

KATHOLIEKE UNIVERSITEIT
LEUVEN

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Electrical Engineering (ESAT)

CLASSIFICATION OF BRAIN TUMORS BASED ON MAGNETIC RESONANCE SPECTROSCOPY

Jan LUTS

Dissertation presented in
partial fulfillment of the
requirements for the degree
of Doctor in Engineering

January 2010

CLASSIFICATION OF BRAIN TUMORS BASED ON MAGNETIC RESONANCE SPECTROSCOPY

Jan LUTS

Jury:
Prof. dr. ir. A. Haegemans, president
Prof. dr. ir. S. Van Huffel, promotor
Prof. dr. ir. J.A.K. Suykens, copromotor
Prof. dr. ir. D. Vandermeulen
Prof. dr. U. Himmelreich
Prof. dr. A.-M. De Meyer
Prof. dr. L.M.C. Buydens (RUN)
Prof. dr. B. Celda (UVEG)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering

January 2010

© Katholieke Universiteit Leuven – Faculty of Engineering
Kasteelpark Arenberg 1, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the publisher.

D/2010/7515/8
ISBN 978-94-6018-169-6

Voorwoord

In de eerste plaats zou ik mijn promotoren prof. dr. ir. Sabine Van Huffel en prof. dr. ir. Johan Suykens willen bedanken. Sabine, bedankt voor de kans om in je onderzoeksgroep te starten en me meteen onder te dompelen in eTUMOUR en HealthAgents. Je raadgevingen en vertrouwen in me maakten van de afgelopen vier jaar een erg positieve ervaring. Johan, bedankt voor de vele verhelderende discussies, nuttige suggesties en je nooit aflatende beschikbaarheid. Dankzij jullie is dit geworden wat het is. Tevens ben ik het Instituut voor de Aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen (IWT-Vlaanderen) erkentelijk voor de financiële ondersteuning door middel van een specialisatiebeurs.

Verder zou ik graag prof. dr. ir. Dirk Vandermeulen bedanken voor zijn expertise en talrijke suggesties. Mede door jouw toedoen heb ik een aantal beeldverwerkingstechnieken in dit werk geïntegreerd. Ze vormen een absolute meerwaarde. I'm also very grateful to prof. dr. Uwe Himmelreich. Thanks for your suggestions during the last year and for helping me with the latest data. I'm sure that very nice, new, results will come out in the near future. Tevens gaat mijn dank uit naar prof. dr. Anne-Marie De Meyer voor haar inbreng als lid van mijn examencommissie en prof. dr. ir. Ann Haegemans voor haar bereidheid de examencommissie voor te zitten.

An important part of the work in this thesis has been performed in the context of two European projects: eTUMOUR and HealthAgents. I would like to thank prof. dr. Lutgarde Buydens and prof. dr. Bernardo Celda for all their efforts, suggestions and comments during these project years and for providing useful comments about my work. These projects created a very nice environment to do research and offered unique opportunities. Thanks for this Juan-Miguel, Javier, Carlos, Elies, Salvador, Miguel, M.Carmen, Daniel, Margarida, Ana Paula, Ivan, Carles, Geert, Bas, Arend, Montse, Magi, Mariola, Xavier, Francesc, Miguel, Carla, Horacio, Theo, Alex, Andrew, Bo, Liang, David, Paul, Sri and many others.

De mensen van onze eigen onderzoekseenheid verdienen uiteraard ook een speciale vermelding. Bedankt Teresa, Diana, Bart, Wim, Anneleen, Bart, Jean-Michel, Mieke, Ivan, Arjan, Maarten, Vanya, Ben, Jean-Baptiste, Mariya, Elisa, Maria Isabel, Anca, Lieven, Katrien, Wouter, Steven, Joachim, Maria, Alexander, Bori, Dominique, Bogdan, Ivan, Thijs, Fabian, Carlos en uiteraard ook Sofie en Flemming. Daarnaast werden dankzij Ida, Ilse, John, Bart en Veerle allerlei administratieve en financiële zaken vlot afgehandeld.

Graag wil ik ook mijn vrienden bedanken en mezelf verontschuldigen voor al die keren dat ik er niet (of te laat) was.

Hoog tijd om de familie te bedanken: bomma's en bompa's, tantes en nonkels, nichten en neef.

Tot slot rest me m'n ouders en Inge te bedanken. Bedankt voor alle kansen, geduld, steun, raad en geluk. Tot u spreekt voortaan een doctor ;-) .

Abstract

Nowadays, diagnosis and treatment of brain tumors is based on clinical symptoms, radiological appearance, and often histopathology. Magnetic resonance imaging (MRI) is a major noninvasive tool for the anatomical assessment of tumors in the brain. However, several diagnostic questions, such as the type and grade of the tumor, are difficult to address using MRI. The histopathology of a tissue specimen remains the gold standard, despite the associated risks of surgery to obtain a biopsy. In recent years, the use of magnetic resonance spectroscopy (MRS), which provides a metabolic profile, has gained a lot of interest for a more detailed and specific noninvasive evaluation of brain tumors. In particular, magnetic resonance spectroscopic imaging (MRSI) is attractive as this may also enable to visualize the heterogeneous spatial extent of tumors, both inside and outside the MRI detectable lesion.

As manual, individual, viewing and analysis of the multiple spectral patterns, obtained by an magnetic resonance (MR) spectroscopy exam, is time-consuming and often needs specific spectroscopic expertise, it is not practical in a clinical environment. Widespread use of MR spectroscopy requires specialized processing and evaluation of the data and easy and rapid display of the results as images or maps for routine clinical interpretation of an exam. In this thesis, different approaches have been developed to process and integrate MRI, MRS and MRSI data for differential diagnosis of brain tumors, and to visualize the obtained results in an attractive manner. This thesis identifies problems that can be encountered during this procedure and provides possible solutions. In particular, feature extraction from MR spectra, (multi-class) classification of MRS(I), integration of MRI and MRSI data and the visual representation of the tissue typing results have been studied.

The conclusions and developments of this thesis have been established within the context of two European projects of the *Sixth Framework Programme for Research and Technological Development*. HealthAgents (*Agent-based distributed decision*

support system for brain tumour diagnosis and prognosis, 2006−2008) and 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) aim to develop decision support systems to assist clinicians in decision making. Different scientific contributions of this thesis have been implemented in these decision support systems. This can potentially expand the use of MR spectroscopy in clinical practice to support diagnosis and prognosis of brain tumors, and it may allow individually optimized therapy planning.

Korte inhoud

De diagnosestelling en behandeling van een hersentumor is gebaseerd op klinische symptomen, radiologische bevindingen en vaak worden hierbij histopathologische bevindingen in acht genomen. Magnetische resonantie beeldvorming (MRB) is een voorname, niet-invasieve techniek voor het beoordelen van de anatomie van tumoren. Een aantal karakteristieken, zoals het type en graad van een tumor, is echter moeilijk te achterhalen met behulp van MRB. Hierdoor blijft histopathologisch onderzoek van een weefselstaal, ondanks het bijhorende risico van een operatie om dit biopt te bekomen, nog steeds de gouden standaard. Anderzijds is er steeds meer toenemende interesse voor het gebruik van magnetische resonantie spectroscopie (MRS) vermits deze methode, welke een metabolisch profiel verschaft, een meer gedetailleerde en specifieke evaluatie van hersentumoren toelaat op niet-invasieve wijze. Magnetische resonantie spectroscopische beeldvorming (MRSB) is in het bijzonder aantrekkelijk daar deze techniek het mogelijk maakt om het heterogene, ruimtelijke karakter van tumoren, zowel binnen als buiten het detecteerbare gebied op MRB, te visualiseren.

Uit een magnetische resonantie (MR) spectroscopisch onderzoek vloeit doorgaans een grote hoeveelheid aan spectroscopische gegevens voort. Vermits het manueel en individueel beschouwen en analyseren van deze hoeveelheid spectrale patronen bijzonder tijdrovend is en specifieke spectroscopische deskundigheid vereist, blijkt de methodologie bijzonder moeilijk werkbaar in de klinische praktijk. Wijdverspreid gebruik van magnetische resonantie spectroscopische technieken vergt immers gespecialiseerde verwerking en evaluatie van gegevens en eenvoudige, snelle weergave van resultaten op basis van afbeeldingen om de bevindingen te beoordelen. In deze thesis ligt de nadruk op het ontwikkelen van methodieken om MRB, MRS en MRSB gegevens te verwerken en te integreren voor diagnosestelling van hersentumoren, en om de bekomen resultaten op een aantrekkelijke manier voor te stellen voor clinici. De thesis besteedt aandacht aan de vele moeilijkheden die opduiken bij het realiseren van deze doelstellingen en stelt mogelijke oplossingen voor. Meer in het bijzonder komen het extraheren van kenmerken uit MR spectra, (meerklasse) classificatie op basis van MRS(B) en het grafisch

voorstellen van een bekomen weefseltypering aan bod.

Het uitgevoerde onderzoek, beschreven in deze thesis, is tot stand gebracht binnen het kader van twee Europese projecten van het *Sixth Framework Programme for Research and Technological Development*. Zowel HealthAgents (*Agent-based distributed decision support system for brain tumour diagnosis and prognosis*, 2006−2008) als 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) streven het ontwerp van beslissingssystemen ter ondersteuning van de clinici na. Verscheidene wetenschappelijke bijdragen van de huidige thesis zijn reeds geïmplementeerd in deze ondersteunende beslissingssystemen. Dit schept de mogelijkheid om het gebruik van MR spectroscopie, ter diagnosestelling en prognose van hersentumoren, verder uit te breiden binnen de klinische praktijk en het opent de deur voor meer geoptimaliseerde, patiëntgerichte therapie-ontwikkeling.

Nomenclature

2D-TSI	two-dimensional turbo spectroscopic imaging
AIDS	acquired immune deficiency syndrome
Ala	alanine
ANN	artificial neural network
ANOVA	analysis of variance
API	application programming interface
AQSES	automated quantitation of short echo time MRS spectra
AQSESGUI	AQSES graphical user interface
ARD	automatic relevance determination
AUC	area under the receiver operating characteristic curve
BAR	balanced accuracy rate
BDK	bi-directional Kohonen
BER	balanced error rate
CBTRUS	Central Brain Tumor Registry of the United States
CCA	canonical correlation analysis
CDP	Centre Diagnostic Pedralbes
CHESS	chemical shift selective
Cho	choline
CNS	central nervous system
Cr	creatine
CSF	cerebrospinal fluid
CSI	chemical shift imaging
CT	computed tomography
dDSS	distributed decision support system
dLDA	LDA with diagonal covariance matrix
DMS	data manipulation software
DNA	deoxyribonucleic acid
dQDA	QDA with diagonal covariance matrix
DSS	decision support system
DTI	diffusion tensor imaging
EBSS	evidence-based search service
ECOC	error correcting output codes

FFT	fast Fourier transform
FID	free induction decay
FLDA	Fisher's linear discriminant analysis
FLENI	Fundacion para la Lucha contra las Enfermedades Neurologicas de la Infancia
FOV	field of view
Gd	Gadolinium
GE	General Electric
Glc	glucose
Gln	glutamine
Glu	glutamate
Glx	glutamate and glutamine
Gly	glycine
GPCho	glycerophosphocholine
GUI	graphical user interface
HLSVD	Hankel Lanczos singular value decomposition
HLSVD-PRO	HLSVD with partial reorthogonalization
HR-MAS	high-resolution magic angle spinning
HSVD	Hankel singular value decomposition
ICA	independent component analysis
ICT	information and communication technologies
IDI	Institut de Diagnostic per la Imatge
JADE	Java agent development
KLR	kernel logistic regression
KNN	k-nearest neighbor
Lac	lactate
LASSO	least absolute shrinkage and selection operator
LCModel	linear combination of model spectra
LDA	linear discriminant analysis
Lip	lipids
LS-SVM	least squares support vector machine
MCD	minimum covariance determinant
mI	myo-inositol
MLP	multi-layer perceptron
MM	macromolecule
MOC	minimal output coding
MP-FIR	maximum phase finite impulse response
MR	magnetic resonance
MRI	magnetic resonance imaging
MRS	magnetic resonance spectroscopy
MRSI	magnetic resonance spectroscopic imaging
ms	milliseconds
MSE	mean squared error
TM	mixing time

MUL	Uniwersytet Medyczny w Lodzi
NAA	N-acetyl-aspartate
NLL	negative log-likelihood
NMR	nuclear magnetic resonance
PCA	principal component analysis
PCho	phosphocholine
PCr	phosphocreatine
PD	proton density
PET	positron emission tomography
ppm	parts per million
PRESS	point-resolved spectroscopy
QDA	quadratic discriminant analysis
QUALITY	quantification improvement by converting lineshapes to the Lorentzian type
QUECC	QUALITY and eddy current correction
QUEST	quantitation based on quantum estimation
RBF	radial basis function
RF	radio frequency
RFE	recursive feature elimination
ROC	receiver operating characteristic
RUN	Radboud Universiteit Nijmegen
SGHMS	St George's Hospital Medical School
SNR	signal-to-noise ratio
STEAM	stimulated echo acquisition mode
SVM	support vector machine
SW	spectral width
TAM	technology acceptance model
Tau	taurine
tCho	total choline
tCr	total creatine
TE	echo time
TR	repetition time
UMCN	University Medical Center Nijmegen
UVEG	Universitat de Valencia-Estudi General
VARPRO	variable projection
VSM	variable-kernel similarity metric
WHO	World Health Organization
XML	extensible markup language

Contents

Voorwoord i

Abstract iii

Korte inhoud v

Nomenclature vii

Contents xi

1 Introduction 1

1.1 Brain tumor diagnosis 1

1.2 Magnetic resonance 2

1.2.1 Basic principles of magnetic resonance 2

1.2.2 Magnetic resonance spectroscopy 7

1.2.3 Magnetic resonance imaging 9

1.2.4 Magnetic resonance spectroscopic imaging 11

1.3 Brain tumors and magnetic resonance 13

1.3.1 Epidemiology 13

1.3.2 Diagnosis 14

1.3.3 Treatment 17

1.3.4 Classification 18

1.3.5 MRI, MRS and MRSI for brain tumor diagnosis and treatment 19
1.4 Aims of the thesis . . . 22
1.5 Organization of the thesis . . . 23
1.6 Personal contributions . . . 25
1.7 Conclusions . . . 26

2 Processing of MR spectra 27
2.1 Preprocessing methods . . . 27
2.2 Classification algorithms . . . 30
2.2.1 K-nearest neighbor . . . 30
2.2.2 Discriminant analysis . . . 31
2.2.3 Multi-layer perceptrons . . . 34
2.2.4 Support vector machines . . . 36
2.2.5 Least squares support vector machines . . . 40
2.2.6 Bayesian least squares support vector machines . . . 41
2.2.7 Kernel logistic regression . . . 43
2.2.8 Canonical correlation analysis . . . 46
2.2.9 Software . . . 47
2.3 Model performance evaluation . . . 47
2.3.1 Performance measures . . . 48
2.3.2 Evaluation strategies . . . 49
2.3.3 Statistical analysis . . . 50
2.4 Conclusions . . . 52

3 Quantification of MR spectra 55
3.1 Modeling MRS(I) data . . . 55
3.2 Peak integration . . . 57
3.3 LCModel . . . 57

3.4 AQSES 58
3.5 QUEST 62
3.6 Conclusions 62

4 Relevance of automatically selected MRS, MRI and MRSI features 65
4.1 Feature selection 65
4.2 Materials 68
4.2.1 MRS data set 68
4.2.2 MRI and MRSI data set 70
4.3 Feature selection methods 73
4.3.1 Fisher discriminant criterion 73
4.3.2 Kruskal-Wallis test 73
4.3.3 Relief-F 74
4.3.4 Automatic relevance determination for LS-SVMs 75
4.3.5 Experimental setup 77
4.4 Results 77
4.5 Discussion 78
4.6 Conclusions 85

5 Feature extraction from MRS and MRSI for brain tumor classification 87
5.1 Feature extraction for MRS(I) data 87
5.2 Materials 89
5.2.1 MRS data set 89
5.2.2 MRSI data set 90
5.3 Feature extraction methods 91
5.3.1 Full spectra 91
5.3.2 Peak integration 92
5.3.3 Peak regions 93

5.3.4 Bagging with Fisher criterion, Kruskal-Wallis test and Relief-F . . . 93
5.3.5 Independent component analysis . . . 93
5.3.6 Linear principal component analysis . . . 94
5.3.7 Stepwise selection . . . 95
5.3.8 AQSES . . . 95
5.3.9 QUEST . . . 96
5.3.10 Experimental setup . . . 96
5.4 Results . . . 96
5.5 Discussion . . . 98
5.6 Conclusions . . . 103

6 Multi-project brain tumor classification using multi-center MRS 105
6.1 Multi-project setting . . . 105
6.2 Materials . . . 106
6.3 Methods . . . 109
6.3.1 Classification and feature extraction . . . 109
6.3.2 Experimental setup . . . 110
6.4 Results . . . 110
6.5 Discussion . . . 113
6.6 Conclusions . . . 118

7 Multi-class system for brain tumor recognition using MRI and MRSI 119
7.1 Multi-class classification . . . 119
7.2 Materials . . . 121
7.3 Multi-class classification methods . . . 121
7.3.1 Pairwise combination of probabilities . . . 122
7.3.2 Price et al. . . . 123
7.3.3 Hastie and Tibshirani . . . 123

7.3.4 Wu et al. − method 1 . . . 125
7.3.5 Wu et al. − method 2 . . . 125
7.3.6 Experimental setup . . . 126
7.4 Results . . . 126
7.5 Discussion . . . 128
7.6 Conclusions . . . 133

8 Nosologic imaging of the brain 135
8.1 Nosologic imaging . . . 135
8.2 Materials . . . 136
8.3 Methods . . . 137
8.3.1 Subject-specific abnormal tissue prior . . . 137
8.3.2 Tumor segmentation based on outlier detection . . . 139
8.3.3 Classification of the abnormal region . . . 140
8.3.4 Experimental setup . . . 141
8.4 Results . . . 141
8.4.1 Analysis of the patient data . . . 141
8.4.2 Class probabilities . . . 147
8.4.3 Subject-specific abnormal tissue prior . . . 147
8.5 Discussion . . . 153
8.6 Conclusions . . . 157

9 Differentiation between brain metastasis and glioblastoma multiforme 159
9.1 Introduction . . . 159
9.2 Materials . . . 161
9.3 Methods . . . 162
9.3.1 Threshold method . . . 162
9.3.2 Nosologic images . . . 163

9.4 Results . . . 163
9.4.1 Metabolite ratios . . . 164
9.4.2 Infiltration . . . 164
9.4.3 Nosologic images . . . 169
9.5 Discussion . . . 169
9.6 Conclusions . . . 179

10 HealthAgents: a distributed DSS for brain tumor diagnosis 181
10.1 Decision support systems . . . 181
10.2 HealthAgents . . . 182
10.2.1 Agent technology . . . 182
10.2.2 Data entry . . . 185
10.2.3 Classification . . . 185
10.3 Classification framework . . . 188
10.3.1 Design of the framework . . . 188
10.3.2 Integration within HealthAgents . . . 192
10.4 Results and evaluation . . . 196
10.5 Conclusions . . . 201

11 Conclusions and further research 203
11.1 Summary . . . 203
11.2 Further research . . . 206

Bibliography 209

Publication list 227

Curriculum vitae 233

Chapter 1

Introduction

This chapter introduces the general concepts of magnetic resonance (MR) and describes its relevance for the diagnosis of brain tumors (Section 1.1). Section 1.2 explains the theoretical foundations of MR and how these principles are used for MR spectroscopy, MR imaging and MR spectroscopic imaging. Section 1.3 provides background information about brain tumors and focuses on the potential of MR for brain tumor diagnosis. The aim of the thesis is defined in Section 1.4 and an overview of the structure of the thesis is provided in Section 1.5. Personal contributions are summarized in Section 1.6.

1.1 Brain tumor diagnosis

A brain tumor is defined as the growth of abnormal cells in the tissues of the brain (National Cancer Institute, http://www.cancer.gov/ , accessed on June 27, 2009). Brain tumors can be benign (noncancerous) or malignant (cancerous). In contrast to normal cells, cancer cells result from uncontrolled cell growth and can grow into adjacent tissue. Although benign tumors can become large and press on healthy organs and tissue, which can potentially affect their functioning, they rarely invade other tissue. Primary brain tumors start from the brain itself, while secondary brain tumors (i.e. metastatic tumors) originate from other parts in the body.

Usually the definite diagnosis of a brain tumor is based on histological examination of tissue samples obtained by a biopsy. A biopsy is highly invasive and has significant risks with an estimated morbidity of 2.4−3.5 % and a death rate of 0.2−0.8 % [58, 57, 84]. In addition, a biopsy will not always be performed as

for instance in very elderly or infirm patients, or in patients with very slowly growing tumors. Moreover, surgery could be avoided for certain pathologies such as lymphomas and brain abscesses [219]. Noninvasive imaging methods, e.g. magnetic resonance imaging (MRI), computed tomography (CT) and positron emission tomography (PET), that could diagnose a brain tumor, would avoid unnecessary surgery. MRI represents an interesting approach for the anatomical assessment of brain tumors since it provides superior soft tissue contrast and high-resolution information. The method is yet widely used in clinical practice, but it fails to detect metabolic changes. Interestingly, magnetic resonance spectroscopy (MRS) and magnetic resonance spectroscopic imaging (MRSI) can provide important information about the metabolism. However, despite these promising potentials MRS and MRSI are not yet widely applied in the clinical environment since specialized processing and analysis of the acquired data are required. This thesis focuses on the application of MRI, MRS and MRSI for the diagnosis of brain tumors.

1.2 Magnetic resonance

Although the first nuclear magnetic resonance (NMR), or magnetic resonance (MR) in short, experiments originate from more than 60 years ago, the principles have only been widely used for medical imaging in clinical centers since the eighties. MR represents a powerful technique that provides the fundamentals for a large variety of applications, e.g. MRI, MRS and MRSI. These noninvasive tools provide in vivo medical images and chemical information about tissues, where in vivo refers to the use on living animals or human beings. Compared to CT or conventional radiography, which use ionizing radiation, MR causes only small patient risk since it is based on the magnetic properties of atoms.

1.2.1 Basic principles of magnetic resonance

Several atomic nuclei, e.g. proton (^{1}H), carbon (^{13}C), fluor (^{19}F) and phosphorus (^{31}P), possess an angular momentum or spin and are potentially useful for application in MR. The spin property can be represented as a rotational motion of the particle around its own axis and this makes it behave like a little magnet (Figure 1.1). The aforementioned atomic nuclei have a nuclear spin quantum number, which depends on the number of protons and neutrons, of $\frac{1}{2}$. According to quantum mechanics, this results in two possible orientations when these atomic nuclei are embedded in an external magnetic field B_0.

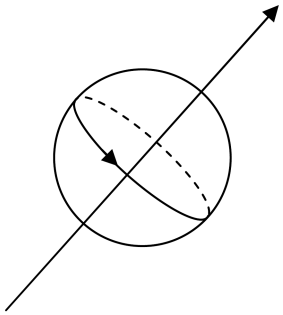


Figure 1.1: *Spinning atomic nucleus creating a magnetic moment.*

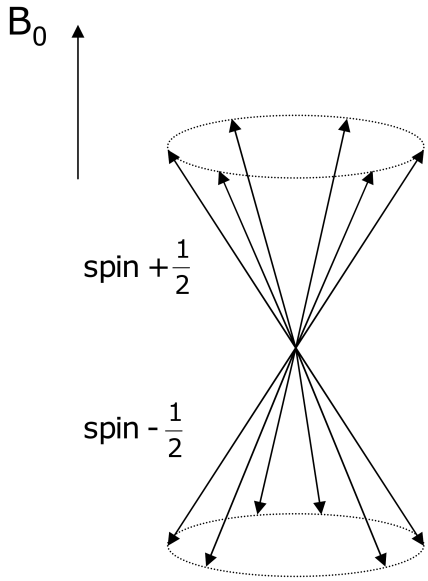


Figure 1.2: *Spins are precessing around an external magnetic field B_0. Spins are aligned along or opposite to B_0.*

The magnetic moments align with the direction of B_0, resulting in parallel (spin $+\frac{1}{2}$) and antiparallel orientations (spin $-\frac{1}{2}$) (Figure 1.2). The spins precess with a specific frequency (i.e. the Larmor frequency f_0) which is proportional to B_0 (e.g. $f_0 = 63.86$ MHz for $B_0 = 1.5$ T) and defined by the Larmor equation

$$f_0 = \frac{\gamma B_0}{2\pi}, \tag{1.1}$$

where γ represents the gyromagnetic ratio, being characteristic for the isotope (e.g. $\gamma = 42.58$, 10.71, 40.08 and 17.25 MHz/T, for ^{1}H, ^{13}C, ^{19}F and ^{31}P, respectively). The two orientations correspond with different energy states, where the antiparallel direction is less stable as it matches with a higher energy state. The difference in energy between both states is proportional to the magnitude of the external magnetic field such that

$$\Delta E = \gamma \hbar B_0,\ \hbar = \frac{h}{(2\pi)}, \tag{1.2}$$

where h is the Planck constant (6.626×10^{-34} J s). In thermal equilibrium the distribution between spin $+\frac{1}{2}$ and spin $-\frac{1}{2}$ is defined by

$$\frac{n_-}{n_+} = e^{-\Delta E/kT}, \tag{1.3}$$

where k is the Boltzman constant (1.381×10^{-23} J/K) and T the temperature in Kelvin. This ratio is very close to unity in normal circumstances because the number of spins in the low energy level is slightly more than the number of spins in the high energy level, causing a net magnetic field (Figure 1.3). The net magnetization, which is represented by a vector, only consists of a longitudinal component along the Z-axis.

To induce transitions between both energy states, an oscillating magnetic field is applied by a coil in a plane perpendicular to the direction of the static external field B_0. The frequency of the oscillating field has to correspond with the energy difference between the two states and matches to the Larmor frequency. The MR signal originates from the process of energy absorption by spins, making a transition from the lower energy state to the higher energy state, and energy emission by spins making a transition from the higher energy state to the lower energy state.

While the net magnetization is only characterized by a longitudinal component at equilibrium, the process of excitation with a 90° radio frequency (RF) pulse introduces a perpendicular transverse component, situated in the $X - Y$-plane

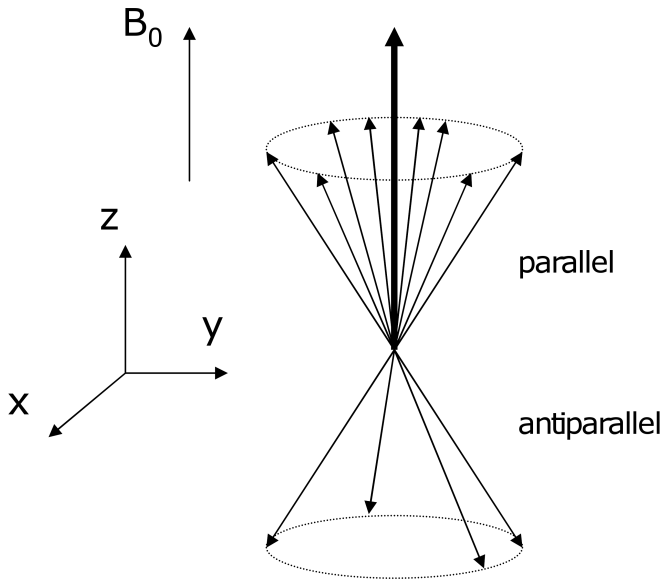


Figure 1.3: *A collection of precessing spins around an external magnetic field B_0 results in a net magnetization in the direction of B_0.*

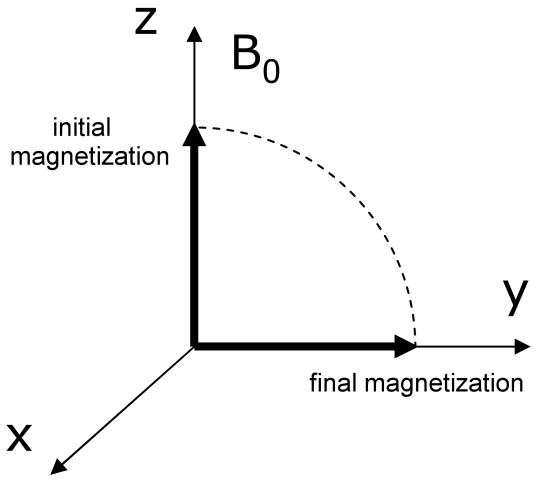


Figure 1.4: *An oscillating magnetic field tilts the net magnetization into the $X-Y$-plane.*

(Figure 1.4). The decrease in longitudinal magnetization is caused by the equal proportions of number of spins in both energy states and the transverse magnetization originates from spins getting into phase coherence.

When the oscillating magnetic field is switched off, spins return to their equilibrium state, causing the net magnetization to precess around the Z-axis until it recovers the original direction (Figure 1.5). During this precession motion, an electromotive force is induced in a receiver coil. The emitted signal is called the free induction decay (FID) signal and it corresponds to an exponentially decaying sinusoid in the time domain. Since the signal is proportional to the population difference between the two states and the ratio in Equation (1.3) is close to unity in normal circumstances, there is low absorption of energy, explaining the low signal-to-noise ratio (SNR) of the emitted signal. Several data acquisitions can be performed in order to increase the SNR of the final signal, which is obtained by averaging, although this increases measurement time. In addition, as Equations (1.2) and (1.3) suggest, imposing a larger external magnetic field can increase signal intensity.

Relaxation refers to the event of returning from the non-equilibrium state to the equilibrium state and this process consists of different aspects. The T_1 relaxation time or spin-lattice relaxation time represents the time constant for the recovery of the magnetization along the direction of B_0. Switching off the oscillating magnetic field results in the realignment of the spins with the external magnetic field B_0 because of the emission of the absorbed energy to the surrounding tissue. The T_2 relaxation time or spin-spin relaxation time represents the time constant for the relaxation in the $X-Y$-plane, which corresponds to the return to equilibrium of the transverse magnetization. This is the result of a dephasing process caused by the fact that spins, contributing to the net magnetization, experience a slightly different magnetic field due to interactions, and they are each associated with a particular Larmor frequency. The longer the elapsed time, the greater these phase differences become. Another factor which decreases the transverse component is the presence of variations in the external magnetic field B_0, called inhomogeneities. This results in a combined time constant called T_2^*, which is usually smaller than T_2 and explains a loss of signal. T_1 relaxation and T_2 relaxation occur simultaneously, but T_2 is always less than or equal to T_1. The parameters T_1 and T_2 are tissue-specific time constants since they are related to the physical interaction phenomena in the tissue, allowing to differentiate between various tissue types.

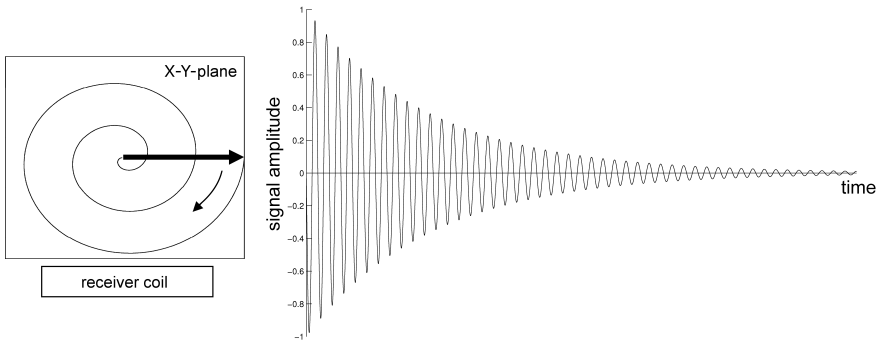


Figure 1.5: *A 90 ° radio frequency pulse results in a net magnetization, located in the $X-Y$-plane, which rotates around the Z-axis and induces an alternating current in the receiver coil.*

1.2.2 Magnetic resonance spectroscopy

MRS enables to obtain in vivo information about the molecular composition of tissue in a noninvasive way. It allows to identify molecular constituents, which are called metabolites. This procedure is based on the fact that nuclei of different atoms can be differentiated because they are characterized by a different γ, which results in a different resonance frequency. Nuclei of the same isotope, but belonging to different molecules, are separable in frequency due to the shielding effect. The existence of an electron cloud constitutes an electronic shield and perturbs the original external magnetic field B_0. The amount of shielding determines the effective magnetic field, which is experienced by the nucleus

$$B_{eff} = B_0(1 - \sigma'), \tag{1.4}$$

where σ' is the shielding constant, being a dimensionless number that depends on the electrical environment of the nucleus. Based on their chemical bonds nuclei resonate at different frequencies

$$f_{eff} = \frac{\gamma}{2\pi} B_0(1 - \sigma'). \tag{1.5}$$

The phenomenon of a change in the resonance frequency of these nuclei is called chemical shift. In addition, the interactions between neighboring nuclei in the same molecule induce a further differentiation in resonance frequencies, called spin-spin coupling or J-coupling. This explains for example the existence of a doublet, which

are two close peaks of the same intensity in the (frequency domain) MR spectrum, which is obtained after Fourier transformation of the time domain signal. The transformation from the time domain to the frequency domain is usually performed for visualization purposes, since it enables to interpret the composition of the signal more easily. In MRS, the chemical shift of a peak in the spectrum is often expressed in terms of the dimensionless unit parts per million (ppm), which is independent of the spectrometer frequency. A metabolite that resonates at a specific ppm-value for a spectrometer frequency of 63.8 MHz, also resonates at the same ppm-value for a spectrometer frequency of 500 MHz. The ppm-value of a resonance is defined by

$$ppm = \frac{f_{Hz}10^6}{f_S} + ppm_{ref} \tag{1.6}$$

where f_{Hz} denotes the resonance frequency in Hz, f_S the spectrometer frequency in Hz and ppm_{Ref} the ppm-value of the reference resonance, which is often 4.7 ppm in ^{1}H MR spectroscopy.

In addition to the strength of the external magnetic field B_0, the sample temperature, the relaxation times T_1, T_2 and T_2^*, there are still other important parameters that have an influence on the acquired MR signal. The timing of different pulses during an MR acquisition is determined by the echo time (TE) and the repetition time (TR). The TE represents the time in milliseconds (ms) between the excitation pulse and the echo signal in a spin echo sequence, while the TR (e.g. 2000 ms) is the time in ms between two excitation pulses. The choice of TR is based on a tradeoff between the desired SNR and the total measurement time. Long TEs ($\geq$ 130 ms) allow an easier extraction of metabolic information of the slowly decaying components, like choline (Cho), creatine (Cr) and N-acetyl-aspartate (NAA) in ^{1}H MR spectroscopy (Figure 1.6). Short TEs ($\leq$ 50 ms) result in larger peaks in the MR spectrum and provide more metabolite information as compared to long TEs because less signal intensity of the metabolites has decayed (Figure 1.7). This allows for instance the observation of myo-inositol (mI), glutamate (Glu) and glutamine (Gln), which are less pronounced at longer TEs. In short TE signals macromolecules (MMs) are prominent, originating from proteins and membrane components. These correspond to very broad peaks that contribute to an underlying and partially unknown baseline. Note that the area below the peak curve is determined by the metabolite concentration.

Single-voxel MR spectroscopy measures a signal from a volume limited to a single voxel. In practice this involves suppression of the water signal and selection of the voxel of interest. Water suppression is important since the amount of hydrogen

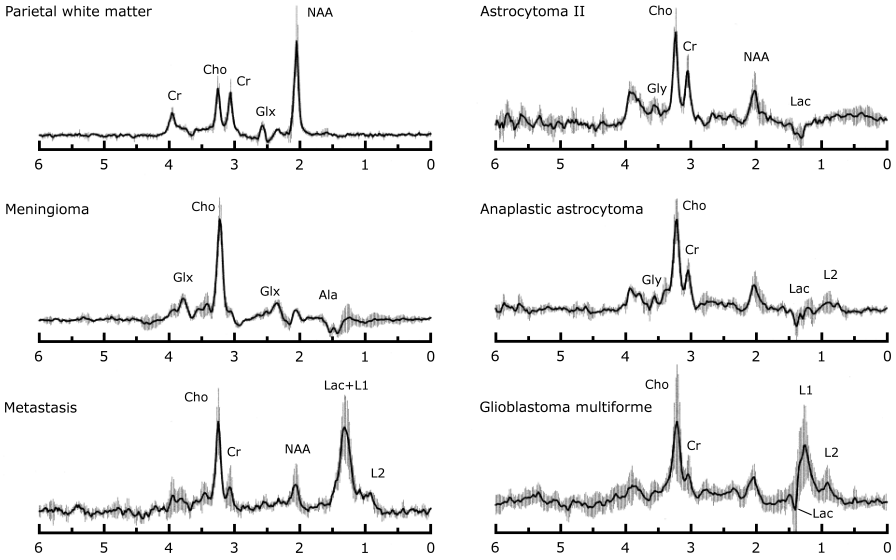


Figure 1.6: *Mean and standard deviations based on n normalized PRESS (TE = 136 ms) ^1H MR spectra for parietal white matter (n = 5), meningioma (n = 4), metastasis (n = 4), astrocytoma grade II (n = 4), anaplastic astrocytoma (n = 6) and glioblastoma (n = 12). Lip1 and Lip2 represent lipids at 0.9 ppm and 1.3 ppm, respectively. Figure adapted from [94].*

nuclei results in a water peak which masks the information from the metabolites of interest. A common method to suppress the water peak is called chemical shift selective (CHESS) [195]. Acquisition of the signal from the selected voxel involves a series of RF pulses. In point-resolved spectroscopy (PRESS) and stimulated echo acquisition mode (STEAM) three consecutive RF pulses are applied such that the signal originates from the selected volume [19, 62].

1.2.3 Magnetic resonance imaging

The basic principles of MR, which are summarized in Section 1.2.1, are widely applied to generate in vivo images of tissues in the human body. A particular application is called MRI and it visualizes the spatial distribution of protons such that morphological information about the examined tissue is obtained. The proton nuclei possess a strong magnetic moment and are present in high abundance in the human body. In particular, the protons of water generate strong signals.

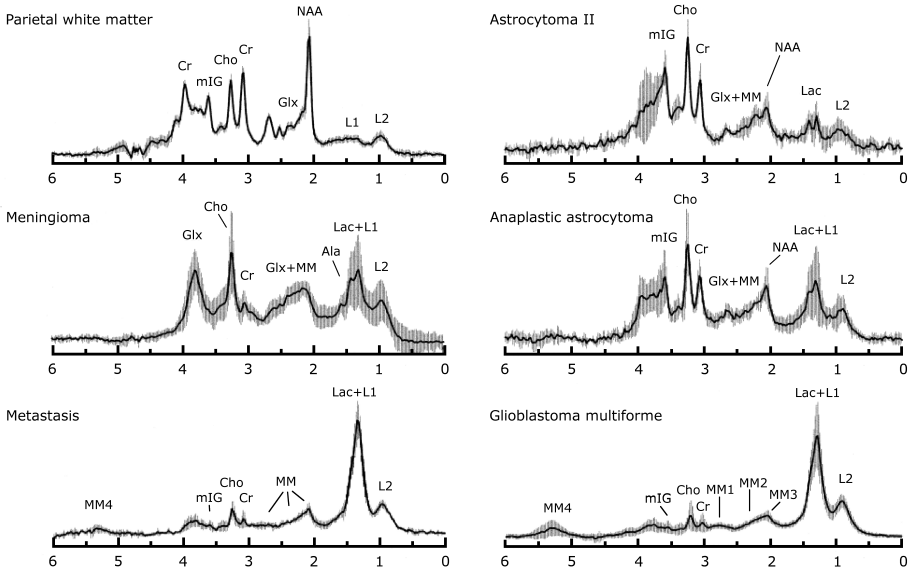


Figure 1.7: *Mean and standard deviations based on n normalized STEAM (TE = 30 ms) ^1H MR spectra for parietal white matter (n = 6), meningioma (n = 8), metastasis (n = 6), astrocytoma grade II (n = 5), anaplastic astrocytoma (n = 7) and glioblastoma (n = 13). Lip1 and Lip2 represent lipids at 0.9 ppm and 1.3 ppm, respectively, while mIG represents myo-inositol + glycine. Figure adapted from [94].*

In order to construct MR images based on the distribution of water molecules, the identification of the spatial origin of the different signal components is required. Using the basic MR principle from Section 1.2.1 would lead to an acquired signal that corresponds to the global response of all spins, lacking information about the spatial distribution. Therefore, gradients are introduced in the system, corresponding to a linear variation in the magnetic field as a function of distance. A different strength of the magnetic field is created for each point inside the magnet. The frequency of each spin is different and characteristic for its position, allowing the identification of the positions of the nuclei. In this way, it is possible to generate high-resolution MR images in any direction without replacing the patient.

The ability to generate images with various contrasts makes MRI a very flexible medical imaging technique (Figure 1.8). Depending on the relaxation processes (i.e. T_1, T_2) and the amount of excited nuclei, some structures appear more bright while others appear more dark. Careful choice of the TE and TR enables to visualize these contrast differences. For instance, T_1-weighted MR images are based on a

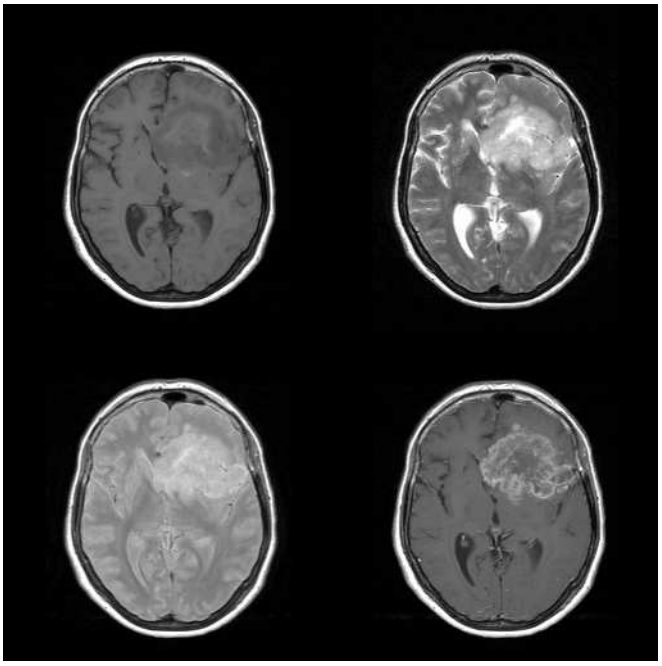


Figure 1.8: *Brain MRI data of a patient with glioblastoma multiforme, acquired on a 1.5 T Siemens scanner. From left to right and top to bottom: T_1-weighted MR image (TE/TR = 15/644 ms), T_2-weighted MR image (TE/TR = 98/3100 ms), proton density-weighted MR image (TE/TR = 16/3100 ms), and gadolinium-enhanced T_1-weighted MR image (15 ml 0.5M Gd-DTPA).*

short TR and a short TE and this results in bright tissues, having short T_1 times, and dark tissues, having long T_1 times. The T_1 contrast follows the order of the T_1 relaxation differences. T_2-weighted MR images originate from a longer TE and TR, while proton density (PD) images are based on a long TR and a short TE. In addition, contrast agents like Gadolinium (Gd) may be used to improve the sensitivity and specificity.

1.2.4 Magnetic resonance spectroscopic imaging

While single-voxel MR spectroscopy provides a signal from a certain volume, MRSI, also called chemical shift imaging (CSI), is able to acquire multiple signals from a grid of voxels (Figure 1.9). MRSI is a powerful technique since it allows to visualize the metabolic state of different regions. The MRSI grid can be two-dimensional

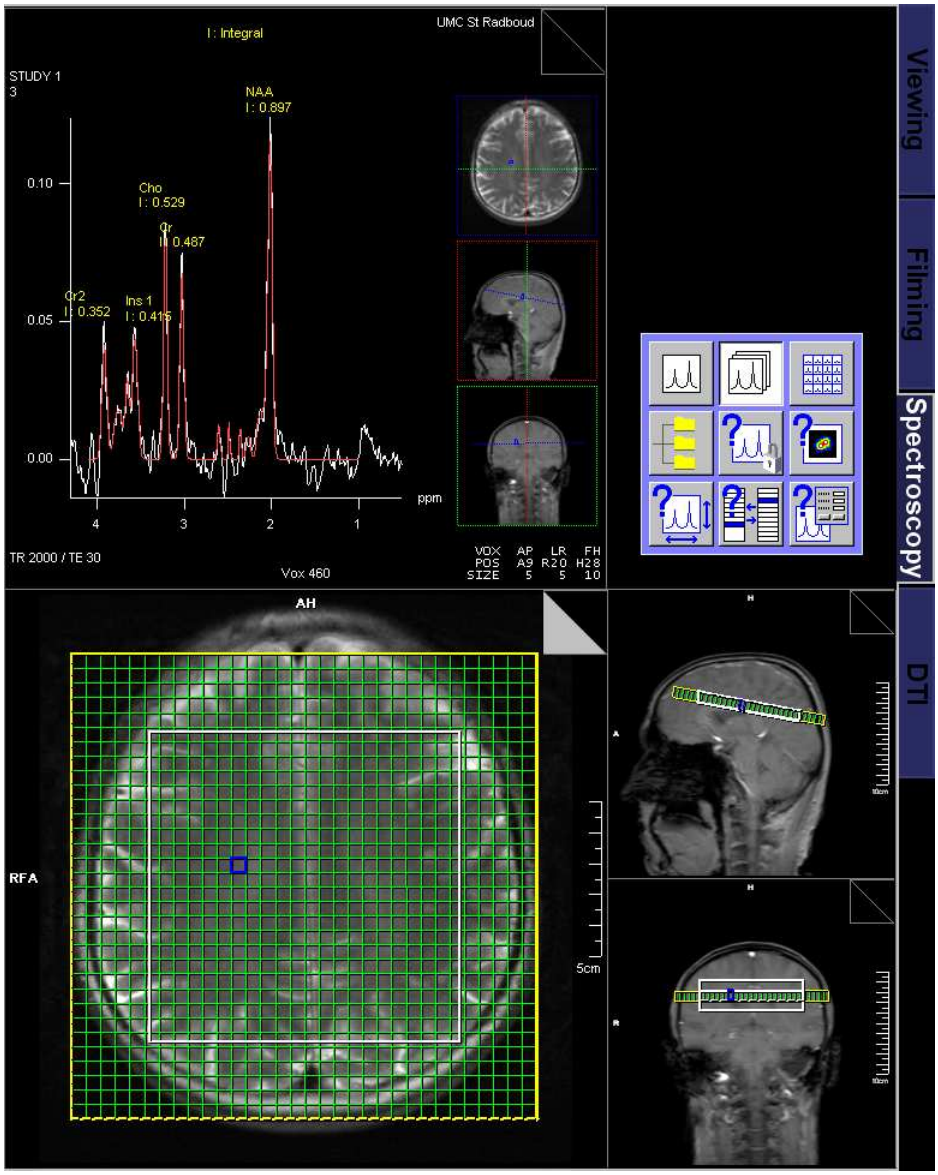


Figure 1.9: *Scanner screenshot of in vivo brain MRSI data acquired with TR = 2000 ms and TE = 30 ms on a 3 T Siemens scanner. The MRSI grid, with dimensions 32 × 32, is superimposed to the T_2-weighted MR image. The MR spectrum from the selected voxel, originating from normal tissue, is visualized.*

or three-dimensional with for example dimensions 16 × 16 (× 16) or 32 × 32 (× 32). Therefore, the acquired FIDs have to be encoded according to their spatial position in the grid. Spatial encoding is performed by the use of field gradients in the spatial directions of the grid, resulting in phase encoded FIDs. The resolution of the MRSI data is limited due to the intrinsic low SNR of the signals, requiring adequately large voxel volumes. The spatial resolution of MRSI is lower than the one of MRI.

1.3 Brain tumors and magnetic resonance

1.3.1 Epidemiology

According to the 2009 statistical report of the Central Brain Tumor Registry of the United States (CBTRUS, http://www.cbtrus.org/ , accessed on June 27, 2009) 98,990 new primary brain and central nervous system (CNS) tumors were diagnosed in the United States from 2004 to 2005. The overall incidence rate for primary brain and CNS tumors was 18.16 per 100,000 for 2004−2005. At the time of diagnosis 7 % of the cases were in individuals less than 20 years of age and 93 % were in individuals 20 years of age or older. The overall incidence rate was 4.58 per 100,000 for children 0−19 years of age (4.47 per 100,000 for children less than 15 years) and 23.62 per 100,000 for adults (20 years of age or older). Males accounted for 44 % of the cases and females for 56 % of the cases. Incidence rates for all primary brain and CNS tumors combined were higher among females (19.16 per 100,000) than males (17.08 per 100,000). Metastatic brain tumors are a common consequence of cancer, occurring in 10 to 20 % of all cancer patients. Among adults, the most common primary sites of brain metastases are lung cancer, breast cancer and melanoma. The number of Americans with brain metastases is estimated to be greater than 150,000 per year [113, 162, 55].

CBTRUS reported that the majority of the tumors (32 %) were located in the meninges, while 24 % of the tumors were situated within the frontal, temporal, parietal, and occipital lobes of the brain. Cerebellum, cerebrum, ventricle, and brain stem tumors accounted for 3 %, 2 %, 1 %, and 2 % of all tumors, respectively. The spinal cord/cauda equina and the cranial nerves accounted for 4 % and 7 % of all tumors, respectively. The pituitary and pineal glands accounted for about 14 % of tumors. Olfactory tumors of the nasal cavity accounted for less than 1 % of all tumors.

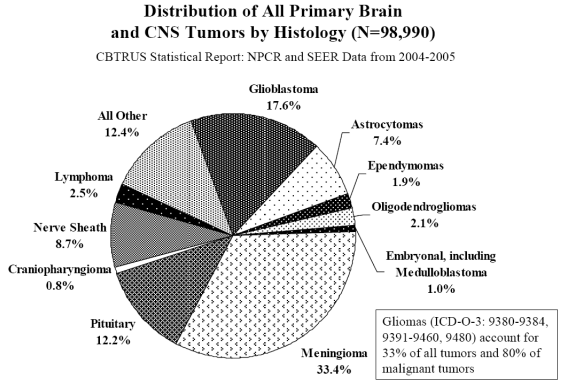
The most frequently reported histology for primary brain tumors was a predominantly nonmalignant (i.e. benign) tumor, meningioma, accounting for over 33 % of all tumors (Figure 1.10). The second most common tumor was glioblastoma (18 %), being an aggressive tumor type, while astrocytomas accounted for 7 % of the cases. The predominately nonmalignant pituitary and nerve sheath tumors accounted for 12 % and 9 % of all tumors, respectively. Acoustic neuromas account for 60 % of all nerve sheath tumors. Glioblastoma accounted for the majority of gliomas, while astrocytoma and glioblastoma combined account for over three quarters of all gliomas. Different histologies were characterized by different age distributions.

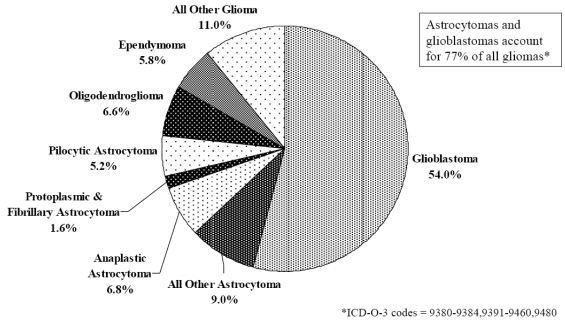
The American Cancer Society (http://www.cancer.org , accessed on June 27, 2009) estimated that 22,070 malignant tumors of the brain or spinal cord (12,010 and 10,060 in men and women, respectively) will be diagnosed during 2009 in the United States. Approximately 12,920 people (7,330 men and 5,590 women) will die from these tumors in 2009. This would account for about 1.5 % of all cancers and 2.3 % of all expected cancer-related deaths in 2008. In children, brain and spinal cord tumors are the second most common cancers. These tumors account for about 21 % of all childhood cancers. Around 3,400 CNS tumors are diagnosed each year in children under the age of 20 and around one fourth of these are considered benign tumors.

1.3.2 Diagnosis

Symptoms of brain tumors, ranging from headache to stroke, may differ among people since they depend on the size and position of the tumor in the brain and the functions that are controlled by that part of the brain. The National Brain Tumor Society (http://www.braintumor.org/ , accessed on June 27, 2009) lists the following symptoms: a new seizure in an adult, gradual loss of movement or sensation in arms or legs, unsteadiness or imbalance (especially if it is associated with headache), loss of vision in one or both eyes (especially if the vision loss is more peripheral), double vision (especially if it is associated with headache), hearing loss with or without dizziness and speech difficulty of gradual onset. In addition, nausea or vomiting, confusion and disorientation, and memory loss may be observed.

Brain tumor diagnosis usually involves several steps. The personal and family medical history is reviewed and a complete physical examination is performed to verify the general signs of health, including signs of disease. A neurological examination commonly evaluates the mental status and it involves testing muscle





Most Common Brain and CNS Tumors by Age

CBTRUS Statistical Report: NPCR and SEER Data from 2004-2005

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

Figure 1.10: *Summary statistics of the 2009 statistical report of the Central Brain Tumor Registry of the United States (CBTRUS). Distribution of all primary brain and central nervous system tumors diagnosed in 2004−2005, all gliomas by histology subtypes and most common brain and central nervous system tumors by age category. Figures taken from the 2009 CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004−2005.*

strength, eye and mouth movement, reflexes, alertness, coordination and other functions. Apart from tumor marker tests, gene testing and lumbar puncture, imaging tests are usually performed. MRI and CT scans are used most often to diagnose brain diseases. Unlike MRI, a CT scan is based on X-rays. The technique provides a series of detailed pictures, which are taken from different angles. A contrast dye may be injected into the veins or swallowed to show more clear differences in the tissues of the brain. During an MRI acquisition, detailed images of the brain are obtained, and contrast agents are used. MR images of brain tumors are in general superior to CT scan images due to the superior soft tissue contrast in MRI. In addition, functional imaging data, identifying functional pathways (e.g. motor, speech, vision), and MR spectroscopy data, providing metabolic information, can be obtained together with the MRI acquisition.

Conventional X-ray methods can also be used for the visualization of tumor vascularization. An angiogram or arteriogram is a series of X-rays taken after a special dye, which flows through the blood vessels of the brain, is injected into an artery. The X-rays provide information about the flow of blood in the brain and can show the tumor and connecting blood vessels.

PET scans provide information about brain activity by measuring levels of injected glucose (Glc) derivatives or amino acids (e.g. methionine). These compounds contain radioactive isotopes (e.g. 18F-fluoro-deoxy-Glc) and are called radioactive tracers. Due to different uptake/metabolism between normal and tumor tissue, contrast is generated in PET images. PET scans are often used to detect the difference between scar tissue, recurring tumor cells and necrosis.

The definite diagnosis of a brain tumor can only be confirmed by histological examination of tumor tissue samples obtained by means of a brain biopsy. Based on microscopic examination of the removed tissue, a pathologist or a neuropathologist determines if the tumor is benign or malignant and what the exact tumor type is. Although histological examination is essential for determining the appropriate treatment and the correct prognosis, it is not always performed, especially when the tumor is located in a part of the brain that would make it hard to biopsy (e.g. brain stem). The two main types of biopsies for brain tumors are stereotactic (needle) biopsy and open surgery. The first is used when the risks of surgery might be too high, which is for instance the case when the tumor is located in vital areas or when it is likely that it cannot be treated. An incision is made in the scalp and a small hole is drilled in the skull. A neurosurgeon uses a hollow needle to remove a small piece of tissue from the tumor. In case of open surgery, small samples of the tumor are immediately examined by the pathologist.

1.3.3 Treatment

Treatment of brain tumors depends on a number of factors such as the type, location and size of the tumor. Other factors that are taken into account are the age and general health of the patient. Several types of treatment may be used for CNS tumors, including surgery, radiotherapy, chemotherapy, and currently developed innovative targeted therapies. In many cases combination therapy is used.

If the location of a brain tumor is accessible for surgery, the neurosurgeon may remove as much as possible of the tumor. Removal of even a portion of the brain tumor may help to reduce signs of elevated intracranial pressure. In addition, surgery can reduce the amount of tumor that needs to be treated by radiation or chemotherapy. Craniotomy is the main type of operation for brain tumor treatment. It involves making an incision in the scalp, folding back the skin and removing the piece of bone over the tumor. The neurosurgeon can remove the tumor or part of it (i.e. debulking) in different ways depending on the amount of blood vessels and how hard or soft it is. Usually the removed piece of bone is brought back to its place and fastened to the skull with metal screws and plates, wires, or stitches. The risks of surgery might be too high for certain tumors when they are located near vital areas in the brain and some tumors that tend to spread diffusely are also not cured by surgery.

Radiotherapy or radiation therapy involves the use of ionizing radiation for cancer treatment to control malignant cells. High-energy rays or particles are often used when surgery is not possible, but it can also be used to treat remaining tumor tissue after surgery. Different approaches exist for radiation therapy. In general, external radiation treatments are fractionated into many small doses and are given over a period of time (e.g. five days a week for several weeks). This schedule depends on the age of the patient and the type and size of the tumor. External radiation may be directed just to the tumor, the surrounding tissue or the entire brain. In the latter case, the patient often receives an extra dose of radiation to the tumor area. Radiation can also be placed inside the body itself. Interstitial radiation therapy or brachytherapy, involves surgically implanting radioactive material directly inside the tumor. The implants loose a certain amount of radioactivity each day. Depending on the material used, the implant may be left in the brain for a short time or permanently. Another way to treat brain tumors is stereotactic radiosurgery which involves a single, high dose of radiation that is delivered to the tumor, minimizing damage to the surrounding brain tissue. High-energy rays are aimed at the tumor from many angles. Usually, patients receive just one treatment. Side effects of radiation therapy may depend on the dose of the radiation, the length of treatment, and the location and size of

the tumor. On the short-term side effects of radiation include fatigue, hair loss, scalp irritation, muffled hearing, and short-term memory loss, whereas long-term side effects include short-term memory loss, radiation necrosis, or neurological impairments associated with the area of brain most radiated.

Chemotherapy represents the use of drugs, which are usually taken orally or intravenously, to treat disease. These drugs interfere with the quickly dividing tumor cells to prevent growth of the tumor. Compared to surgery and radiation therapy, chemotherapy kills cancer cells in the whole body. In this way tumors that have metastasized or spread from other parts of the body can be treated. Chemotherapy is given in cycles, which involves days of treatment followed by recovery periods depending on the drug. A combination of drugs can reduce the chance that the cancer may become resistant to any of the chemotherapy drugs. These decisions depend on the type, location and size of the tumor. Since chemotherapy attacks various (i.e. both good and bad) quickly dividing cells, it can cause hair loss, mouth sores, loss of appetite, nausea and vomiting. Another type of chemotherapy can be placed in the brain during surgery. When (partially) removing the tumor, the neurosurgeon may place disk-shaped wafer implants, which release a chemotherapy drug, in the void left by the tumor. A concentrated dose can be delivered directly to the tumor without increasing side effects.

1.3.4 Classification

Brain tumors are classified based on the type and location of the cells that originate from the tumor, while the degree of malignancy is characterized by the grade. The grade of a tumor is based on how abnormal the cancer cells are and how quickly the tumor is likely to grow and spread. Grade I tumors grow slowly and rarely spread to other tissues. The cells are relatively similar to normal cells and the tumor may be removed by surgery. Grade II tumors grow slowly, but may spread to nearby tissue and may recur. Some of these tumors may become higher-grade tumors. Grade III tumors grow quickly and are likely to spread into nearby tissue. The tumor cells look very different from normal cells. Grade IV tumors grow and spread very quickly. The cells are completely different from normal cells and there may be areas of dead cells. A detailed histological and genetic typing of tumors of the CNS is provided by the World Health Organization (WHO) [133]. Most of the cases that are used in this thesis belong to the class of meningiomas, metastases or gliomas.

Meningiomas grow from the meninges, which are the layers of tissue covering the brain and spinal cord. Meningioma is the most common brain tumor in

adults and it is approximately twice as common in women as in men. As they grow, meningiomas can press on adjacent brain tissue or the spinal cord, causing symptoms as headache or weakness in an arm or leg. Most meningiomas are benign (i.e. grade I), tend to grow slowly, and can be cured by surgery. A small amount (about 20 %) of meningiomas is atypical (i.e. grade II) or malignant (i.e. grade III). The latter may recur or even spread to other parts of the body.

Gliomas comprise a group of tumors that start in glial cells. This category includes astrocytomas, oligodendrogliomas and ependymomas. Ependymomas arise from the ependyma, which is a tissue of the central nervous system, astrocytomas are derived from astrocytes, oligodendrogliomas are derived from oligodendroglial cells, while mixed gliomas are derived from astrocytes and oligodendroglial elements. These cells provide support and nutrition for the nervous system. Patients with benign gliomas may survive for many years [99, 171]. Glioblastoma multiforme is the most common type of glioma (i.e. grade IV) and median survival is less than one year from the time of diagnosis [126]. It grows rapidly and commonly spreads to nearby tissue. This tumor type may have evolved from a lower grade glioma. Standard treatment is surgery followed by radiotherapy or a combination of radiotherapy and chemotherapy.

Metastatic, or secondary, brain tumors are the most common brain tumors. These tumors originate from cancer in another part of the body, known as the primary site. Cancer cells may get from their primary site to the brain by the blood, or may spread from adjacent tissue. Among adults, the most common primary sites of brain metastases are lung cancer, breast cancer, colon cancer, kidney cancer and melanoma. Metastatic tumors are as malignant as the cancerous primary tumor from which it came and most grow quickly. Sometimes multiple metastatic tumors are formed and can be found in different areas of the brain. Metastases are common among middle-aged and elderly men and women and median survival ranges from 2 to 16 months [164, 52]. Routine treatment includes whole-brain radiotherapy, but may also comprise surgery and chemotherapy.

1.3.5 MRI, MRS and MRSI for brain tumor diagnosis and treatment

Since proton is a very sensitive nucleus that is present in many molecules, it is very attractive for medical application. Therefore, this thesis focuses on the use of ^{1}H for in vivo MR imaging and MR spectroscopy. Although the main interest of this research is the application to brain tumors, these techniques can be used for a wide range of medical purposes. Research applications include the study of prostate tumors, neurologic distortions (e.g. epilepsy), acquired immune deficiency

syndrome (AIDS) and Alzheimer disease [102, 33, 244, 54].

MRI has become one of the most important noninvasive aids in the clinical diagnosis of brain tumors. The technique allows to assess the anatomical location, morphology and extent of the tumor with a high spatial resolution. Stacked images are acquired throughout the whole brain for the purpose of tumor localization. MR images with different contrasts are acquired and contrast agents are used (e.g. Gd). Contrast agents allow to assess the integrity of the blood-brain barrier. Brain tumors often result in a disruption of the blood-brain barrier and contrast agent leaks out of the vasculature into brain or tumor tissue, finally, giving rise to abnormal enhancement. The use of Gd tends to modify signal intensities by shortening the T_1 relaxation time in regions to which Gd has access. Normally, tumor is hypointense on a T_1-weighted MR image and hyperintense on a T_2-weighted and PD-weighted MR image, while contrast enhancement usually indicates higher grade tissue for gliomas. Cerebrospinal fluid (CSF) is hypointense on a T_1-weighted MR image, hyperintense on a T_2-weighted MR image, and isointense on a PD-weighted MR image. Necrosis is usually hypointense and shows no contrast enhancement. Although this thesis is restricted to conventional, static MRI, techniques like diffusion and perfusion MR data can provide additional information about brain tissue. Water movement can be analyzed using diffusion-weighted MRI. Abnormality is detected due to restricted water movement. The vascularity of the tissue is examined using perfusion-weighted MRI, allowing to detect alterations in the blood volume or circulation.

Numerous studies have already demonstrated that MRS and MRSI acquisitions allow extraction of metabolite information that is important for the investigation of brain tumors [182, 160]. Metabolite concentrations, which are correlated with the area under the peaks in the MR spectrum, have found to be elevated or decreased in tumor tissue as compared to normal tissue. In addition, the levels of these metabolite allow to discriminate between different types and grades of tumor (Figures 1.6 − 1.7). Relevant metabolites include for example Cho, Cr, NAA, mI, Glu, Gln, glycine (Gly), lactate (Lac), lipids (Lip), taurine (Tau) and alanine (Ala).

As NAA, a singlet at 2.03 ppm, is known to be a marker for neuronal density, reduced levels may reflect neural loss or disfunction [159]. In particular in tumor tissue, the neuronal cells are being replaced by tumor cells [66]. The concentrations of NAA have been found to be equal or greater in gray matter compared to white matter [196]. NAA can be used to distinguish between different grades of gliomas [211]. Total Cho (tCho), around 3.22 ppm, includes contributions from free Cho, glycerophosphocholine (GPCho) and phosphocholine (PCho). Free Cho is involved in the synthesis of the neurotransmitter acetylcholine, as well as

membrane synthesis, PCho is involved in the synthesis of the insoluble membrane phospholipids, while GPCho is a product of membrane degradation [196, 159]. Increased levels of tCho are associated with an increase in membrane breakdown or turnover, myelination or inflammation and has been observed in demyelinating diseases and brain tumors [196]. While NAA correlates with neuronal density, tCho is thought to be associated with cellular density. Total Cr (tCr), including Cr and phosphocreatine (PCr), corresponds to singlets at 3.03 ppm and 3.9 ppm. The metabolite is associated with the energetic status of tissue and may be reduced in some regions of the tumor [159]. The concentration of tCr is reported to be higher in gray matter than in white matter [196]. The concentrations of mI, having multiplets at 3.56 ppm, are normally elevated in newborns, but rapidly decrease afterwards [221]. The metabolite is reported to be located in astrocytes [197]. It is a precursor for inositol lipid synthesis and is a constituent of membrane Lip [196]. High concentrations are reported for low-grade gliomas, while mI is low or absent in meningiomas [94]. Gly is a amino acid that is involved in various metabolic pathways. Its concentration is well regulated in the brain, being primarily synthesized from Glc through serine. Elevated levels of Gly have been reported in glioblastomas and glycine may represent a marker for tumor malignancy [111, 221]. Glutamate and glutamine (Glx), at 3.8 ppm and between 2.0 and 2.4 ppm, are the most abundant amino acids in the brain and are released by 90 % of excitatory neurons [196]. Glx resonances are detected in all highly malignant tumors. Lac, which is the end product of anaerobic glucolysis, is usually detected under pathologic conditions, when the energy metabolism is severely affected [66]. Higher concentrations of Lac, having a doublet around 1.33 ppm, are reported for higher grade tumors, but it is not always detected in glioblastomas and metastases [56]. Mobile Lip, around 1.3 and 0.9 ppm, are thought to correspond to cellular and membrane breakdown, associated with necrosis [159]. Lip and MMs are characteristic for high-grade tumors at short TE [10, 94]. Unless subcutaneous fat is included, normal brain contains very low levels of Lip using in vivo MRS. Tau, having multiplets around 3.42 ppm, is a β-amino acid that is associated with osmoregulation and modulation of the action of neurotransmitters [87]. High concentrations are detected at the time of birth and the metabolite decreases with age [79]. Ala, having a doublet at 1.47 ppm, is a nonessential amino acid that contains a methyl group. Ala is elevated in gliomas and meningiomas [148, 175, 221].

Although single-voxel MRS provides important information about the metabolism, it ignores the spatial heterogeneity of the lesion [158]. On the other hand, based on MRSI, it is possible to visualize the heterogeneity of tissue and to define the spatial extent of the lesion. It permits to inspect the surrounding tissue that may appear as normal on MRI, but shows abnormal metabolism. These factors are known to be important for planning focal treatments such as radiation and surgical resection. Therefore, MRSI is considered to provide unique information,

that is complementary to MRI, for brain tumor diagnosis and treatment.

1.4 Aims of the thesis

The aim of this thesis is to develop methods that process brain MRS and MRSI data and generate output, that is immediately interpretable in a straightforward way, to support clinicians in decision making. Developing a clinical decision support system (DSS) for brain tumor diagnosis using MRS and MRSI data involves a preprocessing step. Relevant information needs to be extracted from the acquired data and inference, evaluation and integration of predictive models (i.e. classifiers) are needed to discriminate among tumor types and grades. This thesis focuses on the validation and evaluation of different strategies to extract features from MR spectra. The next, and main, goal of the thesis is the development of classifiers and classification methods for differential diagnosis of brain tumors. This involves building multi-class classifiers that provide class probabilities. A large amount of classifiers is created based on an extensive repository of patient MR data and evaluated on an independent, newly acquired, repository of patient MR data. An important aspect of the thesis is the combination of different sources of data, in particular MRI and MRSI, for improved diagnosis. Extensive attention goes to the visualization of the obtained classification results based on nosologic images. The important problem of discrimination between glioblastoma multiforme and solitary metastasis is tackled based on the combination of MRI and MRSI. Finally, the integration of the preprocessing techniques and predictive models into software tools is addressed.

Part of the research of this thesis originates from active participation within the HealthAgents project (*Agent-based distributed decision support system for brain tumour diagnosis and prognosis*, 2006−2008, http://www.healthagents.net/ , accessed on June 27, 2009) and the eTUMOUR project(*Web accessible MR decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data*, 2004−2009, http://www.etumour.net/ , accessed on June 27, 2009) [78]. These projects aimed to develop DSSs to assist clinicians in brain tumor diagnosis based on various sources of data (e.g. MRI, MRS, MRSI, high-resolution magic angle spinning (HR-MAS), deoxyribonucleic acid (DNA) microarrays and clinical information). My responsibilities, as the representative for Katholieke Universiteit Leuven, within these projects were related to the development and integration of pattern recognition tools (e.g. classifiers) for brain tumor diagnosis based on MRI, MRS and MRSI.

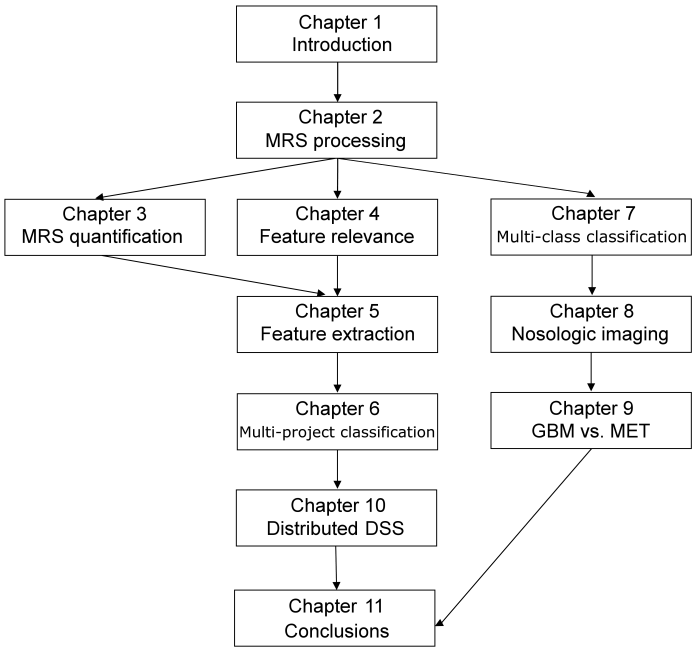


Figure 1.11: *Flowchart of the thesis.*

1.5 Organization of the thesis

This thesis mainly consists of four parts. Chapters 1 and 2 explain the necessary steps to create a brain tumor classification system based on MR techniques: data acquisition, preprocessing of MR data, classifier construction and evaluation. Chapters 3, 4 and 5 mostly deal with feature extraction from MR spectra, while chapters 6, 7, 8 and 9 focus on classification of brain tumors based on MRS, MRSI or MRI. Chapter 10 is devoted to the design of a DSS (Figure 1.11).

Chapter 1 explains the basic principles of MR and introduces the different types of data that are used in this thesis. The characteristics of MRI, MRS and MRSI are outlined by emphasizing on the differences, explaining their advantages and limitations. Important concepts as T_1, T_2, PD, TR, TE, which are used all along the thesis, are addressed in this chapter. Background information (summary statistics, diagnostics, treatment and tumor types) about brain tumors is provided and the potential of MRI, MRS and MRSI for brain tumor diagnosis is covered. The chapter concludes with the aims and structure of the thesis.

Chapter 2 provides a basis for the remaining chapters of the thesis. It gives an overview of the different preprocessing steps (e.g. filtering, normalization) for MR spectra. The chapter summarizes the different methods from statistics and machine learning that are used for classification in the thesis. The aspects of model evaluation are discussed.

Chapter 3 briefly reviews some of the most important methods for obtaining metabolite concentrations from MRS or MRSI. Peak integration, linear combination of model spectra (LCModel), quantitation based on quantum estimation (QUEST) and automated quantitation of short echo time MRS spectra (AQSES) are presented. Together with the methods from Chapter 4, these techniques are used in Chapter 5.

Chapter 4 is devoted to the evaluation of the relevance of MRS, MRSI and MRI features selected by automated feature selection techniques for differential diagnostic purposes. In contrast to the feature extraction methods from Chapter 3, the feature selection methods of this chapter are based on a pattern recognition approach. The selected features have been evaluated using literature and clinical experience.

Chapter 5 studies the effect of different feature extraction approaches for brain tumor classification based on MRSI and multi-center MRS data. The chapter combines the quantification approaches from Chapter 3 with some of the pattern recognition approaches from Chapter 4. The chapter compares these methodologies with the ultimate goal to obtain the best performing classification models.

Chapter 6 presents the results of brain tumor classification based on MRS data. These results have been obtained within the context of the eTUMOUR project. This project provided the unique opportunity to construct classifiers using a large repository of, previously acquired, multi-center data and evaluating them on subsequently acquired multi-center data.

Chapter 7 investigates the use of automated pattern recognition methods on MRI and MRSI data with the ultimate goal to assist clinicians in the diagnosis of brain tumors. Previous work that only uses binary classifiers to assess the type and grade of a tumor is extended to a multi-class classification system, resulting in class probabilities.

Chapter 8 presents a new technique to create nosologic images of the brain based on MRI and MRSI data. The proposed technique applies advanced methods from image processing as well as pattern recognition to segment and classify brain tumors. A registered brain atlas and a subject-specific abnormal tissue prior, obtained from MRSI data, are used for the segmentation. The detected abnormal tissue is classified based on supervised pattern recognition methods, as explained in Chapters 2 and 7.

Chapter 9 describes the combined use of MRI and MRSI to differentiate glioblastoma multiforme and solitary metastases. Metabolite ratios from the enhancing tumor, peri-enhancing region and the non-enhancing/non-T_2 hyperintense area are compared. The infiltrative nature of glioblastoma is visualized and nosologic images of the brain are created based on the technique explained in Chapter 8.

Chapter 10 presents research that has been performed within the context of the HealthAgents project. A framework for providing automated classification functionality has been developed and integrated in the HealthAgents DSS. The combination of the classification framework, which includes classifiers developed in Chapter 6, with a distributed agent system has been the basis of the brain tumor DSS developed in HealthAgents.

Chapter 11 summarizes the work presented in the previous chapters and formulates the main conclusions of the thesis. Important topics for future research are described.

1.6 Personal contributions

The research in chapters 4, 5, 7, 8 and 9 is mainly my own work and based on my own ideas. Chapter 4 and Chapter 5 focus on feature extraction from MRI, MRS and MRSI data. My contribution included the design and implementation of the experimental studies, as well as the analysis. Interpretation of the results in Chapter 4 was in collaboration with University Medical Center Nijmegen (UMCN) and Radboud Universiteit Nijmegen (RUN) and a paper about this research has been submitted to *Computers in Biology and Medicine* [177]. The study in Chapter 5 was published in the international journal *Magnetic Resonance in Medicine* [142]. Chapter 6 describes an extensive study that was performed by the pattern recognition work package within the eTUMOUR project. My contribution to this study, which involved more than 30 people in total, was the development of an extensive number of MRS classifiers, which

also included feature extraction, model performance evaluation and statistical analysis. This multi-project and multi-center research was published in the international journal *Magnetic Resonance Materials in Physics, Biology and Medicine* [70]. The importance of my contribution is reflected in my position as a second author on the paper. My contribution to Chapter 7, dealing with multi-class classification of MRSI data, is the experimental design, implementation and analysis. The study was presented in the international journal *Artificial Intelligence in Medicine* [138]. The nosologic imaging method in Chapter 8 is of my own design, as well as the experimental studies. Analysis and discussion of the results was in collaboration with UMCN. A paper about this topic was published in the international journal *NMR in Biomedicine* [139]. Chapter 9 reports a study that focuses on the differentiation of glioblastoma multiforme and solitary metastases. My contribution was related to the design, implementation and data analysis. Experimental results were discussed in collaboration with Universitat de Valencia-Estudi General (UVEG). The study was published in the conference proceedings of the *22nd IEEE International Symposium on Computer-Based Medical Systems (CBMS)* [140] and a paper about this study has been submitted to the international journal *Neuroradiology* [141]. Chapter 10 provides an overview of research that was performed within the context of the HealthAgents project by the work package that was devoted to the development of the DSS. Being the responsible for Katholieke Universiteit Leuven, my contribution was the implementation and testing of various classifiers and feature extraction methods in Java. A paper was accepted for publication in the international journal *The Knowledge Engineering Review* [198].

1.7 Conclusions

The basic concepts of MR were introduced. This included the physical principle of MR as well as different types of data that could be acquired with the same scanner. Background information about brain tumors was summarized and the potential of MRI, MRS and MRSI for brain tumor diagnosis was discussed. The different goals of the thesis were addressed and a chapter-by-chapter overview, including personal contributions, was provided.

Chapter 2

Processing of MR spectra

While the previous chapter explained the procedure to acquire data, this chapter introduces the foundations for converting the acquired FIDs into classification results in three steps. Section 2.1 outlines the typical procedures for preprocessing the signals. Section 2.2 describes various algorithms for classification of the preprocessed signals. The important issue of evaluating the model performance is discussed in Section 2.3.

2.1 Preprocessing methods

The results of classifiers and classification algorithms are significantly dependent on the procedure that is used to remove irrelevant components from the MRS(I) signals. The different steps that are performed to enhance the key features of the signals, while reducing the influence of the irrelevant information, are called preprocessing. Dominant residual water, artifacts caused by instrumental errors and noise are just a few of the issues that are addressed during preprocessing. A number of important preprocessing steps to improve the quality of the signals is outlined in the following paragraphs.

Zero filling

Often the spectral resolution is too low to resolve the resonances [40]. Increasing the spectral resolution can be achieved by decreasing the spectral width (SW) or increasing the acquisition time. On the other hand, extending the acquisition time can be simulated by extending the acquired FID, which has decayed to zero,

artificially with a set of zeros. The effect of zero filling is an increase in resolution, being equivalent to an interpolation in the frequency domain, which improves spectral visualization.

Apodization

Apodization involves multiplication of the FID with a time domain signal [40]. Depending on the parameters of the time domain signal, the multiplication has a different effect. Typically, an exponentially decaying weighting function is used to increase the SNR of the MR spectrum since the initial points of the FID, which have a larger SNR, get a higher weighting than the points at the end of the FID, which mainly contain noise. The opposite weighting function leads to narrower peaks in the MR spectrum, or a higher spectral resolution. However, the SNR is decreased because of the higher contribution of the points at the end of the FID.

Frequency alignment

The data often suffer from frequency shifts, induced by patient movement for example, which require an appropriate correction. In general, for in vivo MRS(I) data at 1.5 T, as in this thesis, large peaks as NAA, Cho or Cr are used as a reference for alignment. The MR spectra are shifted in the frequency domain and transformed to the time domain using the inverse Fourier transformation.

Phase and lineshape correction

In case an additional water-unsuppressed reference signal is acquired, Klose's method can be carried out to correct for eddy currents [114]. These phenomena, which are caused by instrumental imperfections, can be corrected for by dividing the water-suppressed signal by the phase term of the water-unsuppressed signal. This involves a zero-order phase correction. First-order phase shifts are linearly dependent on the frequency and originate mainly from the delay between the excitation pulse and the first points of the acquisition of the time domain signal. First-order phase correction can be performed to obtain pure absorption spectra. Alternatively, magnitude spectra can be used instead of phased spectra, at the cost of a broadening of the peaks. In addition, lineshape correction methods as quantification improvement by converting lineshapes to the Lorentzian type (QUALITY) and QUALITY and eddy current correction (QUECC) can also correct for distortions in the magnitude of the signal [39, 12]. These methods convert the irregular lineshape to Lorentzian lineshape, which is often used in quantification methods (Chapter 3).

Water filtering

The residual water peak can be removed by subspace-based modeling approaches as Hankel singular value decomposition (HSVD), Hankel Lanczos singular value decomposition (HLSVD) or HLSVD with partial reorthogonalization (HLSVD-PRO) [11, 172, 120]. These methods estimate the water peak and subtract it. For example, the whole signal is fitted as a sum of complex damped exponentials. Afterwards, the components whose frequencies are estimated in the frequency region of the water resonance are subtracted. Other approaches directly use a filter to the signal. Maximum phase finite impulse response (MP-FIR) filters are examples of this category of methods [212]. These filter methods are sometimes embedded in the quantification algorithm [179].

Signal normalization

Normalization compensates for differences between MR spectra due to effects independent of the type of tissue. The process of normalization is even more complicated when the data are acquired by several clinical centers, various scanners or with different acquisition protocols. Usually normalization is performed based on a mathematical norm or the intensity of a particular molecule. Mathematical normalization involves division of each spectral value by the norm (e.g. L_1, L_2) of the MR spectrum [217]. Other studies perform a normalization based on the intensity of a molecule, as for instance Cr [182]. Alternatively, the MR spectrum can be divided by the intensity of water, which is estimated from an MR spectrum without water suppression [225]. This requires an additional measurement of a corresponding water-unsuppressed signal, which is acquired with the same parameters and from the same voxel. Estimation of the intensity of the water peak can be based on the highest value in the time domain, by peak integration in the frequency domain or by modeling using HSVD or HLSVD-PRO for example.

Baseline correction

MR spectra that are acquired with a short TE are characterized by the presence of an underlying and partially unknown baseline, which complicates the assessment of the intensity of the metabolites. In this thesis various methods are used to correct the signal for the baseline. For instance, the baseline can be modeled in a (non)parametric way. This approach is often used within the context of quantification approaches, which aim to estimate the concentration of metabolites [179]. Modeling of the baseline can be done by wavelets, polynomial or spline functions [186, 207]. Another approach represents multiplication of the FID by an exponentially decaying function [203]. Since the baseline consists of broad, fast decaying, components, their main contribution is in the first part of the FID.

The baseline is fitted as the product of the original FID and an exponentially decreasing apodization function, requiring the selection of the decay parameter. Finally, truncation of the initial points of the FID has been proposed in the context of a quantification method [188]. The baseline, being a large component in the MR spectrum, decays very rapidly. Removing the initial part of the FID results in a suppression of the baseline components. However, this procedure may also suppress part of the metabolite contributions and may lead to reduced accuracy of the estimates due to a decreased SNR.

2.2 Classification algorithms

In the machine learning and pattern recognition field, researchers develop methods and computer programs to perform a certain task. In this context, certain learning methods try to optimize a performance criterion using example data or past experience [6, 15]. In case of supervised learning the learning methods are adopted based on these examples (i.e. the learning or training phase), which include output values. When the output values indicate the different categories to which the examples belong, this learning task is called classification. Since the parameters of the classifier are obtained from the example data, they are usually called training data. Furthermore, there exists an important difference between the actual classifier and the classification algorithm [47]. While a classifier is a function or model that assigns an example to a category, the algorithm is the technique that is used to produce the classifier. In the following sections, different algorithms are described, which can produce classifiers for MRI, MRS and MRSI data. The training data set, to be fed to the classification algorithm to construct the classifier, consists of N data points and can be formally denoted as $\{x_i, y_i\}_{i=1}^{N}$, with $x_i \in \mathbb{R}^d$ being an input example of dimension d and $y_i \in \{-1, +1\}$ the corresponding class label for a binary classification problem.

2.2.1 K-nearest neighbor

The k-nearest neighbor (KNN) algorithm is based on a distance function to measure the similarity or difference between data points [50]. In order to determine the class of a new data point x, the distance between x and x_i is computed. A distance function that is often used for this purpose is the Euclidean distance

$$d_i = \sqrt{(x - x_i)^T (x - x_i)}, \ i = 1, ..., N. \tag{2.1}$$

The nearest neighbor rule assigns a new data point x to the class of the data point x_i which is closest to x (Figure 2.1). This approach may produce an unreliable

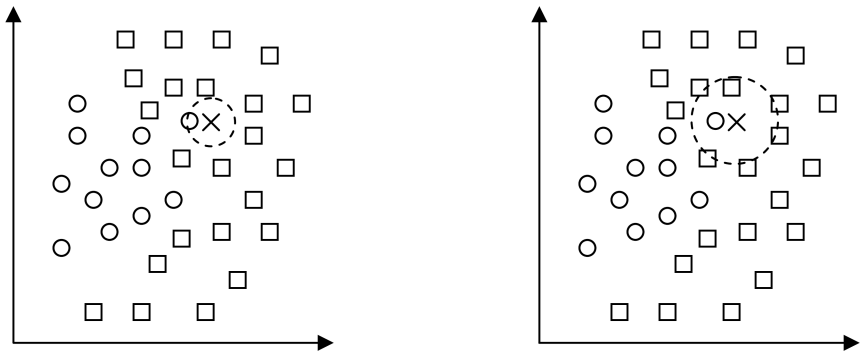


Figure 2.1: *Considering the 7 nearest neighbors instead of only the single nearest neighbor may produce more reliable classification results.*

classification result since the classification rule assigns the new data point based on only a single of the N points. For instance, it may happen to be an outlier that is not representative for its class. Considering k neighbors instead of one, may produce a more reliable result. The KNN algorithm determines the k data points that are the nearest to x and classifies it according to the majority of equal classifications in this group. Data normalization can be used to avoid that the variable with the largest scale dominates the distance measure.

2.2.2 Discriminant analysis

Since Fisher's linear discriminant analysis (FLDA) was introduced in 1936, it has been widely applied for numerous classification problems [59, 50]. The method aims to discriminate between classes in an underlying space by finding a projection (Figure 2.2). The data points are projected

$$z_i = w^T x_i,\ i = 1, ..., N, \tag{2.2}$$

with weights w such that data points from the same class are close to each other and the class means are as far apart as possible. For a binary classification problem the Fisher's linear discriminant is then defined as the linear function $w^T x$ that

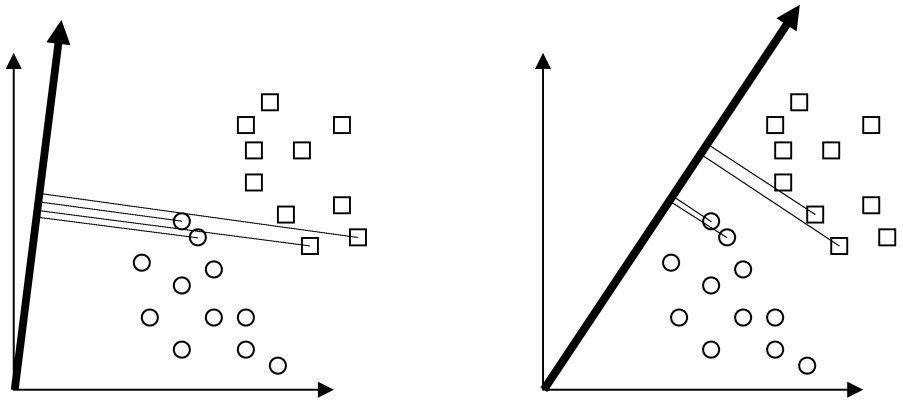


Figure 2.2: *Two projections of the data points for a binary classification problem. The projection on the right provides a better separation between the two classes.*

maximizes the criterion function

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}, \tag{2.3}$$

with class means and within-class scatter

$$\tilde{\mu}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} z_i \quad \text{and} \quad \tilde{\mu}_2 = \frac{1}{N_2}\sum_{i=N_1+1}^{N} z_i, \tag{2.4}$$

$$\mu_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} x_i \quad \text{and} \quad \mu_2 = \frac{1}{N_2}\sum_{i=N_1+1}^{N} x_i, \tag{2.5}$$

$$\tilde{s}_1^2 = \sum_{i=1}^{N_1}(z_i - \tilde{\mu}_1)^2 \quad \text{and} \quad \tilde{s}_2^2 = \sum_{i=N_1+1}^{N}(z_i - \tilde{\mu}_2)^2, \tag{2.6}$$

where it is assumed that the first N_1 data points and the last N_2 data points belong to the first class and second class, respectively [59]. In the original space the criterion function corresponds to

$$J(w) = \frac{w^T S_b w}{w^T S_w w}, \tag{2.7}$$

with the between-class scatter matrix

$$S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T, \tag{2.8}$$

and within-class scatter matrix

$$S_w = S_1 + S_2, \tag{2.9}$$

where

$$S_1 = \sum_{i=1}^{N_1} (x_i - \mu_1)(x_i - \mu_1)^T \quad \text{and} \quad S_2 = \sum_{i=N_1+1}^{N} (x_i - \mu_2)(x_i - \mu_2)^T. \tag{2.10}$$

There exists a natural generalization for the criterion in Equation (2.7) to multi-class classification problems [50]. In case of n_C classes, $n_C - 1$ projection vectors w_j are required which can be arranged as columns into a projection matrix $W = [w_1 w_2 \ ... \ w_{n_C-1}]$. The criterion function becomes

$$J(W) = \frac{W^T S_b W}{W^T S_w W}, \tag{2.11}$$

where S_b and S_w represent the generalized between-scatter matrix and the generalized within-class scatter matrix, respectively. A common way to maximize the criterion in Equation (2.11) is by maximizing the ratio $\frac{\det(W^T S_b W)}{\det(W^T S_w W)}$. If S_w is a nonsingular matrix then the ratio is maximized when the column vectors of the optimal projection matrix, W, are the eigenvectors that correspond to the largest eigenvalues of $S_w^{-1} S_b$. As the rank of S_b is $n_C - 1$ or less, this implies that there are at most $n_C - 1$ nonzero eigenvalues. In order to have a nonsingular

matrix S_w, $d + n_C$ data points are required. This causes problems when dealing with high-dimensional data since a very high amount of points is needed. An additional disadvantage is the high computational complexity of solving $S_w^{-1}S_b$ when working with high-dimensional data.

Quadratic discriminant analysis (QDA) and linear discriminant analysis (LDA) arise as Bayes' rules or maximum likelihood discriminant rules when the data from each class follow a Gaussian distribution. The Bayes' rule is

$$\min_k \mathcal{C}_1(x) = (x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log(|\Sigma_k|) - 2\log(P(y=k)),\ k = 1, ..., n_C, \tag{2.12}$$

where Σ_k denotes the covariance matrix for class k. Using the classification rule in Equation (2.12) results in QDA. Assuming that the covariance matrix of each class is identical, the rule becomes

$$\min_k \mathcal{C}_2(x) = (x-\mu_k)^T\Sigma^{-1}(x-\mu_k) - 2\log(P(y=k)),\ k = 1, ..., n_C, \tag{2.13}$$

for LDA.

2.2.3 Multi-layer perceptrons

The multi-layer perceptron (MLP) is a popular type of artificial neural network (ANN) model, which is known to be a universal approximator [14]. In case of an MLP with one hidden layer the outcome is obtained by feeding the data point $x(\in \mathbb{R}^d)$ to the hidden layer, consisting of H hidden units, which are themselves connected to the outcome unit(s) (Figure 2.3). For hidden neuron h this can be represented by

$$z_h = f(\sum_{j=1}^{d} w_{jh}x_j + b_h),\ h = 1, ..., H, \tag{2.14}$$

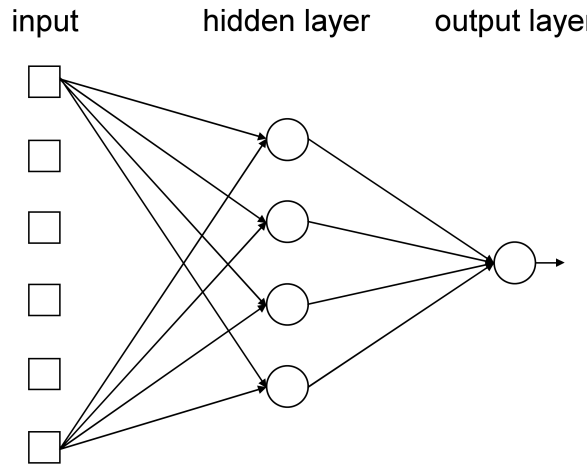


Figure 2.3: *Architecture of an MLP with one hidden layer, containing four hidden units, and one output unit.*

with $f(\cdot)$ an activation function (e.g. tanh), w_{jh} the weight between the input j and hidden neuron h and b_h the bias term of hidden neuron h. The output unit activation $y^*_{k'}$ is modeled as

$$y^*_{k'} = g(\sum_{h=1}^{H} w_{hk'} z_h + b_{k'}), \tag{2.15}$$

where $w_{hk'}$ denotes the weight between hidden neuron h and the output unit, b the bias term of the output unit and $g(\cdot)$ the activation function. For binary classification problems, there is one output unit, $k' = 1$, and $g(\cdot)$ can be represented by a logistic sigmoidal function.

The model parameters consist of all weights and bias terms and can be estimated by minimizing the cross-entropy error function for binary classification. For regularization purposes, an additional term can be added to keep the estimated weights small. There exists a wide variety of methods to optimize the objective function. Typically, these methods involve an iterative minimization of the error function. Each iteration starts with the computation of the derivative of the error function with respect to the weights and bias terms. The backpropagation method is used to propagate the errors backward through the network. Next, the derivatives are used to update the weights and bias terms. Various algorithms can

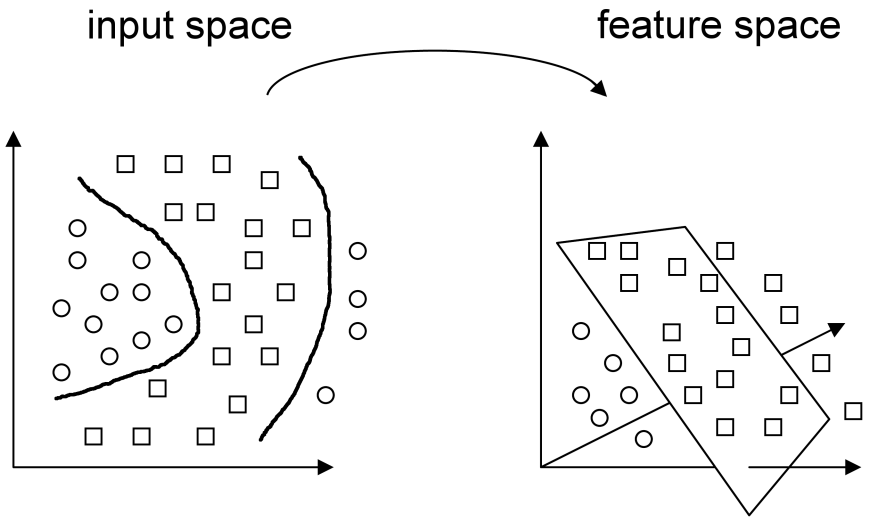


Figure 2.4: *SVMs allow mapping of the data from the input space to a high-dimensional feature space where a linear separation is obtained.*

be used for this purpose: steepest descent, Newton, Levenberg-Marquardt, quasi-Newton and conjugate gradient. It has been proven that MLPs, containing one or more hidden layers, can approximate any continuous nonlinear function arbitrarily well over a compact interval to any degree of accuracy. However, disadvantages of MLPs are the existence of many local minima for the error function and the need to determine the number of hidden units on a trial and error basis.

2.2.4 Support vector machines

Although linear support vector machines (SVMs) originate from almost 50 years ago, a major breakthrough was realized during the early nineties [18]. Originally, the method became popular in the neural networks and machine learning community, and afterwards it was accepted in many other fields. The machine learning technique provides a methodology for (non)linear function estimation and classification problems [234]. SVMs have important advantages compared to other popular methods like for instance ANNs. Contrary to the many local minima in case of ANNs, the training problem is convex for SVMs. In addition, the number of hidden units is automatically obtained for SVMs, in contrast to ANNs.

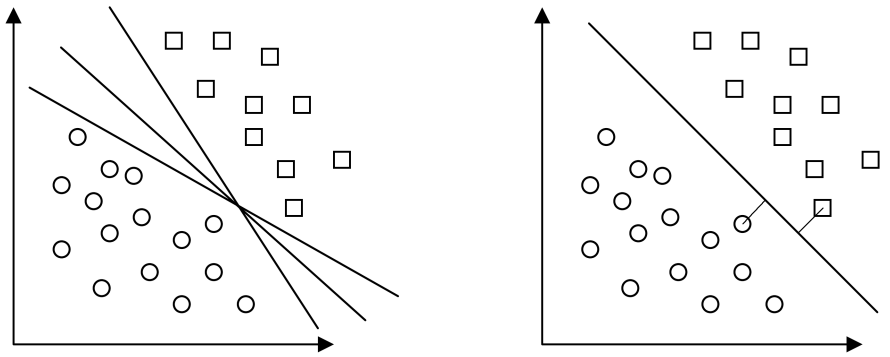


Figure 2.5: *There are different linear separators for the data, but SVMs construct a hyperplane that maximally separates the classes. The data are shown in the input space.*

The key concept of SVMs, which were originally developed for binary classification problems, is the use of hyperplanes to define decision boundaries separating between data points of different classes. SVMs are able to handle both simple, linear, classification tasks, as well as more complex, i.e. nonlinear, classification problems. The idea behind SVMs is to map the original data points from the input space to a high-dimensional, or even infinite-dimensional, feature space such that the classification problem becomes simpler in the feature space (Figure 2.4). SVMs map the d-dimensional input vector x from the input space to the d_h-dimensional feature space using a (non)linear function $\varphi(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^{d_h}$. The separating hyperplane in the feature space is then defined as $w^T\varphi(x) + b = 0$, with bias term $b \in \mathbb{R}$ and w a weighting vector with the same dimension as $\varphi(x)$. A data point x is assigned to the first class if $y(x) = \text{sign}\left(w^T\varphi(x) + b\right)$ equals -1 or to the second class if $y(x)$ equals $+1$.

In case the data are linearly separable, the separating hyperplane can be defined in many ways (Figure 2.5). However, SVMs are based on the maximal margin principle, and aim at constructing a hyperplane with maximal distance between the two classes. The SVM classifier starts from the following formulations

$$w^T\varphi(x_i) + b \leq -1 \text{ for } y_i = -1, \tag{2.16}$$

$$w^T\varphi(x_i) + b \geq +1 \text{ for } y_i = +1, \tag{2.17}$$

equivalent to

$$y_i \left(w^T \varphi(x_i) + b \right) \geq 1, i = 1, ..., N, \tag{2.18}$$

and the classifier is written as

$$y(x) = \text{sign} \left(w^T \varphi(x) + b \right). \tag{2.19}$$

However, in most real life applications data of both classes are overlapping, which makes a perfect linear separation impossible. Therefore, a restricted number of misclassifications should be tolerated around the margin. The resulting optimization problem, where violation of the constraints is penalized, for SVMs is written as

$$\min_{w,\xi,b} \mathcal{J}_1(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i, \tag{2.20}$$

such that

$$y_i \left(w^T \varphi(x_i) + b \right) \geq 1 - \xi_i, \quad i = 1, ..., N, \tag{2.21}$$

$$\xi_i \geq 0, \quad i = 1, ..., N, \tag{2.22}$$

where C is a positive regularization constant. The regularization constant in the cost function defines the tradeoff between a large margin and misclassification error (i.e. empirical risk minimization) and its choice is further discussed in Section 2.3.2. SVMs are based on the principle of structural risk minimization which balances model complexity and empirical error. The first set of constraints corresponds to Equation (2.18), while the second set imposes positive slack variables ξ_i, tolerating misclassifications. The value ξ_i indicates the distance of x_i with respect to the decision boundary:

- $\xi_i \geq 1 : y_i \left(w^T \varphi(x_i) + b \right) < 0$ implies that the decision function and the target have a different sign, indicating that x_i is misclassified,

- $0 < \xi_i < 1 : x_i$ is correctly classified, but lies inside the margin,
- $\xi_i = 0 : x_i$ is correctly classified and lies outside the margin or on the margin boundary.

Typically, the problem formulation in Equations $(2.20) - (2.22)$ is referred to as the primal optimization problem. Equivalently, the optimization problem for SVMs can be written in the dual space using the Lagrangian with Lagrange multipliers $\alpha_i \geq 0$ for the first set of constraints, Equation (2.21). The solution for the Lagrange multipliers is obtained by solving a quadratic programming problem. Finally, the SVM classifier takes the form

$$y(x) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right), \tag{2.23}$$

where the kernel function $K(\cdot, \cdot)$ is positive definite. It satisfies Mercer's condition then, meaning that $K(x, x_i)$ equals $\varphi(x)^T \varphi(x_i)$. This relation is also often called the kernel trick since no explicit construction of the mapping $\varphi(\cdot)$ is needed. This enables SVMs to work in a high-dimensional (or infinite-dimensional) feature space, without actually performing calculations in this space. Various types of kernel functions can be chosen:

- linear SVM: $K(x, z) = x^T z$,
- polynomial SVM of degree d: $K(x, z) = (\tau + x^T z)^d, \quad \tau \geq 0$,
- radial basis function (RBF): $K(x, z) = \exp(- \parallel x - z \parallel_2^2 / \sigma^2)$,
- MLP: $K(x, z) = \tanh(\kappa_1 x^T z + \kappa_2)^d$,

where $K(\cdot, \cdot)$ is positive definite for all σ values in the RBF kernel case and $\tau \geq 0$ values in the polynomial case, but not for all possible choices of κ_1, κ_2 in the MLP case.

The solution to the convex quadratic programming problem in Equations $(2.20) - (2.22)$ is global and unique when a positive definite kernel is used. For a positive semi-definite kernel the solution to the quadratic programming problem is global but not necessarily unique. An important property of the SVM classifier, called sparseness, is that a number of the resulting Lagrange multipliers, α_i, equal zero. As such, the sum in the resulting classifier in Equation (2.23) should only be taken over all nonzero α_i values, i.e. support values, instead of all data points. The

corresponding vectors x_i are referred to as the support vectors. These data points are located close to the decision boundary and contribute to the construction of the separating hyperplane.

2.2.5 Least squares support vector machines

It has been proposed to modify the SVM methodology by introducing a least squares loss function and equality instead of inequality constraints [214]. Instead of solving a quadratic programming problem, the solution is obtained from a set of linear equations, which significantly reduces the complexity and computational effort. The resulting least squares support vector machine (LS-SVM) classifier optimizes the following problem

$$\min_{w,e,b} \mathcal{J}_2(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2, \tag{2.24}$$

subject to

$$y_i(w^T \varphi(x_i) + b) = 1 - e_i, \quad i = 1, ..., N, \tag{2.25}$$

with $e = [e_1 \ ... \ e_N]^T$ a vector of error variables to tolerate misclassifications, $\varphi(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^{d_h}$ a mapping from the input space into a high-dimensional (e.g. infinite-dimensional) feature space of dimension d_h, w a vector of the same dimension as $\varphi(\cdot)$, γ a positive regularization constant and b a bias term. In this equation, the standard SVM formulation is modified by using a least squares loss function with error variables e_i and replacing the inequality constraints with equality constraints. Since, the value 1 in the equality constraints is a target value instead of a threshold value, the method is related to kernel Fisher discriminant analysis [213]. The Lagrangian for the problem in Equations (2.24) − (2.25) is

$$\mathcal{L}(w,b,e,\alpha) = \mathcal{J}_2(w,e) - \sum_{i=1}^{N} \alpha_i \Big(y_i \Big[w^T \varphi(x_i) + b \Big] - 1 + e_i \Big), \tag{2.26}$$

where α_i are the Lagrange multipliers ($i = 1, ..., N$) or the support values. The conditions for optimality yield

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w} & = & 0 \ \rightarrow \ w = \sum_{i=1}^{N} \alpha_i y_i \varphi(x_i), \\ \frac{\partial \mathcal{L}}{\partial b} & = & 0 \ \rightarrow \ \sum_{i=1}^{N} \alpha_i y_i = 0, \\ \frac{\partial \mathcal{L}}{\partial e_i} & = & 0 \ \rightarrow \ \alpha_i = \gamma e_i, \ i = 1, ..., N, \\ \frac{\partial \mathcal{L}}{\partial \alpha_i} & = & 0 \ \rightarrow \ y_i \Big[w^T \varphi(x_i) + b \Big] - 1 + e_i = 0, \ i = 1, ..., N. \end{cases} \tag{2.27}$$

Whereas the primal problem is expressed in terms of the feature map, the linear optimization problem in the dual space is expressed in terms of the kernel function

$$\left[\begin{array}{c|c} 0 & y^T \\ \hline y & \Omega + I/\gamma \end{array} \right] \left[\begin{array}{c} b \\ \hline \alpha \end{array} \right] = \left[\begin{array}{c} 0 \\ \hline 1_{N \times 1} \end{array} \right], \tag{2.28}$$

where $y = [y_1 \ ... \ y_N]^T$, $\alpha = [\alpha_1 \ ... \ \alpha_N]^T$, $1_{N \times 1} = [1 \ ... \ 1]^T$, and $\Omega \in \mathbb{R}^{N \times N}$ is a matrix with elements $\Omega_{ij} = y_i y_j \varphi(x_i)^T \varphi(x_j)$ with $i, j = 1, ..., N$. Provided with an input vector x, the resulting LS-SVM classifier in the dual space is similar to the standard SVM classifier

$$y(x) = \text{sign} \left(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b \right), \tag{2.29}$$

with $K(\cdot, \cdot)$ a positive definite kernel function with $K(x, x_i) = \varphi(x)^T \varphi(x_i)$. The support values α_i are proportional to the error of the corresponding training data point. This implies that usually every training data point is a support vector and no sparseness property remains in the LS-SVM formulation. However, high support values indicate a high contribution of the data point to the decision boundary. The choice of the regularization parameter γ and the kernel hyperparameter σ, in case of an RBF kernel, is further discussed in the following section and Section 2.3.2.

2.2.6 Bayesian least squares support vector machines

A disadvantage of the standard (LS-)SVM classifier, as presented in the previous sections, is the lack of a probabilistic interpretation of the outputs. Therefore, a

Bayesian evidence framework has been established for LS-SVM classifiers [231]. Moreover, an additional advantage is that the Bayesian evidence framework enables to automatically determine the regularization parameter and the kernel hyperparameter σ for the RBF kernel. In case of Bayesian LS-SVMs the posterior distribution is approximated by a Gaussian around the mode, representing the most probable values w_{MP} [144]. Hyperparameters are optimized rather than integrated out. The original LS-SVM model formulation in Equations (2.24) − (2.25) is modified to

$$\min_{w,b,e} \mathcal{J}_3(w,e) = \mu \frac{1}{2} w^T w + \zeta \frac{1}{2} \sum_{i=1}^{N} e_i^2, \tag{2.30}$$

with

$$e_i = y_i - (w^T \varphi(x_i) + b), i = 1, ..., N, \tag{2.31}$$

such that $\gamma = \zeta/\mu$. The Bayesian LS-SVM framework makes use of three different levels of inference [231]. A conceptual interpretation, in case of an RBF kernel, is given by:

- first level: inference of w and b

$$P(w,b|D,\mu,\zeta,H_\sigma) = \frac{P(D|w,b,\mu,\zeta,H_\sigma)}{P(D|\mu,\zeta,H_\sigma)} P(w,b|\mu,\zeta,H_\sigma) \tag{2.32}$$

- second level: inference of hyperparameters μ and ζ

$$P(\mu,\zeta|D,H_\sigma) = \frac{P(D|\mu,\zeta,H_\sigma)}{P(D|H_\sigma)} P(\mu,\zeta|H_\sigma) \tag{2.33}$$

- third level: inference of the kernel parameter σ and model comparison

$$P(H_\sigma|D) = \frac{P(D|H_\sigma)}{P(D)} P(H_\sigma) \tag{2.34}$$

with D being the data and H_σ the model depending on the RBF kernel parameter σ. On the first level of inference, the bias b and weights w (or α in the dual

representation) of the Bayesian LS-SVM are determined. A separable Gaussian distribution is assumed as prior distribution. The prior for w is a Gaussian with zero mean and variance μ^{-1}, while a Gaussian distribution is assumed for the prior for b with zero mean and variance σ_b^2 with $\sigma_b \to \infty$ to approximate a uniform distribution. Using Bayes' rule results in the most probable values for these parameters, b_{MP} and w_{MP}. As the prior distribution and the likelihood function correspond to the regularization term and the sum of squared errors in Equation (2.30), respectively, this relates to solving an LS-SVM classifier. The most probable values for the regularization parameters μ and ζ are calculated on the second level of inference. To this end, uniform priors are assumed. The third level of inference performs model comparison to infer the kernel parameters. A uniform prior $P(H_\sigma)$ is assumed, meaning that all RBF kernel models with different parameters σ are equally likely. The iterative algorithm of Bayesian inference for LS-SVM classifiers consists of the following steps: initialization of μ, ζ and σ in case of an RBF kernel, solving the linear system in Equation (2.28) to obtain b_{MP} and w_{MP}, estimation of μ_{MP} and ζ_{MP} and updating of σ. Class probabilities are obtained by integrating over the posterior distribution for w and b using prior class probabilities, which are often chosen as the proportion of cases from each class in the data set.

2.2.7 Kernel logistic regression

Logistic regression and kernel logistic regression (KLR) have already proven their value in the statistics and the machine learning community [92, 249]. In contrast to the risk minimization approach as employed by standard (LS-)SVMs, logistic regression and KLR directly yield probabilistic outcomes based on a maximum likelihood argument and the methodology can be extended to multi-class classification problems in a straightforward way. In classical binary logistic regression $P(y = 1|x)$, the probability for outcome class 1 given data point x, is based on a linear combination of input variables. The output of the model can be interpreted as a probability by using the logit transformation

$$\log\left(\frac{P(y=1|x)}{1-P(y=1|x)}\right) = \beta_0 + \beta^T x, \tag{2.35}$$

where β_0 is the intercept and β a vector of the weights for each input variable. The intercept and weights are often obtained by maximizing the likelihood of the data based on iteratively re-weighted least squares. Finally, the predicted probability

for a case x is obtained as

$$P(y=1|x) = \frac{\exp(\beta_0 + \beta^T x)}{1+\exp(\beta_0 + \beta^T x)}. \tag{2.36}$$

The binary model can be extended to multi-class logistic regression (multinomial logit) by using baseline category logit equations using a reference class, e.g. class n_C,

$$\begin{cases} P(y=1|x) & = & \frac{\exp(\beta_{01}+\beta_1^T x)}{1+\sum_{j=1}^{n_C-1}\exp(\beta_{0j}+\beta_j^T x)}, \\ P(y=2|x) & = & \frac{\exp(\beta_{02}+\beta_2^T x)}{1+\sum_{j=1}^{n_C-1}\exp(\beta_{0j}+\beta_j^T x)}, \\ ... & & \\ P(y=n_C|x) & = & \frac{1}{1+\sum_{j=1}^{n_C-1}\exp(\beta_{0j}+\beta_j^T x)}, \end{cases} \tag{2.37}$$

where the intercepts and weights are obtained by maximizing the multinomial log-likelihood. The logistic regression model can be extended to KLR by introducing a feature map [104]. In the kernel version of logistic regression the different models are defined by

$$\begin{cases} P(y=1|x) & = & \frac{\exp(\beta_{01}+\beta_1^T \varphi(x))}{1+\sum_{j=1}^{n_C-1}\exp(\beta_{0j}+\beta_j^T \varphi(x))}, \\ P(y=2|x) & = & \frac{\exp(\beta_{02}+\beta_2^T \varphi(x))}{1+\sum_{j=1}^{n_C-1}\exp(\beta_{0j}+\beta_j^T \varphi(x))}, \\ ... & & \\ P(y=n_C|x) & = & \frac{1}{1+\sum_{j=1}^{n_C-1}\exp(\beta_{0j}+\beta_j^T \varphi(x))}. \end{cases} \tag{2.38}$$

KLR is closely related to SVMs, but these methods differ with respect to the loss function [249]. While in SVMs a Hinge loss function is used (i.e. the loss is zero for cases on the correct side of the margin and increases linearly for cases on the wrong side of the margin), KLR uses a negative log-likelihood type of loss function. However, unlike SVMs, KLR by its nature is not sparse and needs all training data points in its final model. Different modifications to the original algorithm were suggested to obtain sparseness [249, 107]. It has been proposed to solve the multi-class KLR problem using an iteratively re-weighted version of LS-SVMs [104]. To account for the intercept define $x_i^* = [1\ (x_i)^T]^T \in \mathbb{R}^{d+1}$, $i = 1, ..., N$, and the outcome y_i indicates to which of the n_C classes the data point belongs. The common method to infer the model parameters, $\beta_j^* = [\beta_{0j}\ (\beta_j)^T]^T, j = 1, ..., n_C -$

1, for the multi-class logistic regression model is using a penalized negative log-likelihood (NLL)

$$\min_{\beta_1^*,...,\beta_{n_C-1}^*} \mathcal{J}_4(\beta_1^*, ..., \beta_{n_C-1}^*) = \text{NLL} + \frac{\gamma}{2} \sum_{j=1}^{n_C-1} (\beta_j^*)^T \beta_j^*, \tag{2.39}$$

with the n_Cth class the reference class. Assuming that $B = [\beta_1^{*T} \ ... \ \beta_{n_C-1}^{*T}] \in \mathbb{R}^{(n_C-1)(d+1)}$, a vector with all model parameters, the Newton-based optimization strategy is used

$$B^{(k)} = B^{(k-1)} + s^{(k)}. \tag{2.40}$$

The vector $B^{(k)}$ contains the parameter estimates at iteration k and the step, $s^{(k)}$, is $-[H^{(k)}]^{-1} g^{(k)}$ where the gradient and Hessian of the penalized NLL are respectively defined as g and H. To express multi-class KLR in terms of iteratively re-weighted LS-SVMs, define $\phi^T \in \mathbb{R}^{(n_C-1)q \times (n_C-1)N}$ as

$$\phi^T = \left[\begin{array}{ccc|ccc|c|ccc} \varphi(x_1^*) & 0 & 0 & \varphi(x_2^*) & 0 & 0 & & \varphi(x_N^*) & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 & \cdots & 0 & \ddots & 0 \\ 0 & 0 & \varphi(x_1^*) & 0 & 0 & \varphi(x_2^*) & & 0 & 0 & \varphi(x_N^*) \end{array}\right], \tag{2.41}$$

where $\varphi(\cdot) : \mathbb{R}^{d+1} \to \mathbb{R}^q$ is a mapping from the input space to the feature space, induced by a positive definite kernel. Next, define the indicator function $I_{(y_i=j)}$, which is 1 when y_i equals j and 0 otherwise, and $r_i = [I_{(y_i=1)} \ ... \ I_{(y_i=n_C-1)}]^T$, $R = [r_1^T \ ... \ r_N^T]^T$, $P^{(k)} = [p_{1,1}^{(k)} \ ... \ p_{n_C-1,1}^{(k)} \ ... \ p_{1,N}^{(k)} \ ... \ p_{n_C-1,N}^{(k)}]^T \in \mathbb{R}^{(n_C-1)N}$ with $p_{j,i}^{(k)}$ the probability that data point i belongs to class j at iteration k and the ith block ($\in \mathbb{R}^{(n_C-1)\times(n_C-1)}$) of the blockdiagonal weight matrix $W^{(k)} = \text{blockdiag}(W_1^{(k)} \ ... \ W_N^{(k)})$ as

$$W_i^{(k)} = \left[\begin{array}{ccc} t_i^{1,1} & t_i^{1,2} & \cdots & t_i^{1,n_C-1} \\ t_i^{2,1} & t_i^{2,2} & & t_i^{2,n_C-1} \\ \vdots & & \ddots & \\ t_i^{n_C-1,1} & t_i^{n_C-1,2} & & t_i^{n_C-1,n_C-1} \end{array}\right], \tag{2.42}$$

with

$$t_i^{a,j} = \begin{cases} p_{a,i}^{(k)}(1-p_{a,i}^{(k)}) & \text{if } a = j, \\ -p_{a,i}^{(k)} p_{j,i}^{(k)} & \text{if } a \neq j. \end{cases} \tag{2.43}$$

Next, within the context of LS-SVMs the following primal problem is minimized

$$\min_{s^{(k)},e^{(k)}} \mathcal{J}_5(s^{(k)}, e^{(k)}) = \frac{1}{2}(e^{(k)})^T W^{(k)} e^{(k)} + \frac{\gamma}{2}(s^{(k)}+B^{(k-1)})^T(s^{(k)}+B^{(k-1)}), \tag{2.44}$$

subject to

$$(W^{(k)})^{-1}(R - P^{(k)}) = \phi s^{(k)} + e^{(k)}, \tag{2.45}$$

with $e^{(k)}$ an error variable at iteration k. In the dual space the solution for the problem in Equations (2.44) $-$ (2.45) can be obtained by solving a linear system. The class probability of a new case x^* for class a, following the multi-class KLR model, can be obtained by

$$P(y = a|x^*) = \frac{\exp\left[(\beta_a^*)^T \varphi(x^*)\right]}{\sum_{j=1}^{n_C} \exp\left[(\beta_j^*)^T \varphi(x^*)\right]}, \tag{2.46}$$

where $(\beta_{n_C}^*)^T \varphi(x^*) = 0$ and $(\beta_a^*)^T \varphi(x^*) = \frac{1}{\gamma} \sum_{i=1, i\in\mathcal{D}_a}^{N} \alpha_{i,a} K(x^*, x_i)$, with $\mathcal{D}_a$ the data belonging to class a, $\alpha_{i,a}$ the support value for the prediction of class a for case i and $K(x^*, x_i) = \varphi(x_i)^T \varphi(x^*)$ a positive definite kernel function (i.e. RBF kernel).

2.2.8 Canonical correlation analysis

Canonical correlation analysis (CCA) is a statistical technique to determine the relationship between two sets of variables [93]. The method is a generalization of ordinary correlation analysis, which quantifies the relationship between two random variables a and u by means of the correlation coefficient

$$\rho = \frac{\sigma_{au}}{\sqrt{\sigma_a^2 \sigma_u^2}}, \tag{2.47}$$

where σ_{au}, σ_a^2 σ_u^2 represent the covariance and variances, respectively. The correlation coefficient ρ, which obtains values between -1 and $+1$, measures the degree of linear dependence between a and u. For variables with zero mean Equation (2.47) is replaced by

$$\rho = \frac{E[au]}{\sqrt{E[a^2]E[u^2]}}, \tag{2.48}$$

where $E[\cdot]$ denotes the expected value. CCA can be applied to two multivariate random vectors, with zero mean, $a = [a_{1t} \ ... \ a_{mt}]^T$ and $u = [u_{1t} \ ... \ u_{nt}]^T$, with $t = 1, ..., d$. The canonical equations represent the linear combinations of the components in a and u, which result in the canonical variates Z_a and Z_u

$$\begin{aligned} Z_a &= \omega_{a_1} a_1 + ... + \omega_{a_m} a_m = \omega_a^T a, \\ Z_u &= \omega_{u_1} u_1 + ... + \omega_{u_n} u_n = \omega_u^T u. \end{aligned} \tag{2.49}$$

CCA computes the linear combination coefficients $\omega_a = [\omega_{a_1} \ ... \ \omega_{a_m}]^T$ and $\omega_u = [\omega_{u_1} \ ... \ \omega_{u_n}]^T$ such that the correlation between Z_a and Z_u is maximum. The solution $\omega_a = \omega_u = 0$ is not allowed. There exist several implementations of CCA. A reliable and fast implementation is based on the interpretation of CCA in terms of principal angles between linear subspaces and has been used for classification purposes [121, 247, 77].

2.2.9 Software

In this thesis Matlab 7.1−7.6 (The MathWorks, Inc., Natick, USA) was used. MLPs were obtained with the Stuttgart Neural Network simulator. SVM classification was performed with the SVM Toolbox Version 1.0 of the Institute for Molecules and Materials of RUN. LS-SVM classifiers were developed using the LS-SVMlab toolbox [169, 213]. KLR and CCA models were obtained using Matlab code from the developers [104, 121].

2.3 Model performance evaluation

A very important aspect of building classifiers or the development of classification algorithms is the evaluation of their performance [7, 2]. Apart from the measure that is used to quantify the classification performance, a specific strategy is required to obtain this estimate. In addition, one is often interested in comparing

the performance of different classification algorithms or classifiers. The following sections provide detailed information about these crucial elements.

2.3.1 Performance measures

The accuracy represents the proportion of cases that is correctly classified. Inversely, the misclassification error (or error rate) denotes the proportion of cases that is not correctly classified. For multi-class problems accuracies can be summarized in a confusion matrix [151]. This matrix structure relates the true classification (on its horizontal axis) and the predicted classification (on its vertical axis) for a data point. Accuracies can be calculated such that the total sum for each actual class outcome becomes 100 %. These measures are not optimal since they can be misleading when data are unbalanced. In case one of two classes accounts for 90 % of the data, a method that always predicts this class already obtains an accuracy of 0.90 (cf. the majority rule). Based on the balanced error rate (BER), which is the average of the error rate for each separate class, an unbalance in the data can be taken into account [81].

Other well-known performance measures are the sensitivity and specificity. The sensitivity is the proportion of cases of the class of interest (e.g. class 1) that are correctly predicted as members of this class. It is the number of true positives (i.e. correctly predicted class 1 cases) divided by the sum of the true positives and false negatives (i.e. incorrectly predicted class 1 cases). The specificity is the proportion of cases from the other class (e.g. class 2) that are correctly predicted. It is the number of true negatives (i.e. correctly predicted class 2 cases) divided by the sum of the number of true negatives and false positives (i.e. incorrectly predicted class 2 cases). Another interesting measure is the positive predictive value, being the probability of belonging to class 1 when this class is predicted. The negative predictive value is the analogue for the other class. Finally, some measures try to quantify the level of discrimination between two classes. A very popular measure for binary problems is the area under the receiver operating characteristic (ROC) curve (AUC) [86]. An ROC curve can be constructed by plotting $1 -$ specificity versus sensitivity by varying the decision threshold over its entire range (Figure 2.6). A classifier or algorithm that gives random predictions results in an AUC of 0.5, while a perfect performance corresponds to an AUC of 1. These values are often interpreted as the probability that a random pair of cases (one from each class) will be correctly discriminated.

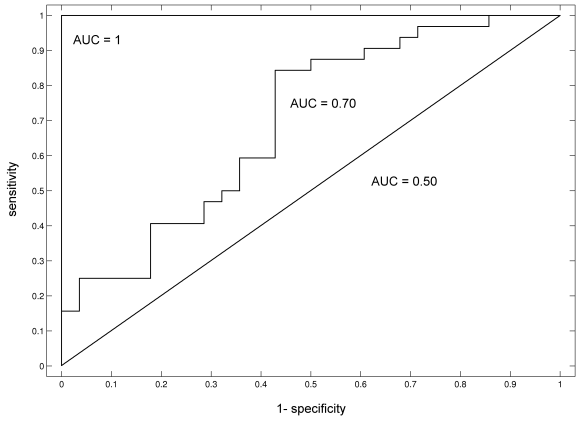


Figure 2.6: *Example of ROC curves.*

For probabilistic classifiers, a suitable performance measure is the Brier score [21]. The Brier score is related to the mean squared error (MSE)

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{n_C}\sum_{j=1}^{n_C}(p_{ji} - t_{ji})^2, \tag{2.50}$$

where t_{ji} is set to 1 if case i is coming from class j and 0 otherwise, while p_{ji} is the predicted posterior probability of class j for case i. This measure takes the amount of uncertainty about the predictions into account.

2.3.2 Evaluation strategies

In this thesis, most algorithms are supervised methods, meaning that a training data set is needed for the construction of the classifier. However, the performance of the classifier or algorithm is evaluated on a set of new, unseen, examples during the testing phase. If the classifier would fit the training data too much, the classification ability on these unseen data may degrade. This behavior is called overfitting. Therefore, the generalization ability of the classifier is determined during this testing phase. Good generalization is an important property, and often the ultimate goal, of a classifier or classification algorithm since it provides information about the quality of the developed classifier or algorithm. As such, in case one aims to evaluate the performance of a classifier, and depending on the number of available data points, it may be required to split the original data

in a training set and a test set. In order to evaluate the performance during training (e.g. to avoid overfitting), an additional validation set may be used. This strategy can for instance be applied to determine hyperparameter (i.e. C, γ, σ) for SVM-based classifiers. If no separate validation set is available, cross-validation techniques can be used for this purpose [115]. In case one is not interested in a classifier, but in an algorithm that produces classifiers, typically, many training and test sets are generated such that multiple classifiers can be generated and evaluated. Data splitting techniques are used to repeatedly obtain a training (and validation) set and a test set. Afterwards, the results on the multiple test sets can be combined, for example by averaging the performance measures.

There exist various techniques to perform data splitting [115, 47]. Applying v-fold cross-validation involves dividing the data into v subsets of approximately equal size. For each of the subsets, the remaining $v - 1$ subsets are combined to form the training data such that the performance can be measured on the remaining subset. This process can be repeated such that different subsets are generated in a random way, after which the results on the test sets can be combined. If one chooses v equal to the number of points in the total data set, this strategy is called leave-one-out cross-validation. On the other hand, repeated random sampling iteratively divides the data in a set for training and a set for testing in a random way. Often 2/3 of the data are used for training and the remaining data points are used for testing. In addition, cross-validation and repeated sampling can be used in a stratified manner [193]. This means that the training sets and test sets are stratified such that they approximately contain the same proportions of labels as the original data. It guarantees a prescribed number of data points from each subpopulation. Finally, a good estimator of the performance can be obtained in a prospective setting, where one tests on data that are collected after the classifier or classification algorithm has been developed [103].

2.3.3 Statistical analysis

Researchers are often interested in comparing the classification performances of different algorithms. Traditionally, these comparisons are performed based on statistical significance testing. The Wilcoxon signed-rank test (for two algorithms) and the Friedman test (for more algorithms) are two nonparametric statistical tests for this purpose [42]. If the null hypothesis of the test is rejected, post-hoc tests for multiple comparison are usually performed [42, 73]. In this thesis the Tukey procedure for multiple comparison is used [223].

The Wilcoxon signed-rank test represents the nonparametric alternative to the paired t-test [42]. The difference in performance between two algorithms for

classification problem i out of I is denoted by d_i. The null hypothesis of the test is that the median of the differences is zero. Therefore, the differences are ranked with respect to their absolute values, while average ranks are assigned in case of ties. Using the corresponding sign of d_i, the sum of positive ranks, R_+, and the sum of negative ranks, R_-, are calculated. Ranks are evenly split between the sums for $d_i = 0$, and one of them is ignored if there is an odd number. The statistic T is calculated as the minimum of

$$R_- = \sum_{i:d_i<0} r_i + \frac{1}{2} \sum_{i:d_i=0} r_i, \tag{2.51}$$

and

$$R_+ = \sum_{i:d_i>0} r_i + \frac{1}{2} \sum_{i:d_i=0} r_i, \tag{2.52}$$

where r_i denotes the rank of d_i. Tables have been developed with critical values for T. For large data sets the statistic

$$z = \frac{T - \frac{1}{4}I(I+1)}{\sqrt{\frac{1}{24}I(I+1)(2I+1)}}, \tag{2.53}$$

is approximately normally distributed.

The Friedman test is a nonparametric alternative to the repeated-measures analysis of variance (ANOVA) [64, 65]. The method ranks the algorithms according to their performances within each block (e.g. a classification problem). If there are ties, average ranks are used. The Friedman test compares the average ranks of the performances of the algorithms

$$r_j = \frac{1}{I} \sum_{i=1}^{I} r_{ij}, \tag{2.54}$$

where r_{ij} denotes the rank of the performance of the jth of k algorithms for the ith of I blocks. Under the null hypothesis all algorithms' performances are assumed

to be equivalent such that their ranks r_j are equal, which results in the following statistic

$$\chi_F^2 = \left[\frac{12}{Ik(k+1)} \sum_{j=1}^{k} r_j^2 \right] - 3I(k+1), \tag{2.55}$$

which is distributed according to the χ^2 statistic with $k-1$ degrees of freedom if I or k are large enough (i.e. $I > 15$ or $k > 4$). For smaller values of I and k, exact critical values should be used. The nonparametric Friedman test has less power than the parametric ANOVA when the assumptions of ANOVA are met, but this does not have to be the case when these assumptions are not met.

The Tukey multiple comparison procedure is used to compare the performances of all algorithms with each other [223]. In case all sample sizes are equal, the family confidence coefficient is $1-\alpha$ and the family significance level is α. The Tukey test is a conservative procedure when the sample sizes are not equal. The confidence limits are $\hat{D} \pm Ts\{\hat{D}\}$, where $\hat{D} = \overline{Y}_{\cdot j} - \overline{Y}_{\cdot j'}$, representing the difference between the estimated means of the obtained measures for the jth and j'th algorithm. $T = \frac{1}{\sqrt{2}} q(1-\alpha; k, I_T - k)$, where q follows a studentized range distribution and I_T denotes the total number of samples. The estimated variance $s^2\{\hat{D}\} = \text{MSE}(\frac{1}{I_j} + \frac{1}{I_{j'}})$, with I_j, $I_{j'}$ the sample size for the jth and j'th algorithm. In addition, a paired comparison plot can provide a graphical representation of the Tukey procedure [161]. Assuming that all sample sizes are equal, it provides an interval for the mean value for the measure of the jth algorithm by $\overline{Y}_{\cdot j} \pm \frac{1}{2} Ts\{\hat{D}\}$. When the intervals for the measures of two algorithms overlap, there is no statistically significant difference between the algorithms. If the intervals are disjoint, there exists a statistically significant difference between the algorithms.

2.4 Conclusions

This chapter described the required steps to transform the acquired data into classification results. This included the main concepts of signal preprocessing, classifier construction and performance assessment. Data preprocessing is an important step since it has a direct influence on all consecutive steps. Enhancement of the key features of the signals, while reducing the influence of the irrelevant information, improves the performance of constructed classifiers. An overview was provided of different algorithms that could be used for classification purposes. Naturally, there exist many more algorithms to perform this task. For this thesis,

methods were selected with clearly different characteristics and, as a consequence, different strengths and weaknesses. Linear techniques (LDA), as well as, nonlinear (QDA, (LS-)SVM) methods were included. Some methods were able to provide class probabilities (Bayesian LS-SVM, KLR), while others ((LS-)SVM) generated 'black-or-white' decisions. High-dimensional data might not directly be fed to some classification methods (LDA), while others (KLR, (LS-)SVM) were able to deal with it. There existed straightforward generalizations to multi-class classification for some approaches (LDA, KLR), while this was less obvious for other methods ((LS-)SVM). Some algorithms (LS-SVM) were computationally more efficient than others (SVM). A common factor was that all these algorithms needed example data to construct classifiers. An important aspect was the evaluation of the classifiers and classification algorithms. There exist many approaches to perform this task; an overview was provided. In this thesis, the evaluation is mainly performed based on cross-validation and stratified repeated sampling using the accuracy, BER, Brier score and AUC. The Wilcoxon signed-rank test, the Friedman test and the Tukey test are used for statistical testing.

Chapter 3

Quantification of MR spectra

This chapter presents an overview of different methods that can be used to estimate the metabolite concentrations from MRS and MRSI data. Section 3.1 describes the models used in this thesis for MRS(I) data. The remaining sections provide details about the quantification methods. Section 3.2, Section 3.3, Section 3.5 and Section 3.4 introduce peak integration, LCModel, AQSES and QUEST, respectively.

3.1 Modeling MRS(I) data

In vivo MRS and MRSI generate time domain signals, which consist of the responses from metabolites (including MMs), partially suppressed water and noise. In theory, the in vivo time domain signals have the shape of a sum of complex damped exponentials. Instead of modeling these individual lineshapes, the in vivo short TE MRS signals can be modeled based on a selected set of metabolite profiles from a database. The database consists of signals from metabolites that are visible during in vivo MRS(I), but these signals can also be measured in vitro or can be simulated using quantum mechanical knowledge. This enables to implicitly incorporate prior knowledge that relates the individual peaks in the in vivo spectrum. The metabolite concentrations are the quantities of interest and the estimation of these values is called quantification. The concentrations can be estimated based on the weighting coefficients (i.e. amplitudes) of the linearly combined in vitro profiles. This linear combination should also allow small corrections in spectral parameters as frequency shifts, damping corrections and phase shifts, since these parameters may vary across different measurements [189]. In addition, a quantification method should also take the presence of a partially

suppressed water resonance and macromolecular baseline into account. The macromolecular baseline overlaps with the metabolites of interest, while the frequency region where the water is located is known.

The database of metabolite profiles is denoted by $\{v_k\}_{k=1}^{K}$, where v_k represents a complex-valued time series of length N. The in vivo measured MRS(I) signal y, which is also a complex-valued time series of length N, is modeled as

$$y(t) = \hat{y}(t) + \varepsilon_t, \;\; t = t_0, ..., t_{N-1}, \tag{3.1}$$

$$\hat{y}(t) = \sum_{k=1}^{K} \alpha_k (\zeta_k)^t (\eta_k)^{t^2} v_k(t) + b(t) + w(t), \tag{3.2}$$

where ε_t, $b(t)$ and $w(t)$ denote an unknown noise perturbation with zero mean, the baseline and the water component, respectively. The unknown parameters $\alpha_k, \zeta_k \in \mathbb{C}$ account for amplitudes of the metabolites in the database and for the necessary corrections of the database signals. The complex amplitudes α_k and the complex values ζ_k, η_k are defined as

$$\begin{aligned} \alpha_k &= a_k \exp(j\phi_k), \\ \zeta_k &= \begin{cases} \exp(-d_k + j2\pi f_k), & \text{for Lorentzian and Voigt lineshapes,} \\ \exp(j2\pi f_k), & \text{for Gaussian lineshapes,} \end{cases} \\ \eta_k &= \begin{cases} \exp(je_k), & \text{for Lorentzian lineshapes,} \\ \exp(-g_k + j2\pi f_k), & \text{for Gaussian and Voigt lineshapes,} \end{cases} \end{aligned} \tag{3.3}$$

where $j = \sqrt{-1}$, a_k are the real amplitudes, ϕ_k are the phase shifts, d_k are damping corrections, f_k are frequency shifts, g_k are Gaussian damping corrections, and e_k are eddy current correction terms.

Long TE MR data are commonly modeled as a linear combination of K exponentially damped sinusoids

$$y(t) = \sum_{k=1}^{K} \alpha_k \, (\zeta_k)^t + w(t) + \varepsilon_t, \tag{3.4}$$

In this case the parameters are not corrections of the metabolite profiles and it is sufficient to characterize the signal.

3.2 Peak integration

A straightforward method to obtain an estimate for the concentration of a metabolite is peak integration [155]. The area under the peak of interest in the frequency domain is simply integrated and no assumptions about the lineshape of the signal have to be made. The advantage of this nonparametric approach is that it is very fast and quite robust. On the other hand, this method does not succeed to disentangle overlapping peaks. This implies that peak integration is not able to extract information from individual peaks or metabolite contributions. Residual baseline and low SNRs also have an influence on the peak-integrated values. In addition, an appropriate phase correction is needed when working with the real part of the frequency domain MRS(I) signal, which is far from trivial. Peak integration is also dependent on the defined bounds for integration. The tail of the peaks is neglected by peak integration and the area under the peaks will therefore be underestimated, possibly by up to 40 % [155]. In this thesis peak integration is performed based on the trapezoidal rule.

3.3 LCModel

LCModel is a semi-parametric approach to model in vivo MR spectra as a linear combination of in vitro metabolite profiles [186] (Figure 3.1). LCModel is a frequency domain method that models the spectrum data points as

$$\hat{Y}(v_p) = \exp(-j(\phi_0 + v_p\phi_1)) \left(\sum_{i=1}^{B} \beta_i B_i(v_p) + \sum_{k=1}^{K} C_k \sum_{n=-S}^{S} S_n M_k(v_{j-n}; \gamma_k, \lambda_k) \right), \tag{3.5}$$

subject to

$$C_k \geq 0, \quad \gamma_k \geq 0, \quad \sum_{n=-S}^{S} S_n = 1, \tag{3.6}$$

where C_k represents the concentration of the kth metabolite, ϕ_0 and ϕ_1 the zero-order and first-order phase corrections, respectively. The in vitro metabolite spectra, $M_k(v;0,0)$, are broadened and shifted with the parameters γ_k and λ_k, respectively, such that

$$M_k(v;\gamma_k,\lambda_k) = \mathrm{FFT}\{m_k(t)\exp(-(\gamma_k + j2\pi\lambda_k))t\}, \tag{3.7}$$

with $m_k(t)$ the inverse fast Fourier transform (FFT) of the model spectrum $M_k(v;0,0)$. The baseline is represented by cubic B-splines, $B_i(v)$ and the lineshape coefficients S_n take field inhomogeneities and eddy currents into account. The concentrations are obtained based on a Marquardt modification of a constrained Gauss-Newton least squares analysis.

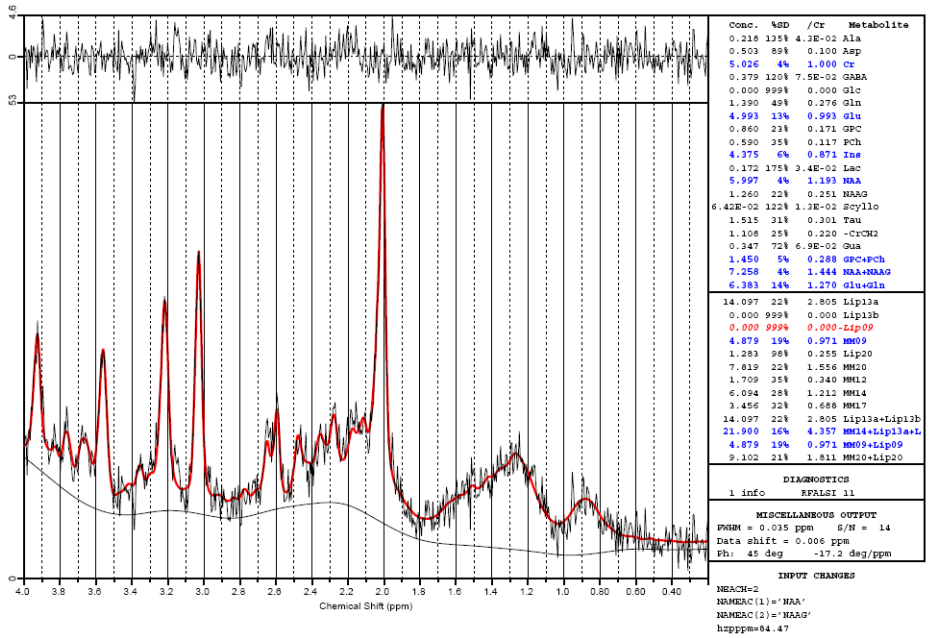


Figure 3.1: *MRS quantification with LCModel. The fitted (red) and original (black) spectrum. Figure adapted from the LCModel user's manual.*

3.4 AQSES

AQSES is a quantification method which fits the in vivo complex signal in the time domain and is implemented in AQSESGUI (http://homes.esat.kuleuven.be/∼

biomed/software.php , accessed on June 27, 2009) [41, 179]. The metabolite amplitudes are the weighting coefficients of a linear combination of corrected in vitro metabolite profiles. The macromolecular baseline b (cf. Equation (3.2)) is characterized by a smooth function. A basis of splines is constructed for nonparametric modeling of the baseline and the discretized splines are put as columns in the matrix A. An arbitrary nonlinear function can be approximated as a linear combination of spline functions. The linear coefficients, $c_1, ..., c_s$ (denoted as the column vector $c \in \mathbb{C}^s$), in the linear combination are unknowns that must be identified. As such, the Fourier transform of the baseline b can be written in matrix notation as the product Ac and using the discrete inverse Fourier transform results in $b = \mathcal{A}c$. AQSES solves the regularized nonlinear least squares criterion

$$\min \ \frac{1}{N} \sum_{t=t_0}^{t_{N-1}} |y(t) - \widehat{y}(t)|^2 + \lambda^2 \|Dc\|^2, \tag{3.8}$$

where $\widehat{y}$ is defined as in Equation (3.2), assuming $b = \mathcal{A}c$. The term $\lambda^2\|Dc\|^2$ imposes a degree of smoothness to the baseline b. The regularization matrix D is used to measure the smoothness of the baseline in the frequency domain and can be taken as a (combination of) discrete derivative operator(s). The regularization parameter λ controls the degree of smoothness and can be tuned via a generalized cross-validation criterion. The water term w in Equation (3.2) is filtered out using an MP-FIR filter in AQSES [212]. The region to be filtered out of the in vivo signal needs to be specified by the user. After neglecting the filter for the sake of simplicity, the nonlinear least squares problem in Equation (3.8) becomes

$$\min_{\alpha_k, \zeta_k, c} \frac{1}{N} \sum_{i=0}^{N-1} \left| y(t_i) - \sum_{k=1}^{K} \alpha_k (\zeta_k)^{t_i} \, v_k(t_i) - (\mathcal{A}c)_i \right|^2 + \lambda^2 \|Dc\|^2, \tag{3.9}$$

and results in

$$\min_{\alpha, \zeta, c} \frac{1}{N} \|y - \Phi(\zeta)\alpha - \mathcal{A}c\|^2 + \lambda^2 \|Dc\|^2, \tag{3.10}$$

with $y = [y(t_0) \ ... \ y(t_{N-1})]^T$, $\alpha = [\alpha_1 \ ... \ \alpha_K]^T$ and $\zeta = [\zeta_1 \ ... \ \zeta_K]^T$. The $N \times K$ matrix $\Phi(\boldsymbol{\zeta})$ is defined as

$$\Phi_{ik} = (\zeta_k)^{t_i} v_k(t_i) = \exp\left((-d_k + j2\pi f_k)t_i\right) v_k(t_i). \tag{3.11}$$

The variable projection (VARPRO) method can be used to solve Equation (3.10). The linear parameters α and c can be projected out of the least squares problem, and only a smaller sized nonlinear least squares problem remains to be solved for the nonlinear variables ζ, via an iterative minimization algorithm of the Levenberg-Marquardt type. In addition, prior knowledge can be imposed in the form of upper and lower bounds on the nonlinear parameters, linear equalities between variables (e.g. d_k or f_k), or constraints on the linear parameters (nonnegative a_k and equal ϕ_k).

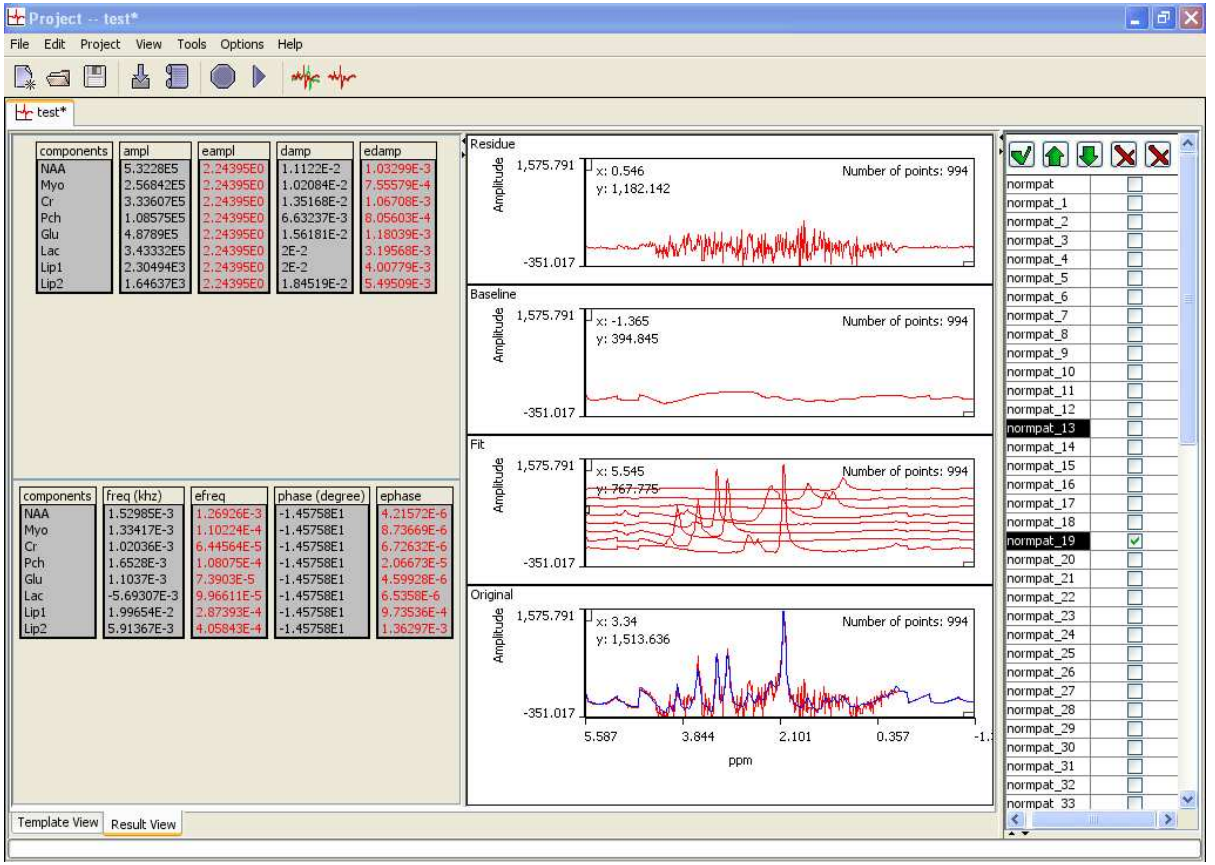


Figure 3.2: *MRS quantification with AQSES. Plots from top to bottom: residue, baseline, corrected individual metabolite profiles and filtered original (red) versus filtered reconstructed (blue) spectra.*

3.5 QUEST

QUEST is a semi-parametric time domain method, which is primarily designed for quantitation of short TE MR spectra [188]. The method uses a simulated or measured metabolite basis set and is implemented in the software package jMRUI-3.0 (http://mrui.uab.es/mrui/ , accessed on June 27, 2009). The nonlinear least squares problem in QUEST can be formulated as

$$\min_{\alpha,\zeta} \frac{1}{N} \left\| y - \Phi(\zeta)\alpha \right\|^2 , \tag{3.12}$$

with the same notations as in Equation (3.10). The nonlinear least squares fit method used in QUEST derives a first-order Taylor expansion of Equation (3.2), where the terms $b(t)$ and $w(t)$ have been removed, with respect to the parameters a_k, d_k, f_k and ϕ_k and iteratively solves for the unknown increments of these parameters by invoking the LAPACK subroutine DGELSS.

Water filtering is not integrated in QUEST. The method is able to deal with the baseline via the subtract procedure. Based on the fast decay of the background signal in the time domain, the background can be subtracted from the raw data after automatic estimation (Algorithm 1).

3.6 Conclusions

Quantification methods aim to estimate the concentrations of metabolites from MRS and MRSI data. These features are interesting since they are able to provide a direct biological interpretation. The advantage of LCModel, AQSES and QUEST is that these methods are based on mathematical models that impose prior knowledge by means of a basis set of metabolite profiles. LCModel is a frequency domain method that is part of a commercial software package. AQSES and QUEST operate in the time domain and are freely available for noncommercial research purposes. On the other hand, compared to these methods, peak integration is a very fast method that does not rely on assumptions about the lineshape of the signal. In this thesis, peak integration, AQSES and QUEST are used for feature extraction.

Algorithm 1 : *Subtract method in QUEST*

Input:
n = number of points to be truncated
$y(t)$ = water filtered signal, $t = 0, \ldots, N-1$

truncate n initial data points in which the MM signal decays below or near the noise level
estimate metabolite parameters in Equation (3.12) with nonlinear least squares fit. The truncation from the previous step may cause loss of metabolite information, resulting in reduced precision of the estimates.
compute the metabolite signal from the estimated parameters and subtract the result from the measured MRS signal. A metabolite-free estimate of the MM signal, including the measurement noise, remains.
model MM signal by a sum of exponentially damped sinusoids with a state space method (e.g. HLSVD-PRO)
subtract state space modeling result from measured MRS signal. A MM-free estimate of the metabolite signal, including the measurement noise, remains.
estimate metabolite parameters α, ζ in Equation (3.12) from signal resulting from previous step with nonlinear least squares fit

Output:
α, ζ = metabolite parameter estimates

Chapter 4

Relevance of automatically selected MRS, MRI and MRSI features

This chapter evaluates the relevance of MR features, which are selected by automated feature selection techniques from machine learning and statistics, for differential diagnosis of brain tumors. The chapter starts with a brief introduction to the feature selection problem (Section 4.1), after which the data (Section 4.2) are described. The in vivo data consist of short TE MRS data from 211 patients with a brain tumor and a data set with combined short TE MRSI data and four types of MRI data from selected voxels of 24 patients with a brain tumor and 4 healthy individuals. These data sets are analyzed with several feature selection methods (Section 4.3). The automatically selected features are summarized in Section 4.4 and evaluated using literature and clinical experience in Section 4.5.

4.1 Feature selection

In recent years, the computational power of computers has significantly increased. One important result is that more and more data are generated. Today, an important problem in machine learning and statistics is keeping track of the most relevant information. For this purpose, feature selection techniques are addressed. The major aims of feature selection for classification are finding a subset of variables (i.e. features) that result in more accurate classifiers and constructing more compact models. Indeed, from a practical point of view, an amount of

variables can be redundant to the classification problem. Feature selection can for example filter out those variables that are irrelevant for the model. In this way, the selection should only capture the relevant features while not overfitting the data. In addition, there is a reduction in the sample size needed for good generalization [17]. In this chapter the focus is mainly on feature weighting and feature selection mechanisms. A technique like principal component analysis (PCA) (Section 5.3), can also lower the dimension of the input space and extract features. However, the present chapter only deals with methods that provide features with a direct biological meaning.

As feature selection is one of the most important topics in pattern recognition, many attempts have been made to develop algorithms. An extensive overview can be found in [240, 116, 16]. Basically, three major types of methods are distinguished [48]. The first category is the filter model [116]. The feature filter model filters the variables independently of the classification algorithm. In this way, an initial analysis is performed on the training data and afterwards the selected features are fed to the classifier. The initial analysis constructs subsets of variables by ranking the features. A simple filtering technique ranks or scores each variable based on some measure like the χ^2-test, information gain criterion, mutual information, cross-entropy measure, Fisher discriminant criterion, Pearson correlation coefficient or the Kruskal-Wallis test. Apart from these simple ranking methods, more advanced methods like FOCUS or Relief exist [5, 112]. FOCUS exhaustively searches through the feature subspace and looks for minimal combinations of features that discriminate between the classes in a perfect way. The method starts with single features and progressively adds features until the stopping criterion is reached. Relief randomly draws cases from the training set and adjusts feature weights based on the nearest neighbor from the same and the opposite class. Both FOCUS and Relief have been used in conjunction with decision trees, Bayesian classifiers [118] and nearest-neighbor classifiers [26].

Because the classification algorithm is never used, the main advantage of the filter model is its low computational cost. On the contrary, the weakest point of the filter method is that it completely ignores the impact of the learning algorithm. In filter models, the performance of a specific feature subset is not tested with the classification technique. Therefore, it has been claimed that the selection procedure should take the learning algorithm into account [100]. This leads to the second category of selection methods: feature selection techniques using the wrapper model [116]. In this class of models, a subset of variables is selected based on a search by actually using the classifier. Different subsets of features are tried on the classification algorithm to estimate the performance of each set, after which the best set is kept. Within this approach, the classifier can be used as a black

box. As an exhaustive search through the input space is not feasible, heuristic search methods using backward, forward or stepwise variable selection are often used [161]. Backward variable selection consecutively removes the feature from the model that is least significant until all variables in the model satisfy a specific level of significance. In forward selection variables are entered until a level of significance is reached, and stepwise selection sequentially adds or removes variables from the model according to a significance level. In addition, more sophisticated methods like best-first search are also able to traverse the space of subsets [76, 116]. To evaluate each subset, v-fold cross-validation or leave-one-out cross-validation can be used. It has been concluded that the wrapper models result in an increased accuracy because possible effects from the classification algorithm are taken into account [116]. Wrapper models consider how the algorithm and the training set interact. The disadvantage of wrapper methods is the high computational cost of the search.

Apart from the filter and wrapper methods, there also exist embedded methods. These methods aim to immediately integrate the variable selection or weighting procedure into the learning algorithm. Least absolute shrinkage and selection operator (LASSO) computes weights for each variable according to their importance using a regularization procedure [220]. Variable-kernel similarity metric (VSM) is a Gaussian RBF technique that uses a set of weights in a weighted distance measure to calculate the distance between new data and training data [134]. By using gradient descent techniques, the weights are computed. The embedded method WINNOW [132] is an online learning algorithm with linear threshold functions that uses feature weights. Based on prior parameter settings in the learning rule, the weights of the relevant features increase in an exponential way while the weights of irrelevant features decrease in an exponential way. Recursive feature elimination (RFE) in conjunction with SVMs is also a popular embedded method [82]. An overview of other integrated techniques can be found in [16].

Metabolic information, as acquired by MRS or MRSI measurements of human brain tumors, was used to build classifiers for diagnostic purposes and the combination of the spectral data and the MRI data was suggested [217, 216]. As these sources of data consist of an enormous amount of variables, automated feature selection is often performed to reduce the number of variables and improve the classification performance. However, the question remains whether the automatically selected features represent relevant biological information to the classification problem. This chapter of the thesis investigates the usefulness of automated feature selection to obtain biologically relevant and promising features for differential diagnosis of brain tumor tissue based on two MR data sets.

4.2 Materials

In vivo data for experiments comprise single-voxel MR spectra and multi-voxel MR spectra with corresponding MR images. All data were acquired within the context of the INTERPRET project (*International network for pattern recognition of tumours using magnetic resonance*, 2000−2002, http://azizu.uab.es/INTERPRET/ , accessed on June 27, 2009) [219].

4.2.1 MRS data set

The MRS data set contains 211 short TE MR spectra from the INTERPRET database [101]. Data were acquired by 6 contributing centers: Centre Diagnostic Pedralbes (CDP), Fundacion para la Lucha contra las Enfermedades Neurologicas de la Infancia (FLENI), Institut de Diagnostic per la Imatge (IDI), Uniwersytet Medyczny w Lodzi (MUL), St George's Hospital Medical School (SGHMS) and UMCN using 1.5 T Siemens, General Electric (GE), and Philips instruments by STEAM or PRESS with TR between 1600 and 2020 ms, TE of 20 or 30−32 ms, SW of 1000−2500 Hz, and 512, 1024, or 2048 data points. Each spectrum was subjected to a strict validation process set up by the INTERPRET Clinical Data Validation Committee and expert spectroscopists [219]. Tumor assignment was performed according to the histological classification of the CNS tumors established by the WHO [133]. Four types of brain tumors were included: glioblastomas (84 patients), meningiomas (57 patients), metastases (37 patients) and low-grade gliomas (33 patients).

Each signal is preprocessed according to the INTERPRET protocol. The fully automated preprocessing pipeline includes eddy current correction, applied to the water-suppressed FID of each case using Klose's method [114]. The residual water resonance is removed based on HLSVD time domain selective filtering using 10 singular values and a water region of [4.33,5.07] ppm. An apodization with a Lorentzian function of 1 Hz of damping is applied. Before transforming the signal to the frequency domain using the FFT, an interpolation is needed in order to increase the frequency resolution of the low resolution spectra to the maximum frequency resolution used in the acquisition protocols. This is carried out with the zero filling procedure. The baseline offset, which was estimated as the mean value of the region [9,11] ∪ [−2,−1] ppm, is subtracted from the spectrum. The normalization of the spectral data vector to the L_2-norm is performed based on the data points in the region [−2.7, 4.33] ∪ [5.07,7.1] ppm. Depending on the SNR and the tumor pattern, an additional frequency alignment check of the spectrum is performed by referencing the ppm-axis to (in order of priority) tCr at 3.03 ppm or to tCho containing compounds around 3.21 ppm or the mobile Lip around 1.29

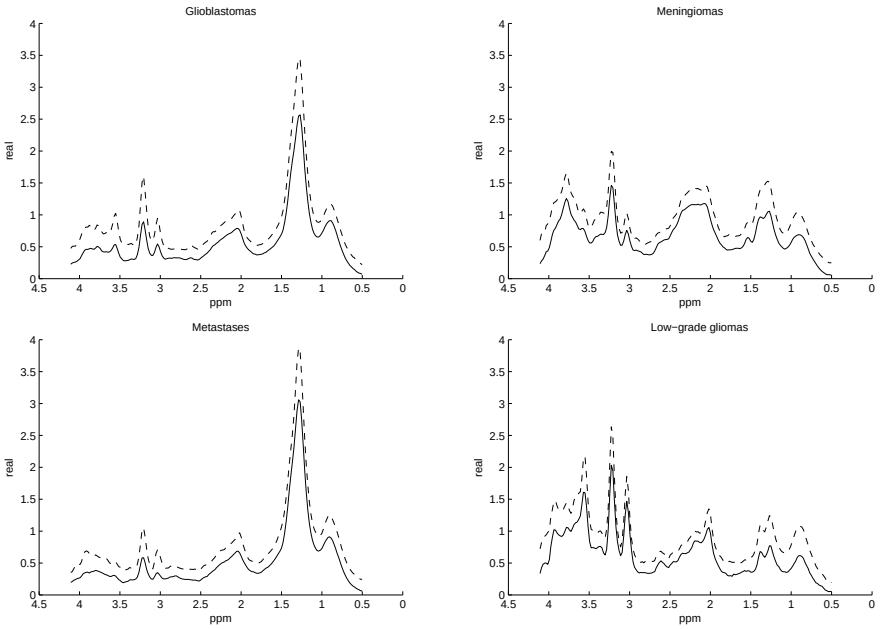


Figure 4.1: *Mean L_2-normalized real MR spectra for glioblastomas, meningiomas, metastases and low-grade gliomas. The solid lines are the means, while the dashed lines are the means plus the standard deviations.*

ppm. Finally, the region of interest is restricted to [0.5,4.1] ppm, obtaining a vector of 190 points for each spectrum. The resonances of the main metabolites are included, while the contribution of the residual water is expected to be minimal (Figure 4.1).

In contrast to quantification approaches as AQSES and QUEST (Chapter 3), the automated feature selection methods, presented in Section 4.3, do not impose biological prior knowledge using in vitro metabolite profiles for example. In order to impose prior knowledge and to start from biologically interpretable features, all resulting signals were first quantified by applying simple peak integration to the real part of the MR spectra. The following regions were integrated [79]: Cr1 ([3.845,3.995] ppm), Glx1 ([3.665,3.815] ppm), Tau + Glc ([3.345,3.495] ppm), mI + Gly ([3.450,3.680] ppm), mI + Tau ([3.195,3.345] ppm), Cho ([3.115,3.265] ppm), Cr2 ([2.955,3.105] ppm), Glx2 ([2.270,2.510] ppm), NAA ([1.935,2.085] ppm), Ala ([1.395,1.545] ppm), Lac ([1.235,1.385] ppm), Lip1 ([1.125,1.275] ppm) and Lip2 ([0.825,0.975] ppm). For the interpretability of the results, feature selection is restricted to these 13 variables, corresponding to areas where metabolite peaks can be present.

4.2.2 MRI and MRSI data set

MRI and MRSI data from 24 patients with a brain tumor and 4 healthy volunteers were acquired in UMCN and selected from the INTERPRET project database. The study was approved by the ethical committee of UMCN and followed the rules of the WHO. All data passed a strict quality control and the tumor type was determined by histopathological consensus. To create a sufficiently large data set for the machine learning techniques and the statistical methods, several voxels, situated in the tumor area, were selected from each patient based on histological, radiological, and spectroscopic information. Voxels for a specific tissue class were only selected from patients for which this tissue type was confirmed by a histopathological consensus, performed by three independent pathologists. Because of the heterogeneity of brain tumors, the MRI data and the spectral information were taken into account during the selection of voxels. For this purpose, for each patient all the high-resolution MR images were plotted together with the MRSI spectra and a segmented image in which voxels from MRSI were clustered by a model-based algorithm using a graphical user interface [239]. Since the clustering provided an objective segmentation, this was considered to be helpful for voxel selection. Based on all these data sources, an expert in MRS selected voxels if they were clearly within the affected brain region, avoiding contamination from normal cerebral tissue, such that spectra under consideration were found to be typical for that pathology. Although this method was subjective, it was chosen because tumors are known to be heterogeneous and because there is no ground truth in the diagnosis of brain tumors at the voxel level. Also, spectra from CSF and normal tissue measured from volunteers and patients were selected. Six classes of tissue were considered: normal brain tissue from volunteers and apparently normal tissue from the contralateral half of the brain of patients (218 voxels from 8 persons), CSF (100 voxels from 8 patients), grade II glioma (176 voxels from 10 patients), grade III glioma (69 voxels from 5 patients), glioblastoma (70 voxels from 7 patients), and meningioma (48 voxels from 3 patients). The data were acquired by a 1.5 T Siemens Vision Scanner with CP-head coil. Four MR images were acquired, namely T_1-weighted MR images (TE/TR = 15/644 ms), T_2-weighted MR images (TE/TR = 98/3100 ms), PD-weighted MR images (TE/TR = 16/3100 ms) and contrast-enhanced T_1-weighted MR images (15 ml 0.5M Gd-DTPA), with matrix size equal to 256×256. Both water-suppressed and water-unsuppressed MR spectroscopic images were measured using a 2D STEAM sequence with the STEAM box positioned totally in the brain (16×16×1024 samples, TR/TE/mixing time (TM) = 2000 or 2500/20/30 ms, slice thickness = 12.5 or 15 mm, field of view (FOV) = 200 mm, SW = 1000 Hz and number of signal averages = 2). Disturbing signals arising from the fat tissue surrounding the skull were avoided. The location of the STEAM box was determined using the contrast-enhanced T_1-weighted MR

image showing the largest tumor area. The MRSI slice was centered around an MRI slice of 5 mm. MR image registration was performed based on the algorithm described in [146]. It was assumed that the MRSI data were registered with the PD-weighted MR image since they were acquired in a consecutive manner.

Preprocessing of MRSI includes filtering of k-space data by a Hanning filter of 50 % using the LUISE software package (Siemens, Erlangen, Germany), zero filling to 32×32, spatial 2D Fourier transformation to obtain time domain signals for each voxel, correction for eddy current effects by a technique which prevents occasional occurrence of eddy current correction induced artifacts [114, 205], frequency alignment, water removal using HLSVD in the region [4.3,5.5] ppm, and a simple baseline correction using an exponential filter with a width of 5 ms followed by subtraction of the residual of the original signal. All first-order phases are corrected by first manually optimizing the mean spectrum, which was calculated from all spectra in the STEAM box of each patient's MRSI data. Next, this correction is applied to each separate signal of the patient's MRSI data. Finally, the spectra are normalized using the water signal. For each metabolite signal, a corresponding water-unsuppressed signal was available, acquired with the same acquisition parameters as the metabolite signal and originating from the same voxel. The amplitude of the water-unsuppressed signal was estimated by a nonlinear least squares algorithm [43]. The water-unsuppressed signal was modeled in the time domain by a Voigt model, with an additional first-order term that corrected for eddy currents. HSVD was used to obtain initial parameter estimates, which were subsequently used as starting values in a nonlinear least squares algorithm to obtain the final parameter estimates. Each spectral value in the preprocessed water-suppressed spectrum was divided by the obtained estimate of the intensity of the water peak. The magnitude spectra are computed and the region of interest is restricted to [0.5,4.0] ppm (Figure 4.2). Part of the MRSI and MRI preprocessing was performed at the Institute for Molecules and Materials, Analytical Chemistry, Chemometrics Research Department of Radboud University Nijmegen.

Similarly to the MRS data set, peak integration is used for quantification to extract biologically interpretable values [79]: Cr1 ([3.885,4.015] ppm), Glx1 ([3.685,3.815] ppm), mI + Gly ([3.495,3.625] ppm), Tau + Glc ([3.375,3.505] ppm), Cho ([3.135,3.265] ppm), Cr2 ([2.955,3.095] ppm), Glx2 ([2.135,2.265] ppm), NAA ([1.955,2.085] ppm), Lip1 + Lac + Ala ([1.265,1.395] ppm) and Lip2 ([0.835,0.965] ppm). Feature selection is applied to the combination of these 10 peak-integrated values and the 4 MRI variables (T_1, T_2, PD, Gd), which are obtained by lowering the resolution of the 4 MR images to the one of the MRSI grid by averaging pixel intensities within each voxel.

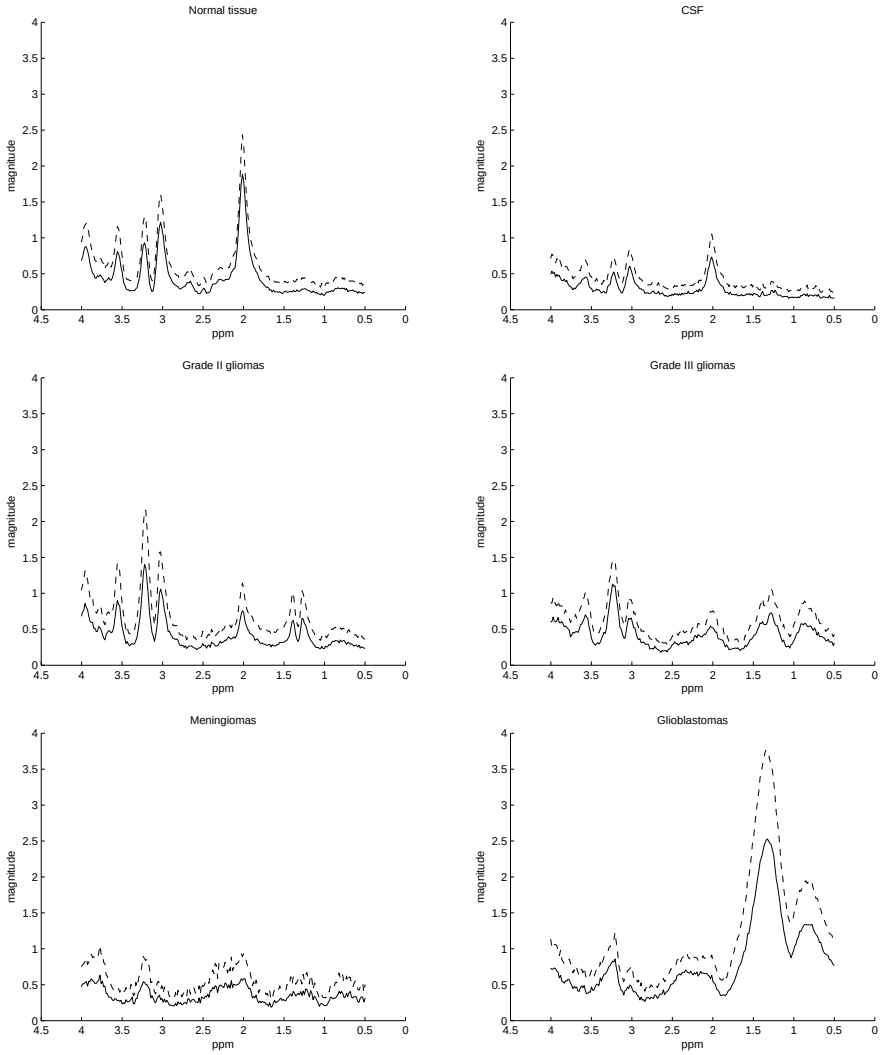


Figure 4.2: *Mean water-normalized magnitude MR spectra for normal tissue, CSF, grade II gliomas, grade III gliomas, meningiomas, and glioblastomas. The solid lines are the means, while the dashed lines are the means plus the standard deviations.*

4.3 Feature selection methods

Feature selection is performed based on three filter models: the Kruskal-Wallis test, the Fisher discriminant criterion and the Relief-F algorithm. Relief-F is an improved version of the original Relief algorithm [194]. Relief-F can be used for multi-class problems, it is more robust and it can handle noisy data. In addition, an embedded method, called automatic relevance determination (ARD), is included. The ARD mechanism that is used in this chapter is established based on Bayesian LS-SVMs [230]. In total, these four methods are used for variable selection in this chapter. In the next paragraphs, the techniques are described in detail and the experimental setup is explained.

4.3.1 Fisher discriminant criterion

Fisher's criterion is a simple measure to compare the discriminability between two classes in one dimension [59]. The criterion, which takes the mean of the classes and the within-class scatter into account, is given by

$$FC = \frac{(\mu_1 - \mu_2)^2}{(\sigma_1^2 + \sigma_2^2)}, \tag{4.1}$$

where μ represents the mean, σ denotes the standard deviation, and the subscripts denote the class numbers. For each variable in the data set a score is obtained and the variables are ranked according to their scores.

4.3.2 Kruskal-Wallis test

The Kruskal-Wallis test is a nonparametric alternative to the well-known one-way independent-samples ANOVA [91, 161]. In contrast to ANOVA, this test does not assume a normal distribution for the data. However, the Kruskal-Wallis test is less powerful than ANOVA when the ANOVA assumptions are met. The null hypothesis of the test is that the samples come from populations with equal medians. The Kruskal-Wallis test statistic is formulated as

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{n_C} n_i (r_i - r)^2, \tag{4.2}$$

or alternatively

$$H = \frac{12}{N(N+1)}(\sum_{i=1}^{n_C} n_i {r_i}^2) - 3(N+1), \tag{4.3}$$

with

$$r_i = \frac{1}{n_i}\sum_{j=1}^{n_i} r_{ij}, \quad r = \frac{1}{n_C}\sum_{i=1}^{n_C} r_i, \tag{4.4}$$

where n_i denotes the number of samples in class i, n_C is the number of classes, r_{ij} represents the rank of sample j in class i after ranking all the cases of all classes and N equals the total number of samples. When two values are equal, the average of the two ranks for which they tie is assigned. The value for the Kruskal-Wallis test should be compared with the χ^2 statistic with $n_C - 1$ degrees of freedom if the sample size within each class is large enough (e.g. > 5). This score is derived for all features and allows to rank them.

4.3.3 Relief-F

In contrast to many heuristic measures for feature selection, Relief-F does not assume conditional independence of the variables [194]. The main idea of Relief-F is to estimate the quality of features based on how good their values discriminate between data points that are close. Consecutively, random data points are drawn from the data set. Each time the k (e.g. 10) nearest neighbors of the same class and the opposite class are determined. Based on these neighboring cases the weights of the attributes are adjusted (Algorithm 2).

If the selected case has different values for an attribute compared to the neighbors of the same class, the weight of that attribute decreases. On the other hand when the attribute value for the specific case is different from the ones of the opposite class neighbors, the weight of the attribute increases. The contribution of all the hit and missing neighbors is averaged in the weight computation rule. The difference measure is defined as

$$\text{diff}(j, x_1, x_2) = \frac{|x_{1j} - x_{2j}|}{\text{range}(j)}, \tag{4.5}$$

Algorithm 2 : *Relief-F*

Input:
m = number of random data points to draw
d = number of features
k = number of nearest neighbors
x = data
y = labels of the data

for $j = 1, \ldots, d$ **do**
  $w_j = 0$
end for

for $t = 1, \ldots, m$ **do**
  randomly select an instance (x_i,y_i)
  compute the k nearest data points, h_a, with similar y_i value
  compute the k nearest data points, m_a, with different y_i value
  for $j = 1, \ldots, d$ **do**
    $w_j = w_j - \sum_{a=1}^{k} \frac{\text{diff}(j,x_i,h_a)}{mk} + \sum_{a=1}^{k} \frac{\text{diff}(j,x_i,m_a)}{mk}$
  end for
end for

Output:
w = weights indicating quality of the features

where x_{1j} and x_{2j} are the values of feature j for instance x_1 and instance x_2, respectively. The range of a feature is defined as the difference between its maximum and minimum value over all data. Analogously to the Fisher discriminant criterion and the Kruskal-Wallis test, the features can be ranked according to their scores, however, the Relief-F algorithm is computationally more expensive than these other approaches.

4.3.4 Automatic relevance determination for LS-SVMs

ARD is implemented within the context of Bayesian LS-SVMs by introducing a diagonal weighting matrix in the kernel function [145, 230]. The strategy of the ARD procedure is to determine a weight for every input of the kernel function. The weighing matrix is defined by

$$U = \text{diag}(u_1 \ ... \ u_d). \tag{4.6}$$

In this chapter, an RBF kernel is used and this implies the following form for the kernel

$$K(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^T U (x_i - x_j)}{\sigma^2}\right), \tag{4.7}$$

which can be reformulated as

$$K(x_i, x_j) = \exp(-(x_i - x_j)^T \overline{U}(x_i - x_j)) \quad \text{with} \quad \overline{U} = \text{diag}(\overline{u}) = U/\sigma^2. \tag{4.8}$$

To infer these weights, model comparison is done on the third level of inference (Equation (2.34)). The algorithm requires good starting values for $\overline{u}$ in order to solve the optimization problem. As all features should be treated in an equal way, equal starting values are chosen. Therefore, σ is first inferred based on $u = [1 \ldots 1]^T$. The point $[1/\sigma^2 \ldots 1/\sigma^2]^T$ serves as the starting point for the optimization problem (Algorithm 3).

Algorithm 3 : *Automatic relevance determination*

1: Input:
2: x = data
3: y = labels of the data
4:
5: normalize the input data x to zero mean and unit variance
6: infer σ from Equation (2.34). Therefore, the first (Equation (2.32)) and second (Equation (2.33)) level of inference need to be optimized to obtain values for μ and ζ.
7: when step 6 was the previous step the initial choice for $\overline{u}_{init}$ is set to $[1/\sigma^2 \ldots 1/\sigma^2]^T$, otherwise feature l is removed from $\overline{u}$ to obtain $\overline{u}_{init}$
8: optimize $\overline{u}$ with starting value $\overline{u}_{init}$ from Equation (2.34)
9: the dimension of the input space is lowered by removing feature l with lowest $\overline{u}$ score
10: return to step 7 as long as the model posterior in Equation (2.34) improves
11:
12: Output:
13: $\overline{u}$ = weights indicating relevance of the features

Relevant features obtain large weights while less relevant features are associated with smaller weights. Instead of performing a backward variable selection procedure based on ARD, only one iteration of the algorithm is performed to reduce the computational cost.

4.3.5 Experimental setup

For each pair of classes in the separate data sets, the filter methods (Fisher discriminant criterion, the Kruskal-Wallis test, Relief-F) and the embedded method (ARD) are applied in a bootstrap aggregating analysis [20]. This so-called bagging procedure involves resampling (i.e. 50 times) with replacement of the data for each pairwise problem. In each of the runs the feature filter methods assign a rank number to each variable, while ARD assigns a weight. These ranks or weights are averaged over the 50 runs, resulting into 4 values for every single combination of variable and classification problem (Table 4.1). The use of bagging within the context of feature selection for MR spectroscopy data was proposed in [135]. For the Kruskal-Wallis test, the Fisher criterion and Relief-F values between 1 and 13 (or 14, depending on the number of features in the data set) were obtained. The smaller this number, the better the feature is able to distinguish between the two tissue classes. For ARD, weights between 0 and 1 are obtained. A weight that is closer to 1 indicates a more relevant feature. A feature was considered to be potentially relevant if the ARD weight was higher than 0.2 or if at least two out of three filter methods obtained a rank smaller than 6 or if the ARD weight was higher than 0.1 and one filter method obtained a rank smaller than 6. A feature was labeled as most relevant if the ARD weight was higher than 0.2 and if all filter methods obtained ranks below 6.

4.4 Results

The data sets enable to investigate a wide range of binary classification problems. Several of these classification problems are clinically important for differential diagnosis (e.g. those involving grade II gliomas, grade III gliomas and glioblastomas). Others are less important, occur less frequently (i.e. involving atypical meningiomas in the differentiation from glioblastomas or metastases [147]), or are not important for differential diagnosis, but are important for the generation of nosologic images (Chapter 8). These binary problems, having no direct clinical significance, served as validation of the obtained results.

The selected input features for the various types of differential diagnosis are provided in the order of clinical importance. The results for the MRS data set are depicted in Tables 4.1 and 4.2. As an example, Table 4.1 provides the averaged rank or weight values, obtained by the feature selection methods, for all features for the separation of glioblastomas and low-grade gliomas. Most relevant metabolites are printed in boldface, while boldface printed weights and ranks refer to features that are potentially relevant. To distinguish glioblastomas from low-grade gliomas mI+Gly, Cr (Cr2, around 3.05ppm), and Lac are detected to be most relevant, and

additionally Glx1, Tau+Glc, Glx2, Lip1 and Lip2 seem to be potentially relevant. In summary, Table 4.2 lists the results for all possible types of differential diagnosis.

For the combined MRSI and MRI data set a similar analysis is performed. Remark that the MRS data set did not contain spectra from healthy tissue, CSF and grade III gliomas, whereas the MRSI and MRI data set did not contain data from metastases, resulting in a partly different set of differential diagnoses. Again, as an example, the results for the differential diagnosis of grade III gliomas and glioblastomas are listed in Table 4.3. Based on the rank and weight values obtained with the filter methods and the embedded method mI+Gly, Lip1+Lac+Ala, and T_1 are found to be most relevant, and additionally, Cho and Gd provide potentially relevant information. Table 4.4 presents the selected features for all possible types of differential diagnosis.

4.5 Discussion

The metabolites and image features, found by the different feature selection methods, are evaluated on their relevance [98, 95, 166, 56, 165].

MRS data set

For the differentiation of low-grade gliomas (of which the majority consisted of grade II astrocytomas) from glioblastomas, most selected features correspond to what is known from literature: mI is mostly found in astrocytes and reduces with grade [28], Cr is reduced in astrocytomas but further reduced in glioblastomas, and Lip1, Lip2 and Lac are characteristics of high-grade tumors (and increase with tumor grade). Glx-related peaks are also reduced in glioblastomas (although not observed in [165]). The selection of Tau+Glc corresponds to visible differences in the mean tissue levels between both classes (although not mentioned in [95]). According to [166] Lip have characteristics that might allow discrimination between metastases and glioblastomas, while Lac is frequently elevated for glioblastomas. The same holds for Cho and mI, whereas Glx was reduced in [56], but slightly elevated in our situation. NAA showed no significant difference in [56], nor in our situation, but was selected. Investigating the data it appeared that NAA, for a considerable number of glioblastoma patients, showed higher tissue levels compared to those of the metastasis class. The selection of Cr corresponds with a higher average tissue level for glioblastomas compared to metastases, similar to [98], although [166] found no clear difference. Regarding Tau and Ala: their choice is supported by small differences in average tissue levels (Tau+Glc slightly

Table 4.1: *Averaged rank and weight values for the features of the MRS data set obtained with various feature selection methods for the differentiation of glioblastomas and low-grade gliomas. The columns refer to the metabolites, the rows depict the feature selection methods (ARD: automatic relevance determination, R-F: Relief-F, FC: Fisher discriminant criterion, K-W: Kruskal-Wallis test). Metabolites in boldface (mI+Gly, Cr2 and Lac) refer to features that are found to be most relevant by the feature selection methods. Numbers in boldface indicate features (Glx1, Tau+Glc, Glx2, Lip1 and Lip2) that are potentially relevant based on the heuristics in Section 4.3.5.*

	Cr1	Glx1	Tau+Glc	**mI+Gly**	mI+Tau	Cho	**Cr2**	Glx2	NAA	Ala	**Lac**	Lip1	Lip2
ARD	0.1944	**0.3161**	**0.4709**	**0.7554**	0.0794	0.0188	**0.4589**	**0.6167**	0.0077	0.0126	**0.6979**	0.0227	**0.4259**
R-F	11.060	7.2600	9.7800	**4.6400**	6.1400	**3.7200**	**3.6000**	10.600	11.040	9.1800	**1.4200**	**3.3400**	9.2200
FC	9.9000	**5.9400**	6.7200	**3.2400**	**5.8600**	7.6600	**4.1200**	11.880	12.320	8.2200	**1.2600**	**2.3200**	11.560
K-W	9.6000	**5.3600**	6.2600	**2.8000**	6.2400	8.4200	**4.3400**	12.080	12.660	8.1200	**1.3000**	**2.9000**	10.920

Table 4.2: *The selected features for each differential diagnosis in the MRS data set. The columns refer to the metabolites, the rows depict the differential diagnosis (GBM: glioblastomas, LGG: low-grade gliomas, MET: metastases, MEN: meningiomas). Features that are labeled as most relevant by the feature selection methods are denoted by '$\underline{x}$', while potentially relevant features are denoted by 'x'. A (potentially) relevant feature that is colored in gray is mentioned in literature as being relevant for the given differentiation.*

Task	Cr1	Glx1	Tau+Glc	mI+Gly	mI+Tau	Cho	Cr2	Glx2	NAA	Ala	Lac	Lip1	Lip2
GBM-LGG		x	x	$\underline{x}$			$\underline{x}$	x			$\underline{x}$	x	x
GBM-MET	x	x	x	x	x	$\underline{x}$	x	$\underline{x}$	x	x	$\underline{x}$	x	x
MEN-MET		$\underline{x}$			x		x	$\underline{x}$		x	$\underline{x}$	$\underline{x}$	x
GBM-MEN		$\underline{x}$	x	x		x	x	$\underline{x}$			$\underline{x}$	x	x
MET-LGG				$\underline{x}$	x	x	$\underline{x}$		x		$\underline{x}$	$\underline{x}$	x
MEN-LGG			x	$\underline{x}$		x	$\underline{x}$	$\underline{x}$		x			

Table 4.3: *Averaged rank and weight values for the features of the combined MRSI and MRI data set obtained with various feature selection methods for the differentiation of grade III gliomas and glioblastomas. The columns refer to the metabolites, the rows depict the feature selection methods (ARD: automatic relevance determination, R-F: Relief-F, FC: Fisher discriminant criterion, K-W: Kruskal-Wallis test). Metabolites in boldface (mI+Gly, Lip1+Lac+Ala and T_1) refer to features that are found to be most relevant by the feature selection methods. Numbers in boldface indicate features (Cho, Gd) that are potentially relevant based on the heuristics in Section 4.3.5.*

	Cr1	Glx1	**mI+Gly**	Tau	Cho	Cr2	Glx2	NAA	**L1+L+A**	Lip2	T_1	T_2	PD	Gd
ARD	0.0059	**0.1261**	**0.4745**	0.0237	**0.1363**	0.0108	0.0116	0.0011	**0.9076**	0.0024	**0.8778**	0.0414	0.0121	**0.1420**
R-F	10.560	12.680	**3.8800**	9.4200	**5.1600**	11.600	8.2600	13.240	**2.3000**	8.2800	**1.0000**	8.3000	6.1800	**4.1400**
FC	8.8000	7.6200	**2.3000**	7.8200	**3.9600**	6.7200	12.380	13.900	**2.8600**	8.9600	**1.0000**	12.540	8.1200	8.0200
K-W	8.5000	7.8400	**3.0000**	7.1200	**4.1000**	7.4000	12.540	13.820	**1.9800**	9.7400	**1.0400**	12.540	6.9200	8.4600

Table 4.4: *The selected features for each differential diagnosis in the combined MRSI and MRI data set. The columns refer to the metabolites, the rows depict the differential diagnosis (NOR: normal tissue, CSF: cerebrospinal fluid, GRII: grade II glioma, GRIII: grade III glioma, GBM: glioblastoma, MEN: meningioma). Features that are labeled as most relevant by the feature selection methods are denoted by 'x', while potentially relevant features are denoted by 'x'. A (potentially) relevant feature that is colored in gray is mentioned in literature as being relevant for the given differentiation.*

Task	Cr1	Glx1	mI+Gly	Tau	Cho	Cr2	Glx2	NAA	L1+L+A	Lip2	T_1	T_2	PD	Gd
GRII-GRIII									x	x	x	x	x	x
GRIII-GBM			x		x				x		x			x
GRII-GBM			x			x			x	x		x		x
NOR-GRII					x	x		x	x			x	x	
NOR-GRIII						x		x	x		x		x	
NOR-GBM			x			x		x	x				x	x
GBM-MEN	x	x		x			x		x	x			x	x
GRIII-MEN			x		x	x	x				x		x	x
CSF-GRII					x		x		x		x	x	x	x
CSF-GRIII					x				x	x			x	x
CSF-GBM				x					x	x	x			x
NOR-MEN			x			x		x	x		x		x	x
GRII-MEN			x			x					x	x	x	x
NOR-CSF						x		x			x	x	x	x
CSF-MEN				x			x		x		x	x		x

higher and Ala slightly lower for glioblastomas). The behavior that most features are selected is likely due to the fact that the differentiation between glioblastomas and metastases based on single-voxel MR spectroscopy is a difficult problem, as their mean spectra suggest. A more promising approach based on multi-voxel data is proposed in Chapter 9. Regarding the other clinically less interesting differential diagnoses it can be stated that most of the selected features also correspond to what is known from literature (e.g. Glx is a prominent feature of meningiomas). The features labeled as most relevant were mentioned in the literature as showing significant differences. For the differentiation of meningiomas from metastases Cr2 is selected, although it has been reported to be low or absent in meningiomas (as well as in metastases). The selection, however, is supported by a clear difference in tissue level (peak area), which is probably due to the applied integration method (including baseline and normalization effects; cf. Figure 4.1). This is also the reason for the selection of mI-related features and Cr2 for the differentiation of glioblastomas from meningiomas. For the differentiation of meningiomas from metastases the overlap in integrated ppm ranges is likely to be the reason for the selection of mI+Tau above the selection of Cho, which is mentioned in the literature.

Combined MRSI and MRI data set

For the multi-voxel data set, MRI information is added (Table 4.4). The information from the images seems to be valuable for most classification problems, similar to [44, 204, 67]. Furthermore, there is considerable correspondence between the features found to be relevant by the filter models and the embedded method (Table 4.3). In the following paragraphs the various differential diagnoses are discussed, including the clinically less interesting differentiations as well, as these can be part of a brain tissue typing system using this kind of data (Chapter 8). Moreover, these problems serve as additional control for the validity of the selected features.

Differentiations of gliomas

Most selected features correspond to what has been reported in literature for the differentiation of the various grades of glial tumors (e.g. the importance of mI+Gly and Lip for the differential diagnosis of grade II gliomas and glioblastomas). With respect to the differentiation of grade II and grade III gliomas, mI(+Gly) is not selected. The mean tissue levels for the metabolite show no difference for the data set. Cho is selected in the differentiation of grade III gliomas from glioblastomas as its tissue level is reduced in glioblastomas, but similar to literature the tissue level difference appears to be small between grade II and grade III gliomas. The filter models and the embedded method do not select this feature for the differentiation

of those grades. The same holds for Glx2 and NAA, which metabolite values are likely to be influenced by the presence of MMs. Regarding the selected image features: high-grade tumors frequently show an enhancement on Gd images whereas they are frequently less intense on T_2-weighted MR images compared to the appearance of grade II gliomas [158]. Grade III gliomas show sometimes contrast enhancement compared to grade II gliomas and frequently show less hyperintense tumor borders compared to glioblastomas. The selection of T_1 in the various differentiations follows from a decrease in voxel intensity going from grade II to grade III gliomas and an increased intensity (compared to the grade II and grade III glioma voxels) in glioblastomas.

Differentiations of the various gliomas from normal tissue

NAA, an indicator of viable neurons, is reduced in most tumors, Lac is frequently observed in tumors, Lac and Lip are often most prominent in high-grade tumors, and Cr is reduced in astrocytes compared to normal brain. It is known that Cho can be increased in lower grade astrocytes compared to healthy tissue. According to [95], Lip2 is a characteristic feature in the differentiation of healthy tissue from glioblastomas, but it is not selected for the present data set. This is likely due to the fact that it has hardly an additional value compared to the already selected more relevant features (e.g Lip1+Lac, GD). Besides, the feature is close to the selection threshold for several methods. The choices of T_1, T_2 and PD are to be expected as the tumor area has frequently a lower intensity on T_1-weighted MR images and is hyperintense on T_2-weighted and PD-weighted MR images.

Differentiations of the various gliomas from CSF

The choices of the metabolites in the CSF versus tumor differentiations mainly correspond to those metabolites (e.g. Cho and Lip) which have in the specific tumor tissue a high (or elevated) tissue level, as all metabolite tissue levels are low or absent in CSF. The chosen image features correspond to the fact that CSF is only enhanced on T_2-weighted MR images, whereas tumors are enhanced on the other image types. The choice of Tau in CSF versus glioblastomas is probably due to the chosen integration method.

Differentiations of meningiomas from other tissues

Regarding the meningioma-related differentiations a number of general observations from literature corresponds to the selected features. Glx is selected in a number of differentiations involving meningiomas, as it is low in all other tissue types. Although the metabolite is not selected in the differentiations involving

normal tissue or grade II gliomas, the ARD method showed elevated weights for Glx2. mI is found to be low or absent in non-glial tumors. Cr (Cr2) and mI are selected in all differentiations except the one involving glioblastomas, because in the latter the Cr2 and mI+Gly level is also low. The selection of Lip1+Lac+Ala (area around 1.3 ppm) in the differentiation from normal tissue and CSF can be explained by the fact that two of the meningiomas are atypical (i.e. grade II). These tumors usually comprise necrotic tissue containing Lip [201, 147]. A macromolecular contribution to the signal in this ppm range has also been reported for meningiomas [165]. The selection of Cho corresponds to the fact that the Cho level is normally low for (benign) meningiomas, although it has been reported that they obtain intensities corresponding to grade III gliomas [165, 147]. In the present data set, despite the two atypical meningiomas, the latter is not observed. The selection of Tau is likely to be data set dependent. The selected image features in the differentiations from gliomas can be explained by comparing the average intensities of both classes, however, meningiomas are normally only hyperintense on contrast-enhanced images as compared to grade II gliomas. In the differentiation from normal tissue and CSF the selected image features correspond to their normal appearance, e.g. hypointense to isointense on T_1, isointense to hyperintense on T_2 and enhanced on contrast-enhanced images [23].

Differentiation of normal tissue from CSF

Concerning the differentiation of normal tissue and CSF, the selected MRSI features include Cr2 and NAA. Most metabolites have lower levels in CSF than in normal tissue and this effect is more pronounced for those metabolites which have higher levels in normal tissue. T_2 shows an hyperintense CSF region, whereas on for example a T_1-weighted MR image, the CSF is more hypointense. On PD images CSF appeared slightly less intense due to the applied relative short TR (i.e. 3100 ms).

4.6 Conclusions

In contrast to quantification approaches, the feature selection methods that were presented in this chapter do not require biological prior knowledge and the techniques can be applied in an automated way to any given set of features. In this chapter the methods were applied to a set of peak-integrated MRS(I) features (and MRI information) to facilitate interpretation of the results. A significant part of the automatically selected features for the single-voxel MR data, as well as the multi-voxel data, corresponded to what is known from literature and clinical experience. In addition, MR image features were found to provide additional information. These findings hold for the clinically more interesting differentiations,

and for the remaining differentiations which were checked as additional support. The results obtained with the various types of feature selection techniques were similar, but a combination of several techniques provided a more complete set of relevant features. Some of the selected features could not directly be accounted for from a biological or clinical perspective, but could be explained from the data set. These features can be data set specific small differences or can give hints to possible interesting, additional, new features. This deserves further research and clinical validation. In conclusion, the automated feature selection approaches from machine learning and statistics represent useful tools to obtain relevant and potentially interesting MRS(I) and MRI features for differential diagnosis of brain tumors and brain tissue segmentation.

Chapter 5

Feature extraction from MRS and MRSI for brain tumor classification

While Chapter 3 provided an overview of different quantification approaches, Chapter 4 focused on feature selection methods from machine learning and statistics. In this chapter, these feature extraction approaches (Section 5.3) are compared with the ultimate goal to obtain the best performing MRS(I)-based brain tumor classifiers. The chapter starts with an introduction to the problem (Section 5.1). The data, including a set of 205 single-voxel MR spectra and a set of 681 multi-voxel MR spectra, are summarized in Section 5.2. A statistical analysis is presented for both data sets in Section 5.4 and discussed in Section 5.5.

5.1 Feature extraction for MRS(I) data

Despite their promising potential for noninvasive evaluation of brain tumors, the introduction of MRS and MRSI into widespread routine clinical practice is hampered by a number of technical limitations of the currently used approaches. One of the reasons for the present situation is that currently MR spectral patterns are mostly inspected and evaluated individually, which is time-consuming and often requires specific spectroscopic expertise. To promote clinical use, more efficient data processing and evaluation techniques are still needed. Several studies have proposed the use of approaches from statistics, machine learning and pattern recognition to overcome this problem. ANNs, LDA, fuzzy techniques, distribution

plots, SVMs and LS-SVMs were applied to MR data [182, 183, 176, 216, 129, 243, 218, 203, 43, 219]. A significant advantage of these methods is that they can be automated and can potentially be integrated into a clinical DSS. However, to effectively introduce a DSS into clinical practice, some requirements have to be considered. Firstly, processing time and execution time should be minimized. In a clinical environment, it is often not acceptable to retrieve the results with a serious delay. Secondly, user interaction should be restricted to a minimum, automation is the keyword. To restrict the computation time and maximize the applicability of a DSS, it is essential to keep the preprocessing protocol of MRS(I) as concise as possible. Therefore, the importance of feature extraction within the protocol needs to be studied. In the past PCA, independent component analysis (ICA), peak integration, peak regions, genetic algorithms, forward selection procedures, wavelet coefficients as well as metabolite quantification approaches like LCModel were used to reduce the dimension of spectral patterns and select the most relevant features [43, 186, 209, 208, 53, 83, 163, 80, 124, 96, 137, 204].

Lukas et al. observed that dimensionality reduction decreased the performance on long TE MRS signals [137]. The use of complete spectra in combination with a kernel-based method yielded the best results. Menze et al. claimed that pattern recognition methods performed $5-10$ % better than a manually supervised quantification approach applied to long TE MRS data [154]. Kelm et al. argued that nonlinear methods operating on the complete spectra yielded the best results for long TE MRSI data of the prostate [110]. Devos et al. found no significant differences between different feature extraction methods applied to short TE MRS signals [43]. Simonetti et al. observed that simple peak integration performed best on short TE MRSI data [204]. Finally, Opstad et al. concluded that quantification by LCModel increased classifier performance compared to the use of whole spectra on single-voxel short TE MR spectra [165].

In this chapter, the effect of a number of feature extraction methods, including quantification approaches, on automated classification techniques for brain tumor diagnosis is studied. Two independent data sets, containing short TE multi-center MRS and MRSI signals, are used for this purpose, while various types of tissue specimen are included. The diverse feature extraction methods are explored using both a kernel-based method and a standard method for classification.

5.2 Materials

5.2.1 MRS data set

Two independent data sets are used in the analysis. The first data set contains 205 single-voxel MR spectra from the short TE INTERPRET database. This data set comprises MR spectra from 6 contributing centers: CDP, FLENI, IDI, MUL, SGHMS and UMCN. Data were acquired at 1.5 T by STEAM or PRESS sequence at an TE between 20 and 32 ms. Detailed information about the acquisition parameters for each center is provided in Table 5.1. Quality control criteria applied to MR system performance and all data in this study passed a strict validation process [228]. Tumor assignment was done according to the rules of the WHO. There were four types of brain tumors included in this data set: glioblastomas (87 patients), meningiomas (57 patients), metastases (39 patients) and grade II astrocytomas (22 patients). These classes are the most prevalent types of tumor occurring in the brain. The distribution of the cases over the different centers is provided in Table 5.2.

Table 5.1: *Overview of the acquisition parameters for the single-voxel MRS data set. For each center the manufacturer of the scanner, sequence, echo time (TE), repetition time (TR), spectral width (SW) and the number (N) of points in the original FID are listed.*

Center	Manufacturer	Sequence	TE(ms)	TR(ms)	SW(Hz)	N
CDP	GE	STEAM	20	1600,2000	2500	2048
SGHMS	GE	STEAM	30	2000,2018,2020	2500	2048
	GE	PRESS	30	2000,2018,2020	2500	2048
FLENI	GE	PRESS	30	2000	2500	2048
IDI	Philips	PRESS	30,31,32	2000	1000	512
UMCN	Siemens	STEAM	20	2000	1000	512
MUL	Siemens	STEAM	20	2000	1000	512

The signals are preprocessed by frequency alignment, and zero-order phase correction is performed with Klose's method [114]. The residual water peak is filtered using HSVD with 15 singular values over the water region of [4.09,5.61] ppm [11]. GE data have been interpolated by piecewise cubic splines in the time domain to match the SW and number of points of Philips and Siemens data. Water normalization is performed using a corresponding water-unsuppressed signal, acquired from the same voxel with equal parameters as the metabolite signal. The amplitude of the water-unsuppressed signal is estimated. A Voigt model is used to model the water-unsuppressed signal in the time domain, with an additional first-order term correcting for eddy currents [149]. HSVD is used to get initial parameter estimates as starting values for a nonlinear least squares algorithm to obtain final parameter estimates. Next, each value in the spectrum is divided by the resulting estimate of the intensity of the water peak. An exponentially decreasing

apodization function is used for baseline correction [203, 43]. Afterwards the baseline model is subtracted from the original signal. Finally, the signals are Fourier transformed.

Table 5.2: *Number of MR spectra per center for each class in the MRS data set (GBM: glioblastomas, MEN: meningiomas, MET: metastases, G2A: grade II astrocytomas).*

Center	GBM	MEN	MET	G2A
CDP	30	12	5	4
SGHMS	18	12	16	10
FLENI	2	0	0	1
IDI	31	30	17	5
UMCN	2	1	0	1
MUL	4	2	1	1
Total	87	57	39	22

5.2.2 MRSI data set

The second data set, containing multi-voxel MR data, is identical to the data that are described in Section 4.2.2. In total, 681 magnitude MR spectra from six tissue classes are included in this data set: normal brain tissue from volunteers and apparently normal tissue from the contralateral half of the brain of patients, CSF from patients, grade II gliomas, grade III gliomas, meningiomas and glioblastomas. For clarity, the distribution of the cases over the classes is summarized in Table 5.3.

Table 5.3: *Number of spectra and number of persons (i.e. patients and/or healthy volunteers) for each class in the MRSI data set.*

Tissue type	Spectra	Persons
Normal tissue	218	8
CSF	100	8
Grade II gliomas	176	10
Grade III gliomas	69	5
Meningiomas	48	3
Glioblastomas	70	7

5.3 Feature extraction methods

This study focuses on standard methods that are often applied in the field of MR spectroscopy (Table 5.4). Quantification methods from Chapter 3 and feature selection approaches from Chapter 4 are included. In addition, approaches as PCA, ICA and forward selection are added to the analysis. As user interaction needs to be minimized, all methods in this chapter are applied in a fully automated manner. For classification, a classical supervised learning algorithm is used, as well as a kernel-based algorithm that handles high-dimensional input vectors; both LDA and LS-SVMs are applied. Based on the results in [44], an RBF kernel is chosen for LS-SVMs in the experiments.

Table 5.4: *The different combinations of feature extraction technique and classification algorithm.*

Label	Feature extraction method	Classifier
1	Bagging+Fisher criterion	LS-SVM
2	Bagging+Relief-F	LS-SVM
3	Bagging+Kruskal-Wallis	LS-SVM
4	Full spectra	LS-SVM
5	ICA	LDA
6	ICA	LS-SVM
7	PCA	LDA
8	PCA	LS-SVM
9	Peak regions	LS-SVM
10	Peak integration	LDA
11	Peak integration	LS-SVM
12	Stepwise selection	LDA
13	AQSES	LS-SVM
14	QUEST	LS-SVM

5.3.1 Full spectra

Only 138 points, which correspond with a region of interest of [0.0,4.17] ppm, are retained for the full spectra in case of the MRS data set. Figure 5.1 depicts the resulting mean magnitude MR spectra with baseline correction for each class, which are referred to as the full spectra for the MRS data set hereafter. By using magnitude spectra, phasing problems can be avoided, while no or only minor differences have been reported between performances based on magnitude spectra and real spectra in [43, 165]. For the MRSI data set a region of interest of [0.5,4.0] ppm results in 230 spectral points. In Figure 4.2 the mean magnitude MR spectra

with baseline correction, which are referred to as the full spectra for the MRSI data set, are depicted for each class. The difference between the number of spectral points (i.e. 138 and 230) for the MRSI and MRS data is caused by the difference in number of points in the original FID.

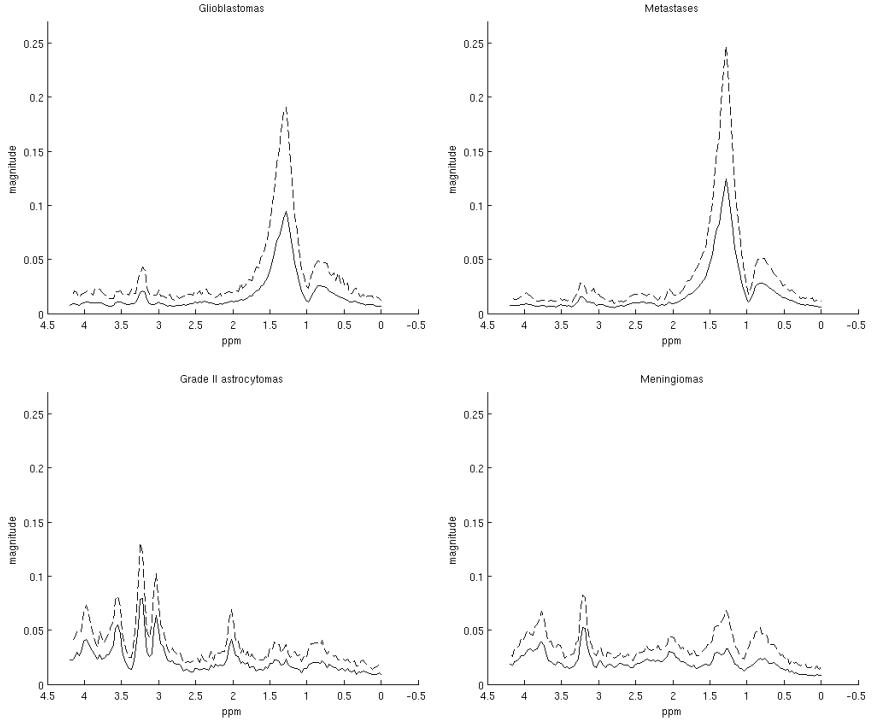


Figure 5.1: *Preprocessed mean magnitude MR spectra after baseline correction for each tissue type in the MRS data set. The solid lines represent the means, dashed lines depict the means plus the standard deviations.*

5.3.2 Peak integration

Peak integration is used to extract features and obtain measures for brain metabolites. For the multi-center MRS data set, 13 peak-integrated values are obtained based on the resonance frequencies reported in [79]. The following metabolites are selected: Cr ([3.845, 3.995] ppm), Glx ([3.665,3.815] ppm), Tau + Glc ([3.345,3.495] ppm), mI + Gly ([3.450, 3.680] ppm), mI + Tau ([3.195,3.345] ppm), Cho ([3.115,3.265] ppm), Cr ([2.955,3.105] ppm), Glx ([2.270,2.510] ppm), NAA ([1.935,2.085] ppm), Ala ([1.395,1.545] ppm), Lac ([1.235,1.385] ppm), Lip

([1.125,1.275] ppm) and Lip ([0.825,0.975] ppm). For the MRSI data, 10 peak-integrated values are computed: Cr ([3.885,4.015] ppm), Glx ([3.685, 3.815] ppm), mI + Gly ([3.495,3.625] ppm), Tau + Glc ([3.375,3.505] ppm), Cho ([3.135, 3.265] ppm), Cr ([2.955,3.095] ppm), Glx ([2.135,2.265] ppm), NAA ([1.955,2.085] ppm), Lip + Lac + Ala ([1.265,1.395] ppm) and Lip ([0.835,0.965] ppm).

5.3.3 Peak regions

Instead of integrating over the frequency regions as in Section 5.3.2, it is possible to aggregate the frequency ranges of the magnitude spectrum in a single pattern and use it as input for the classifier. Theoretically, compared to the full spectra, redundancy generated by spectral noise and artifacts is reduced in this manner. Also, these regions are assumed to contain most of the information necessary to discriminate among tissue types. More specifically, 91 variables are selected for the MRS data. For the MRSI data set, 101 features are included based on the selected peak regions.

5.3.4 Bagging with Fisher criterion, Kruskal-Wallis test and Relief-F

These filter model approaches from statistics and machine learning rank the variables independently of the LDA or LS-SVM algorithm (Chapter 4). This analysis uses the Fisher discriminant criterion, the Kruskal-Wallis test and Relief-F to rank all the frequency points of the magnitude spectrum in the region of interest. Additionally, to assure more robust results, a bagging analysis is incorporated [20]. For each training set, 50 data sets are generated by bagging. Feature ranking is performed on each of the 50 generated sets, and the most relevant features according to the three ranking criteria and the bagged sets are extracted. To have a fair comparison, the number of spectral points to select is equal to the number of variables used for classifying the peak regions as specified in the previous paragraph.

5.3.5 Independent component analysis

ICA is a technique that extracts statistically independent components from a set of measured signals [27]. The method is used for blind source separation and can extract a pure signal from a set of mixtures. As there are partial volume effects present in MRS, a signal from a specific voxel can contain components of different tissue types. This is the main motivation to apply ICA to a set of MRS

or MRSI spectra. Assuming that there are linear mixtures $x_i \in \mathbb{R}^d$, $i = 1, ..., N$, of independent source vectors $s_i \in \mathbb{R}^d$, $i = 1, ..., m$, the noiseless model becomes

$$\begin{aligned} x_1 &= a_{11}s_1 + a_{12}s_2 + ... + a_{1m}s_m, \\ &\vdots \\ x_N &= a_{N1}s_1 + a_{N2}s_2 + ... + a_{Nm}s_m, \end{aligned} \tag{5.1}$$

or in matrix notation

$$X = AS. \tag{5.2}$$

ICA aims to find matrices W and S from X, where the rows of S are as independent as possible, such that

$$S = WX. \tag{5.3}$$

As input for the method, the full region of interest of the magnitude spectrum is used. Based on the joint approximate diagonalization of eigenmatrices implementation, 13 independent components are extracted for the MRS data and 10 for the MRSI data to have a fair comparison with peak-integrated values [27].

5.3.6 Linear principal component analysis

PCA transforms the original features such that a new set of uncorrelated features is obtained [50]. The transformation allows a mapping of the vectors $x_i \in \mathbb{R}^d$ onto low-dimensional vectors $z_i \in \mathbb{R}^m$ with $m < d$. The covariance matrix is estimated by

$$S = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T, \tag{5.4}$$

with $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$. The eigenvectors u_j, $j = 1, ..., d$, and corresponding eigenvalues λ_j of S satisfy

$$Su_j = \lambda_j u_j, \ j = 1, ..., d, \ \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d \geq 0. \tag{5.5}$$

By taking the m largest eigenvalues and their corresponding eigenvectors, the transformed variables can be calculated

$$z_{ij} = u_j^T(x_i - \mu),\ j = 1, ..., m, \tag{5.6}$$

with $z_i = [z_{i1}\ ...\ z_{im}]^T$. These new features are linear combinations of the original variables and they are often ranked according to the variance they retain. Ideally, most of the variance in the original data set can be explained by a limited number of components. The full region of interest of the magnitude spectrum is used as input for the PCA analysis. For comparison with ICA and peak-integrated values, the first 13 and 10 principal components are selected for the MRS data set and MRSI data set, respectively.

5.3.7 Stepwise selection

Stepwise variable selection is a standard wrapper algorithm from statistics and machine learning to perform feature extraction [161]. Since an exhaustive search is impractical in case of high-dimensional data, a greedy hill-climbing approach is applied. The subset of features with the best performance, based on a cross-validation analysis and a measure for quantifying separation, is obtained by stepwise adding or deleting features from the model [74]. This method is applied to the full region of interest of the magnitude spectrum. Similarly to the previous approaches, the upper limit of selected variables is 13 for the MRS data set and 10 for the MRSI data set.

5.3.8 AQSES

The time domain quantitation method AQSES, which was originally designed for short TE MR spectra, is used to obtain metabolite amplitude estimates (Section 3.4). The amplitude estimates can be used as input features for a classifier and, in that sense, AQSES can be considered as a feature extraction method. The quantities of interest are the weighting coefficients of a linear combination of corrected in vitro metabolite profiles. To set up the in vitro basis set, 11 separate metabolite solutions are prepared as described in [179]: mI, Tau, Glc, PCho, Cr, Glu, NAA, Ala, Lac, Lip (1.3 ppm), Lip (0.9 ppm). The in vitro metabolite profiles are acquired by a 1.5 T Siemens scanner using a STEAM sequence with an TE of 20 ms, and a STEAM box of $2 \times 2 \times 2$ cm^3. Instead of HLSVD, the MP-FIR filter in AQSES is applied to suppress the water resonances. The components outside the frequency interval of [0.25,4.2] ppm are filtered out. In contrast to the other feature extraction methods, the baseline correction step is not performed during

preprocessing of the data set, since the macromolecular baseline is incorporated into the fit via nonparametric modeling. The regularization parameter (i.e. λ, Section 3.4) for baseline correction equals 2. No constraints are imposed on the parameters; the amplitudes, damping corrections, phase corrections and frequency shifts are free.

5.3.9 QUEST

QUEST is used for time domain quantitation of the MR spectra using the software package jMRUI-3.0 (Section 3.5). Based on the fast decay of the background signal in the time domain, the background is subtracted from the raw data after automatic estimation. The applied metabolite basis set is identical to the one in the analysis with AQSES. Again, all the correction parameters (amplitude, damping, phase and frequency) are free.

5.3.10 Experimental setup

The same experimental design and evaluation measure are used for the statistical analysis of the MRS and MRSI data set. For each pair of classes in the individual data sets, the feature extraction methods are compared using stratified repeated random sampling (Section 2.3.2). The data from two specific classes is 100 times randomly split in a stratified way in a set for training and validation (i.e. 2/3 of the data), and one for testing (i.e. 1/3 of the data). The training and validation set is used for model selection (e.g. cross-validation is used to obtain the hyperparameters for the LS-SVM classifier), the test set is only used to check the performance. As a performance measure, the AUC is used. For each pair of classes and each feature extraction method, the 100 performance measures on the test set are averaged, resulting in $\overline{\text{AUC}}$. Next, the Friedman test is used on the $\overline{\text{AUC}}$s since the performances for each pair of classes are correlated and can be interpreted as blocks (Section 2.3.3). Additionally, to study the differences between the methods in detail, the Tukey multiple comparison procedure is used.

5.4 Results

Fourteen different combinations of classification algorithm and feature extraction method are compared (Table 5.4). All approaches are performed on both data sets except the quantification algorithms, which are only applied to the MRSI spectra.

For each data set, a figure is presented with paired comparisons and plots of the obtained performances are provided. For the MRS data set, the Friedman test resulted in a p-value of 0.0001, indicating that there are significant differences between the methods. The Tukey multiple comparison procedure results in a paired comparison plot, which gives a global overview of the performance for each method (Figure 5.2). The rank score is indicated on the horizontal axis. The figure contains a comparison interval for the mean rank of the $\overline{\text{AUC}}$ for each method. The higher the mean rank score, the higher the AUC values for that specific method. Since methods perform significantly different if their intervals are disjoint, there exist statistical differences between the methods for the MRS data set. By looking at the intervals, some trends can be observed. Relief-F, the Kruskal-Wallis test, the full spectra, ICA, PCA and peak regions obtain good results that are fairly similar. Using PCA in combination with LS-SVMs or the stepwise selection procedure with LDA leads to a significantly decreased performance. For this data set, there is no clear distinction in performance between the use of peak integration or peak regions and the true pattern recognition approaches (i.e. methods from statistics and machine learning), but LDA appears to perform better in combination with ICA and PCA than with peak-integrated values. The difference between peak regions and the full spectra, or the Kruskal-Wallis test is rather small. When comparing the classification algorithms, i.e. LDA and LS-SVMs, no method outperforms another. Finally, Figure 5.3 illustrates the performances for the different feature extraction methods based on a line plot.

In the examination of the multi-voxel data set, the quantification methods (i.e. AQSES and QUEST) are included. An identical statistical analysis is performed and results in the paired comparison plot, depicted in Figure 5.4. According to the Friedman test, a p-value smaller than 0.0001 is obtained. Compared to the previous analysis, stronger differences are detected. Using the full spectra in the region of interest, Relief-F or ICA, all in combination with LS-SVMs, results in the highest AUCs. Though not significantly, using only peak regions decreases the performance. Additionally, using 10 peak-integrated values with LS-SVMs further decreases the performance. As in the first data set, using LDA and 10 input features, the automatic pattern recognition method ICA performs better than peak-integrated values. Using LS-SVMs, ICA also seems to perform better than peak integration. Relief-F tends to perform better than the Fisher criterion and the Kruskal-Wallis test. Both the sophisticated quantification approaches in combination with LS-SVMs have a strongly decreased performance compared to the use of the full spectra, ICA or Relief-F. The performance of AQSES and QUEST is not better than peak integration. The smallest AUCs are obtained for the stepwise selection with LDA. Concerning the classifier type, in this data set the LS-SVM method outperforms LDA. To conclude, the AUCs of the various feature extraction methods are compared using boxplots in Figure 5.5.

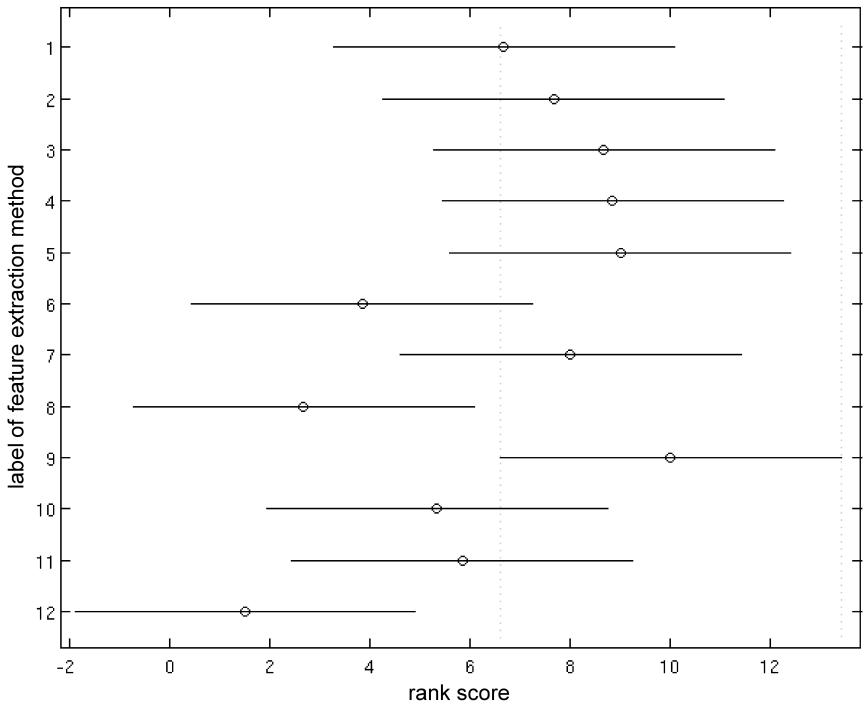


Figure 5.2: *Paired comparison plot for the MRS data set after performing the Friedman test and the Tukey multiple comparison procedure. The horizontal axis indicates the rank score, where higher rank scores correspond to higher AUC values. The different combinations of classification algorithm and feature extraction method are shown on the vertical axis. Relief-F (2), the Kruskal-Wallis test (3), full spectra (4), ICA (5), PCA (7) and peak regions (9) obtain similar performance scores.*

5.5 Discussion

In this chapter multiple binary classifiers are used to compare feature extraction methods. Compared to [43, 137, 204, 154, 110, 165], the results are based on more classification problems and more tissue types are included. In the present analysis, the aim is not to find the best feature extraction method for a specific pairwise problem (e.g. grade II glioma versus grade III glioma), since such a study can only produce reliable results when very extensive data sets are available. The present study examines feature extraction techniques over a wide variety of tissue

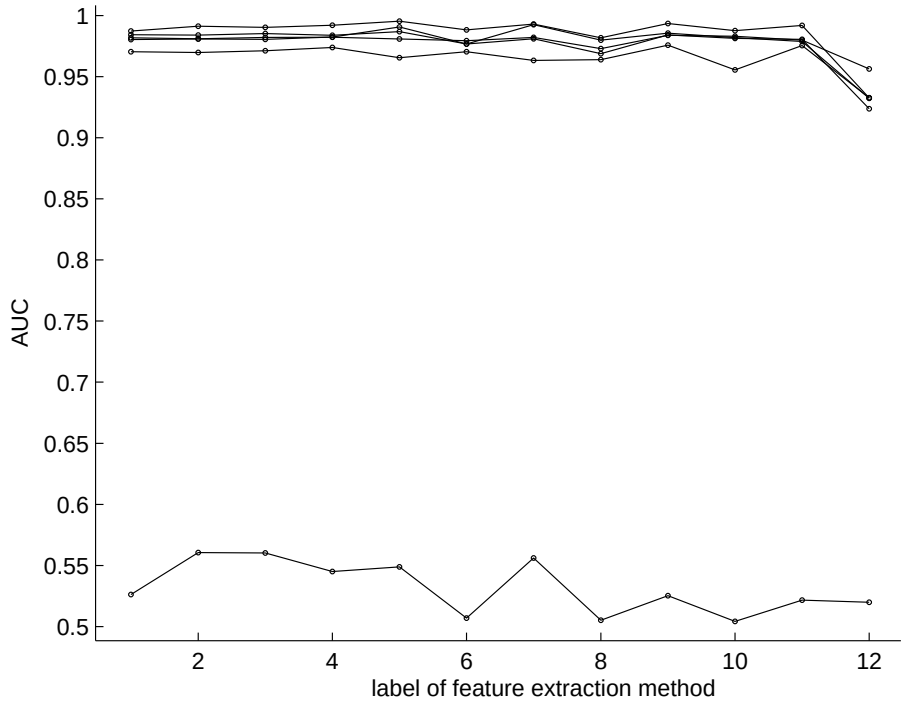


Figure 5.3: *Line plot of the mean AUCs for the MRS data. The mean AUC values on test set are indicated on the vertical axis. The numbers on the horizontal axis refer to the combinations of classification method and feature extraction technique as labeled in Table 5.4. The bottom line corresponds to the difficult separation of glioblastomas and metastases based on single-voxel MR spectroscopy.*

types. As such, the goal is to focus on finding methods that obtain a good overall performance. In the statistical analysis, this reflects in averaging the AUCs over the 100 resamplings for a specific classification problem and feature extraction method.

With respect to the studies by Lukas et al. and Devos et al., the focus is entirely on the effect of feature extraction techniques [137, 43]. Also, compared to [43] and [137], this chapter explores a wide variety of feature extraction methods and verifies the findings on two independent data sets. In the study of Simonetti et al. the data set was split only once into a set for training and for testing [204]. By dividing only once high bias is potentially introduced. This study overcomes

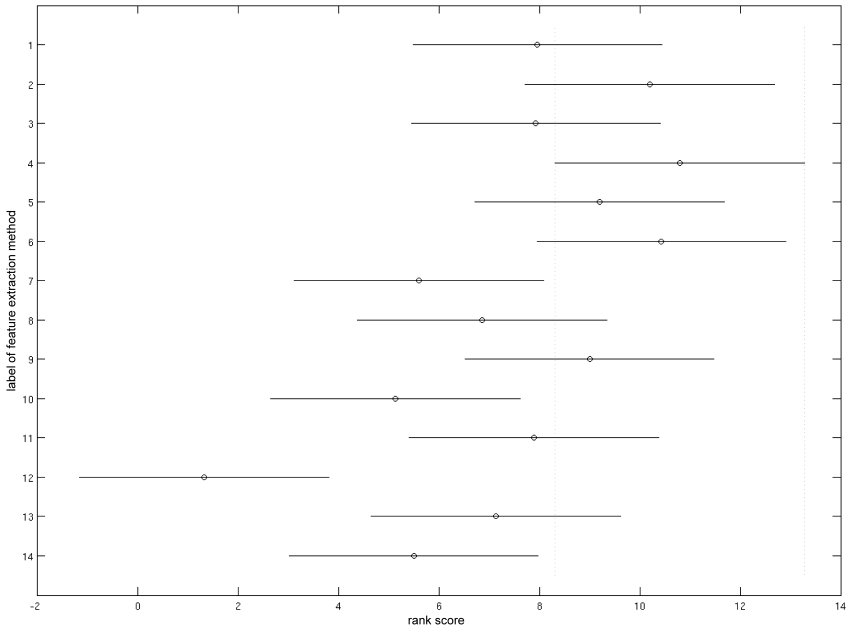


Figure 5.4: *Paired comparison plot for the MRSI data set after performing the Friedman test and the Tukey multiple comparison procedure. The horizontal axis indicates the rank score, where higher rank scores are related to higher AUC values. The different combinations of classification algorithm and feature extraction method are listed on the vertical axis. LS-SVMs with full spectra (4), Relief-F (2) and ICA (6) obtain the highest performances.*

the problem by splitting the data an extensive number of times. Also, no true statistical analysis was performed in [204]. This aspect is further improved in the present study. Compared with the work of Menze et al. and Kelm et al., our focus is only on short TE MR signals [154, 110]. In addition, an important improvement is that many more patients coming from multiple centers are included. This allows us to further increase the number of tissue classes with respect to the amount in [154] and [110] and formulate conclusions that are potentially less dependent on the MR machine or the acquisition parameters. In the study by Opstad et al. only two feature extraction methods were employed [165]. In fact, linear PCA was compared with LCModel based on an LDA classifier. In this way, the authors were not exploiting the power of a method which can truly accept full spectra as input for classification, without the need of reducing dimensionality.

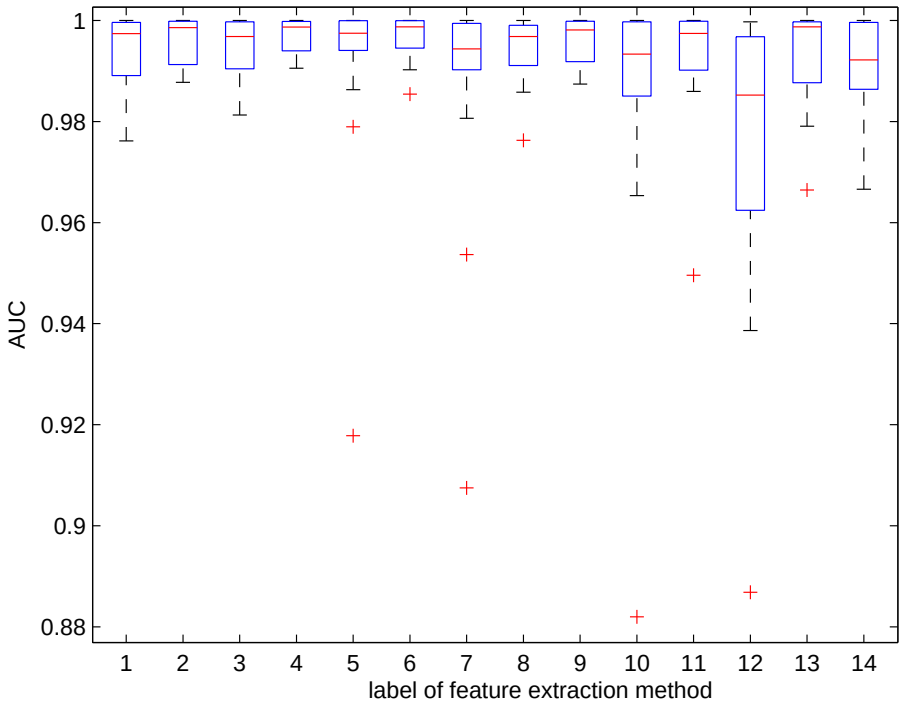


Figure 5.5: *Boxplots for the MRSI data with the mean AUC values on test set on the vertical axis. Pattern recognition approaches, and in particular the use of LS-SVMs and full spectra (4), results in a good performance and small spread. Quantification methods (13,14) show a degraded performance. The various combinations of classification method and feature extraction technique are shown on the horizontal axis as labeled in Table 5.4.*

Since two independent data sets are used in the analysis, results can be verified independently. Despite the differences in acquisition schemes, comparable results are obtained for both data sets. The automated pattern recognition techniques ICA, Relief-F and the full spectra in the region of interest result in the highest overall performances. The use of peak regions might slightly decrease the performance, peak-integrated values and the stepwise selection procedure further reduce the AUCs. In the multi-voxel data set, Relief-F tends to perform better than the Kruskal-Wallis test or the Fisher criterion. This might be explained by the fact that Relief-F does not assume conditional independence of the features.

Using full spectra resulted in the highest performances in [137, 110]. This may suggest to avoid feature extraction for kernel-based methods. However, here, an excellent performance is observed for automated pattern recognition approach based on ICA and Relief-F as well. Additionally, [154] reported good results for automated pattern recognition approaches. One might argue that still more sophisticated methods [230, 135], which can potentially increase the performance for certain classification problems, exist for LS-SVM-based feature extraction. Nevertheless, these algorithms often tend to increase the computation time compared to their gain in accuracy. Certainly in the context of DSSs efficiency is decisive. Compared to its low computational complexity, a fast automated pattern recognition algorithm (i.e. ICA) or the combination of full spectra with LS-SVMs results in a good performance.

The quantification methods AQSES and QUEST perform worse than the full spectra. The pattern recognition approaches Relief-F and ICA, and even peak integration yield higher AUCs. The latter was also reported in [204], where LCModel performed worse than peak integration, ICA or PCA. In [154] one reported an increase of 5 to 10 % in accuracy for automated pattern recognition methods compared to quantification approaches. Full spectra and pattern recognition methods even performed better than manually supervised quantification. Additionally, it was observed that QUEST performed worse than feeding full spectra to SVMs in [110]. On the contrary, the study in [165] reported an improved performance for LCModel compared to full spectra. However, it seems that the authors were comparing LCModel with PCA instead of feeding full spectra to the classifier. Nevertheless, in the present study, it seems that there are still more advanced methods which can further increase classifier performance. The problem with the sophisticated quantification methods is that they are in general hard to automate. For AQSES, it was possible to fix the parameter for controlling the smoothness of the baseline. However in QUEST, where one has to choose the number of points (i.e. n, Section 3.5) for baseline construction in the subtract method, one fixed value did not result in good AUCs for every pairwise problem. For this reason, a baseline correction method is performed prior to quantification with QUEST. With this approach, the results of AQSES and QUEST are comparable.

In medical practice, clinicians and technicians lack time to fit each spectrum separately or to try out different parameters influencing the estimation procedure of the computationally intensive quantification methods. Moreover, it is not clear how robust these methods are with respect to those parameters. The use of the quantification techniques also requires specific expertise, that is not always available. As a consequence, until now it seems that quantification methods are not ideally suited for use in automatic DSSs, certainly when

comparing their performance with results of automated pattern recognition approaches. When a blackbox classification method is satisfactory, automated pattern recognition techniques (e.g. Relief-F, ICA) or full spectra with LS-SVMs seem more appropriate. In case a biological interpretation of MR spectra is required, quantification approaches are needed since automated pattern recognition approaches may have limited abilities to provide this type of information.

5.6 Conclusions

In order to introduce a DSS in clinical practice to assess MR spectra of the human brain, automated procedures are required since user interaction and processing time should be minimized. In this chapter, the focus was on the feature extraction step in the preprocessing protocol of short TE MR spectra prior to classification. Two independent studies on separate data sets, including single-voxel and multi-voxel data, were performed and automated feature extraction methods from pattern recognition, machine learning and statistics (e.g. ICA, Relief-F) or full spectra in combination with a regularization technique, handling high-dimensional data, produced the highest overall performance. Sophisticated quantification methods seem hard to automate in a clinical DSS and are often computationally intensive. When only classification is needed, the use of full spectra or an automated pattern recognition technique seems more appropriate. Additionally, quantification techniques can provide a biological interpretation if needed. These findings suggest to restrict the feature extraction step in the preprocessing protocol, decreasing execution time of a DSS. This may speed up the introduction of DSSs in clinical practice.

Chapter 6

Multi-project brain tumor classification using multi-center MRS

Whereas in the previous chapter brain tumor classifiers were developed and evaluated using a single MRS data set, an extended multi-project MRS study is described in the present chapter. Section 6.1 provides an introduction to the study design. Data acquisition and preprocessing are explained in Section 6.2. Section 6.3 gives an overview of the classification methods and feature extraction techniques. The results of the classifiers are presented in Section 6.4 and discussed in Section 6.5.

6.1 Multi-project setting

Automated brain tumor classification by MRS has been under development for more than a decade [9, 83, 176, 156, 95, 218, 43, 137, 154, 219, 78]. Ultimately, these pattern recognition approaches need to be incorporated in DSSs, which (among others) require a high classification performance, interpretability by means of clinical knowledge and generalization of the performance to new samples that are obtained subsequently in different clinical centers [178, 130, 8, 170]. In this respect, standardization of acquisition conditions and protocols should make data from different centers compatible. This prevents potential bias from single-center or single-machine studies and, additionally, increases the number of available cases

for classifier development and testing purposes.

During the INTERPRET project (*International network for pattern recognition of tumours using magnetic resonance*, 2000−2002, http://azizu.uab.es/INTERPRET/ , accessed on June 27, 2009), a protocol was defined to guarantee compatibility of signals acquired at different centers [219, 101, 227]. As a result, numerous studies developed and evaluated classifiers using long and short TE MRS and MRSI data from INTERPRET [203, 137, 43, 219, 218, 44, 204]. However, the extension towards evaluation over time is recommended in order to provide the clinical community with robust results. The validation of classifiers through subsequent cases can consolidate the confidence of clinicians in the potential applicability of the classifiers. Nevertheless, to the author's knowledge, up to now there are no published evaluations of classifiers with unseen cases that are subsequently acquired in different centers.

Recently, the multi-center eTUMOUR project (*Web accessible MR decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data*, 2004−2009, http://www.etumour.net/ , accessed on June 27, 2009), which builds upon previous expertise from the INTERPRET project, provided this unique opportunity. The INTERPRET acquisition protocols for clinical, radiological, and histopathological data were extended to ex vivo transcriptomic (i.e. DNA microarrays) and metabolomic (i.e. HR-MAS) data acquisition in eTUMOUR. Furthermore, the raw MRS data acquired during INTERPRET were incorporated into the eTUMOUR data set for classifier development. This provides a unique opportunity to evaluate INTERPRET-based classifiers by means of cases from a later acquisition date and from partly different centers with different instrumentation, but obtained using the same or compatible acquisition protocols. The multi-project and multi-center evaluation in this chapter offers a close-up perspective of the conditions that the predictive models may face in a real clinical environment. Based on previous results, good performances are to be expected for most of the differential diagnoses, except for the discrimination of glioblastomas and metastases (cf. Chapter 5).

6.2 Materials

Training data for classifier development include 211 MRS signals, which were acquired at 1.5 T by 6 international centers in the framework of the INTERPRET project. Inclusion of tumor types for pattern recognition was based on the histological classification of the CNS tumors set up by the WHO. Glioblastomas, meningiomas, metastases and low-grade gliomas (grade II astrocytomas, grade

Table 6.1: *Number of training (INTERPRET) and test (eTUMOUR) cases per tumor type. Short TE MRS data were acquired according to a consensus protocol during the INTERPRET (2000−2002) and eTUMOUR (2004−2009) projects (GBM: glioblastomas, MEN: meningiomas, MET: metastases, LGG: low-grade gliomas).*

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

Table 6.2: *Percentage of cases per manufacturer in the training (INTERPRET) and test (eTUMOUR) data.*

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

II oligoastrocytomas, or grade II oligodendrogliomas) were included (Table 6.1). These data are identical to the data set presented in Section 4.2.1.

Test data were provided by 8 international centers in the framework of the eTUMOUR project. The 97 MRS test cases were acquired at 1.5 T (TE = 20 ms for STEAM and TE = 30−32 ms for PRESS) and were validated by expert spectroscopists. Only data with the original histopathology, available before February 28th, 2007, were included. The test cases were acquired after the training cases and originate from partly different centers using instruments of the three main manufacturers. Table 6.2 illustrates that the percentage of cases per manufacturer is similar for training and test data. Table 6.3 provides the percentage of cases per center in the training and test data. For the training set, 40 % of the cases originate from a center that did not provide test data afterwards. In addition, 35 % of test cases belong to three new centers that did not provide any training data.

INTERPRET developed a protocol for fully automated preprocessing of the data (Section 4.2.1). In summary, 211 INTERPRET cases and 47 cases of the eTUMOUR test data set (32 from Philips and 15 from GE) are preprocessed with the automated pipeline. Besides, a semi-automated pipeline is defined for some

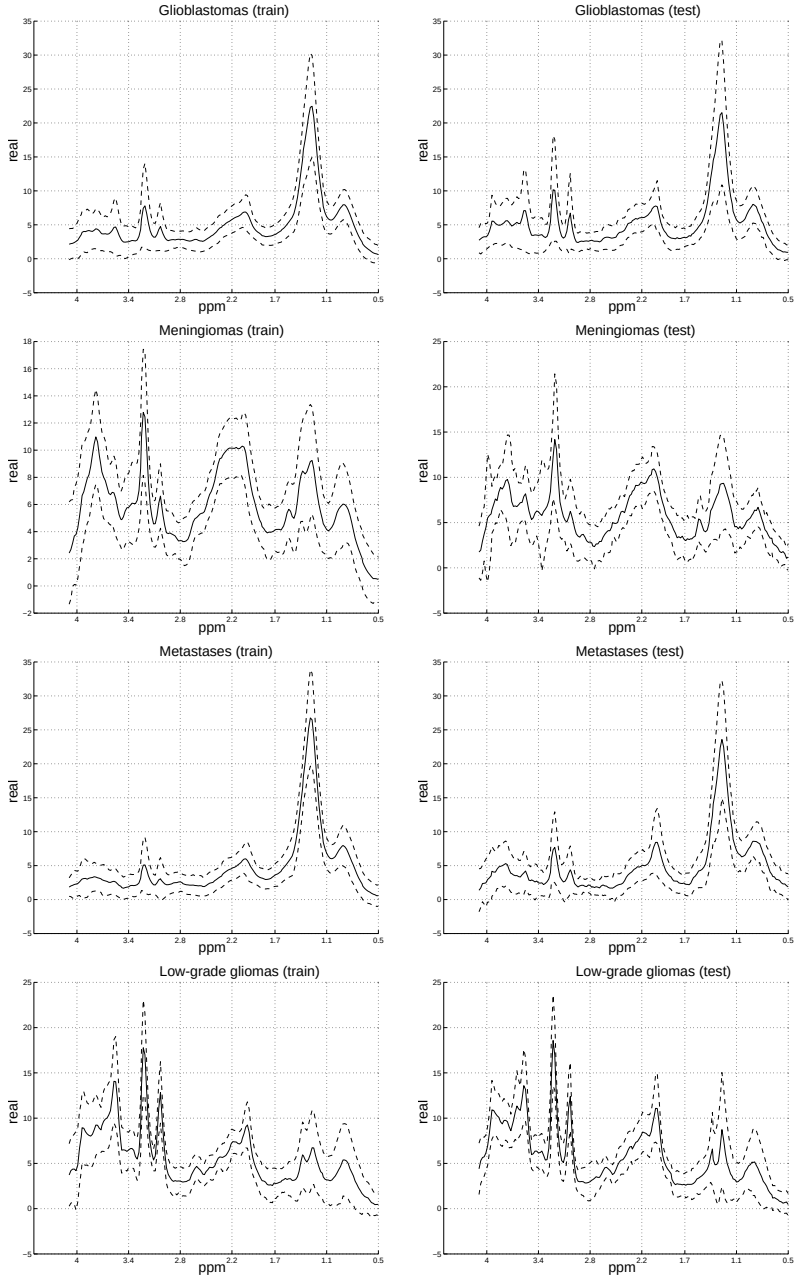


Figure 6.1: *Training (left) and test (right) mean L_2-normalized real MR spectra for glioblastomas, meningiomas, metastases and low-grade gliomas. The solid lines are the means, while the dashed lines are the means plus and minus the standard deviations.*

Table 6.3: *Percentage of cases per acquisition center in training (INTERPRET) and test (eTUMOUR) data.*

Center	INTERPRET (%)	eTUMOUR (%)
UMCN	2.8	1.0
SGHMS	27.0	18.6
MUL	3.8	10.3
FLENI	1.9	6.2
IDI-BELLVITGE	40.3	
CDP	24.2	
CDP + IAT		28.9
IDI-BADALONA		17.5
UNIV. DE VALENCIA		16.5
HOSPITAL SANT JOAN DE DEU		1.0

test cases from GE and Siemens, having new file formats. The design of the semi-automated pipeline ensures compatibility of its output with the automated one. In total, 50 test data points are preprocessed by the semi-automated pipeline that is partially based on jMRUI. Modifications of the semi-automated pipeline, with respect to the automated pipeline, include a phase correction of the water-suppressed FID using the water reference. Additional manual zero-order and first-order phase correction are performed when needed. Residual water is removed by means of HSVD (10 singular values and a water region of [4.33,5.07] ppm). Although the semi-automated preprocessing pipeline is based on a different software implementation, it is compatible with the automated one and provides comparable MR spectra for testing the classifiers (Figure 6.1).

6.3 Methods

6.3.1 Classification and feature extraction

Classification methods include LDA, FLDA, QDA, LDA with diagonal covariance matrix (dLDA), QDA with diagonal covariance matrix (dQDA), SVMs, LS-SVMs, MLPs, bi-directional Kohonen (BDK) networks and single and ensemble classifiers using KNN [153, 152, 226, 89, 50]. Bayesian strategies for regularization are applied for some of the LS-SVM and MLP classifiers [143]. Several feature extraction methods have been applied to the real part of the MR spectra prior to classification. These methods include peak integration, peak regions, PCA, ICA and wavelets. The wavelet basis is coiflet 3 with 9 levels [37]. Finally, some

classification approaches are applied to the full region of interest represented by a data vector of 190 points.

6.3.2 Experimental setup

Classifiers are developed and tested for all pairwise differential diagnoses involving glioblastomas, meningiomas, metastases and low-grade gliomas. For each task, different combinations of feature extraction method and classification technique are applied. In total, 253 pairwise classifiers are inferred for classification among glioblastomas, meningiomas, metastases and low-grade gliomas based on the 211 INTERPRET MR spectra. The selected features for the classifiers are derived from previous experience, other studies, and model validation based on the training data set [43, 71]. For each classifier an estimation of the error rate and the BER is carried out using a 10-fold cross-validation for the INTERPRET data set. Afterwards, the classifiers are tested using 97 MR spectra, which were subsequently acquired during the eTUMOUR project. The statistical analysis includes the Kruskal-Wallis test and the Tukey multiple comparison procedure.

6.4 Results

Table 6.4 summarizes the results for the pairwise classifiers that obtain the best performance on the independent eTUMOUR data. Most of the differential diagnoses among the four classes are solved with high accuracy. The best classifiers among glioblastomas, meningiomas, metastases, and low-grade gliomas achieve an accuracy around 90 %. However, for the differentiation of glioblastomas and metastases the best classifier obtained an accuracy of 78 % on the independent test data, which is strongly degraded compared to results for the other differential diagnoses.

Figure 6.2 provides the boxplots of the performance (i.e. BER on the independent test data) for each differential diagnosis based on the full list (i.e. 253 classifiers) of classifier results. As expected, the distribution for the differentiation of glioblastomas and metastases highly deviates from all other distributions. According to the Kruskal-Wallis test and the Tukey multiple comparison procedure there exists a significant difference between the performances for the separation of glioblastomas and metastases with respect to all other problems. No significant differences are found among these other problems.

Different approaches obtain satisfactory results for the discrimination of glioblastomas and meningiomas. An MLP with full spectra as input features achieves

Table 6.4: *Results of the best performing classifiers for the six pairwise differential diagnoses. The error rate (ERR) and BER for the INTERPRET data (obtained using cross-validation (CV)), and the independent test (IT) set from eTUMOUR. The feature extraction methods and classification algorithms are listed.*

Task	Feature extraction	Classification method	CV		IT	
			ERR	BER	ERR	BER
GBM - MEN	Full spectra	MLP	0.06	0.07	0.07	0.09
GBM - MET	Peak integration	LDA	0.33	0.40	0.22	0.21
GBM - LGG	Peak integration	LS-SVM	0.12	0.18	0.08	0.09
MEN - MET	PCA	MLP	0.05	0.05	0.06	0.07
MEN - LGG	ICA	LS-SVM	0.08	0.09	0.08	0.08
MET - LGG	Peak integration	LDA/FLDA/MLP/LS-SVM	[0.01,0.04]	[0.01,0.04]	0.06	0.07

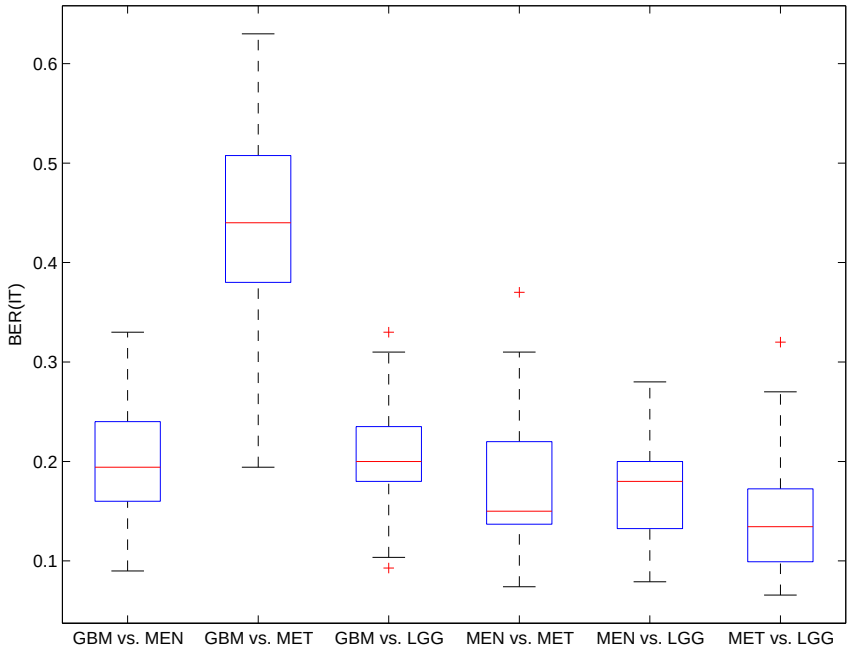


Figure 6.2: *Boxplots of the performances for each differential diagnosis in the eTUMOUR data set based on the results of all 253 pairwise classifiers. Performance is measured as the BER.*

a BER of 0.09. The median of the distribution of the BER for separating glioblastomas and meningiomas is below 0.20. The difficult discrimination between glioblastomas and metastases is clearly observed in the results on the training set, as well as on the test set. For this differential diagnosis, the median BER on the independent test set is around 0.45, and most results range from 0.4 to 0.5. Some classifiers achieve a BER around 0.2, but these values seem to be outliers with respect to most other obtained performances. This makes it hard to ensure reproducibility of the results. The median of the BER for the differential diagnosis of glioblastomas and low-grade gliomas is 0.2. Nevertheless, there is a set of regularized classifiers that obtain a BER around 0.09. The best BER corresponds to an LS-SVM classifier, based on a Bayesian framework, using peak-integrated values. No matter what method is used, the results illustrate that meningiomas can be easily discriminated from metastases. Most of the obtained BER values fall in the interval [0.1,0.2]. The best classifier achieves a BER of 0.07 and is based on PCA and a neural network with Bayesian regularization. For the classification of meningiomas and low-grade gliomas, LS-SVM and LDA with different feature extraction methods achieve a BER of 0.08 and 0.11, respectively. Most of the

results for this classification task are in the interval [0.15,0.25] and the median of the distribution is under 0.2. The low error for this differential diagnosis is also obtained using cross-validation on the INTERPRET data. The distribution of the BER for the separation of metastases and low-grade gliomas indicates good performances (i.e. a BER of 0.1) for most classifiers. Peak integration combined with LDA, FLDA, MLP, or LS-SVM classifiers obtain the best performance on the independent test set. The cross-validation estimates of the error also indicate good performance for the classifiers.

6.5 Discussion

An intuitive approach to compare data sets of signals is by visual inspection of their averaged patterns and corresponding standard deviations (Figure 6.1). In general, the averaged patterns from the training and test set for glioblastomas, meningiomas, metastases and low-grade gliomas are rather similar. The results for the differential diagnosis of metastases and low-grade gliomas and the separation of meningiomas and metastases agree with previous studies [43]. The difficult problem of separating glioblastomas and metastases was already observed in Chapter 5 and previous studies [218, 43, 166, 219]. The necrotic profile of the metastases and of most of the glioblastomas is highly similar. The high lipid peaks mask the other metabolic information. Approaches that combine MRI and MRSI or the acquisition of numerous voxels, adjacent to the brain mass provide relevant additional information for distinguishing between these two types of tumors [25, 98, 122, 122, 140]. The problem is extensively studied in Chapter 9 of the thesis. Devos et al. obtained comparable performance for the differential diagnosis of glioblastomas and low-grade gliomas using LDA and standard LS-SVMs [43]. Significant statistical differences between glioblastomas and low-grade gliomas and between glioblastomas and grade III astrocytomas were also found for different metabolite ratios with respect to Cr and/or water [150, 29]. An increased performance was observed for classifying normal, non-progressive tumors (with radiation injury and stable disease) and brain tumors using long TE MR data and regularized methods [154]. In the present analysis, the smallest non-outlier BER value for separating glioblastomas and low-grade gliomas is higher than the smallest non-outlier observation for the remaining problems. It suggests that this classification problem is more difficult to solve by single-voxel short TE MR data than the other four pairwise problems. For the differential diagnosis of meningiomas and low-grade gliomas, Tate et al. obtained a perfect discrimination in [218], where a three-class discrimination problem was studied: meningiomas versus grade II astrocytomas versus aggressive tumors (which include glioblastomas and metastases). In that study, the confusion matrix of meningiomas and grade II astrocytomas indicated no misclassifications. Identical results were

obtained by extending the three-class classifier to meningiomas versus low-grade gliomas versus aggressive tumor [219].

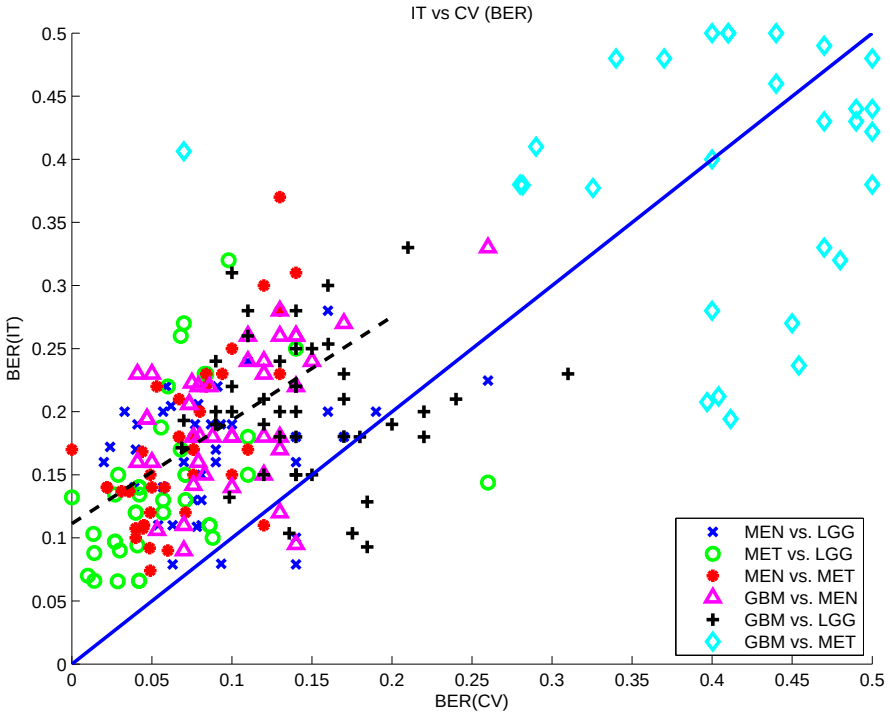


Figure 6.3: *Scatter plot of the BER estimated by the cross-validation (CV) analysis based on INTERPRET training cases and the corresponding BER on the independent test (IT) data, consisting of newly acquired eTUMOUR cases. BER(IT) = BER(CV) is represented by the solid blue line and the trend of the (BER(CV) < 0.2, BER(IT) < 0.3) region is indicated by the black dashed line.*

The BERs of the classifiers on the INTERPRET training data are estimated by cross-validation and compared with BERs on the new eTUMOUR cases (Figure 6.3). Most of the results are situated in the (BER(CV) $<$ 0.2, BER(IT) $<$ 0.3) region, except for the results for the differentiation of glioblastomas and metastases, showing a sparse distribution. The general trend in this region is indicated by the dashed black line. It indicates an underestimation of the BER by the cross-validation analysis on the INTERPRET data. This kind of underestimation is typically observed in pattern recognition challenges and is usually produced by overfitting of the models on the training data set and

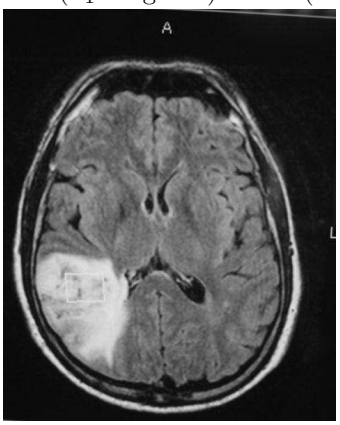
estimation of the error with dependent samples [81, 15]. The evaluation of predictive models using newly acquired multi-center data ensures the independence of training and test data. The results for separating glioblastomas and metastases are scattered in regions of high error. Some overestimations of the cross-validation error are also observed for this differential diagnosis. This confirms the fact that this is a difficult problem and illustrates the randomness in the obtained results.

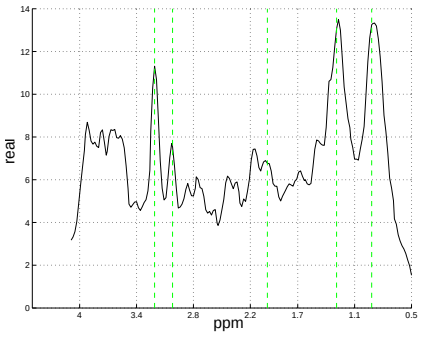
Eight out of 50 semi-automatically preprocessed test cases are misclassified at least once by the pairwise BDK classifiers (thereby omitting glioblastomas versus metastases classifiers). On the other hand, 10 out of 47 of the automatically preprocessed test cases are misclassified at least once by the same classifiers. As such, no difference in the classification of automatically and semi-automatically preprocessed signals is observed. The semi-automated preprocessing pipeline, which is applied to the largest part of the test data, is consistent with the automated pipeline applied to the training data. This has an important implication since it suggests compatibility of different preprocessing software tools, either in an automated or a semi-automated way, for automated classification in DSSs.

A practical result of the present study is that cases which are repeatedly misclassified by different techniques can be flagged for additional revision. For example, these cases may show problems in voxel positioning, acquisition artifacts, contamination from normal tissue, or limitations of the classification methodology (e.g. patterns replicated in non-tumoral diseases, atypical MRS patterns and underrepresented tumor subtypes). In this way, pattern recognition techniques can contribute to the automated validation of cases, even in the absence of a biopsy, thereby assisting experts in the detection of potential sources of error in acquired patient's data.

Figures 6.4 and 6.5 show some misclassified cases from the eTUMOUR data, which may be subjected to revision. Although the original pathologist made a comment about the presence of areas of anaplastic oligodendroglioma (grade III, WHO), the eTUMOUR case et2274 was diagnosed as oligodendroglioma 9450/3 (grade II, WHO). Voxel allocation was carried out following the eTUMOUR acquisition protocol. The spectrum shows an uncommon mobile Lip pattern, with high 0.9 and 1.3 ppm resonances. The disappearance of these resonances at long TE (i.e. 136 ms) discards a significant necrotic contribution (cf. [71]). This kind of pattern was already observed in previous studies [71]. For example, the INTERPRET cases I0450 (oligoastrocytoma) and I0179 (oligodendroglioma) were also misclassified by the INTERPRET DSS. In summary, these facts and its persistent misclassification in the present analysis, seem to indicate that case et2274 is a class outlier. The eTUMOUR case et2206 was originally diagnosed as oligoastrocytoma 9382/3 (grade II, WHO), but there were some discrepancies regarding the glial subtype

et2274 (T_1-weighted): LGG (OD)





et2206 (T_2-weighted): LGG (OA or OD)



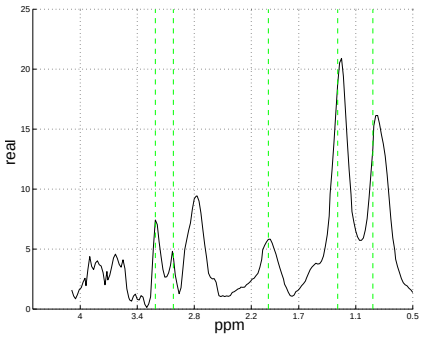
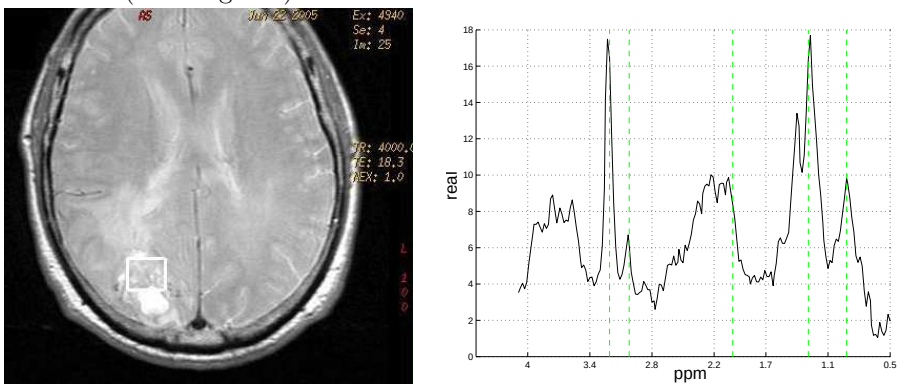


Figure 6.4: *Potential outliers. Case numbering corresponds to the eTUMOUR database entries. For each case, the reference image with voxel location is provided on the left (LGG: low-grade glioma, OD: oligodendroglioma, OA: oligoastrocytoma). The real part of the short TE MR spectrum is shown on the right. Vertical dashed lines indicate the position of the main metabolites: Cho (3.22 ppm), Cr (3.03 ppm), NAA (2.03 ppm), Lip (1.3 ppm) and Lip (0.9 ppm).*

on the validation done by the pathological committee. The case is misclassified by every classifier for separating metastases and low-grade gliomas, and also by some of the classifiers for distinguishing glioblastomas and low-grade gliomas and meningiomas from low-grade gliomas. The mobile Lip pattern of the case is also uncommon, having relatively large 0.9, 1.3 and 2.8 ppm peaks, which are reduced

et2349 (PD-weighted): GBM



et2197 (T_2-weighted): MET

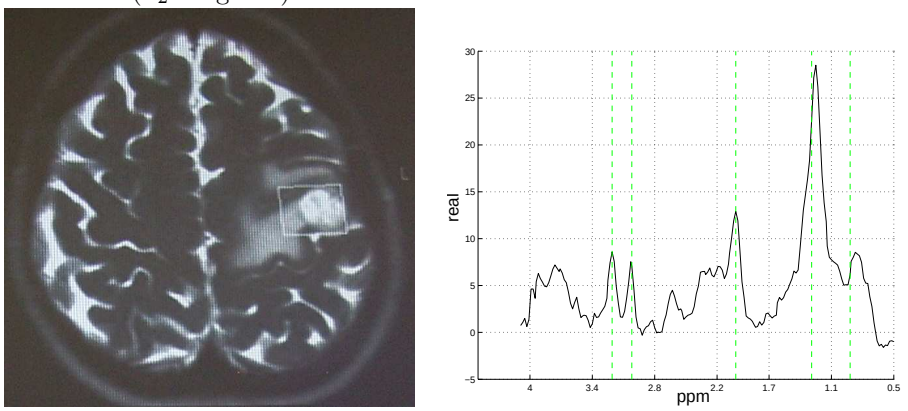


Figure 6.5: *Potential outliers, cf. Figure 6.4 (GBM: glioblastoma, MET: metastasis).*

at long TE, suggesting a non-necrotic origin. The eTUMOUR case et2349 is a glioblastoma without clear visible mobile Lip. The case is misclassified in every classification problem. Expert review did not indicate problems in placement of the voxel, being located in the highly cellular part of the tumor. The eTUMOUR case et2197 is a metastasis and the MR spectrum shows potential contribution from normal brain parenchyma, as can be observed by comparing the voxel size and the extent of the brain lesion. As a consequence, its pattern shows similar Cho and Cr peak heights and relatively high NAA at 2 ppm. On the other hand, the appearance of high Lac and mobile Lip around 1.3 ppm suggests abnormality.

Nevertheless, it is clearly an uncommon spectral pattern for a metastasis tumor.

6.6 Conclusions

This chapter described a multi-project and multi-center evaluation of automated brain tumor classifiers using short TE single-voxel MR spectra. To the author's knowledge, there is no previous work that evaluated predictive models, which are trained with data acquired from a multi-center project, using a new independent test set, that is subsequently acquired from partly different centers. Classifiers were trained with cases acquired by 6 centers during the period 2000−2002. These classifiers were tested with posterior cases, that were acquired by 8 centers during the period 2004−2007. This study design provided a setting that was close to a real environment where similar classifiers, integrated in a DSS, may be used in multiple clinical centers to assist in the diagnosis of new cases. The experiments showed that classifiers, developed from the INTERPRET data set, seemed to be robust enough for predictive classification of prospective cases from eTUMOUR. The pairwise discriminations among glioblastomas, meningiomas, metastases, and low-grade gliomas achieved accuracies of around 90 %. On the other hand, the differentiation of glioblastomas and metastases resulted in an accuracy of 78 %. These results confirmed the conclusions from previous studies on automated brain tumor classification using MRS, but extended it towards a multi-project and multi-center setting. Therefore, an important conclusion is that accurate classification of new cases is feasible using data acquired in different clinical centers, with different instrumentation, but similar acquisition protocols. A well-defined protocol for the acquisition of MRS (e.g. spectral parameters and voxel location), and the application of quality control to MR spectra should allow the reproducibility of such classification rules and the successful use of DSSs, implementing them in a real clinical environment. Additionally, the proposed methodology might also assist in the quality control of new cases. Also, the present approach could be of interest for pediatric brain tumor studies in order to provide predictive information to pediatric neurosurgeons. Hence, the conclusions of this chapter are directly applicable to several of the tasks associated with the development of a DSS for brain tumor diagnosis and prognosis and its deployment in clinical environments.

Chapter 7

Multi-class system for brain tumor recognition using MRI and MRSI

Binary brain tumor classifiers were considered in Chapter 5 and Chapter 6. In this chapter a mechanism is developed to extend binary brain tumor classifiers to a multi-class classifier system, thereby obtaining class probabilities. Section 7.1 provides a general introduction to the problem of obtaining multi-class brain tumor classifiers. The data set, which consists of MRSI features and MRI information, includes samples from 10 different tissue types and is presented in Section 7.2. The strategies for pairwise combination of class probabilities are covered in Section 7.3. The results of these approaches are summarized in Section 7.4 and discussed in Section 7.5

7.1 Multi-class classification

In Chapter 5 and Chapter 6 binary classifiers were developed based on MRS and MRSI data. However, if DSSs need to be developed, the construction of multi-class classifiers is an important issue. While the generalization to the multi-class case is straightforward for LDA for example (Section 2.2.2), the upgrade of binary LS-SVMs to multi-class LS-SVMs is not obvious [1]. Although there exist multi-class generalizations for SVM-based methods, the typical procedure is to split up the multi-class problem into a number of smaller binary problems [1, 213]. The procedure to combine these binary classifiers into a multi-class system can be

performed in many ways and multiple factors have an influence on the choice of the combination algorithm.

In minimal output coding (MOC), each class is represented by a unique binary codeword in a very efficient way. To encode $n_C = 2^k$ classes, the MOC method needs k bits or k classifiers [215]. On the other hand, error correcting output codes (ECOC) is a technique that uses more than the minimal number of bits for encoding. The goal is to increase the Hamming distances between the codewords of different classes [49]. In this way, misclassifications of the binary classifiers are tolerated and the generalization of the multi-class classifier system is enhanced. One-versus-all is a method that constructs n_C binary classifiers for a classification problem involving n_C classes [4]. The n_C classifiers are trained to separate one class from the combination of all the other classes. In this method, n_C bits are needed to produce the codeword. If the output is for instance the second class, then the second bit of the codeword is set to 1 while the others are set to -1. A disadvantage of this approach is that the data set is often very asymmetric after grouping together $n_C - 1$ classes. When using one-versus-one coding the unbalance in the data set is usually less extreme [117]. For the n_C class problem $n_C(n_C - 1)/2$ binary classifiers need to be built. If the number of classes increases to a larger number, this method seems to become cumbersome. However, when the number of classes is not too abundant, each binary classifier needs to be trained using a smaller amount of data such that training of the classifier can actually become faster. To decide the final class for the one-versus-one approach, a simple voting scheme can be used. Voting assigns the case to the class with the highest number of votes using all $n_C(n_C - 1)/2$ binary classifiers. This is in fact a max-wins criterion. Another option is to use a directed acyclic graph to decide about the final class [173].

In this chapter the one-versus-one combination scheme is used to split up the multi-class problem into a number of binary problems. Apart from the fact that the data are more balanced and that the training phase (including determination of the hyperparameters) is less computationally intensive most of the time, there is also another practical reason to choose for one-versus-one coding. In practice, clinicians often have a clue about the diagnosis for a specific patient. Frequently, the clinicians only doubt between two types of tissue such that a binary classification method is sufficient for diagnosing these cases. In fact, one can interpret the binary classifiers as powerful, stand-alone, entities which can also be combined if multi-class classification is needed.

Furthermore, clinicians are interested in a measure of uncertainty when using a DSS. Naturally, it is not sufficient to output a single tumor type for the case to be classified, without a measure of its confidence. Therefore, class probabilities

are calculated for every tissue class in the training data in order to obtain such a measure of uncertainty. All these issues are addressed in the following sections. Four methods to transform one-versus-one pairwise class probabilities in final class probabilities are presented.

7.2 Materials

For building the multi-class classifier system, MRSI and MRI data were selected from the INTERPRET project database. Patient selection and preprocessing of the MRSI and MRI data are identical as performed in Section 4.2.2, but data are grouped in 10 instead of 6 tissue classes. Three subtypes were obtained for both the grade II gliomas and the grade III gliomas. The reason for this is that the differences between the pairwise combination methods become larger when the number of classes in the data set increases [241]. For one patient there was no consensus about the subtype such that data from this patient were removed from the data set. The final data set includes normal brain tissue from volunteers and apparently normal tissue from the contralateral half of the brain of patients (218 voxels from 8 persons), CSF from patients (100 voxels from 8 patients), grade II diffuse astrocytomas (90 voxels from 5 patients), grade II oligoastrocytomas (45 voxels from 2 patients), grade II oligodendrogliomas (22 voxels from 2 patients), grade III astrocytomas (16 voxels from 2 patients), grade III oligoastrocytomas (28 voxels from 1 patient), grade III oligodendrogliomas (25 voxels from 2 patients), meningiomas (48 voxels from 3 patients) and grade IV gliomas (70 voxels from 7 patients) (Table 7.1). Similarly to Section 4.2.2, 10 peak-integrated values and 4 image variables are extracted from MRSI and MRI data, respectively. These 14 features are used as input for the classifiers. No further feature extraction is performed on these features.

7.3 Multi-class classification methods

The multi-class system requires a set of binary classifiers and in this chapter Bayesian LS-SVM are used for this purpose (Section 2.2.6). An RBF kernel is chosen for these binary models since nonlinear classifiers were reported to outperform linear classifiers on such a data set [44]. Although here only Bayesian methods are used to produce posterior class probabilities, standard (LS-)SVMs can also be used to obtain binary class probabilities. Platt proposed a method

Table 7.1: *Class label, number of MR spectra and number of persons (i.e. patients and/or healthy volunteers) for each tissue type in the combined MRSI and MRI data set.*

Label	Tissue type	Spectra	Persons
1	Normal tissue	218	8
2	CSF	100	8
3	Grade II diffuse astrocytomas	90	5
4	Grade II oligoastrocytomas	45	2
5	Grade II oligodendrogliomas	22	2
6	Grade III astrocytomas	16	2
7	Grade III oligoastrocytomas	28	1
8	Grade III oligodendrogliomas	25	2
9	Meningiomas	48	3
10	Glioblastomas	70	7

to map the SVM output into probabilities using a sigmoid function [174]. The following parametric form was used for the sigmoid

$$P(y=1|x) = \frac{1}{1+\exp(a(\sum_{i=1}^{N}\alpha_i y_i K(x,x_i)+b)+c)}, \tag{7.1}$$

where the parameters a and c can be adapted to obtain optimal probability outputs. The values for a and c are estimated based on maximum likelihood estimation from a training set. The training set can be obtained based on cross-validation. Later, the algorithm by Platt was improved in terms of robustness [128].

7.3.1 Pairwise combination of probabilities

Different algorithms were proposed to obtain class probabilities from pairwise combination [191, 184, 63, 88, 241]. In this chapter, the methods by Price et al., Hastie and Tibshirani and two algorithms by Wu et al. are used. The method of Refregier and Vallet and voting are not considered. The reason to omit the algorithm by Refregier and Vallet is that some arbitrary choices about the selection of pairwise probabilities have to be made. It was pointed out by Price et al. and Wu et al. that the results were very sensitive to this choice and that finding the optimal selection was often very expensive. Voting is a very simplistic method and it was illustrated in [241] that the errors were high compared to the other methods. Before introducing the algorithms, the problem is defined in a formal way. Given a data point x and corresponding class label y, the pairwise probabilities $\hat{\mu}_{ij}$ are

denoted as estimates of $\mu_{ij} = P(y_k = i|y_k = i \text{ or } j, x_k)$. The main goal of coupling probabilities is to obtain the probability $p_i = P(y_k = i|x_k)$ based on the $\hat{\mu}_{ij}$ values.

7.3.2 Price et al.

Price et al. developed a method that combined pairwise neural network classifiers with probabilistic outputs for a handwriting recognition system [184]. Although it was originally intended for classification between a limited number of classes, Price et al. showed that the approach was also applicable for problems with more than 10 classes. The method assumes that

$$(\sum_{j:j\neq i} P(y_k = i \text{ or } j|x_k)) - (n_C - 2)P(y_k = i|x_k) = 1, \tag{7.2}$$

and

$$\hat{\mu}_{ij} \approx \mu_{ij} = \frac{P(y_k = i|x_k)}{P(y_k = i \text{ or } j|x_k)}, \tag{7.3}$$

such that the final class probabilities become

$$p_i = \frac{1}{\sum_{j:j\neq i} \frac{1}{\hat{\mu}_{ij}} - (n_C - 2)}. \tag{7.4}$$

Afterwards, the probabilities are normalized such that the sum is exactly 1. From the implementation point of view, this method is very simple. On the other hand, the method does not take the number of cases for each class into account.

7.3.3 Hastie and Tibshirani

Hastie and Tibshirani presented a method for pairwise coupling, which is a special case of the algorithms for generalized Bradley-Terry models [88]. In order to obtain $p = [p_1 \ ... \ p_{n_C}]^T$, the algorithm minimizes the Kullback-Leibler distance criterion such that $\hat{\mu}_{ij}$ approximates $p_i/(p_i + p_j)$

$$l(p) = \sum_{i<j} n_{ij} \left(\hat{\mu}_{ij} \log \frac{\hat{\mu}_{ij}}{\mu_{ij}} + (1 - \hat{\mu}_{ij}) \log \frac{1 - \hat{\mu}_{ij}}{1 - \mu_{ij}} \right), \tag{7.5}$$

where n_{ij} denotes the sum of the number of data points in class i and class j, $\hat{\mu}_{ji} = 1 - \hat{\mu}_{ij}$ and the model is $\mu_{ij} = p_i/(p_i + p_j)$. One needs to estimate p_i such that μ_{ij} is close to $\hat{\mu}_{ij}$. Because there are $n_C - 1$ independent parameters and $n_C(n_C - 1)/2$ equations, it is in general not possible to find p_i such that $\mu_{ij} = \hat{\mu}_{ij}$ for all i, j. To minimize Equation (7.5), one calculates

$$\frac{\partial l(p)}{\partial p_i} = \sum_{j:j\neq i} n_{ij}(-\frac{\hat{\mu}_{ij}}{p_i} + \frac{1}{p_i + p_j}). \tag{7.6}$$

After setting $\frac{\partial l(p)}{\partial p_i} = 0$, $i = 1, ..., n_C$, and multiplying by p_i, Hastie and Tibshirani proposed to find a point that satisfies

$$\sum_{j:j\neq i} n_{ij}\mu_{ij} = \sum_{j:j\neq i} n_{ij}\hat{\mu}_{ij}, \quad \sum_{i=1}^{n_C} p_i = 1, \quad p_i > 0, i = 1, ..., n_C, \tag{7.7}$$

and developed an iterative procedure to calculate such a point (Algorithm 4).

Algorithm 4 : *Pairwise coupling by Hastie and Tibshirani*

Input:
p_i = initial guess for probability for class i

$\mu_{ij} = p_i/(p_i + p_j)$
repeat $(i = 1, 2, ..., n_C, 1, ...)$ 6 and 7 until convergence
$p_i = p_i \frac{\sum_{j\neq i} n_{ij}\hat{\mu}_{ij}}{\sum_{j\neq i} n_{ij}\mu_{ij}}$
renormalize p_i and recompute μ_{ij}
$p = p/\sum p_i$

Output:
p = final class probabilities

The work of Hunter and Zermelo [246, 97] proved that if the $\hat{\mu}_{ij}$ are strictly positive, for any initial point, the sequence of points generated by the algorithm by Hastie and Tibshirani converges to a point satisfying Equation (7.7). This point is also a unique global minimum if $\sum_{i=1}^{n_C} p_i = 1$ and $0 \leq p_i \leq 1$, $i = 1, ..., n_C$. Remark that the method takes the number of cases for each class into account.

7.3.4 Wu et al. – method 1

The first method proposed by Wu et al. makes use of an approximate solution to an identity [241]. The existence of this solution was proven based on finite Markov chains. More specifically, Wu et al. proposed to solve the equations

$$p_i = \sum_{j:j\neq i} \left(\frac{p_i + p_j}{n_C - 1} \right) \hat{\mu}_{ij}, \quad \sum_{i=1}^{n_C} p_i = 1, \quad p_i \geq 0, \tag{7.8}$$

which can be reexpressed as

$$Qp = p, \quad \sum_{i=1}^{n_C} p_i = 1, \quad p_i \geq 0 \quad \text{with } Q_{ij} = \begin{cases} \sum_{s:s\neq i} \hat{\mu}_{is}/(n_C - 1), & \text{if i = j,} \\ \hat{\mu}_{ij}/(n_C - 1), & \text{otherwise.} \end{cases} \tag{7.9}$$

Finally, it was proven that

$$Qp = p, \quad \sum_{i=1}^{n_C} p_i = 1, \tag{7.10}$$

has the same unique solution as Equation (7.9) if the $\hat{\mu}_{ij}$ are strictly positive. The main advantage of this method is that no iterative procedure is required since only the linear system in Equation (7.10) needs to be solved. In contrast with the previous algorithm, this method assumes equal weighting (i.e. equal n_{ij}).

7.3.5 Wu et al. – method 2

The second approach by Wu et al. is an improved version of the method by Refregier and Vallet [241]. Instead of working with Equation (7.8), one rewrites it as

$$\sum_{j:j\neq i} \hat{\mu}_{ji} p_i - \sum_{j:j\neq i} \hat{\mu}_{ij} p_j = 0. \tag{7.11}$$

Based on this formula, Wu et al. suggested to minimize

$$\min_{p} \sum_{i=1}^{n_C} \sum_{j:j\neq i} (\hat{\mu}_{ji} p_i - \hat{\mu}_{ij} p_j)^2 \quad \text{with} \quad \sum_{i=1}^{n_C} p_i = 1, \quad p_i \geq 0. \tag{7.12}$$

It was proven that the nonnegative constraints can be omitted and that Equation (7.12) has a unique solution for p and it can be solved using the following linear system

$$\begin{bmatrix} Q & 1_{n_C\times 1} \\ 1^T_{n_C\times 1} & 0 \end{bmatrix} \begin{bmatrix} p \\ b \end{bmatrix} = \begin{bmatrix} 0_{n_C\times 1} \\ 1 \end{bmatrix} \quad \text{with} \quad Q_{ij} = \begin{cases} \sum_{s:s\neq i} \hat{\mu}^2_{si}, & \text{if i = j,} \\ -\hat{\mu}_{ji}\hat{\mu}_{ij}, & \text{otherwise,} \end{cases} \tag{7.13}$$

where $1_{n_C\times 1}$ and $0_{n_C\times 1}$ are column vectors with respectively n_C ones and n_C zeros.

7.3.6 Experimental setup

The four different methods for pairwise combination of probabilities are compared using stratified repeated random sampling. The data set is repeatedly (i.e. 100 times) randomly split into a set for training (i.e. 2/3 of the data) and a set for testing (i.e. 1/3 of the data). The pairwise Bayesian LS-SVM classifiers are built using the training set. Afterwards, the performance of each multi-class combination scheme is evaluated on the test set. The Friedman test is used to check whether differences in performance are statistically significant. In addition, the Tukey multiple comparison procedure is applied. An additional comparison between the use of Bayesian LS-SVMs and LDA for generating the binary class probabilities is performed using the Wilcoxon signed-rank test (Section 2.3.3). Confusion matrices, the accuracy and the Brier score are used to evaluate the performances (Section 2.3.1).

7.4 Results

The results for the averaged accuracy and the averaged Brier score on test set after 100 times of stratified random sampling are summarized in Table 7.2. According to the test statistic the accuracies for the first method by Wu et al. and the

Table 7.2: *Mean accuracy and Brier score on test set over 100 runs of stratified repeated random sampling. The accuracies of the technique by Hastie and Tibshirani (H-T) and the first method by Wu et al. are significantly higher compared to all others. In contrast to the other approaches, the Brier scores of the technique by Hastie and Tibshirani (if the stopping criterion is sufficiently strict, e.g.* $\leq 10^{-3}$*) and the first method of Wu et al. are not significantly different. Approaches with a performance measure that is not statistically significantly different from the one of best method are printed in boldface.*

	H-T (10^{-1})	H-T (10^{-2})	H-T (10^{-3})	H-T (10^{-4})	H-T (10^{-5})	Price et al.	Wu et al. 1	Wu et al. 2
Accuracy	**0.9826**	**0.9829**	**0.9829**	**0.9829**	**0.9829**	0.9817	**0.9824**	0.9817
Brier score	24.784 10^{-3}	3.7268 10^{-3}	**2.8853** 10^{-3}	**2.8923** 10^{-3}	**2.8941** 10^{-3}	3.1379 10^{-3}	**2.9179** 10^{-3}	2.9640 10^{-3}

technique by Hastie and Tibshirani are not significantly different. Their accuracies are significantly higher than the one of the method by Price et al. and the second approach by Wu et al. By looking at the averaged Brier scores, one can compare the prediction uncertainties for all methods. The results of the approach by Hastie and Tibshirani seem to degrade if the stopping condition is too loose. If the convergence criterion is taken too high, the iterative procedure produces significantly higher Brier scores than the other methods. When this stopping condition is set small enough, the method of Hastie and Tibshirani performs equally well as the first method by Wu et al. Furthermore, the method by Price et al. seems to produce smaller Brier scores than the approach by Hastie and Tibshirani for some stopping criteria, while its accuracy is smaller. By analyzing the results, it was observed that the algorithm by Price et al. predicts more extreme class probabilities. As such, when correct predictions are made, these extreme predictions cause a decrease in the Brier score.

A confusion matrix is provided for each approach (Figures 7.1 − 7.4, Section 2.3.1). The corresponding class label for each tissue type is provided in Table 7.1. Because of the method's performance and computational simplicity, the focus is on the confusion matrix, produced by the first technique by Wu et al. With respect to this data set, all normal tissue cases are classified correctly. For CSF an accuracy of 99.66 % is obtained. The accuracy for grade II diffuse astrocytomas is 96.78 %. Grade II oligoastrocytomas attain an accuracy of 93.97 %. The performance for grade II oligodendrogliomas (92.05 %) is somehow downgraded. Grade III astrocytomas obtain an accuracy of 96 %. The accuracies for grade III oligoastrocytomas and grade III oligodendrogliomas are 99.90 % and 98.80 %, respectively. Meningiomas achieve an accuracy of 95.82 %, while 98.53 % is obtained for glioblastomas. In total, 98.24 % of the cases are classified correctly. The first method by Wu et al. is also used in combination with classical LDA for comparison with Bayesian LS-SVMs. The same experimental setup resulted in an accuracy of 96.31 % for pairwise coupling of LDA classifiers. According to the Wilcoxon signed-rank test this performance is significantly different (i.e. smaller) from the one of the LS-SVM-based approach [91].

7.5 Discussion

In this study MRSI and MRI are combined to construct a multi-class classifier system for brain tumors. To start the discussion, a remark has to be made about the data set. In this chapter the grade II glioma and grade III glioma classes were further split up into three different subtypes (i.e. (diffuse) astrocytomas, oligoastrocytomas, oligodendrogliomas). In this way, the number of cases in each

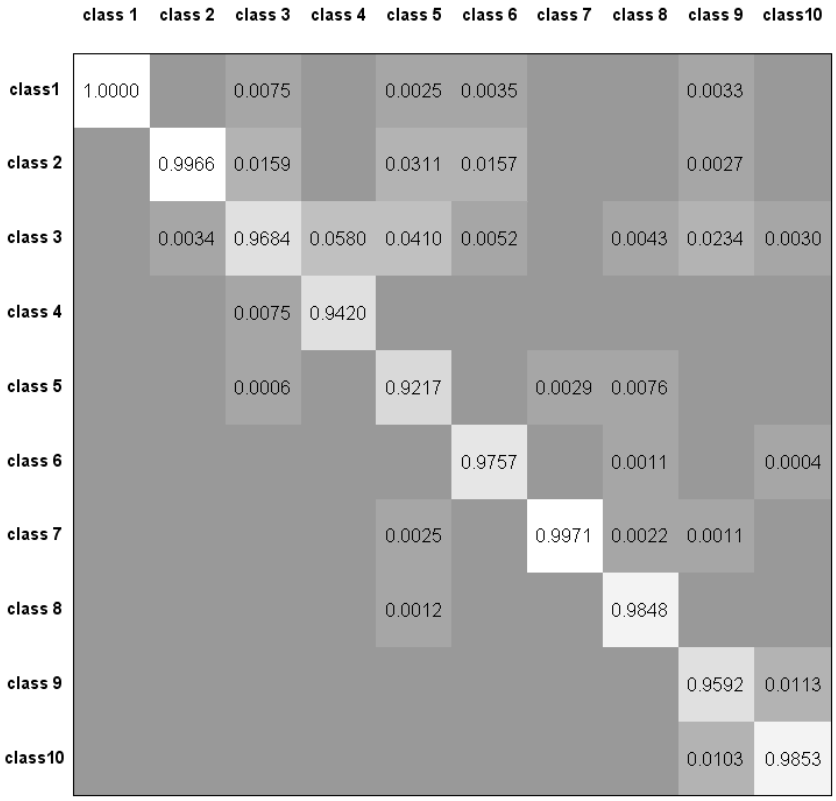


Figure 7.1: *Confusion matrix for the method by Hastie and Tibshirani (with convergence criterion* 10^{-3}*) on test set over 100 runs of stratified random sampling. On the horizontal axis the true classes are indicated, the vertical axis represents the test set predictions. The corresponding class label for each tissue type is provided in Table 7.1.*

class decreases, and also the number of patients per tissue class decreases. This results in the construction of test sets that potentially contain voxels coming from patients from which other voxels were selected for training classifiers. Strictly speaking, these test sets were not always totally independent. One has to keep this in mind when interpreting the classification performances. On the other hand, the goal of this chapter is not to focus on the global performance of the classifier. The aim is to compare different methodologies and show their importance for brain tumor classification. Additionally, it is widely known that SVM-based techniques can handle high-dimensional input spaces and smaller data sets. Moreover, Wu et al. pointed out that the differences between the pairwise combination schemes

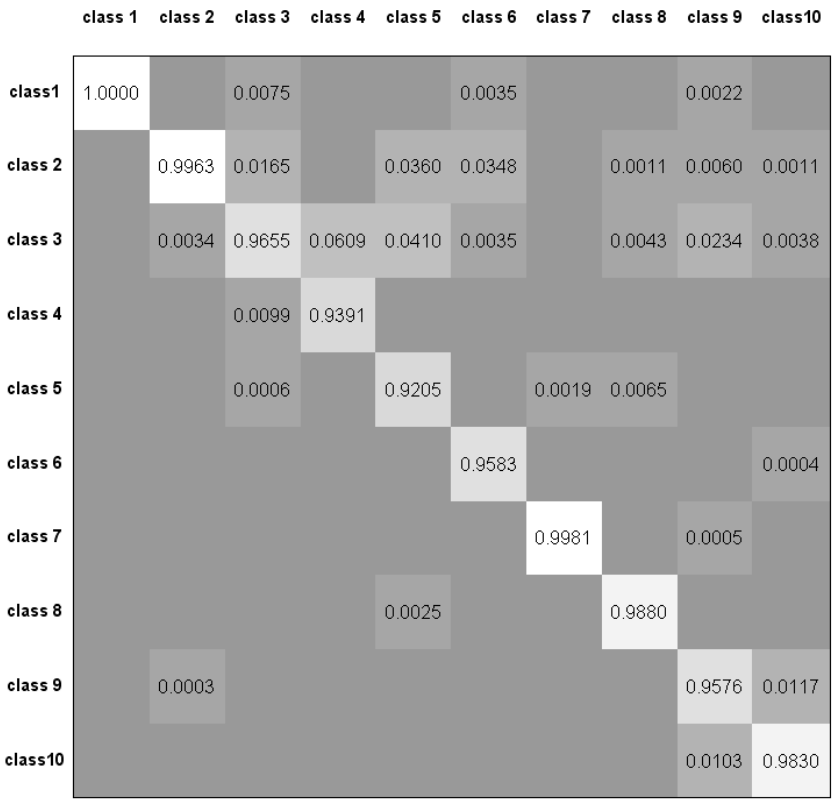


Figure 7.2: *Confusion matrix for the method by Price et al., cf. Figure 7.1.*

become more pronounced with an increasing number of classes [241]. Therefore, it was decided to perform the analysis with 10 classes. In a later phase, it becomes interesting to create and validate the performance of new multi-class classifiers when much more multi-center data become available via acquisition through international projects as HealthAgents or eTUMOUR.

Wu et al. argue that the differences between the algorithms increase in case of a higher number of classes (e.g. $n_C = 10$). In particular, this has a stronger impact on the performance of the method by Hastie and Tibshirani. Like in [241], it is noticed that the results of the approach by Hastie and Tibshirani are dependent on the stopping condition. In our analysis, modifying the stopping criterion led to an improvement in the performance of this method. But, since the choice of this condition is application dependent, it is not clear how to choose an

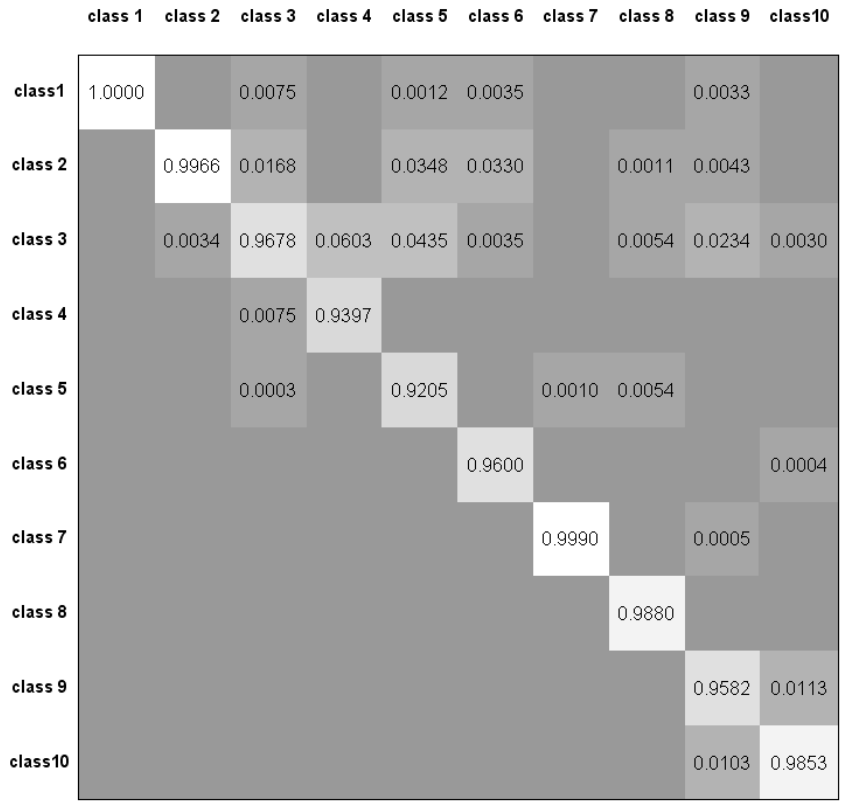


Figure 7.3: *Confusion matrix for the first method by Wu et al., cf. Figure 7.1.*

appropriate stopping criterion in advance, while avoiding an extensive amount of computationally intensive iterations. Since this is a considerable drawback of the method by Hastie and Tibshirani, the first non-iterative procedure by Wu et al. is preferred in the remaining of the thesis. This method can be implemented in a very straightforward way.

As mentioned before, the main aim of this chapter is to introduce new methodologies for the diagnosis of brain tumors. Certain trends can be observed when looking at the confusion matrix obtained by the method by Wu et al. although due to the nature of the data set and the retrospective character of this study, one has to be careful when drawing medical conclusions. Normal tissue is clearly recognized by the classifier. Some tumor classes are mixed with CSF voxels. This can be clarified by the fact that all the CSF voxels are coming from

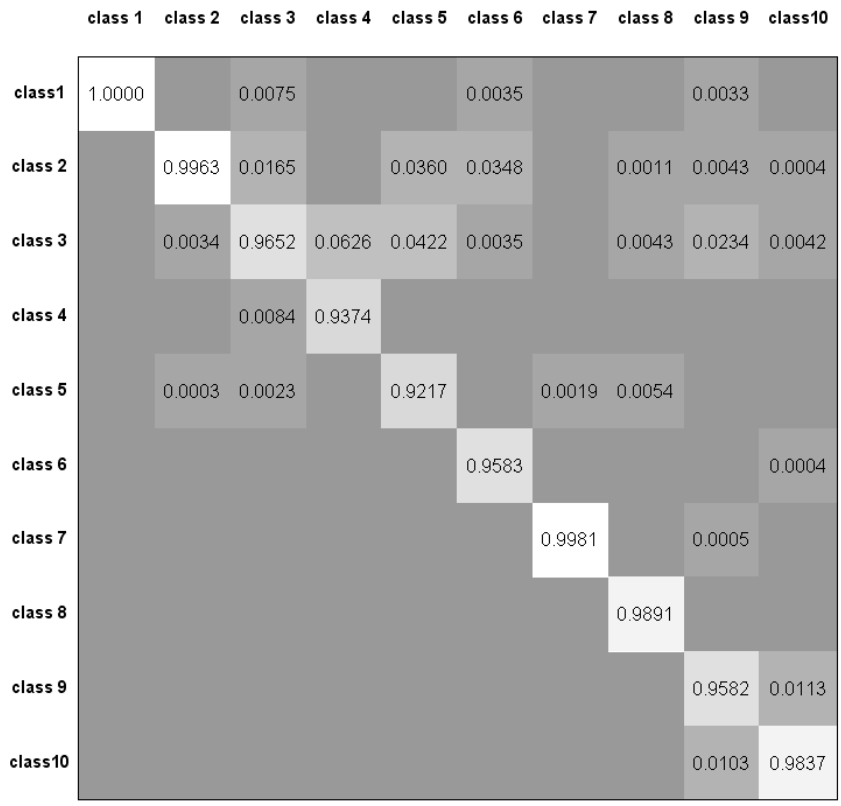


Figure 7.4: *Confusion matrix for the second method by Wu et al., cf. Figure 7.1.*

patients; this may have an influence. Grade II diffuse astrocytomas are classified rather well. Most of the time, this tumor class is mixed with CSF and grade II oligoastrocytomas. A possible explanation for the relatively poor performance of grade II oligodendrogliomas is the small number of cases in this class. Grade III astrocytomas are sometimes mixed with the lower grade diffuse astrocytomas. Since the number of cases for this class is small, more data should be acquired. The accuracies for grade III oligoastrocytomas and grade III oligodendrogliomas are fairly good, however, also more data is needed for additional validation. Meningiomas tend to be mixed with grade II diffuse astrocytomas and grade IV gliomas. Conversely, grade IV gliomas are sometimes confused with grade II diffuse astrocytomas and meningiomas. Compared to classical LDA, the LS-SVM-based approach achieves a significantly higher performance. This was also reported in [44] and can be explained by the fact that certain subtypes of tumors are hard to distinguish with a linear method.

7.6 Conclusions

While the previous chapters exclusively focused on binary classification, the goal of this chapter was to demonstrate that standard Bayesian LS-SVM classifiers can be extended to a multi-class classifier system, thereby providing class probabilities. The one-versus-one combination scheme was used to split up the multi-class problem into a number of binary problems. This approach resulted in more balanced data sets and classifier training might be less computationally intensive. Moreover, in medical practice the binary classifiers can be interpreted as stand-alone building blocks, which can be combined for multi-class classification. Various pairwise coupling methods were compared on a combined MRSI and MRI data set, containing samples from 10 classes. The first combination method by Wu et al. was preferred because of its performance and computational efficiency. The developed multi-class system can be of importance for the construction of DSSs, which assist clinicians in the diagnosis of brain tumors. For example, the developed methodology is used in the following chapter to generate nosologic images of the brain.

Chapter 8

Nosologic imaging of the brain

This chapter presents a new technique to create nosologic images of the brain based on MRI and MRSI. The proposed technique applies methods from image processing as well as classification approaches from Chapter 2 and Chapter 7 to segment and classify brain tumors. Firstly, Section 8.1 introduces nosologic imaging. Section 8.2 describes the MRSI and MRI data for the experiments, while the segmentation and classification approaches are explained in Section 8.3. Examples of nosologic images are presented in Section 8.4. The results are discussed in Section 8.5.

8.1 Nosologic imaging

Automatic brain tumor segmentation based on MRI is a widely studied topic [32, 233, 106, 237, 60, 105, 75, 36, 200]. Techniques based on thresholds, region growing, clustering, Markov random fields, classification algorithms and ANNs were used and accurate segmentation results were reported after comparison with manual tumor segmentation by an expert. However, conventional MRI segmentation results do not provide an answer to several diagnostic questions, such as the type and grade of the tumor. On the other hand, MRSI provides quantitative metabolite maps of the brain, but easy and rapid display of the results as images or maps is needed for routine clinical interpretation of an exam. As such, the term nosologic image was introduced for an image which indicates a specific tissue type, which is determined via MRSI, in a certain color [216]. Szabo de Edelenyi et al. used pattern recognition techniques and combined one specific contrast from MRI with spectroscopic information. Simonetti et al. [203] combined additional image variables with the metabolic information to construct

nosologic images. Both studies used the combination of MRI and MRSI for new, more relevant brain tumor classifiers, but treated each pixel or voxel of the nosologic image independently, only exploiting the spectral information and image intensities. Therefore, Laudadio et al. proposed to apply CCA to MRSI data as a way to include spatial information when producing nosologic images [121, 119]. Afterwards, Kelm et al. [108] suggested to apply conditional random fields to MRSI data since the technique in [121, 119] uses a sliding window which can cancel out effects from isolated voxels and some long-range interactions are lacking because of the fixed window size. Kelm et al. claimed that their approach, based on a global method, is able to increase the performance of single-voxel classifiers by at least 15 %.

This chapter extends previous work to construct improved, high-resolution nosologic images. The new approach benefits from both advanced image processing methods, producing an improved segmented image, and classification methods from pattern recognition. The framework comprises a segmentation step and a classification step where MRSI and MRI data are simultaneously exploited. Furthermore, spatial information is included in both steps and class probabilities are generated.

8.2 Materials

MRSI and MRI data from 24 patients with a brain tumor and 4 healthy volunteers, were selected from the INTERPRET project database. The data set and its preprocessing is similar to the description in Section 4.2.2 and includes T_1-weighted MR images, T_2-weighted MR images, PD-weighted MR images and Gd-enhanced T_1-weighted MR images, with matrix size equal to 256×256. Short TE MRSI data (32×32) are assumed to be registered with the PD-weighted MR image since they were acquired in a consecutive manner. An input pattern for pattern recognition is either a magnitude spectrum (i.e. [0.5,4.0] ppm) or a set of features extracted from the spectrum (i.e. the so-called peak-integrated values Cr1, Glx1, mI + Gly, Tau + Glc, Cho, Cr2, Glx2, NAA, Lip1 + Lac + Ala, and Lip2) in combination with MRI pixel intensities (i.e. T_1, T_2, PD, Gd). Finally, Table 8.1 provides an overview of the patients' data that are used for generating the nosologic images in Section 8.4.

Table 8.1: *Pathology, tumor grade, age and sex for 11 subjects.*

Case	Pathology	Grade	Age	Sex
1	diffuse astrocytoma	II	57	male
2	diffuse astrocytoma	II	58	female
3	diffuse astrocytoma	II	39	male
4	oligoastrocytoma	II	42	male
5	oligodendroglioma	II	48	female
6	oligodendroglioma	II/III	41	female
7	oligodendroglioma	III	30	female
8	glioblastoma	IV	53	male
9	glioblastoma	IV	54	male
10	glioblastoma	IV	50	male
11	meningioma	I	45	female

8.3 Methods

The methodology to create nosologic images is based on a two-step segmentation-classification framework, which is able to simultaneously exploit MRI and MRSI data. In particular, the first step is performed based on the segmentation method presented by Prastawa et al. in [180]. This method requires at least a T_2-weighted MR image. No contrast-enhanced images are needed, but, additional types of MR images can be used to improve the segmentation. More precisely, Prastawa et al. proposed a method consisting of three stages: abnormality detection using a digital brain atlas and outlier detection, edema detection based on the T_2-weighted MR image intensities, and the application of geometric and spatial constraints to the detected edema and tumor regions. The new approach that is presented here includes additional information from MRSI to improve abnormal tissue detection for segmentation (Algorithm 5). Specifically, in addition to the standard brain atlas, a subject-specific abnormal tissue prior is generated by applying a supervised pattern recognition technique to the available MRSI (and MRI) data set. In the second step of the segmentation-classification framework all pixels from the detected abnormal region are classified based on supervised pattern recognition techniques and nosologic images are generated as a final result, where each tissue type is identified by a specific color.

8.3.1 Subject-specific abnormal tissue prior

A digital brain atlas only provides probabilities for the healthy tissue classes (i.e. white matter, gray matter and CSF). In this chapter, a subject-specific prior is

Algorithm 5 : *Brain tumor segmentation using an abnormal tissue prior*

Input:
MR images
MRSI data
digital brain atlas

{Supervised pattern recognition to obtain a subject-specific abnormal tissue prior}
generate an abnormal tissue prior based on the use of LDA on the MRSI data
include the abnormal tissue prior in the original digital brain atlas

{Detection of abnormality: detecting regions that deviate from normal, healthy brain involves finding the intensity parameters for the healthy tissue classes (i.e. white matter, gray matter and CSF) and the abnormal tissue. The first stage of the algorithm involves an iterative procedure that uses a modified digital brain atlas.}
loop
    sample high-confidence regions by using a threshold on the posterior probabilities. In the first iteration the modified atlas probabilities are used instead of the posterior probabilities.
    perform outlier detection based on the MCD method, since the samples for the healthy tissue classes contain unwanted samples from abnormal tissue classes [229]
    estimate probability density functions for all tissue classes based on a nonparametric model using kernel expansions. The initial density for the abnormal tissue class is uniform.
    calculate posterior probabilities following Bayes' rule, while the priors are obtained from the corresponding modified atlas probabilities
    estimate the bias field from the probabilities of the gray and white matter classes [232]. A correction is applied based on the estimated bias field.
end loop

{Reclassification with spatial and geometric constraints: geometric and spatial constraints are included since the previous step generally results in a number of false positives. This involves the pruning of samples before estimating the densities. This stage is iterative and part of it is similar to the previous stage.}
loop
    sample high-confidence regions by using a threshold on the posterior probabilities
    prune abnormal tissue samples using region competition snakes [90]. The abnormal tissue posterior probabilities are used as input for the snake, which is represented by a zero level-set. The snake expands when the boundary is inside the abnormal tissue and shrinks when the boundary encloses part of the regions not part of the abnormal tissue.
    estimate probability density functions based on a nonparametric model using kernel expansions

    calculate posterior probabilities following Bayes' rule
    perform bias field correction
end loop

Output:
image with segmented gray matter, white matter, CSF and abnormal tissue

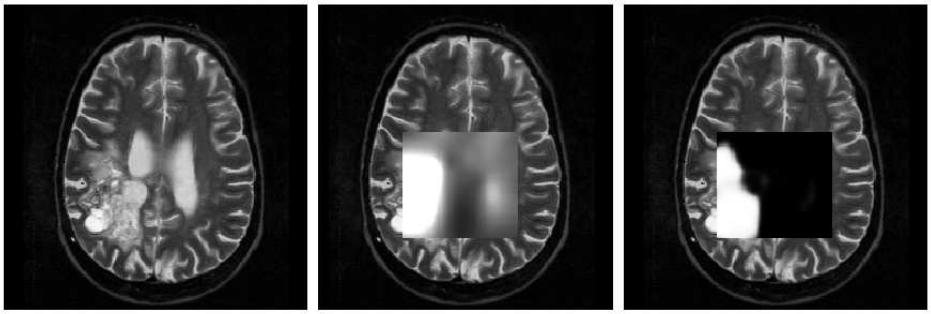


Figure 8.1: *From left to right: the T_2-weighted MRI for case 6, the ratio Cho/Cr and the probabilities generated by the LDA classifier for abnormal tissue. The lighter the map is, the higher the ratio or probabilities.*

generated for the abnormal tissue as well, but based on the result of an LDA classifier. LDA is applied to the grid of feature vectors containing fourteen entries, i.e. ten bicubically interpolated peak-integrated values (such that it matches with the size of the MRI grid) and the four types of MRI pixel intensities. For each pixel within the area of the STEAM box, it is predicted whether there is abnormal or healthy tissue based on all peak-integrated values and the MRI pixel intensities. The right panel of Figure 8.1 depicts the resulting grid that is generated by the LDA classifier for case 6. For illustration purposes, the ratio Cho/Cr, obtained by the interpolated peak-integrated values, is also depicted in the middle.

Next, the output from LDA is integrated into the original digital brain atlas by setting the abnormal prior, if abnormal tissue is predicted by the LDA method, to a probability derived from that classifier. If no abnormal tissue is predicted by LDA, the abnormal prior is set to a fraction of the white matter and gray matter atlas probabilities as in [180]. Since the MRSI information is restricted to the STEAM box, the abnormal tissue prior is only available within that area. Every prior probability outside that box is obtained as in [180].

8.3.2 Tumor segmentation based on outlier detection

After the prior is generated, part of the segmentation approach from [180] is followed. Abnormal tissue is detected using an iterative procedure. Firstly, a threshold is applied to the posterior probabilities, or atlas probabilities in the first iteration, and pixels in high-confidence regions are sampled. Next, the samples from normal tissue that exceed a distance threshold, based on the minimum

covariance determinant (MCD) estimator, are removed. A non-parametric model is used to estimate the probability density functions from the samples. The posterior probabilities are calculated based on the subject-specific atlas priors and the estimated densities. To conclude an iteration of abnormality detection, a bias field correction is performed because of the presence of the image inhomogeneity. The bias field is estimated from the white and gray matter probabilities and the full procedure is repeated. After the separation of abnormality, an edema detection stage was proposed in the framework in [180]. This detection was done using the T_2-weighted MR image, assuming that edema has high fluid content and appears brighter than the tumor. This stage is removed from the present framework since cysts, which are also brighter on the T_2-weighted MRI, can easily be segmented as edema following this approach. Also, it is known that edema in glioma tumors is often infiltrated and therefore this tissue should be detected as abnormal. The last stage of the segmentation step is reclassification with spatial and geometric constraints. For this purpose region competition snakes are used [90]. Firstly, the procedure for the last stage uses a threshold and samples the posterior probabilities. The abnormal tissue samples are pruned based on the level-set evolution after which the densities are estimated. New posterior class probabilities are calculated; the bias field is corrected and the procedure is repeated.

8.3.3 Classification of the abnormal region

After segmentation, white matter, gray matter, CSF and abnormal tissue are separated. To provide additional information about the tumor type that is present within the abnormal tissue region, a supervised pattern recognition step is introduced. As for the subject-specific abnormal tissue prior, the MRSI data (e.g. spectra, peak-integrated values) are interpolated to match the size of the MRI data. Naturally, only pixels within the STEAM box can be classified. More precisely, all pixels, found to be abnormal by the segmentation step, are classified into one of the tumor classes that are represented in the training set for pattern recognition. This means that for this specific analysis, white matter, gray matter, CSF, and, based on spectroscopic profiles, grade II glioma, grade III glioma, grade IV glioma and meningioma can be detected.

Concerning classification, basically any (semi-)supervised pattern recognition method can be incorporated, which makes the framework very flexible. This is illustrated by applying KLR, the multi-class classifier based on Bayesian LS-SVMs (Chapter 7) and CCA. The symmetric 3×3 model without corner voxels is used as spatial model and PCA is applied as subspace model.

8.3.4 Experimental setup

Training data are required to generate the subject-specific abnormal tissue prior with LDA and to perform the classification step with KLR, CCA or Bayesian LS-SVMs. A leave-one-case-out analysis is performed to create an independent training and test set. All data from the test subject are removed from the training set. Model selection is performed on all data except for the voxels coming from the subject under study.

8.4 Results

8.4.1 Analysis of the patient data

The framework is applied to generate nosologic images for 10 patients (Figures 8.2 − 8.8). Light blue reflects white matter, dark blue gray matter, green CSF, yellow grade II glioma, orange grade III glioma, dark red glioblastoma and purple meningioma. All classification results in this section are generated using feature vectors where peak-integrated values are combined with image intensities. To demonstrate the flexibility of the framework, and more specifically the flexibility of the classification step, the use of various pattern recognition methods is illustrated.

Cases 9, 8 and 10 represent patients with glioblastoma multiforme as assigned by the pathologists. The nosologic images, obtained by applying a Bayesian LS-SVM classifier, show that the method is able to detect the tumor lesions and glioblastoma tissue (Figures 8.2 and 8.3). These examples also show that edema and/or tumor infiltration is difficult to assess. In diffuse astrocytoma, the peritumoral edema is presumed to be infiltrated by glioma tumor cells. In these cases the differentiation between pure infiltration and pure edema can not be made with MRI. As discussed before, the segmentation of edema is not consistent and is left out of the procedure such that no edema is shown in the nosologic images. In heterogeneous tumors, like gliomas, this difference indeed can not be made with conventional MRI techniques. Therefore, the most likely assignment for this abnormal tissue is low-grade glioma. Low-grade tumor next to glioblastoma is pathologically possible. In case 10, there are two tumor areas. The posterior part could include a hematoma within the tumor. For the surrounding tissue it is difficult to distinguish edema from infiltrating tumor.

Cases 4, 6 and 7 are good examples of the difficult differentiation between WHO grade II gliomas and grade III gliomas (Figures 8.4 − 8.6). The histopathological

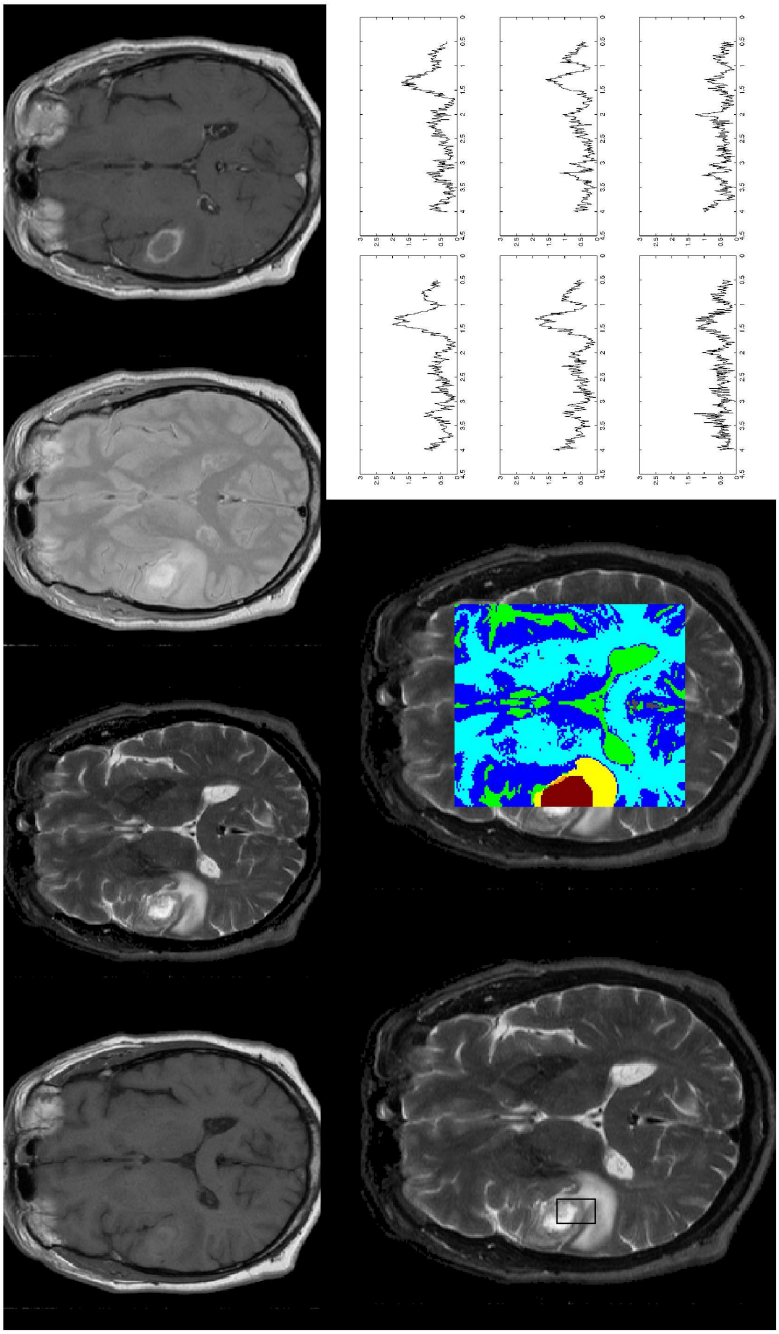


Figure 8.2: *Case 9, glioblastoma. Top: T_1-weighted MR image, T_2-weighted MR image, PD-weighted MR image and Gd-enhanced T_1-weighted MR image. Bottom: T_2-weighted MR image, nosologic image, MR spectra located in the box shown in the T_2-weighted MR image on the left. Light blue reflects white matter, dark blue gray matter, green CSF, yellow grade II glioma, orange grade III glioma, dark red glioblastoma and purple meningioma.*

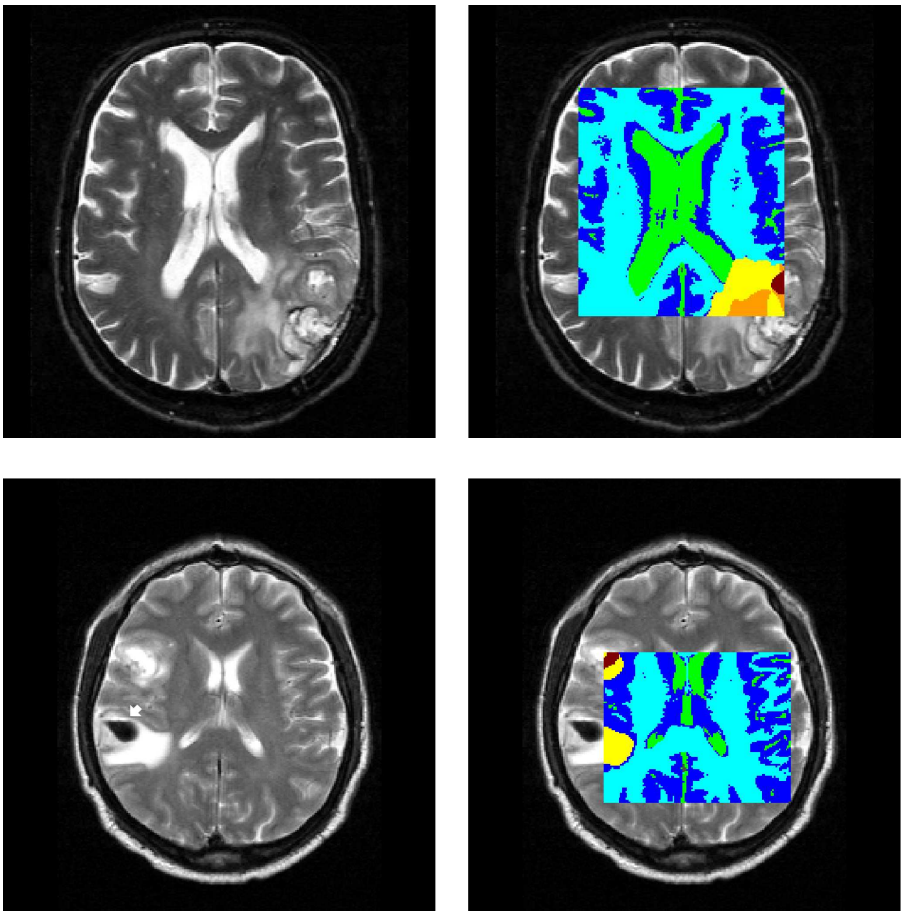


Figure 8.3: *Top: case 8, glioblastoma. T_2-weighted MR image and nosologic image. Bottom: case 10, glioblastoma. T_2-weighted MR image and nosologic image. The hematoma is identified by an arrow.*

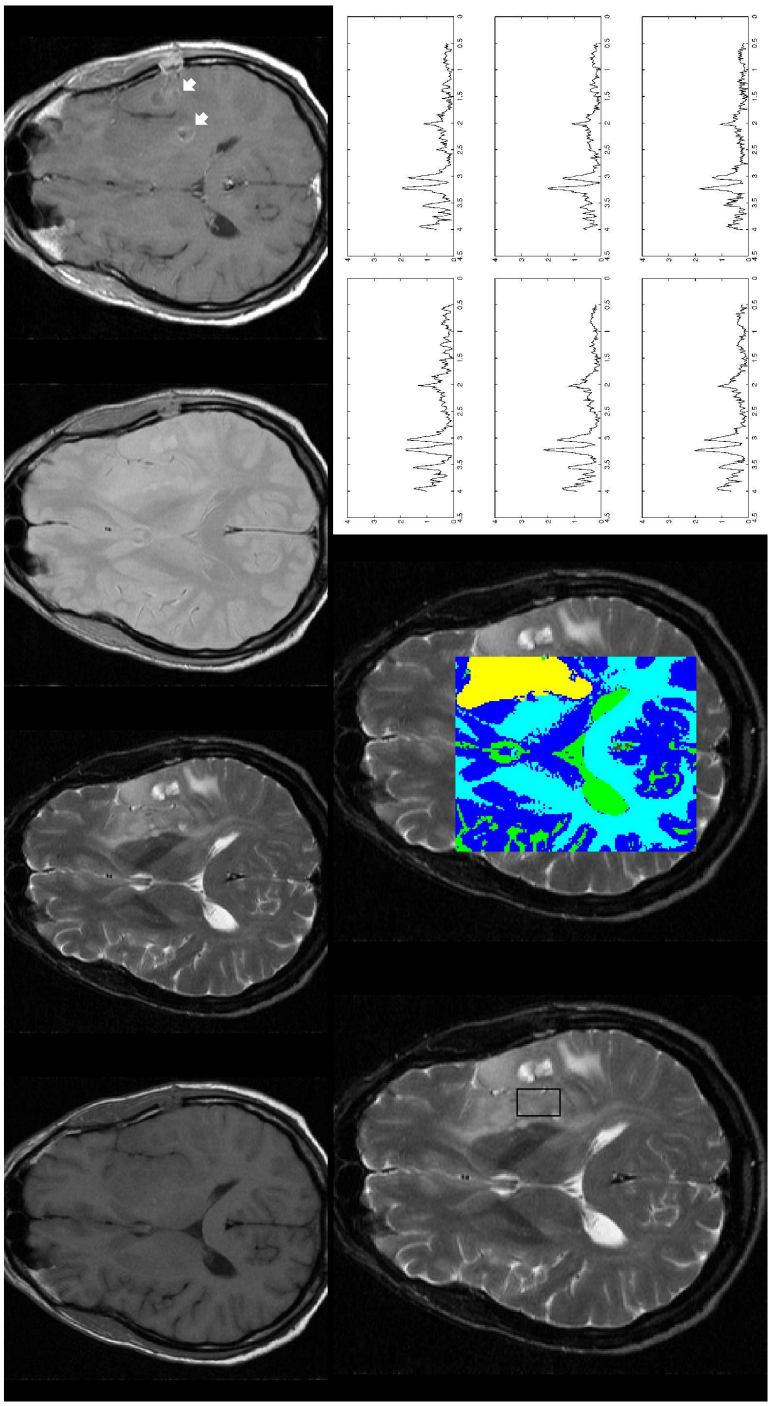


Figure 8.4: *Case 4, grade II glioma. Top: T_1-weighted MR image, T_2-weighted MR image, PD-weighted MR image and Gd-enhanced T_1-weighted MR image. Bottom: T_2-weighted MR image, nosologic image, MR spectra located in the box shown in the T_2-weighted MR image on the left. The arrows identify the enhanced area.*

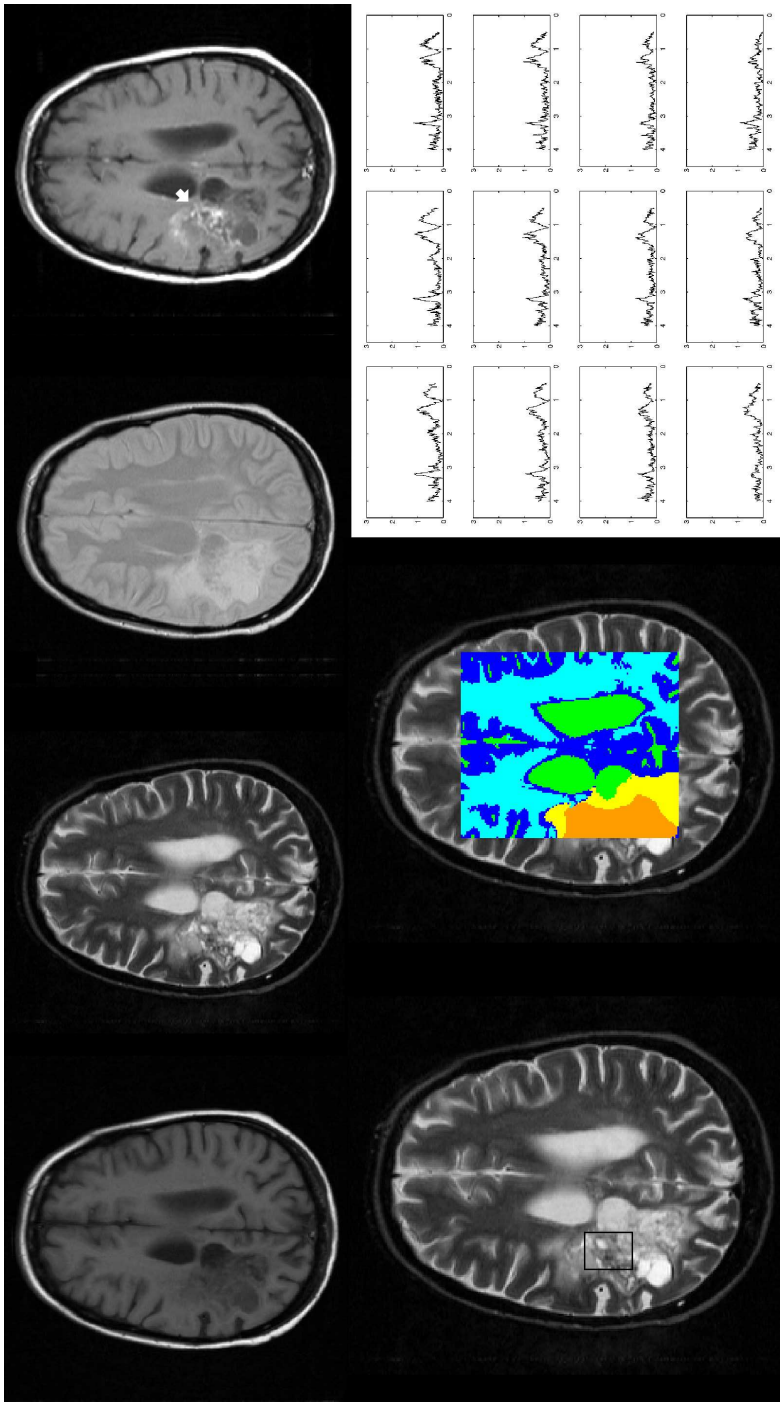


Figure 8.5: *Case 6, grade II/III glioma. Top: T_1-weighted MR image, T_2-weighted MR image, PD-weighted MR image and Gd-enhanced T_1-weighted MR image. Bottom: T_2-weighted MR image, nosologic image, MR spectra located in the box shown in the T_2-weighted MR image on the left. The arrow identifies the enhanced area.*

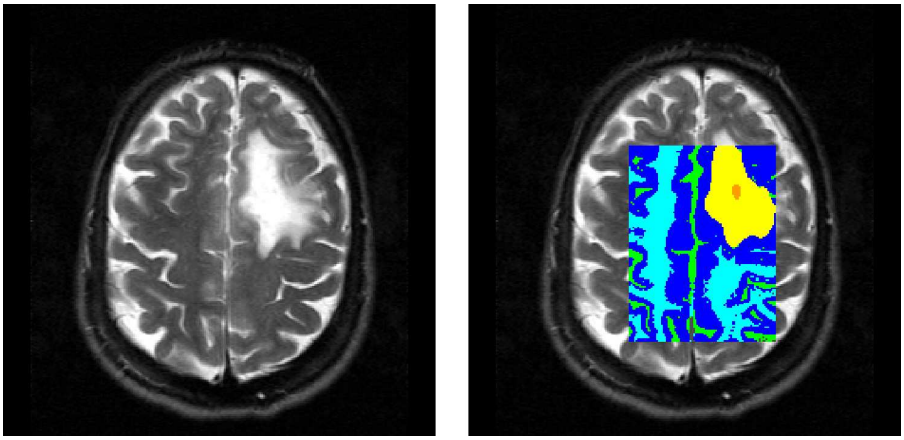


Figure 8.6: *Case 7, grade III glioma. Left: T_2-weighted MR image, right: nosologic image.*

diagnosis for case 4 was grade II glioma. The framework, applying CCA in the classification step, finds the correct tissue type, despite the enhancement after contrast for small parts of the tumor at the occipital parts of the insula (Figure 8.4). For case 6 the pathologists disagreed between a grade II and a grade III glial tumor. The nosologic image, created using a Bayesian LS-SVM classifier, shows a heterogeneous tumor with mainly grade III glioma, according to the spectroscopic classifier, especially in the part with enhancement after Gd (Figure 8.5). The differentiation between grade II glioma and edema at the frontal and medial parts can not be made. For case 7 with an anaplastic oligodendroglioma, CCA is used (Figure 8.6). Almost only low-grade glioma, WHO grade II, is found with a small region of anaplastic glioma, WHO grade III. This would be an underestimation of the real situation. At surgery, a debulking of the tumor was performed. Therefore, the pathological diagnosis was not from only the small part with a nosological anaplastic tumor.

Case 3 is a diffuse low-grade astrocytoma, grade II, with no enhancement after contrast and good segmentation of the tumor/tumor infiltration/edema based on KLR (Figure 8.7). For case 2 the histopathological agreement was grade II glioma, despite the enhancement after Gd (Figure 8.8). The Bayesian LS-SVM method is used and it detects the correct tissue type. However, in this case no edema and/or tumor infiltration is detected which is very reliable around the enhancing spot. The nosologic image for case 11 with a meningioma tumor is obtained by the CCA classifier (Figure 8.8). In this case the small edema part at the occipital part of the tumor is not classified as abnormal tissue. In the case of a low-grade meningioma there is no infiltration.

8.4.2 Class probabilities

The use of class probabilities is illustrated for three patients (Figures 8.9 − 8.10). The histopathology for case 5 was grade II glioma and the result in the middle panel is obtained using a KLR classifier (Figure 8.9). As input for the classifier the magnitude spectra are used. It seems that a region of grade III glioma is found at the top of the tumor. The right panel is constructed by only considering pixels for which the probability of the classification result is more than 60 %. If the probability for a pixel is lower than 60 %, the pixel is left black. In the resulting image one can observe that the classifier is more confident about the grade II glioma pixels and the region of grade III glioma is now entirely dropped. It is possible to generate probability maps for various tissue types. In Figure 8.10 a probability map and a contour plot are depicted for case 6. The lighter the probability map is, the higher the probability for grade III glioma. Blue contour lines show higher grade III glioma probabilities as opposed to red contour lines. The grade III glioma probabilities are higher at the frontal part of the abnormal tissue region with the enhancement after Gd and smaller at the occipital part of the tumor in case 6. Figure 8.10 also depicts the correlation coefficients generated by CCA for grade III glioma in a map and contour plot for case 7. The correlation coefficients are higher in the center and at the medial part of the tumor.

8.4.3 Subject-specific abnormal tissue prior

During the detection of abnormality step a sampling for gray matter, white matter and CSF is done based on the atlas priors (Figure 8.11). On the left, the T_2-weighted MR images for cases 1, 4, 5 and 6 are depicted. The middle images show which pixels are sampled for the healthy tissue classes (i.e. gray matter, white matter and CSF) following the standard digital brain atlas, as proposed in [180]. It can be observed that for all cases samples for healthy tissue are picked from the tumor area. This contamination can be rather strong as for instance for case 4. On the other hand, the images on the right show which samples are picked for the healthy tissue classes after the subject-specific abnormal tissue prior is added to the standard atlas. This clearly illustrates that the extra prior improves the sampling and, as a consequence, reduces the contamination reported in [180]. As a result, improved sampling leads to higher quality prototypes for all the tissue classes and better density estimates in the detection of the abnormality step. Finally, an example (case 4) is provided of the difference between the two approaches with respect to the final segmentation (Figure 8.12). In contrast to the panels on the right, no abnormal tissue prior is added to the digital brain

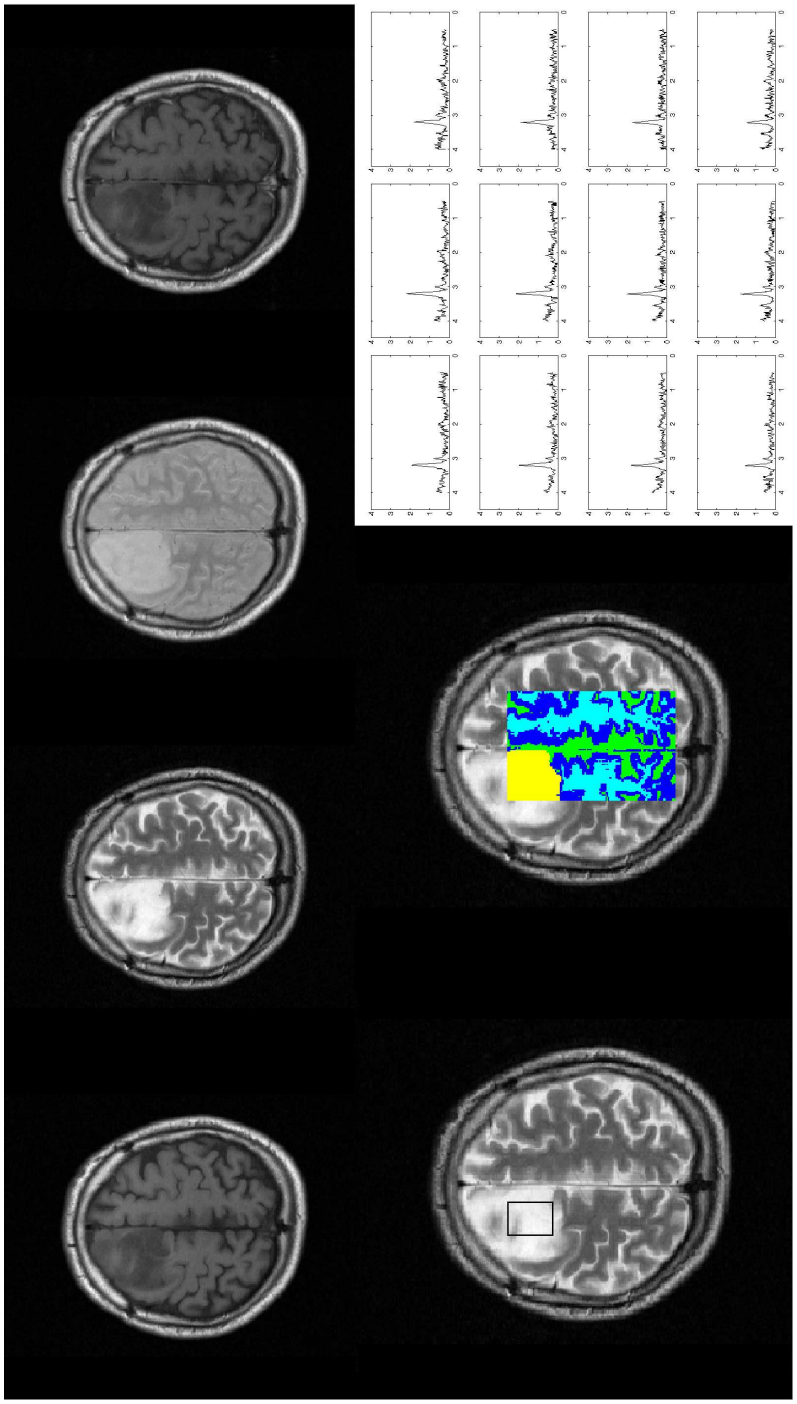


Figure 8.7: *Case 3, grade II glioma. Top: T_1-weighted MR image, T_2-weighted MR image, PD-weighted MR image and Gd-enhanced T_1-weighted MR image. Bottom: T_2-weighted MR image, nosologic image, MR spectra located in the box shown in the T_2-weighted MR image on the left.*

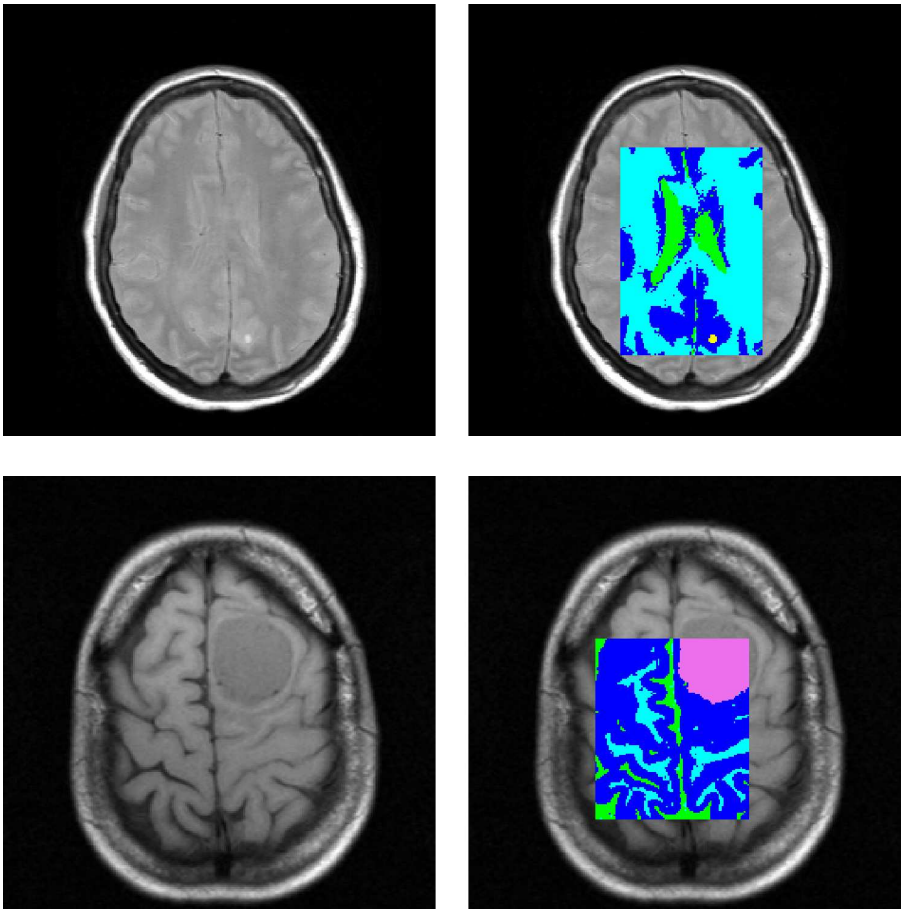


Figure 8.8: *Top: case 2, grade II glioma. PD-weighted MR image and nosologic image. Bottom: case 11, meningioma. T_1-weighted MR image and nosologic image.*

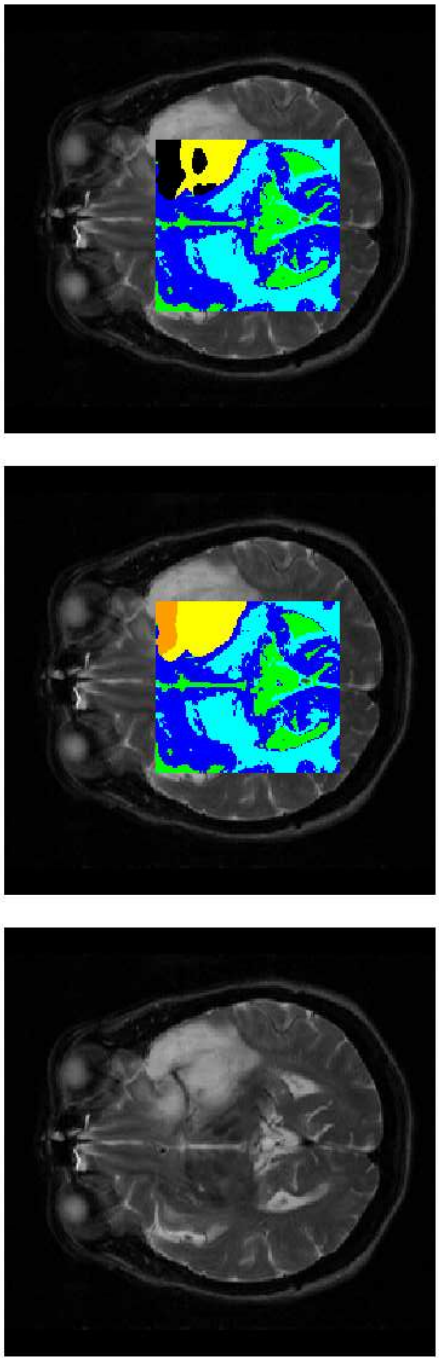


Figure 8.9: *Case 5, grade II glioma. From left to right: T_2-weighted MR image, nosologic image obtained by KLR, nosologic image obtained by using a threshold on the KLR output. The pixels with probabilities higher than 60 % are kept, while the remaining ones are left black.*

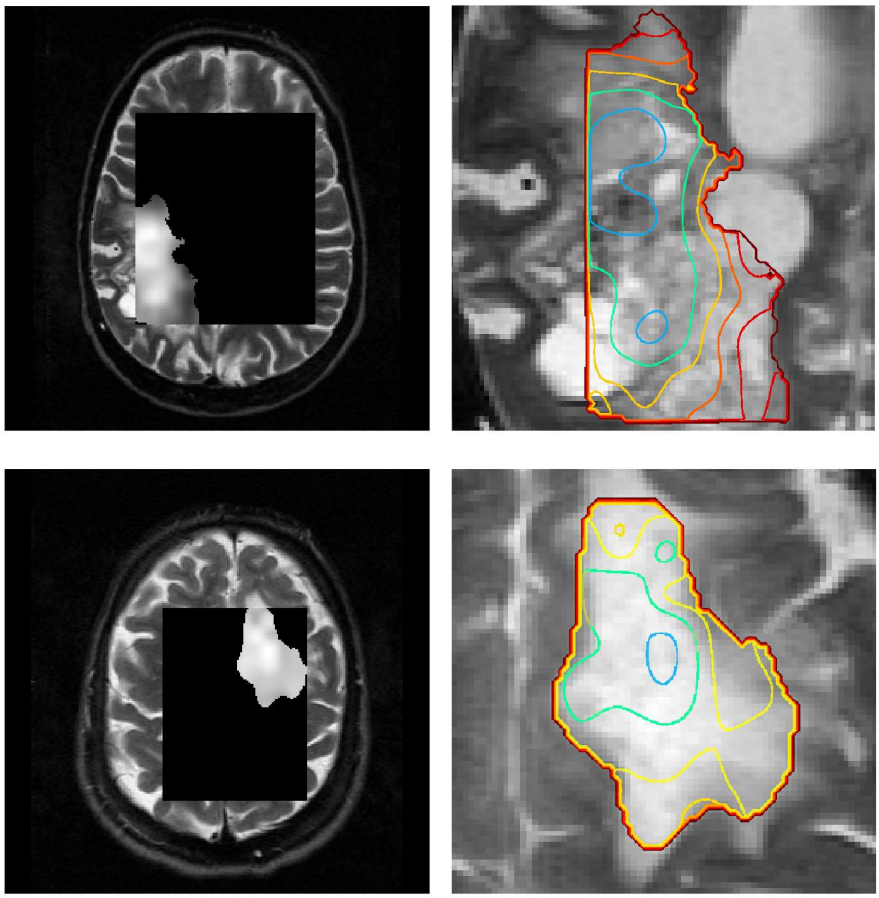


Figure 8.10: *Top: case 6, grade III glioma. Probability map and contour plot. Grade III glioma probabilities are higher at the frontal part of the abnormal tissue region with the enhancement after Gd and smaller at the occipital part of the tumor. Bottom: case 7, grade III glioma. Map of correlation coefficients obtained with CCA and contour plot. The correlation coefficients are higher in the center and at the medial part of the tumor.*

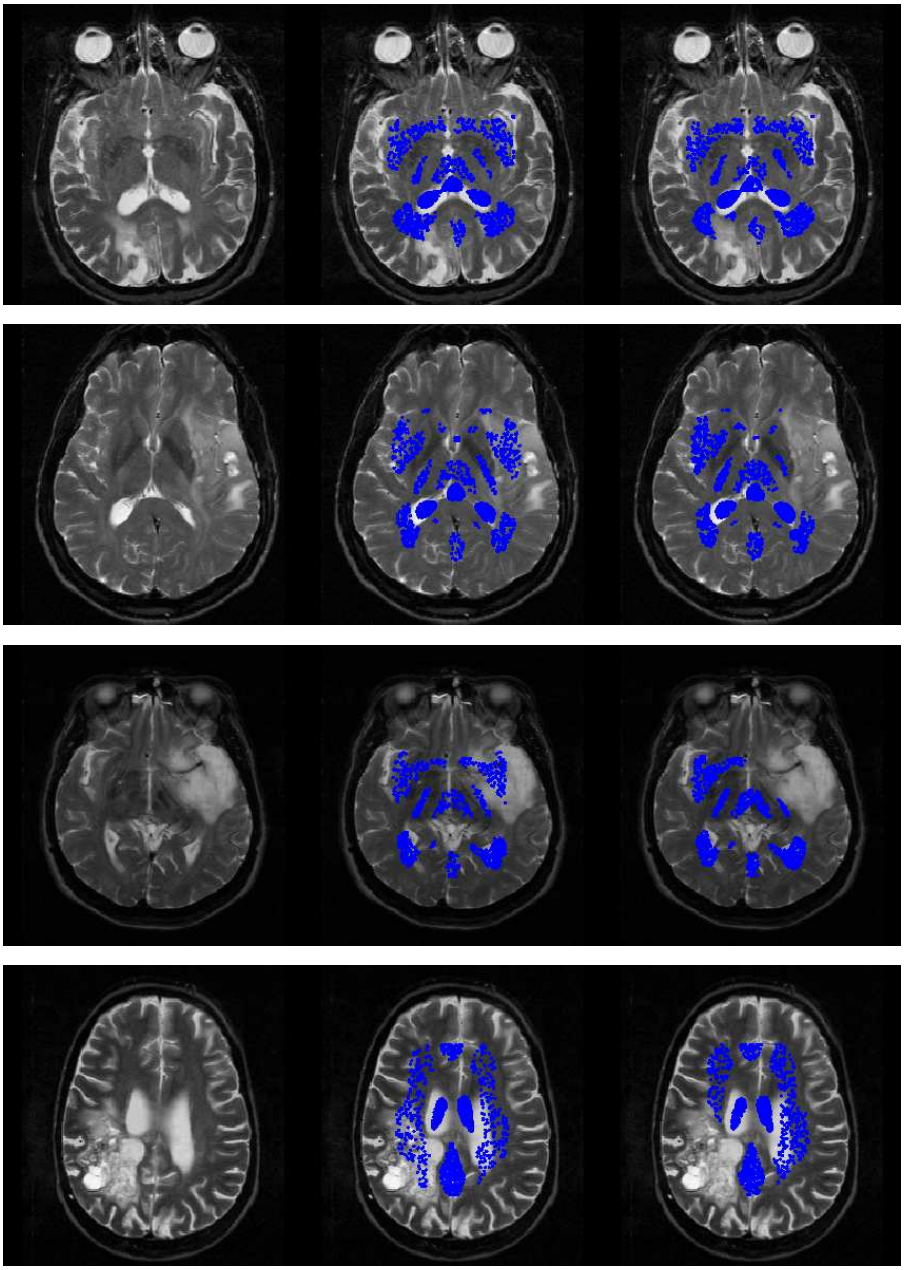


Figure 8.11: *Samples picked for the normal tissue classes (i.e. gray matter, white matter and CSF) using the digital brain atlas during the detection of abnormality step. The T_2-weighted MR images for cases 1, 4, 5 and 6, from top to bottom, are depicted on the left. The images in the middle depict the samples that are chosen based on the original segmentation algorithm described in [180]. The images on the right show the sampling result by including a subject-specific abnormal tissue prior, causing less sample contamination.*

atlas to produce the left images. The top row summarizes the information that is available in the atlas before starting segmentation. This result is generated by choosing the tissue type with the highest prior probability based on the (modified) atlas. The lower row shows the final result after segmentation. CSF is colored green, white matter light blue, gray matter dark blue and abnormal tissue is red. Before starting segmentation, one can observe that there is only a region for which the probabilities of abnormal tissue are higher than the ones of the normal classes in the digital brain atlas that is modified with the subject-specific abnormal tissue prior. After segmentation, and when comparing with the MRI of case 4 (Figure 8.4), only a small part of the abnormal tissue region is detected by the original segmentation algorithm that just includes healthy tissue priors. By introducing the extra prior from MRSI, the abnormal tissue region is better recognized. The MR spectra confirm the existence of abnormal tissue in that area (Figure 8.4).

8.5 Discussion

In MRI-based segmentation of tissue in the brain, a digital brain atlas is an important, additional, tool to resolve the problem of overlap between intensity distributions of different tissue types. Since a digital brain atlas only contains priors for the healthy tissue classes, researchers attempted to include a prior for the abnormal tissue as well. Typically, the contrast-enhanced T_1-weighted MR image is used for this purpose [181]. However, it is known that blood vessels also generally appear enhanced. In addition, tumors may show only partially signal enhancement and some tumors do not enhance at all. For example, in this particular study no enhancement was observed for case 3 (Figure 8.7). Therefore, the MRSI data are used to generate the prior for tissue abnormality in this chapter. Although the spatial resolution of MRSI is low compared to MRI, it appears that MRSI is very useful to define the prior and that it can improve the original segmentation algorithm described in [180]. Indeed, in the study of Prastawa et al. it was explained that the segmentation algorithm could only deal with a limited amount of brain deformation. During the detection of abnormality step the atlas is sampled. The outliers are detected based on the MCD method and the density functions are estimated. However, when there is a serious brain deformation, the brain atlas leads to incorrect sampling since pixels for the healthy tissue classes are sampled from the abnormal tissue regions. In this case the samples are severely contaminated and the MCD algorithm may not yield correct results, leading to incorrect estimated densities. Figures 8.11 and 8.12 illustrate this and show that the use of the abnormal tissue prior, generated from MRSI, is advantageous.

The advantage of the use of Bayesian LS-SVMs [231] or KLR in the classification

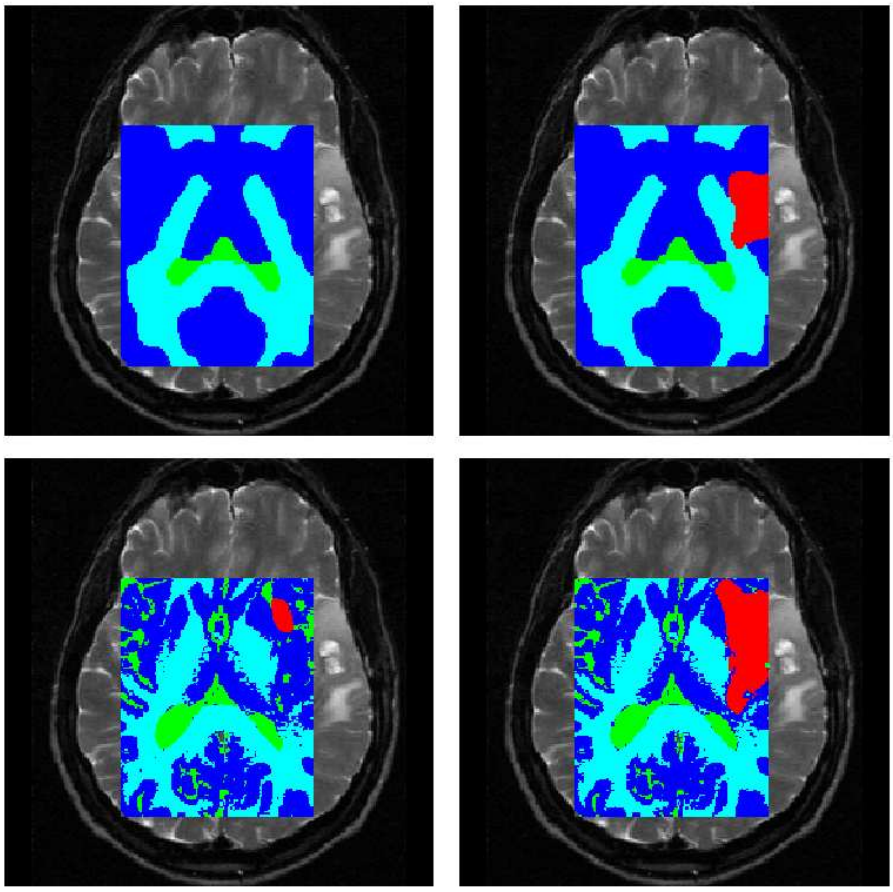


Figure 8.12: *The effect of using a subject-specific abnormal tissue prior based on MRSI for case 4. The results on the left are generated without an abnormal tissue prior in contrast to the results on the right. The top row shows the prior probabilities from the atlas, the bottom row the final segmentation result. The abnormal tissue is only correctly recognized by the segmentation algorithm that includes the extra abnormal tissue prior from MRSI.*

step is that these kernel-based methods are able to provide class probabilities, instead of 'black-or-white' decisions, for each tissue class in the training data. Following the findings in Chapter 7, the combination method of Wu et al. is used to combine pairwise class probabilities from Bayesian LS-SVMs with an RBF kernel [241]. Direct multi-class class probabilities are generated using KLR, but this might be more computationally intensive for large data sets. The advantage of CCA compared to the other techniques is that the method exploits spatial information. Based on the information in the neighboring pixels, a decision is made for the center pixel [121, 119]. Since CCA is applied to the interpolated grid (within the abnormal tissue region), the risk is reduced to average out the effect of isolated tumor voxels by the information that is present in the neighborhood as compared to [121, 119].

In general, the proposed framework highlights the advantages of a method able to exploit and combine different types of information. In particular, the analysis shows the value of adding MRSI over MRI data, of adding MRI over MRSI information, and the improvements achieved in the segmentation and classification approaches with respect to their previously proposed versions. For instance, several diagnostic questions, such as the type and grade of the tumor, are difficult to address using MRI, but, MRSI data enable to take the corresponding metabolic differences into account. Therefore, using MRSI information over just MRI has the additional advantage that nosologic images (specifying the type of the tumor) can be obtained, instead of the results (which only localize the tumor) obtained with classical MRI segmentation approaches as in [180].

The added value of MRI over just MRSI was already evaluated in a number of studies, which concluded that combining MRI with MRSI features improved the performance of the classifiers [44, 204, 67]. For this reason the four MR images are also used in the classification step of the proposed framework. Furthermore, adding MRI over MRSI data, allows to perform the segmentation step based on high-resolution MRI data. This improves the detection of CSF compared to pure segmentation of MRSI data and also gray matter and white matter are detected based on MRI. On the contrary, the CCA approach in [121, 119] only performs a segmentation on the MRSI data. CCA is an interesting method since it is able to simultaneously exploit the spectral-spatial information provided by the MRSI data and it is computationally efficient. On the other hand, the disadvantage of this approach is that the same spatial constraints (i.e. a fixed spatial model) are used for every voxel of the MRSI grid, without taking into account whether the voxel is on the border or the middle of the tumor. This can lead to wrong classifications for rather isolated tumor voxels. In the proposed framework, CCA is applied after the segmentation step, i.e. after the tumor is detected, reducing the risk of misclassification for rather isolated tumor voxels. Therefore, the above

issue is improved by making use of image processing methods, that incorporate more advanced spatial constraints using high-resolution MRI data. The use of these spatial constraints in the current method is also an improvement compared to the approaches where every pixel or voxel was classified separately, ignoring this prior knowledge [216, 203, 138].

On the other hand, as proven by biopsies, outside the abnormal tissue on the conventional MRI infiltrating tumor cells can be found with and without abnormal spectroscopic findings [35, 127, 69, 185]. Therefore, the whole infiltrating tumor is not always taken into account. Price et al. showed a biopsy-guided improved delineation of glioma tumors with diffusion tensor imaging (DTI) [185]. Also, the segmentation of edema in glioma tumors is not straightforward. To distinguish edema from tumor the intensity of a T_2-weighted MR image is not sufficient. This does not always differentiate tumor cysts, resection cysts or cerebral fluid from pure edema. Next to this, in gliomas the peri-tumoral edema is mostly infiltrated by tumor cells and the intensity differences between edema and tumor are on a sliding scale. Ricci et al. showed metabolic changes in peri-tumoral edema with MR spectroscopy [192]. The combination of DTI and MR spectroscopy might be helpful to improve the delineation of gliomas and differentiation of edema and infiltrating tumor. These issues are further studied in Chapter 9.

The difference between WHO grade II and grade III is radiologically very difficult. According to the WHO-2007 classification [133], the main histopathological difference between WHO grade II and III diffuse astrocytic neoplasms is increased mitotic activity in the latter. By including perfusion and diffusion MR data the differentiation may be improved [123, 45].

In this analysis only one of the MRI slices is totally included within the MRSI slice. For this reason, the framework only combines one MRI slice with the MRSI data. However, in practice, depending on its size, multiple MRI slices may be needed to cover the whole tumor. As such, a potential improvement of the existing framework could be the extension to three-dimensional MRSI data with the inclusion of multiple MRI slices. Also, in the present study the standard image orientation protocol in the radiology service is used in order to limit the acquisition time. If a different orientation is used, a better enclosure of the brain tumor mass may be obtained, such that the framework that is presented here provides further improved nosologic images.

8.6 Conclusions

Although MRSI is able to provide metabolite maps of the brain, easy and rapid display of the results for routine clinical interpretation of an exam is still an issue. This chapter proposed a new method to generate nosologic images of the brain by exploiting and combining MRI and MRSI data. Advanced methods from image processing and pattern recognition were combined in a two-step approach. In the first step segmentation was performed using a modified version of an existing method by including an additional subject-specific abnormal tissue prior, derived from MRI and MRSI, in the original brain atlas. The subject-specific abnormal tissue prior was found to improve the segmentation procedure. The segmentation step provided an image of the slice of interest where white matter, gray matter, CSF and the abnormal region were clearly visualized. Next, the abnormal tissue was classified using pattern recognition and class probabilities were generated for the diverse tissue types. The method was applied to various patients with different types of brain tumors. Compared to previous approaches, the use of spatial information and prior knowledge was further improved in the proposed method. The framework offers a new way to visualize tumor heterogeneity and class probabilities in a single nosologic image, assisting clinicians in decision making.

Chapter 9

Differentiation between brain metastasis and glioblastoma multiforme

In the previous chapter a method was developed to create nosologic images of the brain. This methodology is further improved in this chapter by detecting tumor infiltration. The important problem of differentiating between solitary brain metastasis and glioblastoma multiforme is studied (Section 9.1). MRI and MRSI data acquisition protocols and preprocessing are presented in Section 9.2. A technique to visualize tumor infiltration is explained in Section 9.3. Section 9.4 compares different ratios in various parts of the brain for metastasis and glioblastoma multiforme, and presents nosologic images. Finally, these results are discussed in Section 9.5.

9.1 Introduction

Brain metastases and glioblastoma multiforme represent the most aggressive brain tumors of the CNS in adults. It is not always possible to distinguish these very common tumors based on clinical history, the number of lesions and conventional MRI, while both tumors require a different therapeutic approach. MRS was already used in a pattern recognition setting to differentiate glioblastomas from metastases in Chapter 5, Chapter 6 and previous studies [43, 137]. Up till now, these pattern recognition approaches did not provide a reliable differentiation between the two tumor types using short or long TE data. The AUC was 0.60

in [43], 0.64 in [137] and the accuracy was 78 % in Chapter 6. Opstad et al. obtained a sensitivity of 80 % and a specificity of 80 % for the differentiation by considering the lipid peak area ratio (i.e. the total peak area at 1.3 ppm compared to 0.9 ppm), while no significant differences were found in the individual metabolite or lipid concentrations between the classes [166]. Ishimaru et al. found that the absence of Cr in the tumoral area most likely indicates metastasis [98]. Other studies investigated the area surrounding the enhancing region in glioblastoma since these tumors diffusely infiltrate into the normal brain parenchyma. This is not observed in metastases. Infiltrated non-enhancing/non-T_2 hyperintense areas, with abnormal Cho/NAA ratios were reported for gliomas [46]. As such, Fan et al. used single-voxel MR spectroscopy to analyze MR spectra in these regions and found that Cho was elevated in both the tumoral and the surrounding region for gliomas. Also, elevated mI levels were reported within the enhancing foci of gliomas [56].

Law et al. found an elevated Cho/Cr ratio in the enhancing tumor and the peri-enhanced region in glioblastomas compared to metastases using MRSI. No differences were obtained for the NAA/Cr ratio between both tumor classes. The conclusion of that study was that MRSI can be used to distinguish between glioblastomas and metastases [122]. In addition, Chiang et al. also reported a significantly higher Cho/Cr ratio in the peri-enhancing area of glioblastomas compared to metastases, which allowed a differential diagnosis between the two tumor types. On the contrary, no statistical differences were found in the enhancing tumor region for the Cho/Cr ratio [31].

This chapter further investigates the differentiation between glioblastomas and metastases using MRSI and MRI information. More specifically, two-dimensional turbo spectroscopic imaging (2D-TSI) data, T_1-weighted MR images before and after Gd administration and T_2-weighted MR images are used. 2D-TSI data are characterized by reduced acquisition times compared to those required for standard MRSI [51]. The Cho/Cr and Cho/NAA ratios for glioblastomas and metastases are compared in the enhancing tumor, the peri-enhancing area and the non-enhancing/non-T_2 hyperintense region. The infiltration of glioblastomas in the non-enhancing/non-T_2 hyperintense area is detected. The information is included in a nosologic image, which illustrates the ability to generate more realistic abnormal tissue areas compared to previous approaches.

9.2 Materials

Data from 15 patients with a brain tumor are considered, before treatment, for this study (9 men, 6 women; patient age 46−72; mean age 60.1 years). Nine patients were diagnosed with a glioblastoma tumor and six with a solitary metastasis tumor (primary tumor location: colon, two lung, breast, and two unknown), all confirmed by histopathology of surgical resection specimens or biopsy (Table 9.1). Data were acquired at the Radiology Service of Clinica Quirón, Valencia, Spain, and Hospital de La Ribera, Alzira, Spain, as part of the routine clinical preoperative MR imaging and spectroscopy protocol. All patients signed informed consent, in agreement with the guidelines of the University Hospital Ethical Committee.

Table 9.1: *Pathology (primary tumor location), age and sex for 15 patients.*

Case	Pathology (primary tumor location)	Age	Sex
1	glioblastoma multiforme (·)	71	female
2	glioblastoma multiforme (·)	64	male
3	glioblastoma multiforme (·)	64	male
4	glioblastoma multiforme (·)	58	male
5	glioblastoma multiforme (·)	62	male
6	glioblastoma multiforme (·)	67	male
7	glioblastoma multiforme (·)	60	female
8	glioblastoma multiforme (·)	72	male
9	glioblastoma multiforme (·)	71	female
10	metastasis (unknown)	50	male
11	metastasis (colon)	46	male
12	metastasis (lung)	51	male
13	metastasis (breast)	54	female
14	metastasis (unknown)	56	female
15	metastasis (lung)	52	female

Data were acquired at 1.5 T by Philips Achieva scanners. T_1-weighted MR images (slice thickness = 5 mm, matrix size = 256 × 256, TR = 500 ms, TE = 20 ms) before and after Gd injection were acquired prior to the 2D-TSI study. A T_2-weighted MR image package (TR = 4858 ms, TE = 110 ms), with the same FOV and orientation in plane location of the 2D-TSI slice, was acquired before 2D-TSI. Image registration is performed based on the algorithm described in [146]. 2D-TSI (including water-suppressed and water-unsuppressed MR spectra) were acquired using a $90° - 180° - 180°$ pulse sequence PRESS. The multi-voxel spectroscopic images consist of a grid of 24 × 24 voxels with a FOV of 230 × 230 mm and a slice thickness of 20 mm. Each voxel is 9.6 × 9.6 × 20 mm (1.8 ml). 2D-TSI data were acquired with long TE (i.e. 272 ms) and TR = 2000 ms. 256 complex data points

were acquired and zero filled to 512 points. Spectral processing is performed with commercial Philips software. Data processing includes eddy current correction, frequency alignment and removal of the residual water resonance by HLSVD-PRO. Water-unsuppressed MR spectra are used for phase correction, chemical shift correction and correction of field heterogeneity. Magnitude MR spectra are computed to avoid phasing difficulties and peak integration is used to calculate an estimate of the concentration of Cho ([3.1,3.3] ppm), Cr ([2.9,3.1] ppm) and NAA ([1.9,2.1] ppm).

Cho/Cr and Cho/NAA ratios are computed by dividing the metabolite peak integrals in the same MR spectrum. In addition for each patient, the Cho/Cr and Cho/NAA ratios belonging to the voxels of the enhancing tumor, peri-enhancing area (i.e. hyperintense on T_2-weighted MR images, but not enhancing on post-Gd T_1-weighted MR images) and the non-enhancing/non-T_2 hyperintense region are averaged within each of those regions. These areas are manually defined by experienced radiologists. Afterwards, the ratio values are compared in order to differentiate the types of tumor under investigation.

9.3 Methods

9.3.1 Threshold method

The purpose of the threshold method is to stratify the 2D-TSI voxels in two categories: affected and unaffected. The automatic method by Otsu is applied to the Cho/NAA map of each patient in order to calculate an optimal threshold, such that voxels with abnormal ratios become separated from voxels with normal ratios [168]. This approach requires both affected, as well as, unaffected voxels inside the rectangular volume. The goal of the method is to find a threshold that minimizes the within-class variance, which has been shown to be equivalent to maximizing the between-class variance. Using σ_T^2 for the total variance and σ_B^2 for the between-class variance, the following function is maximized

$$t^* = \arg\max_t \frac{\sigma_B^2(t)}{\sigma_T^2}, \tag{9.1}$$

where t^* is the optimal threshold. Once t^* is fixed, the final output of the threshold method is a binary map, indicating voxels with normal and abnormal Cho/NAA ratios. Otsu's threshold selection is an efficient method due to its recursive nature.

The binary map is used to visualize infiltration in nosologic images.

9.3.2 Nosologic images

A nosologic image summarizes the presence of different tissues and lesions in a single image by color coding each voxel according to the tissue type it is assigned to [216]. In Chapter 8, a nosologic imaging technique was proposed combining methods from image processing as well as pattern recognition to segment and classify brain tumors based on MRI and MRSI data. Firstly, a registered digital brain atlas and a subject-specific abnormal tissue prior, obtained from MRSI data, were used for the segmentation of the MRI data. Next, the detected abnormal tissue was classified based on supervised pattern recognition of the spectroscopic data. In this chapter, nosologic images are also generated, but only the output from the first step (i.e. segmentation of MRI) of the nosologic imaging method is used. As a result, the segmentation step produces a segmented MR image that separates abnormal tissue from gray matter, white matter and CSF.

In the present study, the segmentation step is applied to the T_2-weighted MR images such that the segmented abnormal region matches with the T_2 hyperintense region. Next, the binary map, obtained by applying Otsu's threshold method to the Cho/NAA ratio, estimated from the 2D-TSI data, is superimposed to the segmented abnormal region. Abnormal voxels of the binary map that fall outside the lesion, which is obtained by the segmentation step, correspond to infiltration [46].

9.4 Results

MRI and 2D-TSI studies were carried out on 15 patients, 9 patients with glioblastoma and 6 with metastatic disease. The results are presented in three steps. Firstly, distinct differences in metabolite ratios, allowing differentiation, in glioblastomas and metastases are demonstrated. Next, the infiltrative nature of the glioblastoma tumor is detected and, finally, this information is visualized in a nosologic image.

9.4.1 Metabolite ratios

For each patient, the voxels are divided into three groups: voxels assigned/ corresponding to enhancing tumor, peri-enhancing area (i.e. hyperintense on T_2-weighted MR images, non-enhancing on post-Gd T_1-weighted MR images) and the non-enhancing/non-T_2 hyperintense region. Within each group, the Cho/Cr and Cho/NAA ratios are averaged per patient, and secondly, over all patients with the same type of tumor (Table 9.2). The averaged ($\pm$ standard deviation) Cho/Cr ratio in the non-enhancing/non-T_2 hyperintense area of glioblastomas is 1.29 ± 0.22 compared to 1.26 ± 0.17 in metastases. In the peri-enhancing area the ratio is 1.68 ± 0.61 in glioblastomas and 1.14 ± 0.15 in metastases, while ratios equal to 2.76 ± 0.99 and 1.80 ± 0.22 are observed in the enhancing tumor of glioblastomas and metastases, respectively. The averaged ($\pm$ standard deviation) Cho/NAA ratio is 0.43 ± 0.08 in the non-enhancing/non-T_2 hyperintense area of glioblastomas, compared to 0.42 ± 0.08 in metastases. The ratio is 0.83 ± 0.65 in the peri-enhancing area of glioblastomas and 0.51 ± 0.11 in metastases. For the enhancing tumor of glioblastomas the ratio value is 1.81 ± 1.62, while 0.93 ± 0.15 is found for metastases.

A clear difference in Cho/Cr and Cho/NAA is found in the peri-enhancing area and the enhancing tumor between glioblastomas and metastases. Cho/Cr decreases for all metastasis cases when going from the non-enhancing/non-T_2 hyperintense to the peri-enhancing area, while it increases for all glioblastoma cases. In addition, the difference between the averaged ratios from the enhancing tumor and the non-enhancing/non-T_2 hyperintense area is calculated for each patient and a clear separation between metastases and glioblastomas is obtained (Figure 9.1).

9.4.2 Infiltration

The region where the glioblastoma infiltrates normal appearing tissue, is indicated on the MRI. Otsu's threshold selection is applied to the Cho/NAA values of a patient to generate a binary map, indicating voxels with normal and abnormal ratios. The output of the threshold selection is superimposed on the T_2-weighted MR images, included in the 2D-TSI slice, for two patients with glioblastoma (Figures 9.2 $-$ 9.3) and for two patients with metastatic disease (Figures 9.4 $-$ 9.5). Voxels with abnormal Cho/NAA values are indicated in red, while normal ones are indicated in blue.

Case 3 represents a patient with a glioblastoma, confirmed by histopathology, having a satellite lesion (Figure 9.2). Voxels with abnormal Cho/NAA values are found in the T_2 hyperintense region, as well as in adjacent tissue in the

Table 9.2: *Mean and standard deviation of Cho/Cr and Cho/NAA ratio in enhancing tumor, peri-enhancing area and non-enhancing/non-T_2 hyperintense region for glioblastomas (GBM) and metastases (MET). The ratios are higher in the enhancing tumor and peri-enhancing region for glioblastomas compared to metastases.*

		Non-enhancing/non-T_2 hyperintense	Peri-enhancing	Enhancing tumor
Cho/Cr	GBM	1.29±0.22	1.68±0.61	2.76±0.99
	MET	1.26±0.17	1.14±0.15	1.80±0.22
Cho/NAA	GBM	0.43±0.08	0.83±0.65	1.81±1.62
	MET	0.42±0.08	0.51±0.11	0.93±0.15

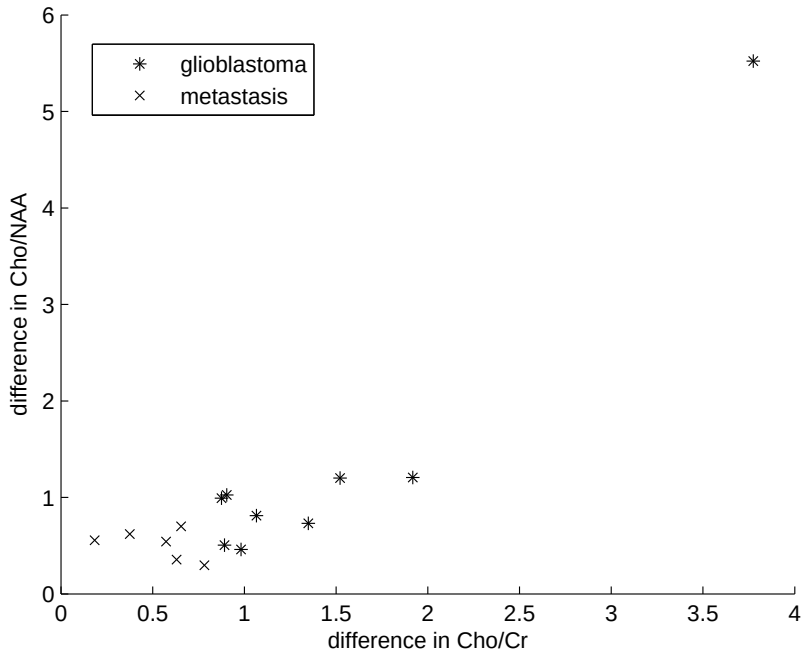


Figure 9.1: *Scatter plot of the difference between the averaged Cho/Cr and Cho/NAA ratio from the enhancing tumor and the non-enhancing/non-T_2 hyperintense area for each patient. Metastases (MET) and glioblastomas (GBM) are separated. The averaged value was 1.48 in glioblastomas and 0.53 in metastases for Cho/Cr. The average was 1.38 in glioblastomas and 0.51 in metastases for Cho/NAA.*

contralateral hemisphere. MR spectrum (a) in Figure 9.2 clearly shows increased Cho compared to NAA and Cr. The voxel is situated outside the T_2 hyperintense region, suggesting that it corresponds to infiltrating tumor. MR spectra from normal tissue (spectrum (b)) and the enhancing tumor area (spectrum (c)) are also provided. Case 4 represents another patient with a glioblastoma tumor (Figure 9.3). The MR spectrum of voxel (b) shows that Cho is increased compared to NAA and Cr. MR spectrum (a), originating from a voxel positioned in the normal contralateral brain and a pure tumor MR spectrum (c) are also depicted.

Case 10 is a patient with metastatic disease, showing no infiltration in the adjacent brain parenchyma based on the 2D-TSI and MRI data (Figure 9.4). The voxels with abnormal Cho/NAA completely correspond to the hyperintense area on

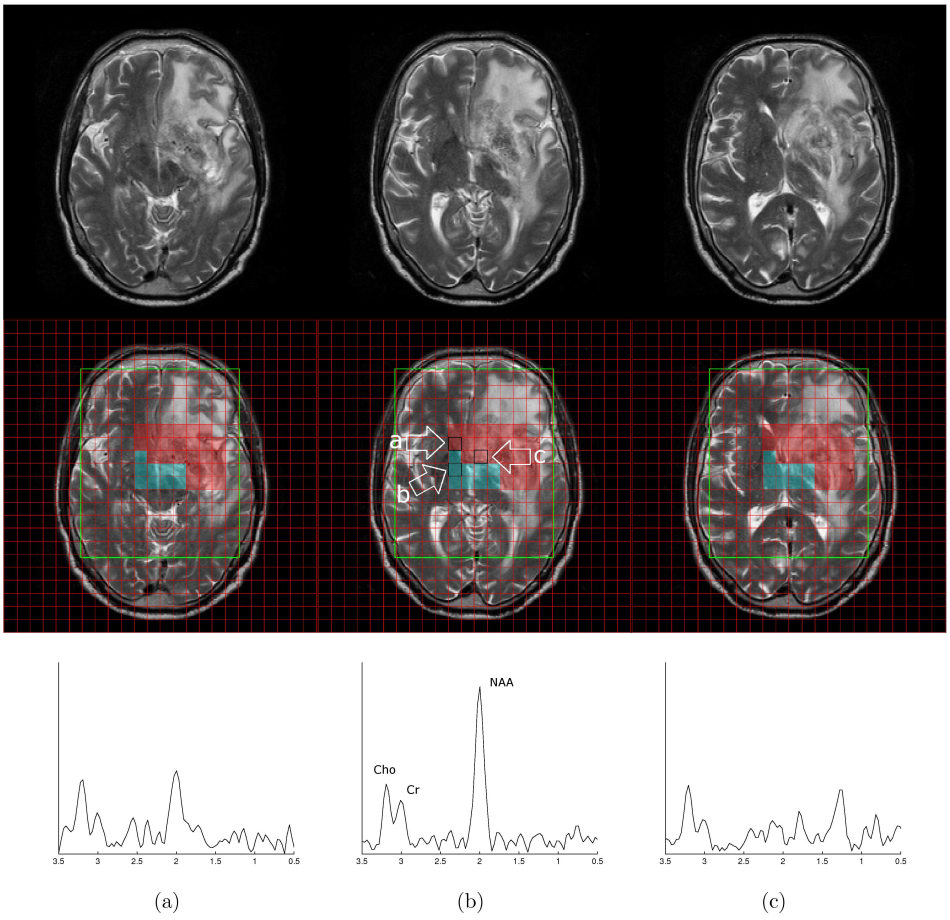


Figure 9.2: *Case 3, patient with a glioblastoma. Top: T_2-weighted MR images of neighboring slices contained in the 2D-TSI box. Middle: T_2-weighted MR images with superimposed Cho/NAA threshold result. Red and blue colors refer to abnormal and normal Cho/NAA levels, respectively. Bottom: MR spectra from voxels that are indicated on the grid of the middle MRI slice. MR spectrum (a) corresponds to a voxel that is situated in the adjacent tissue in the contralateral hemisphere. Cho is increased compared to NAA and Cr, suggesting tumor infiltration. MR spectra from normal (spectrum (b)) and tumor (spectrum (c)) tissue are also represented.*

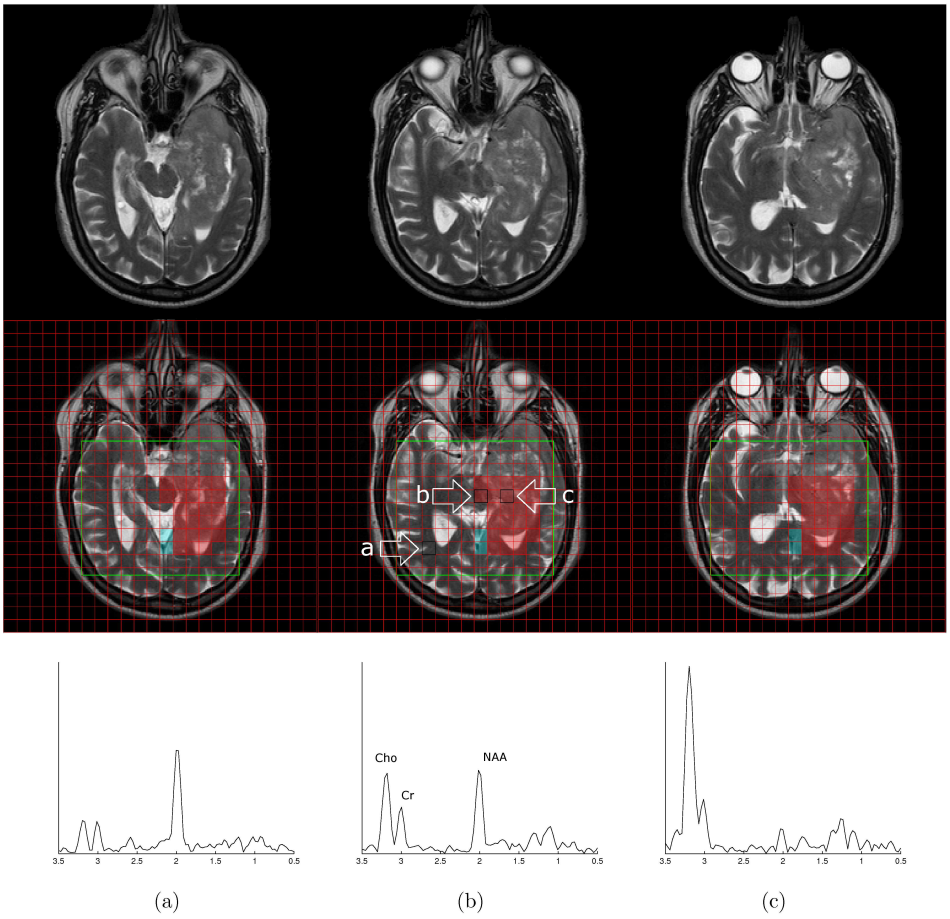


Figure 9.3: *Case 4, patient with a glioblastoma. Top: T_2-weighted MR images of neighboring slices contained in the 2D-TSI box. Middle: T_2-weighted MR images with superimposed Cho/NAA threshold result. Bottom: MR spectra from voxels that are indicated on the grid of the middle MRI slice. MR spectrum (b) shows increased Cho with respect to NAA and Cr, indicating tumor infiltration. MR spectra from the normal contralateral brain (spectrum (a)) and tumor tissue (spectrum (c)) are also depicted.*

the T_2-weighted MR images. MR spectra from different locations are displayed, showing nearly equivalent levels of Cho and Cr and increased NAA. These voxels are located outside the tumoral lesion, in peri-enhancing edema or in normal brain parenchyma. For case 14, having a metastasis, all voxels with abnormal Cho/NAA levels are encapsulated in the hyperintense area on T_2-weighted MR images (Figure 9.5). MR spectra show similar Cho and Cr levels, with elevated NAA.

9.4.3 Nosologic images

The nosologic image for case 1 illustrates the infiltrative nature of a glioblastoma, with a periependimary extension (Figure 9.6). The abnormal tissue area, as obtained from the T_2-weighted MRI, is labeled in red in the nosologic image, while voxels with abnormal Cho/NAA levels that fall outside this area are indicated in white. MR spectra (a) and (b), belonging to these white voxels in the non-enhancing/non-T_2 hyperintense area at the top, are shown (Figure 9.6). Cho is increased with respect to NAA and Cr, suggesting tumor infiltration. Case 2 represents a glioblastoma patient, having satellite lesions (Figure 9.7). The significantly broader white area, as compared to the red region, in the nosologic image suggested tumor infiltration. Green reflects CSF, light blue white matter and dark blue gray matter. MR spectrum (b), situated in the infiltrated area is visualized in Figure 9.7, together with MR spectra from the (enhancing) tumor (spectrum (a)) and normal tissue (spectrum (c)).

Case 13 is a patient with a metastasis tumor (Figure 9.8). The red region, obtained from T_2-weighted MR image segmentation, shows good agreement with the voxels that had abnormal Cho/NAA. MR spectra from different locations at the tumor border show similar Cho and Cr peaks, while NAA is strongly increased. Finally, the segmented abnormal region, obtained from the T_2-weighted MRI, matches with the voxels that have abnormal Cho/NAA levels for case 12, a metastasis (Figure 9.9). MR spectra, positioned at the tumor border, correspond to normal brain parenchyma, suggesting no tumor infiltration. The remaining white areas correspond to voxels that partially contain abnormal tissue.

9.5 Discussion

MRI is the method of choice for diagnosis, evaluation and follow-up of brain tumors [190]. Glioblastomas and brain metastases usually show as irregularly shaped lesions with ring-like zones of contrast enhancement surrounding a central

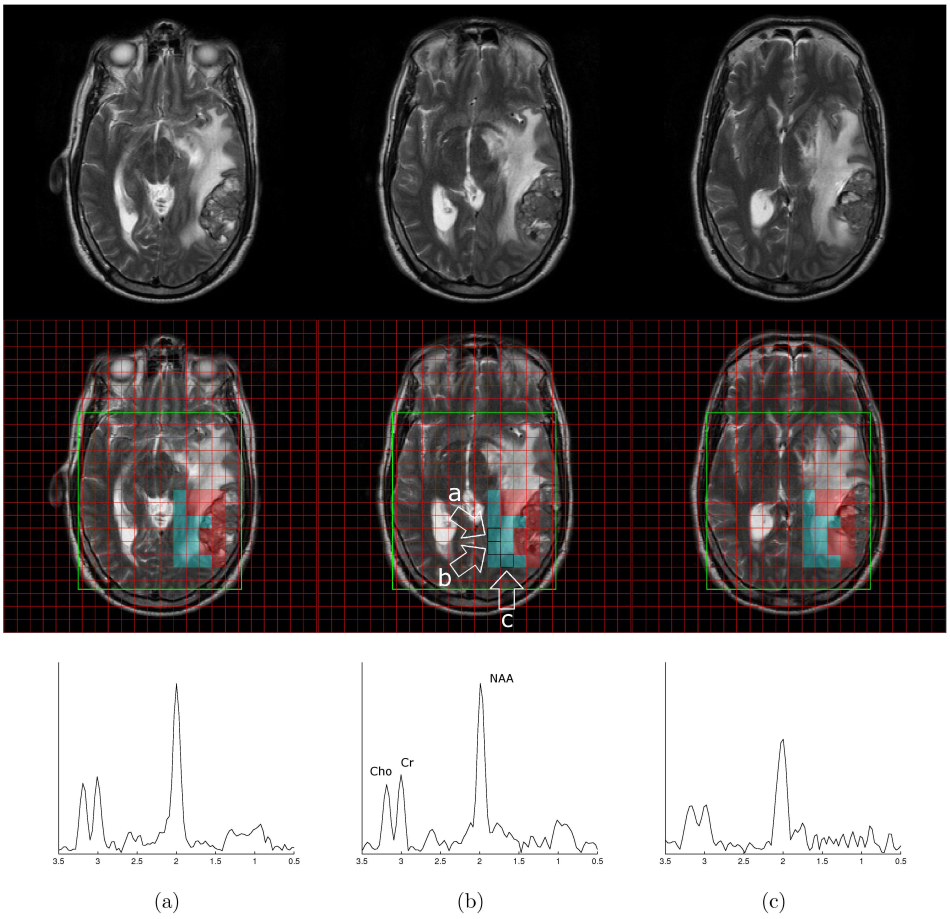


Figure 9.4: *Case 10, patient with a metastatic tumor. Top: T_2-weighted MR images of neighboring slices contained in the 2D-TSI box. Middle: T_2-weighted MR images with superimposed Cho/NAA threshold result. The red voxels completely fall inside the hyperintense area on the T_2-weighted MR images. Bottom: various MR spectra, located outside the tumoral lesion, in peri-enhancing edema or in normal brain parenchyma. Cho and Cr levels are similar, while NAA is increased, indicating no infiltration.*

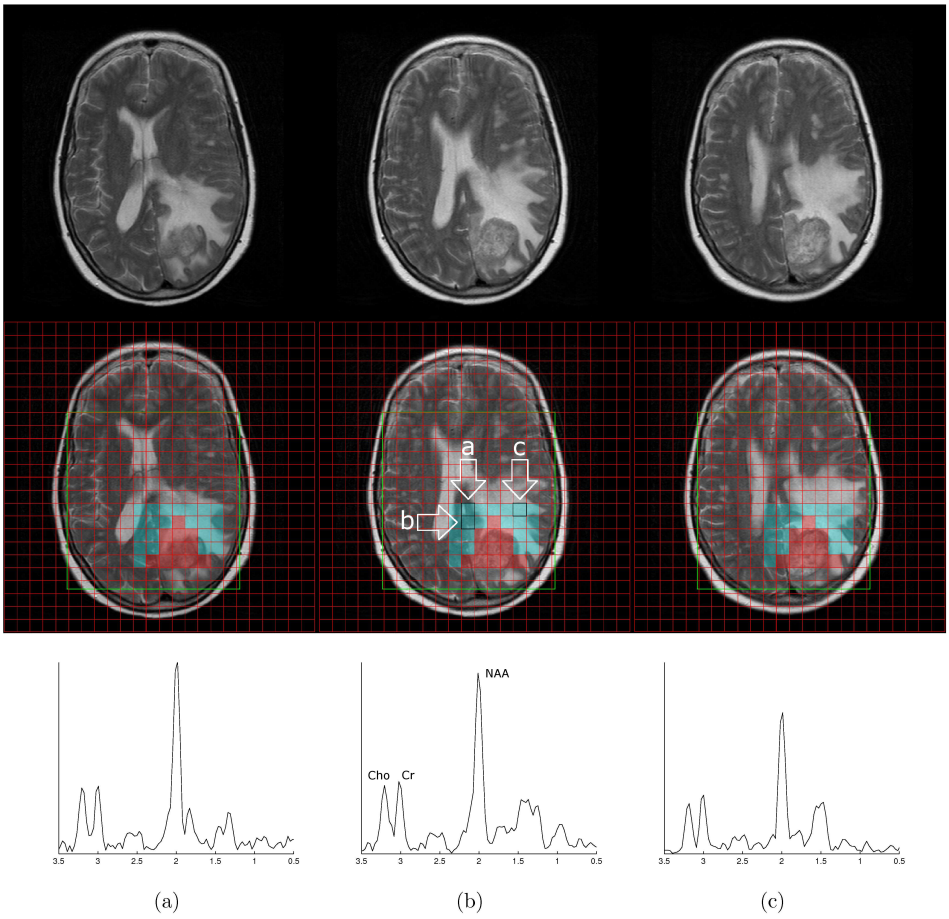


Figure 9.5: *Case 14, patient with a metastatic tumor. Top: T_2-weighted MR images of neighboring slices contained in the 2D-TSI box. Middle: T_2-weighted MR images with superimposed Cho/NAA threshold result. The red voxels completely fall inside the hyperintense area on the T_2-weighted MR images. Bottom: MR spectra showing increased NAA with similar Cho and Cr peaks; no infiltration is detected.*

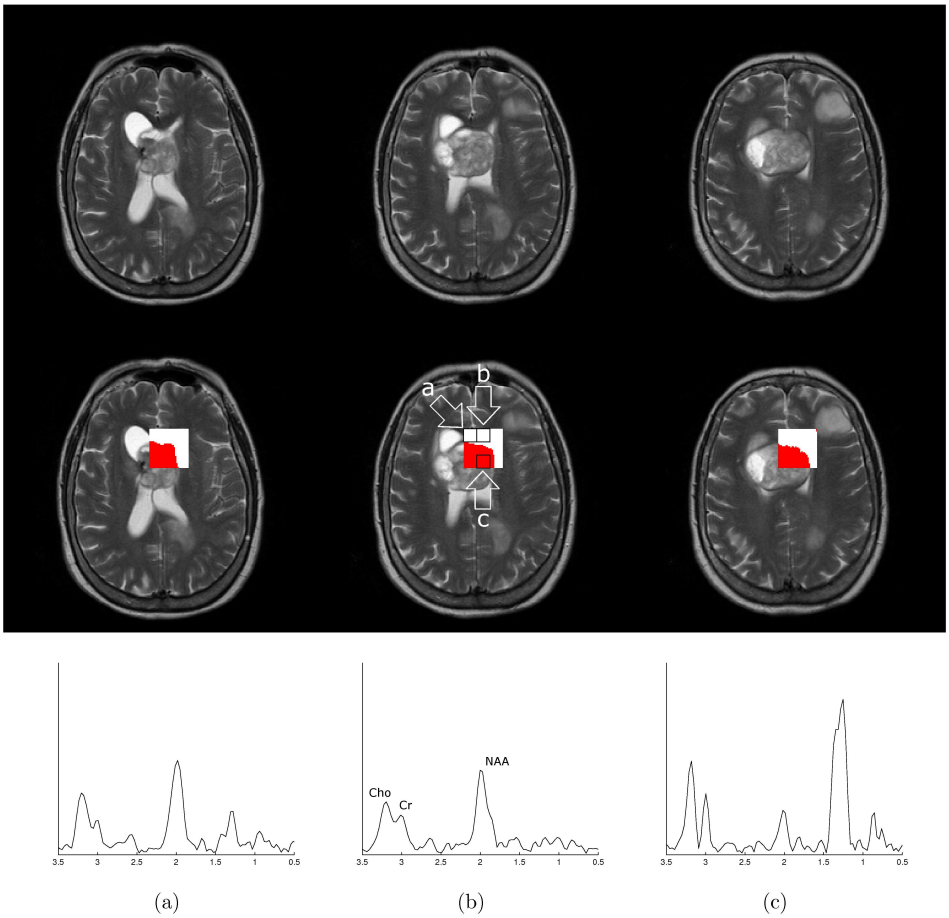


Figure 9.6: *Case 1, patient with a glioblastoma. Top: T_2-weighted MR images of neighboring slices contained in the 2D-TSI box. Middle: T_2-weighted MR images with superimposed nosologic image. Abnormal tissue from MRI is indicated in red. Voxels that contain abnormal Cho/NAA levels and fall outside the abnormal area on T_2, are colored in white. Bottom: MR spectra from voxels that are indicated on the nosologic image of the middle MRI slice. Voxels (a) and (b) are positioned in the white area, suggesting infiltration. The MR spectra show increased Cho with respect to NAA and Cr. Voxel (c) is situated inside the tumor.*

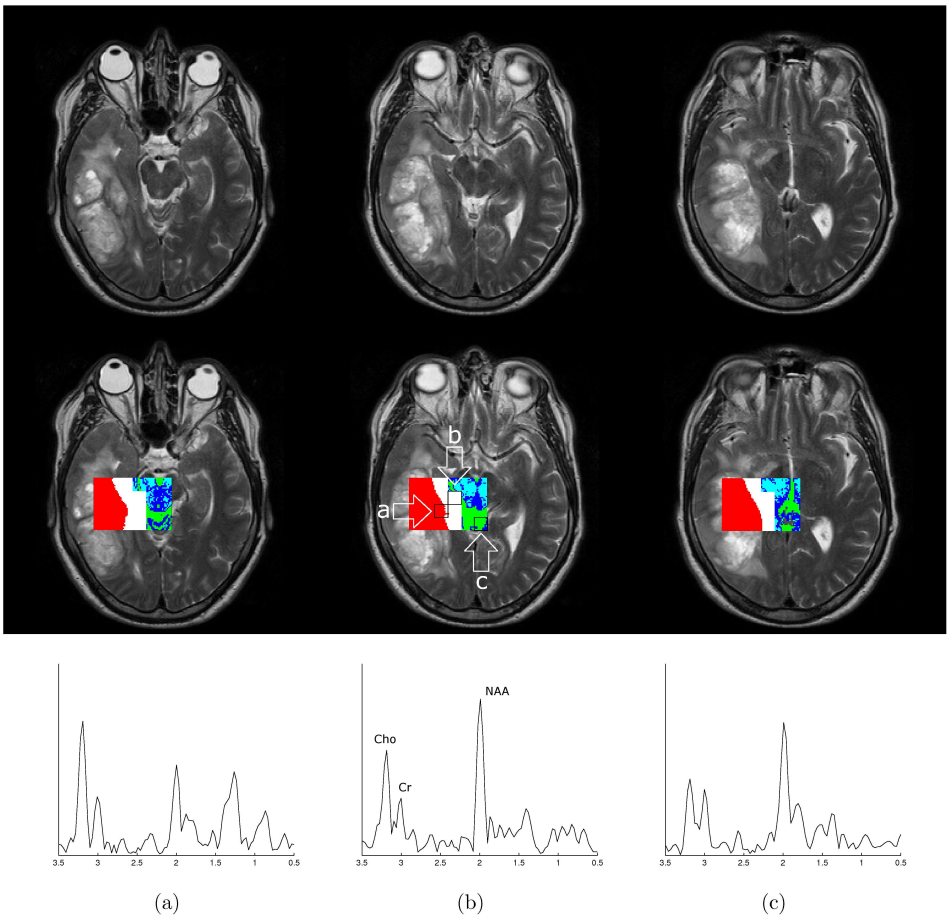


Figure 9.7: *Case 2, patient with a glioblastoma. Top: T_2-weighted MR images of neighboring slices contained in the 2D-TSI box. Middle: T_2-weighted MR images with superimposed nosologic image. Abnormal tissue from MRI is indicated in red, green represents CSF, light blue white matter and dark blue gray matter. Voxels that contain abnormal Cho/NAA levels and fall outside the abnormal area on T_2, are colored in white. Bottom: MR spectra from voxels that are indicated on the nosologic image of the middle MRI slice. MR spectrum (a) originates from tumor tissue while MR spectrum (b) originates from an infiltrated voxel, showing increased Cho with respect to Cr and a higher Cho/NAA ratio. MR spectrum (c) represents normal tissue.*

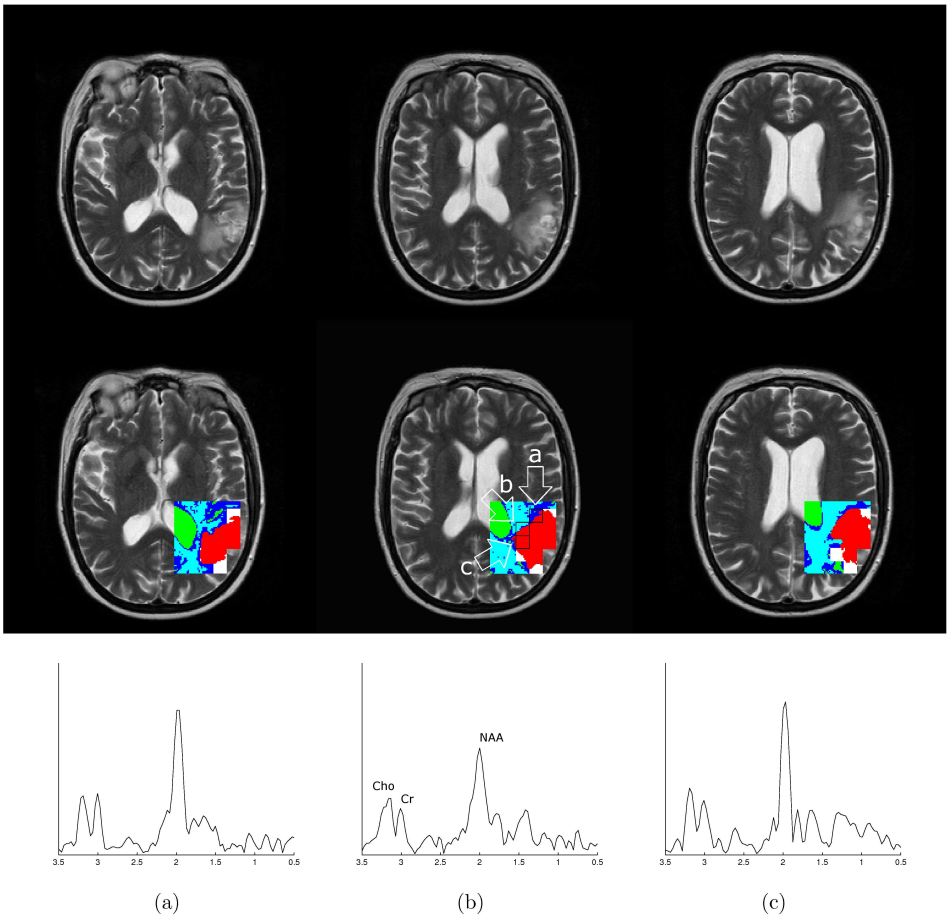


Figure 9.8: *Case 13, patient with a metastatic tumor. Top: T_2-weighted MR images of neighboring slices contained in the 2D-TSI box. Middle: T_2-weighted MR images with superimposed nosologic image. Bottom: MR spectra from voxels, located at the tumor border, are indicated on the nosologic image of the middle MRI slice. No infiltration is detected. Remaining white pixels are due to partial volume effects.*

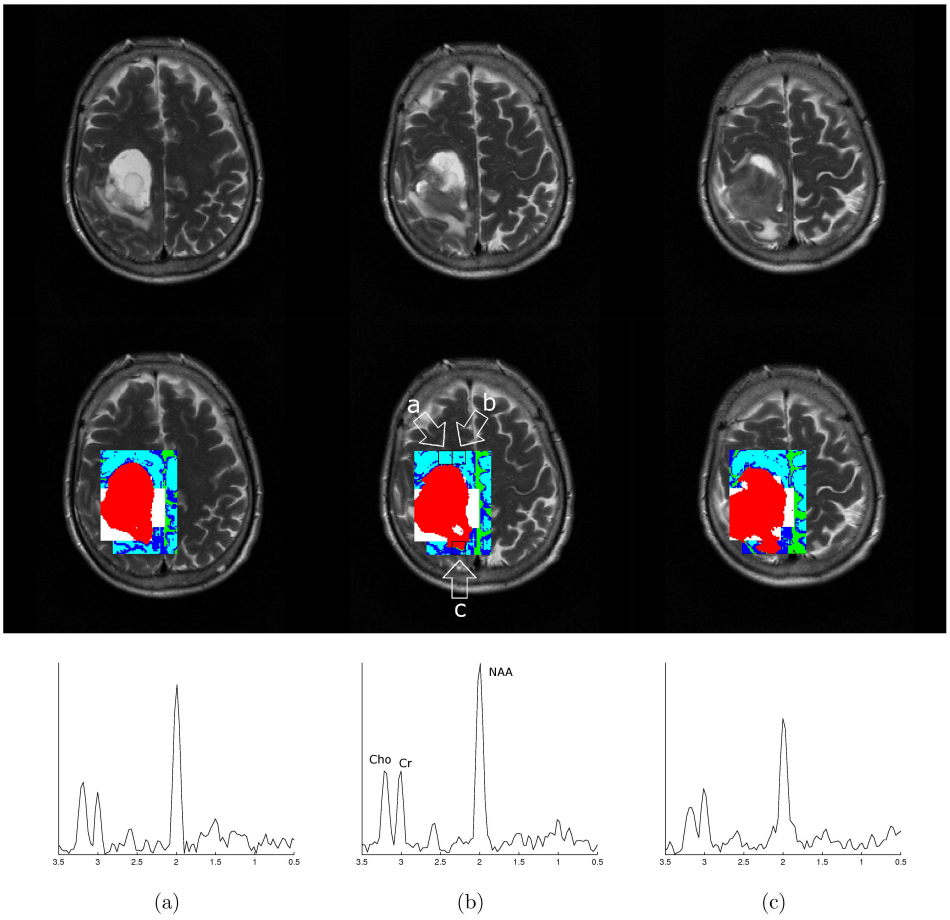


Figure 9.9: *Case 12, patient with a metastatic tumor. Top: T_2-weighted MR images of neighboring slices contained in the 2D-TSI box. Middle: T_2-weighted MR images with superimposed nosologic image. Bottom: MR spectra from voxels, that are indicated on the nosologic image of the middle MRI slice. No infiltration is detected. Remaining white pixels are due to partial volume effects.*

hypointense area of necrosis on Gd-enhanced T_1-weighted MR images [199, 164]. Differentiation between tumor types or between tumor recurrence versus therapy-related changes remains challenging. Based on conventional MRI alone it is often difficult to differentiate between solitary metastasis and glioblastomas. Advanced MRI techniques like MR spectroscopy, diffusion-weighted MRI and perfusion-weighted MRI provide new methods for the assessment of different brain pathologies [245, 206, 238, 85].

Previous studies, differentiating glioblastomas from metastases, applied single-voxel MR spectroscopy, which provides metabolic information from a single location in the brain in a short period of time [166, 56]. In the present study, long TE 2D-TSI data were acquired to provide metabolite maps of the brain. Generally, around 6 minutes are needed to collect a 2D-TSI slice with a 24 $\times$ 24 matrix resolution (turbo factor 3), while 30 minutes are required by standard acquisition sequences with the same spatial resolution. The advantage of this multi-voxel protocol is that the entire tumor, the peri-enhancing region and the non-enhancing/non-T_2 hyperintense brain can be sampled rapidly with relatively high spatial resolution. In this way, the metabolically most active portion of the tumor can be identified with higher likelihood compared to single-voxel MR spectroscopy, where voxel position is essential and often subjective.

Preul et al. averaged long TE (272 ms) 1.5 T multi-voxel MR spectroscopic data within the tumor region and found an elevated Cho level in glioblastomas (24 cases) as compared to metastases (16 cases) [183]. Law et al. compared the maximal Cho/Cr ratio for high-grade gliomas (11 cases) and metastases (6 cases) in the tumoral, peri-tumoral (hyperintense on T_2-weighted MRI, but non-enhancing on post-Gd T_1-weighted MRI) and normal contralateral brain parenchyma using long TE (135 ms) 1.5 T MRSI [122]. The Cho/Cr ratio was 0.96$\pm$0.19 and 0.86$\pm$0.19 in the normal contralateral brain parenchyma, 2.28$\pm$1.24 and 0.76$\pm$0.23 in the peri-tumoral area, and 3.92$\pm$3.31 and 1.84$\pm$1.22 in the enhancing tumor for glioblastomas and metastases, respectively. Law et al. reported significant differences for Cho/Cr in the enhancing tumor and peri-enhancing area. Fan et al. used short TE (45 ms) single-voxel MR spectroscopy (2.0 T) to compare maximal Cho/Cr in similar regions in high-grade gliomas (14 cases of which 6 were glioblastoma) and metastases (8 cases) [56]. Cho/Cr was 0.91$\pm$0.17 and 0.87$\pm$0.24 in the normal contralateral brain parenchyma, 2.65$\pm$0.32 and 1.22$\pm$0.38 in the peri-tumoral area, and 4.98$\pm$1.46 and 1.37$\pm$0.92 in the enhancing tumor for glioblastomas and metastases, respectively. Significant differences were found for Cho/Cr in the enhancing tumor and the peri-tumoral region. Finally, Chiang et al. compared the maximal Cho/Cr ratio using 3 T long TE (270 ms) MRSI between high-grade gliomas (14 cases) and metastases (12 cases) and only found a significant difference in the peri-tumoral area; 0.29$\pm$0.51 in metastases and

1.30±0.45 in glioblastomas [31]. The ratio in the tumoral region was 1.55±0.74 and 1.82±0.66 for metastases and glioblastomas, respectively. An overview is provided (Table 9.3).

This study analyzes the Cho/Cr and Cho/NAA ratios in the enhancing tumor, the peri-enhancing area and the non-enhancing/non-T_2 hyperintense region. The term peri-enhancing is preferred over peri-tumoral, since the margin of a high-grade glioma is not defined by a T_2-weighted MRI because of the highly infiltrative nature of glioma [122, 126]. Ratios from the voxels are averaged within each of the regions. Significant differences in averaged Cho/Cr and averaged Cho/NAA are found in the enhancing tumor and peri-enhancing area between glioblastomas and metastases. The combination of both ratios can be used to extract a decision rule. Although a limited number of patients is used in the analysis, these findings are consistent with what was already published in the literature, supporting the conclusions. In contrast to the previous approaches, this study also focuses on the visualization of tumor infiltration.

Although the comparison between metabolite ratios in the tumor and peri-enhancing region allows distinct differentiation between glioblastomas and metastases, it does not directly show where the potential infiltration is situated. In this chapter, infiltration is automatically detected and visualized using colored maps and nosologic images. A nosologic image offers an interesting way to represent tumor heterogeneity, which may help clinicians in making therapeutic decisions. However, a limitation of the nosologic imaging approach in Chapter 8 is that only MR spectroscopic information within the MRI-based lesions is considered. As such, the resulting nosologic image shows a tumor contour that is similar to the one that can be observed on the MR image. This is not what can be expected in case of diffuse infiltrative tumors (e.g. glioma [126]). In the present chapter, the colored maps, calculated from Cho/NAA, are superimposed to the nosologic images to visualize infiltration since Cho/NAA and NAA are known to be important markers for tumor infiltration in gliomas [69]. MR spectra with an abnormal Cho/NAA ratio were found in the peri-enhancing and non-enhancing/non-T_2 hyperintense region of gliomas, and, in addition, a negative correlation between the NAA level and the degree of tumor infiltration was reported [35, 211, 46]. The peri-enhancing region represents an increase in interstitial water due to alterations in capillary permeability and blood-brain barrier breakdown, caused by the tumor [122]. In case of high-grade gliomas, this region contains tumor cell infiltration along vascular channels. In metastases, signal abnormality in the peri-enhancing region is likely due to pure vasogenic edema. Due to partial volume effects, though, some white pixels are obtained in nosologic images of metastases. These white pixels originate from voxels that, partially, contain enhancing tumor or peri-enhancing tissue on MRI. However, in contrast to all glioblastomas, no white, infiltrated

Table 9.3: *Literature overview. GBM refers to glioblastoma multiforme, HGG to high-grade glioma and MET to metastasis. The number of cases is indicated between parentheses. FS denotes the field strength of the MR scanner.*

Study	TE(ms)	FS(T)	Metabolite	Pathology(#)	Control	Peri-tumoral	Enhancing tumor
Preul et al.	272	1.5	Cho	GBM(24) MET(16)			increased decreased
Law et al.	135	1.5	Cho/Cr	HGG(11) MET(6)	0.96±0.19 0.86±0.19	2.28±1.24 0.76±0.23	3.92±3.31 1.84±1.22
Fan et al.	45	2.0	Cho/Cr	HGG(14) MET(8)	0.91±0.17 0.87±0.24	2.65±0.32 1.22±0.38	4.98±1.46 1.37±0.92
Chiang et al.	270	3.0	Cho/Cr	HGG(14) MET(12)		1.30±0.45 0.29±0.51	1.82±0.66 1.55±0.74

voxels are found outside the peri-enhancing region for the metastases.

In this study, two-dimensional MR spectroscopic data were acquired with the MRSI box positioned to include maximum tumor volume to differentiate solitary metastases from glioblastoma multiforme. In future studies, the use of three-dimensional MRSI will probably increase, since it enables to cover an even larger part of the tumor [167]. This technology can potentially improve the accuracy of MR spectroscopy to differentiate metastases from glioblastomas as it is based on the analysis of the entire tumor volume. Furthermore, various researchers proposed to use other methods like diffusion-weighted MRI, DTI and perfusion-weighted MRI. The combination of those methods with MRSI may further improve their diagnostic potential [122, 123, 31, 222, 185, 45, 30, 236, 235].

9.6 Conclusions

This chapter illustrated the combined use of 2D-TSI and MRI data to differentiate glioblastoma multiforme from solitary metastases. Firstly, it was shown that the Cho/Cr and Cho/NAA ratios were significantly different for glioblastomas compared to metastases in the enhancing tumor and the peri-enhancing area. The calculated difference between the ratio in the enhancing tumor and the ratio in the non-enhancing/non-T_2 hyperintense region was larger for glioblastomas compared to metastases, such that a differentiation between the two tumor types could be made. The infiltrative nature of glioblastomas into the adjacent brain parenchyma, was also detected. Finally, it was shown that the abnormal tissue area of nosologic images could be improved based on this new approach. The combination of the results, described in this chapter, suggests that MRI and 2D-TSI are able to demonstrate the differences between glioblastomas and metastases, which can potentially lead to further therapeutic optimization.

Chapter 10

HealthAgents: a distributed DSS for brain tumor diagnosis

While previous chapters presented classifiers and classification algorithms to perform brain tumor diagnosis, this chapter explains how these techniques can actually be introduced in clinical practice by means of a distributed DSS. The distributed DSS is established within the European HealthAgents project and includes classifiers that were presented in Chapter 6. Section 10.1 provides an introduction to the use of DSSs in clinical practice. The HealthAgents distributed DSS is presented in Section 10.2 and the classification framework is further discussed in Section 10.3.

10.1 Decision support systems

DSSs were already used in clinical practice in the seventies. Leaper et al. developed a DSS for the diagnosis and surgery of acute abdominal pain [125]. In the early eighties, Shortliffe et al. designed a DSS for assisting clinicians in the treatment of cancer patients receiving chemotherapy [202]. Ever since then, specialized DSSs continuously emerged. Most studies focused on solving diagnostic questions related to specific diseases, such as breast, gastrointestinal, hematologic, oral, lung, bladder, brain, or prostate cancer [131, 248, 136, 61, 187, 224, 157, 34, 210, 109, 110]. These systems mainly implemented clinical solutions using a non-distributed design. Few effort was invested in providing generic, distributed, solutions to standardize the incorporation of predictive models in DSSs [72].

A major contribution of the European INTERPRET project (*International network for pattern recognition of tumours using magnetic resonance*, 2000−2002, http://azizu.uab.es/INTERPRET/ , accessed on June 27, 2009) was the development of a DSS for brain tumor diagnosis. Version 2.0 of the INTERPRET DSS offered brain tumor classification by means of two FLDA classifiers using short or long TE MRS data. More recently, the eTUMOUR project (*Web accessible MR decision support system for brain tumour diagnosis and prognosis, incorporating in vivo and ex vivo genomic and metabolomic data*, 2004−2009, http://www.etumour.net/ , accessed on June 27, 2009) developed a non-distributed DSS for brain tumor diagnosis. Compared to INTERPRET, a larger amount of data sources was considered. The eTUMOUR project created a centralized database with MRI, MRS, MRSI, HR-MAS, microarrays and clinical information. INTERPRET and eTUMOUR provided the conceptual basis for the newly designed HealthAgents distributed DSS (dDSS) [78]. The HealthAgents dDSS offers a distributed system where multiple classifiers, which are developed by state-of-the-art pattern recognition techniques using various kinds of data, can be accessed from clinical centers to answer clinical questions in a transparent way. This knowledge-sharing environment allows to continuously improve the capabilities of the dDSS.

10.2 HealthAgents

The major aim of the HealthAgents project is the development of a web-based dDSS which employs metabolic and genomic profiles. The system deploys an agent-based architecture in order to provide a distributed, dynamic, tool for brain tumor diagnosis, implementing a series of automated classifiers based on pattern recognition methodologies. The distributed nature of the system and the protocols for clinical data exchange allow users to preserve their local center policies for sharing information, whilst allowing them to benefit from the use of a distributed data warehouse. Data are stored anonymously and securely through a network of data marts, resulting in a distributed data warehouse. This incipient network grants bona-fide access to qualified organizations in return for its contribution of clinical data to the dDDS. The multi-node HealthAgents architectural implementation is provided in Figure 10.1.

10.2.1 Agent technology

Several modern complex distributed systems are composed of customizable building blocks, known as software agents or, simply, agents. Literature reports

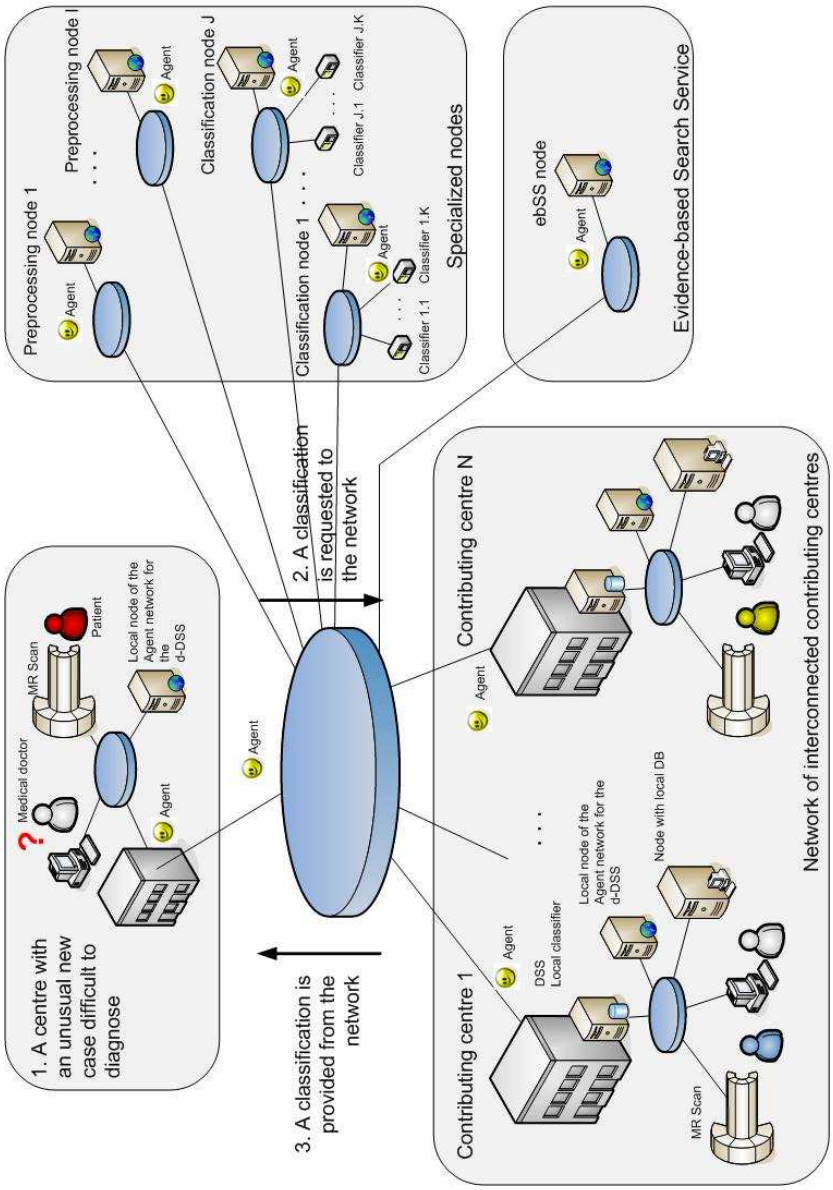


Figure 10.1: *The HealthAgents system architecture.*

four important characteristics of agent technology [22]. Firstly, agents possess an internal knowledge-based state that can be dynamically altered. Secondly, agents have dynamic reasoning capabilities that determine their internal behavior through constraints or goals. Thirdly, they sustain a communication status that enables them to interact with other agents or human entities. Finally, they feature a unique identity that provides roaming and service advertising capabilities. Software agents offer a paradigm for the design and development of complex distributed systems, in which components need to communicate and reason about the information they exchange.

The HealthAgents scenario is one in which distributed data marts are being built by widespread clinical centers in various countries. Typically, individual centers are not able to collect sufficient cases of a particular tumor type, in order to obtain large training data sets for classifier construction. For example, craniopharyngioma is a rare brain tumor of uncertain behavior that occurs at a rate of 1.3 per million person years [24]. A key aim of the HealthAgents system is that data sharing between different centers constitutes a considerable amount of training data, such that robust classifiers can be built for diagnostic purposes. It is not only possible to build local classifiers based solely on a center's own cases, but also global classifiers based on aggregated cases in the distributed system.

The HealthAgents dDDS, which is conceived as an open-source platform, is implemented in Java using the Java agent development (JADE) framework and is supported under Linux and Windows [13]. JADE offers an application programming interface (API) for the development of agent systems. The HealthAgents network consists of different nodes, with at least one active agent. The preprocessing node involves conversion of the raw data to the required data format for classification. For instance, preprocessing of MR spectra is performed by the data manipulation software (DMS) tool, developed in INTERPRET [219]. The classifier node implements the classification functionality. The evidence-based search service (EBSS) node provides contextualized searches, classification-oriented searches and generates an online literature search. The clinical node furnishes the HealthAgents graphical user interface (GUI) to upload the raw data and visualize the results from classification. Finally, there exists a database node in each data providing center. In order to enable agents in the HealthAgents network to find each other, a directory service called yellow pages is used. The yellow pages agent enables agents to advertise their services such that other agents in the system can communicate with the appropriate services. For instance, a dDSS client agent can find classifier agents, which are able to classify specific data, by consulting the yellow pages. Local and general yellow pages run throughout the network in order to improve the general efficiency of the HealthAgents system and to make the directory service robust in case of network dropouts. A typical

operation of the dDSS consists of uploading the raw data, after which the process of data anonymization is performed. Next, the raw data are sent from the clinical center to the preprocessing agent, and the classifiers are invoked. The result from the classifiers is then sent back to the clinical center [78].

10.2.2 Data entry

Data entry for the HealthAgents system is a web front-end for the management of the HealthAgents data (Figure 10.2). The system allows to upload MRS, MRI, MRSI, HR-MAS, microarrays and clinical information and provides additional functionality compared to a simple data repository interface. The data management system allows for example to visualize MR spectra, view MR images and produce reports. In addition, a security system is implemented in HealthAgents [242]. Since centers and countries may vary in their approach to restrict the mobility of data, control of data access policies is required. Therefore, user specific privileges are defined, such that a certain user is allowed to edit data from a specific center, while that same user only has download privileges for data from another center.

10.2.3 Classification

The classification framework, which constitutes the core of the HealthAgents system, offers decision support for brain tumor diagnosis (Figure 10.3). However, because of its generic design it can be adapted in a straightforward way for problems in other fields of application too. Each classifier node in the framework offers a set of classifiers for specific types of data and specific questions, thereby using various feature extraction techniques and classification algorithms. The system enables to create new classifiers and retrain existing classifiers, which then automatically become available throughout the network. Moreover, this architecture allows a clinical center to simultaneously act as a clinical node, as well as a classifier node.

The HealthAgents dDSS benefits from the classification framework and the distributed design since the system ensures that new improvements in pattern recognition and brain tumor classification can be rapidly integrated, due to the extensibility of the framework. New classifiers can be added to the framework by including their specification files, which are created following the provided extensible markup language (XML) specification (cf. [72]). XML is used as specification for customized data or languages and enables developers to describe

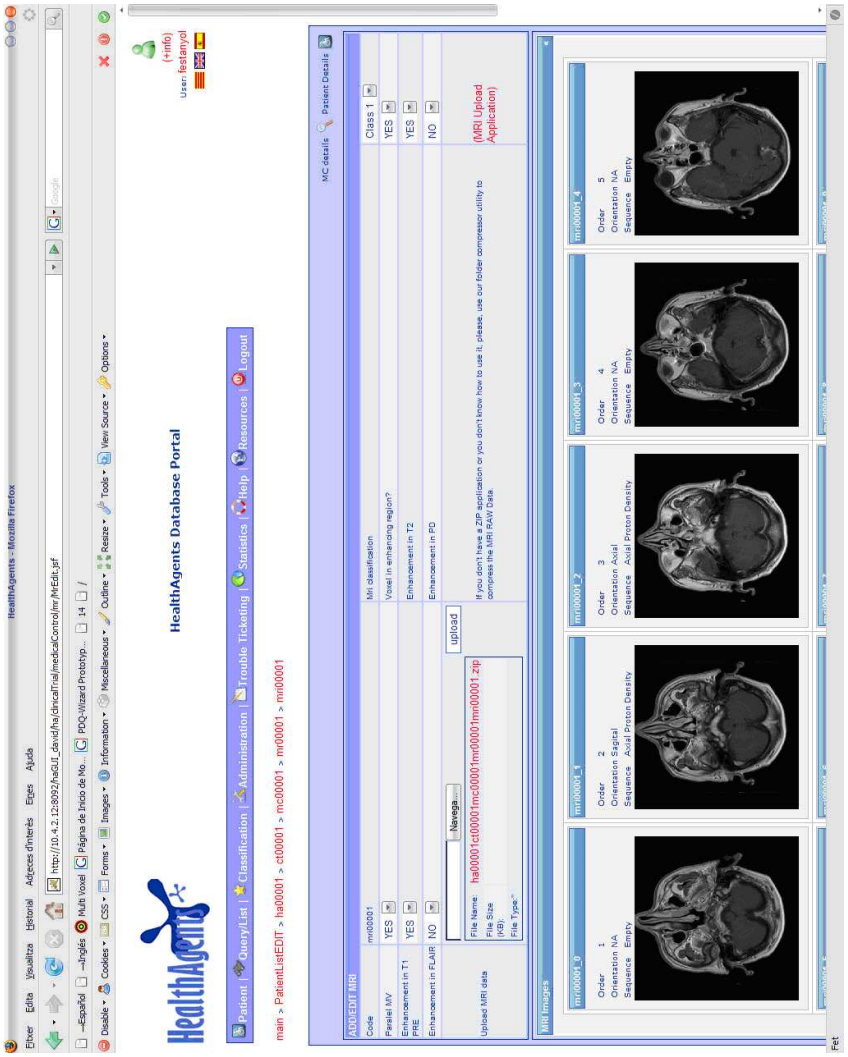


Figure 10.2: *Data entry in the HealthAgents system.*

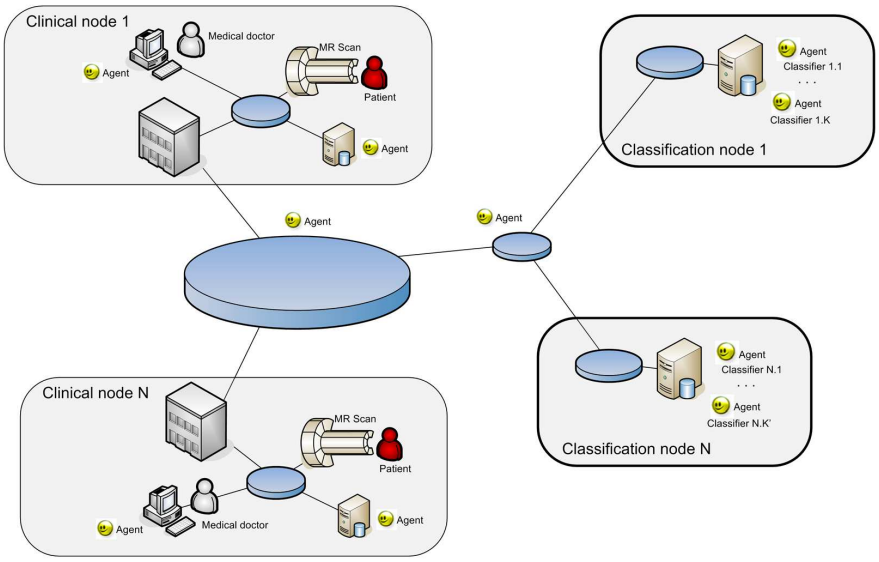


Figure 10.3: *Clinical nodes and classifier nodes within the HealthAgents system.*

standard documents specific to the application. The XML files define the classifier by listing the necessary steps to obtain a classification. The use of the XML standard enables to modify classifier specifications when the framework is improved with new data processing techniques. It also allows to include additional information (e.g. mean MR spectra for brain tumor types) about the field to which a classifier is related. In addition, XML-based documents can be interpreted by many computer tools and applications. This may enable the management of classifiers as well as its sharing with other systems or tools. The classification framework also provides a classifier ranking tool to sort the available classifiers in the network according to a weighted performance measure. The performance of a classifier is weighted with respect to the similarity of the contextual information of the test case and the training data of the classifier. For instance, when comparing two brain tumor classifiers with similar performances, but built from various training data sets having different age distributions, the ranking will give a higher score to the candidate classifier that matches to the age of the test case. This occurs when including data from children in the HealthAgents system. The following section presents a detailed overview of the classification framework and its integration in the HealthAgents dDSS.

10.3 Classification framework

An important requirement of the classification framework is to allow rapid, continuous incorporation of new research advances. For this reason, the software must be scalable, robust and maintainable. Independent components of the system need to be identified and modularized, improving the scalability and reusability of the software. It is essential for this type of software to provide an approach for adding new features, without having to modify the underlying architecture. Object-oriented programming offers such flexibility in software development. Design patterns and, more specifically, object-oriented design patterns provide a set of solutions to problems that commonly occur when developing systems with such requirements [3]. These solutions have been obtained through expertise, have already been tested and provide a guarantee for the developers [68].

10.3.1 Design of the framework

The classification framework represents a software module, or library, which provides classification functionality. While this framework cannot be considered as a stand-alone application, it provides an API that enables its integration into other systems. In parallel, the classification framework's API allows to extend the system with new algorithms. The entry point for client modules requiring classification functionality is the `Classifier` class, which is the main class in the framework. The classification framework separates the specification of classifiers from the implementation of the actual pattern recognition algorithm. A classifier that is added to the system can use any of the algorithms included in the classification framework by means of the XML specification file. This approach enables reuse of algorithms that are already implemented in the system.

Classifiers in the framework consist of different actions, including algorithms for feature extraction, data transformation, and classification itself. Conceptually, these algorithms can be considered as classifier actions, where each algorithm takes an input and provides an output. A classifier can be represented as a pipeline of such actions, which generates the final result. Classifier actions are represented by the `ClassifierAction` abstract class in the classification framework. Incorporation of new classifier actions is straightforward following the strategy design pattern. The `ClassifierAction` abstract class can be extended with new actions (Figure 10.4). Therefore, the developer only needs to implement the method's algorithm.

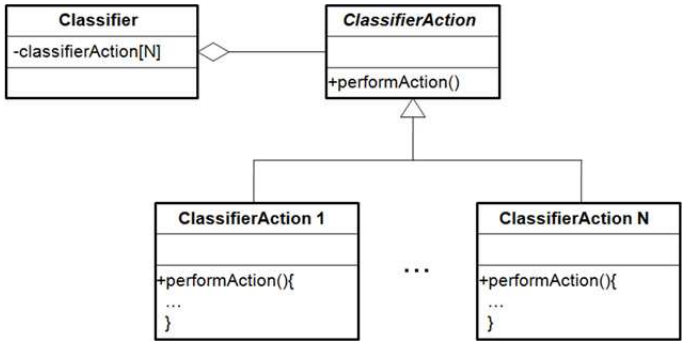


Figure 10.4: *Unified modeling language class diagram of the design of classifier actions in the classification framework. Using the strategy design pattern, classifier action algorithms are encapsulated by the* `performAction()` *method.*

Usually, a data point for pattern recognition consists of an array of features. Most types of medical or biological data can be expressed in this way. Even if classifiers use combinations of different data, a new array can be constructed by combining the two original arrays. The classification framework takes these arrays as input and selects from them the set of features that is required for classification. This approach enables to standardize the input of the framework (Figure 10.5). Because the classification framework is a generic tool, the implemented classification algorithms are able to deal with most kinds of data.

By allowing an open data input format and the possibility of using any classification algorithm, or even a combination of algorithms, the classification framework provides a flexible approach for including classifiers in the system. An XML specification template is defined in the classification framework. To include a classifier in the system, it is only necessary to write the corresponding XML specification and add it to the framework. The XML classifier specification is divided into three separate XML files:

- `description.xml` contains information related to the description of the classifier, such as what kind of input data were used for its construction, the question it solves, where it was developed, and which cases were used for its training. This file provides the necessary information to users such that they can determine whether it is appropriate for their problem or not.
- `profile.xml` contains information related to the classifier's profile: classifier accuracy figures and statistical information about the training data.

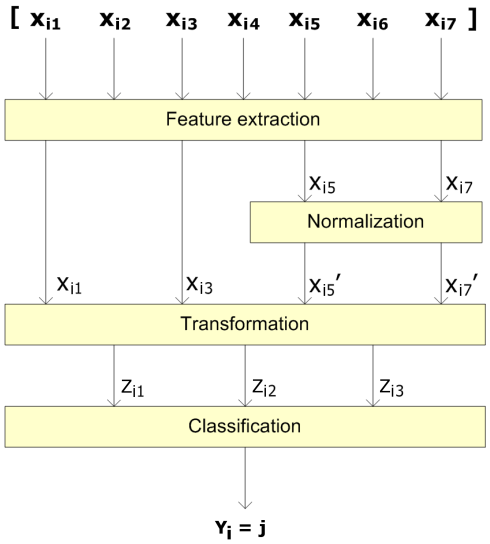


Figure 10.5: *A pipeline of classifier actions to classify an array of features x_i. From the whole set (i.e. x_{i1} to x_{i7}), 4 features are obtained by feature extraction (x_{i1}, x_{i3}, x_{i5} and x_{i7}) and normalization is only applied to x_{i5} and x_{i7}. A transformation is applied to x_{i1} and x_{i3}, and the normalized features x_{i5}' and x_{i7}'. Finally, the output of the transformation is sent to the classification module to obtain the predicted class.*

- `method.xml` contains the specification of the classification pipeline. The XML specification enables the inclusion of the actions, which are required by the classifier. For each classifier action a set of parameters for its execution and the selected input features are defined. Consequently, when a new classifier action is included in the framework, a description of the structure of its XML specification should be provided. The `method.xml` file describes which outputs will be returned by the classifier, and these can be combinations of output from any of the actions. In addition, extra elements such as plots or data visualizations can be included.

A classification scenario using the classification framework consists of a number of steps (Figure 10.6). Firstly, the client module requests the description of the available classifiers to check which fit its needs. Once a valid classifier is selected, it is instantiated by indicating its name. The selected classifier parses the content of its `method.xml` specification and creates an array of classifier actions. Parameters specified in the XML file are loaded for these actions. Now that the classifier is ready to perform a classification, the data point to be classified can be received from the client. Data passes through the pipeline of classifier actions, and generates

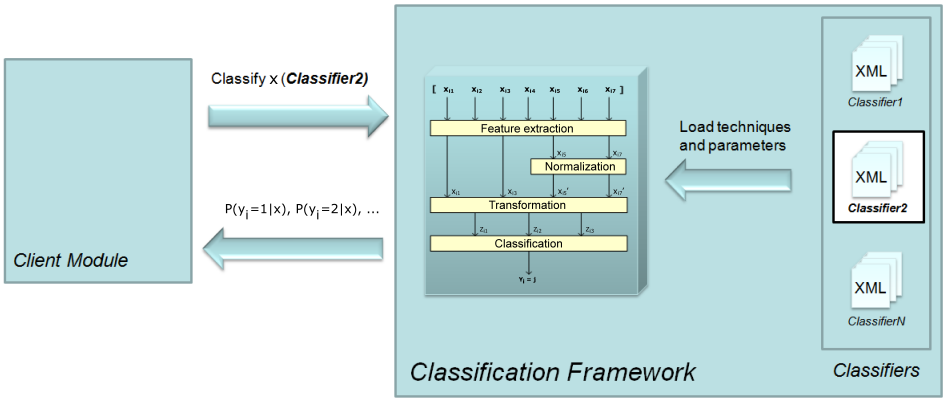


Figure 10.6: *A client module requests the classification of data point x using a classifier called Classifier2. The classifier actions, that are specified in the* `method.xml` *file of Classifier2, are loaded with their corresponding parameters. Class probabilities are sent back to the client module once classification is completed.*

the corresponding outputs at each step. Once the final action is completed, the classifier generates the final output which is sent back to the client module.

The HealthAgents approach results in the existence of multiple classifiers for the same question, each built using different algorithms or trained from different data sets. In addition, due to several factors, such as a poor design, overtraining, or an unrepresentative training set, the level of trust in a classifier may vary over time. Therefore, users may desire a classifier ranking in order to identify which ones are most appropriate for the particular case to be diagnosed. A classifier ranking tool is provided as a module in the classification framework. This tool is also used to rank the obtained results from a set of classifiers. The higher the position in the ranking, the more reliable the result is. In addition, it is also useful when classifiers provide contradictory results, which can happen when a test case is close to the decision boundary. The ranking algorithm is based on a statistical method to compute a score for classifiers. Two types of classifier performance are taken into account, i.e. the static performance and the dynamic performance. The static performance measure is the balanced accuracy rate (BAR) obtained after training, during the evaluation stage of the classifier. It is an initial assessment of the quality of the classifier according to some evaluation, which is usually based on an independent test set, or other techniques such as repeated random sampling or cross-validation (cf. Chapter 6). Dynamic performance is a measure that represents the performance of the classifier over time, based on user validation of previous classifications carried out by the dDSS. Therefore, to

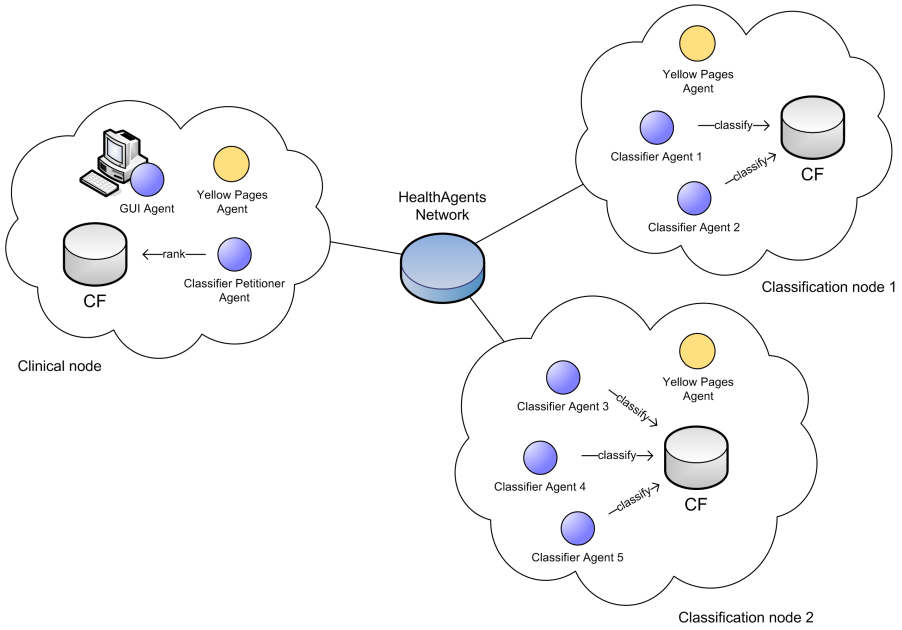


Figure 10.7: *HealthAgents network integrating the classification framework. Classifier agents use the classification framework to perform classifications and the classifier petitioner agent uses it to rank classifiers.*

work properly, feedback about the actual diagnosis of the test cases is required from the clinicians. The classification framework stores the results that classifiers provided in the past. When a case is labeled with the diagnosis obtained from histopathology, the classification framework is notified and, depending on the predictions of the classifiers, the dynamic performance will increase or decrease. The ranking tool also takes into account the appropriateness of a classifier for a case by using its contextual information. The classifier specification defines the distributions of the contextual information for the cases in the training set. In a clinical setting, for example, the patient's age or gender could be used for this purpose. By combining these values, the ranking tool computes a score for each classifier, which is then used for sorting the classifiers.

10.3.2 Integration within HealthAgents

Clinical nodes in the HealthAgents network are able to make secure petitions for classification, which is provided by all available classifiers in the classifier nodes

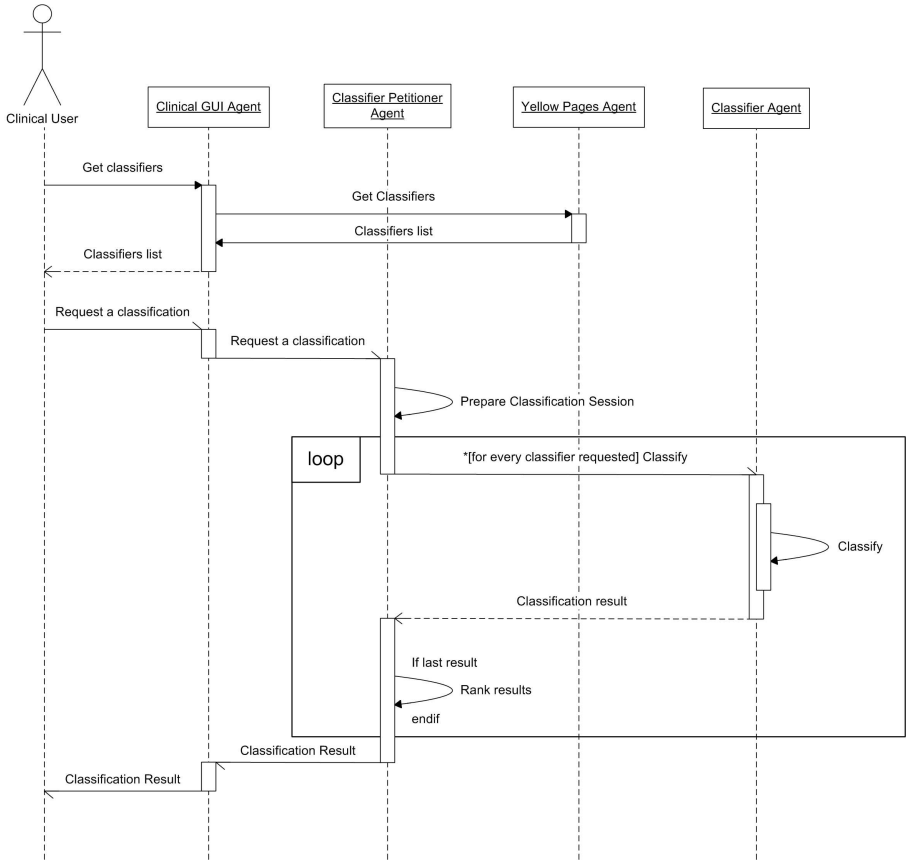


Figure 10.8: *Unified modeling language sequence diagram showing the sequence of messages sent between the agents in order to start classification upon request by a user in a clinical node.*

(Figures 10.7 and 10.8). Two kinds of agents are developed for classification within the HealthAgents system. Firstly, the classifier petitioner agent is responsible for managing classification requests from clinical nodes. Once a clinical node has determined an appropriate list of classifiers, this list and the case to be classified are sent to the classifier petitioner agent. The classifier petitioner agent is designed to handle different classification requests simultaneously because a clinical node might be running several dDSS clients at the same time. When the classifier petitioner agent receives a request, it is resent to every classifier in the list. The results obtained by each classifier arrive asynchronously at the classifier petitioner agent. Therefore, results are mapped to the corresponding classification request on arrival. When the last result is received, the classifier petitioner agent uses the ranking

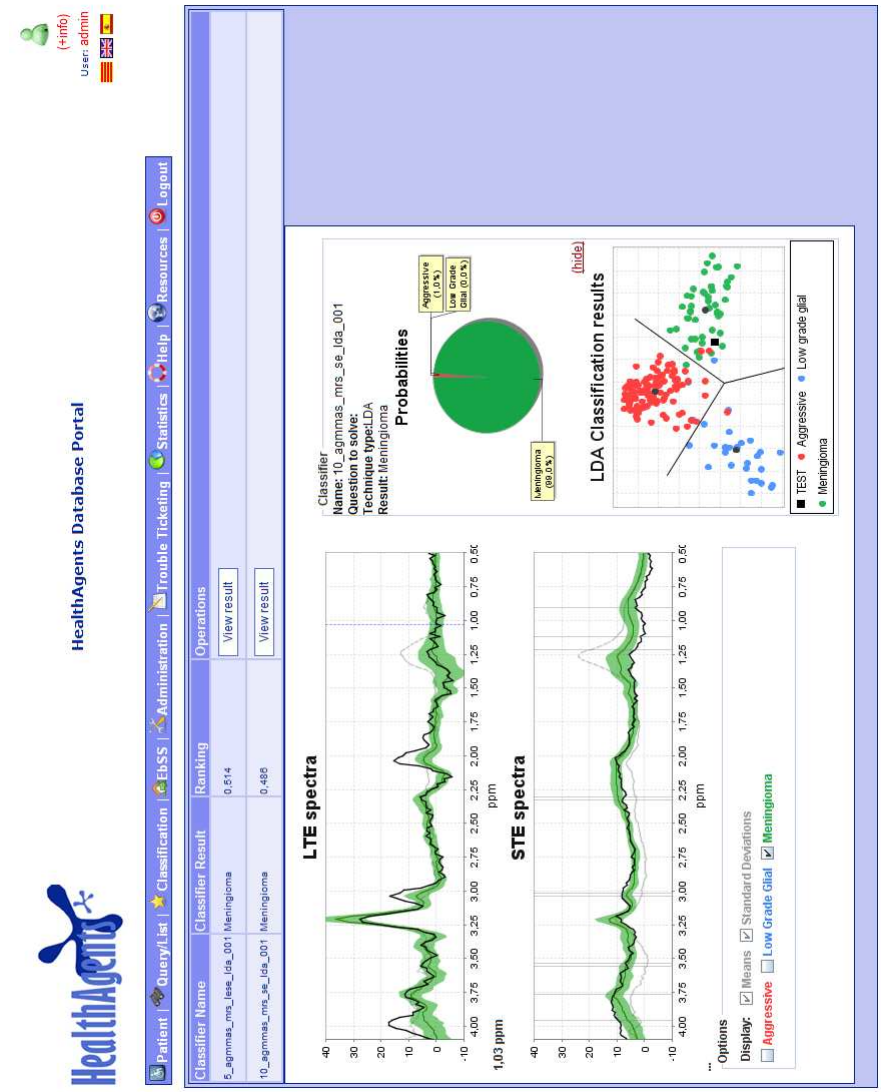


Figure 10.9: *Classification results window of the HealthAgents dDSS.*

tool, provided in the classification framework, to rank the set of classification results. To conclude, the ranked set of results is sent back to the requesting agent on the clinical node. Secondly, the classifier agent represents a classifier in the HealthAgents system. This agent is the link between the classification framework and the HealthAgents dDSS. Classifier agents utilize the API provided by the classification framework to perform classifications and access classifier information. Each classifier agent has an associated classifier in the classification framework. When a classifier agent is started, it loads its classifier description to advertise its main features on the network. This notification is sent to the local yellow pages agent during its registration on the network. An agent requiring service from a classifier agent will be able to find the classifier agent by sending a query to a yellow pages agent. For example, the GUI agent, which runs on a clinical node, can query the nearest yellow pages to obtain the location of all the classifier agents that can classify a particular type of tumor using a particular kind of data. The classifier agent will then receive a message containing the data and classify them using its associated classifier implementation in the classification framework. Once the classifier has provided the result, a reply message is created, containing the classification result, which is send to the original requesting agent.

A web-based GUI provides decision support to clinicians in the HealthAgents system. This GUI is connected to the HealthAgents network through a GUI agent, which is responsible for requesting classifications from the classifier petitioner agent. Only classifiers compatible with the data of clinician's test cases are shown in the GUI. Clinicians decide which classifier they wish to use, but the ranking tool can also provide hints about which classifiers best fit their cases. Once the classification is complete, results are sent to the client and displayed in a results window (Figure 10.9). The clinician can navigate via a tabular form through the results from each classifier. These results are sorted by the ranking tool, such that the results from the most appropriate classifiers are ranked first. By selecting a result in the table, further details can be accessed:

- a first component displays the resulting posterior probabilities in a pie chart together with descriptive information about the classifier.
- a second component enables to visualize the results obtained by the classifier. This component is dependent on the technique that is used for classification. For instance, FLDA classifiers can display the latent space with the training data and the test cases.
- a third component visualizes the means of the MR spectra, and corresponding standard deviations, per tissue class. These are obtained from the training data of the classifier. The test MR spectra can be superimposed to the mean MR spectra, thereby allowing a visual comparison. This

provides additional support when interpreting the results of the classifier. The visualization can also display which features of the spectra are being used for prediction by the classifier. In addition, individual MR spectra from the training set can be superimposed to the test cases if this is permitted by the security policy of the center, contributing the training data.

10.4 Results and evaluation

Presently, as a proof of principle, two classifier nodes and three clinical nodes are running within the HealthAgents system. Clinical nodes are located at Birmingham Children's Hospital, United Kingdom, Universidad de Valencia, Spain, and Universitat Autonoma de Barcelona, Spain. One of the classification nodes is located at Katholieke Universiteit Leuven, Belgium, and another at Universidad Politecnica de Valencia, Spain. The classifier nodes share 25 MRS classifiers for multiple binary and multi-class clinical questions (Tables 10.1 and 10.2). The available classifiers are integrated within the classification framework and use the following pattern recognition algorithms: FLDA, KNN, LS-SVM, PCA and ICA. Classifiers are trained with data acquired during the INTERPRET and HealthAgents projects. Part of them originate from the analysis in Chapter 6. The classifiers are able to detect meningiomas, low-grade gliomas, metastases, unaffected tissue, glioblastomas, embryonal tumors and various combinations of tumor types (e.g. aggressive tumor, gliomas, high-grade tumor, unaggressive tumor). Figures 10.10 and 10.11 provide an experiment to illustrate the distributed use of the classification framework in the HealthAgents dDSS. Two classifiers are included in the classifier node. The tools for monitoring the agents, which are involved in a distributed classification, show messages passing from the GUI agent at a clinical node through the classifier petitioner agent to each classifier agent in the classification node. Finally, the result is received by the GUI agent.

Classifiers can be sorted with respect to an individual test case, by taking into account their contextual information in the ranking tool. As such, the HealthAgents dDSS offers the possibility to classify brain tumors from both adults and children using the provided classifiers (Tables 10.1 and 10.2). Classification results are displayed by means of an interactive GUI, assisting the clinical user in the interpretation (Figure 10.9). The added value of the HealthAgents dDSS was evaluated following the technology acceptance model (TAM) methodology [38]. This evaluation was performed by an HealthAgents partner, the University of Birmingham, United Kingdom. In their study, 26 expert clinicians were interviewed after having used the system. As an overall response, the clinicians experienced the dDSS to be beneficial for improving the quality of brain tumor

Table 10.1: *Eleven MRS brain tumor classifiers integrated into the HealthAgents dDSS. The columns indicate the tumor types the classifier discriminates (AGG: aggressive tumors, MEN: meningiomas, LGG: low-grade gliomas, MET: metastases, HG: high-grade tumors, LG: low-grade tumors, AFF: affected tissue, UNAFF: unaffected tissue, TUM: tumor, NONTUM: non-tumor, NONAGG: nonaggressive tumor, GLM: gliomas, EMB: embryonal tumors, GBM: glioblastomas), whether the classifier is for adults (AD) or children (CH), whether short and/or long TE data are used, the size of the training data set and the distribution of cases per class, the feature extraction method, the classification algorithm, the accuracy (ACC) and standard deviation, BAR and developer of the classifier, respectively.*

Task	Age	TE	Spectra	Feature extraction	Classifier	ACC	BAR	Developer
AGG vs MEN vs LGG	AD	Long + short	125 (71,13,41)	Full spectra	KNN	0.80 (±0.07)	0.78	UPV
AGG vs MEN vs LGG	AD	Long + short	125 (71,13,41)	Full spectra	FLDA	0.75 (±0.07)	0.65	UPV
AGG vs MEN vs LGG	AD	Short	217 (124,58,35)	Full spectra	FLDA	0.94 (±0.03)	0.93	UPV
AGG vs MEN vs LGG	AD	Short	217 (124,58,35)	Peak integration	KNN	0.86 (±0.04)	0.84	UPV
AGG vs MEN vs LGG	AD	Short	217 (124,58,35)	PCA	FLDA	0.91 (±0.04)	0.90	UPV
AGG vs MEN vs LGG	AD	Short	209 (121,55,33)	Peak integration	FLDA	0.90 (±0.04)	0.88	UPV
MEN vs MET	AD	Short	96 (58,38)	PCA	KNN	0.91 (±0.05)	0.91	UPV
MEN vs MET	AD	Short	94 (57,37)	ICA	LS-SVM	0.88 (±0.07)	0.89	KUL
MEN vs MET	AD	Short	94 (57,37)	ICA	LS-SVM	0.90 (±0.06)	0.91	KUL
HG vs LG	AD	Short	237 (134,103)	Peak integration	FLDA	0.87 (±0.05)	0.87	UPV
HG vs LG	AD	Short	262 (147,115)	Peak integration	FLDA	0.85 (±0.04)	0.84	UPV

Table 10.2: *Fourteen MRS brain tumor classifiers integrated into the HealthAgents dDSS, cf. Table 10.1.*

Task	Age	TE	Spectra	Feature extraction	Classifier	ACC	BAR	Developer
HG vs LG	CH	Short	70 (28,42)	PCA	FLDA	0.69 (±0.10)	0.68	UPV
AFF vs UNAFF	AD	Short	304 (22,282)	Peak integration	KNN	0.99 (±0.01)	0.94	UPV
TUM vs NONTUM	AD	Short	282 (274,8)	Peak integration	KNN	0.97 (±0.02)	0.64	UPV
AGG vs NONAGG	AD	Short	244 (135,109)	Peak integration	KNN	0.81 (±0.05)	0.81	UPV
AGG vs NONAGG	CH	Short	79 (30,49)	PCA	KNN	0.72 (±0.10)	0.71	UPV
GLM vs EMB	CH	Short	58 (43,15)	PCA	KNN	0.72 (±0.11)	0.62	UPV
GLM vs EMB	CH	Short	48 (33,15)	PCA	KNN	0.71 (±0.13)	0.66	UPV
GBM vs LGG	AD	Short	117 (84,33)	ICA	LS-SVM	0.83 (±0.07)	0.83	KUL
GBM vs LGG	AD	Short	117 (84,33)	ICA	LS-SVM	0.84 (±0.07)	0.87	KUL
GBM vs MEN	AD	Short	141 (84,57)	ICA	LS-SVM	0.84 (±0.06)	0.79	KUL
GBM vs MEN	AD	Short	141 (84,57)	ICA	LS-SVM	0.91 (±0.07)	0.89	KUL
MEN vs LGG	AD	Short	90 (57,33)	ICA	LS-SVM	0.84 (±0.08)	0.83	KUL
MEN vs LGG	AD	Short	90 (57,33)	ICA	LS-SVM	0.92 (±0.06)	0.92	KUL
MET vs LGG	AD	Short	70 (37,33)	ICA	LS-SVM	0.85 (±0.08)	0.87	KUL

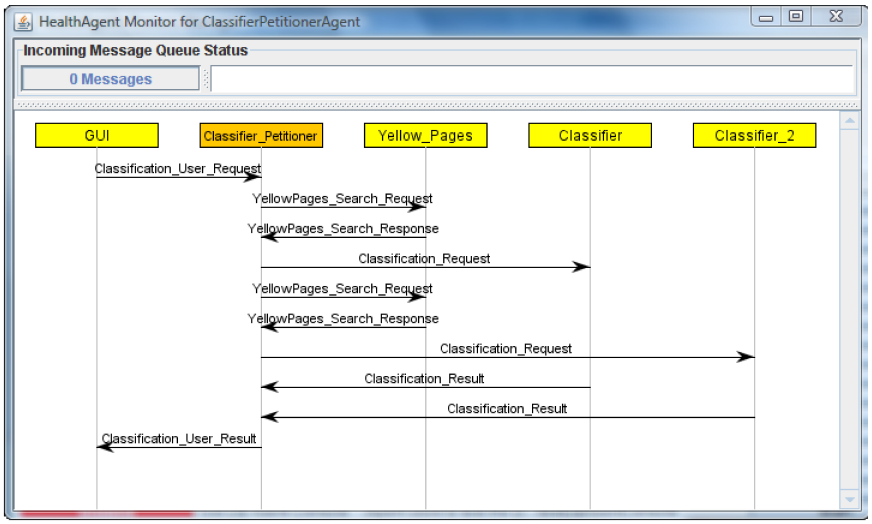


Figure 10.10: *Monitoring of classification performed from a clinical node using two classifiers located in a classification node. The monitor shows the messages sent to/from the classifier petitioner agent in the network. The GUI agent requests a classification to the classifier petitioner agent which asks the yellow pages about the remote classifier agents. The classifier petitioner agent forwards the request to the classifier agents which send their results back. Finally, the classifier petitioner agent collects the results and sends them to the requesting GUI agent.*

diagnosis. The clinicians considered the system to be easy to use, which was an important requirement in the design of the dDSS.

In addition, the performance of the framework was evaluated. Performance is directly related to the hardware the system is running on. A faster host for the classification framework results in a better performance of the classification framework. To obtain a measure of the classification time of the classification framework, a set of tests was carried out on a dual core personal computer running at 2.2 GHz with 2 Gbyte random access memory. Tests consisted of performing 1000 classifications of brain tumor cases while randomly using the 25 developed classifiers. This type of test enabled the execution of various routines within the classification framework and revealed an average of 125 ms per classification task.

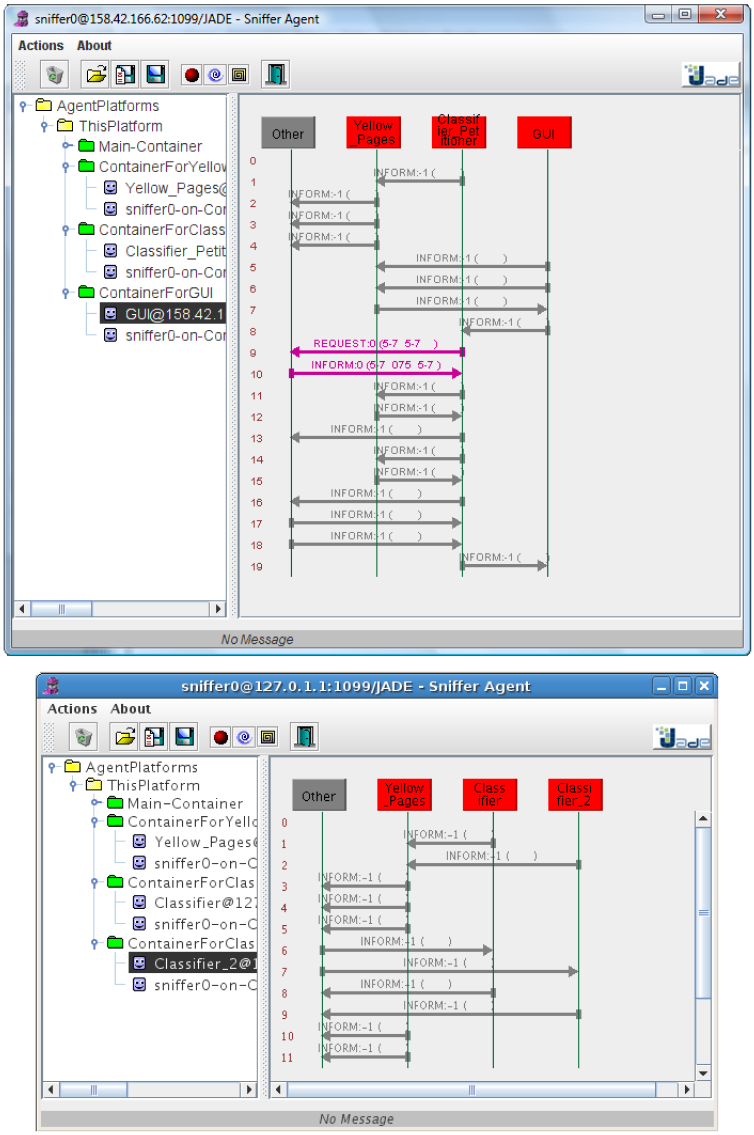


Figure 10.11: *Monitoring of classification performed from a clinical node using two classifiers located in a classification node. Top: JADE sniffer on the clinical node. Bottom: JADE sniffer on the classification node. The JADE sniffer shows the low level agent messages sent between agents on a platform. Messages sent between remote platforms, such as messages sent between the classifier petitioner agent at the clinical node and the classifier agents at the classification node, are displayed on the sniffer as sent or received to/from the other agent.*

10.5 Conclusions

A new framework for providing classification functionality for industrial or research applications was designed, implemented and tested in a real environment. The HealthAgents project offered the opportunity to develop a distributed system for assisting clinicians in the diagnosis of brain tumors using MR data. The system was managed by agent technology and contained databases, which were centered on the users, and classifier nodes. The classification framework, which was the core of the HealthAgents dDSS, illustrated that classifiers, as being developed in the previous chapters, could successfully be transferred and integrated into the real clinical environment. The classification framework offered brain tumor classification on a multidisciplinary research platform. Different centers could create and include their classifiers in the platform and clinical centers could access these classifiers, providing decision support in their day-to-day practice. The framework needed to be scalable and maintainable because of the continuous improvement in pattern recognition algorithms. This scalability was illustrated by incorporating classifiers based on various classification algorithms and developed by different research groups. Additionally, the tests showed the robustness of the classification framework by solving numerous classification problems using various methods. The implementation of the classification framework in the HealthAgents dDSS provides a tool that can improve the process of diagnosing disease, in particular for brain tumors.

Chapter 11

Conclusions and further research

This last chapter summarizes the conclusions of the thesis in Section 11.1 and suggests future research directions in Section 11.2.

11.1 Summary

In Chapter 1 it was explained that noninvasive MR-based techniques are attractive for brain tumor diagnosis. The flexible medical imaging technique MRI enables to assess the anatomical location, morphology and extent of the tumor with a high spatial resolution. Furthermore, MRS provides in vivo information about the molecular composition of tissue in a noninvasive way. Short TEs generate more metabolite information as compared to long TEs, which result in signals that are easier to analyze. MRSI allows to visualize the heterogeneity of the tumor and to observe the more realistic spatial extent of the lesion. Interestingly, these very diverse modalities can be acquired with the same MR scanner. However, analyzing these data often requires a lot of time and expertise which are often not available in clinical centers. Therefore, automated, specialized, processing and analysis techniques are needed. These techniques need to be implemented in DSSs such that clinicians can apply them in a straightforward way.

The machine learning and pattern recognition field develop methods and computer programs to perform certain tasks. Chapter 2 summarized signal preprocessing methods and various techniques, which require example data, to classify the MR

data. These techniques represented the building blocks for the developments in the remaining chapters of the thesis. The algorithms had very different characteristics, strengths and weaknesses. Both linear, as well as, nonlinear methods were included. Some methods generated class probabilities, others only 'black-or-white' decisions. The dimensionality of the data, the issue of binary classification or multi-class classification, and computational efficiency were other important factors to consider.

Chapter 3 reviewed the quantification methods that are popular in the field of MRS(I). These methods aimed to estimate the concentrations of metabolites, which represented interesting features since they provided a direct biological interpretation. The quantification approaches LCModel, AQSES and QUEST were based on mathematical models that imposed prior knowledge by means of a basis set of metabolite profiles. Compared to these approaches, peak integration was a simple and very fast method that did not rely on assumptions about the lineshape of the signal. On the other hand, the feature selection methods from Chapter 4 did not require biological prior knowledge and could be applied in an automated way to any given set of features. Fisher discriminant criterion, the Kruskal-Wallis test, Relief-F and ARD were applied to a set of peak-integrated short TE MRS(I) and MRI features. A significant part of the selected features from MRS and MRSI corresponded to what was known from literature and clinical experience. MRI features were found to provide additional information. Although no biological prior knowledge was imposed, it was concluded that the pattern recognition approach represented a useful tool to obtain relevant and potentially interesting features for differential diagnosis of brain tissue. The introduction of a DSS in clinical practice requires automated procedures since user interaction and processing time should be minimized. Therefore, Chapter 5 focused on the comparison of the feature extraction approaches from the previous chapters, which were part of the preprocessing protocol of short TE MR spectra prior to classification. Quantification and feature extraction methods from machine learning and statistics were compared using MRS and MRSI data. The latter approach, in particular ICA and Relief-F, or full spectra in combination with LS-SVMs for classification produced the highest overall performances. The computationally expensive quantification methods seemed hard to automate in a DSS. In case only classification was needed, the use of full spectra or an automated pattern recognition technique seemed more appropriate. In addition, quantification techniques can provide a biological interpretation if needed. This suggested to restrict the feature extraction step in the preprocessing protocol, decreasing their execution time.

Chapter 6 described a multi-project and multi-center evaluation of automated brain tumor classifiers using short TE MRS data. There exists no previous study

that evaluates classifiers using a new independent test set, that is subsequently acquired from partly different centers. Classifiers were trained with cases acquired by 6 centers during the period 2000−2002 in the INTERPRET project and tested with posterior cases, that were acquired by 8 centers during the period 2004−2007 within the context of the eTUMOUR project. This study design offered a setting that was close to a real environment where similar classifiers, integrated in a DSS, may be used in multiple clinical centers to assist in the diagnosis of new cases. Accurate classification of new cases was feasible using data acquired in different clinical centers, with different instrumentation, but similar acquisition protocols. Although the differentiation between glioblastomas and metastases resulted in an accuracy of 78 %, the remaining pairwise discriminations among glioblastomas, meningiomas, metastases, and low-grade gliomas achieved accuracies of around 90 %. Therefore, a well-defined protocol for the acquisition of MRS and the application of quality control to MR spectra should allow to use such classifiers via DSSs. Chapter 10 focused on the HealthAgents project, which offered the opportunity to develop such a dDSS. The system was managed by agent technology and contained database nodes and classifier nodes. The core of the HealthAgents dDSS was a classification framework, which offered brain tumor classification on the multidisciplinary research platform of HealthAgents. Different centers could create and include their classifiers in the platform and clinical centers could access these classifiers, providing decision support in day-to-day practice.

Chapter 7 demonstrated that standard Bayesian LS-SVM classifiers could be extended to a multi-class classifier system, thereby providing class probabilities. The multi-class problem was tackled by the one-versus-one approach, resulting in more balanced data sets and potentially more computationally efficient classifier training. Various pairwise coupling methods were compared on a combined MRSI and MRI data set, containing samples from 10 tissue types. The first combination method by Wu et al. was preferred because of its performance and computational efficiency. The developed multi-class system was used in Chapter 8 to generate nosologic images of the brain. Nosologic images allow easy and rapid display of the results for routine clinical interpretation of an MR examination. Therefore, an approach that exploited MRI and MRSI data was developed. Methods from image processing and pattern recognition were combined in a two-step approach. MRI segmentation was performed in the first step using a digital brain atlas and a subject-specific abnormal tissue prior, which was derived from MRSI. In the second step the segmented abnormal tissue was classified using pattern recognition and class probabilities were generated for the diverse tissue types. The methodology was applied to various patients with different types of brain tumors. Compared to previous approaches, the use of spatial information and prior knowledge was further improved. The methodology offered a new way to visualize tumor heterogeneity and class probabilities in a single nosologic image, assisting clinicians in decision making. The differentiation between glioblastoma

multiforme and solitary metastases based on long TE 2D-TSI and MRI was studied in Chapter 9. The Cho/Cr and Cho/NAA ratios were significantly different for glioblastomas compared to metastases in the enhancing tumor and the peri-enhancing area. The calculated difference between the ratio in the enhancing tumor and the ratio in the non-enhancing/non-T_2 hyperintense region was larger for glioblastomas compared to metastases. Subsequently, nosologic images and colored MRSI grids were used to visualize the infiltrative nature of glioblastomas into the adjacent brain parenchyma. It was shown that the abnormal tissue area of nosologic images could be improved based on this new approach. The findings illustrated that MRI and 2D-TSI, in contrast to the single-voxel MR spectroscopy approaches in Chapters 4, 5 and 6, are able to demonstrate the differences between glioblastomas and metastases, which can potentially lead to further therapeutic optimization.

11.2 Further research

For a breakthrough into mainstream, further research in the field of MR spectroscopy is necessary. Research should focus on improving data acquisition techniques, signal preprocessing, data processing and, finally, validation. Concerning data acquisition, the use of higher magnetic fields is promising in terms of increasing the SNR and the amount of detectable metabolites. Furthermore, it results in an improved spectral and spatial resolution. Recently, an increase in the number of 3 T scanners has noticeably improved neuroimaging. Interestingly, there are already studies that explored the possibilities of MRSI of the human brain at 7 T. MR spectra at these higher magnetic field strengths can potentially indicate more clear differences between tumor types and grades, increasing the performance of classifiers in a DSS.

The development of classifiers requires good quality MR spectra, which are undistorted by artifacts, line broadening or low SNR. In this thesis, quality control of the MRSI data was done in a manual way by experts. However, more objective, automated, quality control of MRSI is highly desired in a DSS. Objective decision rules, which can be classifiers, need to incorporated in the DSS itself. In this way, large amounts of MR spectra can be processed in a short period of time.

Since metabolite concentrations are much lower than the concentration of water, the SNR of MRSI is much lower as compared to MRI. As a consequence, MRSI data are obtained with a much smaller spatial resolution. Nowadays, there is considerable interest in improving the reconstruction of metabolite maps from MRSI. For example, anatomical spatial information can be incorporated as a prior

to increase the spatial resolution for improved reconstruction of metabolite images. Naturally, a higher spatial resolution for the MRSI data is advantageous for the techniques proposed in Chapter 8 and Chapter 9. Additionally, three-dimensional MRSI can further improve the results in these sections.

The role of quantification approaches as a preprocessing step in the classification of MRSI needs further research. In this thesis the use of quantification approaches, which were applied in a single-voxel sense, as compared to automated feature extraction techniques from machine learning and statistics was not advantageous. However, an interesting research direction is the inclusion of spatial information, as prior knowledge, in MRSI quantification methods. This can prevent the optimization algorithm to converge towards wrong local minima. This prior information is based on the assumption that MR spectra from neighboring voxels are more similar than MR spectra from voxels that are further apart. To this end, an initial clustering of the MRSI data can produce a valuable prior. Improved MRSI features potentially allow to improve nosologic images.

Instead of only using the pixel values from MRI in classifiers, one can extract other (tumor) characteristics from MRI for inclusion in a DSS. For example, the tumor location, heterogeneity, contrast enhancement and the presence of edema, cysts or necrosis are important features to make a differential diagnosis. An interesting approach is to score these features and combine them with MRSI information to create classifiers for nosologic images.

In the eTUMOUR project a centralized database with clinical information, MRI, MRS, MRSI, HR-MAS and microarrays was created. An important research topic is to develop approaches that link these in vivo or ex vivo data in an intelligent way. For instance, this thesis demonstrated that different data sources as MRI and 2D-TSI are complementary. They have a certain amount of information in common, but they also show distinct differences, such that integration of both allows to make differential diagnoses between tumor tissues. The genomic and the metabolic data that are acquired in eTUMOUR provide an extensive amount of valuable, complementary, data, requiring an in-depth analysis. Furthermore, in this thesis clinical information, as age or gender, was, in general, not used as a feature for classification. These factors can be taken into account when creating classifiers or can for instance be used to generate an appropriate prior distribution for Bayesian classifiers. In addition, the use of DTI, diffusion-weighted MRI and perfusion-weighted MRI offers interesting information for improved delineation of tumors.

MRS data for the independent test set in Chapter 6 were selected if they were

available in the eTUMOUR consortium before February 28th, 2007. In the meanwhile, the data providing partners in eTUMOUR acquired much more data, which can potentially be added for additional testing. Moreover, classifiers can be retrained by combining all available data from INTERPRET and eTUMOUR. Future validation studies should include classifier testing, in a prospective setting, in clinical centers outside the INTERPRET, eTUMOUR or HealthAgents consortia.

The HealthAgents project developed a web-based dDSS, as a proof of principle, during the period 2006−2008. In order to increase the quality of the dDSS, the HealthAgents network needs to be extended with additional classifier nodes and clinical nodes. The existing classifiers can be improved by retraining them with additional data. New classifiers should be included based on other types of data such as HR-MAS or microarrays. In addition, the dDSS would benefit from the implementation of new algorithms in the classifier nodes.

Although this thesis focussed on the diagnosis of brain tumors, the techniques can also be used for studying disease progression and response to therapy. MRI provides important morphologic information, but it is for example unable to distinguish between active tumor and radiation necrosis in a reliable way. On the other hand, MRSI allows to verify the effect of surgical resection on the spatial extent of the metabolic lesion. It is an important tool for predicting recurrence, and planning radiation or chemotherapy.

It is clear that MR changed the landscape of modern medicine. Upcoming, further, evolutions of this technology will offer a new era of fascinating research: exciting techniques with attractive applications.

Bibliography

[1] Abe S. *Support vector machines for pattern classification.* Springer, New York, 2005.

[2] Adams N.M. and Hand D.J. Improving the practice of classifier performance assessment. *Neural Computation*, 12:305–311, 2000.

[3] Alexander C., Ishikawa S., Silverstein M., Jacobson M., Fiksdahl-King I., and Angel S. *A pattern language: towns, buildings, construction.* Oxford University Press, Oxford, 1977.

[4] Allwein E., Schapire R., and Singer Y. Reducing multiclass to binary: A unifying approach for margin classifiers. *Journal of Machine Learning Research*, 1:113–141, 2000.

[5] Almuallim H. and Dietterich T.G. Learning with many irrelevant features. In *Proceedings of the 9th National Conference on Artificial Intelligence*, pages 547–552, Anaheim, 1991.

[6] Alpaydin E. *Introduction to machine learning.* MIT Press, Cambridge, 2004.

[7] Altman D.G. and Royston P. What do we mean by validating a prognostic model? *Statistics in Medicine*, 19:453–473, 2000.

[8] Anagnostou T., Remzi M., and Djavan B. Artificial neural networks for decision-making in urologic oncology. *European Urology*, 43(6):596–603, 2003.

[9] Arnold D.L. and De Stefano N. Magnetic resonance spectroscopy in vivo: applications in neurological disorders. *The Italian Journal of Neurological Sciences*, 18:321–329, 1997.

[10] Auer D.P., Gossl C., Schirmer T., and Czisch M. Improved analysis of ^{1}H-MR spectra in the presence of mobile lipids. *Magnetic Resonance in Medicine*, 46:615–618, 2001.

[11] Barkhuijsen H., de Beer R., and van Ormondt D. Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals. *Journal of Magnetic Resonance*, 73:553–557, 1987.

[12] Bartha R., Drost D.J., Menon R.S., and Williamson P.C. Spectroscopic lineshape correction by QUECC: combined QUALITY deconvolution and eddy current correction. *Magnetic Resonance in Medicine*, 44:641–645, 2000.

[13] Bellifemine F., Poggi A., and Rimassa G. JADE - A FIPA-compliant agent framework. In *Proceedings of the 4th International Conference on Practical Application of Intelligent Agents and Multi-Agent Technology*, pages 97–108, London, 1999.

[14] Bishop C.M. *Neural networks and pattern recognition*. Oxford University Press, Oxford, 1995.

[15] Bishop C.M. *Pattern recognition and machine learning*. Springer, New York, 2006.

[16] Blum A. and Langley P. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97:245–271, 1997.

[17] Blumer A., Ehrenfeucht A., Haussler D., and Marmuth M.K. Occam's razor. *Information Processing Letters*, 24:377–380, 1987.

[18] Boser B.E., Guyon I.M., and Vapnik V.N. A training algorithm for optimal margin classifiers. In *Proceedings of the 5th annual workshop on Computational Learning Theory*, pages 144–152, Pittsburgh, 1992.

[19] Bottomley P.A. Spatial localization in NMR spectroscopy in vivo. *Annals of the New York Academy of Sciences*, 508:333–348, 1987.

[20] Breiman L. Bagging predictors. *Machine Learning*, 24:123–140, 1996.

[21] Brier G.W. Verification of forecasts expressed in probabilities. *Monthly Weather Review*, 78:1–3, 1950.

[22] Brugali D. and Sycara K. Towards agent oriented application frameworks. *ACM Computing Surveys*, 32:21–27, 2000.

[23] Buetow M.P., Buetow P.C., and Smirniotupoulos J.G. Typical, atypical and misleading features in meningiomas. *Radiographics*, 11:1087–1106, 1991.

[24] Bunin G.R., Surawicz T.S., Witman P.A., Preston-Martin S., Davis F., and Bruner J.M. The descriptive epidemiology of craniopharyngioma. *Journal of Neurosurgery*, 89:547–551, 1998.

[25] Burtscher I.M., Skagerberg G., Geijer B., Englund E., Stahlberg F., and Holtas S. Proton MR spectroscopy and preoperative diagnostic accuracy: an evaluation of intracranial mass lesions characterized by stereotactic biopsy findings. *American Journal of Neuroradiology*, 21:84–93, 2000.

[26] Cardie C. Using decision trees to improve case-based learning. In *Proceedings of the 10th International Conference on Machine Learning*, pages 25–32, Amherst, 1993.

[27] Cardoso J.-F. and Souloumiac A. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140:362–370, 1993.

[28] Castillo M., Smith J.K., and Kwock L. Correlation of myo-inositol levels and grading of cerebral astrocytomas. *American Journal of Neuroradiology*, 21:1645–1649, 2000.

[29] Celda B., Monleon D., Martinez-Bisbal M.C., Esteve V., Martinez-Granados B., Pinero E., Ferrer R., Piquer J., Marti-Bonmati L., and Cervera J. MRS as endogenous molecular imaging for brain and prostate tumors: FP6 project "eTUMOR". *Advances in Experimental Medicine and Biology*, 587:285–302, 2006.

[30] Cha S., Lupo J.M., Chen M.-H., Lamborn K.R., McDermott M.W., Berger M.S., Nelson S.J., and Dillon W.P. Differentiation of glioblastoma multiforme and single brain metastasis by peak height and percentage of signal intensity recovery derived from dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. *American Journal of Neuroradiology*, 28:1078–1084, 2007.

[31] Chiang I.C., Kuo Y.-T., Lu C.-Y., Yeung K.-W., Lin W.-C., Sheu F.-O., and Liu G.-C. Distinction between high-grade gliomas and solitary metastases using peritumoral 3-T magnetic resonance spectroscopy, diffusion, and perfusion imagings. *Neuroradiology*, 46:619–627, 2004.

[32] Clark M.C., Hall L.O., Goldgof D.B., Velthuizen R., Murtagh F.R., and Silbiger M.S. Automatic tumor segmentation using knowledge-based techniques. *IEEE Transactions on Medical Imaging*, 17:187–201, 1998.

[33] Cohen-Gadol A.A., Pan J.W., Kim J.H., Spencer D.D., and Hetherington H.H. Mesial temporal lobe epilepsy: a proton magnetic resonance spectroscopy study and a histopathological analysis. *Journal of Neurosurgery*, 101:613–620, 2004.

[34] Coppini G., Diciotti S., Falchini M., Villari N., and Valli G. Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms. *IEEE Transactions on Information Technology in Biomedicine*, 7:344–357, 2003.

[35] Croteau D., Scarpace L., Hearshen D., Gutierrez J., Fisher J.L., Rock J.P., and Mikkelsen T. Correlation between magnetic resonance spectroscopy imaging and image-guided biopsies: Semiquantitative and qualitative histopathological analyses of patients with untreated glioma. *Neurosurgery*, 49:823–829, 2001.

[36] Cuadra M.B., Gomez J., Hagmann P., Pollo C., Villemure J.-G., Dawant B.M., and Thiran J.-P. Atlas-based segmentation of pathological brains using a model of tumor growth. In *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 380–387, Tokyo, 2002.

[37] Daubechies I. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[38] Davis F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13:319–340, 1989.

[39] De Graaf A.A., Van Dijk J.E., and Bovée W.M.M.J. QUALITY: quantification improvement by converting lineshapes to the lorentzian type. *Magnetic Resonance in Medicine*, 13:343–357, 1990.

[40] de Graaf R.A. *In vivo NMR spectroscopy: principles and techniques*. Wiley, New York, 1998.

[41] De Neuter B., Luts J., Vanhamme L., Lemmerling P., and Van Huffel S. Java-based framework for processing and displaying short-echo-time magnetic resonance spectroscopy signals. *Computer Methods and Programs in Biomedicine*, 85:129–137, 2007.

[42] Demsar J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[43] Devos A., Lukas L., Suykens J.A.K., Vanhamme L., Tate A.R., Howe F.A., Majos C., Moreno-Torres A., van der Graaf M., Arus C., and Van Huffel S. Classification of brain tumours using short echo time 1H MR spectra. *Journal of Magnetic Resonance*, 170:164–175, 2004.

[44] Devos A., Simonetti A.W., van der Graaf M., Lukas L., Suykens J.A.K., Vanhamme L., Buydens L.M.C., Heerschap A., and Van Huffel S. The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification. *Journal of Magnetic Resonance*, 173:218–228, 2005.

[45] Di Costanzo A., Scarabino T., Trojsi F., Giannatempo G.M., Popolizio T., Catapano D., Bonavita S., Maggialetti N., Tosetti M., Salvolini U., d'Angelo V.A., and Tedeschi G. Multiparametric 3T MR approach to the assessment of cerebral gliomas: tumor extent and malignancy. *Neuroradiology*, 48:622–631, 2006.

[46] Di Costanzo A., Scarabino T., Trojsi F., Popolizio T., Catapano D., Giannatempo G.M., Bonavita S., Portaluri M., Tosetti M., d'Angelo V.A., Salvolini U., and Tedeschi G. Proton MR spectroscopy of cerebral gliomas at 3T: spatial heterogeneity, and tumor grade and extent. *European Radiology*, 18:1727–1735, 2008.

[47] Dietterich T.G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10:1895–1923, 1998.

[48] Dietterich T.G. Machine learning research: four current directions. *The AI Magazine*, 18:97–136, 1998.

[49] Dietterich T.G. and Bakiri G. Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2:263–286, 1995.

[50] Duda R., Hart P., and Stork D. *Pattern classification*. Wiley, New York, 2001.

[51] Duyn J.H. and Moonen C.T.W. Fast proton spectroscopic imaging of human brain using multiple spin-echoes. *Magnetic Resonance in Medicine*, 30:409–414, 1993.

[52] Eichler A.F. and Loeffler J.S. Multidisciplinary management of brain metastases. *The Oncologist*, 12:884–898, 2007.

[53] El-Deredy W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy : a review. *NMR in Biomedicine*, 10:99–124, 1997.

[54] Elderkin-Thompson V., Thomas M.A., Binesh N., Mintz J., Haroon E., Dunkin J.J., and Kumar A. Brain metabolites and cognitive function among older depressed and healthy individuals using 2D MR spectroscopy. *Neuropsychopharmacology*, 29:2251–2257, 2004.

[55] Ewend M.G., Morris D.E., Carey L.A., Ladha A.M., and Brem S. Guidelines for the initial management of metastatic brain tumors: role of surgery, radiosurgery, and radiation therapy. *Journal of the National Comprehensive Cancer Network*, 6:505–513, 2008.

[56] Fan G., Sun B., Wu Z., Guo Q., and Guo Y. In vivo single-voxel proton MR spectroscopy in the differentiation of high-grade gliomas and solitary metastases. *Clinical Radiology*, 59:77–85, 2004.

[57] Favre J., Taha J.M., and Burchiel K.J. An analysis of the respective risks of hematoma formation in 361 consecutive morphological and functional stereotactic procedures. *Neurosurgery*, 50:48–57, 2002.

[58] Field M., Witham T.F., Flickinger J.C., Kondziolka D., and Lunsford L.D. Comprehensive assessment of hemorrhage risks and outcomes after stereotactic brain biopsy. *Journal of Neurosurgery*, 94:545–551, 2001.

[59] Fisher R.A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[60] Fletcher-Heath L.M., Hall L.O., Goldgof D.B., and Murtagh F.R. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial Intelligence in Medicine*, 21:43–63, 2001.

[61] Foran D.J., Comaniciu D., Meer P., and Goodell L.A. Computer-assisted discrimination among malignant lymphomas and leukemia using immunophenotyping, intelligent image repositories, and telemicroscopy. *IEEE Transactions on Information Technology in Biomedicine*, 4:265–273, 2000.

[62] Frahm J., Merboldt K.-D., and Hanicke W. Localized proton spectroscopy using stimulated echoes. *Journal of Magnetic Resonance*, 72:502–508, 1987.

[63] Friedman J. Another approach to polychotomous classification. Technical report, Stanford University, United States, 1996.

[64] Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32:675–701, 1937.

[65] Friedman M. A comparison of alternative tests of significance for the problem of m rankings. *Annals of Mathematical Statistics*, 11:86–92, 1940.

[66] Furuya S., Naruse S., Ide M., Morishita H., Kizu O., Ueda S., and Maeda T. Evaluation of metabolic heterogeneity in brain tumors using ^{1}H-chemical shift imaging method. *Neurosurgery*, 10:25–30, 1997.

[67] Galanaud D., Nicoli F., Chinot O., Confort-Gouny S., Figarella-Branger D., Roche P., Fuentes S., Fur Y.L., Ranjeva J.-P., and Cozzone P.J. Noninvasive diagnostic assessment of brain tumors using combined in vivo MR imaging and spectroscopy. *Magnetic Resonance in Medicine*, 55:1236–1245, 2006.

[68] Gamma E., Helm R., Johnson R., and Vlissides J. *Design patterns: elements of reusable object-oriented software*. Addison-Wesley, San Francisco, 1995.

[69] Ganslandt O., Stadlbauer A., Fahlbusch R., Kamada K., Buslei R., Blumcke I., Moser E., and Nimsky C. Proton magnetic resonance spectroscopic imaging integrated into image-guided surgery: correlation to standard magnetic resonance imaging and tumor cell density. *Neurosurgery*, 56:291–298, 2005.

[70] Garcia-Gomez J.M., Luts J., Julia-Sape M., Krooshof P., Tortajada S., Vicente Robledo J., Melssen W., Fuster-Garcia E., Olier I., Postma G., Monleon D., Moreno-Torres A., Pujol J., Candiota A.-P., Martinez-Bisbal M.C., Suykens J.A.K., Buydens L.M.C., Celda B., Van Huffel S., Arus C., and Robles M. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 22:5–18, 2009.

[71] Garcia-Gomez J.M., Tortajada S., Vidal C., Julia-Sape M., Luts J., Van Huffel S., Arus C., and and Robles M. The effect of combining two echo times in automatic brain tumor classification by MRS. *NMR in Biomedicine*, 21:1112–1125, 2008.

[72] Garcia-Gomez J.M., Vidal C., Marti-Bonmati L., and Robles M. Distributed decision support architecture for soft tissue tumor diagnosis. In *Proceedings of the 11th World Congress on Medical Informatics*, page 1608, San Francisco, 2004.

[73] Garcia S. and Herrera F. An extension on "Statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *Journal of Machine Learning Research*, 9:2677–2694, 2008.

[74] Garczarek U.M. *Classification rules in standardized partition spaces*. PhD thesis, University of Dortmund, 2002.

[75] Gering D.T., Grimson W.E.L., and Kikinis R. Recognizing deviations from normalcy for brain tumor segmentation. In *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 388–395, Tokyo, 2002.

[76] Ginsberg M.L. *Essentials of artificial intelligence*. Morgan Kaufmann, San Francisco, 1993.

[77] Golub G.H. and Van Loan C.F. *Matrix computations*. The Johns Hopkins University Press, Baltimore, 1996.

[78] Gonzalez-Velez H., Mier M., Julia-Sape M., Arvanitis T.N., Garcia-Gomez J.M., Robles M., Lewis P.H., Dasmahapatra S., Dupplaw D., Peet A., Arus C., Celda B., Van Huffel S., and Lluch-Ariet M. HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis. *Applied Intelligence*, 30:191–202, 2009.

[79] Govindaraju V., Young K., and Maudsley A.A. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine*, 13:129–153, 2000.

[80] Gray H.F., Maxwell R.J., Martinez-Perez I., Arus C., and Cerdan S. Genetic programming for classification and feature selection: analysis of 1H nuclear magnetic resonance spectra from human brain tumour biopsies. *NMR in Biomedicine*, 11:217–224, 1998.

[81] Guyon I., Alamdari A.R.S.A., Dror G., and Buhmann J.M. Performance prediction challenge. In *Proceedings of the 19th International Joint Conference on Neural Networks*, pages 1649–1656, Vancouver, 2006.

[82] Guyon I., Weston J., Barnhill S., and Vapnik V.N. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46:389–422, 2002.

[83] Hagberg G. From magnetic resonance spectroscopy to classification of tumors: a review of pattern recognition methods. *NMR in Biomedicine*, 11:148–156, 1998.

[84] Hall W.A. The safety and efficacy of stereotactic biopsy for intracranial lesions. *Cancer*, 82:1749–1755, 1998.

[85] Hamstra D.A., Galban C.J., Meyer C.R., Johnson T.D., Sundgren P.C., Tsien C., Lawrence T.S., Junck L., Ross D.J., Rehemtulla A., Ross B.D., and Chenevert T.L. Functional diffusion map as an early imaging biomarker for high-grade glioma: correlation with conventional radiologic response and overall survival. *Journal of Clinical Oncology*, 26:3387–3394, 2008.

[86] Hanley J.A. and McNeil B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143:29–36, 1982.

[87] Hardy D.L. and Norwood T.J. Spectral editing technique for the in vitro and in vivo detection of taurine. *Journal of Magnetic Resonance*, 133:70–78, 1998.

[88] Hastie T. and Tibshirani R. Classification by pairwise coupling. *The Annals of Statistics*, 26:451–471, 1998.

[89] Hastie T., Tibshirani R., and Friedman J.H. *The Elements of Statistical Learning*. Springer, New York, 2001.

[90] Ho S., Bullitt E., and Gerig G. Level-set evolution with region competition: automatic 3-D segmentation of brain tumors. In *Proceedings of the 16th International Conference on Pattern Recognition*, pages 532–535, Quebec, 2002.

[91] Hollander M. and Wolfe D.A. *Nonparametric Statistical Methods*. Wiley, New York, 1973.

[92] Hosmer D.W. and Lemeshow S. *Applied logistic regression*. Wiley, New York, 2000.

[93] Hotelling H. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

[94] Howe F.A., Barton S.J., Cudlip S.A., Stubbs M., Saunders D.E., Murphy M.M., Wilkins P.R., Opstad K.S., Doyle V.L., McLean M.C., Bell B.A., and Griffiths J.R. Metabolic profiles of human brain tumors using quantitative in vivo ^{1}H magnetic resonance spectroscopy. *Magnetic Resonance in Medicine*, 49:223–232, 2003.

[95] Howe F.A. and Opstad K.S. ^{1}H MR spectroscopy of brain tumours and masses. *NMR in Biomedicine*, 16:123–131, 2003.

[96] Huang Y., Lisboa P.J.G., and El-Deredy W. Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection. *Statistics in Medicine*, 22:147–164, 2003.

[97] Hunter D.R. MM algorithms for generalized Bradley-Terry models. *The Annals of Statistics*, 32:386–408, 2004.

[98] Ishimaru H., Morikawa M., Iwanaga S., Kaminogo M., Ochi M., and Hayashi K. Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. *European Radiology*, 9:1784–1791, 2001.

[99] Janny P., Cure H., Mohr M., Heldt N., Kwiatkowski F., Lemaire J.J., Plagne R., and Rozan R. Low grade supratentorial astrocytomas: management and prognostic factors. *Cancer*, 73:1937–1945, 1994.

[100] John G.H., Kohavi R., and Pfleger K. Irrelevant features and the subset selection problem. In *Proceedings of the 11th International Conference on Machine Learning*, pages 121–129, New Brunswick, 1994.

[101] Julia-Sape M., Acosta D., Mier M., Arus C., Watson D., and The INTERPRET consortium. A multi-centre, web-accessible and quality control-checked database of in vivo MR spectra of brain tumour patients. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 19:22–33, 2006.

[102] Jung J.A., Coakley F.V., Vigneron D.B., Swanson M.G., Qayyum A., Weinberg V., Jones K.D., Carroll P.R., and Kurhanewicz J. Prostate depiction at endorectal MR spectroscopic imaging: investigation of a standardized evaluation system. *Radiology*, 233:701–708, 2004.

[103] Justice A.C., Covinsky K.E., and Berlin J.A. Assessing the generalizability of prognostic information. *Annals of Internal Medicine*, 130:515–524, 1999.

[104] Karsmakers P., Pelckmans K., and Suykens J.A.K. Multi-class kernel logistic regression: a fixed-size implementation. In *Proceedings of the 20th International Joint Conference on Neural Networks*, pages 1756–1761, Orlando, 2007.

[105] Kaus M., Warfield S.K., Nabavi A., Black P.M., Jolesz F.A., and Kikinis R. Automated segmentation of MRI of brain tumors. *Radiology*, 218:586–591, 2001.

[106] Kaus M., Warfield S.K., Nabavi A., Chatzidakis E., Black P.M., Jolesz F.A., and Kikinis R. Segmentation of meningiomas and low grade gliomas in MRI. In *Proceedings of the 2nd International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1–10, Cambridge, 1999.

[107] Keerthi S.S., Duan K.B., Shevade S.K., and Poo A.N. A fast dual algorithm for kernel logistic regression. *Machine Learning*, 61:151–165, 2005.

[108] Kelm B.M. *Evaluation of vector-valued clinical image data using probabilistic graphical models: quantification and pattern recognition.* PhD thesis, University of Heidelberg, 2007.

[109] Kelm B.M., Menze B.H., Neff T., Zechmann C., and Hamprecht F.A. CLARET: A tool for fully automated evaluation of MRSI with pattern recognition methods in *Bildverarbeitung fur die Medizin*, Handels H., Ehrhardt J., Horsch A., Meinzer H.-P., and Tolxdorff T., pages 51–55. Springer, New York, 2006.

[110] Kelm B.M., Menze B.H., Zechmann C.M., Baudendistel K.T., and Hamprecht F.A. Automated estimation of tumor probability in prostate MRSI: pattern recognition vs. quantification. *Magnetic Resonance in Medicine*, 57:150–159, 2007.

[111] Kinoshita Y. and Yokota A. Absolute concentrations of metabolites in human brain tumors using in vitro proton magnetic resonance spectroscopy. *NMR in Biomedicine*, 10:2–12, 1997.

[112] Kira K. and Rendell L.A. A practical approach to feature selection. In *Proceedings of the 9th International Workshop on Machine Learning*, pages 249–256, Aberdeen, 1992.

[113] Klos K.J. and O'Neill B.P. Brain metastases. *The Neurologist*, 10:31–46, 2004.

[114] Klose U. In vivo proton spectroscopy in presence of eddy currents. *Magnetic Resonance in Medicine*, 14:26–30, 1990.

[115] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1137–1145, Montreal, 1995.

[116] Kohavi R. and John G.H. Wrappers for feature subset selection. *Artificial Intelligence*, 97:273–324, 1997.

[117] Kreßel U.H.-G. Pairwise classification and support vector machines in *Advances in kernel methods: support vector learning*, Scholkopf B., Burges C.J.C., and Smola A.J., pages 255–268. MIT Press, Cambridge, 1999.

[118] Kuat M., Flotzinger D., and Pfurtscheller G. Discovering patterns in EEG-signals: Comparative study of a few methods. In *Proceedings of the 6th European Conference on Machine Learning*, pages 366–371, Vienna, 1993.

[119] Laudadio T., Martinez-Bisbal M.C., Celda B., and Van Huffel S. Fast nosological imaging using canonical correlation analysis of brain data obtained by two-dimensional turbo spectroscopic imaging. *NMR in Biomedicine*, 21:311–321, 2008.

[120] Laudadio T., Mastronardi N., Vanhamme L., Van Hecke P., and Van Huffel S. Improved Lanczos algorithms for blackbox MRS data quantitation. *Journal of Magnetic Resonance*, 157:292–297, 2002.

[121] Laudadio T., Pels P., De Lathauwer L., Van Hecke P., and Van Huffel S. Tissue segmentation and classification of MRSI data using canonical correlation analysis. *Magnetic Resonance in Medicine*, 54:1519–1529, 2005.

[122] Law M., Cha S., Knopp E.A., Johnson G., Arnett J., and Litt A.W. High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. *Radiology*, 222:715–721, 2002.

[123] Law M., Yang S., Wang H., Babb J.S., Johnson G., Cha S., Knopp E.A., and Zagzag D. Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging. *American Journal of Neuroradiology*, 24:1989–1998, 2003.

[124] Lean C.L., Somorjai R.L., Smith I.C.P., Russell P., and Mountford C.E. Accurate diagnosis and prognosis of human cancers by proton MRS and a three stage classification strategy in *Annual reports on NMR spectroscopy*, Webb G., pages 71–111. Academic Press, Kent, 2002.

[125] Leaper D.J., De Dombal F.T., Horrocks J.C., and Staniland J.R. Computer-assisted diagnosis of abdominal pain using estimates provided by clinicians. *The British Journal of Surgery*, 59:897–898, 1972.

[126] Lefranc F., Sadeghi N., Camby I., Metens T., Dewitte O., and Kiss R. Present and potential future issues in glioblastoma treatment. *Expert Review of Anticancer Therapy*, 6:719–732, 2006.

[127] Li X., Lu Y., Pirzkall A., McKnight T.R., and Nelson S.J. Analysis of the spatial characteristics of metabolic abnormalities in newly diagnosed glioma patients. *Journal of Magnetic Resonance Imaging*, 16:229–237, 2002.

[128] Lin H.-T., Lin C.-J., and Weng R.C. A note on Platt's probabilistic outputs for support vector machines. *Machine Learning*, 68:267–276, 2007.

[129] Lindon J.C., Holmes E., and Nicholson J.K. Pattern recognition methods and applications in biomedical magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 39:1–40, 2001.

[130] Lisboa P.J.G. and Taktak A.F.G. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Networks*, 19:408–415, 2006.

[131] Lisboa P.J.G., Wong H., Harris P., and Swindell R. A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. *Artificial Intelligence in Medicine*, 28:1–25, 2003.

[132] Littlestone N. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2:285–318, 1988.

[133] Louis D.N., Ohgaki H., Wiestler O.D., and Cavenee W.K. *World Health Organization classification of tumours of the central nervous system*. IARC, Lyon, 2007.

[134] Lowe D.G. Similarity metric learning for a variable-kernel classifier. *Neural Computation*, 7:72–85, 1995.

[135] Lu C., Devos A., Suykens J.A.K., Arus C., and Van Huffel S. Bagging linear sparse Bayesian learning models for variable selection in cancer diagnosis. *IEEE Transactions on Information Technology in Biomedicine*, 11:338–347, 2007.

[136] Lucas P.J., Boot H., and Taal B.G. Computer-based decision support in the management of primary gastric non-Hodgkin lymphoma. *Methods of Information in Medicine*, 37:206–219, 1998.

[137] Lukas L., Devos A., Suykens J.A.K., Vanhamme L., Howe F.A., Majos C., Moreno-Torres A., van der Graaf M., Tate A.R., Arus C., and Van Huffel S. Brain tumour classification based on long echo proton MRS signals. *Artificial Intelligence in Medicine*, 31:73–89, 2004.

[138] Luts J., Heerschap A., Suykens J.A.K., and Van Huffel S. A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection. *Artificial Intelligence in Medicine*, 40:87–102, 2007.

[139] Luts J., Laudadio T., Idema A.J., Simonetti A.W., Heerschap A., Vandermeulen D., Suykens J.A.K., and Van Huffel S. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. *NMR in Biomedicine*, 22:374–390, 2009.

[140] Luts J., Laudadio T., Martinez-Bisbal M.C., Van Cauter S., Molla E., Piquer J., Suykens J.A.K., Himmelreich U., Celda B., and Van Huffel S. Differentiation between brain metastases and glioblastoma multiforme based on MRI, MRS and MRSI. In *Proceedings of the 22nd IEEE International Symposium on Computer-Based Medical Systems*, pages 1–8, Albuquerque, 2009.

[141] Luts J., Laudadio T., Martinez-Bisbal M.C., Van Cauter S., Molla E., Piquer J., Suykens J.A.K., Himmelreich U., Van Huffel S., and Celda B. Differentiation between solitary brain metastasis and glioblastoma multiforme using MRI and two-dimensional turbo spectroscopic imaging data. Submitted to *Neuroradiology*, 2009.

[142] Luts J., Poullet J.-B., Garcia-Gomez J.M., Heerschap A., Robles M., Suykens J.A.K., and Van Huffel S. Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. *Magnetic Resonance in Medicine*, 60:288–298, 2008.

[143] MacKay D.J.C. Bayesian interpolation. *Neural Computation*, 4:415–447, 1992.

[144] MacKay D.J.C. The evidence framework applied to classification networks. *Neural Computation*, 4:720–736, 1992.

[145] MacKay D.J.C. Introduction to Gaussian processes in *Neural networks and machine learning*, Bishop C.M., pages 133–165. Springer, New York, 1998.

[146] Maes F., Collignon A., Vandermeulen D., Marchal G., and Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16:187–198, 1997.

[147] Majos C., Alonso J., Aguilera C., Serrallonga M., Coll S., Acebes J.J., Arus C., and Gili J. Utility of proton MR spectroscopy in the diagnosis of radiologically atypical intracranial meningiomas. *Neuroradiology*, 45:129–136, 2003.

[148] Manton D.J., Lowry M., Blackband S.J., and Horsman A. Determination of proton metabolite concentrations and relaxation parameters in normal human brain and intracranial tumours. *NMR in Biomedicine*, 8:104–112, 1995.

[149] Marshall I., Higinbotham J., Bruce S., and Freise A. Use of Voigt lineshape for quantification of in vivo ^{1}H spectra. *Magnetic Resonance in Medicine*, 37:651–657, 1997.

[150] Martinez-Bisbal M.C., Celda B., Marti-Bonmati L., Ferrer P., Revert-Ventura A.J., Piquer J., Molla E., Arana R., and Dosda-Munoz R. The contribution of magnetic resonance spectroscopy for the classification of high grade glial tumours. The predictive value of macromolecules. *Revista de Neurologia*, 34(309-313), 2002.

[151] McLachlan G.J. *Discriminant analysis and statistical pattern recognition.* Wiley, New York, 1992.

[152] Melssen W., Ustun B., and Buydens L.M.C. SOMPLS: A supervised self-organising map–partial least squares algorithm for multivariate regression problems. *Chemometrics and Intelligent Laboratory Systems*, 86:102–120, 2007.

[153] Melssen W., Wehrens R., and Buydens L.M.C. Supervised kohonen networks for classification problems. *Chemometrics and Intelligent Laboratory Systems*, 83:99–113, 2006.

[154] Menze B.H., Lichy M.P., Bachert P., Kelm B.M., Schlemmer H.-P., and Hamprecht F.A. Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumor. *NMR in Biomedicine*, 19:599–609, 2006.

[155] Meyer R.A., Fisher M.J., Nelson S.J., and Brown T.R. Evaluation of manual methods for integration of in vivo phosphorus NMR spectra. *NMR in Biomedicine*, 1:131–135, 1988.

[156] Moller-Hartmann W., Herminghaus S., Krings T., Marquardt G., Lanfermann H., Pilatus U., and Zanella F.E. Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. *Neuroradiology*, 44:371–381, 2002.

[157] Nayak G.S., Kamath S., Pai K.M., Sarkar A., Ray S., Kurien J., D'Almeida L., Krishnanand B.R., Santhosh C., Kartha V.B., and Mahato K.K. Principal component analysis and artificial neural network analysis of oral tissue fluorescence spectra: classification of normal premalignant and malignant pathological conditions. *Biopolymers*, 82:152–166, 2006.

[158] Nelson S.J. Multivoxel magnetic resonance spectroscopy of brain tumors. *Molecular Cancer Therapeutics*, 2:497–507, 2003.

[159] Nelson S.J. Magnetic resonance spectroscopic imaging. *IEEE Engineering in Medicine and Biology Magazine*, 23:30–39, 2004.

[160] Nelson S.J., Vigneron D.B., and Dillon W.P. Serial evaluation of patients with brain tumors using volume MRI and 3D ^{1}H MRSI. *NMR in Biomedicine*, 12:123–138, 1999.

[161] Neter J., Kutner M.H., Nachtsheim C.J., and Wasserman W. *Applied linear statistical models.* McGraw-Hill, Boston, 1996.

[162] Nguyen T. and DeAngelis L.M. Treatment of brain metastases. *The Journal of Supportive Oncology*, 2:405–416, 2004.

[163] Nikulin A.E., Dolenko B., Bezabeh T., and Somorjai R.L. Near-optimal region selection for feature space reduction: novel preprocessing methods for classifying MR spectra. *NMR in Biomedicine*, 11:209–216, 1998.

[164] Norden A.D., Wen P.Y., and Kesari S. Brain metastases. *Current Opinion in Neurology*, 18:654–661, 2005.

[165] Opstad K.S., Ladroue C., Bell B.A., Griffiths J.R., and Howe F.A. Linear discriminant analysis of brain tumour ^{1}H MR spectra: a comparison of classification using whole spectra versus metabolite quantification. *NMR in Biomedicine*, 20:763–770, 2007.

[166] Opstad K.S., Murphy M.M., Wilkins P.R., Bell B.A., Griffiths J.R., and Howe F.A. Differentiation of metastases from high-grade gliomas using short echo time 1H spectroscopy. *Journal of Magnetic Resonance Imaging*, 20:187–192, 2004.

[167] Osorio J.A., Ozturk-Isik E., Xu D., Cha S., Chang S., Berger M.S., Vigneron D.B., and Nelson S.J. 3D 1H MRSI of brain tumors at 3.0 tesla using an eight-channel phased-array head coil. *Journal of Magnetic Resonance Imaging*, 26:23–30, 2007.

[168] Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9:62–66, 1979.

[169] Pelckmans K., Suykens J.A.K., Van Gestel T., De Brabanter J., Lukas L., Hamers B., De Moor B., and Vandewalle J. LS-SVMlab : a Matlab/C toolbox for least squares support vector machines. Internal Report 02-44, ESAT-SISTA, Katholieke Universiteit Leuven, Belgium, 2002.

[170] Perner P. Intelligent data analysis in medicine-recent advances. *Artificial Intelligence in Medicine*, 37:1–5, 2006.

[171] Piepmeier J., Christopher S., Spencer D., Byrne T., Kim J., Knisel J.P., Lacy J., Tsukerman L., and Makuch R. Variations in the natural history and survival of patients with supratentorial low-grade astrocytomas. *Neurosurgery*, 38:872–879, 1996.

[172] Pijnappel W.W.F., van den Boogaart A., de Beer R., and van Ormondt D. SVD-based quantification of magnetic resonance signals. *Journal of Magnetic Resonance*, 97:122–134, 1992.

[173] Platt J., Cristianini N., and Shawe-Taylor J. Large margin DAGs for multiclass classification. *Advances in Neural Information Processing Systems*, 12:547–553, 2000.

[174] Platt J.C. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods in *Advances in large margin classifiers*, Bartlett P.J., Scholkopf B., Schuurmans D., and Smola A.J., pages 61–74. MIT Press, Cambridge, 2000.

[175] Poptani H., Gupta R.K., Jain V.K., Roy R., and Pandey R. Cystic intracranial mass lesions: possible role of in vivo MR spectroscopy in its differential diagnosis. *Magnetic Resonance Imaging*, 13:1019–1029, 1995.

[176] Poptani H., Kaartinen J., Gupta R.K., Niemitz M., Hiltunen Y., and Kauppinen R.A. Diagnostic assessment of brain tumours and non-neoplastic brain disorders in vivo using proton nuclear magnetic resonance spectroscopy and artificial neural networks. *Journal of Cancer Research and Clinical Oncology*, 125:343–349, 1999.

[177] Postma G., Luts J., Idema A.J., Julia-Sape M., Moreno-Torres A., Gajewicz W., Suykens J.A.K., Heerschap A., Van Huffel S., and Buydens L.M.C. On the relevance of automatically selected single voxel MRS and multimodal MRI and MRSI features for brain tumor differentiation. Submitted to *Computers in Biology and Medicine*, 2010.

[178] Potts H.W.W., Wyatt J.C., and Altman D.G. Challenges in evaluating complex decision support systems: lessons from design-a-trial. In *Proceedings of the 8th Conference on Artificial Intelligence in Medicine in Europe*, pages 453–456, Cascais, 2001.

[179] Poullet J.-B., Sima D.M., Simonetti A.W., De Neuter B., Vanhamme L., Lemmerling P., and Van Huffel S. An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES. *NMR in Biomedicine*, 20:493–504, 2007.

[180] Prastawa M., Bullitt E., Ho S., and Gerig G. A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis*, 8:275–283, 2004.

[181] Prastawa M., Bullitt E., Moon N., Van Leemput K., and Gerig G. Automatic brain tumor segmentation by subject specific modification of atlas priors. *Academic Radiology*, 10:1341–1348, 2003.

[182] Preul M.C., Caramanos Z., Collins D.L., Villemure J.G., Leblanc R., Olivier A., Pokrupa R., and Arnold D.L. Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. *Nature Medicine*, 2:323–325, 1996.

[183] Preul M.C., Caramanos Z., Leblanc R., Villemure J.G., and Arnold D.L. Using pattern analysis of in vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors. *NMR in Biomedicine*, 11:192–200, 1998.

[184] Price D., Knerr S., Personnaz L., and Dreyfus G. Pairwise neural network classifiers with probabilistic outputs. In *Proceedings of the 8th Conference on Neural Information Processing Systems*, pages 1109–1116, Denver, 1994.

[185] Price S.J., Jena R., Burnet N.G., Hutchinson P.J., Dean A.F., Pena A., Pickard J.D., Carpenter T.A., and Gillard J.H. Improved delineation of glioma margins and regions of infiltration with the use of diffusion tensor imaging: An image-guided biopsy study. *American Journal of Neuroradiology*, 27:1969–1974, 2006.

[186] Provencher S.W. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magnetic Resonance in Medicine*, 30:672–679, 1993.

[187] Ratei R., Karawajew L., Lacombe F., Jagoda K., Del Poeta G., Kraan J., De Santiago M., Kappelmayer J., Bjorklund E., Ludwig W.D., Gratama J.W., and Orfao A. 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*, 21:1204–1211, 2007.

[188] Ratiney H., Coenradie Y., Cavassila S., van Ormondt D., and Graveron-Demilly D. Time-domain quantitation of ^{1}H short echo-time signals: background accommodation. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 16:284–296, 2004.

[189] Ratiney H., Sdika M., Coenradie Y., Cavassila S., van Ormondt D., and Graveron-Demilly D. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR in Biomedicine*, 18:1–13, 2005.

[190] Rees J. Advances in magnetic resonance imaging of brain tumours. *Current Opinion in Neurology*, 16:643–650, 2003.

[191] Refregier P. and Vallet F. Probabilistic approach for multiclass classification with neural networks. In *Proceedings of the 1st International Conference on Artificial Neural Networks*, pages 1003–1007, Espoo, 1991.

[192] Ricci R., Bacci A., Tugnoli V., Battaglia S., Maffei M., Agati R., and Leonardi M. Metabolic findings on 3T 1H-MR spectroscopy in peritumoral brain edema. *American Journal of Neuroradiology*, 28:1287–1291, 2007.

[193] Rice J.A. *Mathematical statistics and data analysis*. Duxbury Press, Belmont, 1995.

[194] Robnik-Sikonja M. and Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Machine Learning*, 53:23–69, 2003.

[195] Rosen BR, Wedeen VJ, and Brady TJ. Selective saturation NMR imaging. *Journal of Computer Assisted Tomography*, 8:813–818, 1984.

[196] Ross A.J. and Sachdev P.S. Magnetic resonance spectroscopy in cognitive research. *Brain Research Reviews*, 44:83–102, 2004.

[197] Ross B. and Michaelis T. Clinical applications of magnetic resonance spectroscopy. *Magnetic Resonance Quarterly*, 10:191–247, 1994.

[198] Saez C., Garcia-Gomez J.M., Vicente J., Tortajada S., Luts J., Dupplaw D., Van Huffel S., and Robles M. A generic and extensible framework for providing automatic classification functionality applied for brain tumour diagnosis in HealthAgents. Accepted for publication in *The Knowledge Engineering Review*, 2009.

[199] Schaefer P.W., Budzik R.F. Jr., and Gonzalez R.G. Imaging of cerebral metastases. *Neurosurgery Clinics of North America*, 7:393–423, 1996.

[200] Schmidt M., Levner I., Greiner R., Murtha A., and Bistritz A. Segmenting brain tumors using alignment-based features. In *Proceedings of the 4th International Conference on Machine Learning and Applications*, pages 215–220, Los Angeles, 2005.

[201] Shiino A., Nakasu S., Matsuda M., Handa J., Morikawa S., and Inubushi T. Noninvasive evaluation of the malignant potential of intracranial meningiomas performed using proton magnetic resonance spectroscopy. *Journal of Neurosurgery*, 91:928–934, 1999.

[202] Shortliffe E.H., A.C. Scott, M.B. Bischoff, A.B. Campbell, W. van Melle, and C.D. Jacobs. ONCOCIN: An expert system for oncology protocol management. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, pages 876–881, Vancouver, 1981.

[203] Simonetti A.W., Melsen W., van der Graaf M., Postma G., Heerschap A., and Buydens L.M.C. A chemometric approach for brain tumor classsification using magnetic resonance imaging and spectroscopy. *Analytical Chemistry*, 75:5352–5361, 2003.

[204] Simonetti A.W., Melssen W.J., Szabo de Edelenyi F., van Asten J.J.A., Heerschap A., and Buydens L.M.C. Combination of feature-reduced MR spectroscopic and

MR imaging data for improved brain tumor classification. *NMR in Biomedicine*, 18:34–43, 2005.

[205] Simonetti A.W., Melssen W.J., van der Graaf M., Heerschap A., and Buydens L.M.C. Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction. *Journal of Magnetic Resonance*, 159:151–157, 2002.

[206] Smith E.A., Carlos R.C., Junck L.R., Tsien C.I., Elias A., and Sundgren P.C. Developing a clinical decision model: MR spectroscopy to differentiate between recurrent tumor and radiation change in patients with new contrast-enhancing lesions. *American Journal of Roentgenology*, 192:45–52, 2009.

[207] Soher B.J., Young K., and Maudsley A.A. Representation of strong baseline contributions in ^{1}H MR spectra. *Magnetic Resonance in Medicine*, 45:966–972, 2001.

[208] Somorjai R.L., Dolenko B., Nikulin A.K., Pizzi N., Scarth G., Zhilkin P., Halliday W., Fewer D., Hill N., Ross I., West M., Smith I.C.P., Donnelly S.M., Kuesel A.C., and Briere K.M. Classification of 1H MR spectra of human brain neoplasms: The influence of preprocessing and computerized consensus diagnosis on classification accuracy. *Journal of Magnetic Resonance Imaging*, 6:437–444, 1996.

[209] Somorjai R.L., Nikulin A.E., Pizzi N., Jackson D., Scarth G., Dolenko B., Gordon H., Russell P., Lean C.L., Delbridge L., Mountford C.E., and Smith I.C.P. Computerized consensus diagnosis: a classification strategy for the robust analysis of MR spectra. I. Application to 1H spectra of thyroid neoplasms. *Magnetic Resonance in Medicine*, 33:257–263, 1995.

[210] Spyridonos P., Cavouras D., Ravazoula P., and Nikiforidis G. A computer-based diagnostic and prognostic system for assessing urinary bladder tumour grade and predicting cancer recurrence. *Medical Informatics and the Internet in Medicine*, 27:111–122, 2002.

[211] Stadlbauer A., Gruber S., Nimsky C., Fahlbusch R., Hammen T., Buslei R., Tomandl B., Moser E., and Ganslandt O. Preoperative grading of gliomas by using metabolite quantification with high-spatial-resolution proton MR spectroscopic imaging. *Radiology*, 238:958–969, 2006.

[212] Sundin T., Vanhamme L., Van Hecke P., Dologlou I., and Van Huffel S. Accurate quantification of ^{1}H spectra: from finite impulse response filter design for solvent suppression to parameter estimation. *Journal of Magnetic Resonance*, 139:189–204, 1999.

[213] Suykens J.A.K., Van Gestel T., De Brabanter J., De Moor B., and Vandewalle J. *Least squares support vector machines*. World Scientific, Singapore, 2002.

[214] Suykens J.A.K. and Vandewalle J. Least squares support vector machine classifiers. *Neural Processing Letters*, 9:293–300, 1999.

[215] Suykens J.A.K. and Vandewalle J. Multiclass least squares support vector machines. In *Proceedings of the 11th International Joint Conference on Neural Networks*, pages 900–903, Washington DC, 1999.

[216] Szabo de Edelenyi F., Rubin C., Esteve F., Grand S., Decorps M., Lefournier V., Le Bas J.F., and Remy C. A new approach for analyzing proton magnetic resonance

spectroscopic images of brain tumors: nosologic images. *Nature Medicine*, 6:1287–1289, 2000.

[217] Tate A.R., Griffiths J.R., Martinez-Perez I., Moreno A., Barba I., Cabanas M.E., Watson D., Alonso J., Bartumeus F., Isamat F., Ferrer I., Vila F., Ferrer E., Capdevila A., and Arus C. Towards a method for automated classification of ^{1}H MRS spectra from brain tumours. *NMR in Biomedicine*, 11:177–191, 1998.

[218] Tate A.R., Majos C., Moreno A., Howe F.A., Griffiths J.R., and Arus C. Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. *Magnetic Resonance in Medicine*, 49:29–36, 2003.

[219] Tate A.R., Underwood J., Acosta D.M., Julia-Sape M., Majos C., Moreno-Torres A., Howe F.A., van der Graaf M., Lefournier V., Murphy M.M., Loosemore A., Ladroue C., Wesseling P., Bosson J.-L., Cabanas M.E., Simonetti A.W., Gajewicz W., Calvar J., Capdevila A., Wilkins P.R., Bell B.A., Remy C., Heerschap A., Watson D., Griffiths J.R., and Arus C. Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra. *NMR in Biomedicine*, 19:411–434, 2006.

[220] Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58:267–288, 1996.

[221] Tong Z., Yamaki T., Harada K., and Houkin K. In vivo quantification of the metabolites in normal brain and brain tumors by proton MR spectroscopy using water as an internal standard. *Magnetic Resonance Imaging*, 22:735–742, 2004.

[222] Tsuchiya K., Fujikawa A., Nakajima M., and Honya K. Differentiation between solitary brain metastasis and high-grade glioma by diffusion tensor imaging. *British Journal of Radiology*, 78:533–537, 2005.

[223] Tukey J.W. Comparing individual means in the analysis of variance. *Biometrics*, 5:99–114, 1949.

[224] Tung W.L. and Quek C. GenSo-FDSS: a neural-fuzzy decision support system for pediatric ALL cancer subtype identification using gene expression data. *Artificial Intelligence in Medicine*, 33:61–88, 2005.

[225] Usenius J.-P., Tuohimetsa S., Vainio P., Ala-Korpela M., Hiltunen Y., and Kauppinen R.A. Automated classification of human brain tumours by neural network analysis using in vivo ^{1}H magnetic resonance spectroscopic metabolite phenotypes. *Neuroreport*, 7:1597–1600, 1996.

[226] Valentini G. and Dietterich T.G. Bias-variance analysis of support vector machines for the development of SVM-based ensemble methods. *Journal of Machine Learning Research*, 5:725–775, 2004.

[227] van der Graaf M., Julia-Sape M., Howe F.A., Ziegler A., Majos C., Moreno-Torres A., Rijpkema M., Acosta D., Opstad K.S., van der Meulen Y.M., Arus C., and Heerschap A. MRS quality assessment in a multicentre study on MRS-based classification of brain tumours. *NMR in Biomedicine*, 21:148–158, 2008.

[228] van der Graaf M., Rijpkema M., van der Meulen Y., Majos C., Moreno-Torres A., Ziegler A., Howe F.A., Opstad K.S., and Heerschap A. System quality assurance in INTERPRET. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 15:245–246, 2002.

[229] Van Driessen K. and Rousseeuw P.J. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999.

[230] Van Gestel T., Suykens J.A.K., De Moor B., and Vandewalle J. Automatic relevance determination for least squares support vector machine classifiers. In *Proceedings of the 9th European Symposium on Artificial Neural Networks*, pages 13–18. Bruges, 2001.

[231] Van Gestel T., Suykens J.A.K., Lanckriet G., Lambrechts A., De Moor B., and Vandewalle J. Bayesian framework for least squares support vector machine classifiers, Gaussian processes and kernel Fisher discriminant analysis. *Neural Computation*, 14:1115–1147, 2002.

[232] Van Leemput K., Maes F., Vandermeulen D., and Suetens P. Automated model-based bias field correction of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18:885–896, 1999.

[233] Van Leemput K., Maes F., Vandermeulen D., and Suetens P. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18:897–908, 1999.

[234] Vapnik V.N. *The nature of statistical learning theory*. Springer, New York, 2002.

[235] Wang S., Kim S., Chawla S., Wolf R.L., Zhang W.-G., O'Rourke D.M., Judy K.D., Melhem E.R., and Poptani H. Differentiation between glioblastomas and solitary brain metastases using diffusion tensor imaging. *Neuroimage*, 44:653–660, 2009.

[236] Wang W., Steward C.E., and Desmond P.M. Diffusion tensor imaging in glioblastoma multiforme and brain metastases: the role of p, q, L, and fractional anisotropy. *American Journal of Neuroradiology*, 30:203–208, 2008.

[237] Warfield S.K., Kaus M., Jolesz F.A., and Kikinis R. Adaptive, template moderated, spatially varying statistical classification. *Medical Image Analysis*, 4:43–55, 2000.

[238] Weber M.-A., Giesel F.L., and Stieltjes B. MRI for identification of progression in brain tumors: from morphology to function. *Expert Review of Neurotherapeutics*, 8:1507–1525, 2008.

[239] Wehrens R., Simonetti A.W., and Buydens L.M.C. Mixture modelling of medical magnetic resonance data. *Journal of Chemometrics*, 16:274–282, 2002.

[240] Wettschereck D., Aha D.W., and Mohri T. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*, 11:273–314, Feb 1997.

[241] Wu T., Lin C., and Weng R. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5:975–1005, 2004.

[242] Xiao L., Lewis P., and Gibb A. Developing a security protocol for a distributed decision support system in a healthcare environment. In *Proceedings of the 30th International Conference on Software Engineering*, pages 673–682, Leipzig, 2008.

[243] Ye C.Z., Yang J., Geng D.Y., Zhou Y., and Chen N.Y. Fuzzy rules to predict degree of malignancy in brain glioma. *Medical Biological Engineering Computing*, 40:145–152, 2002.

[244] Yiannoutsos C.T., Ernst T., Chang L., Lee P.L., Richards T., Marra C.M., Meyerhoff D.J., Jarvik J.G., Kolson D., Schifitto G., Ellis R.J., Swindells S., Simpson D.M., Miller E.N., Gonzalez R.G., and Navia B.A. Regional patterns of brain metabolites in AIDS dementia complex. *Neuroimage*, 23:928–935, 2004.

[245] Young G.S. and Setayesh K. Spin-echo echo-planar perfusion MR imaging in the differential diagnosis of solitary enhancing brain lesions: distinguishing solitary metastases from primary glioma. *American Journal of Neuroradiology*, 30:575–577, 2008.

[246] Zermelo E. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29:436–460, 1929.

[247] Zha H. *The singular value decomposition theory, algorithms and applications*. PhD thesis, Pennsylvania State University, 1993.

[248] Zheng M.M., Krishnan S.M., and Tjoa M.P. A fusion-based clinical decision support for disease diagnosis from endoscopic images. *Computers in Biology and Medicine*, 35:259–274, 2005.

[249] Zhu J. and Hastie T. Kernel logistic regression and the import vector machine. *Journal of Computational and Graphical Statistics*, 14:185–205, 2005.

Publication list

Papers in international journals

1. De Neuter B., Luts J., Vanhamme L., Lemmerling P., and Van Huffel S. Java-based framework for processing and displaying short-echo-time magnetic resonance spectroscopy signals. *Computer Methods and Programs in Biomedicine*, 85(2):129−137, 2007.

2. Luts J., Heerschap A., Suykens J.A.K., and Van Huffel S. A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection. *Artificial Intelligence in Medicine*, 40(2):87−102, 2007.

3. Luts J., Poullet J.-B., Garcia-Gomez J.M., Heerschap A., Robles M., Suykens J.A.K., and Van Huffel S. Effect of feature extraction for brain tumor classification based on short echo time 1H MR spectra. *Magnetic Resonance in Medicine* 60(2):288−298, 2008.

4. Garcia-Gomez J.M., Tortajada S., Vidal C., Julia-Sape M., Luts J., Van Huffel S., Arus C., and Robles M. The effect of combining two echo times in automatic brain tumor classification by MRS. *NMR in Biomedicine*, 21(10):1112−1125, 2008.

5. Garcia-Gomez J.M., Luts J., Julia-Sape M., Krooshof P., Tortajada S., Vicente J., Melssen W., Fuster-Garcia E., Olier I., Postma G., Monleon D., Moreno-Torres A., Pujol J., Candiota A.-P., Martinez-Bisbal M.C., Suykens J.A.K., Buydens L.M.C., Celda B., Van Huffel S., Arus C., and Robles M. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 22(1):5−18, 2009.

6. Luts J., Laudadio T., Idema A.J., Simonetti A.W., Heerschap A., Vandermeulen D., Suykens J.A.K., and Van Huffel S. Nosologic imaging of the brain:

segmentation and classification using MRI and MRSI. *NMR in Biomedicine*, 22(4):374−390, 2009.

7. Saez C., Garcia-Gomez J.M., Vicente J., Tortajada S., Luts J., Dupplaw D., Van Huffel S., and Robles M. A generic and extensible framework for providing automatic classification functionality applied for brain tumour diagnosis in HealthAgents. Accepted for publication in *The Knowledge Engineering Review*, 2009.

Papers submitted to international journals

8. Luts J., Laudadio T., Martinez-Bisbal M.C., Van Cauter S., Molla E., Piquer J., Suykens J.A.K., Himmelreich U., Van Huffel S., and Celda B. Differentiation between solitary brain metastasis and glioblastoma multiforme using MRI and two-dimensional turbo spectroscopic imaging data. Submitted to *Neuroradiology*, 2009.

9. Werbrouck E., Veldman J., Luts J., Van Huffel S., Van Schoubroeck D., Timmerman D., and Van den Bosch T. Saline infusion sonography (SIS) versus gel instillation sonography (GIS): a prospective cohort study. Submitted to *Fertility and Sterility*, 2009.

10. Luts J., Ojeda F., Van de Plas R., De Moor B., Van Huffel S., and Suykens J.A.K. A tutorial on support vector machine-based methods for classification problems in chemometrics. Submitted to *Analytica Chimica Acta*, 2009.

11. Pexsters A., Luts J., Van Schoubroeck D., Bottomley C., Van Calster B., Van Huffel S., D'Hooghe T., Lees C., Bourne T., and Timmerman D. Clinical implications of intra and inter-observer reproducibility of first trimester measurements performed with transvaginal ultrasound between 6 and 9 weeks gestation. Submitted to *Human Reproduction*, 2010.

12. Postma G., Luts J., Idema A.J., Julia-Sape M., Moreno-Torres A., Gajewicz W., Suykens J.A.K., Heerschap A., Van Huffel S., and Buydens L.M.C. On the relevance of automatically selected single voxel MRS and multimodal MRI and MRSI features for brain tumor differentiation. Submitted to *Computers in Biology and Medicine*, 2010.

Chapters in books

13. Lluch-Ariet M., Estanyol F., Mier M., Delgado C., Gonzalez-Velez H., Dalmas T., Robles M., Saez C., Vicente J., Van Huffel S., Luts J., Arus

C., Candiota A.P., Julia-Sape M., Peet A., Gibb A., Sun Y., Celda B., Martinez-Bisbal M.C., Valsecchi G., Dupplaw D., Hu B., and Lewis P. On the implementation of HealthAgents: Agent-based brain tumour diagnosis in *Agent-technology and e-health*, Annicchiarico R., Cortes U., and Urdiales C., pages 5−24. Birkhauser Verlag, Basel, 2007.

Papers in international conference proceedings

14. Garcia-Gomez J.M., Tortajada S., Vicente J., Saez C., Castells X., Luts J., Julia-Sape M., Juan-Ciscar A., Van Huffel S., Arus C., and Robles M. Genomics and metabolomics research of brain tumours based on machine learning. In *Proceedings of the 9th International Work-Conference on Artificial Neural Networks (IWANN)*, pages 1012−1019, San Sebastian, 2007.

15. Van Calster B., Luts J., Suykens J.A.K., Condous G., Bourne T., Timmerman D., and Van Huffel S. Comparing methods for multi-class probabilities in medical decision making using LS-SVMs and kernel logistic regression. In *Proceedings of the 17th International Conference on Artificial Neural Networks (ICANN)*, pages 139−148, Porto, 2007.

16. Laudadio T., Luts J., Martinez-Bisbal M.C., Celda B., and Van Huffel S. Differentiation between brain metastasis and glioblastoma using MRI and two-dimensional turbo spectroscopic imaging data. In *Proceedings of the 4th European Medical and Biomedical Engineering Congress (EMBEC)*, pages 1513−1516, Antwerp, 2008.

17. Luts J., Laudadio T., Martinez-Bisbal M.C., Van Cauter S., Molla E., Piquer J., Suykens J.A.K., Himmelreich U., Celda B., and Van Huffel S. Differentiation between brain metastases and glioblastoma multiforme based on MRI, MRS and MRSI. In *Proceedings of the 22nd IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1−8, Albuquerque, 2009.

Abstracts for international conferences

18. Luts J. and Van Huffel S. LS-SVM based multiclass classification of brain tumours using MRI and MRSI. In *Proceedings of the 23rd Annual Scientific Meeting of European Society for Magnetic Resonance in Medicine and Biology (ESMRMB)*, Warsaw, 2006.

19. Luts J., Poullet J.-B., Garcia-Gomez J.M., Suykens J.A.K., Heerschap A., and Van Huffel S. The effect of feature extraction for MRS classification. In

Proceedings of the 15th Scientific Meeting and Exhibition of International Society for Magnetic Resonance in Medicine (ISMRM) − European Society for Magnetic Resonance in Medicine and Biology (ESMRMB), Berlin, 2007.

20. Luts J., Idema A.J., Heerschap A., Vandermeulen D., Suykens J.A.K., and Van Huffel S. Nosologic imaging of the brain: combined MRI and MRSI segmentation and classification. In *Proceedings of the 16th Scientific Meeting and Exhibition of International Society for Magnetic Resonance in Medicine (ISMRM)*, Toronto, 2008.

21. Garcia-Gomez J.M., Luts J., Julia-Sape M., Krooshof P., Tortajada S., Vicente J., Melssen W., Fuster-Garcia E., Olier I., Postma G., Monleon D., Moreno-Torres A., Pujol J., Candiota A.-P., Martinez-Bisbal M.C., Suykens J.A.K., Buydens L.M.C., Celda B., Van Huffel S., Arus C., and Robles M. Multiproject-Multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. In *Proceedings of the 25th Annual Scientific Meeting of European Society for Magnetic Resonance in Medicine and Biology (ESMRMB)*, Valencia, 2008.

22. Luts J., Laudadio T., Martinez-Bisbal M.C., Celda B., and Van Huffel S. Differentiation between brain metastasis and glioblastoma using MRI and two-dimensional turbo spectroscopic imaging data. In *Proceedings of the 17th Scientific Meeting and Exhibition of International Society for Magnetic Resonance in Medicine (ISMRM)*, Honolulu, 2009.

23. Van den Bosch T., Luts J., Bourne T., Bignardi T., Condous G., Epstein E., Leone F., Testa A., Valentin L.I., Van Huffel S., Van Schoubroeck D., and Timmerman D. Comparison of gel instillation sonography (GIS) with unenhanced ultrasound in the diagnosis of uterine intracavity lesions. In *Proceedings of the 19th World Congress on Ultrasound in Obstetrics and Gynecology*, Hamburg, 2009.

24. Van den Bosch T., Luts J., Bourne T., Bignardi T., Condous G., Epstein E., Leone F., Testa A., Valentin L.I., Van Huffel S., Van Schoubroeck D., and Timmerman D. Prediction of intracavity uterine pathology at ultrasound examination using off-line analysis. In *Proceedings of the 19th World Congress on Ultrasound in Obstetrics and Gynecology*, Hamburg, 2009.

25. Van den Bosch T., Luts J., Bourne T., Bignardi T., Condous G., Epstein E., Leone F., Testa A., Valentin L.I., Van Huffel S., Van Schoubroeck D., and Timmerman D. The influence of gel-infusion on the vascularity of endometrial polyps. In *Proceedings of the 19th World Congress on Ultrasound in Obstetrics and Gynecology*, Hamburg, 2009.

26. Van den Bosch T., Luts J., Bourne T., Bignardi T., Condous G., Epstein E., Leone F., Testa A., Valentin L.I., Van Huffel S., Van Schoubroeck D.,

and Timmerman D. Interobserver agreement on reporting uterine intracavity lesions at gel infusion sonography (GIS). In *Proceedings of the 19th World Congress on Ultrasound in Obstetrics and Gynecology*, Hamburg, 2009.

27. Werbrouck E., Van den Bosch T., Veldman J., Luts J., Van Huffel S., Van Schoubroeck D., and Timmerman D. Saline infusion versus gel instillation sonography: a prospective cohort study. In *Proceedings of the 19th World Congress on Ultrasound in Obstetrics and Gynecology*, Hamburg, 2009.

28. Luts J., Laudadio T., Martinez-Bisbal M.C., Van Cauter S., Molla E., Piquer J., Suykens J.A.K., Himmelreich U., Van Huffel S., and Celda B. Solitary brain metastasis and glioblastoma multiforme: differentiation based on MRI and two-dimensional turbo spectroscopic imaging data. In *Proceedings of the 26th Annual Scientific Meeting of European Society for Magnetic Resonance in Medicine and Biology (ESMRMB)*, Antalya, 2009.

29. Luts J., Laudadio T., Martinez-Bisbal M.C., Van Cauter S., Molla E., Piquer J., Suykens J.A.K., Himmelreich U., Van Huffel S., and Celda B. Solitary brain metastasis and glioblastoma multiforme: differentiation based on MRI and two-dimensional turbo spectroscopic imaging data. In *Proceedings of the European Congress of Radiology (ECR)*, Vienna, 2010.

Unpublished workshop contributions

30. Luts J. and Van Huffel S. LS-SVM based multiclass classification of brain tumours using MRI and MRSI. *IAP V/22 Study Day*, Louvain-la-Neuve, 2006.

31. Luts J., Heerschap A., Suykens J.A.K., and Van Huffel S. MRI and MRSI based multiclass LS-SVM classification using class probabilities and feature selection. *International Society for Magnetic Resonance in Medicine (ISMRM) workshop on Data Processing for MR Spectroscopy and Imaging*, Warrenton, 2006.

32. Luts J., Laudadio T., Idema A.J., Heerschap A., Vandermeulen D., Suykens J.A.K., and Van Huffel S. MRI and MRSI segmentation and classification for nosologic imaging of the brain. *LKI Oncoforum*, Leuven, 2008.

33. Luts J., Laudadio T., Idema A.J., Simonetti A.W., Heerschap A., Vandermeulen D., Suykens J.A.K., and Van Huffel S. Nosologic imaging of the brain: segmentation and classification using MRI and MRSI. *HealthAgents workshop, satellite event of the 8th Congress of the European Association of Neuro-Oncology*, Barcelona, 2008.

34. Luts J., Laudadio T., Martinez-Bisbal M.C., Celda B., and Van Huffel S. Differentiation between brain metastasis and glioblastoma using MRI and two-dimensional turbo spectroscopic imaging data. *eTUMOUR workshop 'Towards brain tumour classification by molecular profiling: imaging, metabolomic and genomic tools'*, Valencia, 2008.

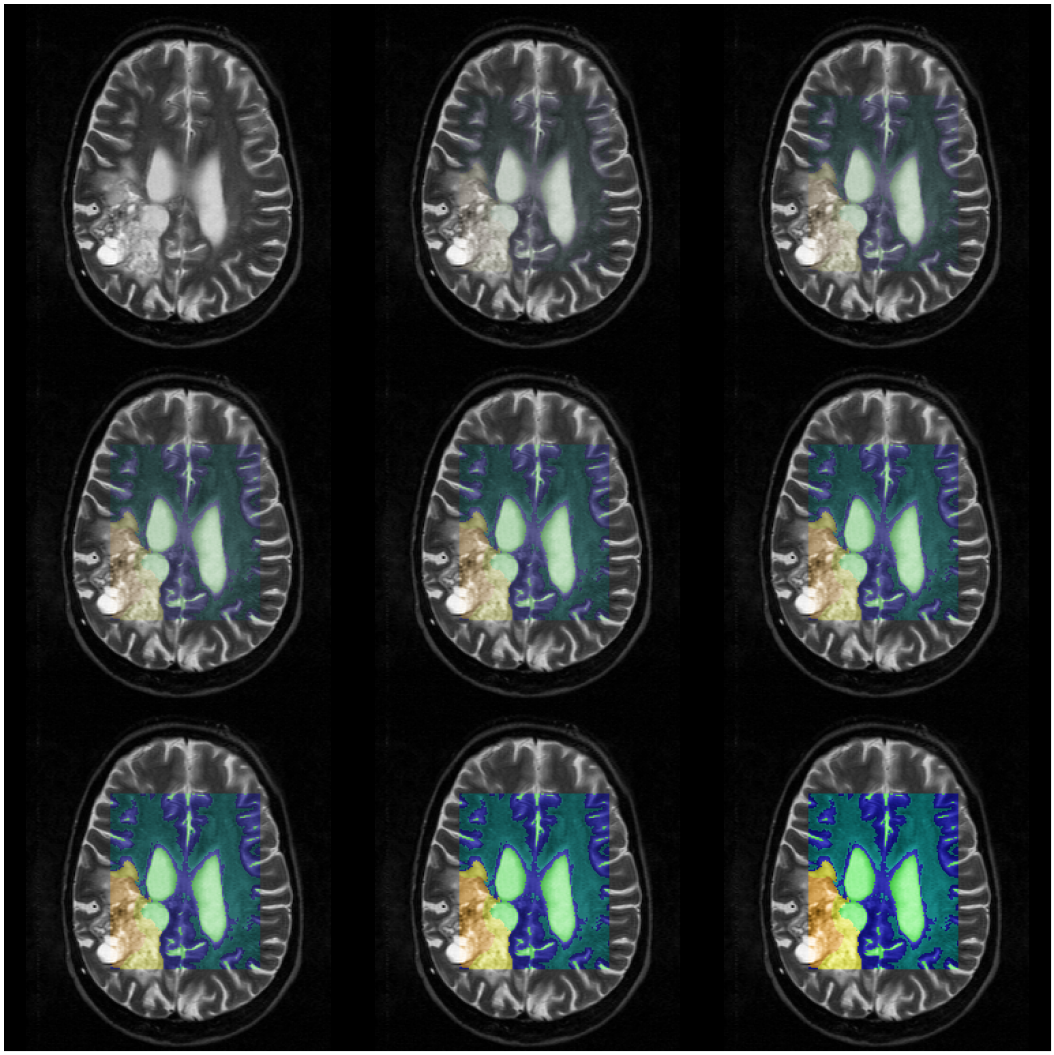
Internal reports

35. Luts J., Heerschap A., Suykens J.A.K., and Van Huffel S. MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs. Internal Report 06−143, ESAT−SISTA, Katholieke Universiteit Leuven, Belgium, 2006.

Curriculum vitae

Jan Luts was born in Diest on June 15, 1981. He received the Master of Information Sciences, option Multimedia, from Universiteit Hasselt, Belgium, in 2003. He received the Master degree in Bio-informatics and the Master degree in Statistics from Katholieke Universiteit Leuven, Belgium, in 2004 and 2005, respectively, and started in the same year the PhD program at the Department of Electrical Engineering (ESAT) of Katholieke Universiteit Leuven under the supervision of prof. dr. ir. Sabine Van Huffel and prof. dr. ir. Johan A.K. Suykens. He received a grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) in 2006. He is an invited speaker at the *International Society for Magnetic Resonance in Medicine (ISMRM) MR of Cancer Study Group Workshop 'Improving Cancer Treatment by Magnetic Resonance'*, Santa Cruz, 2010.

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Electrical Engineering (ESAT)
Research group SCD
Kasteelpark Arenberg 10, B-3001 Leuven (Belgium)



ASSOCIATIE K.U.LEUVEN