coruña, Hospital Universitaire de Toulouse and the INTERPRET consortium.

El más cariñoso recuerdo a mis padres Victoria y Juan, que han sufrido y disfrutado con cada etapa de este trabajo.

Y la más querida dedicatoria a M^a Jesús, que ofrece el amanecer diario a mi vida.

Abstract

When symptoms of cancer are observed, especially in conjunction with genetic predisposition and/or lifestyle risk factors, a non-invasive exploration of the patient's body is required to confirm or refute the presence of a tumor. If there is evidence of a tumor there must be a formal diagnosis of the stage to which the cancer has progressed.

The systemic analysis of the patients' biomedical data, which come from different biological levels, offer greater information for the medical decision process. New biomedical technologies can allow the interpretation of the origin of the illnesses, moving to an evidence-based medicine paradigm.

The recent increase in the complexity of the techniques for acquiring biomedical data as well as their innovative features, make it difficult for them to be incorporated in the clinician's practice. Therefore, it would be convenient to develop automatic data processing and predictive models to objectively assist the experts to interpret the data in the decision workflow of diagnoses, prognoses, and treatments.

This Thesis focus its attention on the computer-assisted support of cancer diagnosis for clinical environments. The aim of the study is to produce results with high accuracy in classification, transparency in relation to the clinical knowledge and capacity to generalize their performance to new samples subsequently obtained in different clinical centers. The technical aspects covered in this Thesis includes the processing, modeling, feature extraction and combination of biomedical data; the inference and evaluation of predictive models for biomedical problems; and the integration of the models into decision support systems for the clinical environments. In order to focus our studies, two medical problems are tackled: Soft Tissue Tumor (STT) diagnosis and, Brain Tumor (BT) diagnosis.

In the STT problem, high efficacy in the discrimination of the benign/malignant character of the tumors was achieved by Pattern Recognition (PR)-based classifiers on Magnetic Resonance (MR) Image findings. These classifiers can help radiologists the confirmation of the diagnoses of new patients, allow the study of suspicious cases, and aid in the education of new radiologists' expertise in tumors of this kind. A new clinical decision-support system (CDSS) for STT has been designed and implemented, which is based on classifiers learned from multicenter datasets. The generic distributed architecture designed for the STT problem has been the basis for posterior developments in the field, such as those adopted by the distributed CDSS of the HEALTHAGENTS project.

This Thesis provides several contributions to the BT medical problem. A new approach that combines MR Spectra of different echo times has been proposed. Significant differences in performance were found when Short TE, Long TE or both spectra combined were used as input. In addition, a probabilistic mixture model and the E(xpectation)-M(aximisation)-based estimation for binned and truncated data with univariate mixture

densities of means relative to a global shift have been proposed for Magnetic Resonance Spectroscopy (MRS) data characterization. The new version of the mixture model keeps the biological information in the model and properly fits the MRS. The discrimination of Brain Tumors based on the parametric space of the probabilistic mixture model is possible with high accuracies, and the combination of MR Spectra for classification can be performed by means of their parametric spaces of the models. With respect to the evaluation of the PR-based models, it has been demonstrated that the prediction of in-vivo MRS cases that are from a later date, from different hospitals, and with different instrumentation, but which are obtained under the same acquisition parameters may be possible by models inferred by multicenter datasets. Our results consolidate, with experiments on subsequently acquired datasets, the conclusions of previous studies on automatic brain tumor classification using MRS. The classification results obtained by the PR-based models have been useful for the validation and review of cases with abnormal profiles in the multicenter datasets of the eTUMOUR project. With respect to the classification of BT by ex-vivo data, a proof of principle was carried out using gene expression profiles to discriminate between glioblastoma and meningioma biopsies. The obtained gene-signatures are in accordance with the expected biological and pathological differences between these two kinds of tumors.

The conclusions and developments of this PhD thesis apply directly to the results of 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 to the results of the HEALTHAGENTS project (*Agent-based Distributed Decision Support System for Brain Tumour Diagnosis and Prognosis, 2006-2008*), which are both European Union projects of the 6th Framework Programme. Hence, based on scientific contributions studied in the Thesis for BT classification with in-vivo data, two practical solutions have been developed to integrate the PR engines in the clinical routine workflow. The first one is a generic interface between the CDSS and the classification module developed for the eTUMOUR project which allows an on-line update of the available classifiers. The second one is an agent-based distributed Decision-Support System, which is the basis for the European HEALTHAGENTS network to assist in brain tumour management.

Resumen

Una vez descubiertos los síntomas que sugieren la aparición de cáncer, especialmente cuando existen factores de riesgo genéticos o ambientales, está indicado llevar a cabo la exploración no-invasiva del cuerpo del paciente para confirmar o rechazar la presencia de un tumor, y si es positiva, indicar el diagnóstico y estado al cual éste ha progresado.

El análisis sistémico de datos biomédicos procedentes de diferentes niveles biológicos abre amplias expectativas en el proceso de toma de decisiones médicas. Las nuevas tecnologías biomédicas permiten la interpretación del origen de las afecciones que sufren los pacientes, trasladando el paradigma de decisión hacia la medicina basada en la evidencia.

Este incremento de técnicas innovadoras de adquisición de datos biomédicos, hacen compleja su accesibilidad al uso clínico. Es comprensible el interés en el desarrollo de métodos de procesamiento y modelado automático que asistan objetivamente a los expertos a interpretar los datos biomédicos durante el flujo de decisiones adoptadas para los diagnósticos, pronósticos o tratamientos de nuevos pacientes.

Esta Tesis centra su atención en la ayuda al diagnóstico del cáncer asistida por ordenador. El objetivo de nuestro estudio es obtener unos resultados de alto acierto en clasificación, que ofrezcan transparencia en su interpretación mediante conocimiento médico y capacidad de generalización cuando se aplican a pacientes procedentes de múltiples centros estudiados con posterioridad. Los aspectos técnicos cubiertos en esta Tesis incluyen el procesamiento, modelado, extracción de características, y combinación de datos biomédicos; así como la inferencia y evaluación de modelos predictivos de dichos datos y la integración de los modelos predictivos en sistemas de ayuda a la decisión para entornos clínicos. Concretamente, estos puntos se abordan para dos problemas médicos: el diagnóstico de Tumores de Partes Blandas (TPB) y, especialmente, el diagnóstico de Tumores Cerebrales (TC).

En los desarrollos realizados para el problema de TPB con hallazgos de imagen se alcanzó una alta eficacia en la clasificación basada en Reconocimiento de Formas de tumores según su carácter benigno o maligno. Los clasificadores desarrollados pueden ser útiles para los servicios de radiología a la hora de confirmar el diagnóstico de nuevos pacientes, estudiar casos sospechosos, y asistir a la preparación de radiólogos noveles en este tipo de tumores. Un sistema de ayuda a la decisión especializado para el problema de TPB fue diseñado e implementado a partir de los clasificadores aprendidos a partir de una base de datos multicéntrica.

Las contribuciones de esta Tesis al estudio de Tumores Cerebrales incluyen el análisis de señales biomédicas in-vivo y ex-vivo del paciente. Ha sido propuesta una nueva aproximación para la combinación de Espectros de Resonancia Magnética (ERM) adquiridos para un mismo paciente con diferentes tiempos de eco (TE corto y TE largo) ha sido

propuesta. En nuestros estudios, se encontraron diferencias significativas al usar el TE corto, el TE largo o la combinación de ambas adquisiciones como entrada de los métodos predictivos. Paralelamente, ha sido propuesto un modelo probabilístico de mixturas para la caracterización de ERM de TE corto, así como su estimación basada en el algoritmo EM para datos pareados y truncados. Esta nueva aproximación utiliza la información conocida de resonancia de las sustancias químicas mediante un modelo de distribuciones gaussianas univariantes de medias conocidas respecto a un desplazamiento global. La discriminación de TC basada en el espacio paramétrico del modelo probabilístico de mixturas obtuvo una alta eficacia, y la combinación de los espacios paramétricos de los ERM de diferentes tiempos de eco es una alternativa para sacar provecho de ambas adquisiciones en la ayuda al diagnóstico. Siguiendo con los objetivos propuestos en la Tesis, se ha realizado una evaluación de los modelos de reconocimiento de formas con el fin de demostrar que clasificadores entrenados con pacientes explorados con diferentes máquinas de RM situadas en diferentes centros médicos siguiendo un mismo protocolo de adquisición, predicen con una buena eficacia datos adquiridos con posterioridad por parte de múltiples centros parcialmente diferentes. Estos resultados consolidan las conclusiones de un gran número de estudios previamente presentados en la literatura, y han resultado de utilidad para la validación y revisión de casos con patrones anormales de la base de datos del proyecto europeo eTUMOUR. Con respecto a la clasificación de TC mediante datos ex-vivo, se ha realizado una prueba de concepto mediante expresión genética para discriminar biopsias de tumores de los tipos glioblastoma y meningioma. Los resultados obtenidos están en consonancia con la clara diferenciación que se conoce por la biología y la patología de las dos estirpes tumorales.

Las conclusiones y desarrollos de esta Tesis Doctoral se han aplicado directamente a los resultados del proyecto 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 a los resultados del proyecto HEALTHAGENTS (*Agent-based Distributed Decision Support System for Brain Tumour Diagnosis and Prognosis*, 2006-2008) del 6º Programa Marco de la UE. A partir de las conclusiones científicas en TC obtenidas en este estudio se han desarrollado dos sistemas de ayuda a la decisión para rutinas clínicas. Se ha desarrollado e integrado en eTUMOUR un interfaz genérico para la incorporación de nuevos modelos predictivos al sistema de ayuda a la decisión y, en el proyecto HEALTHAGENTS, se ha desarrollado una red de ayuda a la decisión de TC mediante sistemas multiagente.

Resum

Una vegada descoberts els símptomes que sugereixen risc de càncer, especialment quan existeixen factor de risc genètics o ambientals, l'exploració no-invasiva del cos del pacient està indicada per a confirmar la presència d'un tumor, i si es positiva, indicar el diagnòstic i estat pel qual aquest ha progressat.

L'anàlisi sistèmic de les dades biomèdiques dels pacients des de diferents nivells biològics obre amplies expectatives en el procés de la toma de decisions mèdiques. Les noves tecnologies biomèdiques permeteixen la interpretació de l'origen de les afeccions que sufriren els pacients, traslladant el paradigma de decisió cap a la medicina basada en la evidència.

Aquest increment de tècniques innovadores d'adquisició de dades biomèdiques, fa complexa la accessibilitat a l'ús clínic. És comprensible l'interès en el desenvolupament de mètodes de processament i modelatge automàtic que assistisquen objectivament als experts a interpretar les dades biomèdiques durant el flux de decisions adoptades per als diagnòstics, pronòstics o tractaments de nous pacients.

Aquesta Tesi centra la seua atenció en l'ajuda assistida per ordinador al diagnòstic del càncer. L'objectiu del nostre estudi és obtenir uns resultats d'alt encert en classificació, que oferisquen transparència en la seua interpretació mitjançant coneixement mèdic i capacitat de generalització en la seua aplicació en pacient multicèntrics estudiats amb posterioritat. Els aspectes tècnics coberts en aquesta Tesi inclouen el processament, modelatge, extracció de característiques, i combinació de dades biomèdiques; així com la inferència i avaluació de models predictius d'aquestes dades i la integració dels models predictius en sistemes d'ajuda a la decisió per a entorns clínics. Concretament, aquests punts s'aborden per a dos problemes mèdics: el diagnòstic de Tumors de Parts Blanes (TPB), i, especialment, el diagnòstic de Tumors Cerebrals (TC).

En els desenvolupaments realitzats per al problema de TPB es va arribar a una alta eficàcia en la classificació basada en Reconeixement de Formes de tumors segons el seu caràcter benigne o maligne. Els classificadors desenvolupats poden ser útils per als serveis de radiologia a l'hora de confirmar el diagnòstic de nous pacients, estudiar casos sospitosos, i assistir a la preparació de radiòlegs novells en aquest tipus de tumors. Un sistema d'ajuda a la decisió especialitzat per al problema de TPB va ser dissenyat i implementat sobre la base dels classificadors apresos a partir d'una base de dades multicèntrica.

Les contribucions d'aquesta Tesi a l'estudi de Tumors Cerebrals inclouen l'anàlisi de senyals biomèdics in-viu i ex-viu del pacient. Una nova aproximació per a la combinació d'Espectres de Ressonància Magnètica (ERM) adquirits per a un mateix pacient amb diferents temps d'eco (ET curt i ET llarg) ha estat proposada. En els nostres estudis, es van trobar diferències significatives a l'usar l'ET curt, l'ET llarg o la combinació d'ambdues

adquisicions com entrada dels mètodes predictius. Paral·lelament, un model probabilístic de mixtures ha estat proposat per a la caracterització de ERM de ET curt, així com la seua estimació basada en l'algorisme EM per a dades aparellats i truncats. Aquesta nova aproximació utilitza la informació coneguda de ressonància de les substàncies químiques mitjançant un model de distribucions gaussianes univariades de mitjanes conegudes respecte a un desplaçament global. La discriminació de TC basada en l'espai paramètric del model probabilístic de mixtures va obtenir una alta eficàcia, i la combinació dels espais paramètrics dels ERM de diferent temps d'eco és una alternativa per a traure profit d'ambdues adquisicions en l'ajuda al diagnòstic. Seguint amb els objectius proposats en la Tesi, s'ha realitzat una avaluació dels models de reconeixement de formes amb la finalitat de demostrar que classificadors entrenats amb pacients explorats amb diferents màquines de RM situades en diferents centres mèdics seguint un mateix protocol d'adquisició, prediuen amb una bona eficàcia dades adquirides amb posterioritat per part de múltiples centres parcialment diferents. Aquests resultats consoliden les conclusions d'un gran nombre d'estudis prèviament presentats en la literatura, i han resultat d'utilitat per a la validació i revisió de casos amb patrons anormals de la base de dades del projecte europeu eTUMOUR. Pel que fa a la classificació de TC mitjançant dades ex-viu, s'ha realitzat una prova de concepte mitjançant expressió genètica per a discriminar biòpsies de tumors dels tipus glioblastoma i meningioma. Els resultats obtinguts estan d'acord amb la clara diferenciació que es coneix sobre la base de la biologia i la patologia de les dues estirps tumorals.

Les conclusions i desenvolupaments d'aquesta Tesi Doctoral s'han aplicat directament als resultats del projecte 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 als resultats del projecte HEALTHAGENTS (*Agent-based Distributed Decision Support System for Brain Tumour Diagnosis and Prognosis, 2006-2008*) del 6^e Programa Marc de la UE. Sobre la base de les conclusions científiques en TC obtinguts en aquest estudi s'han desenvolupat dos sistemes d'ajuda a la decisió per a rutines clíniques. Una interfície genèric per a la incorporació de nous models predictius al sistema d'ajuda a la decisió ha estat desenvolupat i integrat en eTUMOUR i, en el projecte HEALTHAGENTS, s'ha desenvolupat una xarxa d'ajuda a la decisió de TC mitjançant sistemes multiagent.

Glossary

Mathematical Notation

a	scalar a
\mathbf{v}	vector \mathbf{v}
\mathbf{M}	matrix \mathbf{M}
\mathbf{M}^T	transpose of matrix \mathbf{M}
\mathbb{R}	the set of real numbers
$\text{sign}(x)$	sign of x
$\log(x)$	natural logarithm of x
$\exp(x)$	exponential of x , e^x
\mathbf{B}	Magnetic field \mathbf{B}
\mathbf{M}	Magnetic moment \mathbf{M}

Fixed symbols

\hbar	reduced Planck's constant; $6.626e^{-34/2\pi} J_s$
T_1	spin-lattice relaxation time constant
T_2	spin-spin relaxation time constant
^1H	Hydrogen-1
γ	Gyromagnetic constant

Acronyms and Abbreviations

2D	Bi-dimensional
A1	Astrocytoma grade I
A2	Astrocytoma grade II
ACC	Accuracy
AGG	Aggressive tumor: GBM and MET
Ala	Alanine
ANN	Artificial Neural Networks
API	Application Programming Interface
AS	Ability to Separate

a.u.	arbitrary units
AUC	Area Under the Curve
BER	Balanced Error Rate
BDK	Bi-directional Kohonen Networks
BT	Brain Tumor
CADS	Computer Aided Diagnosis System
cDNA	complementary DNA
CDSS	Clinical Decision-Support System
CDSSs	Clinical Decision-Support Systems
CDVC	Clinical Data Validation Committee
Cho	Choline
CNS	Central Nervous System
CPDF	Conditional Probability Density Function
CQCD	Committee for Quality Control of Data
Cr	Creatine
CS	Chemical Shift
CT	Computed Tomography
CV	Cross Validation
dDSS	distributed Decision-Support System
DICOM	Digital Imaging and Communication in Medicine
dLDA	Linear Discriminant Analysis with diagonal covariance matrix
DNA	Deoxyribonucleic acid
dQDA	Quadratic Discriminant Analysis with diagonal covariance matrix
DSS	Decision-Support System
DSSs	Decision-Support Systems
DT	Decision Trees
EbSS	Evidence-based Search Service
ECC	Eddy Current Correction
e.g.	exempli gratia (for example)
EM	Expectation-Maximization

EMBTD	EM for Binned and Truncated Data
EMBTDr	EM for Binned and Truncated Data with related means
ERR	Error rate
eTDB	eTUMOUR database
eTUMOUR	the eTUMOUR EC project
etxxxx	Acronym for a generic patient in the eTUMOUR database
F	Friedman's nonparametric two-way analysis of variance test
FE	Feature Extraction
FID	Free Induction Decay
FLDA	Fisher's LDA
FFT	Fast Fourier Transform
FIPA	Foundation of Intelligent Physical Agents
FWHM	Full Width at Half Maximum
GBM	Glioblastoma
GE	General Electric
GlioMET	Glial (LGG + GBM) + MET
Gly	Glycine
Glx	Glutamate/Glutamine
GMOR	Geometric Mean of Recalls
GUI	Graphical User Interface
HEALTHAGENTS	the HEALTHAGENTS EC project
HAL	HEALTHAGENTS Language
HGG	High grade glial
HIV	Human Immunodeficiency Virus
HLSVD	Hankel-Lanczos Singular Value Decomposition
HR-MAS	High-Resolution Magic Angle Spinning
HSVD	Hankel Singular Value Decomposition
IBIME	Biomedical Informatics group
ICA	Independent Component Analysis
ICT	Information and Communication Technology

i.e	id est (that is)
i.i.d.	independent and identically distributed
INTERPRET	the INTERPRET project
IOTA	International Ovarian Tumor Analysis
IT	Independent Test
ITACA	Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas
Ixxxx	Acronym for a generic patient in the INTERPRET database
JADE	Cardoso's implementation for ICA
jMRUI	Java Magnetic Resonance User Interface
kRSTT	k-Random Sampling Train-Test
KNN	k-Nearest Neighbors
KW	Kruskal-Wallis nonparametric one-way analysis of variance
L2-norm	Euclidean distance
L1	Lipid resonance at 1.29 <i>ppm</i>
L2	Lipid resonance at 0.92 <i>ppm</i>
Lac	Lactate
LCC	Lightweight Coordination Calculus
LDA	Linear Discriminant Analysis
LOO	Leave-One-Out Cross-Validation
LGG	Low-Grade Glial
LSSVM	Least-Squares Support Vector Machines
MCT	Tukey's honestly significant difference criterion for multiple comparison
MEN	Low-grade meningiomas
mMEN	Meningothelial Meningioma
MET	Metastases
ml	myo-Inositol
ML	Mobile lipids
MLP	Multilayer Perceptron
MM	Macromolecules
MiM	Mismatch

MN	Multinomial parametric classifier
MR	(Nuclear) Magnetic Resonance
MRI	Magnetic Resonance Imaging
mRNA	messenger Ribonucleic acid
MRS	Magnetic Resonance Spectroscopy
MRSI	Magnetic Resonance Spectroscopic Imaging
MSE	Mean Squared Error
MV	Multi-voxel
NAA	N-Acetyl Aspartate
NAc	N-Acetyl groups
NMR	Nuclear Magnetic Resonance
NPV	Negative predictive value
OA	Oligoastrocytomas
OD	Oligodendrogliomas
OWL	Web Ontology Language
PR	Pattern Recognition
PC	principal component or loading
PCs	principal components or loadings
PCA	Principal Component Analysis
PCA-KNN	K-nearest neighbours and local feature reduced by PCA
PI	Peak integration
PM	Perfect Match
PPV	Positive predictive value
PRESS	Point-Resolved Spectroscopic Sequence
PPM	Peak height of typical resonances
Q1	Low Grade Meningioma vs. Glioblastoma+Metastasis vs. Low Grade Glial
QC	Quality Control
QDA	Quadratic Discriminant Analysis
QH1	Hierarchical question 1: Meningioma vs. (Glioma and Metastasis)
QH2	Hierarchical question 2: Low grade glial vs. (Glioblastoma and Metastasis)

RDO	Radiological diagnostic orientation
RBF	Radial Basis Function
RDF	Resource Description Framework
ReliefF	ReliefF algorithm for Recursive Elimination of Features
ROC	Receiver operating characteristic
RF	Radio frequency
SNR	Signal-to-Noise Ratio
SNV	Standard Normal Variate
STEAM	Stimulated Echo Acquisition Mode sequence
SOC	Sistema de Orientación Clínica
STT	Soft Tissue Tumor
SV	Single voxel
SVM	Support Vector Machines
SW	Stepwise algorithm for feature selection in classification
Ta2	$T2 * \text{constant}$
TAM	Technology Acceptance Model
Tau	Taurine
TE	Echo Time
TR	Recycling Time
ULN	Unit Length Normalization
VC	Validation Committee
WAV	Wavelet transform
WHO	World Health Organization
XML	eXtensible Markup Language
YP	Yellow Pages

Contents

Abstract	iii
Glossary	ix
Contents	xv
Preface	xvii
1 Introduction	1
1.1 Pattern Recognition	1
1.2 Cancer Research	8
1.3 Decision-Support Systems in Cancer Research	13
1.4 Goals, Contributions, Projects and Partners	15
Bibliography	19
2 Benign/Malignant Classifier of STT Using MRI	25
2.1 Introduction	25
2.2 Materials and methods	26
2.3 Results	32
2.4 Discussion	33
Bibliography	37
3 The combination of two TE in BT by MRS	41
3.1 Introduction	41
3.2 Experimental	43
3.3 Results	47
3.4 Discussion	50
3.5 Conclusions	56
Bibliography	61
4 MRS mixture modelling by binned and truncated data	65
4.1 Introduction	65
4.2 Probabilistic model	66
4.3 EM-based maximum likelihood estimation	68
4.4 Experimental Results	70

4.5	Conclusions and further work	75
	Bibliography	77
5	Multicentric evaluation of BT by MRS	79
5.1	Introduction	79
5.2	Materials and Methods	81
5.3	Results and discussion	84
5.4	Evaluation of the BT classifiers based on EMBTD and EMBTDr	98
5.5	Conclusions	99
	Bibliography	101
6	BT classification using cDNA microarrays	107
6.1	Introduction	108
6.2	Materials and Methods	109
6.3	Results	112
6.4	Discussion	116
6.5	Conclusion	118
	Bibliography	121
7	Clinical Decision Support Systems for STT and BT	129
7.1	Introduction	129
7.2	Distributed Decision Support System for STT Diagnosis	131
7.3	eTUMOUR classification module	139
7.4	HEALTHAGENTS: Distributed multi-agent DSS	141
7.5	Conclusion	145
	Bibliography	147
8	Concluding remarks and recommendations	155
8.1	Conclusions	155
8.2	Recommendations	156
	Bibliography	159
A	EM estimation for EMBTDr	161
A.1	MLE from BTD with Univariable gaussians	161
A.2	MLE from BTD with Relative Gaussians	171
	Bibliography	173
B	Expression differences between GBM and mMEN	175
	List of Figures	181
	List of Tables	183

Preface

This Dissertation includes some applications of the Pattern Recognition (PR) discipline for the analysis of biomedical data. The main focus of the Thesis is on the development of computer-assisted support of cancer diagnosis for clinical environments based on the PR discipline. The aim of the studies described in the Dissertation is to produce results with high accuracy in classification, interpretability by means of clinical knowledge and capacity to generalize the performance to new samples which are subsequently obtained in different clinical centers. The technical aspects covered in the Thesis include the processing, feature extraction and modeling of biomedical data; the inference and evaluation of predictive models; and the integration of the models into decision support systems for the clinical environments. Specifically, we study the application of PR techniques to Soft Tissue Tumor (STT) diagnosis and the Brain Tumor (BT) diagnosis. Figure 1 illustrates the chapters of the Dissertation and establishes the relations between them.

Chapter 1 introduces several concepts from the fields involved in this Dissertation, e.g. Soft Tissue Tumors and Brain Tumors, Magnetic Resonance, Gene-expression biotechnology, Pattern Recognition and Decision Support Systems. This chapter also summarizes the goals of the Thesis and outlines by the research contributions.

Chapter 2 introduces the contribution of the Thesis to the classification of STT using Magnetic Resonance Imaging (MRI) findings from a multicenter European database of patients. In the experiments reported in this chapter, several pattern recognition techniques are studied in the classification of benign and malignant tumors. The results obtained with these classifiers are directly applied to the predictive engines of the Clinical Decision Support System for tumors of this kind (see Chapter 7 for details).

In the automatic classification of BT by means of Magnetic Resonance Spectroscopy (MRS), we investigated three major concerns. Chapter 3 studies the influence of combining two echo times in automatic brain tumor classification by MRS. In this work, we were interested in analyzing whether the combination of single voxel ^1H MRS at two different Echo Time (TE), Short TE (20-32 ms) and Long TE (135-136 ms) improve the classification with respect to using only one echo time. To do this, a new approach that combines the two echo times was used to produce a single data vector from which relevant features were extracted.

Chapter 4 presents the modelling of MRS using mixtures for binned and truncated data. The main goal was to obtain a parametric representation of the signals. Prior knowledge of the relative position of the organic compound contributions in the MRS suggested the development of a probabilistic mixture model and its EM-based Maximum Likelihood Estimation for binned and truncated data. Experiments on MRS from BT demonstrated that the new approach extracts relevant information from the spectra for

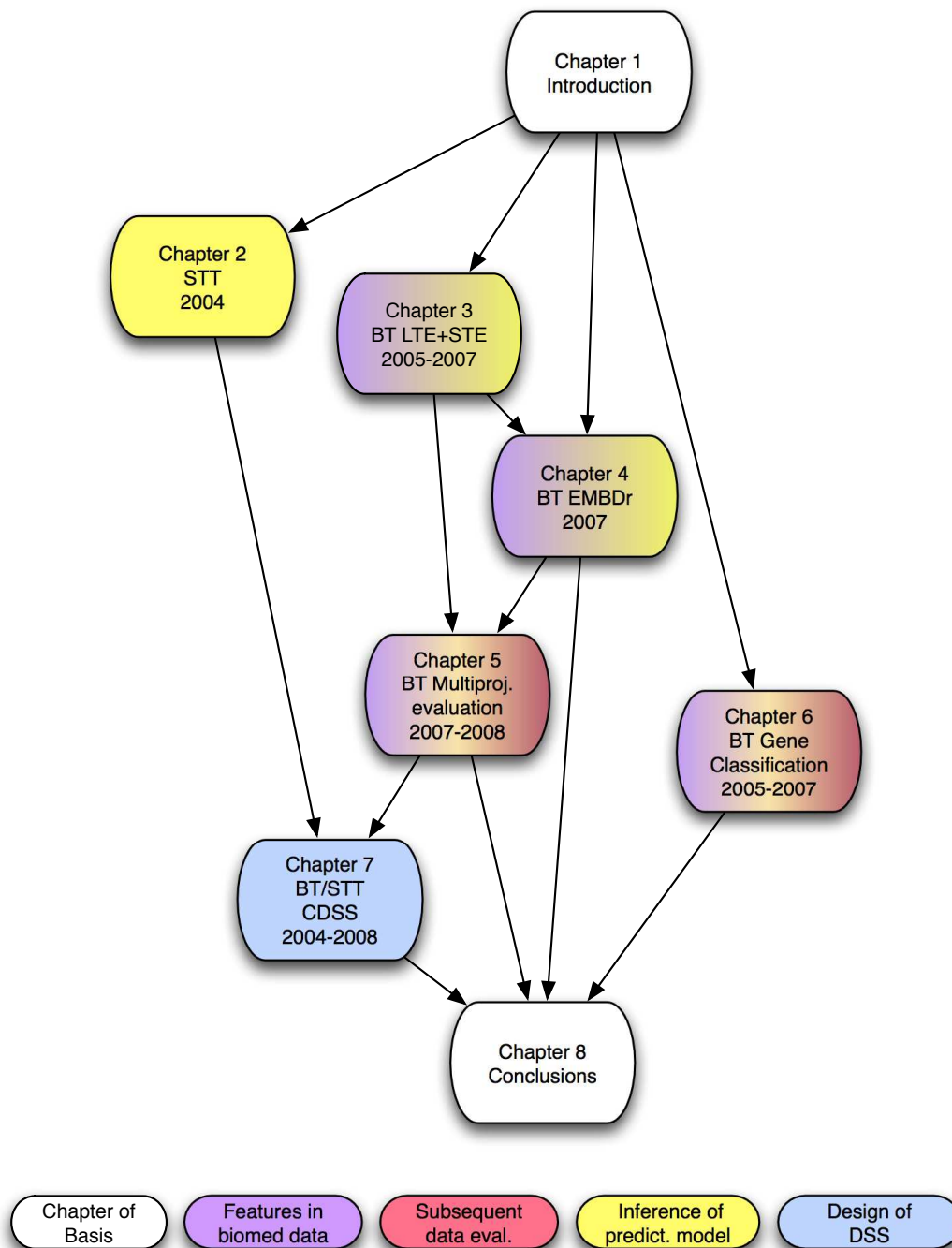


Figure 1: Flowchart of the Dissertation.

characterizing, classifying and combining the (Nuclear) Magnetic Resonance (MR) Spectra.

With regard to the evaluation of the classifiers with subsequently acquired data, Chapter 5 reports the results obtained in a multiproject-multicenter study of automatic BT classification by MRS. In this study, the robustness of brain tumor classifiers is demonstrated on subsequently acquired data independently of the manufacturer of the spectrometer, hospital and preprocessing implementation.

Chapter 6 moves the focus to the automatic classification of Brain Tumors by means of ex-vivo data. This chapter summarizes the automated brain tumor biopsy classification using single-labelling complementary DNA (cDNA) microarrays-based gene expression profiling.

Chapter 7 describes some developments of Decision-Support Systems (DSSs) for clinical environments. Three Clinical Decision-Support Systems (CDSSs) related to the two medical problems analysed in the Dissertation are introduced. First, the Clinical Decision-Support System (CDSS) for Soft Tissue Tumor Diagnosis based on a distributed architecture is described. Next, two CDSSs for Brain Tumors diagnosis are introduced. They are derived from the concepts developed for STT CDSS, however they are more evolved and scaled for their use in two multicenter projects, the eTUMOUR EC project (eTUMOUR) and the HEALTHAGENTS EC project (HEALTHAGENTS).

Finally, Chapter 8 summarizes the conclusions of the dissertation. This chapter also includes recommendations from the experience obtained during the development of the CDSS for STT diagnosis and BT diagnosis.

Chapter 1

Introduction

The systemic analysis of the patients' biomedical data, which come from different biological levels, offer greater information for the medical decision process. New biomedical technologies can allow the interpretation of the origin of the illnesses, moving to an evidence-based medicine paradigm. The recent increase in the complexity of the techniques for acquiring biomedical data as well as their innovative features, make it difficult for them to be incorporated in the clinician's practice. Therefore, it would be convenient to develop automatic data processing and predictive models to objectively assist the experts to interpret the data in the decision workflow of diagnoses, prognoses, and treatments.

The scope of the Thesis is the analysis of biomedical signals and data to assist in cancer diagnosis using Pattern Recognition (PR)-based Decision Support Systems. The people to whom this Thesis is directed might be interested in the application of PR technologies to support Health Systems and Clinical Research.

The first part of this chapter (Sections 1.1 and 1.2) describes the basis of the PR discipline, the data acquisition technologies and clinical problems studied in the Thesis. Section 1.3 reviews the design of Decision-Support Systems (DSSs) in Biomedicine and Cancer Research. Finally, the goals of the Thesis are presented in Section 1.4 followed by an overview of research contributions and partners.

1.1 Pattern Recognition

The PR discipline studies how to recognize an object through regularities that indicate that it belong 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 [1].

The goal of the design of a PR system is to recognize new samples and classify them into different categories using the observed features. To do this, there are two possibilities: 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 [2, 3]. Three major kinds of problems are solved in machine learning: the so-called pattern-recognition problem, the regression estimation problem, and the

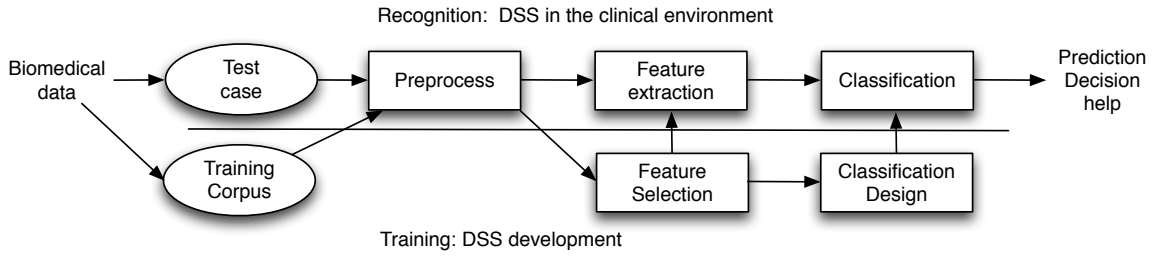


Figure 1.1: Pattern Recognition solution based on the Machine Learning approach.

density estimation problem [4]. The methodology applied in this Thesis solves the Pattern Recognition problems in cancer research using the Machine Learning approach.

The life cycle of a Pattern Recognition problem based on Machine Learning is composed of two main phases: the Training phase and the Recognition phase (see Figure 1.1). 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 preprocessing 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.

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

1.1.1 The Classification problem

In statistical PR, the optimal decision rule when assigning classes to the test cases is defined by the minimum global risk over the sample space. The minimum global risk can be achieved by minimizing the risk for each individual x of a sample.

If the Error rate (ERR) (see section 1.1.4) is defined as the evaluation metric of our classification models, the risk of classifying a sample $\mathbf{x} \in \mathbb{R}^D$ into class c is the probability of error (also called the Bayes' error), i.e. the sum of the posterior probabilities of x for each class except c , or by the probability of the complementary event, $p(error|\mathbf{x}) = 1 - \max_c p(c|\mathbf{x})$.

Therefore, the Bayes' decision rule [5] establishes that the minimization of global risk is obtained when each case \mathbf{x} of the sample is assigned to the class c that maximises the posterior probability $p(c|\mathbf{x})$

$$\hat{c} = \arg \max_c p(c|\mathbf{x}),$$

Bayes' theorem relates the conditional and marginal probabilities of \mathbf{x} and c :

$$p(c|\mathbf{x}) = \frac{p(c)p(\mathbf{x}|c)}{p(\mathbf{x})},$$

where $p(\mathbf{x}) = \sum_c p(c)p(\mathbf{x}|c)$ is constant for each class.

Hence, the Bayes decision rule can be written as

$$\hat{c} = \arg \max_c p(c)p(\mathbf{x}|c),$$

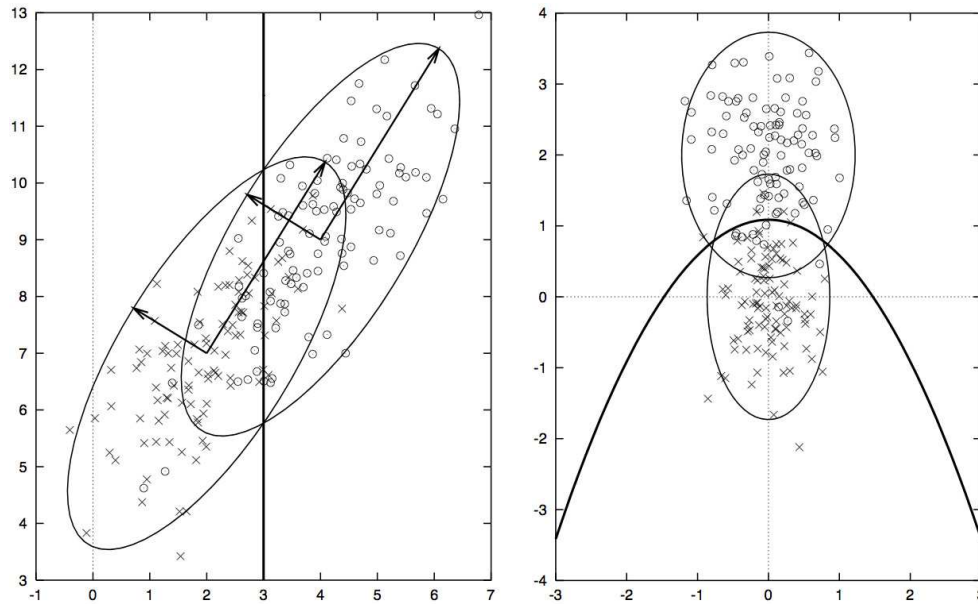


Figure 1.2: The left plot shows the linear decision boundary obtained by LDA for a two-class problem in a 2D space. The right plot shows the quadratic solution obtained by QDA for a similar problem.

where $p(c)$ is the prior probability of class c , and $p(\mathbf{x}|c)$ is the conditional probability density function that describes how likely it is to observe \mathbf{x} in class c .

In practice, the main problem is to approximate the real probability distributions through the observation of a finite sample of cases. To carry out this estimation, different methods have been proposed in the literature. Some statistical-based and non-statistical-based classifiers [6] have been used on the experiments described in this Dissertation. A brief description of them is presented below.

Gaussian Parametric Models

Discriminant analysis techniques are designed to find boundaries between classes [7]. Gaussian parametric models are based on the max-likelihood estimation of the Gaussian distributions for the classes under the study. The most popular method is Linear Discriminant Analysis (LDA), which is based on the assumption of a common variance of the classes (see Figure 1.2-left). In the Quadratic Discriminant Analysis (QDA) the covariances of the classes are independent, obtaining quadratic decision boundaries (see Figure 1.2-right).

The Fisher's LDA (FLDA) is a reduced-rank version of LDA, which projects the variables into the lower-dimensional subspace that maximizes the rate of the between-variance and the within-variance on the training corpus (see Figure 1.3).

K-Nearest neighbors (KNN)

The k-Nearest Neighbors (KNN) is a non-parametric classification method on which the samples are assigned to the classes based on the distances of the test cases to the training

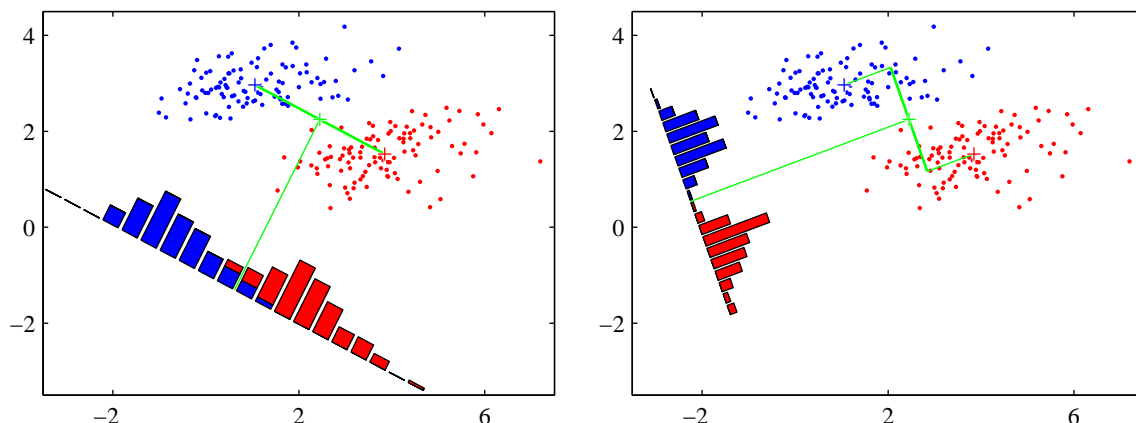


Figure 1.3: C. Bishop illustrates in [1] the basics of FLDA with this bi-dimensional example. The left plot shows samples from two classes along with the histograms resulting from projection onto the line joining the class means. This solution produces a considerable class overlap in the projected space. The right plot shows the corresponding projection based on the FLDA, showing the greatly improved class separation.

corpus in the feature space. KNN is a type of instance-based learning where the function is approximated locally and all computation is deferred until classification [5, 8] (see Figure 1.4).

Multilayer Perceptron (MLP)

The Multilayer Perceptron (MLP) is a connectionist model consisting on a network of simple units called perceptrons [9]. One perceptron computes an output as a non-linear function of the inner-product of the feature vector \mathbf{x} and the weight vector \mathbf{w} . In a MLP for solving a PR problem, the input signal propagates through the perceptrons layer-by-layer obtaining an approximation of the probability distribution of each class (see Figure 1.5). During the Training phase, the output error is propagated to the previous layers to update the weight vector of the perceptrons.

Support Vector Machines (SVM) and Least Squares SVM (LSSVM)

Support Vector Machines (SVM) are classification, nonlinear function estimation, and density estimation methodologies that are defined in the context of statistical learning theory and structural risk minimization [10]. SVM defines the optimal separating hyperplane between two classes with the maximal margin. This margin is the minimum distance of patterns of the training set to the hyperplane. Support Vector Machines represent data in a higher dimensional space where the linear separating hyperplane is built (see Figure 1.6). The explicit construction of a mapping to a higher dimensional space is avoided by using the kernel trick [11].

Least-Squares Support Vector Machines (LSSVM) is a reformulation of the SVM resulting in the solution of a linear system [11]. This characteristic makes LSSVM attractive for solving high dimensionality problems [12].

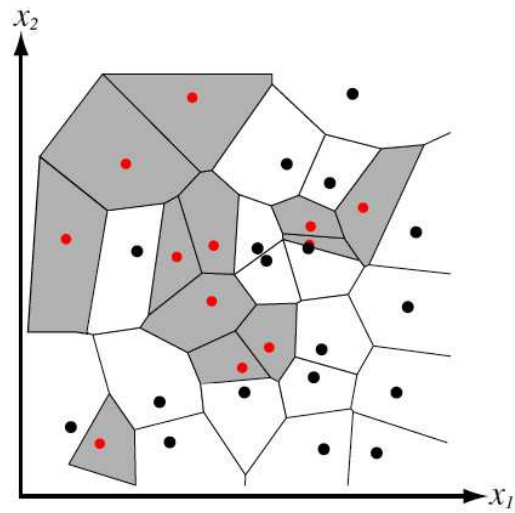


Figure 1.4: In two dimensions, the nearest-neighbor algorithm leads to a partitioning of the input space into cells, each labeled by the category of the training point it contains. From [5].

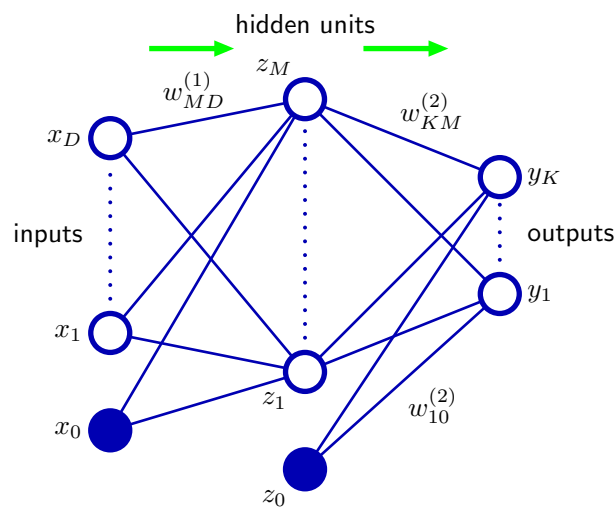


Figure 1.5: Network diagram for the 2-layer MLP. Arrows denote the direction of information flow through the network during forward propagation.

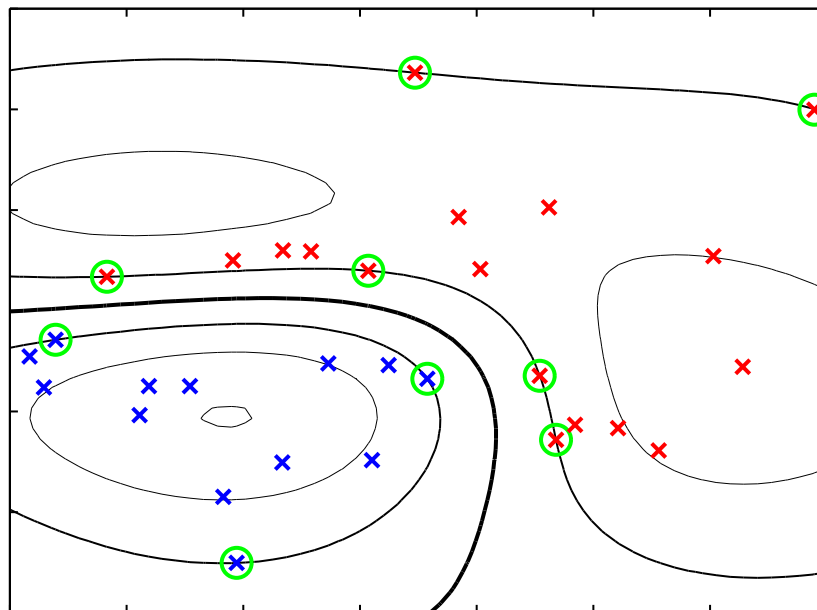


Figure 1.6: Decision boundary (bold line), support vectors (circles), margin boundaries (line defined by the support vectors) for a synthetic dataset from two classes in two dimensions obtained from a support vector machine having a Gaussian kernel function.

Ensemble Models

Ensemble assumes that if a set of individual classifiers are accurate enough and diverse, an ensemble of them performs better than a single predictor in the average [13, 14]. This is based on the reduction of the unbiased variance component of the error [15].

1.1.2 Feature selection and feature extraction

Feature selection and extraction step, the optimal or suboptimal search of features that represents more accurately the objects for the goal of the task is carried out. The feature selection and extraction methods used in our experiments were based on multivariate statistical analysis and signal processing.

Principal Component Analysis (PCA) is a well-known projection method that is commonly used for feature extraction in PR [16, 17]. 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 or loadings (PCs) of the analysis. Hence, feature reduction can be carried out discarding the remaining PCs.

Stepwise algorithm for feature selection in classification (SW) consists on a greedy hill climbing approach where the subset of features with the highest performance measure will be selected in each step and modified in the next step by the addition or deletion of one variable in the model [18].

ReliefF algorithm for Recursive Elimination of Features (ReliefF) is a feature selection

method based on how well features distinguish between instances that are near to each other [19–21]. In classification problems, the estimation of the quality of each variable is calculated by the accumulation of the distance between randomly selected instances and their k -nearest neighbors of a different class minus the distance to the k neighbors of the same class.

Coming from signal processing, the goal of Independent Component Analysis (ICA) is to extract source signals when only a linear mixture of these source signals is available. The most commonly used assumption is that the sources are mutually statistically independent, as well as independent from the noise components [22].

The Wavelet transform (WAV) consists on carrying out translations and scale transformations of a prototypical wavelet function in order to adjust the shape of a signal and to successively obtain the linear expansion of it [23].

1.1.3 Evaluation procedures

There is consensus in the literature that resampling techniques have to be applied when estimating classifier performance in small-sample datasets [24].

In the cross validation evaluation, the data set is divided into k subsets, and a training-test evaluation is repeated k times. Each time, one of the k subsets is used as the test set and the other $k - 1$ subsets are put together to form a training set. Then the average error across all k trials is computed. Every data point gets to be in a test set exactly once, and gets to be in a training set $k - 1$ times. The variance of the resulting estimate is reduced as k is increased. The disadvantage of this method is that the training algorithm has to be rerun from scratch k times.

k -Random Sampling Train-Test (kRSTT) with stratified test sets with k repetitions was the evaluation procedure used for some experiments of the Thesis. k partitions were randomly and independently prepared with the training set composed by 70% of cases of each class. The evaluation was carried out in a nested-loop that covered the feature and model selection in order to avoid underestimation of the true error.

1.1.4 Evaluation metrics

The error rate ERR or classification error rate is the ratio of the misclassified cases and the total number of cases. In a binary classifier A vs. B, ERR is defined by $\frac{e_A + e_B}{n_A + n_B}$, where n_A is the number of cases of the class A, and e_A the number of misclassified cases for this class and n_B is the number of cases of the class B, from which e_B cases were misclassified.

Balanced Error Rate (BER) and Geometric Mean of Recalls (GMOR) are used when the prevalences of the classes are quite different [25]. The BER is defined as $\frac{1}{2}(\frac{e_A}{n_A} + \frac{e_B}{n_B})$. The GMOR is defined as $\sqrt{(1 - \frac{e_A}{n_A})^2 + (1 - \frac{e_B}{n_B})^2}$.

The Receiver operating characteristic (ROC) curve is a graphical technique for assessing the performance of a binary classifier system [26]. The Area Under the Curve (AUC) of the ROC curve measures the discrimination capability of the binary classifiers depending on the sensitivity and specificity when varying the threshold of the latent space obtained by the models.

Berrar et al. in [27] presented some caveats to estimate the credibility intervals appropriately. In this review, $(1 - \alpha)\%$ -credibility intervals (equivalent to confidence intervals in a Bayesian framework) for the true error (τ) are calculated based on Jeffreys' Beta distribution as:

$$\tau \simeq \epsilon + \frac{2(M - 2m)z\sqrt{0.5}}{2M(M + 3)} \pm z\sqrt{\frac{\epsilon(1 - \epsilon)}{M + 2.5}},$$

when $10 \leq M \leq 200$ and $0 \leq m \leq M/2$, with M being the test size and m the number of errors, ϵ is the observed error and $z = 1.96$ for 95% confidence.

1.2 Cancer Research

In this section, we introduce the basis of the acquisition technologies of the biomedical data analyzed in the work: the Magnetic Resonance Imaging and Spectroscopy; and the Gene-expression Microarrays. Afterwards, the two medical problems tackled in this Thesis, the soft tissue tumor diagnosis and the brain tumor diagnosis, are outlined.

1.2.1 Data acquisition

Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) (or MR) is 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 [28].

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.

Magnetic Resonance Spectroscopy (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.

The NMR phenomenon observed can be performed with several sensitive nuclei: ^{31}P , ^{19}F , ^{13}C , and ^1H .

For the protons ^1H , the spin quantum number s associated to the particle angular moment takes the half-integer value $(1/2)$. For this particles, the *secondary spin quantum number* m_s takes the values $m_s = \{-1/2, 1/2\}$. That associates two possible potential energy levels to the ^1H particles (depending on m_s) in the presence of a magnetic field \mathbf{B}_0 , being the energy difference between both states

$$\Delta E_{pot} = -\gamma\hbar|\mathbf{B}_0|,$$

where γ is the gyromagnetic constant and \hbar is the reduced Planck's constant.

These particles do not align exactly with the axis of the external magnetic field \mathbf{B}_0 but precesses around it at a rate given by the Larmor frequency f_0 ,

$$f_0 = \gamma\mathbf{B}_0/(2\pi).$$

In the very first beginning time of the precessing motion around the field, the total magnetic moment \mathbf{M} of material is still near $\mathbf{0}$. As the elements of the molecule have

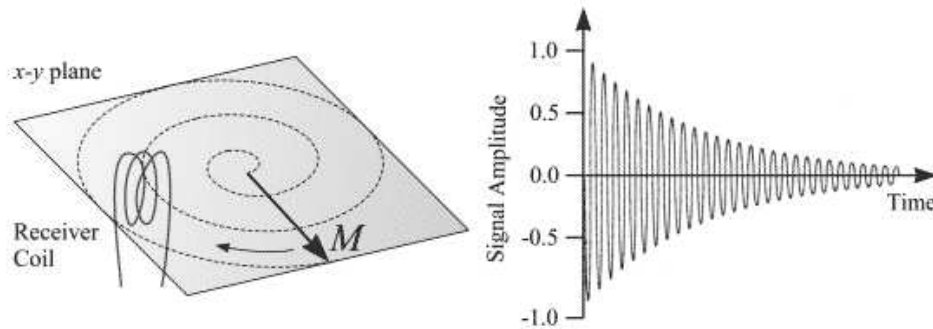


Figure 1.7: Left: M motion projected on the XY -plane when recovering the precession around B_0 . Right: Signal detected in the received coil when recovering the precession around B_0 .

their magnetic momenta, they generate magnetic fields that change with the thermal motion of the environment, so each spin is precessing around a local and changing magnetic field instead of the applied B_0 , so the spins are slowly deviated. The probability of the low energy orientations are slightly higher than the probability of the high energy levels. Hence, when the thermal equilibrium is reached there will be more spins parallel to the B_0 than anti-parallel. Consequently, a total magnetic moment $M \neq 0$ is observed.

If a Radio frequency (RF) pulse is applied to the sample, the spins experiment the influence of two magnetic fields B_0 and B_1 . The first is a static field, the second is an oscillating one of which the frequency is the resonant Larmor's frequency f_0 . A pulse in the y -axis produces a progressive decay of the M vector to the XY -plane. When the pulse is over, the spins return to the precession around the static magnetic field, obtaining, as a result a macroscopic M_{XY} motion similar to the precession of the spins (see Figure 1.7). Transversal to the axis coils can acquire the FID signal produced by the M_{XY} motion.

A noteworthy fact is that the magnetic field for each nucleus depends on the static magnetic field B_0 , but also on the local environment,

$$B_{eff} = B_0(1 - \sigma'),$$

where σ' is the shielding constant that depends on the electrical environment of the nucleus. This results in different frequencies of resonance (or Chemical Shift (CS)) of the same nuclei depending on the molecular environment and the main application of MRS in biochemistry and molecular biology (e.g. see Figure 1.8).

Gene-Expression Microarrays

Another source of biomedical data used in cancer research is the Gene-Expression Microarray biotechnology.

After the human genome sequentiation, the interest changed to the study of complete genomes instead of individual genes [30]. Microarrays measure the amount of messenger Ribonucleic acid (mRNA) transcribed from the genes of cells in a concrete population (e.g. a tissue). The microarray biotechnology allows the quantitation of the transcription of several genes simultaneously.

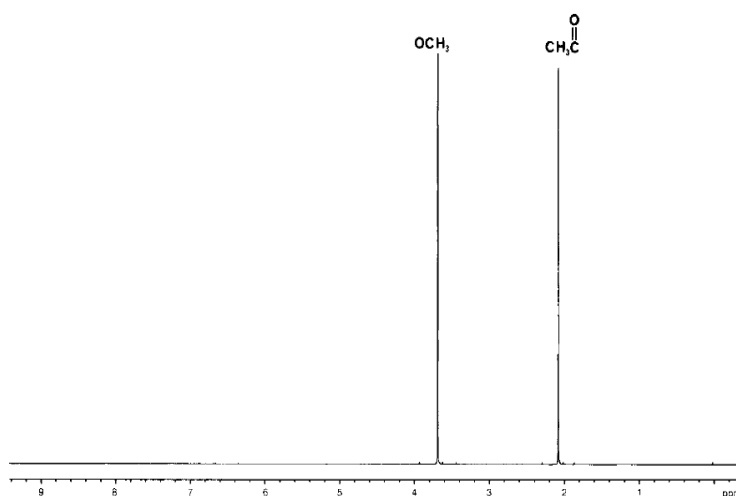


Figure 1.8: Chemical Shift (CS) for OCH_3 and CH_3CO [29].

A microarray is a matrix of nylon, glass or plastic on which single stranded DNA (called probes) are deposited in each cell of the matrix (called spots). In an experiment, the microarray is washed in a solution containing single stranded DNA of the biological sample of study (called target) labeled with a fluorescent dye or a radioactive element. The target sequences hybridize with the complementary probe sequences on the array. The scan of the microarray obtain different intensities in the microarray depending on the number of complementary probe hybridized with the target of each spot.

In gene expression microarrays, the target is usually obtained from the reverse transcriptase reaction from mRNA of tissue samples. Hence, the intensity of each spot is related to the amount of mRNA transcribed in the tissue.

The most spread platforms of microarrays are the *Spotted microarrays* and the *Oligonucleotide microarrays* (Figure 1.9). The oligonucleotide microarrays contain short sequences of nucleotides (from 14 to 20) for each gene in the microarray. Each oligonucleotide is selected to represent a given gene and is directly synthesized in the array by photolithography. Two copies of the oligonucleotide are synthesized in the matrix: the Perfect Match (PM), with the exact sequence of the fragment of the gene and other, the Mismatch (MiM), with a mismatch nucleotide in the middle of the fragment. The MiM copies are used afterwards for the correction of false positive expressed genes.

In spotted microarrays the probes are synthesized prior to the deposition on the array by a robot. The probes are typically cDNA that correspond to mRNA and they can be small sequences, complete genes or even arbitrary sequences. cDNA microarrays are usually configured for comparative analysis of two samples: the experimental samples and the reference. From each sample, the mRNA is extracted, denaturalized and converted to cDNA. The samples are labeled with different fluorescent (Red or Cy5 and Green or Cy3). The microarrays are washed in a dissolution with the mixture of the samples for the hybridization of the targets and the probes. The scan of the microarrays in the specific Red and Green channels allows the posterior quantification of the spots based on their intensities.

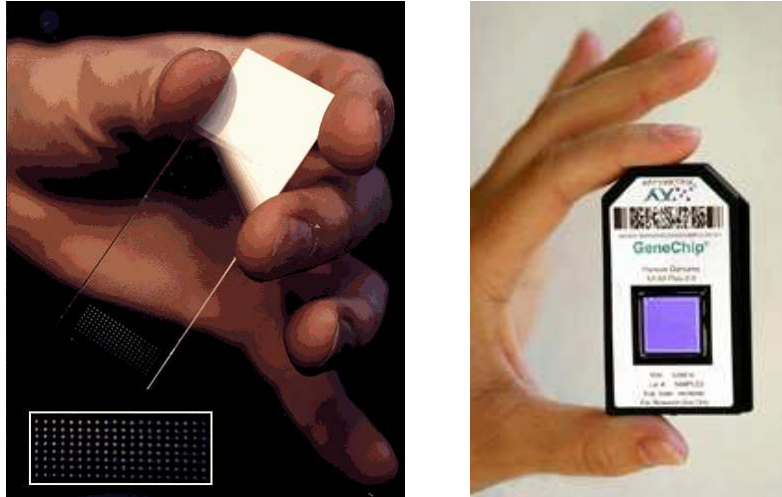


Figure 1.9: A Spotted microarray on the left and a Oligonucleotide microarray on the right.

1.2.2 Soft Tissue Tumors

Soft Tissue Tumor (STT) is an abnormal new growth that appears in supportive tissues of organs and other extrasketal structures [31].

Depending on the aggressivity, the soft tissue tumors are classified as either benign or malignant. Traditionally, tumors have been classified depending on the histological type of tissue: chondral, fibrous, fat, lymphoid, mesenquimal, muscular, nervous, sinovial, vascular, pseudotumor, cyst, and others.

STT is relatively infrequent in the general population^a, e.g. an annual incidence of 18 soft tissue sarcomas per one million masses evaluated. Benign tumor constitutes by far the most common STT, whereas, malignant tumors represent approximately 33% of them [32]. A clear discrimination between benign and malignant tumors becomes crucial in planning the proper diagnostic and surgical procedures.

A mass is the most common symptom of a soft tissue tumor. Most superficial tumors are diagnosed on clinical exploration, but if the lesion is more deeply located, a MRI exploration is indicated. Most of these tumors can be characterized with MR imaging findings because of their main textural and growth pattern. Although MRI protocols have been defined for staging, grading and characterizing for the soft tissue tumor [33], most radiologists and clinicians are not familiar with their appearance and differential diagnosis. Hence, a diagnostic support tool developed to help in making an accurate diagnosis of tumor aggressiveness would improve the correct management of these patients.

^a<http://seer.cancer.gov>

1.2.3 Brain Tumors

Brain tumor is defined^b as the growth of abnormal cells in the tissues of the brain. Brain tumors can be benign (non-cancerous) or malignant (cancerous).

Different classifications are defined in the brain tumors depending on the factor. A first classification is based on the type and location of the cells that originate from the tumor. Primary brain tumors are created by uncontrolled cells 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 tumors are created by cells spread from cancers primarily located in other organs (metastatic tumors).

Based on the malignancy characteristics the brain tumors are classified in grades. Grade I: tumors with a low proliferative potential and a possibility of cure following surgical resection alone. Grade II: generally infiltrating tumors low in mitotic activity, but with a potential to recur. Grade III: tumors with histological evidence of malignancy. Grade IV: generally associated with a rapid pre and postoperative evolution of the disease.

The World Health Organization (WHO) classification [34]^c of Tumors of the Central Nervous System (CNS) describes the diagnoses as a taxonomy. Compatible groups were established according to the studies elaborated by medical experts during the INTERPRET [35] and eTUMOUR [36] projects based on the 1993 WHO Classification.

Brain tumors are the second fastest growing cause of cancer death among people older than 65 years [37], in addition, they are also the second leading cause of cancer death (after leukemia) in children under fifteen years and young adults up to the age of thirty-four.

The most frequent primary brain tumor types are of glial origin (40%), 30% are derived from the meninges and 8% are located in cranial and spinal nerves [38]. Lymphomas and germ cell tumors account for 3% and 1% respectively. Glioblastoma is the most frequent tumor of glial origin (51%).

The distributions of the tumor types by age are not uniform. In early ages (0-19), embryonal/primitive/medulloblastoma and pilocytic astrocytoma are the most common types of tumors. In young adults (20-34), pituitary tumors and meningiomas are the most prevalent types. In older adults, the meningioma and, also, the malignant tumor glioblastoma are the most frequent tumors. The most common brain tumor types by age are summarized in Table 1.1.

There is no specific clinical symptom because the brain does not possess pain receptors. Therefore, a slowly progressive mass effect and intracranial pressure may appear. The presenting symptoms largely depend on the tumor location. Visual disturbance or respiratory arrest, as well as epilepsy, or sudden intracranial hypertension may occur.

When the aforementioned symptoms arise, a radiological examination based on Computed Tomography (CT) or 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 [39, 40]. MRI is widely used for determining tumor extension in surgical and radiotherapy planning, with a sensitivity of 14%-100% depending on tumor type and precision of the wording used by radiologist [41].

^bDictionary of Cancer Terms. National Cancer Institute. <http://www.cancer.gov/dictionary/> (Online; accessed 10-3-2008)

^chttp://www.brainlife.org/who/2007_classification.htm

Table 1.1: Most common brain tumor and CNS types by age (CBTRUS 2005, Statistical Report [38])

Age (years)	Most Common Histology	Second Most Common Histology
0-4	Embryonal/primitive/medulloblastoma	Pilocytic astrocytoma
5-9	Pilocytic astrocytoma	Embryonal/primitive/medulloblastoma
10-14	Pilocytic astrocytoma	Embryonal/primitive/medulloblastoma
15-19	Pilocytic astrocytoma	Pituitary
20-34	Pituitary	Meningioma
35-44	Meningioma	Nerve Sheath
45-54	Meningioma	Glioblastoma
55-64	Meningioma	Glioblastoma
65-74	Meningioma	Glioblastoma
75-84	Meningioma	Glioblastoma
85+	Meningioma	Glioblastoma

Nevertheless, the gold-standard method in the brain tumor diagnosis is the histological examination of tumor 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.

MRS is slowly becoming an additional accurate non-invasive technique for initial examination of brain masses [12, 40], due to its capability to provide useful chemical information of different metabolites for characterizing brain tumors and its complementary role to MRI [42].

Besides, high throughput gene expression techniques provide a transcriptomic signature, which can be used to classify different biological samples or pathological conditions.

1.3 Decision-Support Systems in Cancer Research

The development of new biomedical technologies allows the diagnosis and prognosis of tumors using non-invasive methods. Nowadays, biomedical techniques such as ^1H MRS, MRI, High-Resolution Magic Angle Spinning (HR-MAS), Deoxyribonucleic acid (DNA) microarrays, and anatomopathological images offer complementary points of view for the clinical management of the patients in cancer research. Semi-automatic analyses of these data based on PR are increasingly being used by clinicians in the research of decision making relative to the diagnosis and treatment of new patients.

From a multidisciplinary research standpoint, fast and easy access to the most recent Information and Communication Technology (ICT) developments by users is one of the main requirements to obtain translational results. Currently, the biomedical and Information and Communication Technology (ICT) disciplines have joined their efforts to integrate the heterogeneous data acquired from different levels of biological data in order to analyse the behaviour, origin, and discrimination of the different types of tumors under a systemic paradigm.

A DSS is a computer-based system that aids in the process of decision making [43]. Clinical Decision-Support Systems (CDSSs) are *active knowledge systems which use two or more items of patient data to generate case-specific advice* [44]. CDSSs based on Pattern

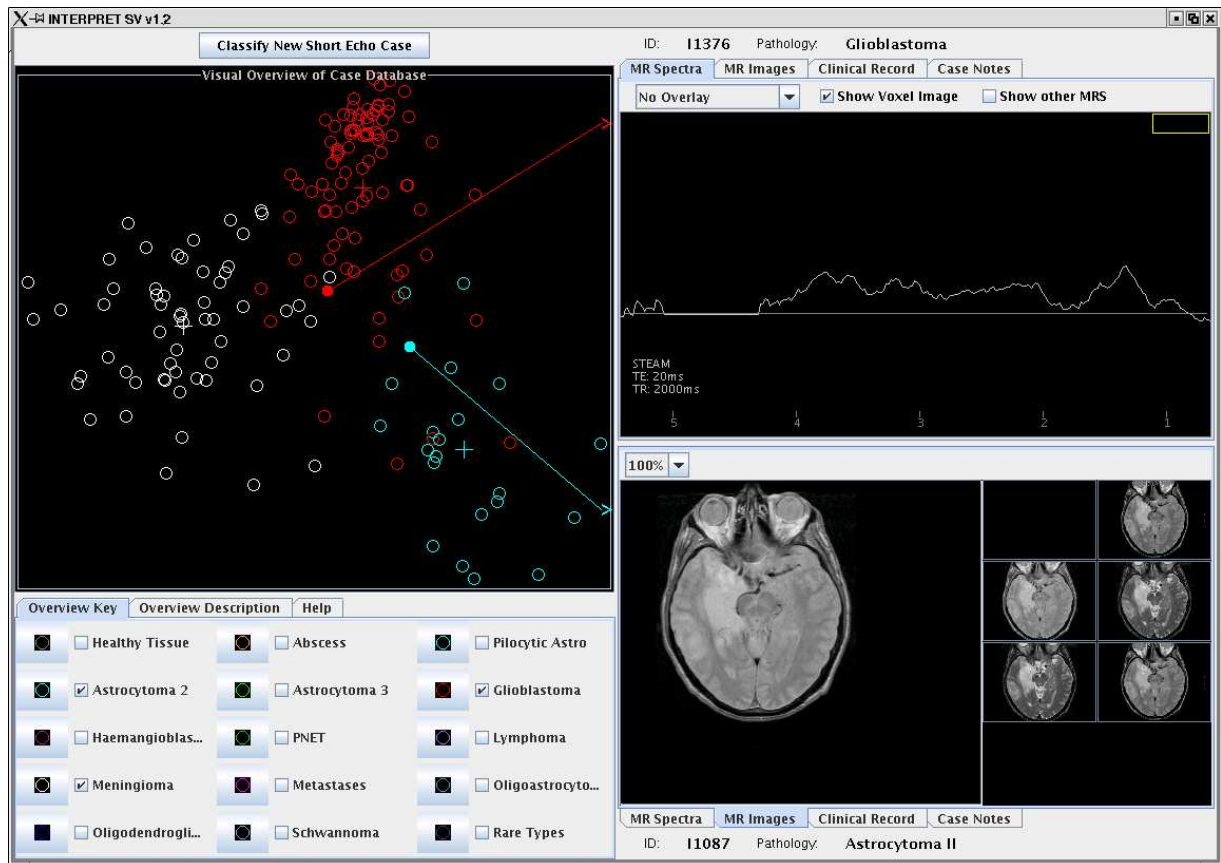


Figure 1.10: SV INTERPRET GUI for SV spectra.

Recognition (PR) have been widely accepted in medical applications due to their capability for optimisation, flexibility, accuracy for predictive inference and interpretability [45].

Clinical Decision-Support Systems facilitate access to high-quality medical practices, which contributes to better health care [46]. Nowadays, the compilation of scientific documents, clinical data and biomedical signals acquired with standardized protocols makes it possible to design the decision support system that are based on them.

Studies on the evaluation of the use of CDSSs indicate a positive effect on clinical practice [47–49] and quality of primary care [50]. The potential benefits of CDSSs are summarized by Coiera and Sintchenko in three major items [51, 52]: improved patient safety, improved quality of care, and improved efficiency in health care delivery.

A good example related to the diagnosis of brain tumors studied in this Thesis is the SV INTERPRET GUI^d based on a LDA latent space projection (see Figure 1.10). The SV INTERPRET GUI provides easy access to a database of spectra, images and clinical information from 304 validated cases of human brain tumors. It is designed to allow the display of classification plots which are useful for automating the classification of tumor spectra [53].

^dhttp://azizu.uab.es/INTERPRET/sv_tutorial/index.php

1.4 Goals, Contributions, Projects and Partners

1.4.1 Goals

The goals of the Thesis are the development and evaluation of PR-based Clinical Decision-Support Systems with high accuracy in classification, interpretability by means of clinical knowledge and generalization of the performance for new samples that are subsequently obtained in different clinical centers. Special attention is given to:

- the processing, extraction and combination of relevant information from the biomedical data;
- the inference of predictive models to discriminate among diagnoses;
- the evaluation of the automatic classification by newly collected data; and
- the integration of the predictive models into decision support systems for clinical environments.

In order to focus our research, two medical problems are tackled: **Soft Tissue Tumor diagnosis** and, **Brain Tumor diagnosis with in-vivo and ex-vivo data**. Each of these problems has its own specific medical protocol for exploring and confirming the diagnosis. Hence, the predictive models and decision support systems developed in the Thesis are focused on the biomedical data used by these medical protocols:

- Findings in Magnetic Resonance Imaging for Soft Tissue Tumors;
- Magnetic Resonance Spectroscopy and Gene Expression Microarrays in Brain Tumors.

1.4.2 Contributions

The scientific results of this Thesis concern the application of Pattern Recognition to cancer 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, Magnetic Resonance, and Molecular Biology.

The research related to the inference of predictive models to discriminate Soft Tissue Tumors (Chapter 2) was published as a journal paper [54]. Work related to this task was also presented in [55–58]. A strong collaboration with radiologists Dr. Luis Martí-Bonmatí and Dr. Joaquín Galant was established in this study for the definition of the image findings, for the compilation of the datasets, and for the interpretation of the results under a clinical point of view.

In the Brain Tumor classification task, the study of feature extraction from two MRS echo times for classification of brain tumor diagnoses (Chapter 3) has been accepted for publication as journal paper [59]. Work related to this study was also presented in [60, 61]. The team of Prof. Dr. Carles Arús contributed to this study in the definition of the hypothesis and the discussion of the results from the point of view of biochemistry. In addition, the development of the model MRS using mixtures for binned and truncated data (Chapter 4) was published in [62].

The multiproject-multicenter evaluation of Brain Tumor classifiers by subsequently acquired MRS has been accepted for publication as journal paper [63]. In this study, a strong collaboration among the partners of the work package 3 of eTumour was established. The teams of Prof. Dr. Sabine Van Huffel and Prof. Dr. Lutgarde Buydens contributed to this study with the preparation of classifiers based on techniques developed by their universities: LSSVM and Bi-directional Kohonen Networks (BDK), respectively. The major contribution of this Thesis to the multiproject-multicenter evaluation was the definition and co-ordination of the collaboration, the development of PR-based models, the evaluation of the classifiers, and the analysis and discussion of the results.

Some contributions have also been made about feature extraction methods for MRS in a journal paper [64] whose main author is Mr. Jan Luts (KULeuven University).

A study about automated brain tumor biopsy prediction using single-labelling cDNA microarrays-based gene expression profiling has been published as a journal paper in [65]. This Thesis contributed to the brain tumor biopsy prediction with the design of the protocol to pre-process the single-labelling cDNA microarrays used in the study and with the development and analysis of the PR experiments for the automatic classification of tumors by means of the gene-signatures. Work related to this task was also presented in [61].

The development of a distributed Decision-Support System (dDSS) for Soft Tissue Tumors was presented in [66, 67], and has also been studied in [58, 68]. Finally, the development of CDSSs for brain tumors was published in [61].

1.4.3 Research Projects and Partners

Part of this Thesis is from the research developed related to the eTUMOUR project and the HEALTHAGENTS project, which was done between 2004 and 2008.

- 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 Universidad Politécnic de Valencia, were focused on two practical outputs. The first output is the development of PR-based engines that offer an objective solution to specific tumor discrimination problems using MR in vivo data. The classification engines reported in Section 7.3 are integrated as the predictive core of the Decision Support System developed in the project. The second output is the contribution to the knowledge of the cancerogenesis of the brain tumors, e.g. generating hypotheses of putative genes related to certain metabolic or functional pathways of the tumor types, such as the study reported in Chapter 6.

- 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 problems is to create a multi-agent DSS to assist in the early diagnosis of brain tumors 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 tumors.

The solution reported in Section 7.4 was adopted as the classification framework for the HEALTHAGENTS agent-based distributed network. In further developments, the distributed network of classifiers incorporates a ranking model of the classifiers, self evaluation and semi-automatic re-training of the classifiers. These systems provide predictions for incoming cases and refine themselves in a continuous auditing process. Furthermore, they can also contribute to improvement in quality control of the biomedical datasets that are included in the distributed networks.

The author of this Thesis performed a research stay at the Dept. of Electrical Engineering (ESAT) of the Katholieke Universiteit Leuven (Belgium), from July 16, 2006 until February 2, 2007. The research stay was performed under the supervision of Prof. Sabine Van Huffel of the SCD division at ESAT.

The new multidisciplinary research line Biomedical-Mining^e was opened in 2003 by the IBIME^f group of the research institute ITACA in the Universidad Politécnic de Valencia. In 2008, this group consisted of five computer engineers, one physicist, and one mathematician who were doing their research in six research projects: two European projects on Brain Tumor automatic classification [36, 69], a Spanish project on prediction of post-partum depression [70], a Spanish project on Genotyping and Psychiatric Genetics [71], a Spanish project on Prostate Cancer Early Detection [72] and a Spanish project on the Evaluation of Decision-Support Systems [73].

^e<http://http://bmg.webs.upv.es>

^f<http://http://www.ibime.upv.es>

Bibliography

- [1] Bishop CM (2006) Pattern Recognition and Machine Learning (Information Science and Statistics). Springer
- [2] Duda RO, Hart PE, Stork DG (2000) Pattern Classification. John Wiley & Sons, New York, 2nd. edition
- [3] Bishop CM (2006) Pattern Recognition and Machine Learning. Information Science and Statistics. Springer-Verlag, New York
- [4] Vapnik VN (1999) The Nature of Statistical Learning Theory. Statistics for Engineering and Information Science. Springer-Verlag, New York, second edition
- [5] Duda R, Hart P, Stork D (2001) Pattern Classification. John Wiley and Sons, inc.
- [6] Jain AK, Duin RPW, Mao J (2000) Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1): 4–37
- [7] Vandeginste B, Massart D, Buydens L, de Jong S, Lewi P, Smeyers-Verbeke J (1998) Handbook of Chemometrics and Qualimetrics: Part B. Science Publishers
- [8] Hastie T, Tibshirani R, Friedman JH (2001) The Elements of Statistical Learning. Springer
- [9] Rosenblatt F (1958) The Perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review 65(6): 386–408
- [10] Vapnik V (1995) The Nature of Statistical Learning Theory. Springer, N.Y.
- [11] Suykens JAK, Vandewalle J (1999) Least Squares Support Vector Machine Classifiers. Neural Process Lett 9(3): 293–300
- [12] Lukas L, Devos A, Suykens JAK, Vanhamme L, Howe FA, Majós C, Moreno-Torres A, Graaf MVD, Tate AR, Arús C, Huffel SV (2004) Brain tumor classification based on long echo proton MRS signals. Artif Intell Med 31: 73–89
- [13] Hansen L, Salamon P (1990) Neural Network Ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(10): 993–1001
- [14] Dietterich TG (2000) Ensemble Methods in Machine Learning. Lecture Notes in Computer Science 1857: 1–15

- [15] Domingos P (2000). A Unified Bias-Variance Decomposition for Zero-One and Squared Loss
- [16] Burges CJ (2004) Geometric Methods for Feature Extraction and Dimensional Reduction: A Guided Tour. Technical report, Microsoft Research, University of Toronto http://research.microsoft.com/~cburgess/tech_reports/tr-2004-55-dimred.pdf
- [17] Fukunaga K (1990) Introduction to statistical pattern recognition (2nd ed.). Academic Press Professional, Inc., San Diego, CA, USA
- [18] Roever C, Raabe N, Luebke K, Ligges U, Szepannek G, Zentgraf M (2006) The klaR Package. Department of Statistics, University of Dortmund
- [19] Kira K, Rendell L (1992) A practical approach to feature selection. In Proceedings of the Ninth International Conference on Machine Learning, pp. 249–256. Morgan Kaufmann, San Francisco, CA
- [20] Kononenko I (1994) Estimating Attributes: Analysis and Extensions of RELIEF. Lecture Notes in Computer Science 784: 171
- [21] Robnikikonja M, Kononenko I (2003) Theoretical and Empirical Analysis of ReliefF and RReliefF. Mach Learn 53(1-2): 23–69
- [22] Comon P (1994) Independent component analysis, a new concept? Signal Processing 36(3): 287–314
- [23] Daubechies I (1992) Ten Lectures on Wavelets (CBMS - NSF Regional Conference Series in Applied Mathematics). Soc for Industrial & Applied Math
- [24] Martin JK, Hirschberg DS (1996) Small Sample Statistics for Classification Error Rates I: Error Rate Measurements. Technical Report ICS-TR-96-22 citeseer.ist.psu.edu/article/martin96small.html
- [25] Guyon I, Alamdari ARSA, Dror G, Buhmann JM (2006) Performance Prediction Challenge. In IJCNN '06. International Joint Conference on Neural Networks, pp. 1649–1656
- [26] Begg CB (1986) Statistical methods in medical diagnosis. Crit Rev Med Inform 1(1): 1–22
- [27] Berrar D, Bradbury I, Dubitzky W (2006) Avoiding model selection bias in small-sample genomic datasets. Bioinformatics 22(10): 1245–1250
- [28] Dodd A (2003) An introduction to the theory of nuclear magnetic resonance. Technical report, University of Melbourne <http://teaching.shu.ac.uk/hwb/chemistry/tutorials/molspec/nmr1.htm>
- [29] Lambert J, Mazzola E (2003) Nuclear Magnetic Resonance Spectroscopy. An Introduction to Principles, Applications, and Experimental Methods. Prentice Hall
- [30] Draghici S (2003) Data Analysis Tools for DNA Microarrays. Chapman & Hall/CRC

-
- [31] Galant J (1998) Magnetic Resonance contribution to the characterization of soft tissue tumor lesions. Ph.D. thesis, Facultad de Medicina
- [32] Weatherall PT (1995) Benign and malignant masses. MR imaging differentiation. *Magnetic Resonance Imaging Clinics of North America* 3(4): 669–694
- [33] De Schepper AM, De Beuckeleer L, Vandevenne J, Somville J (2000) Magnetic resonance imaging of soft tissue tumors. *Eur Radiol* 10(2): 213–223
- [34] Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P (2007) The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 114(2): 97–109
- [35] INTERPRET Consortium (Accessed: 28 April 2008) Interpret web site. <http://azizu.uab.es/INTERPRET>
- [36] 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. Technical report, FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC <http://www.etumour.net>
- [37] BTS (Accessed: 28 January 2007) The Brain Tumor Society. <http://www.tbts.org>
- [38] CBTRUS (2005) (2006) Statistical Report: Primary Brain Tumors in the United States, 1998-2002. Technical report, Central Brain Tumor Registry of the United States <http://www.cbtrus.org/reports/reports.html>
- [39] Lin AP, Tran TT, Ross BD (2006) Impact of evidence-based medicine on magnetic resonance spectroscopy. *NMR Biomed* 19(4): 476–483
- [40] Howe FA, Opstad KS (2003) ¹H MR spectroscopy of brain tumours and masses. *NMR Biomed* 16(3): 123–131
- [41] Julia-Sape M, Acosta D, Majos C, Moreno-Torres A, Wesseling P, Acebes JJ, Griffiths JR, Arús C (2006) Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. *J Neurosurg* 105(1): 6–14
- [42] Galanaud D, Nicoli F, Chinot O, Confort-Gouny S, Figarella-Branger D, Roche P, Fuentes S, Le Fur Y, Ranjeva JP, Cozzone PJ (2006) Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy. *Magnetic Resonance in Medicine* 55(6): 1236–1245
- [43] Finlay P (1995) *Introducing Decision Support Systems* (2nd edition). *The Journal of the Operational Research Society* 46(10): 1282–1283
- [44] Wyatt J, Spiegelhalter D (1991) Field trials of medical decision-aids: potential problems and solutions. *Proc Annu Symp Comput Appl Med Care* pp. 3–7
- [45] Lisboa PJ, Taktak AFG (2006) The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Netw* 19(4): 408–415

- [46] Brand D (2002) Electronic Decision Support for Australia's Health Sector. Technical report, National Electronic Support Taskforce <http://www.health.gov.au>
- [47] Kawamoto K, Houlihan CA, Balas EA, Lobach DF (2005) Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 330(7494): 765
- [48] Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, Sam J, Haynes RB (2005) Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 293(10): 1223–1238
- [49] Kaplan B (2001) Evaluating informatics applications—clinical decision support systems literature review. *Int J Med Inform* 64(1): 15–37
- [50] Delaney BC, Fitzmaurice DA, Riaz A, Hobbs FD (1999) Can computerised decision support systems deliver improved quality in primary care?. Interview by Abi Berger. *BMJ* 319(7220): 1281
- [51] Coiera E (2003) *Guide to Health Informatics*, chapter Clinical Decision Support Systems. Hodder Arnold
- [52] Sintchenko V (2004) 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
- [53] Tate AR, Underwood J, Acosta DM, Julia-Sape M, Majos C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabanas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Remy C, Heerschap A, Watson D, Griffiths JR, Arús C (2006) 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
- [54] Garcia-Gomez JM, Vidal C, Martí-Bonmatí L, Galant J, Sans N, Robles M, Casacuberta F (2004) Benign/malignant classifier of soft tissue tumors using MR imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine* 16(4): 194–201
- [55] García-Gómez JM, Vidal C, Juan A, Martí-Bonmatí L, Galant J, Robles M (2002) Soft tissue tumours pattern classification techniques using MR and epidemiology data. *Magnetic Resonance Materials in Physics, Biology and Medicine* 15: 274–275
- [56] Malassidis E, García-Gómez JM, Vidal C, Martí-Bonmatí L, Robles M, Millet J (2002) El análisis de componentes principales como método de clasificación y visualización de tumores de partes blandas. In *INFORMED-2002*, pp. 213–215
- [57] Vidal C, Garcia-Gomez JM, Marti-Bonmati L, Robles M, Galant J, Millet J (2002) Study of vascular histological group tumors using pattern recognition techniques. In *INFORMED*, pp. 207–211

-
- [58] Garcia-Gomez JM, Vidal C, Vicente J, Martí-Bonmatí L, Robles M (2004) Medical decision support system for diagnosis of soft tissue tumors based on distributed architecture. *Conf Proc IEEE Eng Med Biol Soc* 5: 3225–3228
- [59] Garcia-Gomez J, Tortajada S, Vidal C, Julia-Sape M, Luts J, Van Huffel S, Arús C, Robles M (2008) The influence of combining two echo times in automatic brain tumor classification by Magnetic Resonance Spectroscopy. *NMR in Biomedicine* 21(10): 1112–1125
- [60] Tortajada S, García-Gómez JM, Vidal C, Arús C, Julià-Sapé M, Moreno A, Robles M (2006) Improved classification by pattern recognition of brain tumours combining long and short echo time 1H-MR spectra. In SpringerLink, editor, *Book of Abstracts ESMRMB 2006 - Supplement 1, Journal Magnetic Resonance Materials in Physics, Biology and Medicine*, volume 19, pp. 168–169
- [61] García-Gómez JM, Tortajada S, Vicente J, Sáez C, Castells X, Luts J, Julià-Sapé M, Juan-Císcar A, Van Huffel S, Barcelo A, Ariño J, Arús C, Robles M (2007) Genomics and Metabolomics Research for Brain Tumour Diagnosis Based on Machine Learning. In IWANN, *Lecture Notes in Computer Sciences* 4507, pp. 1012–1019
- [62] Garcia-Gomez JM, Robles M, Van Huffel S, Juan-Císcar A (2007) Modelling of Magnetic Resonance Spectra using mixtures for binned and truncated data. In Springer, editor, *Proceedings of the 1st Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA)*, volume 2 of *Lecture Notes in Computer Science Series 4478*, pp. 266–273. Springer-Verlag, Girona, Spain
- [63] Garcia-Gomez JM, Luts J, Julia-Sape M, Krooshof P, Tortajada S, Robledo JV, Melssen W, Fuster-Garcia E, Olier I, Postma G, Monleon D, Moreno-Torres A, Pujol J, Candiota AP, Martinez-Bisbal MC, Suykens J, Buydens L, Celda B, Van Huffel S, Arus C, Robles M (2009) Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *MAGMA* 22(1): 5–18
- [64] Luts J, Pouillet JB, Garcia-Gomez JM, Heerschap A, Robles M, Suykens JAK, Van Huffel S (2008) Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. *Magnetic Resonance in Medicine* 60(2): 288–298
- [65] Castells X, García-Gómez JM, Navarro A, Acebes JJ, Godino Ó, Boluda S, Barceló A, Robles M, Ariño J, Arús C (2008) Automated brain tumor biopsy prediction using single-labelling cDNA microarrays-based gene expression profiling. *Diagnostic Molecular Pathology* p. Accepted for publication
- [66] García-Gómez JM, Vidal C, Vicente J, Martí-Bonmatí L, Robles M (2004) Medical Decision Support System for Diagnosis of Soft Tissue Tumors based on Distributed Architecture. In *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE*, volume 2, pp. 3225–3228
- [67] García-Gómez JM, Vidal C, Martí-Bonmatí L, Robles M (2004) Distributed decision support architecture for soft tissue tumor diagnosis. In *MEDINFO 2004: 11th World Congress on Medical Informatics*, p. 1608

- [68] Vicente J, García-Gómez JM, Vidal C, Martí-Bonmatí L, del Arco A, Robles M (2004) SOC: A Distributed Decision Support Architecture for Clinical Diagnosis. In Springer, editor, Lecture Notes in Computer Science, volume 3337, pp. 96–104
- [69] Mier M (2005) HealthAgents, Annex 1. Agent-based Distributed Decision Support System for Brain Tumour Diagnosis and Prognosis. Technical report, HealthAgents consortium, IST-2004-27214, VI framework programme, EC <http://www.healthagents.net>
- [70] Convenio con la UVEG FIS Referencia: PI041635 (2006-2008). Vulnerabilidad genético-ambiental a la depresión posparto
- [71] Fundació Privada Centre De Regulació Genòmica (2004-2008). Aplicación de Redes Neuronales Al Proyecto Enfermedades Psiquiatricas
- [72] Ministerio de Industria, Turismo y Comercio (2007-2009). DSSPROSTATA: Sistema de Soporte a la Decisión para Detección Precoz de Cáncer de Próstata
- [73] Ministerio de Industria, Turismo y Comercio (2007-2008). SADE - SISTEMA DE AYUDA A LAS DECISIONES

Chapter 2

Benign/Malignant Classifier of Soft Tissue Tumors Using MRI

This study presents a PR approach to the STT benign/malignant character diagnosis using MRI applied to a large multicentric database. The aim of this work is to develop and test an automatic classifier to differentiate benign from malignant STT by using classical MR imaging findings and epidemiological information. A database of 430 patients (62% benign and 38% malignant) from several European multicentric registers was used. There were 61 different histologies (36 with benign and 25 with malignant nature). Three PR methods (Artificial Neural Networks, Support Vector Machine, K-Nearest Neighbor) were applied to learn the discrimination between benignity and malignancy based on a defined MRI findings protocol. After the system had learned by using training samples (with 302 cases), the clinical decision support system was tested in the diagnosis of 128 new STT cases. An 88–92% accuracy was obtained in a not-viewed set of tumors using the PR techniques. The best results were obtained with a back-propagation artificial neural network. Benign vs. malignant STT discrimination is accurate by using PR methods based on classical MR image findings. This objective tool will assist radiologists in STT grading.

This chapter has been published as journal paper in [1].

2.1 Introduction

Benign tumor constitutes by far the most common STT. Most of these tumors can be characterized with MRI because of their main textural and growth pattern. Soft tissue malignant tumors represent approximately 33% of them [2]. Although, in clinical practice, most superficial tumors are diagnosed solely on the basis of clinical exploration and no further studies are required, more deeply located lesions are nowadays studied with MRI. A clear discrimination between benign and malignant tumors becomes crucial in planning the proper diagnostic and surgical procedures.

As STTs are infrequent in the general population [3], most radiologists and clinicians are not familiar with their appearance and differential diagnosis. Furthermore, their appearance is sometimes misleading. Well-defined malignant tumors are placed on the less aggressive side of the spectrum and have an appearance usually similar to benign lesions, making it difficult on many occasions to distinguish between them. On the other side, some benign tumors can deviate from their typical homogeneous appearance, com-

plicating the differentiation. A diagnostic support tool developed to help in making an accurate diagnosis of tumor aggressiveness would improve the correct management of these patients.

To create such a diagnostic support tool for STT, a large amount of data is required, usually larger than those disposable in an isolated center. The solution to the lack of data has been solved by the merge of several hospital institution cases. The Information and Communication Technologies allow radiologists to share relevant information about the problem and special cases. In this way, other groups, like the Belgian Soft Tissue Neoplasm Registry [4], also recruit MR imaging cases in order to offer support services to professionals and scientists.

To our knowledge, there is no published report in which automatic classification methods, such as Artificial Neural Networks (ANN), KNN, or SVM, have been applied successfully to classify the STT tumors with a MR imaging-findings dataset. A recent article related to soft tissue sarcoma classification using gene expression profile and SVM has been presented by Segal NH et al [5]. However, all of the aforementioned methods have been applied in many medical classification systems with good performance. In tumor classification, KNN has been used in combination with two other classifiers for melanoma diagnosis [6]; SVM has been successfully used for prediction of malignancy of ovarian tumors [7], in analyzing microcalcifications in digital mammograms [8], and in discriminating breast tumors [9–11]; ANN has also been widely and successfully applied in discriminating breast tumors [12–14], in evaluating different features based on defined imaging criteria [14–18]; and in cervical cancer screening [19]. An interesting review of the applications of ANN to the diagnosis, prognosis, and survival analysis in the medical domains of oncology, critical care, and cardiovascular medicine has been presented by Lisboa [20] to assess the evidence of healthcare benefits involving the application of ANN. In this review, the author presents important recommendations for the design and evaluation of ANN in medicine.

Our main purpose was to obtain an automatic classifier of STT aggressiveness. The PR discipline allows the adaptation of models to a specific problem. This adaptation consists of an inference process in which a mathematical mechanism changes its parameters based on a set of supervised training examples. PR methods (ANN, KNN, and SVM) have been applied to design this software because they allow the adaptation of the system to the problem by the use of knowledge contained in the study samples. The classification tool will be used by radiologists and other medical professionals in a clinical decision-support system to aid in the diagnostic report. Our work presents a multidisciplinary study, grouping MRI expertise radiologists and computer science engineers on PR techniques, to discriminate between benign/malignant grading of STT.

2.2 Materials and methods

2.2.1 Patients and equipment

Patients with confirmed musculoskeletal STT examined with MRI were retrospectively selected. Tumors with a biopsy procedure before the imaging studies were excluded. Patients were consecutively recruited at five different European hospital centres.

All malignant and some benign, other than classical hemangioma and lipomas, had his-

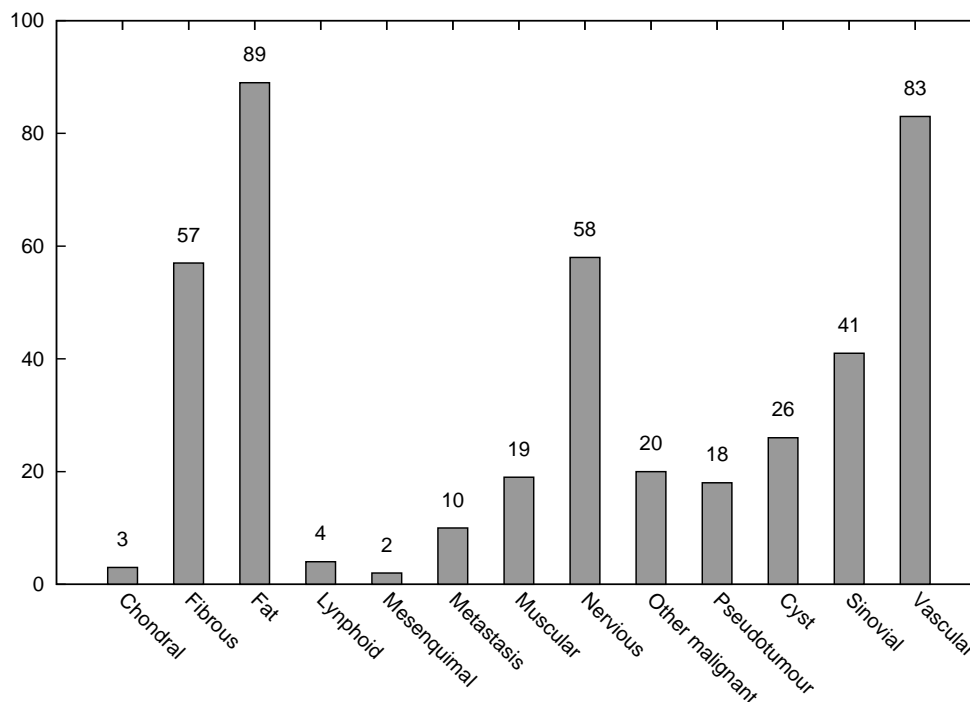


Figure 2.1: Distribution of STT cases into histological groups.

topathological confirmation of their nature. Characteristic multimodality imaging appearance and follow-up studies were employed to confirm the diagnosis of these benign tumors. Finally, there were 430 patients, 267 patients (62.1%) with a benign lesion and 163 patients (37.9%) with a malignant neoplasm. The histological grouping distribution of the cases is shown in Figure 2.1.

Imaging studies were performed on 0.5 T and 1.5 T Gyroscan NT (Philips Medical System, Eindhoven, Netherlands) and 0.5 T Signal (General Electric Medical System, Milwaukee, WI) units. T1-weighted (T1W, TR/TE, 450–650 ms/15–30 ms) and T2-weighted fat-suppression (T2W-FS, TR/TE, 2,800–3,500 ms/80–120 ms) or STIR (TR/TE/TI, 1,400–1,600 ms/40 ms/100–120 ms) images were systematically obtained (Figure 2.2).

2.2.2 Database

The following epidemiological and MR image findings [2, 4, 21–23] were obtained from the clinical records and radiological examinations:

- Age: The age of the patient in years.
- Clinical presentation: Reason for the patient’s consultation (mass, pain, growth, neurological symptoms, skin alterations, asymptomatic).
- Localization: The anatomic compartment where the STT was located (subcutaneous, intramuscular, intermuscular, or intraarticular). Lesions extending to more than one compartment were classified within the compartment where the largest part of the tumor was located.

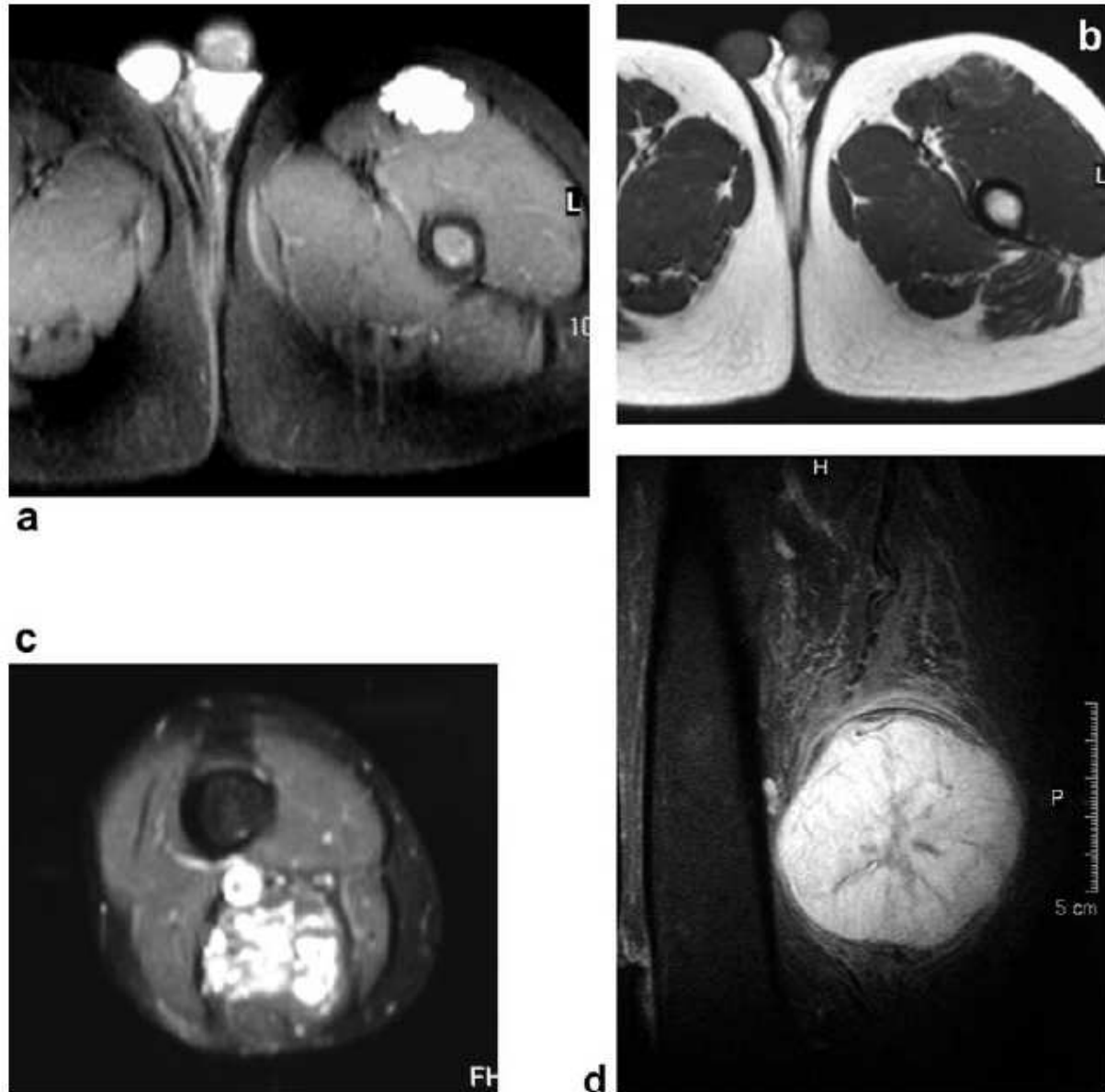


Figure 2.2: Representative MR images. Homogeneous highly hyperintense hemangioma in T2 weighted image **a** with hyperintense tracts in the T1 weighted image **b**. Serpiginous appearance of an intramuscular hemangioma (**c**, fat suppression T2 weighted image). Oval hyperintense heterogeneous neurinoma with inner areas of fibrosis (**d**, fat suppression T2 weighted image)

- Size: The maximum diameter of the lesion, expressed in centimeters (rounded to the nearest unit).
- Shape: The external morphology of the lesion (rounded, oval, lobulated, serpiginous, fusiform, strand, or irregular). Serpiginous lesions were those with tubular structures within them and polylobular external surface (Figure 2.2).
- Signal intensity: Represents the relative signal intensity of the lesions in the images obtained with different weightings (T1 and T2/STIR). Muscle was selected as the reference signal intensity tissue, with an intermediate-low signal intensity in most sequences (T1-weighted, T2-weighted fat suppression, STIR images). There were four categories in the T1-weighted images (hypointense, isointense, hyperintense, and very hyperintense) and three in the T2-weighted images (isointense, slightly hyperintense, and highly hyperintense) (Figure 2.2).
- Margins: The external borders of the lesion. There were three categories: infiltrative (when most of the margins were blurred or the tumor clearly extended at any point to the surrounding tissue), mainly well-defined lesions (with a partial sector of the margin with unclear borders), and noninfiltrative (when the tumor borders were clearly visible without any peripheral infiltration).
- Homogeneity: Presence of different areas of signal intensity variation within the lesion in the different image weightings. Taking into consideration the proportion and signal-intensity differences, there were four categories: very homogeneous (only one signal intensity constituent), homogeneous (an area not more than 25% of the lesion with a slight change in its signal intensity compared with the rest of the lesion), heterogeneous (an area between 25 and 75% with different signal intensities), and very heterogeneous (lesions with more than a 75% of them with different components showing large differences in their signal intensity) (Figure 2.2).
- Edema: A peripheral ill-defined area or halo, hypointense in T1-weighted images but hyperintense in STIR/T2-weighted images with a width larger than 5 mm (no, yes).
- T1 hyperintense tracts: Linear or reticular areas with a high signal intensity in the T1-weighted images without fat suppression (no, yes). (Figure 2.2).
- Multiplicity: Whether the patient had previously or currently presented with similar STT (no, yes).
- Target appearance: Whether the lesion had an inner well-defined centre and concentric peripheral rings with different signal intensities (no, yes).
- Muscular atrophy: A decrease in the diameter of the muscle or group of muscles related to the lesion, with enlargement of the fat planes between the muscular fascicles, and especially if the contralateral extremity was present for comparison (no, yes).
- Intratumoral hemorrhage: Whether heterogeneous areas with hyperintense and hypointense zones were shown in the T1 and STIR/T2 weighted images (no, yes).

- Calcification: Very hypointense areas in all the obtained images and weightings, after excluding vessels and hemosiderin. Phebolits were considered only if a plain film or a CT examination showed a rounded calcification with an inner radiolucent centre (no, yes, phebolits).
- Dependence: Whether an anatomic structure was identified with a very close relationship with the lesion and the lesion seems to originate from that anatomic structure. Special care was taken not to include as dependence a relationship of displacement (none, nerve, tendon, and vessel).
- Intratumoral fat: When an area of signal intensity equal to that of the subcutaneous fat in all the pulse sequences was observed within the lesion. Lesions with fat were further classified regarding the presence of hyperintense zones in the STIR/T2 weighted images (no, fat without hyperintense zones in T2-STIR, fat with hyperintense zones in T2-STIR).
- Fibrosis: Defined as areas of very low signal intensity within the tumor in all the images, mainly if the morphology was irregular or ring-like (no, yes) (Figure 2.2).
- Fascial relationship: The relationship of the subcutaneous lesions with the superficial fascia (no contact, small contact, contact with acute angles between the lesion and the fascia, larger contact with obtuse angles, fascia penetration, and fascial origin of the lesion).
- Bone alterations: If there was an alteration of the underlying bone, it was classified as bone remodeling with periosteal reaction, or bone destruction, including cortical bone permeation (no, yes with erosion-invasion, yes with reshape or reaction).
- Vessels: Whether large vessels constitute the essential part of the lesion (no, yes).

Two experience radiologists, who were masked to the final histological diagnosis, reviewed the studies. Differences were resolved by consensus with a third radiologist.

2.2.3 Methods

A clinical decision support system is a tool to improve the quality in the clinical decision, being defined as an active knowledge system that generates specific advice to each new case [24]. It integrates three main features: medical knowledge that solves the disease cases [25, 26], patient data with specific biomedical information of each patient, and specific advice for each case based on the medical knowledge and the patient data.

There are different methods to design clinical decision-support systems using the artificial intelligence approach. In this study, an inductive strategy, more commonly named PR strategy, was applied. The conclusions made by the clinical decision-support system were inferred by the knowledge captured from a group of samples representing the problem.

The PR methodology [27] includes the next steps. Preprocessing/filtering includes data adaptation to a format recognized by the learning procedures; consequently, it incorporates database connection, queries to get the required registers, and syntactic analysis to transform the formats. The filtering step consisted of a transformation from the Access database format to an ASCII file. Another filter task was the normalization of the data to

avoid artificial order relationships in qualitative variables with a simple local codification, and range transformation to weight correctly the continuous variables into $[0, 1]$ [28]. As an example, variables like dependence, which takes the values none, nerve, tendon and vessel, that doesn't have an order relation between the values, was codified as $(1\ 0\ 0\ 0)$, $(0\ 1\ 0\ 0)$, $(0\ 0\ 1\ 0)$, $(0\ 0\ 0\ 1)$; but variables, such as age or size, that have an important relation in the order of its values, were normalized into $[0, 1]$.

The database was randomly divided into two different sets, one training set used to learn the computational models (70%, 302 cases) and the test set (30%, 128 cases), used to estimate the accuracy of the system. We contemplated the use of the three basic evaluation methods in the experiment planning: Cross Validation (CV), Leave-One-Out Cross-Validation (LOO), and Independent Test (IT). We chose an independent training and test sets because the real independence between both sets will probably give a less optimistic confidence interval of the efficiency (shown in Table 2.1) than the one obtained with nonindependent test sets.

The learning process is the most important step in the clinical decision support system development and its principal goal is to adapt the software system to the requirements of the problem. An iterative procedure takes the relevant information by studying examples to inference the model of the problem. When the learning process was completed, the adaptive system was released and new cases could be studied. The variables of the unseen new cases were analyzed and the generalization knowledge made from the training examples got the associated prediction.

ANN, KNN, and SVM were used as PR methods (see Section 1.1) to design the clinical classifier. These three techniques are nonparametric because they do not require a priori knowledge of the probability distributions. In the STT benign and malignant classification problem, the functionality of a multilayer perceptron can be seen like a nonconnected boundary to separate the regions between each class. From the 22 image characteristics defined in Section 2.2.2, 20 discrete variables were translated to local codification and 2 continuous variables were normalized. A total of 70 variables were used as input of the neural networks. No feature selection or pruning algorithms were applied to keep the clinical meaning of the input space. The learning process with ANN (see Section 1.1.1) consisted in the error minimization between the labels of the training samples and the outputs of the network by means of the back-propagation algorithm with momentum [29]. The topologies studied in the experiment were composed of one or two hidden layers. Multilayer perceptrons with one layer can implement decision boundaries linear at intervals. Besides, multilayer perceptrons with two layers can implement any boundary based on hyperplanes [29]. The number of layers and number of units for each hidden layer were selected empirically using the CV ($k=10$) re-sampling procedure over the training dataset. The number of neurons in the hidden layers varied from 10 to 100. In order to avoid the over-fitting, early stopping was applied. The stop criterium was Mean Squared Error (MSE) < 0.01 , or differential MSE ($|\text{MSE}_i - \text{MSE}_{i-1}| < 0.000001$), or the reach of a number of cycles ≥ 3000 . We have used the Stuttgart Neural Networks Simulator to train the ANN models of the experiments [30].

Besides, in our STT experiments, the KNN classifier (see Section 1.1.1) was tested varying k between 1 and 15 with three distance metrics^a. The best configuration was the five nearest neighbors with the L_1 distance, but its efficiency was worse than that obtained

^a $L_0(x, y) = \max_{1 \leq i \leq D} |x_i - y_i|$, where D is the number of variables.

Table 2.1: Comparison of Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) methods. Efficiency is expressed as value $\pm 95\%$ confidence interval. PPV, positive predictive value; NPV: negative predictive value.

Method	Efficiency (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
ANN	92 \pm 5	86	95	91	93
KNN	88 \pm 6	86	90	84	91
SVM	90 \pm 5	84	94	89	89

Table 2.2: Best results obtained with artificial neural network (ANN). One case was not classified by the network. Distribution of cases into histological groups.

Test / final diagnosis	Malignant	Benign	Unknown	Total
Malignant	42 (86%)	6 (12%)	1 (2%)	49
Benign	4 (5%)	75 (95%)	0 (0%)	79
Total	46	81	1	128

with the other two methods, as will be commented upon in the Results. With respect to the configuration of SVM (see Section 1.1.1) in our experiments, polynomial, Gaussian and radial basis function were tested kernels, and SVM-light was the used implementation [31].

2.3 Results

We created a database with a set of MRI findings of 430 soft tissue tumors (62% benign, 38% malignant) coming from five different European hospital centres, with the purpose of developing an automatic benign/malignant classifier of STT by using a set of PR techniques. The original hospitals of the cases were not available, so, no stratification by hospital was possible in our experiments.

The best results obtained by classifying the test set of 128 stratified samples with ANN, KNN, and SVM are shown in Tables 2.2, 2.3, 2.4. A comparative table including the main evaluation metrics^b estimated for each technique is in Table 2.1.

With ANN, the best efficiency (92%) was obtained with 30 neurons in the first hidden layer and 10 in the second layer, trained with the back-propagation algorithm. Despite the prevalence of benign classes, sensitivity and specificity were quite good (86 and 95%, respectively). The decision thresholds made one malignant case not be classified (Table 2.2). Best efficiency with KNN (88%) has been obtained using $k = 5$ neighbors and L_1 distance. The sensitivity was 86% (Table 2.3) and the specificity 90%. Best efficiency with SVM (90%) was obtained using a polynomial kernel of 7th degree. The sensitivity (84%) was

$$L_1(x, y) = \sum_{i=1}^D |x_i - y_i| \quad L_2(x, y) = \sqrt{\sum_{i=1}^D (x_i - y_i)^2}.$$

^bAccuracy, (TP+TN)/N; N, number of samples; TP, number of true positive samples; TN, number of true negative samples;

Sensitivity, TP/(TP+FN+UN); FN: number of false negative samples; UN, number of unknown answers;

Specificity, TN/(TN+FP+UN); FP: number of false positive samples;

Positive predictive value, PPV=TP/(TP+FP);

Negative predictive value, NPV=TN/(TN+FN)

Table 2.3: Best results obtained with KNN. The number of neighbors was $k = 5$.

Test / final diagnosis	Malignant	Benign	Total
Malignant	42 (86%)	7 (14%)	49
Benign	8 (10%)	71 (90%)	79
Total	50	78	128

Table 2.4: Best results obtained with SVM. They were obtained using a polynomial kernel (degree 7).

Test / final diagnosis	Malignant	Benign	Total
Malignant	41 (84%)	8 (16%)	49
Benign	5 (6%)	74 (94%)	79
Total	46	82	128

worse than that obtained with the other techniques and specificity 94% (Table 2.4).

ANN reached the best results, with more than 90% in accuracy and specificity coefficients (Table 2.1). Classifier fusion has been considered, but we have studied the concrete error cases and only one ANN error is correctly classified by KNN and SVM (Table 2.5). This result indicates that it is not necessary for the classifier fusion, but it must be considered in future improvements.

2.4 Discussion

Magnetic resonance imaging clearly depicts STT due to its high-contrast tissue resolution and multiplanar capability. Correct diagnosis includes the detection, characterization, and staging of these tumors. The MR differentiation between benign and malignant tumors is complicated by the low prevalence of these lesions, radiologist’s low experience in nondedicated hospitals, indirect information of the MR signal intensities, high diversity of histologies, and natural evolution of the lesions. There is much controversy regarding the MR value in the differentiation of benign and malignant STT. Although a subjective analysis has a high sensitivity for malignancy (78%) [23], the statistical combination of individual MR parameters provides both a high sensitivity and specificity, close to 81% (2). Our approach was to develop an automatic technique to help radiologists to characterize benign and malignant STT with high accuracy.

The use of automatic techniques to help in the characterization of STT requires the homogenization of patient databases. A standard MR protocol that incorporates the relevant information to characterize STT is very important in an automatic classification task. For this reason, a standard protocol based on MRI T1-weighted and T2/STIR-weighted images has been used by MR expert radiologists in this study. The automatic systems can retrieve the relevant information of the cases, and good classification was reached because of the quality of the variables extracted by the radiologist. We also did not use the information generated by contrast-enhanced images because of the diversity of the data within centers. The usefulness of gadolinium-enhanced MR images, although widely accepted, did not assist in further narrowing the differential diagnosis compared with plain MR scans in a large series of cases [32]. Moreover, the lack of uniform imaging sequences and parameters after contrast administration within the different centers made

Table 2.5: Errors in benign vs. malignant classification with the Artificial Neural Network (ANN), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) methods.

True diagnosis	ANN	KNN	SVM	Histology
Benign	Benign	Malignant	Malignant	Lymphomatoid granulomatosis
Benign	Malignant	Malignant	Malignant	Granuloma
Benign	Benign	Malignant	Benign	Hemangioma
Benign	Malignant	Malignant	Malignant	Localized pigmented villonodular synovitis
Benign	Benign	Malignant	Benign	Pseudotumor
Benign	Malignant	Malignant	Malignant	Myxoma
Benign	Malignant	Malignant	Malignant	Morton's neuroma
Benign	Benign	Malignant	Benign	Desmoid
Malignant	Benign	Malignant	Malignant	Malignant fibrous histiocyoma
Malignant	Unknown	Malignant	Benign	Malignant fibrous histiocyoma
Malignant	Malignant	Malignant	Benign	Malignant fibrous histiocyoma
Malignant	Benign	Benign	Benign	Metastases
Malignant	Benign	Benign	Benign	Metastases
Malignant	Malignant	Benign	Malignant	Malignant schwannoma
Malignant	Malignant	Malignant	Benign	Myxoid liposarcoma
Malignant	Malignant	Benign	Malignant	Liposarcoma
Malignant	Malignant	Benign	Malignant	Synovial sarcoma
Malignant	Benign	Malignant	Benign	Synovial sarcoma
Malignant	Benign	Benign	Benign	Fibrosarcoma
Malignant	Benign	Benign	Benign	Fibrosarcoma

standardization difficult.

It has been important to make a multicentric consortium to recruit the STT database. A large database was needed to properly develop a general classifier. The use of cases from different places to train and test the decision-support systems makes it useful to generalize our results. However, although the series of STTs used in our study was large, with good results in a multicentric database, the incorporation of new STT categories and presentations to the study will increase the quality of the decision-task process. All classifiers yielded significant prediction efficiency (88-92%), with great consensus between them. The best technique was the ANN with 10 errors over 128 cases (92% of hits).

The similar results of efficiency, sensitivity, and specificity obtained with the three evaluated techniques may prove that some bias may be due to an inherent biological overlap. One possible source of bias is the fact that radiologists' readings were used to train the clinical decision support system. The use of simple MRI variables, as the ones used in this study, by the STT classifier with the adaptive approach offered by the PR discipline will minimize the influence of reading errors from different radiologists.

The practical result of the PR experiments should be a clinical decision-support system in STT discrimination. This tool is useful because it is an objective method to confirm the benign/malignant characterization, allows the investigation of suspicious cases, and has the capability to assist the radiologist's decision in a new case and aids in the education of new radiologists' expertise in STT.

The inductive approximation used in this study enables the system to learn the important features of the cases to make a classification into two categories (benign vs. malignant) defined by the values of the variable or character. This STT problem is a dichotomy classification between two exclusive classes. The main difficulty of this problem is the heterogeneous constitution of the classes. Each class (benign and malignant) is formed by a group of histologies of very different origin. Consequently, there is not just one big cluster per class, making the decision boundary easy to discriminate.

The number of cases in which all the techniques failed is 8 (6% of test set) (Table 2.5). There are various reasons to explain this circumstance: these cases can be very abnormal and the learning process has not enough information to extract the discriminative pattern correctly; the cases are really atypical in their appearance; or the registries have incorrectly filled variables.

The use of the PR approach in medical research is growing more and more because of the new possibilities opened by the digitalization of biomedical information. The disposability of biomedical information in electronic repositories [33] enables the data-mining studies and research by automatic methods to get new and interesting correlations to improve human health. The PR approach can help the search for biomedical pointers of important diseases (like tumors or degenerative diseases) and the development of technological tools applied to clinical and basic medicine research [5, 6, 9, 10, 34, 35].

Our benign/malignant classification is part of a larger study project on STT computer-aided diagnosis, including visualization of tumor data, and feature selection and classification into different histologies [36, 37]. A Web Services layer and a graphical application are being implemented in order to allow the use of clinical decision-support systems developed by the group from distributed application around Internet such as a Web Site, Desktop Application, Clinical Electronic History, or Telemedicine applications.

Bibliography

- [1] Garcia-Gomez JM, Vidal C, Martí-Bonmatí L, Galant J, Sans N, Robles M, Casacuberta F (2004) Benign/malignant classifier of soft tissue tumors using MR imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine* 16(4): 194–201
- [2] Weatherall PT (1995) Benign and malignant masses. MR imaging differentiation. *Magnetic Resonance Imaging Clinics of North America* 3(4): 669–694
- [3] Salgado R, Marck EV (2006) *Imaging of Soft Tissue Tumors*, chapter *Soft Tissue Tumours: the Surgical Pathologist’s Perspective*. Springer
- [4] De Schepper AM, De Beuckeleer L, Vandevenne J, Somville J (2000) Magnetic resonance imaging of soft tissue tumors. *Eur Radiol* 10(2): 213–223
- [5] Segal NH, Pavlidis P, Antonescu CR, Maki RG, Noble WS, DeSantis D, Woodruff JM, Lewis JJ, Brennan MF, Houghton AN, Cordon-Cardo C (2003) Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. *Am J Pathol* 163(2): 691–700
- [6] Sboner A, Eccher C, Blanzieri E, Bauer P, Cristofolini M, Zumiani G, Forti S (2003) A multiple classifier system for early melanoma diagnosis. *Artif Intell Med* 27(1): 29–44
- [7] Lu C, Van Gestel T, Suykens JAK, Van Huffel S, Vergote I, Timmerman D (2003) Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. *Artif Intell Med* 28(3): 281–306
- [8] Bazzani A, Bevilacqua A, Bollini D, Brancaccio R, Campanini R, Lanconelli N, Riccardi A, Romani D (2001) An SVM classifier to separate false signals from microcalcifications in digital mammograms. *Phys Med Biol* 46(6): 1651–1663
- [9] Chang RF, Wu WJ, Moon WK, Chen DR (2003) Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis. *Ultrasound Med Biol* 29(5): 679–686
- [10] Chang RF, Wu WJ, Moon WK, Chou YH, Chen DR (2003) Support vector machines for diagnosis of breast tumors on US images. *Acad Radiol* 10(2): 189–197
- [11] Liu HX, Zhang RS, Luan F, Yao XJ, Liu MC, Hu ZD, Fan BT (2003) Diagnosing breast cancer based on support vector machines. *J Chem Inf Comput Sci* 43(3): 900–907

- [12] Dhawan AP, Chitre Y, Kaiser-Bonasso C (1996) Analysis of mammographic micro-calcifications using gray-level image structure features. *IEEE Trans Med Imaging* 15(3): 246–259
- [13] Sahiner B, Chan HP, Petrick N, Wei D, Helvie MA, Adler DD, Goodsitt MM (1996) Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Trans Med Imaging* 15(5): 598–610
- [14] Chen DR, Chang RF, Kuo WJ, Chen MC, Huang YL (2002) Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. *Ultrasound Med Biol* 28(10): 1301–1310
- [15] Floyd CEJ, Lo JY, Yun AJ, Sullivan DC, Kornguth PJ (1994) Prediction of breast cancer malignancy using an artificial neural network. *Cancer* 74(11): 2944–2948
- [16] Gurney JW, Swensen SJ (1995) Solitary pulmonary nodules: determining the likelihood of malignancy with neural network analysis. *Radiology* 196(3): 823–829
- [17] Baker JA, Kornguth PJ, Lo JY, Floyd CEJ (1996) Artificial neural network: improving the quality of breast biopsy recommendations. *Radiology* 198(1): 131–135
- [18] Abdolmaleki P, Buadu LD, Murayama S, Murakami J, Hashiguchi N, Yabuuchi H, Masuda K (1997) Neural network analysis of breast cancer from MRI findings. *Radiat Med* 15(5): 283–293
- [19] (1999) Assessment of automated primary screening on PAPNET of cervical smears in the PRISMATIC trial. PRISMATIC Project Management Team. *Lancet* 353(9162): 1381–1385
- [20] Lisboa PJG (2002) A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Netw* 15(1): 11–39
- [21] Galant J (1998) Magnetic Resonance contribution to the characterization of soft tissue tumor lesions. Ph.D. thesis, Facultad de Medicina
- [22] Galant J, Marti-Bonmati L, Saez F, Soler R, Alcalá-Santaella R, Navarro M (2003) The value of fat-suppressed T2 or STIR sequences in distinguishing lipoma from well-differentiated liposarcoma. *Eur Radiol* 13(2): 337–343
- [23] Moulton JS, Blebea JS, Dunco DM, Braley SE, Bisset GSr, Emery KH (1995) MR imaging of soft-tissue masses: diagnostic efficacy and value of distinguishing between benign and malignant lesions. *AJR Am J Roentgenol* 164(5): 1191–1199
- [24] van der Lei J, Talmon JL (1997) *Handbook of Medical Informatics*, chapter Clinical Decision-Support Systems. Springer
- [25] van Bommel JH (1997) *Handbook of Medical Informatics*, chapter Methods for Decision Support. Springer
- [26] Shortliffe EH, Perreault LE, editors (2000) *Medical Informatics, Computer Applications in Health Care and Biomedicine*. Springer

-
- [27] Duda R, Hart P, Stork D (2001) Pattern Classification. John Wiley and Sons, inc.
- [28] Bishop CM (1995) Neural Networks for Pattern Recognition. Oxford University Press
- [29] Ripley BD, Hjort NL (1995) Pattern Recognition and Neural Networks. Cambridge University Press, New York, NY, USA
- [30] Zell A, Mamier G, Vogt M, Mache N, Huebner R, Herrmann KU, Soyez T, Schmalzl M, Sommer T, Hatzigeorgiou A, Doering S, Posselt D, Schreiner T (2002) SNNS, Stuttgart Neural Network Simulator, User manual. University of Stuttgart, Comp. Scienc Dept
- [31] Joachims T (2002) SVM-light V4.00: Support Vector Machine classification and learning modules. Technical report, Cornell University <http://svmlight.joachims.org/>
- [32] Maldonado JA, Robles M, Cano C (2001) Integration of healthcare information systems: application of CEN/TC251 ENV 13606. In IEEE, editor, Proceedings of the 23rd Annual International Conference of the IEEE 2001(4), pp. 3731–3734
- [33] May DA, Good RB, Smith PTW D K (1997) MR imaging of musculoskeletal tumors and tumors mimickers with intravenous gadolinium: experience with 242 patients. Skeletal Radiology 26: 2–15
- [34] Underwood J, Tate AR, Luckin R, Majós C, Capdevila A, Howe F, Griffiths J, Arús C (2001) A Prototype Decision Support System for MR Spectroscopy-Assisted Diagnosis of Brain Tumours. In VP et al, editor, MEDINFO 2001, pp. 561–565. IMIA
- [35] Abdolmaleki P, Buadu LD, Naderimansh H (2001) Feature extraction and classification of breast cancer on dynamic magnetic resonance imaging using artificial neural network. Cancer Lett 171(2): 183–191
- [36] Garcia-Gomez JM, Vidal C, Císcar AJ, Martí-Bonmatí L, Galant J, Robles M (2002) Pattern classification techniques applied to soft tissue tumors using MR and epidemiological data. In Proceedings 19th annual meeting of the European Society for Magnetic Resonance in Medicine and Biology, pp. 274–275
- [37] Vidal C, Garcia-Gomez JM, Marti-Bonmati L, Robles M, Galant J, Millet J (2002) Study of vascular histological group tumors using pattern recognition techniques. In INFORMED, pp. 207–211

Chapter 3

The influence of combining two echo times in automatic brain tumor classification by Magnetic Resonance Spectroscopy

^1H MRS is becoming an accurate, non-invasive technique for initial examination of brain masses. We investigated if the combination of single voxel ^1H MRS at 1.5T at two different TE, Short TE (PRESS or STEAM, 20-32 ms) and Long TE (PRESS, 135-136 ms), improve the classification of brain tumors with respect to using only one echo time. A clinically validated dataset of 50 Low-grade meningiomas (MEN), 105 Aggressive tumor: GBM and MET (AGG), and 30 Low-Grade Glial (LGG) was used to fit predictive models based on the combination of features from Short and Long TE spectra. A new approach that combines the two echo times consecutively was used to produce a single data vector from which relevant features of the two TE spectra could be extracted by means of three algorithms: Stepwise, ReliefF, and Principal Components Analysis. Least Squares Support Vector Machines and Linear Discriminant Analysis were applied to fit the pairwise and multiclass classifiers, respectively. Significant differences in performance were found when Short TE, Long TE or both spectra combined were used as input. In our dataset, to discriminate MEN, the combination of the two TE acquisitions produced optimal performance. To discriminate AGG from LGG, the use of Short TE acquisition alone was preferable. The classifier development strategy used here lends itself to automated learning and test performance processes, which may be of use for future web-based multicentric classifier development studies.

This chapter has been published as journal paper in [1].

3.1 Introduction

Brain tumors are the second fastest growing cause of cancer deaths among people older than 65 years^a. Nowadays, the diagnosis and treatment of brain tumors is based on clinical observations, radiological appearance, and often a histopathological diagnosis of

^aBTS. The Brain Tumor Society Accessed: January 14, 2008. <http://www.tbts.org>.

a biopsy. Information provided by non-invasive radiological techniques is highly relevant for evidence-based medicine applied to diagnosis, patient management, and surgery [2].

MRI is widely used for determining tumor extension in surgical and radiotherapy planning. MRI provides a neuroradiological classification with a sensitivity of 14%-100% depending on the tumor type and the precision of the wording used by radiologist [3]. Proton Magnetic Resonance Spectroscopy (^1H MRS) is slowly becoming an additional, accurate, non-invasive technique for initial examination of brain masses [4, 5]. This is due to its capability to provide useful chemical information about different metabolites to characterize brain tumors and its complementary role to MRI [4, 6].

At present, TE used in in-vivo ^1H MRS by most studies range between 18 and 288 ms. A spectrum acquired with a TE < 45 ms is usually considered a Short TE spectrum, and a Long TE spectrum otherwise. Different criteria have been argued in favor and against every option, while a few studies [7–11] and multicentric projects^b have defined protocols based on the acquisition of both Short TE and Long TE spectra from the same patient.

Short TE (20-35 ms) ^1H MRS allows to observe several metabolites and other compounds considered useful for tumor classification. Macromolecules (MM; 5.4ppm, 2.9ppm, 2.25ppm, 2.05ppm, 1.4ppm and 0.87ppm) are mostly observed in a Short TE spectra [2], Myo-Inositol (mI) and Mobile Lipids (ML) are better detected in Short TE [7]. Single voxel (SV) Short TE ^1H MRS is fast (typically 5 min) and robust, so it is very useful for clinical studies [2, 12]. However, Short TE signals show a large number of overlapping peaks, a strong MM-/ML-originated baseline and a certain sensitivity to artifacts [13].

Long TE (about 135 ms) ^1H MRS is less informative than Short TE, because resonances with short T2 may be lost [14]. However, Long TE signals are easier to analyze than Short TE signals [13]. Lipid resonances (1.3 and 0.9ppm) and MM will not be the dominating components at Long TE, making possible the study of the contributions of lactate (Lac, doublet at 1.33ppm) and alanine (Ala, doublet at 1.47ppm) as inverted peaks [2, 15].

Many applications of pattern recognition (PR) to cancer research have been reported in the last two decades [16] and, more specifically, in brain tumor research [5, 17, 18]. In [17], Hagberg summarizes classification of brain tumors with MRS based on pattern recognition and clustering methods. Eight of these studies were applied to brain tumor discrimination from normal tissue or other Central Nervous System (CNS) diseases. All of them were based on Linear Discriminant Analysis (LDA) or Artificial Neural Network (ANN) applied to relative metabolite levels or Principal Components Analysis (PCA) transformations, and they were all evaluated by Leave-One-Out cross-validation. More recent publications have also described results for classification of brain tumors based on the MR data available within the the INTERPRET project (INTERPRET) project [5], where linear and kernel-based methods on MRS features extracted by automatic procedures were applied [19, 20]. While [13] was focused on the classification using Long TE ^1H MRS, publications such as [5, 19, 21, 22] carried out experiments with Short TE ^1H MRS. Based on Least Squares Support Vector Machines (LS-SVMs) [23], Devos, Lukas et al. in [13, 22] developed different classifiers for in-vivo Short TE spectra and Magnetic Resonance Spectroscopic Imaging (MRSI) with good performance. Menze et al. [18] published an extensive benchmark study of quantitation and PR based feature extraction methods combined with learning strategies to discriminate between recurrent and non-recurrent

^bInterpret acquisition protocols 2000. <http://azizu.uab.es/INTERPRET/mrsdata/mrsdata.html>.
eTUMOUR acquisition protocols 2003. <http://www.etumour.net>

brain tumors using Long TE ^1H MRS. They reported that the PR methods perform at least as well as the ones based on manual quantitation (5%-10% higher accuracy).

To date, nobody has combined Short and Long TE spectra in a single classifier for predictive purposes by means of automatic systems. Therefore, our aim has been to improve the automatic classification of brain tumors combining the complementary views of the chemical composition of brain tumors offered by the Short and the Long TE spectra. Majos et al [11] compared the Long TE and Short TE discrimination capacity in clinical use. Based on Majos' results, we are interested in measuring and characterizing if an improvement in automatic tumor classification could be achieved by combining both spectra in a PR model. For this purpose, multiclass and binary classifiers were prepared using the following approaches: Short TE only, Long TE only, and Long TE and Short TE combined. Multiclass and pairwise classifiers for Low-Grade Meningioma (MEN) [3, 5], AGG, and LGG, were prepared to test the Combined approach, in contrast to the Long TE and Short TE approaches. All classifiers were PR based models, that attempted to extract the maximum information of the physiological state of the tissue directly from the pattern of the combined view of the two echo times, instead of resonance line quantifications followed by other approaches [24]. Finally, LDA and LS-SVMs classifiers were based on features extracted from the combined spectra by means of SW, PCA and ReliefF methods (see Section 1.1).

The results of this study may be of interest in determining which acquisition conditions and stages should be taken into account in the development of Decision-Support Systems for in-vivo brain tumor diagnosis.

3.2 Experimental

3.2.1 Data acquisition and pre-processing

Table 3.1: Samples per class used in the study. Short TE and Long TE ^1H MRS data were acquired during INTERPRET according to a consensus protocol

Superclass	Class	Samples	Total (n)
Aggressive (AGG)	Glioblastoma Multiforme (GM)	77	105
	Metastasis (MET)	28	
Low-grade meningioma (MEN)	Low-grade meningioma (MEN)	50	50
Low-grade glial mixture (LGG)	Astrocytoma grade II (A2)	20	30
	Oligodendroglioma (OD)	5	
	Oligoastrocytoma (OA)	5	
TOTAL		185	185

The combination of Short TE and Long TE was evaluated using a dataset provided by six international centers in the framework of the INTERPRET project. Table 3.1 shows a summary of the histopathological classification of tumors included in the study and agreed by a panel of consulting pathologists, based on the histological classification of the CNS tumors set up by the World Health Organization [25]. The cases were grouped into three diagnostic superclasses: AGG, composed of Glioblastoma (GBM) and Metastases (MET),

with 105 patients; MEN, with 50 patients; and LGG, composed of Astrocytomas grade II (A2), Oligodendrogliomas (OD), and Oligoastrocytomas (OA), with 30 patients. The GBM cases were in the age range 9-80, and with mean of 65 years, the MET cases were in the range 29-80, with a mean of 55. With respect the MEN class, all the cases were in the range 39-78, and with a mean of 55 years. For the LGG samples, the range was 18-61 with a mean of 25 years.

There were 185 cases in the INTERPRET database [26] with both SV ^1H MRS at Short TE and at Long TE. Data were acquired^c with SV ^1H MRS at 1.5T, avoiding areas of cysts or necrosis and with minimum contamination from the surrounding non-tumoral tissue. Volume of interest size ranged between $1.5 \times 1.5 \times 1.5 \text{ cm}^3$, (3.4 mL) and $2 \times 2 \times 2 \text{ cm}^3$, (8 mL), depending on tumor dimensions. The aim was to obtain an average spectroscopic representation of the largest possible part of the tumor. In the acquisition of Long TE spectra, the PRESS sequence was used, with a recycling time (TR) between 1500 and 2020 ms, TE of 135 or 136 ms, spectral width of 1000 or 2500 Hz and 512 or 2048 data points. Short TE were acquired using PRESS or STEAM sequences, with TR between 1600 and 2020 ms, TE of 20 or 30 ms, spectral width of 1000 or 2500 Hz and 512 or 2048 data points. Each spectrum used for training and testing had passed a quality control based on the validation of both Long TE and Short TE spectra by the INTERPRET Quality Control (QC) committee [27].

Long TE and Short TE raw data were automatically processed following the well-established ^1H MRS data pre-processing protocol [5] and the quality control procedure developed during the INTERPRET project. Briefly, water-suppressed free induction decays (FIDs) were corrected for Eddy current effects using the Klose algorithm [28]. An expert spectroscopist reviewed the signal obtained after the Klose algorithm, and, for a few cases, manual phase correction was applied. Then, the residual water resonance was removed using the Hankel-Lanczos Singular Value Decomposition (HLSVD) time-domain selective filtering. The resulting signal was transformed to the frequency domain by a Fast Fourier Transform (FFT) with zero-filling to the maximum length of the dataset and a line broadening of 1Hz was applied. The spectra obtained consisted on 512 values from -2.7ppm to 7.1ppm , for this, linear interpolation was applied to have the same frequency sampling for all the signals when needed.

The normalization of the spectral data vectors to the L2-norm was performed based on the data points in the region $[-2.7, 4.33] \cup [5.07, 7.1]\text{ppm}$. The normalization of Short TE and Long TE spectra of each patient were made separately. This approach converts the spectrum in a ratio between the value of each ppm and the whole region; the normalization with respect to the combined spectra was not considered for not translating the influence of high peaks of lipids in Short TE to Long TE in the aggressive class. Normalization with respect to the unsuppressed water peak area was also considered, but preliminary experiments did not show better performance than the results based on L2-norm.

Afterwards, the baseline offset, estimated as the mean value of the region $[11, 9] \cup [-2, -1]\text{ppm}$, was subtracted from the spectrum. A frequency alignment of the spectrum was made referencing the ppm-axis to the total creatine (Cr) at 3.03ppm ; or the choline (Cho) containing compounds at 3.21ppm ; or the ML at 1.29ppm , in order of priority and depending on the signal-to-noise ratio (SNR) and the tumor pattern. Finally, the region

^cInterpret acquisition protocols, 2000. <http://azizu.uab.es/INTERPRET/mrsdata/mrsdata.html>.

of interest was restricted to the region $[0.5, 4.1]ppm$, obtaining a vector of 190 points for each spectrum. See [5, 27] or the INTERPRET web-page^d for further details.

The combination of the Long TE spectrum with the Short TE spectrum was done after the full pre-processing of each spectrum as independent signals.

3.2.2 Combination of the Short TE and Long TE spectra

For this study, we were interested in producing a combination of the Short TE and the Long TE spectra without introducing any a priori restriction or assumption of relationship between them.

For combining the signals, different approaches might have been taken into account, such as ratios between the area-based quantification in Short TE and Long TE of each metabolite, e.g. the calculation of R_2 relaxation rates of metabolites used in [29]. For these approaches, assumptions about the metabolites that appear in the spectra are done when quantifying, for this reason they restrict the information used for classification to the well-known metabolites that appear in both echo times.

Nonetheless, a PR approach for the combination of spectra was followed, i.e., the selection of features is based on their discrimination properties for each classification task. For this reason, we considered the more direct approach, the better, and the Long TE and Short TE vectors were combined by placing the vector of D_s points of the Short TE spectrum after the vector of D_l points of the Long TE spectrum (see Figure 3.1). In posterior PR steps, this joint vector, that represents each case of the dataset, is treated as a $(D_s + D_l)$ -dimensional vector-valued observation of the distribution of the diagnosis. Then, the discrimination functions may choose simultaneously among the features from both spectra to solve the proposed prediction model.

As a result, the Combined TE dataset of 185 samples with 380 data points was obtained. To compare the combined approach with single approaches, we also generated the Short TE dataset composed by 185 samples with the 190 values in the region of interest of the Short TE spectrum, and the Long TE dataset of 185 samples with the 190 values in the region of interest of the Long TE spectrum.

3.2.3 Multiclass and binary classifiers with both spectra

Multiclass and binary classifiers were applied to characterize the behavior of the Combined TE (Long TE + Short TE), in comparison to approaches based on the Short TE or Long TE alone. Two procedures were applied for dimensionality reduction in the multiclass classifiers: the SW procedure based on LDA and PCA-based feature extraction with a LDA classifier. Two procedures were also applied to the binary classifiers: the full spectra used directly in LS-SVM and the ReliefF feature selection procedure with LS-SVM.

Multiclass classifiers

For multiclass classification, the Linear Discriminant Analysis (LDA) technique was chosen (see Section 1.1.1). This method has demonstrated its usefulness in many biomedical applications, having been applied to MRS Decision-Support Systems for brain tumor

^dINTERPRET Consortium. Interpret web site Accessed: January 14, 2008. <http://azizu.uab.es/INTERPRET>.

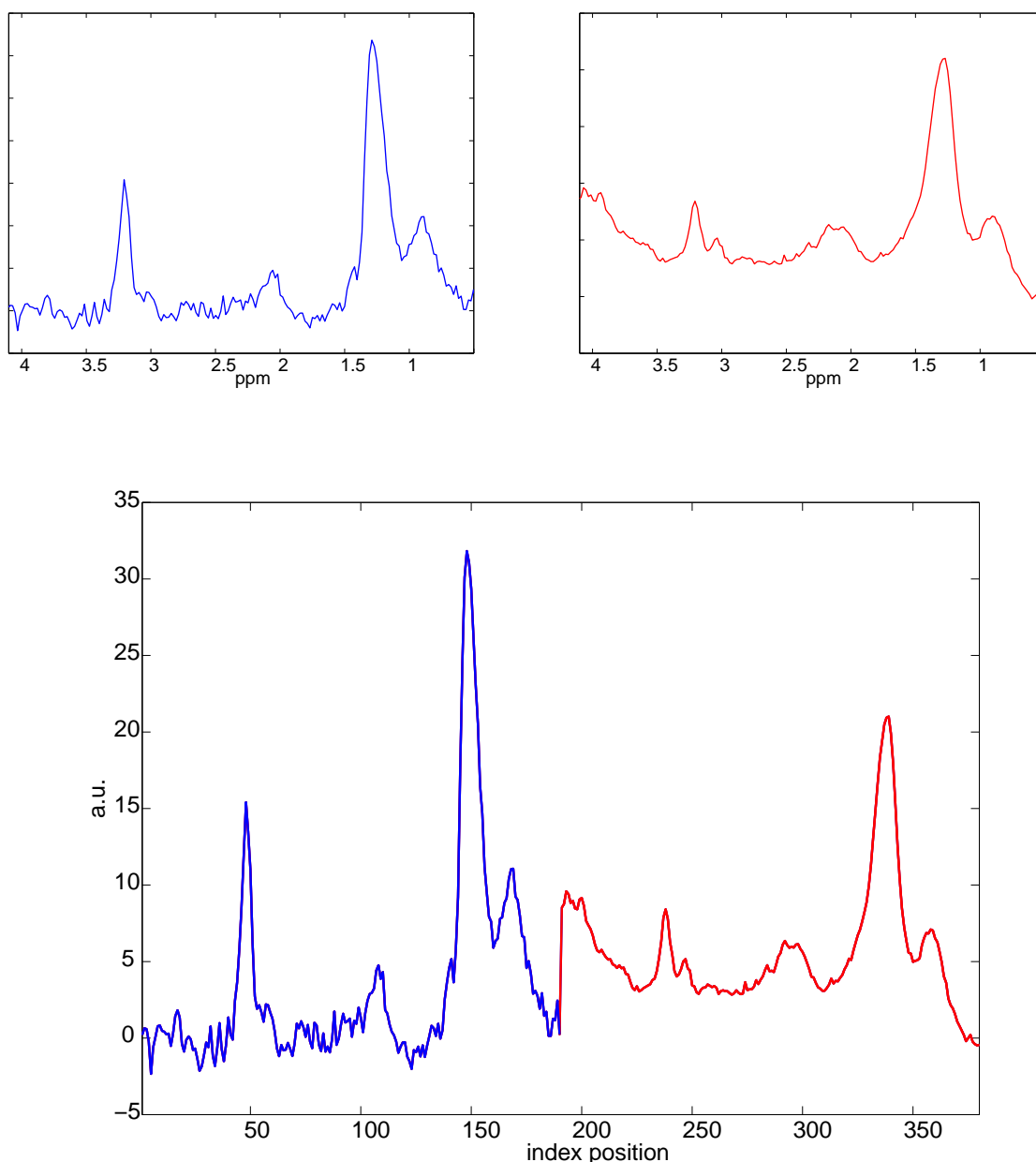


Figure 3.1: The top row (x-axis in ppms, y-axis in arbitrary units (a.u.)) shows the Long TE spectrum (left) and the Short TE spectrum (right) of INTERPRET case I0016 [26] corresponding to a glioblastoma. The second row shows the combined spectra (x-axis in relative position of points in the new data vector, y-axis in a.u.).

diagnosis [5, 24]. One of the advantages of the method is the possibility of plotting the latent space where the variables are projected. When a multiclass task of three classes is solved, the latent space is bi-dimensional (2D), and it could be used to visualize the projection of the samples in a 2D plot.

Before applying LDA, the input space should be reduced in a proper way; for this study we used two methods, the SW algorithm and the PCA for feature selection and

extraction (see Section 1.1.2).

Binary classifiers

The binary classifiers developed in this study were based on LS-SVMs (see Section 1.1.1). Their regularized property allowed us to use the full spectral vector in the classifier development, and compare this with the use of sets of features selected by the ReliefF algorithm (see Section 1.1.2).

In order to compare the Combined TE approach to the Short TE and the Long TE approach, LS-SVM classifiers using the full vector, 10, 50 or 100 features selected by means of the ReliefF algorithm were used.

Evaluation procedures

k-Random Sampling Train-Test (kRSTT) (see Section 1.1.3) with stratified test sets with 150 repetitions was the evaluation procedure used for all the reported experiments. The partitions for repetitions were random and independent among the experiments with the training set composed by 70% of cases of each class. The evaluation was carried out in a nested-loop that covered the feature and model selection in order to avoid underestimation of the true error. Besides, the Accuracy and the AUC of the ROC (see Section 1.1.4) were the evaluation metrics used in this study. In our experiments, credibility intervals (see Section 1.1.4) of the accuracy ($1 - \tau$) were reported in percentage; estimated from the kRSTT evaluation assuming a test size of one repetition.

The significance of differences among the Short TE, the Long TE and the Combined TE approaches were tested by means of Friedman's (F) nonparametric two-way analysis of variance test. The test was applied to the best models of the Combined TE approach and their equivalent models for Long TE approach and Short TE approach for each binary classification. The comparison of the best models for each binary classification was done by means of the Kruskal-Wallis (KW) nonparametric one-way analysis of variance. When multiple comparisons were carried out, the results were analyzed using Tukey's (MCT) honestly significant difference criterion.

3.3 Results

3.3.1 Multiclass classifiers

Multiclass classifiers discriminate among the three aforementioned superclasses (AGG, MEN, and LGG) simultaneously. Two alternatives, SW+LDA and PCA+LDA, were applied to compare the Combined approach with Short TE or Long TE based classifiers.

Table 3.2 shows the kRSTT evaluation of the SW+LDA approach of the multiclass classifiers applied on the Combined TE, the Short TE, and the Long TE datasets. Based on the predictions achieved by the different classifiers, some of the cases from the datasets were singled out for critical review by experts. A case was considered as a potential outlier when most of the classifiers failed the prediction when testing the case during the evaluation process (e.g. less than 20% of successes by at least one of the three approaches in our study). 13 cases were considered potential outliers, four of them are discussed in section 3.4.1.

Table 3.2: Results for the multiclass classification among AGG, MEN and LGG. kRSTT evaluation of the SW approach based on the Ability to Separate (AS) of the predictive Gaussian linear classifier.

Dataset	Accuracy (%)	[Credibility Interval] (%)
Combined TE	88.7	[83.6,92.7]
Short TE	88.8	[83.7,92.8]
Long TE	82.5	[76.6,87.5]

Figure 3.2 shows the accuracy of the PCA+LDA classifiers applied to the Combined TE, the Short TE, and Long TE datasets estimated by kRSTT. The x-axis, the accumulated number of PCs varied from 2 to 15 and in y-axis the accuracy of the PCA+LDA classifiers varied from 0.7 to 1. In general, it was observed that the combined approach achieved the best performance of all the models, when accumulating 7 or more PCs extracted from both Short TE and Long TE spectra.

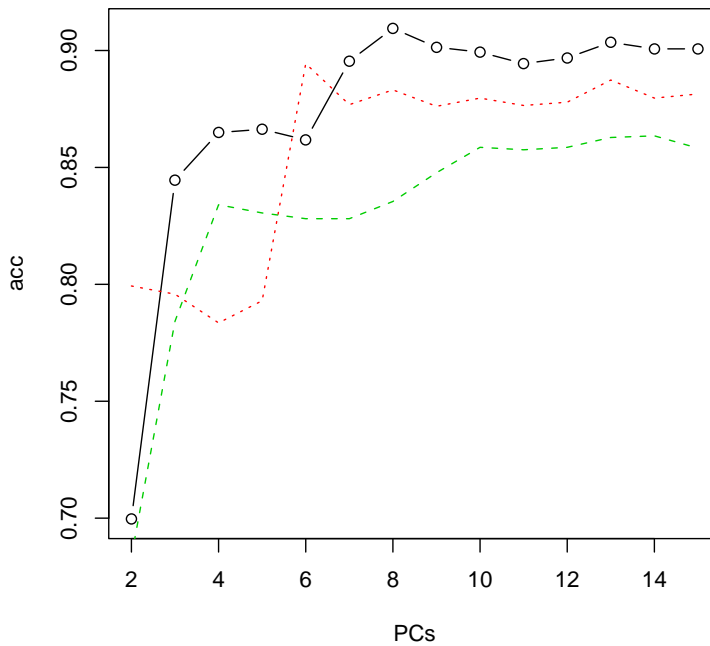


Figure 3.2: The accuracy (acc) of the multiclass approaches using PCA+LDA vs. number of principal components (PCs). The Combined TE approach (black solid line and empty circles) follows the Long TE pattern (green dashed line) in the 3-6 PC region. When more PCs were incorporated, the Combined TE approach improved due to the contributions of the Short TE (red dotted line) and the Long TE.

3.3.2 Binary classifiers

Binary classifiers based on LS-SVMs and ReliefF feature selection were applied to the three pairwise combinations of the superclasses (AGG vs. MEN, AGG vs. LGG, and LGG vs. MEN).

Table 3.3: kRSTT evaluation of the LS-SVM for pairwise classification of AGG, MEN and LGG classes. The percentage of Long TE (LTE) features selected by ReliefF with respect to the total number of features is shown in brackets in the features columns. CI is the credibility interval of the Accuracy.

Task	Dataset	Features	Accuracy [CI] (%)	AUC
AGG vs. MEN	Combined TE	380 [LTE:50%]	95.3 [91.2,97.8]	0.992
	Short TE	100	92.6 [87.8,96.0]	0.982
	Long TE	190	92.2 [87.3,95.7]	0.975
AGG vs. LGG	Combined TE	10 [LTE:0%]	92.6 [87.3,96.1]	0.970
	Short TE	10	92.1 [86.7,95.7]	0.966
	Long TE	10	90.5 [84.8,94.6]	0.95
LGG vs. MEN	Combined TE	50 [LTE:42%]	97.5 [92.6,99.3]	0.996
	Short TE	50	96.0 [90.3,98.7]	0.993
	Long TE	100	94.5 [88.2,98.0]	0.993

Table 3.3 shows in each row the best models obtained for each pairwise classification task by means of the Combined TE, the Short TE and the Long TE approaches. For each result, we report the accuracy with its credibility interval and the AUC of the ROC to compare the behavior in Sensitivity and Specificity of the model. For the best model of the Combined approach, we report in the Features column the percentage of Long TE features used by the model with respect to the total number of features.

An overview of Table 3.3 indicates that the Combined TE approach is the best model when discriminating MEN from AGG or LGG. On the other hand, when discriminating LGG from AGG the Combined TE approach obtains the same performance as the Short TE approach. This may be because the ReliefF algorithm selected a combination of 10 features originating from the Short TE spectra. Table 3.4 summarizes the results of a statistical significance test of the difference among the three models, in which the Friedman test of the Combined TE, the Short TE, and the Long TE approaches reveals significant differences among them. In the pairwise comparison, the Combined TE and the Long TE are significantly different ($p < 0.05$, Friedman’s test); the Combined TE and the Short TE were not significantly different from each other ($p > 0.05$); and the Short TE and the Long TE models were not significantly different either ($p > 0.05$). When comparing the best models of the AGG vs. MEN binary classifiers, the Combined TE approach outperformed the Short TE and Long TE approaches ($p < 0.05$, Kruskal-Wallis test); however, the differences in performance with respect to the Short TE and the Long TE approaches were not statistically different ($p > 0.05$).

The Sensitivity, Specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV) have displayed exactly the same behavior than the accuracy and the AUC for each comparison task. Namely, when discriminating MEN from any other class, there is an increase of the four performance measurements with the Combined TE approach.

Table 3.4: Summary of the significance of difference tests among different approaches. Friedman’s (F) nonparametric two-way analysis of variance test, Kruskal-Wallis (KW) nonparametric one-way analysis of variance and Tukey’s (MCT) honestly significant difference criterion were used.

Significance test ($p < 0.05$)	Three classifiers	AGG vs. MEN	AGG vs. LGG	LGG vs. MEN
	Yes (F)	Yes (KW)	Yes (KW)	Yes (KW)
Combined TE vs. Short TE (MCT)	No	Yes	No	Yes
Combined TE vs. Long TE (MCT)	Yes	Yes	Yes	Yes
Short TE vs. Long TE (MCT)	No	No	Yes	Yes

When discriminating AGG from LGG, the Sensitivity, Specificity, PPV and NPV of the Combined TE approach achieve the same performance as the Short TE approach.

3.4 Discussion

When comparing the multiclass classifiers based on the SW feature selection and LDA, the behavior of the Combined TE approach and the Short TE approach is similar. This may be because in the kRSTT repetitions, the most contributing features selected by the algorithm were mainly from the Short TE spectrum. The accuracy of the Long TE approach is considerably lower than the accuracy of the other two approaches. Even though some Long TE features were selected for the Combined TE models, they did not contribute as much as the Short TE features did.

Figure 3.3 shows the in-vivo Long TE (left) and Short TE (right) average spectra [5] for AGG (solid line), LGG (dotted line) and MEN (dashed line) diagnosis. The first row shows in black dashed-dotted line the number of selected points around each ppm in a $0.1ppm$ window when the SW algorithm is applied to the Combined TE approach. The second row shows in black dashed-dotted line the number of selected points around each ppm in a $0.1ppm$ window when the SW algorithm is applied to the Long TE approach (left) and the Short TE approach (right). It is clearly shown how ppm with contributions of the main metabolites and molecules reported in the MRS literature for brain tumor typing and grading are frequently selected (e.g. Cr, Cho, NAA, Lipids). Both Combined and Single approaches follow similar behaviour for selecting Short TE and Long TE features. Namely, ML and Lac at about $1.3ppm$, alanine at about $1.5ppm$, Glx in the $[2, 2.5]ppm$ range, total Cr at about $3ppm$, Cho containing compounds at about $3.2ppm$, taurine at about $3.4ppm$, mI/Gly at about $3.55ppm$ and Glx/Ala at about $3.76ppm$, in agreement with previous data from extracts [30], biopsies [31] or other in vivo data [2, 32]. Comparing the frequencies of selection in Long TE and Short TE, it is seen that the selection in Long TE is totally focused in the well-known metabolites, but in Short TE the selection is more sparse, confirming the contributions of mixtures of metabolites and molecules useful for classification along the region of interest.

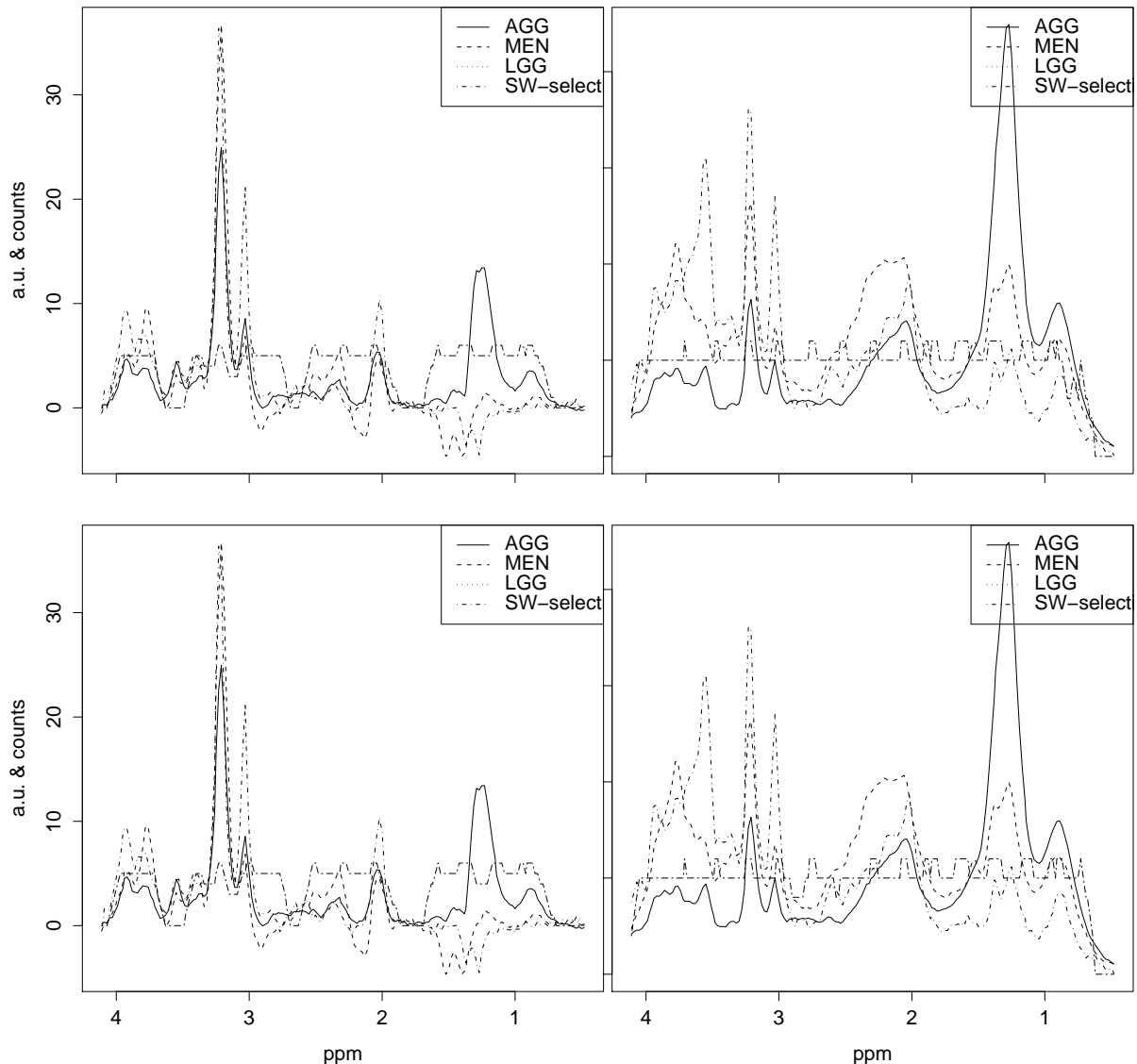


Figure 3.3: In-vivo Long TE (left) and Short TE (right) average spectra [5] for AGG (solid line), LGG (dashed-dotted line) and MEN (dashed line) diagnosis. The first row shows in black dashed-dotted line the number of selected points around each ppm in a 0.1ppm window when the SW algorithm is applied to the Long TE approach (left) and the Short TE approach (right). The second row shows in black dashed-dotted line the number of selected points around each ppm in a 0.1ppm window when the SW algorithm is applied to the Combined TE approach.

When the PCA+LDA methodology was applied, the expected accuracy using the combined spectra was, in general, higher than the accuracy observed when using only the Short TE or the Long TE datasets. It is clearly shown (Figure 3.2) that the combined approach achieved the best performance when accumulating 7 or more PCs extracted from both Short TE and Long TE spectra. In a more detailed analysis of Figure 3.2, we can distinguish three regions in the accuracy curve of the Combined TE approach: 1) when the two first PCs were used, its accuracy was lower than the one achieved by the Short TE approach; 2) when 3 to 6 PCs were cumulatively added in to the model, its

behavior followed the Long TE model, but it had better accuracy due to the contribution of Short TE; 3) when 7 to 15 PCs contributions were cumulatively added, the Combined TE performance increased significantly, immediately after the increase obtained in the Short TE approach at 6 PCs. The Combined TE approach achieved the best accuracy already with 8 PCs. In summary, the main discriminative contribution of the Long TE was accumulated by the PCA in the first PCs of the Combined TE model, while the main discriminative contribution of the Short TE was accumulated in posterior PCs. The good behavior of the Combined TE approach highlights the fact that the simultaneous use of features of the two spectra may improve the predictive capacity of automatic brain tumor classification.

With respect to the binary classifiers, in the AGG vs. MEN classifiers, the Combined TE approach outperformed both the Short TE and Long TE approaches with a significant difference. Both Short TE only and Long TE only models produced good results individually. Hence, the results obtained by the combination classifier were to be expected. In the AGG vs. LGG classifiers, the best Combined TE model used 10 features, from which any of them were Long TE features in any of the 150 repetitions. This may be due to a better individual discriminatory power of the Short TE features, since the ReliefF algorithm selected Short TE features before Long TE features. In the LGG vs. MEN classifiers, the Combined TE approach outperformed the Short TE approach and the Long TE approach with significant difference. Significant differences were also obtained when comparing the Short TE and the Long TE approaches. All the models for distinguishing LGG vs. MEN produced similar good behavior. A large percentage of the Long TE features were used in the 50- and 100-feature models, where improvements were obtained with respect to the Short TE and Long TE models. Both Short TE only and Long TE only models produced good results individually. Hence, the results obtained by the combination classifier were expected.

Our results show that the combination of both Short and Long TE obtains the best results when discriminating between MEN and the other two superclasses (AGG, LGG). On the other hand, the acquisition of Short TE is sufficient to distinguish between AGG and LGG. Besides, Sensitivity, Specificity, PPV and NPV of the models follow exactly the same behaviour than the discussed accuracy and AUC. These results are consistent with those described in Majos et al. [11] who evaluated the possible improvement of using two echo times spectra for clinical classification of brain tumors. In their study, tumor classification was slightly better using Short TE than Long TE, but they obtained the best results when the classification obtained at the two TEs agreed. Also, a preliminary study carried out by Tortajada et al [33] on the INTERPRET database showed an improved accuracy using the Combined TE approach after using LDA and SW feature selection methods for descriptive purposes.

3.4.1 Discovery of outliers by the automatic classifiers

Further analysis of some misclassified cases showed abnormal profiles, although they were validated as usable during INTERPRET. Here we discuss four interesting cases detected as potential outliers by the SW-based multiclass classifiers.

There was one MEN case (I0009, see Figure 3.4, top-left) that showed an abnormal profile for a low-grade meningioma with a clear pattern of lipids at 1.3ppm and 0.9ppm.

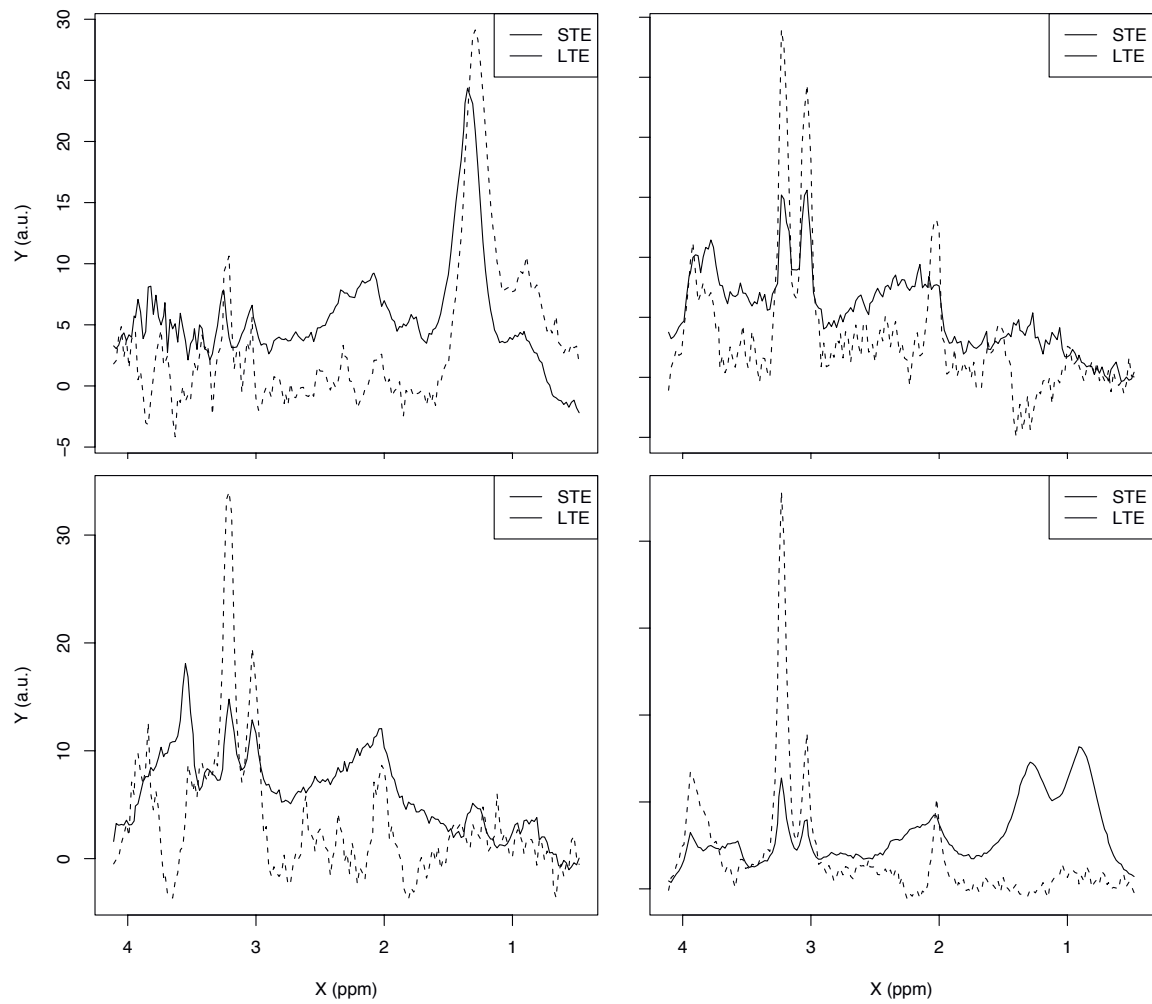


Figure 3.4: Four potential outliers detected by the multiclass classifiers are shown here. The I0009 (Top-left) was validated as MEN [26], the I1390 (top-right) was validated as glioblastoma. The I0063 (bottom-left) was validated as a glioblastoma, and the I0450 (bottom-right) was validated as an oligoastrocytoma. The Short TE is represented by solid lines and the Long TE is represented by dashed lines.

This case was a secretory meningioma in which the representative MRI with the voxel location showed that the possibility of a contamination by subcutaneous fat could not be ruled out (see Figure 3.5, left).

There was also a glioblastoma that was misclassified (I1390, see Figure 3.4, top-right), where normal tissue contribution could potentially contaminate the profile (see Figure 3.5, right). Moreover, phasing of the Short TE was suboptimal at about 1.3ppm while the Long TE showed an atypical profile for a glioblastoma (see Figure 3.4 and [5]), with a Cho/Cr ratio close to 1 and some visible NAA, which might be expected from compromised brain parenchyma but not from a high grade glial tumor.

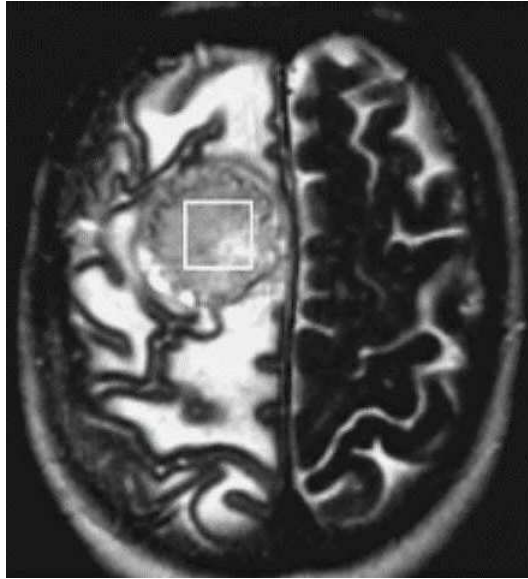
There was another glioblastoma (I0063, see Figure 3.4, bottom-left) that did not show the typical necrotic pattern with clear lipid signals at 0.9 and 1.3ppm that is normally found in most of these tumors. It showed a distinctive mI peak at 3.55ppm instead, which constitutes the strongest signal at Short TE. At Long TE, the signal at 3.55ppm disappeared, confirming the contribution of mI to this resonance [7]. The spectral pattern of this case closely resembled the average pattern of LGG tumors. Although it has also been pointed out [34] that this pattern could correspond to the less common subtype of secondary Glioblastomas, there are not enough representative data of this subtype available in the database for its proper distinction.

Finally, there was also an oligoastrocytoma (I0450, see Figure 3.4, bottom-right) that shows a clear signal from ML at Short TE which disappears at Long TE. Note that the intensity of the 0.9ppm peak was equal to the intensity of 1.3ppm peak, which is fairly unusual. The Mobile lipids (ML) profile observed at Short TE disappeared at Long TE. The reason for this case appearing as a misclassified outlier would be that the Short TE and the Combined TE models were deceived by its atypical, pseudo-necrotic, ML profile. This profile is characterized by a pattern in which the most intense peaks are those of methyl (0.9 ppm) and methylene (1.3 ppm), being of the same amplitude and height, or even with a methyl of higher intensity. On the other hand, Long TE classifiers had much better success (89.4%) in correctly classifying this case. This reinforces the complementary nature of the Short and the Long TE information. It also points to the fact that, in certain instances, such as artifacts or uncommon spectral patterns, it may be necessary to resort to the single TE based prediction results.

3.4.2 Possible use of the combined approach in a Decision Support System

The incorporation of the optimal classifier in a DSS may be of interest to clinicians to reinforce their confidence in the diagnosis of common cases, as well as in cases where the diagnosis is non-typical. An interesting 2D visualization of the "cases universe" [5] for descriptive purposes can be obtained from the latent space provided by the LDA multiclass classifier. Figures 3.6, 3.7, 3.8 show the LDA latent space for the best classifiers obtained by each of the approaches in the kRSTT procedure. Every misclassification of Long TE (cases #8-12 in the figure) is correctly projected on the Combined TE plot. Besides, only two misclassification of Short TE (cases #3,#7) are misclassified in the combination of both TE. Therefore, the projection of the best model obtained from the Combined TE approach provides a visually better discrimination of the training set cases than the best Short TE or Long TE best models. If a new case maps close to typical neighbors,

I0009, Meningioma



I1390, Glioblastoma

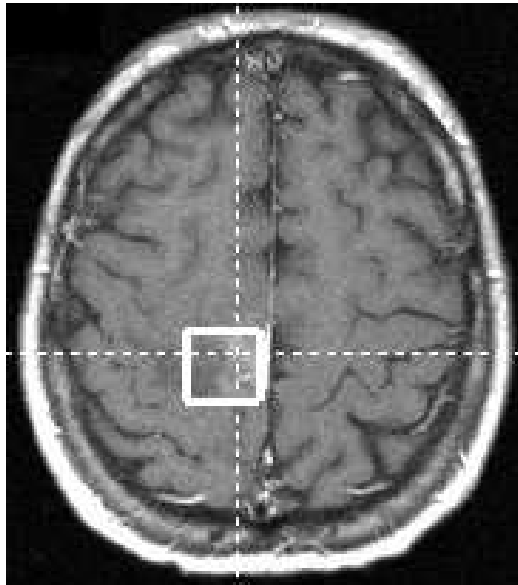


Figure 3.5: MRI scans and voxel localization of potential outliers I0009 (Top, T2-weighted scan) and I1390 (Bottom, T1-weighted post-contrast scan). I0009 case was validated as meningioma and the superior limit of the voxel was close to the calotte, a situation in which contamination by subcutaneous fat is likely to occur. I1390 was validated as glioblastoma and the voxel included an important percentage of non-enhancing parenchyma, apparently normal, which could produce a mixed tumoral/non-tumoral MRS profile.

confidence in the diagnosis will be reinforced. If a new case maps close to non-typical neighbors, new doubts may arise leading to a better diagnosis.

3.4.3 Final considerations

There are still many remaining challenges in brain tumor classification by ^1H MRS. One of the most important ones is the limited number of available spectra per tumor type [3]. This is due to the specific epidemiological distribution of tumors, with classes that have extremely low prevalence [35], and the increasing recognition of brain tumor molecular subtypes [36, 37]. For example, the inclusion of more Low-Grade Glial tumors in the training set might make possible to carry out studies to discriminate among astrocytomas, oligodendrogliomas, and oligoastrocytomas. Moreover, the presence of noise and artifacts in the spectra, even after the elimination of the dominating water peak, affects the "quality" of the information available. The Glioblastoma vs. Metastasis discrimination by means of the MRS is difficult with the use of SV spectroscopy. Approaches, such as MV spectroscopy or acquisition of an additional adjacent voxel to the brain mass should provide convenient information for distinguishing between these two types of tumors [38, 39]. In further works, we will study the predictive capacity of automatic brain tumor classification taking into account the quality of the training and the test samples, in order to characterize the robustness of the classifiers against the appearance of abnormalities in real environments. Two multicentric European projects, eTUMOUR and HEALTHAGENTS [40], are trying to improve the results obtained in the INTERPRET project. These projects include the compilation of an extended collection of ^1H MRS data from brain tumors, with ex-vivo data which may further increase the number of molecular tumor subtypes. Additionally, the ability to automatically train and mathematically validate tumor type classifiers that has been demonstrated in this work may facilitate the compilation of data from centers around the world, as targeted in the HEALTHAGENTS project. This should enable researchers to produce classifiers for tumor molecular subtypes that are currently beyond the reach of traditional databases.

3.5 Conclusions

This study focused on the use of Short TE and Long TE ^1H MRS to improve automatic brain tumor diagnosis of the most prevalent brain tumors in humans [3]. The LDA classifier for the three classes based on features extracted from both TE provided a new latent space representation useful for its incorporation into Decision Support Systems. Additionally, LS-SVM obtained high performance in every binary classification problem. Also, the regularization property of LS-SVM allowed the increase of the number of features in the combination of both TE. The most important finding of this study is that significant differences in the performance of automated classifiers generated with SV ^1H MRS data have been found. Complementary information was obtained by combining the two spectra, which can be used to improve the automatic classification of brain tumors following a PR approach. The Combination of Short TE and Long TE in a single classifier performed better in several instances than the use of one TE ^1H MRS only.

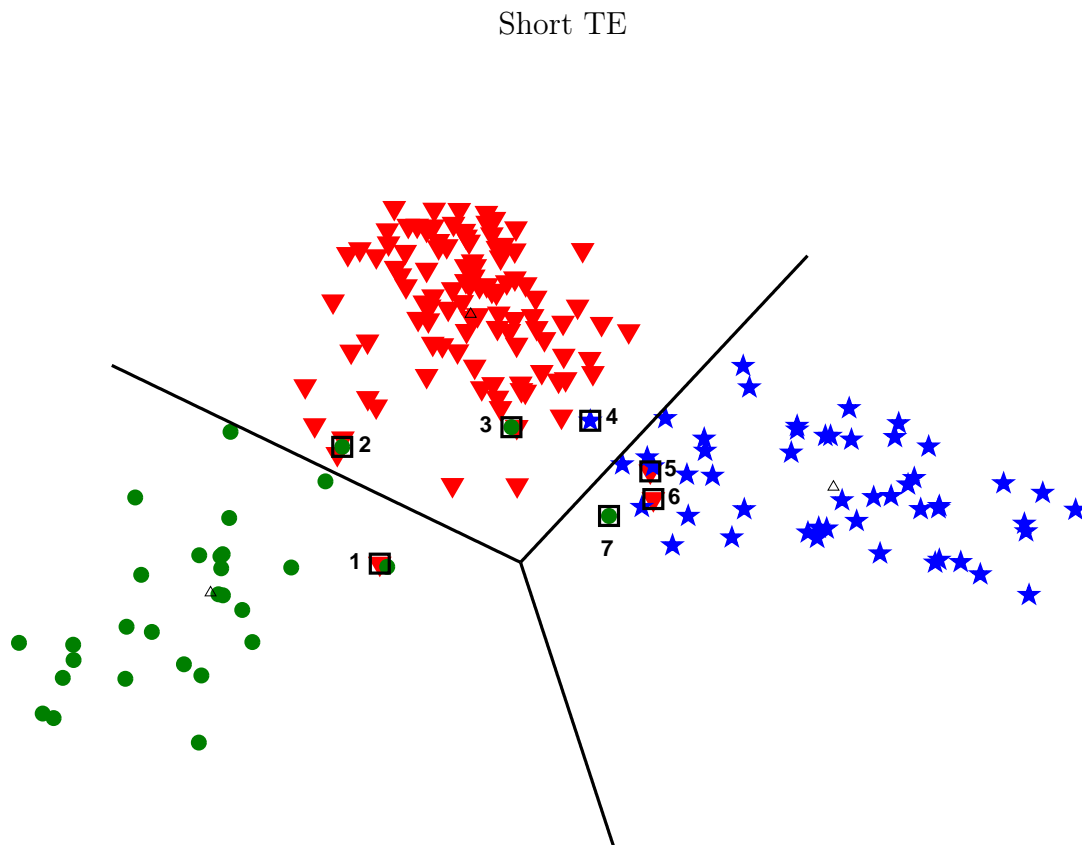


Figure 3.6: Projection of the SW+LDA latent space (or "universe of cases") for the best model obtained in the kRSTT procedure by the Short TE approach for the multiclass classification. A symbol represents a tumor sample projected in the latent space. In the projection, the space is divided into three different regions by the decision boundary and the black triangle labels the class centroid. These regions correspond to each brain tumor class: MEN (blue stars), AGG (red inverted triangles) and LGG (green circles). Any misclassified sample at Short TE is labeled with a black square (numbered #1-7).

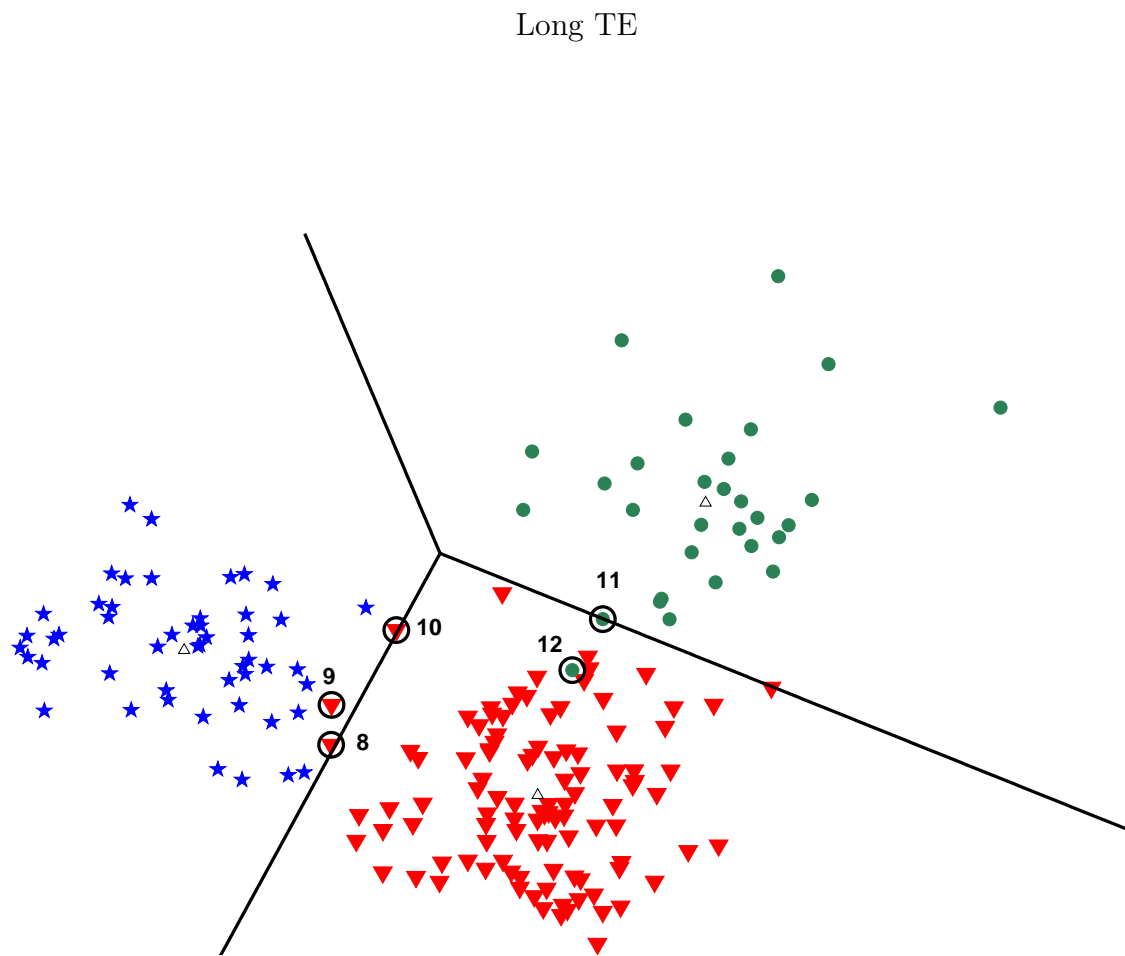


Figure 3.7: Projection of the SW+LDA latent space (or "universe of cases") for the best model obtained in the kRSTT procedure by the Long TE approach for the multiclass classification. Symbols are the same as in Figure 3.6. Any misclassified sample at Long TE is labeled with a black circle (numbered #8-12).

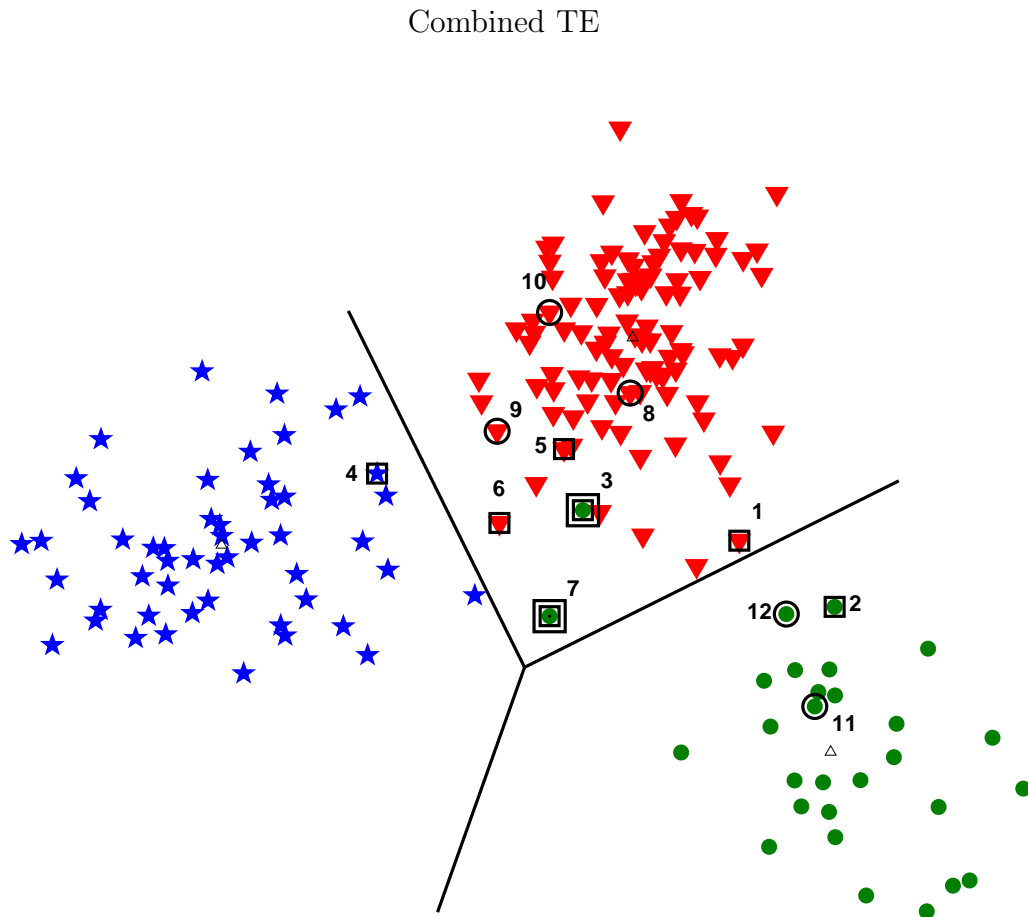


Figure 3.8: Projection of the SW+LDA latent space (or "universe of cases") for the best model obtained in the kRSTT procedure by the Combined TE approach for the multiclass classification. Symbols are the same as in Figure 3.6. Any misclassified sample at the Combined TE is labeled with a double black square(#3,#7). Every misclassification of Long TE (#8-12) (Figure 3.7) is correctly projected on the Combined TE plot. Besides, only two misclassifications at Short TE (#3,#7) (Figure 3.6) are misclassified in the Combination of both TE. Therefore, the projection of the best model obtained by the Combined TE approach provides a better visual discrimination than the best Short TE or the best Long TE models.

Bibliography

- [1] Garcia-Gomez J, Tortajada S, Vidal C, Julia-Sape M, Luts J, Van Huffel S, Arús C, Robles M (2008) The influence of combining two echo times in automatic brain tumor classification by Magnetic Resonance Spectroscopy. *NMR in Biomedicine* 21(10): 1112–1125
- [2] Howe FA, Opstad KS (2003) ¹H MR spectroscopy of brain tumours and masses. *NMR Biomed* 16(3): 123–131
- [3] Julia-Sape M, Acosta D, Majos C, Moreno-Torres A, Wesseling P, Acebes JJ, Griffiths JR, Arús C (2006) Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. *J Neurosurg* 105(1): 6–14
- [4] Lin AP, Tran TT, Ross BD (2006) Impact of evidence-based medicine on magnetic resonance spectroscopy. *NMR Biomed* 19(4): 476–483
- [5] Tate AR, Underwood J, Acosta DM, Julia-Sape M, Majos C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabanas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Remy C, Heerschap A, Watson D, Griffiths JR, Arús C (2006) 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
- [6] Galanaud D, Nicoli F, Chinot O, Confort-Gouny S, Figarella-Branger D, Roche P, Fuentes S, Le Fur Y, Ranjeva JP, Cozzone PJ (2006) Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy. *Magnetic Resonance in Medicine* 55(6): 1236–1245
- [7] Barba I, Moreno A, Martinez-Perez I, Tate AR, Cabanas ME, Baquero M, Capdevila A, Arús C (2001) Magnetic resonance spectroscopy of brain hemangiopericytomas: high myoinositol concentrations and discrimination from meningiomas. *J Neurosurg* 94(1): 55–60
- [8] Saraf-Lavi E, Bowen BC, Pattany PM, Sklar EML, Murdoch JB, Petito CK (2003) Proton MR spectroscopy of gliomatosis cerebri: case report of elevated myoinositol with normal choline levels. *AJNR Am J Neuroradiol* 24(5): 946–951. Case Reports
- [9] Howe FA, Barton SJ, Cudlip SA, Stubbs M, Saunders DE, Murphy M, Wilkins P, Opstad KS, Doyle VL, McLean MA, Bell BA, Griffiths JR (2003) Metabolic profiles of

- human brain tumors using quantitative in vivo ^1H magnetic resonance spectroscopy. *Magn Reson Med* 49(2): 223–232
- [10] Ishimaru H, Morikawa M, Iwanaga S, Kaminogo M, Ochi M, Hayashi K (2001) Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. *Eur Radiol* 11(9): 1784–1791
- [11] Majos C, Julia-Sape M, Alonso J, Serrallonga M, Aguilera C, Acebes JJ, Arús C, Gili J (2004) Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE. *AJNR Am J Neuroradiol* 25(10): 1696–1704
- [12] Kaminogo M, Ishimaru H, Morikawa M, Ochi M, Ushijima R, Tani M, Matsuo Y, Kawakubo J, Shibata S (2001) Diagnostic potential of short echo time MR spectroscopy of gliomas with single-voxel and point-resolved spatially localised proton spectroscopy of brain. *Neuroradiology* 43(5): 353–363
- [13] Lukas L, Devos A, Suykens JAK, Vanhamme L, Howe FA, Majós C, Moreno-Torres A, Graaf MVD, Tate AR, Arús C, Huffel SV (2004) Brain tumor classification based on long echo proton MRS signals. *Artif Intell Med* 31: 73–89
- [14] Kwock L (1998) Localized MR spectroscopy: basic principles. *Neuroimaging Clin N Am* 8(4): 713–731
- [15] Gillies RJ, Morse DL (2005) In vivo magnetic resonance spectroscopy in cancer. *Annu Rev Biomed Eng* 7: 287–326
- [16] Lisboa PJ, Taktak AFG (2006) The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Netw* 19(4): 408–415
- [17] Hagberg G (1998) From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. *NMR in Biomedicine* 11(4-5): 148–156
- [18] Menze BH, Lichy MP, Bachert P, Kelm BM, Schlemmer HP, Hamprecht FA (2006) Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors. *NMR in Biomedicine* 19(5): 599–609
- [19] Simonetti AW, Melssen WJ, Szabo de Edelenyi F, van Asten JJA, Heerschap A, Buydens LMC (2005) Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. *NMR in Biomedicine* 18(1): 34–43
- [20] Huang Y, Lisboa PJG, El-Deredy W (2003) Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection. *Stat Med* 22(1): 147–164
- [21] Tate AR, Majos C, Moreno A, Howe FA, Griffiths JR, Arús C (2003) Automated classification of short echo time in vivo ^1H brain tumor spectra: a multicenter study. *Magnetic Resonance in Medicine* 49(1): 29–36

-
- [22] Devos A, Lukas L, Suykens JAK, Vanhamme L, Tate AR, Howe FA, Majos C, Moreno-Torres A, van der Graaf M, Arús C, Van Huffel S (2004) Classification of brain tumours using short echo time 1H MR spectra. *J Magn Reson* 170(1): 164–175
- [23] Suykens JAK, Vandewalle J (1999) Least Squares Support Vector Machine Classifiers. *Neural Process Lett* 9(3): 293–300
- [24] Opstad KS, Ladroue C, Bell BA, Griffiths JR, Howe FA (2007) Linear discriminant analysis of brain tumour 1H MR spectra: a comparison of classification using whole spectra versus metabolite quantification. *NMR in Biomedicine* 20(8): 763–770
- [25] Kleihues P, Burger PC, Scheithauer BW (1993) The new WHO classification of brain tumours. *Brain Pathol* 3(3): 255–268
- [26] Julia-Sape M, Acosta D, Mier M, Arús C, Watson D (2006) A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. *Magn Reson Mater Phy* 19(1): 22–33
- [27] van der Graaf M, Julia-Sape M, Howe FA, Ziegler A, Majos C, Moreno-Torres A, Rijpkema M, Acosta D, Opstad KS, van der Meulen YM, Arus C, Heerschap A (2008) MRS quality assessment in a multicentre study on MRS-based classification of brain tumours. *NMR Biomed* 21(2): 148–158
- [28] Klose U (1990) In vivo proton spectroscopy in presence of eddy currents. *Magnetic Resonance in Medicine* 14(1): 26–30
- [29] McIntyre DJO, Charlton RA, Markus HS, Howe FA (2007) Long and short echo time proton magnetic resonance spectroscopic imaging of the healthy aging brain. *J Magn Reson Imaging* 26(6): 1596–1606
- [30] Kinoshita Y, Kajiwara H, Yokota A, Koga Y (1994) Proton magnetic resonance spectroscopy of brain tumors: an in vitro study. *Neurosurgery* 35(4): 606–613
- [31] Barton SJ, Howe FA, Tomlins AM, Cudlip SA, Nicholson JK, Bell BA, Griffiths JR (1999) Comparison of in vivo 1H MRS of human brain tumours with 1H HR-MAS spectroscopy of intact biopsy samples in vitro. *Magnetic Resonance Materials in Physics, Biology and Medicine* 8(2): 121–128
- [32] Michaelis T, Merboldt KD, Hanicke W, Gyngell ML, Bruhn H, Frahm J (1991) On the identification of cerebral metabolites in localized 1H NMR spectra of human brain in vivo. *NMR Biomed* 4(2): 90–98
- [33] Tortajada S, García-Gómez JM, Vidal C, Arús C, Julià-Sapé M, Moreno A, Robles M (2006) Improved classification by pattern recognition of brain tumours combining long and short echo time 1H-MR spectra. In SpringerLink, editor, *Book of Abstracts ESMRMB 2006 - Supplement 1, Journal Magnetic Resonance Materials in Physics, Biology and Medicine*, volume 19, pp. 168–169
- [34] Martinez-Bisbal MC, Marti-Bonmati L, Piquer J, Revert A, Ferrer P, Llacer JL, Piotto M, Assemat O, Celda B (2004) 1H and 13C HR-MAS spectroscopy of intact

- biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas. *NMR Biomed* 17(4): 191–205
- [35] Kleihues P, Cavanee W, editors (1997) *Pathology and Genetics of Tumours of the Nervous System*. International Agency for Research on Cancer
- [36] Mischel PS, Shai R, Shi T, Horvath S, Lu KV, Choe G, Seligson D, Kremen TJ, Palotie A, Liau LM, Cloughesy TF, Nelson SF (2003) Identification of molecular subtypes of glioblastoma by gene expression profiling. *Oncogene* 22(15): 2361–2373
- [37] Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, Williams PM, Modrusan Z, Feuerstein BG, Aldape K (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9(3): 157–173. Comparative Study
- [38] Burtscher IM, Skagerberg G, Geijer B, Englund E, Stahlberg F, Holtas S (2000) Proton MR spectroscopy and preoperative diagnostic accuracy: an evaluation of intracranial mass lesions characterized by stereotactic biopsy findings. *AJNR Am J Neuroradiol* 21(1): 84–93
- [39] Law M, Cha S, Knopp EA, Johnson G, Arnett J, Litt AW (2002) High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. *Radiology* 222(3): 715–721
- [40] González-Vélez H, Mier M, Julià-Sapé M, Arvanitis T, García-Gómez J, Robles M, Lewis P, Dasmahapatra S, Dupplaw D, Peet A, Arús C, Celda B, Van Huffel S, Lluch-Ariet M (2007) HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis. *Applied Intelligence* [Article in Press]

Chapter 4

Modelling of Magnetic Resonance Spectra using mixtures for binned and truncated data

MRS provides the biochemical composition of a tissue under study. This information is useful for the in-vivo diagnosis of brain tumors. Prior knowledge of the relative position of the organic compound contributions in the MRS suggests the development of a probabilistic mixture model and its EM-based Maximum Likelihood Estimation for binned and truncated data. Experiments for characterizing and classifying Short TE spectra from brain tumors are reported.

The core of this chapter has been published as a chapter of proceedings in [1]. Section 4.4.2 was included in this chapter to classify brain tumors by means of the parametric spaces of the introduced mixture models for the Short TE, the Long TE, and the combination of both TE. The evaluation procedure carried out here is similar to the one in Chapter 3.

4.1 Introduction

MRS exploits the magnetic properties of ^1H nuclei to provide information about the concentration of the compounds of materials. This makes MRS useful as non-invasive technique for brain tumor diagnosis. The MRS signals are typically interpreted in the frequency domain by visual or automatic procedures to characterize the contribution of the biological compound in the tissue. The amplitude of a compound is proportional to its concentration. This motivates the fitting of MRS spectra by mixture density models.

MRS spectra are typically analyzed by two different approaches. The first approach estimates the underlying model composed by mixtures of components to quantify the concentration of the metabolites. Frequency-domain [2] or time-domain [3] fitting methods based on signal processing are applied to the signals. The second approach extracts features from the spectra using univariate-, multivariate-statistics or PR methods [4] based on their usefulness on discrimination or regression.

This work proposes the definition and estimation of a probabilistic model based on binned and truncated data to fit ^1H magnetic resonance spectra using prior knowledge about the relative position of the components of the organic compounds observed in the

tumoral masses of the brain. The estimated parameters for each spectrum summarize the information from the biological compounds and they are used as features in classification problems of brain tumor diagnosis.

Mixture modelling has been applied in some applications where data are available only in bins and may not be provided along the whole range [5–7]. In [5], red blood cells were collected as volume distributions from a Coulter counter to study the disease status of animals exposed to *Anaplasma marginale*. The problem in MRS is similar to the previous problems in the sense that contributions of a mixture of biological compounds are assumed to be observed as counts of bins in a range of the ppm-axis. We present an adaptation of the Expectation-Maximization (EM) for fitting MR spectra, qualitative results in the characterization of spectra and quantitative results in the classification of brain tumors by the use of the estimated parameters.

The rest of the chapter is organized as follows. In sections 4.2 and 4.3, the probabilistic model and its EM-based Maximum Likelihood Estimation are presented. Then, results using MRS spectra of brain tumors are reported in section 4.4.2.

4.2 Probabilistic model

Let \mathcal{X} be a sample space partitioned into B bins, $\mathcal{X}_1, \dots, \mathcal{X}_B$, of which only the counts on the first B' bins can be recorded, while the counts on the last $B - B'$ can not. For instance, in the univariate case, the first B' bins may be delimited by $B' + 1$ points, $p_0, p_1, \dots, p_{B'}$, such that $p_0 < p_1 < \dots < p_{B'}$ and $X_b = (p_{b-1}, p_b]$, $b = 1, \dots, B'$. N independent samples (draws) from \mathcal{X} are made, but our measuring instrument reports only the number of samples falling in each of these first, *observable* B' bins, but fails to report similar counts for samples out of them, in the $B - B'$ *truncated* regions (e.g. $(-\infty, p_0]$ and $(p_{B'}, \infty)$).

Let $\mathbf{N}' = (N_1, \dots, N_{B'})$ be the vector of observed counts and let $N' = \sum_{b=1}^{B'} N_b$. Clearly, the probability of \mathbf{N}' can be computed by marginalisation of the joint probability of both, observed and truncated counts,

$$p(\mathbf{N}') = \sum_{N_{B'+1}, \dots, N_B} p(\mathbf{N}) \quad (4.1)$$

where $\mathbf{N} = (N_1, \dots, N_{B'}, N_{B'+1}, \dots, N_B)$ is the complete vector of counts. We do not know the truncated counts, nor even the total number of samples N , but we know that \mathbf{N} has a multinomial distribution defined by N samples from B categories,

$$p(\mathbf{N}) = \frac{N!}{\prod_{b=1}^B N_b!} \prod_{b=1}^B p(b)^{N_b} \quad (4.2)$$

where $p(b)$ is the probability for a sample to fall in bin \mathcal{X}_b , $b = 1, \dots, B$.

We assume that (4.2) can also be computed by marginalisation of the joint density for counts and (missing) samples,

$$p(\mathbf{N}) = \int d\mathbf{X} p(\mathbf{N}, \mathbf{X}) \quad (4.3)$$

where $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_B)$ is the whole collection of N independent samples, $\mathbf{X}_b = (\mathbf{x}_{b1}, \dots, \mathbf{x}_{bN_b})$ is the collection of those N_b from bin \mathcal{X}_b ($b = 1, \dots, B$), and

$$p(\mathbf{N}, \mathbf{X}) = \frac{N!}{\prod_{b=1}^B N_b!} \prod_{b=1}^B \prod_{n=1}^{N_b} p(\mathbf{x}_{bn}) \quad (4.4)$$

where $p(\mathbf{x}_{bn})$ is the (unknown) probability density for a sample from bin \mathcal{X}_b .

At this point, we assume that samples come from a common probability density function, irrespective of their originating bins. This density function is a parametric, C -component mixture,

$$p_{\Theta}(\mathbf{x}) = \sum_{c=1}^C \pi_c p_{\Theta'}(\mathbf{x} | c) \quad (4.5)$$

where $\Theta = (\boldsymbol{\pi}, \Theta')$ is the parameter vector of the mixture; $\boldsymbol{\pi} = (\pi_1, \dots, \pi_C)$ is the vector of mixture coefficients, subject to $\sum_c \pi_c = 1$, and Θ' includes the parameters required to define each mixture component $p_{\Theta'}(\mathbf{x} | c)$, $c = 1, \dots, C$. As usual with finite mixtures, we may think of \mathbf{x} as an incomplete component-labelled sample which may be completed by addition of an indicator variable (component label) $\mathbf{z} \in \{0, 1\}^C$ with 1 in the position of the indicated component and zeros elsewhere. Therefore, we can rewrite (4.5) as

$$p_{\Theta}(\mathbf{x}) = \sum_{\mathbf{z}} p_{\Theta}(\mathbf{x}, \mathbf{z}) \quad (4.6)$$

with

$$p_{\Theta}(\mathbf{x}, \mathbf{z}) = \prod_{c=1}^C (\pi_c p_{\Theta'}(\mathbf{x} | c))^{z_c} \quad (4.7)$$

By substitution of (4.7) in (4.6), (4.6) in (4.4) and some straightforward manipulations, we can rewrite (4.4) as

$$p_{\Theta}(\mathbf{N}, \mathbf{X}) = \sum_{\mathbf{Z}} p_{\Theta}(\mathbf{N}, \mathbf{X}, \mathbf{Z}) \quad (4.8)$$

where \mathbf{Z} is the collection of component labels for \mathbf{X} , that is, $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_B)$, with $\mathbf{Z}_b = (z_{b1}, \dots, z_{bN_b})$ and $z_{bn} \in \{0, 1\}^C$ ($b = 1, \dots, B$; $n = 1, \dots, N_b$); and

$$p_{\Theta}(\mathbf{N}, \mathbf{X}, \mathbf{Z}) = \frac{N!}{\prod_{b=1}^B N_b!} \prod_{b=1}^B \prod_{n=1}^{N_b} \prod_{c=1}^C (\pi_c p_{\Theta'}(\mathbf{x}_{bn} | c))^{z_{bnc}} \quad (4.9)$$

Note that we have added the parameter vector Θ as a subscript to the joint densities $p_{\Theta}(\mathbf{N}, \mathbf{X})$ and $p_{\Theta}(\mathbf{N}, \mathbf{X}, \mathbf{Z})$ to emphasize their dependence on the parameters governing the hidden mixture (4.5).

Now, by substitution of (4.8) in (4.3), and (4.3) in (4.1), we can write our probabilistic model as

$$p_{\Theta}(\mathbf{N}') = \sum_{N_{B'+1}, \dots, N_B} \int d\mathbf{X} \sum_{\mathbf{Z}} p_{\Theta}(\mathbf{N}, \mathbf{X}, \mathbf{Z}) \quad (4.10)$$

Note that $p_{\Theta}(\mathbf{N}')$ can be seen as an *incomplete* model which results from marginalisation (many-to-one mapping) of the *complete* model $p_{\Theta}(\mathbf{N}, \mathbf{X}, \mathbf{Z})$.

Obviously, model (4.10) still needs adoption of a particular parametric form for the mixture components. Taking into account the specific application considered in this work, we will assume that samples are drawn from a C -component mixture of univariate normal densities, of means known up to a *global shift* μ_0 , and independent variances $\sigma_1^2, \dots, \sigma_C^2$; that is, for all $c = 1, \dots, C$,

$$p_{\Theta'}(\mathbf{x} | c) \sim N(\mu_0 + \delta_c, \sigma_c^2) \quad (4.11)$$

where δ_c is the known displacement from μ_0 of the c th component mean. Thus, the vector of parameters governing the mixture components is $\Theta' = (\mu_0, \sigma_1^2, \dots, \sigma_C^2)$.

4.3 EM-based maximum likelihood estimation

Maximum likelihood estimation of Θ using the EM algorithm has been previously considered in [5] and [6] for the univariate and multivariate normal cases, respectively.

Let consider $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ to be a sample of N cases randomly drawn from $p(x; \Theta)$. The log-likelihood of Θ w.r.t. the sample would be

$$L(\Theta; X) = \log p(X; \Theta) = \sum_{n=1}^N \log p(\mathbf{x}_n; \Theta).$$

For some interesting forms of $p(x; \Theta)$ (as mixture models), the finding of the estimation $\hat{\Theta}$ that maximizes 4.3 by taking its derivatives may be complicated, and is on these cases where the EM algorithm can help. The EM algorithm [8] considers the sample X to be incomplete data which can be completed by addition of missing data, e.g. Y, Z .

Equation 4.3 is the marginalization over the domain of the hidden continuous variable Y and discrete variable Z that results in the many-to-one mapping from the complete model to the incomplete model.

$$p(X; \Theta) = \int_Y dy \sum_Z p(X, Y, Z; \Theta).$$

The EM algorithm estimates the $\hat{\Theta}$ in a two-steps iterative process till convergence. The E-step computes the expected value of the logarithm of $p(X, Y, Z; \Theta)$ with respect to the observed variables and the (k) estimation of Θ ,

$$Q(\Theta | \Theta^{(k)}) = E(\log p(X, Y, Z; \Theta) | X, \Theta^{(k)})$$

The M-step maximizes Equation 4.3 to obtain a new estimation of Θ ,

$$\Theta^{k+1} = \arg \max_{\Theta} Q(\Theta | \Theta^{(k)}).$$

Our case is similar to, but slightly different from the general, parameter-independent univariate case. More precisely, the general univariate model assumes that component means are independent, while in our model all of them are known up to a global shift. This makes our estimation problem simpler, but the EM algorithm is almost identical. In what follows, we briefly review the EM algorithm for the general model and then we

provide the necessary modifications for our modelling variation. The reader is referred to [5] for more details.

The log-likelihood function of Θ w.r.t. a given \mathbf{N}' is

$$L(\Theta; \mathbf{N}') = \log \sum_{N_{B'+1}, \dots, N_B} \int d\mathbf{X} \sum_{\mathbf{Z}} p_{\Theta}(\mathbf{N}, \mathbf{X}, \mathbf{Z}) \quad (4.12)$$

which is exactly the logarithm of $p_{\Theta}(\mathbf{N}')$ as defined in (4.10), but interpreted as a function of Θ only, and assuming that mixture components are univariate normals. The EM algorithm maximises (4.12) iteratively, through the application of two basic steps in each iteration: the E(xpectation) step and the M(aximisation) step. On the one hand, the E step computes a lower bound of (4.12) for all Θ ; the so-called Q function,

$$Q(\Theta | \Theta^{(k)}) = E[\log p_{\Theta}(\mathbf{N}, \mathbf{X}, \mathbf{Z}) | \mathbf{N}', \Theta^{(k)}] \quad (4.13)$$

that is, the expectation of the logarithm of the complete model, conditional to the incomplete data, \mathbf{N}' , and a current estimation of the model parameters, $\Theta^{(k)}$. On the other hand, the M step obtains a new estimate for Θ , $\Theta^{(k+1)}$, by maximisation of the Q function,

$$\Theta^{(k+1)} = \arg \max_{\Theta} Q(\Theta | \Theta^{(k)}) \quad \text{s.t.} \quad \sum_c \pi_c = 1 \quad (4.14)$$

Given an initial value of the parameters, $\Theta^{(0)}$, these two steps are repeated until convergence to a local maximum of the likelihood function.

Ignoring an additive term not involving Θ , the Q function can be written as

$$Q(\Theta | \Theta^{(k)}) = \sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) (\log \pi_c + \log p_{\Theta'}(x_b | c)) | \mathbf{N}', \Theta^{(k)}] \quad (4.15)$$

where $N_b^{(k)}$ is the expected number of samples drawn from bin \mathcal{X}_b ,

$$N_b^{(k)} = \begin{cases} N_b & \text{if } b \leq B' \\ N' \frac{p(b)^{(k)}}{\sum_{b'=1}^{B'} p(b')^{(k)}} & \text{otherwise} \end{cases} \quad (4.16)$$

with $p(b)^{(k)}$ being the probability for a sample to fall in bin \mathcal{X}_b ,

$$p(b)^{(k)} = \int_{\mathcal{X}_b} dx p_{\Theta^{(k)}}(x) \quad (4.17)$$

The expectation in (4.15) is with respect to a sample x_b from bin \mathcal{X}_b ; i.e., with respect to the truncated density of the bin \mathcal{X}_b

$$p_{\Theta^{(k)}}^{\text{trunc}}(x_b) = \frac{p_{\Theta^{(k)}}(x_b)}{p(b)^{(k)}} \quad (4.18)$$

and involves the posterior probability for x_b to belong to component c of the mixture, given a current parameter estimate $\Theta^{(k)}$,

$$z_c^{(k)}(x_b) = \frac{\pi_c^{(k)} p_{\Theta^{(k)}}(x_b | c)}{p_{\Theta^{(k)}}(x_b)} \quad (4.19)$$

Maximisation of (4.15), as indicated in (4.14), leads to the following re-estimates for each component c ($c = 1, \dots, C$)

$$\pi_c^{(k+1)} = \frac{\sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}]}{\sum_{b=1}^B N_b^{(k)}} \quad (4.20)$$

$$\mu_c^{(k+1)} = \frac{\sum_{b=1}^B N_b^{(k)} E_b[x_b z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}]}{\sum_{b=1}^B N_b^{(k)} E[z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}]} \quad (4.21)$$

$$\sigma_c^{2(k+1)} = \frac{\sum_{b=1}^B N_b^{(k)} E_b[(x_b - \mu_c^{(k+1)})^2 z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}]}{\sum_{b=1}^B N_b^{(k)} E[z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}]} \quad (4.22)$$

where, as in (4.15), all expectations are with respect to the truncated density (4.18). Their derivations were shown by McLachlan and Jones in [5, 9].

Equations (4.20), (4.21) and (4.22) are the basic equations of an EM iteration in the general, parameter-independent univariate case, EM for Binned and Truncated Data (EMBTD). In our case, EM for Binned and Truncated Data with related means (EMBTDr), with means known up to shift μ_0 , the basic equations are (4.20), (4.22) and $\mu_c^{(k+1)} = \mu_0^{(k+1)} + \delta_c$, $c = 1, \dots, C$, where

$$\mu_0^{(k+1)} = \frac{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[(x_b - \delta_c) z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}]}{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}]}, \quad (4.23)$$

4.4 Experimental Results

The mixture models presented in the previous sections were applied to the INTERPRET dataset [10] used in the experiments of the Chapter 3.

The a priori information from biochemical knowledge used in the experiments was the CS (in *ppm* units) of the metabolites Lipid resonance at 0.92 *ppm* (L2) (0.92), Glutamate/Glutamine (Glx) (2.04), Lipid resonance at 1.29 *ppm* (L1) (1.29), Glx(2) (2.46), Lactate (Lac) (1.31), Glx(3) (3.76), Alanine (Ala) (1.48, 1.46), myo-Inositol (mI) (3.26), N-Acetyl groups (NAc) (2.02, 2.00), mI(2) (3.53), Creatine (Cr) (3.03), mI(3) (3.61), Cr(2) (3.92), Taurine (Tau) (3.25), Choline (Cho) (3.19), Tau2 (3.42), Glycine (Gly) (3.55), Ala(2) (3.78).

The initialization of the EMBTD and EMBTDr, in the experiments below, were established in a similar way to make comparable the results between them and with the results obtained in Chapter 3. The initial means of the components for the EMBTD were the typical CS of the metabolites described above. In EMBTDr, *global shift* μ_0 was set to 0 and the displacements of the metabolites were set to their typical CS. For both algorithms, the prior probabilities and variances were equally established for every metabolite.

As a descriptive summary, Figures 4.1, 4.2, 4.3 show the main behaviour of the EMBTD and EMBTDr estimates. In the Figure 4.1, the parameters estimated by both EMBTD and EMBTDr are quite similar, hence the spectrum is fitted in a similar way. In the Figure 4.2, the related means restriction incorporated in the EMBTDr model keeps the position of the compounds better than the EMBTD model according to the underlying biological mixture. In the Figure 4.3, the EMBTD model fits a lipid contribution at 2.75*ppm* not specified in the prior knowledge, but the meaning of the initial components based on biological knowledge is lost.

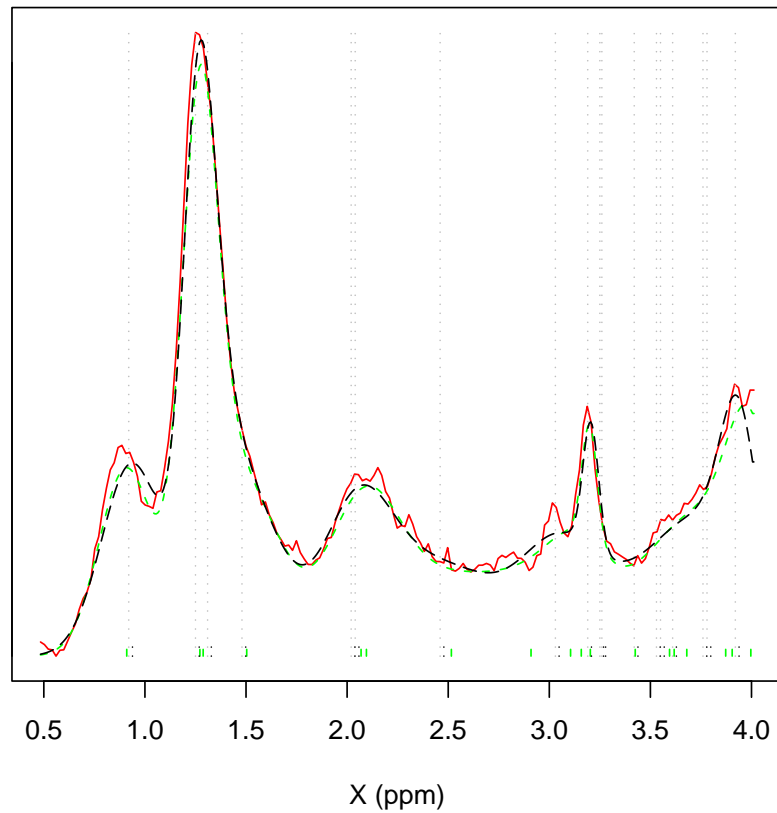


Figure 4.1: Spectrum (1/3) showing the behaviour of the EMBTD and EMBTDr models. Real spectra are drawn in solid red lines, EMBTD models in dashed black lines and EMBTDr in double-dashed green lines. μ_c are marked with dotted gray vertical lines, $\hat{\mu}_c$ of each model are marked with a small green vertical line for EMBTD and two black dots for EMBTDr.

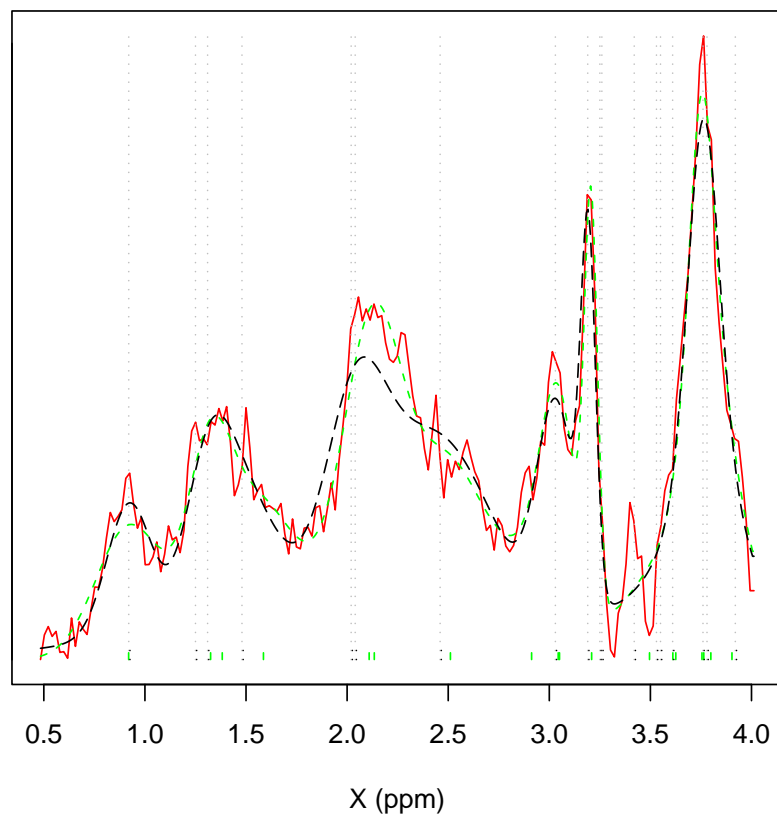


Figure 4.2: Spectrum (2/3) showing the behaviour of the EMBTD and EMBTD_r models. Symbols are the same as in Figure 4.1.

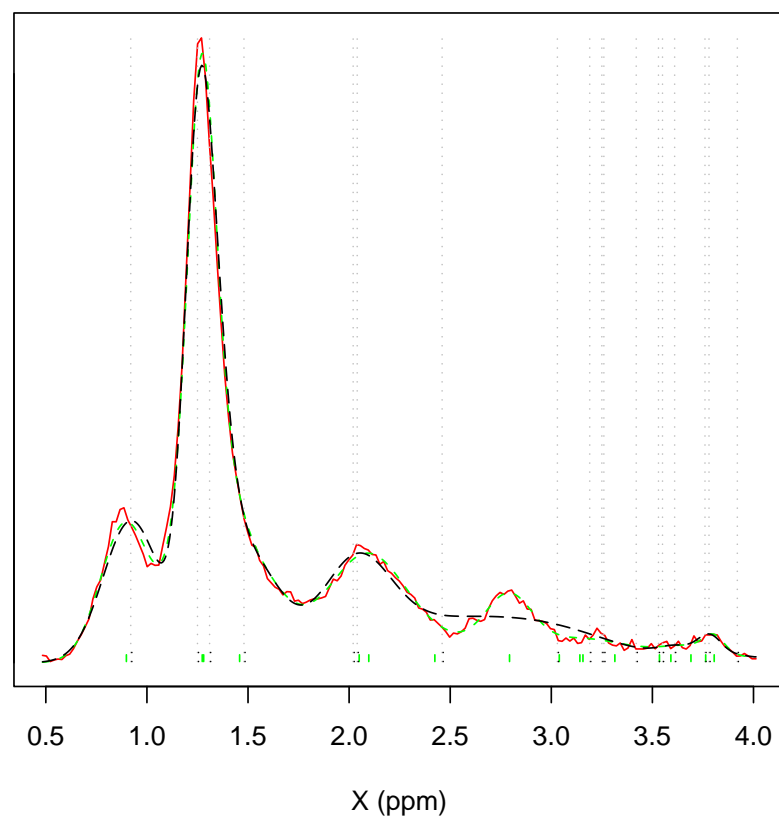


Figure 4.3: Spectrum (3/3) showing the behaviour of the EMBTD and EMBTDr models. Symbols are the same as in Figure 4.1.

Table 4.1: Mean of the difference between the estimated shifting of the components by the EM algorithms with respect to the typical CS

	EMBTDr	EMBTDr
$\mu_c - \hat{\mu}_c$	0.0079	0.0022

Table 4.2: Results for the multiclass EMBTD-parametric-based classification among AGG, MEN and LGG. kRSTT evaluation of the SW approach based on the Ability to Separate (AS) of the predictive Gaussian linear classifier.

Dataset	Accuracy (%)	[Credibility Interval] (%)
Combined TE (EMBTDr)	84.5	[83.7,85.3]
Combined TE (EMBTDr)	83.2	[82.4,84.0]
Short TE (EMBTDr)	85.1	[84.4,85.9]
Short TE (EMBTDr)	83.1	[82.3,83.9]
Long TE (EMBTDr)	68.7	[67.7,69.6]
Long TE (EMBTDr)	62.5	[61.5,63.6]

4.4.1 Fitting of the model to the resonances of the metabolite

In the first experiment, we measured the mean of the differences between the estimated shifting ($\hat{\mu}_c$) of the components with respect to the typical CS (μ_c). None or small shifting is assumed in the spectra of the database. Therefore, the smaller the difference, the closer the estimated component is to the organic compound. Table 4.1 shows the results obtained by EMBTD and EMBTDr on the MRS database. The differences obtained by both models are small, considering that the range is 3.6ppm and the frequency resolution 0.02ppm. However, the difference obtained on average by EMBTD is 3.6 times the difference obtained by EMBTDr. Hence, EMBTDr keeps better the position of the biological compound in the estimated model.

4.4.2 Brain Tumor classification based on the EMBTD and EMBTDr parametric-spaces

In the second study, the ($\hat{\pi}$) parameters estimated by the EMBTD and EMBTDr were the inputs of the classifiers of the multiclass and binary problems studied in Chapter 3. Hence, the combination of Short TE and Long TE spectra by means of the combination of their EMBTD and EMBTDr parametric-spaces were compared with the Short TE and Long TE parametric-spaces alone. The multiclass classifier was based on LDA. Besides, the binary classifiers were based on LSSVM to make comparables these new results with the ones obtained in Chapter 3.

Table 4.2 shows the results for the multiclass parametric-based classifications among AGG, MEN and LGG. The evaluation of the models was kRSTT, similar to the one used in Chapter 3. Generally speaking, the EMBTD-based and EMBTDr-based multiclass classifiers did not outperform the SW-LDA classifiers obtained in Chapter 3. As in the SW-LDA model for the multiclass problem, the Combined TE models follow the Short TE

Table 4.3: kRSTT evaluation of the LS-SVM for the binary classifiers.

Task	Spectra	Features	Accuracy [CI] (%)	AUC
AGG vs. MEN	Combined TE	EMBTDr	90.0	0.952
	Combined TE	EMBTDr	87.0	0.939
	Short TE	EMBTDr	91.1	0.953
	Short TE	EMBTDr	87.0	0.938
	Long TE	EMBTDr	77.4	0.841
	Long TE	EMBTDr	69.9	0.741
AGG vs. LGG	Combined TE	EMBTDr	91.9	0.956
	Combined TE	EMBTDr	90.0	0.958
	Short TE	EMBTDr	92.5	0.962
	Short TE	EMBTDr	90.9	0.951
	Long TE	EMBTDr	82.6	0.853
	Long TE	EMBTDr	80.0	0.853
MEN vs. LGG	Combined TE	EMBTDr	84.1	0.929
	Combined TE	EMBTDr	92.0	0.967
	Short TE	EMBTDr	86.8	0.932
	Short TE	EMBTDr	88.0	0.943
	Long TE	EMBTDr	75.4	0.820
	Long TE	EMBTDr	65.8	0.686

model behaviour. Furthermore, the Combined TE and Short TE models outperformed the Long TE models.

Table 4.3 shows the results for the pairwise classifiers of AGG, LGG and MEN. The Combined TE parametric-space outperformed the Short TE parametric-space only for the MEN vs. LGG classification problem. For this case, the EMBTDr model obtained the best results. For the AGG vs. MEN and the AGG vs. LGG problems, the EMBTD for Short TE was the best approach. In general, The LSSVM classifiers obtained by means of the ReliefF feature selection algorithm outperforms the results obtained by the EMBTD and EMBTDr models. By the contrary, the feature obtained by the EMBTD and, specially, EMBTDr keep the meaning of the resonances of the metabolites and their use in classification achieves good results.

4.5 Conclusions and further work

A probabilistic mixture model for binned and truncated data with univariate mixture densities of means known up to a global shift has been proposed for Magnetic Resonance Spectroscopy data characterization. The model can be efficiently estimated by means of the E(xpectation)-M(aximisation) algorithm. The new version of the algorithm keeps the biological information in the model and fits properly Short TE MR Spectra. The incorporation of the classifier in a CDSS could be of interest for clinicians to decide the diagnosis of routine or special patients. In further work, more applications of the proposed mixture model will be considered in MRS analysis.

Bibliography

- [1] Garcia-Gomez JM, Robles M, Van Huffel S, Juan-Císcar A (2007) Modelling of Magnetic Resonance Spectra using mixtures for binned and truncated data. In Springer, editor, Proceedings of the 1st Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), volume 2 of *Lecture Notes in Computer Science Series 4478*, pp. 266–273. Springer-Verlag, Girona, Spain
- [2] Mierisova S, Ala-Korpela M (2001) MR spectroscopy quantitation: a review of frequency domain methods. *NMR in Biomedicine* 14(4): 247–259
- [3] Vanhamme L, Sundin T, Hecke PV, Huffel SV (2001) MR spectroscopy quantitation: a review of time-domain methods. *NMR in Biomedicine* 14(4): 233–246
- [4] Tate AR, Underwood J, Acosta DM, Julia-Sape M, Majos C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabanas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Remy C, Heerschap A, Watson D, Griffiths JR, Arús C (2006) 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
- [5] McLachlan GJ, Jones PN (1988) Fitting Mixture Models to Grouped and Truncated Data via the EM Algorithm. *Biometrics* 44: 571–578
- [6] Cadez IV, Smyth P, McLachlan GJ, McLaren CE (2002) Maximum Likelihood Estimation of Mixture Densities for Binned and Truncated Multivariate Data. *Mach Learn* 47(1): 7–34
- [7] Same A, Ambroise C, Govaert G (2005) A classification EM algorithm for binned data. *Computational Statistics and Data Analysis*
- [8] Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society* 39: 1–38
- [9] Jones PN, McLachlan GJ (1990) Statistical Algorithms: Algorithm AS 254: Maximum Likelihood Estimation from Grouped and Truncated Data with Finite Normal Mixture Models. *Applied Statistics* 39(2): 273–282
- [10] Julia-Sape M, Acosta D, Mier M, Arús C, Watson D (2006) A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. *Magn Reson Mater Phy* 19(1): 22–33

Chapter 5

Multiproject-multicentric evaluation of automatic brain tumor classification by Magnetic Resonance Spectroscopy

Automatic brain tumor classification by MRS has been under development for more than a decade. Nonetheless, to our knowledge, there are no published evaluations of predictive models with unseen cases that are subsequently acquired in different centers. The multicenter eTUMOUR project (2004-2009), which builds upon previous expertise from the INTERPRET project (2000-2002) has allowed such an evaluation to take place.

253 pairwise classifiers for Glioblastoma, Meningioma, Metastasis, and Low-Grade Glial diagnosis were inferred based on 211 SV Short TE INTERPRET MR spectra obtained at 1.5T (PRESS or STEAM, 20 – 32ms) and automatically pre-processed. Afterwards, the classifiers were tested with 97 spectra, which were subsequently compiled during eTUMOUR. In our results based on subsequently acquired spectra, accuracies of around 90% were achieved for most of the pairwise discrimination problems. The exception was for the Glioblastoma versus Metastasis discrimination, which was below 78%. A more clear definition of metastases may be obtained by other approaches, such as MRSI+MRI.

Therefore, the prediction of the tumor type of in-vivo MRS is possible using classifiers developed from previously acquired data, in different hospitals with different instrumentation under the same acquisition protocols. This methodology may find application for assisting in the diagnosis of new brain tumor cases and for the quality control of multicenter MRS databases.

The core of this chapter has been published as journal paper in [1]. Section 5.4 includes the multiproject-multicenter evaluation of the EMBTD and EMBTD_r-based classifiers introduced in Chapter 4.

5.1 Introduction

MRS is slowly becoming an accurate non-invasive complement to MRI for initial diagnosis exam of brain masses [2], since it provides useful chemical information about metabolites for characterizing brain tumors [3]. To achieve this status, clinical and PR-based classification of brain tumors using MRS data has been thoroughly investigated for more than fifteen years [2, 4–14].

The CDSSs based on PR should be developed in such a way so as to obtain high accuracy in classification, interpretability by means of clinical knowledge and the generalization of the performance to new samples obtained subsequently in different clinical centers [15–18]. Standardization of acquisition conditions and protocols should make data from different hospitals compatible and allow the development and evaluation of joint CDSSs. This standardization prevents possible bias from single-center or single-machine studies and, additionally, increases the number of available cases for classifier development and test purposes.

During INTERPRET [9, 19], a protocol was defined to guarantee the compatibility of the signals acquired at different hospitals [20, 21]. As a result, studies on automated brain tumor classification were carried out using these data. Hence, in previous studies [8, 9, 11, 22], the ability of automatic classifiers based on Short TE MRS to discriminate among different brain tumor diagnoses was demonstrated. In addition, in [12, 14, 22], the automated classification by means of Long TE MRS was also studied and demonstrated. Other studies evaluated the extension of the classifiers towards ^1H MRSI [13, 22–25]. Every study reported above was developed and evaluated using data acquired during the same period of time. Besides, other automated classification studies, such as [3, 14, 26–29], have been reported on single-center MRS datasets of brain masses.

In order to provide the clinical community with robust results of automatic classification, the extension of the evaluation in time is advisable. Hence, the validation of classifiers through subsequent cases can consolidate the confidence of clinicians in the potential applicability of these classifiers. The multicenter eTUMOUR [30] (2004–2009) 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. Furthermore, the raw MRS data acquired during INTERPRET were incorporated into the eTUMOUR dataset for classifier development. This provides a unique opportunity to evaluate INTERPRET-based models by means of cases from a later date from partly different hospitals with different instrumentation, but obtained using the same or compatible acquisition protocols. The multiproject-multicenter evaluation proposed in this study gives a close-up perspective of the conditions that predictive models may face under different real clinical environments.

In this study, six pairwise classifiers for GBM, MEN, MET, and LGG diagnoses were developed and tested on Single voxel (SV) Short TE MRS signals. Short TE MRS is fast (typically 5min) and robust, so it is considered to be appropriate for routine clinical studies [2]. Most major hospitals currently use this acquisition protocol for the MRS evaluation of brain tumors. It has been reported that Short TE spectra contain a larger amount of information than Long TE spectra, e.g. metabolites and other compounds that are considered useful for classification purposes [2, 9, 12]. Hence, Cr (3.02, 3.92ppm), Cho (3.21ppm), N-Acetyl Aspartate (NAA) (2.01ppm), mI and Gly (3.55ppm), mI/Tau (3.26ppm), Glx (2.04, 2.46, 3.78ppm), Lac (1.31ppm), and Ala (1.47ppm) are observed at Short TE. Furthermore, Macromolecules (MM) (5.4ppm, 2.9ppm, 2.25ppm, 2.05ppm, 1.4ppm and 0.87ppm) and ML are also well detected at Short TE [2, 9]. Comparative studies on the use of Short TE vs. Long TE have shown the benefit of using Short TE or the combination of both echo times for automatic classification purposes [31].

Based on previous results from [11, 12, 19, 22], good performance of the PR models

Table 5.1: Number of Training (INTERPRET) and Test (eTUMOUR) cases per class used in the study. Short TE ^1H MRS data were acquired according to a consensus protocol during the INTERPRET (2000-2002) and eTUMOUR (2004-2009) projects.

Class	INTERPRET	eTUMOUR
GBM	84	28
MEN	57	17
MET	37	32
LGG	33	20
	211	97

could be expected for most of the classification problems, except for the discrimination of Glioblastoma and Metastasis [11]. Our performance estimation of models trained with INTERPRET data and tested over eTUMOUR cases confirmed this behaviour. We observed that pairwise discrimination between Glioblastoma, Meningioma, Metastasis, and Low-Grade Glial achieved an accuracy of around 90%. The exception was for the discrimination between Glioblastoma and Metastasis that did not perform better than 78%. This study consolidates the results obtained by previous studies in automatic brain tumor classification using MRS. These results may also increase the confidence of the clinical community in the use of CDSSs that incorporate this kind of classifiers for the interpretation of MRS biomedical signals and the diagnosis of brain tumors.

5.2 Materials and Methods

5.2.1 Data acquisition

The training data used for classifier development were SV MRS signals at $1.5T$ at Short TE (Point-Resolved Spectroscopic Sequence (PRESS) or Stimulated Echo Acquisition Mode sequence (STEAM), $20\text{-}32ms$) that were acquired by international centers in the framework of INTERPRET [19]. The classes considered for inclusion in this study were based on the histological classification of the CNS tumors set up by the WHO [32]: GBM, MEN, MET, and LGG (Astrocytoma gII, Oligoastrocytoma gII, or Oligodendroglioma gII). The number of cases by class is summarized in Table 5.1.

211 SV ^1H MR spectra from the INTERPRET database [20] were included. These signals were acquired with Siemens, General Electric (GE), and Philips instruments by six international centers. The acquisition protocols included PRESS or STEAM sequences, with spectral parameters: Recycling Time (TR) between 1600 and $2020ms$, TE of 20 or $30\text{-}32ms$, spectral width of $1000 - 2500Hz$, and 512 , 1024 , or 2048 data-points, as described in previous studies [20]. Every training spectrum and diagnosis was validated by the INTERPRET Clinical Data Validation Committee (CDVC) and expert spectroscopists [9].

The test data were provided by eight international institutions in the framework of eTUMOUR [30]. The cases with the SV Short TE (STEAM $20ms$, PRESS $30\text{-}32ms$) MRS at $1.5T$) signal validated by the expert spectroscopist of eTUMOUR and with the original histopathology available before February 28th, 2007) were included. Therefore, 97 cases from eTUMOUR were considered for testing in this study. The test cases used to evaluate the performance of the classifiers were acquired from partly different hospitals in

Table 5.2: Breakdown of cases per manufacturer included in the Training (INTERPRET) and Test (eTUMOUR) datasets.

Manufacturer	INTERPRET (%)	eTumour (%)
GE	53.1	54.6
Siemens	6.6	12.4
Philips	40.3	33.0

Table 5.3: Percentage of cases per acquisition center included in the Training (INTERPRET) and Test (eTUMOUR) datasets. Last row indicates the percentage of training cases that belong to centers that did not produce eTUMOUR cases, and the percentage of test cases that belong to centers that did not acquire training data for INTERPRET.

CENTERS	Training from INTERPRET (%)	Test from eTUMOUR (%)
UMC NIJMEGEN	2.8	1.0
ST. GEORGE'S HOSPITAL	27.0	18.6
MEDICAL UNIV. OF LODZ	3.8	10.3
FLENI	1.9	6.2
IDI-BELLVITGE	40.3	
CENTRE DE DIAG. PEDRALBES	24.2	
CENTRE DE DIAG. PEDRALBES + IAT		28.9
IDI-BADALONA		17.5
UNIV. DE VALENCIA		16.5
HOSPITAL SANT JOAN DE DEU		1.0
% OF CASES OF PROJECT		
EXCLUSIVE CENTERS	40.3	35.1

later dates than the training cases and using instruments of the three main manufacturers. Table 5.2 shows that the percentages of cases by manufacturer included in the test data are similar to the percentages in the training data. Table 5.3 shows the percentage of cases by center included in the training and test datasets. Forty per cent of training cases belong to one center that afterwards did not provide test data. Besides, 35% of test cases belong to three new centers that were not providers of training data.

5.2.2 Pre-processing

Each signal was pre-processed according to the INTERPRET protocol. A fully automatic pre-processing pipeline was available for the training data. Besides, a semi-automatic pipeline was defined for some new file formats of the test cases 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) Eddy current correction was applied to the water-suppressed Free Induction Decay (FID) of each case using the Klose

algorithm [33]. 2) The residual water resonance was removed using the Hankel-Lanczos Singular Value Decomposition (HLSVD) time-domain selective filtering using 10 singular values and a water region of $[4.33, 5.07]ppm$. 3) An apodization with a Lorentzian function of $1Hz$ of damping was applied. 4) Before transforming the signal to the frequency domain using the Fast Fourier Transform (FFT), an interpolation was needed in order to increase the frequency resolution of the low resolution spectra to the maximum frequency resolution used in the acquisition protocols (see [9] for details in the acquisition conditions and resolution). This was carried out with the zero-filling procedure. 5) Afterwards, the baseline offset, which was estimated as the mean value of the region $[11, 9] \cup [-2, -1]ppm$, was subtracted from the spectrum. 6) The normalization of the spectral data vector to the L2-norm was performed based on the data-points in the region $[-2.7, 4.33] \cup [5.07, 7.1]ppm$. 7) Depending on the Signal-to-Noise Ratio (SNR) and the tumor pattern, an additional frequency alignment check of the spectrum was performed by referencing the ppm-axis to (in order of priority) the total Cr at $3.03ppm$ or to the Cho containing compounds at $3.21ppm$ or the ML at $1.29ppm$. 8) Finally, the region of interest was restricted to $[0.5, 4.1]ppm$, obtaining a vector of 190 points for each spectrum where, after the pre-processing filters, the resonances of the main metabolites arise and where the contribution of the residual water is expected to be minimal. In summary, 211 INTERPRET cases and 47 cases of the eTUMOUR test dataset (32 from Philips and 15 from GE) were pre-processed with the automatic pipeline.

Semi-automatic pipeline

Due to limitations of the automatic pre-processing software, 50 test samples were pre-processed by a semi-automatic pipeline that was partially based on the Java Magnetic Resonance User Interface (jMRUI) [34]. Some modifications of the semi-automatic pipeline with respect to the automatic pipeline were in the following steps: 1) The phase of the water-suppressed FID was mainly corrected with the reference water. Additional manual zero-order and first-order phase correction was performed when needed. 2) Residual water was removed by means of the jMRUI-implementation of the Hankel Singular Value Decomposition (HSVD) algorithm [35]. The filter was parametrized as in the automatic pipeline. Steps 3-8 remained equivalent to 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 testing the PR models were obtained.

5.2.3 Feature extraction

Several feature extraction methods based on PR (see Section 1.1) were applied to the real part of the spectra prior to any classification approach. These methods included direct spectral Peak integration (PI) on selected metabolite resonance regions [36], Peak height of typical resonances (PPM) [37], PCA [38, 39], ICA [40, 41], and WAV [42, 43]. Finally, some classification approaches were applied to the full region of interest represented by a data vector of 190 points (190). The selected features for the classifiers were derived from previous studies [11, 31] or from model validation based on the training dataset. In some approaches, Standard Normal Variate (SNV) scaling was applied to the obtained features. The wavelet basis used in the experiments was *coiflet 3* with 9 levels [42].

Table 5.4: Best results obtained for the six pairwise classification problems. The ERR and BER estimation based on CV over the INTERPRET data and based on the eTUMOUR IT set are shown. The columns of the table are: Task: classification problem defined by the classes to discriminate by the classifiers; Features: acronym of the feature extraction method, Classif: acronym of the classification method, CV: results estimated by means of a 10-fold CV in the INTERPRET database, IT: results estimated by means of the independent test, with the INTERPRET database as training and the eTUMOUR dataset as test, ERR: error rate, and BER: Balanced Error Rate. [,]: interval within every result falls.

Task	Features	Classif	CV		IT	
			ERR	BER	ERR	BER
GBM vs. MEN	190	MLP	0.06	0.07	0.07	0.09
GBM vs. MET	PI	LDA	0.33	0.40	0.22	0.21
GBM vs. LGG	PI	LSSVM	0.12	0.18	0.08	0.09
MEN vs. MET	PCA	MLP	0.05	0.05	0.06	0.07
MEN vs. LGG	ICA	LSSVM	0.08	0.09	0.08	0.08
MET vs. LGG	PI	LDA/FLDA/MLP/LSSVM	[0.01,0.04]	[0.01,0.04]	0.06	0.07

5.2.4 Classification methods

Ten methods (see Section 1.1) were applied to address the pairwise classifications. These methods included parametric discriminant analysis [44]: LDA, FLDA [45]), QDA, Linear Discriminant Analysis with diagonal covariance matrix (dLDA) and Quadratic Discriminant Analysis with diagonal covariance matrix (dQDA). Kernel-based models (SVM [46] and LSSVM [47]) were also applied. Additionally, Artificial Neural Networks (MLP [48] and BDK [28, 49]) and single and ensemble [50] classifiers using K-nearest neighbours and local feature reduced by PCA (PCA-KNN) [51, 52]) were used.

Bayesian strategies for regularization were also applied in some of the classifiers based on LSSVM [53] and MLP [54].

5.2.5 A measure to evaluate unbalanced classifiers: the Balanced Error Rate (BER).

The performance was measured by means of the ERR and the BER (see Section 1.1.4). BER is useful when one class is underrepresented compared to the other class, e.g. GBM vs. LGG and GBM vs. MET in the INTERPRET dataset and MEN vs. GBM and MEN vs. MET in the eTUMOUR dataset.

5.3 Results and discussion

For each task, different combinations of feature extraction and classification methods were applied in the study. An estimation of the ERR and BER for the INTERPRET dataset using a 10-fold CV was carried out for each model. Afterwards, the estimation of the ERR and BER were obtained on the IT dataset of eTUMOUR. Table 5.4 illustrates the results with the best pairwise classifiers based on the IT estimation.

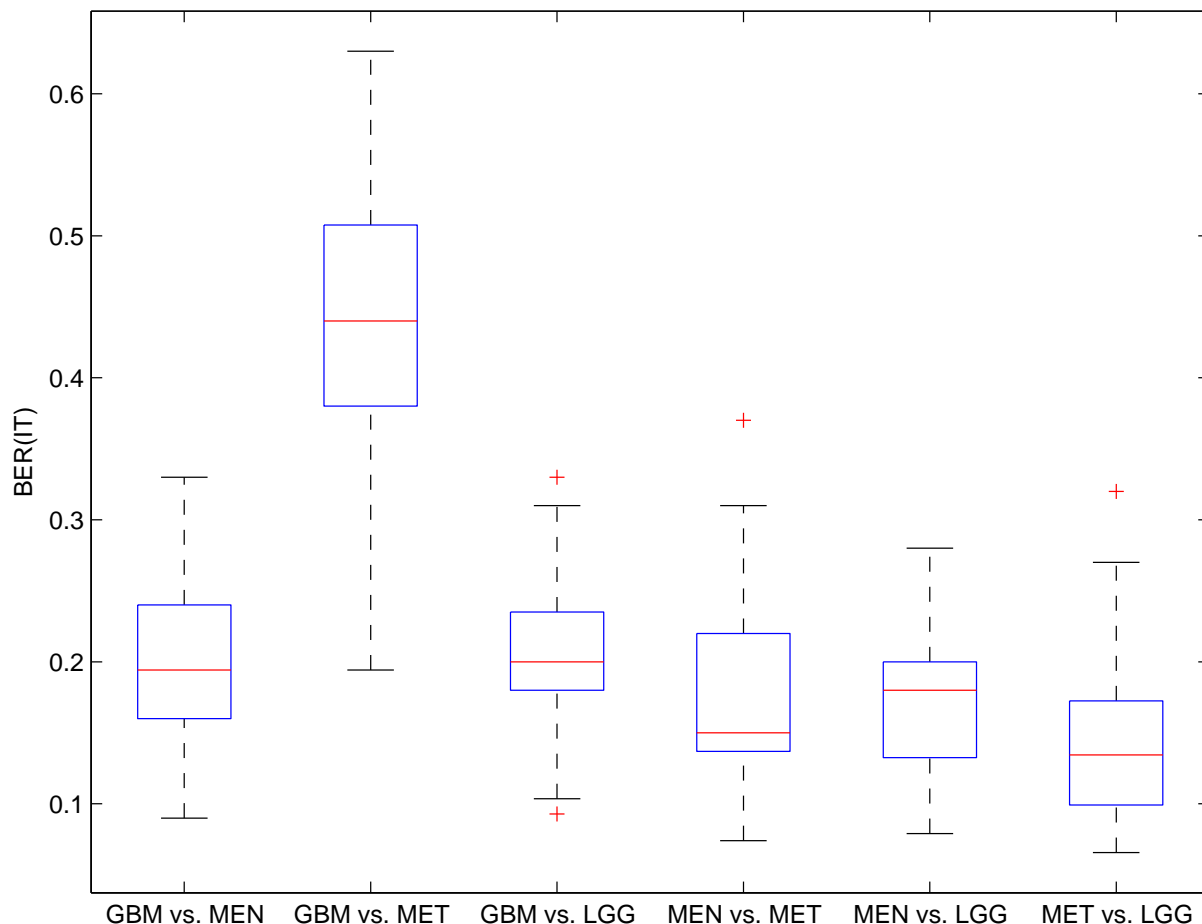


Figure 5.1: Box-whisker plots of the performance for each problem in the eTUMOUR dataset. Performance is measured in BER. The box indicates the region between the lower ($X_{0.25}$) and the upper ($X_{0.75}$) quartiles. The horizontal line inside the box indicates the median of the distribution, and the vertical lines (the "whiskers") extend to at most 1.5 times the box width. Any outlier of the distribution is displayed with a cross (+).

5.3.1 The classification problems

Most of the discrimination problems among the four classes were solved with high accuracy in the eTUMOUR dataset. Table 5.4 shows that most of the best classifiers among GBM, MEN, MET, and LGG achieved an accuracy ($1 - ERR$) of around 90%. Such decision support methodologies with these ratios of accuracy may be useful to be incorporated in integrated CDSSs for clinical purposes. Besides, for GBM vs. MET, the best result was an accuracy of 78% of the independent test, which is far from the accuracy obtained for the other discrimination problems. The Glioblastoma vs. Metastasis discrimination by means of the MRS is difficult with the use of SV spectroscopy alone [8, 9, 55–58]. Other approaches, such as MRSI coupled with MRI or the acquisition of an additional adjacent voxel to the brain mass should provide relevant additional information for distinguishing between these two types of tumors [57–59].

Figure 5.1 shows the box-whisker plot of the performance (BER based on IT) for each problem. Note the high deviation of the distribution for the GBM vs. MET with respect

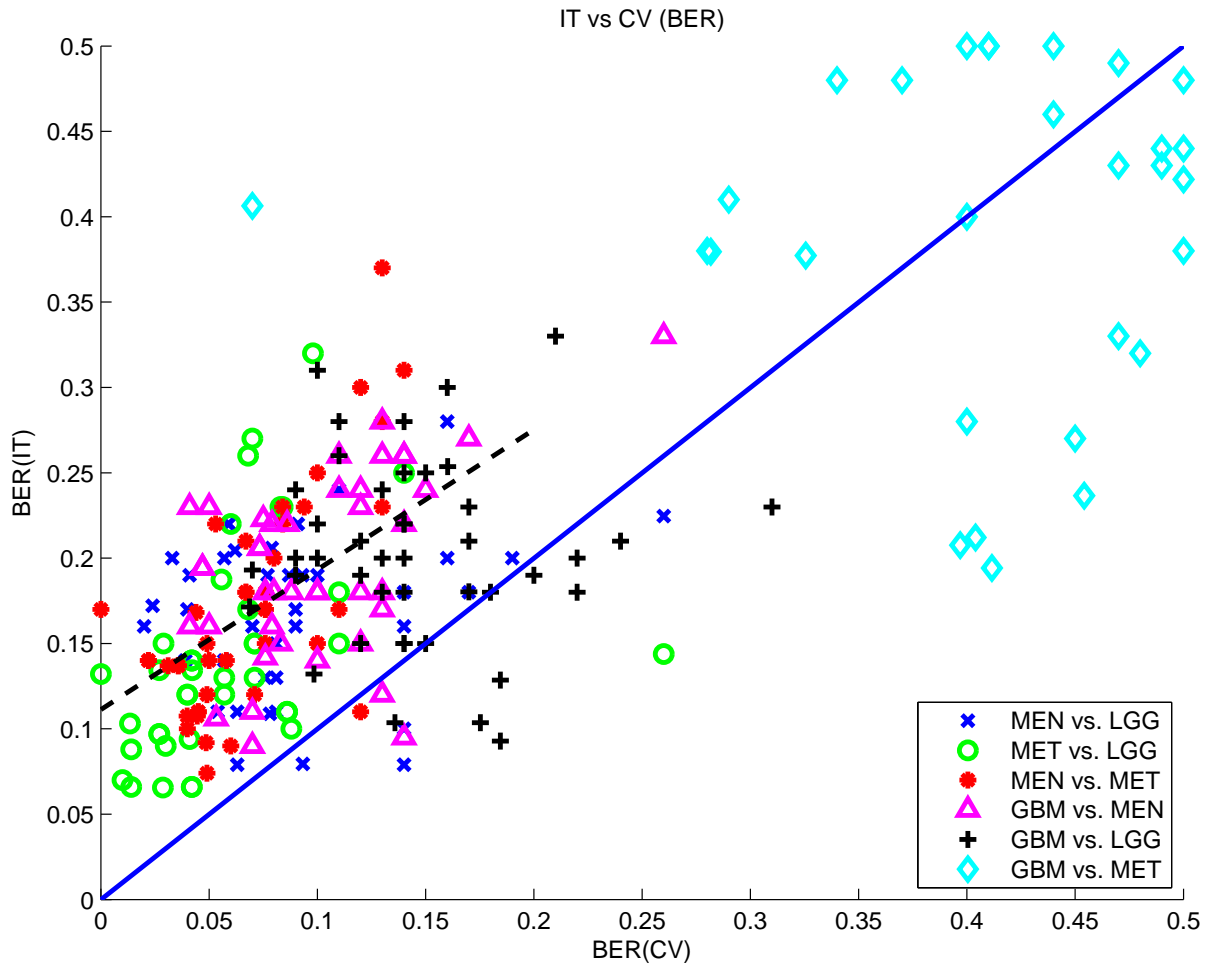


Figure 5.2: Figure 2: Scatter plot of the performance measured in BER estimated by the IT set consisting of new eTUMOUR cases and the BER estimated by the CV using the INTERPRET cases. $BER(IT) = BER(CV)$ is represented by the solid-blue line and the trend of the $(BER(CV) < 0.2, BER(IT) < 0.3)$ region is indicated by the black-dashed line.

to the others. In a multiple comparison at a 0.05 α -level based on the Tukey's honestly significance difference criterion for Kruskal-Wallis nonparametric one-way analysis of variance [60], each problem had a mean rank that was significantly different from the GBM vs. MET problem. The distributions of the other five discrimination problems overlapped among them. Nevertheless, the smallest non-outlier observation of the GBM vs. LGG problem was higher than the smallest non-outlier observation of the remaining problems. This may indicate that the GBM vs. LGG discrimination is more difficult to solve by SV Short TE MRS than the other four discrimination problems.

The different approaches obtained good results for the discrimination of the GBM and MEN classes. A multilayer perceptron with the full spectra achieved a BER of 0.09. The mode of the distribution of BER was below 0.20 for the GBM vs. MEN problem.

The difficulty of the GBM vs. MET discrimination was clearly observed in both CV- and IT-estimations (see Figure 5.2). In the distribution of the IT results for this problem,

the BER mode was 0.5, and the main distribution of the results ranged from 0.4 to 0.55. Some methods achieved a BER of 0.2; nevertheless, the main mass of the distribution was far from this value, which makes it difficult to ensure reproducibility of these performances. These results agree with those already published in previous studies [9, 11]. This is most probably due to the similar necrotic profile (high lipid peaks mask the rest of the metabolic information) of the Metastasis cases and of most of the Glioblastoma cases.

The mode of the BER for the GBM vs. LGG problem was 0.2. Nevertheless, there was a set of regularized classifiers that obtained a BER of around 0.09. To be more precise, the best BER corresponded to the Bayesian framework for LS-SVM using PI values. Devos et al. [11] obtained comparable performances for this problem using LDA and standard LS-SVMs. In studies [26, 61], significant statistical differences between GBM and LGG and between GBM and astrocytoma grade-III were also found for different metabolite ratios with respect to Cr and/or water. In Long TE, Menze et al. [14] observed a better performance with regularized methods compared with the standard ones when classifying normal, non-progressive tumors (with radiation injury and stable disease) and brain tumors.

As expected, our results confirm that MEN can be easily discriminated from MET no matter what method is used. Most of the BER probability mass of the results was in the interval from 0.1 to 0.2. The best result achieved a BER of 0.07, which was based on PCA and a neural network with Bayesian regularization. These results are consistent with [11].

LSSVM and LDA with different feature extraction methods achieved BER of 0.08 and 0.11 for the Meningioma vs. Low-Grade Glial problem. Most of the results for this problem were in the interval from 0.15 to 0.25, and the mode of the distribution was under 0.2. The low error in MEN vs. LGG was also predicted by the CV results on the INTERPRET data. This result is consistent with the performances reported in Tate et al. in [8] on a three-class discrimination problem: MEN vs. LGG vs. AGG (which is composed of GBM and MET). In that study, the confusion submatrix of MEN vs. LGG indicates no missclassifications between them. Identical results were obtained by Tate et al. in [9] when extending the three-class classifier to MEN vs. LGG vs. AGG.

The distribution of BER for MET vs. LGG had a clear trend towards the lower values (BER of 0.1), showing good performance for all the methods studied in this problem. PI combined with LDA, FLDA, MLP, or LSSVM classification methods obtained the best performance for the IT set. The CV estimation of the errors also indicated good performance by the classifiers. These results are also consistent with [11].

Statistical comparison of the results by classification problem

The number of classifiers by classification problem is shown in Table 5.5. Figure 5.3 shows the mean rank of each feature extraction method in a multiple comparison at a 0.05 level based on the Tukey's honestly significance difference criterion for Kruskal-Wallis nonparametric one-way analysis of variance [62].

5.3.2 The pre-processing techniques

Eight out of 50 semi-automatically preprocessed test cases were misclassified at least once by the pairwise BDK classifiers (GBM vs. MET excluded). Also, ten out of 47 of the automatically preprocessed test cases were misclassified at least once by the same

Table 5.5: Number of classifiers obtained for each classification Problem.

Task	Number of classifiers
GBM vs. MEN	41
GBM vs. MET	44
GBM vs. LGG	44
MEN vs. MET	43
MEN vs. LGG	40
LGG vs. MET	41
	253

classifiers. Based on these rates, no differences were observed in the classification of automatic and semi-automatic pre-processed signals. The semi-automatic pre-processing pipeline applied to the larger part of the test dataset was consistent with the automatic pipeline applied on the training set. This is an important practical conclusion because it suggests the compatibility of different pre-processing software tools, either in an automatic or a semi-automatic fashion for automatic classification in CDSSs.

5.3.3 The Feature Extraction methods

All the feature extraction methods applied in this study were based on PR. Therefore, we could not make any comparison between PR and metabolite quantification approaches. Approaches that take advantage of the combination of different TE [26, 27, 31, 63–65] were not considered in order to ensure that results could be compared with previous analyses of this type of data [8, 9, 11, 13, 28, 29, 66–68]. Furthermore, although a feature extraction evaluation is not the aim of the present study and the setup of this study is not designed specifically for it, some effects of the different feature extraction methods are reported.

Figure 5.4 shows the box-whisker plot of the performance (BER) for each Feature Extraction (FE) method. GBM vs. MET classifiers are not included because of their large difference in performance with respect to the other classification problems. The distributions of the results for all FE methods overlap, and no statistical differences were observed. Nevertheless, a noteworthy fact is the trend toward low values of the Peak Integration method compared to other methods. The study of Devos et al. [11] about the same four classes showed similar performances when comparing full region of interest, peak regions and PI. In [13], Simonetti et al. compared PCA, ICA, LCModel [68] and PI for feature extraction on Short TE MRSI data and they also obtained the best results with PI. In a single-center study, Opstad et al. [29] reported that the LCModel quantification obtained better results than PCA for two-step LDA classification. In Long TE spectra, Lukas et al. [12] observed a better performance using the full region of interest rather than using PI or peak region extraction. Finally, Menze et al. [14] and Luts et al. [69] obtained an improvement when PR approaches (e.g. ICA, PCA, binned peak region and WAV) were used in Short or Long TE instead of quantification approaches.

Statistical comparison of the results by Feature Extraction

The number of classifiers by Feature Extraction method is shown in Table 5.6. Figure 5.5 shows the mean rank of each feature extraction method in a multiple comparison at a

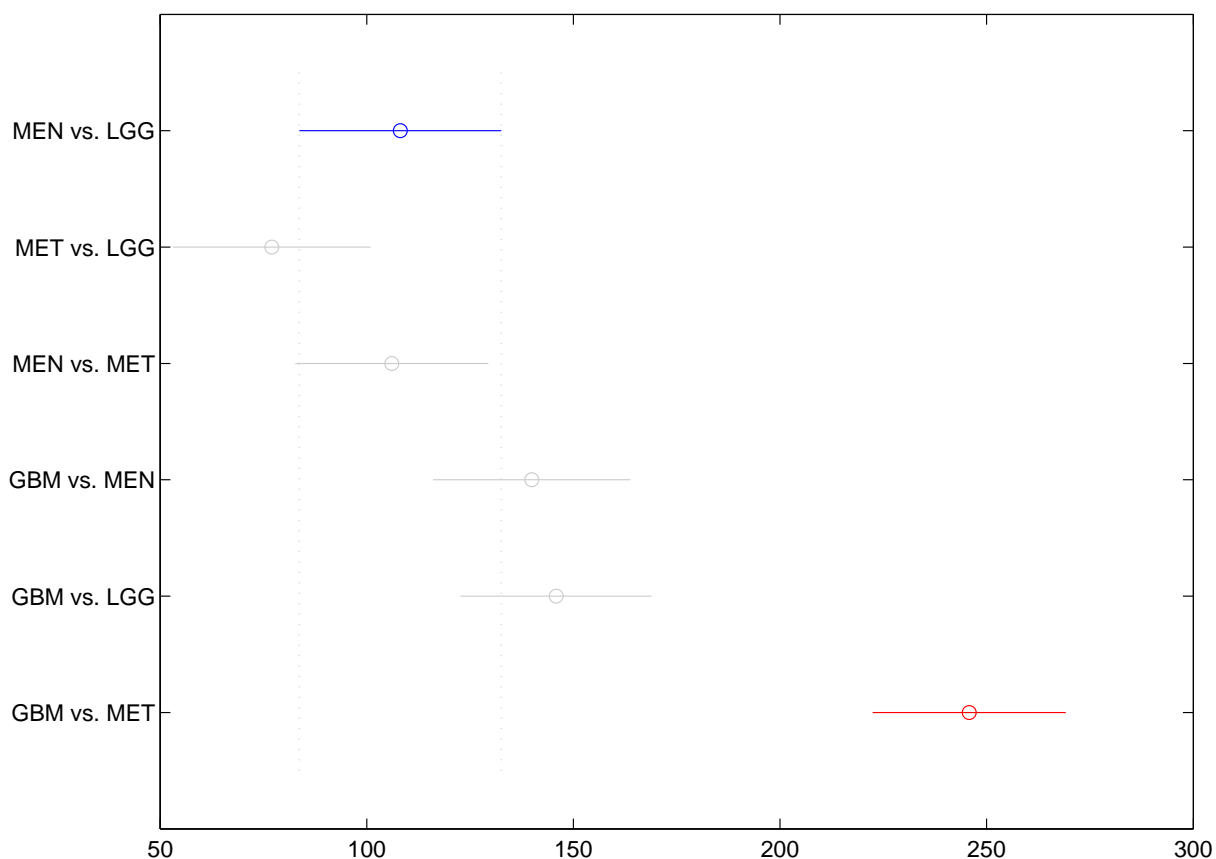


Figure 5.3: Multiple comparison of mean ranks of the classifiers by problem. 5 groups have mean ranks significantly different from GBM vs. MET.

0.05 level based on the Tukey's honestly significance difference criterion for Kruskal-Wallis nonparametric one-way analysis of variance.

5.3.4 The classification methods

The diversity of methods used for classification is broad enough to have a good overview of the effect that this selection has on the performance of the classifiers. Figure 5.6 shows the box-whisker plot of the performance (BER) for each classification method. Analogously to the analysis of FE methods, GBM vs. MET classifiers are not included in the distributions because of their large differences in performance with respect to the other classification methods. As observed in Figure 5.6, the distributions overlap, but in general, lower results of BER were obtained using a BDk. In [28], BDk was used in PI values to discriminate over tumor grades and other tissues in the INTERPRET multi-voxel dataset. The study of Devos et al. [11] observed similar performances of their LDA and LSSVM classifiers based on PI and evaluated by the area under the ROC curves. Tate et al. [8, 9] based their three-class classifiers on the LDA due to the ability of this method for projecting the results in a 2-dimensional space for visualization. Note that FLDA shows similar results

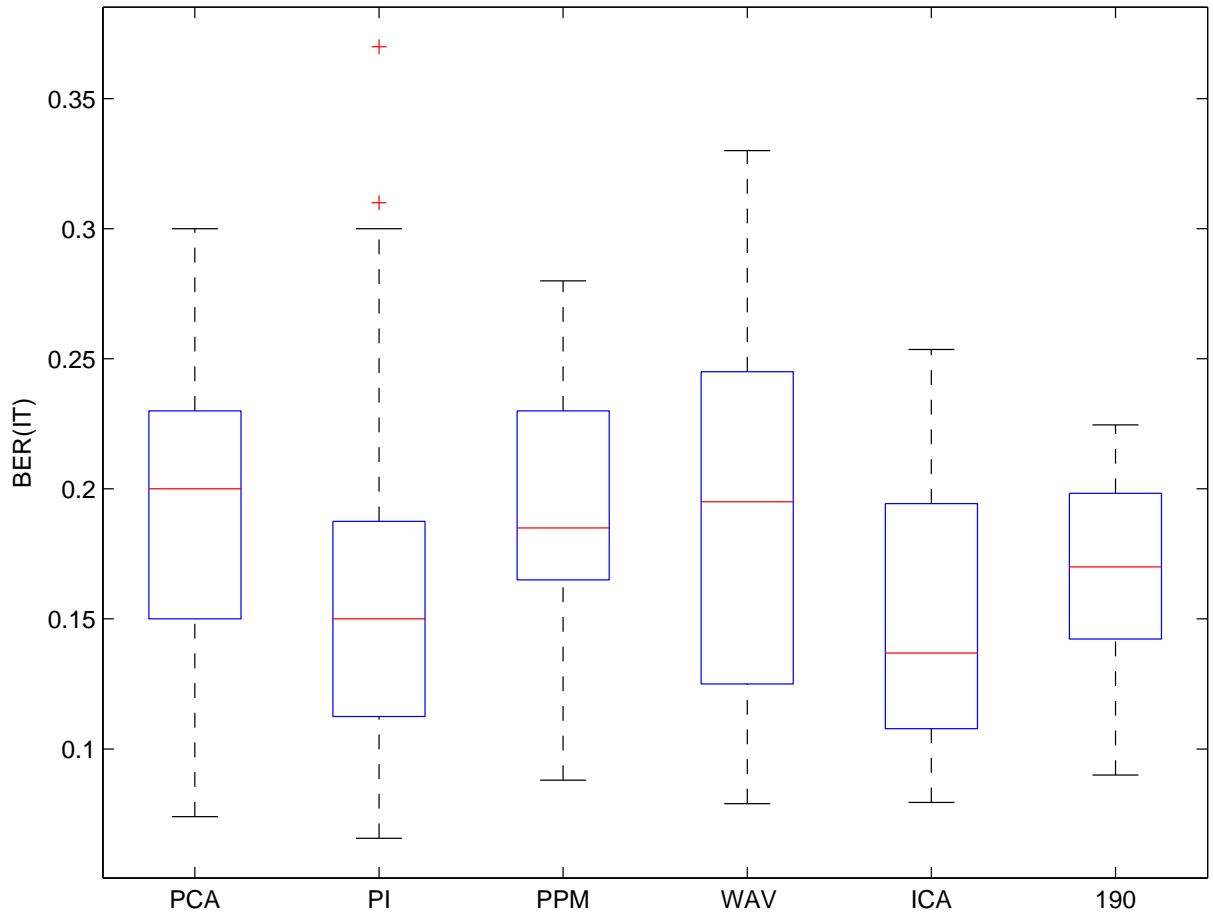


Figure 5.4: Box-whisker plots of the performance for each Feature Extraction method in the eTUMOUR dataset. Performance is measured in BER and the box-whisker characteristics are the same as in Figure 5.1.

when compared with the other methods in average; however, other methods like LSSVM and BDk might be preferable for some discrimination problems (e.g. GBM vs. LGG).

Finally, in Figure 5.2, we summarize and compare the BER estimation obtained by the CV for the INTERPRET training dataset and the IT consisting of the new eTUMOUR cases. Most of the results are in the ($BER(CV) < 0.2$, $BER(IT) < 0.3$) region, except for the GBM vs. MET problem, which had a sparse distribution. The general trend in this region is indicated by the black-dashed line. This indicates an underestimation of the BER by the CV evaluation. The underestimation is typically observed in the PR challenges [70], and it is usually produced by the overfitting of the models on the training dataset and the estimation of the error with non-fully independent samples [71]. A noteworthy feature of our study is the evaluation of the predictive models using the new subsequently acquired multicenter test, that ensures the independence of the training and test sets. With respect to the GBM vs. MET results, they are scattered in regions of larger error. For this problem, some overestimation of the CV error is also observed. This may show the difficulty of the problem and the randomness in the results. The results obtained for the rest of the discrimination problems confirm the expected behaviour of

Table 5.6: Number of classifiers by feature extraction method. GBM vs. MET classifiers are excluded.

Feature Extraction method	Number of classifiers
190 (Full region of interest)	18
ICA	21
PCA	50
PI	80
PPM	42
WAV	42

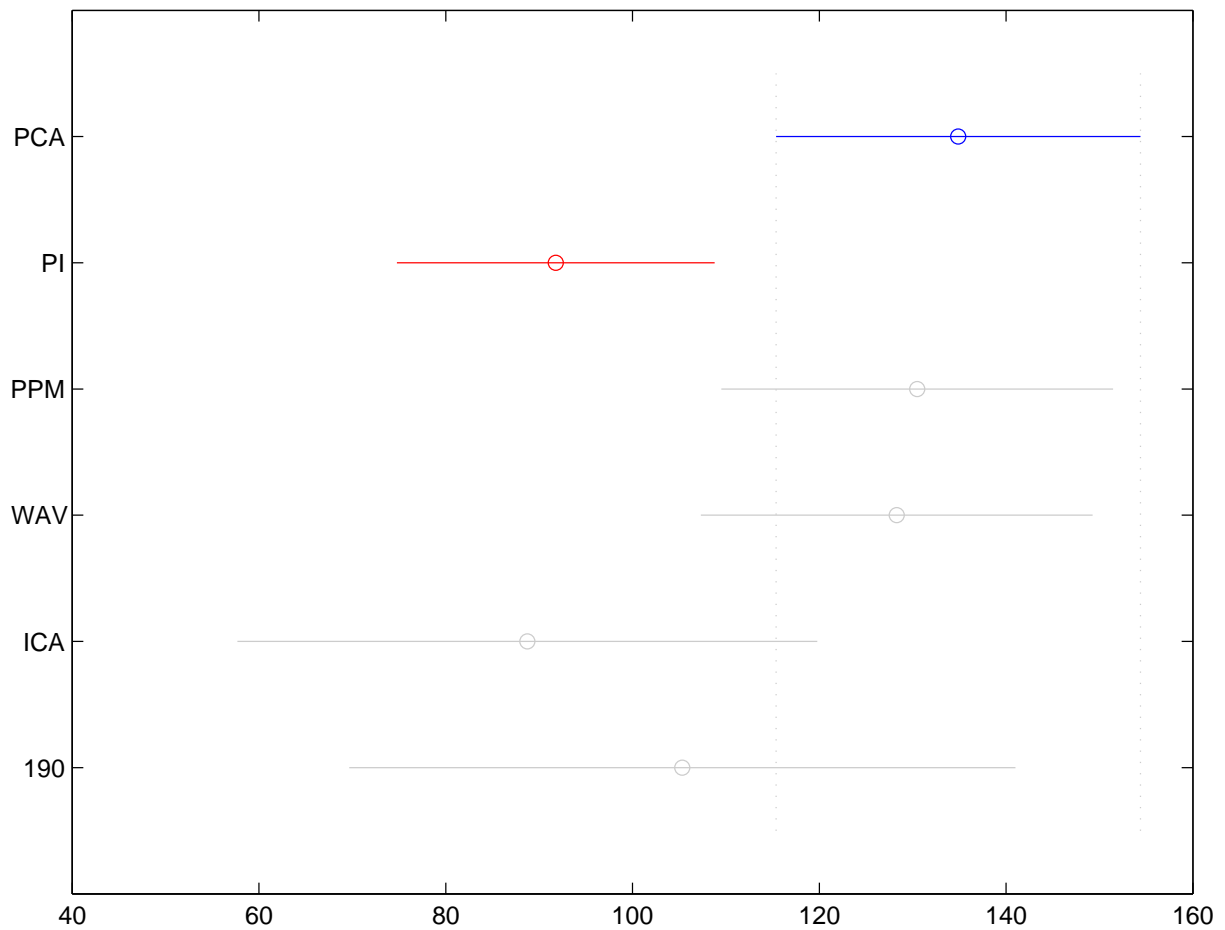


Figure 5.5: Multiple comparison of mean ranks of the classifiers by Feature Extraction method. GBM vs MET classifiers are excluded. 2 groups have mean ranks significantly different from PI.

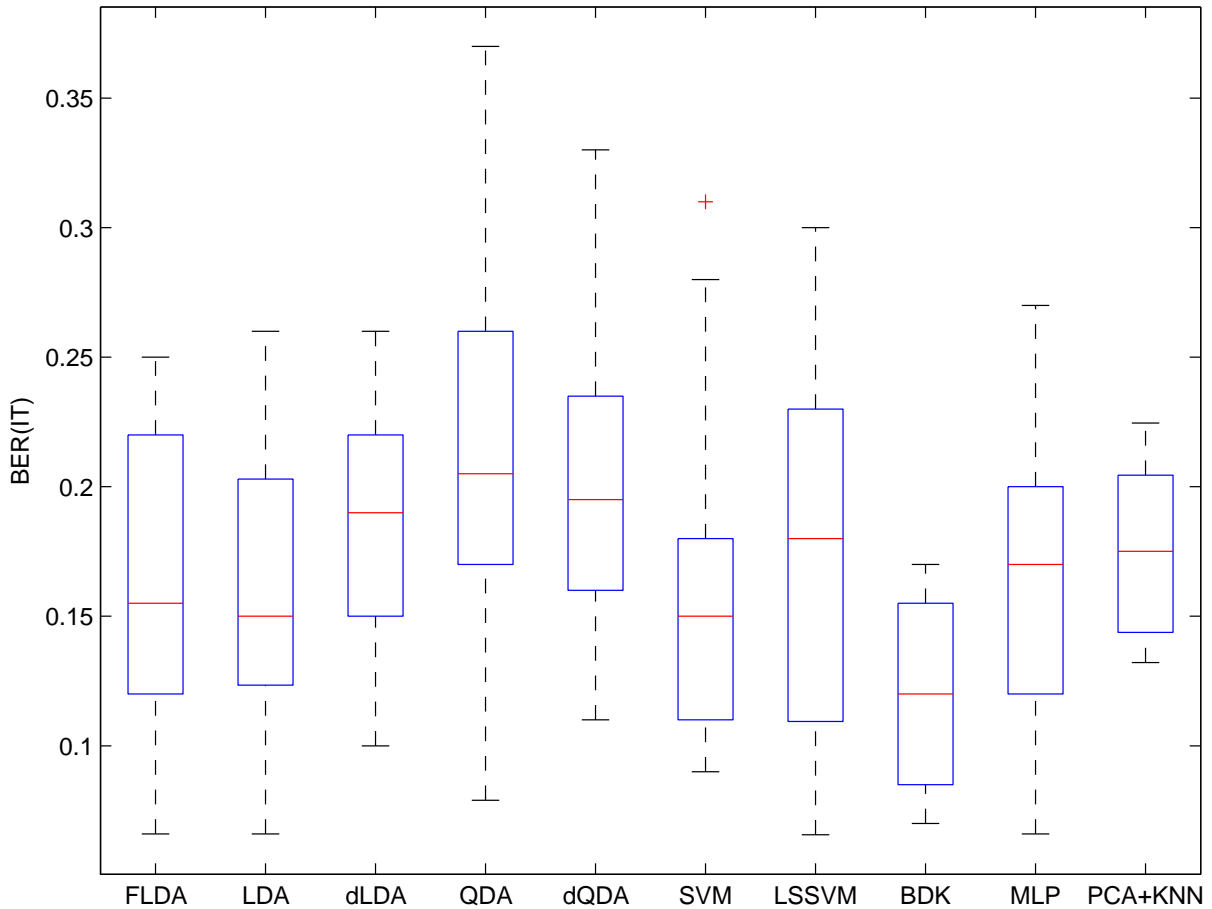


Figure 5.6: Box-whisker plots of the performance for each classification method in the eTUMOUR dataset. Performance is measured in BER and the box-whisker characteristics are the same as in Figure 5.1.

the predictive models.

Statistical comparison of the results by Classification Method

The number of classifiers by classification method is shown in Table 5.7. Figure 5.7 shows the mean rank of each classification method in a multiple comparison at a 0.05 level based on the Tukey's honestly significance difference criterion for Kruskal-Wallis nonparametric one-way analysis of variance.

5.3.5 Use of the study for automatic validation of MRS entries in brain tumour datasets

An intuitive method to compare datasets of signals is the visual inspection of their prototypical patterns. Figure 5.8 shows plots of the unimodal prototypes of the Short TE spectra for the four tumour groups of the training and test datasets. Each prototype is represented by the unsmoothed mean function and the mean function \pm the standard deviation function. The view is zoomed in the $[0.5, 4.1]ppm$ region used in our experi-

Table 5.7: Number of classifiers by classification method. GBM vs. MET classifiers are excluded.

Classification Method	Number of classifiers
BDK	6
FLDA	24
LDA	34
LSSVM	51
MLP	30
PCA+KNN	12
QDA	36
SVM	12
dLDA	24
dQDA	24

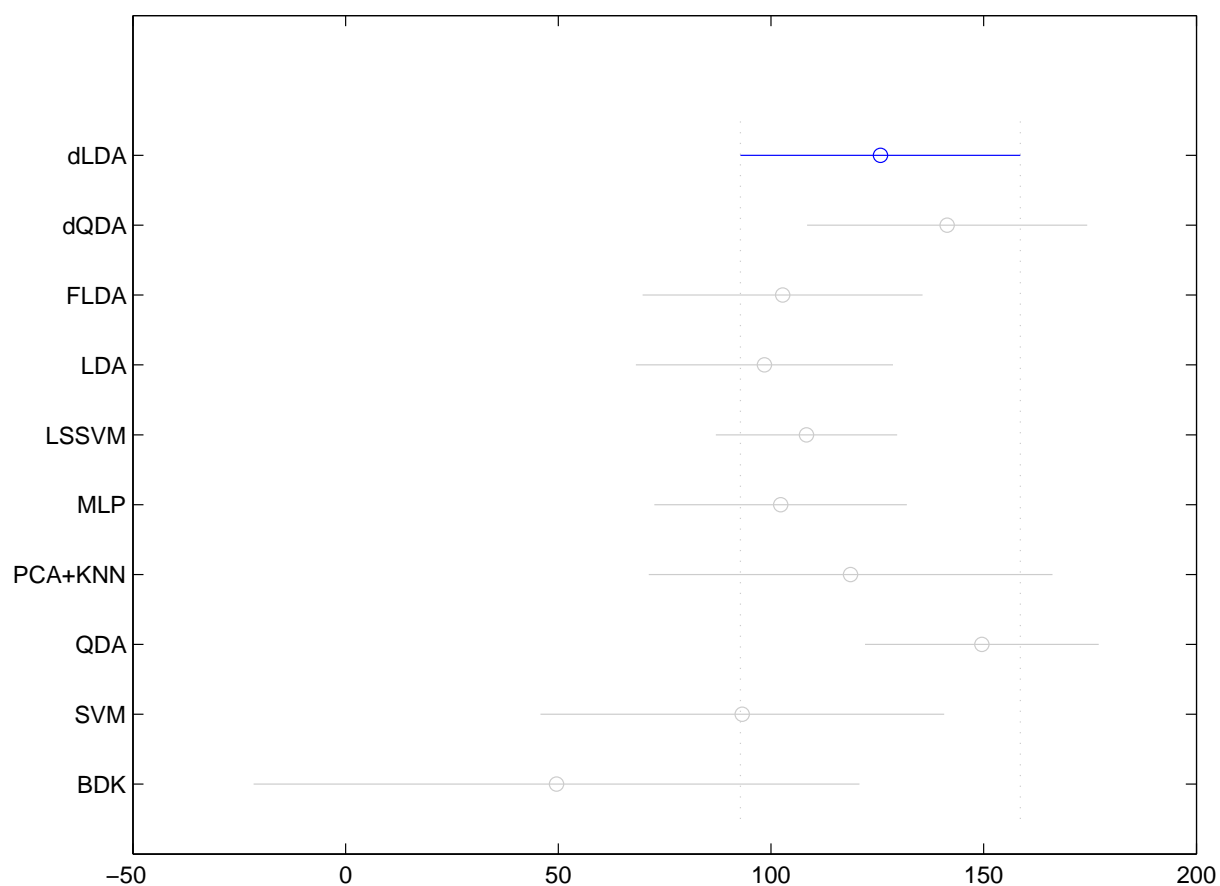


Figure 5.7: Multiple comparison of mean ranks of the classifiers by Classification method. GBM vs MET classifiers are excluded. No groups have mean ranks significantly different from LDA.

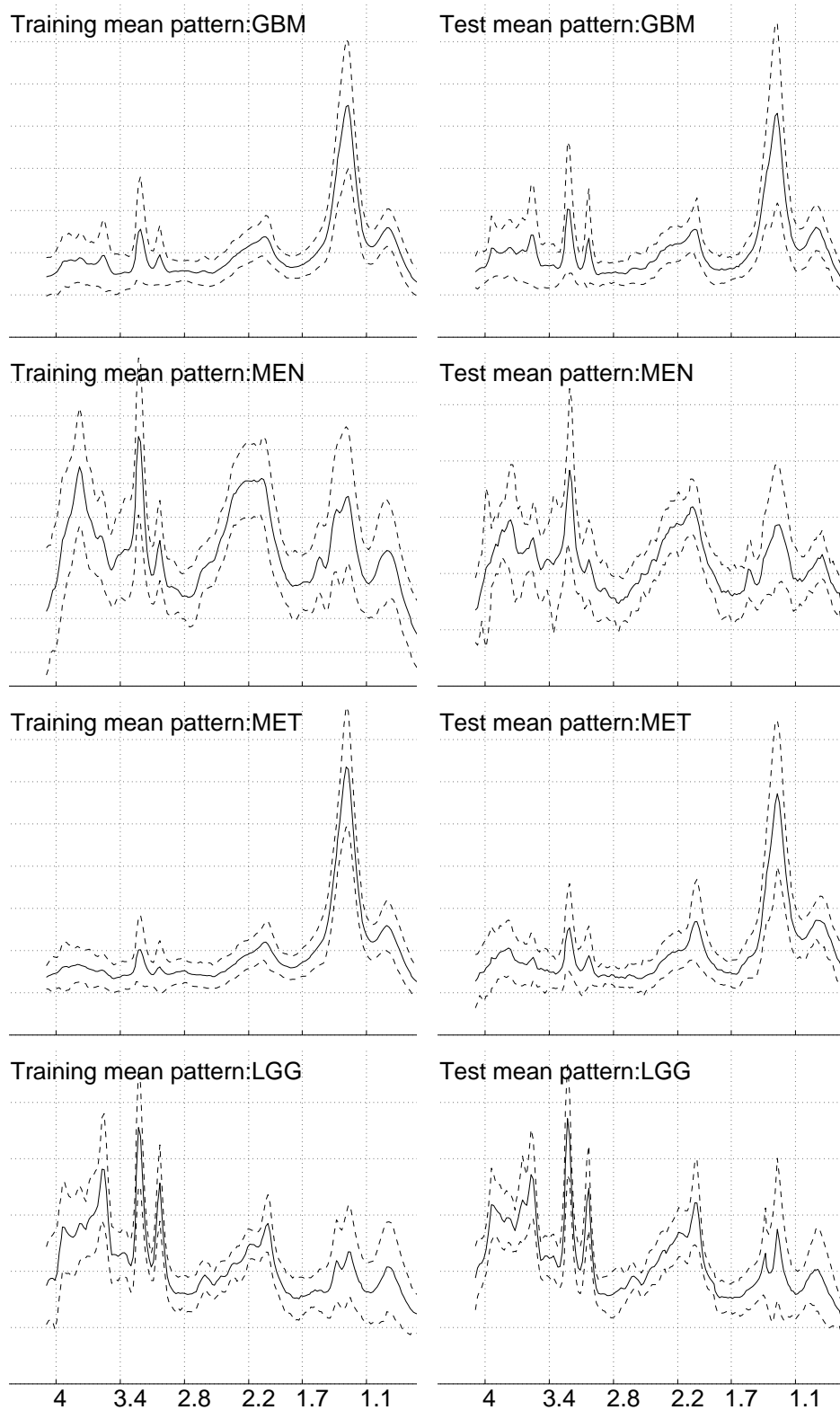


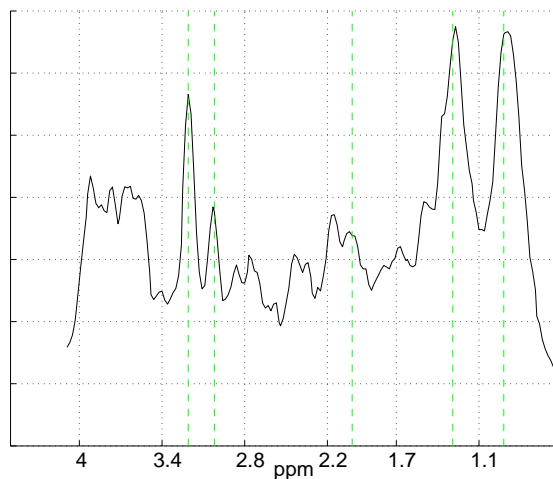
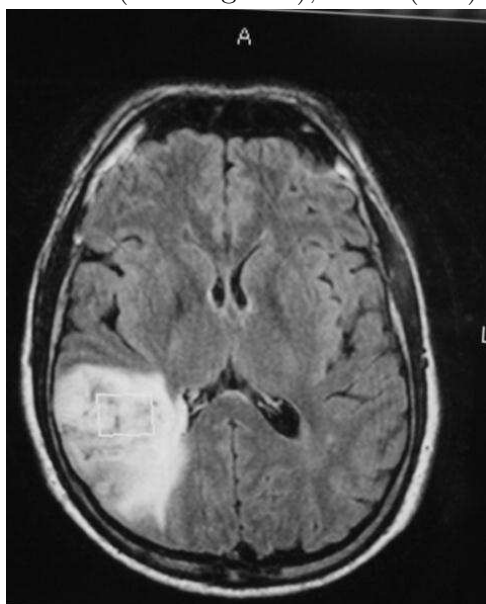
Figure 5.8: Unimodal prototypes of the Short TE spectra for the four tumour groups of the training and test datasets. Each prototype is represented by the unsmoothed mean function and the mean \pm the standard deviation function. The view is zoomed in the $[0.5, 4.1]$ ppm region used in our experiments.

ments. The observed resonances correspond to the main compounds reported in page 80 of Section 5.1. In general, the training and test prototype patterns for GBM, MET and LGG are close to each other, whereas the MEN prototype differs visually more. This may be because of a higher standard deviation on the test dataset around the 3.21ppm peak with respect to the training dataset. Besides, the variation around the 2.2ppm is higher in the test-set mean than in the training one.

A practical result of this study is that cases that are repeatedly misclassified by the different techniques can be flagged as being susceptible of revision for possible problems in voxel positioning, acquisition artifact, normal-tissue contamination, or limitation in the classification methodology (e.g. patterns replicated in non-tumoral diseases, atypical MRS patterns and underrepresented tumor subtypes). In this way, even in the absence of biopsy, PR techniques can contribute to the automatic validation of cases, assisting the specialists on the detection of potential source of errors in the biomedical data acquired from patients.

Figures 5.9 and 5.10 show some eTUMOUR misclassified cases which may be interesting to review. The eTUMOUR case et2274 was diagnosed by the original pathologist as oligodendroglioma 9450/3 (grade II, WHO), although a comment was added to the free text section of the eTUMOUR database (eTDB) making reference to the presence of areas of anaplastic oligodendroglioma (grade III, WHO). Still, the final diagnosis proposed was grade II oligodendroglioma. The voxel allocation was carried out following the eTUMOUR acquisition protocol. The ML pattern is uncommon, as the high 0.9 and 1.3ppm resonances show. The disappearance of these resonances at Long TE (136 ms) discards a significant necrotic contribution (results not shown, but see [31]). This pattern has been observed before [31], for example in the INTERPRET cases I0450 (oligoastrocytoma) and I0179 (oligodendroglioma), which are also misplaced in the Short TE latent space of the INTERPRET DSS 2.0 (<http://azizu.uab.es/INTERPRET>). In summary, et2274 seems to behave as a class outlier and its consistent misclassification in our analysis may be sampling precisely that. The eTUMOUR case et2206 was originally diagnosed as oligoastrocytoma 9382/3 (grade II, WHO), but there were some discrepancies regarding the glial subtype on the validation done by the pathological committee. It was misclassified by every MET vs. LGG classifier, and also by some GBM vs. LGG and MEN vs. LGG classifiers. Its ML pattern at Short TE is also uncommon, having relatively large 0.9, 1.3 and 2.8ppm peaks that are reduced at Long TE (results not shown), which suggests, as well, a non-necrotic origin. The eTUMOUR case et2349 is a GBM without clearly visible ML, which was misclassified in every classification problem. The review of the experts did not indicate problems in the location of the voxel, being this mainly positioned in the highly cellular part of the tumour. The eTUMOUR case et2197 is a MET with possible MRS pattern contribution from normal brain parenchyma, as it could be deduced by the relative difference of size between the voxel used for acquisition and the small brain lesion. Its pattern shows similar Cho and Cr peak heights and relatively high NAA at 2ppm). However, the appearance of high Lac/ML at 1.3ppm at the same time suggests abnormality. Nonetheless, it is clearly an uncommon spectral pattern for a MET.

et2274 (T1-weighted), LGG (OD)



et2206 (T2-weighted), LGG (OA or OD)

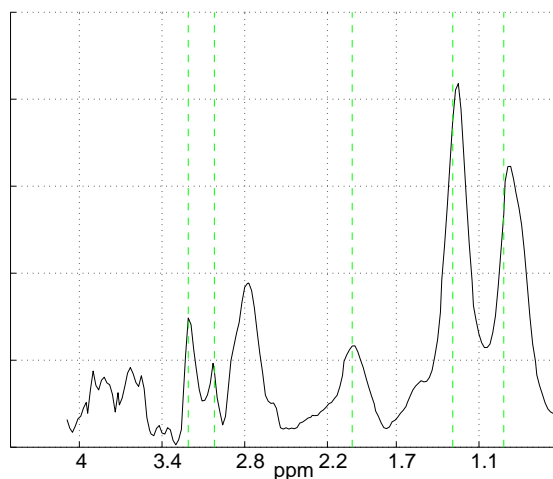
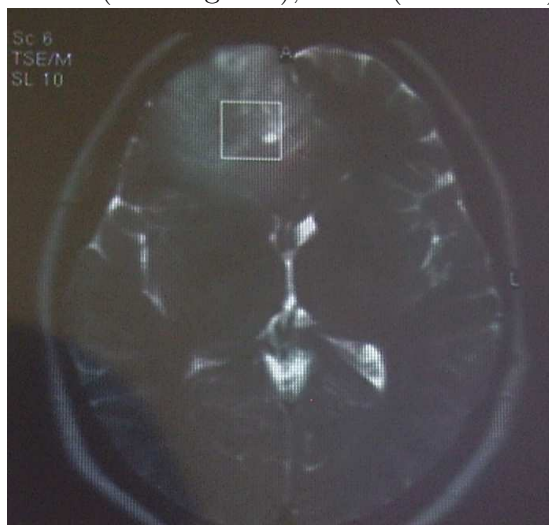


Figure 5.9: Potential outliers (1/2) detected as a consequence of this study. Case numbering corresponds to eTUMOUR database (www.etumour.net) entries). For each case, the reference image and voxel location is shown on the left, and the region of interest of the real part of the Short TE spectrum is shown on the right. For an easier visualization of the spectrum, vertical dashed lines indicates the position of the main resonances: Cho (3.21ppm), Cr (3.02), NAA (2.01ppm), L1 (1.29ppm), L2 (0.92ppm).

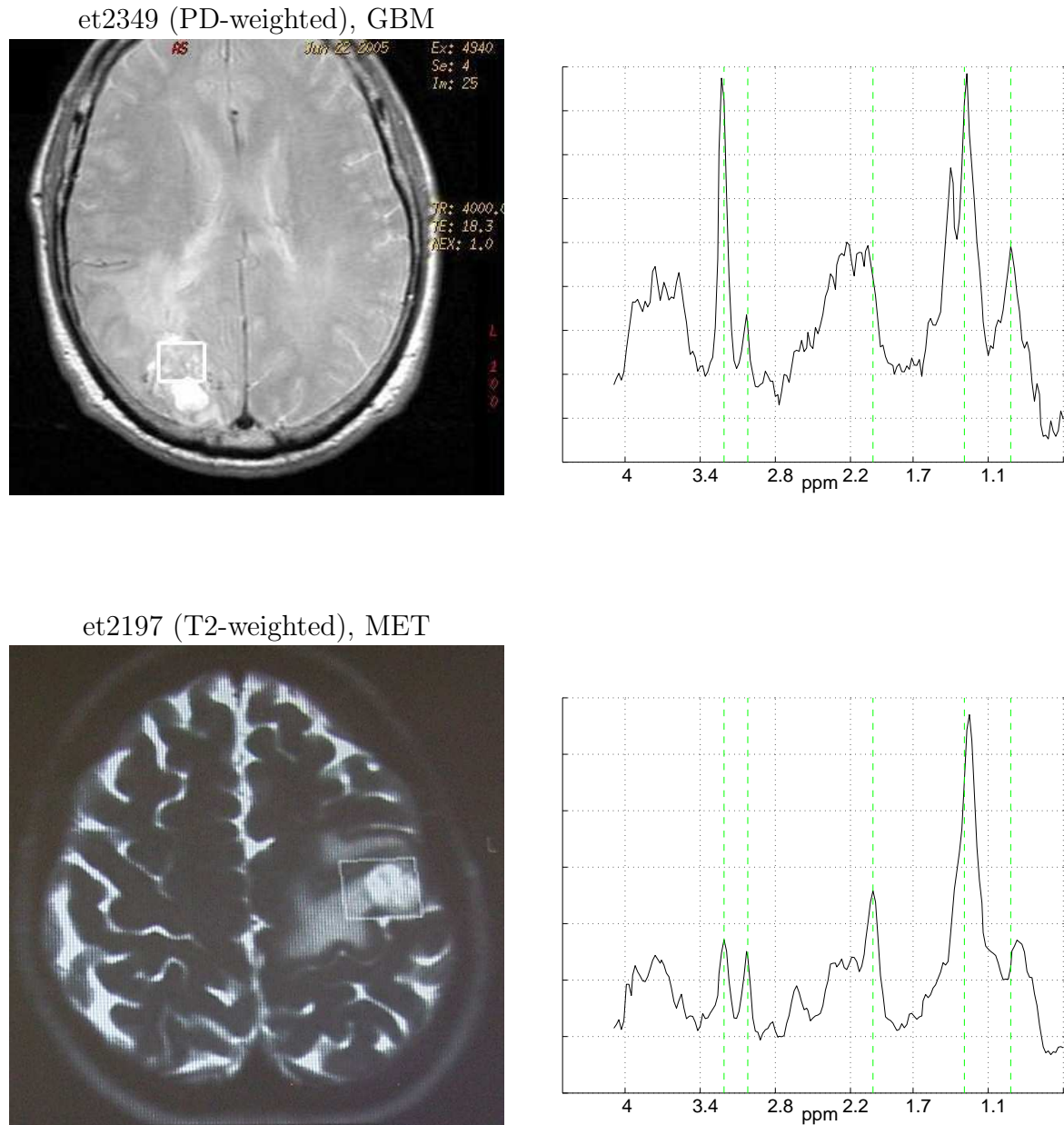


Figure 5.10: Potential outliers (2/2) detected as a consequence of this study. Figure characteristics are the same as in Figure 5.9.

Table 5.8: Results for the six pairwise classification problems based on the EMBTD and EMBTD_r parametric spaces. The Error (ERR) and Balanced Error Rate (BER) estimates based on the cross-validation (CV) over the INTERPRET data and on the independent eTumour test set (IT) are shown. Columns in the table are the same as in Table 5.4.

id	Features	Classif	CV		IT	
			ERR	BER	ERR	BER
GBM vs. MEN						
49	EMBTD	trainb	0.11	0.11	0.13	0.15
50	EMBTD _r	quadra	0.15	0.15	0.16	0.15
GBM vs. MET						
51	EMBTD	linear	0.5	0.49	0.47	0.46
52	EMBTD _r	trainr	0.5	0.51	0.37	0.36
GBM vs. LGG						
53	EMBTD	trainr	0.21	0.22	0.17	0.16
54	EMBTD _r	LS-SVM	0.11	0.15	0.21	0.2
MEN vs. MET						
55	EMBTD	LS-SVM	0.11	0.12	0.2	0.17
56	EMBTD _r	linear	0.064	0.062	0.2	0.17
MEN vs. LGG						
57	EMBTD	linear	0.2	0.23	0.16	0.16
58	EMBTD _r	LS-SVM	0.14	0.17	0.22	0.21
MET vs. LGG						
59	EMBTD	LS-SVM	0.071	0.072	0.12	0.1
60	EMBTD _r	LS-SVM	0.086	0.088	0.13	0.14

5.4 Evaluation of the brain tumor classifiers based on the EMBTD and EMBTD_r parametric spaces

The EMBTD and EMBTD_r estimates, introduced in Chapter 4, obtain the parameters that defines the mixture model for a MRS spectrum. As shown in Chapter 4, it is possible to classify new brain tumor spectra based on the parametric space associated to the EMBTD and EMBTD_r. In our approach, the π_c parameters estimated by EMBTD and EMBTD_r are assumed to incorporate the information of the main resonances of the metabolites that appears in the tumoral tissues.

Here we evaluated the performance of the classifiers based on the EMBTD and the EMBTD_r parametric spaces in new multiproject-multicentric cases. The acquired data and the evaluation protocol for these classifiers was exactly the same as the one introduced in Section 5.2. Table 5.8 shows the best results obtained with the EMBTD and the EMBTD_r parametric spaces on different classification methods for the six classification problems. In general, the best results obtained with the EMBTD and the EMBTD_r parametric spaces are comparable in BER with the best results obtained with the other feature extraction methods (see Table 5.4).

5.5 Conclusions

This study describes a multiproject-multicenter evaluation of automated brain tumor classifiers using single-voxel Short TE MR spectra. To our knowledge, there is no previous work that evaluates predictive models trained with data acquired from a multicenter project using a new independent test set subsequently acquired from partly different centers. Classifiers were trained with cases acquired by six centers during the 2000-2002 period. They were tested with posterior cases acquired by eight institutions during the 2004-2007 period. This strategy provides a view that is close to a real environment where similar classifiers, integrated in a CDSS, may be used in multiple hospitals to assist in the diagnosis of new cases.

Our major conclusion is that accurate classification of those new cases is feasible using data acquired in different hospitals, different instrumentation, but similar acquisition protocols. Specifically, in our experiments, classifiers developed from the INTERPRET dataset seem to be robust enough for predictive classification of prospective cases from eTUMOUR.

The pairwise discrimination between Glioblastoma, Meningioma, Metastasis, and Low-grade Glial achieved accuracies of around 90%. However, the discrimination of Glioblastoma and Metastasis did not achieve a result better than 78% accuracy. Our results consolidate the conclusions of previous studies on automatic brain tumor classification using MRS but with multiproject-multicenter data for training and subsequent test.

In general, the results obtained with the EMBTD and the EMBTD_r parametric spaces are comparable in BER with the best results obtained with the other feature extraction methods.

A well-defined protocol for the acquisition of MRS (e.g. spectral parameters and voxel localization), and the application of quality control to MRS spectra should allow the reproducibility of such classification rules and the successful use of DSSs in clinical environments.

The methodology provided in the present study may also be of use as "automatic flaggers" to help in the quality control of cases during the eTUMOUR multicenter project and beyond. The approach used in this work could be of use for pediatric brain tumour related studies [72] aimed at providing predictive information to pediatric neurosurgeons.

Hence, the conclusions obtained in this study are directly applicable to several of the tasks associated to a CDSS development for brain tumor diagnosis and prognosis and its deployment in clinical environments.

Hence, the conclusions obtained in this study are directly applicable to several of the tasks associated to CDSS development for brain tumor diagnosis and prognosis and its deployment in clinical environments.

Bibliography

- [1] Garcia-Gomez JM, Luts J, Julia-Sape M, Krooshof P, Tortajada S, Robledo JV, Melssen W, Fuster-Garcia E, Olier I, Postma G, Monleon D, Moreno-Torres A, Pujol J, Candiota AP, Martinez-Bisbal MC, Suykens J, Buydens L, Celda B, Van Huffel S, Arus C, Robles M (2009) Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *MAGMA* 22(1): 5–18
- [2] Howe FA, Opstad KS (2003) ¹H MR spectroscopy of brain tumours and masses. *NMR Biomed* 16(3): 123–131
- [3] Galanaud D, Nicoli F, Chinot O, Confort-Gouny S, Figarella-Branger D, Roche P, Fuentes S, Le Fur Y, Ranjeva JP, Cozzone PJ (2006) Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy. *Magnetic Resonance in Medicine* 55(6): 1236–1245
- [4] Arnold DL, De Stefano N (1997) Magnetic resonance spectroscopy in vivo: applications in neurological disorders. *Ital J Neurol Sci* 18(6): 321–329
- [5] Poptani H, Kaartinen J, Gupta RK, Niemitz M, Hiltunen Y, Kauppinen RA (1999) Diagnostic assessment of brain tumours and non-neoplastic brain disorders in vivo using proton nuclear magnetic resonance spectroscopy and artificial neural networks. *J Cancer Res Clin Oncol* 125(6): 343–349
- [6] Moller-Hartmann W, Herminghaus S, Krings T, Marquardt G, Lanfermann H, Pilatus U, Zanella FE (2002) Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. *Neuroradiology* 44(5): 371–381
- [7] Hagberg G (1998) From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. *NMR in Biomedicine* 11(4-5): 148–156
- [8] Tate AR, Majos C, Moreno A, Howe FA, Griffiths JR, Arús C (2003) Automated classification of short echo time in in vivo ¹H brain tumor spectra: a multicenter study. *Magnetic Resonance in Medicine* 49(1): 29–36
- [9] Tate AR, Underwood J, Acosta DM, Julia-Sape M, Majos C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabanas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Remy C, Heerschap A, Watson D, Griffiths JR, Arús C (2006) 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

- [10] González-Vélez H, Mier M, Julià-Sapé M, Arvanitis T, García-Gómez J, Robles M, Lewis P, Dasmahapatra S, Dupplaw D, Peet A, Arús C, Celda B, Van Huffel S, Lluch-Ariet M (2007) HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis. *Applied Intelligence* [Article in Press]
- [11] Devos A, Lukas L, Suykens JAK, Vanhamme L, Tate AR, Howe FA, Majos C, Moreno-Torres A, van der Graaf M, Arús C, Van Huffel S (2004) Classification of brain tumours using short echo time 1H MR spectra. *J Magn Reson* 170(1): 164–175
- [12] Lukas L, Devos A, Suykens JAK, Vanhamme L, Howe FA, Majós C, Moreno-Torres A, Graaf MVD, Tate AR, Arús C, Huffel SV (2004) Brain tumor classification based on long echo proton MRS signals. *Artif Intell Med* 31: 73–89
- [13] Simonetti AW, Melssen WJ, Szabo de Edelenyi F, van Asten JJA, Heerschap A, Buydens LMC (2005) Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. *NMR in Biomedicine* 18(1): 34–43
- [14] Menze BH, Lichy MP, Bachert P, Kelm BM, Schlemmer HP, Hamprecht FA (2006) Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors. *NMR in Biomedicine* 19(5): 599–609
- [15] Potts HWW, Wyatt JC, Altman DG (2001) Challenges in Evaluating Complex Decision Support Systems: Lessons from Design-a-Trial. In *AIME '01: Proceedings of the 8th Conference on AI in Medicine in Europe*, pp. 453–456. Springer-Verlag, London, UK
- [16] Lisboa PJ, Taktak AFG (2006) The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Netw* 19(4): 408–415
- [17] Anagnostou T, Remzi M, Djavan B (2003) Artificial neural networks for decision-making in urologic oncology. *European Urology* 43(6): 596–603
- [18] Perner P (2006) Intelligent data analysis in medicine-recent advances. *Artif Intell Med* 37(1): 1–5
- [19] INTERPRET Consortium (Accessed: 28 April 2008) Interpret web site. <http://azizu.uab.es/INTERPRET>
- [20] Julia-Sape M, Acosta D, Mier M, Arús C, Watson D (2006) A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. *Magn Reson Mater Phy* 19(1): 22–33
- [21] van der Graaf M, Julia-Sape M, Howe FA, Ziegler A, Majos C, Moreno-Torres A, Rijpkema M, Acosta D, Opstad KS, van der Meulen YM, Arus C, Heerschap A (2008) MRS quality assessment in a multicentre study on MRS-based classification of brain tumours. *NMR Biomed* 21(2): 148–158
- [22] Devos A (2005) Quantification and classification of Magnetic Resonance Spectroscopy data and applications to brain tumour recognition. Ph.D. thesis, Faculty of Engineering, K.U.Leuven

-
- [23] Simonetti AW, Melssen WJ, van der Graaf M, Postma GJ, Heerschap A, Buydens LMC (2003) A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy. *Anal Chem* 75(20): 5352–5361
- [24] Devos A, Simonetti AW, van der Graaf M, Lukas L, Suykens JAK, Vanhamme L, Buydens LMC, Heerschap A, Van Huffel S (2005) The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. *J Magn Reson* 173(2): 218–228
- [25] Laudadio T, Martinez-Bisbal M, Celda B, Van Huffel S (2007) Fast nosological imaging using canonical correlation analysis of brain data obtained by two-dimensional turbo spectroscopic imaging. *NMR in Biomedicine* 21(4): 311–321
- [26] Martinez-Bisbal MC, Celda B, Marti-Bonmati L, Ferrer P, Revert-Ventura AJ, Piquer J, Molla E, Arana R, Dosda-Munoz R (2002) The contribution of Magnetic Resonance Spectroscopy for the classification of high grade glial tumours. The predictive value of macromolecules. *Revista de Neurología* 34(309-313)
- [27] Martinez-Bisbal MC, Ferrer-Luna R, Martinez-Granados B, Monleón D, Esteve V, Piquer J, Revert AJ, Mollá E, Martí-Bonmatí L, Celda B (2005) Glial tumours grading by a combination of (1)H MR short and medium echo time single voxel located by spectroscopic imaging. *Magnetic Resonance Materials in Physics* 18(S1): S68
- [28] Melssen W, Wehrens R, Buydens L (2006) Supervised Kohonen networks for classification problems. *Chemometrics and Intelligent Laboratory Systems* 83(2): 99–113
- [29] Opstad KS, Ladroue C, Bell BA, Griffiths JR, Howe FA (2007) Linear discriminant analysis of brain tumour 1H MR spectra: a comparison of classification using whole spectra versus metabolite quantification. *NMR in Biomedicine* 20(8): 763–770
- [30] 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. Technical report, FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC <http://www.etumour.net>
- [31] Garcia-Gomez J, Tortajada S, Vidal C, Julia-Sape M, Luts J, Van Huffel S, Arús C, Robles M (2008) The influence of combining two echo times in automatic brain tumor classification by Magnetic Resonance Spectroscopy. *NMR in Biomedicine* 21(10): 1112–1125
- [32] Kleihues P, Burger PC, Scheithauer BW (1993) The new WHO classification of brain tumours. *Brain Pathol* 3(3): 255–268
- [33] Klose U (1990) In vivo proton spectroscopy in presence of eddy currents. *Magnetic Resonance in Medicine* 14(1): 26–30
- [34] Naressi A, Couturier C, Castang I, de Beer R, Graveron-Demilly D (2001) Java-based graphical user interface for MRUI, a software package for quantitation of in vivo/medical magnetic resonance spectroscopy signals. *Comput Biol Med* 31(4): 269–286

- [35] Cabanes E, Confort-Gouny S, Le Fur Y, Simond G, Cozzone PJ (2001) Optimization of residual water signal removal by HLSVD on simulated short echo time proton MR spectra of the human brain. *J Magn Reson* 150(2): 116–125
- [36] Hoch JC, Stern AS (1996) *NMR Data Processing*. John Wiley and Sons, Inc., New York, NY
- [37] Preul MC, Caramanos Z, Collins DL, Villemure JG, Leblanc R, Olivier A, Pokrupa R, Arnold DL (1996) Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. *Nat Med* 2(3): 323–325
- [38] Burges CJ (2004) *Geometric Methods for Feature Extraction and Dimensional Reduction: A Guided Tour*. Technical report, Microsoft Research, University of Toronto http://research.microsoft.com/~cburges/tech_reports/tr-2004-55-dimred.pdf
- [39] Fukunaga K (1990) *Introduction to statistical pattern recognition* (2nd ed.). Academic Press Professional, Inc., San Diego, CA, USA
- [40] Comon P (1994) Independent component analysis, a new concept? *Signal Processing* 36(3): 287–314
- [41] J-F Cardoso, A Souloumiac (1993) Blind beamforming for non Gaussian signals. *IEE Proceedings-F* 140(6): 362–370
- [42] Daubechies I (1992) *Ten Lectures on Wavelets* (CBMS - NSF Regional Conference Series in Applied Mathematics). Soc for Industrial & Applied Math
- [43] Panagiotacopulos N, Lertsuntivit S, Savidge L, Lin A, Shic F, Ross B (2000) Wavelet Analysis of Brain Tumors in Clinical MRS. In *Advances in Physics, Electronics and Signal Processing Applications*, pp. 290–296
- [44] Krzanowski WJ, editor (1988) *Principles of multivariate analysis: a user’s perspective*. Oxford University Press, Inc., New York, NY, USA
- [45] Fisher RA (1925) *Statistical methods for research workers*. Oliver and Boyd, Edinburgh, UK
- [46] Vapnik V (1995) *The Nature of Statistical Learning Theory*. Springer, N.Y.
- [47] Suykens JAK, Vandewalle J (1999) Least Squares Support Vector Machine Classifiers. *Neural Process Lett* 9(3): 293–300
- [48] Rosenblatt F (1958) The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 65(6): 386–408
- [49] Melssen W, Ustun B, Buydens L (2007) SOMPLS: A supervised self-organising map–partial least squares algorithm for multivariate regression problems. *Chemometrics and Intelligent Laboratory Systems* 86(1): 102–120
- [50] Valentini G, Dietterich TG (2004) Bias-Variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods. *Journal of Machine Learning Research* 5: 725–775

-
- [51] Hastie T, Tibshirani R, Friedman JH (2001) *The Elements of Statistical Learning*. Springer
- [52] Duda R, Hart P, Stork D (2001) *Pattern Classification*. John Wiley and Sons, inc.
- [53] Van Gestel T, JAK Suykens, Lanckriet G, Lambrechts A, De Moor B, Vandewalle J (2002) Bayesian Framework for Least Squares Support Vector Machine Classifiers, Gaussian Processes and Kernel Fisher Discriminant Analysis. *Neural Computation* 14: 1115–1147
- [54] MacKay DJC (1992) Bayesian Interpolation. *Neural Computation* 4(3): 415–447
- [55] Ishimaru H, Morikawa M, Iwanaga S, Kaminogo M, Ochi M, Hayashi K (2001) Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. *Eur Radiol* 11(9): 1784–1791
- [56] Opstad KS, Murphy MM, Wilkins PR, Bell BA, Griffiths JR, Howe FA (2004) Differentiation of metastases from high-grade gliomas using short echo time 1H spectroscopy. *J Magn Reson Imaging* 20(2): 187–192
- [57] Law M, Cha S, Knopp EA, Johnson G, Arnett J, Litt AW (2002) High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. *Radiology* 222(3): 715–721
- [58] Burtscher IM, Skagerberg G, Geijer B, Englund E, Stahlberg F, Holtas S (2000) Proton MR spectroscopy and preoperative diagnostic accuracy: an evaluation of intracranial mass lesions characterized by stereotactic biopsy findings. *AJNR Am J Neuroradiol* 21(1): 84–93
- [59] Laudadio T, Luts J, Martinez-Bisbal M, Celda B, Huffel SV (2008) Differentiation between brain metastasis and glioblastoma using MRI and two-dimensional turbo spectroscopic imaging data. In Proc. of the 4th European Medical and Biomedical Engineering congress, p. Accepted for publication.
- [60] Hochberg Y, Tamhane AC (1987) *Multiple comparison procedures*. John Wiley & Sons, Inc., New York, NY, USA
- [61] Celda B, Monleon D, Martinez-Bisbal MC, Esteve V, Martinez-Granados B, Pinero E, Ferrer R, Piquer J, Marti-Bonmati L, Cervera J (2006) MRS as endogenous molecular imaging for brain and prostate tumors: FP6 project "eTUMOR". *Adv Exp Med Biol* 587: 285–302
- [62] Gibbons J (1971) *Nonparametric Statistical Inference*
- [63] Tortajada S, García-Gómez JM, Vidal C, Arús C, Julià-Sapé M, Moreno A, Robles M (2006) Improved classification by pattern recognition of brain tumours combining long and short echo time 1H-MR spectra. In SpringerLink, editor, *Book of Abstracts ESMRMB 2006 - Supplement 1, Journal Magnetic Resonance Materials in Physics, Biology and Medicine*, volume 19, pp. 168–169

- [64] García-Gómez JM, Tortajada S, Vicente J, Sáez C, Castells X, Luts J, Julià-Sapé M, Juan-Císcar A, Van Huffel S, Barcelo A, Ariño J, Arús C, Robles M (2007) Genomics and Metabolomics Research for Brain Tumour Diagnosis Based on Machine Learning. In IWANN, Lecture Notes in Computer Sciences 4507, pp. 1012–1019
- [65] McIntyre DJO, Charlton RA, Markus HS, Howe FA (2007) Long and short echo time proton magnetic resonance spectroscopic imaging of the healthy aging brain. *J Magn Reson Imaging* 26(6): 1596–1606
- [66] Majos C, Julia-Sape M, Alonso J, Serrallonga M, Aguilera C, Acebes JJ, Arús C, Gili J (2004) Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE. *AJNR Am J Neuroradiol* 25(10): 1696–1704
- [67] Julia-Sape M, Acosta D, Majos C, Moreno-Torres A, Wesseling P, Acebes JJ, Griffiths JR, Arús C (2006) Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. *J Neurosurg* 105(1): 6–14
- [68] Provencher SW (2001) Automatic quantitation of localized in vivo ^1H spectra with LCMoDel. *NMR in Biomedicine* 14(4): 260–264
- [69] Luts J, Pouillet JB, Garcia-Gomez JM, Heerschap A, Robles M, Suykens JAK, Van Huffel S (2008) Effect of feature extraction for brain tumor classification based on short echo time ^1H MR spectra. *Magnetic Resonance in Medicine* 60(2): 288–298
- [70] Guyon I, Alamdari ARSA, Dror G, Buhmann JM (2006) Performance Prediction Challenge. In IJCNN '06. International Joint Conference on Neural Networks, pp. 1649–1656
- [71] Bishop CM (2006) *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer
- [72] Davies N, Wilson M, Harris L, Natarajan K, Lateef S, Macpherson L, Sgouros S, Grundy R, Arvanitis T, Peet A (2008) Identification and characterisation of childhood cerebellar tumours by in vivo proton MRS. *NMR in Biomedicine* 21(8): 908–918

Chapter 6

Automated brain tumour biopsy classification using single-labelling cDNA microarrays-based gene expression profiling

Gene-signatures obtained from microarray experiments may be of use to improve prediction of brain tumor diagnosis. Nevertheless, automated and objective prediction with accuracy comparable or better than the gold-standard should be convincingly demonstrated for possible clinician uptake of the new methodology. Herewith we demonstrate that primary brain tumor (PBT) types can be discriminated using microarray data in an automated and objective way.

Postsurgical biopsies from 35 patients (17 GBM and, 18 Meningothelial Meningioma (mMEN)) were stored in liquid nitrogen, total RNA extracted, cDNA labelled with Cy3 fluorochrome and hybridised onto a cDNA-based microarray containing 11500 cDNA clones representing 9300 loci. Scanned data was pre-processed, normalised, and used for predictor development. The predictive functions were fitted to a subset of samples and their performance evaluated with an independent subset. Expression results were validated by means of RT-PCR.

Some gene expression-based predictors achieved 100% of accuracy both in training re-sampling validation and independent testing. One of them composed of GFAP, PTPRZ1, GPM6B and PRELP produced 100% prediction accuracy for both training and independent test dataset. Furthermore, gene-signatures obtained: increased cell detoxification, motility and intracellular transport in GBM and increased cell adhesion and cytochrome-family genes in mMEN agree well with expected biological and pathological characteristics of the studied tumors.

The ability of gene-signatures to automate prediction of brain tumors through a fully objective approach has been demonstrated. Comparison of gene expression profiles between GBM and mMEN may provide additional clues about patterns associated to each tumor type.

This chapter has been published as journal paper in [1]. The study to predict GBM and MEN using single-labelling cDNA microarrays data was performed through collaboration between UAB and its associated clinical centres and UPVLC. The article derived from

this study and accepted for publication in Diagnostic Molecular Pathology, displays Xavier Castells and Juan-Miguel García-Gómez, both as first co-authors. To the effect of their respective PhD thesis, both co-authors agree in that Xavier Castells performed the RNA isolation of biopsies, hybridization of isolated RNA, scanning and analysis of microarrays, the RT-PCR experiments and the functional analysis of differentially detected genes. They also agree in that Juan Miguel García-Gómez performed the statistical analysis of data including background correction and normalisation, detection of differentially expressed genes and development of a prediction formula.

6.1 Introduction

Primary Brain Tumors (PBT) are among the most aggressive and intractable tumors [2]. They are the leading cause of death from cancer in children under the age of 15, and the second one in young people from 15 to 34 [3]. In 2002, the worldwide incidence of PBT was 189485 cases, and 141650 the estimated mortality [4]. Diagnoses of these tumors are devastating for the patient, and prognoses are yet difficult to determine [5–7]. Currently, neurological examination and imaging techniques are commonly used to approximate in vivo the diagnosis and the prognosis of the PBT, which is usually confirmed in vitro by histological examination of a biopsy [6]. In spite of the widespread use of histopathology, the gold standard for diagnosis, its results are not completely accurate [6]. For example, they are limited with respect to predicting the clinical course of the patient and response to therapy, probably because they cannot fully account for the underlying molecular mechanisms of tumor progression [8]. Furthermore, classification of tumors by histopathological examination produces a subjective output that can be controversial, since a tumor sample can be differently classified depending on the clinician [9, 10].

High throughput gene expression techniques provide a transcriptomic signature, which can be used to characterize different biological samples or pathological conditions. In the oncological context, discovering genomic signatures of tumors allows a better understanding on the molecular mechanisms of tumorigenesis [11–13]. Currently, several studies using DNA microarrays have described novel genes involved in brain tumor appearance and progression [14–17], improving classification of brain tumors [18–21] and revealing molecular markers that identify a given brain tumor type [22–26]. For such objective, machine learning tools (MLT) have been adopted as a relevant computer technology in microarray data analysis to characterise and discriminate among classical WHO types of human brain tumors or among new molecular subtypes. Microarray literature analysis reveals that unsupervised class discovery problems in PBT often benefit from MLT use [18, 19]. Similarly, supervised problems (where class labels are used for the training of the predictors) have also been addressed by MLT [24, 27].

Most work using DNA microarray for PBT typing has focused on distinction between different grades of a certain histological PBT type. Despite the complete prediction of distinct histopathological PBT using gene signatures from gene expression microarray has been rarely addressed, such a proof of principle is essential in order to compare the outcome of data derived from high throughput technologies with the gold standard, which still is histopathological examination. Subsequent to the proof of principle, the attempt of classifying subtypes of a determined histological PBT may be more convincing, although large dataset analysis are usually needed for molecular sub-type investigations [20, 26, 28,

29]. Furthermore, histopathological examination is always subjective, while microarray-based classification lends itself to fully automated operator-bias independent analysis.

Hence, our goal in this work has been to demonstrate the ability of automated methods based on gene expression profiling to discriminate GBM from mMEN. These tumors are histological and pathologically highly distinct PBT. GBM display glial cell lineage, they are the most aggressive and infiltrative PBT, and a poor prognosis is associated with its diagnosis [6]. In contrast, mMEN display meningotheial cell lineage, they usually show well defined edges and they remain at the benign stage [30]. Although GBM and mMEN are usually correctly diagnosed in vivo through imaging techniques by an expert radiologist and recognized with few exceptions ex-vivo by the pathologist, their in vivo differential diagnosis is not exempt of problems [31–34]. Therefore, the reported confusion about the in vivo differential diagnosis of some GBM and mMEN makes it relevant to screen for gene-signatures that could help to better diagnose PBT when detectable as characteristic patterns by in vivo MRS [35]. For this, proper recognition of the characteristic differential molecular signatures of the in vivo confusable GBM and mMEN may be relevant, although beyond our immediate goal for this work. All this taken into account, our experimental approach was expected to achieve a 100% of prediction success with an operator-independent approach using microarray-based gene expression profiling.

Single-labelling cDNA microarrays were used in this study. Having overcome the initial drawback of batch reproducibility [36], this technology has several advantages over both oligonucleotide and double-labelling cDNA microarray strategies. First, this approach is economically more accessible for researchers. Second, quantification of gene expression obtained can be ascribed to a tumor type without the comparison to a common RNA reference, which may not always be available for posterior experiments. Finally, manufacturing these microarrays is a more flexible process than what is available from other commercial products, allowing custom-redesign of the microarray to characterise only a subset of potentially more relevant genes.

This study shows a successful automated prediction of PBT tumors based on single-labelling cDNA microarrays. This work has been developed within the context of two EU funded projects (eTUMOUR and HealthAgents), which aim to improve diagnosis and prognosis in vivo of PBT tumors subtypes by combining transcriptomic (microarrays), metabolomic (high resolution magic angle spinning NMR) and in vivo magnetic resonance spectroscopy (MRS) data.

6.2 Materials and Methods

6.2.1 Collection, storage and histopathology analysis of samples

Collection of 35 biopsies (17 glioblastomas and 18 meningiomas) was carried out at the IDIBELL-Hospital Universitari de Bellvitge (L’Hospitalet de Llobregat). The study was approved by the local Ethics Committee and informed consent was obtained from all patients. Tumor samples were fixed in 4% buffered formalin and embedded in paraffin. For routine histological examination 4 μ m thick sections were stained with hematoxylin and eosin (HE). The WHO 2000 Nervous System Classification criteria were used for diagnosis [37]. All GBM cases were primary glioblastomas except for two of them which developed from a lower grade astrocytoma. None of them had extensive small cell compo-

ment or were of the giant cell, or gliosarcoma histological variants. All meningiomas were of the meningothelial subtype.

6.2.2 RNA Isolation

Total RNA from frozen biopsies stored in liquid nitrogen was isolated following the procedure indicated by the manufacturer using the Rneasy midi kit (Qiagen, Duesseldorf, Germany). RNA was characterised using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, USA). Absence of protein contamination was monitored by the 260 nm/280 nm ratio of absorbance, and samples with a ratio ranging between 1.6 and 2.0 were accepted for further processing. Integrity of the RNA was assessed by using the capillary electrophoretic system 2100 Bioanalyzer (Agilent, Santa Clara, USA). Only samples producing a 28S/18S ratio higher than 1.1 or an RNA integrity number (RIN) number higher than 5 were used for further analysis.

6.2.3 Single-Cy3 cDNA Microarray Labelling

cDNA labelling was performed using the Cy3-fluorescent dye and the CyScribe First Strand labelling kit (Amersham, Buckinghamshire, UK). The starting material was approximately $14\mu\text{g}$ of total RNA. Starting RNA was copied into cDNA using a reverse transcriptase and an oligo(dT) primer incorporating Cy3-dUTP into the growing cDNA sequence. Alkaline treatment was performed to eliminate the RNA template. Then, the cDNA labelled product was purified from the reaction mixture using the CyScribe™ GFX™ purification kit. Labelled cDNA was resuspended in $100\mu\text{l}$ of the hybridisation solution, composed of 50% deionised formamide, 5x sodium saline citrate (SSC) and 0.1% SDS. Two μl human COT1-DNA ($1\mu\text{g}/\mu\text{l}$), 2 μl polyadenilic acid ($6\mu\text{g}/\mu\text{l}$) and $0.4\mu\text{l}$ salmon sperm DNA ($10\mu\text{g}/\mu\text{l}$) were added to avoid unspecific hybridisations. The final solution was denatured for 2 minutes at 95°C and immediately placed in ice. The solution containing the labelled cDNA was hybridised onto a pre-hybridised human CNIO oncochip for an overnight period in an incubator ArrayBooster (Advalytix, Munich, Germany). The human CNIO oncochip is a 12K cDNA microarray produced at the Spanish National Cancer Research Centre (CNIO Genomics Unit, Madrid, Spain, ArrayExpress acc. no. A-MEXP-261) that contains 11500 cDNA clones representing 9300 loci [38]. After incubation, slides were washed and Cy3-dye fluorescence was measured using a ScanArray 4000 (Perkin Elmer, Waltham, USA) detection system. Signal was quantified by the Genepix 6.0 software (Molecular Devices, Sunnyvale, USA).

6.2.4 Data pre-processing

Prior to the development of the computational experiments to obtain the predictor on our dataset, a pre-processing step to make the expression values comparable among microarrays was performed. Due to the specific protocol used in this study (single-labelling cDNA-based microarrays), non-standard pre-processing methods derived from adaptations of the Affymetrix pre-processing methods [39, 40] were set up and applied to our data.

First, a pre-filtering step, based on a visual inspection of the scan images, discarded genes having an artefactual signal in at least one microarray experiment. Foreground values were corrected using the background smoothing procedure defined by Edwards in [41].

Genes with negative intensity signal (foreground minus background) in more than 20% of cases of both GBM and mMEN groups were also removed. Data was normalized using the average reference loess method [42]. Afterwards, a post-filtering step was performed and genes that were not validated by the microchip manufacturer by PCR evidence (single band) and sequence verification were removed. Finally, signals corresponding to genes spotted more than once in the microarray were averaged.

6.2.5 Feature selection and sample classification

Statistical significance was assessed by the non-parametric Mann-Whitney (MW) test on data from the training dataset (10 GBM and 11 mMEN). Afterwards, p-values of the MW test were corrected for the false discovery rate (FDR) control obtaining the so called q-values [43].

Genes with q-value lower than 0.02 were considered to be differentially expressed. Starting from the set of differentially expressed genes, selection of 3 or 4 genes through a SW procedure was performed.

Linear models based on Rank Reduced Linear Discriminant Analysis (LDA) were fitted to our data. Hence, given a set of samples a projection that maximized the separation between projected values of both classes was searched for. Prediction accuracy was evaluated by randomly sampling the training dataset 200 times [44]. That is, 15 samples, following the distribution frequency of the classes in the dataset, were selected to train the predictor and 6 samples to validate its result. Such a resampling procedure provided an estimation of the prediction accuracy. The final evaluation of the predictors' performance was carried out in a totally independent test dataset (7 GBM and 7 mMEN) with the labels blind to the testers (J.M. G-G., A.N. and M.R.). The ability to produce a single predictor for direct use at the bench was demonstrated by generating an LDA-based predictor with the four most selected genes across the 200 iterations. Such an LDA-based predictor was developed over the training dataset (10 GBM and 11 mMEN) and its performance tested over the independent dataset (7 GBM and 7 mMEN).

6.2.6 Functional analysis of gene signatures

Aiming to determine a gene signature that may characterize each tumor type based on the expression levels, a hierarchical cluster was performed with the 629 genes with the q-value lower than 0.02. Furthermore, the selected gene subset was submitted to the web-based Database for Annotation, Visualization and Integrated Discovery tool (DAVID) [45] with the purpose of detecting statistically significant functional gene groups. In our study, we chose the highest stringency level among the five stringency levels provided by DAVID for a set of genes to be called a functional group.

6.2.7 Real Time-PCR (RT-PCR) validation

The expression of 10 selected genes (see Section 6.3) was monitored by RT-PCR. Total RNA (100ng) was used as starting template RNA for reverse transcription. We used the validated primers Quantitect Primer Assays (Qiagen) and the one-step Quantitect SYBR Green RT-PCR kit (Qiagen), on a Smart Cycler (Cepheid) system. The GBM/mMEN ratio was calculated using the $2^{-\Delta Ct}$ method [46].

6.3 Results

6.3.1 Pre-processing and classification results

The pre-processing methodology described in the data pre-processing section was applied to both training and test datasets to attenuate the effect of possible variability due to nonbiological causes. In our experiments, the total number of analysed probes per microarray was 27648. After pre-filtering, 23652 features remained in the expression matrix. The next step consisted in discarding a gene from further processing when more than 20% of samples produced negative signals (foreground minus background) in each group, GBM and mMEN. After background correction and the normalization steps, post-filtering was performed, yielding 15584 features. Averaging of feature replicates gave rise to a final expression matrix of 7218 features for the 35 investigated samples (training and test dataset).

Starting from this number of genes, those differentially expressed in each tumor type were investigated. Considering that GBM and mMEN are both histological and pathologically highly distinct brain tumor types, simple graphical discrimination of these two tumor types by plotting values of differentially expressed genes was expected. For this purpose, the GBM/mMEN ratio for each gene was computed and genes displaying the highest and lowest GBM/mMEN ratios were used to create a graphical classifier (see Figure 6.1A and 6.1B). Interestingly, the gene with the highest GBM/mMEN ratio was the glial fibrillar acidic protein (GFAP) with a value higher than 400. The protein encoded by this gene is a well known biological marker of glial cells 40. Furthermore, desmoplakin (DSP) showed the lowest GBM/mMEN ratio, being 250-fold more expressed in mMEN compared to GBM (GBM/mMEN=0.004, see Table B.1 of the Appendix B).

A predictor based on LDA was developed (see Figure 6.1C). Our aim was two-fold, in the first place, to benefit from the panoply of genes available to build a better and potentially more robust predictor than the expression of a single gene product (GFAP or DSP). Secondly, we were interested in detection of gene signatures providing biological information about the underlying mechanisms, which may characterise GBM and mMEN in such a pairwise comparison. The predictor was built by splitting the full dataset in twenty-one samples (10 GBM and 11 mMEN) for training and validation purposes. Additionally, fourteen totally independent and blinded samples (7 GBM and 7 mMEN) for the testers were used for the final evaluation purposes. Statistical significance was computed by using the Mann-Whitney nonparametric test on the training set and genes with a corrected q-value less than 0.02 were selected. As a result 629 genes were found differentially expressed from the initial 7218 genes.

This set of genes was used to generate a prediction model with three or four randomly selected genes from a SW selection procedure. Two hundred repetitions were performed, from which two hundred LDA predictors were obtained. As a result, a 95% prediction accuracy mean was obtained. Concerning our blinded test dataset of fourteen samples, an observed accuracy of 100% with a [70%, 100%] confidence interval was obtained. A noteworthy fact is that the confidence intervals of the evaluation measures depend on the number of test samples; hence, wide intervals are obtained in the evaluation of our experiments.

In order to demonstrate the ability to produce a predictor that could be used in an automated way once developed, all training samples (21) were used to fit the final models.

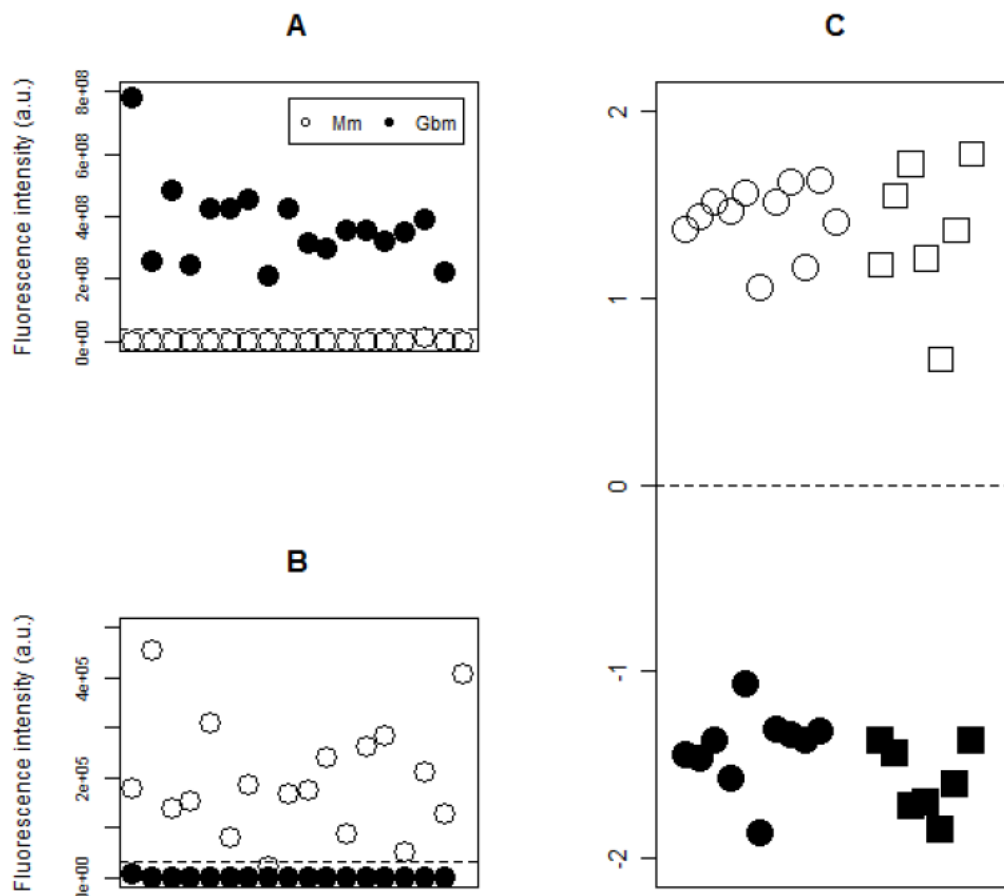


Figure 6.1: Graphical representation of normalised data and LDA-based predictor. Solid symbols correspond to GBM samples, while empty symbols denote mMEN samples. A) and B) show the scatter plots of normalised expression values of genes showing the highest and lowest GBM/mMEN ratio, GFAP and DSP respectively for all GBM and mMEN cases (see Table B.1 of the Appendix B for abbreviations meaning). Samples were arbitrarily distributed along the x axis, while along the y axis fluorescence intensity signals (a.u.) were plotted. C) shows the discriminant scores obtained from the LDA-based predictor generated using normalised expression values from GFAP, PTPRZ1, GPM6B and PRELP genes, those most selected across the 200 iterations, which perfectly separated these two tumors class members. Both empty and solid circles are training samples, while empty and solid squares are test samples. Along the y axis discriminant scores at the latent space are shown.

Therefore, the final predictor shown in Figure 6.1C was calculated selecting the four most selected genes across the 200 prediction iterations: GFAP, PTPRZ1, GPM6B and PRELP (see Table B.1 of the Appendix B). Such a predictor produces an objective and automated prediction result by simply introducing the pre-processed and normalised gene expression values into the LDA Formula 6.1.

$$y = -0.394GFAP - 0.397PTPRZ1 - 0.397GPM6B + 0.365PRELP. \quad (6.1)$$

The cut-off point at 0 enables prediction between the two tumor types, being GBM and mMEN for negative and positive values, respectively (see appendix). Using this predictor, a 100% prediction of the independent test set was obtained.

6.3.2 Molecular characterization of expression differences between GBM and mMEN

Aiming to detect a gene-signature that could differentiate GBMs and mMENs, the subset of genes with q-value lower than 0.02, was used to perform the hierarchical cluster shown in Figure 6.2. Genes were initially grouped in sixteen clusters clearly defining a specific profile for each tumor type. Furthermore, determination of functionally-related groups of genes was assessed by subjecting the mentioned gene subset to the DAVID tool. A total of eleven functional groups with p-value lower than 0.05 accounting for eighty genes were obtained. Interestingly, three out the eleven clusters were exclusively composed by genes either overexpressed in GBM or in mMEN (see Tables B.2, B.3 and B.4 of the Appendix B). The functional group 2 contained genes overexpressed in mMEN that belong to the family of the small leucine rich proteoglycans (SLRPs): FMOD (J), PRELP (G), OMD (G), BGN (J) and OGN (G). Similarly, the functional group 11 was composed of five members belonging to the cytochrome family: CYP1B1 (J), CYP4Z1 (B), CYB5 (J), CYP4B1 (B) and CYP3A5 (B). On the other hand, the functional group 6 was composed of several isoforms of genes encoding tubulins overexpressed in GBM: TUBA1 (I), TUBA2 (I), TUBA3 (F), TUBB (F), TUBB2 (F), TUBB4 (I), TUBA4A (I) and TUBA8 (I). Also, detected by the DAVID tool when setting the medium stringency level, a large number of metallothionein isoforms were highly expressed in GBM (see Figure 6.2): MT1H (K), MT1F (I), MT1X (K), MT2A (F and I) and MT3 (I).

Concerning the remaining nine functional groups composed of genes overexpressed in both GBM and mMEN, the cluster with highest and lowest statistical significance harboured a collection of cadherin and cytochrome isoforms, respectively (see Tables B.2, B.3 and B.4 of the Appendix B). Curiously, functional groups 7, 8, 9 and 10 were mainly composed of genes encoding proteins somewhat related to cell signalling: signal receptors (group 7), G-protein receptors (group 8), Ras proteins (group 9) and tyrosine kinases (group 10). Groups 3, 4 and 5 were apparently enriched with proteins related to the extra-cellular matrix and the cell-cell adhesion complexes.

6.3.3 RT-PCR expression results

A subset composed of 6 samples, 3 GBM and 3 mMEN, were subjected to RT-PCR for validation purposes. Selection of these samples was based on RNA integrity and

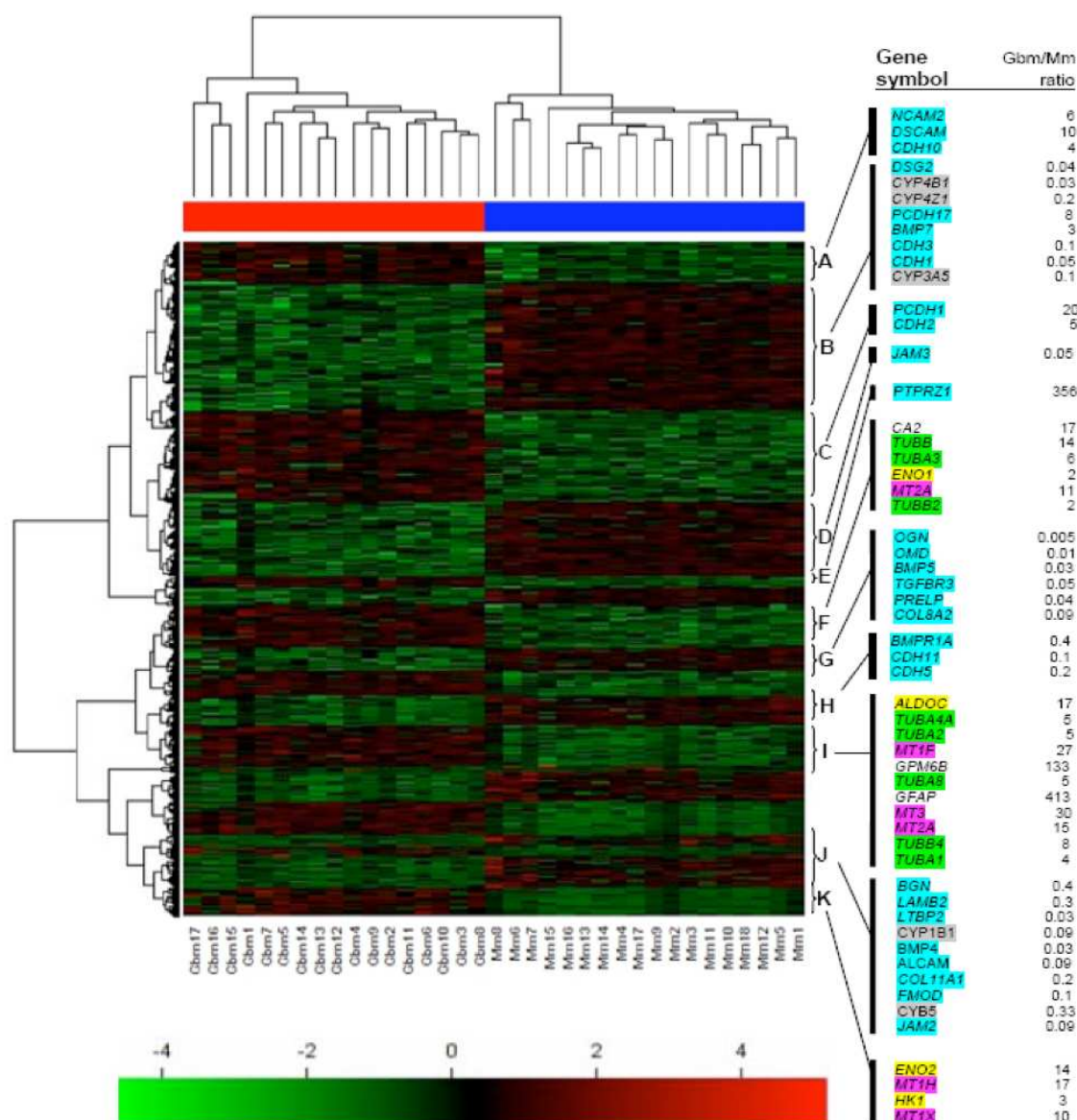


Figure 6.2: Hierarchical cluster of differentially expressed genes from the training set. Graphical illustration of the hierarchical cluster performed across samples and genes computed using Euclidean distance. Columns are samples and rows are genes. The 629 genes with a p-value lower than 0.02 were used to generate this cluster with the 35 samples of the full dataset. At the top of the figure, GBM and mMEN samples are denoted by red and blue bars respectively. At the right margin, letters (A-K) indicate clusters of genes belonging to functional families or related to different signalling pathways. Groups or gene-families are specifically coloured: metallothioneins (pink), tubulins (green), glycolysis-related genes (yellow), cell membrane and/or the extracellular matrix (blue) and cytochrome-related genes (grey). GFAP, PTPRZ1, GPM6B and PRELP are also included within the hierarchical cluster. See Tables B.2, B.3 and B.4 of the Appendix B for gene abbreviations meaning.

concentration, as well as absence of protein contamination. Among differentially expressed genes in our study, a subset to be subjected to RT-PCR was selected based on four criteria: maximal or minimal GBM/mMEN ratio and minimal FDR corrected p-value, maximal selection at the re-sampling of the training set and biological meaning. Using such criteria selected genes were: GFAP, PTPRZ1, GPM6B, MT3, CA2, TUBB, APM2, PRELP, OGN and DSP. The six first genes were far more expressed in GBM than in mMEN, and the four last genes showed opposite behaviour. As observed in the Table B.5 of the Appendix B, RT-PCR confirmed the expression profile of the mentioned genes according to the type of tumor evaluated.

6.4 Discussion

6.4.1 Development of an automated classifier based on gene signatures of brain tumors

Several studies have demonstrated usefulness of data generated from gene-expression based microarray to classify brain tumors when they cannot be properly discriminated by using histological and image-based morphologic examinations alone [19, 21, 24, 47]. These studies have focused on recognizing tumor molecular subtypes of GBM and classification of histologically distinct mMEN, but no automated predictor for classical histological WHO types was developed. This issue was considered highly relevant within the context of the European project eTUMOUR, as a required proof of principle prior to attempting automated and objective recognition of tumor types or grades difficult to ascertain by classical histology (i.e. glioblastoma multiforme molecular subtypes) [6–8]. In our study we have demonstrated that 100% successful automated prediction between glioblastoma multiforme and meningotheial meningioma tumors is achievable without subjective data judgement. Furthermore, such a prediction was performed using singlelabelling cDNA microarrays. Being a protocol of lower cost and greater design flexibility than other genomic technologies, it results in an attractive experimental option for routine use.

The particular type of microarray protocol (single-labelling cDNA microarrays) forced us to develop a novel algorithm for data pre-processing to correct background and normalise experiments. Discrimination of GBM and mMEN was first performed by plotting only normalised intensity values of genes with highest or lowest GBM/mMEN ratios (Figure 6.1A and 6.2B). Separation of GBM and mMEN samples is better defined when plotting GFAP expression values rather than when plotting those from DSP. This fact is not surprising since GFAP is specifically expressed in astrocytes and astrocytomas, and it is considered the molecular marker reference of this cellular type in the central nervous system [48, 49], whereas such well defined cell type specific marker does not exist for mMEN. Regarding GFAP, variable expression among samples seen in Figure 6.1A could be justified by a heterogeneous population of glioblastoma in our study concerning both tumor progression stage and cell stemness characteristics, since expression of GFAP seems to be modulated depending on these factors in GBM tumors[50, 51].

Nevertheless, a search for a potentially more robust prediction formula by using linear discriminant analysis (LDA) of the most differentially expressed genes was performed. Fully successful prediction of the blindly analysed independent test set was achieved. The obtained result provides evidence of a 100% of sensitivity and specificity by means of a

completely objective method, in which intervention of operator biases is strongly reduced. To our knowledge, complete discrimination of two PBT by developing an LDA predictor based on gene-signatures arisen from microarray data had not been reported until now.

6.4.2 Molecular signature characteristics of GBM and mMEN

Development of an automated-predictor based on microarray experiments was also useful to study the gene signatures underlying the biology of both GBM and mMEN. For this purpose, an unsupervised hierarchical cluster for each pre-processing procedure using as input those genes within a threshold q -value lower than 0.02 was generated (see Figure 6.2). This computation corroborated the previous LDA-based prediction because all samples were correctly clustered within its tumor group. In addition, a clear gene-signature was achieved for both GBM and mMEN tumor types, which was functionally characterized by subjecting these genes to the DAVID tool. From our results, tubulins seem to play a needed role in GBM, and SLRPs and cytochrome-related genes in mMEN, since a large group of these gene-families products were found clearly more expressed in only one of the two tumor types.

Functional group 2 is a paradigmatic finding of genes belonging to the SLRPs family. FMOD and PRELP genes are located at the 1q chromosome, OGN and OMD at the 9q chromosome and BGN at the Xq chromosome. However, a common promoter region has been described for all members [52]. Noticeably, the described involvement of the SLRPs family in collagen fibrillogenesis, cellular growth, differentiation and migration revealed the relevance of this family in extracellular matrix modelling [52]. Concretely, cleavage of OGN precursors by the bone morphogenetic-1 protein (BMP1) producing the mature OGN forms, was proposed as a mechanism by which formation of collagen fibrils is controlled [53]. Curiously, we found a set of BMP genes among those of q -value lower than 0.02, although it was not detected by the DAVID tool as a functional significant group. The gene coding for receptor BMP1R, BMP4 and BMP5 were found overexpressed in mMEN. Furthermore, OGN has been recently described as one of the main components of the human amniotic membrane that promotes the development of limbal stem cell niches [54]. Also, interaction of SLRPs members with TGF- β facilitates signal transduction inside the cell, resulting in an increase of SLRPs gene-expression [52, 55, 56]. In our study, we found overexpressed in mMENs two genes encoding SLRP proteins involved in the recruitment of TGF- β from the extracellular space to the membrane, FMOD and BGN [55], a receptor of TGF- β , TGFBR2, and a protein modulating the secretion and activation of TGF- β , LTBP2 [56]. From these results, an apparent modulation of the extracellular matrix through SLRPs may characterize tumorigenesis of mMENs. Incidentally, LTBP2 is downregulated when benign meningiomas progress into atypical or anaplastic stages [47].

Functional group 11 contains a set of genes encoding cytochrome proteins overexpressed in meningiomas. Interestingly, there is a consolidated bibliography describing the involvement of cytochrome P450 in cancer drug metabolism [57], and their crucial role in sterol and androgen synthesis, as well as in retinoic acid metabolism [57]. Nevertheless, the cytochrome P450 isoforms found overexpressed in our mMEN samples, had not apparently been associated in the biology of this tumor, although other isoforms implication had been linked to both meningioma and glioma progression [58–60].

With respect to functional group 6, tubulins are structural components of microtubules, which take part in cell motility and intracellular transport, which seem involved in malig-

nant progression of gliomas. Nitration of tubulins is more acute in grade IV than in grade I gliomas [61]. Specifically, Gene expression of TUBA3 is induced by PI3K in human glioblastoma cells under stimulation with KCl, a well known differentiation inducer [62].

The functional group with highest statistical significance harbours a set of cadherins, among which one half are overexpressed in mMENs and the other half overexpressed in GBMs. Such a result may suggest a cadherin sub-type link with the tumor grade and/or histological type. Strikingly, E-cadherin was found overexpressed in mMENs, which corroborates previous findings of E-cadherin detection by histochemistry in meningiomas [63, 64]. Furthermore, Ecadherin is normally not expressed in gliomas, since in gliomas the neuronal isoform, Ncadherin, appears to be more specifically found [63, 65, 66], in agreement with our results (see Tables B.2, B.3 and B.4 of the Appendix B). Therefore, we show here a tumor type specific expression of cadherins that in our study may be first exemplified by the differential expression of E- and N-cadherin, and secondly, by those isoforms specifically overexpressed in each tumor type. Likewise, the functional group 5 with a set of genes related to cell adhesion would also sustain the important and tumor specific role of the extracellular matrix in cancer. Being significantly detected by the DAVID tool when setting the medium stringency level, metallothioneins are involved in cell detoxification, growth and redox balance, among other cellular roles [67], and were previously found overexpressed in GBM compared to mMEN by immunohistochemistry [68]. Likewise, genes related to glycolytic metabolism were also overexpressed in GBM (see Figure 6.2), in agreement with references [69, 70], and also in agreement with the well known correlation between glycolytic phenotype and malignity [71, 72].

Considering single genes, the finding of the protein tyrosine phosphatase ζ -1 receptor-type (PTPRZ1) highly expressed in GBM (see Table B.1 of the Appendix B) is also coherent with studies showing that RNA interference targeting of this receptor has recently been shown to suppress glioblastoma growth in vitro and in vivo [73]. Moreover, PTPRZ1 has been proposed as a potential target for therapy of GBM based on its inhibition by specific antibodies [74]. Furthermore, the increased expression of carbonic anhydrase type II, a cytosolic isoform, which may enhance electrogenic sodium bicarbonate export from cells [75], may be correlated with the recently described increased presence of CA2 detected by immunohistochemical analysis in GBM as compared to low or intermediate astrocytic grade tumors [76]. Concerning mMENs, the high expression of DSP, a “plaque” component of desmosome associated proteins, is concordant with the histological characteristics of this tumor type.

6.5 Conclusion

This study establishes the proof of principle about the discrimination of primary brain tumors by means of gene signatures. That is, glioblastoma and meningioma biopsies can be predicted in an automated and objective way with 100% sensitivity and specificity by using a gene expression profile-predictor based on only four genes (GFAP, PTPRZ1, GPM6B and PRELP). Accordingly, we now feel confident with proceeding with the automation of microarray-based brain tumors biopsy recognition of more complex predictions, like glial tumor grading, GBM molecular subtypes, or correlation with survival. For this, a much larger dataset will be needed and it is presently being accrued within the eTUMOUR and HEALTHAGENTS EU projects. Future work will also contemplate the validation

of the discriminant genes detected in this work between the benign and malignant gene signatures by comparing our results with those from additional datasets obtained using different analysis strategies and platforms (Affymetrix, Agilent and others).

Bibliography

- [1] Castells X, García-Gómez JM, Navarro A, Acebes JJ, Godino Ó, Boluda S, Barceló A, Robles M, Ariño J, Arús C (2008) Automated brain tumor biopsy prediction using single-labelling cDNA microarrays-based gene expression profiling. *Diagnostic Molecular Pathology* p. Accepted for publication
- [2] Oliver TG, Wechsler-Reya RJ (2004) Getting at the root and stem of brain tumors. *Neuron* 42(6): 885–888
- [3] Baldwin RT, Preston-Martin S (2004) Epidemiology of brain tumors in childhood—a review. *Toxicol Appl Pharmacol* 199(2): 118–131
- [4] J Ferlay et al (2004) *GLOBALCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide*. IARC CancerBase ed Lyon
- [5] Armstrong TS, Cohen MZ, Eriksen LR, Hickey JV (2004) Symptom clusters in oncology patients and implications for symptom research in people with primary brain tumors. *J Nurs Scholarsh* 36(3): 197–206
- [6] Behin A, Hoang-Xuan K, Carpentier AF, Delattre JY (2003) Primary brain tumours in adults. *Lancet* 361(9354): 323–331
- [7] Noble M, Dietrich J (2004) The complex identity of brain tumors: emerging concerns regarding origin, diversity and plasticity. *Trends Neurosci* 27(3): 148–154
- [8] Mischel PS, Cloughesy TF, Nelson SF (2004) DNA-microarray analysis of brain cancer: molecular classification for therapy. *Nat Rev Neurosci* 5(10): 782–792
- [9] Giannini C, Scheithauer BW, Weaver AL, Burger PC, Kros JM, Mork S, Graeber MB, Bauserman S, Buckner JC, Burton J, Riepe R, Tazelaar HD, Nascimento AG, Crotty T, Keeney GL, Pernicone P, Altermatt H (2001) Oligodendrogliomas: reproducibility and prognostic value of histologic diagnosis and grading. *J Neuropathol Exp Neurol* 60(3): 248–262
- [10] Rorke LB (1997) Pathologic diagnosis as the gold standard. *Cancer* 79(4): 665–667
- [11] Lander ES (1999) Array of hope. *Nat Genet* 21(1 Suppl): 3–4
- [12] Ntzani EE, Ioannidis JPA (2003) Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 362(9394): 1439–1444
- [13] Segal E, Friedman N, Kaminski N, Regev A, Koller D (2005) From signatures to models: understanding cancer using microarrays. *Nat Genet* 37 Suppl: 38–45

- [14] Furumoto H, Ying H, Chandramouli GVR, Zhao L, Walker RL, Meltzer PS, Willingham MC, Cheng SY (2005) An unliganded thyroid hormone beta receptor activates the cyclin D1/cyclin-dependent kinase/retinoblastoma/E2F pathway and induces pituitary tumorigenesis. *Mol Cell Biol* 25(1): 124–135
- [15] Hui AM, Zhang W, Chen W, Xi D, Purow B, Friedman GC, Fine HA (2004) Agents with selective estrogen receptor (ER) modulator activity induce apoptosis in vitro and in vivo in ER-negative glioma cells. *Cancer Res* 64(24): 9115–9123
- [16] Ma D, Nutt CL, Shanehsaz P, Peng X, Louis DN, Kaetzel DM (2005) Autocrine platelet-derived growth factor-dependent gene expression in glioblastoma cells is mediated largely by activation of the transcription factor sterol regulatory element binding protein and is associated with altered genotype and patient survival in human brain tumors. *Cancer Res* 65(13): 5523–5534
- [17] Tatenhorst L, Puttmann S, Senner V, Paulus W (2005) Genes associated with fast glioma cell migration in vitro and in vivo. *Brain Pathol* 15(1): 46–54. Comparative Study
- [18] Markert JM, Fuller CM, Gillespie GY, Bubien JK, McLean LA, Hong RL, Lee K, Gullans SR, Mapstone TB, Benos DJ (2001) Differential gene expression profiling in human brain tumors. *Physiol Genomics* 5(1): 21–33
- [19] Mischel PS, Shai R, Shi T, Horvath S, Lu KV, Choe G, Seligson D, Kremen TJ, Palotie A, Liau LM, Cloughesy TF, Nelson SF (2003) Identification of molecular subtypes of glioblastoma by gene expression profiling. *Oncogene* 22(15): 2361–2373. Comparative Study
- [20] Nutt CL, Mani DR, Betensky RA, Tamayo P, Cairncross JG, Ladd C, Pohl U, Hartmann C, McLaughlin ME, Batchelor TT, Black PM, von Deimling A, Pomeroy SL, Golub TR, Louis DN (2003) Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Res* 63(7): 1602–1607
- [21] Watson MA, Gutmann DH, Peterson K, Chicoine MR, Kleinschmidt-DeMasters BK, Brown HG, Perry A (2002) Molecular characterization of human meningiomas by gene expression profiling using high-density oligonucleotide microarrays. *Am J Pathol* 161(2): 665–672
- [22] Neben K, Korshunov A, Benner A, Wrobel G, Hahn M, Kokocinski F, Golanov A, Joos S, Lichter P (2004) Microarray-based screening for molecular markers in medulloblastoma revealed STK15 as independent predictor for survival. *Cancer Res* 64(9): 3103–3111
- [23] Park PC, Taylor MD, Mainprize TG, Becker LE, Ho M, Dura WT, Squire J, Rutka JT (2003) Transcriptional profiling of medulloblastoma in children. *J Neurosurg* 99(3): 534–541
- [24] Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, Williams PM, Modrusan Z, Feuerstein BG, Aldape

- K (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9(3): 157–173. Comparative Study
- [25] Tso CL, Freije WA, Day A, Chen Z, Merriman B, Perlina A, Lee Y, Dia EQ, Yoshimoto K, Mischel PS, Liau LM, Cloughesy TF, Nelson SF (2006) Distinct transcription profiles of primary and secondary glioblastoma subgroups. *Cancer Res* 66(1): 159–167
- [26] Lamszus K (2004) Meningioma pathology, genetics, and biology. *J Neuropathol Exp Neurol* 63(4): 275–286
- [27] Kim S, Dougherty ER, Shmulevich I, Hess KR, Hamilton SR, Trent JM, Fuller GN, Zhang W (2002) Identification of combination gene sets for glioma classification. *Mol Cancer Ther* 1(13): 1229–1236
- [28] Liang Y, Diehn M, Watson N, Bollen AW, Aldape KD, Nicholas MK, Lamborn KR, Berger MS, Botstein D, Brown PO, Israel MA (2005) Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci U S A* 102(16): 5814–5819
- [29] Phillips HS, Kharbanda S, Chen R, Forrester WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, Williams PM, Modrusan Z, Feuerstein BG, Aldape K (2006) Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* 9(3): 157–173
- [30] Lamszus K (2004) Meningioma pathology, genetics, and biology. *J Neuropathol Exp Neurol* 63(4): 275–286
- [31] Nestler U, Schmidinger A, Schulz C, Huegens-Penzel M, Gamerdinger UA, Koehler A, Kuchelmeister KW (2007) Glioblastoma simultaneously present with meningioma—report of three cases. *Zentralbl Neurochir* 68(3): 145–150
- [32] Julia-Sape M, Acosta D, Majos C, Moreno-Torres A, Wesseling P, Acebes JJ, Griffiths JR, Arús C (2006) Comparison between neuroimaging classifications and histopathological diagnoses using an international multicenter brain tumor magnetic resonance imaging database. *J Neurosurg* 105(1): 6–14
- [33] Miguel PS, Fernandez G, Vasallo FJ, Hortas M, Lorenzo JR, Rodriguez I, Ortiz-Rey JA, Anton I (2006) Neurobrucellosis mimicking cerebral tumor: case report and literature review. *Clin Neurol Neurosurg* 108(4): 404–406
- [34] Taillibert S, Pedretti M, Sanson M (2004) [Current classification of gliomas]. *Presse Med* 33(18): 1274–1277
- [35] Celda B, Monleon D, Martinez-Bisbal MC, Esteve V, Martinez-Granados B, Pinero E, Ferrer R, Piquer J, Marti-Bonmati L, Cervera J (2006) MRS as endogenous molecular imaging for brain and prostate tumors: FP6 project "eTUMOR". *Adv Exp Med Biol* 587: 285–302

- [36] Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu TM, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, Walker SJ, Zhang L, Hurban P, de Longueville F, Fuscoe JC, Tong W, Shi L, Wolfinger RD (2006) Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol* 24(9): 1140–1150. Comparative Study
- [37] Kleihues P, Cavenee WK (2000) World Health Organization Classification of Tumours: Pathology and Genetics: Tumours of the Nervous System. Oxford University Press
- [38] Tracey L, Villuendas R, Ortiz P, Dopazo A, Spiteri I, Lombardia L, Rodriguez-Peralto JL, Fernandez-Herrera J, Hernandez A, Fraga J, Dominguez O, Herrero J, Alonso MA, Dopazo J, Piris MA (2002) Identification of genes involved in resistance to interferon-alpha in cutaneous T-cell lymphoma. *Am J Pathol* 161(5): 1825–1837
- [39] Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20(3): 307–315
- [40] Schinke-Braun M, Couget JA (2007) Expression profiling using affymetrix genechip probe arrays. *Methods Mol Biol* 366: 13–40
- [41] Edwards D (2003) Non-linear normalization and background correction in one-channel cDNA microarray studies. *Bioinformatics* 19(7): 825–833
- [42] Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2): 185–193
- [43] Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ Method. *Methods* 25(4): 402–408
- [44] Markowetz F, Spang R (2005) Molecular diagnosis. Classification, model selection and performance evaluation. *Methods Inf Med* 44(3): 438–443
- [45] Eng LF, Ghirnikar RS, Lee YL (2000) Glial fibrillary acidic protein: GFAP-thirty-one years (1969-2000). *Neurochem Res* 25(9-10): 1439–1451
- [46] Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ Method. *Methods* 25(4): 402–408
- [47] Wrobel G, Roerig P, Kokocinski F, Neben K, Hahn M, Reifenberger G, Lichter P (2005) Microarray-based gene expression profiling of benign, atypical and anaplastic meningiomas identifies novel genes associated with meningioma progression. *Int J Cancer* 114(2): 249–256
- [48] Eng LF, Ghirnikar RS, Lee YL (2000) Glial fibrillary acidic protein: GFAP-thirty-one years (1969-2000). *Neurochem Res* 25(9-10): 1439–1451

-
- [49] Baba H, Nakahira K, Morita N, Tanaka F, Akita H, Ikenaka K (1997) GFAP gene expression during development of astrocyte. *Dev Neurosci* 19(1): 49–57
- [50] Lee J, Kotliarova S, Kotliarov Y, Li A, Su Q, Donin NM, Pastorino S, Purow BW, Christopher N, Zhang W, Park JK, Fine HA (2006) Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer Cell* 9(5): 391–403
- [51] Zhou R, Skalli O (2000) TGF- α differentially regulates GFAP, vimentin, and nestin gene expression in U-373 MG glioblastoma cells: correlation with cell shape and motility. *Exp Cell Res* 254(2): 269–278
- [52] Tasheva ES, Klocke B, Conrad GW (2004) Analysis of transcriptional regulation of the small leucine rich proteoglycans. *Mol Vis* 10: 758–772
- [53] Ge G, Seo NS, Liang X, Hopkins DR, Hook M, Greenspan DS (2004) Bone morphogenetic protein-1/tolloid-related metalloproteinases process osteoglycin and enhance its ability to regulate collagen fibrillogenesis. *J Biol Chem* 279(40): 41626–41633
- [54] Baharvand H, Heidari M, Ebrahimi M, Valadbeigi T, Salekdeh GH (2007) Proteomic analysis of epithelium-denuded human amniotic membrane as a limbal stem cell niche. *Mol Vis* 13: 1711–1721
- [55] Droguett R, Cabello-Verrugio C, Riquelme C, Brandan E (2006) Extracellular proteoglycans modify TGF- β bio-availability attenuating its signaling during skeletal muscle differentiation. *Matrix Biol* 25(6): 332–341
- [56] Oklu R, Hesketh R (2000) The latent transforming growth factor beta binding protein (LTBP) family. *Biochem J* 352 Pt 3: 601–610
- [57] Purnapatre K, Khattar SK, Saini KS (2008) Cytochrome P450s in the development of target-based anticancer drugs. *Cancer Lett* 259(1): 1–15
- [58] Elexpuru-Camiruaga J, Buxton N, Kandula V, Dias PS, Campbell D, McIntosh J, Broome J, Jones P, Inskip A, Aldersea J (1995) Susceptibility to astrocytoma and meningioma: influence of allelism at glutathione S-transferase (GSTT1 and GSTM1) and cytochrome P-450 (CYP2D6) loci. *Cancer Res* 55(19): 4237–4239
- [59] Khalid MH, Tokunaga Y, Caputy AJ, Walters E (2005) Inhibition of tumor growth and prolonged survival of rats with intracranial gliomas following administration of clotrimazole. *J Neurosurg* 103(1): 79–86
- [60] Wundrack I, Meese E, Mullenbach R, Blin N (1994) Debrisoquine hydroxylase gene polymorphism in meningioma. *Acta Neuropathol* 88(5): 472–474
- [61] Fiore G, Di Cristo C, Monti G, Amoresano A, Columbano L, Pucci P, Cioffi FA, Di Cosmo A, Palumbo A, d’Ischia M (2006) Tubulin nitration in human gliomas. *Neurosci Lett* 394(1): 57–62

- [62] Kim SJ, Chung TW, Jin UH, Suh SJ, Lee YC, Kim CH (2006) Molecular mechanisms involved in transcriptional activation of the human Sia- α 2,3-Gal- β 1,4-GlcNAc-R: α 2,8-sialyltransferase (hST8Sia III) gene induced by KCl in human glioblastoma cells. *Biochem Biophys Res Commun* 344(4): 1057–1064
- [63] Shimada S, Ishizawa K, Hirose T (2005) Expression of E-cadherin and catenins in meningioma: ubiquitous expression and its irrelevance to malignancy. *Pathol Int* 55(1): 1–7
- [64] Panagopoulos AT, Lancellotti CLP, Veiga JCE, de Aguiar PHP, Colquhoun A (2008) Expression of cell adhesion proteins and proteins related to angiogenesis and fatty acid metabolism in benign, atypical, and anaplastic meningiomas. *J Neurooncol* 89(1): 73–87
- [65] Barami K, Lewis-Tuffin L, Anastasiadis PZ (2006) The role of cadherins and catenins in gliomagenesis. *Neurosurg Focus* 21(4): E13
- [66] Utsuki S, Sato Y, Oka H, Tsuchiya B, Suzuki S, Fujii K (2002) Relationship between the expression of E-, N-cadherins and beta-catenin and tumor grade in astrocytomas. *J Neurooncol* 57(3): 187–192
- [67] Miles AT, Hawksworth GM, Beattie JH, Rodilla V (2000) Induction, regulation, degradation, and biological significance of mammalian metallothioneins. *Crit Rev Biochem Mol Biol* 35(1): 35–70
- [68] Maier H, Jones C, Jasani B, Ofner D, Zelger B, Schmid KW, Budka H (1997) Metallothionein overexpression in human brain tumours. *Acta Neuropathol* 94(6): 599–604
- [69] Altenberg B, Greulich KO (2004) Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* 84(6): 1014–1020
- [70] Kumanishi T, Watabe K, Washiyama K (1985) An immunohistochemical study of aldolase C in normal and neoplastic nervous tissues. *Acta Neuropathol* 67(3-4): 309–314
- [71] Gatenby RA, Gillies RJ (2007) Glycolysis in cancer: a potential target for therapy. *Int J Biochem Cell Biol* 39(7-8): 1358–1366
- [72] Gatenby RA, Gillies RJ (2004) Why do cancers have high aerobic glycolysis? *Nat Rev Cancer* 4(11): 891–899
- [73] Ulbricht U, Eckerich C, Fillbrandt R, Westphal M, Lamszus K (2006) RNA interference targeting protein tyrosine phosphatase zeta/receptor-type protein tyrosine phosphatase beta suppresses glioblastoma growth in vitro and in vivo. *J Neurochem* 98(5): 1497–1506
- [74] Muller S, Lamszus K, Nikolich K, Westphal M (2004) Receptor protein tyrosine phosphatase zeta as a therapeutic target for glioblastoma therapy. *Expert Opin Ther Targets* 8(3): 211–220

- [75] Pushkin A, Abuladze N, Gross E, Newman D, Tatishchev S, Lee I, Fedotoff O, Bondar G, Azimov R, Ngyuen M, Kurtz I (2004) Molecular mechanism of kNBC1-carbonic anhydrase II interaction in proximal tubule cells. *J Physiol* 559(Pt 1): 55–65
- [76] Haapasalo JA, Nordfors KM, Hilvo M, Rantala IJ, Soini Y, Parkkila AK, Pastorekova S, Pastorek J, Parkkila SM, Haapasalo HK (2006) Expression of carbonic anhydrase IX in astrocytic tumors predicts poor prognosis. *Clin Cancer Res* 12(2): 473–477

Chapter 7

Clinical Decision Support Systems for Soft Tissue Tumour Diagnosis and Brain Tumour Diagnosis

The final purpose of the experiments reported in previous chapters is to improve the clinical management of patients with soft tissue tumours or brain tumours. The design of Decision-Support Systems (DSSs) is the practical solution for integrating PR developments in clinical routine workflows.

After a review of previous CDSSs in the literature, the first system for cancer diagnosis described in this chapter is the distributed Decision-Support System (dDSS) that is aimed at helping radiologists in STT. It is based on a generic distributed architecture for decision support systems. This architecture has been the basis of posterior developments, such as those carried out in the dDSS of HEALTHAGENTS. Below, we present developments made in eTUMOUR and in HEALTHAGENTS - and which are based on a different conceptual approach for obtaining CDSSs for patients with brain tumours. We introduce the interface and classification modules designed to integrate the classifiers into the CDSS developed for the eTUMOUR project. We also describe the agent-based prototype developed in the HEALTHAGENTS context.

Sections of this chapter have appeared as scientific contributions in [1–4].

7.1 Introduction

7.1.1 Pattern recognition and decision support systems in clinical environments and cancer research

The first CDSSs used in clinical practice were designed in the 1970s. Leaper et al. [5, 6] developed a CDSS for the diagnosis and surgery of acute abdominal pain based on a naive Bayesian approach. In the early 80s, Shortliffe et al [7, 8] developed a CDSS for assisting physicians with the treatment of cancer patients receiving chemotherapy.

Early detection by screening [9] is common in the developed world, and breast cancer screening is one of the research areas in which many PR-based CDSS are found [10–19].

DSSs in gastrointestinal cancer help in diagnosis by using video endoscopy videos or

images [20, 21]. CDSSs were also reported for gastric [22, 23], esophageal [23] and oral cancer [24, 25]. In colon and colorectal cancer, genetic and gene-expression based systems were developed for prognosis [26–28].

Leukemia and lymphoma have been the main hematologic cancers studied by means of pattern recognition and decision support systems. The goals for these studies are focused on patient management [29, 30] and practitioner training [30]; or biomedical data as morphological recognition of cellular images [31, 32]; flow cytometry analysis [33]; or gene-expression [27, 34, 35].

Skin cancers are usually studied by means of image processing techniques in combination with classification algorithms. Melanoma has been studied using techniques based on textures [36], and other feature extraction methods [37] using ultrasound images [38, 39]; digital epiluminescence microscopy [38, 40]; or multispectral imaging [41]. Classification systems for melanoma gene-expression profiling [34] and patient management [42] were also studied.

The main problems in lung cancer are related to the treatment decision [43], and nodules in radiograms and computed tomography images [44–46]. Epidemiological studies related to smoking habits [47]; the evaluation of clinical management of cases; and the evaluation of care services [48–50] were also modelled using PR.

Diagnosis, grading, prognosis, and therapy of bladder cancer using PR methods have been studied using clinical, pathological, and histological factors in [51–55].

DSSs for prostate cancer are focused on the detection and grading of tumour tissue. Pattern recognition on prostate patients has been applied to in-vivo MRS [56, 57]; ultrasound images [57]; and cytometry images [58].

CDSSs are reported as potential diagnosis tools for ovarian cancer using mass spectroscopy [59]; for improving patient satisfaction [60]; gene-expression based classifiers [34, 61]; and ultrasonography-based classification [62]. Recently, Van Holsbeke et al. have reported a prospective validation of models to predict malignancy in adnexal masses using the International Ovarian Tumor Analysis (IOTA) protocols [63].

A screening-based study of over 100.000 women using PAPNET commercial software and based on ANN was carried out for cervical cancer [64].

As a general overview, the design of a translational CDSS should be accomplished with some caveats [65–67]: (1) its accuracy in solving the specific problem should be demonstrated. An important subsequent step is: (2) the adequate evaluation of the capability to generalize the performance on new cases in the problem domain [65, 67, 68]. Clearly, (3) the more transparency shown in relation to clinical knowledge, the more easily the CDSS will be accepted by the clinical community. Finally, (4) easy access to the system; and (5) integration with clinical routine workflow are key points in the use of the system in a clinical environment [66, 69].

7.1.2 Decision Support Systems for Soft Tissue Tumour

Tumours constitute the largest group within soft-tissue masses. Pre-surgical diagnosis of histological groups, as well as benign and malignant tumours is crucial in planning surgical procedures. Benign tumour constitutes by far the most common STT. These tumours can be characterized with MRI because of their main textural and growth pattern. Malignant tumours in this group represent approximately 33% [70]. Although in clinical practice,

most superficial tumours are diagnosed solely on the basis of clinical exploration and no further studies are required. Most deep lesions are studied with MRI.

However, STT are rare in the general population. Most radiologists and clinicians are unfamiliar with their appearance and differential diagnosis. Furthermore, their appearances are sometimes misleading. Malignant well-differentiated tumours are on the less aggressive side of the spectrum and usually appear similar to benign lesions, making it often difficult to distinguish between them. Some benign tumours can deviate from a typical homogeneous appearance, and so complicating the task of differentiation. The incorporation of DSSs in the diagnosis of STT may yield two benefits: assistance for the radiologist when deciding in a new case; and help in the training of new radiologists without expertise in STT.

7.1.3 Decision support systems in brain tumour research

The clinical and PR-based classification of brain tumours using MRS data has already been widely researched since the 1990s. [71]. Initial studies were based on LDA or ANN and focused on the tumour vs. normal tissue, or other central nervous system diseases. They were applied to relative metabolites, or PCA transformations, and evaluated using the LOO method [71].

The INTERPRET project [72] (1999-2002) was the first large-scale effort in the systematization of MRS multi-centric acquisitions - with the aim of developing a complete CDSS [72, 73] for brain tumours. The project defined a protocol to guarantee the compatibility of MR signals acquired at different hospitals. Furthermore, workflows including the compilation mechanism, quality control protocols, automatic pre-processing methods, high-level analysis, and model evaluation were also defined to manage the large volume of heterogeneous data. As a result, a CDSS for brain tumour classification based on SV MRS spectra was reported [74], in addition to a large improvement in PR methods for MRS analysis and brain tumour classification [75–82].

Two European projects, eTUMOUR [83] and HEALTHAGENTS [3], began research on the scientific and the technological aspects of CDSS brain tumour management. Both projects start from the output of INTERPRET eTUMOUR (2004-2009) resulted in a *web accessible MR decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data* which extended the study of brain tumours to ex-vivo data by means of HR-MAS MRS and gene-expression microarrays. HEALTHAGENTS (2006-2008) *produced an agent-based distributed decision support system for brain tumour diagnosis and prognosis* and which focused on developing a multi-agent DSS linked to a distributed data warehouse.

7.2 Distributed Decision Support System for Soft Tissue Tumour Diagnosis

The Sistema de Orientación Clínica (SOC) prototype for helping in the diagnosis of STT was designed. To create such a diagnostic support tool for STT, a large amount of data is required, usually more than is available in an isolated center. The solution to the lack of data is merging data from several hospitals. For this reason, a location-independent

system with the capability of sharing information and resources between professionals or institutions can solve the main problem in STT. The main objectives of the dDSS for STT are to make powerful pattern recognition and statistical tools available to experts and young radiologists, and bring together multidisciplinary experts in pattern recognition and radiology fields.

The DSSs are based on pattern recognition engines and classify benign/malignant characteristics and histological groups with good estimated efficiency. This dDSS builds upon the conclusions obtained in Chapter 2 and published in [1, 84–87] for the 2002-2004 period.

The patients treated in this development were confirmed musculoskeletal STT examined with MRI. Patient registers used to train the pattern recognition classifiers were obtained at five different European hospital centers.

Imaging studies were performed on 0.5 T and 1.5 T Gyroscan NT (Philips Medical System) and 0.5 T Signal (General Electric Medical System) units. T1-weighted (T1W, TR/TE, 450-650 ms/15-30 ms) and T2-weighted fat-suppression (T2W-FS, TR/TE, 2800-3500 ms/ 80-120 ms) or STIR (TR/TE/TI, 1400-1600 ms/ 40 ms / 100-120 ms) images were systematically obtained.

The DSS uses the following MRI findings obtained from the radiological examination [84]: age, clinical presentation, localization, size, shape, signal intensity, margins, homogeneity, edema, T1-hyperintense tracts, multiplicity, target appearance, muscular atrophy, intratumoural hemorrhage, calcification, dependence, intratumoural fat, fibrosis, fascial relationship, bone alterations, and vessels.

Discrimination between benign and malignant tumours, as well as classification between different histological groups were carried out to provide the DSS with efficient computer engines to help the radiologist. Experiments used PR techniques such as ANN, SVM, Decision Trees (DT), Multinomial parametric classifier (MN), and KNN [88]. Efficiency and other reliability parameters were measured from the classifiers using unseen patient registers.

Distributed decision support architecture

The introduced system is based on a distributed architecture with specialized nodes: radiologist visual interface, information system, and decision support web-services. The visual interface is the radiologist and clinician point of access to local and remote STT registers, statistical analysis tools, and distributed pattern recognition engines. A location-independent and multi-platform system has been developed to bring together hospitals and institutions to research useful tools in clinical and laboratory environments. Node maintenance and upgrades are automatically controlled by the architecture.

Three independent nodes comprise the STT distributed decision system. The nodes are specialized in three main groups and there is also web server infrastructure (see Figure 7.1):

- STT radiologist visual interface

Provides the access point to the system for radiologists and clinicians. These experts can obtain decision support and statistical information regarding the soft tissue tumours registers.

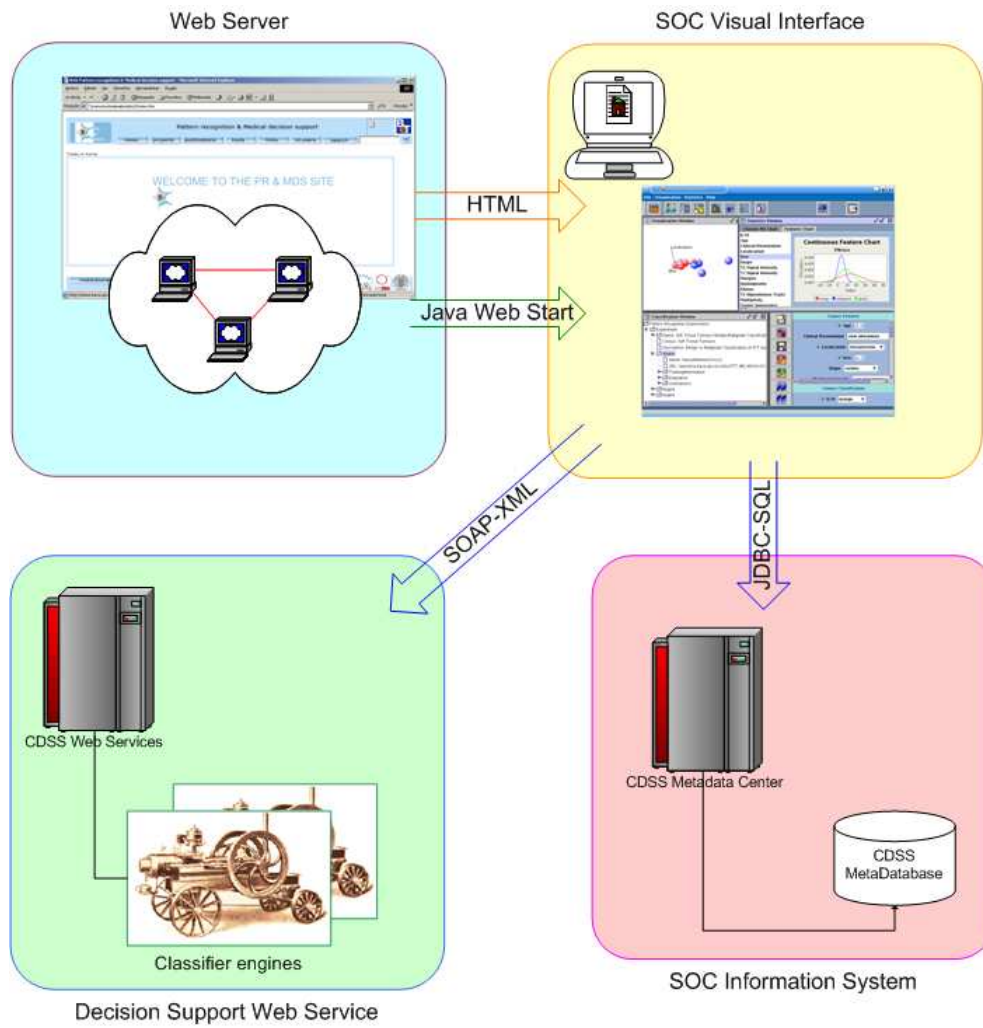


Figure 7.1: STT-dDSS architecture.

Table 7.1: Accuracy (expressed in %) obtained by multinomial (MN), SVM, decision trees (DT), and nearest neighbours (KNN) to discriminate between vascular (V), cyst (C), fat (F), nervous (N), sinovial (S) and fibrous (Fi).

Method	V-C-F (%)	V-C-F-N (%)	V-C-F-N-S (%)	V-C-F-N-S-Fi (%)
MN	98	92	87	83
SVM	97	94		
DT	94	87	82	76
KNN	95	89	84	78

- STT information system

The information system contains patient registers with the variables and diagnosis already accomplished. Also incorporates the meta-data of the variables that enables the generic use of the system, and the specific statistical analysis.

- STT decision support web-services

Web services provide classifier engines that have been developed with pattern recognition technology. They can be distributed through the Internet and incorporate heterogeneous technologies for prediction.

The STT radiologist visual interface can connect to local or distributed registers to analyze registered patients or incorporate new data. The visual interfaces can show statistical information processed locally, or invoke STT decision support web-services to make decision support for diagnosis. Connections between nodes are secure and nodes continuously identify themselves. The maintenance and upgrade process of the STT radiologist visual interface nodes is automatically controlled by the STT-dDSS architecture web-based system.

Soft tissue tumour classifiers

PR-based engines for classifying benign and malignant tumours were tested with a set of 128 unseen samples. The best results were obtained by ANN, KNN, and SVM, and are shown in Chapter 2 [84]. Additional engines were developed to classify histological groups. The histological classification was performed for the six most prevalent groups: Vascular (V), Cyst (C), Fat (F), Nervous (N), Sinovial (S) and Fibrous (Fi). Multiclass classifiers of 3, 4, 5, and 6 classes were developed based on Multinomial distributions (MN), SVM, C4.5 Decision Trees (DT), and KNN [88]. The classifiers were based on the same image findings and cases defined in Chapter 2. Table 7.1 [89] shows the accuracies obtained for the four multiclass classifiers using the four techniques.

STT decision support web-services

Engines are invoked by the STT radiologist visual interface using a web-service method to assist in the diagnostic process. All the PR engines are specified by an eXtensible Markup Language (XML) file with details of the training process, corpus, evaluation methods, results, and audit statistics. The visualization detail of the XML information is shown in Figure 7.2.

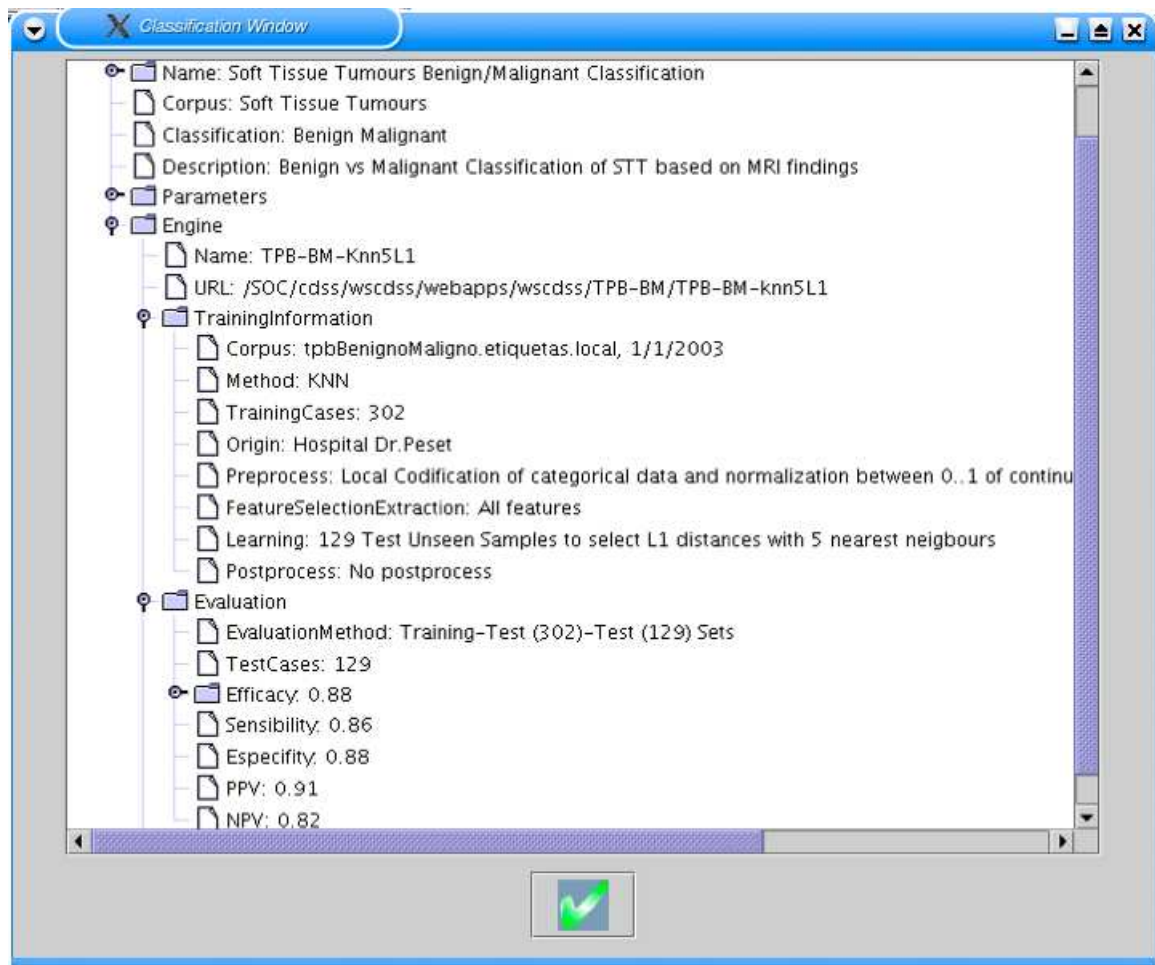


Figure 7.2: Information regarding the KNN-based engine for benign/malignant classification shown in the STT-dDSS.

Visual interface functionalities

Radiologist access to the STT-dDSS is enabled using the STT radiologist visual interface. This Graphical User Interface (GUI) contains four windows that offer access to the main system functionalities:

- Access to STT registers
Access to a local or distributed STT information system that contains the features under study (see Figure 7.3).
- Statistical analysis
Provides a graphical representation and reports on basic statistical and evaluation information such as probability distributions per class, correlation studies, and ROC curves (see Figure 7.4).
- Graphical representation of STT registers
It provides graphical representation of the Universe of cases projecting in a three-dimensional space the original image findings by means of PCA-transformation, LDA-transformation, or feature selection (see Figure 7.5).
- Automatic classification
Access to STT classifiers over distributed web-services developed with pattern recognition technology.

As a result, the global appearance of the STT radiologist visual interface is shown in Figure 7.6. A location-independent and multi-platform system has been developed to provide decision support services for radiologists specialized in soft tissue tumour diagnosis. The system architecture handles access from multiple clinical sites to distributed data servers and classification engine servers. Node maintenance and upgrades are automatically controlled by the architecture.

The most difficult problem in developing decision support systems is collecting enough patient data to infer good PR models. The distributed architecture of STT-dDSS brings hospitals and research institutions together in developing useful tools for clinical and laboratory environments.

Feedback of clinicians

A feedback of the clinicians about the benefits of the DSS for STT was obtained through a personal interview with Luis Martí-Bonmatí, head of the Radiology Sections of the Dr. Peset Hospital and the Quirón Hospital in Valencia. For this interview, a questionnaire based on the Technology Acceptance Model (TAM) methodology [90] was prepared to predict the attitudes of the user in their willingness to use a DSS for STT. TAM was designed by Davis to identify the Perceived Usefulness and the Perceived Ease of Use, as measures of user acceptance in evaluating a software package in comparison with performing the same task through an alternative approach. In answering these questions the scale ranges from Very Likely to Very Unlikely. The value 1 is associated with Very Unlikely, a value of 7 with Very Likely and Neither Case identified with a value of 4.

Case Window

Current case | All cases

Case Prova

Complete patient name: Miquel Oltra Sastre

CORPUS FEATURES

<input type="checkbox"/> Age	0	?
<input type="checkbox"/> Clinical Presentation	Mass	?
<input type="checkbox"/> Location	Subcutaneous	?
<input type="checkbox"/> Size	0	?
<input type="checkbox"/> Shape	Rounded	?
<input type="checkbox"/> T1 Signal Intensity	Hypointense	?
<input type="checkbox"/> T2 Signal Intensity	Hypointense	?
<input type="checkbox"/> Margins	Infiltrative	?
<input type="checkbox"/> Homogeneity	Very heterogeneous	?
<input type="checkbox"/> Edema	No	?
<input type="checkbox"/> T1 Hyperintense Tracts	No	?
<input type="checkbox"/> Multiplicity	One	?
<input type="checkbox"/> Target Appearance	No	?
<input type="checkbox"/> Dependence	No	?
<input type="checkbox"/> Muscular Atrophy	No	?
<input type="checkbox"/> Calcification	No	?
<input type="checkbox"/> Vessels	No	?
<input type="checkbox"/> Intratumoral Hemorrhage	No	?
<input type="checkbox"/> Fat	No	?
<input type="checkbox"/> Fibrosis	No	?
<input type="checkbox"/> Fascial Relationship	No contact	?
<input type="checkbox"/> Bone Alteration	No	?

CORPUS CLASSIFICATION

<input type="checkbox"/> BM	Benign	?
<input type="checkbox"/> Histological Group	Linfoid	?

Create | Classification | Validate | Delete

Figure 7.3: STT Register window in the STT-dDSS visual interface, a non-real patient shown.

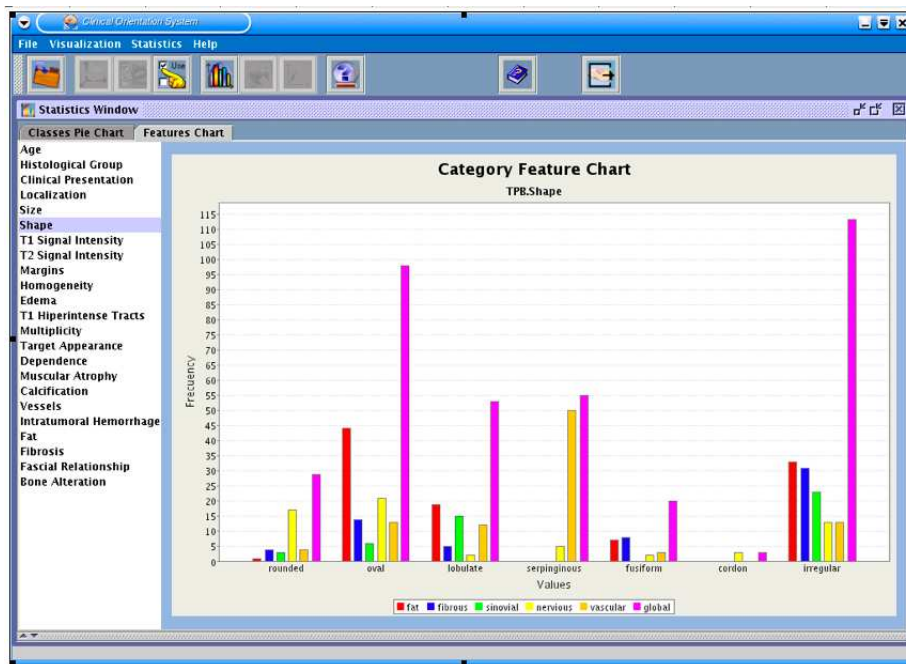


Figure 7.4: Visualization of the statistical distributions of the categorical features in the STT dataset.

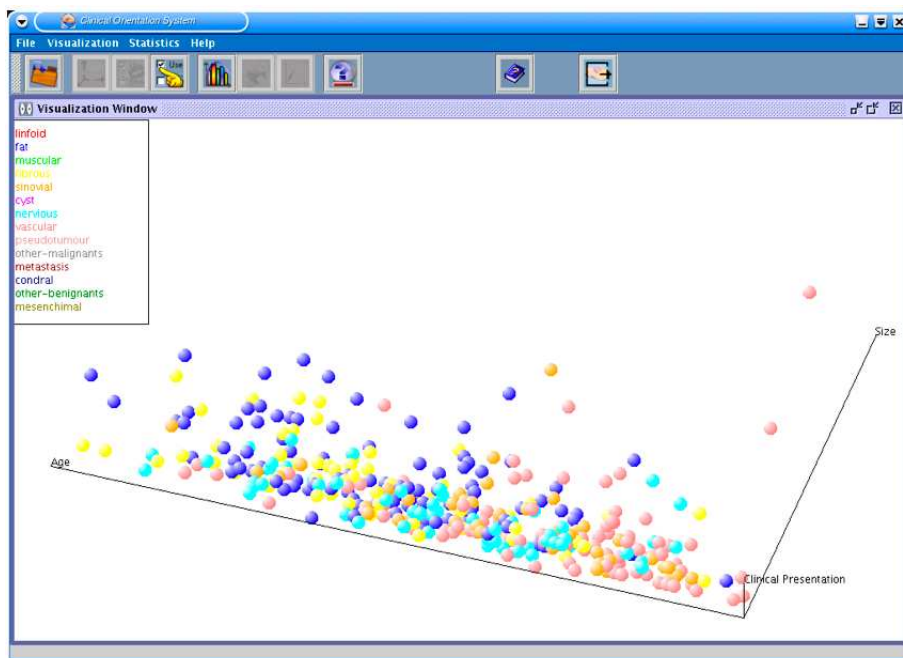


Figure 7.5: Visualization of the Universe of cases for the STT-dDSS.

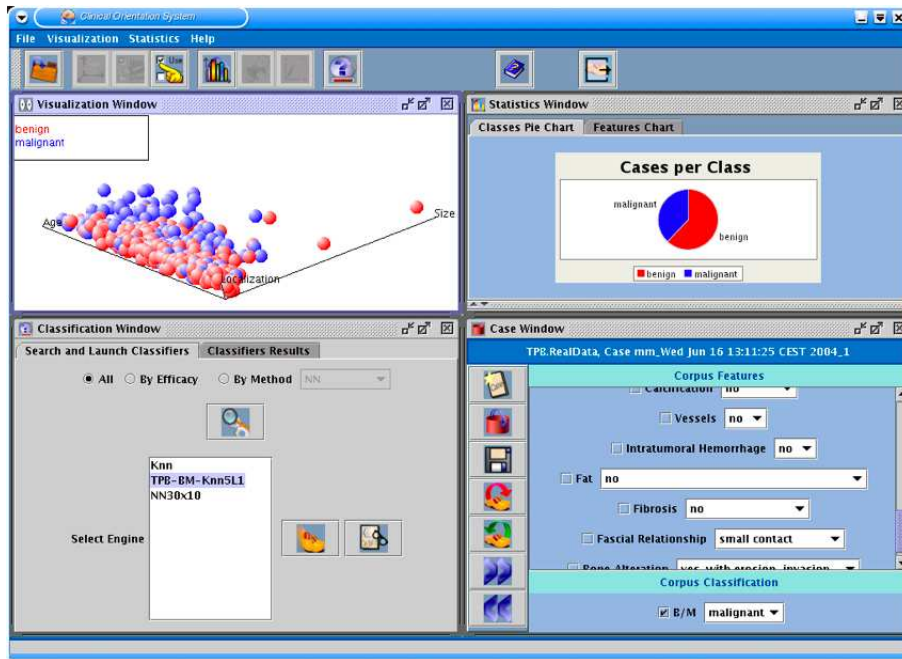


Figure 7.6: STT radiologist visual interface appearances.

The answer to the questions about the User perspective of the user's own view on tasks using the decision support system, indicated that the system improves efficiency, and ease-of-use of the task (average of 6.5 in a scale range from 1-Very Unlikely to 7-Very Likely). The clinician agreed that the DSS would improve the quality of the diagnostic decisions and the productivity (average of 5.3). The clinician believed that their hospital environments would very likely benefit from the use of the Decision Support System (average of 7). The user identified the system as being quite wise, quite positive, extremely good, and extremely beneficial. Finally, the overall reaction to the DSS was quite positive (5.8 on average in a range of [1..7]).

The ICT company, CSD (<http://www.csd.com.es>), is developing a DSS for STT based on the concepts described in this Thesis. The evaluation of the added value of the DSS is planned to be carried out in two hospitals where the system will be deployed.

7.3 eTUMOUR classification module

The Computer Aided Diagnosis System (CADs) is the CDSS for brain tumour diagnosis developed in eTUMOUR. It is based on a client-server architecture, and the main design and implementation of the system is carried out by the research group lead by C. Segebarth at INSERM (France).

The development of the CADs and the training and testing of the classifiers were carried out simultaneously during the eTUMOUR project. Furthermore, the dynamism of the dataset available for training the classifiers made the training of the classifiers an iterative process, in which new versions of the classifiers can be periodically made.

The interconnection between the CADs and the classifiers for brain tumour diagnosis was achieved using a generic interface, and a well-defined specification language for pattern

recognition techniques. The adopted solution separates the main module of the CADS from the specific implementation of the classifiers. The interconnection module enables the tumour classification to be performed independently of the rest of the system. An interface for developing classifiers makes the system scalable because adding a new classifier, or replacing an existing classifier, is an easy programming task.

The ClassifierToCADS module enables the CADS to obtain the available questions and perform the classification of new cases with the selected engines. The typical use of the module is: (1) obtain list of available questions, followed by adjustment for the question and the available biomedical signals (MR spectra at specific echo times); (2) perform the classification - the module then receiving the MRS data and the question to solve. This is followed by classification through the selected classifier. The posterior probabilities and the assigned class is then returned to the CADS to be interpreted and displayed.

To ensure scalability and provide an easy protocol for integrating new classifiers, an abstract Java class was defined in the ClassifierToCADS. Using polymorphism, this abstract class is inherited when a new classifier is created for the CADS. The abstract class forces the implementation of a common entry point for each classifier, so leaving the internal implementation open.

Nine classifiers based on Short TE, Long TE and Combined TE were developed for the CADS software. The questions to be solved by the classifiers were selected to reflect the preferences of the clinicians in an inquiry made during the eTUMOUR project. The training and test datasets consisted of the MRS data available in the eTUMOUR project on November, 2007. The data was preprocessed by Dr. Peggy Provent from INSERM using the CADS pre-processing pipeline to ensure compatibility with the future use of the system.

The three problems solved for the CADS software:

1. QH1: Is this tumor MEN or Glial (LGG + GBM) + MET (GliOMET)?
2. QH2: Is this GliOMET an AGG tumor or an LGG?
3. Q1: Is this tumor an MEN, an LGG, or an AGG?

Based on the results reported in Chapters 3 and 5, LDA using as inputs peak integrated values of the 15 main resonances observed in Short TE, and the nine main resonances observed in Long TE, was selected as methodologies for the classifiers. Table 7.2 shows the evaluation results of the three approaches (Long TE, Short TE, and Combined TE) for the three discrimination questions - Hierarchical question 1: Meningioma vs. (Glioma and Metastasis) (QH1); Hierarchical question 2: Low grade glial vs. (Glioblastoma and Metastasis) (QH2); and Low Grade Meningioma vs. Glioblastoma+Metastasis vs. Low Grade Glial (Q1). The best results were always obtained using the Combined TE approach, but the most relevant differences were observed in the QH1 and the Q1 classifiers. Long TE classifiers achieved significant lower performances than Short TE and Combined TE. In question QH2, Short TE and Combined TE approaches achieved similar performances.

The performances seen in the classifiers were consistent with those reported in previous studies. For MEN discrimination, our classifiers obtained a better performance when features from the two echo times were combined [91], and for LGG and AGG discrimination, similar behaviour was observed by the Short TE and the combination of signals. The good

Table 7.2: GMOR for the three approaches in the questions to solve.

Question	Long TE	Short TE	Combined TE
QH1	0.78	0.92	0.94
QH2	0.84	0.92	0.92
Q1	0.75	0.87	0.90

performance of the predictive models evaluated over the combination of datasets from different multicenter projects is also noteworthy. The specific experiments and specifications of the classifiers are included in the final technical report of the pattern recognition work package in the eTUMOUR project [92]. The conclusions of the multicentric-multiproject study (Chapter 5) agreed with the CADS classification module solution - the study used the same MRS data but with a different preprocessing implementation and aimed at solving partially different problems.

Finally, the three multiclass classifiers for the Q1 question were selected for the CADS version for clinical validation. In the final implementation, the Short TE, the Long TE or the Combined TE classifier is used depending on the available signals. Preference was given to the Combined TE classifier if both signals were acquired.

7.4 HEALTHAGENTS: Distributed multi-agent DSS

The HEALTHAGENTS scenario is one in which distributed datamarts are being built by hospital groups in various countries throughout Europe. Individual hospitals do not typically collect sufficient cases of a particular tumour type to form a training set that is sufficiently large for developing robust software-based tumour classifiers that can provide reliable diagnoses when presented with non-invasive imaging and spectroscopic data.

The solution adopted in the dDSS of HEALTHAGENTS is based on software agents. The use of agent technology and ontologies is not new. Several authors have described systems in medical and bio-informatics domain and elsewhere [93, 94]. The characteristics of this technology fit with the needs of dDSS [95]: (1) dynamic reasoning capabilities are provided by the classification engines; (2) there is a strong interaction among different types as database agents (classifier agents, preprocessing agents, GUI agents, Evidence-based Search Service (EbSS), and Yellow Pages (YP) agents); (3) unique identities enable roaming with the YP agents; (4) agents possess an internal knowledge-based state which can be dynamically altered by a dynamic evaluation, or ranking of the classifiers.

An ontology of the specific project domain was defined to ensure a correct interoperability between software agents. This ontology structures the knowledge implicit in the application data, and also provides a platform for reasoning. The languages included in the HEALTHAGENTS project cover the domain of brain tumours, medical imaging, biomedical data, clinical and histopathological domains (e.g. based on the WHO taxonomy), anatomical location, pattern recognition methods, evaluation strategies, and the communication language used by agents.

The architecture of the dDSS of the HEALTHAGENTS network is composed of software agents with several specific functionalities:

- Database agents to handle input and output of cases to and from hospital datamarts.

- Preprocessing agents to convert imaging data to canonical form.
- GUI interface agents for handling interaction with medical users at hospital nodes.
- YP agents to keep track of system resources: including the location of case data, classifiers and their profiles, performance, and reputations.
- Classifier agents to provide tumour class predictors based on case data.
- Petitioner agents to invoke appropriate classifiers, gather new data and rank the classifiers based on their performance.
- EbSS agents to provide context-sensitive information searching.

The general activity scenario in the distributed HEALTHAGENTS system starts in the clinical centers, where patients are diagnosed and treated in different units. MRI and MRS raw data is available from the radiological unit. Decision support, using the dDSS, for the radiological diagnosis based on the MRS is considered to be the most common use of the system. After surgery, if an histopathological diagnosis confirmation is available, it may be possible to share the case with the system for training purposes. In both cases, the cases are link anonymised and copied to the hospital's local HEALTHAGENTS datamart.

To visualize and classify the patient data (especially the MR spectra), the data should be pre-processed by following the well established protocols (see Section 5.2.2). The GUI consults nearby YP to establish the availability of preprocessing agents. The communication between the GUI agent and the preprocessing agent is established and, as a result, the patient data is transformed into canonical form.

At the request of a medical user, the availability of appropriate classifiers is consulted on the YP, on the basis of tumour types, and the contextual patient information (age, gender, tumour location, and geographical location). The classifiers may be located at classifier nodes anywhere on the HEALTHAGENTS network, in which case the data to be classified may be sent from the hospital to the remote classifier nodes for classification. If the hospital does not allow data to leave the local node, classifiers may be run locally. Results from the different classifiers are gathered, ranked, and returned to the user via the GUI to support the user's decision making processes. All potentially suitable classifiers in HEALTHAGENTS are invoked to classify the current case, and the various factors influencing classifier choice are used to rank the results. Classifier results are also recorded in the system so that, when a confirmed diagnosis is available for a case, an estimate of the performance and reputation of classifiers can be updated.

When a new classifier (or version) is developed, it is deployed in a classification node of the system and its existence, initial performance, reputation, training profile, and test data are published in the HEALTHAGENTS YP.

In addition to the above classification processes, the HEALTHAGENTS system provides an EbSS which seeks, in a context-sensitive way, articles from the medical literature to assist the medical user in the current task. An automatic search mode, based on the keywords, may be triggered by the classification processes being undertaken.

Conceived as an open-source platform, the HEALTHAGENTS dDSS is implemented using the Jade agent development environment (Java, Ant, and D2RQ are supported under Windows and Linux platforms. The aim is a distribution in four different types

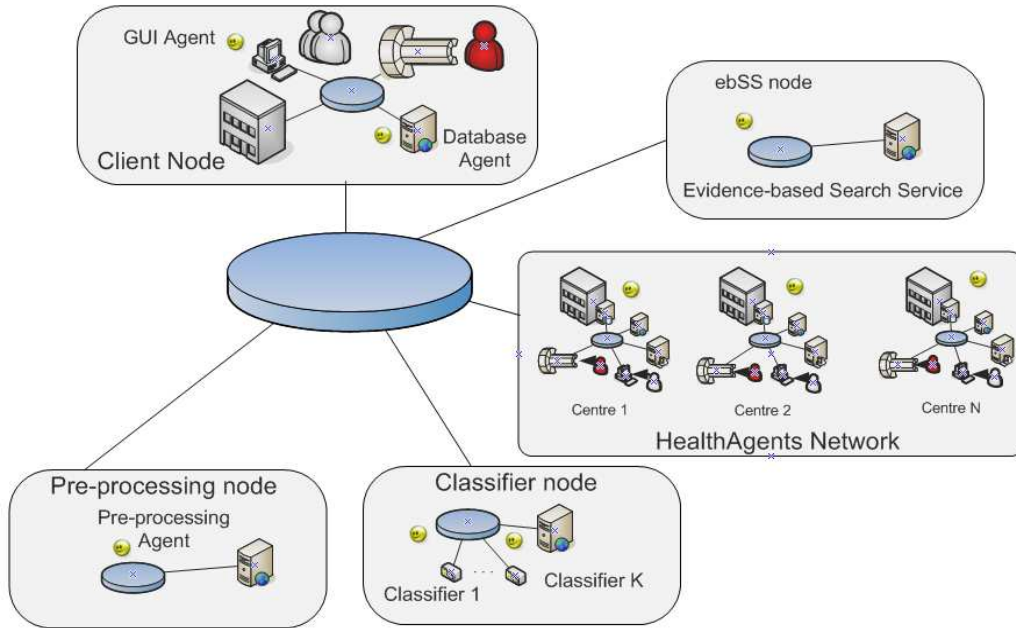


Figure 7.7: The multi-node HEALTHAGENTS architectural implementation

of computing nodes (GUI node, ebSS node, Classifier node, Pre-processing node) with at least one active agent, as depicted in Figure 7.7.

Figure 7.8 shows the classification window of the GUI for the radiological units. The ranked list of appropriate classifiers for the test case is shown at the top of the window. Each row shows the classification result, the ranks of the classifiers, and the dynamic evaluation of the classifiers. A pie chart on the right shows the posterior probability of the case for each class obtained by the classifier. On the left of the window, a visual inspection of the MR spectra of the test case is shown. Vertical lines indicate the relevant features used by the selected classifier. It is possible to compare the spectra with the mean plus/minus standard deviation patterns obtained from the training dataset for each class. When a classifier with reduced-rank projections (as the FLDA) is selected, the latent space is shown on the bottom right.

A Prospective Study of the added value of the use of the dDSS for brain tumour diagnosis was carried out in the HEALTHAGENTS project. It was performed in the project led by a team from the University of Birmingham^a following the TAM methodology [90]. In this study, 26 clinicians were interviewed after using the dDSS for brain tumor diagnosis. The average amount of clinical experience was 13 years. All clinicians had a major part of their practice caring for patients with brain tumours and included radiologists, oncologists, neurosurgeons and histopathologists.

In their answers, they positively believed that the use of the DSS would improve the quality of their diagnostic decisions (average of 5.2 in a scale range of [1..7]), and especially the non-invasive diagnostic accuracy of their patients (average of 5.7). Additionally, they believe that the HEALTHAGENTS DSS will not provide them the flexibility to do their

^aFull study available in Deliverable D27 (A Prospective Study of the added value of the use of the d-DSS for brain tumour diagnosis) of HEALTHAGENTS. Alex Gibb, 2009. www.healthagents.net

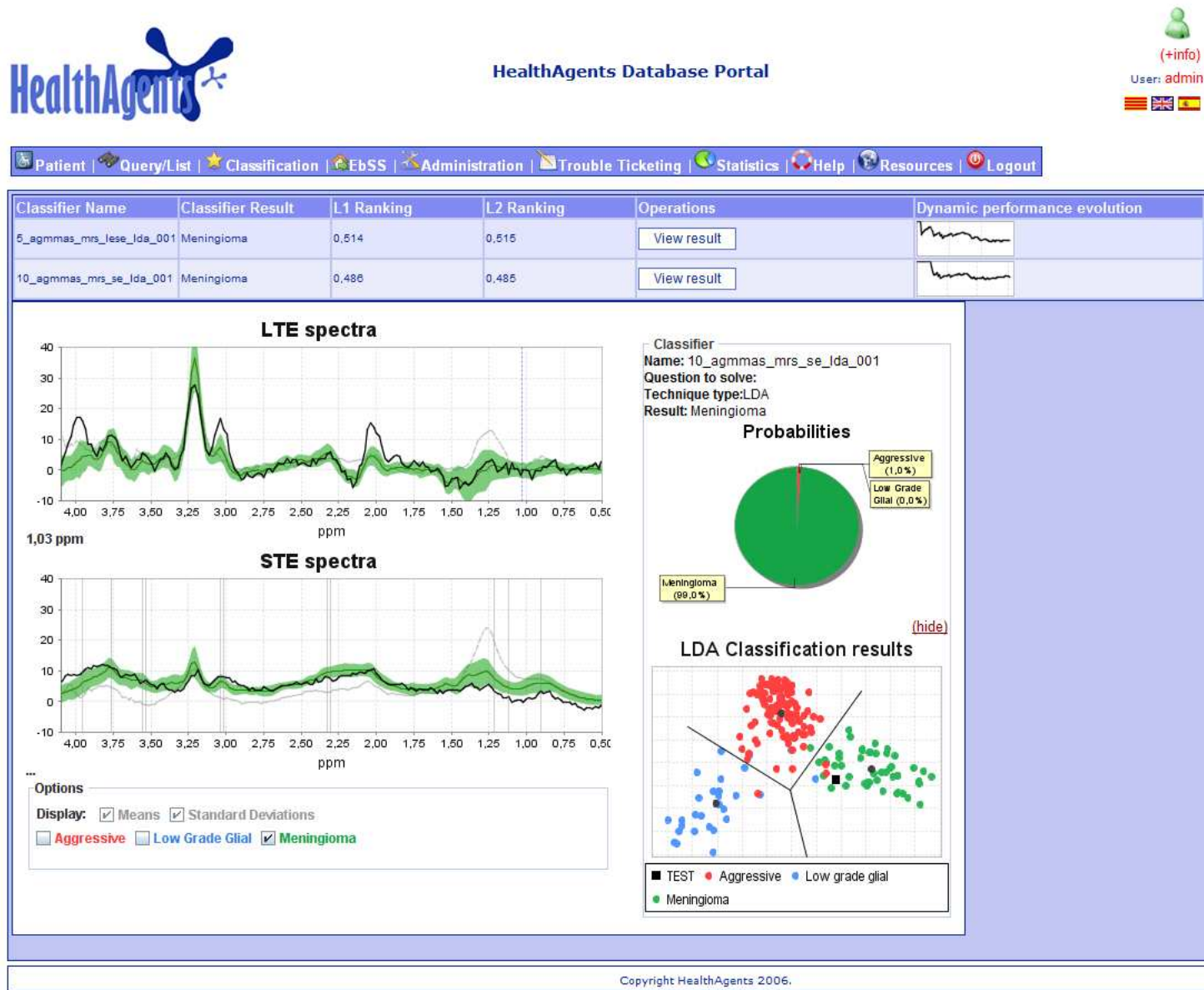


Figure 7.8: Classification window of the GUI for the radiological units.

task in the way they want (average of 4.5). However, they consider the system easy to use in their task (average of 7). Finally, the clinicians believed that their hospital environments would more than likely benefit from the use of the Decision Support System.

The overall response, from the TAM, has been the belief that the system would be beneficial for the clinicians in performing diagnostic and prognostic evaluations for patients with brain tumours. Specific comments from the clinicians identified that the practical application of the system would be more useful for less common types of tumours, as many common diagnoses can be achieved from MRI. In addition the system would be useful for brain stem tumours, as these cannot be biopsied, and for following the progression of the treatment plans. Finally, the identification of a particular tumour type given the choice of two or more for a particular location was especially appealing.

7.5 Conclusion

The practical results of the pattern recognition experiments and software development applied in this thesis are three DSSs for the diagnosis of STT and BT.

The dDSS for STT could offer radiologists novel and powerful methods for soft tissue tumour diagnosis. It provides access to distributed data, statistical analysis, graphical representation, and pattern recognition classification. This tool will be useful because it offers an objective methodology to assist the radiologist in evaluating a new case, and can help train new radiologists without expertise in STT. The introduced architecture enables experts to audit and upgrade pattern recognition engines and improve the diagnosis decision task. The main ideas developed for this system were adopted in HEALTHAGENTS to develop the agent-based dDSS for brain tumour diagnosis.

Two practical solutions for integrating the pattern recognition engines in the clinical routine workflow were carried out for the eTUMOUR and HEALTHAGENTS decision support systems in brain tumour diagnosis. For the eTUMOUR project, a set of classifiers and a generic interface between the CDSS and the classification module was developed and integrated in the CADS software for clinical validation of the system. Classification agents are the core of the dDSS of the HEALTHAGENTS European network.

A generic Decision Support System, Curiam [96], featuring an assembled view of predictive models for clinical and radiological data incorporates automatic classifiers for both STT and the BT tasks as a desktop software solution (<http://bmg.webs.upv.es/>).

Bibliography

- [1] Garcia-Gomez JM, Vidal C, Vicente J, Martí-Bonmatí L, Robles M (2004) Medical decision support system for diagnosis of soft tissue tumors based on distributed architecture. *Conf Proc IEEE Eng Med Biol Soc* 5: 3225–3228
- [2] García-Gómez JM, Vidal C, Martí-Bonmatí L, Robles M (2004) Distributed decision support architecture for soft tissue tumor diagnosis. In *MEDINFO 2004: 11th World Congress on Medical Informatics*, p. 1608
- [3] González-Vélez H, Mier M, Julià-Sapé M, Arvanitis T, García-Gómez J, Robles M, Lewis P, Dasmahapatra S, Dupplaw D, Peet A, Arús C, Celda B, Van Huffel S, Lluch-Ariet M (2007) HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis. *Applied Intelligence* [Article in Press]
- [4] García-Gómez JM, Tortajada S, Vicente J, Sáez C, Castells X, Luts J, Julià-Sapé M, Juan-Císcar A, Van Huffel S, Barcelo A, Ariño J, Arús C, Robles M (2007) Genomics and Metabolomics Research for Brain Tumour Diagnosis Based on Machine Learning. In *IWANN, Lecture Notes in Computer Sciences 4507*, pp. 1012–1019
- [5] Leaper DJ, De Dombal FT, Horrocks JC, Staniland JR (1972) Computer-assisted diagnosis of abdominal pain using estimates provided by clinicians. *Br J Surg* 59(11): 897–898
- [6] Leaper DJ, Horrocks JC, Staniland JR, De Dombal FT (1972) Computer-assisted diagnosis of abdominal pain using "estimates" provided by clinicians. *Br Med J* 4(5836): 350–354
- [7] Shortliffe EH, Scott AC, Bischoff MB, Campbell AB, va Melle W, Jacobs CD (1981) ONCOCIN: An expert system for oncology protocol management. In *Seventh International Joint Conference on Artificial Intelligence*. Vancouver
- [8] Shortliffe EH (1981) ONCOCIN: An aid for the outpatient management of cancer patients. In *Proceedings of the Society for Computer Medicine and the Society for Advanced Medical Systems*. Washington, DC
- [9] Smith RA, Saslow D, Sawyer KA, Burke W, Costanza ME, Evans WPr, Foster RSJ, Hendrick E, Eyre HJ, Sener S (2003) American Cancer Society guidelines for breast cancer screening: update 2003. *CA Cancer J Clin* 53(3): 141–169
- [10] Lisboa PJG, Wong H, Harris P, Swindell R (2003) A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. *Artif Intell Med* 28(1): 1–25

- [11] Tan KC, Yu Q, Heng CM, Lee TH (2003) Evolutionary computing for knowledge discovery in medical diagnosis. *Artif Intell Med* 27(2): 129–154
- [12] Andrews RJ, Mah RW (2003) The NASA Smart Probe Project for real-time multiple microsensor tissue recognition. *Stereotact Funct Neurosurg* 80(1-4): 114–119
- [13] Nattkemper TW, Arnrich B, Lichte O, Timm W, Degenhard A, Pointon L, Hayes C, Leach MO (2005) Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods. *Artif Intell Med* 34(2): 129–139
- [14] Dasmahapatra S, Dupplaw D, Hu B, Lewis PH, Shadbolt N (2005) Ontology-Mediated Distributed Decision Support for Breast Cancer. In S Miksch, J Hunter, ET Keravnou, editors, *AIME 2005: 10th Conf on Artificial Intelligence in Medicine*, volume 3581 of *Lect. Notes Comput. Sc.*, pp. 221–225. Springer-Verlag, Aberdeen, UK
- [15] Markey MK, Tourassi GD, Margolis M, DeLong DM (2006) Impact of missing data in evaluating artificial neural networks trained on complete data. *Comput Biol Med* 36(5): 516–525
- [16] Biganzoli EM, Boracchi P, Ambrogi F, Marubini E (2006) Artificial neural network for the joint modelling of discrete cause-specific hazards. *Artif Intell Med* 37(2): 119–130
- [17] Leinsinger G, Schlossbauer T, Scherr M, Lange O, Reiser M, Wismuller A (2006) Cluster analysis of signal-intensity time course in dynamic breast MRI: does unsupervised vector quantization help to evaluate small mammographic lesions? *Eur Radiol* 16(5): 1138–1146
- [18] Abidi SR, Abidi SSR, Hussain S, Shepherd M (2007) Ontology-based modeling of clinical practice guidelines: a clinical decision support system for breast cancer follow-up interventions at primary care settings. *Medinfo* 12(Pt 2): 845–849
- [19] Hassanien AE (2007) Fuzzy rough sets hybrid scheme for breast cancer detection. *Image Vision Comput* 25(2): 172–183
- [20] Iakovidis DK, Maroulis DE, Karkanis SA (2006) An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy. *Comput Biol Med* 36(10): 1084–1103
- [21] Zheng MM, Krishnan SM, Tjoa MP (2005) A fusion-based clinical decision support for disease diagnosis from endoscopic images. *Comput Biol Med* 35(3): 259–274
- [22] Lucas PJ, Boot H, Taal BG (1998) Computer-based decision support in the management of primary gastric non-Hodgkin lymphoma. *Methods Inf Med* 37(3): 206–219
- [23] Mofidi R, Deans C, Duff MD, de Beaux AC, Paterson Brown S (2006) Prediction of survival from carcinoma of oesophagus and oesophago-gastric junction following surgical resection using an artificial neural network. *Eur J Surg Oncol* 32(5): 533–539

-
- [24] Kawazu T, Araki K, Kanda S (2001) Application of neural networks to the prediction of lymph node metastasis in oral cancer. In *CARS*, pp. 1295–1296
- [25] Nayak GS, Kamath S, Pai KM, Sarkar A, Ray S, Kurien J, D’Almeida L, Krishnanand BR, Santhosh C, Kartha VB, Mahato KK (2006) Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra: classification of normal premalignant and malignant pathological conditions. *Biopolymers* 82(2): 152–166
- [26] Kumar A, Yip YL, Smith B, Grenon P (2006) Bridging the gap between medical and bioinformatics: an ontological case study in colon carcinoma. *Comput Biol Med* 36(7-8): 694–711
- [27] Futschik ME, Reeve A, Kasabov N (2003) Evolving connectionist systems for knowledge discovery from gene expression data of cancer tissue. *Artif Intell Med* 28(2): 165–189
- [28] Anand SS, Smith AE, Hamilton PW, Anand JS, Hughes JG, Bartels PH (1999) An evaluation of intelligent prognostic systems for colorectal cancer. *Artif Intell Med* 15(2): 193–214
- [29] van Oosterhout EM, Talmon JL, De Clercq PA, Schouten HC, Jansen MP, Hasman A (2003) The PropeR way to support medical doctors in daily practice. Developing the protocol based DSS. *Stud Health Technol Inform* 95: 340–345
- [30] Belacel N, Wang Q, Richard R (2005) Web-integration PROAFTN methodology for acute leukemia diagnosis. *Telemed J E Health* 11(6): 652–659
- [31] Foran DJ, Comaniciu D, Meer P, Goodell LA (2000) Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy. *IEEE Trans Inf Technol Biomed* 4(4): 265–273
- [32] Chen W, Meer P, Georgescu B, He W, Goodell LA, Foran DJ (2005) Image mining for investigative pathology using optimized feature extraction and data fusion. *Comput Methods Programs Biomed* 79(1): 59–72
- [33] Ratei R, Karawajew L, Lacombe F, Jagoda K, Del Poeta G, Kraan J, De Santiago M, Kappelmayer J, Bjorklund E, Ludwig WD, Gratama JW, Orfao A (2007) 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* 21(6): 1204–1211
- [34] Covell DG, Wallqvist A, Rabow AA, Thanki N (2003) Molecular classification of cancer: unsupervised self-organizing map analysis of gene expression microarray data. *Mol Cancer Ther* 2(3): 317–332
- [35] Tung WL, Quek C (2005) GenSo-FDSS: a neural-fuzzy decision support system for pediatric ALL cancer subtype identification using gene expression data. *Artif Intell Med* 33(1): 61–88

- [36] Yuan X, Yang Z, Zouridakis G, Mullani N (2006) SVM-based texture classification and application to early melanoma detection. *Conf Proc IEEE Eng Med Biol Soc* 1: 4775–4778
- [37] Sboner A, Eccher C, Blanzieri E, Bauer P, Cristofolini M, Zumiani G, Forti S (2003) A multiple classifier system for early melanoma diagnosis. *Artif Intell Med* 27(1): 29–44
- [38] Debeir O, Decaestecker C, Pasteels JL, Salmon I, Kiss R, Van Ham P (1999) Computer-assisted analysis of epiluminescence microscopy images of pigmented skin lesions. *Cytometry* 37(4): 255–266
- [39] Coleman DJ, Silverman RH, Rondeau MJ, Boldt HC, Lloyd HO, Lizzi FL, Weingeist TA, Chen X, Vangveeravong S, Folberg R (2004) Noninvasive in vivo detection of prognostic indicators for high-risk uveal melanoma: ultrasound parameter imaging. *Ophthalmology* 111(3): 558–564
- [40] Binder M, Kittler H, Dreiseitl S, Ganster H, Wolff K, Pehamberger H (2000) Computer-aided epiluminescence microscopy of pigmented skin lesions: the value of clinical data for the classification process. *Melanoma Res* 10(6): 556–561
- [41] Carrara M, Bono A, Bartoli C, Colombo A, Lualdi M, Moglia D, Santoro N, Tolomio E, Tomatis S, Tragni G, Santinami M, Marchesini R (2007) Multispectral imaging and artificial neural network: mimicking the management decision of the clinician facing pigmented skin lesions. *Phys Med Biol* 52(9): 2599–2613
- [42] Tsai CA, Chen DT, Chen JJ, Balch CM, Thompson JF, Soong SJ (2007) An integrated tree-based classification approach to prognostic grouping with application to localized melanoma patients. *J Biopharm Stat* 17(3): 445–460
- [43] Wigren T, Kolari P (1994) Evaluation of a decision-support system for inoperable non-small cell lung cancer. *Methods Inf Med* 33(4): 397–401
- [44] Coppini G, Diciotti S, Falchini M, Villari N, Valli G (2003) Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms. *IEEE Trans Inf Technol Biomed* 7(4): 344–357
- [45] Campadelli P, Casiraghi E, Artioli D (2006) A fully automated method for lung nodule detection from postero-anterior chest radiographs. *IEEE Trans Med Imaging* 25(12): 1588–1603
- [46] Matsopoulos GK, Mouravliansky NA, Asvestas PA, Delibasis KK, Kouloulis V (2005) Thoracic non-rigid registration combining self-organizing maps and radial basis functions. *Med Image Anal* 9(3): 237–254
- [47] Bala M (2004) [Determining the possibility of collecting reliable data for use in decision making in health care on the example of cost-effectiveness analysis of methods used in smoking cessation]. *Przegl Lek* 61(10): 1180–1183
- [48] Lenhard REJ, Waalkes TP, Herring D (1983) Evaluation of the clinical management of cancer patients. A pilot study. *JAMA* 250(24): 3310–3316

-
- [49] Shi H, Lyons-Weiler J (2007) Clinical decision modeling system. *BMC Med Inform Decis Mak* 7: 23
- [50] McCowan IA, Moore DC, Nguyen AN, Bowman RV, Clarke BE, Duhig EE, Fry MJ (2007) Collection of cancer stage data by classifying free-text medical reports. *J Am Med Inform Assoc* 14(6): 736–745
- [51] Spyridonos P, Cavouras D, Ravazoula P, Nikiforidis G (2002) A computer-based diagnostic and prognostic system for assessing urinary bladder tumour grade and predicting cancer recurrence. *Med Inform Internet Med* 27(2): 111–122
- [52] Papageorgiou EI, Spyridonos PP, Stylios CD, Ravazoula P, Groumpos PP, Nikiforidis GN (2006) Advanced soft computing diagnosis method for tumour grading. *Artif Intell Med* 36(1): 59–70
- [53] Sylvester RJ, van der Meijden APM, Oosterlinck W, Witjes JA, Bouffieux C, Denis L, Newling DWW, Kurth K (2006) Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur Urol* 49(3): 466–465
- [54] Tachibana M, Miyakawa A, Deguchi N, Baba S, Murai M, Tazaki H (1994) A new scoring system based on the histological behavior and proliferative activity of tumor cells for grading the malignant potential of bladder cancers. *Int J Urol* 1(1): 37–42
- [55] Papageorgiou EI, Stylios CD, Groumpos PP (2003) An integrated two-level hierarchical system for decision making in radiation therapy based on fuzzy cognitive maps. *IEEE Trans Biomed Eng* 50(12): 1326–1339
- [56] Kelm BM, Menze BH, Zechmann CM, Baudendistel KT, Hamprecht FA (2007) Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: pattern recognition vs quantification. *Magn Reson Med* 57(1): 150–159
- [57] Zhu Y, Williams S, Zwiggelaar R (2006) Computer technology in detection and staging of prostate carcinoma: a review. *Med Image Anal* 10(2): 178–199
- [58] Mattfeldt T, Trijic D, Gottfried HW, Kestler HA (2004) Classification of incidental carcinoma of the prostate using learning vector quantization and support vector machines. *Cell Oncol* 26(1-2): 45–55
- [59] Vlahou A, Schorge J, Gregory B, Coleman R (2003) Diagnosis of Ovarian Cancer Using Decision Tree Classification of Mass Spectral Data. *J Biomed Biotechnol* 2003(5): 308–314
- [60] Armstrong K, Weber B, Ubel PA, Peters N, Holmes J, Schwartz JS (2005) Individualized survival curves improve satisfaction with cancer risk management decisions in women with BRCA1/2 mutations. *J Clin Oncol* 23(36): 9319–9328
- [61] Tan TZ, Quek C, Ng GS (2005) Ovarian cancer diagnosis by hippocampus and neocortex-inspired learning memory structures. *Neural Netw* 18(5-6): 818–825

- [62] Szpurek D, Moszynski R, Smolen A, Sajdak S (2005) Artificial neural network computer prediction of ovarian malignancy in women with adnexal masses. *Int J Gynaecol Obstet* 89(2): 108–113
- [63] Van Holsbeke C, Van Calster B, Testa AC, Domali E, Lu C, Van Huffel S, Valentin L, Timmerman D (2009) Prospective internal validation of mathematical models to predict malignancy in adnexal masses: results from the international ovarian tumor analysis study. *Clin Cancer Res* 15(2): 684–691
- [64] Nieminen P, Hakama M, Viikki M, Tarkkanen J, Anttila A (2003) Prospective and randomised public-health trial on neural network-assisted screening for cervical cancer in Finland: results of the first year. *Int J Cancer* 103(3): 422–426
- [65] Lisboa PJ, Taktak AFG (2006) The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Netw* 19(4): 408–415
- [66] Shortliffe E (2006) Medical thinking: what should we do? In *Conference on Medical Thinking*
- [67] van Bommel JH (1997) *Handbook of Medical Informatics*, chapter *Methods for Decision Support*, pp. 233–260. Springer
- [68] Kaplan B (2001) Evaluating informatics applications—clinical decision support systems literature review. *Int J Med Inform* 64(1): 15–37
- [69] Malhotra S, Jordan D, Shortliffe E, Patel VL (2007) Workflow modeling in critical care: piecing together your own puzzle. *J Biomed Inform* 40(2): 81–92
- [70] Weatherall PT (1995) Benign and malignant masses. MR imaging differentiation. *Magnetic Resonance Imaging Clinics of North America* 3(4): 669–694
- [71] Hagberg G (1998) From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. *NMR in Biomedicine* 11(4-5): 148–156
- [72] INTERPRET Consortium (Accessed: 28 April 2008) Interpret web site. <http://azizu.uab.es/INTERPRET>
- [73] Julia-Sape M, Acosta D, Mier M, Arús C, Watson D (2006) A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. *Magn Reson Mater Phy* 19(1): 22–33
- [74] Tate AR, Underwood J, Acosta DM, Julia-Sape M, Majos C, Moreno-Torres A, Howe FA, van der Graaf M, Lefournier V, Murphy MM, Loosemore A, Ladroue C, Wesseling P, Luc Bosson J, Cabanas ME, Simonetti AW, Gajewicz W, Calvar J, Capdevila A, Wilkins PR, Bell BA, Remy C, Heerschap A, Watson D, Griffiths JR, Arús C (2006) 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
- [75] Devos A, Lukas L, Suykens JAK, Vanhamme L, Tate AR, Howe FA, Majos C, Moreno-Torres A, van der Graaf M, Arús C, Van Huffel S (2004) Classification of brain tumours using short echo time 1H MR spectra. *J Magn Reson* 170(1): 164–175

-
- [76] Tate AR, Majos C, Moreno A, Howe FA, Griffiths JR, Arús C (2003) Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. *Magnetic Resonance in Medicine* 49(1): 29–36
- [77] Simonetti AW, Melssen WJ, Szabo de Edelenyi F, van Asten JJA, Heerschap A, Buydens LMC (2005) Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. *NMR in Biomedicine* 18(1): 34–43
- [78] Devos A (2005) Quantification and classification of Magnetic Resonance Spectroscopy data and applications to brain tumour recognition. Ph.D. thesis, Faculty of Engineering, K.U.Leuven
- [79] Lukas L, Devos A, Suykens JAK, Vanhamme L, Howe FA, Majós C, Moreno-Torres A, Graaf MVD, Tate AR, Arús C, Huffel SV (2004) Brain tumor classification based on long echo proton MRS signals. *Artif Intell Med* 31: 73–89
- [80] Menze BH, Lichy MP, Bachert P, Kelm BM, Schlemmer HP, Hamprecht FA (2006) Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors. *NMR in Biomedicine* 19(5): 599–609
- [81] Simonetti AW, Melssen WJ, van der Graaf M, Postma GJ, Heerschap A, Buydens LMC (2003) A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy. *Anal Chem* 75(20): 5352–5361
- [82] Devos A, Simonetti AW, van der Graaf M, Lukas L, Suykens JAK, Vanhamme L, Buydens LMC, Heerschap A, Van Huffel S (2005) The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. *J Magn Reson* 173(2): 218–228
- [83] 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. Technical report, FP6-2002-LIFESCIHEALTH 503094, VI framework programme, EC <http://www.etumour.net>
- [84] Garcia-Gomez JM, Vidal C, Martí-Bonmatí L, Galant J, Sans N, Robles M, Casacuberta F (2004) Benign/malignant classifier of soft tissue tumors using MR imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine* 16(4): 194–201
- [85] García-Gómez JM, Vidal C, Juan A, Martí-Bonmatí L, Galant J, Robles M (2002) Soft tissue tumours pattern classification techniques using MR and epidemiology data. *Magnetic Resonance Materials in Physics, Biology and Medicine* 15: 274–275
- [86] Malassidis E, García-Gómez JM, Vidal C, Martí-Bonmatí L, Robles M, Millet J (2002) El análisis de componentes principales como método de clasificación y visualización de tumores de partes blandas. In *INFORMED-2002*, pp. 213–215
- [87] Vidal C, Garcia-Gomez JM, Marti-Bonmati L, Robles M, Galant J, Millet J (2002) Study of vascular histological group tumors using pattern recognition techniques. In *INFORMED*, pp. 207–211

- [88] Duda R, Hart P, Stork D (2001) *Pattern Classification*. John Wiley and Sons, inc.
- [89] Garcia-Gomez JM, Vidal C, Císcar AJ, Martí-Bonmatí L, Galant J, Robles M (2002) Pattern classification techniques applied to soft tissue tumors using MR and epidemiological data. In Proceedings 19th annual meeting of the European Society for Magnetic Resonance in Medicine and Biology, pp. 274–275
- [90] Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* 13(3): 319–340
- [91] Garcia-Gomez J, Tortajada S, Vidal C, Julia-Sape M, Luts J, Van Huffel S, Arús C, Robles M (2008) The influence of combining two echo times in automatic brain tumor classification by Magnetic Resonance Spectroscopy. *NMR in Biomedicine* 21(10): 1112–1125
- [92] García-Gómez JM (2008) Final considerations of the Pattern Recognition workpackage. Technical report, eTumour project www.etumour.net
- [93] Luck M, Merelli E (2005) Agents in bioinformatics. *Knowl Eng Rev* 20(2): 117–125
- [94] Merelli E, Armano G, Cannata N, Corradini F, d’Inverno M, Doms A, Lord P, Martin A, Milanese L, Möller S, Schroeder M, Luck M (2006) Agents in bioinformatics, computational and systems biology. *Brief Bioinform* In press. doi:10.1093/bib/bbl014
- [95] Brugali D, Sycara K (2000) Towards agent oriented application frameworks. *ACM Computing Surv* 32(1): 21–27
- [96] Sáez C, García-Gómez J, Vicente J, Tortajada S, Esparza M, Navarro A, Fuster-García E, Robles M, Martí-Bonmatí L, Arús C (2008) A generic Decision Support System featuring an assembled view of predictive models for Magnetic Resonance and clinical data. In *ESMRMB 2008: 25th Annual Scientific Meeting*. Springer

Chapter 8

Concluding remarks and recommendations

8.1 Conclusions

This thesis deals with the application of Pattern Recognition (PR) to biomedical data for the development of Clinical Decision-Support Systems (CDSSs) for two cancer pathologies: Brain Tumor (BT) and Soft Tissue Tumor (STT). In the following, the main conclusions are presented.

- The PR-based classifiers for distinguishing the benign/malignant character of an STT achieved an accuracy of 90% when evaluated in a multicenter dataset. A new distributed architecture that incorporates these classifiers has been designed for a STT CDSS. Some concepts developed in this architecture were useful for posterior developments in the distributed Decision-Support System (dDSS) for HEALTHAGENTS.
- A new approach that combines MRS echo times has been proposed for BT classification based on MRS. Significant differences in performance were found when Short TE, Long TE, or a combination of both spectra were used for classification. In our dataset, the combination of the Long and Short TE acquisitions yielded optimal performance in the discrimination of Low-grade meningiomas (MEN). For the discrimination of Aggressive tumor: GBM and MET (AGG) from Low-Grade Glial (LGG), the use of Short TE acquisition alone was preferable.
- A probabilistic mixture model and the e(xpectation)-m(aximisation)-based estimation of parameters for binned and truncated data with univariate mixture densities of means relative to a global shift has been proposed for Short TE MRS data characterization. The new approach keeps the biological information in the model and correctly fits the MR spectra. A highly accurate discrimination of brain tumors based on the parametric space of the probabilistic mixture model is possible.
- It has been demonstrated that reliable prediction of BT types using in-vivo MRS acquired at later dates, from different hospitals, or using different instrumentation; but obtained under similar acquisition conditions, may be possible using models inferred by multicenter datasets. Our results consolidate the conclusions of previous

studies on automatic brain tumor classification using MRS - but using multicenter-multiproject data and subsequently acquired cases. The classification results obtained by the PR-based models can be useful for the validation and review of cases with abnormal profiles in the datasets of multicenter projects.

- A proof of principle in brain tumor classification of Glioblastoma (GBM) and Meningothelial Meningioma (mMEN) biopsies using gene expression profile produced an accuracy of 100% in a blinded independent test dataset. The gene-signatures obtained is in accordance with the expected biological and pathological differences in the tumors: glycolytic phenotype, bicarbonate export capability, cell detoxification and intracellular transport in GBM; and cell adhesion, and TGF- β signalling pathways in mMEN.
- Two practical solutions to integrate the PR engines in the clinical routine workflow were proposed for the eTUMOUR and HEALTHAGENTS CDSSs for BT diagnosis. The former is a generic interface between the main module of the software and the classification module that was developed for the eTUMOUR project, and which allows the inclusion of new classifiers at execution time. The latter is the classification framework for the agent-based dDSS, which is the core of a European network to assist in BT management and assist in the compilation of specific biomedical data for cancer research.

8.2 Recommendations

The design of DSSs that are useful for clinical environments and cancer research should be carried out as a multidisciplinary development. PR provides the practical mathematical background to obtain inferred models that are fitted to the specific medical problems.

The use of multicentric cases avoids the bias to specific systematic conditions of the clinical centres. This ensures translational results of the CDSSs to centers that are not involved in the design process. The evaluation strategy with subsequently acquired data is a close perspective to the conditions that the predictive models will confront in real clinical environments.

Cooperation with clinical experts for specific problems should increase confidence in CDSSs. The comparative analysis of the results obtained by automatic systems with clinical knowledge is useful for checking the consistency of the databases, detecting possible outliers, and studying diagnosis patterns.

The design of a distributed Decision-Support System, which is integrated in clinical routine workflow may be the best strategy to reduce the gap between ICT and the clinical community. Access to a CDSS that includes the multidisciplinary expertise of clinicians and biomedical engineers should help convince more clinical institutions to share their biomedical data related to cancer research. An increased data set would increase the robustness of the systems and facilitate new discoveries in specific biomedical domains.

Future work in the application of PR for biomedical data in computer-assisted cancer research could include the following:

- To keep improving BT classification using PR-approaches we propose several recommendations. The Glioblastoma vs. Metastasis discrimination by means of MRS

is difficult with the use of SV spectroscopy. Other approaches, such as Multi-voxel (MV) spectroscopy, or the acquisition of an additional adjacent voxel to the brain mass, should provide the necessary information for differentiating these two types of tumours. Additionally, molecular evidence may be used for sub-dividing glioblastoma tumors into molecular subgroups. Also, the patterns observed in in-vivo MRS within each diagnosis are usually heterogeneous. Unsupervised methodologies may be interesting for separating these patterns and so enabling analysis of the cancerous tissues. The extension of the classifiers to the diagnosis of childhood tumors could be carried out in the following three approaches: (a) evaluation of the compatibility of the classifiers for adults with the patterns shown in children; (b) preparation of specific predictive models; and (c) inclusion of the age factor in the predictive models. Finally, when designing CDSSs for BT based on MRS, it would be interesting to expand the types of cases that can be predicted. To achieve this, one-class approaches may be useful [1].

- Functional data analysis approaches, such as that proposed in [2], can be applied to model MRS and temporal series of MRS. Also, the introduction of 3T MR machines in clinical centers makes it necessary to test the compatibility of CDSSs based on 1.5T.

Additionally, the use of SV-based classifiers for MRSI data would increase the amount of data available for evaluating the robustness of nosologic images for brain tumour probability maps [3].

- The probabilistic mixture model for binned and truncated data, featuring univariate mixture densities of means that are relative to a global shift, can be extended to a mixture of patterns that describe each type of tumour.
- The involvement of more international clinics in the collection of multicenter datasets may increase the number of cases, especially for tumours with low epidemiological prevalence (lymphomas, PNET, oligodendrogliomas, oligoastrocytomas, etc). A practical solution to generate the clinical datasets should follow the recommendations of standards (such as ISO 13606 or HL7 v3) for the representation and communication of electronic health records, in order to ensure the semantic interoperability between datasources and CDSSs.
- Clinical evaluation of the CDSS should be carried out to measure the added value for STT and BT diagnosis. The characterization of the performance of classifiers with respect to the quality control levels of the MRS or microarrays could be of interest to evaluate the influence of the CDSS in clinical environments. Finally, a dynamic evaluation of predictive models in the continuous use of classifiers could be useful for auditing CDSS deployed in clinical environments.

Bibliography

- [1] Tax D (2001) One-class classification; Concept-learning in the absence of counter-examples. Technical report, Delft University of Technology <http://www-ict.et.tudelft.nl/~davidt/papers/thesis.pdf>
- [2] Ramsay J, Silverman BW (2005) Functional Data Analysis (Springer Series in Statistics). Springer
- [3] De Edelenyi FS, Rubin C, Esteve F, Grand S, Decorps M, Lefournier V, Le Bas JF, Remy C (2000) A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. *Nat Med* 6(11): 1287–1289

Appendix A

Derivation of the EM steps and implementation aspects of the Mixture model for binned and truncated data

A.1 Maximum Likelihood Estimation from Binned and Truncated Data with Independent Univariate Gaussians

The complete log-likelihood function

$$L_c(\Theta; \mathbf{N}, \mathbf{X}, \mathbf{Z}) = \log p_{\Theta}(\mathbf{N}, \mathbf{X}, \mathbf{Z}) \tag{A.1}$$

$$= \sum_{b=1}^B \sum_{n=1}^{N_b} \log \prod_{c=1}^C (\pi_c p_{\Theta'}(\mathbf{x}_{bn} | c))^{z_{bnc}} \tag{A.2}$$

$$= \sum_{b=1}^B \sum_{n=1}^{N_b} \sum_{c=1}^C (z_{bnc} \log \pi_c p_{\Theta'}(\mathbf{x}_{bn} | c)), \tag{A.3}$$

$$\tag{A.4}$$

allows the application of the EM steps iteratively.

E-Step In the iteration $k+1$, the so-called function $Q(\Theta | \Theta^{(k)})$, calculates the expected value of the complete log-likelihood A.1, conditional to the parameters $\Theta^{(k)}$ of the previous iteration and the observed data \mathbf{N}' . The E step computes a lower bound of the incomplete log-likelihood function 4.12.

$$Q(\Theta | \Theta^{(k)}) = E\left[\sum_{b=1}^B \sum_{n=1}^{N_b} \sum_{c=1}^C z_{bnc} \log(\pi_c p_{\Theta'}(x_{bn} | z_{bnc})) | \mathbf{N}', \Theta^{(k)}\right] \quad (\text{A.5})$$

$$= E\left[\sum_{c=1}^C \sum_{b=1}^B \sum_{n=1}^{N_b} z_{bnc} \log(\pi_c p_{\Theta'}(x_{bn} | z_{bnc})) | \mathbf{N}', \Theta^{(k)}\right], \quad (\text{A.6})$$

$$(\text{A.7})$$

from where $z_{bnc}^{(k)}$ is the expected value of z_{bnc} in the (k) -iteration

$$z_{bnc}^{(k)} = E[z_{bnc} | x_{bn}; \Theta^{(k)}] \quad (\text{A.8})$$

$$= \sum_{z_{bnc}} z_{bnc} p_{\Theta^{(k)}}(z_{bnc} | x_{bn}) \quad (\text{A.9})$$

$$= p_{\Theta^{(k)}}(z_{bnc} = 1 | x_{bn}) \quad (\text{A.10})$$

$$= \frac{\pi_c p_{\Theta^{(k)}}(x_{bn} | z_{bnc} = 1)}{p_{\Theta^{(k)}}(x_{bn})} \quad (\text{A.11})$$

$$= \frac{\pi_c p_{\Theta^{(k)}}(x_{bn} | z_{bnc} = 1)}{\sum_{c'=1}^C \pi_{c'} p_{\Theta^{(k)}}(x_{bn} | z_{bnc'} = 1)}. \quad (\text{A.12})$$

The sum of $z_{cbn}^{(k)}$ for every n of the bin b can be expressed as

$$\sum_{n=1}^{N_b} z_{bnc}^{(k)} = \sum_{n=1}^{N_b} p_{\Theta^{(k)}}(z_{bnc} = 1 | x_{bn}) \quad (\text{A.13})$$

$$= N_b^{(k)} E_b[p_{\Theta^{(k)}}(z_{bnc} = 1 | x_{bn}) | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.14})$$

$$= N_b^{(k)} E_b[z_c(\mathbf{x}_b)^{(k)} | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.15})$$

The expectation in (A.15) is with respect to a sample x_b from bin \mathcal{X}_b ; i.e., with respect to the truncated density of the bin \mathcal{X}_b

$$p_{\Theta^{(k)}}^{\text{trunc}}(x_b) = \frac{p_{\Theta^{(k)}}(x_b)}{p^{(b)}(x_b)} \quad (\text{A.16})$$

and involves the posterior probability for x_b to belong to component c of the mixture, given a current parameter estimate $\Theta^{(k)} : z_c^{(k)}(x_b)$.

$$Q(\Theta | \Theta^{(k)}) = \sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c(\mathbf{x}_b)^{(k)} \log(\pi_c p_{\Theta^{(k)}}(\mathbf{x}_b | c)) | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.17})$$

where

$$N_b^{(k)} = \begin{cases} N_b & \text{if } b \leq B' \\ N' \frac{p^{(b)}(x_b)}{\sum_{b'=1}^{B'} p^{(b')}(x_b)} & \text{otherwise,} \end{cases} \quad (\text{A.18})$$

and

$$p(b)^{(k)} = \int_{\mathcal{X}_b} dx p_{\Theta^{(k)}}(x) \quad (\text{A.19})$$

M-Step. In the M-Step, the parameters $\Theta^{(k+1)}$ are updated for maximizing $Q(\Theta \mid \Theta^{(k)})$, s.t. $\sum_{c=1}^C \pi_c = 1$.

First, we define the Lagrangian as

$$\mathcal{L}(\Theta, \lambda) = Q(\Theta \mid \Theta^{(k)}) - \lambda \left(\sum_{c=1}^C \pi_c - 1 \right) \quad (\text{A.20})$$

$$= \sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) \log(\pi_c p_{\Theta^{(k)}}(\mathbf{x}_b \mid c) \mid N', \Theta^{(k)})] \quad (\text{A.21})$$

$$- \lambda \left(\sum_{c=1}^C \pi_c - 1 \right), \quad (\text{A.22})$$

M-Step. Updates of π_c in $(k+1)$. To calculate the updates of $\pi_c^{(k+1)}$ in each M-Step, we derivate the Lagrangian \mathcal{L} with respect to π_c and λ .

$$\left(\frac{\partial \mathcal{L}(\Theta, \lambda)}{\partial \pi_c} \right)_{c=(1, \dots, C)} = \mathbf{0} \text{ y } \frac{\partial \mathcal{L}(\Theta, \lambda)}{\partial \lambda} = 0. \quad (\text{A.23})$$

Hence,

$$\frac{\partial \mathcal{L}(\Theta, \lambda)}{\partial \pi_c} = \frac{\partial}{\partial \pi_c} \sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) \log \pi_c \mid N', \Theta^{(k)}] - \lambda \left(\sum_{c=1}^C \pi_c - 1 \right) \quad (\text{A.24})$$

$$= \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) \frac{\partial}{\partial \pi_c} \log \pi_c \mid N', \Theta^{(k)}] - \lambda \frac{\partial}{\partial \pi_c} \pi_c \quad (\text{A.25})$$

$$= \sum_{b=1}^B N_b^{(k)} E_b \left[\frac{z_c^{(k)}(x_b)}{\pi_c} \mid N', \Theta^{(k)} \right] - \lambda \quad (\text{A.26})$$

\Downarrow

$$\lambda = \sum_{b=1}^B N_b^{(k)} E_b \left[\frac{z_c^{(k)}(x_b)}{\pi_c} \mid N', \Theta^{(k)} \right] \quad (\text{A.27})$$

$$\Downarrow \quad (\text{A.28})$$

$$\pi_c = \frac{\sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) \mid N', \Theta^{(k)}]}{\lambda} \quad (\text{A.29})$$

$$(\text{A.30})$$

and

$$\begin{aligned}
 \frac{\partial \mathcal{L}(\Theta, \lambda)}{\partial \lambda} &= \sum_{c=1}^C \pi_c - 1 \\
 &\Downarrow \\
 \sum_{c=1}^C \pi_c &= 1 \\
 \text{and by the substitution of } \pi_c \text{ by A.29} &\Downarrow \\
 \sum_{c=1}^C \frac{\sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) | N', \Theta^{(k)}]}{\lambda} &= 1 \\
 &\Downarrow \\
 \sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) | N', \Theta^{(k)}] &= \lambda \\
 \text{and the substitution of } \lambda \text{ in A.29} &\Downarrow \\
 \pi_c &= \frac{\sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) | N', \Theta^{(k)}]}{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) | N', \Theta^{(k)}]}, \\
 \text{as } \sum_{c=1}^C E_b[z_c^{(k)}(x_b) | N', \Theta^{(k)}] &= 1 \Downarrow \\
 \pi_c &= \frac{\sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) | N', \Theta^{(k)}]}{\sum_{b=1}^B N_b^{(k)}}
 \end{aligned}$$

Hence, the update of $\pi_c^{(k+1)}$ is:

$$\pi_c^{(k+1)} = \frac{\sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}]}{\sum_{b=1}^B N_b^{(k)}}, \forall c : 1, \dots, C \quad (\text{A.31})$$

M-Step. Updates of the rest of the parameters. Generically, in the M-Step, we maximize the Lagrangian with respect every parameter ξ obtaining the roots of

$$\begin{aligned}
 \frac{\partial \mathcal{L}(\Theta, \lambda)}{\partial \xi} &= \frac{\partial \sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) \log(\pi_c p_{\Theta^{(k)}}(\mathbf{x}_b | c) | N', \Theta^{(k)})]}{\partial \xi} \\
 &\quad - \frac{\lambda(\sum_{c=1}^C \pi_c - 1)}{\partial \xi}
 \end{aligned}$$

M-Step. Mixture of independent univariate gaussian distributions Given a (multinomial) mixture model of C independent univariate gaussian distributions $(N(\mu_c, \sigma_c))$, the samples x follow the probability density function:

$$p(x) = \sum_{c=1}^C \pi_c p(x | c; \mu_c, \sigma_c)$$

where

$$p(x | c; \mu_c, \sigma_c) = \frac{1}{(2\pi)^{1/2} \sigma_c^{(k)}} e^{-\frac{1}{2} \left[\frac{(x - \mu_c)}{\sigma_c} \right]^2}$$

By substitution of A.1 in A.22.

$$\begin{aligned} \mathcal{L}(\Theta, \lambda) &= \\ &= \sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) \log\left(\pi_c \frac{1}{(2\pi)^{1/2} \sigma_c^{(k)}} e^{-\frac{1}{2} \left[\frac{(x_b - \mu_c)}{\sigma_c} \right]^2} \right) | \mathbf{N}', \Theta^{(k)}] \\ &\quad - \lambda \left(\sum_{c=1}^C \pi_c - 1 \right) \end{aligned}$$

The roots of the partial derivative of \mathcal{L} with respect to μ_c give the update of $\mu_c^{(k+1)}$.

$$\frac{\partial \mathcal{L}(\Theta, \lambda)}{\partial \mu_c} = 0, c = 1..C;$$

hence,

$$\frac{\partial \mathcal{L}(\Theta, \lambda)}{\partial \mu_c} = \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{\partial}{\partial \mu_c} \log\left(\frac{1}{(2\pi)^{1/2} \sigma_c} e^{-\frac{1}{2} \left[\frac{(x_b - \mu_c)}{\sigma_c} \right]^2} \right) | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.32})$$

$$= \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{-1}{2(\sigma_c)^2} 2(x_b - \mu_c)(-1) | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.33})$$

$$= \sum_{b=1}^B N_b E_b[z_c^{(k)}(\mathbf{x}_b) \frac{1}{(\sigma_c)^2} (x_b - \mu_c) | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.34})$$

$$= 0 \quad (\text{A.35})$$

$$(\text{A.36})$$

↓

$$\sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{1}{(\sigma_c)^2} \mu_c | \mathbf{N}', \Theta^{(k)}] = \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{1}{(\sigma_c)^2} x_b | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.37})$$

$$\mu_c \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) | \mathbf{N}', \Theta^{(k)}] = \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) x_b | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.38})$$

$$\mu_c = \frac{\sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) x_b | \mathbf{N}', \Theta^{(k)}]}{\sum_{b=1}^B N_b E_b[z_c^{(k)}(\mathbf{x}_b) | \mathbf{N}', \Theta^{(k)}]} \quad (\text{A.39})$$

hence,

$$\mu_c^{(k+1)} = \frac{\sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b)x_b \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}]}{C_c(\boldsymbol{\Theta}^{(k)})}, \forall c : 1, \dots, C; \quad (\text{A.40})$$

where

$$C_c^{(k)}(\boldsymbol{\Theta}^{(k)}) = \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}], \forall c : 1, \dots, C \quad (\text{A.41})$$

To calculate the update of $(\sigma_c^{(k+1)})^2$, we will obtain the roots of the derivative of \mathcal{L} with respecto to σ_c .

$$\frac{\mathcal{L}(\boldsymbol{\Theta}, \lambda)}{\sigma_c}.$$

Hence,

$$\frac{\partial \mathcal{L}(\boldsymbol{\Theta}, \lambda)}{\partial \sigma_c} = \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{\partial}{\partial \sigma_c} \log\left(\frac{1}{(2\pi)^{1/2} \sigma_c} e^{-\frac{1}{2} \left[\frac{x_b - \mu_c}{\sigma_c}\right]^2} \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}\right)] \quad (\text{A.42})$$

$$= \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{\partial}{\partial \sigma_c} \left[-\log(\sigma_c) - \frac{1}{2} \left[\frac{x_b - \mu_c}{\sigma_c}\right]^2 \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}\right)] \quad (\text{A.43})$$

$$= \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \left[-\frac{1}{\sigma_c} - \frac{1}{2}(-2) \frac{(x_b - \mu_c)^2}{\sigma_c^3}\right] \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}] \quad (\text{A.44})$$

$$= \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \left[-\frac{1}{\sigma_c} + \frac{(x_b - \mu_c)^2}{\sigma_c^3}\right] \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}] \quad (\text{A.45})$$

$$= 0 \quad (\text{A.46})$$

↓

$$\frac{1}{\sigma_c^3} \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b)(x_b - \mu_c)^2 \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}] = \frac{1}{\sigma_c} \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}] \quad (\text{A.47})$$

$$\frac{1}{\sigma_c^2} \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b)(x_b - \mu_c)^2 \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}] = \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}] \quad (\text{A.48})$$

$$(\sigma_c)^2 = \frac{\sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b)(x_b - \mu_c)^2 \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}]}{\sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}]} \quad (\text{A.49})$$

hence,

$$(\sigma_c^{(k+1)})^2 = \frac{\sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b)(x_b - \mu_c^{(k+1)})^2 \mid \mathbf{N}', \boldsymbol{\Theta}^{(k)}]}{C_c(\boldsymbol{\Theta}^{(k)})}, \forall c : 1, \dots, C \quad (\text{A.50})$$

Calculus of the expected values $E_b[g(x_b) | \mathbf{N}', \Theta^{(k)}]$. Let the probability function of a distribution conditioned to c ,

$$P(x | c) = \int_{-\text{inf}}^x p(x | c)dx,$$

and let the probability function of a distribution conditional to c in an interval from x_ϵ be defined as $[p_{b-1}, p_b]$

$$P(x_\epsilon | c) = \int_{p_{b-1}}^{p_b} p(x | c)dx = P(p_b | c) - P(p_{b-1} | c)$$

The unconditioned probability function of x is obtained by marginalization,

$$P(x) = \sum_{c=1}^C \pi_c P(x | c)dx,$$

and, in the interval x_ϵ ,

$$p(b) = P(p_b) - P(p_{b-1}).$$

The probability function of x conditioned to belong to the interval x_ϵ is

$$P^{trunc}(x_b) = P(x_0 \leq x_b | p_{b-1} < x_0 < p_b) = \frac{\int_{p_{b-1}}^{x_b} p(x_b)dx_b}{p(b)}, \quad (\text{A.51})$$

and, hence, the probability density function truncated to the interval x_ϵ is

$$p^{trunc}(x_b) = \frac{dP^{trunc}(x_b)}{dx_b} = \frac{1}{p(b)}p(x_b) \quad (\text{A.52})$$

The expected value of a function $g(x)$, when the variable x that follows a distribution with parameters $\Theta^{(k)}$, when x is conditioned to be in the interval x_ϵ is defined as:

$$E_b[g(x_b) | \mathbf{N}', \Theta^{(k)}] = \frac{1}{p_{\Theta^{(k)}}(b)} \int_{x_\epsilon} p_{\Theta^{(k)}}(x)g(x)dx$$

Let define each component of the mixture model as a independent univariate gaussian with the probability density function A.1. The calculus of the updates in the iteration k of the parameters π_c , μ_c and σ_c^2 are performed using the expressions A.31, A.40 and A.50. For solving these expressions, it is necessary to solve some expected value E_b conditioning x to be in the interval x_ϵ .

For simplicity in the notation, we define

$$A_{e,cb}^{(k)} = E_b[z_c^{(k)}(x_b)x_b^e | \mathbf{N}', \Theta^{(k)}], \forall e = (0, 1)$$

and

$$A_{2,cb}^{(k)} = E_b[z_c^{(k)}(x_b)(x_b - \mu_c^{(k)})^2 | \mathbf{N}', \Theta^{(k)}].$$

$A_{0,cb}^{(k)}$ is needed for solving $\pi_c^{(k+1)}$,

$$A_{0,cb}^{(k)} = E_b[z_c^{(k)}(x_b) \mid \mathbf{N}', \Theta^{(k)}] \quad (\text{A.53})$$

$$= \frac{1}{\int_{x_s} p_{\Theta^{(k)}}(x_b) dx_b} \int_{x_s} p_{\Theta^{(k)}}(x_b) \frac{\pi_c p_{\Theta^{(k)}}(x_b \mid c)}{p_{\Theta^{(k)}}(x_b)} dx_b \quad (\text{A.54})$$

$$= \frac{\pi_c \int_{x_s} p_{\Theta^{(k)}}(x_b \mid c) dx_{bs}}{\int_{x_s} p_{\Theta^{(k)}}(x_b) dx_{bs}} \quad (\text{A.55})$$

$$= \frac{\pi_c (P^{(k)}(p_b \mid c) - P^{(k)}(p_{b-1} \mid c))}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} \quad (\text{A.56})$$

$$(\text{A.57})$$

$A_{1,cb}^{(k)}$ is needed for solving $\mu_c^{(k+1)}$,

$$A_{1,cb}^{(k)} = E_b[z_c^{(k)}(x_b)x_b \mid \mathbf{N}', \Theta^{(k)}] \quad (\text{A.58})$$

$$= \frac{1}{\int_{x_s} p_{\Theta^{(k)}}(x_b) dx_b} \int_{x_s} p_{\Theta^{(k)}}(x_b) \frac{\pi_c p_{\Theta^{(k)}}(x_b \mid c)}{p_{\Theta^{(k)}}(x_b)} x_b dx_b \quad (\text{A.59})$$

$$= \frac{\pi_c \int_{x_s} p_{\Theta^{(k)}}(x_b \mid c) \int_{x_s} p_{\Theta^{(k)}}(x_b \mid c) dx_b x_b dx_b}{\int_{x_s} p_{\Theta^{(k)}}(x_b) dx_b} \quad (\text{A.60})$$

$$= \frac{\pi_c (P^{(k)}(p_b \mid c) - P^{(k)}(p_{b-1} \mid c)) \int_{x_s} x_b p_{\Theta^{(k)}}(x_b \mid c) dx_b}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} \quad (\text{A.61})$$

$$= \frac{\pi_c (P^{(k)}(p_b \mid c) - P^{(k)}(p_{b-1} \mid c))}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} \quad (\text{A.62})$$

$$\frac{(\mu_c^{(k)} + \frac{(\sigma_c^{(k)})^2}{(P^{(k)}(p_b \mid c) - P^{(k)}(p_{b-1} \mid c))} (p_{\Theta^{(k)}}(p_b \mid c) - p_{\Theta^{(k)}}(p_{b-1} \mid c)))}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} \quad (\text{A.63})$$

$$= \frac{\pi_c (\mu_c^{(k)} (P^{(k)}(p_b \mid c) - P^{(k)}(p_{b-1} \mid c)) + (\sigma_c^{(k)})^2 (p_{\Theta^{(k)}}(p_b \mid c) - p_{\Theta^{(k)}}(p_{b-1} \mid c)))}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} \quad (\text{A.64})$$

$$(\text{A.65})$$

$A_{2,cb}^{(k)}$ is needed for solving $(\sigma_c^{(k+1)})^2$,

$$A_{2,cb}^{(k)} = E_b[z_c^{(k)}(x_b)(x_b - \mu_c^{(k)})^2 \mid \mathbf{N}', \Theta^{(k)}] \quad (\text{A.66})$$

$$= \frac{\pi_c ((\sigma_c^{(k)})^2 [H_{0,cb}^{(k)} + (2\mu_c^{(k+1)} - \mu_c^{(k)}) H_{1,cb}^{(k)} - H_{2,cb}^{(k)}] + (\mu_c^{(k+1)} - \mu_c^{(k)})^2 H_{0,cb}^{(k)})}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}, \quad (\text{A.67})$$

$$(\text{A.68})$$

where

$$H_{0,cb}^{(k)} = P^{(k)}(p_b | c) - P^{(k)}(p_{b-1} | c) \quad (\text{A.69})$$

$$H_{1,cb}^{(k)} = p^{(k)}(p_b | c) - p^{(k)}(p_{b-1} | c) \quad (\text{A.70})$$

$$H_{2,cb}^{(k)} = p_b p^{(k)}(p_b | c) - p_{b-1} p^{(k)}(p_{b-1} | c) \quad (\text{A.71})$$

$$(\text{A.72})$$

Summarizing, for each interval $x_b, \forall b = (1, \dots, v)$,

$$A_{e,cb}^{(k)} = \frac{\pi_c^{(k)} G_{e,cb}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}, \forall e = (0, \dots, 2) \quad (\text{A.73})$$

where

$$G_{0,cb}^{(k)} = H_{0,cb}^{(k)} \quad (\text{A.74})$$

$$G_{1,cb}^{(k)} = \mu_c^{(k)} H_{0,cb}^{(k)} - (\sigma_c^{(k)})^2 H_{1,cb}^{(k)} \quad (\text{A.75})$$

$$G_{2,cb}^{(k)} = (\sigma_c^{(k)})^2 [H_{0,cb}^{(k)} + (2\mu_c^{(k+1)} - \mu_c^{(k)}) H_{1,cb}^{(k)} - H_{2,cb}^{(k)}] + (\mu_c^{(k+1)} - \mu_c^{(k)})^2 H_{0,cb}^{(k)}. \quad (\text{A.76})$$

$$(\text{A.77})$$

Efficient implementation of $\pi_c^{(k+1)}, \mu_c^{(k+1)}$ y $(\sigma_c^{(k+1)})^2$. Jones and McLachlan published in 1990 [1] the Fortran implementation of the algorithm *AS 254* for the max-likelihood estimation by binned and truncated data with univariate finite mixture models. The development of $\pi_c^{(k+1)}, \mu_c^{(k+1)}$ y $(\sigma_c^{(k+1)})^2$ by means of A.31, A.40 and A.50 and using A.73 for the expected values ($A_{e,cb}^{(k)}$) conditioning x_b to be in the region x_b :

$$\pi_c^{(k+1)} = \frac{\sum_{b=1}^B N_b^{(k)} A_{0,cb}^{(k)}}{\sum_{b=1}^B N_b^{(k)}} \quad (\text{A.78})$$

$$= \frac{\sum_{b=1}^B N_b^{(k)} \frac{\pi_c^{(k)} P^{(k)}(p_b | c) - P^{(k)}(p_{b-1} | c)}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}}{\sum_{b=1}^B N_b^{(k)}} \quad (\text{A.79})$$

$$= \pi_c^{(k)} \frac{\sum_{b=1}^B N_b^{(k)} \frac{P^{(k)}(p_b | c) - P^{(k)}(p_{b-1} | c)}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}}{\sum_{b=1}^B N_b^{(k)}} \quad (\text{A.80})$$

$$= \pi_c^{(k)} \frac{\sum_{b=1}^B \frac{N_b^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} H_{0,cb}^{(k)}}{\sum_{b=1}^B N_b^{(k)}}, \quad (\text{A.81})$$

$$(\text{A.82})$$

$$\begin{aligned}
 \mu_c^{(k+1)} &= \frac{\sum_{b=1}^B N_b^{(k)} A_{1,cb}^{(k)}}{\sum_{b=1}^B N_b^{(k)} A_{0,cb}^{(k)}} \\
 &= \frac{\sum_{b=1}^B N_b^{(k)} \frac{\pi_c[\mu_c^{(k)}(P^{(k)}(p_b|c) - P^{(k)}(p_{b-1}|c)) - (\sigma_c^{(k)})^2(f^{(k)}(p_b|c) - f^{(k)}(p_{b-1}|c))]}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}}{\sum_{b=1}^B N_b^{(k)} \frac{\pi_c(P^{(k)}(p_b|c) - P^{(k)}(p_{b-1}|c))}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}} \\
 &= \mu_c^{(k)} \frac{\sum_{b=1}^B N_b^{(k)} \frac{P^{(k)}(p_b|c) - P^{(k)}(p_{b-1}|c)}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}}{\sum_{b=1}^B N_b^{(k)} \frac{P^{(k)}(p_b|c) - P^{(k)}(p_{b-1}|c)}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}} \\
 &\quad - (\sigma_c^{(k)})^2 \frac{\sum_{b=1}^B N_b^{(k)} \frac{f^{(k)}(p_b|c) - f^{(k)}(p_{b-1}|c)}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}}{\sum_{b=1}^B N_b^{(k)} \frac{P^{(k)}(p_b|c) - P^{(k)}(p_{b-1}|c)}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}} \\
 &= \mu_c^{(k)} - (\sigma_c^{(k)})^2 \frac{\sum_{b=1}^B \frac{N_b^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} H_{1,cb}^{(k)}}{\sum_{b=1}^B \frac{N_b^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} H_{0,cb}^{(k)}},
 \end{aligned}$$

and a similar expression is obtained for $(\sigma_c^{(k)})^2$.

The final expressions are:

$$\pi_c^{(k+1)} = \pi_c^{(k)} R_{0c}^{(k)} \quad (\text{A.83})$$

$$\mu_c^{(k+1)} = \mu_c^{(k)} - (\sigma_c^{(k)})^2 R_{1c}^{(k)} \quad (\text{A.84})$$

$$(\sigma_c^{(k+1)})^2 = (\sigma_c^{(k)})^2 [1 + (2\mu_c^{(k+1)} - \mu_c^{(k)}) R_{1c}^{(k)} - R_{2c}^{(k)}] + (\mu_c^{(k+1)} - \mu_c^{(k)})^2 \quad (\text{A.85})$$

where

$$R_{0c}^{(k)} = \frac{\sum_{b=1}^B t_b^{(k)} H_{0,cb}^{(k)}}{\sum_{b=1}^B N_b^{(k)}} \quad (\text{A.86})$$

$$R_{1c}^{(k)} = \frac{\sum_{b=1}^B t_b^{(k)} H_{1,cb}^{(k)}}{\sum_{b=1}^B t_b^{(k)} H_{0,cb}^{(k)}} \quad (\text{A.87})$$

$$R_{2c}^{(k)} = \frac{\sum_{b=1}^B t_b^{(k)} H_{2,cb}^{(k)}}{\sum_{b=1}^B t_b^{(k)} H_{0,cb}^{(k)}} \quad (\text{A.88})$$

$$t_b^{(k)} = \frac{N_b^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})} \quad (\text{A.89})$$

A.2 Maximum Likelihood Estimation from Binned and Truncated Data with c Univariate Gaussians relative to a global shift $\mu_c = \mu_0 + \Delta\mu_c, \forall c = 1..C$

Given a (multinomial) mixture model of C univariate gaussian distributions ($N(\mu_0 + \Delta\mu_c, \sigma_c)$), which means μ_c are relative to a global mean μ_0 with a fixed shifting of $\Delta\mu_c$, the samples x follow the probability density function:

$$p(x) = \sum_{c=1}^C \pi_c p(x | c; \mu_0 + \Delta\mu_c, \sigma_c)$$

where

$$p(x | c; \mu_0 + \Delta\mu_c, \sigma_c) = \frac{1}{(2\pi)^{1/2} \sigma_c^{(k)}} e^{-\frac{1}{2} \left[\frac{x - (\mu_0^{(k)} + \Delta\mu_c)}{\sigma_c^{(k)}} \right]^2}.$$

M-Step. Substituting A.2 in A.22 we obtain the Lagrangian for the new model

$$\mathcal{L}_o(\theta, \lambda) = \tag{A.90}$$

$$= \sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b[z_c^{(k)}(x_b) \log(\pi_{cbs} \frac{1}{(2\pi)^{1/2} \sigma_c^{(k)}} e^{-\frac{1}{2} \left[\frac{x_b - (\mu_0^{(k)} + \Delta\mu_c)}{\sigma_c^{(k)}} \right]^2}) | \mathbf{N}', \Theta^{(k)}] \tag{A.91}$$

$$- \lambda \left(\sum_{c=1}^C \pi_c - 1 \right) \tag{A.92}$$

For this Lagrangian, the expression to maximize π_c in the iteration $(k+1)$ is similar to the previous case A.31.

In order to calculate the new estimation of $\mu_0^{(k+1)}$, we should obtain the maximum of

$$\frac{\partial \mathcal{L}_o(\theta, \lambda)}{\partial \mu_0} = 0,$$

hence,

$$\frac{\partial \mathcal{L}_o(\theta, \lambda)}{\partial \mu_0} = \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{\partial}{\partial \mu_0} \log\left(\frac{1}{(2\pi)^{1/2} \sigma_c} e^{-\frac{1}{2} \left[\frac{x - (\mu_0 + \Delta\mu_c)}{\sigma_c} \right]^2} \right) | \mathbf{N}', \Theta^{(k)}] \tag{A.93}$$

$$= \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{-1}{2(\sigma_c)^2} 2(x_b - \mu_0 - \Delta\mu_c)(-1) | \mathbf{N}', \Theta^{(k)}] \tag{A.94}$$

$$= \sum_{b=1}^B N_b E_b[z_c^{(k)}(x_b) \frac{1}{(\sigma_c)^2} (x_b - \mu_0 - \Delta\mu_c) | \mathbf{N}', \Theta^{(k)}] \tag{A.95}$$

$$= 0 \tag{A.96}$$

$$\tag{A.97}$$

\Downarrow

$$\sum_{c=1}^C \sum_{b=1}^B N_b E_b [p(c | x_b) \frac{\mu_0}{(\sigma_c)^2} | \mathbf{N}', \Theta^{(k)}] = \sum_{c=1}^C \sum_{b=1}^B N_b E_b [p(c | x_b) \frac{x_b - \Delta\mu_c}{(\sigma_c)^2} | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.98})$$

$$\mu_0 \sum_{c=1}^C \sum_{b=1}^B N_b E_b [p(c | x_b) | \mathbf{N}', \Theta^{(k)}] = \sum_{c=1}^C \sum_{b=1}^B N_b E_b [p(c | x_b) (x_b - \Delta\mu_c) | \mathbf{N}', \Theta^{(k)}] \quad (\text{A.99})$$

$$\mu_0 = \frac{\sum_{c=1}^C \sum_{b=1}^B N_b E_b [p(c | x_b) (x_b - \Delta\mu_c) | \mathbf{N}', \Theta^{(k)}]}{\sum_{c=1}^C \sum_{b=1}^B N_b E_b [p(c | x_b) | \mathbf{N}', \Theta^{(k)}]} \quad (\text{A.100})$$

Hence, for iteration $k + 1$,

$$\mu_0^{(k+1)} = \frac{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b [p(c | x_b) (x_b - \Delta\mu_c) | \mathbf{N}', \Theta^{(k)}]}{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} E_b [p(c | x_b) | \mathbf{N}', \Theta^{(k)}]} \quad (\text{A.101})$$

To obtain the updating expression for $(\sigma_c^{(k+1)})^2$, we use $\mu_c = \mu_0 + \Delta\mu_c, \forall c = 1..C$ in A.50.

$$(\sigma_c^{(k+1)})^2 = \frac{\sum_{b=1}^B N_b^{(k)} E_b [p(c | x_b) (x_b - (\mu_0^{(k+1)} + \Delta\mu_c))^2 | x_b \in \mathbf{N}', \Theta^{(k)}]}{C_c(\Theta^{(k)})} \quad (\text{A.102})$$

Expanding the conditioned expected values in A.31, A.101 and A.102, the final expressions for the M-Step implementation are:

$$\pi_c^{(k+1)} = \pi_c^{(k)} R_{0c}^{(k)} \quad (\text{A.103})$$

$$\mu_0^{(k+1)} = \frac{\sum_{c=1}^C \pi_c^{(k)} (\mu_c^{(k)} - \Delta\mu_c) \sum_{b=1}^B N_b^{(k)} \frac{H_{0,cr}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}}{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} \frac{\pi_c^{(k)} H_{0,cr}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}} \quad (\text{A.104})$$

$$= \frac{\sum_{c=1}^C \pi_c^{(k)} (\sigma_c^{(k)})^2 \sum_{b=1}^B N_b^{(k)} \frac{H_{1,cr}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}}{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} \frac{\pi_c^{(k)} H_{0,cr}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}} \quad (\text{A.105})$$

$$= \frac{\sum_{c=1}^C \pi_c^{(k)} ((\mu_c^{(k)} - \Delta\mu_c) \sum_{b=1}^B N_b^{(k)} \frac{H_{0,cr}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})})}{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} \frac{\pi_c^{(k)} H_{0,cr}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}} \quad (\text{A.106})$$

$$= \frac{(\sigma_c^{(k)})^2 \sum_{b=1}^B N_b^{(k)} \frac{H_{1,cr}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}}{\sum_{c=1}^C \sum_{b=1}^B N_b^{(k)} \frac{\pi_c^{(k)} H_{0,cr}^{(k)}}{P^{(k)}(p_b) - P^{(k)}(p_{b-1})}} \quad (\text{A.107})$$

$$(\sigma_c^{(k+1)})^2 = (\sigma_c^{(k)})^2 [1 + (2\mu_0^{(k+1)} - \mu_0^{(k)} + \Delta\mu_c) R_{1c}^{(k)} - R_{2c}^{(k)}] + (\mu_0^{(k+1)} - \mu_0^{(k)})^2 \quad (\text{A.108})$$

$$(\text{A.109})$$

Bibliography

- [1] Jones PN, McLachlan GJ (1990) Statistical Algorithms: Algorithm AS 254: Maximum Likelihood Estimation from Grouped and Truncated Data with Finite Normal Mixture Models. *Applied Statistics* 39(2): 273–282

Appendix B

Tables of the expression differences between GBM and mMEN

Table B.1: Genes with highest discriminant capacity. We show the fifty most selected genes across the 200 iterations of the SW resampling approach over the training dataset to estimate the prediction accuracy. Interestingly, the two genes harbouring the highest GBM/mMEN gene-expression ratios are the most selected genes in the training.

Gene symbol	Accession number	Gene description	GBM/mMEN ratio	Fold selected
GFAP	AA069414	Glial fibrillary acidic protein	413	16
PTPRZ1	AA476460	Protein tyrosine phosphatase, receptor-type, Z polypeptide 1	356	9
GPM6B	AA284329	Glycoprotein M6B	133	9
PRELP	AA131664	Proline/arginine-rich end leucine-rich repeat protein	0,042	9
FABP7	W72051	Fatty acid binding protein 7, brain	220	7
EGFL3	AA975413	EGF-like-domain, multiple 3	0,299	7
PDE4B	AA453293	Phosphodiesterase 4B, cAMP-specific (phosphodiesterase E4 duncce homolog, Drosophila)	26	5
OMD	N32201	Osteomodulin	0,011	5
LAPTM4A	AA398233	Lysosomal-associated protein transmembrane 4 alpha	0,346	5
USP25	AA479313	Ubiquitin specific peptidase 25	0,11	5
NFATC3	AA293819	Nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3	0,399	4
CTGF	AA598794	Connective tissue growth factor	0,106	4
PIGT	H83225	Phosphatidylinositol glycan, class T	0,508	4
FLJ39155	R08141	Hypothetical protein FLJ39155	0,019	4
DSP	H90899	Desmoplakin	0,004	4
GAS1	AA025819	Growth arrest-specific 1	0,084	4
PLK1	AA629262	Polo-like kinase 1 (Drosophila)	4,3	4
NEK6	AA463188	NIMA (never in mitosis gene a)-related kinase 6	3,9	4
TNXB	T58430	Similar to tenascin XB isoform 1; tenascin XB1; tenascin XB2; hexabrachion-like [Pan troglodytes]	0,118	4
LHX2	AA018276	LIM homeobox 2	36	4
MGC21621	W52061	MAS-related GPR, member F	0,041	4
PDGFD	A1005125	Platelet derived growth factor D	0,034	4
IL27RA	A1088984	Interleukin 27 receptor alpha	2,1	4
-	A1249137	Transcribed locus	0,205	4
CCND1	R81200	Cyclin D1	0,109	4
NCAM2	AI306467	Neural cell adhesion molecule 2	6,2	4
SH3GL3	A1359676	SH3-domain GRB2-like 3	0,079	4
MT2A	BF131311	Metallothionein 2A	11	4
NUDT1	AA443998	Nudix (nucleoside diphosphate linked moiety X)-type motif 1	2,7	4
RARRES2	AA481944	Retinoic acid receptor responder (tazarotene induced) 2	0,196	3
SMARCD3	AA035796	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3	2,1	3
HYAL1	AA464791	Hyaluronoglucosaminidase 1	0,159	3
CDK2AP1	R78607	CDK2-associated protein 1	2,0	3
CTNND2	H04985	Catenin (cadherin-associated protein), delta 2 (neural plakophilin-related arm-repeat protein)	44	3
CYB5	R92281	Cytochrome b5 type A (microsomal)	0,333	3
CA2	H23187	Carbonic anhydrase II	16	3
OAT	AA446819	Ornithine aminotransferase (gyrate atrophy)	0,383	3
GPM6A	AA448033	Glycoprotein M6A	90	3
HSPC195	R63735	CXXC finger 5	4,2	3
ZMYM6	W81504	Zinc finger, MYM-type 6	0,567	3
PPARGC1A	N89673	Peroxisome proliferative activated receptor, gamma, coactivator 1, alpha	0,257	3
TEK	H02848	TEK tyrosine kinase, endothelial (venous malformations, multiple cutaneous and mucosal)	0,14	3
APM2	AA478298	Chromosome 10 open reading frame 116	0,034	3
FGL2	H56349	Fibrinogen-like 2	0,022	3
CDH2	W49619	Cadherin 2, type 1, N-cadherin (neuronal)	20	3
CANPL1	H15456	Calpain 1, (mu/I) large subunit	0,532	3
GPR17	R44664	G protein-coupled receptor 17	45	3
LOC119504	AA004832	Chromosome 10 open reading frame 104	0,504	3
-	W52340	-	0,355	3
DHRS3	AA171606	Dehydrogenase/reductase (SDR family) member 3	0,217	3

Table B.2: Functional analysis of genes with q-value lower than 0.02. The eleven functional clusters arising from the DAVID tool are depicted. The mentioned tool enables a stringency range for a set of genes to be considered a differentially expressed functional group. The depicted table was computed using the highest stringency and selecting those groups with p-value lower than 0.05. (1/3).

Functional Group 1		P-value: 0.000047	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
CDH3	AA425556	CADHERIN 3, TYPE 1, P-CADHERIN (PLACENTAL)	0,114
DSG2	W37448	DESMOGLEIN 2	0,041
PCDH17	AA669075	PROTOCOLADHERIN 17	8,365
CDH2	W49619	CADHERIN 2, TYPE 1, N-CADHERIN (NEURONAL)	19,699
PCDH1	R77512	PROTOCOLADHERIN 1 (CADHERIN-LIKE 1)	5,324
CDH10	R14164	CADHERIN 10, TYPE 2 (T2-CADHERIN)	4,273
PCDH9	R38168	PROTOCOLADHERIN 9	16,891
CDH5	H02884	CADHERIN 5, TYPE 2, VE-CADHERIN (VASCULAR EPITHELIAL)	0,2
CDH1	AI671174	CADHERIN 1, TYPE 1, E-CADHERIN (EPITHELIAL)	0,052
CDH11	AA136983	CADHERIN 11, TYPE 2, OB-CADHERIN (OSTEOBLAST)	0,219
Functional Group 2		P-value: 0.000080	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
FMOD	AA486471	FIBROMODULIN	0,135
PRELP	AA131664	PROLINE/ARGININE-RICH END LEUCINE-RICH REPEAT PROTEIN	0,042
OMD	N32201	OSTEOMODULIN	0,011
BGN	BE262957	BIGLYCAN	0,409
OGN	AA045327	OSTEOGLYCIN (OSTEOINDUCTIVE FACTOR, MIMECAN)	0,005
Functional Group 3		P-value: 0.000090	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
DSCAM	N64532	DOWN SYNDROME CELL ADHESION MOLECULE	9,673
GHR	AA775738	GROWTH HORMONE RECEPTOR	0,34
EST	BM674517	INTERLEUKIN 6 SIGNAL TRANSDUCER (GPI30, ONCOSTATIN M RECEPTOR)	0,517
IFNGR1	BE973918	INTERFERON GAMMA RECEPTOR 1	0,212
LEPR	H51066	LEPTIN RECEPTOR	0,345
Functional Group 4		P-value: 0.00014	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
AGTRL1	R58969	ANGIOTENSIN II RECEPTOR-LIKE 1	20,807
GPM6A	AA448033	GLYCOPROTEIN M6A	89,864
GPR4	AI492409	G PROTEIN-COUPLED RECEPTOR 4	0,152
TM4SF7	AA100696	TETRASPANIN 4	0,127
SDC2	H64346	SYNDECAN 2 (HEPARAN SULFATE PROTEOGLYCAN 1, CELL SURFACE-ASSOCIATED, FIBROGLYCAN)	0,042
TM4SF13	W86202	TETRASPANIN 13	13,86
Functional Group 5		P-value: 0.00017	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
NCAM2	AI306467	NEURAL CELL ADHESION MOLECULE 2	6,234
DSCAM	N64532	DOWN SYNDROME CELL ADHESION MOLECULE	9,673
ALCAM	R13558	ACTIVATED LEUKOCYTE CELL ADHESION MOLECULE	0,092
JAM2	AA410345	JUNCTIONAL ADHESION MOLECULE 2	0,094
EST	H73479	JUNCTIONAL ADHESION MOLECULE 3	0,065

Table B.3: Functional analysis of genes with q-value lower than 0.02. (2/3).

Functional Group 6		P-value: 0.00027	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
TUBB4	BX100915	MELANOCORTIN 1 RECEPTOR (ALPHA MELANOCYTE STIMULATING HORMONE RECEPTOR)	7,806
TUBA8	BF195571	TUBULIN, ALPHA 8	5,523
TUBA1	AA180912	TUBULIN, ALPHA 1 (TESTIS SPECIFIC)	4,26
TUBA2	AA426374	TUBULIN, ALPHA 2	5,2
TUBA2	AA626698	ALPHA-TUBULIN ISOTYPE H2-ALPHA	4,922
TUBB	A1672565	TUBULIN, BETA 2A	14,003
TUBB2	A1000256	TUBULIN, BETA 2C	2,334
TUBA3	AA865469	TUBULIN, ALPHA 3	6,407
Functional Group 7		P-value: 0.00083	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
INSR	T47312	INSULIN RECEPTOR	0,174
PDGFRA	H23235	PLATELET-DERIVED GROWTH FACTOR RECEPTOR, ALPHA POLYPEPTIDE	19,359
BMPR1A	AA991180	BONE MORPHOGENETIC PROTEIN RECEPTOR, TYPE IA	0,415
KIT	N20798	V-KIT HARDY-ZUCKERMAN 4 FELINE SARCOMA VIRAL ONCOGENE HOMOLOG	0,231
TEK	H02848	TEK TYROSINE KINASE, ENDOTHELIAL (VENOUS MALFORMATIONS, MULTIPLE CUTANEOUS AND MUCOSAL)	0,14
EPHA7	N91461	EPH RECEPTOR A7	0,031
ERBB3	AA664212	V-ERB-B2 ERYTHROBLASTIC LEUKEMIA VIRAL ONCOGENE HOMOLOG 3 (AVIAN)	13,96
RAGE	N77779	RENAL TUMOR ANTIGEN	0,475
RYK	T77810	RYK RECEPTOR-LIKE TYROSINE KINASE	0,272
TYRO3	BM665421	TYRO3 PROTEIN TYROSINE KINASE	4,212
FGFR4	AA446994	FIBROBLAST GROWTH FACTOR RECEPTOR 4	0,178
Functional Group 8		P-value: 0.0026	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
GPR4	AI492409	G PROTEIN-COUPLED RECEPTOR 4	0,152
EST	R44664	G PROTEIN-COUPLED RECEPTOR 17	45,13
P2RY5	R91539	PURINERGIC RECEPTOR P2Y, G-PROTEIN COUPLED, 5	0,355
FZD4	AA677200	FRIZZLED HOMOLOG 4 (DROSOPHILA)	0,131
CCRL2	AI288845	CHEMOKINE (C-C MOTIF) RECEPTOR-LIKE 2	7,673
AGTRL1	R58969	ANGIOTENSIN II RECEPTOR-LIKE 1	20,807
EST	H71474	FRIZZLED HOMOLOG 7 (DROSOPHILA)	0,119
GPR153	AA777493	G PROTEIN-COUPLED RECEPTOR 153	3,139
MGC21621	W52061	MAS-RELATED GPR, MEMBER F	0,041
RAMP1	BE262882	RECEPTOR (CALCITONIN) ACTIVITY MODIFYING PROTEIN 1	18,624
Functional Group 9		P-value: 0.0044	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
ARL7	N35301	ADP-RIBOSYLATION FACTOR-LIKE 4C	11,492
ARF4L	AA878652	ADP-RIBOSYLATION FACTOR-LIKE 4D	0,081
EST	H98534	RAB9A, MEMBER RAS ONCOGENE FAMILY	2,157
RAB31	AA432084	RAB31, MEMBER RAS ONCOGENE FAMILY	1,795
RAB33A	AI360342	RAB33A, MEMBER RAS ONCOGENE FAMILY	5,832
RRAS2	R21415	RELATED RAS VIRAL (R-RAS) ONCOGENE HOMOLOG 2	0,206
RASD1	BM674708	RAS, DEXAMETHASONE-INDUCED 1	18,885
ARHN	AI027909	RHO FAMILY GTPASE 2	12,461
RALB	W15297	V-RAL SIMIAN LEUKEMIA VIRAL ONCOGENE HOMOLOG B (RAS RELATED; GTP BINDING PROTEIN)	2,324

Table B.4: Functional analysis of genes with q-value lower than 0.02. (3/3).

Functional Group 10		P-value: 0.011	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
PLK1	AA629262	POLO-LIKE KINASE 1 (DROSOPHILA)	4,297
RIPK1	AA426324	RECEPTOR (TNFRSF)- INTERACTING SERINE- THREONINE KINASE 1	0,631
NEK6	AA463188	NIMA (NEVER IN MITOSIS GENE A)-RELATED KINASE 6	3,883
PRKCN	AA417816	PROTEIN KINASE D3	0,373
PRKACB	AA459980	PROTEIN KINASE, CAMP- DEPENDENT, CATALYTIC, BETA	5,018
RAGE	N77779	RENAL TUMOR ANTIGEN	0,475
PRKCM	N53380	PROTEIN KINASE D1	5,168
PRKCD	AA496360	PROTEIN KINASE C, DELTA	0,386
PRKCH	AA128274	PROTEIN KINASE C, ETA	0,187

Functional Group 11		P-value: 0.036	
Gene symbol	Accession number	Gene description	GBM/mMEN ratio
CYP1B1	AA448157	CYTOCHROME P450, FAMILY 1, SUBFAMILY B, POLYPEPTIDE 1	0,093
CYP4Z1	H21977	CYTOCHROME P450, FAMILY 4, SUBFAMILY Z, POLYPEPTIDE 1	0,223
CYB5	R92281	CYTOCHROME B5 TYPE A (MI- CROSOMAL)	0,333
CYP4B1	AA291484	CYTOCHROME P450, FAMILY 4, SUBFAMILY B, POLYPEPTIDE 1	0,028
CYP3A5	BF062953	CYTOCHROME P450, FAMILY 3, SUBFAMILY A, POLYPEPTIDE 5	0,127

Table B.5: Comparison of gene-expression values between cDNA microarrays and RT-PCR. This table shows the GBM/mMEN ratios derived from gene-expression values from cDNA microarrays and RT-PCR. Ratios were obtained from expression values originating from the three GBM and three mMEN selected samples to be subjected to RT-PCR. For cDNA microarrays the ratio was obtained by dividing the average of expression values from the 3 GBM by that from the 3 mMEN. In contrast, for RT-PCR this ratio was computed as the power of 2 to the negative difference between the Ct average from the 3 GBM and 3 mMEN samples subjected to RT-PCR ($2^{-(Ct(GBM)-Ct(MEN))}$).

Gene Symbol	Accession number	cDNA	RT-PCR
		GBM/MEN ratio	GBM/MEN fold-change
GFAP	AA069414	813	50419
PTPRZ1	AA476460	378	517
GPM6B	AA284329	149	157
MT3	A1362950	27	61
CA2	H23187	17	28
TUBBB	A1672565	9	42
PRELP	AA131664	0.022	0.04
APM2	AA478298	0.019	0.03
DSP	H90899	0.0016	0.0007
OGN	AA045327	0.00076	0.0005

List of Figures

1	Flowchart of the Dissertation.	xviii
1.1	Pattern Recognition solution based on the Machine Learning approach. . .	2
1.2	LDA and QDA decision boundaries	3
1.3	FLDA class separation in a 2D space.	4
1.4	Nearest-neighbor partitioning of a 2-dimensional space.	5
1.5	Diagram for the 2-layer MLP.	5
1.6	Support Vectors and margin boundaries of a SVM.	6
1.7	\mathbf{M} motion when recovering the precession around \mathbf{B}_0	9
1.8	Chemical Shift (CS) for OCH_3 and CH_3CO [29].	10
1.9	Spotted and Oligonucleotide microarrays.	11
1.10	SV INTERPRET GUI for SV spectra.	14
2.1	Distribution of STT cases into histological groups.	27
2.2	Representative MRI of STT.	28
3.1	Long TE, Short TE and Combined spectra.	46
3.2	ACC vs. PCs. for the PCA+LDA approach.	48
3.3	Selected points of the SW approach in Short TE and Long TE.	51
3.4	Four potential outliers.	53
3.5	MRI scans and voxel localation of two potential outliers.	55
3.6	LDA latent space for Short TE.	57
3.7	LDA latent space for Long TE.	58
3.8	LDA latent space for Combined TE.	59
4.1	Spectrum (1/3) modeled by EMBTD and EMBTD _r	71
4.2	Spectrum (2/3) modeled by EMBTD and EMBTD _r	72
4.3	Spectrum (3/3) modeled by EMBTD and EMBTD _r	73
5.1	Box-whisker plots of the performance for each problem.	86
5.2	Scatter plot $BER(CV)$ vs. $BER(IT)$	87
5.3	Multiple comparison of mean ranks by problem.	89
5.4	Box-whisker plots of the performance for each FE method.	90
5.5	Multiple comparison of mean ranks by FE method.	91
5.6	Box-whisker plots of the performance by classification method.	92
5.7	Multiple comparison of mean ranks by classification metod.	93
5.8	Prototypes by class in training and test datasets.	94
5.9	Potential outliers (1/2) detected as a consequence of this study.	96
5.10	Potential outliers (2/2) detected as a consequence of this study.	97

6.1	Plot of normalised data and LDA-based predictor.	113
6.2	Hierarchical cluster of differentially expressed genes.	115
7.1	STT-dDSS architecture.	133
7.2	KNN-based engine for Benign/Malignant in the STT-dDSS.	135
7.3	Register Window of the STT-dDSS.	137
7.4	Visualization of statistical distributions of the categorical features.	138
7.5	Visualization of the Universe of cases for the STT-dDSS.	138
7.6	STT radiologist visual interface appearances.	139
7.7	The HEALTHAGENTS architecture	143
7.8	Classification window of the GUI for the radiological units.	144

List of Tables

1.1	Brain tumor and CNS types by age.	13
2.1	Comparison of ANN, KNN, and SVM.	32
2.2	Best results with ANN.	32
2.3	Best results with KNN.	33
2.4	Best results with SVM.	33
2.5	Errors in STT Bening/Malignant classification.	34
3.1	Samples per class used in the study.	43
3.2	Multiclass classification among AGG, MEN and LGG.	48
3.3	LS-SVM for pairwise classification of AGG, MEN and LGG classes.	49
3.4	Significance of difference tests among the three approaches.	50
4.1	Difference between the estimated shifting and the typical CS.	74
4.2	EMBTD-based classification among AGG, MEN and LGG.	74
4.3	Multiclass EMBTD-based pairwise classification.	75
5.1	Training and Test Corpora in the Multicenter evaluation.	81
5.2	Training and Test cases by manufacturer.	82
5.3	Training and Test cases by acquisition center.	82
5.4	Best results for the multicenter evaluation.	85
5.5	Number of classifiers obtained for each classification Problem.	88
5.6	Number of classifiers by FE method.	91
5.7	Number of classifiers by classification method.	93
5.8	Multicenter evaluation of EMBTD and EMBTD _r	98
7.1	Results using MN, SVM, DT and KNN for histological classification.	134
7.2	GMOR for the three approaches in the questions to solve.	141
B.1	Selection of differentially expressed genes.	176
B.2	Functional analysis of genes with q-value lower than 0.02. (1/3)	177
B.3	Functional analysis of genes with q-value lower than 0.02. (2/3).	178
B.4	Functional analysis of genes with q-value lower than 0.02. (3/3).	179
B.5	Gene-expression values on cDNA and RT-PCR.	179

