

KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

QUANTIFICATION AND CLASSIFICATION OF MAGNETIC RESONANCE SPECTROSCOPY DATA AND APPLICATIONS TO BRAIN TUMOUR RECOGNITION

Promotoren:
Prof. dr. ir. S. VAN HUFFEL
Prof. dr. P. VAN HECKE

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Andy DEVOS

Mei 2005



KATHOLIEKE UNIVERSITEIT LEUVEN
FACULTEIT TOEGEPASTE WETENSCHAPPEN
DEPARTEMENT ELEKTROTECHNIEK
Kasteelpark Arenberg 10, 3001 Leuven (Heverlee)

QUANTIFICATION AND CLASSIFICATION OF MAGNETIC RESONANCE SPECTROSCOPY DATA AND APPLICATIONS TO BRAIN TUMOUR RECOGNITION

Jury:
Prof. dr. ir. G. De Roeck, voorzitter
Prof. dr. ir. S. Van Huffel, promotor
Prof. dr. P. Van Hecke, co-promotor (Faculteit Geneeskunde)
Prof. dr. ir. J. Vandewalle
Prof. dr. ir. D. Vandermeulen
Prof. dr. ir. R. Pintelon (ELEC, VUB)
Prof. dr. L.M.C. Buydens (RU Nijmegen)
Prof. dr. C. Arús (UA Barcelona)

Proefschrift voorgedragen tot
het behalen van het doctoraat
in de toegepaste wetenschappen

door

Andy DEVOS

U.D.C. 616.831-073 Mei 2005

© Katholieke Universiteit Leuven – Faculteit Toegepaste Wetenschappen
Arenbergkasteel, B-3001 Heverlee (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag 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/2005/7515/34

ISBN 90-5682-603-4

Voorwoord

Dit werk zou nooit tot stand gekomen zijn zonder de steun en hulp van een aantal mensen. Bij het beëindigen van mijn doctoraat ben ik hen dan ook veel dank verschuldigd.

In de eerste plaats zou ik mijn promotor prof. Sabine Van Huffel willen bedanken. Ze gaf mij een unieke kans om boeiend onderzoek te verrichten binnen de Biomed-groep. Het was dankzij haar enthousiasme en haar vele inspirerende ideëen dat mijn onderzoek resultaten opbracht. Zelfs in moeilijkere tijden wist ze me te motiveren om mij te blijven inzetten voor het onderzoek. De voorbije periode onder haar begeleiding betekende een leuke en verruimende ervaring.

Mijn dank gaat uit naar mijn co-promotor prof. Paul Van Hecke. Hij was steeds bereid tijd vrij te maken om een rapport na te lezen of kritische opmerkingen te maken. En dit ondanks het feit dat hij een geruime tijd geduld heeft moeten oefenen om de lang verwachte resultaten te ontvangen.

Ik zou graag prof. Joos Vandewalle, prof. Rik Pintelon en prof. Dirk Vandermeulen willen bedanken voor het beoordelen van mijn proefschrift en het geven van waardevolle opmerkingen en suggesties. Prof. Lutgarde Buydens ben ik heel erkentelijk omdat ze bereid was deel uit te maken van mijn jury. It is a great pleasure for me that prof. Carles Arús wants to participate in my jury. Tenslotte gaat mijn dank uit naar prof. Guido De Roeck voor het waarnemen van het voorzitterschap van de jury.

Tevens gaat een enorme dank uit naar Leentje Vanhamme voor de begeleiding van mijn onderzoek. Door haar ervaring was ze de ideale persoon om me de juiste weg te wijzen in de verwerking van spectroscopische data. Ze stond altijd klaar voor vragen, discussies of om mijn programma's te doorzoeken naar fouten.

Het was een groot plezier om met Lukas samen te werken; veel van de algoritmes werden samen met hem ontwikkeld. Ik zal ook de vele toffe momenten en de ellenlange discussies nooit vergeten. Een speciale vermelding verdienen zeker Johan Suykens, René in 't Zandt, Jos de Brabanter en Philippe Lemmerling.

I had the nice opportunity to be involved in the European INTERPRET-

project. The contributions of all partners were very valuable. I especially would like to thank prof. Carles Arús and Rosemary Tate, not only for their contribution to the project, but also for their large expertise and their many valuable comments on the research. I am grateful to Franklyn Howe, Marinette van der Graaf, Angelo Moreno-Torres and Carles Majós who provided me with a large amount of crucial information. I would like to express my gratitude to Margarida Julià-Sapé as data manager and all acquisition centres for providing a valuable dataset to the project.

Het was heel aangenaam om samen te werken met de onderzoeksgroepen in Nijmegen begeleid door prof. Lutgarde Buydens en prof. Arend Heerschap. Veel dank gaat uit naar Arjan Simonetti en Marinette van der Graaf voor de lange discussies en vele terechte opmerkingen.

Niki Bergans en Tom Dresselaers van de Biomedische NMR Eenheid ben ik zeer erkentelijk voor alle informatie omtrent de glycogeen signalen en alle kritische opmerkingen ter verduidelijking van het onderzoek.

Alle collega's van SISTA en Biomed wil ik bedanken voor hun collegialiteit en de toffe werksfeer. Speciaal denk ik aan mijn bureaugenoten: Giansalvo, Nicola, Vasile, Chuan, Cynthia, Jean-Michel, Kristiaan, Lieveke, Jos, Lukas, Bart, Arjan, Bart, Jean-Baptiste, Pieter en Teresa. Een speciaal woordje van dank gaat uit naar Ida, Bart, Ilse en Pela voor het afhandelen van allerlei praktische zaken. De systeembeheerders waren steeds bereid hulp te bieden voor computer- of netwerkproblemen. Bedankt!

Tevens ben ik de K.U. Leuven en het I.W.T.-Vlaanderen erkentelijk voor de financiële steun.

Tenslotte gaat mijn grootste dank uit naar mijn familie. Mijn ouders, broers en zussen spaarden nooit enige moeite om mij te helpen. Ik kan hen dan ook niet genoeg bedanken voor hun liefde, vriendschap en steun.

Els en Mattias, met geen woorden kan ik mijn liefde en dank voldoende uitdrukken. De afgelopen maanden had ik nauwelijks tijd voor jullie, maar er volgen nu nog talrijke mooie jaren dat we samen van het leven zullen genieten!

Abstract

The medical diagnosis of brain tumours is one of the main applications of Magnetic Resonance. Magnetic Resonance consists of two main branches: Imaging and Spectroscopy. Magnetic Resonance Imaging is very well-known as the radiologic technique applied to produce high-quality images of tissues, such as the brain tissue, for diagnostic purposes. Magnetic Resonance Spectroscopy provides chemical information about all the molecules present in the brain, such as their concentrations. Both Imaging and Spectroscopy can be exploited for the grading and typing of brain tumours, also called the classification of brain tumours.

As first topic, this thesis mainly studied the contribution of Spectroscopy for automated classification and the influence of several factors on the classification performance. It was found that a few preprocessing steps did not have a large impact on the classification results. This implies that several preprocessing steps can be ignored for diagnostic purposes and that acquisition schemes can be simplified. Furthermore, classification can be based on a restricted number of extracted parameters that cover the most characteristic information, which simplifies the computation and reduces the computation time. In addition, it was observed that spectroscopic and imaging parameters can provide complementary information for the typing of brain tumours.

A second topic was the processing of experimental glycogen signals, which provides a better understanding of the biochemical processes involved in diseases such as diabetes. However, the spectral processing needs improvement. A common problem is that most methods assume the number of components present to be known exactly, but in practice this knowledge is lacking. Especially for so-called multi-exponential signals, in which different components occur at the same spectral frequency but with different linewidths. A framework was applied that determined the number of components for glycogen signals and that was able to confirm the presence of a multi-exponential pattern in glycogen signals.

Korte inhoud

De medische diagnose van hersentumoren is een van de belangrijkste toepassingen van Magnetische Resonantie. Magnetische Resonantie bestaat hoofdzakelijk uit twee takken: Beeldvorming en Spectroscopie. Magnetische Resonantie Beeldvorming is welbekend als de radiologische techniek voor het genereren van hoge-kwaliteitsbeelden van levende organismen voor het bepalen van de medische diagnose, zoals bij hersentumoren. Met Magnetische Resonantie Spectroscopie wordt chemische informatie beschikbaar gesteld over de moleculen aanwezig in de hersenen. Zowel Beeldvorming als Spectroscopie kunnen worden toegepast voor het bepalen van de graad en het type van een hersentumor, ook wel de classificatie van hersentumoren genoemd.

Een eerste onderwerp handelt over de bijdrage van Spectroscopie voor de geautomatiseerde classificatie en de invloed van een aantal factoren op de classificatieresultaten. Zo bleek dat een aantal van de voorbewerkingsstappen geen grote invloed hadden op de performantie. Indien de analyse uitgevoerd wordt voor diagnostische doeleinden, betekent dit dat verschillende voorbewerkingsstappen genegeerd kunnen worden en dat de meetprocedure vereenvoudigd kan worden. Bovendien kan classificatie gebaseerd zijn op een beperkt aantal parameters die de meest relevante informatie bevatten, wat de berekening vereenvoudigt en versnelt. Resultaten toonden eveneens aan dat informatie uit Spectroscopie en Beeldvorming elkaar kunnen aanvullen voor de typering van hersentumoren.

Als tweede onderwerp werd de analyse van experimentele glycogeen signalen bestudeerd. Door de studie van glycogeensignalen kunnen de biochemische processen die een rol spelen bij ziektes zoals diabetes beter begrepen worden. Dit vereist echter een verbetering van de spectrale analyse. Een veel voorkomend probleem is het gebrek aan kennis over het aantal spectrale componenten, terwijl dit belangrijk is voor een nauwkeurige spectrale analyse. Dit treedt typisch op bij zogenaamde multi-exponentiële signalen, met verschillende componenten op dezelfde frequentie maar bij een verschillende lijnbreedte. In dit proefschrift werd een methode toegepast voor het bepalen van het aantal componenten die in staat was om het multi-exponentiële patroon bij glycogeensignalen te bevestigen.

Glossary

This section lists symbols and acronyms that occur frequently in this thesis.

Remark. The notation used in this thesis allows one to distinguish among scalars, vectors and matrices. Lower case characters represent scalar values. Boldface lower case characters are used for vectors. Boldface capitals represent matrices.

Mathematical Notation

a	scalar a
$\mathbf{v}$	vector $\mathbf{v}$
$\mathbf{M}$	matrix $\mathbf{M}$
$\mathbf{M}^T$	transpose of matrix $\mathbf{M}$
$\mathbf{M}^{-1}$	inverse of matrix $\mathbf{M}$
$\mathbf{M}^H$	complex conjugate of matrix $\mathbf{M}$
$\mathbf{M}^\dagger$	pseudo–inverse of matrix $\mathbf{M}$
$\det(\mathbf{M})$	determinant of matrix $\mathbf{M}$
$\mathbf{M}(m,n)$	element on the m–th row and n–th column of matrix $\mathbf{M}$
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	set of real n–dimensional vectors
$\hat{x}$	estimate of x
$\hat{\mathbf{A}}$	estimate of matrix $\mathbf{A}$
$\lvert x\rvert$	absolute value of x
$\lvert\lvert\mathbf{x}\rvert\rvert_2$	L_2–norm of $\mathbf{x}$
$a \ll b$	a is much smaller than b
$a \gg b$	a is much larger than b
$a \approx b$	a is approximately equal to b
$\text{sign}(x)$	sign of x
$\log(x)$	natural logarithm of x
$\text{trace}(\mathbf{A})$	trace of the matrix $\mathbf{A}$

$K(\cdot,\cdot)$	kernel function

Fixed Symbols

σ'	shielding constant
σ_e^2	error or noise variance
σ	kernel width of the RBF kernel
B_0	static magnetic field
T_1	spin-lattice relaxation time constant
T_2	spin-spin relaxation time constant
t_n	the n-th time point
N	number of time points or datapoints
y_n	the n-th measured signal sample
$\bar{y}(\cdot)$	time-domain model function
a_k	amplitude of resonance k
d_k	damping of resonance k
f_k	frequency of resonance k
ϕ_k	phase of resonance k
ϱ_k	nonlinear parameter vector of resonance k for MRS quantification
c_k	complex amplitude of resonance k
$\mathbf{\Gamma}$	matrix with the nonlinear part of the model function
$\mathbf{H}$	Hankel matrix
M_i	model with i components at the resonance frequency of glycogen
θ	parameter vector
$d_{eff,i}$	effective number of parameters of model M_i
C_p	Mallows' C_p
α	Lagrange multiplier
γ	regularization constant (chapter 3)
C	regularization constant for support vector machines
Ω	feature space
$\mathcal{D}$	original input space

Acronyms and Abbreviations

aggr	aggressive tumours
AIC	Akaike's Information Criterion
AIC_c (or AICC)	bias-corrected version of AIC
Ala	alanine
AMARES	Advanced Method for Accurate, Robust, and Efficient Spectral fitting
astroII	astrocytomas of grade II
ATP	adenosine triphosphate
AUC	area under the ROC curve
BIC	Bayesian Information Criterion
CDVC	Clinical Data Validation Committee of the INTERPRET-project
Cho	choline
CNS	Central Nervous System
Cr	creatine
CSF	cerebrospinal fluid
CT	Computed Tomography or Computerized Tomography
FFT	Fast Fourier Transform
FID	Free Induction Decay
FOV	Field Of View
GAIC	Generalized AIC (nonlinear models)
GAICC	Generalized AICC (nonlinear models)
GBIC	Generalized BIC (nonlinear models)
GD	Gadolinium enhanced
GE	General Electric
GIClog	GPE with $\lambda = \log(\log(N))$
glio	glioblastomas
Glu	glutamate
Gln	glutamine
Glx	glutamine and glutamate
Gly	glycine
GLY	glycogen
GPE	Generalized Prediction Error (nonlinear models)
HLSVD	Hankel Lanczos Singular Value Decomposition
HSVD	Hankel Singular Value Decomposition
HTLS	Hankel Total Least Squares
INTERPRET	International Network for Pattern Recognition of Tumours using Magnetic Resonance (IST-1999-10310)
K-L	Kullback-Leibler distance
Lac	lactate
LDA	Linear Discriminant Analysis
Lin	linear

LOOCV	leave-one-out cross-validation
LS	least squares
LS-SVM	Least Squares Support Vector Machine
meni	meningiomas
meta	metastases
mI	myo-Inositol
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MRSI	Magnetic Resonance Spectroscopic Imaging
NAA	N-acetyl-aspartate
NAc	N-acetyl containing compounds
NLLS	nonlinear least squares
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
PCr	phosphocreatine
P_i	inorganic phosphate
PD	proton density
PET	Proton Emission Tomography
ppm	parts per million
PRESS	Point RESolved Spectroscopy
QP	Quadratic Programming
RBF	Radial Basis Function
RF	radio frequency
ROC	Receiver Operating Characteristic
SE	standard error on the AUC
SNR	signal-to-noise ratio
SPECT	Single Photon Emission Computerized Tomography
STEAM	STimulated Echo Acquisition Mode
SVD	Singular Value Decomposition
SVM	Support Vector Machine
T	Tesla
Tau	taurine
TE	echo time
TLS	total least squares
TR	repetition time
VARPRO	VARiable PROjection
WHO	World Health Organisation

Contents

Voorwoord **i**

Abstract **iii**

Korte inhoud **v**

Glossary **vii**

Contents **xi**

Samenvatting **xvii**

1 Introduction **1**

1.1 Basic concepts of Magnetic Resonance 1

1.1.1 Magnetic Resonance Spectroscopy 2

1.1.2 Magnetic Resonance Imaging 4

1.1.3 Magnetic Resonance Spectroscopy and Spectroscopic Imaging 4

1.1.4 Relaxation 6

1.1.5 Applications of MRS 7

1.2 NMR for brain tumour diagnosis 8

1.2.1 Brain tumours 9

1.2.2 NMR versus other medical imaging techniques 9
1.2.3 Characteristic information available in medical NMR . . 10
1.2.4 MRS characteristics of brain tumour types 12
1.3 Goals of the thesis 13
1.4 Outline of the thesis 15
1.5 Contributions and research partners 18
1.6 Conclusions 20

2 Quantification of MRS signals **21**
2.1 Modelling of an MRS signal 21
2.2 Prior knowledge 22
2.3 Quantification methods 23
2.3.1 Frequency-domain methods 24
2.3.2 Time-domain methods 25
2.4 Model selection 28
2.4.1 Problem statement 28
2.4.2 Model selection criteria 28
2.5 Conclusion 30

3 Pattern recognition of brain tumours using MR data **33**
3.1 Pattern recognition 33
3.1.1 What is meant by pattern recognition? 33
3.1.2 The importance of feature selection 34
3.2 The development of a pattern recognition system 35
3.2.1 Data acquisition and collection 35
3.2.2 Preprocessing of NMR signals 37
3.2.3 The construction and testing of a classifier 41

3.2.4 Performance measures 42
3.3 Pattern recognition techniques 47
3.3.1 Unsupervised techniques 47
3.3.2 Supervised techniques 49
3.4 Conclusions 60

4 Classification of brain tumours using long echo time ^{1}H MRS 63
4.1 Introduction 64
4.2 Material 65
4.3 Methods 70
4.3.1 Input features 70
4.3.2 Classification techniques 70
4.3.3 Experimental setting 71
4.4 Results 74
4.4.1 Classification using complete spectra 74
4.4.2 Classification using selected frequency regions 74
4.4.3 Classification using peak integration 74
4.4.4 Multiclass approach 75
4.5 Discussion 77
4.5.1 Limitations 77
4.5.2 Glioblastomas versus metastases 79
4.5.3 Classification techniques 79
4.5.4 Influence of dimensionality reduction 80
4.5.5 Multiclass classification 81
4.6 Conclusions 86

5 Classification of brain tumours using short echo time ^{1}H MRS 95

5.1 Introduction 96
5.2 Material 96
5.3 Methods 98
5.3.1 Input features 98
5.3.2 Experimental setting 100
5.4 Results 100
5.4.1 Binary classification 100
5.4.2 Multiclass approach 101
5.5 Discussion 101
5.5.1 Classification techniques 104
5.5.2 Influence of input features 105
5.5.3 Influence of dimensionality reduction 106
5.5.4 Multiclass classification 106
5.5.5 Clinical practice 106
5.6 Conclusions 107

6 The use of multivariate MR Imaging intensities versus metabolic data from MR Spectroscopic Imaging for brain tumour classification 111

6.1 Introduction 112
6.2 Material 113
6.3 Methods 115
6.3.1 Input features 115
6.3.2 Experimental approach 120
6.4 Results 121
6.5 Discussion 126
6.5.1 Classification techniques 126
6.5.2 Imaging intensities versus metabolic data from MRSI . 127

6.5.3 Several classification problems 128

6.6 Conclusions 128

7 Model selection for quantification of multi-exponential MR spectra 131

7.1 Introduction 132

7.2 Material 133

7.2.1 Glycogen synthesis in perfused rat liver 133

7.2.2 Data 134

7.3 Methods 134

7.3.1 Prior knowledge 135

7.3.2 Model selection for MRS quantification 136

7.3.3 Application on Monte-Carlo simulations 137

7.3.4 Application on experimental signals 139

7.4 Results 140

7.4.1 Monte-Carlo simulations 140

7.4.2 Experimental signals 143

7.5 Discussion 145

7.5.1 Evaluation of the model selection criteria 145

7.5.2 Model selection applied to quantification of ^{13}C glycogen signals 146

7.6 Conclusions 147

8 Conclusions and suggestions for future research 155

8.1 Summary 155

8.2 Future research 157

Appendix 159

A Medical background about brain tumours 159

B Medical terminology concerning brain tumours 160

Bibliography **164**

Publication list **179**

Biography **182**

Kwantificatie en classificatie van Magnetische Resonantie Spectroscopie data en toepassingen voor de patroonherkenning van hersentumoren

Nederlandse samenvatting

Hoofdstuk 1: Inleiding

Magnetische Resonantie (MR) is een techniek die frequent toegepast wordt om klinische beelden te maken van het menselijk lichaam (Magnetische Resonantie Beeldvorming of MRB). De MR beelden leveren informatie over de structuur en de heterogeniteit van het onderzochte weefsel of orgaan en dit op een niet-invasieve manier, namelijk zonder de huid binnen te dringen. Daarnaast is het mogelijk om concentraties van chemische stoffen of metabolieten te bepalen (Magnetische Resonantie Spectroscopie of MRS). MRS signalen bevatten metabolische informatie die beschouwd wordt als een soort vingerafdruk nuttig voor het typeren van bijvoorbeeld tumoren en neurologische ziekten. Oorspronkelijk werd met MRS maar voor één voxel of volume-element een signaal opgemeten. Een techniek die de ruimtelijke en spectroscopische informatie combineert is Magnetische Resonantie Spectroscopische Beeldvorming (MRSB). Deze spec-

troscopische techniek meet signalen op van een matrix van voxels en maakt het mogelijk om een ruimtelijke spreiding van de metabolische informatie te visualiseren.

De kern van sommige atomen, zoals proton (^{1}H) en carbon (^{13}C), bezit een zogenaamde "spin", te visualiseren als een rotatie van de kern om eigen as, zoals een tol. In een statisch magneetveld kunnen de kernen van bijvoorbeeld waterstof zich in twee toestanden bevinden; georiënteerd met of tegen het magnetisch veld in. De kernen hebben een lichte voorkeur voor de oriëntatie volgens het magneetveld omwille van de lagere energietoestand. Deze evenwichtstoestand kan verstoord worden door het aanleggen van een korte puls met een bepaalde energie waardoor het systeem van spins geëxciteerd wordt. Onmiddellijk na de puls relaxeren de spins terug naar de evenwichtstoestand en zenden magnetische golven uit met dezelfde energie als de golven gebruikt voor excitatie. Deze golven genereren dan het MR signaal, wiskundig te beschrijven als een exponentieel in de tijd afnemende sinusoïde.

Proton MRS wordt veelvuldig gebruikt voor de karakterisatie van hersentumoren. In tumorweefsel zijn de concentraties van een aantal metabolieten toe- of afgenomen ten opzichte van deze in gezond weefsel. Zo zijn de concentraties van *NAA* (N-acetyl-aspartaat) en *Cr* (creatine) afgenomen en deze van *Cho* (choline) toegenomen. Elk van deze stoffen heeft een bepaalde functie in het metabolisme (of de stofwisseling) van de hersenen; *NAA* wordt gebruikt bij de neurale activiteit en een afname van *NAA* wordt verbonden met neuraal verlies of neurale schade. Een toename van *Cho* kan veroorzaakt worden door een hogere celdichtheid en membraanmetabolisme en een versnelde tumorgroei. *Cr* speelt een rol bij het energiemetabolisme, waarbij een verlaagd *Cr*-gehalte verklaard kan worden door de aanwezigheid van een kwaadaardige tumor met een hoge metabolische activiteit die de energievoorraden opgebruikt. Verschillende van deze metabolietconcentraties kunnen bepaald worden met proton MRS. Proton signalen kunnen worden onderverdeeld volgens de waarde van de echotijd (TE), een belangrijke acquisitieparameter. De echotijd is de tijd tussen het aanleggen van de puls en het opmeten van het MR signaal. Zogenaamde lange echotijd proton MR signalen bevatten minder componenten en zijn gemakkelijk te verwerken, maar leveren in sommige gevallen onvoldoende informatie om goede diagnoses voor patiënten te stellen. Korte echotijd proton MR signalen daarentegen bevatten meer informatie dan de lange echotijd MRS signalen, maar zijn echter moeilijker te analyseren.

Met het oog op het gebruik van MRS in een klinische omgeving is het noodzakelijk een betrouwbaar medisch beslissingssysteem te ontwikkelen voor het bepalen van het type en de graad van een hersentumor. De doelstelling van deze thesis is om een aantal aspecten te bestuderen en een bijdrage te leveren tot de ontwikkeling van een betrouwbaar beslissingssysteem:

- We willen nagaan of een aantal voorbewerkingsstappen nodig zijn om de

interessante informatie aanwezig in MRS signalen te versterken en om een hoge performantie te bekomen.

- Een belangrijke vraag is of MRS en MRB elkaar aanvullen voor de karakterisatie van tumoren of het gebruik van één van beide voldoende is.
- Intuitief wordt verwacht dat een grotere hoeveelheid aan informatie de differentiatie van tumoren ten goede komt. Een groter aantal parameters verhoogt ook de dimensie en daarmee ook de rekencomplexiteit. Dit benadrukt de bijdrage van dimensiereductie die tracht uit een grote hoeveelheid kenmerken de meest informatieve parameters te extraheren.
- Tenslotte onderzoeken we welke techniek het meest aangewezen is voor de classificatie van hersentumoren op basis van MR data.

Een tweede onderwerp betreft de wiskundige analyse en parameterbepaling van MRS data, ook wel MRS kwantificatie genoemd. Vele kwantificatiemethodes veronderstellen dat de modelorde of het aantal aanwezige componenten gekend is. In experimentele situaties is dit echter meestal niet het geval. Het voorgestelde werk heeft als doel een methode toe te passen voor de correcte bepaling van de modelorde bij experimentele ^{13}C MRS signalen.

Hoofdstuk 2: Kwantificatie van MRS signalen

Hoofdstuk 2 beschrijft een aantal wiskundige technieken die gebruikt worden voor kwantificatie en modelselectie van MRS signalen. Het MRS signaal wordt vaak gemodelleerd als een som van Lorentzianen

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} (a_k e^{j\phi_k})\, e^{(-d_k + j2\pi f_k)t_n} + e_n, \quad n = 0, 1, ..., N-1, \quad (0.1)$$

met K het aantal verschillende componenten. De parameters van het signaal leveren rechtstreekse informatie over elk aanwezig metaboliet: elke spectrale component wordt geidentificeerd door de frequentie f_k, de demping d_k wordt bepaald door de mobiliteit van het metaboliet, de amplitude a_k (intensiteit) is rechtstreeks evenredig met de concentratie van het metaboliet en ϕ_k is de fase. De reële en imaginaire delen van de ruis e_n worden verondersteld ongecorreleerd te zijn en gelijke varianties te hebben.

Voorkennis is een belangrijk concept voor de nauwkeurige gegevensverwerking van MRS data. Bepaalde modelparameters zijn namelijk met elkaar verbonden. Zo kan via kwantummechanische principes berekend worden dat de frequenties en amplitudes van twee componenten van sommige metabolietspectra verbonden zijn. De opgelegde modelorde is eveneens een voorbeeld van voorkennis.

Het gebruik van voorkennis reduceert het aantal onafhankelijk te schatten variabelen en leidt tot een verhoogde nauwkeurigheid en flexibiliteit.

Kwantificatie kan uitgevoerd worden in het tijds- of frequentiedomein. Een eenvoudige frequentiedomeinmethode is piekintegratie die de amplitude van elke metaboliet schat door de oppervlakte onder de piek in het frequentiedomein te bepalen. Deze methode maakt geen modelveronderstellingen, is heel robuust, maar heeft een lage nauwkeurigheid.

Tijdsdomeinmethodes kunnen ingedeeld worden in zwarte-doos en interactieve methodes. Zwarte-doos methodes vereisen maar een minimale gebruikersinteractie, wat de gebruikersonafhankelijkheid verhoogt. Het opleggen van voorkennis is maar in beperkte mate mogelijk en de schattingen zijn suboptimaal.

De interactieve methodes minimaliseren het kwadratisch verschil tussen de data en de modelfunctie

$$||\mathbf{y} - \bar{\mathbf{y}}||^2. \tag{0.2}$$

Om het niet-lineaire kleinste kwadratenprobleem op te lossen wordt de modelfunctie opgesplitst in een lineair en een niet-lineair deel. Eerst worden de niet-lineaire parameters f_k en $d_k, k = 1, \ldots, K$ bepaald, waarbij K de opgelegde modelorde voorstelt. Daarna worden de lineaire parameters berekend als de oplossing van een lineair kleinste kwadratenprobleem. Het bekomen model kan dan voorgesteld worden in termen van de gegeven data:

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}, \tag{0.3}$$

waarbij de matrix $\mathbf{S}$ de "hoedmatrix" wordt genoemd en afhankelijk is van de geschatte modelparameters. De interactieve methodes zijn heel flexibel en laten toe om heel eenvoudig voorkennis over het model in te bouwen. Het belangrijkste nadeel is de vereiste interactie met de gebruiker. Voor elk signaal is het immers nodig om startwaarden te bepalen die de convergentie van de methode versnellen. De methode AMARES veronderstelt, zoals de meeste methodes, dat de modelorde of het aantal componenten exact gekend is. Voor experimentele signalen is dit echter meestal niet het geval.

Een nauwkeurige schatting van de modelorde kan uitgevoerd worden door een criterium uit de informatietheorie. Een veralgemeende criterium toepasbaar voor niet-lineaire modellen M_i, uit een verzameling van modellen $\mathcal{A}$, wordt gegeven door:

$$\frac{RSS_i}{N} + \lambda \frac{d_{eff,i}}{N} \hat{\sigma}^2, \tag{0.4}$$

waarbij RSS_i de residuele som van de kwadraten is en N het aantal datapunten. Het effectieve aantal parameters wordt voorgesteld door $d_{eff,i}$ en $\hat{\sigma_e}^2$ is een onvertekende schatter voor de foutvariantie σ_e^2. De optimale modelorde komt dan overeen met de laagste criteriumwaarde.

Afhankelijk van de waarde van λ kunnen andere criteria gedefinieerd worden, zoals het veralgemeende Akaike criterium (GAIC, $\lambda = 2$), veralgemeende Bayes criterium (GBIC, $\lambda = \log(N)$) en GIClog ($\lambda = \log(\log(N))$). Eveneens werd een verbeterde versie van Akaike's criterium voorgesteld, die de vertekening van Akaike's criterium verkleint. Voor niet-lineaire modellen wordt de veralgemeende versie van dit criterium genoteerd als GAICC.

De informatiecriteria kunnen toegepast worden voor niet-lineaire modellen waarbij het aantal effectieve parameters bepaald kan worden via de "hoedmatrix" (Vergelijking (0.3)). Bovendien is de optimale modelorde snel te bekomen via een eenvoudige formule.

Hoofdstuk 3: Patroonherkenning van hersentumoren op basis van MR data

Patroonherkenning wordt beschouwd als een proces waarbij levende wezens of computers een patroon of een aantal karakteristieken herkennen en relateren aan een bepaald object, idee of concept. Zo heeft elk object een aantal eigenschappen die patroonherkenning of classificatie mogelijk maakt, een bal bijvoorbeeld heeft een ronde vorm, heeft een bepaalde kleur en kan gebruikt worden als speelgoed.

Verschillende aspecten van patroonherkenning spelen een belangrijke rol bij het proces voor de ontwikkeling van een classificatiesysteem en meer specifiek voor het automatisch herkennen van hersentumoren. We beschouwen zowel het opmeten en de voorbewerking van de data, als het toepassen van de classificatietechniek en het bepalen van performantie.

Voorbewerking

De voorbewerking houdt onder meer de volgende bewerkingen in:

- Fasering. Het MRS signaal bestaat uit een reëel en een imaginair signaal, die 90° uit fase staan tegenover elkaar, wat een eventuele faseverschuiving kan veroorzaken. Fasering is nodig om het reële signaal correct voor te stellen, maar fasering van experimentele signalen is moeilijk automatiseerbaar. Een alternatief is om magnitude spectra te bepalen, wat het signaal fase-onafhankelijk maakt.
- Normalisatie. Spectra, opgenomen op verschillende scanners of onder een verschillend meetprotocol, kunnen een verschillend patroon bevatten door effecten onafhankelijk van het onderzochte weefseltype. Een normalisatiemethode tracht deze effecten op te vangen zodat spectra met

elkaar vergelijkbaar zouden zijn. Normalisatie kan bijvoorbeeld uitgevoerd worden volgens een wiskundige norm, zoals de kwadratische norm of L_2-norm. Een andere veel gebruikte werkwijze is de normalisatie door deling met de intensiteit van een bepaald molecule, zoals *Cr* of water.

- Correctie voor de basislijn. Korte echotijd metingen bevatten sterk overlappende brede componenten, macromoleculen genoemd en het zijn deze brede pieken die een belangrijke bijdrage leveren aan een deels onbekende en gegolfde basislijn. Vele methodes zijn beschikbaar die het signaal corrigeren voor de basislijn. Veelal zijn deze methodes moeilijk te evalueren, aangezien de basislijn sterk afhankelijk kan zijn van de pathologische conditie en dus niet geweten is wat dé basislijn is.

Patroonherkenning kan gebaseerd zijn op verschillende geformaliseerde eigenschappen of parameters; zoals het volledige spectrum of enkele geëxtraheerde parameters. De doelstelling van een extractiemethode is om de relevante parameters te weerhouden. Dit zorgt eveneens voor een lagere invoerdimensie en vereenvoudigt de berekening. Discriminerende parameters bij lange echotijd omvatten onder andere resonanties van vetten, *Lac* (lactaat), *Ala* (alanine), *NAA*, *Cr*, *Cho* en *Gly* (glycine). Bij korte echotijd kunnen eveneens resonanties beschouwd worden die bij lange echotijd minder zichtbaar zijn, zoals *Glx* (glutamate/glutamine), *mI* (myo-inositol) en *Tau* (taurine).

Een veel gebruikte extractiemethode in MRS maakt gebruik van amplitudeschattingen, voor de hierboven vermelde resonanties, bekomen via een eenvoudige kwantificatiemethode zoals piekintegratie. Daarnaast kunnen ook spectrale parameters geselecteerd worden uit de interessante frequentieregios.

Classificatietechnieken

Classificatietechnieken worden hoofdzakelijk ingedeeld in gesuperviseerde en ongesuperviseerde technieken, afhankelijk van hun benadering tot het leerproces. Gesuperviseerde technieken hebben een virtuele "leraar" ter beschikking tijdens een training- of leerfase om een classificator te construeren. De bekomen classificator kan dan getest worden tijdens een zogenaamde testfase.

Ongesuperviseerde technieken daarentegen hebben geen afzonderlijke training- en testfase. Een klassieke ongesuperviseerde techniek is lineaire Principale Component Analyse (PCA), wat veelal wordt toegepast voor parameterextractie om bepaalde parameters te selecteren. De techniek transformeert de oorspronkelijke parameters naar een orthogonale ruimte, waarbij de parameters een maximale variantie verklaren in de data. De parameters die de meeste variantie verklaren worden weerhouden om zo goed mogelijk de oorspronkelijke data te beschrijven.

De volgende gesuperviseerde technieken toegepast in dit proefschrift worden ook beschreven:

- Lineaire Discriminant Analyse (LDA) projecteert de data $x_k \in \mathbb{R}^n$ van de oorspronkelijke dataruimte op een variabele $z_k \in \mathbb{R}$ in een ééndimensionale ruimte. De discriminatie gebeurt dan op basis van de geprojecteerde variabele z_k. Deze methode tracht, gegeven meerdere klassen, de varianties tussen de klassen te maximaliseren en de varianties binnen één klasse te minimaliseren.

- Support Vector Machines (SVM) is een techniek die de laatste jaren reeds voor vele classificatietoepassingen met succes werd uitgevoerd. De oorspronkelijke invoerdata $x_k \in \mathbb{R}^n$ worden getransformeerd van de invoerruimte naar een hoog-dimensionale kenmerkenruimte. Verschillende kernfuncties kunnen gebruikt worden voor deze transformatie; zoals lineaire- en Gaussiaanse ("Radial Basis Functions" of RBF) kernfuncties. In de kenmerkenruimte wordt een hypervlak geconstrueerd dat de gegeven klassen scheidt. Dit hypervlak is oplossing van een kwadratisch programmeringsprobleem, waarbij aan ongelijkheidsbeperkingen moet voldaan zijn. Classificatie wordt uitgevoerd in de invoerruimte door het hypervlak terug te transformeren naar de invoerruimte.

- Least Squares SVM (LS-SVM) is een kleinste kwadraten versie van SVM, waarbij het optimalisatieprobleem wordt geherformuleerd als een lineair programmeringsprobleem. De ongelijkheidsbeperkingen worden eveneens vervangen door gelijkheidsbeperkingen, wat de berekening van de oplossing aanzienlijk vereenvoudigt en versnelt.

Performantiematen

Het is niet alleen belangrijk om classificatoren te ontwikkelen, maar ook om ze te evalueren of anders gezegd om hun performantie te bepalen. Performantiematen leveren een schatting van de nauwkeurigheid van de classificatieresultaten, wat toelaat verschillende technieken met elkaar te vergelijken. De performantie kan bijvoorbeeld bekomen worden met de analyse van de Ontvanger Operatie Characteristiek (OOC), veelal gebruikt in de medische wereld om diagnostische methodes objectief te beoordelen. De methodes in kwestie worden vergeleken op basis van de oppervlakte onder de OOC curve (OOOC), wat een globaal beeld geeft van de klinische efficiëntie over de volledige OOC curve.

Hoofdstuk 4: Classificatie van hersentumoren op basis van lange echotijd ^{1}H MRS

Een betrouwbaar idee over het type en de graad van de hersentumor bepaalt grotendeels de te volgen therapie voor de patiënt. Er is dan ook nood aan een betere classificatie van tumoren. De laatste jaren werd reeds aangetoond dat MRS, als nauwkeurige, niet-invasieve techniek, ons gedetailleerde chemische informatie kan geven over metabolieten aanwezig in de eventuele hersentumor.

Methodes

Op basis van MRS signalen werden verschillende classificatietechnieken vergeleken voor de discriminatie van hersentumoren op basis van de lange echotijd signalen afkomstig van het Europese INTERPRET-project (IST-1999-10310). Een databank van 183 lange echotijd signalen opgemeten in 4 centra was beschikbaar, waarbij 4 veel voorkomende types hersentumoren werden beschouwd: glioblastoma's, meningioma's, metastases en astrocytoma's van graad II.

Invoerparameters

Als invoerparameters werden de volgende drie mogelijkheden beschouwd:

1. volledige spectra. Het L_2-genormaliseerde magnitude spectrum werd beschouwd in het interessante spectrale gebied.
2. geselecteerde spectrale gebieden. Uit de magnitude spectra werden specifieke spectrale gebieden geselecteerd, met de informatie over de belangrijkste metabolieten: *Cho*, *Cr*, *NAA*, *Lac*, *Ala* en vetten.
3. piekintegratie waarden. De amplitude van elk geselecteerd metaboliet werd berekend over een specifiek spectraal gebied in het magnitude spectrum. Pieken van dezelfde metabolieten als hierboven werden beschouwd.

Experimentele opstelling

Binaire classificatie levert inzicht in de distributies van elke klasse en hun onderlinge overlap. Gegeven de 4 types hersentumoren, werden 6 binaire classificatoren opgesteld die alle klassen paarsgewijs scheidden. Als lineaire technieken werden LDA, SVM en LS-SVM met lineaire kernfuncties gebruikt, terwijl LS-SVM met een RBF-kernfunctie als niet-lineaire techniek werd toegepast. LDA werd toegepast op de parameters geselecteerd door PCA, behalve bij de piekintegratie waarden omdat hierbij reeds een sterke dimensiereductie werd uitgevoerd.

Het experiment werd uitgevoerd door willekeurige splitsing van de data in een training databank (2/3 van de data) en een test databank (overige 1/3). De performantie van de binaire classificatoren werd geschat aan de hand van de oppervlakte onder de OOC curve (OOOC). Deze splitsing werd 200 keer herhaald en de gemiddelde OOOC en de gepoolde standaardfout werden berekend.

Classificatie van meerdere klassen

De medische praktijk vereist de ontwikkeling van classificatoren die meerdere klassen ineens kunnen beschouwen. In dit geval werd een nieuw spectrum aanzien als één van de vier tumortypes en zou dus tot één hiervan moeten worden toegekend door de classificator.

Het classificatieprobleem met 4 klassen werd geherformuleerd als een combinatie van binaire classificatoren, waarbij in twee stappen werd gewerkt. Aangezien glioblastoma's en metastases erg moeilijk scheidbaar zijn, werden deze tumortypes samengevoegd in een klasse van aggressieve tumoren. In een eerste stap werden de binaire classificatoren gecombineerd die meningioma's, astrocytoma's van graad II en aggressieve tumoren paarsgewijs van elkaar scheidden. In een tweede stap werden de aggressieve tumoren ingedeeld in glioblastoma's en metastases.

Resultaten

Deze uitgebreide vergelijkende studie op basis van lange echotijd proton MRS signalen toonde aan dat verschillende tumortypes paarsgewijs scheidbaar zijn met MRS. Behalve voor het ingewikkelde binaire probleem glioblastoma's versus metastases, bereiken binaire classificatoren op basis van de volledige spectra een gemiddelde OOOC hoger dan 0.9. Gebruik makende van de beschikbare hoeveelheid data, werden geen statistisch significante verschillen geobserveerd tussen de classificatietechnieken, zowel voor binaire classificatie als classificatie met meerdere klassen. Bij binaire classificatie is de classificatietechniek met de hoogste performantie afhankelijk van zowel het type invoerdata als de beschouwde klassen.

Dimensiereductie werd verondersteld de storende factoren in het spectrum te reduceren. Niettemin resulteerde de beschreven aanpak in een lagere performantie. Hieruit blijkt dat eveneens waardevolle informatie, aanwezig in de niet-geselecteerde spectrale gebieden, gereduceerd werd. Deze informatie kan belangrijk zijn voor het discrimineren van de tumortypes. Meer specifiek denken we aan resonanties die prominenter voorkomen bij korte echotijd, zoals *mI*, *Glu* en *Gln*. Extra discriminerende informatie afkomstig van een groter aantal metabolieten zou de performantie kunnen verhogen. Dit motiveert verder onderzoek naar classificatie op basis van korte echotijd proton MRS signalen, aangezien daar meerdere metabolieten aanwezig zijn.

Hoofdstuk 5: Classificatie van hersentumoren op basis van korte echotijd ^{1}H MRS

Deze studie onderzocht de toepassing van verschillende technieken en het effect van een aantal voorbewerkingsstappen op de performantie. Terwijl vorig hoofdstuk gebruik maakte van lange echotijd data, betreft het hier een toepassing op korte echotijd data, die inherent meer informatie bevatten.

Methodes

De korte echotijd databank van het Europese INTERPRET-project (IST-1999-10310) bevatte 205 signalen uit 6 centra, waarbij dezelfde types hersentumoren werden beschouwd als in de vorige studie: glioblastoma's, meningioma's, metastases en astrocytoma's van graad II. De indeling volgens tumortype, de selectieprocedure en de voorbewerking verliepen op dezelfde wijze als bij de studie over lange echotijd MRS.

De invloed van fasering (magnitude versus reële spectra), normalisatie (L_2-versus water normalisatie) en de basislijn (correctie voor de basislijn) op de classificatieperformantie werd onderzocht. Classificatie werd uitgevoerd met de lineaire technieken LDA en LS-SVM met een lineaire kernfunctie en vergeleken met LS-SVM met een RBF-kernfunctie. De invloed van dimensiereductie werd eveneens nagegaan op een analoge wijze als in de vorige studie. Het gebruik van volledige spectra werd vergeleken met het gebruik van geselecteerde spectrale gebieden en piekintegratie waarden. In deze studie werden echter een aantal relevante resonanties bijgevoegd, overeenstemmend met metabolieten die duidelijker aanwezig zijn bij korte echotijd.

Resultaten

Het gebruik van LDA, als eenvoudige lineaire techniek, resulteerde in een goede discriminatie van hersentumoren. De techniek leverde bijvoorbeeld op de L_2-genormaliseerde volledige magnitude spectra zonder correctie voor de basislijn een gemiddelde OOOC van minstens 0.95, behalve voor de scheiding van glioblastoma's en metastases. Geen significante verschillen tussen classificatietechnieken werden gedetecteerd. De performanties waren gelijkaardig aan deze in de vorige studie, alhoewel dit niet rechtstreeks statistisch werd vergeleken.

Faseringsproblemen optredend bij reële spectra kunnen worden vermeden door het gebruik van magnitude spectra. De berekening van magnitude spectra resulteert echter in bredere pieken en een grotere piekoverlap, wat de spectrale informatie kan verstoren. Uit onze studie werd geen significante invloed van de fasering op de classificatieperformantie gedetecteerd.

Het patroon van korte echotijd spectra wordt sterk beinvloed door de aanwezigheid van de basislijn. Correctie voor de basislijn was verondersteld te resulteren in een duidelijker piekpatroon en een hogere performantie. Een verbetering werd echter niet gevonden, wat verklaard kan worden door de mogelijke, maar onduidelijke bijdrage van het patroon in de basislijn tot de biomedische karakterisatie van tumoren.

Spectrale normalisatie is bedoeld voor de consistente schaling van metingen. Het vergelijken van twee eenvoudige normalisatiemethodes resulteerde in enkele significant slechtere resultaten voor water normalisatie tegenover L_2-normalisatie, vooral bij het gebruik van lineaire technieken.

Dimensiereductie, zoals met het selecteren van spectrale gebieden of piekintegratie, kan leiden tot een lagere complexiteit van het probleem en een versnelde berekening. In vergelijking met volledige spectra, werd op basis van geselecteerde spectrale gebieden en piekintegratie waarden een gelijkwaardige performantie bekomen, wat aanduidt dat de geëxtraheerde parameters de belangrijkste informatie bevatten.

Hoofdstuk 6: Het gebruik van multivariate MR Beeldvormingsintensiteiten versus metabolische data van MR Spectroscopische Beeldvorming voor classificatie van hersentumoren

In de klinische praktijk wordt de diagnose van hersentumoren in grote mate bepaald door de morfologische en ruimtelijke informatie beschikbaar in een medisch beeld, zoals bekomen met MRB. Het rechtstreeks gebruik van MRB informatie in een patroonherkenningssysteem zou dan leiden tot een meer betrouwbare classificatie. In deze studie werden intensiteiten afkomstig van MRB en metabolische data van MRSB zowel afzonderlijk als in combinatie met elkaar gebruikt voor de scheiding van hersentumoren. Alhoewel de ruimtelijke informatie niet werd beschouwd, werd reeds een goed idee verkregen van de relevante bijdrage van beide types invoerparameters.

Methodes

Een databank van 25 patiënten met een hersentumor en 4 gezonde vrijwilligers werd ontwikkeld in het kader van het Europese INTERPRET-project (IST-1999-10310), opgenomen in het Universitair Medisch Centrum Nijmegen (UMCN) te Nijmegen, Nederland. De databank bevatte zowel MR beelden als korte echotijd MRSB signalen. De MR beelden werden opgenomen bij vier ver-

schillende contrasten en de positie van de MRSB snede en van de MR beelden werd gekozen zodat de opeenvolgende beelden afkomstig waren van dezelfde ruimtelijke locatie. De databank bevatte 4 tumortypes, namelijk 10 glioma's van graad II, 4 van graad III en 7 van graad IV en 3 meningioma's. Voor elke patiënt werden zoveel mogelijk volume-elementen uit het gebied van de tumor geselecteerd. Daarnaast waren ook data uit niet-tumor gebieden beschouwd; namelijk cerebraal spinale vloeistof (CSV) van patiënten, als gezond weefsel van vrijwilligers en patiënten. Data met dezelfde pathologie werden gecombineerd, zodat 7 groepen werden bekomen.

Classificatie werd uitgevoerd op de volledige spectra (231 variabelen), piekintegratie waarden (10 variabelen), beeldintensiteiten (4 variabelen) en een combinatie van de laatste twee parametertypes (14 variabelen). Een analoge experimentele opstelling werd gebruikt als bij de classificatiestudie over korte echotijd data, inclusief technieken, indeling training- en testfase en performantiebepaling. Dit werd toegepast om de volgende binaire problemen te evalueren, waarbij tumoren van graad II laaggradig en tumoren van graad III of IV hooggradig waren en alle meningioma's tot de laaggradige tumoren behoorden:

- gezond versus tumorweefsel.
- laag- versus hooggradige tumoren.
- laag- versus hooggradige glioma's.
- glioma's versus meningioma's.
- glioma's van graad II versus van graad III.

Resultaten

LDA bereikte voor de meeste problemen een gelijkwaardige performantie als de kernfunctie-gebaseerde lineaire LS-SVM. Voor het onderscheid tussen laag- en hooggradige glioma's was de performantie van LDA significant lager dan deze van LS-SVMs. Eveneens de lineaire LS-SVM presteerde niet steeds even goed als de niet-lineaire LS-SVM, zoals bij laag- en hooggradige tumoren.

Alhoewel MRB een courant toegepaste techniek is voor de diagnose van tumoren, werd met MRB intensiteiten voor enkele binaire problemen een lagere performantie bereikt dan met metabolische informatie, zoals bij laag- en hooggradige tumoren. Op basis van metabolische data werd een gelijkaardige performantie bekomen als in de twee vorige studies. Het gecombineerd gebruik van MRS en MRB intensiteiten verbeterde de classificatieresultaten, zoals voor de scheiding van laag- en hooggradige glioma's. We merken echter op dat in de klinische praktijk veel meer informatie wordt gebruikt om een diagnose op te stellen. Dit betreft zowel anatomische informatie, zoals de locatie en de homogeniteit van de tumor, als klinische informatie, zoals klachten van de patiënt.

Hoofdstuk 7: Model selectie voor kwantificatie van multi-exponentiële MR spectra

Kwantificatie wordt veelal gebruikt voor fundamenteel onderzoek op basis van de bekomen parameterschattingen. In vele gevallen heeft de kwantificatie van MRS signalen dan belangrijke implicaties op de interpretatie van biologische of biomedische processen. Zo wordt er verondersteld dat het aantal componenten aanwezig in het signaal op voorhand gekend is, wat veelal niet het geval is. Een typisch voorbeeld hiervan is een signaal met meerdere componenten op dezelfde frequentie maar met verschillende dempingen, een multi-exponentieel signaal genoemd. Multi-exponentiële signalen werden reeds geobserveerd in ^{13}C glycogeensignalen tijdens een zogenaamd "pulse-chase" experiment op een geperfuseerde rattenlever. Het experiment bestaat hoofdzakelijk uit 2 "pulse" fasen, met tussenin een "chase" fase. Tijdens de "pulse" fasen wordt er ^{13}C-gemerkt glucose aangediend, dat omgezet wordt in ^{13}C-gemerkt glycogeen (GLY). Dit maakt het mogelijk om glycogeen te detecteren met ^{13}C MRS. Tijdens de "chase" fase wordt ongemerkt glucose gegeven, zodat de vorming van glycogeen verder gezet wordt. De doelstelling was om een methode toe te passen die de modelorde van glycogeensignalen betrouwbaar kan bepalen.

Methodes

Kwantificatie werd uitgevoerd met AMARES, op zowel gesimuleerde als experimentele ^{13}C glycogeensignalen. Elk van de signalen werd gemodelleerd door een model met 1, 2 of 3 componenten voor de glycogeenpiek, bij dezelfde resonantie, maar met een verschillende demping en respectievelijk genoteerd als M_1, M_2 en M_3. De modelselectie criteria GAIC,GBIC, GIClog en GAICC werden toegepast op gesimuleerde signalen om te onderzoeken welk criterium het best geschikt was voor de beoogde doelstelling. Bij gesimuleerde signalen met 2 GLY-pieken werd de invloed van de verschillende parameterwaarden bekeken door resultaten bij een groot en klein verschil tussen de dempingen/amplitudes van de twee pieken met elkaar te vergelijken. Gaussiaanse ruis van verschillende ruisniveaus werd aan de simulatiesignalen toegevoegd.

Het optimale criterium werd daarna toegepast op GLY-signalen van gevaste en gevoede ratten, bij verschillende SRVs en bij de 3 experimentele fasen ("chase" fase en beide "pulse" fasen).

Resultaten

Alle informatiecriteria leverden gelijkaardige resultaten, waarbij GBIC eerder de voorkeur gaf aan een lagere modelorde en het minst goed presteerde bij de detectie van multi-exponentiële signalen. GAIC, GAICC en GIClog daaren-

tegen selecteerden gemakkelijker een hogere modelorde, waarbij GAICC iets beter presteerde bij de analyse van signalen met maar één GLY-piek. Bij toenemende SRV was het eenvoudiger om het correcte aantal pieken te detecteren, waarbij de minimaal vereiste SRV voor het observeren van een bi-exponentieel signaal afhangt van de relatieve parameterwaarden. Bij een gunstige situatie waarbij de twee GLY-pieken een even grote intensiteit hadden, was een SRV vereist van slechts 16dB.

Met het criterium GAICC was het mogelijk om verscheidene observaties van vorige studies te versterken. Signalen van de "chase" en de tweede "pulse" fasen van het experiment bij gevaste ratten werden best benaderd als de superpositie van een smalle en een brede piek. Bovendien gaf het tri-exponentiële model eveneens een goede benadering, wat duidde op de aanwezigheid van zelfs drie GLY-pieken. Een bi-exponentieel patroon werd eveneens geobserveerd voor signalen van de "chase" en de tweede "pulse"bij gevoede ratten, terwijl voor de eerste "pulse" fase in geen enkel geval een indicatie werd gevonden voor het bestaan van meer dan één GLY-piek.

Daarnaast werd ook geëxperimenteerd met het toevoegen van een aantal losse beperkingen op de dempingen van de verschillende GLY-componenten, dit door de demping waarden te laten aannemen in een bepaald interval. Hiermee werd convergentie van een aantal modellen voor ogen gehouden die zonder deze beperkingen niet convergeerden. In deze studie werd onder deze beperkingen convergentie bekomen, maar werden echter veelal schattingen voor de dempingen bekomen die zich dichtbij één van de grenzen van het opgelegde interval bevonden. Dit wijst erop dat het gebruik van losse beperkingen nog niet optimaal werd uitgebouwd.

Hoofdstuk 8: Besluiten en nieuwe uitdagingen

Besluiten

- Voor korte echotijd signalen werd de invloed van een aantal voorbewerkingsstappen op de classificatieresultaten onderzocht. Het gebruik van magnitude of reële spectra, met of zonder basislijncorrectie, had geen significante invloed op de performantie. Spectrale normalisatie bleek wel een duidelijke invloed te hebben, aangezien normalisatie ten opzichte van de intensiteit van water slechter presteerde dan een eenvoudige kwadratische normalisatie. Dit wijst erop dat, indien veralgemeenbaar, moeilijke voorbewerkingsstappen als fasering, correctie voor de basislijn en normalisatie overgeslaan kunnen worden met het oog op de diagnose van hersentumoren. Eveneens impliceert dit dat het opmeten van een water referentiesignaal niet meer cruciaal is en het meetprotocol vereenvoudigd kan worden.

- Het toepassen van dimensiereductie tracht de relevante parameters te extraheren en de invloed van ruis te onderdrukken. Op basis van korte echotijd MRS en MRSB signalen werden gelijkaardige resultaten bekomen voor piekintegratie waarden als voor volledige spectra, ondanks het feit dat piekintegratie een eenvoudige en onnauwkeurige kwantificatiemethode is. Echter op basis van lange echotijd MRS signalen werd een lagere performantie bekomen voor piekintegratie waarden. Dit verschil is mogelijks te wijten aan het feit dat bij korte echotijd spectra een grotere hoeveelheid informatie beschikbaar is over een groter aantal variabelen, wat kan leiden tot een betere scheiding.

- Het gecombineerd gebruik van intensiteiten bekomen via MRB en MRSB verhoogde de performantie in vergelijking tot het gebruik van enkel MRB intensiteiten. Dit versterkt de idee dat spectroscopie en beeldvorming complementaire informatie kunnen leveren voor de karakterisatie van hersenweefsel en -tumoren.

- Alhoewel in de studies over lange en korte echotijd MRS PCA voorafgaande aan LDA enkel 75% van de variantie verklaart, bereikt de klassieke LDA veelal een hoge performantie. Voor enkele problemen, zoals de scheiding tussen laag- en hooggradige tumoren, werd een (gedeeltelijk) niet-lineair patroon vastgesteld.

- Criteria uit de informatietheorie gaven veelbelovende resultaten voor het bepalen van de modelorde uit een verzameling van mogelijke waarden. Dit werd aangetoond zowel via gesimuleerde als experimentele glycogeensignalen. Bovendien werd het multi-exponentiëel gedrag bij verschillende experimentele glycogeensignalen bevestigd.

- Het algoritme AMARES bereikte niet altijd een convergentie van het model zonder het opleggen van bepaalde bijkomende beperkingen. Het is dan ook aan te raden experimentele MRS signalen met aandacht te analyseren.

Nieuwe uitdagingen

- Het gebruik van ruimtelijke informatie. In dit proefschrift werden MRB intensiteiten gebruikt waarbij gegevens van elk voxel afzonderlijk werden beschouwd. Dit impliceert het verlies van belangrijke morfologische en ruimtelijke informatie, wat een grote rol speelt bij de diagnose in de klinische praktijk uitgevoerd door expert-radiologen. Een vergelijk van MRB en MRSB met alle beschikbare informatie zou dan rechtstreeks een indicatie geven over hun relatieve bijdrage tot de karakterisatie van tumoren.

- Combinatie van verscheidene MR technieken. Naast conventionele MRB, werden er de laatste jaren verschillende nieuwe veelbelovende modaliteiten ontwikkeld zoals diffusie- en perfusie-gewogen MRB en functionele

MRB. Elk van deze technieken bevat nuttige informatie voor de medische diagnose van tumoren. De informatie is echter verspreid. Het is dan niet evident om een duidelijk overzicht van de informatie te behouden, wat het rechtstreeks bepalen van een betrouwbare diagnose bemoeilijkt. Het automatisch bepalen van een optimale combinatie van allerlei MR technieken zou een realistische oplossing zijn.

- Bijdrage MR parameters. Het vorige punt verhoogt logischerwijze de belangrijkheid van dimensiereductie. Aangezien meer informatie wordt beschouwd, is het des te belangrijker de bijdrage van elke parameter te bepalen en enkel de meest relevante parameters te behouden. De integratie van een statistisch kader zou hierin een grote stap vooruit kunnen betekenen.

- Binaire classificatie en classificatie met meerdere klassen. Alhoewel in dit proefschrift vooral aandacht werd geschonken aan binaire classificatie, zou de implementatie van een betrouwbaar systeem met meerdere klassen klinisch een veel grotere draagwijdte hebben. De medische diagnose vergt namelijk een zo nauwkeurig mogelijke specificatie van het type en de graad van de tumor.

- Kwantificatie voorafgaande aan classificatie. Amplitudes werden geschat met behulp van de eenvoudige piekintegratie methode. Verder onderzoek zou moeten uitwijzen of het vooraf toepassen van modelgebaseerde kwantificatie al dan niet tot een hogere classificatieperformantie zou leiden.

- Convergentie voor multi-exponentiële modellen. Convergentieproblemen traden op voor verscheidene signalen bij benadering door multi-exponentiële modellen. Het gebruik van losse beperkingen kan hierbij helpen en is een mogelijk onderwerp voor verder onderzoek.

- Bepaling van de modelorde voorafgaand aan kwantificatie. In dit proefschrift werden eerst een aantal modellen bepaald, waaruit dan het optimale model werd geselecteerd met behulp van een informatiecriterium. Deze aanpak verhoogt echter de rekentijd naarmate meer modellen worden beschouwd, wat de techniek in praktijk niet steeds bruikbaar maakt. De bepaling van de modelorde voorafgaand aan of geintegreerd in de kwantificatiemethode zou tot een duidelijke verbetering leiden.

- Betrouwbaarheidsintervallen voor de geschatte parameters. Voor onderzoeksdoeleinden is het belangrijk dat de kwantificatie zo nauwkeurig en betrouwbaar mogelijk wordt uitgevoerd. Met het oog op de evaluatie van de kwantificatieresultaten, is de uitwerking van een betrouwhaarheidsinterval rond de parameters van groot belang. De methode dient specifiek toepasbaar te zijn voor multi-exponentiële signalen.

Chapter 1

Introduction

The purpose of this chapter is to provide the necessary background information about the main aspects of Magnetic Resonance (MR), including the basic terminology used further in this thesis. Section 1.1 explains the principles of MR and how the technique is used to produce MR Spectroscopy (MRS) as well as Magnetic Resonance Imaging (MRI) data. The relevance of spectroscopy for biomedical clinical applications is illustrated. Section 1.2 provides background information about brain tumours and focuses on the potential of MR for the diagnosis of brain tumours. Further, several challenges are mentioned and the structure of the thesis is given. The chapter closes with an overview of the contributions.

1.1 Basic concepts of Magnetic Resonance

Nuclear Magnetic Resonance (NMR), or Magnetic Resonance (MR) in short, has become an extensively used technique to obtain clinical images and to study metabolic processes non-invasively. NMR consists of two main branches: Magnetic Resonance Imaging (MRI) and Magnetic Resonance Spectroscopy (MRS). However, both techniques have many aspects in common and have evolved from the same traditional spectroscopic technique. The following paragraphs introduce the basic principles of MRS and MRI and give an overview of the applications.

1.1.1 Magnetic Resonance Spectroscopy

A compass needle normally lines up with the Earth's magnetic field with the north-seeking end pointing north. Suppose the needle could be twisted around with the fingers so that it is pointed south, lining it up opposed to the Earth's magnetic field. As soon as it is let free again, the needle will flip back pointing north, as it is very unstable opposed to the Earth's field. The same principle is encountered during an NMR experiment.

Certain atomic nuclei, like ^{1}H (proton), ^{31}P (phosphorus) and ^{13}C (carbon), possess an inherent angular momentum or spin. This can be visualized as a spinning (i.e. rotational) motion of the nucleus around its own axis (Figure 1.1). These nuclei with a spin behave just like little magnets. The nuclei can either be aligned along an external static magnetic field or opposed to it. The alignment where it is opposed to the field is less stable, as this requires a higher energy state. It is possible to make it flip from the more stable alignment to the less stable one by supplying exactly the right amount of energy.

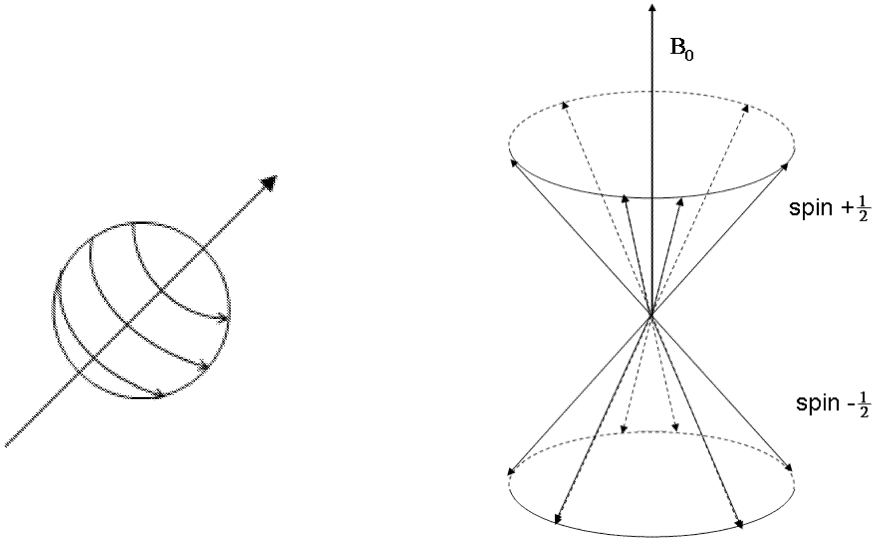

Figure 1.1: The spin of a nucleus can be visualized as a rotational motion around its own axis, corresponding to the axis of its magnetic moment (left). When applying a static magnetic field B_0, the spins are precessing around B_0. The spin can be aligned along or opposite the magnetic field (right).

^{1}H, ^{13}C or ^{31}P all have a spin quantum number of $\frac{1}{2}$, corresponding to two different spin or energy states and the energy difference between both states is proportional to the applied magnetic field

$$\Delta E = \gamma \hbar B_0,\ \hbar = h/(2\pi), \tag{1.1}$$

with γ the gyromagnetic ratio, characteristic of the isotope and h the Planck constant. To induce transitions between these energy states, an oscillating magnetic field is applied by a coil in a plane perpendicular to the static field B_0. The magnetic moment does not align exactly with the axis of the external magnetic field but precesses around it at a rate given by the Larmor relationship

$$f_0 = \gamma B_0/(2\pi). \tag{1.2}$$

The amount of energy absorbed by the nuclei (and emitted in a later stage) depends on the population difference between the two energy states. The population density of the nuclei in the different energy states is determined by their Boltzmann distribution. In thermal equilibrium at absolute temperature T, the relative amounts n^- and n^+ of nuclei in the highest (spin -$\frac{1}{2}$) and the lowest (spin +$\frac{1}{2}$) energy states are given by

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

where k is the Boltzmann constant. The ratio n^-/n^+ is very close to unity in normal circumstances, resulting in a low absorption of energy. This explains the inherently low signal-to-noise ratio (SNR) of the signal emitted when the spin system returns to equilibrium after excitation. The emitted signal is called a free induction decay (FID) signal and corresponds to an exponentially decaying sinusoid in the time domain.

The intensity of the signal is proportional to the number of nuclei that contribute to it. In order to increase the SNR, a number of data acquisitions is performed consecutively and the final signal is the average of all measured signals. As shown in Equations (1.1) and (1.3), increasing the field strength B_0 results in a higher energy difference between both states. This will increase the amount of emitted energy and therefore gives signals with higher SNR. Decreasing the temperature T will also increase the SNR, but this is not feasible for experiments on human beings.

Nuclei of different atoms can easily be distinguished from one another since they have a different γ and consequently differ widely in resonance frequency. Nuclei of the same isotope which are not chemically equivalent, are separable in frequency due to the effect of shielding. The amount of shielding, experienced by a nucleus, determines the effective magnetic field,

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

where σ' is the shielding constant, a dimensionless unit that depends on the electrical environment of a nucleus. Therefore, nuclei in a different chemical

environment emit signals with different frequencies. This phenomenon makes MRS a very attractive tool since it allows to differentiate between molecular structures. Spin-spin (or J-) coupling, the interactions between neighbouring spins in the same molecule, induces a further differentiation in resonance frequencies. The separation of resonance frequencies from an arbitrarily chosen reference frequency is termed the chemical shift and is expressed in terms of the dimensionless units of parts per million (ppm).

In its simplest form an MRS experiment consists of applying a radio frequency (RF) pulse with a bandwidth chosen to excite all the nuclei within the required frequency range and acquiring the FID signal as the response. This process is repeated until the desired SNR is obtained.

1.1.2 Magnetic Resonance Imaging

The most well-known feature of the NMR principle has been Magnetic Resonance Imaging (MRI), which is utilized extensively in the medical radiology field to produce images of tissues in the human body, such as the brain. MRI visualizes the spatial distribution of protons and provides morphological information about the examined tissue or organ.

Using the NMR principle directly as described in section 1.1.1, the acquired signal would correspond to the global response of all spins around the coil. The main feature that is still needed here to construct images of the distribution of water molecules is the identification of the spatial origin of the different signal components. Lauterbur (1973) solved this problem by introducing gradients in the magnetic field. This corresponds to a linear variation in the magnetic field as a function of the distance, while the spins are precessing around the magnetic field. The idea introduces a different strength of the magnetic field for each point inside the magnet.

The frequency of each spin is different and characteristic for its position in the magnet, which enabled the development of MRI and the introduction of the MR scanner in clinical radiologic imaging. MRI acquires high-resolution images from any direction without replacing the patient, even without any danger of radiation or ionization.

1.1.3 Magnetic Resonance Spectroscopy and Spectroscopic Imaging

The concept of the spatial localization also enabled the development of localized spectroscopy for biomedical purposes. Magnetic Resonance Spectroscopy (MRS) or single voxel spectroscopy (Figure 1.2) acquires one signal from a certain volume-element, also called voxel, while Magnetic Resonance Spectroscopic Imaging (MRSI) or multi voxel spectroscopy acquires simultaneously

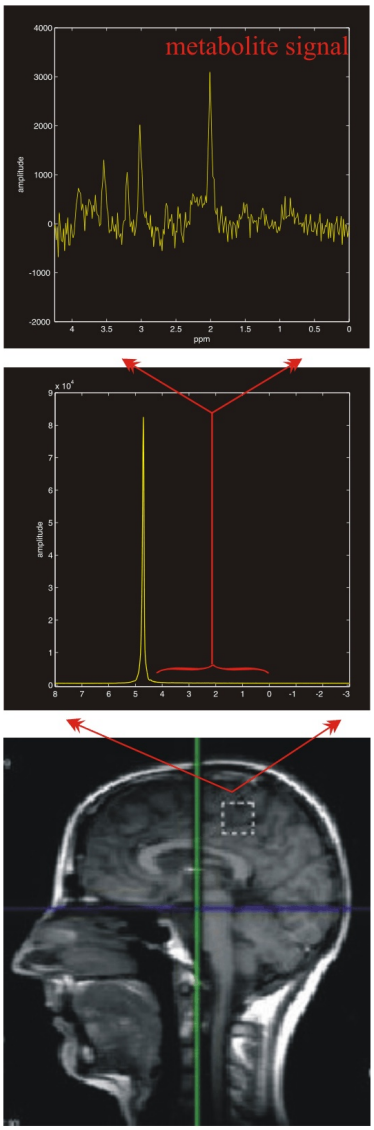

Figure 1.2: Single voxel spectroscopy. Spatial localization techniques enable the selection of a small volume-element or voxel (bottom) to acquire a spectrum of the selected voxel (middle). The large peak (middle) corresponds to water, while the relevant resonance peaks (top) are more than 1000 times smaller and are situated at the right hand side of the water peak.

signals from a twodimensional grid of voxels. MRSI can facilitate the identification of heterogeneity of a tumourous region, since spatial variations of the tissue characteristics can be assessed at metabolite level. For each of the voxels, the intensity of the biochemically relevant metabolites can be determined. This enables the development of metabolic imaging (Figure 1.3), which constructs a spatial distribution or map of the intensity of any relevant metabolite over the selected brain region, while MRI only visualizes the spatial distribution of water. Therefore, MRSI has a large potential for clinical applications as it provides spectroscopic as well as spatial information of the brain tissue.

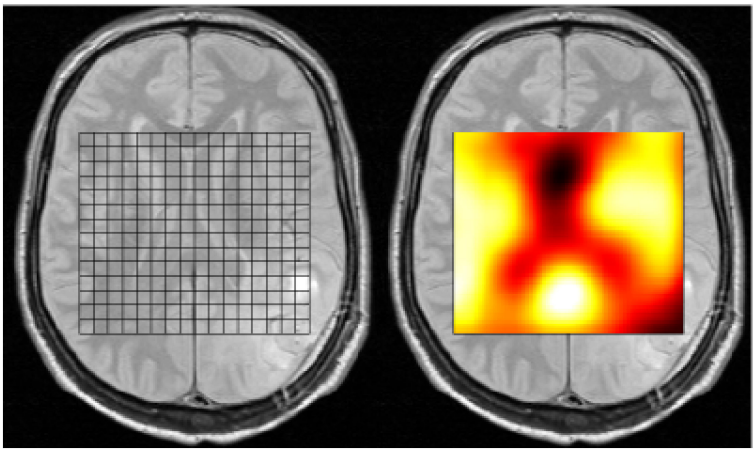

Figure 1.3: MR enables the spectroscopic visualization of the whole brain. The left part shows a MR image (water content) of 256 $\times$ 256, while the right figure displays an MRS metabolite image (NAA; N-acetyl aspartate) of 16 $\times$ 16.

1.1.4 Relaxation

As mentioned in section 1.1.1, an MR signal is acquired by exploiting the magnetic properties of nuclei. The signal intensity and appearance of MR spectra and images can be influenced by so-called relaxation processes. Suppose a sample containing nuclei of spin $\frac{1}{2}$ is placed in a magnetic field B_0, the nuclear magnets take one of the two spin orientations with respect to the field (Figure 1.4), with a slight preference for the orientation of lower energy. The sample becomes slightly magnetized producing a net magnetic moment along the direction of the field. The radiofrequency field B_1 is applied perpendicular to B_0, which causes the magnetization to be perturbed (Figure 1.4). After perturbation, the nuclear magnetization returns to its initial equilibrium state by relaxation processes, which are characterized by the biological parameters T_1 and T_2. These parameters are tissue-specific time constants, introducing

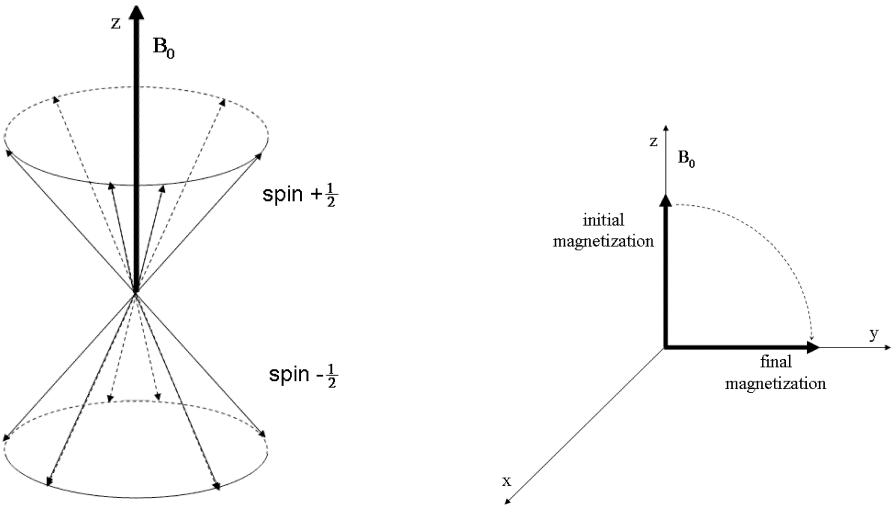

Figure 1.4: In the static situation, the net magnetization component (thick arrow) is along the static magnetic field B_0 (left). Applying the field B_1 tilts the magnetization into the xy-plane (right).

the possibility to separate different tissue types. T_1, called the spin-lattice relaxation time, is the time constant for the recovery of magnetization along the direction of B_0, where as T_2, called the spin-spin relaxation time, is the time constant for the relaxation process in the plane perpendicular to B_0. T_1 and T_2 are related to the physical interaction phenomena in the tissue.

1.1.5 Applications of MRS

MRS measures signals of chemical substances, called metabolites, and is applied for fundamental research as well as clinical studies. MRS can be applied on biological systems containing atoms that are sensitive to the NMR principle. MRS applied to living animals and human beings is called *in vivo* MRS. The measured signals typically have a low signal-to-noise ratio (SNR) due to the limited measurement time and the low metabolite concentrations. *Ex vivo* signals are obtained from an organ taken out of its natural environment and artificially kept alive using perfusion. *Ex vivo* signals usually have a better quality as they can be obtained at higher field strength. In biological research, MRS can also be applied on fluids, tissue extracts, cells. Such signals are called *in vitro* MRS signals. Since there are no limitations on the measurement time and the field strength, a high SNR can be achieved.

Proton MRS is the most commonly applied technique for *in vivo* medicine. The

technique is often applied to obtain metabolic information for clinical diagnosis of patient suspected to have lesion and for follow-up of treated patients. Medical application areas include brain tumours (e.g. Nelson, 2003; Leclerc et al., 2002) and prostate tumours (e.g. Jung et al., 2004; Pickett et al., 2004) neurologic distortions such as epilepsy (e.g. Cohen-Gadol et al., 2004; Hammen et al., 2003), AIDS (e.g. Yiannoutsos et al., 2004), disease of Alzheimer (e.g. Elderkin-Thompson et al., 2004). For brain studies, excellent field homogeneity can often be obtained, such that even at relatively low field strengths of 1.5-3.0 T adequate spectral resolution can be achieved for many brain signals. Additionally, the normal human brain generates little signal from lipids (fat) and this facilitates the detection of resonances, that overlap in frequency with lipids. However, the detection of lipids would already indicate the presence of a lesion.

The natural abundance of ^{13}C (carbon) is very low and ^{13}C has a much lower intrinsic NMR sensitivity than ^{1}H. Therefore, ^{13}C MRS can not directly observe atoms. However, specific metabolic pathways can be investigated by the use of ^{13}C-labelling, which introduces ^{13}C as a tracer to follow intermediary metabolism. This technique is employed to investigate the glycogen metabolism in diabetes patients (Bergans, 2003).

MRS can be applied on several other nuclei such as ^{31}P (phosphorus) and ^{19}F (fluorine). ^{31}P is less sensitive than ^{1}H, but includes dominant signals of ATP, *PCr* (phosphocreatine) and P_i (inorganic phosphate), which all play a central role in the energy metabolism. ^{31}P MRS is often used for studies of the skeletal muscles, the energy metabolism and pH measurements (Gadian, 1995).

^{19}F has a high sensitivity and a high natural abundance. As there are no endogenous compounds that give detectable *in vivo* ^{19}F signals, ^{19}F MRS is used to monitor exogenous compounds. Applications include the study of blood flow and oxygenation, the monitoring of drug metabolism and the study of anaesthetics (de Graaf, 1998).

1.2 NMR for brain tumour diagnosis

In this section the use of medical imaging techniques is discussed for brain tumour diagnosis and several characteristics of brain tumours are mentioned. Proton is the most sensitive nucleus and is present in many molecules, which makes ^{1}H NMR very attractive for medical examinations.

1.2.1 Brain tumours

Yearly, about 29,000 persons are diagnosed with a brain tumour in Europe (1998 estimates) (CaMon-project, 1998), a rate of 7 in every 100,000. Accurate diagnosis of these tumours is desirable. Prognosis largely depends on the tumour type and grade of malignancy; ranging from a median survival of 7-10 years for tumours of grade II, 2-5 years for grade III tumours and less than 1 year for grade IV tumours (Aldape et al., 2003; Leibel et al., 1994; Prados, 1991).

A few of the most common adult brain tumours are briefly mentioned. Some more medical background is given in the Appendices A and B. The most common tumour type is secondary or metastatic brain tumours, which is cancer that spreads from other parts of the body to the brain. They occur in 10-15% of people with cancer. Primary brain tumours do not metastasize to other parts of the body and are classified according to the cells of origin. Gliomas represent about half of all primary brain tumours and grow from glial cells (The Brain Tumor Society, 2005), which are the supporting cells of our nervous system. There exist several types of supporting cells, e.g. astrocytes and oligodendrocytes. Depending on the type of glial cell, the tumour is given another name; astrocytomas develop from astrocytes, oligodendrogliomas from oligodendrocytes and mixed gliomas relate to more than one type of glial cell.

Pathologists use a grading system for astrocytomas (Kernohan et al., 1949), ranging from the least aggressive, grade I, to the most aggressive, Grade IV. The four types of astrocytomas are pilocytic astrocytomas (Grade I), astrocytoma (Grade II), anaplastic astrocytoma (Grade III) and glioblastoma multiforme (Grade IV). Low-grade astrocytomas tend to grow slowly and are usually localized in one part of the brain. Higher-grade astrocytomas may invade surrounding tissue, grow quickly, and usually require more aggressive treatment. Throughout this thesis, grade I and II tumours will be considered to be low-grade, while grade III and IV will be regarded as high-grade.

Meningiomas are the most common type of primary brain tumours (The Brain Tumor Society, 2005). Meningiomas develop from meninges, the thin, protective membranes that cover the brain and spinal cord. They usually grow slowly, generally do not invade surrounding normal tissue, and rarely spread to other parts of the central nervous system or body.

1.2.2 NMR versus other medical imaging techniques

Patients with symptoms of a brain tumour will be sent to a neuroscience unit for investigation, which involves a number of specialized tests to diagnose brain tumours. A neurological examination investigates the performance of all brain functions, like eye movement, sensory-motoric and muscular responses, reflexes,

hearing, tactile and sensory sensation, movement, balance and coordination. Furthermore, medical imaging techniques can be applied to detect and localize a possible abnormal brain mass.

Computed Tomography (CT) uses X-ray radiation to construct detailed images of the brain from different angles. It can be performed with or without injecting a contrast agent. CT scans provide low levels of radiation, but can be regulated to provide the minimum amount of radiation exposure needed to produce the image.

Proton Emission Tomography (PET) and Single Photon Emission Computerized Tomography (SPECT) images are obtained by measuring local levels of injected contrast agents, such as glucose or methionine, that have been labelled with a radioactive tracer. SPECT is based on the principle of blood flow in the brain, which is directly related to brain activity. Areas of increased blood flow take up more radioactive tracer than areas of less blood flow. PET assesses changes in the function, circulation and metabolism of body organs and has emerged as a method of measuring body function and guiding disease treatment.

MRI applies a magnetic field to create high resolution images of the brain and will be discussed further in the next paragraphs. This technique has several advantages as it does not use X-rays nor any other type of "ionizing" radiation and has a high potential to accurately distinguish between healthy and abnormal tissue. MRI in combination with MR Spectroscopic Imaging (MRSI) is able to provide anatomical as well as chemical information.

1.2.3 Characteristic information available in medical NMR

MRI has become one of the most important non-invasive aids in clinical diagnosis of brain tumours, because it enables the radiologist to assess the anatomical location, morphology and size of the tumour tissue with a relatively high spatial resolution. Normally, stacked images are recorded throughout the total volume of the brain in order to localize the tumour optimally. MR images with different contrasts are recorded, as the acquisition of only one type of image may not show the desired information. Different contrasts are generated in MRI by exploiting T_1- and T_2-dependent effects. More specifically, there exist variations in the relaxation constants T_1 and T_2 of water, firstly between different tissues and secondly between normal and abnormal tissue. The contrast is obtained with specific acquisition parameters, resulting in T_1-weighted, T_2-weighted and proton density weighted images. Tissues with a short T_1 will appear bright on a T_1-weighted image, while tissues with a long T_1 will yield a low signal appearing dark on a T_1-weighted image. Tissues having a short T_2 will relax rapidly, generating very little signal on a T_2-weighted image and

as such will appear dark, while a long T_2 tissue will appear bright on a T_2-weighted image. A proton density weighted image will reflect the distribution of the proton density, which is measured by the number of protons per unit of tissue. The higher the number of protons in a given unit of tissue, the brighter the signal on the proton density contrast weighted image. These images are often followed by a T_1-weighted image after intravenous administration of a contrast agent like Gadolinium-DTPA (Gd-DTPA), to assess the blood brain barrier (BBB) viability. Gadolinium will tend to change signal intensities by shortening the T_1 value in its surroundings.

Next to this, also more modern MRI methods may be used, like diffusion- and perfusion-weighted imaging and functional MRI. During diffusion-weighted imaging, the water movement within the brain tissue is analyzed at the molecular level. An abnormality can be detected and visualized on diffusion-weighted images, since the normal water movement is restricted. Perfusion-weighted MRI examines the vascularity of the tissue, which enables to detect alterations in the blood volume or circulation. With functional MRI it is possible to visualize which specific brain structure participates in a specific function, and is therefore important in the assessment of the function of the brain tissue surrounding a tumour.

MRS can be performed with several MR sensitive nuclei. The use of ^{31}P MRS has been found important in studies of energy metabolism. It can also be used to calculate intracellular pH (Negendank, 1992), which is an important parameter in the study of tumours. Also the signals of phosphomonoesters in tumours can be measured by ^{31}P MRS. These signals are found to be raised in tumours compared with normal tissue.

The metabolic information obtained by ^{1}H MRS has proven to be important in the investigation of brain tumours (Nelson et al., 1999; Preul et al., 1996). Several metabolite concentrations have found to be lowered or increased in tumours compared with normal tissue. This change in concentration can be used for specific tumour determination. As the proton nucleus has a high MR sensitivity and a high natural abundance, a large number of brain metabolites contain protons that generate signals at specific frequencies in the MR signal. The number of signals is so high that these actually crowd the MR spectrum and the overlap of signals is manifest. To influence the presence of signals from metabolites in the proton MR spectrum, acquisition protocols can be used that use different acquisition parameters. One of these parameters, the echo time (TE), has a direct influence on the signal of metabolites in the MR spectrum. After excitation of the proton spins, the signal is decaying by the T_1- and T_2-relaxation processes. These parameters, together with the concentration and the number of protons in the molecule, determine the height and shape of each resonance in the MR spectrum. The longer the TE, the more the signal has attenuated before acquisition. Hence, a short echo time spectrum (TE $\leq$ 50ms) has larger peaks than a long echo time spectrum (TE

$\geq$ 130ms) (Figure 1.5). A short echo time spectrum also contains more peaks, as resonances with a small T_2 value (rapid T_2 relaxation) or complex coupling pattern, like *mI* (myo-Inositol), *Glu* (glutamate) and *Gln* (glutamine) are less pronounced at longer echo times. At short echo time signals macromolecules are prominent, originating from proteins and membrane components. They have very broad peaks with a large contribution to an underlying and partially unknown baseline.

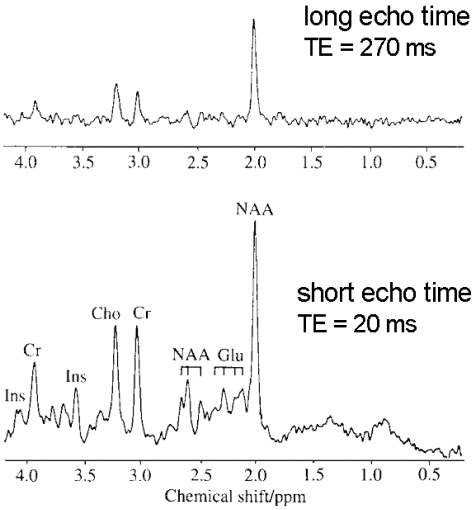

Figure 1.5: Single-voxel ^{1}H MR spectra at long echo time (top, TE=270 ms) and short (bottom, TE=20ms). At short echo time more resonances can be detected, however peaks are largely overlapping and an underlying baseline is present.

1.2.4 MRS characteristics of brain tumour types

Although there exists a high degree of variability in human brain tumour types (Mukherji, 1998; Smith and Stewart, 2002), some general characteristic differences between normal and brain tumour tissues have been reported, illustrated on Figure 1.6 for short TE spectra: *NAA* (N-acetyl-aspartate) and *Cr* (creatine) levels are reduced, while *Cho* (choline) levels are increased. *NAA* in the brain is commonly believed to be a marker of neuronal density and reduced *NAA* levels may reflect neuronal dysfunction or loss (Govindaraju et al., 2000). An elevation of *Cho* indicates increased cell density and membrane metabolism, which could be related to rapid tumour growth. *Cr* plays a role in the maintenance of energy metabolism. Malignant tumours have a high metabolic activity and unload the energy storage. However this metabolite profile is not exclusively specific to brain tumours. Several authors (Leclerc et al., 2002; Mukherji, 1998;

Govindaraju et al., 2000; Hakumäki et al., 1997; Martin et al., 2001) discuss more on the function of several metabolites and their relation with various tumoural and nontumoural processes.

In tumour diagnosis, accurate determination of the grade of malignancy is one of the most important factors for some types of tumours. Generally, fast growing tumours are much more likely to emerge again after treatment than slowly growing tumours. On the other hand, high-grade tumours do respond better to radiotherapy and chemotherapy. Lipids, *Lac* (lactate) and macromolecules are the most significant metabolic markers of tumour malignancy (Howe et al., 2003).

1.3 Goals of the thesis

This thesis concerns two main topics of processing Magnetic Resonance data, namely classification and quantification. The main challenge for classification of MR data is to develop a reliable decision support system that provides a reliable separation of brain tumours. If the system would be implemented in a practical clinical system, a confident diagnosis can only be achieved when the system gives a reliable indication of the type and the grade of the tumour. However, the development and integration of such a highly performing system in clinical practice is not yet reached. The aim of this study is to facilitate the road towards a confident decision support system. This includes several aspects:

- Preprocessing steps. Many preprocessing steps exist to increase the spectral quality and to enhance the interesting spectral features. The issue is whether, for the purpose of brain tumour diagnosis based on MRS data, these preprocessing steps are required for achieving a high performance.

- Dimensionality reduction. Classification can be based on the spectra, but also on extracted features. Feature extraction has the advantage to lower the dimension and to simplify the computation. It may also result in the loss of important information and as such lower the resulting performance. The question is whether the same results can be obtained by applying dimensionality reduction.

- MRSI and MRI parameters. MRI is widely accepted as diagnostic tool providing spatial information about the water content of the examined tissue. During MRSI, information is obtained about the present metabolites. Both techniques each provide valuable information for the typing of brain tumours. Then it would be of high interest whether the integration of parameters from both techniques improves classification.

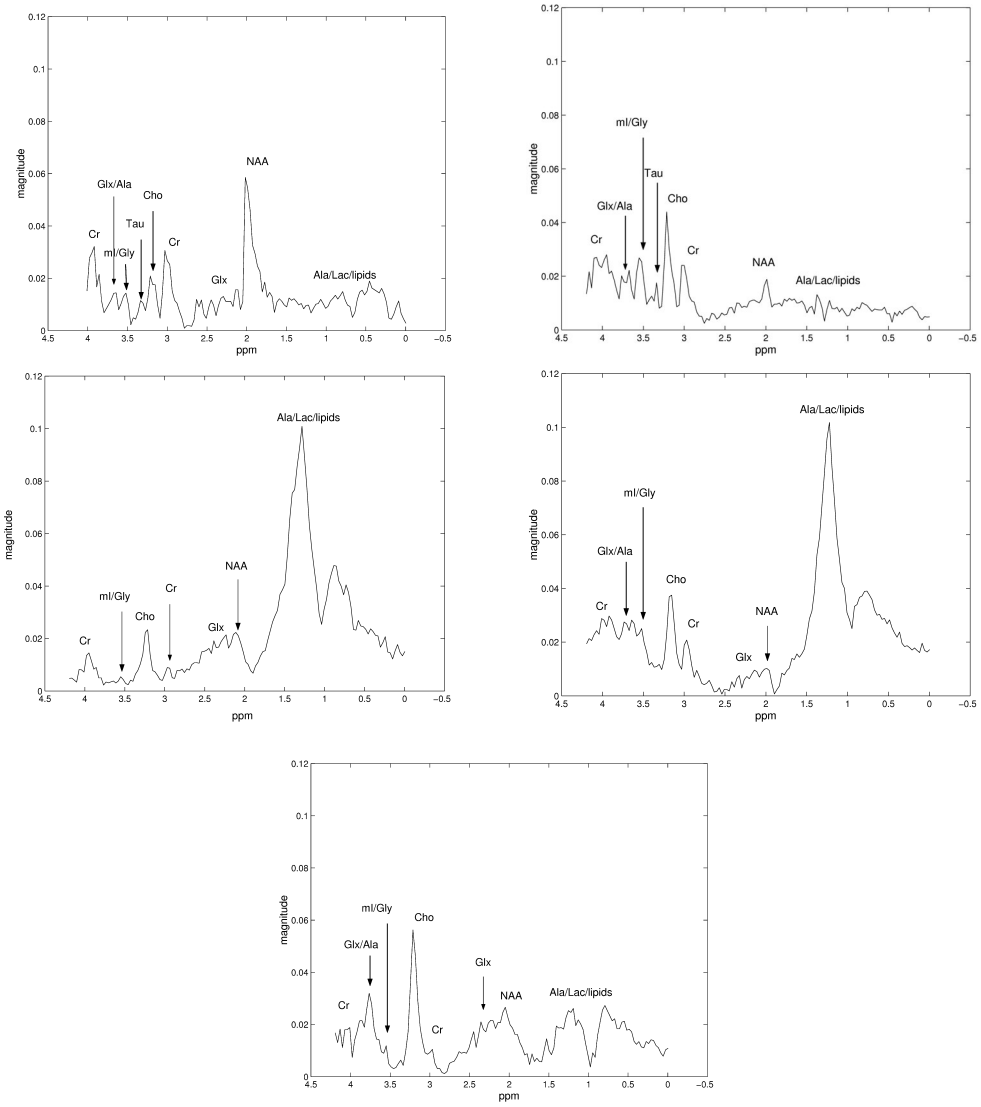

Figure 1.6: Typical water normalized magnitude ^{1}H MR spectra (TE=31 ms) of the normal tissue of a volunteer (top-left), an astrocytoma of grade II (top-right), a glioblastoma (middle-left), a metastasis (middle-right) and a meningioma (bottom). Data are provided by IDI (Barcelona, Spain), within the framework of INTERPRET (IST-1999-10310).

- Classification technique. There exists a large amount of techniques applicable to NMR; based on several principles, simple or more complicated, linear or nonlinear. Each technique has its advantages and drawbacks, but the issue is which technique would yield the best results for brain tumour classification using NMR.

Classification studies are all performed in order to investigate which input types and which techniques are most appropriate for brain tumour diagnosis using MR data. In addition to these goals, the problems of a low amount of variables and automation of the analysis should be addressed. To achieve a sufficiently high performance, most algorithms require a substantially large data set. However, the amount of experimental data is restricted, which makes generalization more difficult. Automation is not only important for the processing of the large-dimensional data but also to ensure that results are reproducible and independent of the user. In clinical practice it is required that data processing and generation of the results should be as efficient as possible.

The second topic concerns MRS quantification. MRS signals can be analyzed by a time domain model to retrieve estimates of the model parameters. The technique is often employed as a tool for fundamental research about biological and biomedical processes. The parameters that describe the MRS signal are then related to biochemically relevant properties; for example the amplitude corresponds to the intensity of the investigated molecule, while the damping is determined by the mobility of the molecule. MRS data are characterized by a low SNR, overlapping peaks and in some cases model distortions and the presence of artefacts. Despite these problems, it is important to process the signals as reliably as possible. Many existing algorithms assume a certain model order of the signals. But for experimental signals it is not straightforward to impose a certain model order. The goal of this thesis for quantification is to implement a method for selecting the correct model order and to apply this for reliable quantification of *ex vivo* ^{13}C MRS signals from biological experiments.

1.4 Outline of the thesis

During the past five years the author performed research on classification and quantification of MR data. Figure 1.7 illustrates the structure and visualizes different relations between the chapters of this thesis.

This **chapter 1** introduces several concepts of Magnetic Resonance and its applications. After a brief explanation of the physical principle of NMR and its main branches MRS and MRI, an overview is given of a few applications for fundamental research and clinical examinations. MR for diagnosis of brain tumours is the first main focus in this thesis. Some medical background about

brain tumours is outlined as well as the current clinical practice in diagnosing them and the potential value of MRS in this context.

Chapter 2 presents two related areas, namely quantification and model selection. First the concept of MRS quantification is introduced, which is the second main focus of this thesis. Different approaches can be taken to quantification and they can be divided into frequency and time domain methods. A few of these methods are explained and their advantages and disadvantages are mentioned. Quantification of MRS data is not always straightforward. One important issue is that most methods assume the number of signal components to be known a priori, while in practice this knowledge is not available. Then we come upon model selection which is the second topic explained in this chapter. An information theoretic approach is outlined for model selection of MRS signals. The model selection methods mentioned are based on some criterion and are designed to select the model order from a set of possible model orders.

Chapter 3 is dedicated to the broad research topic of classification or pattern recognition. An interpretation is given on the concept of pattern recognition and the development process of a pattern recognition system is described. Several steps are distinguished in the development of such a system and some guidelines are provided for each of these steps. Classical and more recent techniques are explained applicable to classification of MRS data.

Chapter 4 concerns classification of brain tumours based on long echo time proton MRS data. A pattern recognition process was carried out following the scheme described in chapter 3. The issue of dimensionality reduction prior to classification is investigated. This is for example carried out by quantification. Quantification provides parameter estimates related to the level of metabolites such as NAA and Cho, which could then be supplied to the classifier. Classification of data from two classes (binary classification) as well as from more classes (multiclass classification) are considered. Classification techniques introduced in chapter 3 are applied and their performances are compared.

Chapter 5 reports on the use of short echo time proton MRS data for brain tumour classification. The influence of preprocessing on the performance is tested, in order to make the pattern recognition based on MR data more efficient. In addition, short echo time spectra provide more information than long echo time spectra, but processing implies also more difficulties to analyze signals.

Chapter 6 explores the value of metabolic information from MRSI and of intensities obtained from MRI for separating types from healthy tissue and brain tumour. Parameters of both input types are given to the pattern recognition system. Results using MRSI and MRI intensities individually are confronted with those using a combination of both input types. This provides an illustration of the diagnostic value of both input types.

Chapter 7 applies model (order) selection to models obtained by quantification. Quantification is performed in the time domain by AMARES, while prior knowledge is supplied to the algorithm to increase efficiency, convergence rates and overall accuracy. Model selection is implemented and tested for simulated "multi-exponential" signals, which consist of narrow as well as broad components at the same resonance frequency. The method is applied further on experimental carbon signals of glycogen, for which a multi-exponential pattern has already been observed. The objective of model order selection is to provide objective information about the multi-exponential nature of glycogen signals. Implications on the time course of relevant parameters during the biological experiment is observed.

Chapter 8 mentions conclusions about the main topics; classification of brain tumours using MRS and MRI data as well as model selection applied to quantification of multi-exponential signals. To end this thesis, suggestions are given for future research.

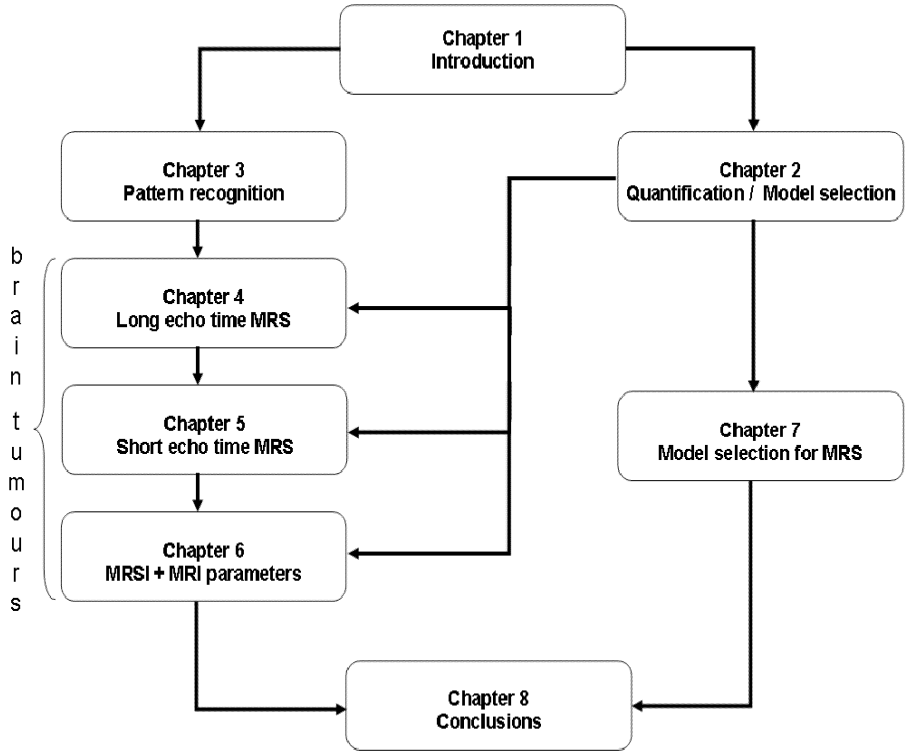

Figure 1.7: Flowchart of the thesis. Classification and quantification of MRS are respectively represented at the left and right sides of the scheme.

While the introductory chapter provides a background of MR useful for all chapters, chapters 2 and 3 each introduce the cornerstones of MRS quantification and brain tumour pattern recognition using MR data. Chapter 3 explains many concepts which are exploited later on in chapters 4, 5 and 6. The study described in chapter 4 provides a framework that is employed for chapters 5 and 6 in a similar way. Quantification results are also available as input types for classification, which illustrates the value of chapter 2 for classification. Chapter 2 also describes the necessary mathematical tools to perform model selection for a set of models obtained by quantification, as carried out in chapter 7. The concluding chapter 8 is based on results obtained in all studies mentioned in this thesis.

1.5 Contributions and research partners

Contributions

The scientific results of this PhD thesis concern two main applications of MR. The first topic concerns the value of *in vivo* ^{1}H MR Spectroscopy and MR Imaging parameters for diagnosing brain tumours. The focus lies on the development of algorithms to separate different tumour types. Contributions were made for each of the purposes mentioned in section 1.3. The effect of dimensionality was one of the main topics of chapters 4 (long echo time MRS) and 5 (short echo time MRS). For short echo time signals, the influence of several preprocessing steps on the classification performance was investigated (chapter 5). The integration of particular parameter values from MRS and MRI for brain tumour recognition was the main focus of chapter 6. For each of these studies classification was performed by linear as well as nonlinear techniques and performance results are shown in terms of Receiver Operating Characteristic (ROC) analysis (chapter 3), which is commonly used for medical diagnostic problems. The presented material on classification for brain tumours has been published as journal papers (Lukas et al., 2004; Devos et al., 2004, 2005b), while chapter 3 mainly corresponds to a chapter submitted for publication in a book about pattern recognition (Devos et al., 2005c).

The second topic concerns the use of processing tools for the estimation of the model order and the reliable analysis of glycogen signals by using *ex vivo* ^{13}C MRS signals (chapter 7). A framework was implemented for model selection, applied to simulation signals and experimental glycogen ^{13}C MRS signals. The simulation study investigated whether the method is applicable to multi-exponential signals, while experimental signals were analyzed to test for the multi-exponential nature of experimental glycogen signals. The results described in chapter 7 are included in Devos et al. (2005a).

Next also the quantification of the long echo time ^{1}H MRS data from the database of the INTERPRET-project (1999-2002) was automated. This study

has been described in two technical reports (Devos et al., 2002a,b), but was not included in this text.

Research partners

The studies mentioned in this thesis are the results of a collaboration with many persons. The work was mainly inspired by my supervisors Sabine Van Huffel and Paul Van Hecke, who lead me through 5 years of research always providing critical comments. Many inspiring ideas for quantification and classification were given by one of my closest collaborators and advisor Leentje Vanhamme. The three classification studies were carried out together with Lukas, who implemented many of the algorithms, and Johan Suykens, who motivated to apply kernel-based methods. The third classification study (chapter 6) was a result of a collaboration with Arjan Simonetti and Marinette van der Graaf, who provided a large amount of knowledge about the acquisition and preprocessing of the MRSI and MRI data. Preprocessing was mainly carried out by Arjan Simonetti. A related study was published as Simonetti et al. (2004).

Without any brain tumour data the presented classification studies would not have existed. All data were acquired at several centres in the framework of the INTERPRET-project (1999-2002). Many discussions were held about the MR (pre)processing, design and automation of the classification procedure. Related work includes that of Tate et al. (2003), Szabo De Edelenyi et al. (2000), Simonetti (2004) and Majós et al. (2003). The study was performed in collaboration with Carles Arús, Rosemary Tate and the representatives of each of the partners acting as acquisition centres; Franklyn Howe, Marinette van der Graaf, Angelo Moreno-Torres and Carles Majós.

The software implementation consisted of three main programs for preprocessing, classification and statistical testing respectively and was adapted where necessary for each study. My contributions covered the preprocessing and preparation of the data, part of the classification tests, in particular for the studies on short echo time MRS and on the combination of MRSI and MRI parameters. The software for the application of the techniques was mainly written by Lukas in several subprograms, while I managed the main programs and some other subprograms. The statistical tests were mainly carried out by myself. Also several classification tests were carried out by myself in order to reproduce results of Tate et al. (2003). The experimental strategy was based on my initiative and discussions with Lukas, Leentje Vanhamme, Jos de Brabanter, Johan Suykens and Sabine Van Huffel. Reporting on the results, especially for the last two classification studies, was mainly my work. I was responsible for the intensive correspondence with all collaborating INTERPRET partners concerning the scientific reporting of the results. This also included checking and updating the information of the available data during the development process of the INTERPRET-database, according to the quality control and validation guidelines and decisions of the Clinical Data Validation Committee

(CDVC) of INTERPRET.

The quantification study (chapter 7) was carried out together with the Biomedische NMR Eenheid of the KULeuven under supervision of Paul Van Hecke. The centre acquired the necessary experimental data. Many critical remarks and fruitful discussions were provided by Tom Dresselaers and Niki Bergans. Guidance about statistical aspects was given by Jos de Brabanter of the department of Electrical Engineering.

The previous paragraphs illustrate my role in the development of classification algorithms within the INTERPRET-project (1999-2002). My responsibility within the project also consisted of the software development for the automated and reliable quantification of long echo time MR spectra and the parameter estimation of the metabolites (Devos et al., 2002b,a). This research topic required a phantom study on the signals acquired at different acquisition centres. The study was done in collaboration with Marinette van der Graaf and Angelo Moreno-Torres, who provided a large amount of necessary information about the acquisition and the quality control of the MRS data.

1.6 Conclusions

All basic concepts of Magnetic Resonance have been introduced, which are important for this thesis. This included the physical principle of MR as well as three types of MR data that can be acquired using the same scanner. Next, the potential of MR was stressed as an examination tool in clinical diagnosis, especially of human brain tumours. Several purposes for this thesis were addressed and related challenges that will be encountered were mentioned. To continue, a chapter-by-chapter overview and the main contributions of the thesis were provided.

Chapter 2

Quantification of MRS signals

This chapter introduces the mathematical processing performed on an MR signal which is also called quantification. The purpose is to obtain accurate estimates of the quantities characterizing each of the components of the signal. The discussion is mainly restricted to the methods employed in this thesis. Section 2.1 mentions the most important model used for quantification. The concept of prior knowledge about relations between model parameters is outlined. Prior knowledge can lead to a more reliable estimation of the model. Section 2.3 outlines the different approaches to MRS quantification and mentions the signal processing tools applied in this thesis: one frequency- and two time-domain methods. Most approaches assume a priori knowledge about the model order, which is the number of signal components. In practice this model order is unknown, which motivates to apply a method for estimation of the model order. Section 2.4 describes several criteria based on information theory and their use to select the optimal model from a set of models.

2.1 Modelling of an MRS signal

As the FID signal is acquired in the time domain, it seems quite obvious to analyze the signal by a time-domain model. The FID is often modelled as a Lorentzian model, which is a sum of complex damped exponentials:

$$y_n = \bar{y}_n + e_n = \sum_{k=1}^{K} (a_k e^{j\phi_k})\, e^{(-d_k + j2\pi f_k)t_n} + e_n \quad n = 0, 1, ..., N-1, \quad (2.1)$$

where the model order K corresponds to the number of different resonances and $j = \sqrt{-1}$. The term $\bar{y}_n$ is the noiseless signal and e_n is the Gaussian noise term. The noise term e_n is supposed to be circular complex, which means that the real and imaginary parts of the noise are uncorrelated and have equal variances. The frequency f_k identifies the biochemical component, the amplitude a_k is proportional to the concentration of the corresponding molecule, the damping d_k characterizes the mobility of the molecule and the phase ϕ_k consists of a so-called zero-order phase ϕ_0 and an additional phase factor ϕ'_k. The acquisition of the signal begins after a time t_0. The sampling times are determined by $t_n = n\Delta t + t_0$, with Δt the sampling interval. The local field inhomogeneities caused by variations in the tissue homogeneity and operator shim introduce distortions of this theoretical model.

2.2 Prior knowledge

In vivo MRS signals are characterized by a low SNR, the presence of unwanted spectral components like the residual water peak in proton MRS and distortions with respect to the expected model function. Additionally, in short echo time proton spectra different peaks are strongly overlapping, making it difficult to determine the model parameters.

In order to facilitate a more reliable and accurate quantification, the use of prior knowledge is exploited. Prior knowledge typically deals with several mathematical relations between model parameters, which reduces the number of independent model parameters. This prior knowledge can in some cases be derived from quantum mechanics, illustrated by the ^{31}P spectrum containing adenosine triphosphate (ATP) (Figure 2.1). Although the ATP signal is not analyzed in this thesis, it is a traditional example of the concept of prior knowledge. The ATP molecule contains three phosphorus atoms, called P_α, P_β and P_γ, each at a different position, experiencing a different chemical environment and thus a different shielding. As a result the three ^{31}P atoms will resonate at different frequencies. Moreover, spin-spin coupling between neighbouring atoms will split the peaks of each group into a multiplet of lines. The groups P_α and P_γ are split into two peaks and P_β consists of three peaks (Figure 2.2).

The quantum mechanical knowledge reflects in the following relations between the model parameters of the ATP signal:

- The frequency differences between the individual resonances within a multiplet are equal and known: $f_1 - f_2 = f_3 - f_4 = f_5 - f_6 = f_6 - f_7 = \Delta f$.
- The dampings and phases of all peaks are equal: $d_1 = d_2 = \ldots = d_7$, $\phi_1 = \phi_2 = \ldots = \phi_7$.

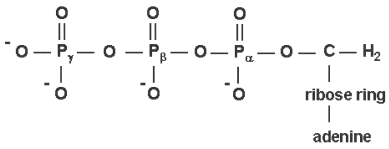

Figure 2.1: Chemical structure of ATP. Three phosphorus atoms have different positions within the molecule, leading to a different shielding.

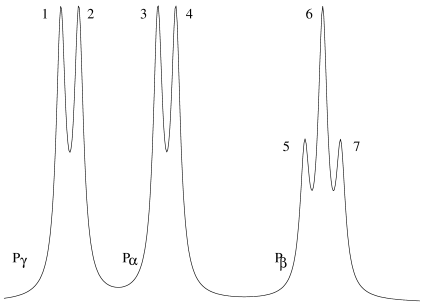

Figure 2.2: Theoretical ^{31}P spectrum of ATP. Three groups of spectral lines can be identified as resonances from P_γ, P_α and P_β. Moreover, each group is split into a multiplet.

- The amplitudes of the doublet peaks of the P_α and P_γ atoms and the central peak of the P_β atom are equal. The amplitudes of the outer peaks of P_β are half the amplitude of the central peak. In summary: $a_1 = a_2 = a_3 = a_4 = 2a_5 = a_6 = 2a_7$.

2.3 Quantification methods

Quantification of MRS signals can either be done in the original time domain or in the frequency domain. Since the spectroscopic data are acquired in the time domain, it is most natural to process the data in the time domain. Visual interpretation of the MRS signal and quantification results is more intuitive in the frequency domain. In the time domain the concentration of a metabolite is proportional to the amplitude of the signal, while in the frequency domain the concentration is proportional to the area under the metabolite peak.

2.3.1 Frequency-domain methods

Two main approaches exist for MRS quantification in the frequency domain: nonparametric and parametric quantification.

Nonparametric quantification can be done by peak integration. The method is the oldest and simplest method for quantification in the frequency domain and is based on the integration of the area under the peak of interest in the frequency domain signal (Meyer et al., 1988). This method is relatively robust, but has a low estimation accuracy. Moreover, the method requires appropriate phasing. The method is explained below.

Parametric quantification uses a model to approach the spectrum. Mostly a superposition of Lorentzian lineshapes is used as in Equation (2.1). The parameters can be estimated in the frequency domain using a nonlinear least squares (NLSS) algorithm, which minimizes the squared difference between the spectrum and the model function. Theoretically, this approach is equivalent to the time-domain modelling. A semi-parametric method called LCModel has been proposed by Provencher (1993) and is the current standard frequency-domain method. It models the MRS signal as superposition of *in vitro* metabolite solutions, which has the advantage that a maximal amount of prior knowledge is already provided. Moreover, model distortions due to acquisition effects can also be taken into account. But the method requires the recording of a basis set of metabolite spectra under exactly the same settings as the signal under investigation and the model order should be provided prior to quantification.

Peak integration

The integral under the peak can be estimated by the sum rule, trapezoid rule and Simpson's rule, respectively written as:

$$a_{\text{sum}} = \sum_{f_{\min} \le f_l \le f_{\max}} I(f_l) \tag{2.2}$$

$$a_{\text{trapezoid}} = \sum_{f_{\min} \le f_l \le f_{\max}} \frac{I(f_l) + I(f_{i+1})}{2} \tag{2.3}$$

$$\begin{aligned} a_{\text{Simpson}} = \; & \frac{1}{3}(I(f_{\min}) + 4I(f_{\min+1}) + 2I(f_{\min+2}) + \dots \\ & +2I(f_{\max-2}) + 4I(f_{\max-1}) + I(f_{\max})), \end{aligned} \tag{2.4}$$

with $I(f_l)$ the intensity at frequency f_l. This simple approach is nonparametric, does not make any model assumptions and is quite robust. The method is also non-iterative and very fast, such that an estimate will always be obtained. Nevertheless, the method has a few disadvantages. The accuracy is highly dependent on the integration interval $[f_{\min}, f_{\max}]$ chosen. In theory, a resonance peak will be spread over the complete frequency domain and to obtain an

estimate of the total signal intensity for each peak, the complete spectral range should be taken as integration interval. This is not feasible due to the peak overlap of different resonances, which inherently causes the method to miss part of the peak's area. Most often an integration interval is taken symmetric around the theoretical resonance frequency of the peak. Under experimental conditions, the resonances can be misaligned, due to variations in temperature, pH or physiological abnormalities. This implies that the resonance frequency can deviate from the theoretical one, which might result in a less accurate estimate of the peak's area. Additionally, serious problems can be caused by the possibly low resolution of experimental signals. Due to a low resolution it might be impossible to resolve two or more largely overlapping resonances.

2.3.2 Time-domain methods

Time-domain methods can be divided into interactive and black-box methods (Vanhamme et al., 2001). In contradiction to interactive methods, the black-box methods are able to provide parameter estimates after one single quantification step, without the need for starting values. However, it is more difficult to built in prior knowledge for black-box methods. This is an important shortcoming when analyzing signals for which the use of prior knowledge is needed to obtain confident results. Interactive methods have the ability to incorporate different types of prior knowledge which make them the favorable approach for MRS quantification. Unfortunately, user interaction is required to provide starting values and no guarantee can be given that the algorithm will converge to a global optimum.

AMARES as an interactive time-domain method

AMARES is an algorithm that belongs to the family of interactive time-domain methods. Interactive methods are flexible and result in reliable solutions. The algorithm can easily incorporate prior knowledge, which reduces the number of free model parameters, thereby significantly improving parameter accuracy (Vanhamme, 1999). When the underlying assumptions concerning the noise distribution are satisfied (section 2.1), obtaining maximum likelihood (ML) estimates is equivalent to minimizing the squared difference between the data and the model function. The cost function can be written in vector notation:

$$||\mathbf{y} - \bar{\mathbf{y}}||^2 \tag{2.5}$$

where $||\cdot||$ denotes the Euclidean vector norm and $\mathbf{y} = [y_0, \ldots, y_{N-1}]^T$, $\bar{\mathbf{y}} = [\bar{y}_0, \ldots, \bar{y}_{N-1}]^T$ are respectively the acquired time-domain points and the model.

The estimates can then be obtained by minimizing the variable projection functional, which is a functional that enables to split the model function into

a linear and a nonlinear part. To derive this functional, the model in Equation (2.1) is rewritten as:

$$\bar{y}_n = \sum_{k=1}^{K} c_k \gamma_k(\varrho_k, n),\ n = 0, \ldots, N-1 \tag{2.6}$$

where c_k are the complex amplitudes and $\gamma_k(\varrho_k, n)$ are independent functions of the nonlinear parameter vector ϱ_k. In the case of a Lorentzian model, $\varrho_k = [f_k d_k t_0]$, $c_k = a_k e^{j\phi_k}$ and $\gamma_k(\varrho_k, n) = e^{(-d_k + j\,2\pi f_k)n\Delta t}$. Using the matrix notation, Equation (2.6) becomes

$$\bar{\mathbf{y}}_n = \mathbf{\Gamma c} \tag{2.7}$$

where $\mathbf{c} = [c_1, \ldots, c_K]^T$ and

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1(\varrho_1, 0) & \ldots & \gamma_K(\varrho_K, 0) \\ \vdots & \ddots & \vdots \\ \gamma_1(\varrho_1, N-1) & \ldots & \gamma_K(\varrho_K, N-1) \end{bmatrix} \tag{2.8}$$

is a $N \times K$ matrix of full rank. If the nonlinear parameters $\varrho_k, k = 1, \ldots, K$ would be known, then the matrix $\mathbf{\Gamma}$ could be computed and the linear parameters c_k could be estimated as the solution of a linear least squares (LS) problem. The estimates of the linear parameters can then be calculated as $\hat{\mathbf{c}} = \mathbf{\Gamma}^{\dagger}\mathbf{y}$, where $\mathbf{\Gamma}^{\dagger}$ denotes the pseudo-inverse of $\mathbf{\Gamma}$, i.e. $\mathbf{\Gamma}^{\dagger} = (\mathbf{\Gamma}^H\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^H$ and $\mathbf{\Gamma}^H$ is the complex conjugate of the matrix $\mathbf{\Gamma}$. The cost function in Equation (2.5) is rewritten as:

$$||\mathbf{y} - \mathbf{\Gamma}\mathbf{\Gamma}^{\dagger}\mathbf{y}||^2. \tag{2.9}$$

This procedure eliminates the linear parameters from the cost function in Equation (2.5). The minimization problem is now described in fewer variables. Estimates $\hat{\varrho}_k$ for the nonlinear parameters are obtained by minimizing the cost function in Equation (2.9). These estimates are then used to calculate $\hat{\mathbf{\Gamma}}$ and estimates $\hat{c}_k$ for the linear parameters are obtained by solving the linear LS problem:

$$\hat{\mathbf{\Gamma}}\mathbf{c} \approx \mathbf{y}, \tag{2.10}$$

or the estimates can be computed as $\hat{\mathbf{c}} = \hat{\mathbf{\Gamma}}^{\dagger}\mathbf{y}$. Golub and Pereyra (1973) have proven that both cost functions (Equations (2.5) and (2.9)) provide the same estimates for the nonlinear parameters. The estimated signal is then related to the original data as:

$$\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}, \tag{2.11}$$

where $\mathbf{S} = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}^{\dagger}$ is called the "hat" matrix.

The nonlinear least squares (NLLS) problem can be solved using local or global optimization algorithms. The use of global optimization prevents the method to end in local minima. However, these optimization techniques are computationally inefficient. By using local optimization, it is possible to provide good

starting values of the frequencies and dampings by peak picking. Peak picking is a method to manually indicate the central frequency and the damping for each peak by visual inspection of the spectrum in the frequency domain. Since this step involves user interaction, these model-based quantification methods are called interactive methods.

The first well-known interactive method for quantification of MRS data was VARPRO (van der Veen et al., 1988), which is based on a modified version of the Levenberg-Marquardt algorithm (Osborne, 1972). VARPRO was further improved by AMARES (Vanhamme et al., 1997) in terms of efficiency, robustness and flexibility. AMARES stands for "Advanced Method for Accurate, Robust and Efficient Spectral fitting" and computes the NLLS solution by using a more sophisticated algorithm. Additionally, the method also enables to impose more types of prior knowledge and to model the signal by non-Lorentzian model functions.

Black-box methods

Within the black-box methods, two main categories can be identified: linear prediction methods (Stephenson, 1988; Koehl, 1999) and state-space methods (Barkhuysen et al., 1987; Kung et al., 1983; de Beer and van Ormondt, 1992). In this section we restrict our overview to state-space methods.

State-space methods order the data $y_n, n = 0, \ldots, N-1$ into a Hankel matrix as follows:

$$\mathbf{H} = \begin{bmatrix} y_0 & y_1 & \cdots & y_M & \\ y_1 & y_2 & \cdots & y_{M+1} & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{N-M-1} & y_{N-M} & \cdots & y_{N-1} & \end{bmatrix}, \tag{2.12}$$

with $M \geq K, N - M > K$. The number M is chosen such that the matrix $\mathbf{H}$ is as square as possible. The method of Kung et al. (1983), introduced by Barkhuysen et al. (1987) in the field of MRS, computes the Singular Value Decomposition (SVD) of $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. The matrix Σ contains the singular values $\sigma_l, l = 1, \ldots, p$ with $\sigma_1 \geq \ldots \geq \sigma_p$ and p the rank of the matrix. The matrix is truncated to a selected number of singular values K:

$$\mathbf{H}_K = \mathbf{U}_K \mathbf{\Sigma}_K \mathbf{V}_K^H, \tag{2.13}$$

where $\mathbf{U}_K$, $\mathbf{V}_K$ contain the first K columns of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}_K$ is the upper-left $K \times K$ submatrix of $\mathbf{\Sigma}$. Using $\mathbf{U}_K$

$$\mathbf{U}_K^{\uparrow} \approx \mathbf{U}_{K\downarrow} \mathbf{Q}, \tag{2.14}$$

where the up (down) arrow stands for deleting the top (bottom) row respectively. In the method called HSVD (Barkhuysen et al., 1987; Kung et al.,

1983) an estimate of the matrix Q is obtained by solving Equation (2.14) in a LS sense. The eigenvalues of Q are the estimates $\hat{z}_k, k = 1, \ldots, K$ of the signal poles $z_k = e^{(-d_k + j\,2\pi f_k)n\Delta t}, k = 1, \ldots, K$. In order to compute the amplitudes and phases, these estimates are backsubstituted in the model equations for Equation (2.1) and fitted to the original data y_n with a LS method. Many variants of the HSVD algorithm exist. An improved version is the Hankel Total Least Squares (HTLS) method (Van Huffel et al., 1994) which computes the solution of Equation (2.14) in a total least squares (TLS) sense. This leads to a more accurate parameter estimation. The Lanczos version of HSVD (HLSVD) (Pijnappel et al., 1992) computes only the K first singular values of $\mathbf{\Sigma}$, as the singular values σ_l with $l > K$ are not considered and do not have to be computed. This scheme leads to a computationally more efficient version.

HSVD methods can be applied for MRS quantification as well as filtering of unwanted components in the signals. This latter can for example be important for removal of the residual water peak in proton MRS signals. For this purpose a filter region and a number of expected components K need to be defined a priori. The signal is modelled as a superposition of K peaks by HSVD or HLSVD. The reconstructed signal in the filter region is then subtracted from the original signal in order to retrieve a filtered signal. The subtraction of the reconstruction results in the removal of the often large tails of the water peak, but with only a limited influence on the interesting metabolite peaks.

2.4 Model selection

2.4.1 Problem statement

AMARES is often used for parametric quantification of MRS signals in the time domain. However, the method supposes the model order to be known exactly, which in practice is not the case. When imposing a too large model order, the model adapts itself too closely to the data, including the noisy components and will not generalize well. However, when choosing a too low model order, the model will underfit the data, causing a large bias and also resulting in poor generalization. In order to determine the model order, several techniques exist to perform model selection, such as cross-validation, bootstrapping techniques and information theoretic criteria. The information theoretic criteria applied in this thesis are described in the following sections.

2.4.2 Model selection criteria

Consider the selection of a model from a set of models $M_i \in \mathcal{A}$, where each M_i represents a nonlinear model for experimental data:

$$y_n = g_i(t_n, \theta_i) + e_n,\ n = 0, \ldots, N-1. \tag{2.15}$$

The vector θ_i contains all model parameters and $g_i(t_n, \theta_i)$ is the model function depending on t_n and determined by θ_i, and e_k are independently and identically distributed errors with mean 0 and constant variance σ_e^2.

The model M_i is selected that minimizes the following cost function:

$$\frac{RSS_i}{N} + \lambda \frac{d_{eff,i}}{N} \hat{\sigma_e}^2 \tag{2.16}$$

in which RSS_i is the residual sum of squares, N the number of data points used to obtain the model M_i, $d_{eff,i}$ the effective number of parameters (for nonlinear and complex models) and $\hat{\sigma_e}^2$ is an estimate of the error variance σ_e^2. In order to minimize Equation (2.16), a trade-off should be made between a low residual error (first term) and a low model complexity (second term). Under the above assumptions concerning the distribution of the error term e_n, an unbiased estimator of the error variance σ_e^2 can be calculated (Gasser et al., 1986):

$$\hat{\sigma_e}^2 = \frac{1}{N-2} \sum_{n=2}^{N-1} \left[\frac{y_n - y_{n-1}\zeta(n) - y_{n+1}\eta(n)}{\sqrt{1 + \zeta(n)^2 + \eta(n)^2}} \right]^2, \tag{2.17}$$

with $\zeta(n) = \frac{t_{n+1} - t_n}{t_{n+1} - t_{n-1}}$ and $\eta(n) = \frac{t_n - t_{n-1}}{t_{n+1} - t_{n-1}}$. In case of equidistant samples in the time domain, as in MRS, $\zeta(n) = \eta(n) = \frac{1}{2}$. Equation (2.17) provides a consistent and nonparametric estimate of the error variance and is independent of the model M_i, which enables to estimate the model and error variance independently.

Equation (2.16) is called the Generalized Prediction Error (GPE) and generalizes well-known criteria such as Akaike's Information Criterion (AIC) (Akaike, 1973) to nonlinear models. If $\lambda = 2$, Equation (2.16) corresponds to the Generalized version of AIC (GAIC), applicable for model selection of nonlinear models (Hastie et al., 2001):

$$\frac{RSS_i}{N} + 2 \frac{d_{eff,i}}{N} \hat{\sigma_e}^2. \tag{2.18}$$

The criterion only depends on the model and the data and does not require any knowledge about the error variance σ_e^2. In case the error variance $\sigma_e^2 = \hat{\sigma_e}^2$ is assumed to be known, AIC is equivalent to C_p (Mallows, 1973). When the error variance is unknown, it should be estimated for example by Equation (2.17).

The effective number of parameters d_{eff} generally differs from the true number of parameters d and depends on the nonlinearity of the model, the prior model constraints and the amount of model bias (Moody, 1992). For linear models d_{eff} equals the number of parameters d. For nonlinear parametric models d_{eff} generally is not equal to d.

If $\lambda = \log(N)$, then Equation (2.16) corresponds to the Generalized version of the Bayesian Information Criterion (GBIC):

$$\frac{RSS_i}{N} + \log(N)\frac{d_{eff,i}}{N}\hat{\sigma_e}^2. \tag{2.19}$$

In the context of autoregressive model order determination (Hannon and Quinn, 1979), also the criterion with $\lambda = \log(\log(N))$ has been proposed. In this thesis the generalized version for nonlinear models is simply abbreviated as GIClog.

Akaike's criterion originally was designed as an estimate of the Kullback-Leibler (K-L) distance between the parametric model function $g_i(t_n, \theta_i)$ and the "true" model function $h(t_n)$. The K-L distance is interpreted as the amount of information lost when g_i is used to approximate h. For linear regression and time series models (Hurvich and Tsai, 1989), AIC may perform poorly if the number of parameters is too large in relation to the number of data points. For such cases Hurvich and Tsai (1989) proposed a corrected version of AIC, called AICC or AIC_c, which was found to be less biased than AIC. The Generalized version of AICC (GAICC) for nonlinear models is given by:

$$GAICC = \frac{RSS_i}{N} + \left(1 + \frac{2 * (d_{eff,i} + 1)}{(N - d_{eff,i} - 2)}\right)\hat{\sigma_e}^2.$$

McQuarrie and Tsai (1998) found that more complex penalty terms suited for certain error distributions did not yield improved performance over the simpler term in AICC.

In contrast to cross-validation (CV) and bootstrapping techniques, the information criteria do require an estimate of the error variance σ_e^2 and the computation of the "hat" matrix (Equation (2.11)) in order to determine the effective number of parameters (Moody, 1992). However, interactive quantification methods like VARPRO and AMARES directly provide the "hat" matrix, which makes it straightforward to apply the above mentioned generalized information criteria to MRS data quantification. Quantification by AMARES using a Lorentzian model, as in Equation (2.1), results in a specific model function which corresponds to the general formula in Equation (2.15). Moreover, the assumptions about the error term are fulfilled for MRS signals, such that an estimate for the error variance can be obtained. In addition, results of the information criteria are obtained very fast, which is in contrast to cross-validation and bootstrapping techniques.

2.5 Conclusion

This chapter introduced the concept of MRS data quantification, which tries to approximate the MRS signals as a model. Mathematical signal processing tools

were described that can be applied for quantification. This included frequency as well as time-domain methods relevant for this thesis. The concept of prior knowledge about the parameters is of great value to improve the accurateness and reliability of the model. This not only includes knowledge about relations between the model parameters, but also about the model order, which is often unknown in experimental situations. Therefore, the concept of model selection was introduced which selects the appropriate model from a set of given models. The background of a few specific criteria from information theory was outlined and the connection with MRS quantification was briefly touched.

Chapter 3

Pattern recognition of brain tumours using MR data

In this chapter an overview is given of the different concepts of a typical pattern recognition system applied to MR data of brain tumours. The purpose is to stress the important steps needed in order to obtain reliable results that lead to a confident medical diagnosis. In section 3.1 the main idea behind pattern recognition is illustrated. In section 3.2 the development of a pattern recognition system is outlined, including all steps, from data collection over data preprocessing, construction and testing of the system, till the performance assessment. For each of these steps several possible approaches are described. Section 3.3 describes a few unsupervised as well as supervised techniques applicable to medical classification problems.

3.1 Pattern recognition

3.1.1 What is meant by pattern recognition?

Recognition is a process that human beings perform at about every instance in life and is considered as a first sign of intelligence. The recognition ability is developed from learning activities processed in our brain. A baby, a few months of age, begins to recognize faces of people around. At 2 years, a child can be taught to learn a few alphabets and numerals, based on the shapes. The child also learns to distinguish between objects of the same general shape, such as a ball, an apple and the sun. Perhaps at the age of 3 years, the human brain uses contextual information to recognize objects such as: a circular object on a plate is probably an apple, whereas a circular object in the dark sky is a moon

and a circular object on the playground is a ball.

Recognition is assumed to be based on certain patterns. A pattern can be defined as a representation or model which can be used to recognize or to generate things (of any kind such as objects, ideas, words), especially if these things have enough in common for the underlying pattern to be inferred. Such a representation can be formalized as a set of rules, which is especially needed when problems become more complex. These rules are, in a mathematical sense, based on measurements of so-called features such as compactness, height, color, temperature, volume, humidity. A feature is any kind of formalized property which can be used. Choosing discriminating and independent features is an important key to any pattern recognition process being successful in recognition.

In the field of pattern recognition the aim is to learn a computer by examples to recognize patterns in data sets. The recognition process performed by a computer is also called classification. The recognition of patterns can be applied to signals from many research areas, such as speech signals, medical images, seismic data, character data, satellite images.

3.1.2 The importance of feature selection

Assume several objects should be assigned to either one of the assumed classes. It is necessary to choose the features appropriately, to achieve a representation that enables a successful and reliable classification. A representation is considered to be good when the structural relationships among the components are simple and as such the true, but unknown, model of the patterns can be expressed. Patterns can be represented as vectors of real numbers (e.g. using numerical features, like length), as lists of attributes (e.g. using symbolic features, like color), as descriptions of parts or their relations (e.g. using contextual information, like position). In pattern recognition, a representation is searched for in which the patterns of objects belonging to the same class are “close” to one another, but “far” from those of objects belonging to other classes. The extent to which we learn or generate an appropriate presentation and how we quantify “close” and “far” determines the performance of the classifier. Additionally a few other requirements could be important for the presentation. We might wish to select features that are robust against noise and artefacts. Also we might wish to select only a small number of features that enables to work in a lower-dimensional space and that could be easier to train and could lead to simpler decision boundaries. It is important that the extracted features are relevant to the particular task at hand.

3.2 The development of a pattern recognition system

Pattern recognition may be suited for several purposes. By extracting characteristic information, the data may be significantly reduced in size and noise. It may also serve as basis for a more comprehensive visualization of the multidimensional data and it can be used for statistical interpretation. Lastly, pattern recognition can be applied for classification, which assigns a sample to a certain group.

The development of a pattern recognition system involves the following steps, as illustrated on Figure 3.1: (1) data acquisition and collection, (2) preprocessing, (3) construction and testing of the classifier, (4) measuring the performance. Each of these steps is discussed in the next paragraphs.

3.2.1 Data acquisition and collection

The results of pattern recognition are very dependent on the quality of the data. It is of utmost importance to develop a detailed acquisition protocol in advance in order to obtain reproducible high quality data. Depending on the purposes of the study, single or multivoxel data, at short or long echo time spectra should be acquired. If, in the area of MR, a research purpose is to investigate the combined use of spectral and clinical features, it is advisable to collect as many samples as possible that include these features.

Issues in spectral quality and quality control are in general underestimated as there is no worldwide consensus or set of detailed criteria for quality control of MR spectra. Spectral quality can for example be degraded by poor water suppression, peak doubling, ghost artifacts, outer volume signal. A detailed overview of artifacts lies out of the scope of this thesis, but several topics in quality assessment are discussed by Kreis (2004), including the reproducibility of spectra, the signal-to-noise ratio and the suppression of several artifacts. A multicentre collaboration was carried out for quality control assessment; Bovée et al. (1995) focusses on the protocols for quality assessment, Leach et al. (1995) discusses how an appropriate phantom can be developed, while Keevil et al. (1995) investigated the multicentre trial of several test cases and the validation of the quality assessment. For the EU-funded INTERPRET project (IST-1999-10310), all partners reached a consensus (van der Graaf, 2001) about the acquisition protocol for MRS and MRSI measurements. In addition, within INTERPRET, bimonthly quality control measurements were carried out using a phantom based upon the design by Leach et al. (1995). This facilitated the combined use of data acquired at different centers and using MR scanners from different manufacturers.

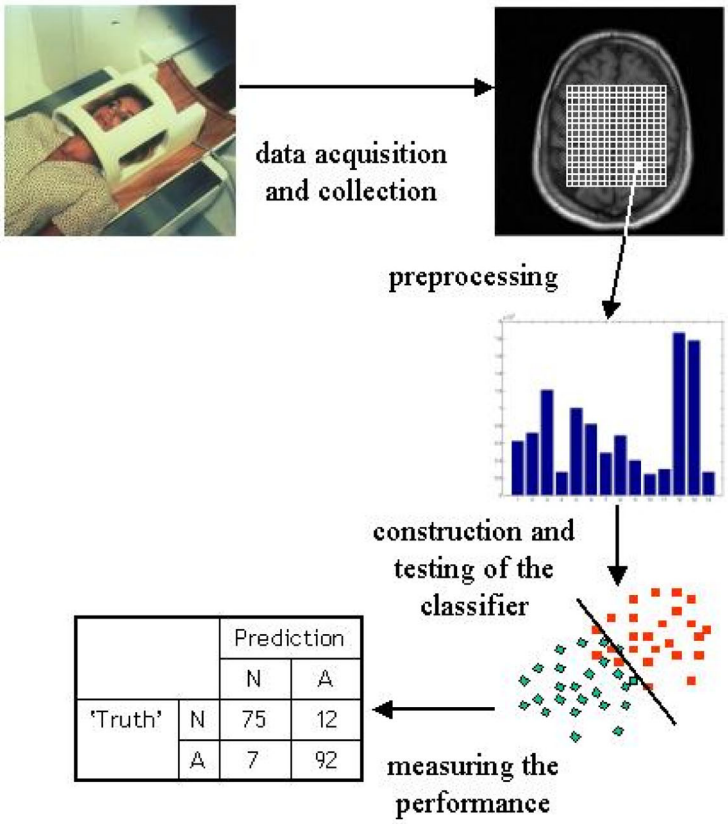

Figure 3.1: Scheme representing the different steps of the pattern recognition process based on MR data. Firstly, using an MR scanner (top-left) a large database of high quality signals is acquired; here a T_1-weighted image of a patient with grade III oligodendroglioma is displayed (top-right). Secondly, the MR signals and other medical data are preprocessed in order to enhance the biomedically relevant information; here several MR intensities (middle-right) are shown from one selected voxel on the grid. Based on these features a classifier (bottom-right) can be optimally constructed and tested. The classification result is evaluated using a particular performance measure; here using a contingency table (section 3.2.4). The T_1-weighted image was provided by UMCN (Nijmegen, The Netherlands).

3.2.2 Preprocessing of NMR signals

The aim of the preprocessing step is to remove the irrelevant information, while enhancing the key features in order to simplify the pattern recognition task. A few of the preprocessing steps are noise and artifact filtering and normalization. Phase correction, line broadening and baseline removal are preprocessing steps specifically developed for the field of MRS. Also feature extraction and selection can be regarded as a kind of preprocessing, that find a small amount of features with as much valuable information as possible. Typically, feature extraction can be performed by so-called unsupervised techniques (section 3.2.3), as clustering and principal component analysis techniques.

Spectral preprocessing

MRS signals are affected by the presence of artifacts, instrumental errors, noise and other unwanted components. The spectral quality can be considerably improved by appropriate manipulation of the data prior to Fourier transformation. The sensitivity can be enhanced by line broadening or apodization. This technique multiplies the time domain signal by a decaying exponential function, which broadens the resonance peaks in the frequency domain. This also enables to mimimize possible leakage, caused by the discontinuous nature of the signal Hoch and Stern (1996).

In case a water unsuppressed signal is available together with the water suppressed signal, the method of Klose (1990) can be carried out. This corrects for eddy currents (caused by instrumental imperfections) by dividing the water suppressed signal by the phase term of the water unsuppressed signal. This operation involves a zero-order phase correction and frequency alignment of the signal.

The proton MRS signal is composed of many resonances including several unwanted components, like the dominant residual water peak, artifacts and noise. Filtering the MRS signal to eliminate the unwanted features is advised prior to further processing. Residual water removal can be performed by a subspace-based modelling approach (Barkhuysen et al., 1987; Laudadio et al., 2002) or by filtering (Sundin et al., 1999; Kuroda et al., 1989). A discussion of the application of several methods for residual water removal can be found in Vanhamme et al. (2000).

In the next paragraphs a few preprocessing aspects are further discussed in detail.

Normalization of spectra

According to Cousins (1995), there are two approaches to interpret MRS signals: quantitative and qualitative. Pattern recognition can either be based on

metabolite concentrations (quantitative approach) or on spectra (qualitative). The first approach requires absolute concentrations using an external or internal reference signal to which the metabolite concentrations can be calibrated (Ernst et al., 1993; Henriksen, 1995; Kreis et al., 1993). Absolute quantification involves corrections for T_1- and T_2-relaxation behaviour. As T_1- and T_2-values differ between different brain regions (Mlynárnik et al., 2000) and water T_2-values are elevated in several brain tumour tissues compared to normal white matter (Howe et al., 2003), assessment of good estimates of T_1- and T_2-values for all relevant types of pathological tissues is very difficult. For most *in vivo* studies this is a serious disadvantage which motivates the use of the second approach in which the spectral values are fed to a classification algorithm, as discussed in this section.

Nevertheless, also in the qualitative approach a normalization procedure should be applied that tries to compensate for signal intensity differences between spectra due to effects independent of the type of tumour tissue. Two main normalization approaches can be applied to MRS signals; normalization with respect to a particular mathematical norm and normalization by division by an intensity of a particular molecule.

- Mathematical normalization: if the spectrum is considered as a vector $\mathbf{v}$, then the L_1-norm is defined as the sum of the absolute form of the elements in that vector, which corresponds to the integral of the spectrum. Each spectral value is then divided by the L_1-norm of the spectrum, in order to obtain an L_1-normalized spectrum. The spectral vector can also be divided by its L_2-norm $||\mathbf{v}||_2$, which is defined as the square root of the sum of the squares of the elements in that vector (Tate et al., 1996, 1998, 2003; Usenius et al., 1996). After division, the spectrum has unity norm with respect to the L_2-norm.
- Normalization by an intensity of a molecule: a common normalization approach in NMR quantification is normalization with respect to the amplitude of *Cr* (Preul et al., 1996) which is assumed to reach constant values over the complete brain region in normal tissue as well as in many pathologies. Nevertheless, Li et al. (2003) found that this normalization method increases variability. An alternative method divides the spectrum by the intensity of water, estimated from a spectrum without water suppression (Usenius et al., 1996). This requires an additional measurement of a corresponding water unsuppressed signal, acquired with the same acquisition parameters and originating from the same voxel. The intensity of the water peak can be estimated for example by taking the highest value in the time domain signal, by peak integration in the frequency domain (Meyer et al., 1988), by modelling the water unsuppressed signal using HSVD (Barkhuysen et al., 1987), HLSVD-PRO (Laudadio et al., 2002) or other modelling approaches.

Magnitude or real spectra

An MRS signal is typically represented by its real and imaginary part, and the process of "phasing" is required to correctly produce the real part of the spectrum for classification. Although automated phasing techniques exist, they are less suitable for *in vivo* data with large background signals. An alternative to using phased spectra is to calculate the magnitude spectrum, which makes the signal phase independent. However, magnitude calculation yields a significant broadening of the peaks and results in a larger overlap of the individual peaks (Hoch and Stern, 1996).

Baseline correction

Short echo time ^{1}H MRS signals are characterized by the presence of a partially unknown broad baseline underlying the sharper resonances of the metabolites of interest, that hinders the assessment of the intensity (i.e. quantification) of low weight metabolites. In order to correct the signal for the baseline, most approaches first obtain an approximation of the baseline and then subtract that approximation from the original signal. A few of these methods are mentioned below:

- Measurement of the baseline and subtraction from the original FID (Bartha et al., 1999; Behar et al., 1994). This assumes that the baseline has a similar pattern independent of the pathology, which is in general not the case.
- Parametric and nonparametric approaches that model the macromolecular baseline (Hofmann et al., 1999, 2002; Lemmerling et al., 2002; Provencher, 1993; Seeger et al., 2003; Soher et al., 2001). Most of these techniques are integrated as part of a quantification approach to estimate the MRS parameters. Baseline modelling can be done for example by wavelets (Young et al., 1998; Soher et al., 2001), polynomial or spline functions (Provencher, 1993; Lemmerling et al., 2002). This requires the selection of several spectral points to define the fitted baseline, a process which is heavily user-dependent.
- Multiplication of the FID by an exponentially decaying function. The baseline is fitted as the product of the original FID and an exponentially decreasing apodization function (Campbell et al., 1973; Simonetti et al., 2003), which requires only the decay as parameter to be selected. As the baseline consists of broad fast decaying components, their main contribution is in the first part of the FID. After multiplication with a decaying function, the resulting function will mainly contain the broad baseline components.
- Truncation of the initial points in the FID (Ratiney et al., 2004). By removing the initial part of the FID, the baseline components are sepa-

rated from the metabolite contributions. This approach also assumes the baseline corresponds to the first time points in the FID.

Feature selection and extraction applied to brain tumour diagnosis

Based on MR data, several features can be used for brain tumour diagnosis. Pattern recognition can be based on the complete spectrum or on selected or extracted features. The goal of feature selection and feature extraction is to retain the most characteristic information, to minimize redundancy and to exclude noise and artefacts in the spectrum. The process of feature selection and extraction has even more advantages as it performs dimensionality reduction, decreases the amount of complexity and simplifies the calculation. Feature extraction can be carried out by for example Principal Component Analysis (PCA, section 3.3.1). However, also biomedically relevant features can be exploited, obtained by e.g. quantification.

It is well known that characteristic resonance peaks correspond to important brain metabolites (Mukherji, 1998; Smith and Stewart, 2002; Howe et al., 2003; Leclerc et al., 2002; Majós et al., 2002; Murphy et al., 1993). It seems reasonable then, that these peaks might be used as discriminatory features to distinguish tumour types, in particular for those spectral regions which are clearly different between spectra of different tumour types. Thus, as an alternative to using complete spectra, selected frequency regions or parameter estimates obtained by quantification can be used which are assumed to contain most of the information. Hence, the redundancy generated by spectral noise and artefacts in the spectrum could be reduced. Selected frequency regions are taken as intervals around the resonance frequencies (Govindaraju et al., 2000) of several characteristic metabolites. Estimates of the amplitudes of several metabolite resonances can be obtained by peak integration (Hoch and Stern, 1996) around the selected resonances or by a more accurate model-based approach (Lemmerling et al., 2002; Provencher, 1993; Vanhamme et al., 1997).

At long echo time (e.g. Szabo De Edelenyi et al., 2000), the metabolite peaks of interest are *L2* (lipids at $\approx$ 0.90 ppm); *L1* (lipids at $\approx$ 1.25 ppm); *Lac* ($^3CH_3$-group at central frequency 1.31 ppm); *Ala* (alanine; $^1CH_3$-group at central frequency 1.47 ppm); *NAc* (N-acetyl containing compounds; $^2CH_3$-group 2.01 ppm); *Cr* (total creatine, $N(CH_3)$-group at 3.03 ppm and $^2CH_2$-group at 3.91 ppm); *Cho* (choline containing compounds, $N(CH_3)_3$-group at 3.18 ppm); *Gly* (glycine; $^2CH_2$-group at 3.54 ppm). The notation of the molecular groups (between brackets) corresponds to the notation of Govindaraju et al. (2000) and follows the IUPAC nomenclature (IUPAC, 2000).

At short echo time (e.g. Simonetti et al., 2004), an extended set of metabolite peaks can be selected, because several additional metabolites (e.g. *Glx* (glutamate/glutamine), *mI*, *Tau* (taurine)) are more visible in short echo time

spectra. This additional set covers metabolite peaks of for example *Glx* ($^3CH_2$-, $^4CH_2$-groups; at several resonances between 2.03 and 2.45 ppm; ^{2}CH-groups at 3.75 ppm); *mI* (^{5}CH-, ^{1}CH-, ^{3}CH-, ^{4}CH- and ^{6}CH-groups at 3.26, 3.52 and 3.61 ppm); *Tau* ($^2CH_2$- and $^1CH_2$-groups at 3.24 and 3.42 ppm); *Ala* (^{2}CH-group at 3.77 ppm).

Furthermore, also amplitude ratios can be provided as input features to discriminate brain tumours. Opstad et al. (2004) calculated the lipid peak area (LPA) ratio as the ratio of the amplitude of the peaks around 1.25 ppm over that of the peaks around 0.90 ppm (in high-grade tumours mainly containing lipids), resulting in a good classification of glioblastomas and metastases.

Tate et al. (1998) proposed another approach based on correlation analysis to select the most important features. The method searches for the resonance frequencies with the highest correlation coefficient between their intensity value (in the frequency domain) and their class label. Related resonance frequencies are avoided by selecting a threshold on the intercorrelation coefficient between their intensities. The feature selection is performed for a binary classification problem, but also multiclass classification can be performed using a common set of features. Although this approach is statistically based, the results could be largely affected by the possible misalignment of spectra.

3.2.3 The construction and testing of a classifier

In pattern recognition a distinction is made between unsupervised and supervised techniques, based on their approach to the learning process. Unsupervised techniques try to group samples based on their similarities between each other, without knowing the exact grouping. The similarity measure can be based on the distance of the samples in the multi-dimensional feature space constructed by the observed data features. Clustering techniques (Duda et al., 2001), Self-Organizing Maps (SOMs) (Kohonen, 2001), Principal Component Analysis (PCA) (Duda et al., 2001) are a few common unsupervised techniques.

Supervised techniques explicitly exploit the concept of a training and test phase. The learning or training process is guided by a virtual "teacher" providing the target values for each of the training samples. Training is performed to construct a classifier that separates the training data in the feature space in an optimal manner according to some criterion. The resulting classifier is then tested on a test set of unseen data and the test results are evaluated with respect to the correct test targets. Typically 2/3 of the samples are randomly selected as training set and the remainder as independent test set. K-Nearest Neighbours (Duda et al., 2001) and Artificial Neural Networks (Bishop, 1995) are a few of the most well-known supervised techniques.

Although it is desirable to build a classifier with the highest performance, one

should avoid overfitting. In case of overfitting the classifier almost perfectly fits the training data, but can not generalize on an independent test set. This leads to a low performance upon the insertion of new samples. A common setup to avoid overfitting is by introducing a validation set which is used to evaluate the classification performance during training (Figure 3.2). In this scheme the total number of data is divided into a training set, a validation set and a test set. The classifier is trained with the training set, while the validation set is used to decide when to stop training (i.e. when a minimal error on this validation set is reached). The test data are only used after the training phase in order to measure the generalization performance on unseen data.

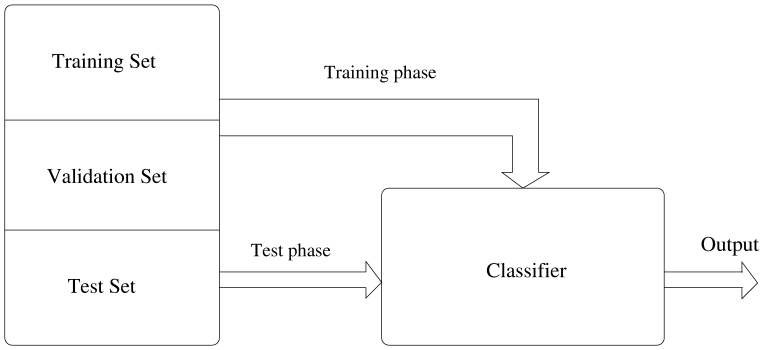

Figure 3.2: The data set is divided into a training, validation and test set. The validation set is used in the training process to obtain a good generalization, e.g. to avoid overfitting. The test set is used to measure the performance of the classifier on the independent or unseen data.

3.2.4 Performance measures

Once a classifier is developed, it should be possible to validate the classifier and to compare the results of several classifiers. In order to assess the classification performance, many performance measures can be utilized. A performance measure estimates the classification accuracy which should objectively reflect the predictive quality of the classifier in assigning unseen data samples to the correct classes.

The correct classification rate shows the proportion of correctly assigned test samples to the total number of test samples, while the error rate is defined as the proportion of incorrectly assigned samples to the total number of samples. The correct classification rate can be calculated using several methods; for example using a subset of independent test samples, leave-one-out cross-validation (LOOCV) or V-fold cross-validation. Another approach is to exploit Receiver Operating Characteristic (ROC) curve analysis, which is often used in medicine

to assess the quality of statistical techniques.

Correct classification rate

The correct classification rate could be estimated by either one of the following methods:

- Training performance rate. The correct classification rate is estimated when using the training samples as test samples. This measure, also called the apparent error rate, is known to be biased towards the training samples as the classifier is constructed and tested according to the same samples and it is likely to overestimate the performance.

- Correct classification rate on a subset of independent test samples (cf. section 3.2.3). The data set is randomly split into a training set, from which the classifier is constructed and a test set from which the correct classification rate estimate is determined. Often 2/3 of the samples is selected for training and the remaining for test. This process can be carried out iteratively and the average performance can be computed to reduce variance. This iterative randomization is also called repeated subsampling (Kohavi, 1995), repeated learning-testing (Burman, 1990) or random subsampling (Weiss and Kulikowski, 1991).

- V-fold cross-validation. The data set is randomly divided into V subsets of approximately equal size. For each of the subsets, the remaining $V-1$ subsets are combined to form a training set and the resulting classifier's correct classification rate is estimated on the remaining subset. The cross-validation estimate is the average of the V correct classification rate estimates.

- Leave-one-out cross-validation (LOOCV). The leave-one-out estimate is obtained by V-fold cross-validation, where V equals the size N of the data set. LOOCV can be used as a method to estimate the classifier performance in an unbiased manner.

In stratified random sampling, the training and test sets are stratified such that they contain approximately the same proportions of labels as the original data set. In an analogous manner, stratified cross-validation implies that each fold contains approximately the same proportion of labels. The use of stratified random sampling guarantees a prescribed number of observations from each subpopulation, where as the use of a simple random sampling can result in the underrepresentation of some subclasses. Furthermore, using stratification makes the stratified sample mean more precise than the mean of a simple random sample (Rice, 1995).

Receiver Operating Characteristic

In clinical medicine, there does not exist any perfect test. Every test will have an error rate and on occasion will either fail to identify an abnormality, or identify an abnormality which is not present. It is common to describe this error rate by the terms *true and false positive* and *true and false negative* as follows :

True Positive (TP) : the number of positive test results in the presence of the clinical abnormality.

True Negative (TN) : the number of negative test results in the absence of the clinical abnormality.

False Positive (FP) : the number of positive test results in the absence of the clinical abnormality.

False Negative (FN) : the number of negative test results in the presence of the clinical abnormality.

A tabular format, called contingency table is illustrated in Table 3.1. As the presentation of test results as true and false positives and true and false negatives can be difficult to understand, statisticians have introduced terms like sensitivity and specificity to describe the efficiency of a clinical test. Table 3.2 defines many of these terms based on TP, TN, FP, FN. Sensitivity and specificity are the basic measures of accuracy of a diagnostic test. The sensitivity is the proportion of all cases of clinical abnormality which have an abnormal test result, while the specificity is the proportion of all cases with no clinical abnormality which have a normal test result. The actual proportions of true positives and true negatives detected by a diagnostic test are arbitrarily set according to the position of the cut-off point above or below which the result of a data sample becomes negative or positive. There is a tradeoff between sensitivity and specificity in determining the position of this cut-off point as the sensitivity increases while the specificity decreases, or vice versa.

	Predicted Group	
Actual Group	Normal	Abnormal
Normal	TN	FP
Abnormal	FN	TP

Table 3.1: Contingency table commonly used in clinical medicine to assess the classifier's performance in correctly predicting the clinical abnormality to the true diagnosis.

Receiver Operating Characteristic (ROC) curve analysis (Obuchowski, 2003;

Sensitivity	$TP/(TP+FN)$
Specificity	$TN/(FP+TN)$
Predictive value (positive)	$TP/(FP+TP)$
Predictive value (negative)	$TN/(FN+TN)$
Relative risk	$\frac{TP/(FP+TP)}{FN/(FN+TN)}$
Likelihood ratio	$\frac{TP/(TP+FN)}{FP/(TN+FP)}$ or $\frac{\text{Sensitivity}}{1-\text{Specificity}}$
Accuracy	$(TP+TN)/total$

Table 3.2: The definition of various terms used to describe clinical efficiency of a test

Swets, 1979) is an objective and highly effective technique for assessing the performance in binary classification or a diagnostic test. A test sample is assigned to one of two groups or classes (e.g. normal and abnormal) depending on the output value of the diagnostic test relative to a certain cut-off point. ROC curve analysis provides measures that are independent of the position of the cut-off point, in contradiction to the sensitivity and the specificity.

The ROC curve is a graphical presentation of sensitivity versus 1-specificity (or false positive rate) as the cut-off point is varying. Each point on the graph is generated by using a different cut-off point (Obuchowski, 2003). The area under the ROC curve (AUC) gives then a global measure of the clinical efficiency over a range of test cut-off points on the ROC curve. For example as shown on Figure 3.3 a test with an AUC of 1.0 is perfectly accurate as the sensitivity is 1.0 when the specificity is 1.0 (perfect test). In contrast, a test with an AUC of 0.0 is perfectly inaccurate. The line segment from (0,0) to (1,1) has an area of 0.5 and is called the chance diagonal (Figure 3.3). Tests with an AUC value larger than 0.5 have at least some discrimination ability. The closer the AUC reaches 1.0, the better the diagnostic test. The AUC can also be interpreted as the average value of sensitivity for all possible values of specificity (Obuchowski, 2003). Hence, the AUC is independent of the cut-off points used or the prevalence of disease and is therefore a good summary measure of test accuracy.

ROC analysis and the AUC calculation can be used to choose the best of a host of different candidate diagnostic tests or to establish a tradeoff between the number of false positives and false negatives.

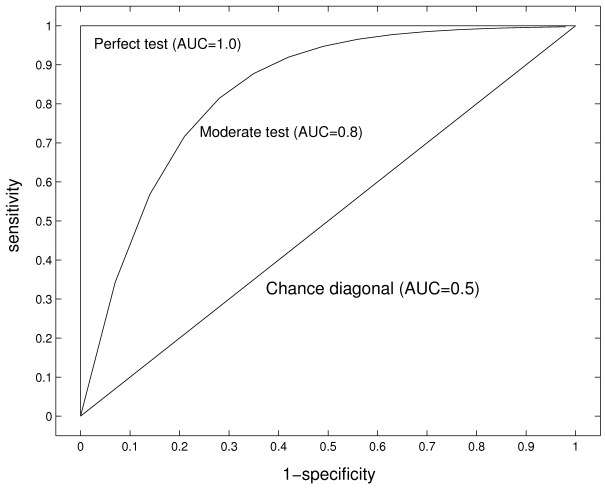

Figure 3.3: ROC curve for different cases. Area under ROC curve (AUC) serves as a well-established index of diagnostic accuracy. ROC following diagonal line results in AUC=0.5 (chance diagonal), whereas the maximum value of 1.0 corresponds to perfect assignment (unity sensitivity for all values of specificity).

Statistical analysis

This section describes the statistical comparison of the areas under the ROC curves derived from the same set of patients by taking into account the correlation between the areas that is induced by the paired nature of the data. This can be used to compare the performance of different classifiers which are applied on M runs of stratified randomized splittings of the data set into a training and test set (stratified random subsampling).

Consider two classifiers C_1 and C_2 that handle the same input data; e.g. C_1 is classifier 1 and C_2 is classifier 2 applied to separate input into class 1 and 2. Let the AUC of each classifier $C_i, i = 1, 2$ be $A_{i,l}$ with standard error $SE_{i,l}, i = 1, 2; l = 1, ..., M$, with M the number of stratified randomizations ($M = 100$). The pooled statistics are then given by $\overline{A_i}$ and SE_i ($i = 1, 2$), where n_l is the amount of samples for the stratified randomization $l = 1, ..., M$ (Rice, 1995; Montgomery and Runger, 1994):

$$\overline{A_i} = \frac{1}{M} \sum_{l=1}^{M} A_{i,l}, \tag{3.1}$$

$$SE_i = \sqrt{\frac{1}{N-M}\sum_{l=1}^{M}(n_l-1)SE_{i,l}^2} \quad (3.2)$$

$$= \sqrt{\frac{1}{M}\sum_{l=1}^{M}SE_{i,l}^2} \quad (3.3)$$

The last line is only satisfied if the test set contains an equal amount of samples for each stratified randomization, i.e. $\forall\, l\ n_l = n, N = \sum_{l=1}^{M} n_l$.

A general approach to statistically test whether the areas under two ROC curves derived from the same samples differ significantly from each other is then given by the critical ratio z, defined as (Hanley and McNeil, 1983):

$$z = \frac{\overline{A_1} - \overline{A_2}}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}},$$

in which r is a quantity representing the correlation introduced between the two areas by studying the same samples. The z-value can be calculated using Equations (3.1)-(3.3) based on the pooled statistics $\overline{A_i}, SE_i, i = 1, 2$ from M runs. If the resulting z-value satisfies $z \geq 1.96$, then $\overline{A_1}$ and $\overline{A_2}$ are statistically different. The cutoff-value 1.96 is taken as the quantity for which, under the hypothesis of equal AUCs ($\overline{A_1} = \overline{A_2}$), $z \geq 1.96$ occurs with a probability of $\alpha = 0.05$ assuming a normal distribution.

This ROC analysis can be performed for binary classification. Although ROC analysis has been extended to multiclass classification (Srinivasan, 1999), the result is generally non-intuitive and computationally expensive. This motivates the use of the correct classification rate as performance measure for multiclass classification.

3.3 Pattern recognition techniques

3.3.1 Unsupervised techniques

The aim of unsupervised techniques can be manifold. For instance, it might be of interest to use a large amount of unlabelled data to train and only use supervision to label the groupings found. This may be appropriate for large "data mining" applications, in which the contents of a large database is not known a priori. Unsupervised techniques can also be valuable to perform exploratory data analysis and gain insight into the structure or nature of the data. This might lead to the discovery of subclasses or certain patterns, distinct from the expected characteristics, which may influence the approach to

designing the classifier. A final important purpose of unsupervised techniques is to find characteristic features and continue classification by only retaining the most valuable information. This might be beneficial in several research areas, such as *in vivo* NMR, in which data can be generated at a relatively high dimension with respect to the amount of data, involving a high complexity. Such an approach minimizes the redundancy and noise and decreases the amount of complexity. In this section we will discuss Principal Component Analysis (PCA) as a typical unsupervised technique. For an overview of other unsupervised techniques we refer to Duda et al. (2001).

Principal Component Analysis

Linear Principal Component Analysis (PCA) is one of the most common statistical methods to reduce high-dimensional data. It consists of a transformation of the original variables into a new set of uncorrelated variables, called principal components (PCs). These new variables are linear combinations of the original ones, ranked based on the variance they retain. If the original variables are highly correlated, then the first few PCs will account for most of the variation and the remaining PCs can be discarded with slight loss of information. The more PCs are used the better the reconstructed data fit the original noisy data.

Basically, PCA projects data onto an orthogonal space, where the axes of this new coordinate system lie along the direction of maximum variance of the original data. By this transformation, it allows the mapping of vectors $\mathbf{x} \in \mathbb{R}^n$ onto lower-dimensional vectors $\mathbf{z} \in \mathbb{R}^m$ with $m < n$. The covariance matrix can be estimated by

$$\hat{\mathbf{S}} = \frac{1}{N-1} \sum_{k=1}^{N} (\mathbf{x}_k - \mu)(\mathbf{x}_k - \mu)^T, \tag{3.4}$$

where $\mu = \frac{1}{N} \sum_{k=1}^{N} \mathbf{x}_k$, such that $(\mathbf{x}_k - \mu), k = 1 \ldots, N$ represent the centered data. From a symmetric matrix such as $\hat{\mathbf{S}}$, we can calculate an orthogonal basis by finding its eigenvalues and eigenvectors. The eigenvectors $\mathbf{u}_i, i = 1, \ldots, n$ and the corresponding eigenvalues λ_i are the solutions of the following equation

$$\hat{\mathbf{S}}\mathbf{u}_i = \lambda_i \mathbf{u}_i,\ i = 1, \ldots, n,\ \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0, \tag{3.5}$$

which corresponds to the eigenvalue decomposition of $\hat{\mathbf{S}}$. The eigenvector $\mathbf{u}_1$ associated with the largest eigenvalue λ_1 has the same direction as the first principal component, while the eigenvector $\mathbf{u}_2$ with eigenvalue λ_2 determines the direction of the second principal component. The sum of the eigenvalues equals the trace of the covariance matrix. Taking the m largest eigenvalues

and the corresponding eigenvectors, the scores or transformed variables can be calculated using

$$z_i = \mathbf{u}_i^T(\mathbf{x} - \mu),\ i = 1, ..., m, \tag{3.6}$$

such that $\mathbf{z} = (z_1, \ldots, z_m)^T$ is a point in the orthogonal coordinate system defined by the eigenvectors. The components $z_i, i = 1, ..., m$, are the coordinates in the orthogonal basis.

PCA is commonly applied in signal processing and data analysis to eliminate a significant number of dimensions by partitioning the feature space into subspaces of signal and noise. This assumes that most of the variance in the data can be explained by the retained components and small contributions by the noise. For every application, a decision should be made on the amount of principal components to retain in order to efficiently summarize the data. Rencher (1995) mentions four possible criteria as guidelines:

1. Retain those components that account for more variance than the average variance of the individual variables.
2. Take the largest components that explain a certain prespecified percentage of total variance, e.g. 80% or 85%.
3. Use a scree graph, a plot of λ_i versus i. This attempts to visualize a gap between the large eigenvalues and the small eigenvalues.
4. Test the significant contribution of the components corresponding to the larger eigenvalues. This can be done by testing whether the last eigenvalues are small and equal to each other.

3.3.2 Supervised techniques

The supervised approach uses a "training set" to build a statistical classifier which classifies the samples in an optimal manner and is tested on an independent "test set". There exist several types of supervised techniques that can be applied to MR data. For an overview of supervised techniques like neural networks, k-nearest neighbour analysis and decision trees we refer to Duda et al. (2001) and Lindon et al. (2001). This section mainly focuses on Linear Discriminant Analysis (LDA) as a classical technique and Support Vector Machines (SVM) and Least-Squares Support Vector Machines (LS-SVM) as kernel-based techniques.

Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) originally was designed by Fisher (1936) for taxonomic classification, but has been widely applied for several classification purposes. Consider the problem of projecting data from n dimensions onto a one-dimensional space. Even if the samples formed well-separated, compact clusters in n-space, their projection onto an arbitrary line probably will produce a confused mixture of samples from all classes and thus produce no separation ability. By rotating the line around, there might be an orientation from which a good separation of the projected samples is possible. Hence, the goal of LDA is to find the underlying space that best discriminate among classes.

Assume we have a set of n-dimensional samples $\mathbf{x}_1^1, \mathbf{x}_2^1, \ldots, \mathbf{x}_{N_1}^1, \mathbf{x}_1^2, \mathbf{x}_2^2, \ldots, \mathbf{x}_{N_2}^2$, $N = N_1 + N_2$ (Figure 3.4), of which the first N_1 belong to class 1, and the last N_2 to class 2. We seek to obtain a scalar z by projecting the samples $\mathbf{x}$ onto a line: $z_k^l = \mathbf{w}^T \mathbf{x}_k^l$, $k = 1, \ldots, N_l, l = 1, 2$. Of all possible lines we would like to select the one that maximizes the separability of the scalars z_k^l. In order to find a good projection, a measure of separation between the projected samples is defined. The mean values of the $\mathbf{x}_k^l$ and $z_k^l, k = 1, \ldots, N_l, l = 1, 2$ samples are:

$$\mu_l = \frac{1}{N_l} \sum_{k=1}^{N_l} \mathbf{x}_k^l,$$

$$\tilde{\mu}_l = \frac{1}{N_l} \sum_{k=1}^{N_l} z_k^l \, l = 1, 2. \tag{3.7}$$

(3.8)

We could define the distance between the projected means as our objective function:

$$J(\mathbf{w}) = |\tilde{\mu}_2 - \tilde{\mu}_1| = |\mathbf{w}^T(\mu_2 - \mu_1)|. \tag{3.9}$$

However, to obtain a good separation of the projected data, this distance does not take into account the standard deviation within the classes. The solution proposed by Fisher is to maximize a function representing the difference between the projected means relative to some measure of the standard deviation for each class. We define the within-class scatter, an equivalent of the variance, as the sum of $\tilde{s}_1^2$ and $\tilde{s}_2^2$, given by (Duda et al., 2001):

$$\tilde{s}_l^2 = \sum_{k=1}^{N_l} (z_k^l - \tilde{\mu}_l)^2, \; l = 1, 2. \tag{3.10}$$

The Fisher linear discriminant is then defined as the linear function $\mathbf{w}^T\mathbf{x}$ that maximizes the criterion function:

$$J(\mathbf{w}) = \frac{|\tilde{\mu}_2 - \tilde{\mu}_1|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}. \tag{3.11}$$

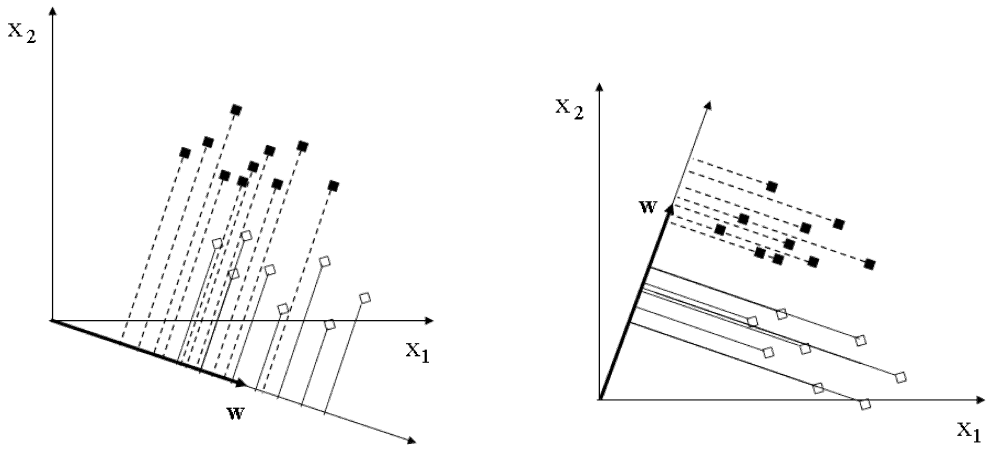

Figure 3.4: Two possible projections of the same samples from two classes, represented by the black and white squares, onto two different lines along the vector $\mathbf{w}$ are shown. The projection on the right enables better separation of the classes.

Therefore, LDA searches for a projection where examples from the same class are projected very close to each other and, at the same time, the projected means are as far apart as possible. In order to find the optimal $\mathbf{w}$, called $\mathbf{w}^*$, the function $J(\mathbf{w})$ is defined as an explicit function of $\mathbf{w}$. The equivalent of the scatter in the projection is defined in the original multivariate feature space, corresponding to the scatter matrices:

$$\mathbf{S}_l = \sum_{k=1}^{N_l} (\mathbf{x}_k^l - \mu_l)(\mathbf{x}_k^l - \mu_l)^T,\ l = 1, 2,$$
$$\mathbf{S}_w = \mathbf{S}_1 + \mathbf{S}_2. \tag{3.12}$$

The matrix $\mathbf{S}_w$ is called the within-class scatter matrix, which is related to the within-class scatter of the projection as:

$$\tilde{s}_1^2 + \tilde{s}_2^2 = \mathbf{w}^T \mathbf{S}_w \mathbf{w}, \tag{3.13}$$

and similarly the between-class scatter $\mathbf{S}_b$ is defined as:

$$\mathbf{S}_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T, \tag{3.14}$$

and is related to the distance between the projected means as

$$(\tilde{\mu}_2 - \tilde{\mu}_1)^2 = \mathbf{w}^T \mathbf{S}_b \mathbf{w}, \tag{3.15}$$

the Fisher criterion can then be expressed in terms of the within-class and between-class scatter matrices:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}. \tag{3.16}$$

For the c-class problem, the natural generalization of Fisher's LDA involves $c-1$ discriminant functions. The projection is from an n-dimensional space to a $(c-1)$-dimensional space, where it is assumed that $n > (c-1)$. The within-class scatter matrix is generalized as (Duda et al., 2001):

$$\begin{aligned} \mathbf{S}_w &= \sum_{l=1}^{c} \mathbf{S}_l, \\ \mathbf{S}_l &= \sum_{k=1}^{N_l} (\mathbf{x}_k^l - \mu_l)(\mathbf{x}_k^l - \mu_l)^T, \\ \mu_l &= \frac{1}{N_l} \sum_{k=1}^{N_l} \mathbf{x}_k^l,\ l = 1, \ldots, c. \end{aligned} \tag{3.17}$$

In order to generalize the between-class scatter matrix, the generalization of the total scatter matrix is defined as the sum of the within-class and the between-class scatter matrices:

$$\begin{aligned} \mathbf{S}_t &= \mathbf{S}_w + \mathbf{S}_b, \\ \mathbf{S}_t &= \sum_{l=1}^{c} \sum_{k=1}^{N_l} (\mathbf{x}_k^l - \mu)(\mathbf{x}_k^l - \mu)^T, \\ \mu &= \frac{1}{N} \sum_{l=1}^{c} (N_l)\mu_l,\ l = 1, \ldots, c. \end{aligned} \tag{3.18}$$

From Equations (3.17) and (3.18), the formula for the between-scatter matrix can be derived as:

$$\mathbf{S}_b = \sum_{l=1}^{c} N_l (\mu_l - \mu)(\mu_l - \mu)^T. \tag{3.19}$$

$$\tag{3.20}$$

When dealing with c classes, we seek for $(c-1)$ projection vectors $\mathbf{w}_i$, which are arranged as columns into a projection matrix $\mathbf{W} = [\mathbf{w}_1|\mathbf{w}_2|\ldots|\mathbf{w}_{c-1}]$ so that $\mathbf{z} = \mathbf{W}^T\mathbf{x}$. It can then be derived that the criterion function becomes (Duda et al., 2001):

$$J(\mathbf{W}) = \frac{\mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\mathbf{W}^T \mathbf{S}_w \mathbf{W}}. \tag{3.21}$$

A common way to maximize Equation (3.21) is by maximizing the ratio $\frac{\det(\mathbf{S}_b)}{\det(\mathbf{S}_w)}$. It has been proven (Fisher, 1936) that if $\mathbf{S}_w$ is a nonsingular matrix then this ratio is maximized when the column vectors of the optimal projection matrix, $\mathbf{W}$, are the eigenvectors that correspond to the largest eigenvalues of $\mathbf{S}_w^{-1}\mathbf{S}_b$. However, as $\mathbf{S}_b$ is the sum of c matrices of rank one or less, and only $c-1$ of these are independent, $\mathbf{S}_b$ is of rank $c-1$ or less. This implies there are at most $c-1$ nonzero eigenvalues. We also require at least $n+c$ samples to guarantee that $\mathbf{S}_w$ does not become singular. Problems arise when dealing with high-dimensional data, which requires a very high amount of samples in order to avoid that the within-class scatter matrix becomes singular. Another disadvantage is the high computational complexity of solving $\mathbf{S}_w^{-1}\mathbf{S}_b$ when working in a high-dimensional input space. One common solution to solve this problem is to first project the data into an intermediate space and then perform classification based on the transformed features. PCA is often utilized for this task; it reduces the dimensionality by selecting the most important parameters and removing the noise. LDA is applied to the PCA lower-dimensional feature space. This PCA+LDA approach was shown to be efficient by empirical testing and was proven to have a theoretical foundation by Yang and Yang (2003).

The combination PCA+LDA requires the implementation of an appropriate criterion to retain a number of components. For example, when selecting a number of components that explain a certain percentage of total variance, a threshold for the total variance should be fixed. On the one hand, a too low threshold could give too few components and a large loss of information. On the other hand, if the threshold is fixed too high, one risks to include components that are specific to the data (Rencher, 1995) and might not be generalizable to independent data. Moreover, this implies that a larger amount of components included does not always guarantee a better test performance. Also in combination with LDA, retaining many components while having only a restricted number of data samples, might result into rank deficiency problems.

Support Vector Machines

Support Vector Machines (SVM) have been introduced by Vapnik (1995) as a powerful machine learning technique for solving classification and nonlinear function estimation problems. The method has already been successfully used for a wide variety of problems such as speech recognition, forecasting, prediction of traffic speed, image classification, protein structure detection, cancer prognosis.

The optimal separating hyperplane

SVMs are based on the concept of hyperplanes defining decision boundaries that separate between samples belonging to different classes. Figure 3.4 illustrates a classical linear classification problem in which the samples of two classes 1

and 2 are easily separated by a linear line. However, most classification tasks are not that simple and often more complex structures are needed in order to make an optimal separation, i.e. correctly classify test samples on the basis of the training samples. It is clear that a correct separation of the classes requires a nonlinear curve. Classification tasks based on drawing separating lines to distinguish between samples of different classes are known as hyperplane classifiers. Support Vector Machines are particularly suited to handle such tasks.

Figure 3.5 illustrates the basic idea behind kernel-based techniques, like SVMs. The original samples are mapped from the input space into a high-dimensional or even infinite-dimensional feature space. In the feature space the classification problem becomes much simpler. Let us describe the construction of such optimal hyperplane in more detail.

Consider the training samples $\{\mathbf{x_k}, y_k\}_{k=1}^{N}$, $\mathbf{x_k} \in \mathbb{R}^n$, $y_k = \pm 1$, either corresponding to class 1 ($y_k = -1$) or class 2 ($y_k = 1$). SVM maps the input vector $\mathbf{x}$ into the high-dimensional feature space Ω through a (non)linear mapping $\varphi : \mathcal{D} \rightarrow \Omega$. It is important to note that the dimension of this space Ω is not specified, it can even be infinite dimensional. A class of hyperplanes can then be given by $\mathbf{w}^T\varphi(\mathbf{x}) + b$, with $b \in \mathbb{R}$ and $\mathbf{w}$ a weighting vector with the same dimension as $\varphi(\mathbf{x})$.

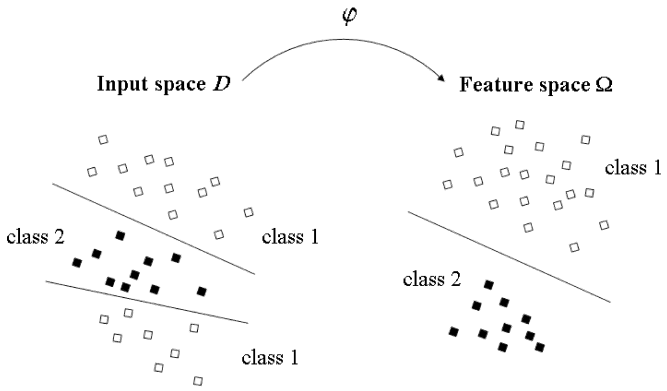

Figure 3.5: In the input space classes 1 and 2 are nonlinearly separable. By mapping the input space to a high-dimensional feature space a linear separation can me made.

In this feature space, we consider the linear function

$$f(\mathbf{x}) = \text{sign}[\mathbf{w}^T\varphi(\mathbf{x}) + b], \tag{3.22}$$

in which a sample $\mathbf{x}$ is assigned to class 1 or 2, if $f(\mathbf{x}) = -1$ or $f(\mathbf{x}) = 1$, respectively.

In case of linearly separable data, a separating hyperplane can be constructed in many ways (Figure 3.6). In order to find the optimal separating hyperplane, a criterion should be added.

The idea of a training phase is to recognize the pattern of a class l by examining the training samples corresponding to that class. This assumes that new samples of that class l have a similar pattern or are lying close to the examined training data. A hyperplane should be chosen such that a small shift of the data does not dramatically alter their class assignment. However, if the distance between the separating hyperplane and the training samples becomes too small, even test samples very close to the training samples may be classified incorrectly. Hence, the generalization ability depends on the distance of the training samples to the hyperplane. SVMs are based on this so-called maximal margin principle (Fig. 3.6), which searches for a hyperplane that maximally separates the classes.

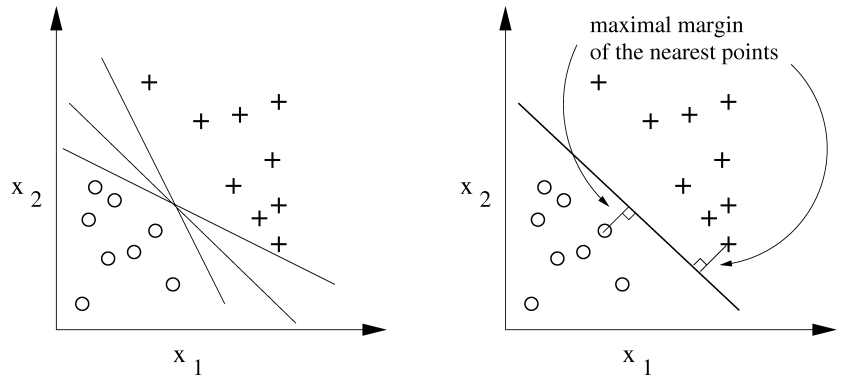

Figure 3.6: Illustration of the maximal margin principle. Linearly separable data do not have a unique separating hyperplane (left). But by using the maximal margin principle between the hyperplane and the closest points a unique solution can be found.

In most real life applications, experimental data are more complex which will make a linear separation impossible. This implies that also a restricted number of misclassifications should be tolerated in the margin, i.e. the region where both classes are overlapping. A violation of the constraints will be penalized.

The SVM classifier starts from the following formulation

$$\mathbf{w}^T\varphi(\mathbf{x}_k) + b \geq +1 \quad \text{for} \quad y_k = +1 \tag{3.23}$$
$$\mathbf{w}^T\varphi(\mathbf{x}_k) + b \leq -1 \quad \text{for} \quad y_k = -1 \tag{3.24}$$

which is equivalent to

$$y_k[\mathbf{w}^T\varphi(\mathbf{x}_k) + b] \geq 1, \ k = 1, ...N \tag{3.25}$$

in the case of separable data. The classification can be written as

$$y(\mathbf{x}) = \text{sign}[\mathbf{w}^T\varphi(\mathbf{x}) + b]. \tag{3.26}$$

Basically, the optimization problem can be written as:

$$\min_{\mathbf{w},b,\xi} \mathcal{J}_P = \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{k=1}^{N}\xi_k \tag{3.27}$$
$$\text{such that} \quad y_k[\mathbf{w}^T\varphi(\mathbf{x}_k) + b] \geq 1 - \xi_k, \ k = 1, ..., N$$
$$\xi_k \geq 0, \ k = 1, ..., N. \tag{3.28}$$

in which $C \in \mathbb{R}$ is a regularization constant. The first set of constraints is based on the maximal margin principle, while the second set of constraints imposes the variables ξ_k to be positive in order to tolerate misclassifications. Using these constraints, the value ξ_k indicates where $\mathbf{x}_k$ lies compared to the separating hyperplane:

- $\xi_k \geq 1 : y_k[\mathbf{w}^T\varphi(\mathbf{x}_k) + b] < 0$, implies that the decision function $f(\mathbf{x})$ and the target y_k of $\mathbf{x}_k$ have a different sign or, in other words, $\mathbf{x}_k$ is misclassified.

- $0 < \xi_k < 1 : \mathbf{x}_k$ is classified correctly, but lies inside the margin.

- $\xi_k = 0 : \mathbf{x}_k$ is classified correctly, and lies outside the margin or exactly on the margin boundary.

The Lagrangian for this problem is:

$$\mathcal{L}(\mathbf{w}, b, \xi; \alpha, \nu) = \mathcal{J}_P(\mathbf{w}, \xi) - \sum_{k=1}^{N}\alpha_k(y_k[\mathbf{w}^T\varphi(\mathbf{x}_k) + b] - 1 + \xi_k) - \sum_{k=1}^{N}\nu_k\xi_k \tag{3.29}$$

with Lagrange multipliers $\alpha_k \geq 0, \nu_k \geq 0$ for $k = 1, ..., N$. The solution is given by the saddle point of the Lagrangian:

$$\max_{\alpha,\nu} \min_{\mathbf{w},b,\xi} \mathcal{L}(\mathbf{w}, b, \xi; \alpha, \nu). \tag{3.30}$$

One obtains

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 & \rightarrow \quad \mathbf{w} = \sum_{k=1}^{N} \alpha_k y_k \varphi(\mathbf{x}_k) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 & \rightarrow \quad \sum_{k=1}^{N} \alpha_k y_k = 0 \\ \frac{\partial \mathcal{L}}{\partial \xi_k} = 0 & \rightarrow \quad 0 \leq \alpha_k \leq C,\ k = 1, ..., N, \end{cases} \tag{3.31}$$

which gives the following dual quadratic programming (QP) problem after substituting (3.31) into (3.29):

$$\begin{aligned} \max_{\alpha} \mathcal{J}_D(\alpha) = \quad & -\frac{1}{2} \sum_{k,l=1}^{N} y_k y_l K(\mathbf{x}_k, \mathbf{x}_l) \alpha_k \alpha_l + \sum_{k=1}^{N} \alpha_k \\ \text{such that} \quad & \sum_{k=1}^{N} \alpha_k y_k = 0 \\ & 0 \leq \alpha_k \leq C,\ k = 1, ..., N. \end{aligned} \tag{3.32}$$

Interestingly, no explicit construction of the nonlinear mapping $\varphi(\mathbf{x})$ is needed. Instead, in this optimization problem the so-called kernel function $K(\mathbf{x}, \mathbf{z})$ is introduced, which is related to $\varphi(\mathbf{x})$. The inner product in the feature space $\varphi(\mathbf{x}_k)^T \varphi(\mathbf{x}_l)$ can be replaced with the corresponding kernel $K(\mathbf{x_k}, \mathbf{x_l})$ satisfying Mercer's condition:

$$K(\mathbf{x}_k, \mathbf{x}_l) = \varphi(\mathbf{x}_k)^T \varphi(\mathbf{x}_l), \tag{3.33}$$

for $k = 1, ..., N$. This relation is often called the kernel trick. It enables to work in a high-dimensional space without explicitly performing calculations in this space. SVMs start from a formulation in the primal weight space (Equation (3.22)) with a high-dimensional feature space by the transformation $\varphi(.)$. Computations are not done in the primal weight space, but in the dual space of Lagrange multipliers after applying the kernel trick. Finally the nonlinear SVM classifier takes the form

$$y(\mathbf{x}) = \text{sign}\left[\sum_{k=1}^{N} \alpha_k y_k K(\mathbf{x}, \mathbf{x}_k) + b\right] \tag{3.34}$$

with α_k positive real constants which are the solution to the QP problem.

Properties of the SVM

Several properties of nonlinear SVM classifiers and their solutions are mentioned in Suykens et al. (2002), of which a few are listed here:

- Choice of kernel function.
  Four common choices of kernels are

  1. $K(x, z) = x^T z$ (linear SVM)
  2. $K(x, z) = (\tau + x^T z)^d$ (polynomial SVM of degree d)
  3. $K(x, z) = \exp(-\|x - z\|_2^2/\sigma^2)$ (RBF kernel)
  4. $K(x, z) = \tanh(\kappa_1 x^T z + \kappa_2)$ (Multilayer Perceptron/MLP)

  Mercer's condition holds for all σ values in the RBF kernel case and τ values in the polynomial case, but not for all possible choices of κ_1, κ_2 in the MLP case.

- Global and unique solution.
  When chosing a positive definite kernel for $K(\cdot, \cdot)$, the solution to the convex QP problem is global and unique. This choice guarantees that the matrix involved in the QP problem is positive definite as well, and the kernel trick is applicable. For a positive semi-definite kernel the solution to the QP problem is global but not necessarily unique.

- Sparseness.
  An interesting property is that many of the resulting α_k values are equal to zero. Hence the obtained solution is sparse. This means that in the resulting classifier the sum should be taken only over non-zero α_k values instead of all training data points:

$$y(\mathbf{x}) = \text{sign}[\sum_{k=1}^{\#SV} \alpha_k y_k K(\mathbf{x}, \mathbf{x}_k) + b]$$

  where $\#SV$ is the number of support vectors $\mathbf{x}_k$ of the training set. The corresponding nonzero α_k values are called the support values.

- Geometric meaning of the support vectors.
  The vectors $x_k \in \mathbb{R}^n$ with corresponding nonzero α_k values are called the support vectors. They are located in the boundary margin and contribute to the construction of the separating hyperplane.

Least Squares Support Vector Machines

Suykens and Vandewalle (1999) have proposed to modify the SVM of Vapnik by introducing a least squares term in the cost function. This transformed the

problem from solving a QP problem to solving a set of linear equations, which significantly reduces the complexity and computation time.

The LS-SVM formulation considers the following optimization problem:

$$\min_{\mathbf{w},b,e} \mathcal{J}_P = \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + \gamma \sum_{k=1}^{N} e_k^2 \tag{3.35}$$

$$\text{such that} \quad y_k[\mathbf{w}^T\varphi(\mathbf{x}_k) + b] = 1 - e_k,\ k = 1, ..., N, \tag{3.36}$$

which corresponds to the classifier written as in Equation (3.26). Note that the inequality constraints of Vapnik's SVM are modified to the equality constraints, where the value 1 at the right hand side is considered as a target value, instead of a threshold value. Some tolerance of misclassification is allowed by the error variable e_k in the case of overlapping distributions, which is in analogy to the function of the variable ξ_k in the standard SVM. The objective function contains a squared term of this error variable.

The Lagrangian can be constructed from Equation (3.35):

$$\mathcal{L}(\mathbf{w}, b, e; \alpha) = \mathcal{J}_P(\mathbf{w}, e) - \sum_{k=1}^{N} \alpha_k \{y_k[\mathbf{w}^T\varphi(\mathbf{x}_k) + b] - 1 + e_k\} \tag{3.37}$$

where α_k are the Lagrange multipliers. The conditions for optimality yield:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 \quad \rightarrow \quad \mathbf{w} = \sum_{k=1}^{N} \alpha_k y_k \varphi(\mathbf{x}_k) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 \quad \rightarrow \quad \sum_{k=1}^{N} \alpha_k y_k = 0 \\ \frac{\partial \mathcal{L}}{\partial e_k} = 0 \quad \rightarrow \quad \alpha_k = \gamma e_k,\ k = 1, ..., N \\ \frac{\partial \mathcal{L}}{\partial \alpha_k} = 0 \quad \rightarrow \quad y_k[\mathbf{w}^T\varphi(\mathbf{x}_k) + b] - 1 + e_k = 0,\ k = 1, ..., N. \end{cases} \tag{3.38}$$

These can be written as the linear system (Fletcher, 1987):

$$\left[\begin{array}{ccc|c} \mathbf{I} & 0 & 0 & -\mathbf{Z}^T \\ 0 & 0 & 0 & -\mathbf{Y}^T \\ 0 & 0 & \gamma\mathbf{I} & -\mathbf{I} \\ \hline \mathbf{Z} & \mathbf{Y} & \mathbf{I} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{w} \\ b \\ \mathbf{e} \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline \mathbf{1} \end{array}\right] \tag{3.39}$$

where $\mathbf{Z} = [\varphi(\mathbf{x}_1)^T y_1 ... \varphi(\mathbf{x}_N)^T y_N]^T$, $\mathbf{Y} = [y_1 ... y_N]^T$, $\mathbf{1} = [1 ... 1]^T$, $\mathbf{e} = [e_1 ... e_N]^T$, $\alpha = [\alpha_1 ... \alpha_N]^T$. Elimination of $\mathbf{w}$ and $\mathbf{e}$ gives

$$\left[\begin{array}{c|c} 0 & \mathbf{Y}^T \\ \hline \mathbf{Y} & \mathbf{Z}^T\mathbf{Z} + \mathbf{I}/\gamma \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline \mathbf{1} \end{array}\right]. \tag{3.40}$$

Mercer's condition is applied to the matrix $\Omega = Z^T Z$ with

$$\begin{aligned} \Omega_{kl} &= y_k y_l \, \varphi(\mathbf{x}_k)^T \varphi(\mathbf{x}_l) \\ &= y_k y_l \, K(\mathbf{x}_k, \mathbf{x}_l), \; k, l = 1, ...N. \end{aligned} \tag{3.41}$$

The resulting classifier in the dual space is given by

$$y(\mathbf{x}) = \text{sign}\left[\sum_{k=1}^{N} \alpha_k y_k K(\mathbf{x}, \mathbf{x}_k) + b\right], \tag{3.42}$$

similar to the standard SVM classifier.

Properties of the LS-SVM
We mention the following properties of the LS-SVM classifier:

- Choice of kernel function.
  A chosen kernel function should be positive definite and satisfy the Mercer condition. The comments on the kernel functions in the SVM also apply to the kernels used in the LS-SVM formulation. The choice of the regularization parameter γ can be done by using a Bayesian framework (Suykens et al., 2002). Nevertheless, the hyperparameters are most often chosen based on a leave-one-out cross-validation procedure, while bounding the grid of possible hyperparameter values to avoid overfitting.
- Global and unique solution.
  The dual problem for linear and nonlinear LS-SVMs corresponds to solving a linear system which is a square system with a unique solution when the matrix has full rank.
- Lack of sparseness.
  Each α_k value is proportional to the error of the corresponding sample. Hence, nonzero support values α_k are spread over all training samples and every training sample is a support vector. No sparseness property raises as in the standard SVM formulation. But, interestingly, in the LS-SVM case one can relate a high support value to a high contribution of the training sample on the decision line.

3.4 Conclusions

In this chapter an overview was given of a typical pattern recognition process applied to MR data of brain tumours. This included the main concepts of pattern recognition, such as data collection, preprocessing and feature selection, construction and testing of the classifier and performance assessment. Although

there is no guarantee for success, e.g. when dealing with difficult problems, an appropriate design of each of these steps should facilitate the development of a reliable and successful classification tool. It can be favorable to perform a prior exploratory data analysis in order to gain more insight in the data, to detect the underlying structure of the data or to detect the important features. This may significantly alter the classification strategy and improve the resulting performance.

Also the classification techniques, unsupervised and supervised, were explained that are applied in this thesis. Linear Discriminant Analysis (LDA) is a common supervised linear technique to categorize data into different classes. However, in certain practical applications the dimensionality of the data in the problem space may be very high and classification can become computationally very expensive. In order to tackle this problem, dimensionality reduction is often applied prior to LDA classification. Principal Component Analysis (PCA), as unsupervised technique, is a common technique for dimensionality reduction in order to achieve a better classification. Alternatively, more advanced kernel-based techniques, such as SVMs and LS-SVMs can be applied.

Support Vector Machines (SVM) is a method of calculating the optimal separating hyperplane in the feature space. The optimal separating hyperplane is defined as the maximal margin hyperplane in the higher-dimensional feature space. Vapnik suggested using the kernel trick, which causes the algorithm to operate in a different space than the input space.

Least Squares SVM (LS-SVM) simplifies the formulation by replacing the inequality constraint in SVM with equality constraints. This approach significantly reduces the complexity and the computation time, solving a set of linear equations instead of by solving a QP problem.

Chapter 4

Classification of brain tumours using long echo time ^{1}H MRS

In this chapter, an objective comparison of several classification techniques is described for the discrimination of brain tumours. Meningiomas, glioblastomas, astrocytomas grade II and metastases are considered, which are four common brain tumour types. Linear and nonlinear classifiers are compared based on their classification results and advantages and disadvantages. Classification is performed by Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and Least Squares SVM (LS-SVM) with a linear kernel as linear techniques and LS-SVM with a Radial Basis Function (RBF) kernel as a nonlinear technique. The analysis includes binary as well as multiclass classification and follows several guidelines mentioned in chapter 3.

Section 4.1 motivates the study and stresses the main contributions of the study. Section 4.2 describes the data used for classification, while section 4.3 describes the experimental setting for classification. Classification is performed using spectra and extracted features such as selected frequency regions and peak integration values. Section 4.4 summarizes the results of binary classification to discriminate brain tumours. To continue, results of the multiclass classification approach are also mentioned. Afterwards, section 4.5 discusses the classification performance of the classifiers, the limitations of the data set and the influence of dimensionality reduction. Finally, section 4.6 presents the conclusions.

4.1 Introduction

In comparison to *in-vitro* spectroscopy, *in-vivo* spectroscopy signals are more difficult to analyze because of their broader resonances, strongly overlapping peaks, lower signal-to-noise ratio and higher number of artefacts. Cousins (1995) discusses the influence of the echo time TE on the spectral pattern of a MRS signal. Short echo time ^{1}H MRS signals contain more information about several metabolites (e.g. *Glx*, *mI*, *Tau*) that are more visible in short echo time spectra. However, short echo time ^{1}H MRS signals are more difficult to analyze than long echo time ^{1}H MRS signals due to a higher number of overlapping peaks, a stronger baseline and a higher sensitivity to artefacts. In comparison, long echo time ^{1}H MRS signals are poorer in information but they allow a more reliable analysis and testing of classification methods.

Many studies have been performed to classify MRS signals. Lindon et al. (2001) overviewed pattern recognition methods and their applications in biomedical magnetic resonance. But most of the studies investigated only one classification method and restrict data collection to one center only. As performance measure either the training performance or test performance on a specifically selected set was considered. In our study we have measured the binary classification performance based on the Receiver Operating Characteristic (ROC) curve analysis (section 3.2.4) over 200 stratified random samplings of training and test set. The area under the ROC curve (AUC) gives a global measure of the clinical efficiency over a range of test cut-off points on the ROC curve. This is in contrast to performance measures like the correct classification rate estimated on a test set (e.g. Ye et al., 2002) which is only based on a single cut-off point (e.g. for one specific value of the false-positive rate). Various clinical studies focus on the prediction of the malignancy of tumours, more specifically for brain gliomas (e.g. Herminghaus et al. (2003); Ye et al. (2002)). Thereby, they consider only two classes: low-grade and high-grade gliomas. In our study, astrocytomas of grade II and glioblastomas (also called astrocytomas of grade IV) are included, which are large subtypes of respectively low-grade and high-grade gliomas. Additionally, we consider two other common brain tumour types, namely metastases and meningiomas.

Moreover, this chapter reports on the results of a comparative study on a multi-centre data set of MRS signals. This data set was developed in the framework of the EU funded INTERPRET project (INTERPRET-project, 1999-2002). Several INTERPRET partners (Simonetti et al., 2003; Tate et al., 1998, 2003; Ladroue et al., 2002; Lukas et al., 2002a,b; Szabo De Edelenyi et al., 2002; Tate et al., 2002) have already published results for classification of brain tumours based on MR data available within the project. The papers (Simonetti et al., 2003; Szabo De Edelenyi et al., 2002) focus on the use of ^{1}H MRSI data, while others consider the use of short or long echo time ^{1}H MRS. Nevertheless, most of these studies are based on a previous version of the data set or focus on

a specific technique. For example in Tate et al. (2003), 144 short echo time ^{1}H MRS spectra from 3 contributing centers were used, originating from three groups of brain tumours; meningiomas, low-grade astrocytomas and aggressive tumours. The latter group includes glioblastomas and metastases. Note that these groups correspond to the same four tumour groups as considered in this chapter. But Tate et al. (2003) selected a specific training and test set; the data from two centers formed the training set (94 spectra) and the data from the third center were used for testing (50 spectra). Based on this specific test set an accuracy of 96% was obtained using LDA.

In this study we applied several methods on all histopathologically validated long echo time ^{1}H MRS data from four common brain tumour types available in the INTERPRET database. We mention three additional differences with previous classification studies within the framework of the INTERPRET project. First of all, we investigated what can be obtained as typical performance on a representative test set. Therefore, we constructed 200 different combinations of training and independent test set. Secondly, the discrimination ability was judged by the AUC, which is, in contradiction to the correct classification rate, a global measure. Thirdly, four different techniques were applied for classification; linear as well as nonlinear techniques. We investigated binary as well as multiclass classification. Moreover, this analysis included optimal input variable selection and (hyper-) parameter estimation.

4.2 Material

Long echo time ^{1}H MRS data were provided by CDP (Centre Diagnòstic Pedralbes, Barcelona, Spain), IDI (Institut de Diagnòstic per la Imatge, Barcelona, Spain), SGHMS (St. George's Hospital Medical School, London, UK) and UMCN (University Medical Center Nijmegen, Nijmegen, The Netherlands). Table 4.1 reports the most important acquisition parameters. The partners of the EU-funded INTERPRET project (IST-1999-10310) reached a consensus (van der Graaf, 2001) about the acquisition protocol for MRS and MRSI measurements. Quality control criteria have been applied to MR system performance and data from all centres (van der Graaf et al., 2002). Data were acquired at 1.5T using a PRESS sequence, at an echo time (TE) of 135 or 136 ms, repetition time (TR) between 1598 and 2020 ms, spectral width (SW) of 1000 Hz or 2500 Hz, using 512 or 2048 datapoints.

Four common brain tumour types are considered: glioblastomas (glio), meningiomas (meni), metastases (meta) and astrocytomas grade II (astroII), which for the purpose of grouping are labelled as class 1, 2, 3 and 4, respectively (Table 4.2). Although the INTERPRET database does contain data from other types of brain tumours there are much fewer spectra of these. In order to allow

centre	manufacturer	sequence	TE(ms)	TR(ms)	SW(Hz)	N
CDP	GE	PRESS	135	1598,1620	2500	2048
SGHMS	GE	PRESS	136	1998,2000,2020	2500	2048
IDI	Philips	PRESS	136	2000	1000	512
UMCN	Siemens	PRESS	136	2000	1000	512

Table 4.1: Overview of the acquisition parameters of the long echo time ^{1}H MRS data from all centres. The columns respectively correspond to the name of the centre, the manufacturer of the scanner, the sequence used, the echo time, the repetition time, the spectral width and the number of points in the original FID.

Center (acquisition scheme)	1	2	3	4	total
CDP (PRESS, TE=135ms)	38	16	5	6	65
IDI (PRESS, TE=136ms)	28	27	16	6	77
SGHMS (PRESS, TE=136ms)	10	9	11	7	37
UMCN (PRESS, TE=136ms)	1	1	0	2	4
total	77	53	32	21	183

Table 4.2: Number of long echo time ^{1}H MRS data of glioblastomas (class 1), meningiomas (2), metastases (3) and astrocytomas grade II (4). The rows correspond to the acquisition center, while the columns mention the type of brain tumour. The acquisition scheme is a PRESS sequence and TE denotes the echo time.

statistically valid comparisons and to prevent unwanted bias due to large differences in the numbers of spectra within each group, we focused the analysis on the brain tumour types with more than 20 spectra available. Moreover, the selected tumour types are the major tumour groups occurring in the brain and include both benign/low grade (meningiomas and astrocytomas grade II) and malignant (glioblastomas and metastases) tumour types, and tumours of the same original cell type (glioblastomas vs. astrocytomas grade II). All data have passed a quality control and validation process, which was regulated by strict rules agreed on by all INTERPRET partners (van der Graaf, 2000). This includes spectroscopic as well as clinical aspects; the spectral quality control was carried out by expert spectroscopists, while for the clinical validation a Clinical Data Validation Committee (CDVC) was formed. Concerning the spectral quality control, spectra with either large water linewidths (> 8 Hz), a low signal-to-noise ratio (SNR < 10) or obvious artefacts were discarded. The spectral water linewidth was obtained by fitting a Lorentzian lineshape to the real part of the water unsuppressed spectrum after eddy current correction of the water unsuppressed FID using the phase information of the FID itself. The

SNR was defined as the ratio of the highest signal intensity in the range 0-3.4 ppm and the standard deviation of the noise in the range 9-11 ppm of the water suppressed spectrum.

Spectra approved by one expert spectroscopist, were checked by another expert spectroscopist. If both expert spectroscopists agreed, the spectrum was approved. In case of disagreement, a third expert spectroscopist made the final decision. The clinical validation was done by the clinical committee CDVC. Each case was examined and based upon the quality of the available clinical information, biopsy and MRI scans, the case was validated or not. Cases with inadequate data (for example the spectral localization was not within the site of the brain tumour) or cases where there was no histological agreement about the diagnosis were discarded. The histopathological classification of the brain tumours was performed by two pathologists (three in case of disagreement between the first two, similar as for the spectral quality control). These class assignments were based on the histological classification of Tumours of the Central Nervous System (CNS) set up by the World Health Organization (WHO) (Kleihues et al., 2002).

The raw data were acquired in the time domain at the aforementioned centers. A few preprocessing steps were carried out: frequency alignment and phase correction with Klose's method (Klose, 1990) and filtering of the dominating residual water peak using HSVD (Barkhuysen et al., 1987), with 15 singular values over the water region of 4.09-5.31 ppm. The initial point of the time domain signal was removed, because it was often affected by artefacts. The FIDs acquired with a GE-scanner had a different spectral width and number of points than the Siemens- and Philips-data (Table 4.1). In order to match these parameters, the GE-data were interpolated using piecewise cubic splines in the time domain. Hence, after preprocessing, all FIDs had 511 points (512 minus the deleted first point) which provided spectra with a final spectral width of 1000 Hz. The interpolation of the GE-data actually resulted in a four times lower input dimension than the original GE-data, which is more appropriate for classification. The resulting signal was then transformed to the frequency domain by FFT (Fast Fourier Transform). No additional referencing of the spectra was made, apart from the frequency alignment with Klose's method, leaving a slight misalignment (< 0.08 ppm). For each signal the L_2-normalized magnitude spectrum (of unit length) is considered only in the frequency region of interest (4.17 to 0.94 ppm), corresponding to 108 input variables. Figures 4.1 and 4.2 depict the mean magnitude frequency spectra of the four considered classes.

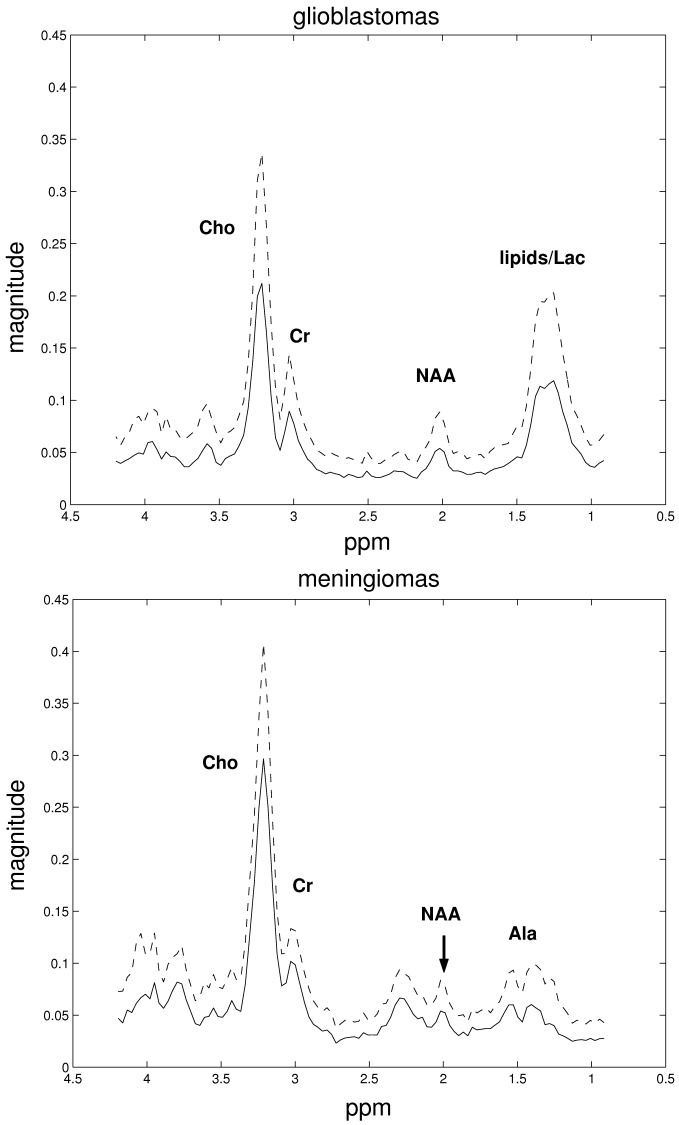

Figure 4.1: Mean L_2-normalized magnitude frequency spectra of the four considered classes: class 1 (top), class 2 (bottom) correspond to the glioblastomas, meningiomas, respectively. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

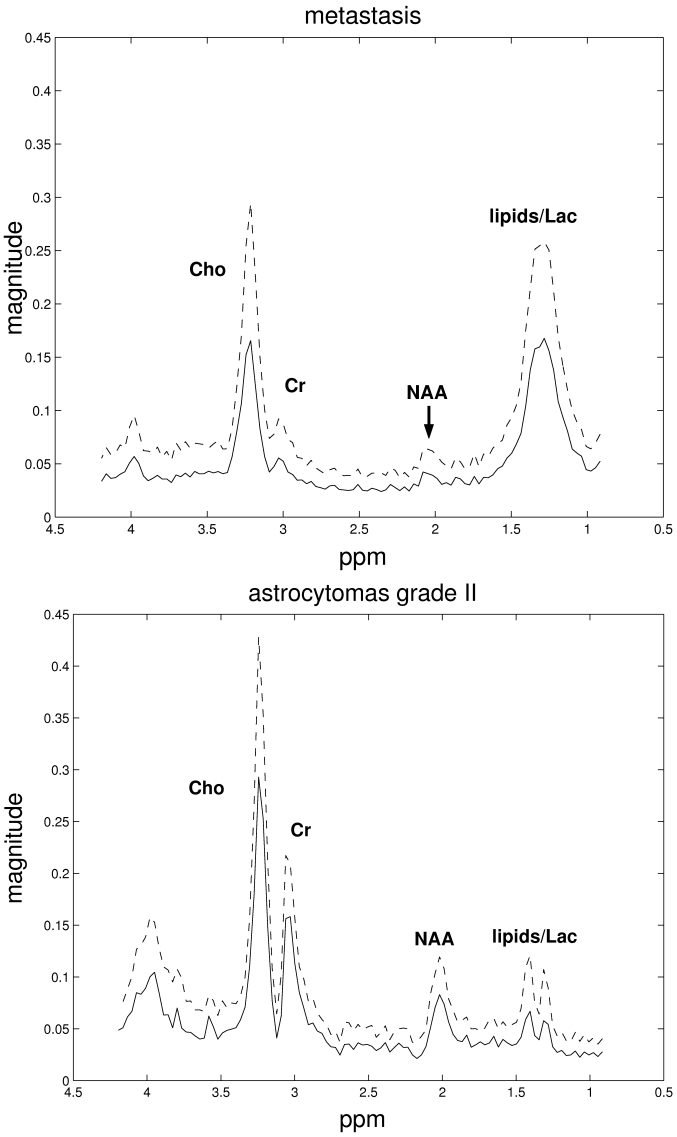

Figure 4.2: Mean L_2-normalized magnitude frequency spectra of the four considered classes: class 3 (top) and class 4 (bottom) correspond to metastases and astrocytomas (grade II) respectively. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

4.3 Methods

4.3.1 Input features

As explained in section 3.2.2, several resonance peaks of important brain metabolites can be exploited as discriminatory features to distinguish tumours of different types and grades. Instead of using complete spectra as input variables to the classifier, selection of the most explanatory input features can be used. In this study, three approaches were compared in order to investigate the influence of the input features:

- Complete spectra. This approach uses the L_2-normalized magnitude spectra (spectral region 4.17 to 0.94 ppm) as preprocessed in section 4.2.

- Selected frequency regions. The spectral variables are selected within certain regions or intervals of the magnitude spectrum which are assumed to contain most of the information. The characteristic resonances can be observed in the following regions: *Cho* and *Cr* (2.95 - 3.3 ppm); *NAc* (1.95 - 2.1 ppm); *Lac*, *Ala* and *L1* (1.15 - 1.55 ppm); *L2* (0.9 - 1.0 ppm).

- Peak integration. The amplitude of a resonance is proportional to the integral of the corresponding peak in the spectrum. However, precise estimation of the peak integrals is difficult due to several factors, including nonzero baseline, peak overlap, noise and also the discrete nature of the spectrum. Peak integration is performed here by using the trapezoidal rule. For each selected resonance the area under the frequency peak in the magnitude spectrum is calculated. These regions cover: *Cho* (3.1 - 3.3 ppm); *Cr* (2.95 - 3.05 ppm); *NAc* (1.95 - 2.1 ppm); *Lac* and *Ala* (1.25 - 1.55 ppm); and *L1* (1.1 - 1.25 ppm).

Note that the selected regions and resonances used for peak integration are based on the metabolites that are assumed to be most characteristic according to prior knowledge available from field experts participating in this study. Nevertheless this selection is still subjective as the size of the regions could be altered or some other resonances (e.g. from metabolites that typically are less visible at long echo times; *mI*, *Gln*, *Gly*,...) could also have been included.

4.3.2 Classification techniques

Several classification techniques were compared to separate the given MR spectra. The applied techniques were chosen so that linear as well as nonlinear methods were considered: LDA, SVM and LS-SVM, which were explained in section 3.3.2.

LDA has the strength of being simple and computationally fast, and works well for data that are linearly separable. The dimension of the input features to LDA is reduced by PCA. The number of principal components is determined by the number of components that account for 75% of the total variance of the given data. Note that PCA is not used when peak integration values are taken as input features, as peak integration already significantly reduces the dimension.

Quite often, different classes do not have equally distributed datapoints and their distributions are also overlapping among classes, which causes the problem to be linearly nonseparable. Here, two kernel-based classifiers SVM and LS-SVM are assessed. SVM and LS-SVM with a linear kernel can be regarded as regularized linear classifiers, while LS-SVM with an RBF kernel is regarded as a regularized nonlinear classifier. One of the interesting advantages of (LS-)SVM is its ability to learn and generalize processing of high-dimensional data but without dimensionality reduction, which is important for the present application.

To achieve a high level of performance in SVMs, some hyperparameters must be tuned. These adjustable hyperparameters include: a regularization parameter (C for SVM, γ for LS-SVM), which determines the tradeoff between minimizing the training errors and minimizing the model complexity. In case of an RBF kernel, also a kernel parameter (the width σ) must be selected. We have selected the value of hyperparameters C for SVM, γ for LS-SVM with a linear kernel and (σ, γ) for LS-SVM with an RBF kernel through leave-one-out (LOO) cross-validation for a grid of hyperparameter values, while bounding the search to avoid overfitting.

The MRS spectra were classified using Steve Gunn's SVM toolbox (Gunn, 1997; MATLAB Support Vector Machines toolbox, Version 2.0-1998) and KULeuven's LS-SVMlab toolbox (Suykens et al., 2002; MATLAB/C LS-SVMlab toolbox, Version 1.4-2002; Pelckmans et al., 2002) for LS-SVM classification with both linear and RBF kernels.

4.3.3 Experimental setting

Binary Classification

Binary classification can be used to distinguish two different tumour types. Instead of using a one-against-all scheme, the classes are pairwise compared by means of a binary classifier. Consider four types of brain tumours, then six binary classifiers could be constructed to separate the following pairs:

- glioblastomas vs meningiomas (glio-meni),
- glioblastomas vs metastases (glio-meta),

- glioblastomas vs astrocytomas grade II (glio-astroII),
- meningiomas vs metastases (meni-meta),
- meningiomas vs astrocytomas grade II (meni-astroII), and
- metastases vs astrocytomas grade II (meta-astroII).

By classifying in pairs, more information was obtained about:

1. the distribution of two classes and their overlap,
2. the balance of the data distribution of the classes, and
3. the performance of the classifier which could be measured using ROC analysis.

The experiment consisted of the following steps:

1. the data were divided in a training set (2/3 of the data) and a test set (remainder) using stratified random sampling,
2. train the classifiers and use the test set to evaluate the performance,
3. the index of the misclassified spectra was noted.

This randomization was repeated 200 times to avoid bias possibly introduced by selection of a specific training and test set. This repeated sampling method has the preference over a single randomization and is very often applied for classification. In this way we tried to obtain a representative performance on the test set. Rather than using the correct classification rate as performance measure, like in repeated subsampling, ROC (Swets, 1979) analysis was used to evaluate the binary classifiers. The performance was then measured by the mean AUC and its pooled standard error calculated from 200 randomizations.

Multiclass Classification

In the framework of binary classification it was assumed that a new MRS spectrum belongs to one of the two considered classes. Nevertheless, in medical practice, the number of possible tumour types is mostly not restricted to two types. This motivates the development of multiclass classifiers, that handle all classes in one construction, which extends the classifiers mentioned in the previous section. With this setup, the classifier is expected to classify a certain spectrum as one of the four tumour types.

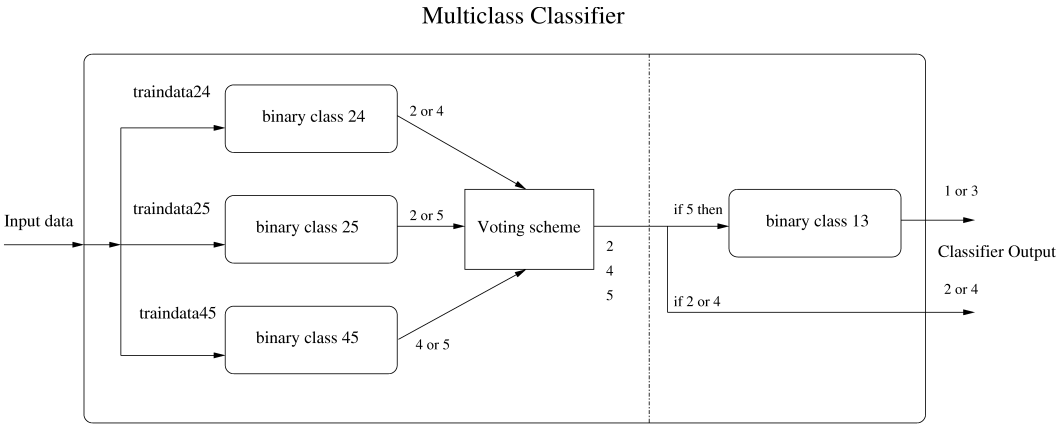

Figure 4.3: Two-steps classification. The left part shows step one, classification of three tumour classes: meningiomas (class 2), astrocytomas of grade II (4), and aggressive tumours (5). The right part, or step two, further refines the classification if the output is class 5 and assigns the spectra of this class either to glioblastomas (class 1) or metastases (3).

Various pattern recognition techniques have been tried to distinguish MRS spectra of class 1 (glio) and class 3 (meta), but none gives satisfactory results (Tate et al., 1998; Poptani et al., 1999). Alternatively, as was suggested by Tate et al. (2003), we can merge these two classes, obtaining a new group called class 5, containing only aggressive (aggr) tumours. This scheme is depicted as step one shown on the left part of the Figure 4.3. A voting scheme is applied to decide which class is chosen based on the three outputs of the contributing binary classes. With a minimum two-out-of-three vote, a certain class is taken if two of the binary classifiers give the same output, otherwise the classifier considers the output as undecided. Step two is carried out, as illustrated in the right part: if the output of step one is class 5, then further classify the spectrum either into class 1 (glio) or 3 (meta) using the binary classifier glio-meta. If the output is class 2 or 4, then the output of step two is the same as the output of step one. Four binary classifiers are the building blocks of this multiclass classifier: binary classifiers meni-astroII and glio-meta were explained in the previous section, additionally two binary classifiers are required:

- meningiomas vs. aggressive tumours (meni-aggr),
- astrocytomas grade II vs. aggressive tumours (astroII-aggr).

Statistical Analysis

From 200 runs, the mean AUCs ($\overline{\text{AUC}}$) are listed in the tables, as well as the standard error SE on the AUC. For each binary classifier C, the mean and standard error of the AUC is calculated. In section 3.2.4 a framework was given to statistically test whether the areas under two different ROC curves derived from the same samples differ significantly from each other and was generalized for the mean AUCs and their SE resulting from M stratified randomizations. In this chapter, this framework ($M = 200$) was applied to test whether two binary classifiers have a significantly different performance. Although ROC analysis has been extended to multiclass classification (Srinivasan, 1999), the result is generally non-intuitive and computationally expensive. This motivates the use of the correct classification rate as performance measure for multiclass classification.

4.4 Results

4.4.1 Classification using complete spectra

In the following paragraphs the classification performance of LDA, SVM, and LS-SVM (using linear and RBF kernels) are reported. The results using the complete spectra are summarized in Table 4.3, while Figures 4.4, 4.5 and 4.6 show the boxplots corresponding to the same cases. Note that the boxplots display the median of the AUC values and the Interquartile Range (IQR), while the tables display the mean and the standard error of the AUC values. The latter can be used to calculate the z-value (section 3.2.4).

4.4.2 Classification using selected frequency regions

By selecting the values within specific frequency ranges in the spectra, the number of datapoints is reduced from 108 into 30. For the LDA classifier, PCA is applied to further reduce this input dimension, covering at least 75% of the variance. These input variables are different from those obtained for the complete spectrum, due to the higher degree of freedom in the latter case. The results of 200 runs of stratified random samplings of the L_2-normalized magnitude MRS spectra are shown in Table 4.4 and Figures 4.7, 4.8 and 4.9.

4.4.3 Classification using peak integration

Table 4.5 and Figures 4.10, 4.11 and 4.12 show the results of the ROC analysis for classification using peak integration. Five peak integration values are used as input of the classifiers. The linear classifier LDA is used without applying PCA.

4.4.4 Multiclass approach

As mentioned above, two additional binary classifiers are constructed by merging glioblastomas and metastases into one class of aggressive tumours. Table 4.6 shows the performance of these classifiers using the complete spectra as input.

Training performance

One way to train the multiclass classifier is by feeding all the spectra to the classifier and train each binary classifier with the corresponding classes. For example, use the spectra of class 2 and class 4 to train the binary classifier 24, and similarly for the others.

Table 4.7 shows a comparison of the multiclass classifier performance. The first row shows the percentage of correctly classified spectra in the class of *meningiomas*, *astrocytomas grade II* and *aggressive tumours*. One undecided case arose when using PCA/LDA with complete spectra classification, fifteen when using PCA/LDA and one when using LS-SVM classification both with the selected frequency regions as input variables.

In the second step, we have used classifier 13 to further subclassify the aggressive class. Using this subclassification, the multiclass classifier's performance is shown in Table 4.8.

Test performance

Besides using all spectra to choose the hyperparameters and to train the classifiers, one can also select 2/3 of the data set as training set and use the remainder as test set. This stratified random sampling was repeated for 200 runs. The results are shown in Table 4.9 for one-step classification, which assigned the spectra to one of the 3 following classes: 2, 4 or 5. Table 4.10 shows the classifier performance after two-steps classification, which assigned the spectra to one of the 4 following classes: 1, 2, 3 or 4. Each spectrum of class 5 in step one, was either assigned to class 1 or class 3 in step two. Tables 4.9 and 4.10 mention the mean correct classification rate, the mean misclassification rate and the mean percentage of undecided cases and their standard deviation. The correct classification rate is defined as the percentage of correctly classified spectra, while the misclassification rate is the percentage of misclassified cases.

Classes	PCA/LDA	SVM lin	LS-SVM lin	LS-SVM RBF
glio-meni	$0.9528 \pm 0.0306_{(8)}$	0.9519 ± 0.0335	0.9506 ± 0.0338	0.9560 ± 0.0304
glio-meta	$0.5926 \pm 0.1036_{(6)}$	0.6323 ± 0.0942	0.6431 ± 0.0983	0.5851 ± 0.1037
glio-astroII	$0.9180 \pm 0.0627_{(7)}$	0.9159 ± 0.0565	0.9351 ± 0.0524	0.9385 ± 0.0486
meni-meta	$0.9605 \pm 0.0375_{(5)}$	0.9642 ± 0.0337	0.9711 ± 0.0307	0.9701 ± 0.0306
meni-astroII	$0.9313 \pm 0.0725_{(10)}$	0.9661 ± 0.0390	0.9581 ± 0.0482	0.9595 ± 0.0456
meta-astroII	$0.9612 \pm 0.0533_{(4)}$	0.9695 ± 0.0418	0.9740 ± 0.0393	0.9721 ± 0.0377

Table 4.3: Classification using **complete spectra**. Average performance on the test set from 200 runs of stratified random samplings of the L_2-normalized magnitude MRS spectra. As performance measure the mean AUC and its pooled standard error SE was used. The number between the brackets mentions the number of principal components used.

Classes	PCA/LDA	SVM lin	LS-SVM lin	LS-SVM RBF
glio-meni	$0.7643 \pm 0.0722_{(2)}$	0.8532 ± 0.0575	0.8922 ± 0.0494	0.9187 ± 0.0413
glio-meta	$0.6381 \pm 0.1004_{(2)}$	0.5081 ± 0.1044	0.6368 ± 0.0998	0.5576 ± 0.1030
glio-astroII	$0.8319 \pm 0.0776_{(2)}$	0.8692 ± 0.0713	0.8849 ± 0.0660	0.9012 ± 0.0594
meni-meta	$0.9212 \pm 0.0525_{(2)}$	0.9098 ± 0.0594	0.9339 ± 0.0475	0.9534 ± 0.0374
meni-astroII	$0.9079 \pm 0.0645_{(3)}$	0.9592 ± 0.0410	0.9619 ± 0.0422	0.9617 ± 0.0411
meta-astroII	$0.9173 \pm 0.0689_{(2)}$	0.9459 ± 0.0549	0.9698 ± 0.0389	0.9642 ± 0.0429

Table 4.4: Classification using **selected frequency regions**. Average performance on the test set from 200 runs of stratified random samplings of the L_2-normalized magnitude MRS spectra. As performance measure the mean AUC and its pooled standard error SE was used. The number between the brackets mentions the number of principal components used.

Classes	LDA	SVM lin	LS-SVM lin	LS-SVM RBF
glio-meni	0.8504 ± 0.0586	0.8561 ± 0.0577	0.8448 ± 0.0593	0.8677 ± 0.0550
glio-meta	0.6252 ± 0.1007	0.6236 ± 0.1005	0.6434 ± 0.1006	0.6264 ± 0.0988
glio-astroII	0.8773 ± 0.0635	0.8916 ± 0.0571	0.8787 ± 0.0628	0.8818 ± 0.0631
meni-meta	0.9103 ± 0.0628	0.9113 ± 0.0618	0.9191 ± 0.0585	0.9357 ± 0.0473
meni-astroII	0.8441 ± 0.0858	0.8297 ± 0.0926	0.8485 ± 0.0851	0.8281 ± 0.0921
meta-astroII	0.9592 ± 0.0461	0.9727 ± 0.0376	0.9597 ± 0.0453	0.9521 ± 0.0528

Table 4.5: Classification using **peak integration values**. Average performance on the test set from 200 runs of stratified random samplings of the L_2-normalized magnitude MRS spectra. As performance measure the mean AUC and its pooled standard error SE was used.

Classes	LDA	SVM lin	LS-SVM lin	LS-SVM RBF
meni-aggr	$0.9433 \pm 0.0306_{(6)}$	0.9409 ± 0.0343	0.9620 ± 0.0279	0.9110 ± 0.1121
astroII-aggr	$0.9230 \pm 0.0674_{(6)}$	0.9343 ± 0.0458	0.9416 ± 0.0502	0.9129 ± 0.1137

Table 4.6: Classification using **complete spectra**. Average performance on the test set from 200 runs of stratified random samplings of the L_2-normalized magnitude MRS spectra. As performance measure the mean AUC and its pooled standard error SE was used. The number between the brackets mentions the number of principal components used.

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
compl. spec.	84.6995	93.9891	97.8142
disc. feat.	65.0273	84.6995	90.1639
peak integ.	75.9563	77.0492	80.8743

Table 4.7: One-step classification using **complete spectra**, selected frequency regions and peak integration values on respectively the first, second and third rows. Percentage of correctly classified spectra using all L_2-normalized magnitude MRS spectra to assess the training performance.

4.5 Discussion

In this section we discuss various issues concerning the results we obtained using the available long echo time ^{1}H MRS data.

4.5.1 Limitations

MRS signals of brain tumours contain chemical information about metabolites characteristic for the type of tumour. Nevertheless, there are still some factors making it hard to construct a classifier which is able to discriminate between

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
compl. spec.	71.0383	78.1421	83.6066
disc. feat.	50.2732	68.8525	74.8634
peak integ.	61.7486	62.2951	67.7596

Table 4.8: Two-steps classification using **complete spectra**, selected frequency regions and peak integration values on respectively the first, second and third rows. Percentage of correctly classified spectra using all L_2-normalized magnitude MRS spectra to assess the training performance.

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
correct	80.1855 ± 4.2853	82.7823 ± 3.3449	83.5726 ± 3.5058
misclass	14.0887 ± 4.0665	13.6532 ± 3.1140	12.5565 ± 3.3290
undecided	5.7258 ± 2.6110	3.5645 ± 2.0870	3.8710 ± 2.1144

Table 4.9: One-step classification using **complete spectra**. Average performance on the test set from 200 runs of stratified random samplings (2/3 of the data used for training, 1/3 for testing). The first, second and third rows give, respectively, the mean correct classification rate, the mean misclassification rate and the mean percentage of undecided cases, each with their standard deviation.

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
correct	63.1532 ± 4.7255	65.7984 ± 3.3449	66.8145 ± 3.5058
misclass	31.1210 ± 4.6858	30.6371 ± 3.1706	29.3145 ± 3.5954
undecided	5.7258 ± 2.6110	3.5645 ± 2.0870	3.8710 ± 2.1144

Table 4.10: Two-steps classification using **complete spectra**. Average performance on the test set from 200 runs of stratified random samplings (2/3 of the data used for training, 1/3 for testing). The first, second and third rows give, respectively, the mean correct classification rate, the mean misclassification rate and the mean percentage of undecided cases, each with their standard deviation. Note that the number of undecided cases is equal to that for the one-step classifier.

different brain tumours using MRS signals:

1. the limited number of available spectra per type of tumour (see Table 5.2). Especially the amount of available metastases and astrocytomas grade II is low. This makes it difficult to construct a classifier with a high generalization capacity.

2. the presence of noise and artefacts in the spectra. Even after elimination of the dominating water peak, remaining artefacts might affect important peaks in the spectra.

3. the large variances within each class and the overlap between spectra of different brain tumour types (Figures 4.1 and 4.2). For example the mean spectra of glio and meta show very similar characteristic patterns, which makes the discrimination between glio and meta a very hard problem. This problem is also observed in Tate et al. (1998) and Poptani et al. (1999). Further discussion about this is addressed below (section 4.5.2).

4.5.2 Glioblastomas versus metastases

Although we obtained a low performance for distinguishing glioblastomas (glio) and metastases (meta), there are indications that these tumour types might be separable based on MR. Szabo De Edelenyi et al. (2000) introduced the so-called nosologic images, which is an approach to analyze ^{1}H MRSI data of brain tumours. It is a tool that assigns the spectroscopic data of each voxel in the spectroscopic image to a histopathological class. Classification was carried out by LDA applied on 6 metabolite values obtained from long echo time ^{1}H MRSI spectra (TE=272ms), together with the unsuppressed water area. Their study included 77 images, of which 24 high-grade gliomas and 10 metastases, for which they obtained a training performance of 87% following a leave-one-out (LOO) procedure. For the high-grade gliomas and metastases, respectively 19 and 6 spectra were correctly assigned.

Researchers (Ishimaru et al., 2001; Law et al., 2002; Opstad et al., 2004) have found a few metabolite peaks or ratios which might contribute to the discrimination of high-grade gliomas and metastases. Law et al. (2002) concluded out of a study based on MRSI that, despite the small size of their data set (11 high-grade gliomas, 6 metastases), the *Cho/Cr* ratio was significantly higher in high-grade gliomas than in metastases; this was the case for the tumoural region as well as the peritumoural region. Also based on perfusion-weighted MRI they have found different characteristics. Opstad et al. (2004) have considered short echo time ^{1}H MRS spectra (TE=30ms) from 25 glioblastomas and 34 metastases. Based on these data, they were able to find a significant difference in the ratio of the 1.3 ppm and the 0.9 ppm lipid/macromolecule peaks between these two tumour groups. This lipid peak area (LPA) ratio was 2.6 ± 0.6 for glioblastomas and 3.8 ± 1.4 for metastases ($P \leq 0.0001$). Based on ^{1}H MRS, Ishimaru et al. (2001) have shown that the absence of *Cr* might indicate a diagnosis of metastasis, while in short echo time the absence of lipids may exclude metastasis. In Figure 4.2 we do indeed notice a large mean lipid peak in metastases, but also in glioblastomas (Figure 4.1). This latter might be due to the occurrence of necrotic tissue in part of the glioblastomas. This partially explains the large variation especially observed within this class and the similarity with the class of metastases.

4.5.3 Classification techniques

In general, LDA as linear classifier, preceded by PCA (except for peak integration) performs quite well in solving the brain tumour classification problem. This is in correspondence with Tate et al. (2003). However, due to its linear boundary, overlapping classes are very difficult to handle.

As stated above, the small data set available also forms a limitation for training. Therefore, the discrimination boundary will strongly correlate with the training

set. Especially LDA requires a significant amount of datapoints to be able to draw a linear separating line between overlapping classes. In addition, it is possible that the separating line is very dependent on the selected training set.

Kernel-based classifiers are less sensitive to the amount of datapoints; although the dimension is larger than the number of datapoints, these classifiers could draw an optimal separating boundary, without applying any dimensionality reduction (such as PCA). Kernel-based classifiers, SVM and LS-SVM, feature the advantage of detecting automatically important characteristics independently of the input pattern.

Based on the statistical analysis, described in section 3.2.4, no statistically significant difference was found between the AUC values for any of the classification techniques applied to the available long echo time ^{1}H MRS data. The highest z-value (1.72) was obtained, when comparing PCA/LDA with LS-SVM with a linear kernel based on the frequency selected regions; this is still lower than the cutoff value (1.96).

The best performing technique depends on the considered classes and the type of input. To be more specific, when comparing the classification techniques we can group them in two ways (e.g. consider only the results with the complete spectra as input):

- linear (LDA, SVM lin, LS-SVM lin) vs nonlinear techniques (LS-SVM RBF): in the cases glio-meni, glio-astroII the mean AUC values were slightly higher for the nonlinear technique, while in the other cases the AUC values were in the same range or slightly lower.
- LDA vs kernel-based techniques (SVM lin, LS-SVM lin, LS-SVM RBF): in the cases meni-meta, meni-astroII and meta-astroII the kernel-based techniques performed slightly better.

Additionally, from the comparison of SVM vs LS-SVM, we mainly observed that the best performing technique is also quite dependent on the case.

4.5.4 Influence of dimensionality reduction

For classification using selected frequency regions (section 4.4.2) and peak integration values (section 4.4.3) the input dimension was reduced by selecting only spectral regions which contain resonances of important metabolites. The underlying idea for this dimensionality reduction was to remove any redundant input features and reduce the influence of noise and artefacts. Hence, this tried to enhance the discriminatory chemical information present in the spectra.

In contradiction to our expectations, we have observed that the results in sections 4.4.2 and 4.4.3 on average were worse than in section 4.4.1. This seems to

imply that this approach to dimensionality reduction also reduces part of the valuable information, that is present in the excluded frequency regions, which was important to explain the variance between the brain tumour classes. More specifically, in this study we did not consider resonances which typically have a low apparent intensity at long echo times (because of a small T2-value or cancellation due to J-modulation) (Ernst and Hennig, 1991; Govindaraju et al., 2000): *mI* (e.g. with triplets and multiplet at 3.26 and 3.57 ppm), *Glu* (e.g. multiplets at 2.33 ppm and 3.74 ppm), *Gln* (e.g. multiplets at 2.43 ppm and 3.75 ppm), *Gly* (singlet at 3.55 ppm). In Figures 4.1 and 4.2 we indeed noticed a few small peaks around the specified resonances.

4.5.5 Multiclass classification

Multiclass classifiers handle all classes in one construction. We reduced this problem to a set of four binary classification problems, as explained in section 4.3.3. Hence, we obtained four separating functions instead of one (one for each binary problem). Multiclass classifiers with the proposed scheme showed a high learning capability. This was illustrated by the correct classification rates for the first step using the complete spectra as input: 84.7% (PCA/LDA), 93.9% (LS-SVM lin) and 97.8% (LS-SVM RBF). Given an independent test set as input, the classifiers on average gave a quite good generalization performance: 80.2% (PCA/LDA), 82.8% (LS-SVM lin) and 83.6% (LS-SVM RBF).

In the second step of the multiclass classifier the output of the first step was combined with the binary classifier that separated glioblastomas and metastases. The test performances reduced to 63.1% correct classification (LDA/PCA), 65.8% (LS-SVM lin), 66.8% (LS-SVM RBF). This could be explained by the hard binary problem glioblastomas versus metastases. As observed in the binary classification, these two classes were very similar. Therefore separating them from one single class 5 into class 1 and class 3 deteriorated the total performance of the classifier.

Although no ROC analysis for multiclass classification was performed in order to test for significant differences between the classification techniques, the following indications could be noted, without drawing a general conclusion. The results, after step one as well as step two, yielded a clearly higher training performance for the kernel-based methods than for LDA. Moreover, the kernel-based methods on average performed slightly better on an independent test set than LDA. This was clear from the mean percentage of correctly classified spectra which was a few percentages higher for LS-SVM (see Tables 4.9, 4.10 and Figure 4.13). Also the mean percentage of undecided cases differed slightly in favor of LS-SVM. This indicates that kernel-based methods can generalize at least as well as LDA based on a small data set.

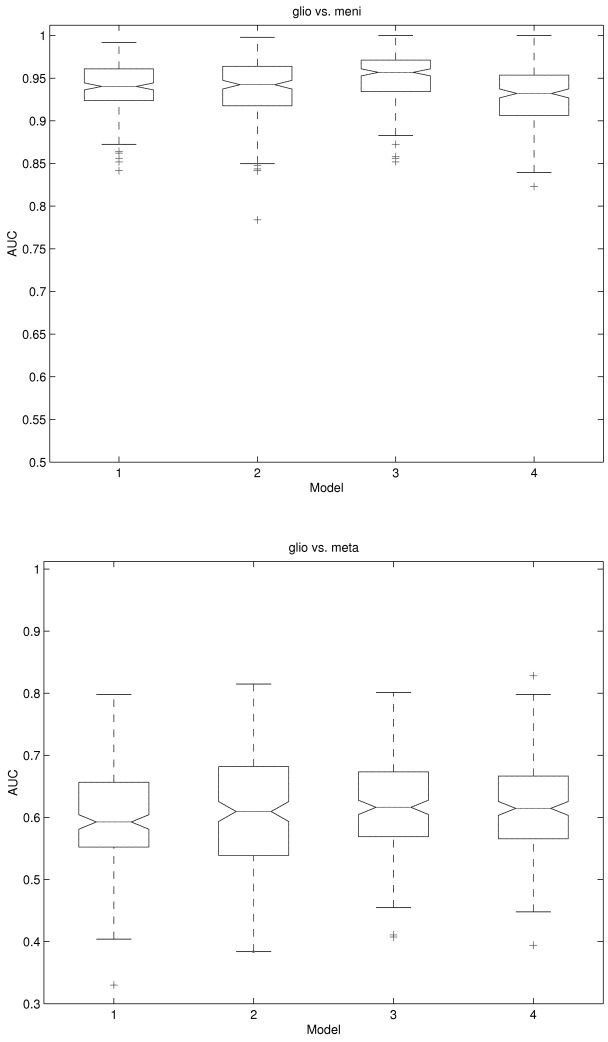

Figure 4.4: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) PCA/LDA, (2) SVM with a linear kernel, (3) LS-SVM with a linear kernel, (4) LS-SVM with an RBF kernel. The figures correspond to the binary classifiers using complete spectra: glio vs. meni (top), glio vs. meta (bottom). Each box has lines at the lower quartile (lowest line), median (line in the middle), and upper quartile (highest line) values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. Outliers are data with values beyond the ends of the whiskers.

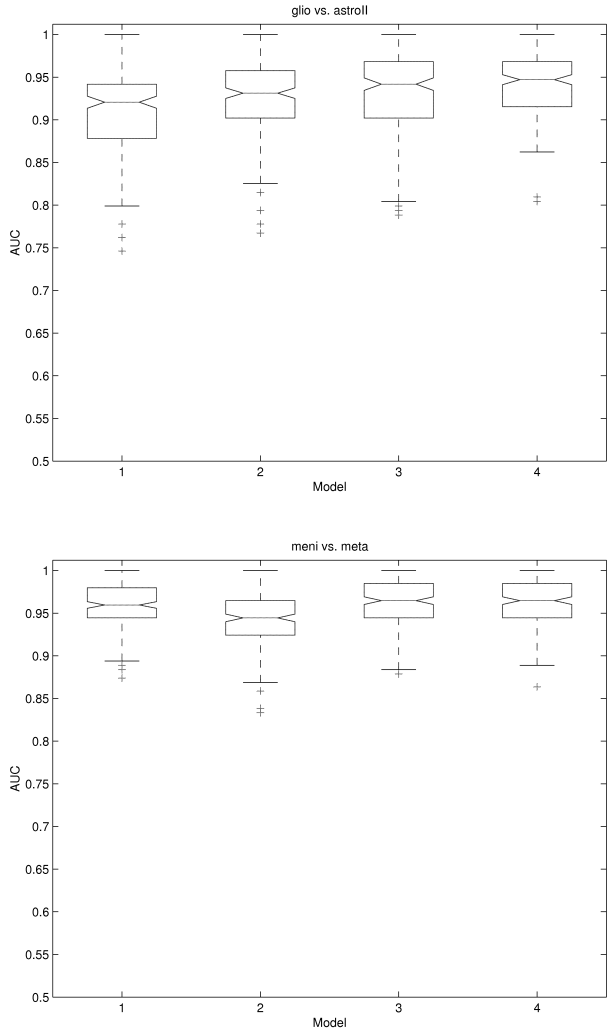

Figure 4.5: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) PCA/LDA, (2) SVM with a linear kernel, (3) LS-SVM with a linear kernel, (4) LS-SVM with an RBF kernel. The figures correspond to the binary classifiers using complete spectra: glio vs. astroII (top). meni vs. meta (bottom). For further explanation about boxplot, see Figure 4.4.

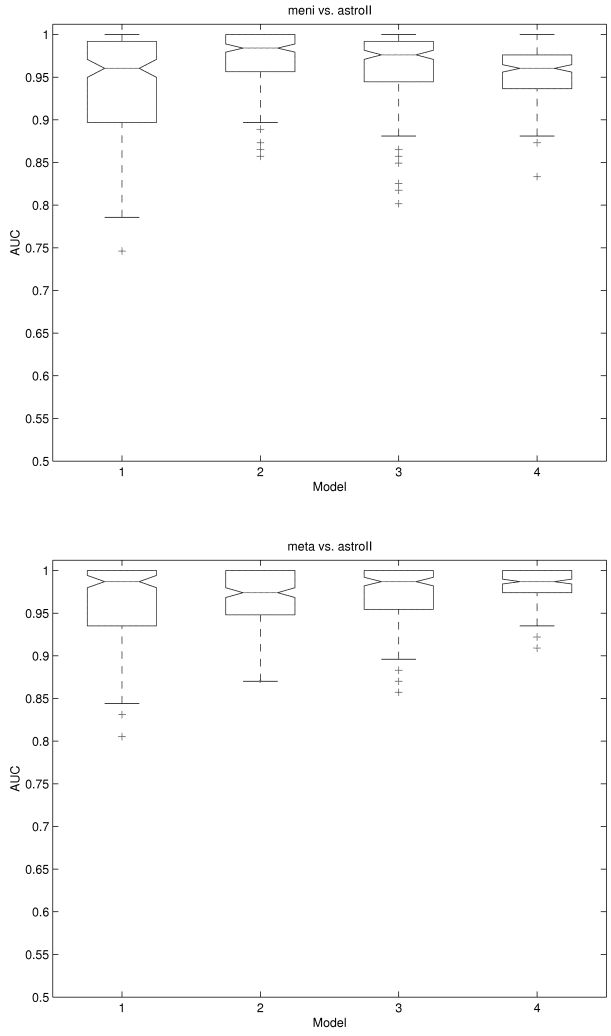

Figure 4.6: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) PCA/LDA, (2) SVM with a linear kernel, (3) LS-SVM with a linear kernel, (4) LS-SVM with an RBF kernel. The figures correspond to the binary classifiers using complete spectra: meni vs. astroII (top) and meta vs. astroII (bottom). For further explanation, see Figure 4.4.

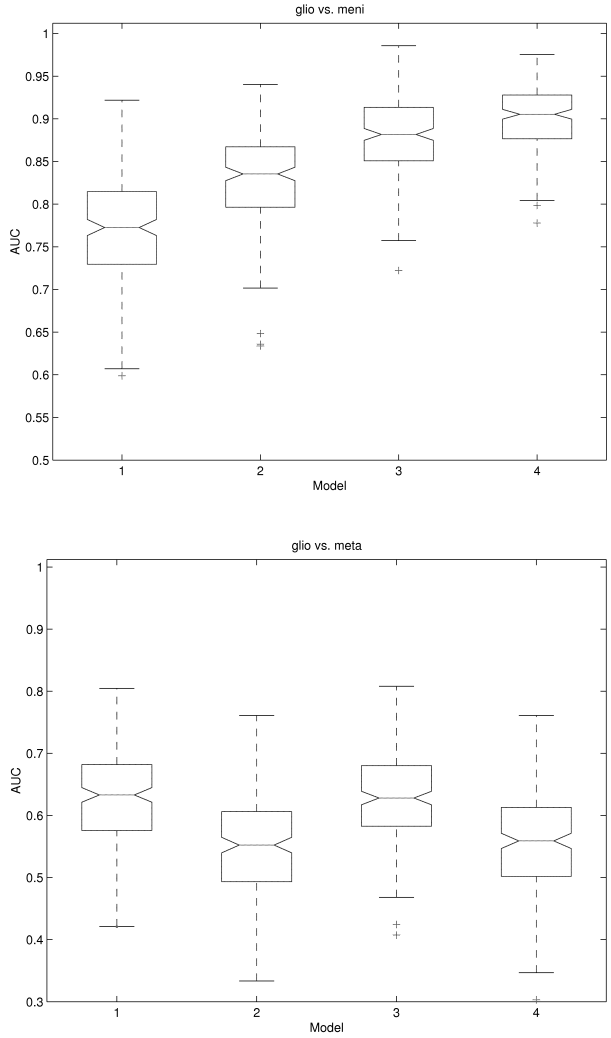

Figure 4.7: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) PCA/LDA, (2) SVM with a linear kernel, (3) LS-SVM with a linear kernel, (4) LS-SVM with an RBF kernel. The figures correspond to the binary classifiers using selected frequency regions: glio vs. meni (top), glio vs. meta (bottom). For further explanation, see Figure 4.4.

4.6 Conclusions

This chapter showed a comparative study of brain tumour classification based on long echo time ^{1}H MRS signals. Linear as well as nonlinear classifiers were compared. All techniques were applied automatically, including (hyper-) parameter selection, training and testing. Also for use in clinical practice, all techniques are easy to automate for analysis of independent data.

Binary classification gives more insight on the distributions of each class and their overlap. Except for the hard case glioblastomas vs. metastases, all classifiers based on the complete magnitude spectra are able to reach an AUC of more than 0.9. Based on the available data, we were not able to statistically prove any difference in performances between the classification techniques, for binary as well as multiclass classification. This indicates that kernel-based methods and LDA statistically perform as well for classification of brain tumours based on a small set of long echo time ^{1}H MRS data.

However, each of the applied techniques has its characteristics. LDA is a simple technique, but requires a prior dimensionality reduction of input variables (e.g. by applying PCA), while dimensionality reduction is done automatically in kernel-based methods.

We expected that dimensionality reduction, by selecting frequency regions or peak integration, would reduce the disturbing noise and artefacts in the spectra. However, the described approach for selecting resonance peaks of long echo time ^{1}H MRS spectra resulted in a lower performance. It might be necessary to include additional spectral information to increase classification performance. This also motivates further research in learning the peak pattern of short echo time ^{1}H MRS, for which data are also provided within the INTERPRET-project.

By using magnitude spectra, phasing problems are avoided. Nevertheless, with respect to real spectra, in magnitude spectra there occurs more peak overlap. Also, in real spectra at long echo time TE (TE=135, 136 ms) the peaks of *Ala* and *Lac* are inverted. This might reduce the ability to distinguish tumour types based on subtle differences in the spectral pattern. In order to test for this effect, in a future study also real spectra could be included as input features.

Discriminating aggressive tumour types, glioblastomas and metastases, using long echo time ^{1}H MRS spectra clearly is a very hard problem due to the highly similar pattern of the spectra from both classes, possibly due to the presence of some necrotic tissue. In order to address this problem, the discriminatory information present in the ^{1}H MRS spectrum should be enhanced, potentially by considering other types of MR signals. In particular, improvements could be expected when processing short echo time ^{1}H MRS signals, since more metabolites are visible in these spectra. Moreover, this spectral information is spread

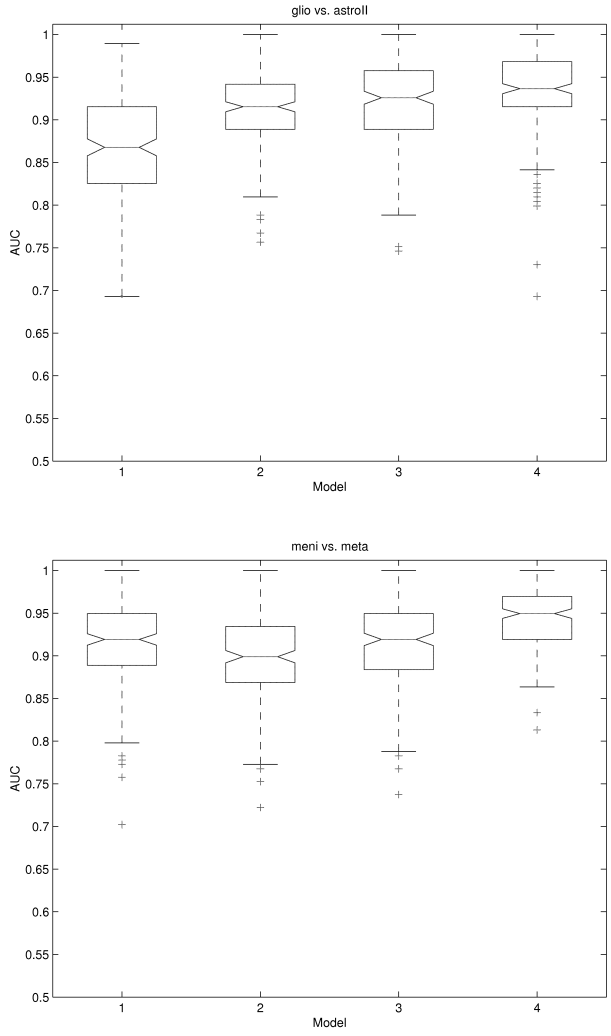

Figure 4.8: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) PCA/LDA, (2) SVM with a linear kernel, (3) LS-SVM with a linear kernel, (4) LS-SVM with an RBF kernel. The figures correspond to the binary classifiers using selected frequency regions: glio vs. astroII (top), meni vs. meta (bottom). For further explanation, see Figure 4.4.

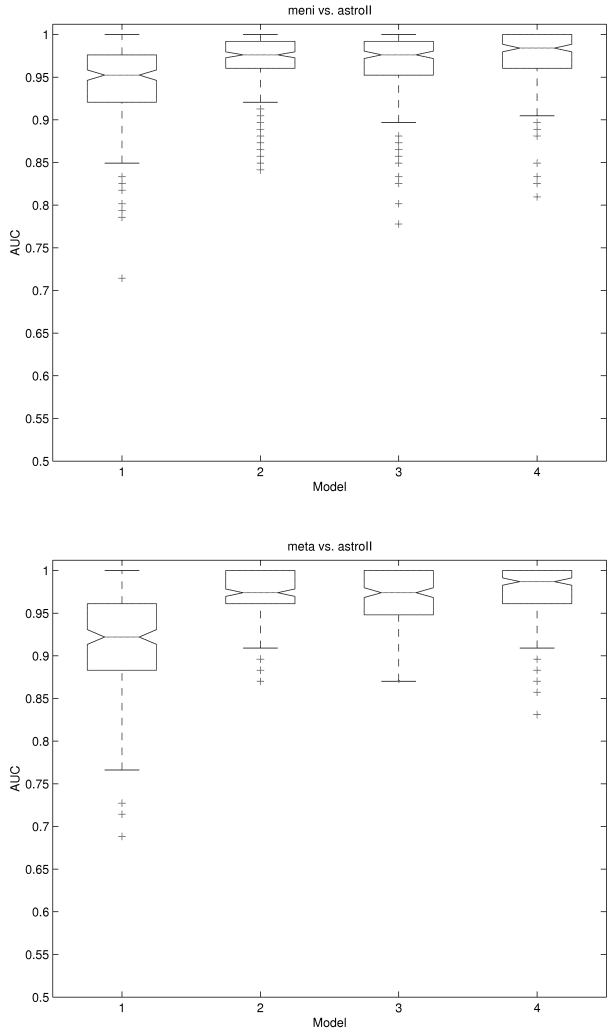

Figure 4.9: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) PCA/LDA, (2) SVM with Linear kernel, (3) LS-SVM with Linear kernel, (4) LS-SVM with RBF kernel. The figures correspond to the binary classifiers using selected frequency regions: meni vs. astroII (top) and meta vs. astroII (bottom). For further explanation, see Figure 4.4.

out over a larger amount of peaks, thereby enlarging the number of possible discriminatory features. In the next chapter the use of short echo time signals is investigated.

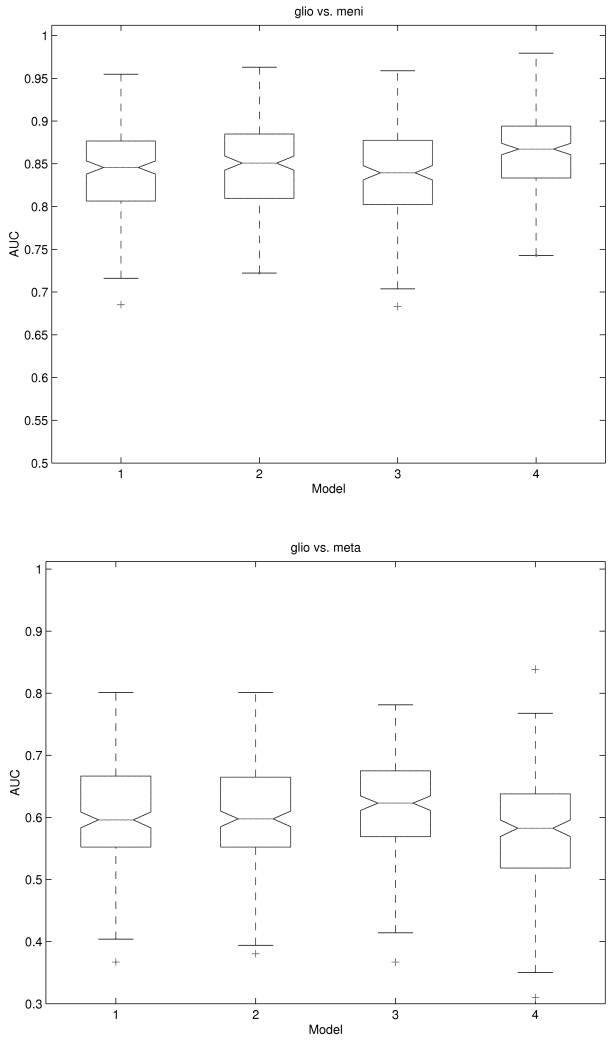

Figure 4.10: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) LDA, (2) SVM with Linear kernel, (3) LS-SVM with Linear kernel, (4) LS-SVM with RBF kernel. The figures correspond to the binary classifiers using peak integration: glio vs. meni (top), glio vs. meta (bottom). For further explanation, see Figure 4.4.

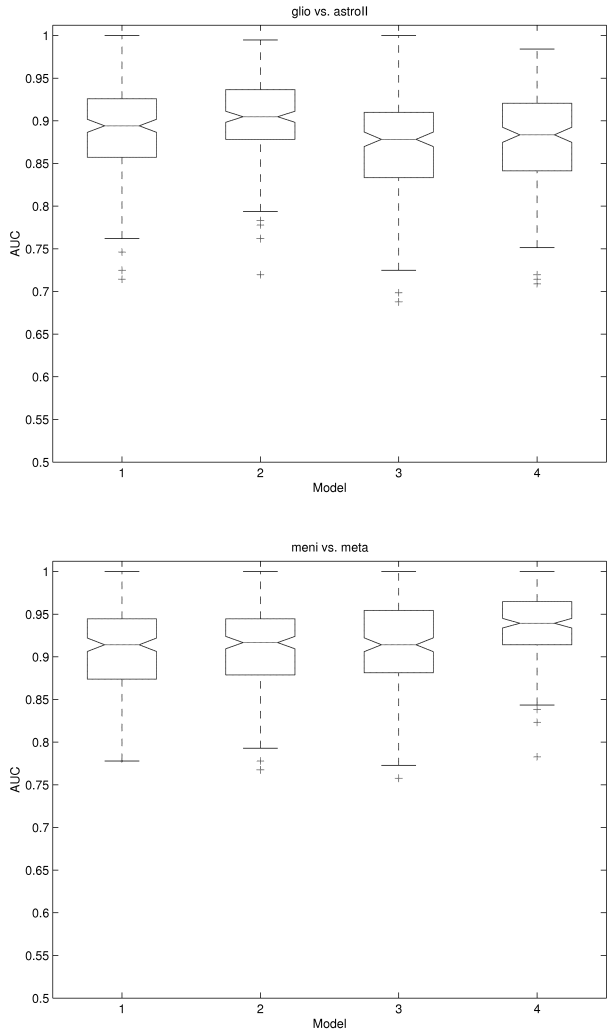

Figure 4.11: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) LDA, (2) SVM with Linear kernel, (3) LS-SVM with Linear kernel, (4) LS-SVM with RBF kernel. The figures correspond to the binary classifiers using peak integration: glio vs. astroII (top), meni vs. meta (bottom). For further explanation, see Figure 4.4.

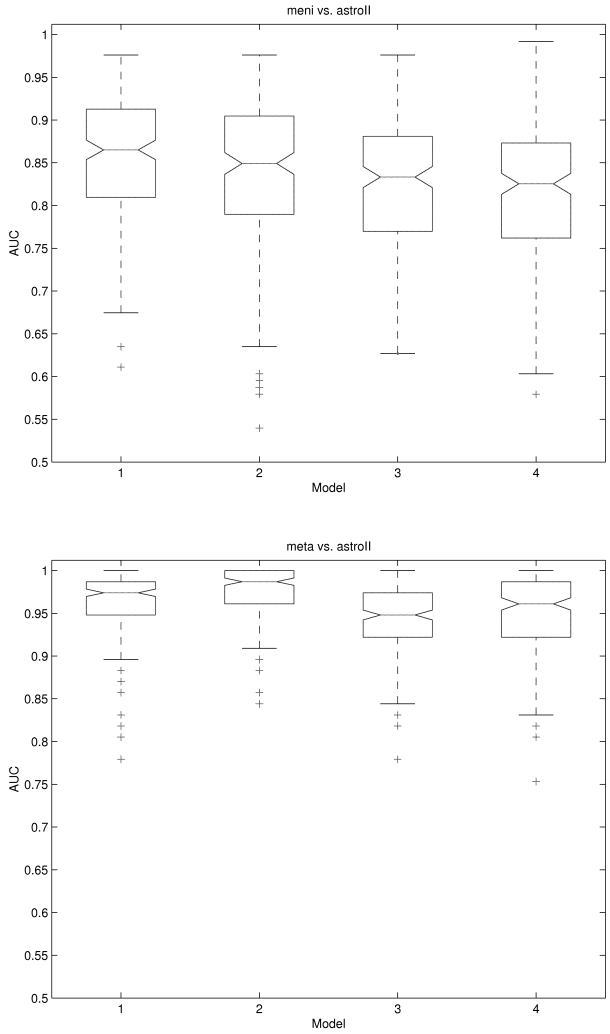

Figure 4.12: Boxplots of the area under ROC curves (AUC) on 200 stratified randomly sampled test sets of the four models (1) LDA, (2) SVM with Linear kernel, (3) LS-SVM with Linear kernel, (4) LS-SVM with RBF kernel. The figures correspond to the binary classifiers using peak integration: meni vs. astroII (top) and meta vs. astroII (bottom). For further explanation, see Figure 4.4.

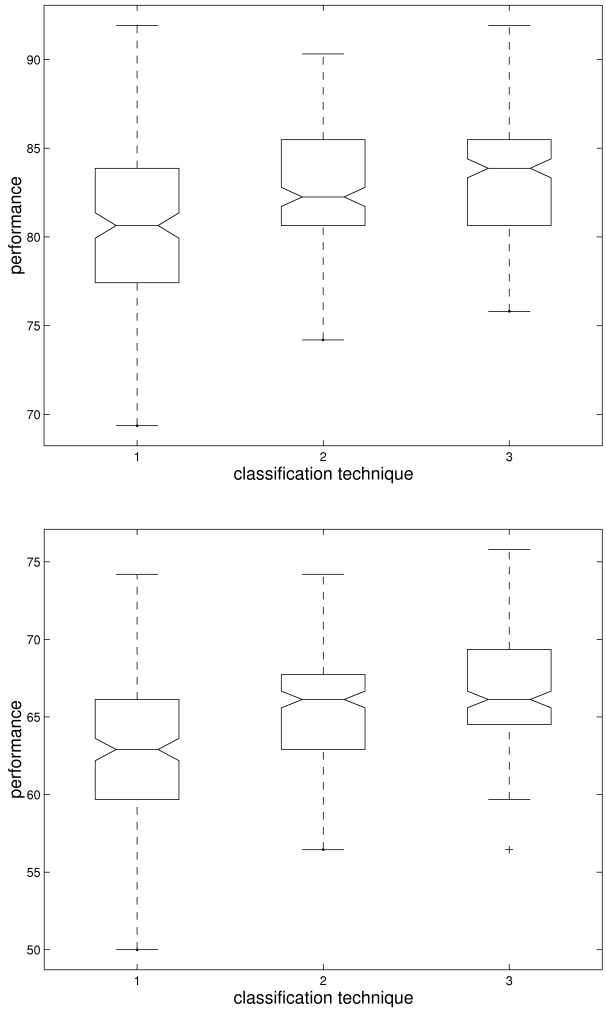

Figure 4.13: Boxplots of the correct classification rate on the test set from 200 runs of stratified random samplings (2/3 of the data for training, 1/3 for testing) of the three models (1) LDA, (2) LS-SVM with Linear kernel, (3) LS-SVM with RBF kernel. The two figures correspond to the multiclass classifiers using complete spectra: output of step-one classifier (top), output of step-two classifier (bottom). For further explanation, see Figure 4.4.

Chapter 5

Classification of brain tumours using short echo time ^{1}H MRS

Short echo time proton spectra are expected to provide more information than long echo time proton spectra, but are more difficult to analyze. While chapter 4 described a classification study on long echo time spectra, this chapter involves the use of short echo time spectra for classification of brain tumours. The purpose of this chapter is to investigate the influence of several factors on the classification performance, including several preprocessing steps and dimensionality reduction.

Section 5.2 gives a description of the data and its preprocessing, while section 5.3 explains the use of several types of input features and the experimental setting. Similarly as in the previous chapter, algorithms are designed to extract the most important features which are then used to classify each spectrum into the corresponding tumour type. As classification is required to be objective and user-friendly, all techniques are automated. In section 5.4 the results of binary and multiclass classification are tabulated, including the test results for the influence of dimensionality reduction and the use of input features. The performance results of binary as well as multiclass classifiers and the influence of dimensionality reduction are discussed in section 5.5. Similar results are obtained for all classification techniques and input features except for water normalized spectra, where classification performance is lower. This indicates that data acquisition and processing can be simplified for classification purposes, excluding the need for separate water signal acquisition, baseline correction or phasing, as is concluded in section 5.6.

5.1 Introduction

The purpose of this chapter was to objectively compare the application of several techniques and the use of several input features for brain tumour classification using short echo time ^{1}H MRS data. Several well-established classification and feature selection techniques were applied to the multicentre data set of short echo time ^{1}H MRS signals. The spectra were acquired on scanners from GE, Philips and Siemens, the three leading manufacturers of these MR scanners, using the major acquisition sequences PRESS and STEAM.

In the previous chapter on long echo time magnitude spectra, automated binary classifiers reached a mean AUC of more than 0.90 except for the most difficult binary classification to discriminate glioblastomas from metastases. In comparison with long echo time spectra, short echo time spectra are more difficult to analyze due to a higher number of overlapping peaks, a more prominent baseline and a higher sensitivity to artefacts. However, short echo time spectra are richer in information than long echo time spectra, as several resonances diminish in apparent intensity at longer echo times (Barba et al., 2001; Ernst and Hennig, 1991), because of a small T_2-value or signal cancellation due to J-modulation (Ernst et al., 1993; Govindaraju et al., 2000). For example, resonances of *mI* (triplets and multiplet at 3.26 and 3.57 ppm), *Glu* (multiplets at 2.33 ppm and 3.74 ppm), *Gln* (multiplets at 2.43 ppm and 3.75 ppm) are much less pronounced in spectra at longer echo times. Here, we report our results of an extended study based on short echo time spectra. We investigated whether the applied pattern recognition techniques were able to exploit the large amount of information available in short echo time MR spectra, despite the problems in analyzing these data, in order to discriminate several types of brain tumours.

5.2 Material

The short echo time INTERPRET database contained single voxel ^{1}H MR spectra from 6 centres: CDP (Centre Diagnòstic Pedralbes), Barcelona (Spain); FLENI (Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia), Buenos Aires (Argentina); IDI (Institut de Diagnòstic per la Imatge), Barcelona (Spain); MUL (Uniwersytet Medyczny w Łodzi), Łodzi (Poland); SGHMS (St George's Hospital Medical School), London (United Kingdom) and UMCN (Universitair Medisch Centrum Nijmegen), Nijmegen (The Netherlands). Table 5.1 reports the most important acquisition parameters. Quality control criteria have been applied to MR system performance and data from all centres (van der Graaf et al., 2002) and data were acquired according to a consensus protocol (van der Graaf, 2001), like described in chapter 4. Data were acquired at 1.5T either by a STEAM- or PRESS-sequence, at an echo

time (TE) between 20 and 32 ms. Note that TE can differ somewhat from one centre to another, or even between data from the same centre, due to variations in the optimised acquisition protocols supplied by the manufacturers. This range of TE will mostly affect the appearance of signals from metabolites with coupled spins (e.g. *Glx*, *Lac*, *Ala*) and the intensity of signals from macromolecules which have a short T_2. The resonance pattern of signals from coupled spins is also affected by the choice of the acquisition sequence used (PRESS or STEAM). Nevertheless, Tate et al. (2003) concluded that pattern recognition is less sensitive to the effects of differences in acquisition parameters than had been expected.

centre	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
		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

Table 5.1: Overview of the acquisition parameters of the short echo time ^{1}H MRS data from all centres. The columns respectively correspond to the name of the centre, the manufacturer of the scanner, the sequence used, the echo time, the repetition time, the spectral width and the number of points in the original FID.

Centre	acq. scheme	1	2	3	4	total
CDP	STEAM	30	12	5	4	51
SGHMS	STEAM	12	4	5	6	27
	PRESS	6	8	11	4	29
FLENI	PRESS	2	0	0	1	3
IDI	PRESS	31	30	17	5	83
UMCN	STEAM	2	1	0	1	4
MUL	STEAM	4	2	1	1	8
	total	87	57	39	22	205

Table 5.2: Number of short echo time ^{1}H MRS data of glioblastomas (class 1), meningiomas (2), metastases (3) and astrocytomas grade II (4). The rows correspond to the acquisition centre, while the columns show the type of brain tumour.

Four common brain tumour types are considered in our study: glioblastomas, meningiomas, metastases and astrocytomas grade II, which for the purpose of

grouping are labelled as class 1, 2, 3 and 4, respectively (Table 5.2). The tumour types are selected using the same selection, quality control and validation procedures as mentioned in chapter 4. Tumour assignment was based on the histological classification of Tumours of the Central Nervous System (CNS) set up by the World Health Organization (WHO) (Kleihues et al., 2002).

Preprocessing of the raw MRS data prior to classification was performed using the same procedure as applied to the long echo time spectra (see section 4.2). The resulting FID was then transformed to the frequency domain by FFT (Fast Fourier Transformation). For classification, only the data points in the frequency region of interest (4.17 to 0 ppm), corresponding to 138 input variables, were used. Figures 5.1 and 5.2 depict the mean L_2-normalized magnitude spectra without baseline correction of the four classes.

5.3 Methods

5.3.1 Input features

The influence of several factors on the classification performance has been tested: L_2- vs. water normalization, magnitude vs. real spectra and baseline correction. The effect of input feature reduction was also investigated by using only the selected frequency regions containing the most discriminatory information, and peak integration values.

Normalization

Two normalization approaches were compared in order to test for the influence of normalization on brain tumour classification. Both normalization methods were mentioned in 3.2.2 and were elaborated in the following way:

- L_2-normalization. Mathematical normalization was considered in which each spectral value was divided by the L_2-norm of the spectrum.
- Water normalization. Normalization was performed with respect to the intensity of water, calculated from the water unsuppressed signal. The amplitude of the water unsuppressed signal was estimated with KULeuven ESAT-SCD software developed by Leentje Vanhamme (personal communication). The water unsuppressed signal was modelled in the time domain by a Voigt model (Marshall et al., 1997), with an additional first order term that corrects 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.

Within the INTERPRET-project, data were acquired using standard protocols, (i.e. short echo time PRESS or STEAM), to obtain MRS data measured under acquisition parameters that are as uniform as possible (Table 5.1). Water normalization compensates for differences in voxel size and receiver coil sensitivity variations, assuming that different brain regions and tissues have the same water content. However, this assumption is not completely valid (Howe et al., 2003). Absolute quantification (Ernst et al., 1993; Henriksen, 1995; Kreis et al., 1993) might solve this problem, if used with an external reference signal, but is more time consuming and less practical for routine clinical situations.

Magnitude versus real spectra

The use of real and magnitude spectra (see section 3.2.2) were compared for classification of short echo time signals.

Baseline correction

Baseline correction was applied by multiplication of the preprocessed FID by an exponentially decaying apodization function (see 3.2.2), given by $g_n = e^{-\beta n \Delta t}, n = 0, \ldots, N-1$, with N the number of points in the FID and a parameter β. The resulting FID is a model for the broad baseline. The value $\beta = 0.15$ was chosen to avoid, on the one hand a too flat baseline and hence overestimation of several metabolite resonances because of baseline contributions to the metabolite peaks (β too high), and on the other hand the underestimation of several metabolite resonances (β too low). The chosen baseline model was subtracted from the original signal to obtain the baseline corrected signal. The performances of classification using the baseline and non-baseline corrected spectra were compared.

Dimensionality reduction by prior knowledge

Three approaches were assessed to investigate the influence of dimensionality reduction, in a similar manner as in section 4.3.1:

- Complete spectra. The spectra, obtained after preprocessing were used as input features.
- Selected frequency regions. By selecting only the spectral regions of ± 0.075 ppm around the resonances (Govindaraju et al., 2000) of several characteristic metabolites, the number of variables was reduced to 71. The following regions of the MRS spectrum were selected: *L2* (lipids at $\approx$ 0.9 ppm; 0.825 - 0.975 ppm); *L1* (lipids at $\approx$ 1.2 ppm; 1.125 - 1.275 ppm); *Lac* ($^3CH_3$-group; 1.235 - 1.385 ppm); *Ala* ($^1CH_3$-group;

1.395 - 1.545 ppm); *NAc* ($^2CH_3$-group; 1.935 - 2.085 ppm); *Glx* ($^4CH_2$-group; 2.270 - 2.510 ppm); *Cr* ($N(CH_3)$-group; 2.955 - 3.105 ppm); *Cho* ($N(CH_3)_3$-group; 3.115 - 3.265 ppm); *mI* (^{5}CH-group) + *Tau* ($^2CH_2$-group) (3.195 - 3.345 ppm); *mI/Gly* (^{1}CH-, ^{3}CH-, ^{4}CH- and ^{6}CH-groups; 3.450 - 3.680 ppm); *Tau* ($^1CH_2$-group; 3.345 - 3.495 ppm); *Glx* + *Ala* (^{2}CH-groups; 3.665 - 3.815 ppm); *Cr* ($^2CH_2$-group; 3.845 - 3.995 ppm). This set of metabolite peaks was an extension of the one used in the previous study on long echo time MRS (section 4.3.1), because several additional metabolite peaks (e.g. *Glx*, *mI*, *Tau*) are more visible in short echo time MR spectra.

- Peak integration. For each selected metabolite the area under the frequency peak in the spectrum was calculated by using the trapezoidal rule. We have used the same frequency ranges as for the selected frequency regions above, and so reduced the number of input variables to 13.

5.3.2 Experimental setting

As in chapter 4, the standard LDA technique was compared to LS-SVM with linear and RBF kernels. The same experimental setup was used as mentioned in section 4.3.3, except that in this study no standard SVM was applied and that the stratified random sampling was repeated 100 times, instead of 200 times. All other aspects of the experiment for binary as well multiclass classification, including the performance assessment was analogous to that applied in chapter 4.

5.4 Results

This section summarizes the results of binary classification using complete spectra, selected frequency regions and peak integration values, consecutively; comparing L_2- with water normalization, magnitude and real spectra, with and without baseline correction. Results are then shown using a multiclass classification approach.

5.4.1 Binary classification

The test performance of the classifiers using the L_2-normalized complete magnitude spectra without baseline correction is shown in Table 5.3. Tables 5.4-5.8 each differ from Table 5.3 in one single aspect. By comparing Table 5.3 with Tables 5.4-5.8, the individual effect on the classification performance of normalization, baseline correction, real vs. magnitude spectra, dimensionality reduction by selected frequency regions and peak integration, respectively can be

determined. In order not to overload the chapter, the results for the other cases are not tabulated, but can be obtained by correspondence with the author. In this chapter only the tabulated results will be discussed in detail. However the classification results for the other input features all showed a similar behaviour.

In general, all techniques achieved a good performance with mean AUCs up to 0.9971, with only slightly and non-significantly different results when mutually comparing classification techniques. However, the discrimination of glioblastomas and metastases remains a very difficult binary problem, resulting in a very poor performance (with a mean AUC between 0.5054 and 0.6085 and a high pooled SE of up to 0.0984) for all the classification techniques investigated. A slightly lower performance was observed for the discrimination of glioblastomas and metastases when applying PCA/LDA as compared to LS-SVMs (Tables 5.4, 5.5 and 5.8), but for all other binary classification problems a high performance was reached with a mean AUC of more than 0.95. A few exceptions occurred when using water normalized spectra, for which almost all results were worse with respect to using L_2-normalized complete spectra. The discrimination of classes 1 and 2 and of classes 2 and 4 yielded a significantly lower performance for water normalization, using PCA/LDA (glio-meni and meni-astroII) and using LS-SVM with a linear kernel (glio-meni). Although less striking, when using selected frequency regions a significantly lower performance was obtained for the classification of classes glio-meni when applying PCA/LDA as compared to using the complete spectra.

5.4.2 Multiclass approach

After merging of glioblastomas and metastases into one class of aggressive tumours, two additional binary classifiers were constructed in addition to those previously discussed. Table 5.9 shows the performance of these classifiers using the complete spectra as input. Tables 5.10 (first step) and 5.11 (second step) report the multiclass test results using complete spectra. A mean correct test classification rate of 85.1% (PCA/LDA), 87.3% (LS-SVM lin), and 86.6% (LS-SVM RBF) has been reached after the first step for all classification techniques. After the second step, the test performance has been reduced to 67.6% (PCA/LDA), 69.4% (LS-SVM lin), and 68.5% (LS-SVM RBF) because of the difficulty in discriminating classes 1 and 3.

5.5 Discussion

We next discuss various issues concerning the results we obtained using the short echo time ^{1}H MRS data available.

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
glio-meni	$0.9556 \pm 0.0278_{(4)}$	0.9725 ± 0.0210	0.9730 ± 0.0204
glio-meta	$0.5907 \pm 0.0967_{(4)}$	0.5917 ± 0.0939	0.5873 ± 0.0942
glio-meni	$0.9660 \pm 0.0292_{(3)}$	0.9635 ± 0.0305	0.9615 ± 0.0314
meni-meta	$0.9540 \pm 0.0441_{(4)}$	0.9770 ± 0.0263	0.9781 ± 0.0254
meni-astroII	$0.9965 \pm 0.0093_{(11)}$	0.9959 ± 0.0100	0.9902 ± 0.0168
meta-astroII	$0.9856 \pm 0.0248_{(2)}$	0.9916 ± 0.0178	0.9903 ± 0.0194

Table 5.3: Binary classification using L_2-normalized complete magnitude spectra without baseline correction. As performance measure the mean test AUC and its pooled standard error SE was used from 100 runs of stratified random samplings. The number between the brackets mentions the number of principal components used (column PCA/LDA).

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
glio-meni	$0.8399 \pm 0.0607_{(1)}$	0.9294 ± 0.0397	0.9581 ± 0.0281
glio-meta	$0.5492 \pm 0.0935_{(1)}$	0.5917 ± 0.0928	0.5950 ± 0.0935
glio-meni	$0.9485 \pm 0.0403_{(1)}$	0.9560 ± 0.0339	0.9598 ± 0.0326
meni-meta	$0.8924 \pm 0.0618_{(1)}$	0.9616 ± 0.0470	0.9704 ± 0.0309
meni-astroII	$0.7427 \pm 0.1199_{(1)}$	0.9930 ± 0.0139	0.9793 ± 0.0303
meta-astroII	$0.9784 \pm 0.0297_{(1)}$	0.9757 ± 0.0386	0.9837 ± 0.0284

Table 5.4: Binary classification using water normalized complete magnitude spectra without baseline correction. For further explanation we refer to Table 5.3.

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
glio-meni	$0.9591 \pm 0.0265_{(5)}$	0.9625 ± 0.0271	0.9616 ± 0.0269
glio-meta	$0.5762 \pm 0.0939_{(5)}$	0.5914 ± 0.0931	0.6002 ± 0.0930
glio-meni	$0.9650 \pm 0.0295_{(4)}$	0.9584 ± 0.0335	0.9587 ± 0.0334
meni-meta	$0.9607 \pm 0.0388_{(5)}$	0.9685 ± 0.0308	0.9736 ± 0.0281
meni-astroII	$0.9965 \pm 0.0099_{(13)}$	0.9973 ± 0.0082	0.9960 ± 0.0100
meta-astroII	$0.9725 \pm 0.0411_{(3)}$	0.9813 ± 0.0309	0.9836 ± 0.0282

Table 5.5: Binary classification using L_2-normalized complete magnitude spectra with baseline correction. For further explanation we refer to Table 5.3.

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
glio-meni	$0.9543 \pm 0.0279_{(6)}$	0.9743 ± 0.0196	0.9742 ± 0.0199
glio-meta	$0.5628 \pm 0.0984_{(4)}$	0.5655 ± 0.0970	0.5753 ± 0.0973
glio-meni	$0.9755 \pm 0.0240_{(5)}$	0.9717 ± 0.0264	0.9691 ± 0.0278
meni-meta	$0.9562 \pm 0.0425_{(6)}$	0.9702 ± 0.0319	0.9734 ± 0.0288
meni-astroII	$0.9959 \pm 0.0102_{(13)}$	0.9947 ± 0.0119	0.9917 ± 0.0154
meta-astroII	$0.9924 \pm 0.0167_{(4)}$	0.9893 ± 0.0210	0.9875 ± 0.0233

Table 5.6: Binary classification using L_2-normalized complete real spectra without baseline correction. For further explanation we refer to Table 5.3.

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
glio-meni	$0.9060 \pm 0.0430_{(2)}$	0.9748 ± 0.0202	0.9733 ± 0.0200
glio-meta	$0.5959 \pm 0.0932_{(2)}$	0.5898 ± 0.0945	0.6085 ± 0.0930
glio-meni	$0.9500 \pm 0.0358_{(1)}$	0.9636 ± 0.0318	0.9620 ± 0.0315
meni-meta	$0.9495 \pm 0.0468_{(2)}$	0.9712 ± 0.0316	0.9683 ± 0.0336
meni-astroII	$0.9958 \pm 0.0103_{(7)}$	0.9956 ± 0.0105	0.9971 ± 0.0086
meta-astroII	$0.9947 \pm 0.0133_{(1)}$	0.9924 ± 0.0172	0.9929 ± 0.0165

Table 5.7: Binary classification using selected frequency regions of the L_2-normalized magnitude spectra without baseline correction. For further explanation we refer to Table 5.3.

Classes	LDA	LS-SVM lin	LS-SVM RBF
glio-meni	0.9675 ± 0.0243	0.9704 ± 0.0221	0.9713 ± 0.0210
glio-meta	0.5054 ± 0.0940	0.5762 ± 0.0956	0.6073 ± 0.0935
glio-meni	0.9580 ± 0.0335	0.9616 ± 0.0328	0.9539 ± 0.0355
meni-meta	0.9641 ± 0.0364	0.9640 ± 0.0378	0.9632 ± 0.0379
meni-astroII	0.9968 ± 0.0090	0.9965 ± 0.0092	0.9875 ± 0.0201
meta-astroII	0.9892 ± 0.0219	0.9938 ± 0.0147	0.9938 ± 0.0146

Table 5.8: Binary classification using peak integration values of the L_2-normalized magnitude MR spectra without baseline correction. Note that in this case no PCA prior to LDA was applied, because peak integration already involves dimensionality reduction. For further explanation we refer to Table 5.3.

Classes	PCA/LDA	LS-SVM lin	LS-SVM RBF
meni-aggr	$0.9570 \pm 0.0245_{(4)}$	0.9803 ± 0.0148	0.9801 ± 0.0148
astroII-aggr	$0.9701 \pm 0.0232_{(3)}$	0.9728 ± 0.0225	0.9725 ± 0.0226

Table 5.9: Binary classification using complete spectra for two additional binary problems, needed to implement the multiclass classification. Average performance on the test set from 100 runs of stratified random samplings of the L_2-normalized magnitude MR spectra without baseline correction. As performance measure the mean AUC and its pooled standard error SE was used.

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
correct	85.0882 ± 4.2939	87.2794 ± 3.6466	86.6324 ± 3.4857
misclass	9.3676 ± 3.6722	8.7941 ± 3.5409	9.8971 ± 3.4336
undecided	5.5441 ± 2.4183	3.9265 ± 2.0592	3.4706 ± 1.9102

Table 5.10: One-step multiclass classification using complete spectra. Average test performance from 100 runs of stratified random samplings of the L_2-normalized magnitude MR spectra without baseline correction. The first, second and third rows give, respectively, the mean correct classification rate, the mean misclassification rate and the mean percentage of undecided cases, each with their standard deviation.

5.5.1 Classification techniques

LDA has been widely used and has obtained good results for many applications including brain tumour classification (Tate et al., 2003; Hagberg et al., 1995). The method has the strength of being simple and computationally fast, and works well for data that are linearly separable. Nevertheless, LDA also has some disadvantages. These include the nonuniqueness of the derived discrimination function (Lindon et al., 2001) and the potentially poor performance in the case of severe nonlinearity. Moreover, in the case of a high number of variables relative to the number of training data (e.g. in the study presented), LDA requires the data to be reduced in dimensionality, as by PCA for example. In contrast to classical techniques, LS-SVM as a kernel-based technique is also suited for linearly nonseparable problems and is able to find a global solution (Suykens et al., 2002). Kernel-based methods are also less sensitive to the number of data. Although the dimension is larger than the number of data, these classifiers have the advantage to robustly learn the peak pattern and draw an optimal separating boundary, even without applying any prior dimensionality reduction. Linear as well as nonlinear classifiers can be selected by choosing an appropriate kernel. However, in contrast to classical techniques, LS-SVM is a more complex method that requires to select a set of hyperparameters in order to obtain a high performance.

	PCA/LDA (%)	LS-SVM lin (%)	LS-SVM RBF (%)
correct	67.6176 ± 4.9418	69.4118 ± 3.6082	68.5441 ± 4.0417
misclass	26.8382 ± 4.5043	26.6618 ± 3.6122	27.9853 ± 4.0984
undecided	5.5441 ± 2.4183	3.9265 ± 2.0592	3.4706 ± 1.9102

Table 5.11: Two-step multiclass classification using complete spectra. Average test performance from 100 runs of stratified random samplings of the L_2-normalized magnitude MR spectra without baseline correction.

LDA, as a linear classifier, performed quite well in discriminating brain tumours using complete spectra. Except for the most difficult binary classification to discriminate glioblastomas and metastases, classification of L_2-normalized non-baseline corrected complete magnitude spectra reached a mean AUC of at least 0.95. Using the same input features as LDA, LS-SVMs with a linear and an RBF kernel both reached a mean AUC of at least 0.96, which is similar to the performance of LDA. Although the tables indicate a few differences in the classification performance, no statistical significance was found between the AUC values from different classification techniques. Due to the limited number of data available (i.e. limited in terms of classification), nonlinear techniques, in this case LS-SVMs with an RBF kernel, could not fully exploit the advantage of a nonlinear method.

These findings are in agreement with our previous study on long echo time spectra reported in the previous chapter and results described in Tate et al. (2003). Direct comparison of the performances even reveals overall slightly higher mean AUC values with respect to the use of long echo time spectra (confront Tables 5.3, 5.7 and 5.8 with Tables 4.3, 4.4 and 4.5), but this was not statistically tested because different datasets were used.

5.5.2 Influence of input features

The purpose of normalization is to obtain consistent scaling of the measurement data. Using the complete magnitude spectra without baseline correction, a lower performance was found for water normalization compared to L_2-normalization. This included a few significant differences when using PCA/LDA and LS-SVM with a linear kernel. Only slightly different results were observed due to the normalization method when using nonlinear techniques.

Although Klose's method was used to phase the spectra, there still exists residual phase variation in the real spectra, which may have reduced the accuracy of classification. Phasing problems can be avoided by using magnitude spectra, but this yields a significant broadening of the peaks (Hoch and Stern, 1996). Hence, magnitude calculation results in a larger overlap of the individual peaks, which potentially distorts the information contained in the spectrum. However,

no significant differences in classification performance were found between the use of complete real and magnitude spectra without baseline correction.

In short echo time spectra the appearance of the metabolite peaks is heavily influenced by peak overlap and the presence of a macromolecular baseline, hence a baseline correction was applied in an attempt to improve resolution of the metabolite peaks. Spectra with baseline correction were expected to have clearer peak patterns and to result in a more robust classification, but no improvement was found. Although our baseline correction may simplify the identification of metabolite peaks, macromolecule signals included in the baseline may contribute to the biochemical fingerprint of the tumours. Hence, baseline correction may result in loss of some valuable information, which may explain why classification does not improve after baseline correction.

5.5.3 Influence of dimensionality reduction

In classifying MR spectra using selected frequency regions and peak integration, all techniques achieved at least a mean AUC of 0.90 except for the binary classification problem to discriminate glioblastomas and metastases. In comparison to classification using complete spectra, classification using selected frequency regions or peak integration values reached a similar performance. This suggests that the selected metabolite regions or integrated signal areas include most of the information that correctly distinguishes between types of brain tumour. Only for the discrimination of glioblastomas and meningiomas PCA/LDA performed significantly worse using selected frequency regions than using the complete spectra.

5.5.4 Multiclass classification

Although no statistical analysis has been carried out for multiclass classification, we can make some observations from the results. All classification techniques reached a high test performance of 85% to 87% for distinguishing classes 2, 4 and 5 in the first step. In the second step, the test performances reduced with about 18% because of the difficulties in separating glioblastomas from metastases, which share very similar peak patterns. Therefore, separating class 5 (aggressive tumours) into class 1 (glioblastomas) and 3 (metastases) deteriorated the total performance of the classifier.

5.5.5 Clinical practice

We have performed an objective comparison of several classification techniques, based on several types of input features. This requires the selection of a training set to construct a classifier and an independent test set to evaluate that

classifier. Therefore we implemented an experimental setting of 100 stratified randomizations, which resulted each time in another classifier.

The author is aware of the fact that, from a clinical point of view, a practical classification system requires only one specific classifier. The classification system is then used to assign new MRS data to a certain class, corresponding to the type of the tumour and its grade. However, our objective comparison makes it possible to obtain an overview of the performance of several techniques, with their advantages and disadvantages, and to evaluate the influence on classification of the type of input features used. From such results it should be possible to make an informed choice of classification technique, as well as of the most appropriate input features and implement this in a practical classification system. For example, this selection could depend on the computational complexity and on the types of brain tumours the system typically should be able to discriminate. It should be noted that all techniques are applied automatically, including (hyper-)parameter selection, training and testing. Hence, for use in clinical practice, all techniques are easy to automate for analysis of independent data.

5.6 Conclusions

In general, no significant differences were found with respect to the use of classification techniques. Linear techniques found the best separating linear boundary for most of the cases. Kernel-based methods have the advantage of robustly learning the peak pattern and were able to reach a high mean AUC (except for glioblastomas vs. metastases), even without the use of any feature reduction. Kernel-based methods are also more robust against different types of normalization than LDA, which may be a result of data information loss due to PCA.

From our analysis, we found that the best classification could be obtained using L_2-normalized magnitude spectra without baseline correction and simply using peak integration or PCA for dimensionality reduction. This is an important result if it would be generally true as it means that neither the difficult task of accurate spectral phasing, or acquisition of a water reference for lineshape correction and zero order phasing would be required. Thus data acquisition times can be reduced and preprocessing protocols can be simplified. This would be particularly useful for spectroscopic imaging data, for which there is a significant time penalty in acquiring a full water reference data set.

Additionally, although peak integration may not accurately quantify overlapping metabolite signals (Vanhamme et al., 2001) and PCA only expresses up to 75% of the data, automated dimensionality reduction by peak integration or PCA speeds up the computation and does not appear to compromise the

classification efficiency for the comparison made in this study. Further studies are still needed to assess whether other approaches to dimensionality reduction, such as more accurate model-based quantification methods (Provencher, 1993; Vanhamme et al., 2001), yield any significant improvement in classification performance.

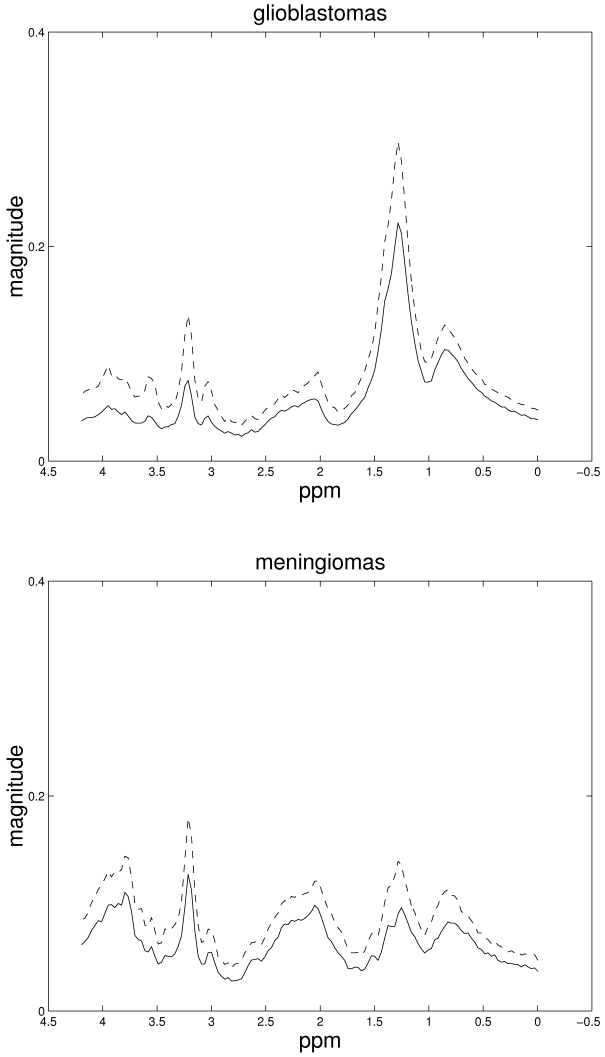

Figure 5.1: Mean L_2-normalized magnitude complete MRS spectra without baseline correction of the four considered classes: class 1 (top), class 2 (bottom) correspond to the glioblastomas, meningiomas respectively. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

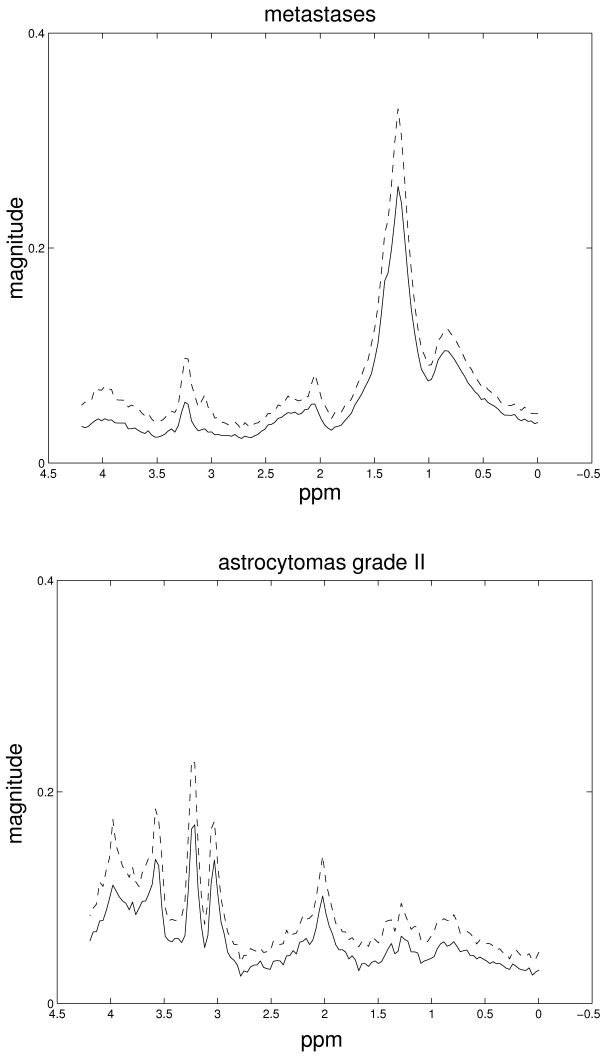

Figure 5.2: Mean L_2-normalized magnitude complete MRS spectra without baseline correction of the four considered classes: class 3 (top) and class 4 (bottom) correspond to the metastases and astrocytomas (grade II) respectively. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

Chapter 6

The use of multivariate MR Imaging intensities versus metabolic data from MR Spectroscopic Imaging for brain tumour classification

Chapters 4 and 5 investigated the use of long and short echo time MRS and illustrated the potential of MRS for brain tumour classification. However, in clinical practice diagnosis is largely based on morphological and spatial information obtained from imaging techniques such as MRI. A pattern recognition system that merges spectroscopic and imaging data should be able to provide a more reliable classification than one that is only based on either spectroscopic or imaging data. Moreover, a more reliable system would then increase the confidence of the medical diagnosis. This motivates to study the influence of using imaging and spectroscopic information for brain tumour classification.

In this chapter MRI intensities and metabolic data from MRSI are used separately and in combination with each other for the discrimination of brain tumours. Although indeed spatial information is lost by using the imaging intensities only, an indication is already given about the relative importance of each of the input types for automatic classification.

The purpose of this chapter is twofold:

- *To investigate the discriminatory value of MRI intensities and metabolic data extracted from MRSI for automated brain tumour diagnosis. This analysis also provides the typical AUC values achievable for several relevant diagnostic problems of brain tumours.*
- *To apply and compare several classification techniques, including the investigation of the influence of the input features used.*

The introduction situates the study with respect to related work about the value of MRI as well as MRSI. Next, section 6.2 describes the data and the preprocessing, while section 6.3 explains the processing of the input features and the experimental setting, which is similar to that of chapters 4 and 5. Section 6.4 mentions an overview of the results of each of the binary problems, which also includes several significant differences. To continue, section 6.5 compares the different techniques and discusses the use of MRI and MRSI and the performance for all classification problems. The results of this study indicate that imaging intensities and metabolic data provide complementary information for the accurate discrimination between several brain tissue types.

6.1 Introduction

Magnetic Resonance Imaging (MRI) is an important non-invasive tool for identifying the location and size of brain tumours, because it yields morphological and anatomical information about the brain tissue. However, conventional MRI has a limited specificity in determining the underlying type of brain tumour and grade (Dean et al., 1990; Earnest et al., 1988). More recently developed MR techniques like diffusion-weighted MRI, perfusion-weighted MRI and Magnetic Resonance Spectroscopic Imaging (MRSI) are promising new techniques in the characterization of brain tumours (Nelson and Cha, 2003; Rees, 2003). Diffusion-weighted MRI visualizes the tissue structure and is useful for assessing tumour cellularity, while perfusion-weighted MRI provides images that reflect changes in tumour vasculature and tumour grading. MRSI or multivoxel Magnetic Resonance Spectroscopy (MRS) provides chemical information about metabolites present in normal and abnormal tissue (Nelson, 2003; Leclerc et al., 2002; Mukherji, 1998; Nelson et al., 2002). The differentiation of abnormal brain tissues, including brain tumours, from normal brain forms a potentially major clinical application of these new techniques. In general, diagnosis of brain tumours is based on the microscopic examination of tissue obtained by a biopsy, which includes risks associated with anesthesia and surgery. It would be very beneficial to the patient if the invasive biopsy could be guided or even avoided by the use of non-invasive techniques like diffusion-weighted MRI, perfusion-weighted MRI and MRS(I). In this study we combined the use of conventional MRI intensities and one of the new techniques, more specifically

MRSI.

Several studies (Herminghaus et al., 2003; Huang et al., 2003; Preul et al., 1996; Simonetti et al., 2003; Szabo De Edelenyi et al., 2000; Tate et al., 1998, 2003; Ye et al., 2002) have shown progress in automated pattern recognition for brain tumour classification using MRI or MRS(I). However, currently only a few studies (Simonetti et al., 2003; Szabo De Edelenyi et al., 2000) have used a combination of MRI and MRSI features for classification of brain tumours. In order to enhance the diagnostic capabilities in clinical practice, we investigated whether the combined use of MR imaging intensities and metabolic data from MRSI could improve the discrimination between several brain tumour and normal brain tissue types. Although the radiologist also uses spatial and morphological information present in the MR images, these features were not taken into account in this study, as they are difficult to quantify. Classification was applied to several input features, such as short echo time magnitude spectra, imaging intensities, peak integration values obtained from the spectra and a combination of the latter two. By comparing the results obtained, the strength was evaluated of both, MR imaging intensities and metabolic data from MRSI, in discriminating brain tissue types.

6.2 Material

Data from 25 patients with a brain tumour and 4 volunteers were selected from the database developed in the framework of the EU funded (INTERPRET-project, 1999-2002). All data were provided by the acquisition center UMCN (University Medical Center Nijmegen), Nijmegen (The Netherlands). Each case was clinically validated. The patients' tumour type was determined by a central consensus histopathological validation. For one of the 25 patients no consensus was reached. Therefore, the data from the tumour region of this patient were not used.

The data set contained MR images as well as MR spectra, acquired and preprocessed as described in Simonetti et al. (2003). For each subject, stacked MR images of cross-sections of the whole brain at four contrasts were acquired: T_1- and T_2-weighted images, a proton density weighted image and a gadolinium enhanced (Gd-DTPA) T_1-weighted image (256x256, FOV = 200 mm, slice thickness = 5 mm). The image values will further be labelled as T_1, T_2, PD and GD. No Gd-DTPA administration was applied to the healthy volunteers. Besides MR images, also ^{1}H MRSI data were acquired for each subject, both with and without water suppression using a 16x16 2D STEAM ^{1}H MRSI sequence with acquisition parameters TR = 2000 or 2500 ms, TE = 20 ms, slice thickness 12.5 or 15 mm, FOV = 200 mm, SW = 1000 Hz, 1024 data points. The position of the MRSI slice was chosen according to the slice position of

the GD image which showed the largest GD enhancement.

To ensure that image pixels from subsequent images originate from the same spatial location, the images were co-aligned (Simonetti et al., 2003). All MRSI data were semi-automatically preprocessed (cf. Simonetti et al. (2003)), which involved:

- Filtering of the k-space data by a Hanning filter of 50 % using the LUISE software package (Siemens, Erlangen, Germany).
- Zero filling to 32x32, which involved an increase of the apparent spatial resolution with a factor of 2.
- Spatial 2D Fourier transformation to obtain time domain signals for each voxel.
- Correction for eddy current effects in the MR spectra using a method which prevents the occasional occurrence of eddy current correction induced artefacts (Simonetti et al., 2002). This process resulted in a frequency alignment and zero order phasing of the MR spectra.
- Removal of the dominating residual water using HLSVD (Pijnappel et al., 1992), with 12 singular values and 4.0-6.0 ppm as residual water region.
- Frequency alignment was performed semi-automatically. First, the position of the NAA peak ($^2CH_3$-group) in the mean spectrum of the MRSI data set was set to 2.02 ppm. The obtained shift was used to reset each spectrum of the data set in the time domain automatically.
- First order phase correction was also manually performed on the mean spectrum of a data set. The obtained first order time instant was used to automatically correct each spectrum in the data set.
- Fourier transformation was applied to the time domain data to obtain frequency spectra.

The set of validated data consisted of 10 grade II, 4 grade III and 7 grade IV gliomas and 3 meningiomas, which gave us 4 classes of brain tumours. From each patient, data were selected from several voxels, identified as lying in the tumour area. Besides brain tumour data, also data from cerebro spinal fluid (CSF) from patients and normal brain tissue from volunteers and patients were selected. Data from all subjects with the same pathology were combined into one class, which resulted in the following 7 classes of pathologies (Table 6.1):

1. normal tissue from volunteers: normal brain tissue from healthy persons,
2. normal brain tissue from patients: apparently normal tissue from the contralateral half of the brain of patients,

3. cerebro spinal fluid (CSF): CSF from patients, where the ventricles were clearly visible and the voxels were located as far from the tumour as possible. Unfortunately, no CSF voxels could be selected from volunteers, since the MRSI slices of volunteers did not include the ventricles.
4. grade II gliomas: diffuse astrocytomas (90 voxels from 5 patients), oligodendrogliomas (22 voxels from 2 patients) and mixtures (64 voxels from 3 patients),
5. grade III gliomas: anaplastic astrocytomas (4 voxels from 1 patient), oligodendrogliomas (25 voxels from 2 patients) and mixtures (28 voxels from 1 patient),
6. grade IV gliomas: glioblastomas,
7. meningiomas.

Label	Pathology	No. of data	No. of subjects
1	normal from volunteers	142	4
2	normal from patients	76	4
3	CSF	100	8
4	gliomas, grade II	176	10
5	gliomas, grade III	57	4
6	gliomas, grade IV	70	7
7	meningiomas	48	3
	Total	669	29

Table 6.1: Number of data for each type of brain tissue (brain tumour or healthy tissue). The first and second columns give the label of the classes (from 1 till 7) and the pathology. The third and fourth columns display the number of data and subjects for each class. Note that the total number of subjects was not simply the summation of the number of subjects per class, because for several patients data were available from brain tumour as well as from healthy tissue.

6.3 Methods

6.3.1 Input features

In this study an input pattern was either an MR spectrum, a set of quantified values from the MR spectrum, a set of imaging intensities or a combination of the latter two input types. This enabled us to investigate whether the combination of imaging and spectroscopic information can improve the performance for pattern recognition of brain tumours. The following input features were considered:

- Water normalized magnitude spectra (Figures 6.1, 6.2, 6.3 and 6.4). The amplitude of the water unsuppressed signal was estimated as described in section 5.3. Then each spectral value in the preprocessed water suppressed spectrum was divided by the resulting estimate of the intensity of the water peak. Only the spectral values in the region of interest (0.5-4.0 ppm) were used as input features.
- Metabolite amplitudes obtained by peak integration. In order to remove the broad baseline a simple baseline correction (Simonetti et al., 2003; Campbell et al., 1973) was applied as additional preprocessing step prior to the first order phase correction. This was performed as described in Simonetti et al. (2003). Amplitude estimates were then obtained from the baseline corrected frequency spectra using peak integration within a spectral range of 0.13 ppm, around resonances from metabolites and lipids that are assumed to be characteristic to distinguish between tumour types (Leclerc et al., 2002; Mukherji, 1998; Smith and Stewart, 2002; Howe et al., 2003; Majós et al., 2002; Murphy et al., 1993). As short echo time ^{1}H MR *in vivo* spectra are characterized by substantial peak overlap and a relatively low spectral resolution at the clinical field strength of 1.5 T, a particular region might cover resonances of more than one metabolite. Such regions are (Govindaraju et al., 2000): *L2* (lipids at $\approx$ 0.9 ppm; 0.835-0.965 ppm), *L1* (lipids at $\approx$ 1.2 ppm) + *Lac* ($^3CH_3$-group) + *Ala* ($^1CH_3$-group; 1.265-1.395 ppm), *NAA* ($^2CH_3$-group; 1.955-2.085 ppm), *Glx* ($^3CH_2$-group; 2.135-2.265 ppm), *Cr* ($N(CH_3)$-group; 2.955-3.095 ppm), *Cho* ($N(CH_3)_3$-group; 3.135-3.265 ppm), *Tau* ($^1CH_2$-group; 3.375-3.505 ppm), *mI* (^{1}CH-, ^{3}CH-, ^{4}CH- and ^{6}CH-group) + *Gly* ($^2CH_2$-group; 3.495-3.625 ppm), *Glx* + *Ala* (^{2}CH-groups; 3.685-3.815 ppm), *Cr* ($^2CH_2$-group; 3.885-4.015 ppm). The resulting peak integration values were then water normalized as described above.
- Imaging intensities (Figures 6.5 and 6.6). For each of the images (T_1, T_2, PD and GD), an image value was extracted, producing four additional variables. In order to obtain the same spatial resolution of both MRI and MRSI data, the resolution of the MRI data was lowered to that of the MRSI grid by averaging the image pixels within each spectroscopic voxel. Each intensity value was divided by the highest intensity in the corresponding downsampled image and scaled to the same range as the spectral data. No GD information was available for the volunteers. Instead the T_1-image was used, under the assumption that no GD enhancement would occur in the brain tissue of volunteers. However, this is an approximation as Gd-DTPA typically causes an increase in intensity in the blood vessels.

Selection of the data used was based on the visual inspection of the low and high resolution images and MR spectra. To obtain data for a specific histopathological class, the following procedure was used for each patient within this class.

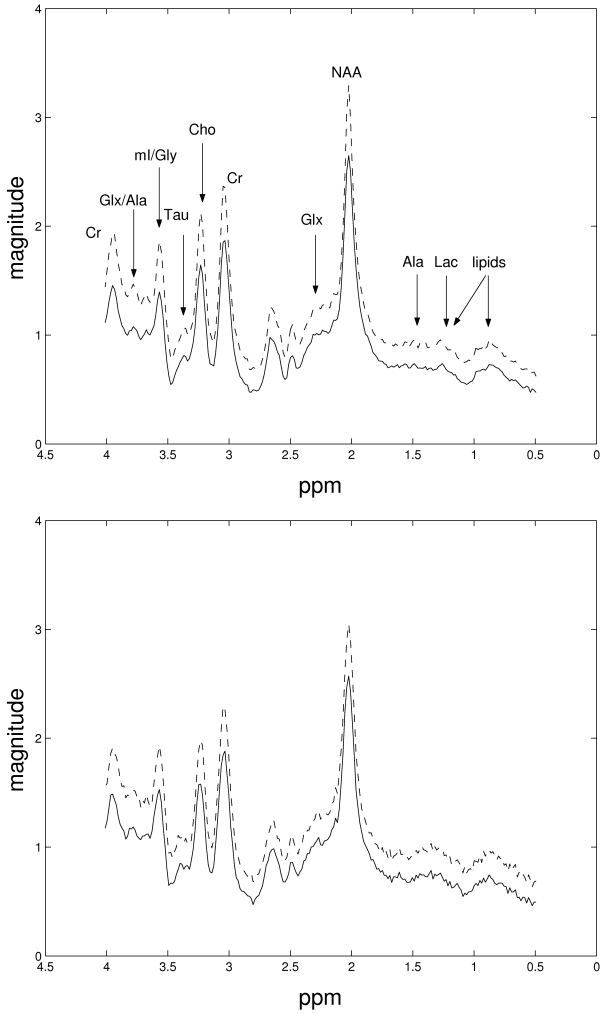

Figure 6.1: Mean water normalized magnitude ^{1}H MR spectra (TE=20 ms) of the considered classes: class 1 (top), class 2 (bottom) correspond to the normal tissue of volunteers, normal tissue of patients. The solid lines are the means, while the dotted lines are the means plus the standard deviations of each class.

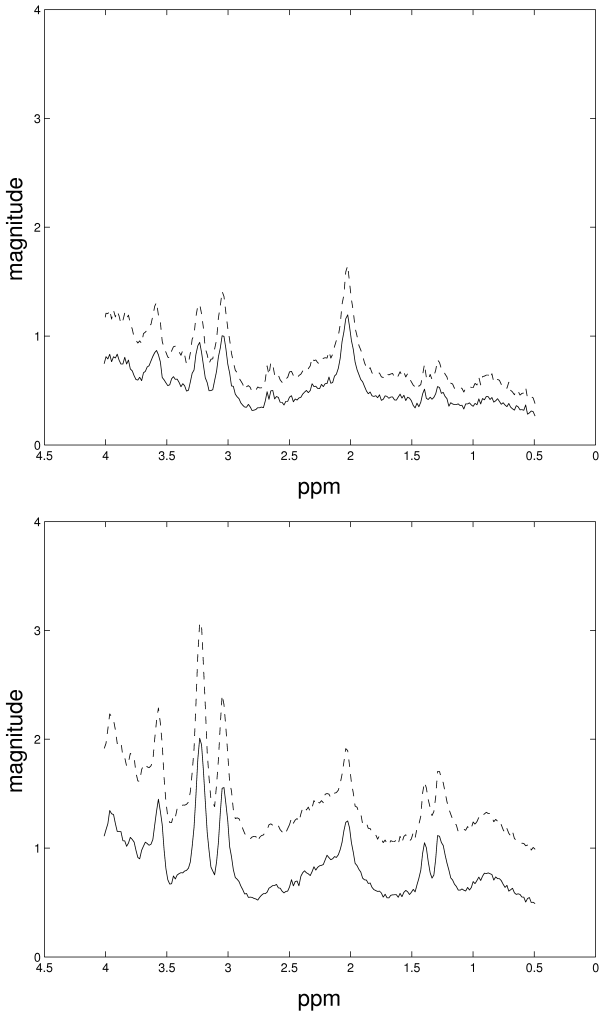

Figure 6.2: Mean water normalized magnitude ^{1}H MR spectra (TE=20 ms) of the considered classes: class 3 (top), class 4 (bottom) correspond to CSF and gliomas of grade II. For further explanation see Figure 6.1.

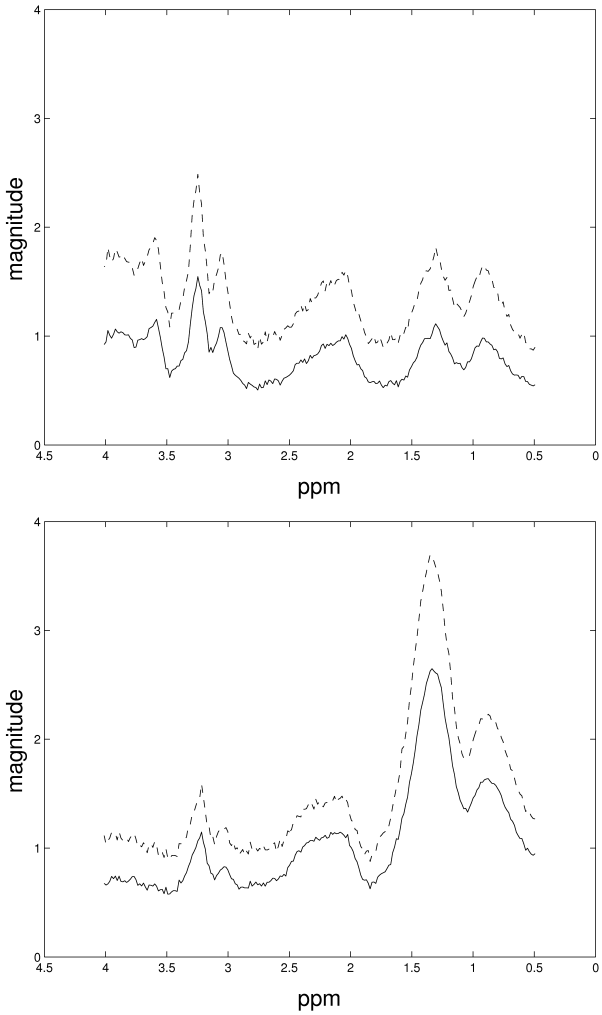

Figure 6.3: Mean water normalized magnitude ^{1}H MR spectra (TE=20 ms) of the considered classes: class 5 (top), class 6 (bottom) correspond to gliomas of grade III and grade IV. For further explanation see Figure 6.1.

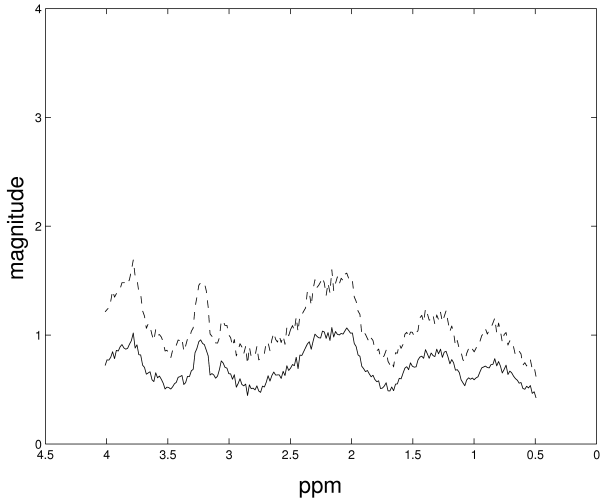

Figure 6.4: Mean water normalized magnitude ^{1}H MR spectra (TE=20 ms) of class 7 (meningiomas). For further explanation see Figure 6.1.

The four MR images were plotted together with a segmented image in which voxels were clustered using a model-based clustering algorithm (Fraley and Raftery, 1998) as described in Wehrens et al. (2002). The clustering provides an objective segmentation based on similarities obtained from the MR images as well as MR spectroscopic features, and thus is considered to be helpful in the selection of voxels. For each pathology, only voxels were included in the data set if their corresponding MR spectra were found to be typical for that pathology by an expert in MR spectroscopy. Since tumours are known to be heterogeneous, this approach was considered better than for example taking all spectra from one segment. Neighbouring voxels were not selected in order to avoid too much mutual correlation and as such to ensure that samples were as independent as possible. Although the method of class selection is subjective, we think it is appropriate since in tumour diagnosis a "ground truth" is not available and the number of patients for each specific class is low.

6.3.2 Experimental approach

Binary classification was performed by LDA, LS-SVM with linear and RBF kernel, which were also applied in two previous extensive studies for classification based on long (chapter 4) and short echo time ^{1}H MRS (chapter 5). The same experimental setup was used as mentioned in section 5.3.2, except that in this study another criterion was chosen to select the number of PCs, prior to LDA classification, which is motivated below.

In order to perform meaningful data-analysis in a high-dimensional space, a sufficiently large amount of training data is required. This limitation can be overcome e.g. by dimensionality reduction, which decreases the amount of complexity and risk of overfitting and also simplifies the calculation. In fact, peak integration (section 5.3.1) is a feature extraction method that reduces the input dimension and therefore uses prior knowledge about the most discriminatory features in the spectrum. As the input dimension is already small when using imaging intensities (dimension 4), peak integration values (dimension 10) or the combination of both (dimension 14), dimensionality reduction is only required when using magnitude spectra (dimension 231).

Especially LDA suffers from this dimensionality problem, while LS-SVM is able to work without any prior dimensionality reduction, thanks to the primal-dual aspects of the model, even in case of a relatively low number of input data. Therefore, the use of PCA as a feature extraction technique was only needed prior to LDA. The 231 given spectral variables were reduced by PCA that retained those components that account for a larger variance than the average over all individual components (Rencher, 1995). This strategy was different to the one taken in chapters 4 and 5, in which only the largest components were selected that explain a 75% of total variance. More specifically in the previous chapters it was unfeasible to retain more components due to rank deficiency problems related to a too small number of training samples. Nevertheless, if feasible, it is more appropriate to retain a certain number of components, as performed by the strategy taken in this study.

6.4 Results

We evaluated the following binary classifications of brain tissue types:

- Healthy vs. tumour tissue (classes 1,2 and 3; 318 data vs. 4,5,6 and 7; 351 data; Table 6.2). The resulting mean AUC was for all techniques and input patterns higher than 0.95. The performance based on imaging intensities alone was significantly lower with respect to using peak integration values (e.g. when using LDA, $p < 0.01$) or the combination of imaging intensities and peak integration values (e.g. when using LDA, $p < 0.001$). Also using magnitude spectra was significantly worse than using the combination of imaging intensities and peak integration values for LDA and LS-SVM with a linear kernel ($p < 0.05$). In addition, using solely imaging intensities, LDA and LS-SVM with a linear kernel reached a significantly lower performance than LS-SVM with an RBF kernel (both $p < 0.05$).
- Low- vs. high-grade tumours (classes 4 and 7; 224 data vs. 5 and 6; 127 data; Table 6.3). Classification based on imaging intensities alone was

poor, while discrimination based on peak integration values was significantly better, compared to classification using imaging intensities alone (with, $p < 10^{-8}$, $< 10^{-8}$ and < 0.001 respectively, in case of LDA, LS-SVM lin and LS-SVM RBF). Also the performances obtained when using magnitude spectra and the combination of imaging intensities and peak integration values were significantly higher than those obtained using imaging intensities alone. LS-SVM RBF achieved a significantly higher performance with respect to LDA and LS-SVM lin, when using imaging intensities ($p < 0.001$), peak integration ($p < 0.05$) and the combination of imaging intensities and peak integration values ($p < 0.001$). Based on magnitude spectra, LS-SVM RBF was significantly better than LDA using 9 PCs ($p < 0.05$). Based on imaging intensities, peak integration and their combination, LS-SVM RBF also achieved a significantly higher performance than LDA and LS-SVM with a linear kernel.

- Low- vs. high-grade gliomas (class 4; 176 data vs. 5 and 6; 127 data; Table 6.4). The classification problem resembles that of low- vs. high-grade tumours, but reached in general a slightly higher performance. This might be partially due to the lower heterogeneity in the two compared classes, while in low- vs. high-grade tumours also non-gliomas (namely meningiomas) were included. Once more, peak integration and the combination of imaging intensities and peak integration provided a significantly better performance than using imaging intensities alone. For example when comparing the use of peak integration with that of imaging intensities we obtained $p < 0.001$ (LDA, LS-SVM lin) and $p < 0.01$ (LS-SVM RBF) and the significance was even stronger for the combination of imaging intensities and peak integration values. For this problem, classification with LS-SVMs using magnitude spectra also reached a very high mean AUC, for which $p < 0.001$ (LDA), $p < 0.00001$ (LS-SVM lin) and < 0.01 (LS-SVM RBF) with respect to imaging intensities. From Figures 6.5 and 6.6 we remark that the imaging intensities of class 4 indeed highly resemble those of classes 5 and 6, which explains the low performance based on the imaging intensities. LS-SVMs achieved a significantly better result than LDA when using magnitude spectra, while LS-SVM RBF provided a significantly higher AUC than LDA and LS-SVM lin when using imaging intensities.

- Gliomas vs. meningiomas (classes 4, 5 and 6; 303 data vs. class 7; 48 data; Table 6.5). Note that the same data were used as in low- vs. high-grade tumours, but now gliomas were differentiated from meningiomas. At least a mean AUC of 0.99 was reached when based on peak integration or the combination of imaging intensities and peak integration (AUC > 0.85 when using imaging intensities alone). For LDA and LS-SVM lin, using peak integration, combined with imaging intensities or not, gave a significantly better result than using imaging intensities ($p < 0.001$ for all cases). However, no significant differences were found for LS-SVM RBF

with respect to the input type used.

- Grade II vs. grade III gliomas (class 4; 176 data vs. class 5; 57 data; Table 6.6). With respect to the significant influence of the input features, we observed the same as for low- vs. high-grade gliomas, except when using LS-SVM RBF. In case of LS-SVM RBF, only the combination of imaging intensities and peak integration values gave a significantly higher mean AUC than imaging intensities alone ($p < 0.05$). Linear classification techniques gave a significantly lower performance than LS-SVM RBF when using imaging intensities ($p < 0.05$).

Classifier	MRI	peak integration	MRI/peak integration	magnitude spectra
LDA	0.9569 ± 0.0128	0.9912 ± 0.0042	0.9991 ± 0.0009	$0.9828 \pm 0.0066_{(10;89.7\%)}$
LS-SVM lin	0.9570 ± 0.0128	0.9921 ± 0.0039	0.9991 ± 0.0008	0.9755 ± 0.0097
LS-SVM RBF	0.9858 ± 0.0065	0.9974 ± 0.0036	0.9998 ± 0.0003	0.9851 ± 0.0076

Table 6.2: Average test performance for classification of healthy vs. tumour tissue from 100 runs of stratified random splittings. *MRI*: imaging intensities; *peak integration*: peak integration values; *MRI/peak integration*: imaging intensities and peak integration values; *magnitude spectra*: water normalized magnitude spectra. Performance measure for each classifier is the mean AUC and its pooled standard error. When using magnitude spectra, PCA was applied prior to LDA. The number of principal components that was given as input to LDA and the amount of variance explained by those components is mentioned between brackets (row LDA, column magnitude spectra).

Classifier	MRI	peak integration	MRI/peak integration	magnitude spectra
LDA	0.6239 ± 0.0543	0.9210 ± 0.0243	0.9195 ± 0.0255	$0.9193 \pm 0.0265_{(9;92.8\%)}$
LS-SVM lin	0.6239 ± 0.0543	0.9328 ± 0.0220	0.9260 ± 0.0240	0.9573 ± 0.0189
LS-SVM RBF	0.8469 ± 0.0385	0.9827 ± 0.0124	0.9918 ± 0.0072	0.9797 ± 0.0161

Table 6.3: Classification of low- vs. high-grade tumours. For further explanation we refer to Table 6.2.

Classifier	MRI	peak integration	MRI/peak integration	magnitude spectra
LDA	0.7429 ± 0.0517	0.9452 ± 0.0223	0.9563 ± 0.0203	$0.9339 \pm 0.0240_{(9;93.4\%)}$
LS-SVM lin	0.7431 ± 0.0517	0.9453 ± 0.0222	0.9589 ± 0.0193	0.9809 ± 0.0114
LS-SVM RBF	0.8774 ± 0.0354	0.9774 ± 0.0158	0.9920 ± 0.0084	0.9896 ± 0.0081

Table 6.4: Classification of low- vs. high-grade gliomas. For further explanation we refer to Table 6.2.

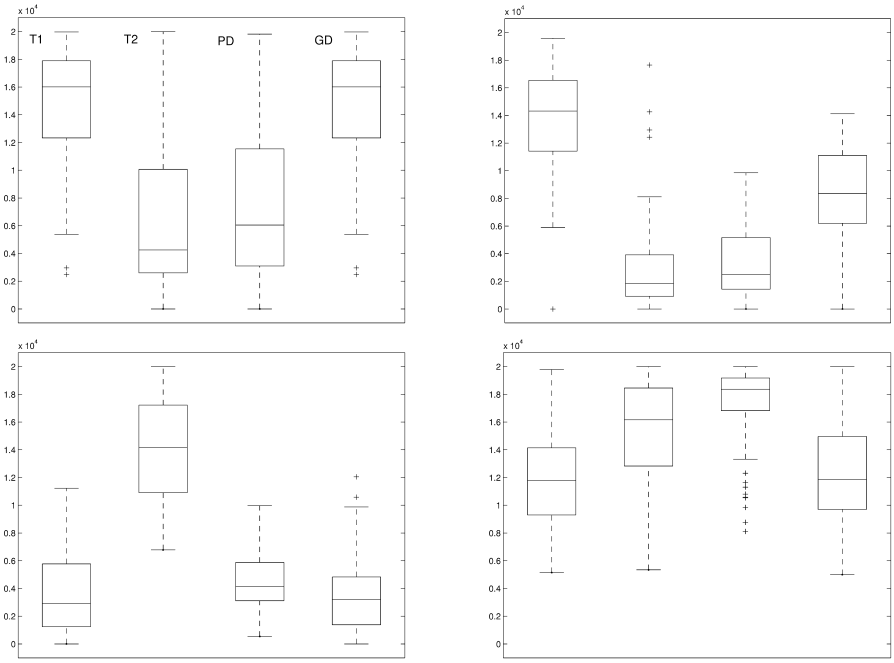

Figure 6.5: Boxplots of the imaging intensities of the considered classes: class 1 (top-left), class 2 (top-right), class 3 (bottom-left) and class 4 (bottom-right) correspond to the normal tissue of volunteers, normal tissue of patients, CSF and gliomas of grade II.

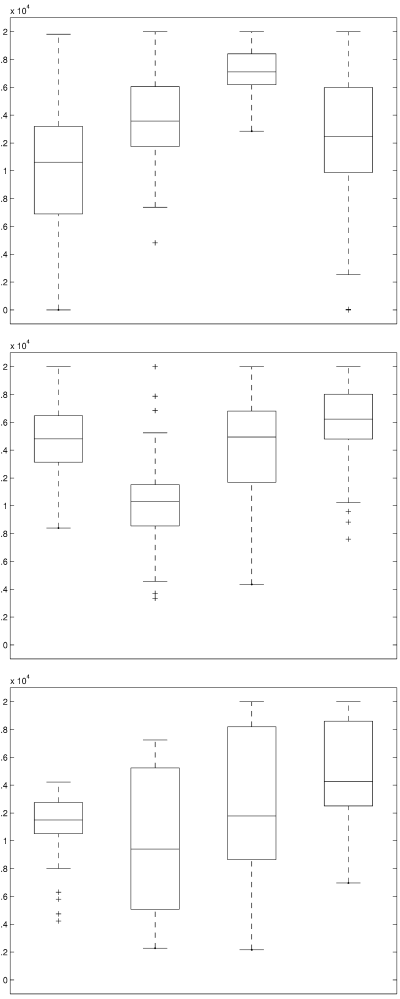

Figure 6.6: Boxplots of the imaging intensities of the considered classes: class 5 (top-left), class 6 (top-right) and class 7 (bottom-left) correspond to gliomas of grade III, grade IV and meningiomas.

Classifier	MRI	peak integration	MRI/peak integration	magnitude spectra
LDA	0.8593 ± 0.0406	0.9945 ± 0.0056	0.9961 ± 0.0040	$0.9890 \pm 0.0083_{(9;92.8\%)}$
LS-SVM lin	0.8590 ± 0.0407	0.9947 ± 0.0055	0.9964 ± 0.0039	0.9889 ± 0.0092
LS-SVM RBF	0.9520 ± 0.0303	0.9985 ± 0.0025	0.9989 ± 0.0019	0.9912 ± 0.0092

Table 6.5: Classification of gliomas vs. meningiomas. For further explanation we refer to Table 6.2.

Classifier	MRI	peak integration	MRI/peak integration	magnitude spectra
LDA	0.7799 ± 0.0689	0.9278 ± 0.0310	0.9480 ± 0.0313	$0.9339 \pm 0.0274_{(9;93.1\%)}$
LS-SVM lin	0.7799 ± 0.0688	0.9290 ± 0.0306	0.9488 ± 0.0308	0.9669 ± 0.0196
LS-SVM RBF	0.9018 ± 0.0410	0.9706 ± 0.0217	0.9907 ± 0.0098	0.9861 ± 0.0141

Table 6.6: Classification of grade II vs. grade III gliomas. For further explanation we refer to Table 6.2.

6.5 Discussion

6.5.1 Classification techniques

Tables 6.2 to 6.6 show that all classification techniques performed very well, especially based on the peak integration values and the combination with imaging intensities. In general, LDA is shown to be competitive with linear LS-SVM classifiers, even when based on the first principal components of the magnitude spectra. Using all PCs that explain more variance than the average, LDA based on magnitude spectra reached a similar performance as based on peak integration or the combination of imaging intensities and peak integration values.

Nevertheless, in several cases a significant difference was found, e.g. for the discrimination of low- and high-grade gliomas, based on magnitude spectra both kernel-based techniques reached a significantly higher performance than PCA/LDA. However, not only the classical linear LDA technique, but also the kernel-based linear LS-SVM did not always reach the performance of the non-linear LS-SVM. This occurred for the problem low- versus high-grade tumours based on peak integration and the combination of imaging intensities and peak integration.

For several problems, the unbalanced situation (e.g. gliomas vs. meningiomas) or the relatively small number of data available forms a limitation for training (e.g. grade II vs. grade III gliomas). Therefore, the discrimination boundary might strongly correlate with the training set. Especially LDA requires a significant amount of data to be able to draw a linear separating line between overlapping classes. Kernel-based techniques are less sensitive to the amount of

data and the input dimension and are able to detect automatically important characteristics independently of the input pattern. Hence, these techniques are able to obtain a high performance even without any prior dimensionality reduction, although dimensionality reduction may further improve the results.

6.5.2 Imaging intensities versus metabolic data from MRSI

Although MRI is an established technique for the characterization of brain tumours, it has a few limitations that produce uncertainty for an accurate assessment of the presence and extent of the tumour. In practice, the contrast-enhancing lesion on an MR image is often much smaller than the region of abnormal metabolism (Manton et al., 2000; Nelson et al., 1999). From our results we observed that, for some problems the use of imaging intensities alone reached a significantly lower performance, e.g. for the discrimination between low- and high-grade tumours. This is in correspondence to the conclusion of Manton et al. (2000), Nelson et al. (1999) and Vigneron et al. (2001), that the MR imaging intensities are unable to fully explain the metabolic heterogeneity of brain tumours.

Scaling of the MR data is necessary in order to correct for effects independent of the tissue characteristics. As a result of the scaling procedure, MR imaging intensities from different subjects or acquired under different conditions, should be more comparable. Although the applied scaling method might not be fully appropriate, the procedure yielded good classification results for healthy versus tumour tissue and gliomas versus meningiomas. Scaling with respect to healthy tissue might be an alternative method that possibly could improve results, but - in contrast to the applied method - would involve processing that is difficult to automate.

The obtained test performances based on metabolic data were in agreement with our previous studies on short and long echo time spectra reported in chapters 4 and 5. The results were also similar with respect to those of other MRS(I) studies (Herminghaus et al., 2003; Preul et al., 1996; Simonetti et al., 2003; Szabo De Edelenyi et al., 2000; Tate et al., 2003), although these authors used other performance measures than the test AUC. This confirms the statement that spectroscopy is able to add valuable information about the metabolic status of brain tumours, which is in agreement with a few clinical combined MRSI/MRI studies (Manton et al., 2000; Nelson et al., 1999; Vigneron et al., 2001). As such multivoxel MRS could be very helpful for the diagnosis of brain tumours in combination with conventional MRI.

However, automated discrimination between different tumour types is still difficult, partially due to the fact that only intensity values are used and no anatomical (e.g. the location and homogeneity of the tumour) or clinical information (e.g. complaints of the patient) is included in the input features. In

order to make an accurate diagnosis a neuroradiologist exploits such clinical features - in contradiction to most classification studies - as they could be very specific for certain types of tumour. For example, meningiomas are a type of tumour of the meninges and are not really a type of brain tumour, while gliomas are intracerebral tumours that start in glial cells. In an MRI classification study (Ye et al., 2002) also several other diagnostic factors were found to be important for prediction of brain glioma like age, oedema, blood supply, calcification and haemorrhage. Hence, adding important anatomical and clinical information as input features is expected to further improve automated diagnosis of brain tumours.

6.5.3 Several classification problems

If a classification technique is developed for diagnostic purposes, then the technique should be able to distinguish healthy tissue from tumour tissue. Several metabolic differences between normal cells and various tumour types are reflected in MR spectra: the *NAA* and *Cr* levels are lower in tumour spectra, while the *Cho* level is higher. The applied techniques were able to extract and exploit this differing pattern available in imaging intensities and MR spectra. Healthy brain and tumour tissue could be distinguished almost perfectly from each other.

Classification of low- versus high-grade gliomas reached a high mean AUC (> 0.93) using any input (except imaging intensities or magnitude spectra with a low number of PCs given to LDA). A similar observation could be made for gliomas versus meningiomas (AUC > 0.98), low- versus high-grade tumours (AUC > 0.91) and grade II versus grade III gliomas (AUC > 0.92).

Using the imaging intensities alone for discriminating gliomas of grade II and grade III, a linear classifier performed poorly. This might partially be due to the unbalanced distribution of the data (176 data of class 4 and 57 of class 5). A significantly higher performance was obtained by using peak integration or combining the imaging intensities and peak integration. Also for the discrimination between gliomas and meningiomas, the use of peak integration and the combination of peak integration and imaging intensities gave a significant improvement with respect to using imaging intensities alone. This was not the case when using LS-SVM with an RBF kernel, since the performance was already high.

6.6 Conclusions

This study investigated the use of several classification techniques for classification of brain tumours using MR imaging intensities and metabolic data from

MRSI. The nonlinear technique LS-SVM with an RBF kernel reached in several specific problems a significantly better performance than the linear techniques. Although linear classifiers also performed well, based on these data this indicates that a few diagnostic problems seem to have a nonlinear behaviour.

From current studies it is clear that ^{1}H MRSI is an important adjunct to the clinical imaging techniques for non-invasive diagnosis of viable tumours. For most problems, binary classification based on imaging intensities and metabolic information from MRSI is very well possible. The combined use of MR imaging intensities and metabolic information significantly increased the performance with respect to imaging intensities alone. Also with respect to metabolic data alone, the combination of imaging intensities and metabolic data reached a higher, although not significantly, performance. The results of this study strengthen the statement that imaging intensities and metabolic data provide complementary information for the accurate discrimination between several brain tissue types. Therefore, we motivate the integration of MRSI into a standard clinical examination which is performed for the diagnosis of brain tumours. However, in order to enhance the quality of automated diagnosis, it would be benificial that classification datasets also include several other relevant anatomical and clinical parameters. This would enable pattern recognition people to develop and test classifiers that use as much information as relevant.

Chapter 7

Model selection for quantification of multi-exponential MR spectra

When applied for fundamental research, it is important to process the MRS signal as reliably as possible based on numerical processing tools. MRS signals can be analyzed by a time-domain model to retrieve estimates of the model parameters. Most algorithms require prior knowledge about the model order, which is not the case for experimental signals in which peaks are largely overlapping. The problem can even be more difficult when signals are a superposition of peaks occurring at the same resonance frequency, but with different damping values. Such a signal is called a multi-exponential signal and is observed in ^{13}C glycogen signals. Specifically for these type of signals, several model selection criteria are developed, applicable to the quantification of MRS signals.

In this study four model selection criteria from section 2.4 are tested on Monte-Carlo simulation signals. The optimal criterion is further applied for selecting the model order of experimental glycogen signals. For each signal, several models are determined with a different model order. The models are obtained by spectral processing using AMARES described in section 2.3.

Section 7.1 outlines the different purposes of this study, while section 7.2 describes the background about the NMR experiment. Further, several concepts of the processing are mentioned, including the provided prior knowledge and model selection. This analysis is then applied to simulation signals as well as

experimental signals from the glycogen experiment. The simulation and experimental results are described in section 7.4 and discussed in section 7.5. This includes the adequacy of the model selection criteria, the convergence of the time domain algorithm, the retrieved optimal model order as well as implications on the time course of the glycogen intensity and linewidth as biochemical parameters. Results are also compared to a previous study by Bergans (2003). From the analysis it can be concluded that model selection criteria are suited for finding the optimal model order of glycogen signals. It is also found that several multi-exponential signals can be detected during a glycogen experiment. However, quantification of ^{13}C glycogen signals is not straightforward as convergence problems can arise at low SNR and research is further needed to increase the reliability of the analysis.

7.1 Introduction

Quantification of Magnetic Resonance Spectroscopic (MRS) signals can be done parametrically in the time domain by modelling the signal as a superposition of exponentially damped sinusoids. This supposes the model order to be known exactly, which in practice is not the case. The determination of the model order is particularly difficult in the case that some of the exponentially damped sinusoids have the same frequency. This problem is encountered in the analysis of ^{13}C MRS signals where glycogen (GLY) is known to be a superposition of an unknown number of exponentially damped sinusoids with the same frequency but with different damping factors; a so-called multi-exponential signal. In order to determine the model order, a model selection criterion is implemented that evaluates models with a different model order. This criterion would allow an objective evaluation of the model order rather than the operator-biased evaluation of the residue. Several criteria were evaluated on simulated signals to determine the optimal criterion for the experimental glycogen signals.

The glycogen signals were obtained during a so-called ^{13}C-1 pulse-chase experiment (section 7.2), which follows the glycogen synthesis in a perfused rat liver and mainly consists of two pulse phases and a chase phase. During the pulse phases is possible to estimate the rate of glycogen synthesis, while the rate of glycogen breakdown is followed in the chase phase. A detailed overview of the biochemical background of glycogen synthesis and the ^{13}C-1 NMR experiments designed to studying this process can be found in (Bergans, 2003).

Processing of these glycogen signals yielded that the glycogen signals apparently were multi-exponential signals, i.e. consisting of a sum of several peaks at the same resonance frequency, but with different dampings (Bergans et al., 2003; Overloop et al., 1996). During the experiment, an increase of the damping of the signal has been observed, due to either a gradual line broadening

or an increasing contribution of the broad components or both. Hence, reliable processing by modelling the glycogen signal as one single signal becomes difficult. From these observations, several purposes were formulated:

1. To implement and apply a method that determines the model order of glycogen signals in different situations; fed and starved rats, chase, first and second pulse phases.
2. To find out whether the observations from experiments can be confirmed or not. Therefore, results were compared with those of Bergans et al. (2003) and Bergans (2003).
3. To test which SNR is required in order to detect a multi-exponential signal.
4. To investigate whether certain soft constraints on the model can facilitate the reliable processing.
5. To assess the time course of certain model parameters, more specifically the amplitude and damping of the glycogen peaks. This should improve the insight in the biomedical processes during the glycogen synthesis in a perfused rat liver.

7.2 Material

7.2.1 Glycogen synthesis in perfused rat liver

Male Wistar-derived pfd rats (Louvain) of 200 to 300 g were used. The animals were observed using standard procedures (National Institutes of Health guideline, NIH 80-23) and the experiments were approved by the Ethical Committee for Animal Experiments of the KULeuven. Experiments were carried out on a perfused rat liver of either fed or starved rats, of which food was withdrawn for 24 hours prior to the experiment. On the morning of the study, 10-15 minutes prior to prelevation of the livers, rats were anaesthetized. Livers were perfused as described in Bergans (2003).

Glycogen synthesis from glucose in perfused livers of both fed and starved rats was studied by ^{13}C NMR. A so-called ^{13}C-1 pulse-chase experiment was carried out and followed by NMR. A ^{13}C-1 pulse-chase experiment involves the exposure to ^{13}C enriched glucose, noted as ^{13}C-1 glucose, followed by a chase phase in which glucose was replaced by ^{12}C-1 glucose (unlabelled). The experiment consisted of the following phases:

- In the first pulse phase (60 min) 99% enriched ^{13}C-1 glucose was given in recirculating mode. The natural abundance of ^{13}C NMR is only 1.1%

(98.9% is ^{12}C) hence ^{13}C MRS one can then detect the incorporation of the ^{13}C-label into other compounds, for example the conversion of ^{13}C-glucose molecules into labelled glycogen, a polymer of glucose units.

- During the wash-out phase (8 min), perfusion continued with a buffer containing unlabelled glucose in a single-pass mode to wash out the ^{13}C-labelled glucose. Unlabelled glycogen was formed. The ^{12}C-glucose subunits that are being added to the glycogen polymers do not result in an NMR signal.
- In the chase phase (60 min), perfusion continued in recirculating mode. Unlabelled glucose was given such that glycogen formation continued.
- During the second pulse phase (60 min), the first phase was repeated in order to verify the continuous glycogen synthesis during the complete experiment.

7.2.2 Data

Spectra were acquired and preprocessed as described in Bergans et al. (2003), on a 4.7T, 30 cm wide bore horizontal magnet equipped with a Biospec spectrometer (Bruker Spectrospin, Karlsruhe). Waltz broad-band ^{1}H-decoupled relaxed ^{13}C NMR spectra were acquired at 50.4 MHz in blocks of 512 scans using 90° pulses (pulse length 75 μs), a dwell time of 45 μs, TR=846 ms and a 270 μs dead time.

Experimental signals were available from fed and starved rats for the two pulse phases and the chase phase, at several noise levels. The SNR was increased by averaging the spectra from different livers (fed: 4 experiments, starved: 7 experiments) corresponding to the same time point in the experiment. To further increase the SNR, in certain cases a so-called sliding rule was applied in which three consecutive spectra within the sequence were averaged. The central time point was shifted (slided) over the consecutive spectra. The sliding rule was not applied at the transition between different phases of the experiment, in order not to mix signals from different phases.

7.3 Methods

Ideally, a method should be able to select the correct model order. But in practice most methods are not always able to find the number of components, as the signal can consist of a large number of peaks at the same frequency but with different dampings possibly out of a continuous set of dampings. This motivates to restrict the maximum model order to three and to consider only models with one, two and three peaks for the glycogen resonance frequency,

noted as M_1, M_2 and M_3, respectively. All models were obtained by time-domain modelling using AMARES (section 2.3).

7.3.1 Prior knowledge

Quantification of the signals was carried out in order to approximate the resonances from glycogen as well as from α- and β-glucose. The parameters of the different components are labelled as a_k (amplitude), d_k (damping), f_k (frequency) and ϕ_k (phase), where $k = 1, \ldots, i$ ($i = 1$ for M_1, $i = 2$ for M_2, or $i = 3$ for M_3) are the peaks of glycogen and $k = 4, 5$ are the peaks of β- and α-glucose. The following prior knowledge was used:

- The first 5 points of the FID were truncated to filter out ringing (dwell time is 45 μs).
- The model function was purely Lorentzian.
- The begin time was set to be -37.5 ms. The zero order phase was estimated by AMARES, but was constrained to be the same for all peaks.
- The amplitudes of the glycogen peak were unconstrained, for each M_i.
- The amplitude of β- and α-glucose were to be estimated.
- The dampings $d_4 = d_5$ of β- and α-glucose were set to be equal. For the chase phase soft constraints can be imposed on the dampings (linewidth between 46 en 52 Hz). The linewidth lw_k of a Lorentzian is related to the damping as $lw_k = d_k/\pi$.
- The frequency f_4 of β-glucose was constrained between 95.5 ppm and 97.5 ppm, while that of α-glucose (f_5) was imposed to be between 91.5 ppm and 93.5 ppm.

Depending on M_i, additional information was given:

- M_1 (glycogen as one single Lorentzian): the phase was set to be equal to that of the other peaks. The amplitude and damping were unconstrained, while the frequency was constrained between 100 and 102 ppm.
- M_2 (glycogen as a sum of 2 Lorentzians at the same frequency, but different dampings): the phases were set to be equal to that of the other peaks. The amplitudes and dampings were unconstrained, while frequencies were imposed to be equal and to be between 100 en 102 ppm.
- M_3 (glycogen as a sum of 3 Lorentzians at the same frequency, but different dampings): the phases were set to be equal to that of the other peaks. The amplitudes and dampings were unconstrained, while frequencies were imposed to be equal and to be between 100 en 102 ppm.

In general no additional constraints were used, unless otherwise mentioned. For several signals convergence problems occurred without imposing soft constraints on the dampings of the glycogen peaks, especially when computing M_3. When mentioned in section 7.4, we experimented with the use of the following soft constraints on the dampings, based on parameter values obtained from experimental data for which models converged without imposing soft constraints. For livers of fed and starved rats different soft constraints were imposed according to Table 7.1. For M_2, the effect of soft constraints was only tested for the experimental signals.

fed/non-fed	Lorentzian	M_2		M_3	
		b_l (Hz)	b_u (Hz)	b_l (Hz)	b_u (Hz)
non-fed	1	55	60	55	60
	2	365	490	85	145
	3			460	700
fed	1	57	62	57	62
	2	100	250	64	68
	3			135	250

Table 7.1: Soft constraints on the linewidths of Lorentzians of the GLY peaks of models M_2 and M_3, imposed in several tests on experimental signals. The second column mentions the number of the components $k = 1, \ldots, i$ for each model M_i, while b_l, b_u are the lower and upper bounds on the linewidths).

7.3.2 Model selection for MRS quantification

Four model selection criteria from information theory were applied to simulation signals in order to investigate which criterion would be most appropriate for the processed signals. This criterion was then used for model order selection of the experimental signals.

GPE, as explained in section 2.4, is a mathematically sound method based on information theory to determine the appropriate model order of a multi-exponential MRS signal. Using different λ-values, we considered GAIC, GBIC and GIClog, as well as the bias-corrected version of GAIC, called GAICC.

For each of the criteria, the best model was given as the one with the lowest criterion value. In case Equation (2.16) is applied to models obtained by AMARES, d_{eff} is equal to $4\,\mathrm{trace}(\mathbf{S}) + 2 - d_C$, where $\mathrm{trace}(\mathbf{S})$ is the trace of the hat matrix $\mathbf{S} = \hat{\mathbf{\Gamma}}\hat{\mathbf{\Gamma}}^{\dagger}$, which relates the model and the data as shown in Equation (2.11). The number $d_{C,i}, i = 1, 2, 3$ is the number of hard constraints (reflecting the prior knowledge about the model parameters) imposed for the model M_i prior to the computation of AMARES. The bias correction involved in GAICC is known to be important for a small number of data points N rel-

ative to d_{eff}. In our application d_{eff} was of the order of 10 to 15, depending on the model order, while each signal consisted of 512 data points.

7.3.3 Application on Monte-Carlo simulations

The four different criteria GAIC, GBIC, GAICC and GIClog were applied on simulated ^{13}C signals with resonances of glycogen and α- and β-glucose. Signals were simulated with one, two as well as three Lorentzians at the resonance frequency of glycogen, but with different dampings. Each of these simulation signals can be approximated as a model M_1, as well as a multi-exponential model M_2 and M_3. The amplitude a_4 of β-glucose was set to be 4/10 of the total amplitude of glycogen, unless mentioned otherwise. The amplitude of α-glucose was set to be $a_5 = 2/3a_4$, which corresponds to the equilibrium situation of D-glucose in solution.

GLY-peak: one Lorentzian

Two different situations were considered for the noiseless simulation signals containing only one glycogen peak:

1. $a_1 = 1 * 10^4, lw_1 = 105$Hz.
2. $a_1 = 1 * 10^4, lw_1 = 148$Hz.

GLY-peak: two Lorentzians

Assume a signal with 2 components at the same frequency. Intuitively, the model selection method is expected to have a higher performance when the difference between dampings is larger or the intensity of the broad component, relative to that of the small component, is larger. The influence of the difference between the dampings as well as the amplitude ratio was tested by considering four different combinations of dampings and amplitudes.

1. Large difference between the dampings ($lw_2 \gg lw_1$) and a relatively large amplitude a_2 ($a_2/a_1 = 1$): $lw_1 = 85$Hz, $lw_2 = 400$Hz; $a_1 = 0.5 * 10^4, a_2 = 0.5 * 10^4$.
2. $lw_2 \gg lw_1$, $a_2/a_1 \approx 0$: $lw_1 = 85$Hz, $lw_2 = 400$Hz; $a_1 = 0.9 * 10^4, a_2 = 0.1 * 10^4$.
3. $lw_2 > lw_1$, $a_2/a_1 = 1$: $lw_1 = 63$Hz, $lw_2 = 148$Hz; $a_1 = 0.5 * 10^4, a_2 = 0.5 * 10^4$.
4. $lw_2 > lw_1$, $a_2/a_1 \approx 0$: $lw_1 = 63$Hz, $lw_2 = 148$Hz; $a_1 = 0.9 * 10^4, a_2 = 0.1 * 10^4$.

5. $lw_2 \gg lw_1, a_2/a_1 = 1, a_4 \approx 0$: $lw_1 = 85$Hz, $lw_2 = 400$Hz; $a_1 = 0.5 * 10^4, a_2 = 0.5 * 10^4$ and $a_4 = 0.01 * 10^4$. This test was performed to investigate whether the presence of the glucose-peaks had an influence on the model order selection.

GLY-peak: three Lorentzians

Additionally, in order to investigate whether AMARES was able to detect three components of glycogen, the following two situations were constructed:

1. $lw_1 = 63$Hz, $lw_2 = 148$Hz, $lw_3 = 637$Hz; $a_1 = 0.4{*}10^4, a_2 = 0.3{*}10^4, a_3 = 0.3 * 10^4$.
2. $lw_1 = 63$Hz, $lw_2 = 148$Hz, $lw_3 = 637$Hz; $a_1 = 0.4{*}10^4, a_2 = 0.3{*}10^4, a_3 = 0.3 * 10^4, a_4 = 0$.

When quantifying these simulation signals using M_3, soft constraints were imposed on the dampings, similar as those mentioned in Table 7.1 for starved rats, to guarantee convergence of the algorithm.

Noise realizations

The following experimental setting was implemented in order to test the four criteria and to investigate the influence of the noise level:

- Twelve different SNRs (4, 6, 8, 12, 16, 24, 32, 48, 64, 80, 100 and 128) were considered. The SNR was determined after 15 Hz exponential filtering and was defined as 2.5 times the ratio of the height of the GLY-peak and the peak-to-peak height of the noise (Bergans et al., 2003).
- 500 Gaussian noise realizations were created for each SNR.
- Each signal was modelled by M_i, $i = 1, 2, 3$, the model estimated by AMARES, that approaches the GLY-peak as a superposition of i Lorentzians.
- The criteria were tested by comparing the estimated model order with the true model order. However, by considering 500 noise realizations this can be done more statistically based on the Wilcoxon signed rank test of equality of medians. This test returns the significance for a test of the null hypothesis that the median difference between two sample series is zero. M_1 is a statistically better model than M_2 if the median criterion value of M_1 is lower than that of M_2 under a significance level of 0.05. For each group of 50 signals, the Wilcoxon test was applied, resulting in 10 significance tests for each SNR.

7.3.4 Application on experimental signals

From the results of the Monte-Carlo simulations, the optimal criterion was selected, in order to apply to the experimental glycogen signals. Quantification was carried out, using M_1, M_2 and M_3, on signals of perfusion livers of starved and fed rats, at different noise levels. Results were compared with those obtained by Bergans et al. (2003). The effect of imposing soft constraints on the dampings of the GLY-peaks was also investigated.

In Bergans (2003) the noise levels were computed for the different signals. For example in the first pulse phase, signals from one single experiment had an SNR between 5 and 13 (starved rats) and between 20 and 30 (fed rats). By averaging the spectra from different livers (note that Bergans (2003) used the summed spectra instead), the SNR increased to 40 and 60 for starved and fed rats. For the sliding rule averaging, the SNR still increased with a factor of $\sqrt{3}$.

The estimates of the glycogen intensity of different models were compared using the parametric one-tailed t-test (MATLAB, The Mathworks). A significance level of 0.05 was used.

SNR	GAIC			GBIC			GAICC			GIClog		
4	5	2	3	10	0	0	4	3	3	2	2	6
6	0	0	10	10	0	0	0	1	9	0	0	10
8	2	8	0	10	0	0	2	7	1	1	9	0
12	0	10	0	1	0	9	1	9	0	0	10	0
16	0	10	0	1	9	0	0	10	0	0	10	0
$\geq$24	0	10	0	0	10	0	0	10	0	0	10	0

Table 7.2: Model M_1 versus M_2 for simulation signals as described in 7.3.3, two Lorentzians, item 1. Wilcoxon signed rank test results for the values of GAIC, GBIC, GAICC and GIClog. Each criterion corresponds to three columns containing from left to right, for each SNR, the number of times M_1 was (incorrectly) chosen as the best model, the number of times M_2 was (correctly) chosen as the best model and the number of times no significant difference could be found.

7.4 Results

7.4.1 Monte-Carlo simulations

GLY-peak: one Lorentzian

1. $a_1 = 1 * 10^4, lw_1 = 105$Hz. All criteria selected the correct model order (M_1) in more than 90% of the signals. For each SNR, GBIC selected M_1 (100% correct), while GAICC gave M_1 the lowest value for each SNR, except for SNR=6, for which M_1 was the best model for 9 out of 10 signals. GAIC and GIClog both selected M_1 as the best model for SNR= 8, 12, 16 and SNR$\geq$64, while for all other SNR values 90% correctness was reached, which was slightly worse than the performance of GAICC and GBIC.

2. $a_1 = 1 * 10^4, lw_1 = 148$Hz. M_1 has the lowest GBIC value for each SNR. And without going into details, GAIC, GAICC and GIClog also consider M_1 as the best model for most SNR values.

No significant results with respect to model M_3 were shown, as no soft constraints were imposed on the dampings, which could be needed to guarantee convergence of M_3. However, for these signals we did not focus on M_3.

SNR	GAIC			GBIC			GAICC			GIClog		
4	9	1	0	10	0	0	9	1	0	8	2	0
6	9	1	0	10	0	0	9	1	0	8	2	0
8	10	0	0	10	0	0	10	0	0	9	1	0
12	8	2	0	10	0	0	9	1	0	8	2	0
16	9	1	0	10	0	0	10	0	0	9	1	0
24	9	1	0	10	0	0	8	2	0	9	1	0
32	5	4	1	9	1	0	7	2	1	3	2	5
48	0	1	9	10	0	0	0	0	10	3	1	6
64	0	10	0	9	1	0	0	10	0	0	10	0
80	0	10	0	4	2	4	0	10	0	0	10	0
100	0	10	0	1	4	5	0	10	0	0	10	0
128	0	10	0	0	10	0	0	10	0	0	10	0

Table 7.3: Model M_1 versus M_2 for simulation signals as described in 7.3.3, two Lorentzians, item 2. We refer to Table 7.2 for further explanation.

GLY-peak: two Lorentzians

1. $lw_2 \gg lw_1$, $a_2/a_1 = 1$ (Table 7.2 and Figure 7.1) . For SNR$\geq$12, all criteria gave an almost perfect result.

SNR	GAIC			GBIC			GAICC			GIClog		
4	8	2	0	10	0	0	9	1	0	8	2	0
6	9	1	0	10	0	0	8	2	0	8	2	0
8	7	1	2	10	0	0	7	2	1	5	2	3
12	0	0	10	10	0	0	1	0	9	0	0	10
16	2	6	2	10	0	0	1	6	3	1	8	1
24	0	10	0	4	0	6	1	9	0	0	10	0
32	0	10	0	2	8	0	0	10	0	0	10	0
$\geq$48	0	10	0	0	10	0	0	10	0	0	10	0

Table 7.4: Model M_1 versus M_2 for simulation signals as described in 7.3.3, two Lorentzians, item 3. We refer to Table 7.2 for further explanation.

SNR	GAIC			GBIC			GAICC			GIClog		
4	9	1	0	10	0	0	9	1	0	9	1	0
6	10	0	0	10	0	0	10	10	0	9	1	0
8	10	0	0	10	0	0	10	0	0	10	0	0
12	8	2	0	10	0	0	9	1	0	8	2	0
16	9	1	0	10	0	0	10	0	0	9	1	0
24	9	1	0	10	0	0	9	1	0	9	1	0
32	6	4	0	10	0	0	8	2	0	6	4	0
48	5	1	4	10	0	0	6	1	3	2	2	6
64	1	1	8	10	0	0	0	2	8	1	4	5
80	1	9	0	10	0	0	0	9	1	0	10	0
100	1	9	0	8	2	0	1	9	0	1	9	0
128	0	10	0	0	1	9	0	10	0	0	10	0

Table 7.5: Model M_1 versus M_2 for simulation signals as described in 7.3.3, two Lorentzians, item 4. We refer to Table 7.2 for further explanation.

2. $lw_2 \gg lw_1$, $a_2/a_1 \approx 0$ (Table 7.3). All criteria, except GBIC, resulted in the correct model order for SNR$\geq$64.

3. $lw_2 > lw_1$, $a_2/a_1 = 1$ (Table 7.4). All criteria resulted in the correct model order for SNR$\geq$48; GAIC and GIClog even gave perfect results for SNR$\geq$24.

4. $lw_2 > lw_1$, $a_2/a_1 \approx 0$ (Table 7.5). GBIC failed to select the correct model order, while all other criteria gave a good result for SNR$\geq$80.

5. $lw_2 \gg lw_1$, $a_2/a_1 \approx 0, a_4 \approx 0$. Although the correct model order was already selected at a slightly lower SNR compared to item 2, the glucose-peaks do not seem to have a large impact.

SNR	GAIC			GBIC			GAICC			GIClog		
≤32	10	0	0	10	0	0	10	0	0	10	0	0
48	8	2	0	10	0	0	9	1	0	8	2	0
64	0	0	10	10	0	0	0	0	10	0	2	8
80	4	5	1	9	1	0	1	6	3	3	6	1
100	1	9	0	8	1	1	1	9	0	1	9	0
128	0	10	0	0	4	6	0	10	0	0	10	0

Table 7.6: Model M_2 versus M_3 for simulation signals as described in 7.3.3, three Lorentzians, item 1. Wilcoxon signed rank test results for the values of GAIC, GBIC, GAICC and GIClog. Each criterion corresponds to three columns containing from left to right, for each SNR value, the number of times M_2 was (incorrectly) chosen as the best model, the number of times M_3 was (correctly) chosen as the best model and the number of times no significant difference could be found.

For the results of M_3 similar remarks can be made as in the previous paragraph.

SNR	GAIC			GBIC			GAICC			GIClog		
≤32	10	0	0	10	0	0	10	0	0	10	0	0
48	7	2	1	10	0	0	8	2	0	7	2	1
64	1	2	7	10	0	0	0	2	8	1	2	7
80	5	4	1	10	0	0	2	7	1	4	6	0
100	0	10	0	2	0	8	0	10	0	0	10	0
128	0	10	0	0	10	0	0	10	0	0	10	0

Table 7.7: Model M_2 versus M_3 for simulation signals as described in 7.3.3, two Lorentzians, item 2. We refer to Table 7.6 for further explanation.

GLY-peak: three Lorentzians

1. $lw_1 = 63\text{Hz}, lw_2 = 148\text{Hz}, lw_3 = 637\text{Hz}; a_1 = 0.4{*}10^4, a_2 = 0.3{*}10^4, a_3 = 0.3 * 10^4$ (Table 7.6). GBIC once more had most problems to select the correct model order. For the other criteria also a high SNR≥100 was needed to obtain 90% correct results.

2. $lw_1 = 63\text{Hz}$, $lw_2 = 148\text{Hz}$, $lw_3 = 637\text{Hz}$; $a_1 = 0.4 * 10^4, a_2 = 0.3 * 10^4, a_3 = 0.3{*}10^4, a_4 = 0$ (Table 7.7). In comparison to the previous item, similar results were obtained. Nevertheless, now also GBIC obtained 100% correctness at SNR=128, and also the performance of the other criteria increased slightly.

From the simulation results obtained, we observed that GAICC was clearly less

biased than GAIC and GIClog and overall seemed to be the optimal criterion out of the considered ones for this problem. For the analysis on experimental multi-exponential glycogen signals, we decided to use GAICC for model order selection.

7.4.2 Experimental signals

Convergence of the algorithm

For various noise levels the convergence behaviour of the different models obtained by AMARES was investigated. Glycogen could always be modelled with M_1, so as a single Lorentzian, independent of the SNR. When modelling with M_2, convergence behaviour depended on the SNR and the experimental phase. For all (8 out of 8) available signals of the chase phase (starved state), convergence was reached for any noise level when using M_2. AMARES also converged when using model M_2 for all signals of the second pulse phase (starved state) (Figure 7.2). Only at low SNR (1 experiment), no model M_2 was obtained for 2 out of 8 signals. Even for model M_3 convergence was reached for a few signals of the second pulse phase at high SNR (7 experiments: 5 out of 8 signals, sliding rule: 5 out of 6 signals) (Figure 7.2), while in the chase phase only for 2 out of 6 signals a model M_3 was obtained at a high SNR (sliding rule). A similar convergence behaviour was observed for signals of fed rats, although when modelling with M_3 the algorithm did not reach convergence for any signal of the fed state. Convergence of models M_2 and M_3 was not easily reached for signals of the first pulse phase (starved and fed state).

In order to reach convergence in these cases, the use of soft constraints on the dampings was investigated as specified in Table 7.1. This facilitated convergence of M_2 and M_3 for the signals of the chase phase and second pulse phase (fed and starved state). However, the resulting damping estimates were systematically lying close to one of the borders of the soft constraints, which indicated that the soft constraints were not optimized yet.

Optimal model order

All results mentioned in this paragraph were based on the analysis without imposing soft constraints on the dampings. Analysis of the high SNR signals (sliding rule) in the starved state yielded that GAICC selected M_1 as the optimal model for the first pulse phase. For this phase, models M_2 and M_3 did not converge well (either two or three of the glycogen peaks had the same damping or one or more of the glycogen peaks had zero amplitude).

At high SNR (sliding rule), model selection of the signals in the starved state resulted in M_2 as the best model for all time points of the chase phase. Even at lower SNR (6 experiments) a multi-exponential signal was detected as M_2

was the optimal model for all 8 signals. Even at low SNR (1 experiment) M_2 had the lowest criterion value for 5 out of 8 signals, more specifically at the first four and the last time point. This analysis for signals of 1 experiment was carried out only for one starved rat.

Although for high SNR signals of the second pulse phase, M_2 was selected as optimal model for the last 5 time points (sliding rule), the analysis with M_3 also resulted in a better model than M_1 for the last 4 time points. Even at lower SNR (7 experiments) multi-exponential signals were detected, as M_2 had the lowest GAICC value for 6 out of 8 signals and M_3 had a lower GAICC value for one signal.

Also in the fed state, signals of the first pulse phase (sliding rule) were best modelled by only one Lorentzian. However, for high SNR (sliding rule) signals from the chase phase and second pulse phase, GAICC selected M_2 as the optimal model. Two components were found for the last 5 (out of 6) time points of the chase phase and for all 6 (out of 6) time points of second pulse phase.

Time course of the signal intensity

The resulting time course of the amplitude and linewidth estimates for the glycogen signal (sliding rule) is given in Figures 7.3 (starved rats) and Figure 7.4 (fed rats). Results were only shown for signals that converged to an appropriate model, e.g. results of M_3 were discarded when a model with only two glycogen peaks with non-zero amplitude was obtained. In the chase phase, the glycogen intensity of M_2 clearly was higher than that of M_1 and significant for the chase phase ($p < 10^{-6}$). This effect was less expressed for the second pulse phase, for which the glycogen intensities of M_2 and M_3 were not significantly different.

When using M_2, the time course of signal intensity was discontinuous at the transition between the chase and second pulse phase. The estimated signal intensity was lower in the second pulse phase than at the end of the chase phase. This tendency also seemed to occur for the glycogen estimates of M_3, although this could be due to numerical aspects as model M_3 resulted in an appropriate model for only a restricted number of time points. More specifically, a convergent model M_3 was not obtained for the last 4 time points of the chase phase and the first time point of the second pulse phase.

During both pulse phases, the signal intensity was linearly increasing. This was valid for starved (Figure 7.3) as well as for fed (Figure 7.4) rats. The signal intensity estimated by M_1 clearly declined during the chase phase and this decreasing rate was similar for starved and fed rats. When using M_2 this rate was less expressed and not monotonous.

The time course of the resulting estimates was also investigated for the chase phase (starved rats) at different noise levels (Figure 7.5). At lower SNR, the

time course clearly showed a larger variation in the estimate. However, also at medium (6 experiments) and even low SNR (1 experiment), the signal intensity of model M_1 decreased linearly.

Time course of the linewidth

In the first pulse phase, signals were difficult to model with multiple Lorentzians (M_2 and M_3) without additional constraints on the linewidth. Even at high SNR (sliding rule) no appropriate fit was obtained for the starved as well as the fed state. The average estimates of the linewidths as plotted in Figures 7.3 and 7.4 were tabulated in Table 7.8.

For signals in the starved state, there were a few clear differences between the linewidths of different peaks. In the chase phase, M_2 resulted in two Lorentzians, of which the first one had a smaller linewidth than the Lorentzian of M_1, while the second one was 4 to 5 times broader. The same observation was made for the linewidths of the first two Lorentzians of M_3, while an additional Lorentzian had a slightly higher linewidth than the second one.

For signals in the second pulse phase, a similar pattern was observed, although there was a stronger difference between the second and third peak of M_3. The estimates of the linewidths of the first two components of M_2 and M_3 were lower than their estimates in the chase phase. Especially, the second component of M_2 and M_3 was much smaller than in the chase phase.

In the fed state, differences between the small and the broad peaks of M_2 were less expressed. However, we observed that the peaks were smaller than those in the starved state.

In the chase phase, the linewidth of the second component of M_2 showed a slightly increasing time course for the signals in the starved as well as in the fed state.

7.5 Discussion

7.5.1 Evaluation of the model selection criteria

The Monte-Carlo simulation results confirmed that all tested criteria gave similar results, which could be expected as they all correspond to the same general criterion form. GBIC was more in favor of a lower model order and had the worst performance when analyzing multi-exponential signals. GAIC and GIClog selected more often a higher model order. GAICC had similar results as GAIC, although it was slightly more in favor of a lower model order and as such performed better when analyzing signals with only one glycogen peak, while

fed/starved	model	k	linewidth (Hz)		
			1st pulse phase	chase phase	2nd pulse phase
starved	M_1	1	70.6 ± 1.2	107.0 ± 3.6	79.7 ± 1.0
	M_2	1		84.4 ± 3.6	67.7 ± 1.8
		2		397.9 ± 58.5	219.7 ± 45.8
	M_3	1		79.6 ± 1.0	56.9 ± 5.6
		2		301.4 ± 86.6	92.8 ± 7.2
		3		363.9 ± 1.8	369.5 ± 54.4
fed	M_1	1	53.3 ± 5.0	72.4 ± 3.0	70.7 ± 1.5
	M_2	1		60.8 ± 2.3	60.5 ± 1.8
		2		126.2 ± 24.8	183.7 ± 13.8

Table 7.8: Averages and their standard deviations of the estimates of the linewidth using models M_1, M_2 and M_3 for all signals for which an appropriate fit was obtained. The first three columns denote the physiological state of the rat, the model used and the number of each of the glycogen peaks in the model, respectively. The last three columns mention the estimated linewidths for each of the phases.

multi-exponential signals were also easily selected. This motivated the decision to apply GAICC for model order selection of glycogen signals, for which a multi-exponential pattern had already been observed (Overloop et al., 1996).

The higher the SNR of the simulation signals, the stronger the tendency towards the correct model order. In a favorable condition of the parameter values of a multi-exponential signal -e.g. components that clearly differ in damping and/or with an equally large intensity value- GAICC as well as GAIC and GIClog were able to select the correct model order even at low SNR values. For our simulation signals, these three criteria only required an SNR of about 16 dB (equal intensities) or about 12 dB (equal intensities and large difference in damping). For signals with components having only a large difference in damping, higher SNR values were required. For a restricted set of simulation signals consisting of three glycogen peaks, the algorithm AMARES needed an SNR of 100 in order to detect all three components.

7.5.2 Model selection applied to quantification of ^{13}C glycogen signals

The criterion GAICC was able to confirm most of the results and observations of previous studies on ^{13}C glycogen signals (Overloop et al., 1996; Bergans, 2003). During the chase phase of the experiment (starved state), the glycogen signal was best modelled as a superposition of a narrow and a broad component. For a few signals of the second pulse phase, the residue (first term of the criterion)

of a tri-exponential model was even slightly lower than that of a bi-exponential one, but the complexity term caused the criterion to select the bi-exponential model. The time course of the glycogen intensity revealed an abrupt decrease of the intensity estimated by a bi-exponential model at the transition between the chase and second pulse phase, which might imply that such model still underestimates the total glycogen intensity. This strengthens the indication that the glycogen signal of the second pulse phase also consists of at least three components. Also for signals of the chase phase and the second pulse phase in the fed state, a bi-exponential pattern was detected. For signals of the first pulse phase (starved and fed state) no clear indication was found for the presence of a multi-exponential glycogen signal.

These results were in agreement with Bergans et al. (2003), except that Bergans et al. (2003) obtained convergence of M_2 and M_3 for a few more signals. In this study using additional soft constraints on the dampings in general enabled convergence of the algorithm, but often resulted in estimates of the linewidth at the border of the soft constraints interval. This indicates that the use of soft constraints still needs to be investigated.

The higher the noise level, the more difficult it would be to detect a multi-exponential signal. However, for several signals of the chase phase (starved state) at medium (SNR=30-40) and low SNR (SNR=13) values, the algorithm was able to detect a bi-exponential model. The obtained model for the glycogen signal had a narrow and a broad component with an amplitude ratio much larger than zero, which relatively corresponded to a favorable situation, already considered in the simulation signals. This indeed yielded that SNR=12-16 was the minimal SNR value in order to detect a bi-exponential signal, which was in agreement with the analysis of the low SNR signals.

The overall time course of the total glycogen intensity and linewidth corresponded to the results of Bergans (2003). However, small differences were noted, especially for the linewidth of the multi-exponential signals in the chase phase. Therefore, experiments with the use of some well constructed (soft or hard) constraints might further clarify these different results.

7.6 Conclusions

Several model selection criteria from information theory were applied for selecting the correct model order by taking the model with the lowest criterion value. Models were obtained with three different model orders by the quantification algorithm AMARES. Simulation results indicated that information theoretic criteria were suited for model order selection of multi-exponential MRS signals. Results depended on the SNR and the differences in damping and intensity of the multiple glycogen peaks; especially at medium to high

SNR or in favorable situations of glycogen peaks with equal intensities and large damping differences, the correct model order was found. However these favorable situations also occur in experimental signals.

GAICC as optimal criterion for simulations was applied to experimental glycogen signals. In general, the conclusions of Bergans (2003) were strengthened by the implementation of a model selection method. For several signals, e.g. at low SNR, convergence of AMARES was not guaranteed for bi- and tri-exponential models. This corresponded to Bergans (2003), although the convergence performance was worse for tri-exponential models, for which additional constraints might facilitate convergence. For signals with convergent models, the analysis with multi-exponential models is needed to approximate glycogen signals during a pulse-chase experiments and to capture the time course of the parameter estimates as accurately as possible.

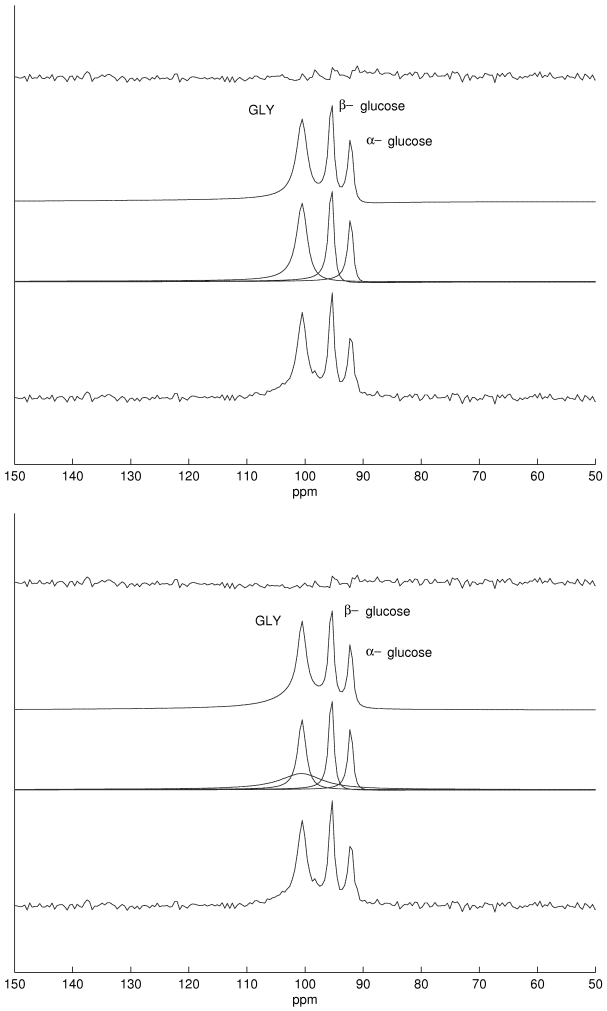

Figure 7.1: Quantification result of AMARES on a simulated glycogen ^{13}C NMR signal at SNR=24. The signal contained two glycogen peaks and a β- and α-glucose peak and was best modelled by M_2. The glycogen peaks had equal amplitudes and largely differing linewidths. Each figure contains respectively (from bottom to top) the original signal, the different model components, the model and the residue.

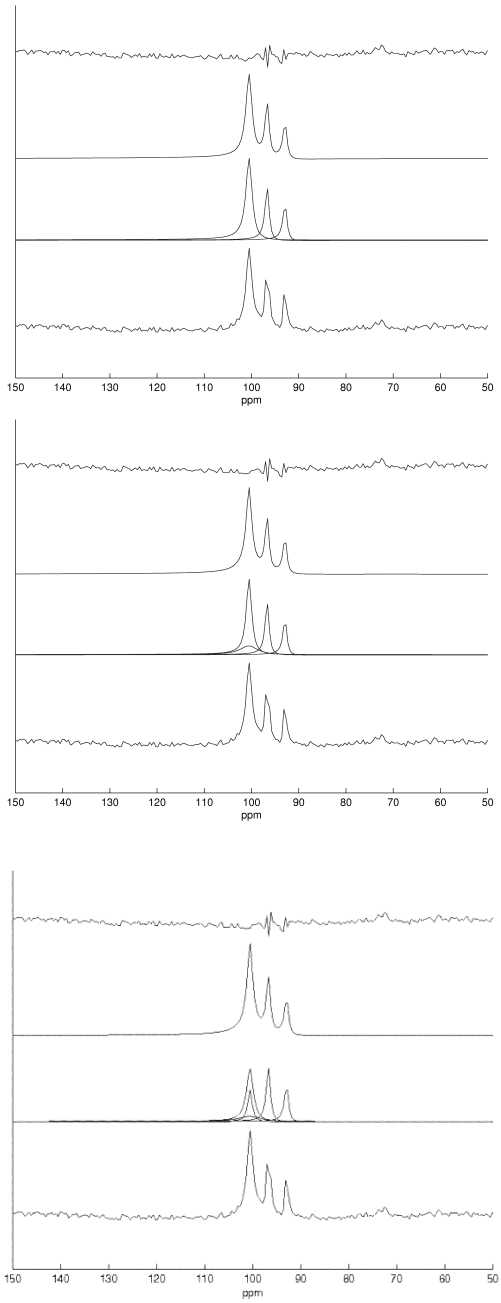

Figure 7.2: Resulting models of a signal of the second pulse phase at high SNR (sliding rule) from starved rats. From top to bottom models M_1, M_2 and M_3 are shown. The signal was best modelled by M_2. Each figure contains (from bottom to top) the original signal, the different model components, the model and the residue respectively.

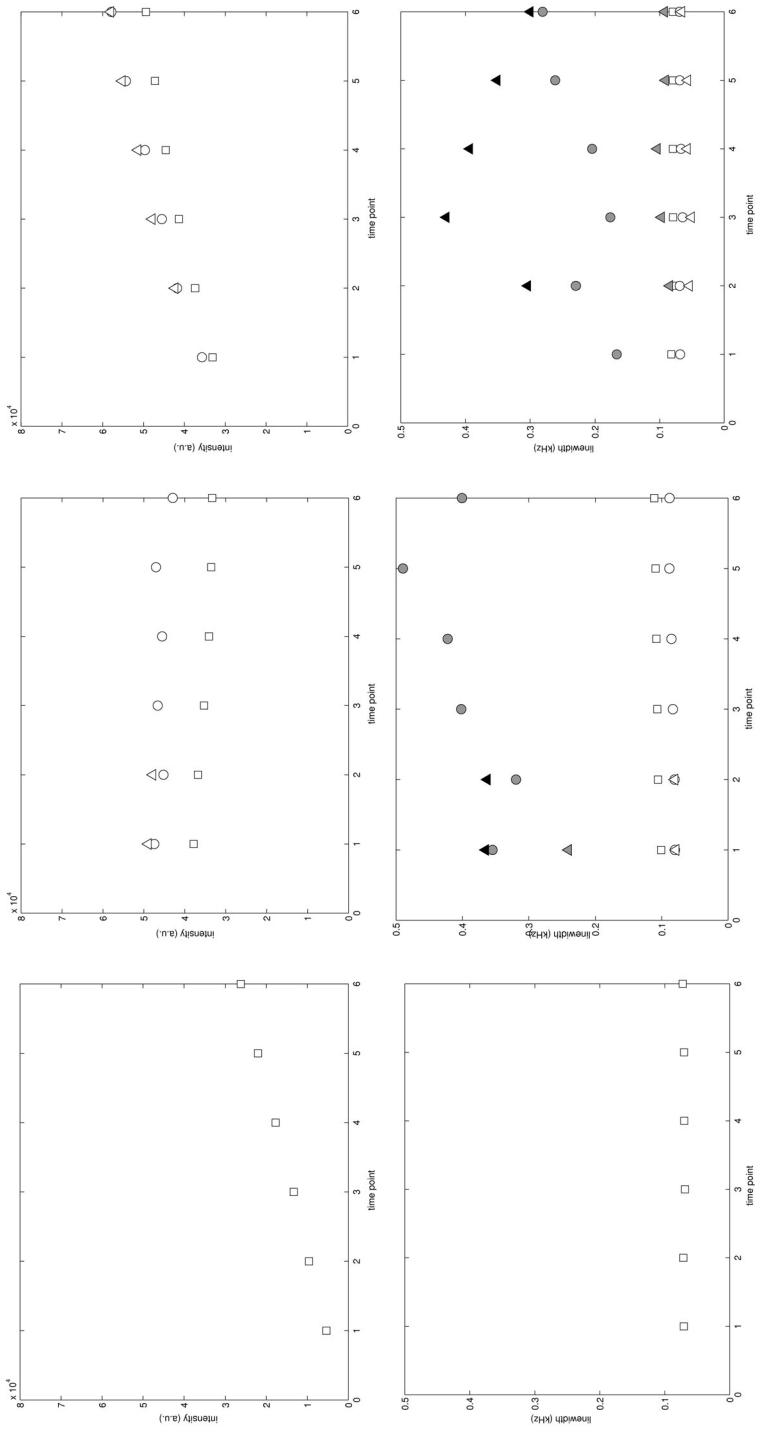

Figure 7.3: Estimates of the total glycogen intensity (top) and linewidths (bottom) of all glycogen peaks according to the three models M_1 ('□'), M_2 ('○'), M_3 ('△') for the first pulse phase (left), chase phase (middle) and second pulse phase (right) signals of starved rats based on a sliding rule. The total signal intensity (top) is represented by the empty symbols. For each model M_i, i symbols are needed to represent the linewidths of the i components; the empty $(M_i, i = 1, 2, 3)$, gray $(M_i, i = 2, 3)$ and black $(M_i, i = 3)$ symbols respectively correspond to the first, second and third components. The estimates were only displayed when an appropriate fit was obtained.

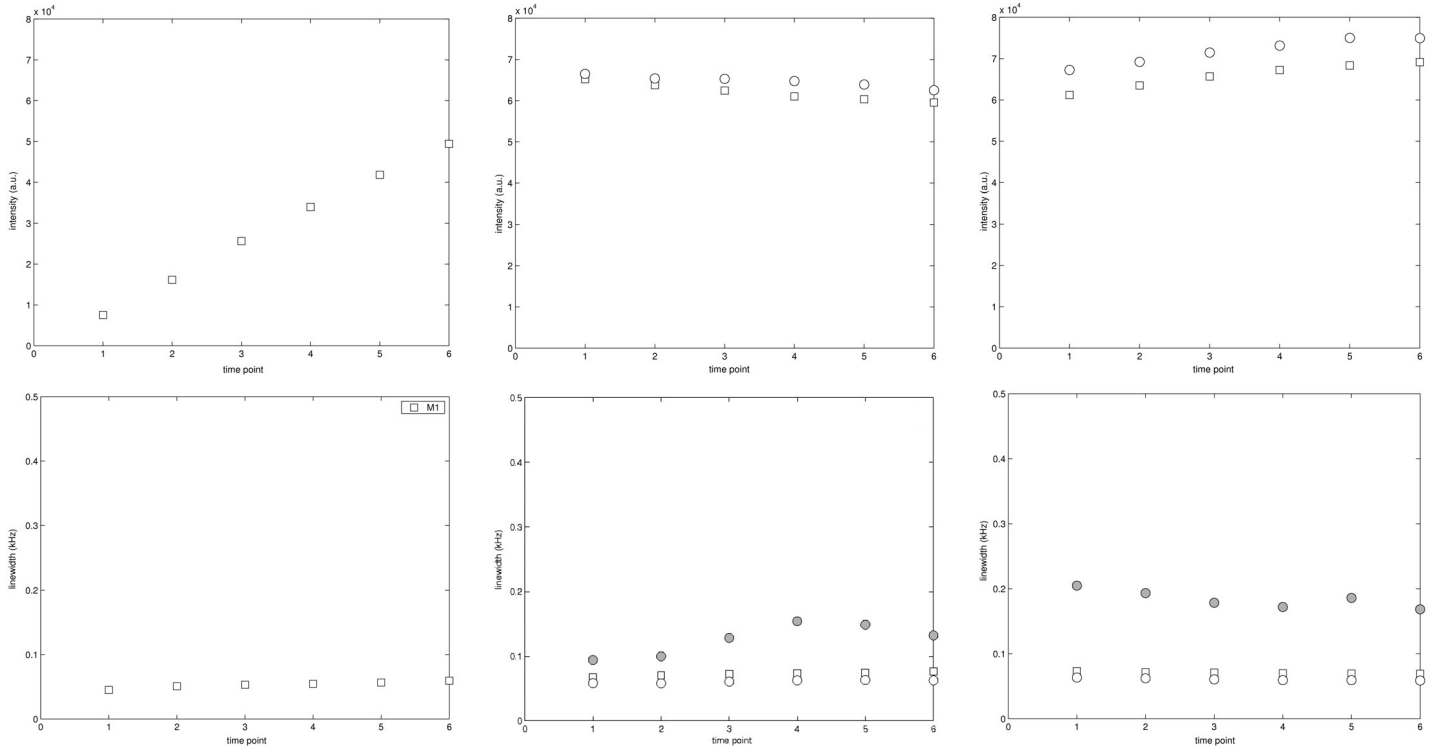

Figure 7.4: Estimates of the total glycogen intensity (top) and linewidths (bottom) of all glycogen peaks according to the three models M_1 ('□'), M_2 ('◯'), M_3 ('△') for all first pulse phase (left), chase phase (middle) and second pulse phase (right) signals of fed rats based on a sliding rule. For further explanation we refer to Figure 7.3.

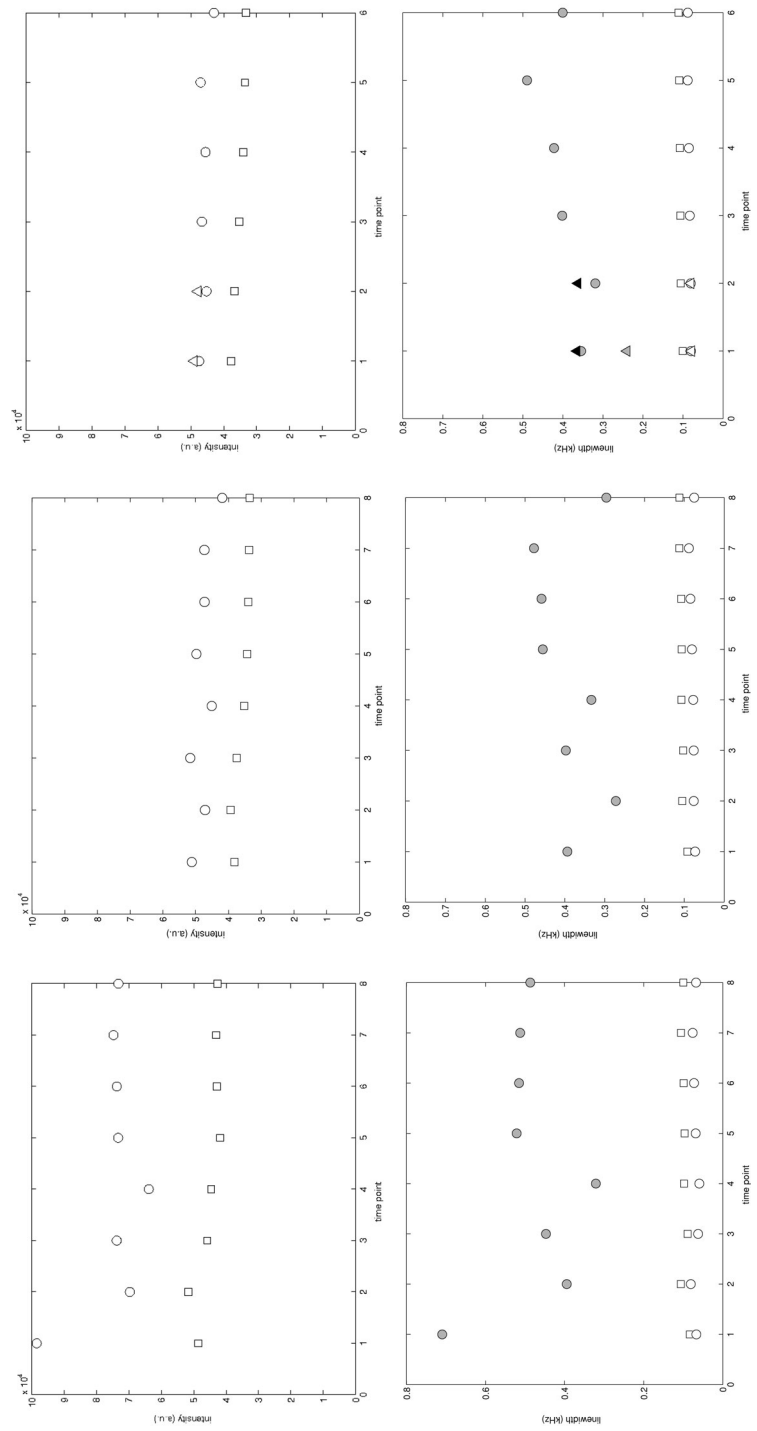

Figure 7.5: Influence of the noise level on model order selection. Estimates of the total glycogen intensity (top) and linewidths (bottom) of all glycogen peaks according to the three models M_1 ('□'), M_2 ('○'), M_3 ('△') for signals of the chase phase of starved rats. Results are shown for signals at low SNR (left), medium SNR (middle) and high SNR (right). For further explanation we refer to Figure 7.3.

Chapter 8

Conclusions and suggestions for future research

At the end of the journey through this thesis, it is time to look back at what has been achieved and to look ahead at what can be done further on. First, an outline of the major contributions is given for classification of brain tumours and MRS quantification. Next, suggestions for future research are pointed out in the second part of this chapter.

8.1 Summary

A short overview of the main conclusions and contributions is given, mentioning what has been done for each of the purposes in section 1.3. For classification this concerns the following aspects:

- Preprocessing steps. For short echo time signals, the influence of several preprocessing steps on the classification performance was investigated. This included phasing (magnitude or real spectra), removal of the underlying baseline (baseline correction) and spectral normalization (normalization with respect to the quadratic norm or normalization by the intensity of water). Classification results using magnitude or baseline corrected spectra yielded no significant effects on the performance with respect to real or non-baseline corrected spectra. Differences were observed for normalization, as the overall performance of water normal-

ization was worse. If generally true, the difficult tasks of phasing, baseline correction and spectral water normalization could be ignored for the purpose of brain tumour diagnosis. This makes the acquisition of a water unsuppressed signal not essential anymore and MRS acquisition protocols can be simplified.

- Dimensionality reduction. The approach to use extracted features such as selected frequency regions or peak integration values possibly reduces the influence of disturbing noise and artefacts in the spectra. For short echo time signals (chapter 5), peak integration enables to reach a similar performance as using complete spectra. Although peak integration is a simple and inaccurate quantification method, it yields a relatively high classification performance. This indicates that the selected regions are most informative to differentiate brain tumours and classification can be done on a limited amount of intensity variables. This conclusion was in agreement with results obtained on short echo time MRSI (chapter 6), but in disagreement with several results on long echo time MRS (chapter 4) although implementing the same experimental setting. A possible explanation is that spectra at short echo time contain much more information which is spread over a larger number of resonances. This enables to use more relevant parameters, which might provide a better separability.

- MRSI and MRI parameters. The combined use of MR imaging intensities and metabolic information (chapter 6) increased the resulting performances with respect to imaging intensities alone, and, although not significantly, also with respect to metabolic data alone. The results strengthen the statement that imaging intensities and metabolic data provide complementary information for the accurate discrimination between several brain tissue types.

- Classification technique. Although PCA prior to LDA only expresses up to 75% of the data in chapters 4 and 5, this does not appear to compromise the classification efficiency of the classical technique LDA. For a few classification problems (e.g. for the grading of gliomas, chapter 6), nonlinear techniques performed slightly better than the linear ones, which indicates a partially nonlinear behaviour. Nonlinear techniques are worth to investigate further for diagnostic problems.

Research has also been done on quantification for ^{13}C glycogen MRS signals. The results can be summarized as follows:

- Information theoretic criteria have shown promising results for selecting the model order out of a set of possible values. Tests on simulation as well as experimental signals illustrated the ability of model selection to detect multi-exponential signals. Results depended on the noise level and the signal parameters of the different glycogen components.

- In a few particular situations (e.g. when using a multi-exponential model for signals of the first pulse phase), AMARES did not always converge without imposing additional model constraints. This indicates that analysis of experimental signals should be performed with caution.
- The overall results of Bergans et al. (2003) were confirmed and in particular ^{13}C glycogen MRS signals seem to express a multi-exponential behaviour.

8.2 Future research

Further research is necessary in the field of Magnetic Resonance processing. The final goal of pattern recognition applied to brain tumour data is to develop a decision support system integrable in clinical practice. Such a software tool would be very valuable for the non-invasive diagnosis of patients. Pattern recognition techniques are needed to manage the enormous and still growing amount of data generated in a clinical environment. Suggestions for research topics about pattern recognition include:

- Spatial information. Parameters from MR images and MRSI data have been merged for classification of brain tumours. In this thesis the morphological and spatial information in the images was not exploited as each voxel was treated as a separate sample. But each voxel has its own surrounding voxels which could be highly informative for determining the class of the voxel under investigation. Exploiting all tissue information available in an MR image would then be better comparable to the approach expert-radiologists take to interpret the image for medical diagnostic purposes. Moreover, the abilities of both conventional MRI and MRSI would then be directly comparable.
- Fusion of several MR modalities. Conventional MRI is widely applied for medical diagnostic purposes. But MRI has much more potential and also covers more recent modalities such as diffusion- and perfusion-weighted MRI and functional MRI. Each of these techniques has already illustrated its value for diagnosis of brain tumour patients. The information is spread over several techniques, which makes it difficult to derive a final diagnosis. Therefore, merging information from several MR techniques should achieve an optimal combination of anatomical, clinical and metabolic information.
- Contribution of MR parameters. The fusion of several MR techniques would increase the importance of dimensionality reduction. Feature selection should then be restricted to the most relevant parameters. Moreover, diagnosis based on a decision support system requires a high con-

fidence about the results of such a tool. Better insight into the value of the different parameters obtained by several MR techniques would be required. The integration of a probabilistic framework that informs about the contribution of each input feature to the classification is highly recommended.

- Binary and multiclass classification. The focus of the reported studies was mainly on binary classification, which assumes only two classes are available for diagnosis. However, there exists a high diversity in brain tumour types, which restricts the direct practical use of binary classification. The development of a multiclass system would be of high interest, possibly based on the available binary classifiers.
- MRS quantification prior to classification. Parameter estimates were provided as input features for classification. These estimates were obtained by the simple, but inaccurate peak integration method. Further studies are still needed to assess whether other more accurate model-based quantification methods (Provencher, 1993; Vanhamme et al., 2001) yield any significant improvement.
- Selection of hyperparameters. The choice of the hyperparameter(s) of kernel-based classifiers is quite crucial to achieve a high performance. More research can be done in finding the optimal hyperparameter values.

Suggestions for future research on quantification of MRS data include:

- Convergence for multi-exponential models. The quantification algorithm had convergence problems for several glycogen signals when modelling by multi-exponential models. The use of soft constraints might have an influence and is worth to investigate further.
- Model order selection prior to parameter estimation. Chapter 7 applied model selection by selecting a model out of a set of available candidates; in this case three models were considered. This implies that several models need to be computed, increasing the computation time with the number of models computed. Such an approach would not be feasible for practical situations where more candidate models should be included. Model order selection should then be applied prior to or embedded in the quantification procedure.
- Confidence intervals in case of multi-exponential signals. Quantification of glycogen signals needs to be performed as reliably and accurately as possible. In order to evaluate the parameter estimates, confidence bounds on the parameter estimates should be computed. Further research is still needed to develop such a method that is applicable to multi-exponential signals.

Appendix

A Medical background about brain tumours

Brain tumours account for 85% to 90% of all primary central nervous system (CNS) tumours. Worldwide, approximately 176,000 new cases of brain and other CNS tumours were diagnosed in the year 2000, with an estimated mortality of 128,000 (Parkin et al., 2001).

A brain tumour is a mass of unnecessary and abnormal cells growing in the brain. For treatment, brain tumours are classified by the type of cell in which the tumour begins, the location of the tumour in the central nervous system, and the grade of the tumour. A primary brain tumour is one that starts in the brain. They may spread within the central nervous system (brain and spinal cord), but rarely spread to other parts of the body. Glioblastomas and astrocytomas are examples of primary brain tumours. If tumour cells have spread from another cancerous part of the body, they are known as secondary brain tumours or metastatic brain tumours. Common cancers that spread to the brain are those of the lung, breast, melanoma, colon or an unknown primary site.

A detailed histological classification is done based on the World Health Organization (WHO) classification of nervous system tumours (Kleihues et al., 2002). The WHO approach incorporates markers concerning biological aspects such as the structure and form of the tumour, the structure and function of cells and mitosis (cell division) in an attempt to construct a cellular classification that is universally applicable and prognostically valid.

The grade of a tumour refers to how abnormal the cancer cells look under a microscope and how quickly the tumour is likely to grow and spread (Kleihues et al., 1993). The pathologist determines the grade of the tumour using tissue removed for biopsy. The following grading system may be used for adult brain tumours (National Cancer Institute, 2005). Grade I tumours grow slowly, have cells similar to normal cells, and rarely spread to nearby tissues. It may be possible to remove the entire tumour by surgery. Grade II tumours also grow slowly, but may spread into nearby tissue and may evolve to a higher grade. Grade III tumours grow quickly and are likely to spread into nearby tissue. The tumour cells look very different from normal cells. Grade IV tumours are very aggressively growing tumours, have cells that look very different from normal ones and are difficult to treat successfully.

The chance of recovery (prognosis) and choice of treatment depend on the type, grade, and location of the tumour and whether cancer cells remain after surgery and/or have spread to other parts of the brain.

B Medical terminology concerning brain tumours

This list of commonly used medical terms is compiled from (British Brain and Spine Foundation, 2001; National Cancer Institute, 2005; The Brain Tumor Society, 2005).

- **Anaplastic astrocytoma**: astrocytoma of WHO grade III, also known as malignant astrocytoma and high-grade astrocytoma. The tumour may arise from a grade II astrocytoma.
- **Aggressive tumour**: a quickly growing cancerous tumour.
- **Astrocytoma**: a type of brain tumour that begins in the brain or spinal cord in small, star-shaped cells called astrocytes.
- **Benign tumour**: a tumour that is not cancerous.
- **Biopsy**: removal of a small piece of living tissue from a part of the body for microscopic examination.
- **Blood brain barrier**: filtering mechanism made up of blood vessels and glial cells, that protects the brain by keeping out many harmful substances. Only certain types of chemotherapy can effectively cross this barrier to reach a brain tumour.
- **Calcification**: the deposit of calcium that is associated with certain types of tumours, such as meningiomas, astrocytomas, or oligodendrogliomas.
- **Cancer**: a term for diseases in which abnormal cells divide without control.
- **Cerebrospinal fluid (CSF)**: a clear and colorless fluid that surrounds the central nervous system. CSF has several functions including the protection of the brain from damage by buffering.
- **Chemotherapy**: the prevention or treatment of a tumour by the use of chemical substances.
- **Diffuse astrocytoma**: astrocytoma (WHO grade II) that grows slowly, but often spreads to nearby tissues. They can progress to a higher grade and occur most often in young adults.
- **Gd-DTPA or gadolinium-DTPA**: contrast agent frequently in MR imaging. The agent shortens the T_1 relaxation time which results in signal enhancement.
- **Glial tissue**: the special connective tissue of the central nervous system.

- **Glioblastoma**: malignant astrocytoma (WHO grade IV), also known as glioblastoma multiforme.
- **Glioma**: a tumour of the brain that begins in glial cells (cells that surround and support nerve cells). Also known as glial tumour.
- **Haemorrhage**: the escape of blood from a ruptured blood vessel, externally or internally.
- **High grade tumour**: a tumour that grows quickly over a period of a few months.
- **Histology**: the study of tissues and cells under a microscope.
- **Histopathology**: the study of the microscopic structure, composition and function of diseased tissue
- **Lesion**: area of tissue with impaired function as a result of damage by disease or wounding.
- **Low grade tumour**: a tumour that develops slowly over a number of years.
- **Malignant tumour**: a tumour that invades and destroys the tissue where it originates and which can spread to other sites in the body.
- **Membrane**: a thin layer of tissue surrounding an organ, lining a cavity or separating adjacent structures or cavities.
- **Meninges**: the three connective tissue membranes that line the skull and vertebral canal and cover and protect the brain and spinal cord.
- **Meningioma**: a type of tumour that occurs in the meninges. Meningiomas can have WHO grades I to III and usually grow slowly.
- **Metabolism**: the chemical changes in living cells by which energy is provided for vital processes and activities and new material is assimilated.
- **Metabolite**: a substance or molecule essential to the metabolism of a particular organism.
- **Metastasis**: the spread of cancer from one part of the body to another. A tumour formed from cells that have spread is called a secondary tumour, a metastatic tumour or a metastasis.
- **Mitosis**: the type of cell division that makes new body cells.
- **Mixed glioma**: brain tumour that contains more than one type of cell. The prognosis is dependent on the cell type with the highest grade present in the tumour.

- **Necrosis**: pathologic death of cells, tissue, or an organ due to irreversible damage.
- **Nerve cell**: the basic functional unit of the nervous system. These specialized cells transmit chemicals and electrical impulses and as such carry information from one part of the body to another.
- **Nervous system**: the vast network of nerve cells which carries information to and from all parts of the body in order to bring about bodily activity. It is classically divided into the brain, spine and peripheral nervous system.
- **Neurological**: refers to conditions occurring in the nervous system, including the brain, spine and all the peripheral nerves.
- **Neuron**: another term for nerve cell.
- **Non-invasive**: not involving penetration of the skin of an intact organism.
- **Oedema**: an accumulation of an excessive amount of watery fluid in cells, tissues or serous cavities.
- **Oligo-astrocytoma**: mixed glioma that grows slowly and has cells that look like astrocytes and oligodendrocytes.
- **Oligodendroglioma**: glioma that starts in the brain cells called oligodendrocytes, which support and nourish the nerve cells. Oligodendroglioma can have WHO grade II or grade III (anaplastic oligodendroglioma).
- **Oncologist**: a specialist in the treatment of tumours.
- **Pathologist**: a doctor who specialises in the study of disease, its causes, mechanisms and effects on the body.
- **Primary tumour**: tumour that only occurs in the organ in which it started.
- **Prognosis**: the assessment of the future course and outcome of a patient's condition.
- **Radiotherapy**: a therapeutic procedure in which conditions are treated with energy in the form of waves or particles.
- **Secondary tumour**: the spread of cancer from one part of the body to another.
- **Tumour**: any abnormal swelling in or on any part of the body. This term is usually used when the swelling is as a result of an overgrowth of cells.

- **Ventricles**: the fluid-filled cavities in the heart of the brain containing CSF.

Bibliography

Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov, B., Csaki, F. (Eds.), Second International Symposium on Information theory. Akademiai Kiado, Budapest, pp. 267–281.

Aldape, K., Okcu, M., Bondy, M., Wrensch, M., 2003. Molecular epidemiology of glioblastoma. Cancer J 9 (2), 99–106.

Barba, I., Moreno, A., Martinez-Perez, I., Tate, A., Cabanas, M., Baquero, M., Capdevila, A., Arús, C., 2001. Magnetic resonance spectroscopy of brain hemangiopericytomas: high myoinositol concentrations and discrimination from meningiomas. J. Neurosurg. 94 (1), 55–60.

Barkhuysen, H., de Beer, R., van Ormondt, D., 1987. Improved algorithm for noniterative time-domain model fitting to exponentially damped Magnetic Resonance Signals. J. Magn. Reson. 73, 553–557.

Bartha, R., Drost, D., Williamson, P., 1999. Factors affecting the quantification of short echo in-vivo ^{1}H MR spectra: prior knowledge, peak elimination, and filtering. NMR Biomed. 12, 205–216.

Behar, K., Rothman, D., Spencer, D., Petroff, O., 1994. Analysis of macromolecule resonances in ^{1}H NMR spectra of human brain. Magn. Reson. Med. 32, 294–302.

Bergans, N., May 2003. Biochemical and ^{13}C NMR spectroscopy study of pathways involved in glycogen turnover. Ph.D. thesis, Faculty of Medicine, Katholieke Universiteit Leuven.

Bergans, N., Dresselaers, T., Vanhamme, L., Van Hecke, P., Van Huffel, S., Vanstapel, F., 2003. Quantification of the glycogen ^{13}C-1 NMR signal during glycogen synthesis in perfused rat liver. NMR Biomed. 16, 36–46.

Bishop, C., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Bovée, W., Keevil, S., Leach, M., Podo, F., 1995. Quality assessment in in vivo NMR spectroscopy: II. A protocol for quality assessment. EEC Concerted Research Project. Magn. Reson. Imag. 13 (1), 123–129.

British Brain and Spine Foundation, 2001. Brain tumour: a guide for patients and carers.
URL `http://www.bbsf.org.uk/`

Burman, P., 1990. Estimation of optimal transformations using v-fold cross validation and repeated learning-testing methods. Sankhya: The Indian Journal of Statistics. Series A 52 (3), 314–345.

CaMon-project, 1998. Comprehensive Cancer Monitoring Programme in Europe.
URL `http://www-dep.iarc.fr/hmp/camon.htm`

Campbell, I., Dobson, C., Williams, R., Xavier, A., 1973. Resolution enhancement of protein PMR spectra using the difference between a broadened and a normal spectrum. J. Magn. Reson. 11, 172–181.

Cohen-Gadol, A., Pan, J., Kim, J., Spencer, D., Hetherington, H., 2004. Mesial temporal lobe epilepsy: a proton magnetic resonance spectroscopy study and a histopathological analysis. Journal of Neurosurgery 101 (4), 613–620.

Cousins, J., 1995. Clinical MR spectroscopy: fundamentals, current applications, and future potential. AJR Am J Roentgenol 164, 1337–1347.

de Beer, R., van Ormondt, D., 1992. NMR Basic Principles and Progress. Vol. 26. Springer-Verlag, Berlin/Heidelberg, Ch. Analysis of NMR data using time-domain fitting procedures, pp. 201–248.

de Graaf, R., 1998. *In-vivo* NMR spectroscopy. Wiley, England.

Dean, B., Drayer, B., Bird, C., Flom, R., Hodak, J., Coons, S., Carey, R., 1990. Gliomas: classification with MR imaging. Radiology 174, 411–415.

Devos, A., Bergans, N., Dresselaers, T., de Brabanter, J., Vanhamme, L., Van Huffel, S., Vanstapel, F., Van Hecke, P., 2005a. Model selection for quantification of a multi-exponential MR spectrum. Internal Report 05-39, ESAT-SISTA, K.U.Leuven, Leuven, Belgium.

Devos, A., Lukas, L., Suykens, J., Vanhamme, L., Tate, A., Howe, F., Majós, C., Moreno-Torres, A., van der Graaf, M., Arús, C., Van Huffel, S., 2004. Classification of brain tumours using short echo time ^{1}H MR spectra. J. Magn. Reson. 170 (1), 164–175.

Devos, A., Simonetti, A., van der Graaf, M., Lukas, L., Suykens, J., Vanhamme, L., Buydens, L., Heerschap, A., Van Huffel, S., 2005b. The use of

multivariate MR Imaging Intensities versus metabolic data from MR Spectroscopic Imaging for brain tumour classification. J. Magn. Reson. 173 (2), 218–228.

Devos, A., Van Huffel, S., Simonetti, A., van der Graaf, M., Heerschap, A., Buydens, L., 2005c. Classification of brain tumours by pattern recognition of Magnetic Resonance Imaging and Spectroscopic data. submitted as chapter in "A Multi-Perspective View on Outcome Prediction in Cancer", A. Taktak (ed.), Studies in Multidisciplanarity series, Elsevier 05-40, ESAT/SCD-SISTA, KULeuven.

Devos, A., Vanhamme, L., in 't Zandt, R., Van Huffel, S., Van Hecke, P., 2002a. Deliverable : automated quantification of in-vivo long echo proton magnetic resonance spectroscopic signals. Internal Report 02-209, ESAT-SISTA, K.U.Leuven, Leuven, Belgium.

Devos, A., Vanhamme, L., Van Huffel, S., 2002b. Deliverable : Number of acquisition blocks of MRS-signals. Internal Report 02-11, ESAT-SISTA, K.U.Leuven, Leuven, Belgium.

Duda, R., Hart, P., Stork, D., 2001. Pattern Classification, 2nd ed. John Wiley & Sons.

Earnest, F., Kelly, P., Scheithauer, B., Kall, B., Cascino, T., Ehman, R., Forbes, G., Axley, P., 1988. Cerebral astrocytomas: histopathologic correlation of MR and CT contrast enhancement with stereotactic biopsy. Radiology 166, 823–827.

Elderkin-Thompson, V., Thomas, M., Binesh, N., Mintz, J., Haroon, E., Dunkin, J., Kumar, A., 2004. Brain metabolites and cognitive function among older depressed and healthy individuals using 2D MR spectroscopy. Neuropsychopharmacology 29 (12), 2251–2257.

Ernst, T., Hennig, J., 1991. Coupling effects in volume selective 1H spectroscopy of major brain metabolites. Magn. Reson. Med. 21, 82–96.

Ernst, T., Kreis, R., Ross, B., 1993. Absolute quantitation of water and metabolites in the human brain. I. Compartments and water. J. Magn. Reson. B 102, 1–8.

Fisher, R., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179–188.

Fletcher, R., 1987. Practical methods of optimization. Wiley.

Fraley, C., Raftery, A., 1998. How many clusters? which clustering method? - answers via model-based cluster analysis. Computer J 41, 578–588.

Gadian, D., 1995. NMR and its applications to living systems, 2nd Edition. Oxford Science publishers.

Gasser, T., Sroka, L., Jennen-Steinmetz, C., 1986. Residual variance and residual pattern in non-linear regression. Biometrika 73 (3), 625–633.

Golub, G., Pereyra, V., 1973. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. SIAM J. Numer. Anal. 10 (2), 413–432.

Govindaraju, V., Young, K., Maudsley, A., 2000. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR Biomed. 13, 129–153.

Gunn, S., 1997. Support Vector Machines for Classification and Regression. Tech. rep., Image Speech and Intelligent Systems Research Group, University of Southampton.

Hagberg, G., Burlina, A., Mader, I., Roser, W., Radue, E., Seelig, J., 1995. In vivo proton MR spectroscopy of human gliomas: definition of metabolic coordinates for multi-dimensional classification. Magn. Reson. Med. 34, 242–252.

Hakumäki, J., Ala-Korpela, M., Kauppinen, R., 1997. ^{1}H nuclear magnetic resonance spectroscopy : noninvasive neurochemistry for basic research and clinical applications. Current topics in neurochemistry 1, 59–85.

Hammen, T., Stefan, H., Eberhardt, K., W-Huk, B., Tomandl, B., 2003. Clinical applications of 1H-MR spectroscopy in the evaluation of epilepsies–what do pathological spectra stand for with regard to current results and what answers do they give to common clinical questions concerning the treatment of epilepsies? Acta Neurol. Scand. 108, 223–238.

Hanley, J., McNeil, B., 1983. A method of comparing the areas under Receiver Operating Characteristic curves derived from the same cases. Radiology 148, 839–843.

Hannon, E., Quinn, B., 1979. The determination of the order of autoregression. Journal of the Royal Statistical Institute Series B 41, 190–195.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer-Verlag, Heidelberg.

Henriksen, O., 1995. *In vivo* quantitation of metabolite of concentrations in the brain by means of proton MRS. NMR Biomed. 8, 139–148.

Herminghaus, S., Dierks, T., Pilatus, U., Möller-Hartmann, W., Wittsack, J., Marquardt, G., Labisch, C., Lanfermann, H., Schlote, W., Zanella, F., 2003. Determination of histopathological tumor grade in neuroepithelial brain tumors by using spectral pattern analysis of in vivo spectroscopic data. J. Neurosurg. 98, 74–81.

Hoch, J., Stern, A., 1996. NMR data processing. John Wiley & Sons.

Hofmann, L., Slotboom, J., Boesch, C., Kreis, R., 1999. Model fitting of ^{1}H-MR spectra of the human brain: incorporation of short-T_1 components and evaluation of parametrized vs. non-parametrized models. In: Proceedings of the 7th Scientific Meeting & Exhibition (ISMRM99). Philadelphia, USA, p. 586.

Hofmann, L., Slotboom, J., Jung, B., Maloca, P., Boesch, C., Kreis, R., 2002. Quantitative ^{1}H-magnetic resonance spectroscopy of human brain: influence of composition and parameterization of the basis set in linear combination model-fitting. Magn. Reson. Med. 48, 440–453.

Howe, F., Barton, S., Cudlip, S., Stubbs, M., Saunders, D., Murphy, M., Wilkins, P., Opstad, K., Doyle, V., McLean, M., Bell, B., Griffiths, J., 2003. Metabolic profiles of human brain tumors using quantitative in vivo ^{1}H magnetic resonance spectroscopy. Magn. Reson. Med. 49, 223–232.

Huang, Y., Lisboa, P., El-Deredy, W., 2003. Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection. Statist. Med. 22, 147–164.

Hurvich, C., Tsai, C., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.

INTERPRET-project, 1999-2002. International network for Pattern Recognition of Tumours Using Magnetic Resonance (IST-1999-10310).
URL `http://carbon.uab.es/INTERPRET/`

Ishimaru, H., Morikawa, M., Iwanaga, S., Kaminogo, M., Ochi, M., Hayashi, K., 2001. Differentiation between high-grade glioma and metastatic brain tumor using single-voxel proton MR spectroscopy. Eur. Radiol. 11, 1784–1791.

IUPAC, 2000. International Union of Pure and Applied Chemistry (IUPAC).
URL `http://www.chem.qmul.ac.uk/iupac/`

Jung, J., Coakley, F., Vigneron, D., Swanson, M., Qayyum, A., Weinberg, V., Jones, K., Carroll, P., Kurhanewicz, J., 2004. Prostate depiction at endorectal MR spectroscopic imaging: Investigation of a standardized evaluation system. Radiology 233 (3), 701–708.

Keevil, S., Barbiroli, B., Collins, D., Danielsen, E., Hennig, J., Henriksen, O., Leach, M., Longo, R., Lowry, M., Moore, C., Moser, E., Segebarth, C., Bovée, W., Podo, F., 1995. Quality assessment in in vivo NMR spectroscopy: IV. A multicentre trial of test objects and protocols for performance assessment in clinical NMR spectroscopy. Magn. Reson. Imag. 13 (1), 139–157.

Kernohan, J., Mabon, R., Svien, K., Adson, A., 1949. A simplified classification of gliomas. Mayo Clinic Proceedings 24, 54–64.

Kleihues, P., Burger, P., Scheithauer, B., 1993. The new WHO Classification of brain tumours. Brain Pathol. 3 (3), 255–268.

Kleihues, P., Louis, D., Scheithauer, B., Rorke, L., Reifenberger, G., Burger, P., Cavanee, W., 2002. The WHO Classification of Tumors of the Nervous System. J. Neuropathol. Exp. Neurol. 61 (3), 215–225.

Klose, U., 1990. *In-vivo* proton spectroscopy in presence of eddy currents. Magn. Reson. Med. 14, 26–30.

Koehl, P., 1999. Linear prediction spectral analysis of NMR data. Prog. in NMR Spectrosc. 34, 257–299.

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In: IJCAI. pp. 1137–1145.
URL `citeseer.ist.psu.edu/kohavi95study.html`

Kohonen, T., 2001. Self-Organizing Maps, 3rd Edition. Vol. 30 of Springer Series in Information Sciences. Springer.

Kreis, R., 2004. Issues of spectral quality in clinical ^{1}H-magnetic resonance spectroscopy and a gallery of artifacts. NMR Biomed. 17, 361–381.

Kreis, R., Ernst, T., Ross, B., 1993. Absolute quantitation of water and metabolites in the human brain. II. Metabolite concentrations. J. Magn. Reson. B 102, 9–19.

Kung, S., Arun, K., Bhaskar Rao, D., 1983. State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem. J. Opt. Soc. Am. 73 (12), 1799–1811.

Kuroda, Y., Wada, A., Yamazaki, T., Nagayama, K., 1989. Postacquistion data processing method for suppression of the solvent signal. J. Magn. Reson. 84, 604–610.

Ladroue, C., Tate, A., Howe, F., Griffiths, J., August 2002. Exploring magnetic resonance data with independent component analysis. In: 19th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB02). Cannes, France, pp. 147–148.

Laudadio, T., Mastronardi, N., Vanhamme, L., Van Hecke, P., Van Huffel, S., 2002. Improved Lanczos algorithms for blackbox MRS data quantitation. J. Magn. Reson. 157, 292–297.

Lauterbur, P., 1973. Image formation by using local interactions: examples employing nuclear magnetic resonance. Nature 242, 190–191.

Law, M., Cha, S., Knopp, E., Johnson, G., Arnett, J., Litt, A., 2002. High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. Radiology 222, 715–721.

Leach, M., Collins, D., Keevil, S., Rowland, I., Smith, M., Henriksen, O., Bovée, W., Podo, F., 1995. Quality assessment in in vivo NMR spectroscopy: III. Clinical test objects: design, construction, and solutions. Magn. Reson. Imag. 13 (1), 131–137.

Leclerc, X., Huisman, T., Sorensen, A., 2002. The potential of proton magnetic resonance spectroscopy (^{1}H-MRS) in the diagnosis and management of patients with brain tumors. Current Opinion in Oncology 14, 292–298.

Leibel, S., Scott, C., Loeffler, J., 1994. Contemporary approaches to the treatment of malignant gliomas with radiation therapy. Semin. Oncol. 21, 198–219.

Lemmerling, P., Vanhamme, L., in 't Zandt, H., Van Huffel, S., Van Hecke, P., 2002. Time-domain quantification of short-echo-time proton MRS. MAGMA 15, 178–179.

Li, B., Wang, H., Gonen, O., 2003. Metabolite ratios to assumed stable creatine level may confound quantification of proton brain MR spectroscopy. Magn. Reson. Imag. 21, 923–928.

Lindon, J., Holmes, E., Nicholson, J., 2001. Pattern recognition methods and applications in biomedical magnetic resonance. Prog. Nucl. Magn. Reson. Spectrosc. 39, 1–40.

Lukas, L., Devos, A., Suykens, J., Vanhamme, L., Howe, F., Majós, C., Moreno-Torres, A., van der Graaf, M., Tate, A., Arús, C., Van Huffel, S., 2004. Brain Tumour Classification based on long echo proton MRS signals. Artif. Intell. Med. 31 (1), 73–89.

Lukas, L., Devos, A., Suykens, J., Vanhamme, L., Van Huffel, S., Tate, A., Majós, C., Arús, C., October 2002a. The use of LS-SVM in the classification of brain tumors based on ^{1}H-MR Spectroscopy signals. In: Proceedings of the IEE Symposium on Medical applications of signal processing. Savoy Place, London, UK, pp. 15/1–5.

Lukas, L., Devos, A., Suykens, J., Vanhamme, L., Van Huffel, S., Tate, A., Majós, C., Arús, C., April 2002b. The use of LS-SVM in the classification of brain tumors based on magnetic resonance spectroscopy signals. In: Proceedings of the European Symposium for Artifical Neural Networks (ESANN02). Bruges, Belgium, pp. 131–135.

Majós, C., Alonso, J., Aguilera, C., Serrallonga, M., Acebes, J., Arús, C., Gili, J., 2002. Adult primitive neuroectodermal tumor: Proton MR spectroscopic findings with possible application for differential diagnosis. Radiology 225, 556–566.

Majós, C., Alonso, J., Aguilera, C., Serrallonga, M., Pérez-Martín, J., Acebes, J., Arús, C., Gili, J., 2003. Proton magnetic resonance spectroscopy (^{1}H MRS) of human brain tumours: assessment of differences between tumour types and its applicability in brain tumour categorization. Eur. Radiol. 13, 582–591.

Mallows, C., 1973. Some comments on Cp. Technometrics 15, 661–675.

Manton, D., Lowry, M., Rowland-Hill, C., Crooks, D., Mathew, B., Turnbull, L., 2000. Combined proton MR spectroscopy and dynamic contrast enhanced MR imaging of human intracranial tumours *in vivo*. NMR Biomed. 13, 449–459.

Marshall, I., Higinbotham, J., Bruce, S., Freise, A., 1997. Use of Voigt lineshape for quantification of *in-vivo* ^{1}H spectra. Magn. Reson. Med. 37, 651–657.

Martin, E., Capone, A., Schneider, J., Hennig, J., Thiel, T., 2001. Absence of N-acetylaspartate in the human brain: Impact on neurospectroscopy? Ann Neurol 49(4), pp. 518–521.

MATLAB Support Vector Machines toolbox, Version 2.0-1998. MATLAB Support Vector Machines toolbox.
URL `http://www.isis.ecs.soton.ac.uk/isystems/kernel`

MATLAB/C LS-SVMlab toolbox, Version 1.4-2002. MATLAB/C LS-SVMlab toolbox.
URL `http://www.esat.kuleuven.ac.be/sista/lssvmlab`

McQuarrie, A., Tsai, C., 1998. Regression and time series model selection. World Scientific Publishing, Singapore.

Meyer, R., Fisher, M., Nelson, S., Brown, T., 1988. Evaluation of manual methods for integration of in vivo phosphorus NMR spectra. NMR Biomed. 1 (3), 131–135.

Mlynárnik, V., Gruber, S., Starčuk, Z., Starčuk, Z. J., Moser, E., 2000. Very short echo time proton MR spectroscopy of human brain with a standard transmit/receive surface coil. Magn. Reson. Med. 44, 964–967.

Montgomery, D., Runger, G., 1994. Applied statistics and probability for engineers. John Wiley & Sons, Inc.

Moody, J., 1992. The effective number of parameters: an analysis of generalization and regularization in non-linear learning systems. In: Advances in Neural Information Processing Systems 4, Proceedings of the 1991 NIPS Conference. Morgan Kaufmann Publishers, San Mateo, CA, pp. 847–854.

Mukherji, S. (Ed.), 1998. Clinical Applications of Magnetic Resonance Spectroscopy. Wiley-Liss.

Murphy, M., Loosemore, A., Clifton, A., Howe, F., Tate, A., Cudlip, S., Wilkins, P., Griffiths, J., Bell, B., 1993. The contribution of proton magnetic resonance spectroscopy (^{1}HMRS) to clinical brain tumour diagnosis. Magn. Reson. Med. 30, 518–519.

National Cancer Institute, 2005.
URL `http://www.nci.nih.gov/`

Negendank, W., 1992. Studies of human tumors by MRS: a review. NMR Biomed. 5, 302–324.

Nelson, S., 2003. Multivoxel magnetic resonance spectroscopy of brain tumors. Molecular Cancer Therapeutics 2 (5), 497–507.

Nelson, S., Cha, S., 2003. Imaging glioblastoma multiforme. Cancer J 9 (2), 134–145.

Nelson, S., Graves, E., Pirzkall, A., Li, X., Chan, A., Vigneron, D., McKnight, T., 2002. In Vivo Molecular Imaging for Planning Radiation Therapy of Gliomas: an Application of 1H MRSI. J. Magn. Reson. Imag. 16, 464–476.

Nelson, S., Vigneron, D., Dillon, W., 1999. Serial evaluation of patients with brain tumors using volume MRI and 3D ^{1}H MRSI. NMR Biomed. 12, 123–138.

Obuchowski, N., 2003. Receiver operating characteristic curves and their use in radiology. Radiology 229, 3–8.

Opstad, K., Murphy, M., Wilkins, P., Antholny Bell, B., Griffiths, J., Bell, B., Howe, F., 2004. Differentiation of Metastases from High-Grade Gliomas Using Short Echo Time ^{1}H Spectroscopy. J. Magn. Reson. Imag. 20, 187–192.

Osborne, M., 1972. Some aspects of non-linear least squares calculations. In: Lootsma, F. (Ed.), Numerical methods for non-linear optimization. Academic Press, London, pp. 171–189.

Overloop, K., Van Hecke, P., Vanstapel, F., Chen, H., Van Huffel, S., Knijn, A., van Ormondt, D., 1996. Evaluation of signal processing methods for the quantification of a multi-exponential signal: the glycogen ^{13}C-1 NMR signal. NMR Biomed. 9, 315–321.

Parkin, D., Bray, F., Ferlay, J., Pisani, P., 2001. Estimating the world cancer burden: Globocan 2000. Int. J. Cancer 94 (2), 153–156.

Pelckmans, K., Suykens, J., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., De Moor, B., Vandewalle, J., 2002. LS-SVMlab Toolbox User's Guide. Internal Report 02-145.

Pickett, B., Kurhanewicz, J., Coakley, F., Shinohara, K., Fein, B., Roach, M., 2004. Use of MRI and spectroscopy in evaluation of external beam radiotherapy for prostate cancer. International Journal of Radiation Oncology Biology Physics 60 (4), 1047–1055.

Pijnappel, W., van den Boogaart, A., de Beer, R., van Ormondt, D., 1992. SVD-based quantification of magnetic resonance signals. J. Magn. Reson. 97, 122–134.

Poptani, H., Kaartinen, J., Gupta, R., Niemitz, M., Hiltunen, Y., Kauppinen, R., 1999. Diagnostic assessment of brain tumours and non-neoplastic brain disorders in vivo using proton nuclear magnetic resonance spectroscopy and artificial neural networks. J. Cancer Res. Clin. Oncol. 125, 343–349.

Prados, M., 1991. Treatment strategies for patients with recurrent brain tumors. Semin. Radiat. Oncol. 1, 62–68.

Preul, M., Caramanos, Z., Collins, D., Villemure, J., Leblanc, R., Olivier, A., Pokrupa, R., Arnold, D., 1996. Accurate, noninvasive diagnosis of human brain tumors by using magnetic resonance spectroscopy. Nat. Med. 2 (3), 323–325.

Provencher, S., 1993. Estimation of metabolite concentrations from localized *in-vivo* proton NMR spectra. Magn. Reson. Med. 30, 672–679.

Ratiney, H., Coenradie, Y., Cavassila, S., van Ormondt, D., Graveron-Demilly, D., 2004. Time-domain quantitation of ^{1}H short echo-time signals: background accommodation. MAGMA 16, 284–296.

Rees, J., 2003. Advances in magnetic resonance imaging of brain tumours. Curr Opin Neurol 16, 643–650.

Rencher, A., 1995. Methods of multivariate analysis. Wiley series in probability and mathematical statistics.

Rice, J., 1995. Mathematical Statistic and Data Analysis. Duxbury Press.

Seeger, U., Klose, U., Mader, I., Grodd, W., Nägele, T., 2003. Parametrized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases. Magn. Reson. Med. 49, 19–28.

Simonetti, A., 2004. Investigation of brain tumor classification and its reliability using chemometrics on MR spectroscopy and MR imaging data. Phd thesis, Catholic University Nijmegen.

Simonetti, A., Melssen, W., Szabo de Edelenyi, F., van Asten, J., Heerschap, A., Buydens, L., 2004. Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification. NMR Biomed. 17, 1–10.

Simonetti, A., Melssen, W., van der Graaf, M., Heerschap, A., Buydens, L., 2002. Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction. J. Magn. Reson. 159, 151–157.

Simonetti, A., Melssen, W., van der Graaf, M., Postma, G., Heerschap, A., Buydens, L., 2003. A New Chemometric Approach for Brain Tumor Classification using Magnetic Resonance Imaging and Spectroscopy. Anal. Chem. 75 (20), 5352–5361.

Smith, I., Stewart, L., 2002. Magnetic resonance spectroscopy in medicine: clinical impact. Prog. Nucl. Magn. Reson. Spectrosc. 40, 1–34.

Soher, B., Young, K., Maudsley, A., 2001. Representation of strong baseline contribution in ^{1}H MR spectra. Magn. Reson. Med. 45, 966–972.

Srinivasan, A., 1999. Note on the location of optimal classifiers in n-dimensional ROC space. Tech. Rep. Technical Report PRG-TR-2-99, Oxford University Computing Laboratory, Oxford, England.

Stephenson, D., 1988. Linear prediction and maximum entropy methods in NMR spectroscopy. Prog. in NMR Spectrosc. 20, 515–626.

Sundin, T., Vanhamme, L., Van Hecke, P., Dologlou, I., Van Huffel, S., 1999. Accurate quantification of ^{1}H spectra: from finite impulse response filter design for solvent suppression to parameter estimation. J. Magn. Reson. 139, 189–204.

Suykens, J., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J., 2002. Least Squares Support Vector Machines. World Scientific Publishing Co., Singapore.

Suykens, J., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neur. Proc. Lett. 9 (3), 293–300.

Swets, J., 1979. ROC analysis applied to the evaluation of medical imaging techniques. Invest. Radiol. 14 (2), 109–121.

Szabo De Edelenyi, F., Estève, F., Rémy, C., Buydens, L., August 2002. Application of independent component analysis to ^{1}H MR spectroscopic imaging exams of brain tumors. In: 19th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB02). Cannes, France, p. 91.

Szabo De Edelenyi, F., Rubin, C., Estève, F., Grand, S., Décorps, M., Lefournier, V., Le Bas, J., Rémy, C., 2000. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. Nat. Med. 6, 1287–1289.

Tate, A., Crabb, S., Griffiths, J., Howells, S., Mazucco, R., Rodrigues, L., Watson, D., 1996. Lipid Metabolite Peaks in Pattern Recognition Analysis of Tumour *in vivo* MR Spectra. Anticancer Res. 16, 1575–1580.

Tate, A., Griffiths, J., Martinez-Pérez, I., Moreno, A., Barba, I., Cabañas, M., Watson, D., Alonso, J., Bartumeus, F., Isamat, F., Ferrer, I., Villa, F., Ferrer, E., Capdevila, A., Arús, C., 1998. Towards a method for automated classification of ^{1}H MRS spectra from brain tumours. NMR Biomed. 11, 177–191.

Tate, A., J.R. Griffiths, F. H., Pujol, J., Arús, C., April 2002. Differentiating Types of Human Brain Tumours by MRS. A Comparison of Pre-processing Methods and Echo Times. In: Proceedings of the Ninth Scientific Meeting & Exhibition (ISMRM01). Glasgow, Scotland, p. 2284.

Tate, A., Majós, C., Moreno, A., Howe, F., Griffiths, J., Arús, C., 2003. Automated classification of short echo time in in vivo ^{1}H brain tumor spectra: a multicenter study. Magn. Reson. Med. 49, 29–36.

The Brain Tumor Society, 2005.
URL `http://www.tbts.org`

Usenius, J., Tuohimetsa, S., Vainio, P., Ala-Korpela, M., Hiltunen, Y., Kauppinen, R., 1996. Automated classification of human brain tumours by neural network analysis using in vivo 1H magnetic resonance spectroscopic metabolite phenotypes. Neuroreport 7 (10), 1597–1600.

van der Graaf, M., 2000. Quality Control of MR spectra in the INTERPRET-validated database.
URL `http://carbon.uab.es/INTERPRET/mrs_data/docs/MRS_QC.pdf`

van der Graaf, M., 2001. Deliverable Number 4. D2.2: Data Protocols (MRS) (EU-IST-1999-10310).
URL `http://carbon.uab.es/INTERPRET/mrs_data/mrs_data.html`

van der Graaf, M., Rijpkema, M., van der Meulen, Y., Majós, C., Moreno, A., Ziegler, A., Howe, F., Opstad, K., Heerschap, A., 2002. System quality assurance in INTERPRET, a multicenter study on brain tumor MR spectroscopy. MAGMA 15 (suppl. 1), 245–246.

van der Veen, J., de Beer, R., Luyten, P., van Ormondt, D., 1988. Accurate Quantification of *in-vivo* ^{31}P NMR Signals Using the Variable Projection Method and Prior Knowledge. Magn. Reson. Med. 6, 92–98.

Van Huffel, S., Chen, H., Decanniere, C., Van Hecke, P., 1994. Algorithm for time domain NMR data fitting based on total least squares. J. Magn. Reson. A 110, 228–237.

Vanhamme, L., 1999. Advanced time-domain methods for nuclear magnetic resonance spectroscopy data analysis. Ph.D. thesis, Dept. of Electr. Eng., Katholieke Universiteit Leuven, downloadable from ftp.esat.kuleuven.ac.be/sista/members/vanhamme/reports/phd.ps.gz.

Vanhamme, L., Sundin, T., Van Hecke, P., Van Huffel, S., 2001. MR spectroscopy quantitation: a review of time-domain methods. NMR Biomed. 14, 233–246.

Vanhamme, L., Sundin, T., Van Hecke, P., Van Huffel, S., Pintelon, R., 2000. Frequency-selective quantification of biomedical magnetic resonance spectroscopy data. J. Magn. Reson. 143 (1), 1–16.

Vanhamme, L., van den Boogaart, A., Van Huffel, S., 1997. Improved Method for Accurate and Efficient Quantification of MRS Data with Use of Prior Knowledge. J. Magn. Reson. 129, 35–43.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Vigneron, D., Bollen, A., McDermott, M., Wald, L., Day, M., Moyher-Moworolski, S., Henry, R., Chang, S., Berger, M., Dillon, W., Nelson, S., 2001. Three-dimensional magnetic resonance spectroscopic imaging of histologically confirmed brain tumors. Magn. Reson. Imag. 19, 89–101.

Wehrens, R., Simonetti, A., Buydens, L., 2002. Mixture modelling of medical magnetic resonance data. J. Chemometrics 16, 274–282.

Weiss, S., Kulikowski, C., 1991. Computer Systems that Learn: Classification and Prediction Methods From Statistics, Neural Nets, Machine Learning, and Expert Systems. Morgan Kaufmann, San Mateo, CA.

Yang, J., Yang, J., 2003. Why can LDA be performed in PCA transformed space? Pattern Recognition 36 (2), 563–566.

Ye, C., Yang, J., Geng, D., Zhou, Y., Chen, N., 2002. Fuzzy rules to predict degree of malignancy. Med. Biol. Eng. Comput. 40, 145–152.

Yiannoutsos, C., Ernst, T., Chang, L., Lee, P., Richards, T., Marra, C., Meyerhoff, D., Jarvik, J., Kolson, D., Schifitto, G., Ellis, R., Swindells, S., Simpson, D., Miller, E., Gonzalez, R., Navia, B., 2004. Regional patterns of brain metabolites in AIDS dementia complex. Neuroimage 23 (3), 928–935.

Young, K., Soher, B., Maudsley, A., 1998. Automated spectral analysis II: application of wavelet shrinkage for characterization of non-parameterized signals. Magn. Reson. Med. 40, 816–821.

Publication list

Articles in (inter)national journals

1. L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, F.A. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, A.R. Tate, C. Arús, S. Van Huffel, 2003. Brain Tumour Classification based on long echo proton MRS signals, Artif. Intell. Med. 31 (1), 73-89.

2. A. Devos, L. Lukas, J.A.K. Suykens, L. Vanhamme, A.R. Tate, F.A. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, C. Arús, S. Van Huffel, 2004. Classification of brain tumours using short echo time ^{1}H MR spectra, J. Magn. Reson. 170 (1), 164-175.

3. A. Devos, A.W. Simonetti, M. van der Graaf, L. Lukas, J.A.K. Suykens, L. Vanhamme, L.M.C. Buydens, A. Heerschap, S. Van Huffel, 2005. The use of multivariate MR Imaging Intensities versus metabolic data from MR Spectroscopic Imaging for brain tumour classification, J. Magn. Reson. 173 (2), 218-228.

4. L. Vanhamme, A. Devos, P. Van Hecke, S. Van Huffel, 2005. Van spectroscopisch signaal tot klinische diagnose, Het Ingenieursblad 3/2005, 52-58.

Chapters in books

1. A. Devos, S. Van Huffel, A.W. Simonetti, M. van der Graaf, A. Heerschap and L.M.C. Buydens, Classification of brain tumours by pattern recognition of Magnetic Resonance Imaging and Spectroscopic data, Internal Report 05-40, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), submitted as chapter in "A Multi-Perspective View on Outcome Prediction in Cancer", A. Taktak (ed.), Studies in Multidisciplanarity series, Elsevier, 2005.

Articles in conference proceedings

1. L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, S. Van Huffel, A.R. Tate, C. Majós, C. Arús, The use of LS-SVM in the classification of brain tumors based on Magnetic Resonance Spectroscopic signals, in Proc. of the European Symposium on Artificial Neural Networks (ESANN'2000), Bruges, Belgium, April 2000, pp. 131-136.

2. L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, S. Van Huffel, A.R. Tate, C. Majós, C. Arús, The use of LS-SVM in the classification of brain tumors based on ^{1}H-MR Spectroscopy signals. *in* "Proceedings of the IEE Symposium on Medical applications of signal processing", Savoy Place, London, UK, Oct. 2002, 15/1-5.

3. A. Devos, N. Bergans, J. De Brabanter, L. Vanhamme, F. Vanstapel, P. Van Hecke and S. Van Huffel, Model selection for quantification of a multi-exponential MRS signal, in Proc. of the 2nd European Medical and Biological Engineering Conference (EMBEC'02), Vienna, Austria, Dec. 2002, pp. 1066-1067.

4. A. Devos, L. Lukas, A.W. Simonetti, M. van der Graaf, L.M.C. Buydens, A. Heerschap, S. Van Huffel, Does the combination of Magnetic Resonance Imaging and Spectroscopic Imaging improve the classification of brain tumours?, in Proc. of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), San Francisco, California, Sep. 2004.

Abstracts for (inter)national conferences

1. A. Devos, L. Vanhamme, L. Lukas, S. Van Huffel, J.A.K. Suykens, Least Squares Support Vector Machines for pre-operative pattern recognition of brain tumors using Magnetic Resonance Spectroscopy. 9th annual meeting of the Belgian Statistical Society, Oostende, Belgium, Oct. 12-13, 2001.

2. L. Lukas, A. Devos, J.A.K. Suykens, L. Vanhamme, F.A. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, A.R. Tate, C. Arús, S. Van Huffel, Classification of brain tumours using short echo time ^{1}H MRS spectra, 3rd Belgian Day on Biomedical Engineering, Oct. 17, 2003.

3. A. Devos, L. Lukas, J.A.K. Suykens, L. Vanhamme, F.A. Howe, C. Majós, A. Moreno-Torres, M. van der Graaf, A.R. Tate, C. Arús, S. Van Huffel, Brain tumour classification using short echo time 1H MRS. Objective comparison of classification techniques (LDA, LS-SVM)., in Proc. of the 12th Scientific Meeting and Exhibition of the International Society for Magnetic Resonance in Medicine (ISMRM04), Kyoto, Japan, May 15-21, 2004, 384 p.

4. A. Devos, L. Lukas, J.A.K. Suykens, L. Vanhamme, S. Van Huffel, Classification of brain tumours using short echo time ^{1}H MRS spectra, 1st Flanders PhD symposium, Dec. 11, 2003.

5. A.W. Simonetti, A. Devos, A. Heerschap, S. Van Huffel, L.M.C. Buydens, Does brain tumor classification improve on fusion of MRI and short echo time MRSI?, 13th Scientific Meeting in Miami Beach, Florida (ISMRM05), USA, May 7-13, 2005.

Internal reports (including submitted papers)

1. A. Devos, L. Vanhamme and S. Van Huffel, Deliverable : Number of acquisition blocks of MRS-signals, Internal Report 02-11, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2002.

2. A. Devos, L. Vanhamme, R. in 't Zandt, S. Van Huffel and P. Van Hecke, Deliverable : automated quantification of in-vivo long echo proton magnetic resonance spectroscopic signals, Internal Report 02-209, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2002.

3. C. Lu, A. Devos, J.A.K Suykens, C. Arús, S. Van Huffel, Bagging linear sparse Bayesian learning models for variable selection in cancer diagnosis, Internal Report 04-176, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2004.

4. A. Devos, N. Bergans, T. Dresselaers, J. de Brabanter, L. Vanhamme, S. Van Huffel, F. Vanstapel and P. Van Hecke, Model selection for quantification of a multi-exponential MR spectrum, Internal Report 05-39, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2005.

Andy Devos was born on November 18, 1977, in Kortrijk. He received the degree of Mathematics, option Astronomy and Mathematical Physics from the Katholieke Universiteit Leuven, Belgium in September 1999. He received the degree in Artificial Intelligence at the same university in June 2000. In August 2000 he started as a researcher at the Department of Electrical Engineering (ESAT) of the K.U. Leuven, first on a K.U. Leuven grant and from October 2001 on an I.W.T.-Vlaanderen (Institute for the Promotion of Innovation through Science and Technology in Flanders) grant.