KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

UNIVERSITÉ DE LYON
Lyon 1

École Doctorale Interdisciplinaire Sciences Santé
EDISS - ED205
Faculté des Sciences et Technologies

Machine learning for classifying abnormal brain tissue progression based on multi-parametric Magnetic Resonance data

Adrian Ion-Mărgineanu

Supervisors:
Prof. dr. ir. S. Van Huffel
M.Cf. dr. ir. D. Sappey-Marinier
Prof. dr. ir. F. Maes
Dr. D.M. Sima

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

October 2017

Machine learning for classifying abnormal brain tissue progression based on multi-parametric Magnetic Resonance data

Adrian ION-MĂRGINEANU

Examination committee:
Prof. dr. ir. J. Vandewalle, chair
Prof. dr. ir. S. Van Huffel, supervisor
M.Cf. dr. ir. D. Sappey-Marinier, supervisor
Prof. dr. ir. F. Maes, co-supervisor
Dr. D.M. Sima, co-supervisor
Prof. dr. ir. J. Suykens
Prof. dr. A. Heerschap
(RUNMC, Nijmegen)
Assoc. Prof. dr. ir. C. Frindel
(INSA-Lyon, Lyon)
M.Cf. dr. D. Maucort-Boulch
(UCBL, Lyon)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

October 2017

© 2017 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Adrian Ion-Mărgineanu, Kasteelpark Arenberg 10, box 2446, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, 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, electronic or any other means without written permission from the publisher.

Dedic această carte Directorului Programului de Master ITEMS, Prof. Vasile Buzuloiu, şi bunicilor mei, Elisabeta Mărgineanu, Petre Mărgineanu, şi Aurel Zainea.

Preface

To start, I want to thank three people who became my academic family, because they made all this research possible and helped me tremendously along these four years. First, I want to thank my KU Leuven promoter, Prof. Sabine Van Huffel, for believing in me and accepting me in two exceptional research groups: BioMed and TRANSACT. Thank you for the academic guidance, the nice words of encouragement during bad times, and for filling my (and the whole BioMed's) doctoral years with warm social activities. Second, I want to thank my Université de Lyon promoter, Prof. Dominique Sappey-Marinier, for giving me the opportunity to study under his supervision and for all the insights into the French lifestyle. It has been a pleasure to be part of your group and I will miss all the interesting discussions, both academic as well as non-academic. Last but definitely not least, I want to thank my daily advisor, Dr. Diana Sima, for .. everything. It was an honour meeting you and a privilege having you by my side. No words can describe the immense gratitude that I hold for all three of you, because you helped me develop into a better researcher and a better man.

To continue, I want to thank Dr. Sofie Van Cauter, Prof. Uwe Himmelreich and Prof. Frederik Maes for the close collaboration during our meetings. Your support and knowledge contributed heavily to my personal development and to this thesis.

Next, I want to thank my Examination Committee for accepting to review this thesis: Prof. Joos Vandewalle, Prof. Sabine Van Huffel, Prof. Dominique Sappey-Marinier, Prof. Frederik Maes, Dr. Diana Sima, Prof. Delphine Maucort-Boulch, Assoc. Prof. Carole Frindel, Prof. Arend Heerschap and Prof. Johan Suykens.

Personally, I want to thank Bogdan for all the time spent together, for all the coffees and especially the beers. It wouldn't have been the same without you. And the same goes for Delia, as well ;) Va pup !

BioMed, I hope you know that I love you all, old and new: Alex 'Le Chef', Abhi, Alexander, Amir, Anca, Amalia, Bharath, Bori, Carolina, Dzemila, Dorien, Dries, Frederik, Griet, Ivan, Jasper, Javier, John, Jonathan, Kaat, Laure, Lieven, Martijn, Mario, Margot, Matthieu, Neetha, Ninah, Nico, Nicolas, Ofelie, Otto, Rob, Simon, Sibasankar, Stijn, Tim, Thomas, Vanya, Vladimir, Ying, Yipeng, and Yissel. Thanks for all the activities, laughs, parties, lunch discussions, football .. and everything in between .. I will miss you more than you (or I) think I will.

Special thanks to the NMR office, and in particular to Nicolas and Bharath, for their patience and every small discussion we ever had. It was a pleasure having you by my side.

Being an Early Stage Research Fellow of the European Union TRANSACT network literally changed my life in ways I could not imagine. I got to meet a lot of very friendly international people, who formed a very warm group. A big 'thank you' to all ESRs: Pruthvi, Nuno, Nassim, Sana, Veronika, Mihal, Saurabh, Iveta, and Ross. Special shout-outs to my spiritual brothers Victor, Akila, Miguel, and to our manager Aldona. I am very happy to have met you, and hope to see you again very soon!

Claudio, we were basically brothers-in-arms these last 3 years .. always between Leuven and Lyon. It was a real privilege meeting you and spending time together. I will never forget you <3

Gabriel, a massive 'thank you' for all the help. You made my stay in Lyon a lot more enjoyable and I am glad to call you my friend.

A big 'thank you' also to the rest of my colleagues from sunny Lyon: Monica, Radu, Jamila, Ilaria, Violaine, Nikos, and Salem.

In the end, I want to thank all my Romanian friends for keeping me (in)sane: Victor, Bogdan, Delia, Liana, Roxana, Adi H., Jeni, Horia, Andreea, Mihai, Cristina, Adi G., Carmen .. si bineinteles Gabi!

Mama, Tata, Miru .. va iubesc! Multumesc ca m-ati crescut, ajutat, si iubit.

Abstract

Machine learning is a subdiscipline in the field of artificial intelligence, which focuses on algorithms capable of adapting their parameters based on a set of observed data, by optimizing an objective or cost function. Machine learning has been the subject of large interest in the biomedical community because it can improve sensitivity and/or specificity of detection and diagnosis of any disease, while increasing the objectivity of the decision-making process. With the late increase in volume and complexity of medical data being collected, there is a clear need for applying machine learning algorithms in multi-parametric analysis for new detection and diagnostic modalities.

Biomedical imaging is becoming indispensable for healthcare, as multiple modalities, such as Magnetic Resonance Imaging (MRI), Computed Tomography, and Positron Emission Tomography, are being increasingly used in both research and clinical settings. The non-invasive standard for brain imaging is MRI, as it can provide structural and functional brain maps with high resolution, all within acceptable scanning times. However, with the increase of MRI data volume and complexity, it is becoming more time consuming and difficult for clinicians to integrate all data and make accurate decisions.

The aim of this thesis is to develop machine learning methods for automated preprocessing and diagnosis of abnormal brain tissues, in particular for the follow-up of glioblastoma multiforme (GBM) and multiple sclerosis (MS). Current conventional MRI (cMRI) techniques are very useful in detecting the main features of brain tumours and MS lesions, such as size and location, but are insufficient in specifying the grade or evolution of the disease. Therefore, the acquisition of advanced MRI, such as perfusion weighted imaging (PWI), diffusion kurtosis imaging (DKI), and magnetic resonance spectroscopic imaging (MRSI), is necessary to provide complementary information such as blood flow, tissue organisation, and metabolism, induced by pathological changes. In the GBM experiments our aim is to discriminate and predict the evolution of patients treated with standard radiochemotherapy and immunotherapy based

on conventional and advanced MRI data. In the MS experiments our aim is to discriminate between healthy subjects and MS patients, as well as between different MS forms, based only on clinical and MRSI data.

As a first experiment in GBM follow-up, only advanced MRI parameters were explored on a relatively small subset of patients. Average PWI parameters computed on manually delineated regions of interest (ROI) were found to be perfect biomarkers for predicting GBM evolution one month prior to the clinicians.

In a second experiment in GBM follow-up of a larger subset of patients, MRSI was replaced by cMRI, while PWI and DKI parameter quantification was automated. Feature extraction was done on semi-manual tumour delineations, thereby reducing the time put by the clinician for manual delineating the contrast enhancing (CE) ROI. Learning a modified boosting algorithm on features extracted from semi-manual ROIs was shown to provide very high accuracy results for GBM diagnosis.

In a third experiment in GBM follow-up of an extended subset of patients, a modified version of parametric response maps (PRM) was proposed to take into account the most likely infiltration area of the tumour, reducing even further the time a clinician would have to put for manual delineating the tumour, because all subsequent MRI scans were registered to the first one. Two types of computing PRM were compared, one based on cMRI and one based on PWI, as features extracted with these two modalities were the best in discriminating the GBM evolution, according to results from the previous two experiments. Results obtained within this last GBM analysis showed that using PRM based on cMRI is clearly superior to using PRM based on PWI.

As a first experiment in MS follow-up, machine learning algorithms for binary classification problems were tuned on multiple types of data, such as metabolic features, clinical data (e.g. patient age, disease age), and lesion load. Classification results for discriminating healthy control subjects from MS patients were not satisfactory, even though statistically significant differences between the two groups were observed. Classification results for discriminating between different MS forms based only on MRSI features were moderate, while high classification results were found only when incorporating clinical data.

A second experiment was done in order to extract higher level MRSI features and used state of the art machine learning algorithms, such as convolutional neural networks, but results obtained with more complex classifiers did not outperform the ones obtained with classical algorithms trained on more simple MRSI features.

Beknopte samenvatting

Machine learning is een onderdeel van de studie naar artificiele intelligentie dat zich richt op het aanpassen van parameters in algoritmen (gebaseerd op verkregen data) om een functie te optimaliseren. Binnen de biomedische wetenschappen hebben machine learning methoden grote interesse verworven omdat deze de sensitivieteit en/of specificiteit van diagnoses in ziektes verbeteren en tegelijkertijd de objectiviteit van deze beslissingen verhogen.

Biomedische beeldvorming is van groot belang in de gezondheidszorg aangezien meerdere methoden, zoals bijvoorbeeld magnetische resonantie imaging (MRI), computertomografie en positron emissie tomografie in toenemende mate worden gebruikt voor onderzoek en klinische diagnosen. MRI is op dit moment de standaard voor niet-invasieve hersenbeeldvorming omdat het structurele en functionele hersenbeelden met hoge resolutie biedt met aanvaardbare scantijden. Echter, door het toenemen van de hoeveelheid MRI data en complexiteit kost het de clinici steeds meer tijd en expertise om deze data te interpreteren om correcte medische beslissingen te maken.

Het doel van dit proefschrift is het ontwikkelen van machinale leermethoden voor geautomatiseerde preprocessing en diagnose van abnormale hersenweefsels, met name voor het opvolgen van glioblastoma multiforme (GBM) en multiple sclerose (MS). Huidige conventionele MRI (cMRI) technieken zijn zeer nuttig bij het opsporen van de belangrijkste kenmerken van hersentumoren en MS-letsels, zoals grootte en locatie, maar zijn onvoldoende om de graad of evolutie van de ziekte te specificeren. Daarom is de acquisitie van geavanceerde MRI, zoals perfusiegewogen beeldvorming (PWI), diffusie kurtosis imaging (DKI), en magnetische resonantie spectroscopische beeldvorming (MRSI) nodig om complementaire informatie te verschaffen zoals bloedstroom, weefselorganisatie en metabolisme, die kenmerkend zijn voor pathologische veranderingen. In de GBM experimenten is het ons doel om de evolutie van patiënten te voorspellen en onderscheid te maken tussen de patiënten die behandeld worden met standaard radiochemotherapie en immunotherapie op basis van conventionele en

geavanceerde MRI data. In de MS-experimenten is het ons doel om onscheid te maken tussen gezonde personen en MS-patiënten, evenals tussen verschillende MS-vormen, beiden uitsluitend op basis van klinische en MRSI-gegevens.

In het eerste experiment voor GBM-follow-up werden alleen geavanceerde MRI-parameters onderzocht op een relatief kleine subset van patiënten. Gemiddelde PWI parameters berekend op manuele aflijningen van 'regio's van belangstelling' (ROI) bleken perfecte biomarkers te zijn voor het voorspellen van GBM evolutie een maand eerder dan de clinici.

In een tweede experiment met een grotere deelgroep patiënten voor GBM-follow-up werd MRSI vervangen door cMRI en de PWI en DKI parameter kwantificering werd geautomatiseerd. Kenmerk-extractie werd gedaan op basis van semi-manuele tumoraflijningen, waardoor de tijd die de clinicus nodig heeft voor manuele aflijning van het contrastverhoging (CE) ROI korter wordt. Hoge classificatie nauwkeurigheid werd aangetoond op basis van een aangepast boosting-algoritme, toegepast op kenmerken die uit semi-manuele ROI's werden geëxtraheerd.

In een derde experiment voor GBM-follow-up met een uitgebreide subset van patiënten werd een gewijzigde versie van parametrische responsbeelden (PRM) voorgesteld om rekening te houden met het meest waarschijnlijke infiltratiegebied van de tumor. Door alle daaropvolgende MRI-scans te refereren aan de eerste, zou de tijd die een clinicus doorbrengt met manuele aflijningen nog verder dalen. Twee PRM varianten werden vergeleken, één gebaseerd op cMRI en één op basis van PWI, aangezien deze kenmerken de beste waren in de discriminatie van de GBM evolutie. Zoals bleek uit de resultaten van de vorige twee experimenten. Resultaten laten zien dat het gebruik van PRM op basis van cMRI duidelijk beter is dan het gebruik van PRM op basis van PWI.

Als eerste experiment voor MS-follow-up werden machine learning algoritmes voor binaire classificatieproblemen afgestemd op meerdere soorten data zoals: metabolische kenmerken, klinische data (bijvoorbeeld patiënt leeftijd, ziekte leeftijd) en laesie volume. Classificatie tussen gezonde subjecten en MS patiënten waren niet bevredigend, alhoewel er wel statistisch significante verschillen tussen de twee groepen werden waargenomen. Classificatie tussen verschillende MS vormen gebaseerd op MRSI kenmerken waren redelijk, terwijl hoge classificatie resultaten alleen gevonden zijn bij het gebruiken van klinische gegevens.

Een tweede experiment werd uitgevoerd om diepere informatie uit de MRSI te halen en state-of-the-art machine learning methoden, zoals convolutionele neurale netwerken, te gebruiken. De resultaten die werden verkregen met deze complexere classifiers waren niet beter dan die verkregen met klassieke algoritmen die werken met eenvoudigere MRSI kenmerken.

Résumé

«Machine Learning» est un champ d'étude de l'intelligence artificielle qui se concentre sur des algorithmes capables d'adapter leur paramètres en se basant sur les données observées par l'optimisation d'une fonction objective ou d'une fonction de cout. Cette discipline a soulevé l'intérêt de la communauté de la recherche biomédicale puisqu'elle permet d'améliorer la sensibilité et la spécificité de la détection et du diagnostic de nombreuses pathologies tout en augmentant l'objectivité dans le processus de prise de décision thérapeutique.

L'imagerie biomédicale est devenue indispensable en médecine, puisque plusieurs modalités comme l'imagerie par résonance magnétique (IRM), la tomodensitométrie et la tomographie par émission de positron sont de plus en plus utilisées en recherche et en clinique. L'IRM est la technique d'imagerie non-invasive de référence pour l'étude du cerveau humain puisqu'elle permet dans un temps d'acquisition raisonnable d'obtenir à la fois des cartographies structurelles et fonctionnelles avec une résolution spatiale élevée. Cependant, avec l'augmentation du volume et de la complexité des données IRM, il devient de plus en plus long et difficile pour le clinicien d'intégrer toutes les données afin de prendre des décisions précises.

Le but de cette thèse est de développer des méthodes de « machine learning » automatisées pour la détection de tissu cérébral anormal, en particulier dans le cas de suivi de glioblastome multiforme (GBM) et de sclérose en plaques (SEP). Les techniques d'IRM conventionnelles (IRMc) actuelles sont très utiles pour détecter les principales caractéristiques des tumeurs cérébrales et les lésions de SEP, telles que leur localisation et leur taille, mais ne sont pas suffisantes pour spécifier le grade ou prédire l'évolution de la maladie. Ainsi, les techniques d'IRM avancées, telles que l'imagerie de perfusion (PWI), de diffusion (DKI) et la spectroscopie par résonance magnétique (SRM), sont nécessaires pour apporter des informations complémentaires sur les variations du flux sanguin, de l'organisation tissulaire et du métabolisme induits par la maladie.

Dans une première étude de suivi de patients GBM, seuls les paramètres d'IRM avancés ont été explorés dans un relativement petit sous-groupe de patients. Les paramètres de PWI moyens, mesurés dans les régions d'intérêts (ROI) délimités manuellement, se sont avérés être d'excellents marqueurs, puisqu'ils permettent de prédire l'évolution du GBM en moyenne un mois plus tôt que le clinicien.

Dans une seconde étude, réalisée sur un échantillon plus important que la précédente, la SRM a été remplacée par l'IRMc et la quantification de la PWI et du kurtosis de diffusion (DKI) a été réalisée de manière automatique. L'extraction des paramètres d'imagerie a été effectuée sur des segmentations semi-automatiques des tumeurs, réduisant ainsi le temps nécessaire au clinicien pour la délimitation du ROI de la partie de la lésion rehaussée au produit de contraste (CE-ROI). L'application d'un algorithme modifié de «boosting» sur les paramètres extraits des ROIs a montré une grande précision pour le diagnostic du GBM.

Dans une troisième, une version modifiée des cartes paramétriques de réponse (PRM) est proposée pour prendre en compte la région d'infiltration de la tumeur, réduisant toujours plus le temps nécessaire pour la délimitation de la tumeur par le clinicien, puisque toutes les images IRM sont recalées sur la première. Deux façons de générer les RPM ont été comparées, l'une basée sur l'IRMc et l'autre basée sur la PWI, ces deux paramètres étant les meilleurs pour la discrimination de l'évolution du GBM, comme le montrent les deux études précédentes. Les résultats de cette étude montrent que l'emploi de PRM basés sur l'IRMc permet d'obtenir des résultats supérieurs à ceux obtenus avec les PRM basés sur la PWI.

Dans une première étude de suivi de patients SEP, des algorithmes de « machine learning » permettant une classification binaire, ont été adaptés à différents types de données, telles que les paramètres métaboliques, les données cliniques (âge du patient, durée de la maladie, etc.) et la charge lésionnelle. Les résultats de la classification pour la discrimination des patients SEP des sujets contrôles n'étaient pas satisfaisants, bien que des différentes significatives soient observées pour ces différents paramètres entre les deux groupes. Les résultats de la classification pour la discrimination des différentes formes cliniques de la maladie, basée sur les paramètres de MRS uniquement étaient modérés, bien que l'ajout des données cliniques améliore considérablement ces résultats.

Une seconde étude a été réalisée pour extraire des paramètres de MRS de plus haut niveau, utilisant les réseaux de neurones conventionnels. Les résultats obtenus avec ces paramètres de MRS de haut niveau n'ont pas surpassé ceux obtenus avec des algorithmes de classification classiques entrainés sur des paramètres plus simples de MRS.

List of Abbreviations

AD	Axial Diffusivity
ADC	Apparent Diffusion Coefficient
AIF	Arterial Input Function
AK	Axial Kurtosis
ASL	Arterial Spin Labelling
BAR	Balanced Accuracy Rate
BBB	Blood Brain Barrier
BER	Balanced Error Rate
CBF	Cerebral Blood Flow
CBV	Cerebral Blood Volume
Cho	Choline
CIS	Clinically Isolated Syndrome
cMRI	Conventional Magnetic Resonance Imaging
CNN	Convolutional Neural Network
CNS	Central Nervous System
Cre	Creatine
CRLB	Cramer-Rao Lower Bound
CSF	Cerebro-Spinal Fluid
DCE-MRI	Dynamic Contrast Enhanced Magnetic Resonance Imaging
DKI	Diffusion Kurtosis Magnetic Resonance Imaging
DSC-MRI	Dynamic Susceptibility Contrast Magnetic Resonance Imaging
DTI	Diffusion Tensor Magnetic Resonance Imaging
DWI	Diffusion Weighted Magnetic Resonance Imaging
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
FLAIR	FLuid-Attenuated Inversion Recovery
FOV	Field Of View
GBM	Glioblastoma Multiforme

GE	Gradient Echo
Gln	Glutamine
Glu	Glutamate
Glx	Glutamine+glutamate
Gly	Glycine
Lac	Lactate
LDA	Linear Discriminant Analysis
Lip	Lipids
LOPOCV	Leave One Patient Out Cross Validation
MD	Mean Diffusivity
mI	myo-Inositol
MK	Mean Kurtosis
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRS	proton Magnetic Resonance Spectroscopy
MRSI	proton Magnetic Resonance Spectroscopic Imaging
MS	Multiple Sclerosis
MTT	Mean Transit Time
NAA	N-Acetyl-Aspartate
NAWM	Normal Appearing White Matter
OS	Overall Survival
PFS	Progression Free Survival
PP	Primary Progressive
PRESS	Point RESolved Spectroscopy
PRM	Parametric Response Map
PWI	Perfusion-Weighted Magnetic Resonance Imaging
RANO	Response Assessment in Neuro-Oncology
RD	Radial Diffusivity
RF	Random Forest
RK	Radial Kurtosis
ROI	Region Of Interest
RR	Relapsing Remitting
SE	Spin Echo
SNR	Signal to Noise Ratio
SP	Secondary Progressive
STEAM	STimulated Echo Acquisition Mode
SVM	Support Vector Machines
SVS	Single Voxel proton Magnetic Resonance Spectroscopy

T1pc	T_1-weighted Magnetic Resonance Imaging post contrast enhancing
TE	Echo Time
TI	Inversion Time
TNP	True Negative Rate
TPR	True Positive Rate
TR	Repetition Time
VOI	Volume Of Interest

Contents

Abstract v

List of Abbreviations xiii

Contents xv

List of Figures xxi

List of Tables xxvii

1 Introduction 1

1.1 Machine Learning 1

1.1.1 Support Vector Machines 2

1.1.2 Random Forests 4

1.1.3 Deep learning 4

1.1.4 Cross-Validation and Performance measures 10

1.2 Magnetic Resonance Imaging 12

1.2.1 Principles of MRI 12

1.2.2 Conventional MRI 16

1.2.3 Perfusion weighted MRI 19

1.2.4 Diffusion MRI 22

1.2.5 Magnetic Resonance Spectroscopic Imaging 25
1.3 Glioblastoma Multiforme 27
1.3.1 Glioblastoma Multiforme Overview 27
1.3.2 Advanced MRI in the post-operative GBM follow-up . . 28
1.3.3 UZ Leuven post-operative GBM dataset 30
1.4 Multiple Sclerosis 33
1.4.1 Multiple Sclerosis Overview 33
1.4.2 Advanced MRI in the longitudinal MS follow-up 35
1.4.3 AMSEP longitudinal dataset 37
1.5 Objectives of the thesis and main contributions 39
1.6 Outline of the thesis 40
1.7 Conclusion 41

2 Tumour relapse prediction using multi-parametric MR data recorded during follow-up of GBM patients 43

2.1 Introduction 44
2.2 Materials and Methods 45
2.2.1 Study setup 45
2.2.2 MRI acquisition and processing 45
2.2.3 Classifiers 48
2.2.4 In-house imputation method 50
2.2.5 Performance indices 51
2.3 Results and Discussion 52
2.3.1 Results 52
2.3.2 Discussion 54
2.4 Conclusions 57

3 Classifying glioblastoma multiforme follow-up progressive vs. responsive forms using multi-parametric MRI features 59

3.1 Introduction . . . 60

3.2 Materials and Methods . . . 61

3.2.1 Study setup . . . 61

3.2.2 MRI acquisition and processing . . . 62

3.2.3 Classifiers . . . 66

3.3 Results . . . 67

3.4 Discussion . . . 75

3.5 Conclusions . . . 77

4 Classification of Recurrent Glioblastoma using modified Parametric Response Maps of contrast-enhanced T1-weighted MRI and Perfusion MRI 79

4.1 Introduction . . . 80

4.2 Materials and Methods . . . 81

4.2.1 Patient population . . . 81

4.2.2 MRI acquisition and processing . . . 81

4.2.3 MRI Co-registration . . . 82

4.2.4 Feature extraction: Parameter Response Map . . . 83

4.2.5 Feature selection: Minimum Redundancy Maximum Relevance . . . 84

4.2.6 Classifiers . . . 86

4.2.7 Performance measures . . . 86

4.3 Results . . . 87

4.4 Discussion . . . 88

4.5 Conclusions . . . 91

5 Machine learning approach for classifying Multiple Sclerosis courses by combining clinical data with lesion loads and Magnetic Resonance metabolic features 93

5.1 Introduction 94

5.2 Materials and Methods 95

5.2.1 Patient population 95

5.2.2 Longitudinal MS data 95

5.2.3 MRI acquisition and processing 95

5.2.4 Feature extraction 96

5.2.5 Training approach 96

5.2.6 Performance measures and statistical testing 97

5.2.7 Classifiers 98

5.3 Results 98

5.4 Discussion 101

5.5 Conclusions 104

6 A comparison of Machine Learning approaches for classifying Multiple Sclerosis courses using MRSI and brain segmentations 107

6.1 Introduction 108

6.2 Materials and Methods 108

6.2.1 Patient population 108

6.2.2 Magnetic Resonance data acquisition and processing . . 109

6.2.3 Classification tasks and performance measures 109

6.2.4 Feature extraction models 110

6.2.5 Classifiers 111

6.3 Results and Discussion 112

6.4 Conclusions 113

7 Conclusions 115

7.1 General conclusions 115
7.2 Future perspectives 117

A Appendix **119**

Bibliography **139**

Curriculum Vitae **165**

List of publications **167**

List of Figures

1.1 SVM: finding the best separation plane 2

1.2 Random forests: majority voting. Figure adapted from [163]. . 5

1.3 Deep learning growth in the last years. Source: Nvidia website [107]. 5

1.4 Fully connected network with 2 hidden layers. Source: [164]. . . 6

1.5 Schematic representation of the neuron as it is used in neural networks. Source: [122]. 7

1.6 Activation functions: tanh (left) and ReLU (right). Source: [122]. 7

1.7 Convolution operation applied on an input image. Source: [87]. 8

1.8 Max-pooling. Source: [122]. 9

1.9 Dropout during a randomly selected training epoch of a fully connected neural network with 2 hidden layers. Source: [213]. . 9

1.10 Dropout during training. Source: [213]. 10

1.11 Precessing spins in external field $\mathbf{B_0}$ form the Magnetization vector $\mathbf{M}$. Figure adapted from [197]. 13

1.12 Radio-frequency pulse flips the magnetization in the transversal x-y space. Figure adapted from [197]. 16

1.13 Envelope of the FID signal in the transversal space. On the x-axis there is time, and on the y-axis there is the relative amplitude of the transversal FID signal's envelope. 17

1.14 Envelope of the recovered signal in the z direction. On the x-axis there is time, and on the y-axis there is the relative amplitude of the recovered FID signal's envelope in the z direction. 18

1.15 Details of a spin-echo sequence. Figure adapted from [126]. . . 19

1.16 Details of a gradient-echo sequence. Figure adapted from [126]. 20

1.17 Perfusion MRI CBV quantification after correcting for contrast agent leakage. Source: Cha et al., Radiology, 2002 [31] 21

1.18 Types of diffusion. Source: [154] 22

1.19 Pulsed Gradient Spin Echo diffusion weighted acquisition sequence. Figure adapted from [126]. 23

1.20 Diffusion tensor. Figure adapted from [126]. 24

1.21 MRS acquisition sequences: PRESS and STEAM, where MT is the Mixing Time. Image adapted from [13]. 26

1.22 Short and long TE MRS spectra (right and left columns) for a healthy subject and a patient (top and bottom rows) suffering from progressive multifocal leukoencephalopathy, scanned at 1.5 Tesla. Image adapted from [117]. 27

1.23 Conventional MRI of a post-operative GBM patient. 31

1.24 Post-operative GBM perfusion MRI parameter maps obtained using the DSCoMAN plugin [18]. 32

1.25 DKI parameter maps of a post-operative GBM patient. 33

1.26 Multiple Sclerosis global incidence. Source: [6]. 34

1.27 Multiple Sclerosis disease progression. Source: [74]. 35

1.28 MRSI grid (red) superimposed on T1pc of a Multiple Sclerosis patient. From left to right: coronal, sagittal, and axial view. . . 38

1.29 Thesis outline. 41

2.1 Brain tumour delineations on T1pc MRI. Green - necrosis, Red - contrast enhancing region of interest, Blue - edema. 46

3.1 Left - T1pc. Center - Manual delineations on top of T1pc. Right - Semi-manual delineations on top of T1pc. In red there is the contrast enhancing region (CER), while in blue it is the non-enhancing region (NER). 63

3.2 Example of co-registration results to T1pc for all multi-parametric magnetic resonance maps. 69

3.3 Rank estimates and confidence intervals for all combinations of classifiers, delineations, and MR modalities. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. The highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different than the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different than the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap. Each group has 10 BAR values, corresponding to 10 different features. 70

3.4 Rank estimates and confidence intervals for all combinations of delineations and MR modalities. CER - contrast enhancing region, NER - non-enhancing region. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. The highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different than the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different than the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap. Each group has 70 BAR values, corresponding to 10 features and 7 classifiers. 71

3.5 Rank estimates and confidence intervals for all combinations of delineations and classifiers. CER - contrast enhancing region, NER - non-enhancing region. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. The highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different than the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different than the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap. Each group has 40 BAR values, corresponding to 4 MR datasets and 10 features. 72

3.6 Rank estimates and confidence intervals for all combinations of delineations and varying number of features. CER - contrast enhancing region, NER - non-enhancing region. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. The highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different than the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different than the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap. Each group has 28 BAR values, corresponding to 4 MR datasets and 7 classifiers. 73

3.7 Maximum classification results over all MR modalities using 1 to 10 features. On y-axis are BAR values, and on x-axis the number of features used for classification. BAR - balanced accuracy rate, CER - contrast enhancing region, NER - non-enhancing region. 74

4.1 MRI Co-registration: on the top row there are baseline MRI maps, while on the bottom one there are MRI maps from the second time point. On both rows there are 5 columns, from left to right: (1) T1pc, (2) Reunion of total tumour ROIs from all time points (red) and NAWM ROI (blue) superimposed on T1pc, (3) FLAIR, (4) CBV, (5) CBF. 83

4.2 Comparison of T1pc-PRM$_+$ and CBV-PRM$_+$. First row, left corner: T1pc difference map between time point 2 and baseline, white/dark limits are +0.14 and -0.14. First row, right corner: T1pc-PRM$_+$ on top of the T1pc difference map. Second row, left corner: CBV difference map between time point 2 and baseline, white/dark limits are +1.3 and -0.7. Second row, right corner: CBV-PRM$_+$ on top of the CBV difference map. PRM$_+$ - positive parametric response map. 85

4.3 Area under the curve (AUC) values obtained by training SVM-lin and SVM-rbf using conventional and perfusion MRI (cpMRI) features extracted separately by the two positive parametric response maps, T1pc-PRM$_+$ and CBV-PRM$_+$. Training of the classifiers was done with an increasing number of features from 1 to 16, sorted using minimum-redundancy-maximum-relevance (mRMR). 87

4.4 Area under the curve (AUC) values obtained by training SVM-lin and SVM-rbf using only conventional MRI (cMRI) features extracted separately by the two positive parametric response maps, T1pc-PRM$_+$ and CBV-PRM$_+$. Training of the classifiers was done with an increasing number of features from 1 to 8, sorted using minimum-redundancy-maximum-relevance (mRMR). 88

4.5 Area under the curve (AUC) values obtained by training SVM-lin and SVM-rbf using only perfusion MRI (PWI) features extracted separately by the two positive parametric response maps, T1pc-PRM$_+$ and CBV-PRM$_+$. Training of the classifiers was done with an increasing number of features from 1 to 8, sorted using minimum-redundancy-maximum-relevance (mRMR). 89

5.1 Box-plots of magnetic resonance metabolic features and lesion loads extracted from healthy controls (HC) and multiple sclerosis (MS) patients: A. NAA/Cho; B. NAA/Cre; C. Cho/Cre; D. Lesion load (LL). The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 99

A.1 Classification results on CER & NER semi-manual delineations, using 1 to 10 features assigned by rank products per each dataset. On y-axis are BAR values, and on x-axis the number of features used for classification. CER - contrast enhancing region, NER - non-enhancing region. 128

A.2 Classification results on Total manual delineations, using 1 to 10 features assigned by rank products per each dataset. On y-axis are BAR values, and on x-axis the number of features used for classification. 129

A.3 Healthy Controls (HC) vs. Multiple Sclerosis (MS) groups in 2-D feature space: x-axis is NAA/Cho and y-axis is NAA/Cre. The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 132

A.4 Comparison of Multiple Sclerosis (MS) groups in 2-D feature space: x-axis is NAA/Cho and y-axis is NAA/Cre. The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 133

A.5 Comparison of Multiple Sclerosis (MS) groups in 2-D feature space: x-axis is disease age and y-axis is Cho/Cre. The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 134

A.6 Comparison of Multiple Sclerosis (MS) groups in 2-D feature space: x-axis is lesion load and y-axis is EDSS. The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 135

List of Tables

1.1 General binary confusion matrix. 10

1.2 Typical relaxation times and water percentages of the most important brain tissues at 1.5 Tesla. 16

1.3 Patient population: Age - average value (standard deviation); Disease duration - average value (standard deviation); EDSS - median (minimum - maximum); Lesion Load - average value (standard deviation). 37

2.1 Supervised and semi-supervised classifiers tested in this chapter. 49

2.2 Detailed BER results for each time point when training the best 6 classifiers on complete features for all MR modalities. The decision (i.e. labelling) moment 'L' is highlighted. Some time points do not have results because there were no complete measurements. 53

2.3 Detailed BER results for each time point for the best 6 classifiers when trained on imputed data for all MR modalities. The decision (i.e. labelling) moment 'L' is highlighted. 54

2.4 Weighted BER for the best 6 supervised classifiers trained on complete data for each MR modality separately. PWI and DKI features were extracted from both CE and ED ROI. MRSI features were extracted only from CE voxels. 54

2.5 Weighted BER for the best classifiers trained on imputed features from each MR modality separately. PWI and DKI features were extracted from both CE and ED ROI. MRSI features were extracted only from CE voxels. 55

2.6 Weighted BER comparison between our in-house method of imputing missing values and built-in imputation strategy of different supervised classifiers. 55

5.1 Patient population: Age - average value (standard deviation); Disease duration - average value (standard deviation); EDSS - median (minimum - maximum); Lesion Load - average value (standard deviation). The four multiple sclerosis (MS) groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 95

5.2 Adjusted p-values for multiple comparisons between multiple sclerosis (MS) groups modelled by linear mixed effects model, tested using the "multcomp" package in 'R' (* for $p < 0.05$ and ** for $p < 0.01$). The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 99

5.3 F1-scores for all nine classification tasks (rows) after training LDA using only metabolic ratios. Values above 75 are coloured in light gray. HC - healthy controls, CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 100

5.4 F1-scores for classification tasks (columns) involving only multiple sclerosis (MS) patients. Abbreviations: M = all three average metabolic ratios; Age = patient age; DD = disease duration; LL = lesion load; EDSS = Expanded Disability Status Scale. Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, while values larger than 85 are coloured in dark gray. CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 101

6.1 Multiple sclerosis (MS) patient population details. CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 108

6.2 Multiple sclerosis (MS) metabolite ratios - mean (standard deviation). CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 110

6.3 Area under the curve (AUC), Sensitivity, and Specificity values for all classifiers, feature extraction models (M1-M4), and classification tasks. Dimensionality of the models: M1 - 81 (metabolic spectra), M2 - 3 (metabolic features), M3 - 6 (3 metabolic and 3 tissue percentages), M4 - CNN - input image is 128×57. CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive. 112

A.1 Weighted BER for supervised and semi-supervised classifiers trained on complete and imputed data. We highlight the best 6 classifiers. 120

A.2 Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on complete perfusion features. The decision (i.e. labelling) moment 'L' is highlighted. Some time points do not have results because there were no complete perfusion measurements. 121

A.3 Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on complete diffusion features. The decision (i.e. labelling) moment 'L' is highlighted. Some time points do not have results because there were no complete diffusion measurements. 122

A.4 Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on complete spectroscopy features. The decision (i.e. labelling) moment 'L' is highlighted. Some time points do not have results because there were no complete spectroscopy measurements. . . 123

A.5 Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on imputed perfusion features. The decision (i.e. labelling) moment 'L' is highlighted. 123

A.6 Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on imputed diffusion features. The decision (i.e. labelling) moment 'L' is highlighted. 124

A.7 Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on imputed spectroscopy features. The decision (i.e. labelling) moment 'L' is highlighted. 124

A.8 Number of data points acquired at each time point. The decision (i.e. labelling) moment 'L' is highlighted. 125

A.9 Number of features per MRI modality and delineation. 125

A.10 Supervised classifiers used in Chapter 3 and their software implementations. 125

A.11 Maximum BAR of all MR modalities over all classifiers. CER - contrast enhancing region, NER - non-enhancing region. 126

A.12 Top 10 selected features according to rank products for each dataset 127

A.13 Balanced accurary rates (BAR), sensitivity (TPR), and specificity (TNR) values, for all 9 classification tasks (rows) after training LDA using only metabolic ratios. Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, values between 85 and 89 are coloured in dark gray, while values higher than 90 are coloured in very dark gray. 130

A.14 BAR values for classification tasks involving only MS patients (columns). Abbreviations: M = all three average metabolic ratios; Age = patient age; DD = disease duration; LL = lesion load; EDSS = Expanded Disability Status Scale.Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, values between 85 and 89 are coloured in dark gray, while values higher than or equal to 90 are coloured in very dark gray. 130

A.15 Sensitivity values for classification tasks involving only MS patients (columns). Abbreviations: M = all three average metabolic ratios; Age = patient age; DD = disease duration; LL = lesion load; EDSS = Expanded Disability Status Scale. Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, values between 85 and 89 are coloured in dark gray, while values higher than or equal to 90 are coloured in very dark gray. 131

A.16 Specificity values for classification tasks involving only MS patients (columns). Abbreviations: M = all three average metabolic ratios; Age = patient age; DD = disease duration; LL = lesion load; EDSS = Expanded Disability Status Scale. Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, values between 85 and 89 are coloured in dark gray, while values higher than or equal to 90 are coloured in very dark gray. 131

A.17 Performance measures computed with SVM-lin and SVM-rbf trained on an increasing number of features from 1 to 16 for Conventional and Perfusion MRI (cpMRI), extracted using the two positive parametric response maps (PRM_+), T1pc-PRM_+ and CBV-PRM_+. Values over 90% are highlighted in gray. . . 132

A.18 Performance measures computed with SVM-lin and SVM-rbf trained on an increasing number of features from 1 to 8 for cMRI and PWI separately, features extracted using the two positive parametric response maps (PRM_+), T1pc-PRM_+ and CBV-PRM_+. Values over 90% are highlighted in gray. 133

A.19 Conventional and Perfusion MRI (cpMRI) features selected with minimum-redundancy-maximum-relevance (mRMR) after applying separately the two positive parametric response maps (PRM_+), T1pc-PRM_+ and CBV-PRM_+, where “F" stands for FLAIR. Features are ‘X’-percentile, where ‘X’ can be 50, 70, 90, and 99. 136

A.20 Conventional MRI features selected with minimum-redundancy-maximum-relevance (mRMR) after applying separately the two positive parametric response maps (PRM_+), T1pc-PRM_+ and CBV-PRM_+. Features are ‘X’-percentile, where ‘X’ can be 50, 70, 90, and 99. 136

A.21 Perfusion MRI features selected with minimum-redundancy-maximum-relevance (mRMR) after applying separately the two positive parametric response maps (PRM_+), T1pc-PRM_+ and CBV-PRM_+. Features are ‘X’-percentile, where ‘X’ can be 50, 70, 90, and 99. 137

Chapter 1

Introduction

1.1 Machine Learning

Machine learning is an area of computer science which has been constantly growing in popularity during the last 30 years. It is being used in a very wide range of applications, like search engines, computer vision, anti-spam software, financial market analysis, bioinformatics, astronomy, and many more. Its main purpose is to find meaningful patterns in data, therefore it is a very interesting field to explore, especially now when there is an explosion of data in the world. According to [84], the volume of all digital information in 2020 will have grown 300 times compared to 2005, up to approximately 40 trillion gigabytes. For an extended analysis of machine learning methods the reader is referred to Hastie et al. [78], one of the many open-access books available online.

This work will focus only on supervised machine learning for binary classification tasks, meaning that there are only two labels (e.g. 0 and 1, negative and positive) that have to be differentiated. The general problem statement in supervised learning is: given a training set with $\mathbf{N}$ labelled data points having the form $\{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_N, \mathbf{y}_N)\}$, find the function $h : \mathbf{X} \to \mathbf{Y}$ such that the predicted labels $\hat{\mathbf{y}}_i = h(\mathbf{x}_i)$ ideally match the real labels, where $\mathbf{x}_i$ is the feature vector of the i-th training example, $\mathbf{y}_i$ is its label, $\mathbf{X} = \mathbb{R}^{\mathbf{d}}$ is the feature space, and $\mathbf{Y} = \{-\mathbf{1}, \mathbf{1}\}$ is the label space. After the training stage is complete, the learned function h is verified against a test set independent from the training set, to see how it performs against unseen examples. Three main supervised classification methods are further discussed: Support Vector Machines (SVM), which is probably the most widely used classifier in the last 20 years, Random Forests

(RF), one of the best off-the-shelf classifier, and a brief introduction into the current growing paradigm of deep learning.

1.1.1 Support Vector Machines

The original SVM algorithm was invented by Cortes and Vapnik in 1995 [39], and is based on the assumption that there exists an optimal separation plane between data points belonging to different classes. Since then, many variants have been proposed, one of the most important ones being least squares support vector machines (LSSVM) [222]. A simple 2-D graphical representation of data points from two classes is shown in Figure 1.1. SVM will find the best separation line and its associated margin marked by the two lines parallel to the separation line. The support vectors are highlighted with a green border and lie within the margin. Mathematically, the SVM binary classifier is a maximum-margin

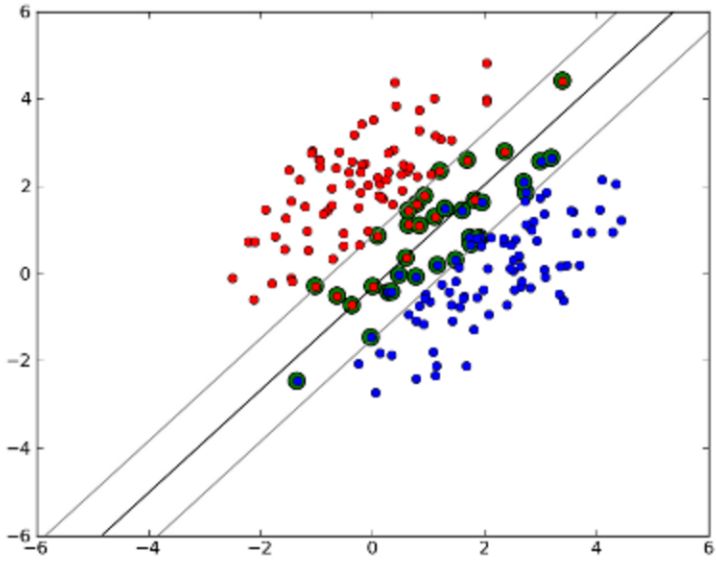


Figure 1.1: SVM: finding the best separation plane

linear model of the form:

$$h(\mathbf{x}) = \begin{cases} 1 & \text{if } \ b + \sum_{j=1}^{d} x_j w_j > 0 \\ -1 & \text{otherwise} \end{cases} \tag{1.1}$$

where $\mathbf{w} = [w_1, \ldots, w_d]$ is the vector containing all weights, and b is the intercept term. These two terms, $\mathbf{w}$ and b, define the function h. The purpose of SVM is to find these two terms, given the training data set $\{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_N, \mathbf{y}_N)\}$. For the non-separating case, the learning phase is done by solving the following primal optimization problem:

$$\begin{aligned} \min_{\mathbf{w},\xi_i} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\ \textit{subject to} \quad & y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1 \ldots N \end{aligned} \tag{1.2}$$

where ξ_i are called slack variables, and C is a hyper-parameter that controls the degree of misclassification. The optimization problem in its Lagrangian dual form is:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \textit{subject to} \quad & C \geq \alpha_i \geq 0, \quad i = 1 \ldots N \\ & \sum_{i=1}^{N}\alpha_i y_i = 0 \end{aligned} \tag{1.3}$$

Having solved the dual optimization problem, namely finding the α_i, the weights and intercept are computed by:

$$\begin{aligned} \hat{\mathbf{w}} &= \sum_{i=1}^{N}\alpha_i y_i \mathbf{x}_i \\ b &= \hat{\mathbf{w}}\cdot\mathbf{x}_k - y_k, \quad \text{for} \quad \alpha_k \geq 0 \end{aligned} \tag{1.4}$$

The SVM methodology can easily be extended to non-linear classification tasks. By replacing each data vector $\mathbf{x}_i$ with a non-linear mapping $\phi(\mathbf{x}_i)$, and defining a kernel as $K(\mathbf{x}, \mathbf{z}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{z})$, the optimization problem in its Lagrangian dual form can be rewritten as:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \textit{subject to} \quad & C \geq \alpha_i \geq 0, \quad i = 1 \ldots N \\ & \sum_{i=1}^{N}\alpha_i y_i = 0 \end{aligned} \tag{1.5}$$

One of the most commonly used kernels in the literature is the gaussian or radial basis function kernel, defined as:

$$K(\mathbf{x}, \mathbf{z}) = \exp(-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2}) \tag{1.6}$$

1.1.2 Random Forests

The original RF algorithm was invented by Breimann in 2001 [21], and is based on the assumption that a collection of weak classifiers outperforms a single weak classifier, namely a weak decision tree [22].

Decision trees are very attractive classifiers to use because they can handle heterogeneous data (ordered, categorical, or a mix of both), they intrinsically implement variable selection, they are robust to outliers, and most importantly, they are easily interpretable. However, it can be proved that they suffer from the "high variance" problem, meaning they risk to overfit the training data. Breimann solved this problem, by combining bagging [20] with random variable selection at each node.

Bagging stands for **b**ootstrap **agg**regating, meaning each individual decision tree of the forest will learn a different classification model based on a bootstrap sample of the original training set. One bootstrap sample has approximately 63% of the original training data points, sampled with replacement, while the remaining 37% form the out-of-bag data. Each time a tree is added to the forest, the out-of-bag data is used as internal validation data for estimating classification error and variable importance.

For classification tasks, each tree will learn a model only on $\sqrt{d}$ variables, where d is the dimension of any data point.

It is widely recognized that random forests are an excellent off-the-shelf machine learning algorithm, with a great overview given by Louppe [145].

A simplified graphical explanation is provided in Figure 1.2.

1.1.3 Deep learning

Deep learning is the hottest topic of machine learning, with an exponential growth in the last 5 years, which can be targeted at improving every field of our daily lives, from healthcare to finance and to space exploration, as seen in Figure 1.3. It is widely known that huge companies such as Google, Microsoft, Facebook, Amazon, Instagram, Baidu, IBM, Tesla, and many more, either use

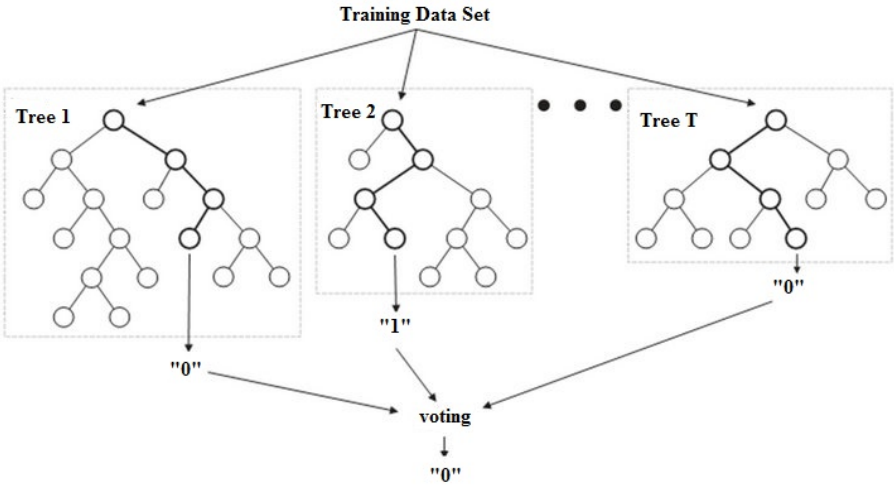


Figure 1.2: Random forests: majority voting. Figure adapted from [163].

Nvidia's graphical processing units in their programs or they are building their own patented deep learning chips.

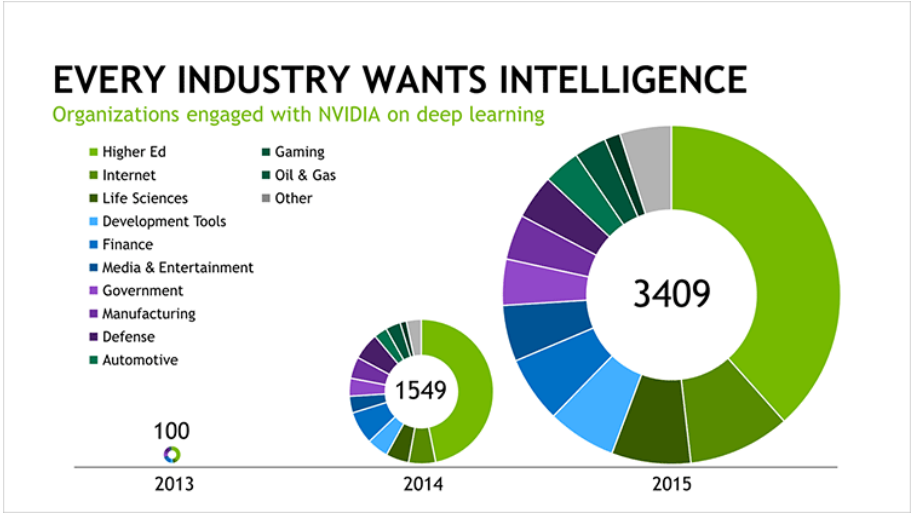


Figure 1.3: Deep learning growth in the last years. Source: Nvidia website [107].

The most influential paper is the one by Krizhevsky et al. [131] from 2012, where they first describe deep convolutional networks, in the most difficult

visual recognition challenge, ImageNet Large-Scale Visual Recognition Challenge. Currently, deep learning models are capable of impressive feats, as they perform better than humans at object recognition and classification [97, 113], speech recognition [99], and have recently beaten the European number 1 master in the game of Go [206], and afterwards also the world number 1, an achievement previously thought to be at least a decade away. After this success, Google and Facebook announced that they will focus also on mastering the popular computer game Starcraft (Blizzard Entertainment), which will give valuable insights into real world adversarial situations.

Considering the huge amount of work on deep learning, it would be impossible to properly make an introduction within this thesis. Instead, we will only give an overview of each layer of the Convolutional Neural Network (CNN) used in Chapter 6, and for detailed mathematical formulation we will refer the reader to the "Deep Learning" book of Goodfellow, Bengio, and Courville [87].

Most CNNs architectures are usually made of the following types of layers: convolutional (conv), pooling (either maximum or average), dropout, fully-connected (FC), and activation (e.g. rectified linear unit (ReLU)). The most important layers are the convolutional and the fully connected, called weight layers. The paper from Krizhevsky only had 8 weight layers (5 conv and 3 FC), while modern architectures have gone up to 152 [98].

Layers are usually connected sequentially, starting from an input layer to an output layer, while all layers in-between are called "hidden". An example of a neural network with 2 fully connected hidden layers is given in Figure 1.4. The

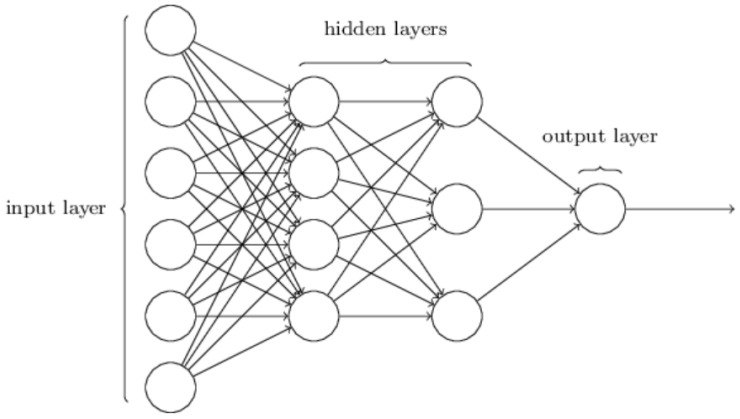


Figure 1.4: Fully connected network with 2 hidden layers. Source: [164].

input layer can be of any dimension, either a signal, a 2-D gray image, a colour

image with three channels (e.g. RGB), or even more. However, CNNs are built especially for images, both colour and gray, so we will focus on them.

The building block of the original neural networks was the neuron, depicted in Figure 1.5. A neuron takes as input a number of points from the previous layer,

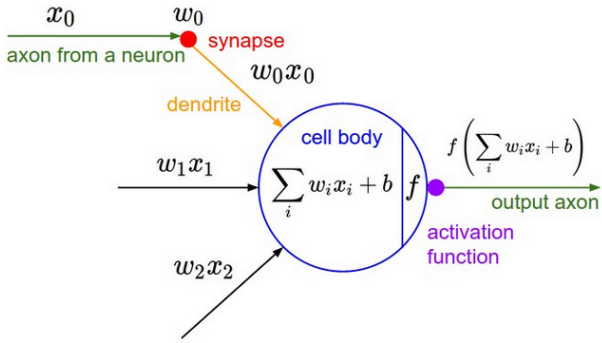


Figure 1.5: Schematic representation of the neuron as it is used in neural networks. Source: [122].

which are multiplied by individual weights, and finally performs an activation. During training, weights are iteratively updated following an optimisation procedure. In CNNs two main activation functions are used, hyperbolic tangent (tanh) and ReLU, as shown in Figure 1.6. However, ReLU was found to be 6 times faster than tanh [131].

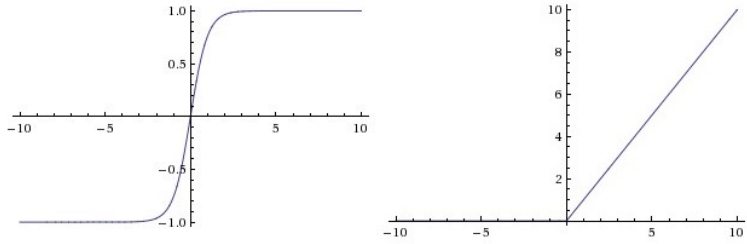


Figure 1.6: Activation functions: tanh (left) and ReLU (right). Source: [122].

Convolutional layers apply the convolution operation to the input, passing the result to the next layer. Mathematically, the convolution operation is noted as $*$ and is described by

$$(f * g)(t) = \int_{-\infty}^{+\infty} f(\tau) \cdot g(t-\tau) d\tau = \int_{-\infty}^{+\infty} f(t-\tau) \cdot g(\tau) d\tau \qquad (1.7)$$

Graphically, a convolutional kernel slides through the image and simple operations are performed, as seen in Figure 1.7. The amount of pixels that

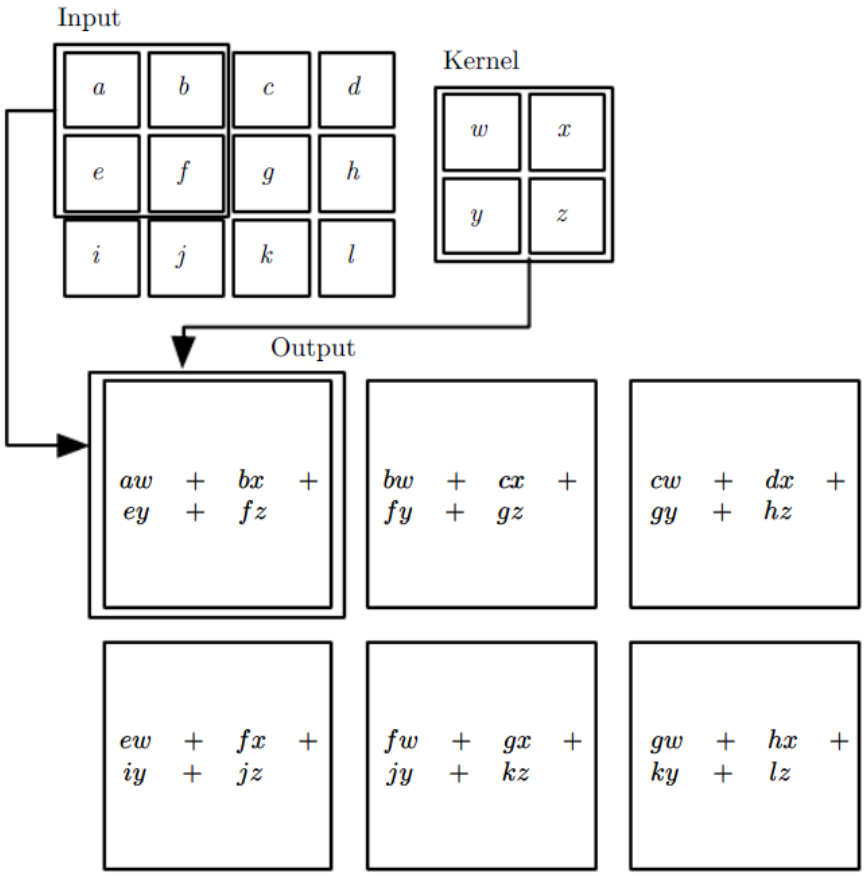


Figure 1.7: Convolution operation applied on an input image. Source: [87].

the convolution filter (or kernel) slides between two operations is called stride, and is equal to 1 in our graphical example. Convolutional layers usually have multiple kernels designed to detect specific image features (e.g. vertical edges), and are typically followed by ReLU activation layers, previously described, and maximum pooling (MP) layers, as described in Figure 1.8. Because of the large amount of parameters that are optimised for specific data, CNN overfitting is very likely, even if the learning data is split into training and validation. To prevent it from happening, Srivastava et al. [213] introduced the dropout layer which is especially useful between fully connected layers. During the

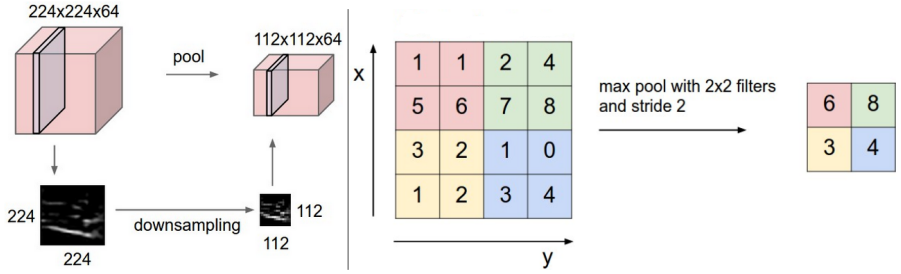
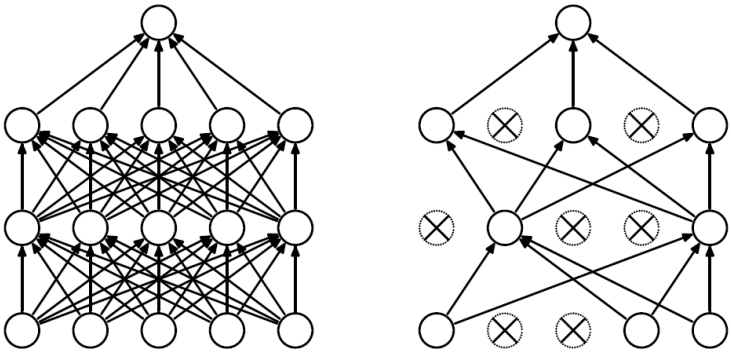


Figure 1.8: Max-pooling. Source: [122].

training phase, incoming and outgoing connections to a dropped-out neuron are randomly removed with a probability 1-p, as shown in Figure 1.9. During

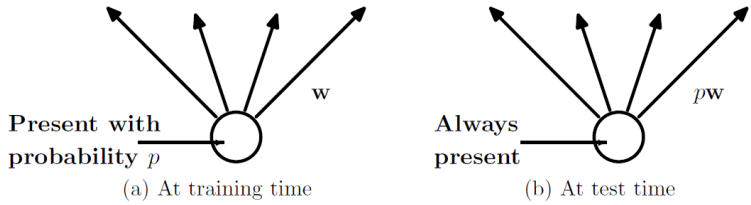


(a) Standard Neural Net (b) After applying dropout.

Figure 1.9: Dropout during a randomly selected training epoch of a fully connected neural network with 2 hidden layers. Source: [213].

test phase, all neurons are present and their weights are multiplied by their specific probability to be present in the network during training, as shown in Figure 1.10.

A typical CNN architecture has a few main building blocks of [conv-ReLU-MP] with convolutional kernels of size 5×5 or 7×7, followed by a few FC layers, and a final activation layer, usually tanh. Simonyan and Zisserman [207] have shown the benefits of modifying the building block by adding extra convolutional and ReLU layers, [conv-ReLU-conv-ReLU-MP], but using only small convolutional kernels of size 3×3. The CNN architecture that we built in Chapter 6 is inspired by their work.



(a) At training time (b) At test time

Figure 1.10: Dropout during training. Source: [213].

1.1.4 Cross-Validation and Performance measures

In order to quantify the quality of a machine learning model, different performance measures can be computed using the predicted labels. Because the focus of this thesis is on biomedical applications, a leave-one-patient-out cross-validation (LOPOCV) scheme was mostly used, except for Chapter 6 when a 2-fold cross-validation scheme was used for testing CNNs. In a LOPO-CV scheme, data points from one patient are assigned to the test set, while data points from the rest of the patients are assigned to the training set. In this way it is made sure that the test set is always independent from the training set. Patients are assigned one by one to the test set and all predicted labels are stored. The comparison between predicted labels and real labels is done at the end, after each patient was tested once.

Multiple performance measures can be computed based on the confusion matrix, which is presented in Table 1.1 in a general way for a binary classification task. All measures described below can take values between 0 and 1.

Confusion matrix		predicted condition	
		negative	positive
true condition	negative	True Negative (TN)	False Positive (FP)
	positive	False Negative (FN)	True Positive (TP)

Table 1.1: General binary confusion matrix.

Sensitivity

Sensitivity, also called recall or true positive rate (TPR), measures the amount of positives recognized as such from the total amount of positives:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1.8}$$

Specificity

Specificity, also called true negative rate (TNP), measures the amount of negatives recognized as such from the total amount of negatives:

$$Specificity = \frac{TN}{TN + FP} \tag{1.9}$$

Precision

Precision measures the amount of positives recognized as such from the total amount of predicted positives:

$$Precision = \frac{TP}{TP + FP} \tag{1.10}$$

Balanced accuracy rate

Balanced accuracy rate (BAR) is the average between sensitivity and specificity:

$$BAR = \frac{Sensitivity + Specificity}{2} \tag{1.11}$$

Balanced error rate

Balanced error rate (BER) is defined as:

$$BER = 1 - BAR \tag{1.12}$$

F_1 score

F_1 score is the harmonic mean between precision and recall, and can be reduced to:

$$F_1 = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{1.13}$$

1.2 Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is primarily a medical investigation technique based on the theory of nuclear magnetic resonance, which states that certain atomic nuclei can absorb (and emit) radio frequency energy when placed in an external magnetic field. The first MRI scan of a human patient was done in 1977, and since then, the number of MRI scans in developed countries has increased up to an average of 52/1000 inhabitants/year [110]. MRI has a wide range of applications, especially in medical diagnosis, as approximately 36000 scanners are estimated to be in use worldwide [185]. Because MRI scanners use only magnetic fields, they are much safer than any other scanners using ionizing radiation, such as Computed Tomography (CT), Positron Emission Tomography (PET), or Single-Photon Emission Computed Tomography (SPECT). MRI is the tool of choice for neuroimaging, as it provides high resolution 3-D brain and neck images that offer great contrast of soft tissues (i.e. gray and white matter).

1.2.1 Principles of MRI

In quantum mechanics, nuclei with an even atomic number, such as hydrogen (^{1}H), carbon (^{13}C), fluor (^{19}F), and phosphorus (^{31}P), possess a spin, or angular momentum. Due to the large amount of protons in the human body, ^{1}H-MRI is the most commonly used and developed technique, and it will be the unique subject throughout this thesis. Because protons are moving electrical charges, they also have magnetic moments. Therefore, when protons are put in an external magnetic field $\mathbf{B_0}$, they align along it either in a parallel (spin $+\frac{1}{2}$) or anti-parallel state (spin $-\frac{1}{2}$), as seen in Figure 1.11. The resulting magnetic vector $\mathbf{M}$ is called Magnetization, is aligned parallel to $\mathbf{B_0}$ and its amplitude depends on the density of protons. One of the most important formula in magnetic resonance (MR) is the Larmor equation, which shows that the precession (or Larmor) frequency ω of any proton magnetic moment is proportional to the external magnetic field B_0 and the gyromagnetic ration γ, which for protons is equal to $2.675 \cdot 10^8 rad \cdot s^{-1} \cdot T^{-1}$. In this thesis, the external magnetic field B_0 is either 1.5 Tesla (used in scanning multiple sclerosis (MS) patients) or 3 Tesla (used in scanning glioblastoma multiforme (GBM)

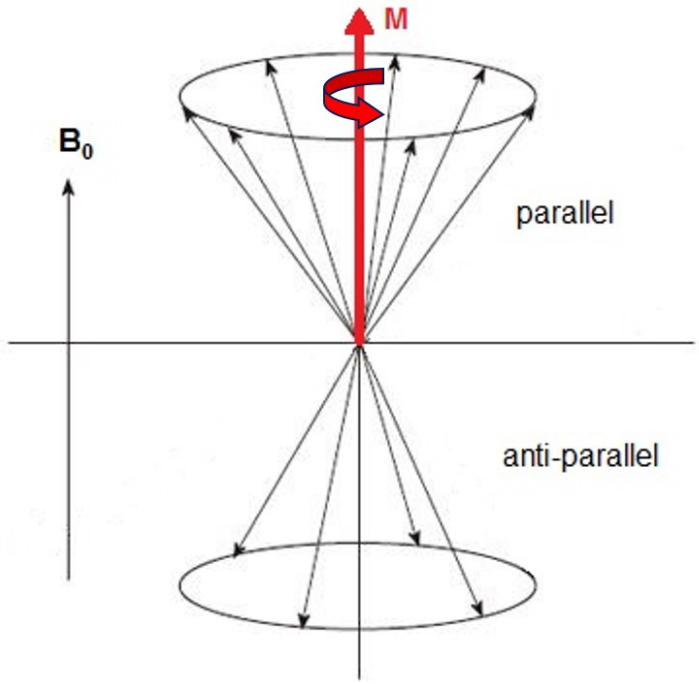


Figure 1.11: Precessing spins in external field $\mathbf{B_0}$ form the Magnetization vector **M**. Figure adapted from [197].

patients).

$$\begin{aligned} \omega_0 =& \gamma \cdot B_0 \\ & \text{or} \\ \nu_0 =& \frac{\gamma}{2\pi} \cdot B_0 \end{aligned} \tag{1.14}$$

The energy difference between the two states is:

$$\Delta E = h\nu_0 = \frac{h\gamma}{2\pi} \cdot B_0 \tag{1.15}$$

The ratio between the two states can be computed using the Boltzmann equation:

$$\frac{n_{+\frac{1}{2}}}{n_{-\frac{1}{2}}} = \exp(\frac{\Delta E}{kT}) = \exp(\frac{h\gamma B_0}{2\pi kT}), \tag{1.16}$$

where k is the Boltzmann constant, T is the temperature in Kelvin, and kT is the thermal energy. For an external magnetic field of 1.5 Tesla, at a body temperature of 37 degree Celsius or 310.15 Kelvin, the ratio will be:

$$\frac{n_{+\frac{1}{2}}}{n_{-\frac{1}{2}}} = 1.00000988191692 = \alpha. \tag{1.17}$$

This means that for every million protons in the anti-parallel state, there are approximately 10 more protons in the parallel state. To better understand this number, water will be considered as soft tissue, and an approximate number of excess protons will be computed. Using common chemical formulations, we know that

$$\begin{gathered}1 \text{ mol of water} \ldots 18 \text{ grams} \ldots 2 \text{ grams of } {}^1H \ldots 6.023 \cdot 10^{23} \text{molecules} \\ \text{leads to} \\ 1 \text{ gram of water} \ldots 0.11 \text{ grams of } {}^1H \ldots 3.346 \cdot 10^{22} \text{molecules}\end{gathered} \tag{1.18}$$

Therefore, in 1 gram of water there are approximately 3.34×10^{22} water molecules, or a total amount N_t of protons of 6.68×10^{22}. This leads to an excess of

$$n_{+\frac{1}{2}} - n_{-\frac{1}{2}} = \frac{\alpha - 1}{\alpha + 1} \cdot N_t = 3.3 \cdot 10^{17} \tag{1.19}$$

protons contributing to a maximum net magnetic dipole $\mathbf{m}$ of amplitude

$$m_0 = \frac{\gamma^2 h^2 B_0 N_t}{16\pi^2 kT} = 4.65 \cdot 10^{-9} Am^2 \tag{1.20}$$

and after considering water density we obtain the magnetization, $\mathbf{M} = \mathbf{m}/\mathbf{V}$, with maximum amplitude

$$M_0 = 4.65 \cdot 10^{-3} A/m. \tag{1.21}$$

The huge number of protons contributing to the magnetization help us understand that the microscopic quantum effects sum up and produce macroscopic effects, therefore facilitating the transition between quantum theory and classical electro-magnetic theory. The fundamental Bloch equation behind MRI phenomenology have been mathematically proven using both quantum and classical theory:

$$\frac{d\mathbf{M}(t)}{dt} = \gamma \mathbf{M}(t) \times \mathbf{B}(t) \tag{1.22}$$

If we introduce relaxation terms T_1 for spin-lattice interactions and T_2 for spin-spin interactions, and we consider that the external magnetic vector $\mathbf{B}$

typically has only a z-component $B_z = B_0$, while $B_x = B_y = 0$ the Bloch equations can be re-written

$$\begin{aligned}
\frac{dM_x(t)}{dt} &= \gamma(\mathbf{M}(t) \times \mathbf{B}(t))_x - \frac{M_x(t)}{T_2} \\
\frac{dM_y(t)}{dt} &= \gamma(\mathbf{M}(t) \times \mathbf{B}(t))_y - \frac{M_y(t)}{T_2} \\
\frac{dM_z(t)}{dt} &= \gamma(\mathbf{M}(t) \times \mathbf{B}(t))_z - \frac{M_z(t) - M_0}{T_1}
\end{aligned} \tag{1.23}$$

with solutions in the x-y space

$$\begin{aligned}
M_x(t) =& [M_x(0)\cos(\gamma Bt) - [M_y(0)\sin(\gamma Bt)]\exp(-\frac{t}{T_2}) \\
M_y(t) =& [-M_x(0)\sin(\gamma Bt) + [M_y(0)\cos(\gamma Bt)]\exp(-\frac{t}{T_2})
\end{aligned} \tag{1.24}$$

and in the z-direction

$$M_z(t) = M_0 + (M_z(0) - M_0)\exp(-\frac{t}{T_1}). \tag{1.25}$$

By applying a radio-frequency pulse $\mathbf{B}_1$ along the x-direction, the magnetization $\mathbf{M}$ will flip along the y-direction, as shown in Figure 1.12. In this case, $M_z(0)$ and $M_x(0)$ will be equal to 0, and the Bloch solutions will be

$$\begin{aligned}
M_x(t) =& - M_0\sin(\gamma Bt)\exp(-\frac{t}{T_2}) \\
M_y(t) =& + M_0\cos(\gamma Bt)\exp(-\frac{t}{T_2}) \\
M_z(t) =& M_0(1 - \exp(-\frac{t}{T_1})).
\end{aligned} \tag{1.26}$$

When the radio-frequency pulse is switched off, the magnetization will oscillate to equilibrium along B_0, and two types of signals can be observed: a recovery signal along z and free induction decay (FID) signals in the transversal x-y space. Using proper coils, real in-vivo T_1 and T_2 can be measured, as shown in Figure 1.13 and Figure 1.14. Because the magnetic field is not perfectly homogeneous in real-life applications, it will affect the relaxation time in the transversal space such that the measured time will be T_2^*, where

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{inhomogen}} \tag{1.27}$$

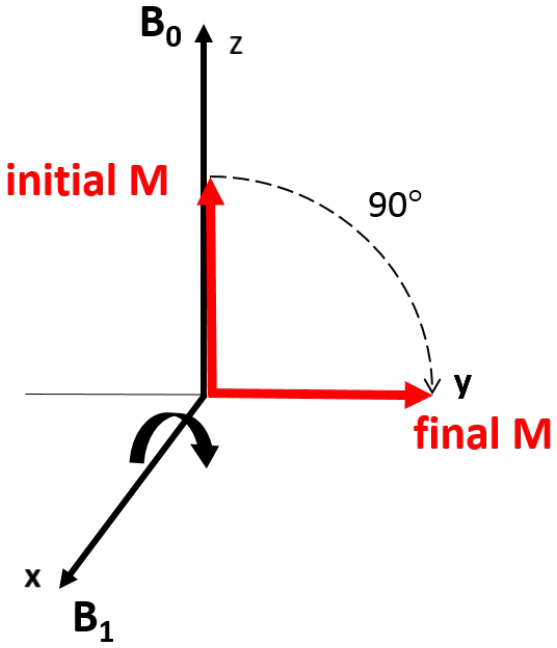


Figure 1.12: Radio-frequency pulse flips the magnetization in the transversal x-y space. Figure adapted from [197].

1.2.2 Conventional MRI

For neurological applications, MRI is the standard imaging technique used around the world. It has a high resolution and provides excellent contrast between different brain tissues. This is possible because different brain tissues have different relaxation times, as shown in Table 1.2.

Brain tissue	T_1 (ms)	T_2 (ms)	$\%H_2O$
Cerebro-Spinal Fluid	4000	2000	97
Gray Matter	900	100	80
White Matter	780	90	72
Fat	260	80	0

Table 1.2: Typical relaxation times and water percentages of the most important brain tissues at 1.5 Tesla.

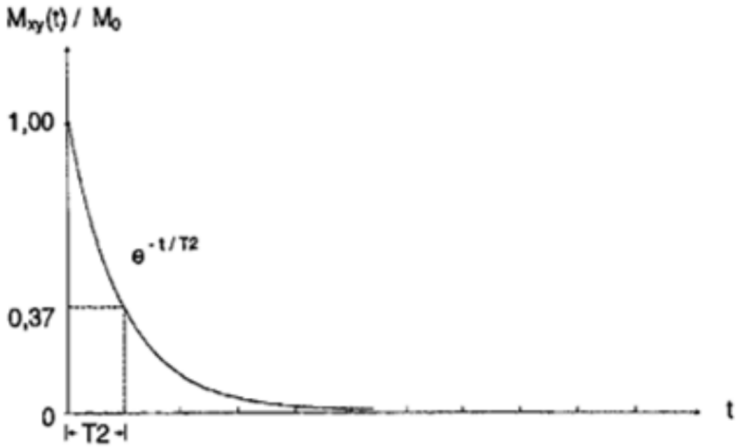


Figure 1.13: Envelope of the FID signal in the transversal space. On the x-axis there is time, and on the y-axis there is the relative amplitude of the transversal FID signal's envelope.

MRI sequences are characterized by two main parameters: echo time (TE) and repetition time (TR). The echo time is the duration between the radio-frequency pulse application and the acquisition of the main (echo) MR signal, and its optimization allows imaging the T_2 contrast between tissues. The repetition time is the duration between two successive radio-frequency pulses, and its optimization allows imaging the T_1 contrast between tissues.

T_2-weighted MRI

MRI sequences that use the 90° flip of the magnetization in the transversal x-y space are called spin-echo (SE). A classical SE sequence is shown in Figure 1.15, and is mainly characterized by long TE and long TR. The second radio-frequency pulse is called the re-focusing 180° pulse and it corrects for any magnetic field inhomogeneities, therefore SE MRI acquisitions are mainly used for detecting T_2 contrast between tissues.

T_1-weighted MRI

For T_1 acquisitions, another type of MR sequence is mainly used, called gradient echo (GE). A classical GE sequence is shown in Figure 1.16, and is mainly characterized by short TE and short TR. This type of sequences do not uses

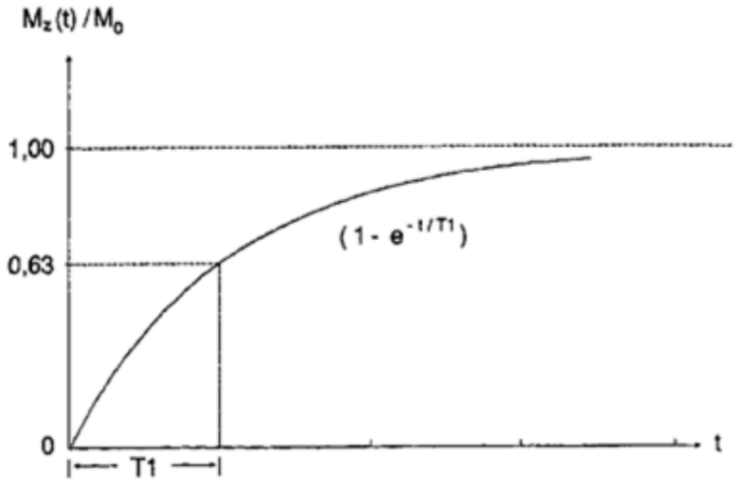


Figure 1.14: Envelope of the recovered signal in the z direction. On the x-axis there is time, and on the y-axis there is the relative amplitude of the recovered FID signal's envelope in the z direction.

any re-focusing 180° radio-frequency pulse, which means that the magnetization in x-y space will be affected by local magnetic field inhomogeneities, making GE sequences unsuitable for T_2 acquisitions. However, using only one radio-frequency pulse lowers TE, which lowers TR, and therefore improves T_1 contrast between tissues. Moreover, in GE sequences the flip angle α of the radio-frequency pulse is usually less than 90°, which means that $M_z(0)$ has a positive value greater than 0, allowing for a faster recovery of the magnetization along the z-direction, and a shorter TR. To improve contrast of vascular tissues and to detect breakdown in blood-brain barrier (BBB), while performing a T_1-weighted MRI scan, patients are injected with a non-toxic paramagnetic contrast enhancement agent, typically a gadolinium compound [28], and a post contrast T_1-weighted MRI (T1pc) is acquired.

FLAIR MRI

Fluid attenuated inversion recovery (FLAIR) is one of the most commonly used MRI sequence to suppress the cerebro-spinal fluid (CSF). It is based on the spin-echo sequence but it has an additional 180° radio-frequency pulse before the 90° radio-frequency pulse, separated in time by the inversion time (TI), which ensures cancelling the CSF signal when reading the echo. Typical FLAIR sequences have very long TR, TE, and TI.

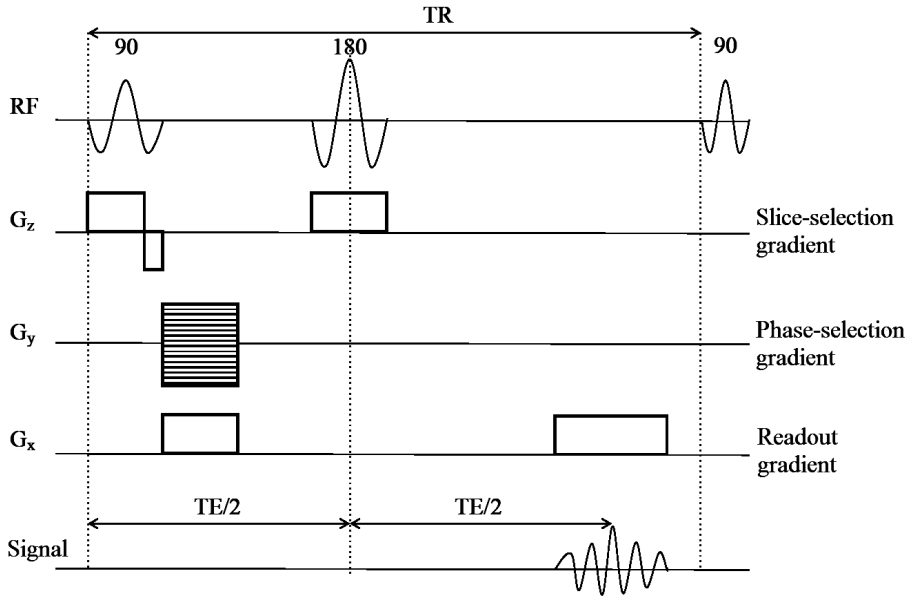


Figure 1.15: Details of a spin-echo sequence. Figure adapted from [126].

1.2.3 Perfusion weighted MRI

Perfusion weighted MRI (PWI) complements the anatomical information acquired with conventional MRI, as it is mainly applied for investigating vascular and tumoral brain pathologies. It provides measurements of the parameters of cerebral micro-vascularization, such as cerebral blood volume (CBV), cerebral blood flow (CBF), and mean transit time (MTT). It relies on using either an endogenous or exogenous tracer, followed by measurements of the transient signal through the brain during the first pass of the tracer. Arterial spin labelling (ASL) uses magnetically labelled blood as an endogenous tracer, while Dynamic Contrast Enhanced MRI (DCE-MRI) and Dynamic Susceptibility Contrast MRI (DSC-MRI) use a gadolinium chelate as an exogenous tracer. Both DCE- and DSC-MRI are based on the acquisition of serial MR images before, during, and after the administration of a contrast agent. DCE-MRI relies on acquiring and processing T_1-weighted images, which leads to quantification of permeability parameters (e.g. transfer constant (k^{trans}), rate constant (k_{ep})). DSC-MRI relies on acquiring and processing either T_2- or T_2^*-weighted images, but because the first one requires a larger amount of contrast agent and is less sensitive to magnetic susceptibility effects, T_2^*-weighted MRI became the standard. One of the main parameters quantified with DSC-MRI is CBV, which is extremely

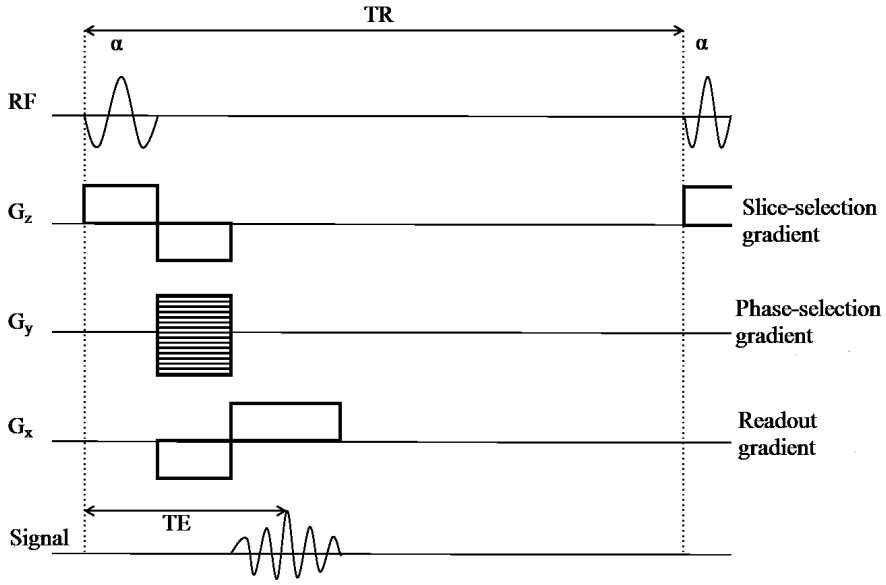


Figure 1.16: Details of a gradient-echo sequence. Figure adapted from [126].

important in brain tumour detection and diagnosis. Therefore, in this thesis, only DSC-MRI has been used to measure changes in brain perfusion. The vast majority of DSC-MRI acquisitions are based on the gradient-echo echo planar imaging (EPI) sequence, mainly because of its low acquisition time due to the short TE and the EPI readout [31]. Typical scan times are between 1 and 5 minutes, during which around 50 to 60 3-D brain images are taken. This allows to follow the evolution in time of a pixel's signal, which can be modelled by

$$S(t) = S(0) \cdot \exp(-\frac{TE}{\Delta T_2^*}). \tag{1.28}$$

It has been shown [252] that in the absence of recirculation and contrast agent leakage, CBV is proportional to the area under the contrast agent concentration–time curve. The contrast agent concentration $C(t)$ depends linearly with the change in relaxation rate [187]($\Delta R_2^* = \frac{1}{\Delta T_2^*}$), which can be calculated after re-writing equation 1.28:

$$C(t) = k\Delta R_2^* = -\frac{k}{TE} \ln(\frac{S(t)}{S(0)}), \tag{1.29}$$

where k is a constant specific to tissue, field strength, and pulse sequence.

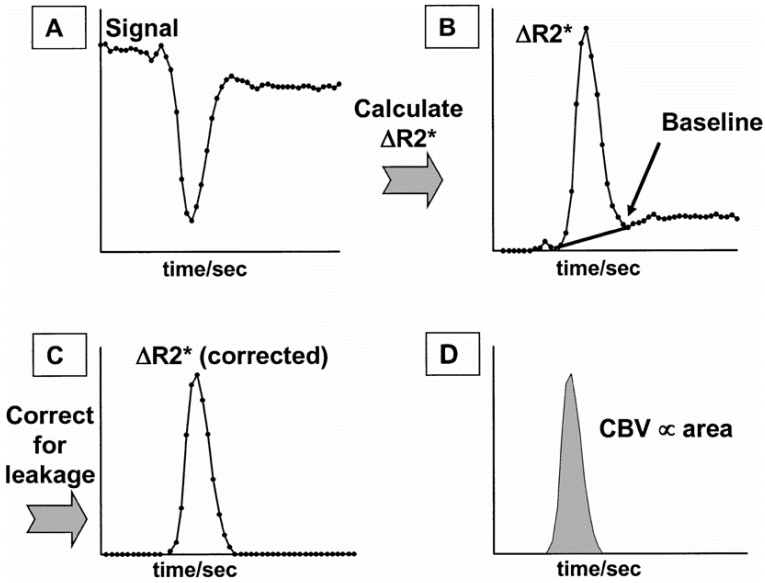


Figure 1.17: Perfusion MRI CBV quantification after correcting for contrast agent leakage. Source: Cha et al., Radiology, 2002 [31]

The typical workflow for CBV quantification is shown in Figure 1.17. In practice, relative CBV (rCBV) values are extracted by dividing the CBV values to an average CBV usually calculated over normal appearing white matter (NAWM), $rCBV = \frac{CBV}{CBV_{NAWM}}$. CBF calculation is done using the indicator dilution theory [168], which states that the contrast agent concentration is proportional to the convolution between an arterial input function (AIF) and a tissue residue function R:

$$C(t) = \alpha \cdot CBF \cdot (AIF(t) * R(t)). \tag{1.30}$$

The residue function describes the fraction of contrast agent remaining in tissue at time t, following an ideal injection (i.e. delta function), meaning that $R(t = 0) = 1$. Therefore, after deconvolving $C(t)$ and $AIF(t)$, we obtain a curve whose amplitude at $t = 0$ is proportional to CBF. To obtain reliable CBF values, a large brain vessel is usually selected by hand for the AIF concentration curve. In the end, MTT can be calculated using the central volume theorem [215], $MTT = \frac{CBV}{CBF}$.

1.2.4 Diffusion MRI

Diffusion MRI is an advanced non-invasive imaging technique which provides information about the water movement in biological environments, where it can interact with macromolecules, fibres, membranes, or other cellular structures. Usually, inside the brain, we can separate three types of diffusion: free isotropic, encountered mostly in CSF, restricted isotropic, encountered mostly in gray matter, and anisotropic, encountered mostly in white matter. A simple graphical representation of these diffusion types is shown in Figure 1.18. The first diffusion-

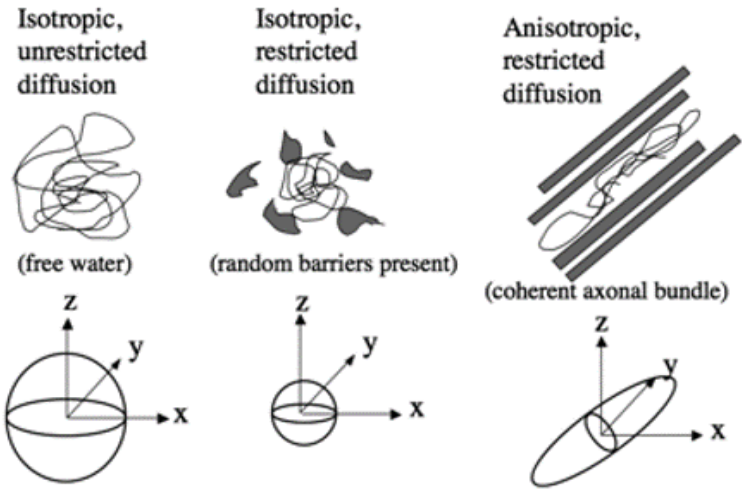


Figure 1.18: Types of diffusion. Source: [154]

weighted MRI (DWI) of the human brain were published in 1986 [139], and used a SE sequence. Because of the short acquisition time induced by EPI readout development, measuring diffusion signals is presently done using SE-EPI sequences. The most commonly used Pulsed Gradient Spin-Echo (PGSE) acquisition sequence is shown in Figure 1.19. The main novelty introduced by the PGSE sequence is the diffusion gradient, which can be applied in any direction. The diffusion gradient has two pulses in a given direction, equal in duration and amplitude, but opposite in phase. If there is no water diffusivity, the effects of these two pulses cancel each other. If there is water diffusivity, the re-phasing of water molecules that have moved will be incomplete and it will lead to a decrease of the diffusion signal measured in a voxel, which can be modelled by the following diffusion equation [214]:

$$S = S_0 \cdot \exp(-\gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}) D) \tag{1.31}$$

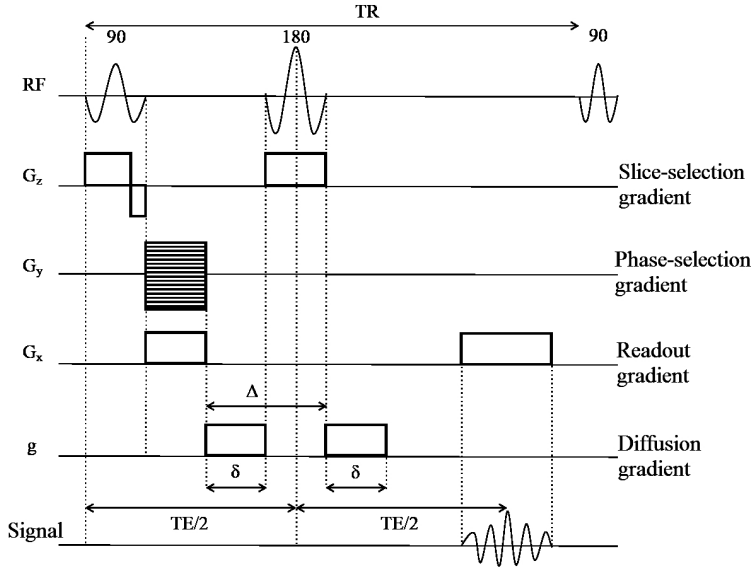


Figure 1.19: Pulsed Gradient Spin Echo diffusion weighted acquisition sequence. Figure adapted from [126].

where S is the acquired signal, S_0 is the spin-echo signal, G is the diffusion gradient's amplitude, δ is the duration of one pulse, Δ is the time between the two pulses, and D is the diffusion coefficient. According to Le Bihan et al. [139], the diffusion signal can be re-written as a function of the b coefficient:

$$S = S_0 \cdot \exp(-b \cdot ADC), \tag{1.32}$$

where ADC is the Apparent Diffusion Coefficient, and

$$b = -\gamma^2 G^2 \delta^2 (\Delta - \frac{\delta}{3}). \tag{1.33}$$

In DWI, it is common to acquire only two signals at different b-values, then to compute only one measure:

$$ADC = \frac{1}{b_2 - b_1} \cdot \ln(\frac{S_1}{S_2}) \tag{1.34}$$

In Diffusion Tensor Imaging (DTI) [10], the diffusion process is considered a *symmetric tensor* with different elongations along three main directions, x, y, and z. Seven measurements, one at $b = 0$ and six independent measurements at $b \neq 0$ in non-collinear directions, are sufficient to properly determine the

tensor shape, given by D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{xz}, and D_{yz}. Then, the diffusion equation can be re-written as:

$$\ln(\frac{S(b)}{S(b=0)}) = -\sum_{i,j\subset(x,y,z)} b_{ij}D_{ij}, \tag{1.35}$$

where

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix}, b = \begin{pmatrix} b_{xx} & b_{xy} & b_{xz} \\ b_{xy} & b_{yy} & b_{yz} \\ b_{xz} & b_{yz} & b_{zz} \end{pmatrix} \tag{1.36}$$

After performing eigenvalue decomposition on the diffusion tensor, three eigenvalues ordered from highest to lowest, λ_1, λ_2, and λ_3, are obtained across the three main directions, as shown in Figure 1.20. Multiple diffusion

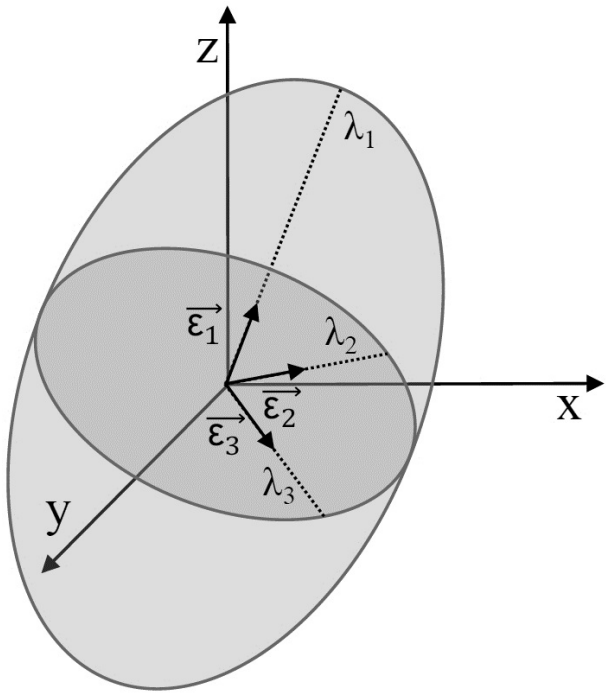


Figure 1.20: Diffusion tensor. Figure adapted from [126].

measures can be computed using the eigenvalues, such as: Axial Diffusivity (AD), defined as the largest eigenvalue, $AD = \lambda_1$; Radial Diffusivity (RD), defined as $RD = \frac{\lambda_2+\lambda_3}{2}$; Mean Diffusivity (MD), defined as the mean of all

three eigenvalues, $MD = \frac{\lambda_1+\lambda_2+\lambda_3}{3}$, and Fractional Anisotropy (FA), defined as

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - MD)^2 + (\lambda_2 - MD)^2 + (\lambda_3 - MD)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{1.37}$$

Both DWI and DTI assume the diffusion process is a random process, so it can be modelled using a Gaussian distribution. However, in biological tissues, the diffusion process has also a non-Gaussian component [120], which can help model an extended diffusion equation:

$$\ln(\frac{S(b)}{S(b=0)}) = -bD + \frac{1}{6}b^2D^2K \tag{1.38}$$

where the kurtosis K is a fourth-order tensor (K_{ijkl}) which quantifies the deviation from a Gaussian distribution. Similar to DTI, multiple diffusion kurtosis MRI (DKI) measures can be computed [108,175], such as Axial Kurtosis (AK), Radial Kurtosis (RK), Mean Kurtosis (MK), and Kurtosis Anisotropy (KA).

1.2.5 Magnetic Resonance Spectroscopic Imaging

Protong Magnetic Resonance Spectroscopy (MRS) is an advanced MR acquisition modality which provides biochemical information about a sample. Using MRS it is possible to identify a large number of molecules involved in the metabolic process of a cell, called metabolites. The basic principle behind MRS acquisition is the chemical shift effect, which states that in a perfectly homogeneous magnetic field B_0, protons from different molecules will experience a different magnetic field $B_0(1 - \sigma)$, where σ depends only on the chemical structure of that molecule. Therefore, protons from different metabolites will oscillate at different frequencies, given by $\omega = \gamma B_0(1 - \sigma)$, which are visible in the frequency domain of the acquired MR signal. Because the metabolite frequency shift is independent of the external magnetic field, it is commonly known as chemical shift and is not measured in Hertz, but in parts per million (ppm).

Following the radio-frequency excitation pulse in a large magnetic field, metabolites will emit signals several orders of magnitude lower than the water's signal, which has by far the highest strength due to the high natural abundance. In order to acquire MRS signals of sufficient Signal to Noise Ratio (SNR), it is necessary to excite bigger volumes compared to T1 MRI or T2 MRI. The acquisition of a Single Voxel MRS (SVS) is very common in clinical settings because it takes only a couple of minutes and provides high SNR. However, Multi Voxel MRS, also known as Chemical Shift Imaging (CSI) or Magnetic

Resonance Spectroscopic Imaging (MRSI), offers a larger excitation volume (2D or 3D) with a relatively better resolution compared to SVS, but in a longer time. Many reviews [13, 177] discussing advantages and disadvantages of SVS compared to MRSI have been published in the last decade.

Two commonly used localization sequences for both SVS and MRSI are Point RESolved Spectroscopy (PRESS) [17] and STimulated Echo Acquisition Mode (STEAM) [75], as shown in Figure 1.21. STEAM is usually adapted to short TE acquisitions (TE<50ms), while PRESS is adapted for both short and long TE (TE>100) acquisitions and provides double the SNR compared to STEAM.

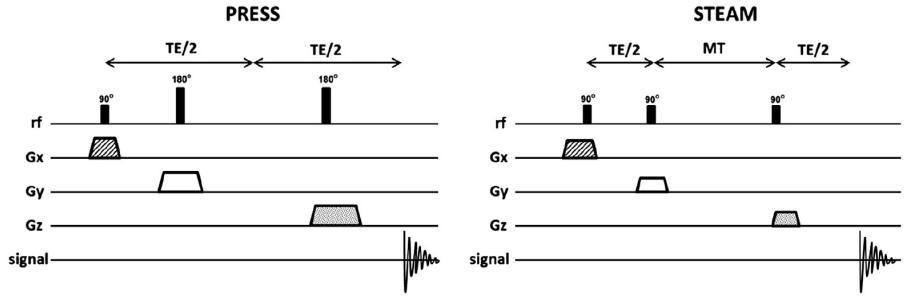


Figure 1.21: MRS acquisition sequences: PRESS and STEAM, where MT is the Mixing Time. Image adapted from [13].

The main difference between spectra acquired by short and long TE, as shown in Figure 1.22, is that at long TE only a limited number of metabolites can be identified and quantified: N-Acetyl-Aspartate (NAA), Choline (Cho), Creatine (Cre), and Lactate (Lac), while short TE allows for more metabolites to be quantified, such as Glutamate (Glu), Glutamine (Gln), Glu+Gln (Glx), Glycine (Gly), myo-Inositol (mI), and Lipids (Lip). However, spectra acquired by long TE MRS are less prone to errors, because signals from the underlying baseline of macro-molecules will drop close to zero at long echo times.

Quantifying metabolites of interest can be done by extracting information from either time or frequency domain. Most algorithms use a database containing MRS signals acquired from individual metabolites, either in-vitro or generated synthetically, which are then combined to provide the best fit in time or frequency domain. Two widely known methods for fitting MRS signals in time-domain are AQSES [180](Accurate Quantitation of Short Echo time domain Signals) and QUEST [183](QUantitation based on quantum ESTimation), while the most widely known method for fitting signals in frequency domain is LCModel [181]. In order to improve relevant information of MRS spectra, several preprocessing steps can be used, such as zero-filling, frequency alignment, phase correction,

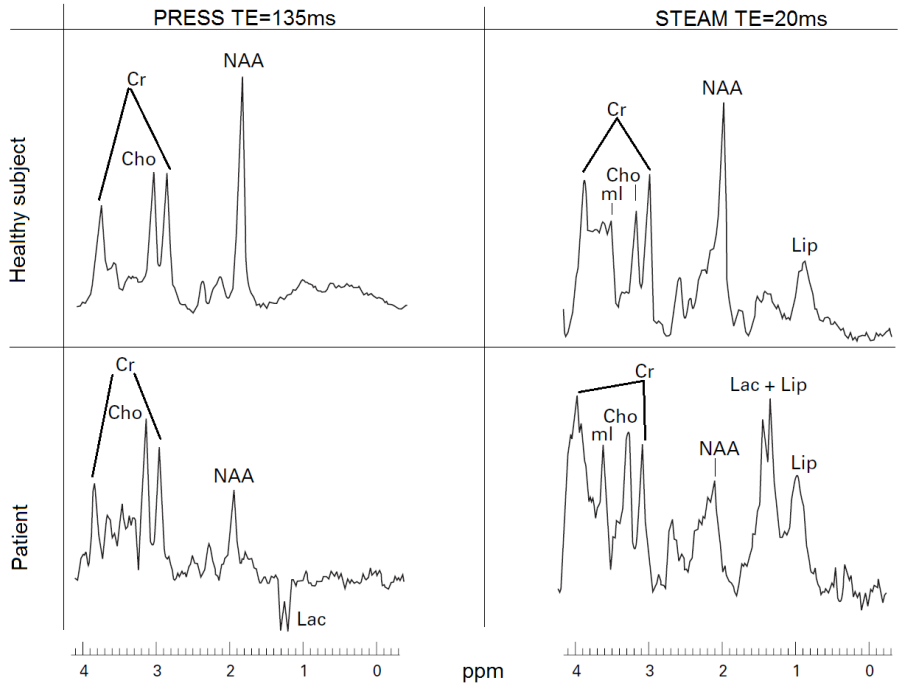


Figure 1.22: Short and long TE MRS spectra (right and left columns) for a healthy subject and a patient (top and bottom rows) suffering from progressive multifocal leukoencephalopathy, scanned at 1.5 Tesla. Image adapted from [117].

water filtering, signal normalization, and baseline correction.

1.3 Glioblastoma Multiforme

1.3.1 Glioblastoma Multiforme Overview

Glioblastoma Multiforme, or simply Glioblastoma (GBM), is the most common and aggressive primary brain and Central Nervous System (CNS) tumour, and corresponds to the highest tumour grade (IV) based on World Health Organization (WHO) Classification [144]. It accounts for 45% of malignant primary brain and CNS tumours, 54% of all gliomas, and 16% of all primary brain and CNS tumours [169]. The average incidence rate is 3.19/100,000 population, with a higher incidence in men and individuals of white race and

non-Hispanic ethnicity [224]. GBM is an incurable disease, with a median age of diagnosis of 64 years, and a median survival of 15 months [129, 228].

GBMs comprise of primary and secondary subtypes which evolve through different genetic pathways, affecting patients at different ages and have differences in outcomes [125]. Primary GBMs account for 80% of GBMs and occur in older patients (mean age 62 years), while secondary GBMs develop from lower-grade astrocytoma or oligodendrogliomas and occur in younger patients (mean age 45 years) [167]. Treatment is complex and initially consists of maximal-safe surgical resection followed by radiation therapy with concurrent temozolomide (TMZ) chemotherapy followed by six cycles of maintenance TMZ [216]. Many environmental and lifestyle factors including several occupations, environmental carcinogens, and diet have been reported to be associated with an elevated glioma risk, but the only factor unequivocally associated with an increased risk is therapeutic X-irradiation [166]. Several favourable clinical prognostic factors were also identified, such as younger age at diagnosis, cerebellar location, high Karnofsky Performance Status, and maximal tumour resection [135].

Although all GBMs are classified as WHO grade IV, they exhibit significant genetic heterogeneity. Various prognostic molecular markers can be identified based on specific genetic alterations [125, 135], including methylation status of the gene promoter for O6-methylguanine-DNA methyltransferase (MGMT), isocitrate dehydrogenase enzyme 1/2 (IDH1/2) mutation, epidermal growth factor receptor (EGFR) over-expression and amplification, glioma-CpG island methylator phenotype (G-CIMP), tumour protein (TP53) mutation and genetic losses of chromosomes. Several research groups also investigate immunotherapy, with a very recent study showing 7 out of 16 patients survived disease free after 5 years, 6 patients survived disease free after 7.7 years, and 3 were still disease free after 8 years [172].

1.3.2 Advanced MRI in the post-operative GBM follow-up

The current standard for assessing GBM response to therapy is called the Response Assessment in Neuro-Oncology (RANO) criteria [239], and it is based entirely on conventional MRI measurements (T1pc, T2, FLAIR) of contrast enhancing lesions. According to the RANO criteria, a GBM patient should be assigned into one of the following four groups:

- **Progressive Disease** - MRI shows an increase of $\geq$25% in the sum of the products of perpendicular diameter of enhancing lesions compared

to the smallest tumour measurement obtained either at baseline or best response.

- **Complete Response** - MRI shows disappearance of all measurable and non-measurable disease sustained for at least 4 weeks.
- **Partial Response** - MRI shows (1) a decrease of $\geq$50% compared with baseline in the sum of products of perpendicular diameters of all measurable contrast-enhancing lesions sustained for at least 4 weeks, and (2) stable or improved non-enhancing lesions compared with baseline scan.
- **Stable Disease** - does not qualify for partial response, complete response, nor progressive disease.

However, an apparent increase or decrease of enhancing lesions can also be misidentified as *pseudoprogression* or *pseudoresponse* [217, 230]. Therefore, one of three main directions of GBM follow-up research is discriminating between Treatment Related Changes (TRC) (e.g. radiation injury, pseudoprogression) and true progression (TPr), with an extended review given by Da Cruz et al. [46]. The other two main directions of GBM follow-up research are (1) studying overall survival (OS) and progression free survival (PFS), and more recently, (2) computing parametric response maps (PRM) based on diffusion and perfusion MRI maps for studying OS, PFS, and TRC vs. TPr.

There is a general consensus in the literature that features extracted from advanced MRI modalities, such as diffusion MRI [16, 30, 105, 236], perfusion MRI [5, 105, 161, 217, 236], and MR Spectroscopy [242], are helpful in discriminating TRC and TPr with high accuracy, while only one study [86] shows that volumetric features of the tumour could be predictors of treatment outcome. All literature states that higher values of rCBV and rCBF are associated with angiogenesis, therefore they are indicative of tumour growth and infiltration. Lower values of MD or ADC and higher values of FA are also associated to true tumour progression.

For survival analysis there is the same general consensus: diffusion metrics [148], with an emphasis on ADC [32, 90, 152, 176, 240, 245], but also perfusion metrics [104, 148, 199], are helpful in discriminating short and long OS and PFS.

Parametric Response Map (PRM) is a voxel-wise approach for image analysis and quantification of alterations during treatment that has been shown to have predictive power of OS, PFS, and discriminative power between TPr and TRC. It was initially developed for ADC maps by Moffat et al. [61–63, 90, 91, 152, 247], but later it was adapted to rCBV maps by Galban et al. [3, 82, 83, 229, 247].

Recent studies have shown that advanced machine learning algorithms can predict early recurrence areas [1] based on multi-parametric MRI, and also treatment outcome [250] based on bio-molecular gene expression.

1.3.3 UZ Leuven post-operative GBM dataset

Patients population In this thesis, there is a total of 29 GBM patients followed up post-operative at the University Hospitals of Leuven (UZ Leuven). Sixteen of them have relapse-GBM and were treated according to the HGG-IMMUNO-2003 protocol [51, 52, 194, 234], meaning they received immune therapy as sole treatment after surgery. The rest of thirteen patients had primary GBM and were treated according to the HGG-IMMUNO-2010 protocol [234], meaning that after surgery they were split into two groups. The first group consisting of 6 patients received radiochemotherapy and the immune therapy vaccine. The second group consisting of the remaining 7 patients received just radiochemotherapy for the first six months after surgery, and afterwards all of them received radiochemotherapy plus the immune therapy vaccine. We refer to the first group as "HGG-IMMUNO-2010 vaccine" and to the second group as "HGG-IMMUNO-2010 placebo".

All 29 patients were offered monthly MRI follow-up, but after six months under immune therapy all patients switched to a three monthly scanning schedule. The local ethics committee approved this study and informed consent was obtained from every patient before the first imaging time point. Based on radiological evaluation of the follow-up MRI scans using the RANO criteria [239], each time-point of a patient was labelled as one of the following:

- **Unlabeled**, if the MRI scan was labelled as "Stable" according to the RANO criteria
- Labeled as **Progressive**, if the MRI scan was labelled as "Progressive Disease" according to the RANO criteria
- Labeled as **Responsive**, if the MRI scan was labelled as "Partial Response" or "Complete Response" according to the RANO criteria

MRI acquisition and processing Magnetic resonance scanning was performed on a clinical 3 Tesla MRI system (Philips Achieva, Best, The Netherlands), using a body coil for transmission and a 32-channel head coil for signal reception. The scanning protocol at UZ Leuven consisted of standard anatomical imaging (T1pc, T1, T2 and FLAIR) and advanced MRI (DSC-MRI, DKI-MRI, and MRSI).

Conventional MRI Conventional MRI (cMRI) were acquired as described in [231, 233, 235]. In brief, an axial spin echo T2-weighted MR image (TR/TE: 3000/80 msec, slice/gap: 4/1 mm, field of view (FOV): 230×184 mm^2, turbo factor (TF): 10, acquisition matrix: 400×300), a FLAIR image (TR/TE/TI: 11000/120/2800 msec, slice/gap: 4/1 mm, acquisition matrix: 240×134) and a T1-weighted 3D spoiled GE scan (fast field echo - FFE, TR/TE: 9.7/4.6 msec, flip angle: 8°, turbo field echo factor: 180, acquisition voxel size: 0.98×0.98×1 mm^3, 118 contiguous partitions, inversion time: 900 msec) with and without contrast administration were acquired as high-resolution anatomical reference images. Examples of conventional MRI are shown in Figure 1.23.

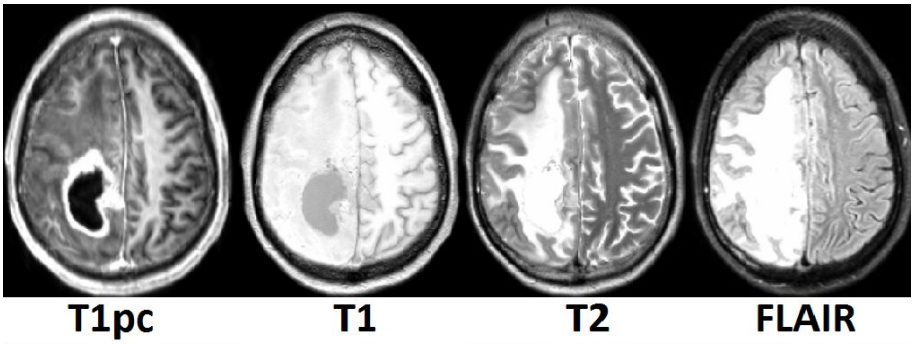


Figure 1.23: Conventional MRI of a post-operative GBM patient.

DSC-MRI acquisition Perfusion images were obtained using a standard DSC-MRI protocol consisting of a GE-EPI sequence, TR/TE: 1350/30 msec, section thickness/gap: 3/0 mm, dynamic scans: 60, FOV: 200×200 mm^2, matrix: 112×109, number of slices: 23, scan time: 1 minute 26 seconds. EPI data were acquired during the first pass following a rapid injection of a 0.1 mmol/kg body weight bolus of megluminegadoterat (Dotarem, Guerbet, Villepinte, France) via a mechanical pump at a rate of 4 ml/sec, followed by a 20 ml bolus of saline. Pre-load dosing was performed in order to correct for T1-weighted leakage (pre-load dose 0.1 mmol/kg megluminegadoterat, incubation time 10 min) [103].

Examples of perfusion MRI parameter maps obtained using the DSCoMAN plugin [18] for ImageJ [200], which takes into consideration the leakage correction, are shown in Figure 1.24.

DKI-MRI acquisition and processing DKI data were acquired as described in [231, 233], using a SE-EPI-DWI sequence with TR/TE: 3200/90 msec, δ/Δ:

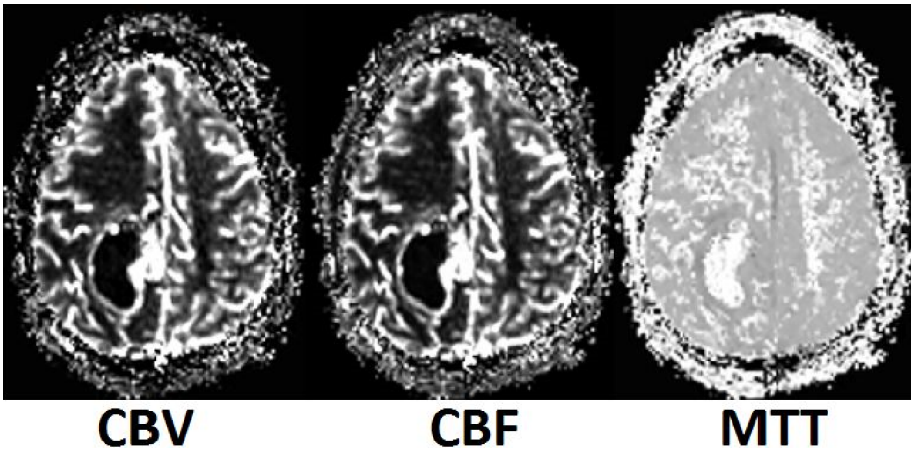


Figure 1.24: Post-operative GBM perfusion MRI parameter maps obtained using the DSCoMAN plugin [18].

20/48.3 msec; FOV: 240×240 mm^2, matrix: 96×96, number of slices: 44, 1 signal average acquired, section thickness/gap: 2.5/0 mm, b-values: 700, 1000 and 2800 sec/mm^2 in 25, 40 and 75 uniformly distributed directions respectively [175]. The DKI data were processed as described in [231].

Examples of DKI parameter maps are shown in Figure 1.25.

MRSI acquisition and processing A 2D-MRSI short echo time PRESS protocol with $TR/TE = 2000/35$ msec was used as validated in [232]. The volume of interest (VOI) is positioned on the slice of the transverse reconstruction of the T1-weighted 3D-FFE sequence with the largest section of contrast enhancement. The slice thickness of the VOI is 10 mm and the VOI is 80 × 80 × 10 mm^3, with each voxel being 5 × 5 × 10 mm^3 (16 × 16 voxels in total). If the contrast-enhancing lesion was smaller than 2 cm^3 or the contrast-enhancement is located in areas with large susceptibility differences e.g. the basal forebrain or the anterior temporal lobes, a SVS technique was performed (TR/TE: 2000/35 msec, minimal volume: 1 cm^3).

MR spectra were processed using the MATLAB 2010b environment (MathWorks, Massachusetts, U.S.A.) with the SPID graphical user interface [212], as described in [232]. Nine metabolites were quantified using the AQSES method [44]: NAA, Gln, Glu, Cre, phosphorylcholine (PCh), glycerophosphorylcholine GPC, mI, and lipids (Lips) at 0.9 and 1.3ppm, referred to as Lip1 and Lip2 respectively. Two different sums, Glu+Gln and PCh+GPC, were reported as Glx and tCho

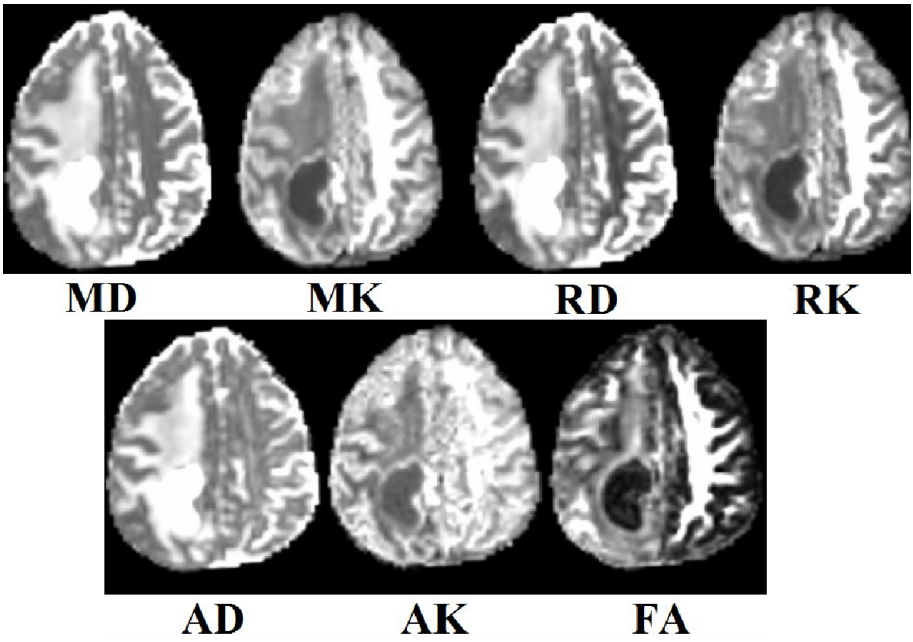


Figure 1.25: DKI parameter maps of a post-operative GBM patient.

(total choline), respectively. For each metabolite, AQSES reported metabolite concentrations in institutional units and their error estimates as Cramer-Rao Lower Bound (CRLB) [29].

1.4 Multiple Sclerosis

1.4.1 Multiple Sclerosis Overview

Multiple Sclerosis (MS) is an inflammatory neurodegenerative disorder of the human CNS that damages axons and the myelin protective sheath covering the axons, which typically generates focal lesions in white and grey matter. MS affects approximately 2.5 million people worldwide, with an average onset age of 30 years. Global incidence rates are shown in Figure 1.26, with women cases doubling those of men [6].

The exact cause of inflammation is unclear, but the most likely suspects are either an autoimmune response directed towards the CNS, or failure of myelin-

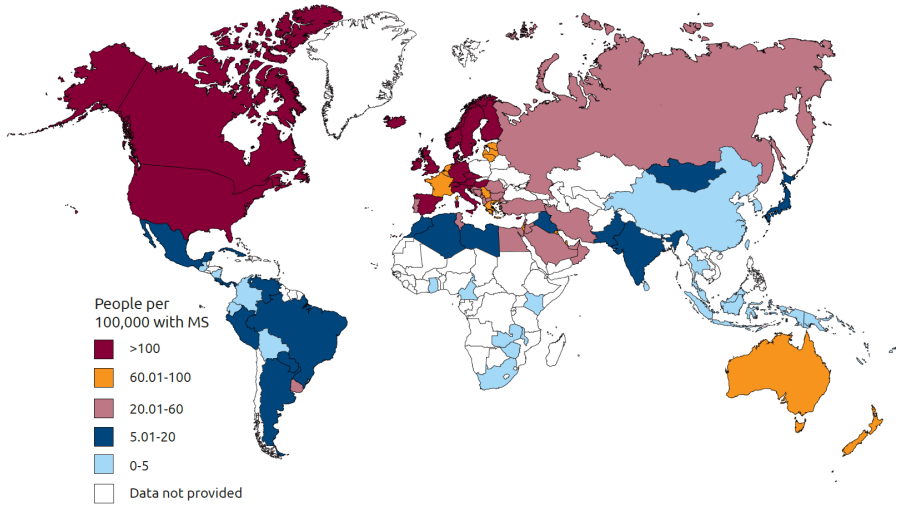


Figure 1.26: Multiple Sclerosis global incidence. Source: [6].

producing cells [156]. The majority (85%) of MS patients usually experience a first attack defined as Clinically Isolated Syndrome (CIS), while the rest of 15% will start directly with a primary progressive (PP) form. All CIS patients will develop a relapsing-remitting (RR) form [151], with two thirds of RR patients developing a secondary progressive (SP) form, and the other third having a benign course [198]. A simplified graphic of MS evolution is presented in Figure 1.27.

Diagnosing MS forms is a complicated topic which was originally described by McDonald in 2001 [150], and revised by Polman in 2005 [174] and 2011 [173]. State of the art MS diagnosis criteria include clinical and paraclinical laboratory assessments emphasizing the need to demonstrate lesion dissemination in space and time and to exclude alternative diagnoses.

Because the underlying cause of MS is still under debate, treatment has recently been shifted towards 'no evidence of disease activity' or NEDA, which is a composite of three measures: no relapses, no disability progression, and no MRI activity (new or enlarging T2 lesions or T1pc lesions) [8, 85, 94]. Two standard clinical scores used to measure disability and decrease in cognitive functions are the Expanded Disability Status Scale (EDSS) [133] and the Multiple Sclerosis Functional Composite [45]. Conventional MRI (T1, T1pc, T2, FLAIR) plays a crucial role in MS diagnosis due to high sensitivity for visualizing MS lesions, and MRI measures such as T2 lesion volume, T2 lesion count, T1pc lesion count, and brain volume loss (BVL), have been extensively investigated. In an

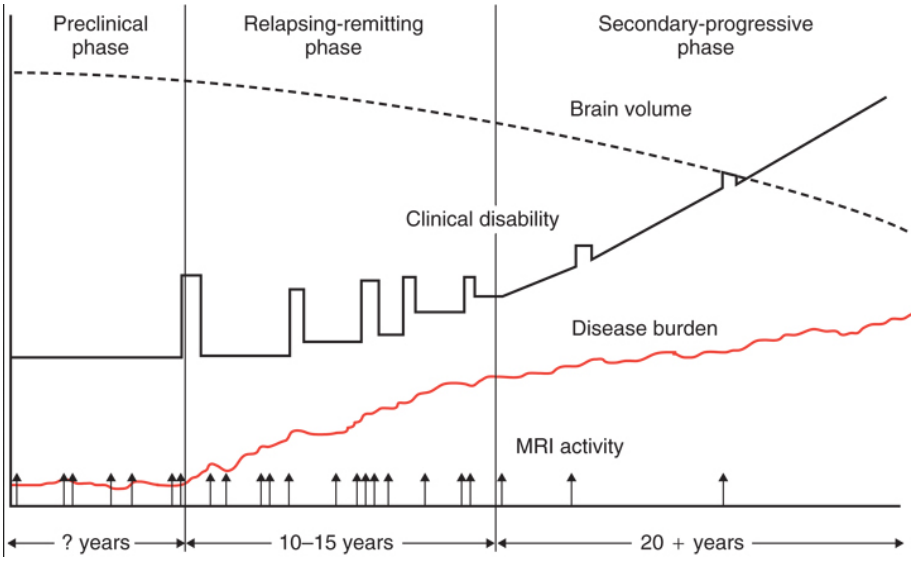


Figure 1.27: Multiple Sclerosis disease progression. Source: [74].

effort to improve the classical NEDA criteria, which is weighted towards focal inflammatory disease activity, Kapos et al. [121] propose BVL as a measure to capture neurodegeneration effects.

1.4.2 Advanced MRI in the longitudinal MS follow-up

Conventional MRI provides a high sensitivity in detecting new lesions and in computing volumetric measurements of the brain and the lesions, but it comes with a poor specificity in identifying the real stage of MS lesions. Indeed, conventional MRI can discriminate between active and non-active lesions, but this separation is based on the evidence of BBB breakdown, as indicated only by contrast enhancement. However, there are multiple factors influencing the imaging of T1pc, such as treatment effects, contrast enhancement dosage, MR acquisition parameters, and time delay between dosage and imaging [67].

Histopathological and MRI data suggest that lesions evolve differently during early versus chronic MS phases, going through several stages of demyelinating activity: early active, late active, inactive, early remyelinated, and late remyelinated [24]. Active lesions predominate in relapsing MS and might be the cause of clinical attacks. Inactive or slowly expanding lesions predominate in progressive MS, therefore they might contribute to progression. Remyelinated

lesions are found in both relapsing and progressive MS [72]. Moreover, it has been shown that in progressive MS, inflammation and active tissue injury can occur in the absence of BBB breakdown [100, 138].

Therefore, advanced MR modalities have been explored, such as DTI, Magnetization Transfer Imaging (MTI), and MRS, due to their ability to assess tissue damage both inside the lesions and outside, in the NAWM and in the normal appearing gray matter (NAGM) [65, 124].

Indeed, multiple studies have shown higher ADC or MD and lower FA values in lesions than in areas of NAWM, but conflicting results have been achieved when comparing ADC or MD values in enhancing vs non-enhancing lesions or between lesions with different patterns of enhancement [7, 58, 68, 70, 192, 241]. Another way of analysing diffusion data is by generating white matter fiber tracts (or tractography), which give information about the structural connectivity of the brain. It has been proven that MS patients have lower brain connectivity than healthy controls, and it correlates with disability [203]. Additional longitudinal diffusion studies are needed to investigate how much of tissue disruption in enhancing lesions is permanent (i.e., related to axonal loss) and how much is transient (i.e., related to demyelination and remyelination). Two extended reviews [111, 189] provide additional information on the use of diffusion MRI in MS.

MTI is an MRI acquisition that can explore non-water components in tissue. Protons bound to larger molecules have relaxation properties that make them invisible in a conventional MR measurement. However, the number of bound protons can be indirectly quantified by sequences that use the energy exchange between bound protons and water protons, commonly termed Magnetization Transfer (MT). The relative magnitude of MT, MT ratio (MTR), is very attractive for clinical use, as it is fast, easy to acquire, reproducible, and comparable among subjects, provided the same sequence is used [186]. Reduced MTR values compared to healthy controls have been measured in NAWM and NAGM from patients with different MS forms, including those at the earliest stage [71]. MTR abnormalities from NAGM correlated with disease duration, disability, cognitive impairment, and were more pronounced in patients with progressive MS forms [2, 71, 95, 182, 190]. MTR was also consistent with demyelination and remyelination that followed different temporal evolutions and were ongoing in different lesion regions for at least 3 years after new lesion formation [35].

Proton MRS metrics provide high MS pathological specificity as well as high sensitivity to biochemical changes. Decrease of NAA and increase in Cho and Cre contents were observed in lesions and NAWM of MS patients, compared to healthy controls, indicating axonal loss, demyelination, and cell proliferation [191,

195,223]. A recent longitudinal study [124] showed that pre-lesional tissue giving rise to new lesions had higher Choline and Creatine values compared to NAWM. Moreover, they observed a positive correlation between changes in moderately hypointense lesions volume and Choline values, indicating that changes in lesion size are mediated by the level of chronic inflammation.

1.4.3 AMSEP longitudinal dataset

Patient population In this thesis, eighty-seven MS patients (12 CIS, 30 RR, 28 SP and 17 PP) part of the AMSEP project were studied against 18 volunteers without any neurological disorders, who served as healthy control (HC) subjects. Diagnosis and disease course were established according to the McDonald criteria [146, 150]. This prospective study was approved by the local ethics committee (CPP Sud-Est IV) and the French national agency for medicine and health products safety (ANSM) and written informed consents were obtained from all patients and control subjects prior to study initiation. More details for each MS group, such as average age at first scan, average disease duration, median EDSS and average lesion loads can be found in Table 1.1. The MS patients involved in this study were scanned multiple times over a different period for each patient, ranging from 2.5 to 6 years. The minimum number of scans is 3, while the maximum is 10. The gap between two consecutive scans is either 6 months or 1 year.

	CIS	RR	PP	SP
Number of patients (Male/Female)	12 (6/6)	30 (6/24)	17 (6/11)	28 (17/11)
Age at first scan [years]	31.8 (6.4)	33.2 (7)	39.5 (6)	41.1 (4.8)
Disease duration [years]	2.9 (1.9)	8.3 (4.8)	7.5 (2.9)	14.9 (6.1)
EDSS median [range]	1 (0-4)	2 (0-5.5)	4 (2-7.5)	5 (3-8.5)
Lesion Load [ml]	6.6 (3.5)	16.7 (12.6)	20.8 (13)	31 (12.9)
Total number of scans	62	226	125	206

Table 1.3: Patient population: Age - average value (standard deviation); Disease duration - average value (standard deviation); EDSS - median (minimum - maximum); Lesion Load - average value (standard deviation).

All patients and control subjects underwent MR examination using a 1.5 Tesla MR system (Sonata Siemens, Erlangen, Germany) and an 8 elements phased-array head-coil.

Conventional MRI acquisition Conventional MRI protocol consisted of a 3 dimensional T1-weighted MPRAGE (magnetization prepared rapid gradient echo) sequence with TR/TE/TI= 1970/3.93/1100 ms, flip angle= 15°, matrix

size= 256×256, FOV= 256×256 mm, slice thickness= 1 mm, voxel size= 1×1×1 mm, acquisition time= 4.62 min, and a FLAIR sequence with TR/TE/TI= 8000/105/2200 ms, flip angle= 150°, matrix size= 192 × 256, FOV= 240 × 240 mm, slice thickness= 3 mm, voxel size= 0.9 × 0.9 × 3 mm, and an acquisition time of approximately 5 min.

MRSI acquisition and processing MRSI data was acquired from one slice of 1.5 cm thickness, placed above the corpus callosum and along the anterior commissure - posterior commissure (AC-PC) axis, encompassing the centrum semioval region, and took 5 minutes and 20 seconds. A PRESS with TR=1690 ms and TE=135 ms was used to select a volume of interest (VOI) of 105×105×15 mm^3 during the acquisition of 24 × 24 (interpolated to 32 × 32) phase-encodings over a FOV of 240 × 240 mm^2.

MRSI data processing was performed using SPID [178, 179] in MatLab 2015a (MathWorks, Natick, MA, USA). AQSES [44, 180] was used to quantify NAA, Choline, and Creatine, using a synthetic basis set, which incorporates prior knowledge of the individual metabolites in the quantification procedure. MPFIR (maximum-phase finite impulse response) filtering [220] was included in the AQSES procedure for residual water suppression, with a filter length of 50 and spectral range from 1.7 to 4.2 ppm. A band of two voxels at the outer edges of each VOI was discarded in order to avoid chemical shift displacement artifacts and lipid contamination artifacts.

In Figure 1.28 there are examples of MS lesions on the axial view (right), brain atrophy on the sagittal view (center), and a full MRSI grid superimposed on the T1pc map.



Figure 1.28: MRSI grid (red) superimposed on T1pc of a Multiple Sclerosis patient. From left to right: coronal, sagittal, and axial view.

1.5 Objectives of the thesis and main contributions

The main objective of this thesis is to develop supervised classification models based on multi-parametric MRI data in order to discriminate and predict the evolution of two major pathologies affecting the brain: GBM and MS. To that end, the thesis has two parallel tracks: analysing GBM MRI data acquired at UZ Leuven, Belgium, and analsing MS MRI data acquired at CERMEP, Lyon, France.

The main objective in analysing the GBM MRI data was to study the discrimination between progressive and responsive patients, labelled as described in Section 1.3.3, with the secondary objective of predicting in advance their evolution from unlabelled (or stable) to labelled (progressive or responsive).

To some extent, both objectives have reached their end-goals in Chapter 2, when perfect discrimination and even perfect prediction were possible with one month in advance, but only on a limited number of 18 GBM patients, because we restricted the study to comparing only advanced MRI modalities (PWI, DKI, and MRSI). Perfect results were obtained by learning classifiers without any parameter tuning (e.g. Random Forest with 100 trees), trained only on perfusion MRI features that were easy to extract, such as average rCBV and rCBF values computed inside the manually delineated Contrast Enhancing (CE) region of interest (ROI) of the tumour.

In Chapter 3 we replaced the low quality MRSI data with the high quality conventional MRI data, and proposed a semi-manual method for delineating CE ROIs. As a result, more patients and more data points were added to the study, and perfect prediction and discrimination between progressive and responsive patients was not possible anymore. Therefore, more complex feature extraction and selection methods were required to achieve high discriminative results. This increase in overall complexity of the classification models was possible only after developing an automated preprocessing pipeline for quantifying PWI and DKI parameters. Learning a modified boosting algorithm on features extracted from semi-manual ROIs was shown to provide very high accuracy results for GBM diagnosis.

As a last effort in analysing GBM data, a modified PRM approach was developed in Chapter 4, which takes into account the most likely infiltration area of the tumour, greatly reducing the manual tumour delineation time of a clinician, because all subsequent MRI scans were registered to the first one. Two types of computing PRM were compared, one based on T1pc and one based on rCBV, as features extracted with these two modalities were the best in discriminating the GBM evolution, according to results from the previous two chapters. Results obtained within this last GBM analysis showed that using PRM based on cMRI

is clearly superior to using PRM based on PWI.

The main objective in analysing the MS MRI data was to study the metabolic differences between all four MS courses (CIS, RR, PP, and SP), and a second objective was to compare metabolic features extracted from healthy control subjects and MS patients. Due to the difficulty of correctly diagnosing MS courses even when more information is available, only binary classification models were investigated.

Classifiers for different MS courses have been learned and tuned independently on multiple types of data, such as clinical data (e.g. patient age, disease age, EDSS), lesion loads, metabolic ratios (e.g. NAA/Cre, NAA/Cho, Cho/Cre), but also on all of them combined. These results are summarized and presented in Chapter 5. Furthermore, state of the art classifiers (e.g. CNN) were developed and their architecture was optimized for higher level MRS features. These results are summarized and presented in Chapter 6. Results for discriminating healthy control subjects from MS patients were not satisfactory, even though statistically significant differences between the two groups were observed. Finally, training CNNs on MRS spectrograms did not improve classification of MS courses over less complex machine learning algorithms, such as SVM-rbf, trained on more simple features, such as metabolic ratios.

1.6 Outline of the thesis

This thesis is structured as shown in Figure 1.29.

In **Chapter 1** basic concepts of ML, MRI, GBM, and MS, are described. In **Chapter 2** a first study is presented on classifying and predicting the evolution of 18 post-operative GBM patients based only on advanced MRI data. In **Chapter 3** a second study, more complex than the first one, is presented on classifying 29 post-operative GBM patients based on conventional and advanced MRI data. In **Chapter 4** a third study is presented on classifying and predicting the evolution of 29 post-operative GBM patients using a modified PRM approach on conventional and perfusion MRI data. In **Chapter 5** a first study is presented on classifying four different MS forms of 87 MS patients from 18 HC subjects, based on clinical and MRI data. In **Chapter 6** a second study used state of the art classifiers (CNN) for discriminating the main four MS forms of 87 MS patients based only on MRSI data and brain tissue segmentations. Finally, in **Chapter 7**, the main limitations, achievements, general conclusions of our studies, as well as possible future ideas are discussed.

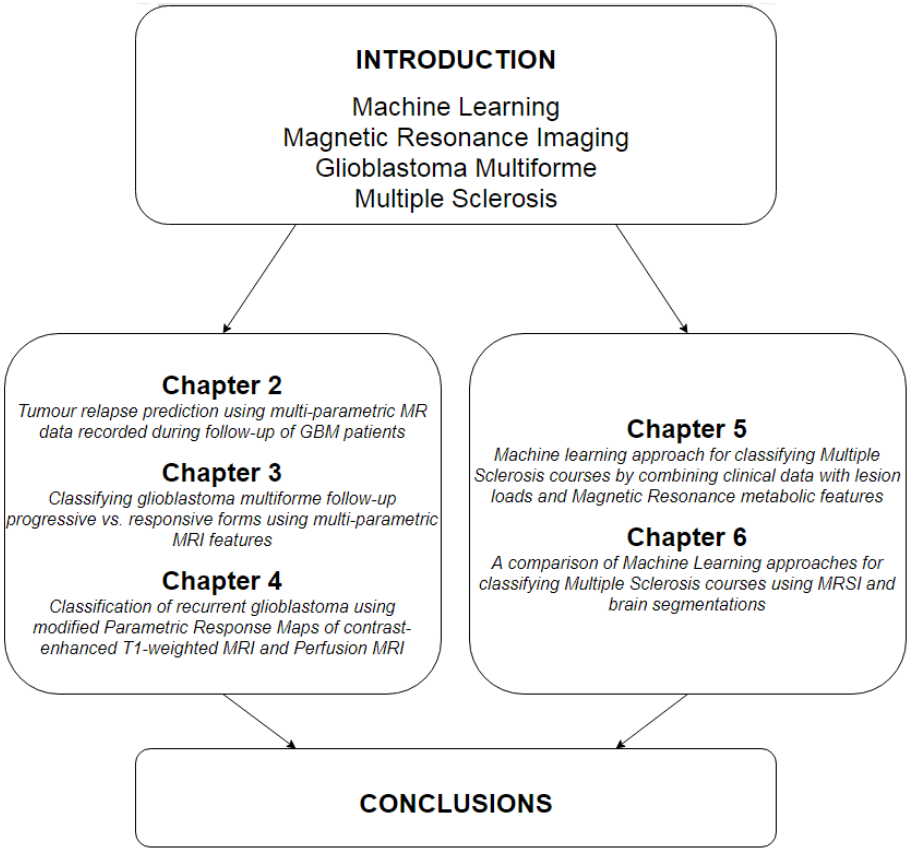


Figure 1.29: Thesis outline.

1.7 Conclusion

Conventional MRI is the standard modality for detecting the main features of abnormal brain tissues, such as brain tumours or lesions, but it has limited specificity in detecting the grade or evolution of both diseases, therefore more advanced MRI modalities need to be acquired for better brain tissue characterization. For GBM follow-up prediction and diagnosis there is an increasing interest in research and clinical settings to acquire and analyse DKI, MRSI, and PWI, due to remarkable properties of imaging tissue organisation, metabolism, blood perfusion and neo-angiogenesis. Also for MS follow-up prediction and diagnosis there is an increasing interest in both research and clinical settings to acquire and analyse MRSI, MTI, and DWI. With this

increasing MRI data volume and complexity, machine learning fits perfectly the need to compact and compare different types of MRI, analyse data and extract meaningful features that can be used for validating different models, with the ultimate goal of improving diagnosis accuracy and reducing time consuming tasks of clinicians such as manual tumour delineation.

Chapter 2

Tumour relapse prediction using multi-parametric MR data recorded during follow-up of GBM patients

In this chapter the difference between progressive and responsive GBM patients was analyzed based only on advanced MRI data: PWI, DKI, and MRSI. For each scanning session from a total of 178 sessions, three regions of interest were manually delineated by an expert clinician: contrast enhancing, edema, and necrosis. Average values of the first two ROI, contrast enhancing and edema, were then extracted from three parametric perfusion maps (cerebral blood volume, cerebral blood flow, decrease ratio), and from three parametric diffusion maps (fractional anisotropy, mean diffusivity, mean kurtosis). Therefore, six features were extracted from PWI (three from contrast enhancing and three from edema), and six more features were extracted from DKI (three from contrast enhancing and three from edema). Ten features were extracted from MRSI, assessed only within the contrast enhancing voxels. Five additional features were extracted and always used as input to machine learning algorithms: the three ROI volumes (contrast enhancing, edema, necrosis) and two categorical parameters, one indicating if the patient had a total tumour resection and another one indicating the treatment group. In total there are 27 features: six perfusion parameters, six diffusion parameters, 10 spectroscopic parameters, three tumour compartments volumes, and two additional parameters. Although

the MR data was acquired from 29 patients, only 18 patients had a complete data set. The larger incomplete dataset has 178 data points (56 labelled and 122 unlabelled), while the smaller complete dataset has 45 data points (18 labelled and 27 unlabelled). A method was developed to fill-in the missing data and a comparison was done between results based on complete data and results based on imputed data. The performance of twenty-three supervised and three semi-supervised classifiers was measured using a leave-one-patient-out cross-validation scheme. Training of the classifiers was done only on labelled data, while testing was done also on unlabelled data. The work presented in this chapter was published in [116] in a slightly modified version.

2.1 Introduction

GBM is the most common and malignant intracranial tumour [26], representing as much as 30% of primary brain tumours with increasing incidence in some geographic regions [56]. The patients have a median survival of only 10 to 14 months after diagnosis with only 3 to 5% of patients surviving more than three years. Recurrence is universal, and at the time of relapse, the median survival is only five to seven months despite therapy [193]. The current standard of care is surgical resection followed by radiotherapy and concomitant and adjuvant temozolomide chemotherapy [216].

MRI is the most widely used medical imaging technique for identifying the location and size of brain tumours. However, cMRI has a limited specificity in determining the underlying type of brain tumour and tumour grade [53, 60]. More advanced MR techniques like DWI, PWI, and MRSI, are promising in the characterization of brain tumours as they give potentially more physiological information [162, 184, 235].

DWI and DKI visualize the tissue structure and are useful for assessing tumour cellularity, because they give information about the water movement inside different tissues including biological barriers. Typical parameters related to diffusion are ADC, MD, MK and FA. MD is a general parameter that accounts for the mean diffusivity in all directions, MK might be a specific parameter for tissue structure [120], and FA is a general index of anisotropy, with a value of zero corresponding to isotropic diffusion and a value of one corresponding to diffusion only in one direction.

PWI provides measurements that reflect changes in blood flow, volume, and angiogenesis. Hypervascularity due to glioma-induced neoangiogenesis may show up as high rCBV while necrosis of different tissues may show up as low

rCBV [147]. MRSI provides information about metabolites present in normal and abnormal tissue [153].

We have studied patients with GBM that had the tumour surgically removed and afterwards were treated according to two different protocols developed for evaluating dendritic cell immuno-therapy: HGG-IMMUNO-2003 [51,52,194,234], and HGG-IMMUNO-2010 [234].

The focus of our this chapter is finding a map between multiparametric MR data acquired during GBM follow-up and the brain tumour relapse after surgery, as described by the clinically accepted RANO criteria [239]. In order to do this, we test different families of classifiers on multiparametric MR data, starting from simple ones, e.g. k-Nearest Neighbours (k-NN) and Linear Discriminant Analysis (LDA), and moving to non-linear classifiers, e.g. random forests and neural networks, using a total of 27 features extracted from PWI, DKI and MSRI data.

2.2 Materials and Methods

2.2.1 Study setup

Study setup was described in Section 1.3.3. Each MRI time point for each patient was assessed based on the RANO criteria, and assigned to two major groups: either unlabelled for all time points before the decision moment, or labelled as "responsive" or "progressive" for all time points at and after the decision moment.

2.2.2 MRI acquisition and processing

MRI was performed on a clinical 3 Tesla MR imaging system (Philips Achieva, Best, The Netherlands), as previously described in Section 1.3.3.

Anatomical magnetic resonance imaging

Conventional MRI data was acquired as previously described in Section 1.3.3. Regions of interest (ROI) were manually drawn around the solid contrast-enhancing (CE) region if present, avoiding areas of necrosis or cystic components such as the surgical cavity. A second ROI was manually drawn around the total lesion, i.e. sum of contrast-enhancement (CE) and perilesional oedema

(ED). The ROI containing the perilesional oedema was obtained by subtracting the CE ROI from the total lesion. Finally, a separate ROI was drawn around the contralateral NAWM to standardize the PWI hemodynamic measurements. The CE ROI was delineated on T1pc, while the total lesion was delineated on T2.

The manual delineations were drawn by a radiologist (Dr. Sofie Van Cauter) with 5 years experience of brain tumour MR imaging. An example of delineations on T1pc can be seen in figure 2.1, where green is the necrosis, red is CE and blue is ED.

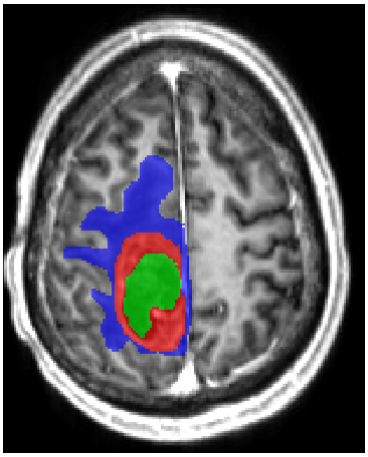


Figure 2.1: Brain tumour delineations on T1pc MRI. Green - necrosis, Red - contrast enhancing region of interest, Blue - edema.

Magnetic resonance spectroscopy

MRSI data was acquired and processed as previously described in Section 1.3.3. Sixty-six per cent (66 %) of all spectroscopic time points are not included in this study. There are two reasons for this: (1) quantification was not possible for all time points, mainly because of movement artefacts, and (2) the rest of them did not pass the quality control recommended by Kreis [130], as listed below.

Reject data if:

- full-width at half-maximum peak height of metabolites > 0.07–0.1 ppm

- relative Cramer-Rao lower bounds $> 50\%$
- unexplained features in residuals—reject, if artifact or expand model, if unexpected metabolite peaks doubled or patient moved (post-acquisition MRI)
- lineshape strongly asymmetric after eddy correction outer volume ghosts or other artifacts present (at least exclude metabolites overlaid with artifact)

Perfusion Weighted MRI

PWI data was acquired as previously described in Section 1.3.3. DSC data were analysed using DPTools (www.fmritools.org), as described in [231]. The mean values of the considered perfusion parameters were retrieved in the CE, ED, and NAWM regions. We report relative rCBV (rrCBV), relative rCBF (rrCBF) and relative DR (rDR) of tumoural tissue by using the corresponding parameter value in the contralateral NAWM as internal reference. Although quantification was possible for all time points, after quality assessment done by visual inspection by SVC, 30% of them were not included in this study.

Diffusion kurtosis imaging

DKI data was acquired as previously described in Section 1.3.3. The DKI data were processed as described in [231]. Fractional anisotropy (FA), mean diffusivity (MD) and mean kurtosis (MK) were derived from the tensors [108, 120]. A nonlinear registration of the parameter maps to the anatomical MR imaging data was performed to minimize the local misalignment between the EPI distorted DKI data and the anatomical data on which the ROIs were manually positioned. MK, MD and FA were determined in the CE and ED regions. Although quantification was possible for all time points, after quality control according to [175], 44% of them were not included in this study.

Summary of MRI acquisition and processing

In total, from 29 patients, there are 178 data points of follow-up MR imaging sessions, and each of these ones has 27 features:

- 3 volumes - contrast enhancement (CE), oedema (ED), necrosis (N)
- 6 perfusion features - rrCBV, rrCBF, rDR for CE and ED

- 6 diffusion features - MK, MD, FA for CE and ED
- 10 spectroscopic features - from CE - NAA/tCho, NAA/sum, tCho/sum, NAA/Cre, Lips/tCho, tCho/Cre, Myo/sum, Cre/sum, Lips/Cre, Glx/sum
- a parameter (0 or 1) for total resection of the tumour
- a parameter (0,1, or 2) to describe the group of the patient - HGG-IMMUNO-2003, HGG-IMMUNO-2010 placebo or HGG-IMMUNO-2010 vaccine

Out of all 178 data points, if we extract just the ones with complete features, it will result in a subset of 18 patients with 45 data points. This implies that more than 75% of the measurements have at least one feature missing. Five features are always present: the three volumes, the parameter for tumour resection, and the parameter for different groups.

2.2.3 Classifiers

We have used several supervised and semi-supervised classifiers, as presented in Table 2.1, with the goal of testing whether the unlabelled data could have been reliably labelled before the actual decision was taken in the clinic according to the RANO criteria.

The list of classifiers in Table 2.1 is representative for the most important families of classification methods, starting from simple classical ones such as LDA and k-NN up to more complex non-linear classifiers such as RF and neural networks.

All classifiers are implemented in Matlab R2013a (MathWorks, Massachusetts, U.S.A.). All classifiers except Least Squares Support Vector Machines (LSSVMs) and the semi-supervised ones are part of the Statistics Toolbox and Neural Networks Toolbox of Matlab R2013a.

k-NN [42] is one of the basic classifers in machine learning. The class label of a new testing point is given by the most common class among its k neighbours. We used the default MATLAB R2013a (Statistics Toolbox) function “knnclassify” to run a grid search for the best combination of number of neighbours (k) and type of distance. We varied k between 1 and 11 and the distance was either “euclidean”, “cityblock”, “cosine” or “correlation”. We found the best results for the combination of 3 neighbours and the “correlation” distance.

Diagonal LDA (dLDA [201]) is a simple modification of Linear Discriminant Analysis, which implies that we use the pseudoinverse of the covariance

Supervised Classifiers	Handles missing values
Random Forests	✓
Classification Tree	✓
Boost ensembles	✓
Neural networks	–
SVM	–
LSSVM	–
k-NN	–
dLDA	–
Semi-supervised classifiers	
LDS	–
SMIR	–
S4VM	–

Table 2.1: Supervised and semi-supervised classifiers tested in this chapter.

matrix instead of the actual inverse. We used the default MATLAB R2013a implementation "classify" from the Statistics Toolbox.

SVMs [39, 43] are among the most popular machine learning models because they are easy to understand: given a training set with points that belong to two classes, we try to find the best hyperplane to differentiate between the two types of points. We can try this in the original space or we can map the points to another space by using the kernel trick. We used the default MATLAB R2013a (Statistics Toolbox) implementations "svmtrain" and "svmclassify". We used different types of kernel: linear, polynomial, radial basis function and multi-level perceptron.

Classification tree [22] is an algorithm commonly used in machine learning. Like in a real tree there are leaves which represent class labels, and branches. At each node of a tree a single feature is used to discriminate between different branches. We used the default MATLAB R2013a (Statistics Toolbox) implementation "classregtree".

Neural networks [88, 101, 143, 188] are built on interconnected layers of artificial "neurons" that try to map an input vector to its specific output. There are three types of layers: input, hidden and output. The weights between different neurons are trained until a maximum number of iterations or a minimum error is reached. We used the default MATLAB R2013a (Neural Network Toolbox) implementation "net" with 10 hidden neurons. We tested four types of neural networks: pattern net, feed forward net, cascade forward net and fit net.

Random forests [20, 21] are part of the ensemble methods for classification that use a collection of decision trees. Each decision tree learns a rule and then it can classify a new point. The new point is assigned to the class voted by the majority of the decision trees. We used the default MATLAB R2013a (Statistics Toolbox) implementation "TreeBagger" with 100 trees.

Boosting algorithms [76, 77, 79, 202] start with a collection of weak classifiers (e.g. decision trees) and with each iteration they try to improve the overall classification by learning what was misclassified at the previous step. We used the default MATLAB R2013a (Statistics Toolbox) implementation "fitensemble" with 100 trees. We tested seven types of boosting algorithms: AdaBoost, LogitBoost, GentleBoost, RobustBoost, LPBoost, TotalBoost and RUSBoost.

LSSVMs [221, 222] are a powerful machine learning technique. We downloaded LSSVMlab from [66] and followed the instructions from [49] to tune the parameters. We used different types of kernel: linear, polynomial, radial basis function, and also the Bayesian approach on LSSVM.

The semi-supervised classifiers used in this chapter are Low Density Separation (LDS [34]), Squared-loss Mutual Information Regularization (SMIR [165]), and Safe Semi-Supervised Support Vector Machine (S4VM [141, 251]). In the last years there has been a steady increase in the use and development of semi-supervised classifiers, as they take into account information from unlabeled data also, not just from labeled data. This makes them powerful machine learning tools. The implementation for semi-supervised classifiers was downloaded from [218], [136] and [33].

Classifiers were tested first with all features described in section 2.2.2 taken as input, but then also by selecting subsets of the available features as input, i.e., only the features pertaining to a single modality (perfusion, diffusion and spectroscopy). Additionally, classifiers were tested first on the smaller dataset containing 45 time points with a complete set of features and then on the larger dataset containing 178 time points where missing values have been imputed according to section 2.2.4, presented below.

2.2.4 In-house imputation method

Some classifiers have built-in strategies of handling missing values, but other classifiers do not handle missing values (see table 2.1). This is why we developed our own in-house imputation method, so the handling of missing values will be the same for all classifiers.

Our method is based on the volumes of contrast enhancement and oedema

regions, in the sense that if the volume of a tumour region is zero, that missing tissue is considered healthy tissue. If we have values of any modality (perfusion, diffusion, spectroscopy) that are missing from CE or ED, and the volume of CE or ED corresponding to that measurement is zero, then we assume that those missing values belong to a normal type of tissue. For perfusion, because we normalize every parameter to the normal appearing white matter value, the missing values will be replaced by 1's. For diffusion and spectroscopy, the missing values will be replaced by the average of the features taken over the measurements which were labeled as responsive, because we consider that these measurements are recorded from a healthy tissue. If we have missing values without association to zero volume for CE or ED, they will be replaced by the average taken over all the labeled measurements.

2.2.5 Performance indices

Leave one patient out (LOPO)

Classifiers are trained on labelled data from all patients except one who is the test patient. Each patient in turn is selected as test patient. All time points that belong to the test patient are classified independently. Results for each classifier are averaged per time point over all patients relative to the time point at which the clinical decision was made. This way of testing is intuitive from a medical point of view and provides us with information about how good is the classification when we approach the decision time. In this way we can look at the temporal evolution of the classification for each patient.

We compute the balanced error rate (BER) at each time point before and after the decision, using the clinical decision assigned to each patient as expected label for all time points of this patient. BER is computed as

$$BER_i = \frac{ERR_i^{resp} + ERR_i^{prog}}{2}, \tag{2.1}$$

where

$$ERR_i^{resp} = \frac{\text{Number of responsive patients misclassified as progressive}}{\text{Total number of responsive patients}} \tag{2.2}$$

and

$$ERR_i^{prog} = \frac{\text{Number of progressive patients misclassified as responsive}}{\text{Total number of progressive patients}}. \tag{2.3}$$

For each classifier we have a total of 17 time points, due to the fact that there are patients with up to 6 time points after the decision and other patients with

up to 11 time points before the decision. In order to compare the classifiers by using just one error number instead of 17, we compute a weighted average for each classifier's time response. This performance measurement is denoted by "weighted BER (wBER)" in the results section.

We use two sets of weights:

- one for the temporal response - the classifier should perform better when we approach the labelling time point and after it $W_i^t = 1$, if i $\geq$ decision time point $W_i^t = 1 - \frac{0.5}{11} \cdot i$, if i $<$ decision time point
- one for patient population - the time points with more patients get a higher weight (see table A.8 from Appendix) $W_i^p = \frac{\text{Number of patients at time point i}}{\text{Total number of patients}}$

The equation of wBER is

$$wBER = \frac{\sum W_i^p \cdot W_i^t \cdot BER_i}{\sum W_i^p \cdot W_i^t} \tag{2.4}$$

2.3 Results and Discussion

2.3.1 Results

LOPO when using all modalities

Table A.1 from Appendix shows how different classifiers perform on complete and imputed features when trained on all MR modalities. We selected the best 6 classifiers (highlighted in table A.1) and present their detailed BER results for each time point in table 2.2.

Table 2.3 shows the detailed BER results for each time point for the best 6 classifiers trained on imputed data.

LOPO when using each modality

Table 2.4 shows the best six supervised classifiers' performance trained on complete features of each MR modality separately.

Tables A.2, A.3, A.4 from Appendix list the best classifiers' performance trained on complete features of PWI, DKI, and MRSI, respectively. Table 2.5 shows the performance of the best classifiers trained on separate imputed features. PWI

Table 2.2: Detailed BER results for each time point when training the best 6 classifiers on complete features for all MR modalities. The decision (i.e. labelling) moment 'L' is highlighted. Some time points do not have results because there were no complete measurements.

BER	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+5	–	–	–	–	–	–
L+4	–	–	–	–	–	–
L+3	0	0	0	0	0	0
L+2	0	0	0	0	0	0
L+1	0	0	0	0	0	0
L	0	0.1	0.217	0	0	0.1
L-1	0	0.125	0	0	0	0.125
L-2	0.25	0.25	0.5	0.25	0.25	0.25
L-3	0.5	0.5	1	0.5	0.5	0.25
L-4	1	1	1	1	1	0.5
L-5	0.25	0.25	0.25	0.25	0.25	0.25
L-6	0.5	0	0	0.5	0.5	0
L-7	1	0	1	1	1	0
L-8	–	–	–	–	–	–
L-9	0	0	0	0	0	0
L-10	–	–	–	–	–	–
L-11	0	0	1	0	0	0
wBER	0.148	0.172	0.276	0.148	0.148	0.136

and DKI features were extracted from both CE and ED ROI. MRSI features were extracted only from CE voxels.

Tables A.5, A.6, A.7 from Appendix list the performance of the best supervised classifiers (highlighted in table A.1) when using, respectively, perfusion, diffusion or spectroscopy data separately, considering imputed features only.

In-house imputation strategy vs. Built-in imputation strategy

Table 2.6 shows how different classifiers perform with our in-house imputation of missing values (section 2.2.4) versus the built-in imputation strategy of missing values for the classifiers marked in table 2.1.

Table 2.3: Detailed BER results for each time point for the best 6 classifiers when trained on imputed data for all MR modalities. The decision (i.e. labelling) moment 'L' is highlighted.

BER	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+5	0	0	0	0	0	0
L+4	0	0	0	0	0	0
L+3	0	0	0	0	0	0
L+2	0.125	0.25	0.125	0.125	0.125	0
L+1	0.171	0.071	0.071	0.171	0.171	0.071
L	0.105	0.022	0.149	0.188	0.105	0.359
L-1	0.214	0.065	0.130	0.3	0.192	0.192
L-2	0.444	0.417	0.194	0.444	0.472	0.5
L-3	0.418	0.382	0.282	0.418	0.418	0.482
L-4	0.475	0.413	0.388	0.475	0.413	0.475
L-5	0.688	0.438	0.563	0.688	0.688	0.688
L-6	0.368	0.467	0.3	0.567	0.567	0.567
L-7	0.375	0.375	0.75	0.5	0.75	0.625
L-8	0.5	0.333	0.583	0.5	0.75	0.333
L-9	0.333	0.333	0.833	0.333	0.833	0.5
L-10	0.5	0.75	0.75	0.5	1	0.75
L-11	0.5	0.5	1	0.5	0.5	0.5
wBER	0.294	0.216	0.242	0.335	0.325	0.352

Table 2.4: Weighted BER for the best 6 supervised classifiers trained on complete data for each MR modality separately. PWI and DKI features were extracted from both CE and ED ROI. MRSI features were extracted only from CE voxels.

Weighted BER	Random forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
PWI	0.148	0.256	0.220	0.148	0.148	0.193
DKI	0.358	0.259	0.255	0.367	0.367	0.349
MRSI	0.571	0.561	0.600	0.609	0.623	0.629

2.3.2 Discussion

A first conclusion that we can draw from a comparative analysis of the different classifiers is that we obtain the lowest error when training classifiers on data with complete features and not on data with imputed features, no matter the imputation method (our in-house method or the built-in method). In order to improve the performance of classifiers, improving the quality of the data would help.

Table 2.5: Weighted BER for the best classifiers trained on imputed features from each MR modality separately. PWI and DKI features were extracted from both CE and ED ROI. MRSI features were extracted only from CE voxels.

Weighted BER	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
PWI	0.294	0.311	0.275	0.289	0.265	0.282
DKI	0.277	0.327	0.322	0.277	0.277	0.380
MRSI	0.412	0.401	0.423	0.423	0.408	0.415

Table 2.6: Weighted BER comparison between our in-house method of imputing missing values and built-in imputation strategy of different supervised classifiers.

Weighted BER	Our method	Built-in method
Random forests	0.294	0.423
AdaBoost	0.324	0.333
LogitBoost	0.335	0.241
GentleBoost	0.308	0.245
RobustBoost	0.325	0.296
LPBoost	0.256	0.369
TotalBoost	0.289	0.323
RUSBoost	0.308	0.361
Decision Tree	0.346	0.651

The lowest error when using complete features is around 0.14 (SVM-mlp - 0.136), while if we use imputed features the lowest error is 0.216 (dLDA). The best classifiers on complete features are ensemble classifiers (random forests and boosting algorithms), dLDA and SVM, while the best classifiers on imputed features are dLDA, SVM-lin and random forests.

If we compare the results of single MR modalities when training classifiers on data with complete features, we can say that the use of spectroscopy only leads to the worst results with a minimum error of 0.561. The single use of perfusion generates better results than using only diffusion data, especially when using ensemble methods (random forests, LogitBoost and RobusBoost), with a minimum error of 0.148 compared to 0.255. When using imputed features, the minimum error almost doubles.

An interesting aspect when looking at detailed measurements on complete features (table 2.2) is the fact that we have error equal to zero (perfect classification) one time point before the actual labeling according to the RANO

criteria, when using random forests, LogitBoost or RobustBoost. This means that we can predict the patient outcome (progressive, responsive) with 100% accuracy one time point (i.e. about 1 month in our study) earlier than the actual clinical decision was made. When looking at each MR modality separately (tables A.2, A.3, A.4) we notice that the same result could have been obtained by using solely the perfusion data. This is a very important finding, mainly because perfusion is very fast to measure (2-3 minutes) and it has the lowest rate of missing data, which makes it reliable. Our study is not the only one that shows that perfusion parameters are very reliable when it comes to differentiating between tumour tissues and other tissues. Multiple studies (among others Barajas et. al. [9] and Hu et. al. [104]) prove that perfusion parameters are strongly correlated to tumour progression and overall survival. The main reason behind this strong correlation is the fact that tumours grow very fast, so they require large amounts of nutrients to develop, which is reflected in the angiogenesis of the tumour. This increase in angiogenesis is visualised and measured using perfusion imaging.

When comparing the two methods of imputing missing values, our in-house method (section 2.2.4) and the classifier-dependent built-in strategies, the difference between them is not important with respect to the performance of the classifiers.

Using machine learning for classification of brain tumoral tissue is a field with an increasing amount of work.

In [105] Hu et. al used a support vector machine approach on multiparametric MRI (perfusion, diffusion and anatomical MRI) to automatically differentiate between radiation necrosis voxels and progressive tumour voxels coming from patients with resected GBM. They optimize a One Class SVM based on the Area under Receiver Operator Curve from 6000 training voxels manually delineated from 8 patients and then tested on manually delineated voxels from 8 new patients. Their results show that perfusion and diffusion have a high discrimination rate between radiation necrosis and tumour progression.

In [9] Barajas et al. used perfusion MR imaging to investigate which parameters can be used to differentiate between recurrent GBM and radiation necrosis. Their study was based on 57 patients, they used Welch t test to compare measurements between groups, and they found that all perfusion parameters (relative CBV, peak height, percentage of signal intensity recovery) are strongly correlated to tumour progression.

In [104] Hu et al. used perfusion metrics on contrast enhancement lesions (CBV mean, mode, maximum, width, and a new thresholding metric called fractional tumour burden (FTB)) to see how they correlate to overall survival

(OS). Their study was based on 25 patients with recurrent GBM and found that all parameters are strongly correlated to OS.

In [242] Weybright et. al used MRSI to differentiate voxels with tumour recurrence and radiation injury. Their study was based on 29 patients and they had high quality data for 28 of them (97%). They found that the Cho/NAA and Cho/Cr ratios may be the best numerical discriminators between tumour recurrence and radiation injury.

A number of factors limited this study and will try to be addressed in the future. MRSI feature extraction was done only from contrast enhancing voxels, however the edema contains infiltrative tumour growth and is expected to show changes in metabolites. Moreover, MRSI acquisition was done at the end of the scanning session, therefore making it difficult for patients to resist, and more prone to movement artifacts. The imputation method assumed normal type of tissue if the contrast enhancing or edema ROI was missing, but there could be cases of pseudoresponse where it is possible for tumour to be present. Also, filling in with average values has its disadvantage, as any information concerning the patient was missing, e.g. tumour location or scanning session.

Although we cannot compare our results directly to the ones from previous studies because they are studying a different classification problem, it is becoming more obvious that a learning algorithm based on multiparametric MR data will evolve in the near future and will help clinicians in differentiating between progressive tumour tissue and other types (necrotic or normal).

2.4 Conclusions

In this chapter we study the classification problem between progressive and responsive GBM patients, by comparing different supervised and semi-supervised machine learning algorithms. We train them on multi-parametric MR data with complete and imputed features, acquired from 29 GBM follow-up patients. We found that on a subset of 18 patients, the same label according to the RANO criteria could have been put one month earlier with 100% accuracy, if we trained RF on data with complete features, and more importantly, the same result was achieved by the same classifiers using only complete perfusion data. This conclusion is in contrast to the conception that multi-parametric MRI (e.g. perfusion and diffusion) is always better than single-parametric MRI (e.g. perfusion). For future work we plan on using the temporal evolution of the features when classifying different MR sessions and also allow updating of the class labels in time.

Chapter 3

Classifying glioblastoma multiforme follow-up progressive vs. responsive forms using multi-parametric MRI features

In this chapter the difference between progressive and responsive GBM patients was analyzed based on conventional and advanced MRI data: cMRI, PWI, and DKI. Conventional MRI consisted of four parameter maps, perfusion MRI consisted of five parameter maps, and diffusion MRI consisted of seven parameter maps. The same data from the previous chapter was used: follow-up MR data acquired from 29 patients, summing up to 183 data points, 56 labelled and 127 unlabelled. For each scanning session from the total of 183, only one ROI (total tumour) was manually delineated by an expert clinician, as opposed to the previous chapter, where three ROI were manually delineated. Afterwards, based on the total tumour T1pc voxels, contrast enhancing and non-enhancing ROI were automatically delineated. Twenty-six statistical and 3-D texture features were extracted from each MR parameter ROI, summing up to 416 multi-parametric MR features from one ROI: 104 conventional MR features (26 features $\times$ 4 maps), 130 perfusion MR features (26 features $\times$ 5 maps), and 182 diffusion MR features (26 features $\times$ 7 maps). In this chapter the complete labelled dataset had 55 data points, compared to the previous chapter where it

had only 18 data points. Therefore, the focus was only on the complete dataset, and results based on the single manually delineated total tumour ROI were compared to results based on the two semi-manually delineated ROI, contrast-enhancing and non-enhancing. The performance of seven supervised classifiers was measured using a leave-one-patient-out cross-validation scheme, after the best ten features were selected based on the rank products of six independent feature selection algorithms. The work presented in this chapter was published in [115] in a slightly modified version.

3.1 Introduction

GBM is the most common and malignant intracranial tumour [26], representing as much as 30% of primary brain tumours with increasing incidence in some geographic regions [56]. The patients have a median survival of only 10 to 14 months after diagnosis with only 3 to 5% of patients surviving more than three years. Recurrence is universal, and at the time of relapse, the median survival is only five to seven months despite therapy [193]. The current standard of care is surgical resection followed by radiotherapy and concomitant adjuvant temozolomide chemotherapy [216].

MRI is the most widely used medical imaging technique for identifying the location and size of brain tumours. However, cMRI has a limited specificity in determining the underlying type and grade of the brain tumour [53, 60]. More advanced MR techniques like PWI and DKI are promising in the characterization of brain tumours as they give potentially more physiological information [162, 184, 235]. DKI visualize the tissue structure and are useful for assessing tumour cellularity, as it gives information about the water movement inside different tissues including biological barriers. Typical parameters related to diffusion are FA, MD, and MK. FA is a general index of anisotropy, with a value of zero corresponding to isotropic diffusion and a value of one corresponding to diffusion only in one direction. MD is also a general parameter that accounts for the mean diffusivity in all directions, while MK might be a specific parameter for tissue structure [120]. PWI provides measurements that reflect changes in blood flow and blood volume. Hypervascularity due to glioma-induced neoangiogenesis may show up as high rCBV while necrosis of different tissues may show up as low rCBV [147].

We studied patients with GBM that had their tumour surgically removed and afterwards were treated according to two different protocols developed for evaluating dendritic cell immuno-therapy: HGG-IMMUNO-2003 [51,52,194,234], and HGG-IMMUNO-2010 [234].

The focus of this chapter is the same as in the previous chapter, namely finding a map between multi-parametric MRI data acquired during the follow-up of GBM patients and the relapse of brain tumour after surgery, as described by the clinically accepted RANO criteria [239]. We were motivated to conduct this study because of our excellent previous results where we could differentiate, based only on PWI features, between progressive and responsive follow-up GBM patients with 100% accuracy one month before the patients were labelled according to the RANO criteria. A drawback of our previous results was the small sample size (18 patients, 27 data points), so in this study we want to confirm our findings on an extended dataset of 29 patients, which includes data from the previous study.

Additionally, we present two main improvements: (1) semi-manually delineating contrast enhancing regions (CER) and non-enhancing regions (NER) and (2) extracting additional texture and histogram features, with the purpose of improving classification performance. The first improvement is to automatically delineate CER, based on the manually delineated total tumour region. Delineating CER by hand is a time consuming process and requires the radiologist's full attention to make sure that necrosis or non-enhancing regions are not in CER. We select CER based on the T1pc main property of imaging the contrast agent's leakage into the active tumour, which determines high intensity areas where the active tumour is located. The second improvement is extracting histogram and texture features and selecting those with high differentiating power. In the previous chapter we used only the average parameter values from CER, NER, and total, while in this one we extract six histogram features and twenty 3-D texture features based on the gray level co-occurence matrix (GLCM), as described in Section 3.2.2. We do feature selection using six of the most widely known features selection algorithms and combine feature rankings using the rank product method, as described in Section 3.2.2. In Section 3.3, linear and non-linear classifiers are tested on a varying number of features, and their results are combined into separate groups, which are used as input to non-parametric statistical tests to discover which combination of delineation, MR modality, and classifier, achieves the highest rank.

3.2 Materials and Methods

3.2.1 Study setup

Study setup was described in Section 1.3.3. Each MRI time point for each patient was assessed based on the RANO criteria, and assigned to two major groups: either unlabelled for all time-points before the decision moment, or

labelled as "responsive" or "progressive" for all time-points at and after the decision moment. In total there are 183 time points, 56 are labelled and 127 are unlabelled.

3.2.2 MRI acquisition and processing

The MR images were acquired on a clinical 3 Tesla MR imaging system (Philips Achieva, Best, The Netherlands), using a body coil for transmission and a 32-channel head coil for signal reception. The imaging protocol consisted of cMRI, PWI, and DKI.

Conventional MRI

Four types of conventional MR images were acquired as previously described in Section 1.3.3.

Perfusion MRI

Perfusion MRI data were acquired as previously described in Section 1.3.3. PWI were processed using the DSCoMAN plugin [18] for ImageJ [200], which takes into consideration the leakage correction and can easily be automated. For each PWI acquisition, five parameter maps were extracted: corrected cerebral blood volume (CBV), cerebral blood flow (CBF), mean transit time (MTT), time to peak (TTP), and R^2 (Rsquare).

Diffusion MRI

DKI data were acquired as previously described in Section 1.3.3. For each DKI acquisition, seven parameters maps were derived from the tensors [108, 120]: fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), mean kurtosis (MK), axial kurtosis (AK), radial kurtosis (RK).

Delineations

A ROI was manually drawn around the Total tumour lesion, avoiding areas of necrosis or cystic components such as the surgical cavity. A separate ROI was drawn around the contra-lateral NAWM to standardize measurements extracted

from the tumour region. The Total and NAWM ROIs were drawn by Dr. Sofie Van Cauter, a radiologist with 5 years experience.

To automatically split the Total region in two ROIs, CER and NER, a threshold was set at the 90^{th} percentile of T1pc Total voxels. In this way, two semi-manual ROIs were made for each patient based on the T1pc intensities selected from Total: CER, containing very high T1pc intensity Total voxels, and NER, containing the rest of Total voxels. The 90^{th} percentile threshold was selected after visually inspecting T1pc maps of multiple patients.

A typical example of manual and semi-manual ROIs on T1pc can be seen in Figure 3.1, where red is CER and blue is NER.

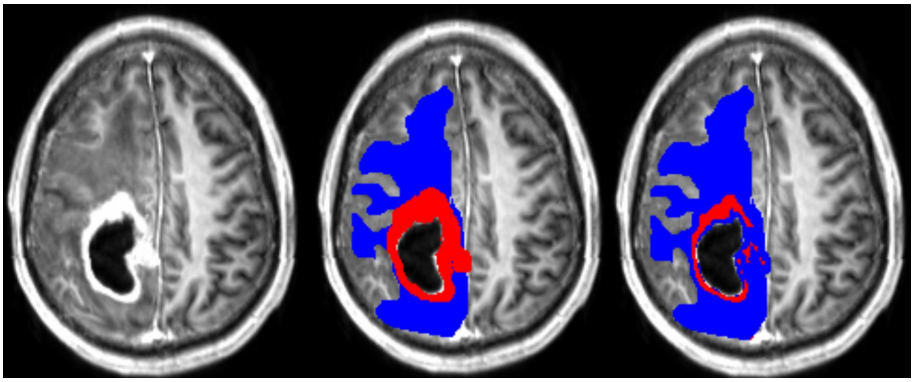


Figure 3.1: Left - T1pc. Center - Manual delineations on top of T1pc. Right - Semi-manual delineations on top of T1pc. In red there is the contrast enhancing region (CER), while in blue it is the non-enhancing region (NER).

Co-registration

All four cMRI maps, T1, T1pc, T2, and FLAIR, were first skull-stripped using FSL-BET with default parameters [209]. Afterwards, affine co-registration of skull-stripped T1, T2, and FLAIR to skull-stripped T1pc was done using NiftyReg [170] with default parameters. Three affine transformation matrices were saved and used to re-sample the corresponding original T1, T2 and FLAIR to the T1pc space.

To co-register PWI a similar protocol was used. Each PWI scan has 60 T2* brain volumes that can be selected to be co-registered to T1pc. In this study we select the first PWI brain volume and assume that the rest of them are

well aligned with it, ignoring any artifacts. After skull-stripping the first PWI volume using FSL-BET with default parameters, affine co-registration to skull-stripped T1pc was done using NiftyReg with default parameters. We obtain one affine transformation that is used to co-register all five PWI parameter maps (described in Subsection 3.2.2) to T1pc.

To co-register DKI a similar protocol was used. Each DKI scan has 10 T2 brain volumes that can be selected to be co-registered to T1pc. In this study we select the first DKI brain volume and assume that the rest of them are well aligned with it, ignoring any artifacts. After skull-stripping the first DKI volume using FSL-BET with default parameters, affine co-registration to skull-stripped T1pc was done using NiftyReg with default parameters. We obtain one affine transformation that is used to co-register all seven DKI parameter maps (described in Subsection 1.3.3) to T1pc.

Visual inspection of the tumour's centre in the axial plane of all maps for all patients after co-registration to T1pc was done by AIM and found no major misalignments. An example of all 16 maps for a random patient can be seen in Figure 3.2.

Feature extraction

After co-registering all maps to T1pc, the three ROIs (Total, CER, NER) are used as separate 3-D masks on each map to extract histogram and texture features. On each map, voxel intensities from each mask were normalized to the average value computed from the corresponding NAWM ROI. For each mask six histogram measures are computed: mean, coefficient of variation, 90^{th} percentile, 10^{th} percentile, skewness, and kurtosis. Additionally, twenty texture features are extracted from the 3-D GLCM [93]. To compute the GLCM, each map has been rescaled such that the voxel intensities are integers varying from 1 to 64.

The GLCM computation was done using the function *graycomatrix* implemented in Matlab R2015a (MathWorks, Massachusetts, U.S.A.) with distance set to 1, the $'Symmetric'$ flag set to *true*, and 4 values of $'Offset'$ set to the four main directions: 0°, 45°, 90°, and 135°. Twenty 3-D texture features, as described in [93], [211], and [37], were extracted from GLCM: autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of squares: variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information measure of correlation (IMC) 1 and 2, inverse difference normalized (IDN), and inverse difference moment normalized (IDMN).

In the end 416 features are extracted from each ROI: 26 histogram and texture features for each of the 16 maps. All features have been rescaled between 0 and 1, using all 183 labelled and unlabelled time points. In total, our dataset has 56 data points and 1248 features.

Datasets comparison

In this study we analyse two main influences: (1) features extracted from CER&NER vs. features extracted from Total tumour; (2) cMRI vs. PWI vs. DKI vs. all multi-parametric MRI features (cPD). Therefore, we split the original dataset according to the two main influences and create eight smaller datasets. These eight datasets were built using *complete* labelled time points, meaning labelled time points which have all MRI data available. Only complete time points were selected because in order to have a fair comparison between different MRI modalities, the number of points must not differ between them. All eight datasets have acquisitions from 29 patients and 55 time points. The ratio between progressive and responsive time points is slightly unbalanced, 34 to 21, or 62% vs. 38%. Table A.9 from Appendix shows the number of features in each dataset.

Feature selection methods

In order to avoid the curse-of-dimensionality [12], the number of data points should be much larger than the number of features. Because there are only 55 data points, different feature selection methods were used to reduce the dimension between 1 and 10 [106]. In this study six of the most widely known feature selection algorithms were used: minimum redundancy maximum relevance (mRMR) [55], RELIEFF [128], information gain (InfoGain) [246], Pearson's Chi^2 [246], random forest - mean decrease in accuracy (RF-MDA) [21], and random forest - mean decrease in Gini (RF-MDG) [21, 27]. The first four methods were run using the WEKA [89] application program interface (API) in Matlab R2015a. The last two methods were run using a random forest (RF) of 10000 trees in the statistical environment R [142]. Principal component analysis or other dimension reduction methods are not used because the biomedical meaning of the extracted features is lost.

Cross validation and performance measure

Given the fact that multiple data points per patient were acquired from 29 patients, a LOPOCV setup is used. In this way 29 folds are created in which

the test patient is always independent of the training patients: in each fold data points from one patient are considered test points, while data points from the remaining 28 patients are used for training.

In this study, two LOPOCV rounds are done, one for feature selection, and another one for classification using fixed feature sets.

In the first round, feature rankings are learned on the training sets, with the most important features at the top, and the least important at the bottom. Because each fold will have 6 different feature rankings as outputs from mRMR, Relieff, InfoGain, Chi^2, RF-MDA, and RF-MDG, each dataset will have 174 feature rankings. Combining different rankings is done by computing rank products [23] of each feature that appears at least once in top 10 of any feature ranking. The output of the first round of LOPOCV is a fixed set of 10 features selected by rank products per dataset shown in Table A.12 from Appendix.

In the second round, increasing number of features from 1 to 10 were used to classify data points. Classifiers are trained on the training set of each fold, then they assign a label to each testing data point from the test set. The assigned labels are compared to the true labels by measuring the BAR of all 55 test points. BAR, defined as the average between sensitivity and specificity, was preferred as a performance measure because the interest is in classifying correctly both labels (responsive and progressive). BAR can take values between 0 and 1, 1 pointing to a perfect classification and 0 to a completely wrong classification. A random classifier should give a BAR value of 0.5.

3.2.3 Classifiers

Several supervised classifiers have been used, with the goal of testing if data labelled according to the RANO criteria could have been reliably labelled using histogram and 3-D texture measures extracted from multi-parametric MRI. The list of classifiers as well as their software implementation environment is presented in Table A.10 from Appendix .

The list of classifiers in Table A.10 is representative for most of the classification algorithms, starting from simple linear ones such as Linear Discriminant Analysis (LDA) and Support Vector Machines with linear kernel (SVM-lin) up to more complex non-linear classifiers such as RF and Stochastic Gradient Boosting (SGB).

Fisher's LDA [73] is a classifier that finds a linear combination of input features that best separates the two classes. It is also very easy to use as there are no parameters that need to be set. Support Vector Machines [39, 43] is among

the most popular machine learning models because of its simplicity: given a training set with points from two classes, it tries to find the best hyperplane to differentiate between the two types of points. It can be used in the original feature space or the points can be mapped to another space by using kernel transformations. Two types of SVM kernel were used in this study: linear (SVMlin) and radial basis function (SVMrbf) with default settings ('C' and 'sigma' set to 1). Random forests [20, 21] are part of the ensemble methods for classification that use a collection of decision trees. Each decision tree learns a rule on a bootstrap sample of the original dataset and then it can classify a new point. The new point is assigned to the class voted by the majority of the trees. In this study RF was run with 1000 trees on all input features and 'class weight' set to 'balanced subsample', to adjust weights for each data point inversely proportional to class frequencies for each individual tree. Boosting algorithms [77,202] start with a collection of weak classifiers, in this case decision trees, and with each iteration they try to improve the overall classification by learning what was misclassified at the previous step. In this chapter the boosting algorithms had the following parameters: AdaBoost was run with 1000 trees; SGB was run with 1000 trees, 'learning rate' set to 0.1 (default) and 'subsample' set to 0.5, as suggested in [80]; Random Under Sampling Boosting (RUSBoost) was run with 1000 trees and 'LearnRate' set to 0.1.

3.3 Results

We compared seven classifiers, four MR modalities, two types of delineations, with a variable number of features from one to ten, summing up to a total of 560 BAR values. These BAR values were grouped in several ways (e.g. 56 groups of 10 values), and then a non-parametric comparison was made to account for statistical differences between groups. Multiple Kruskal-Wallis [132] rank tests were run in MATLAB R2015a to determine if all groups originate from the same distribution, followed by Dunn-Šidák's post-hoc test [59,205] to determine which group had the highest rank (better accuracy results). The relationship between BAR values and ranks is as follows: each BAR value is assigned a rank value, ignoring group membership. The assigned rank should be an integer, except for the case when there are multiple equal BAR values, then the assigned rank is the average of the individual ranks. For example, the BAR sequence [0.5, 0.7, 0.7, 0.9] is transformed into the rank sequence [1, 2.5, 2.5, 4].

Figures 3.3, 3.4, 3.5, and 3.6, show rank estimates and intervals of different groups. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. In each figure, the highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different

from the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different from the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap.

Figure 3.3 shows 56 groups, each group containing 10 BAR values coming from test runs with varying only the number of features. In the upper part there are 28 groups using CER&NER features, while in the lower part there are 28 groups using Total tumour features. The highest ranked group is CER&NER-cMRI-RUSBoost and its rank is significantly higher than 18 out of 28 groups achieved with Total tumour features.

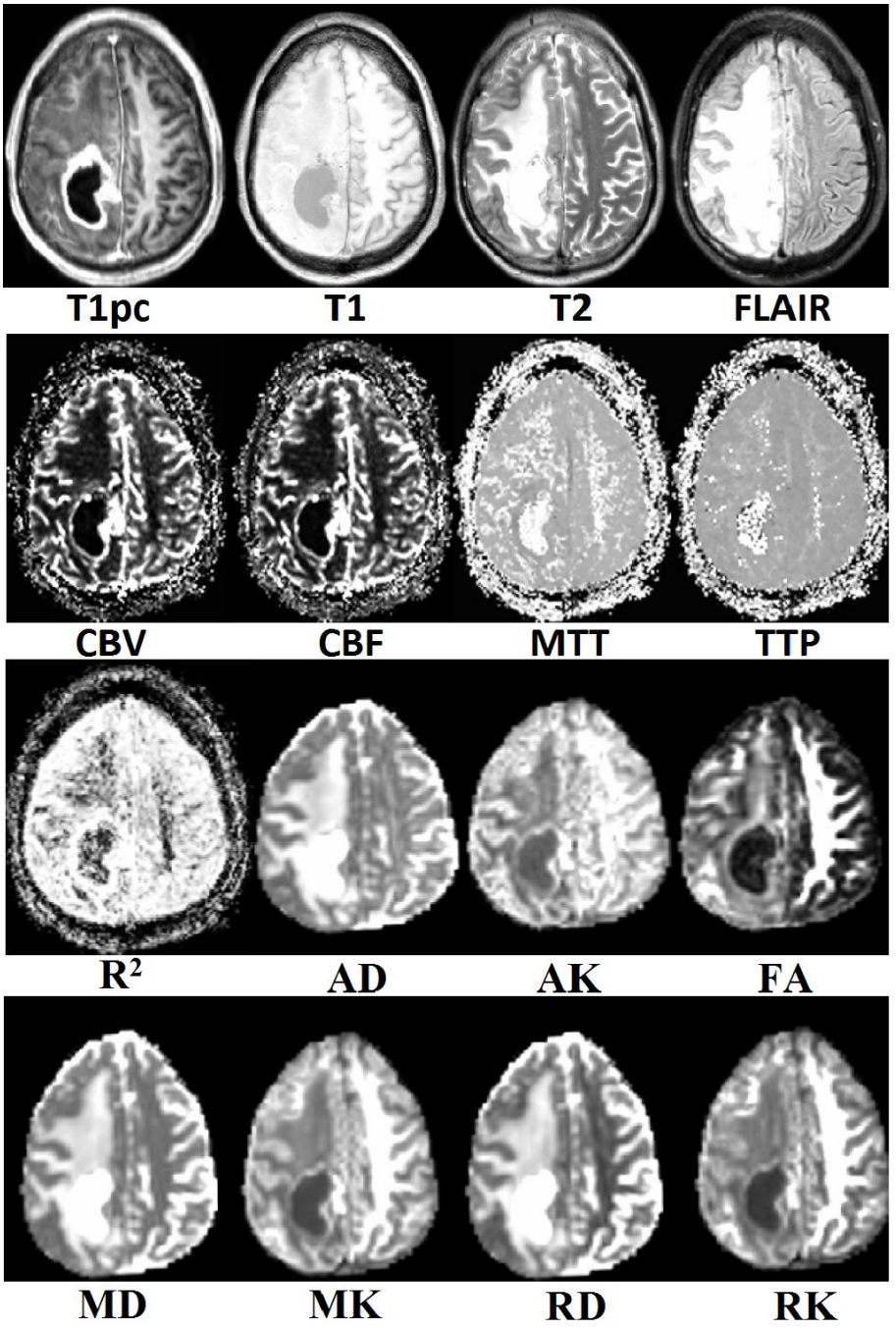


Figure 3.2: Example of co-registration results to T1pc for all multi-parametric magnetic resonance maps.

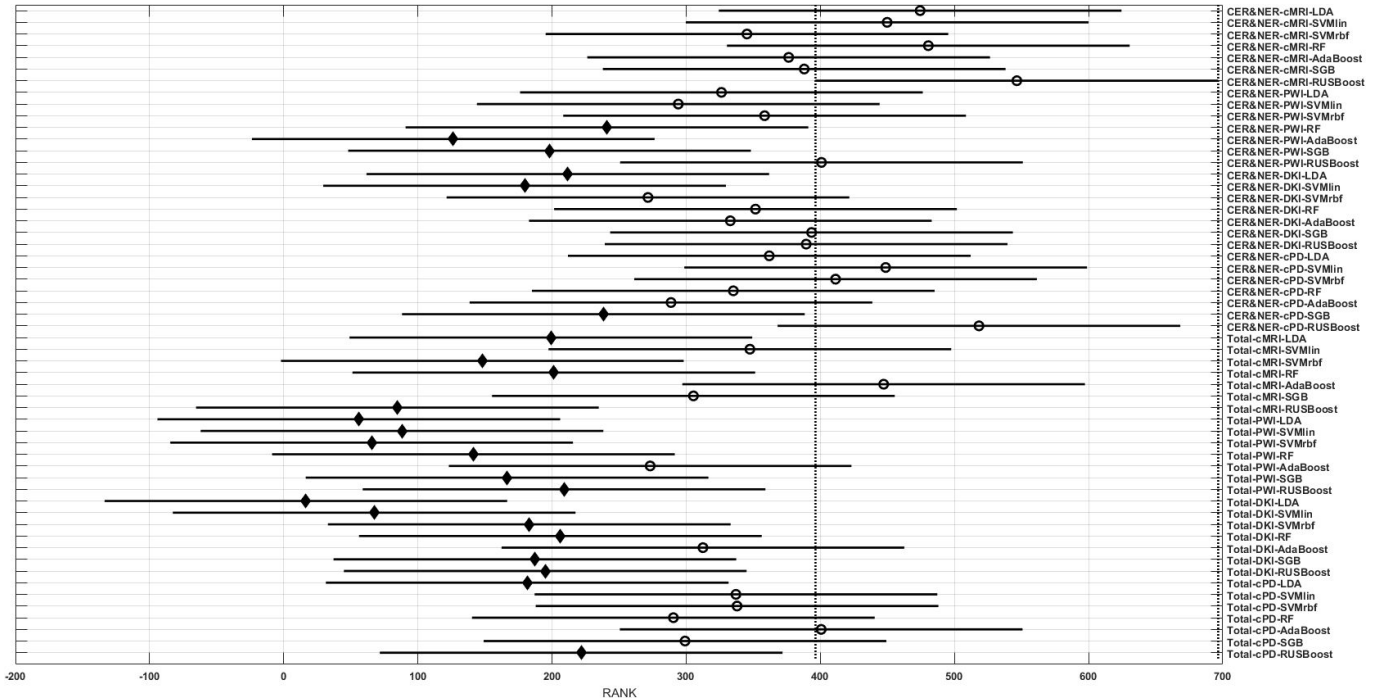


Figure 3.3: Rank estimates and confidence intervals for all combinations of classifiers, delineations, and MR modalities. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. The highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different than the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different than the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap. Each group has 10 BAR values, corresponding to 10 different features.

Figure 3.4 shows 8 groups, each group containing 70 BAR values coming from test runs with varying classifiers and number of features. In the upper part there are 4 groups using CER&NER features, while in the lower part there are 4 groups using Total tumour features. The highest ranked group has CER&NER-cMRI features and its rank is significantly higher than all groups using Total tumour features. Moreover, the CER&NER-cMRI group has a significantly higher rank than CER&NER-PWI and CER&NER-DKI. This means that classification based only on conventional MRI features performs better than the classification based only on perfusion or diffusion features.

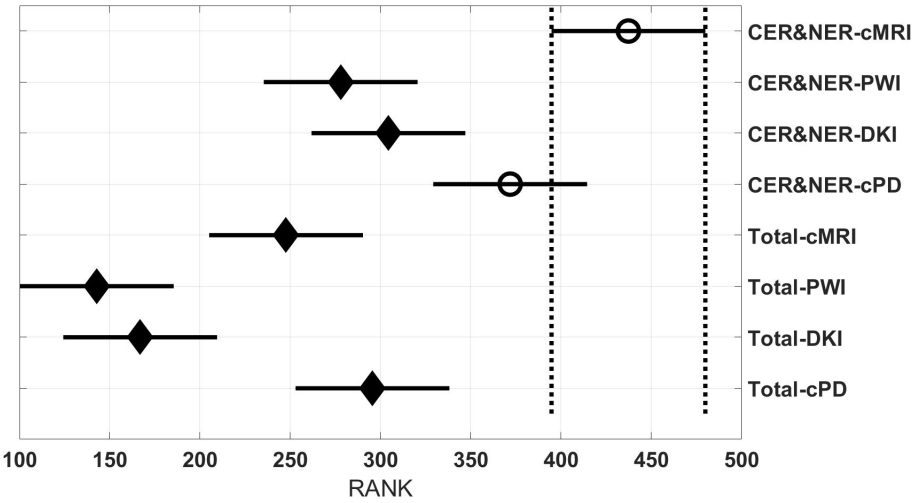


Figure 3.4: Rank estimates and confidence intervals for all combinations of delineations and MR modalities. CER - contrast enhancing region, NER - non-enhancing region. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. The highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different than the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different than the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap. Each group has 70 BAR values, corresponding to 10 features and 7 classifiers.

Figure 3.5 shows 14 groups, each group having 40 BAR values coming from test runs with varying MR modalities and number of features. In the upper part there are 7 groups using CER&NER features, while in the lower part there are 7 groups using Total tumour features. The highest ranked group is

CER&NER-RUSBoost and its rank is significantly higher than all groups using Total tumour features except AdaBoost.

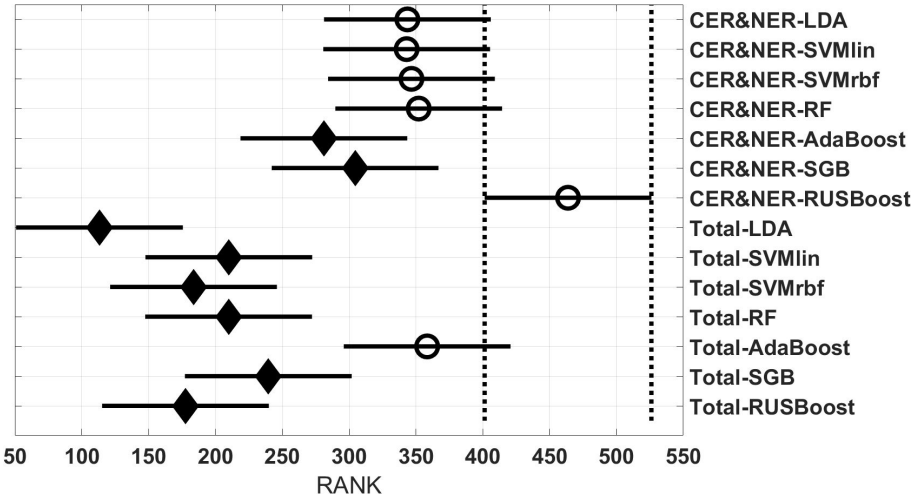


Figure 3.5: Rank estimates and confidence intervals for all combinations of delineations and classifiers. CER - contrast enhancing region, NER - non-enhancing region. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. The highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different than the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different than the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap. Each group has 40 BAR values, corresponding to 4 MR datasets and 10 features.

Figure 3.6 shows 20 groups, each group having 28 BAR values coming from test runs with varying MR modalities and classifiers. In the upper part there are 10 groups using CER&NER features, while in the lower part there are 10 groups using Total tumour features. The highest ranked group is CER&NER-Number of features:3 and its rank is significantly higher than all groups using Total tumour features except Total-Number of features:4.

Figure 3.7 shows the maximum BAR over all MR modalities for CER&NER and Total tumour ROIs, when varying the number of features from 1 to 10. In Appendix, associated with Figure 3.3, there is Table A.11. Figure A.1 and Figure A.2 from Appendix show results of each classifier when varying the

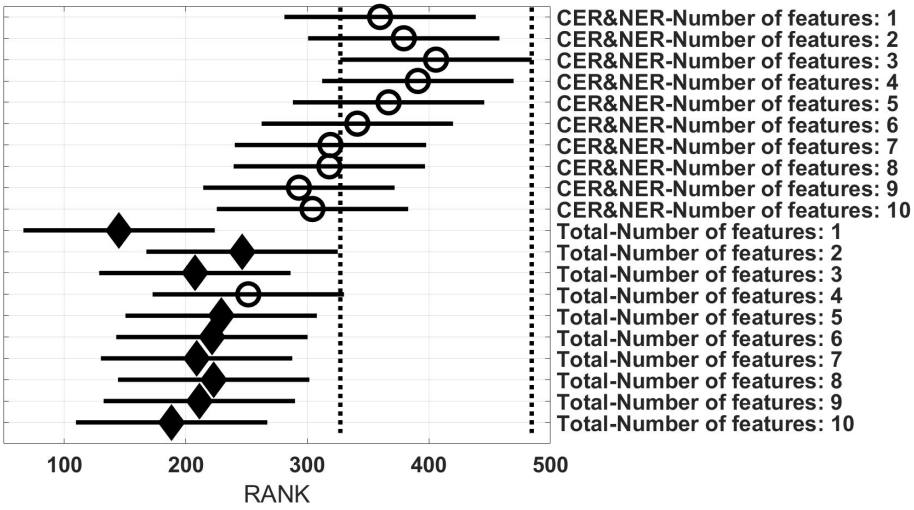


Figure 3.6: Rank estimates and confidence intervals for all combinations of delineations and varying number of features. CER - contrast enhancing region, NER - non-enhancing region. Intervals are shown as horizontal lines, while rank estimates are in the middle of the intervals. The highest ranked group has its interval limited by two vertical dotted lines. Groups that are significantly different than the highest ranked group have a filled diamond marker in the middle of their interval, while groups that are not significantly different than the highest ranked group have an empty circular marker in the middle of their interval. Two groups are significantly different if their intervals are disjoint; they are not significantly different if their intervals overlap. Each group has 28 BAR values, corresponding to 4 MR datasets and 7 classifiers.

number of features from 1 to 10 for each MR modality, for CER&NER and Total tumour ROIs, respectively.

Multiple remarks can be made after analysing the figures and tables previously presented. One of the most important remarks is that there is no combination of features, classifier, and delineations, that gives a perfect BAR value of 1. Although all post-hoc tests show the superiority of features extracted from contrast-enhancing and non-enhancing regions, the highest BAR value is achieved using total tumour features. To compare, the highest BAR value obtained using total tumour features is 0.956, with AdaBoost on cMRI, between 7 and 9 features, while the highest BAR value obtained using contrast-enhanced and non-enhanced features is 0.947, with SVMrbf trained on the first 3 cPD features. It is interesting to note that the maximum BAR value using total tumour features is achieved using only T1pc features, while the maximum BAR

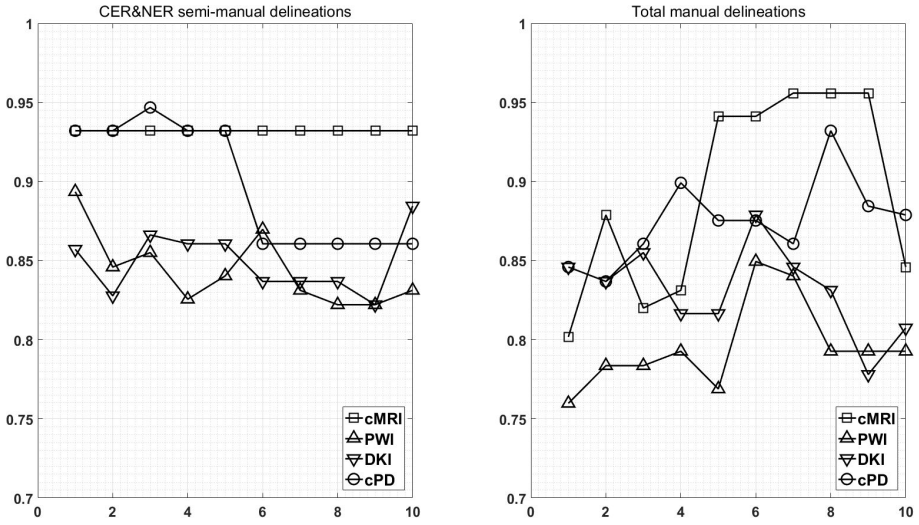


Figure 3.7: Maximum classification results over all MR modalities using 1 to 10 features. On y-axis are BAR values, and on x-axis the number of features used for classification. BAR - balanced accuracy rate, CER - contrast enhancing region, NER - non-enhancing region.

value using contrast-enhanced and non-enhanced features is achieved only after combining perfusion features with T1pc features. To be more specific, the maximum BAR value using contrast-enhanced and non-enhanced features is achieved only after training SVMrbf on the following three features: T1pc-average-CER, T1pc-90^{th} percentile-CER, and CBV-90^{th} percentile-CER.

Another notable result is the fact that RUSBoost trained only on just 1 feature (T1pc-Average-CER) achieves a remarkable BAR value of 0.932. This BAR value is maintained by RUSBoost trained on all 10 contrast-enhanced and non-enhanced conventional MRI features, indicating a robust classification. The same BAR value of 0.932 is achieved by RUSBoost also on the first 5 CER&NER-cPD features, which are histogram features extracted from T1pc, CBV and CBF. When we add the sixth cPD feature, a diffusion texture feature (ak-IMC2-CER), the BAR value drops to 0.86.

The post-hoc results reflect the consistent high BAR values of RUSBoost trained on contrast-enhanced and non-enhanced features, placing it at the two highest ranked positions in Figure 3.3. Although no group is significantly higher than the rest of groups, CER&NER-cMRI-RUSBoost is ranked significantly higher than most of Total-PWI and Total-DKI classifiers. Top 2 classifiers using Total

tumour features that are not significantly different than CER&NER-cMRI-RUSBoost are Total-AdaBoost-cMRI and Total-AdaBoost-cPD.

Figure 3.4 shows a surprising result: the best ranked group, compared to all but one, is CER&NER-cMRI. This is surprising because it means that there is no need to acquire perfusion or diffusion MRI, one can obtain high BAR values using only conventional MRI features. Although we are aware that the highest BAR value using contrast-enhanced and non-enhanced features was obtained after combining perfusion and conventional MRI features, BAR values obtained only on CER&NER-PWI or CER&NER-DKI features were ranked significantly lower than CER&NER-cMRI. BAR values obtained using Total tumour features with any kind of MR modality were also ranked significantly lower than CER&NER-cMRI.

Analysing Figure 3.5 we can make two remarks: RUSBoost is the best classifier using contrast-enhanced and non-enhanced features and AdaBoost is the best classifier using total tumour features. Analysing Figure 3.6 we see that for CER&NER, increasing the number of features above a threshold of 3 will decrease the BAR values, although not significantly.

It is clear from these four figures, Fig. 3.3, Fig. 3.4, Fig. 3.5, and Fig. 3.6, that splitting the total tumour into two regions, contrast-enhancing and non-enhancing, using a simple threshold like 90^{th} percentile, can improve significantly the classification accuracy.

3.4 Discussion

In Table A.12 we can see that feature selection for contrast-enhanced and non-enhanced features revealed an interesting result: only one feature was selected from the non-enhancing group (for DKI, 8^{th} position). This means that contrast-enhancing features are very meaningful and we could rely only on them in future works involving classification or statistical analysis.

When comparing the number of features selected from histogram or texture, even though we extracted more texture features (20 compared to 6), we can see a relative balance in all MR modalities except DKI. In top 10 CER&NER-cMRI features, 4 come from histogram; in top 10 CER&NER-PWI features, 4 come from histogram; in top 10 CER&NER-cPD features, 6 come from histogram; in top 10 CER&NER-DKI features, only 1 comes from histogram. In top 10 Total-cMRI features, 4 come from histogram; in top 10 Total-PWI features, 2 come from histogram; in top 10 Total-cPD features, 5 come from histogram; in top 10 Total-DKI features, none come from histogram. These feature rankings

strongly suggest that if only DKI data is available, one should definitely extract texture features to assess tumour recurrence.

When selecting cPD features, for both CER&NER and Total tumour ROIs, we can see in top 5 the same two features: T1pc-90^{th} percentile and CBV-90^{th} percentile. This selection comes as a confirmation of the majority of literature articles showing that contrast enhancement areas and CBV values are strongly correlated to tumour progression [9, 104, 105, 147]. The main reason behind this strong correlation is the fact that tumours grow uncontrollably, so they require more nutrients compared to surrounding tissue, which is reflected in the tumour's angiogenesis. The increase in angiogenesis is visualised and measured using T1pc and PWI.

There are multiple studies that focus on predicting the treatment outcome of follow-up GBM patients using multi-parametric MR data. The majority focuses mainly on overall survival, true progression vs. pseudo-progression or true progression vs. radiation injury. Some recent examples are the following: Elson et al. [64] show using DKI data from 52 patients that Apparent Diffusion Coefficient (ADC) values strongly correlate to overall survival; Smets et al. [208] conclude on 24 patients that absence of contrast enhancement on immediately post-operative T1pc correlates to an increase in overall survival; Zhank et al. [248] developed a new feature selection method using DKI data from 79 patients which gives an area under the receiver operating characteristic (ROC) curve (AUC) of 0.86 for separating true from pseudo-progression, without any manual segmentation; Bulik et al. [25] found significant differences in ADC and spectroscopic metabolites values between patients with true and pseudo-progression; Di Costanzo et al. [54] show, using data from 29 patients, that LDA trained on ADC, CBV, and normalized Choline gives a 96.6% accuracy in differentiating patients with true progression vs. radiation injury; Khalifa et al. [123] show that the fraction of hypoperfused tumour volume gives a 79.2% accuracy in anticipating tumour relapse at the next follow-up point.

Our study is, to our best knowledge, the only one that tries to classify progressive vs. responsive follow-up GBM patients based on multi-parametric MR data acquired at 3T. In the previous chapter we showed, using data from 18 patients, that PWI is a very powerful predictor of tumour recurrence, obtaining 100% accuracy in predicting the label one month before the label was put according to RANO criteria. In this chapter we used data acquired from 29 patients, therefore the classification problem is more difficult because of the increasing overlap between the classes. However, we still obtained maximum BAR values higher than 0.85 for each dataset: (i) Contrast-enhancing and Non-enhancing features - cMRI-0.932, PWI-0.894, DKI-0.885, cPD-0.947; (ii) Total tumour features - cMRI-0.956, PWI-0.85, DKI-0.879, cPD-0.932. Although the maximum value is achieved using features extracted from Total tumour ROI and not CER&NER

(0.956 vs. 0.947), we showed using non-parametric multiple comparison tests that it is recommended to use features from CER&NER, which could be defined by a simple threshold like the 90^{th} percentile on T1pc-Total ROI.

3.5 Conclusions

We proposed an automatic pipeline for processing multi-parametric MR data acquired at 3T and validated it after extracting histogram and GLCM 3-D texture features. We determined the added value of extracting features from semi-manually delineated contrast-enhancing and non-enhancing ROIs compared to features extracted from manual total tumour ROIs using non-parametric multiple comparison tests. We showed that AdaBoost, RUSBoost and SVM-rbf trained mainly on features extracted from T1pc and CBV maps achieve the highest ranked performance in classifying progressive vs. responsive follow-up GBM patients. Finally, our results suggest that using only conventional MRI features is better than using only perfusion or diffusion MRI features in the same classification problem.

Chapter 4

Classification of Recurrent Glioblastoma using modified Parametric Response Maps of contrast-enhanced T1-weighted MRI and Perfusion MRI

In this chapter the difference between progressive and responsive GBM patients was analyzed based on conventional and perfusion MRI data, each of them consisting of two MR parametric maps, T1pc and FLAIR for conventional MRI, and CBV and CBF for perfusion MRI. For each scanning session from a total of 268, the total tumour ROI was manually delineated by an expert clinician. For each patient out of 29, all delineations from different scanning sessions were grouped together, such that there is only one delineation per patient. The main differences compared to the previous chapter are the fact that previously there was one delineation per scanning session, as opposed to one delineation per patient in the current chapter, and the increased number of scanning sessions, from 183 to 268. Moreover, the number of labelled data points increased from 55 to 85. The new delineations were then used to follow the tumour evolution at the pixel level. Two separate masks were derived based on the new delineations

and using T1pc and CBV, indicating the most enhancing spots in time around the tumour area, called positive parametric response maps. Four features were extracted from each MR parametric map, summing up to 16 multi-parametric MR features for one mask. The focus was again only on the complete dataset, and results based on both masks, the positive parametric response T1pc map and the positive parametric response CBV map, were compared. The performance of SVM with linear and gaussian kernel was measured using a leave-one-patient-out cross-validation scheme, after the features were ranked according to the minimum-redundancy-maximum-relevance sorting algorithm. The work presented in this chapter is under review.

4.1 Introduction

GBM is the most common and malignant intracranial tumor [26], representing almost 30% of primary brain tumours with increasing incidence [56]. Patients have a median survival around 10 months after diagnosis, with maximum 5% of patients surviving more than three years. Recurrence is universal, and at the time of relapse, the median survival is five to seven months despite therapy [193]. The current standard of care is surgical resection followed by radiotherapy and concomitant temozolomide chemotherapy [216].

Identifying the location and size of brain tumours is done mainly by Magnetic Resonance Imaging (MRI). However, conventional MRI (cMRI) has limited specificity in determining the underlying type of brain tumour and tumour grade [53, 60]. Therefore, more advanced complementary magnetic resonance (MR) acquisition techniques such as DWI and PWI have been studied and shown to provide better tumour characterization [231]. DWI is a non-invasive imaging technique which provides information about water movement in biological environments, where it can interact with macromolecules, fibres, membranes, or other cellular structures. PWI provides measurements that reflect changes in blood volume, angiogenesis, and also glioma-induced neoangiogenesis, which show up as high relative cerebral blood volume (rCBV) and relative cerebral blood flow (rCBF) [147].

Parametric response map (PRM) is a voxel-wise approach for image analysis and quantification of alterations during treatment of high-grade tumours, which was proved to be a good predictor for the radiological response, OS, and PFS of GBM patients. The PRM technique was first developed for the diffusion parameter ADC, therefore it was first known as "functional diffusion map" (fDM). There is a large body of research regarding fDM for glioma patients [61–63, 90, 91, 152, 247], all articles showing that the amount of tumour volume which

shows an increased ADC compared to baseline is a predictor of radiological response, time to progression, OS, and PFS. PRM was then computed also for rCBV and rCBF [82], not only for ADC, and showed a lot of success in predicting overall survival, progression free survival, or discriminating true progressive from pseudoprogressive forms of GBM [3, 83, 229].

In the previous two chapters we showed that discriminating between responsive and progressive GBM patients treated with radio-chemotherapy and immunotherapy could be done with high accuracy using features extracted mainly from T1pc and PWI. Therefore, in this chapter we investigate the use of conventional and perfusion MRI features extracted by enhancing PRM masks computed using T1pc and rCBV, on a larger dataset than in our previous work.

4.2 Materials and Methods

4.2.1 Patient population

Study setup was described in Section 1.3.3. Each MRI time point for each patient was assessed based on the RANO criteria, and assigned to two major groups: either unlabelled for all time-points before the decision moment, or labelled as "responsive" or "progressive" for all time-points at and after the decision moment. In total there are 85 time points from 29 patients, with an approximate average of 3 time points per patient, out of which 36 are responsive and 49 are progressive, providing a relatively balanced dataset (42% vs. 58%).

4.2.2 MRI acquisition and processing

Magnetic Resonance (MR) scanning was performed on a clinical 3 Tesla MR imaging system (Philips Achieva, Best, The Netherlands), using a body coil for transmission and a 32-channel head coil for signal reception.

Conventional MRI

Conventional MR images were acquired as previously described in Section 1.3.3, but in this work we studied only T1pc and FLAIR. Two ROIs were manually drawn for each MRI session: a total tumour ROI on T2 maps around the entire lesion, avoiding areas of necrosis or cystic components (e.g. the surgical cavity), and a second ROI around the contralateral NAWM on T1pc to standardize the MRI measurements. All manual delineations were drawn by a radiologist (Sofie

Van Cauter) with 5 years experience on brain tumour MR imaging. For the modified PRM analysis, one large tumour delineation was used as the reunion of all tumour (RET) delineations from all time points, for each patient.

Perfusion MRI

PWI data were acquired as previously described in Section 1.3.3. Perfusion MR images were processed using the DSCoMAN plugin [18] for ImageJ [200], which takes into consideration leakage correction. We study only the two most important perfusion parameter maps, cerebral blood volume (CBV) and cerebral blood flow (CBF).

4.2.3 MRI Co-registration

For each patient, co-registration of all MRI maps was done with respect to the baseline T1pc map, following four steps of pre-processing: denoising, skull-stripping, bias field correction, and affine co-registration. To co-register the perfusion maps, only the first T2* brain volume was selected from a total of 60 T2* volumes that were used for computing the perfusion parameter maps, assuming the rest of the volumes are well aligned with the first one and ignoring any artifacts.

In the first phase, the three MRI brain volumes (T1pc, FLAIR, T2*) were denoised using the optimized blockwise nonlocal means denoising filter for 3D magnetic resonance images [40, 41], implemented in the *MRIdenoisingPackage* Matlab package (Matlab R2016a, MathWorks, Massachusetts, U.S.A.). In the second phase, skull-stripping was performed using FSL-BET with default parameters [119, 209, 210, 244]. In the third phase, bias field correction was performed using FSL-FAST with default parameters [249]. In the fourth and final phase, affine co-registration with 12 degrees of freedom was done using NiftyReg [170] with default parameters. The T2* affine transformation is then used to co-register the two PWI parameter maps (CBV and CBF) to the baseline T1pc.

The intensities of all four MRI maps were normalized by dividing to the corresponding NAWM ROI average. Visual inspection of the tumour's centre in the axial plane of all maps for all patients after co-registration was done by one author (AIM) and found no major misalignments. An example of 4 maps from the baseline and the second time point for a random patient can be seen in Figure 4.1.

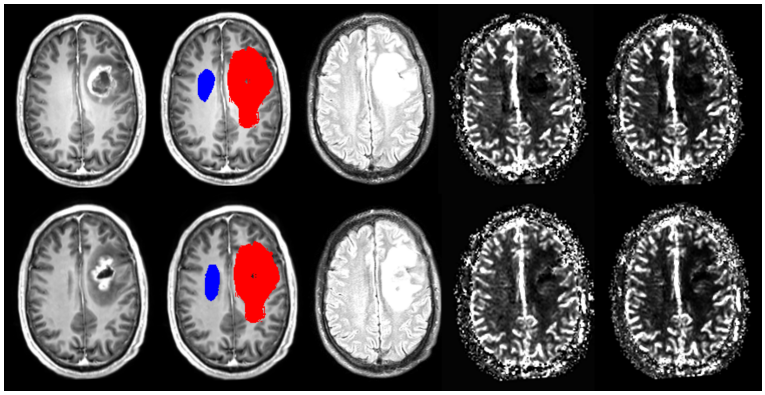


Figure 4.1: MRI Co-registration: on the top row there are baseline MRI maps, while on the bottom one there are MRI maps from the second time point. On both rows there are 5 columns, from left to right: (1) T1pc, (2) Reunion of total tumour ROIs from all time points (red) and NAWM ROI (blue) superimposed on T1pc, (3) FLAIR, (4) CBV, (5) CBF.

4.2.4 Feature extraction: Parameter Response Map

The classic PRM approach [82, 152] can be described by the following steps:

- *CE ROI intersection* - compute the Intersection of CE ROIs (ICE) from all time points
- *Building linear model on NAWM voxels* - fit a linear model using the NAWM voxels from the baseline of a few randomly selected patients as predictors for the NAWM voxels from the following time points
- *Compute the 95% confidence interval* - compute the 95% confidence interval (CI): [Low Threshold (LT), High Threshold (HT)]
- *Positive and Negative PRM* - Split the ICE tumour voxels in three groups based on the 95% CI: PRM_0 - unchanged voxels, PRM_+ and PRM_- if tumour voxel values are higher or lower than HT or LT, respectively
- *PRM features* - use as features the ICE volume fractions: $\frac{N_{PRM_+}}{N_{ICE}}$, $\frac{N_{PRM_-}}{N_{ICE}}$, $\frac{N_{PRM_-+PRM_+}}{N_{ICE}}$, where N_x is the number of voxels in a given ROI 'x'

We use a modified version of PRM, adapted for follow-up of post-operative patients with variable CE ROIs. The classic PRM approach typically gives a

small volume for follow-up, but there are patients who present little to none CE volume, therefore there is a need to extend this approach to follow the whole tumour volume, not only the CE volume. By doing so, we are also exploring the possible transformation of non-enhancing tissue into active tumour. The modified PRM approach can be described by the following steps, applied for each patient:

- *Total tumour ROI reunion* - compute the REunion of Total tumour ROIs (RET) from all time points
- *Building linear model on NAWM voxels* - fit a personalized linear model using the NAWM voxels from the baseline of each patient as predictors for the NAWM voxels from the following time points
- *Compute the 99% CI* - compute the 99% CI: [Low Threshold (LT), High Threshold (HT)]. The higher confidence value will ensure that only the most representative voxels will be selected from the extended tumour area
- *Positive PRM* - we study only the positive PRM, PRM_+, when tumour voxel values are higher than HT. We apply PRM_+ as a mask on four types of MRI difference maps: $T1pc_k$-$T1pc_{baseline}$, $FLAIR_k$-$FLAIR_{baseline}$, CBV_k-$CBV_{baseline}$, and CBF_k-$CBF_{baseline}$, where k is any time point different than the baseline
- *PRM features* - From each MRI difference map masked by PRM_+, we extract four percentiles that describe the relative increase to the baseline maps: 50^{th}, 70^{th}, 90^{th}, and 99^{th} percentiles, referred as 50p, 70p, 90p, and 99p, respectively.

We compare results obtained with the positive mask (PRM_+) computed using T1pc values (T1pc-PRM_+) and also using CBV values (CBV-PRM_+), because these two MRI maps are the most relevant for the follow-up of GBM patients, as shown in our earlier works [115, 116]. Examples of these two masks can be seen in Figure 4.2.

4.2.5 Feature selection: Minimum Redundancy Maximum Relevance

After extracting features as described above, using the PRM_+ mask on either T1pc or CBV, we will have a total of 16 features, from 4 difference MRI maps (T1pc, FLAIR, CBV, CBF) $\times$ 4 percentiles. In order to obtain the best classification results, we will apply a feature selection algorithm on three

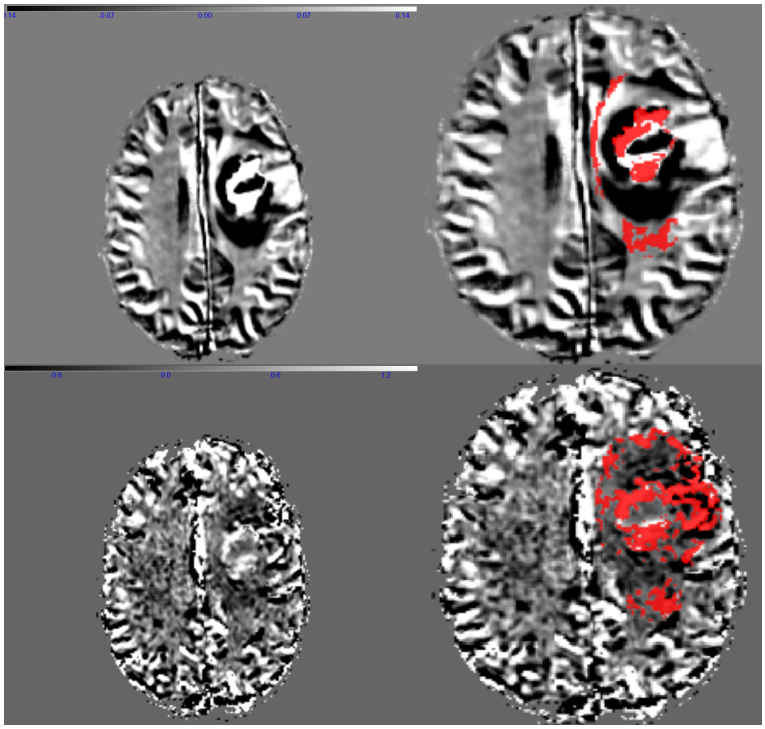


Figure 4.2: Comparison of T1pc-PRM$_+$ and CBV-PRM$_+$. First row, left corner: T1pc difference map between time point 2 and baseline, white/dark limits are +0.14 and -0.14. First row, right corner: T1pc-PRM$_+$ on top of the T1pc difference map. Second row, left corner: CBV difference map between time point 2 and baseline, white/dark limits are +1.3 and -0.7. Second row, right corner: CBV-PRM$_+$ on top of the CBV difference map. PRM$_+$ - positive parametric response map.

groups of MRI maps: cMRI (8 features), PWI (8 features), and cMRI+PWI, or cpMRI (16 features). Minimum Redundancy Maximum Relevance (mRMR) [55] is one of the most widely known feature selection algorithm, combining two major ideas: select features that are highly correlated to the class labels, or the maximum relevance features, while in the same time provide complementary information to each other, ensuring minimum redundancy. After using mRMR we will get a sorted list of the most relevant features for each dataset (cMRI, PWI, cpMRI).

4.2.6 Classifiers

For classifying responsive vs. progressive patients we use one of the most popular machine learning algorithm, support vector machines (SVM) [39]. Given a training set with data points from two classes, SVM tries to find the best hyperplane to differentiate between the two types of points. It can be used in the original feature space or the points can be mapped to another space by using kernel transformations. In this study we used two kernels, linear (SVM-lin) and radial basis function (SVM-rbf), implemented in Python 2.7.11 with scikit-learn 0.17.1 [171]. We tuned their parameters by optimizing the accuracy over a 5-fold cross validation on the training set within a logarithmic grid search: for SVM-lin we choose the best misclassification cost C between [0.001, 0.002, 0.005, 0.01, 0.02, 0.05, .., 100, 200, 500, 1000], while for SVM-rbf we choose the best pair of two parameters, the misclassification cost C, within the same boundaries as before, and γ, which is proportional to the inverse of a support vector's radius of influence, within the same boundaries as before. The small class unbalance was also adjusted by setting the *class weight* parameter to *balanced*, for both SVM-lin and SVM-rbf.

4.2.7 Performance measures

For measuring the performance of different methods, we use a LOPOCV setup, meaning that there are 29 folds and in each fold we learn a different classifier and test it on all time points of a patient. Feature scaling was learned on the training dataset from each fold and then applied to the test dataset. All time points from each patient will be only once in the test set, and 28 times they will be in the training set. In the end we aggregate the results from all 29 folds, which are probabilities of being in one class or the other, and compute AUC (**A**rea **U**nder receiver operating characteristic (ROC) **C**urve) [19]. Then, we find the optimal operating point of the ROC curve, and for that specific point (i.e. threshold) we compute three additional measures: sensitivity (TPR), specificity (TNR), and BAR, defined in Section 1.1.4. In our case, the positive class is the "progressive" label, while the negative class is the "responsive" label. The ROC curve can be created when the classification model gives probability values of test points belonging to one class, by plotting Sensitivity (y-axis) against 1-Specificity (x-axis) at various probability thresholds. A random classifier has an AUC of 0.5 or 50%, while a perfect classifier will have an AUC of 1 or 100%.

4.3 Results

AUC values obtained after training SVM-lin and SVM-rbf with features extracted from cpMRI, cMRI, and PWI, can be visualized in Figure 4.3, Figure 4.4, and Figure 4.5, respectively. For each figure, we compare results of classifiers trained on an increasing number of features from 1 to 8 or from 1 to 16, features extracted after using the two different masks, T1pc-PRM$_+$ and CBV-PRM$_+$.

Corresponding AUC, BAR, TPR, and TNR values associated with the 3 figures can be found in Table A.17 and Table A.18 in Appendix. Features selected by mRMR are listed in Table A.19, Table A.20, and Table A.21, also in Appendix.

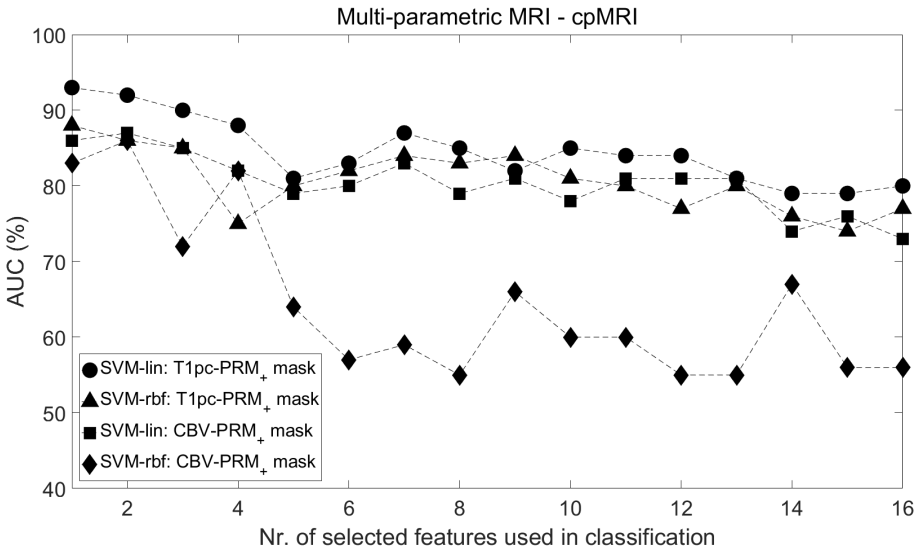


Figure 4.3: Area under the curve (AUC) values obtained by training SVM-lin and SVM-rbf using conventional and perfusion MRI (cpMRI) features extracted separately by the two positive parametric response maps, T1pc-PRM$_+$ and CBV-PRM$_+$. Training of the classifiers was done with an increasing number of features from 1 to 16, sorted using minimum-redundancy-maximum-relevance (mRMR).

The maximum AUC value for the multi-parametric MRI case (93%) is obtained after training SVM-lin only on T1pc-99p, extracted using the T1pc-PRM$_+$ mask. BAR, sensitivity, and specificity values associated to the maximum AUC are 91%, 85%, and 97%, respectively.

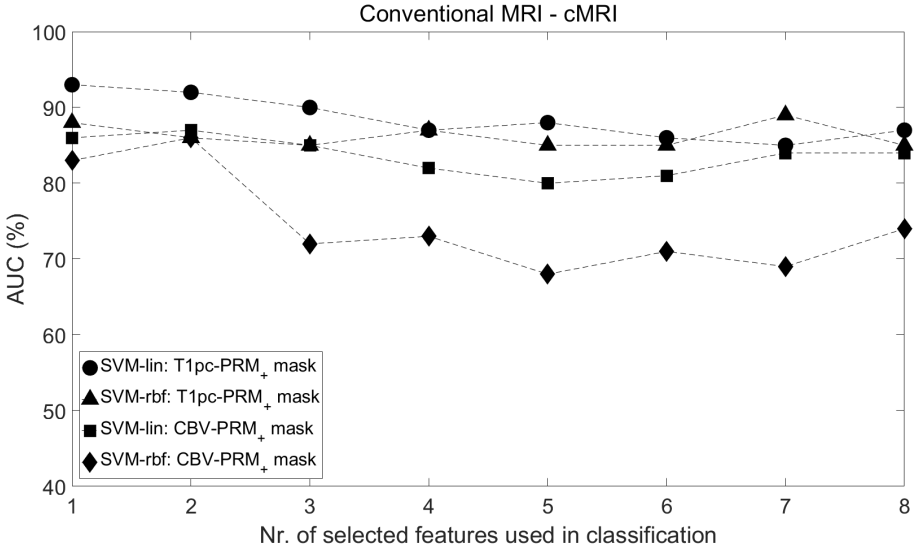


Figure 4.4: Area under the curve (AUC) values obtained by training SVM-lin and SVM-rbf using only conventional MRI (cMRI) features extracted separately by the two positive parametric response maps, T1pc-PRM$_+$ and CBV-PRM$_+$. Training of the classifiers was done with an increasing number of features from 1 to 8, sorted using minimum-redundancy-maximum-relevance (mRMR).

The maximum AUC value for the conventional MRI case (93%) is obtained as above: after training SVM-lin only on T1pc-99p, extracted using the T1pc-PRM$_+$ mask, with the same values for BAR, sensitivity, and specificity.

The maximum AUC value for the perfusion MRI case (73%) is obtained after training SVM-lin only on CBV-99p, extracted using the T1pc-PRM$_+$ mask, with BAR, sensitivity, and specificity values of 74%, 63%, and 84%.

4.4 Discussion

In this work we present a new PRM approach, different from what was already studied in the literature. Most of the published articles analyse only voxels from the intersection of contrast enhancing ROIs delineated on T1pc at separate time points, some of them imposing a minimum CE volume of 4 mL. Ellingson et al. [62] are the first ones that validate ADC-PRM on larger contrast enhancing ROIs delineated on FLAIR images. We extend this approach, by computing PRM even though there is no enhancement at baseline on T1pc, FLAIR, or T2.

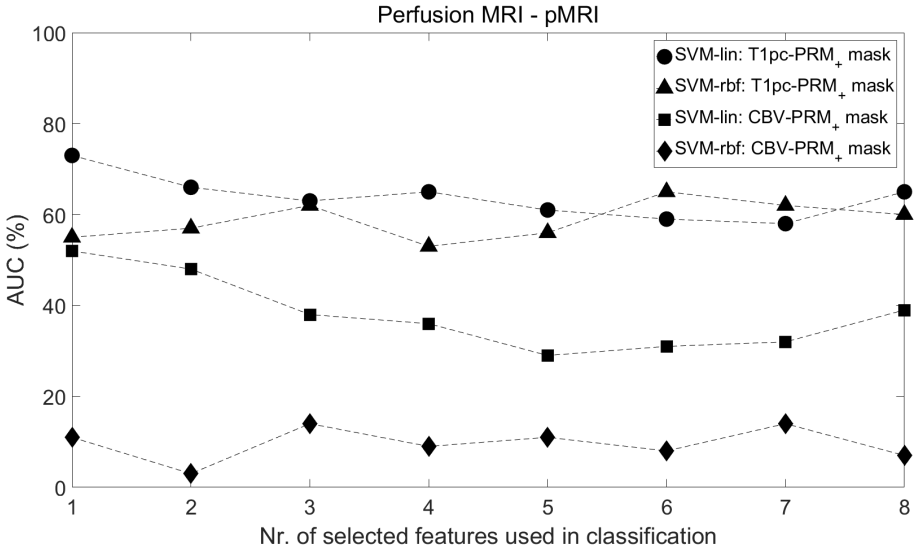


Figure 4.5: Area under the curve (AUC) values obtained by training SVM-lin and SVM-rbf using only perfusion MRI (PWI) features extracted separately by the two positive parametric response maps, T1pc-PRM$_+$ and CBV-PRM$_+$. Training of the classifiers was done with an increasing number of features from 1 to 8, sorted using minimum-redundancy-maximum-relevance (mRMR).

For this reason, the ROIs that we analyse could be regarded as the most likely regions for tumour infiltration. Another difference is how we fit the linear model and how we choose the confidence interval. While in the literature most authors use a single linear regression model for all patients, we propose a custom model designed for each patient. Because we use the 'reunion of total tumour ROI' (RET T2) which is clearly larger than the 'intersection of contrast enhancement ROI' (ICE T1pc), we also increased the confidence value of the prediction to 99%, in order to select only the most representative voxels out of a larger ROI.

In [63] Ellingson et al. used ICE FLAIR ROIs to extract the rate of change in ADC-PRM and report very high AUC values of 94% and 98%, for predicting progressive disease for 30 patients treated with radiochemotherapy and 20 patients treated with anti-angiogenic agent, respectively. Other studies also showed high AUC values when using ICE T1pc ROIs for computing ADC-PRM: in [247] the authors found AUC values of 90% and 80% for detecting progressive disease at 3 months and 1 year after baseline, respectively; in [91] the authors found an AUC of 72% for detecting progressive disease on 55 patients; in [90] the authors found an AUC of 83% for detecting progressive disease on 29

patients, while in [152] the authors found 100% accuracy for detecting stable and progressive disease on 20 patients.

In the four studies of CBV-PRM that we are aware of [3,82,83,229], there are no AUC values reported, only statistical analysis and linear models fit for OS and PFS. In [229] the authors report on 27 GBM patients that there are statistical significant differences between pseudoprogressive forms and true progressive forms using rCBV-PRM on ICE T1pc ROIs.

After analysing the figures presented in this study, we can safely conclude that T1pc-PRM helps in the extraction of consistently better features when compared to CBV-PRM. We also found the overall maximum AUC value of 93% using T1pc-PRM, while the maximum AUC value using CBV-PRM was 87%. After analysing Table A.17 and Table A.18 there are two clear conclusions: (1) using only 1 to 3 features yields the highest classification results, and (2) we obtain consistently high specificity values (i.e. 95%), and lower sensitivity values (i.e. 85%), with both T1pc-PRM and CBV-PRM. Therefore our new PRM approach is designed to detect mainly responsive GBM patients, and not progressive ones.

We do not use as features the classical percentages of change in tumour ROI because the RET T2 volume is significantly larger than the classical ICE T1pc volume, and there would not be enough information in the classical PRM features to discriminate progressive from responsive GBM patients. Therefore, we used histogram features expected to have high discriminative power, and sorted them using mRMR, a state of the art feature selection algorithm. In our previous work we showed that high values of CE T1pc and CBV can classify with very high accuracy progressive from responsive GBM patients. In this work we showed that maximum AUC values are obtained with T1pc-99p for T1pc-PRM and with CBV-99p for CBV-PRM.

Our study has several limitations that need to be discussed, considering that PRM is a voxel-wise approach and any distortion between baseline and labelled time points induces errors. First, only a relatively low number of patients are included in the analysis, which could be increased by performing a multi-center study. Second, the scanner's magnetic field may not be perfectly homogeneous, contributing to voxel-wise errors in each MRI acquisition. Third, quantification of CBV and CBF is done taking account the leakage, but accurate values are hard to obtain. Fourth, and last, coregistration effects are the ones that induce the largest errors due to the nature of GBM, because tumour infiltration and expansion are difficult to model. Therefore, we used a 12 degrees of freedom affine coregistration, just as in previous studies. We reduced as much as possible preprocessing errors by normalizing maps to their NAWM average, and coregistration errors by applying denoising, skull-stripping, and bias field

correction prior to coregistration, and checking the results visually.

Future studies are required, aimed on a larger number of patients, on minimizing normal structural differences between time points, with the ultimate goal of expanding the PRM analysis to the whole brain.

4.5 Conclusions

In this study we present a modified PRM approach, which is based on larger ROIs than classical PRM, taking into consideration a likely infiltrating area of the tumour and not only the contrast enhancing ROI. We test it on T1pc and CBV maps, and extract histogram features from 4 difference MRI maps: T1pc, FLAIR, CBV, and CBF. We sort features using mRMR and tune SVM-lin and SVM-rbf to discriminate progressive from responsive GBM patients. We report maximum AUC values of 93% and 87% with features extracted from T1pc-PRM and CBV-PRM, respectively. Therefore, we conclude that for GBM patients treated with radiochemotherapy and immunotherapy, T1pc-PRM may yield better results than CBV-PRM.

Chapter 5

Machine learning approach for classifying Multiple Sclerosis courses by combining clinical data with lesion loads and Magnetic Resonance metabolic features

In this chapter the difference between the four main multiple sclerosis courses (CIS, RR, PP, SP) was analyzed based on clinical and multi-parametric MRI data (conventional MRI and MRSI). A total of seven features were analyzed: clinical data consisted of three features (patient age, disease age, and EDSS), conventional MRI consisted of only one feature (lesion load), computed based on T1 and FLAIR, while MRSI consisted on three metabolic features (NAA/Cho, Cho/Cre, NAA/Cre). Eighty seven patients were scanned multiple times, summing up to 619 scans (data points), but because of missing data, in the end there were 592 scans with full complete data. Results based on different combinations of features are compared after training three supervised classifiers using a leave-one-patient-out cross-validation scheme. The work presented in this chapter was published in [114] in a slightly modified version.

5.1 Introduction

Multiple sclerosis (MS) is an inflammatory disorder of the brain and spinal cord in which focal lymphocytic infiltration leads to damage of myelin and axons [38]. MS affects approximately 2.5 million people worldwide, with an onset age commonly between 20 and 40 years, and an incidence more than twice as high in women compared to men [149].

The majority of MS patients (85%) usually experience a first attack defined as CIS, and will develop an RR form [151]. Two thirds of the RR patients will develop an SP form, while the other third will follow a benign course [198]. The rest of MS patients (15%) will start directly with a PP form.

The criteria to diagnose MS forms was originally formulated by McDonald in 2001 [150] and revised by Polman in 2005 [174] and 2011 [173]. They all rely on cMRI (T1, T1pc, T2, FLAIR) due to the high sensitivity for visualizing MS lesions. Conventional MRI is also used for quantifying lesion load (LL), a marker of inflammation process but only a moderate predictor of MS evolution [69].

More recently, advanced magnetic resonance techniques such as MRSI, DTI and MTI have been shown [191] to provide a better characterization of NAWM and thus a better understanding of the pathological mechanisms of MS. MTI metrics reflect the demyelination and remyelination processes and have been shown to predict the evolution of MS lesions. DTI metrics are very sensitive to the MS pathology and have been shown to be mainly affected by myelin loss and decreased neuronal integrity. MRS metrics provide high MS pathological specificity as well as high sensitivity to biochemical changes. Decrease of *N*-acetyl-aspartate (NAA) was observed in both chronic lesions and NAWM, reflecting a neuronal integrity loss [191]. Choline (Cho) and Creatine (Cre) contents were found to be increased in WM lesions and in NAWM, indicating the presence of severe demyelination and cell proliferation in relation with inflammatory processes [195, 223].

Therefore, in this chapter we investigate the added value of magnetic resonance metabolic features (NAA/Cho, NAA/Cre, Cho/Cre) combined with routinely collected clinical MS data (e.g. patient age, disease duration (DD), Expanded Disability Status Scale (EDSS)) and lesion load values (LL). To this purpose, we build multiple binary classifiers to automatically discriminate between different clinical forms of MS patients, by training each classifier on combinations of clinical data, lesion loads and metabolic features.

5.2 Materials and Methods

5.2.1 Patient population

The patient population was described in Section 1.4.3, while the most relevant information is condensed in Table 5.1. Eighty-seven MS patients (12 CIS, 30 RR, 28 SP and 17 PP) part of the AMSEP project were studied against 18 volunteers without any neurological disorders, who served as healthy control (HC) subjects. Diagnosis and disease course were established according to the McDonald criteria [146, 150].

	CIS	RR	PP	SP
Number of patients (Male/Female)	12 (6/6)	30 (6/24)	17 (6/11)	28 (17/11)
Age at first scan [years]	31.8 (6.4)	33.2 (7)	39.5 (6)	41.1 (4.8)
Disease duration [years]	2.9 (1.9)	8.3 (4.8)	7.5 (2.9)	14.9 (6.1)
EDSS median [range]	1 (0-4)	2 (0-5.5)	4 (2-7.5)	5 (3-8.5)
Lesion Load [ml]	6.6 (3.5)	16.7 (12.6)	20.8 (13)	31 (12.9)
Total number of scans (data points)	62	226	125	206

Table 5.1: Patient population: Age - average value (standard deviation); Disease duration - average value (standard deviation); EDSS - median (minimum - maximum); Lesion Load - average value (standard deviation). The four multiple sclerosis (MS) groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

5.2.2 Longitudinal MS data

In total there are 619 MS scans, but because of missing lesion loads and metabolic features, there are 592 (95.6%) scans with full complete data, leading to an average of 6-7 complete scans/patient.

5.2.3 MRI acquisition and processing

All patients and control subjects underwent MR examination using a 1.5 Tesla MR system (Sonata Siemens, Erlangen, Germany) and an 8 elements phased-array head-coil.

Conventional MRI

Conventional MRI data (T1 and FLAIR) was acquired as previously described in Section 1.4.3.

MRSI acquisition and processing

MRSI data was acquired and processed as previously described in Section 1.4.3, resulting in the quantification of three metabolites: NAA, Choline, and Creatine.

Quality control

Voxels with CRLB lower than 10% for *each* of the three metabolites (NAA, Cho, and Cre) were kept as having "good quality" for the next step of feature extraction. If the number of "good quality" voxels is lower than 50% of the total amount of voxels in the MRSI grid, then the acquisition is discarded. All 18 Control subjects had MRSI data with a number of "good quality" voxels higher than 50% of the total amount of voxels, and 606 out of 619 (97.9%) MRSI data from MS patients had good quality as defined earlier.

5.2.4 Feature extraction

In this study we use three types of features: clinical (e.g. patient age, disease duration, and EDSS), lesion loads, and metabolic features. The clinical features are routinely acquired in the hospital. The lesion loads were computed based on T1 and FLAIR, using the MSmetrix software [118] developed by ico**metrix** (Leuven, Belgium). The computation of metabolic features was performed in two steps: three metabolic ratios (NAA/Cho, NAA/Cre, Cho/Cre) were computed for each "good quality" voxel and then averaged, leading to three metabolic features extracted from each MRSI grid.

5.2.5 Training approach

Nine binary classification tasks were studied: HC vs. CIS, HC vs. RR, HC vs. PP, HC vs. RR+SP, HC vs. PP+SP, CIS vs. RR, CIS vs. RR+SP, RR vs. PP, RR vs. SP. The first three tasks investigated differences between HC and the starting MS forms (CIS, RR, and PP). The next task investigated differences between HC and MS patients that are likely to evolve or had evolved into secondary

progressive form (RR+SP). Afterwards, we investigated differences between HC and definite progressive forms (PP+SP). The next two tasks investigated differences between CIS patients and the most likely progression of CIS, namely RR and RR+SP. From a neurological point of view, the last two tasks were the most intriguing, as they were discriminating between the most common inflammatory MS form (RR) and the two progressive forms, PP and SP.

For each task, data normalization was performed. We use a LOPOCV setup combined with 100 random patient-based bootstrap selections for the training set. In this way, the test set has all data points of one patient, while the training set has $n-1$ data points corresponding to $n-1$ patients, where n is the total number of patients, different for each classification task (e.g. for HC vs. CIS, $n = 30$). Basically, to construct the training set, we randomly select one data point from each patient assigned to the training set. The test set always includes all data points of the test patient. We repeat the procedure 100 times and store the results. Each data point from the test set will be assigned 100 times to either class 1 or class 2, and in the end it will be assigned to one of the classes according to majority voting. This procedure is repeated until all patients from each classification task have been tested.

By using this random patient-based bootstrap selection, the two classes in the training set have a more balanced distribution of points (18 HC, 12 CIS, 30 RR, 17 PP, 28 SP), compared to using the total number of points of each class (18 HC, 61 CIS, 214 RR, 121 PP, 196 SP).

5.2.6 Performance measures and statistical testing

For each task we computed and reported four measures, in percentage: F1-score, sensitivity, specificity, and BAR, which were explained in Section 1.1.4.

Throughout our study the positive class was the first class from each of the nine binary classification tasks: HC for the first 5 tasks, CIS for the 6^{th} and 7^{th} tasks, and RR for 8^{th} and 9^{th} tasks.

In order to correctly assess if there are significant differences between the four MS groups, we built several linear mixed effects models which were able to incorporate the temporal evolution of each patient's MS course. We used five fixed effects and two random effects. The fixed effects are: MS course, gender, disease onset age, disease duration, and the interaction between MS course and disease duration. The random effects are set for each patient allowing an individual starting point and an individual disease evolution. The most interesting fixed effect for this study is the first one, which represents the average of the response variable at the beginning of the MS course, or when 'disease

duration' = 0. We built four linear mixed effects models, one for each response variable: NAA/Cho, NAA/Cre, Cho/Cre, and lesion load. All statistical models were built in the 'R' software using the "lme4" package [11], statistical testing was done using the "lmerTest" package [134] and post-hoc analysis was done using the "multcomp" package [102]. All tests were done for a significance level (α) of 0.05.

5.2.7 Classifiers

Three supervised classifiers implemented in Python 2.7.11 with scikit-learn 0.17.1 [171] have been used: LDA, SVM, and RF. We tuned each classifier's parameters by optimizing the F1-score over a 5-fold cross validation on the training set within a grid search of individual parameters, specified further for each particular classifier. Fisher's LDA [73] is based on a linear combination of input features, with three possible solvers: singular value decomposition, least squares solution, and eigenvalue decomposition. Tuning involved choosing between the first solver and the last two solvers combined with shrinkage varying from 0 to 1 in steps of 0.1. Class unbalance was adjusted by setting the *priors* parameter equal to class probabilities. We use SVM [39, 43] with a radial basis function kernel (SVM-rbf), defined by two parameters: C, or the misclassification cost, and γ, which is proportional to the inverse of a support vector's radius of influence. We tuned C and γ by performing a logarithmic grid search between 0.00001 and 100000. Class unbalance was adjusted by setting the *class weight* parameter to *balanced*. Random Forests [21] is based on a group of decision trees. We tune the number of decision trees between 200, 400, 600, 800, and 1000. Class unbalance was adjusted by setting the *class weight* parameter to *balanced subsample*.

5.3 Results

Figure 5.1 shows boxplots comparing MR metabolic features (A, B, C) and lesion loads (D) extracted from HC and each MS course. Boxplots are drawn using default style in MatLab, meaning the middle line inside the box represents the median value, the vertical limits are the 25^{th} and 75^{th} percentiles (q_1 and q_3), each whisker covers 1.5 the interquartile range (i.e. $q_3 - q_1$), and the crosses outside the whiskers represent outliers. Figures A.3, A.4, A.5, and A.6 from Appendix show the MS data points in various 2-D feature spaces.

Using the previously described (Section 5.2.6) linear mixed-effects models we found that the fixed effect MS course is statistically significant in the evolution

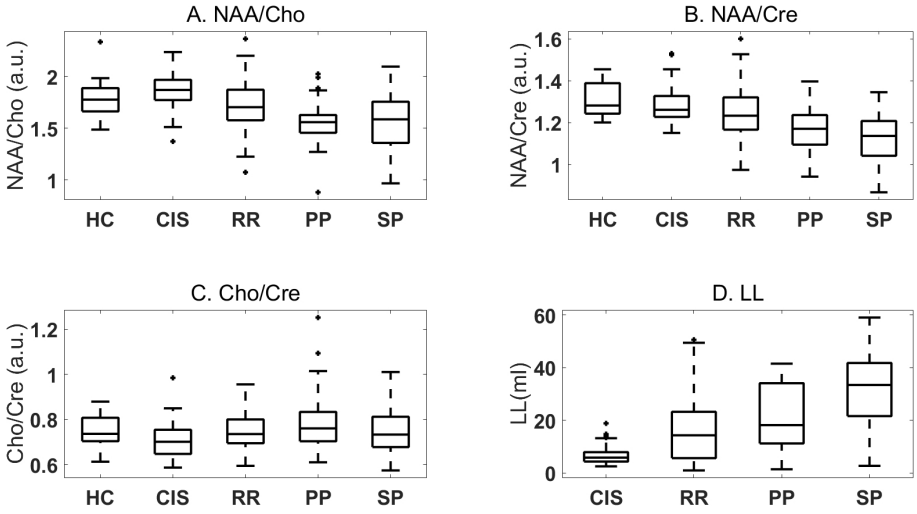


Figure 5.1: Box-plots of magnetic resonance metabolic features and lesion loads extracted from healthy controls (HC) and multiple sclerosis (MS) patients: A. NAA/Cho; B. NAA/Cre; C. Cho/Cre; D. Lesion load (LL). The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

of NAA/Cho, NAA/Cre, Cho/Cre, and LL, with corresponding p-values of: $3.4 \times 10^{-6}, 2 \times 10^{-4}$, 2.3×10^{-2}, and 2.6×10^{-4}. Table 5.2 provides adjusted p-values for multiple comparisons between the MS groups.

	CIS - RR	RR - PP	RR - SP
NAA/Cho	-	**	**
NAA/Cre	-	-	*
Cho/Cre	-	-	-
LL	-	-	*

Table 5.2: Adjusted p-values for multiple comparisons between multiple sclerosis (MS) groups modelled by linear mixed effects model, tested using the "multcomp" package in 'R' (* for $p < 0.05$ and ** for $p < 0.01$). The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

Table 5.3 shows F1-scores after training LDA using only metabolic ratios, as clinical data and lesion loads were not available for healthy controls. Corresponding BAR, sensitivity and specificity values of this table can be

found in Table A.13 in Appendix. If F1-scores are missing, then the classifier assigned all data points to the negative class (second MS group).

	NAA/Cho	NAA/Cre	Cho/Cre	All 3 metabolic ratios
HC vs. CIS	35	33	43	36
HC vs. RR	6	16	-	14
HC vs. PP	47	45	19	49
HC vs. RR+SP	8	19	-	16
HC vs. PP+SP	21	26	-	28
CIS vs. RR	15	-	-	21
CIS vs. RR+SP	3	-	-	19
RR vs. PP	75	78	75	74
RR vs. SP	60	67	58	69

Table 5.3: F1-scores for all nine classification tasks (rows) after training LDA using only metabolic ratios. Values above 75 are coloured in light gray. HC - healthy controls, CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

Surprisingly, the F1-scores for separating HC from any MS course are very low, and the same holds true for separating very early MS form (CIS) and the most likely MS evolution, RR and RR+SP. In contrast, for RR vs. PP we find that all three metabolic ratios have F1-scores higher than 75, with a maximum of 78 for NAA/Cre. For RR vs. SP the F1-scores are lower, with a maximum of 69 after combining all metabolic features.

Table 5.4 shows F1-scores of classification tasks involving only MS patients. Training was done on seven different combinations of features to evaluate the classification power of clinical data, lesion loads, and metabolic features. Corresponding BAR, sensitivity, and specificity values can be found in Appendix in Tables A.14, A.15, andA.16, respectively. If F1-scores are missing, then the classifier assigned all data points to the negative class (second MS group).

The highest F1-scores for CIS vs. RR and CIS vs. RR+SP, respectively 71 and 72, were achieved by SVM-rbf trained on clinical data and lesion loads. Training any classifier only on metabolic features yielded very low F1-scores.

The highest F1-score for RR vs. PP (85) was achieved by LDA using patient age, disease age, and EDSS. Adding all spectroscopic information maintained the F1-score at 85, while adding lesion load lowered the F1-score at 79. LDA outperformed SVM-rbf and RF in all RR vs. PP cases, always achieving an F1-score higher than 70.

The highest value for RR vs. SP (87) was first achieved after training SVM-rbf on clinical and metabolic features, but also with LDA trained on all

	CIS vs. RR			CIS vs. RR+SP			RR vs. PP			RR vs. SP		
	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF
M	21	48	11	19	31	-	74	52	73	69	70	67
LL	-	51	27	-	40	24	71	19	73	75	77	68
Age + DD	48	58	51	44	56	50	79	64	74	76	75	71
Age + DD + EDSS	55	65	49	57	66	48	85	81	79	84	85	84
Age + DD + EDSS + LL	67	71	59	63	72	60	79	75	79	86	86	86
Age + DD + EDSS + M	56	59	48	60	59	51	85	83	80	86	87	85
Age + DD + EDSS + LL + M	65	64	57	65	63	57	83	81	78	87	87	86

Table 5.4: F1-scores for classification tasks (columns) involving only multiple sclerosis (MS) patients. Abbreviations: M = all three average metabolic ratios; Age = patient age; DD = disease duration; LL = lesion load; EDSS = Expanded Disability Status Scale. Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, while values larger than 85 are coloured in dark gray. CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

features combined (clinical data, lesion loads, and metabolic features). SVM-rbf outperformed LDA in the majority RR vs. SP cases, but only with 1 to 2%.

5.4 Discussion

In this chapter, we present results for nine binary classification problems using clinical data, lesion loads and metabolic features extracted from MS patients and healthy controls. We focused on metabolic features as numerous studies showed significant metabolic alterations in MS patients of different MS forms. It has been demonstrated that metabolic abnormalities in MS patients are not restricted to lesions alone [57, 81, 96, 109, 158, 160, 196] and NAWM tissue is well known to be altered in MS [50, 157]. Concentrations of NAA in NAWM were shown to be significantly lower in MS patients [14, 15, 112, 219, 226, 237, 238]. Concentrations of Cho and Cre in NAWM were shown to be significantly higher in MS patients [112, 158, 219, 223, 227]. Concentrations of NAA/Cre in NAWM were shown to be significantly lower in MS patients [140, 159]. Multiple studies also report significant differences between metabolite concentrations in lesions vs. NAWM of HC: lower NAA and increased Cho and Cre [4, 47, 48, 96, 137, 158, 243].

Our findings are in agreement with these previous reports as decreased NAA and increased Cho and Cre contents were measured in NAWM and lesions of MS patients. After building linear mixed-effects models to properly analyse the statistical difference between the four clinical courses, we observed significant differences at the disease starting points of all MS courses using four response variables, namely the lesion load, NAA/Cre, NAA/Cho, and Cho/Cre ratios. A cross-sectional study [92] based on a subset of our MRSI data found statistical

differences in the NAA/Cre and NAA/Cho ratios between HC and RR, PP, SP, and RR+PP+SP patients. To our knowledge, there is only one study that reports sensitivity and specificity values for classifying healthy controls from MS patients based on spectroscopic features. Inglese et al. show in [112] that absolute values of choline in NAWM can differentiate 9 controls and 10 out of 11 RR patients.

Other MS classification studies are [155] and [127], both based on diffusion features. The first one reports a classification accuracy of 97% between 20 CIS and 33 RR patients. The second one analyses classification tasks based on DTI data from a cross-sectional subset of our database. They found very high F1-scores (91.8% for both HC-CIS and CIS-RR) after training SVM-rbf on six global brain connectivity metrics. For RR vs. PP their maximum F1-score was 75.6%, which is lower than our results based on metabolic features, while for RR vs. SP, their maximum F1-score was 85.5%, which is comparable to our results. It is also worth mentioning that they did not use any clinical data, which might improve their results.

In this study, we analysed the added value of combining standard clinical data with quantitative magnetic resonance features. To this purpose, we trained linear and non-linear classifiers only on advanced MR features, and then only on clinical data. Afterwards we train the classifiers on clinical data combined with lesion loads and metabolic features.

Although MS patients are expected to have significantly different WM metabolism compared to healthy controls, this difference was not reflected in the metabolic average obtained from "good quality" voxels (Figure A.3, A and B). This result is not entirely surprising, considering that we averaged over a high number of voxels, and the subtle lesion information could be lost in the average. However, we can visually see in Figure A.3:C&D that the two progressive MS courses tend to have lower NAA/Cho and NAA/Cre ratios than healthy controls.

CIS and RR patients' distribution over the NAA/Cho and NAA/Cre feature space do not differ much, as seen in Figure A.4:A. Disease duration interval for RR patients is much larger than for CIS patients, as most of CIS patients have a disease duration lower than 5 years, which can be seen in Figure A.5:A. Because RR patients have more relapses than CIS patients, the number of lesions will be higher and the lesion volume as well, while EDSS scores are mainly in the same range, as seen in Figure A.6:A. BAR values in Table A.14 show a maximum of 85, when combining patient age, disease duration, EDSS, and lesion load. However, the corresponding maximum F1-score of 71 is much lower because the dataset is unbalanced (61 CIS vs. 214 RR), heavily influencing the classifier's precision. In this case the F1-score reflects better than BAR the difficulty of

discriminating CIS from RR forms.

CIS and SP patients' distribution over different features is visible in Figure A.4:B, Figure A.5:B, and Figure A.6:B, and it is clear that these two are the least and most advanced forms of MS. Because RR patients will eventually evolve into SP forms during their lifetime, we grouped together RR and SP patients for a separate classification task versus CIS patients. BAR values in Table A.14 show a maximum of 92, when combining patient age, disease duration, EDSS and lesion load. The same discussion as for CIS vs. RR apply: the corresponding maximum F1-score is only 72 because the dataset is very unbalanced (61 CIS vs. 410 RR+SP) and the precision will be very low.

RR and PP patients can be discriminated using only EDSS by visually inspecting Figure A.6:C. Training a linear classifier on clinical data (patient age, disease duration, and EDSS) gives the maximum F1-score of 85. Adding the 3 metabolic features keeps the score at 85, while adding lesion load information lowers the score to 79. This drop in the F1-score suggests that lesion load is not useful in differentiating RR from PP patients. Indeed, these two MS forms have the closest lesion load averages (16.7 ml and 20.8 ml), as shown in Table 5.1. In contrast, the clinical status of RR and PP patients are very different, as reflected by the EDSS values of 2 for RR and 4 for PP. Moreover, training LDA on individual metabolic features always provided higher F1-scores than lesion load, therefore we can conclude that for RR vs. PP, metabolic features have a higher discrimination power than LL. BAR values in Table A.14 are also closer to the F1-scores in Table 5.4 because the dataset is more balanced compared to previous cases.

RR and SP patients can also be discriminated using only EDSS by visually inspecting Figure A.6:D. Our results showed that EDSS is very important in differentiating RR patients from primary or secondary progressive patients. We also report consistent higher F1-scores for classifiers trained only on lesion load compared to classifiers trained only on metabolic features. Furthermore, it is clearly visible in Table 5.3 that we obtain higher F1-scores for this classification task using multiple features, compared to the rest of 8 tasks. These findings suggest that in the future it might be possible to build a decision support system using clinical data combined with lesion loads and metabolic features.

However, this study suffers from a few limitations caused by the low scanning frequency of only 1.5 Tesla. Firstly, it is known that the sensitivity of lesion load segmentation is improved by scanning at higher frequencies [204]. Therefore, our LL values may not reflect entirely the pathological changes inside the brain. Secondly, the signal to noise ratio of MRSI is proportional to the scanning frequency, meaning our metabolites' quantification is not entirely accurate. In order to obtain true metabolites values, we would have to measure T1 and T2

relaxation times of water for each patient, which would be impossible in clinical practice. Moreover, spectroscopic signal scales can differ from patient to patient, resulting in large metabolite variations. To overcome some of these limitations, we use as features all three metabolite ratios (NAA/Cho, NAA/Cre, Cho/Cre). By doing so, we expect to retain the most valuable information.

When comparing classification tasks from a computational point of view, LDA is clearly the winner as the training period last only 3 hours using a computer with 8 threads. Training both SVM-rbf and RF took around 20 days in total and it was done using 60 threads, meaning LDA is approximately 600 times faster than SVM-rbf or RF. Also, the maximum F1-scores for RR vs. PP and RR vs. SP were obtained with LDA and SVM-rbf, suggesting that a linear classifier performs equally good as a non-linear classifier in these cases.

This study is a proof of concept that investigates the added value of MR metabolites combined with clinical data and lesion loads, in classifying MS patients and healthy controls. Clinical data is routinely collected by doctors, lesion load is a known marker of neurodegeneration, while MR metabolites have been shown to provide high specificity of MS pathology. In order to better understand the underlying MS pathological mechanisms, we used three different machine learning methods, one linear and two non-linear, and had a strict quality control for extracting metabolic features. Despite all our efforts, averaging metabolite ratios over "good quality" voxels provides only moderate biomarkers for discriminating MS groups (i.e. RR vs. PP). In general, combining patient age, disease duration, EDSS, and averaged metabolic ratios, leads to the highest classification results. We believe extracting metabolic information from specific brain sub-regions of the MRSI grid (e.g NAWM) should provide a more detailed view of MS pathology and help the classification tasks. Therefore, further investigations about the MS patients' evolution will be done in the future on sub-regions metabolite quantification, DTI-based brain connectivity metrics, patient treatment, and multi-class classification.

5.5 Conclusions

In this chapter we performed nine binary classification tasks and report F1-scores and BAR values after learning linear and non-linear classifiers on combinations of clinical data, lesion loads, and metabolic features. We presented a simple method to compute metabolic features by averaging metabolite ratios over "good quality" voxels of an MRSI grid. Using linear mixed-effects models we found that the MS course is statistically significant in the evolution of four response variables: Lesion Load, NAA/Cre, NAA/Cho, and Cho/Cre ratios.

Our results showed that the best classifier for discriminating CIS from RR or RR+SP is SVM-rbf trained on clinical data and lesion loads. We also showed that discriminating RR from PP or SP with high accuracy is possible when training LDA on clinical data. For RR vs. PP, adding metabolic features will not change the results, while for RR vs. SP, adding metabolic features and lesion loads will slightly improve the results.

Chapter 6

A comparison of Machine Learning approaches for classifying Multiple Sclerosis courses using MRSI and brain segmentations

In this chapter the difference between the four main multiple sclerosis courses (CIS, RR, PP, SP) was analyzed based only on multi-parametric MRI data (conventional MRI and MRSI). Four feature extraction models (M1, M2, M3, M4) were compared after training classifiers using a two-fold cross-validation scheme. The four models had different input dimensionality: (M1) the whole spectroscopic Fourier domain (81 features), (M2) only three features computed based on the spectroscopic Fourier domain, (M3) six features: the three spectroscopic features computed in M2 and three conventional MRI features, and (M4) the spectrogram of the time-domain spectroscopic signal, of size 128×57. The same dataset as in the previous chapter was used, consisting of eighty-seven patients scanned multiple times, summing up to approximately 51000 spectroscopy voxels, or data points. The work presented in this chapter was accepted for publication as a conference paper at "The 26th International Conference on Artificial Neural Networks, ICANN2017", in a slightly modified version.

6.1 Introduction

Multiple sclerosis (MS) is an inflammatory disorder of the brain and spinal cord [38], affecting approximately 2.5 million people worldwide.

The majority of MS patients (85%) usually experience a first attack defined as Clinically Isolated Syndrome (CIS), and will develop a relapsing-remitting (RR) form [151]. Two thirds of the RR patients will develop a secondary progressive (SP) form, while the other third will follow a benign course [198]. The rest of MS patients (15%) will start directly with a primary progressive (PP) form.

The criteria to diagnose MS forms were originally formulated by McDonald in 2001 [150] and revised by Polman in 2005 [174] and 2011 [173]. They all rely on using conventional magnetic resonance imaging techniques (MRI), such as T1 and FLAIR, due to high sensitivity in visualizing MS lesions. More recently [191], ^{1}H-Magnetic Resonance Spectroscopic Imaging (MRSI) has been shown to provide a better understanding of the pathological mechanisms of MS.

The objective of this study is to fully explore the potential of MRSI for automatic classification of MS courses. To this purpose we use four different machine learning approaches to classify individual spectroscopic voxels inside the brain. We start by using simple machine learning methods (i.e. Linear Discriminant Analysis (LDA)) trained on low-level features commonly used in MRSI, and advance up to state-of-the-art methods (e.g. Convolutional Neural Networks (CNN)) trained on high-level MRSI features.

6.2 Materials and Methods

6.2.1 Patient population

The patient population was described in Section 1.4.3, while the most relevant information is condensed in Table 6.1.

	CIS	RR	PP	SP
Number of patients	12	30	17	28
Total number of scans	60	212	117	192
Total number of voxels	5916	18682	10830	17377

Table 6.1: Multiple sclerosis (MS) patient population details. CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

6.2.2 Magnetic Resonance data acquisition and processing

All patients underwent MR examination using a 1.5 Tesla MR system (Sonata Siemens, Erlangen, Germany) and an 8 elements phased-array head-coil.

MRI acquisition and processing

Conventional MRI data (T1 and FLAIR) was acquired as previously described in Section 1.4.3. Three tissues of the brain, gray matter (GM), white matter (WM), and lesions, were segmented based on T1 and FLAIR, using the MSmetrix software [118] developed by ico**metrix** (Leuven, Belgium).

MRSI acquisition and processing

MRSI data was acquired and processed as previously described in Section 1.4.3, resulting in the quantification of three metabolites, NAA, Cho, and Cre.

Quality control

For each voxel inside a grid, we performed three outlier detections, corresponding to each metabolite, using the median absolute deviation filtering. Next step keeps only voxels with a maximum CRLB error of 20% for each metabolite, preserved by all three outlier detection mechanisms. In the end, the average voxel exclusion rate was 31%, with 6% standard deviation, and only 2 out of 581 spectroscopy grids had an exclusion rate higher than 50%.

6.2.3 Classification tasks and performance measures

We study four binary classification tasks, relevant from a clinical point of view: CIS vs. RR, CIS vs. PP, RR vs. PP, and RR vs. SP. For each task we set the less represented class between the two to be the positive class, or the class of interest. Therefore, we set the positive class to CIS, CIS, PP, and SP, corresponding to each task. When classifying, we perform a 2-fold stratified cross-validation at the patient level, meaning that each patient will be assigned once to training, and once to testing. The training dataset includes all voxels from all patients assigned to training. When testing, a voxel will be assigned to one of the two classes. For each grid, we compute the probability to be assigned to the positive class by measuring the percentage of voxels assigned to the positive class. We

measure and report the area under receiver operating characteristic curve (AUC) and sensitivity and specificity at the optimal operating point of ROC.

6.2.4 Feature extraction models

Model nr.1 (M1)

We use the absolute values of the complex frequency spectrum cut by a pass-band filter between 1.2 and 4.2 ppm, so that we retain the most useful information. In order to have a perfect alignment of all spectra for all patients, we detect the highest peak in the low frequencies (NAA) and shift to the NAA peak of a randomly assigned reference voxel. In this case, each voxel is represented by the filtered frequency vector, which has 81 points. We normalize each vector to its L_2-norm.

Model nr.2 (M2)

We use the three quantified metabolite concentrations (NAA, Cho, Cre) to compute three ratios: NAA/Cho, NAA/Cre, and Cho/Cre. Mean values and standard deviations for each MS group can be found in Table 6.2.

	CIS	RR	PP	SP
NAA/Cho	2.21 (0.24)	2.02 (0.25)	1.83 (0.18)	1.86 (0.32)
NAA/Cre	1.36 (0.1)	1.35 (0.11)	1.27 (0.11)	1.22 (0.12)
Cho/Cre	0.63 (0.07)	0.69 (0.08)	0.72 (0.1)	0.69 (0.1)

Table 6.2: Multiple sclerosis (MS) metabolite ratios - mean (standard deviation). CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

Model nr.3 (M3)

For each voxel, we measure the percentage of each tissue of the brain (GM, WM, lesions). In this case, each voxel is represented by 6 features: three metabolic ratios and three tissues percentages.

Model nr.4 (M4)

For each voxel, we compute the spectrogram of its time-domain signal. First, we interpolate the time-domain signal from 512 to 1024 points. We compute the spectrogram using a moving window of 128 points, with an overlap of 112 points. In the end, each voxel will be represented by a 128×57 image. These values have been especially selected such that the final image is large enough to be used as input in CNNs.

6.2.5 Classifiers

For each classification task and for each of the first three feature extraction models, we used three supervised classifiers: (1) LDA [73] without adjusting for class unbalance, (2) Random Forest [21] with 1000 trees, adjusted for class unbalance by setting the *class weight* parameter to *balanced subsample*, and (3) SVM-rbf [39], adjusted for class unbalance by setting the *class weight* parameter to *balanced*, and tuned the misclassification cost "C" by searching between [0.1, 1, 10, 100] over a 5-fold cross-validation loop. The *gamma* parameter was set to *auto*. All classifiers were built in Python 2.7.11 with scikit-learn 0.17.1 [171]. Feature scaling was performed using the training set, only for the second and third model.

For the last feature extraction model and for each classification task, we built a CNN inspired by [207] using the Keras package [36] based on Theano [225]. Our architecture consists of 8 weighted layers: 6 convolutional (conv) and 2 fully connected (FC). All convolutional layers have a receptive field of 3×3 and the *border mode* parameter set to 'same'. All weighted layers are equipped with the rectification non-linearity (ReLU). Spatial pooling is carried out by 3 max-pooling (MP) layers over a 2×2 window with stride 2. The first FC layer has 64 channels, while the second one has only 2, because it performs the two-class classification. The final layer is the sigmoid layer. To regularise the training, we used a Dropout layer (D) between the two FC layers, with ratio set to 0.8. A simplified version of our architecture is (conv-conv-MP-conv-conv-MP-conv-conv-MP-FC(64)-D(0.8)-FC(2)-Sigmoid). When training each CNN, we used the 'adadelta' optimizer, the 'categorical crossentropy' loss function, and we split the training dataset into 70-30 training-validation data. We stopped training after 200 epochs, and for each classification task, validation accuracy was at a stable value over 85%, signalling that training was performed correctly.

6.3 Results and Discussion

All performance measures can be found in Table 6.3. Maximum AUC values for each classification task are highlighted in gray.

Percentage [%]		M1			M2			M3			M4
		LDA	RF	SVM-rbf	LDA	RF	SVM-rbf	LDA	RF	SVM-rbf	CNN
CIS vs. RR	AUC	65	50	63	53	55	66	63	76	77	71
	Sensitivity	0	0	38	2	0	13	2	28	25	17
	Specificity	100	100	83	100	100	99	100	96	100	98
CIS vs. PP	AUC	89	92	88	87	90	90	88	91	95	83
	Sensitivity	68	68	63	67	72	78	65	77	83	73
	Specificity	93	95	94	91	90	89	91	87	90	82
RR vs. PP	AUC	66	62	68	64	64	68	55	54	57	68
	Sensitivity	21	17	50	29	37	56	0	0	0	28
	Specificity	93	94	78	87	82	76	100	100	100	92
RR vs. SP	AUC	72	72	73	73	71	72	73	71	71	69
	Sensitivity	60	54	57	40	43	48	51	38	29	56
	Specificity	75	84	77	90	86	81	82	92	97	75

Table 6.3: Area under the curve (AUC), Sensitivity, and Specificity values for all classifiers, feature extraction models (M1-M4), and classification tasks. Dimensionality of the models: M1 - 81 (metabolic spectra), M2 - 3 (metabolic features), M3 - 6 (3 metabolic and 3 tissue percentages), M4 - CNN - input image is 128×57. CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

For CIS vs. RR we obtain a maximum AUC of 77% when combining metabolite ratios with GM, WM, and lesions percentage. The increase in AUC for both SVM-rbf and RF is higher than 10% when we compare M3 to M1 or M2, therefore we can conclude that adding GM, WM, and lesions percentage, is indeed beneficial when classifying CIS vs. RR courses. This is most probably due to the fact that RR patients have more lesions than CIS patients. It is worth mentioning that the CNN, which takes as input only the MRSI spectrogram, performs better than all other classifiers based on spectroscopic features.

For CIS vs. PP we obtain a maximum AUC of 95% when combining metabolite ratios with GM, WM, and lesion percentages in each voxel. The increase in AUC for SVM-rbf is higher than 5% when we compare M3 to M1 or M2. This task is not too interesting from the medical point of view, because we know that PP patients have a more aggressive form of MS and a higher lesion load than CIS patients. Our results confirm the clinical background and provide an accurate classification with high sensitivity for PP.

For RR vs. PP we obtain the lowest AUC value of the four classification tasks, only 68%. It is interesting to see that adding GM, WM, and lesion percentages did not improve the results, but on the contrary. This indicates an opposing

effect between brain segmentation percentages and metabolic ratios. Another interesting fact is that maximum results obtained with M1, M2, or M4, are exactly the same, indicating that spectroscopy is not sensitive enough to classify these two MS courses.

For RR vs. SP we obtain a maximum AUC value of 73%, if we use M1, M2, or M3. The are two main observations to be made: (1) LDA trained on metabolic ratios can be regarded as the best classifier for this task, due to a simple feature extraction model and high computational speed, and (2) adding brain segmentation percentages did not improve the results.

The main goal of this study was to compare different levels of extracting information from the MRSI voxels. To that extent, at the low-level we used only 3 metabolite ratios, at the mid-level we used the entire absolute frequency spectrum of 81 points, and at the high-level we used the MRSI spectrograms, of size 128×57. To boost the low-level features, we added the brain tissue segmentations percentages of WM, GM, and lesions. We used spectrograms as input to state of the art classifiers (e.g. CNNs), and compared the results with widely used machine learning algorithms (e.g. LDA, RF, SVM-rbf) trained on features commonly used in MRSI. We observe that results obtained with CNNs are not significantly worse or better than the rest. It is worth noting that overall, training CNN provides a pipeline for both feature extraction and classification, and is less time-consuming than all MRSI pre-processing and quantification needed for classical machine learning algorithms.

Our results show that there is an inherent limitation of our particular MRSI protocol to classify MS courses. Combining low-level MRSI features with brain tissue segmentations percentages can improve classification between the least aggressive MS course (CIS) and the moderate-severe courses (RR and PP). However, there are obvious limitations on any level of the MRSI features when classifying moderate (RR) from severe MS courses (PP and SP). Recent results [127] have shown that features extracted from MR diffusion brain maps can improve classification between RR and SP. In the future we will also incorporate diffusion features and perform multi-class classification.

6.4 Conclusions

In this chapter we performed four binary classification tasks for discriminating between MS courses. We report AUC, sensitivity, and specificity values, after training simple and complex classifiers on four different types of features. We show that combining metabolic ratios with brain tissue segmentation percentages can improve classification results between CIS and RR or PP patients. Our

best results are always obtained with SVM-rbf, there we can conclude that our CNN architecture does not add any improvement over classical machine learning methods.

Chapter 7

Conclusions

7.1 General conclusions

In this thesis we proved the applicability of supervised machine learning algorithms to multi-parametric MRI data in order to predict and discriminate the evolution of abnormal brain injuries, such as GBM and MS, by always obtaining high classification accuracies.

In Chapter 1 we gave a soft introduction into the fields of machine learning and MRI. We described three of the most popular classifiers' concepts: (1) SVM, the most widely used algorithm in the last 20 years, (2) RF, the best off-the-shelf algorithm, and (3) deep learning, which is the current paradigm when working on big data and will probably shape the future of every possible field. We also described basic physical concepts of MRI, as well as the complementarity of conventional MRI (T1,T2, FLAIR, T1pc) and advanced MRI (PWI, DKI, MRSI). Afterwards, MRI acquisition sequences and data preprocessing pipelines were described for two large clinical centres, UZ Leuven (Belgium) and CERMEP (France), investigating two of the most important brain diseases, GBM and MS, respectively.

Then, the thesis was branched into two parallel tracks, one for GBM data analysis and the other for MS data analysis.

In Chapter 2 we analysed the difference between progressive and responsive GBM patients based only on advanced MRI data: PWI, DKI, and MRSI. The main conclusion was that it is possible, using ensemble classifiers (e.g. Random Forest) trained only on PWI data, to perfectly separate between progressive and

responsive GBM patients, and moreover, perfect classification was possible with one month prior to the doctor's decision. However, this study was limited by three factors: (1) a relatively low number of patients (i.e. 18), mainly influenced by the low quality of the MRSI data, (2) extraction of multi-parametric MRI features was done separately for each MRI map, therefore it was very time consuming, and (3) a clinician had to manually delineate for each map two ROIs, contrast enhancing and total tumour.

In Chapter 3 we focused on the same problem as in the previous chapter, but tried to solve its limitations. Therefore, instead of MRSI data we used conventional MRI data, we developed an automated preprocessing pipeline for PWI and DKI, and we proposed a new way of delineating CE ROIs based on the T1pc map and the total tumour ROI. With these modifications we increased the number of patients to 29, we were able to do batch preprocessing, with the final MRI parameter maps not showing any major errors on a visual validation, and we theoretically reduced by half the time spent by clinician for delineations. We did a thorough comparison between the impact of manual ROIs and semi-manual ROIs by searching for the best histogram and texture features, then training classifiers on subsets of the best 10 features, separately for cMRI, PWI, and DKI, but also on all of them combined. Following a post-hoc statistical analysis, we came to three main conclusions: (1) results obtained with features extracted from semi-manual ROIs are better than the ones obtained with features extracted from manual ROIs, (2) the best machine learning algorithm suited for this problem was, again, an ensemble classifier (i.e. RUSBoost), and (3) conventional MRI features gave the best results. Some limitations remained still unsolved, such as the moderate number of patients, which could not be changed, and the fact that for each MR scanning session a clinician still had to delineate a total tumour ROI.

In Chapter 4 we focused only on the last limitation from the previous study and analysed only the most sensitive MRI modalities for GBM detection, grading, and follow-up: T1pc, FLAIR, CBV, and CBF. Therefore, we defined a new total tumour ROI by reuniting all tumour delineations done on T2 from all time points of a patient. This new large area can be regarded as the most likely infiltration area of the tumour. We coregistered all MRI data of each patient to its baseline T1pc and performed a PRM anaylsis, which is a voxel-wise selection of enhancing areas between two time points. We compared the effects of computing PRM using T1pc and CBV separately, by extracting histogram features and training SVM with linear and gaussian kernels. Our final results show that using the T1pc-PRM mask is clearly superior to using CBV-PRM.

In Chapter 5 we analysed the difference between 18 healthy control subjects and 87 patients affected by four different MS forms (CIS, RR, PP, SP), based on three types of features: (1) clinical data (e.g. patient age, disease age,

EDSS), (2) lesion loads, and (3) metabolic ratios averaged over a MRSI grid (NAA/Cho, NAA/Cre, Cho/Cre). After training LDA separately on the three metabolic ratios, we found only moderate discrimination results between the 18 healthy subjects and any MS group, although statistical differences were observed. After training linear and non-linear classifiers on different feature combinations, we determined the following: CIS form is very difficult to detect from RR, or RR+SP; RR form could be discriminated from PP or SP only based on clinical data; lesion load does not help in discriminating RR from PP, while metabolic ratios could help; both lesion loads and metabolic ratios could help in discriminating RR from SP; LDA performed equally good as SVM-rbf or RF, but in a much shorter time. The limitations of this study were mainly related to the scanner frequency, because lesion loads quantification and metabolites quantification would have been more reliable if the scanner frequency was larger than 1.5 Tesla. However, this was a hard limit as the data was already acquired and could not be overcome.

In Chapter 6 we focused on the MRSI voxel-wise analysis of MS patients, without the use of any clinical data. We estimated the percentage of white matter, gray matter, CSF, and lesions, inside every MRSI voxel of every grid, and used them as new features. We trained simple and complex classifiers on four different types of features: (1) the full spectroscopy signal, (2) the estimated NAA/Cho, NAA/Cre, Cho/Cre, (3) the estimated metabolic ratios and the estimated brain tissue percentages, and (4) the spectrograms of the time-domain spectroscopic signal. Three classical machine learning algorithms (LDA, RF, and SVM-rbf) were trained on the first three types of features, while state of the art convolutional neural networks were trained on the fourth type. We determined that combining metabolic ratios with brain tissue percentages can improve discrimination results between CIS and RR or PP patients, while for RR and PP or SP we obtained only moderate results, lower than in the previous study. Moreover, our best results for classifying MS forms were always obtained with SVM-rbf, so we can conclude that our CNN design does not add any improvement over classical machine learning methods.

7.2 Future perspectives

In this thesis, classical and state of the art machine learning algorithms were utilised to discriminate between GBM progressive and responsive patients, and between healthy subjects and four MS courses.

The GBM analysis has shown that machine learning can clearly help clinicians with their task, at the same time cutting down the time required for delineating

tumours. There is also a big parallel focus on using machine learning for tumour segmentation, and it is only a matter of time until full trust of clinical centres will be gained to use machines smartly, with the latest development in deep learning. However, a machine can learn a task only if the input information is reliable, meaning the MR scanner should provide medical images of sufficient quality, or available images could be preprocessed and their quality improved. If this step is not met, then any further discussion is irrelevant.

Perspectives in the near future of GBM research should include the following ideas: detecting progression using deep learning methods, which could be CNN or long short term memory blocks; upscaling low-resolution MRI (e.g. MRSI, PWI) by using mutual information from multi-parametric MRI; creating an online repository for anonymised MRI data collection; augmenting MR data.

In the distant future, with the increase of the magnetic field of the scanner, sharper images with better resolution and better contrast will be available. Therefore, the focus will change on detecting small changes at the voxel level, and not on the ROI level. Methods like PRM analysing the whole brain, combined with deep learning, will probably become more relevant and will be implemented in hospitals and clinics. Partially intrusive MR techniques like T1pc and perfusion MRI, which require the injection of a contrast agent, will probably become less utilised in favour of ASL and diffusion MRI.

The same discussion applies also for the MS analysis: the use of a greater magnetic field will help for better lesion segmentations and metabolic quantifications. Diffusion tensor imaging will definitely benefit from a higher magnetic field, but also from better diffusion models, for resolving fiber crossings and computing graph based measures.

As a distant end, for the major purpose of individual personalised medicine, there is a need for combining radiomics (MRI, PET, CT) with genomics and proteomics in a greater anonymised public database, where biomedical feature interactions can be better studied.

Appendix A

Appendix

In the next tables we use balanced error rate (BER), weighted balanced error rate (wBER), and balanced accuracy rate (BAR) to present the performance of the classifiers. BER and wBER are numbers between 0 and 1, 0 being perfect classification and 1 being total misclassification. BAR is also a number between 0 and 100, 100 being perfect classification and 0 being total misclassification.

Weighted BER	Complete features	Imputed features	Average
dLDA	0.172	0.216	0.194
SVM-lin	0.276	0.242	0.259
SVM-poly	0.285	0.334	0.310
SVM-rbf	0.493	0.520	0.507
SVM-mlp	0.136	0.352	0.244
Bayesian LSSVM	0.371	0.469	0.420
LSSVM-lin	0.452	0.280	0.366
LSSVM-poly	0.462	0.362	0.412
LSSVM-rbf	0.408	0.320	0.364
Random forests	0.148	0.294	0.221
AdaBoost	0.505	0.324	0.415
LogitBoost	0.148	0.335	0.242
GentleBoost	0.296	0.308	0.302
RobustBoost	0.148	0.325	0.237
LPBoost	0.505	0.256	0.381
TotalBoost	0.505	0.289	0.397
RUSBoost	0.281	0.308	0.295
Classification Tree	0.268	0.346	0.307
3-NN (correlation)	0.357	0.428	0.392
Pattern Net	0.449	0.288	0.366
Feed Forward Net	0.399	0.411	0.405
Cascade forward Net	0.586	0.485	0.535
Fit Net	0.535	0.350	0.443
LDS	0.442	0.534	0.488
SMIR	0.278	0.436	0.357
S4VM	0.456	0.473	0.465

Table A.1: Weighted BER for supervised and semi-supervised classifiers trained on complete and imputed data. We highlight the best 6 classifiers.

Table A.2: Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on complete perfusion features. The decision (i.e. labelling) moment 'L' is highlighted. Some time points do not have results because there were no complete perfusion measurements.

BER on PWI	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+5	–	–	–	-	-	-
L+4	–	–	–	-	-	-
L+3	0	0	0	0	0	0
L+2	0	0	1	0	0	0
L+1	0	0	1	0	0	0
L	0	0.217	0.05	0	0	0.05
L-1	0	0.187	0.187	0	0	0.187
L-2	0.25	0.25	0.375	0.25	0.25	0.25
L-3	0.5	0.5	0.5	0.5	0.5	0.5
L-4	1	1	1	1	1	0.5
L-5	0.25	0.25	0.25	0.5	0.5	0.5
L-6	0.5	0.5	0.5	0.5	0.5	0.5
L-7	1	1	1	1	1	1
L-8	–	–	–	–	–	–
L-9	0	0	0	0	0	0
L-10	–	–	–	–	–	–
L-11	0	0	0	0	0	0

Table A.3: Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on complete diffusion features. The decision (i.e. labelling) moment 'L' is highlighted. Some time points do not have results because there were no complete diffusion measurements.

BER on DKI	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+5	–	–	–	–	–	–
L+4	–	–	–	–	–	–
L+3	0	0	0	0	0	1
L+2	0	0.25	0	0	0	0.5
L+1	0	0	0	0	0	0
L	0.217	0.1	0.1	0.217	0.217	0.267
L-1	0.562	0.25	0.125	0.562	0.562	0.562
L-2	0.5	0.25	0.5	0.5	0.5	0.375
L-3	0.5	0.75	0.75	0.5	0.5	0.25
L-4	0.5	1	0.5	0.5	0.5	0.5
L-5	0.25	0.25	0.5	0.5	0.5	0
L-6	0.5	0	0.5	0.5	0.5	0
L-7	0	0	0	0	0	0
L-8	–	–	–	–	–	–
L-9	1	1	1	1	1	0
L-10	–	–	–	-	-	-
L-11	1	1	1	1	1	0

Table A.4: Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on complete spectroscopy features. The decision (i.e. labelling) moment 'L' is highlighted. Some time points do not have results because there were no complete spectroscopy measurements.

BER on MRSI	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+5	–	–	–	–	–	–
L+4	–	–	–	–	–	–
L+3	0	0	0	0	0	0
L+2	1	0.75	0.75	1	1	1
L+1	1	1	1	1	1	0
L	0.55	0.583	0.632	0.6	0.55	0.583
L-1	0.562	0.562	0.813	0.5	0.562	0.687
L-2	0.625	0.625	0.25	0.625	0.75	0.875
L-3	0.25	0.5	0.25	0.5	0.5	0.25
L-4	0.5	0.5	1	0.5	0.5	1
L-5	0.5	0.5	0	1	1	1
L-6	0.5	0	0.5	0.5	0.5	0.5
L-7	0	0	1	0	0	0
L-8	–	–	–	–	–	–
L-9	1	1	1	1	1	0
L-10	–	–	–	-	-	-
L-11	1	1	1	1	1	0

Table A.5: Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on imputed perfusion features. The decision (i.e. labelling) moment 'L' is highlighted.

BER on PWI	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+5	0	0	0	0	0	0
L+4	0	0	0	0	0	0
L+3	0	0.25	0	0	0	0.25
L+2	0.125	0	0	0.125	0	0.125
L+1	0.171	0.071	0.071	0.171	0.071	0
L	0.127	0.109	0.043	0.127	0.043	0.109
L-1	0.130	0.196	0.152	0.214	0.130	0.279
L-2	0.444	0.528	0.472	0.389	0.444	0.417
L-3	0.418	0.464	0.418	0.373	0.418	0.281
L-4	0.475	0.475	0.475	0.412	0.475	0.512
L-5	0.687	0.687	0.687	0.625	0.687	0.562
L-6	0.567	0.567	0.567	0.567	0.567	0.567
L-7	0.5	0.5	0.5	0.5	0.5	0.5
L-8	0.5	0.5	0.5	0.5	0.5	0.5
L-9	0.333	0.5	0.5	0.333	0.333	0.333
L-10	0.5	0.5	0.5	0.5	0.5	0.25
L-11	0.5	0.5	0.5	0.5	0.5	0

Table A.6: Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on imputed diffusion features. The decision (i.e. labelling) moment 'L' is highlighted.

BER on DKI	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+5	0	0	0	0	0	0
L+4	0	0	0	0	0	0
L+3	0	0	0	0	0	0.25
L+2	0	0.125	0	0	0	0
L+1	0.1	0.243	0.243	0.1	0.1	0.314
L	0.105	0.297	0.192	0.105	0.105	0.420
L-1	0.254	0.257	0.257	0.254	0.254	0.424
L-2	0.361	0.25	0.25	0.361	0.361	0.278
L-3	0.282	0.473	0.473	0.282	0.282	0.436
L-4	0.45	0.637	0.637	0.45	0.45	0.387
L-5	0.562	0.5	0.562	0.562	0.562	0.437
L-6	0.433	0.367	0.533	0.433	0.433	0.433
L-7	0.5	0.5	0.5	0.5	0.5	0.75
L-8	0.667	0.167	0.667	0.667	0.667	0.667
L-9	0.667	0.667	0.667	0.667	0.667	0.5
L-10	0.75	0.75	0.75	0.75	0.75	0.75
L-11	1	0.5	1	1	1	1

Table A.7: Detailed BER results for each time point for the best 6 supervised classifiers when using the leave-one-patient-out method on imputed spectroscopy features. The decision (i.e. labelling) moment 'L' is highlighted.

BER on MRSI	Random Forests	dLDA	SVM-lin	LogitBoost	RobustBoost	SVM-mlp
L+5	0	0	0	0	0	0
L+4	0	0	0	0	0	0
L+3	0	0	0	0	0	0.25
L+2	0.25	0.25	0.125	0.25	0.25	0.25
L+1	0	0	0	0	0	0
L	0.562	0.504	0.609	0.587	0.543	0.569
L-1	0.293	0.337	0.380	0.315	0.293	0.359
L-2	0.389	0.389	0.389	0.389	0.389	0.389
L-3	0.436	0.436	0.381	0.436	0.436	0.336
L-4	0.55	0.55	0.612	0.55	0.55	0.55
L-5	0.687	0.687	0.562	0.687	0.687	0.687
L-6	0.433	0.433	0.533	0.6	0.433	0.433
L-7	0.75	0.75	0.875	0.75	0.75	0.75
L-8	0.667	0.667	0.667	0.667	0.667	0.667
L-9	0.667	0.167	0.667	0.667	0.667	0.667
L-10	0.75	0.75	0.25	0.75	0.75	0.75
L-11	1	1	1	0.5	1	1

Table A.8: Number of data points acquired at each time point. The decision (i.e. labelling) moment 'L' is highlighted.

	Nr. of complete data points	Nr. of total data points (complete+imputed)
L+5	0	2
L+4	0	2
L+3	1	3
L+2	3	8
L+1	1	12
L	13	29
L-1	9	29
L-2	6	24
L-3	3	16
L-4	2	13
L-5	2	12
L-6	2	8
L-7	1	6
L-8	0	5
L-9	1	4
L-10	0	3
L-11	1	2

Table A.9: Number of features per MRI modality and delineation.

	CER&NER				Total ROI			
	cMRI	PWI	DKI	cPD	cMRI	PWI	DKI	cPD
Number of features	208	260	364	832	104	130	182	416

Table A.10: Supervised classifiers used in Chapter 3 and their software implementations.

Classifiers	Software
Linear Discriminant Analysis (LDA)	Matlab R2015a - Statistics and Machine Learning Toolbox
Support Vector Machine (SVM)	Matlab R2015a - Statistics and Machine Learning Toolbox
Random Forests (RF)	Python 2.7.6 - sklearn.ensemble.RandomForestClassifier
Adaptive Boosting (AdaBoost)	Python 2.7.6 - sklearn.ensemble.AdaBoostClassifier
Stochastic Gradient Boosting (SGB)	Python 2.7.6 - sklearn.ensemble.GradientBoostingClassifier
Random Under Sampling Boosting (RUSBoost)	Matlab R2015a - Statistics and Machine Learning Toolbox

Table A.11: Maximum BAR of all MR modalities over all classifiers. CER - contrast enhancing region, NER - non-enhancing region.

Dataset	Nr of features	cMRI	PWI	DKI	cPD
CER&NER	1	0.9321	0.8936	0.8571	0.9321
	2	0.9321	0.8459	0.8277	0.9321
	3	0.9321	0.8550	0.8662	0.9468
	4	0.9321	0.8256	0.8606	0.9321
	5	0.9321	0.8403	0.8606	0.9321
	6	0.9321	0.8697	0.8368	0.8606
	7	0.9321	0.8312	0.8368	0.8606
	8	0.9321	0.8221	0.8368	0.8606
	9	0.9321	0.8221	0.8221	0.8606
	10	0.9321	0.8312	0.8845	0.8606
Total ROI	1	0.8018	0.7598	0.8459	0.8459
	2	0.8789	0.7836	0.8368	0.8368
	3	0.8200	0.7836	0.8550	0.8606
	4	0.8312	0.7927	0.8165	0.8992
	5	0.9412	0.7689	0.8165	0.8754
	6	0.9412	0.8494	0.8789	0.8754
	7	0.9559	0.8403	0.8459	0.8606
	8	0.9559	0.7927	0.8312	0.9321
	9	0.9559	0.7927	0.7780	0.8845
	10	0.8459	0.7927	0.8074	0.8789

Table A.12: Top 10 selected features according to rank products for each dataset

Dataset	cMRI	PWI	DKI	cPD
CER&NER	T1pc-average-CER	CBV-average-CER	ak-IMC2-CER	T1pc-average-CER
	T1pc-90th_percentile-CER	MTT-homogeneity-CER	ak-IMC1-CER	T1pc-90th_percentile-CER
	T1pc-10th_percentile-CER	CBV-90th_percentile-CER	fa-autocorrelation-CER	CBV-90th_percentile-CER
	FLAIR-cluster shade-CER	MTT-IDN-CER	fa-sum variance-CER	CBV-average-CER
	FLAIR-skewness-CER	MTT-maximum probability-CER	fa-sum of squares-CER	CBF-90th_percentile-CER
	FLAIR-sum of squares:variance-CER	CBF-90th_percentile-CER	fa-average-CER	ak-IMC2-CER
	T2-cluster shade-CER	MTT-dissimilarity-CER	fa-sum average-CER	MTT-homogeneity-CER
	FLAIR-autocorrelation-CER	Rsquare-homogeneity-CER	ad-cluster prominence-NER	CBF-average-CER
	FLAIR-sum variance-CER	MTT-difference entropy-CER	mk-IMC2-CER	MTT-IDN-CER
	FLAIR-sum average-CER	CBF-average-CER	mk-IMC1-CER	ak-IMC1-CER
Total ROI	T1pc-90th_percentile	CBV-90th_percentile	ad-cluster prominence	ad-cluster prominence
	T1pc-coefficient of variation	Rsquare-IMC2	md-cluster prominence	md-cluster prominence
	T1pc-cluster shade	Rsquare-IMC1	ad-cluster shade	T1pc-90th_percentile
	T1pc-sum average	CBF-90th_percentile	rd-sum entropy	T1pc-coefficient of variation
	T1pc-skewness	CBV-IMC2	md-cluster shade	CBV-90th_percentile
	T1pc-sum variance	MTT-correlation	md-sum entropy	ad-cluster shade
	T1pc-autocorrelation	CBF-IMC2	rd-cluster prominence	T1pc-skewness
	T1pc-sum of squares:variance	Rsquare-contrast	rd-cluster shade	rd-sum entropy
	T1pc-maximum probability	Rsquare-dissimilarity	ak-cluster prominence	T1pc-sum average
	T1pc-average	Rsquare-IDMN	rk-sum entropy	T1pc-cluster shade

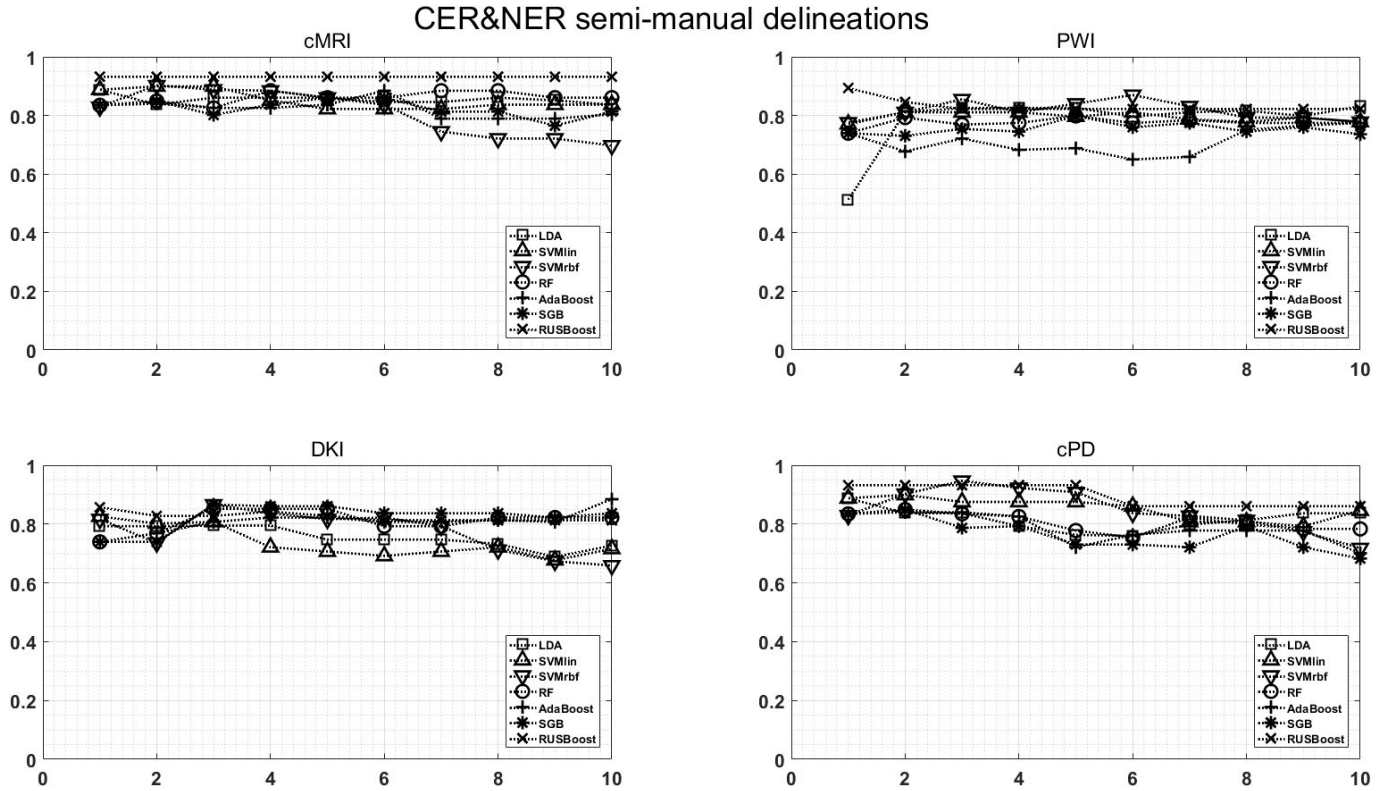


Figure A.1: Classification results on CER & NER semi-manual delineations, using 1 to 10 features assigned by rank products per each dataset. On y-axis are BAR values, and on x-axis the number of features used for classification. CER - contrast enhancing region, NER - non-enhancing region.

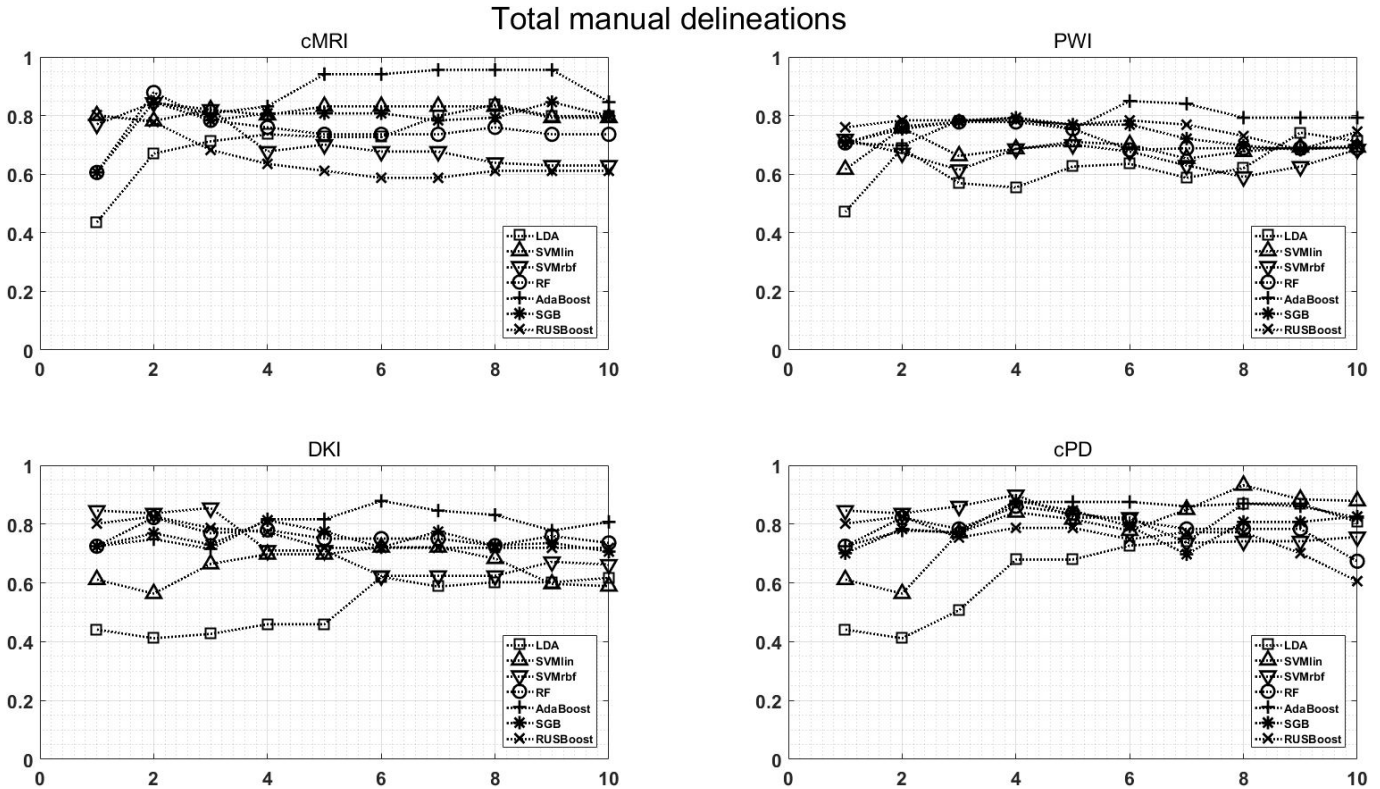


Figure A.2: Classification results on Total manual delineations, using 1 to 10 features assigned by rank products per each dataset. On y-axis are BAR values, and on x-axis the number of features used for classification.

	NAA/Cho			NAA/Cre			Cho/Cre			All 3 metabolites		
	BAR	TNR	TPR	BAR	TNR	TPR	BAR	TNR	TPR	BAR	TNR	TPR
HC vs. CIS	47	0	94	46	15	78	61	39	83	53	39	67
HC vs. RR	50	94	6	55	82	28	50	100	0	52	76	28
HC vs. PP	76	80	72	78	72	83	45	29	61	77	82	72
HC vs. RR + SP	52	98	6	60	92	28	50	100	0	59	90	28
HC vs. RR + PP	61	89	33	66	88	44	50	100	0	52	88	16
CIS vs. RR	52	95	10	50	100	0	50	99	0	52	88	16
CIS vs. RR + SP	51	100	2	49	99	0	50	100	0	54	94	15
RR vs. PP	59	37	81	63	38	88	48	2	95	63	49	77
RR vs. SP	57	53	62	65	62	69	39	0	79	66	62	70

Table A.13: Balanced accurary rates (BAR), sensitivity (TPR), and specificity (TNR) values, for all 9 classification tasks (rows) after training LDA using only metabolic ratios. Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, values between 85 and 89 are coloured in dark gray, while values higher than 90 are coloured in very dark gray.

	CIS vs. RR			CIS vs. RR+SP			RR vs. PP			RR vs. SP		
	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF
M	52	68	49	54	63	49	63	28	59	66	66	63
LL	48	70	52	50	73	56	43	12	58	74	75	68
Age + DD	66	75	68	66	83	70	67	38	62	75	76	71
Age + DD + EDSS	71	80	67	77	89	69	81	78	70	84	85	84
Age + DD + EDSS + LL	79	85	73	81	92	76	71	72	69	86	86	85
Age + DD + EDSS + M	72	76	66	81	82	70	80	81	71	86	87	84
Age + DD + EDSS + LL + M	78	80	71	82	83	73	78	78	68	86	86	86

Table A.14: BAR values for classification tasks involving only MS patients (columns). Abbreviations: M = all three average metabolic ratios; Age = patient age; DD = disease duration; LL = lesion load; EDSS = Expanded Disability Status Scale.Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, values between 85 and 89 are coloured in dark gray, while values higher than or equal to 90 are coloured in very dark gray.

	CIS vs. RR			CIS vs. RR+SP			RR vs. PP			RR vs. SP		
	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF
M	16	79	8	15	67	0	77	56	78	70	75	72
LL	0	80	30	0	80	23	87	16	78	77	80	67
Age + DD	41	77	49	36	84	46	84	75	78	74	70	70
Age + DD + EDSS	56	82	44	62	92	43	84	75	80	80	83	81
Age + DD + EDSS + LL	69	87	56	69	93	57	81	69	83	85	84	85
Age + DD + EDSS + M	59	74	41	74	79	44	84	76	82	84	85	84
Age + DD + EDSS + LL + M	67	79	49	72	77	51	83	75	81	87	87	86

Table A.15: Sensitivity values for classification tasks involving only MS patients (columns). Abbreviations: M = all three average metabolic ratios; Age = patient age; DD = disease duration; LL = lesion load; EDSS = Expanded Disability Status Scale. Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, values between 85 and 89 are coloured in dark gray, while values higher than or equal to 90 are coloured in very dark gray.

	CIS vs. RR			CIS vs. RR+SP			RR vs. PP			RR vs. SP		
	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF	LDA	SVM-rbf	RF
M	88	57	89	94	59	98	49	0	40	62	56	54
LL	96	60	75	100	66	89	0	7	37	70	70	69
Age + DD	91	74	87	96	83	94	50	0	46	76	82	72
Age + DD + EDSS	87	79	89	91	87	95	78	81	60	89	87	86
Age + DD + EDSS + LL	89	83	90	92	90	95	60	75	55	87	87	85
Age + DD + EDSS + M	85	78	91	89	86	95	75	86	60	88	88	84
Age + DD + EDSS + LL + M	88	81	93	92	89	96	74	82	56	85	86	85

Table A.16: Specificity values for classification tasks involving only MS patients (columns). Abbreviations: M = all three average metabolic ratios; Age = patient age; DD = disease duration; LL = lesion load; EDSS = Expanded Disability Status Scale. Values between 75 and 79 are coloured in light gray, values between 80 and 84 are coloured in medium gray, values between 85 and 89 are coloured in dark gray, while values higher than or equal to 90 are coloured in very dark gray.

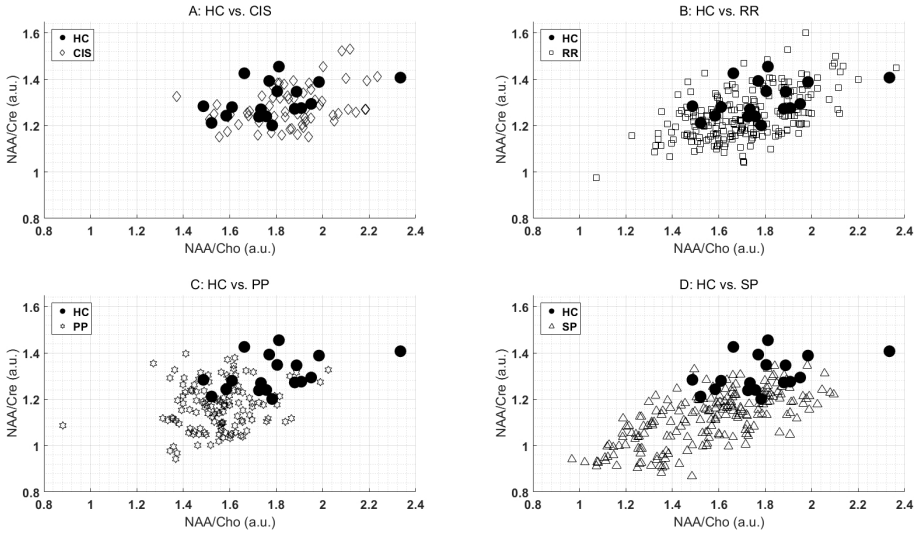


Figure A.3: Healthy Controls (HC) vs. Multiple Sclerosis (MS) groups in 2-D feature space: x-axis is NAA/Cho and y-axis is NAA/Cre. The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

			Conventional and Perfusion MRI (cpMRI)															
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T1pc-PRM$_+$	SVM-lin	AUC	93	92	90	88	81	83	87	85	82	85	84	84	81	79	79	80
		BAR	91	89	90	86	79	83	86	82	82	83	83	85	80	80	81	81
		TPR	85	83	83	81	69	77	85	75	69	77	71	73	71	63	65	65
		TNR	97	95	97	92	89	89	86	89	95	89	95	97	89	97	97	97
	SVM-rbf	AUC	88	86	85	75	80	82	84	83	84	81	80	77	80	76	74	77
		BAR	87	84	90	72	74	78	78	80	80	76	73	75	73	76	75	78
		TPR	83	85	83	87	87	73	83	71	81	81	87	77	87	77	75	67
		TNR	92	84	97	57	62	84	73	89	78	70	59	73	59	76	76	89
CBV-PRM$_+$	SVM-lin	AUC	86	87	85	82	79	80	83	79	81	78	81	81	81	74	76	73
		BAR	86	87	85	85	78	80	83	81	82	78	78	77	78	74	74	73
		TPR	77	83	75	75	62	67	71	69	71	62	62	71	77	63	73	62
		TNR	95	92	95	95	95	92	95	92	92	95	95	84	78	84	76	84
	SVM-rbf	AUC	83	86	72	82	64	57	59	55	66	60	60	55	55	67	56	56
		BAR	85	79	73	77	68	62	63	64	64	64	59	58	59	61	58	57
		TPR	81	83	73	81	98	92	94	92	92	100	87	83	94	94	96	87
		TNR	89	76	73	73	38	32	32	35	35	27	32	32	24	27	19	27

Table A.17: Performance measures computed with SVM-lin and SVM-rbf trained on an increasing number of features from 1 to 16 for Conventional and Perfusion MRI (cpMRI), extracted using the two positive parametric response maps (PRM$_+$), T1pc-PRM$_+$ and CBV-PRM$_+$. Values over 90% are highlighted in gray.

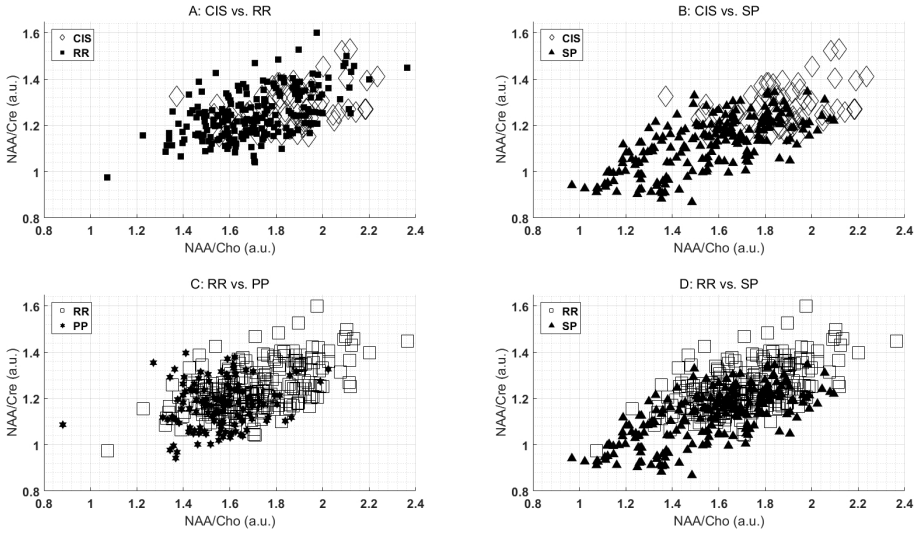


Figure A.4: Comparison of Multiple Sclerosis (MS) groups in 2-D feature space: x-axis is NAA/Cho and y-axis is NAA/Cre. The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

			cMRI								PWI							
			1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
T1pc-PRM$_+$	SVM-lin	AUC	93	92	90	87	88	86	85	87	73	66	63	65	61	59	58	65
		BAR	91	89	87	83	84	80	80	84	74	71	69	69	67	67	67	73
		TPR	85	83	83	71	73	77	67	79	63	56	38	58	40	40	40	52
		TNR	97	95	92	95	95	84	92	89	84	86	100	81	95	95	95	95
	SVM-rbf	AUC	88	86	85	87	85	85	89	85	55	57	62	53	56	65	62	60
		BAR	87	84	81	80	79	81	82	81	67	66	69	60	65	64	66	62
		TPR	83	85	87	77	77	69	77	73	56	65	62	77	62	85	65	62
		TNR	92	84	76	84	81	92	86	89	78	68	76	43	68	43	68	62
CBV-PRM$_+$	SVM-lin	AUC	86	87	85	82	80	81	84	84	52	48	38	36	29	31	32	39
		BAR	86	87	85	81	80	80	82	80	63	63	46	43	34	48	39	50
		TPR	77	83	75	67	67	67	77	77	58	62	52	48	37	38	42	44
		TNR	95	92	95	95	92	92	86	84	68	65	41	38	32	57	35	57
	SVM-rbf	AUC	83	86	72	73	68	71	69	74	11	3	14	9	11	8	14	7
		BAR	85	79	73	66	70	72	66	73	60	34	22	23	32	22	33	24
		TPR	81	83	73	87	81	96	92	62	58	58	44	38	50	38	58	48
		TNR	89	76	73	46	59	49	41	84	62	11	0	8	14	5	8	0

Table A.18: Performance measures computed with SVM-lin and SVM-rbf trained on an increasing number of features from 1 to 8 for cMRI and PWI separately, features extracted using the two positive parametric response maps (PRM$_+$), T1pc-PRM$_+$ and CBV-PRM$_+$. Values over 90% are highlighted in gray.

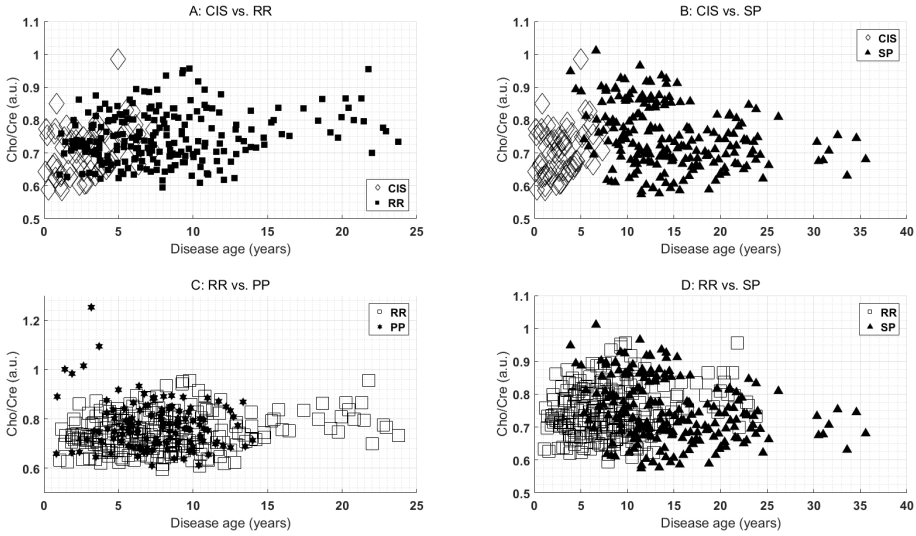


Figure A.5: Comparison of Multiple Sclerosis (MS) groups in 2-D feature space: x-axis is disease age and y-axis is Cho/Cre. The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

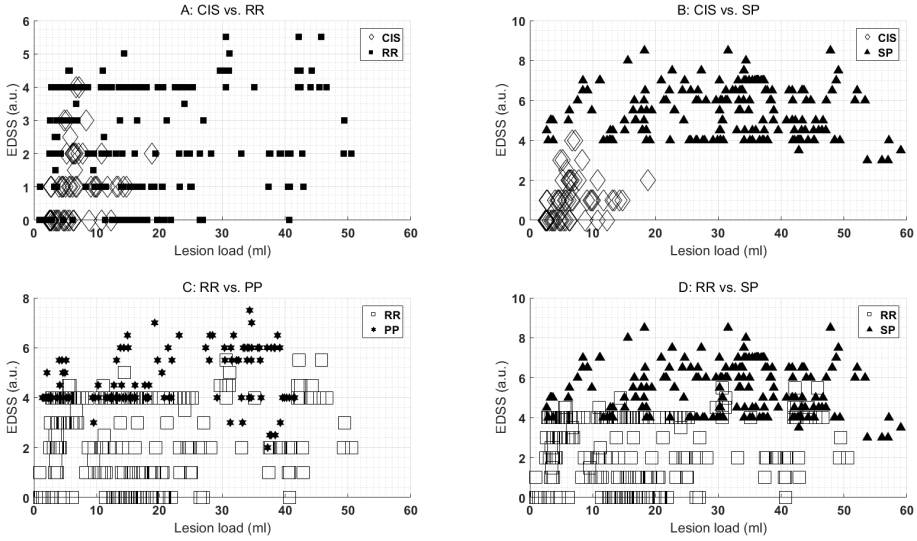


Figure A.6: Comparison of Multiple Sclerosis (MS) groups in 2-D feature space: x-axis is lesion load and y-axis is EDSS. The four MS groups are: CIS - clinically isolated syndrome, RR - relapsing-remitting, PP - primary progressive, SP - secondary progressive.

	Conventional and Perfusion MRI (cpMRI)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T1pc-PRM$_+$	T1pc-99p	T1pc-50p	CBF-70p	F-50p	F-90p	CBV-90p	T1pc-70p	F-99p	F-70p	CBV-70p	T1pc-90p	CBV-50p	CBF-90p	CBF-50p	CBV-99p	CBF-99p
CBV-PRM$_+$	T1pc-99p	F-70p	T1pc-70p	CBF-50p	F-99p	T1pc-50p	F-50p	T1pc-90p	F-90p	CBV-50p	CBV-90p	CBV-70p	CBF-90p	CBF-70p	CBV-99p	CBF-99p

Table A.19: Conventional and Perfusion MRI (cpMRI) features selected with minimum-redundancy-maximum-relevance (mRMR) after applying separately the two positive parametric response maps (PRM$_+$), T1pc-PRM$_+$ and CBV-PRM$_+$, where “F" stands for FLAIR. Features are ‘X’-percentile, where ‘X’ can be 50, 70, 90, and 99.

	Conventional MRI							
	1	2	3	4	5	6	7	8
T1pc-PRM$_+$	T1pc-99p	T1pc-50p	FLAIR-50p	FLAIR-90p	T1pc-70p	FLAIR-99p	FLAIR-70p	T1pc-90p
CBV-PRM$_+$	T1pc-99p	FLAIR-70p	T1pc-70p	FLAIR-99p	T1pc-90p	T1pc-50p	FLAIR-50p	FLAIR-90p

Table A.20: Conventional MRI features selected with minimum-redundancy-maximum-relevance (mRMR) after applying separately the two positive parametric response maps (PRM$_+$), T1pc-PRM$_+$ and CBV-PRM$_+$. Features are ‘X’-percentile, where ‘X’ can be 50, 70, 90, and 99.

	Perfusion MRI							
	1	2	3	4	5	6	7	8
T1pc-PRM$_+$	CBV-99p	CBF-50p	CBV-70p	CBV-50p	CBF-70p	CBF-90p	CBV-90p	CBF-99p
CBV-PRM$_+$	CBV-99p	CBV-50p	CBF-50p	CBV-70p	CBF-70p	CBV-90p	CBF-90p	CBF-99p

Table A.21: Perfusion MRI features selected with minimum-redundancy-maximum-relevance (mRMR) after applying separately the two positive parametric response maps (PRM$_+$), T1pc-PRM$_+$ and CBV-PRM$_+$. Features are 'X'-percentile, where 'X' can be 50, 70, 90, and 99.

Bibliography

[1] Akbari, H., Macyszyn, L., Da, X., Bilello, M., Wolf, R. L., Martinez-Lage, M., Biros, G., Alonso-Basanta, M., O'Rourke, D. M., and Davatzikos, C. Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma. *Neurosurgery 78*, 4 (2016), 572–580.

[2] Amato, M., Portaccio, E., Stromillo, M., Goretti, B., Zipoli, V., Siracusa, G., Battaglini, M., Giorgio, A., Bartolozzi, M., Guidi, L., et al. Cognitive assessment and quantitative magnetic resonance metrics can help to identify benign multiple sclerosis. *Neurology 71*, 9 (2008), 632–638.

[3] Aquino, D., Di Stefano, A. L., Scotti, A., Cuppini, L., Anghileri, E., Finocchiaro, G., Bruzzone, M. G., and Eoli, M. Parametric response maps of perfusion MRI may identify recurrent glioblastomas responsive to bevacizumab and irinotecan. *PLoS One 9*, 3 (2014), e90535.

[4] Arnold, D., De Stefano, N., Narayanan, S., and Matthews, P. Proton MR spectroscopy in multiple sclerosis. *Neuroimaging clinics of North America 10*, 4 (2000), 789–98.

[5] Artzi, M., Liberman, G., Nadav, G., Blumenthal, D. T., Bokstein, F., Aizenstein, O., and Bashat, D. B. Differentiation between treatment-related changes and progressive disease in patients with high grade brain tumors using support vector machine classification based on DCE MRI. *Journal of neuro-oncology 127*, 3 (2016), 515–524.

[6] Atlas of MS 2013: Mapping Multiple Sclerosis Around the World. London: Multiple Sclerosis International Federation. `http://www.msif.org/about-ms/publications-and-resources/`. [Online; accessed 05-June-2017].

[7] Bammer, R., Augustin, M., Strasser-Fuchs, S., Seifert, T., Kapeller, P., Stollberger, R., Ebner, F., Hartung, H.-P., and Fazekas, F. Magnetic resonance diffusion tensor imaging for characterizing diffuse and focal white matter abnormalities in multiple sclerosis. *Magnetic Resonance in Medicine 44*, 4 (2000), 583–591.

[8] Banwell, B., Giovannoni, G., Hawkes, C., and Lublin, F. Editors' welcome and a working definition for a multiple sclerosis cure. *Multiple sclerosis and related disorders 2*, 2 (2013), 65–67.

[9] Barajas Jr, R. F., Chang, J. S., Segal, M. R., Parsa, A. T., McDermott, M. W., Berger, M. S., and Cha, S. Differentiation of recurrent glioblastoma multiforme from radiation necrosis after external beam radiation therapy with dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging 1. *Radiology 253*, 2 (2009), 486–496.

[10] Basser, P. J., Mattiello, J., and LeBihan, D. MR diffusion tensor spectroscopy and imaging. *Biophysical journal 66*, 1 (1994), 259–267.

[11] Bates, D. M. *lme4: Mixed-effects modeling with R*. Springer New York, 2010.

[12] Bellman, R. E. *Adaptive control processes: a guided tour*. Princeton university press, 2015.

[13] Bertholdo, D., Watcharakorn, A., and Castillo, M. Brain proton magnetic resonance spectroscopy: introduction and overview. *Neuroimaging Clinics of North America 23*, 3 (2013), 359–380.

[14] Bitsch, A., Bruhn, H., Vougioukas, V., Stringaris, A., Lassmann, H., Frahm, J., and Brück, W. Inflammatory CNS demyelination: histopathologic correlation with in vivo quantitative proton MR spectroscopy. *American Journal of Neuroradiology 20*, 9 (1999), 1619–1627.

[15] Bjartmar, C., Kinkel, R. P., Kidd, G., Rudick, R. A., and Trapp, B. D. Axonal loss in normal-appearing white matter in a patient with acute ms. *Neurology 57*, 7 (2001), 1248–1252.

[16] Bobek-Billewicz, B., Stasik-Pres, G., Majchrzak, H., and Zarudzki, Ł. Original article differentiation between brain tumor recurrence and radiation injury using perfusion, diffusion-weighted imaging and mr spectroscopy. *Folia Neuropathologica 48*, 2 (2010), 81–92.

[17] Bottomley, P. A. Spatial localization in NMR spectroscopy in vivo. *Annals of the New York Academy of Sciences 508*, 1 (1987), 333–348.

[18] Boxerman, J., Schmainda, K., and Weisskoff, R. Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not. *American Journal of Neuroradiology 27*, 4 (2006), 859–867.

[19] Bradley, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition 30*, 7 (1997), 1145–1159.

[20] Breiman, L. Bagging predictors. *Machine learning 24*, 2 (1996), 123–140.

[21] Breiman, L. Random forests. *Machine learning 45*, 1 (2001), 5–32.

[22] Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. *Classification and regression trees*. CRC press, 1984.

[23] Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS letters 573*, 1 (2004), 83–92.

[24] Brück, W., Porada, P., Poser, S., Rieckmann, P., Hanefeld, F., Kretzschmarch, H. A., and Lassmann, H. Monocyte/macrophage differentiation in early multiple sclerosis lesions. *Annals of neurology 38*, 5 (1995), 788–796.

[25] Bulik, M., Kazda, T., Slampa, P., and Jancalek, R. The Diagnostic Ability of Follow-Up Imaging Biomarkers after Treatment of Glioblastoma in the Temozolomide Era: Implications from Proton MR Spectroscopy and Apparent Diffusion Coefficient Mapping. *BioMed research international 2015* (2015).

[26] Burger, P. C., Vogel, F. S., Green, S. B., and Strike, T. A. Glioblastoma multiforme and anaplastic astrocytoma pathologic criteria and prognostic implications. *Cancer 56*, 5 (1985), 1106–1111.

[27] Calle, M. L., and Urrea, V. Letter to the editor: stability of random forest importance measures. *Briefings in bioinformatics 12*, 1 (2010), 86–89.

[28] Caravan, P., Ellison, J. J., McMurry, T. J., and Lauffer, R. B. Gadolinium (iii) chelates as mri contrast agents: structure, dynamics, and applications. *Chemical reviews 99*, 9 (1999), 2293–2352.

[29] Cavassila, S., Deval, S., Huegen, C., van Ormondt, D., and Graveron-Demilly, D. Cramer Rao bounds: an evaluation tool for quantitation. *NMR in Biomedicine 14*, 4 (2001), 278–283.

[30] Cha, J., Kim, S., Kim, H.-J., Kim, B.-j., Kim, Y., Lee, J., Jeon, P., Kim, K., Kong, D.-s., and Nam, D.-H. Differentiation of tumor progression from pseudoprogression in patients with posttreatment glioblastoma using multiparametric histogram analysis. *American Journal of Neuroradiology 35*, 7 (2014), 1309–1317.

[31] Cha, S., Knopp, E. A., Johnson, G., Wetzel, S. G., Litt, A. W., and Zagzag, D. Intracranial mass lesions: Dynamic contrast-enhanced susceptibility-weighted echo-planar perfusion mr imaging 1. *Radiology 223*, 1 (2002), 11–29.

[32] Chang, K., Zhang, B., Guo, X., Zong, M., Rahman, R., Sanchez, D., Winder, N., Reardon, D. A., Zhao, B., Wen, P. Y., et al. Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab. *Neuro-oncology 18*, 12 (2016), 1680–1687.

[33] Chapelle, O. `http://olivier.chapelle.cc/lds/`. [Online; accessed 09-August-2014].

[34] Chapelle, O., and Zien, A. Semi-supervised classification by low density separation. In *AISTATS* (2005), pp. 57–64.

[35] Chen, J. T., Collins, D. L., Atkins, H. L., Freedman, M. S., and Arnold, D. L. Magnetization transfer ratio evolution with demyelination and remyelination in multiple sclerosis lesions. *Annals of neurology 63*, 2 (2008), 254–262.

[36] Chollet, F. Keras. `https://github.com/fchollet/keras`, 2015.

[37] Clausi, D. A. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing 28*, 1 (2002), 45–62.

[38] Compston, A., and Coles, A. Multiple sclerosis. *The Lancet 372*, 9648 (Oct. 2008), 1502–1518.

[39] Cortes, C., and Vapnik, V. Support-vector networks. *Machine learning 20*, 3 (1995), 273–297.

[40] Coupé, P. Mri denoising package. `http://sites.google.com/site/pierrickcoupe/softwares/denoising-for-medical-imaging/mri-denoising/mri-denoising-software`. Accessed: 2017-04-05.

[41] Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., and Barillot, C. An optimized blockwise nonlocal means denoising filter for

3-d magnetic resonance images. *IEEE transactions on medical imaging 27*, 4 (2008), 425–441.

[42] Cover, T., and Hart, P. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on 13*, 1 (1967), 21–27.

[43] Cristianini, N., and Shawe-Taylor, J. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[44] Croitor-Sava, A. R., Sima, D. M., Poullet, J.-B., Wright, A. J., Heerschap, A., and Van Huffel, S. Exploiting spatial information to estimate metabolite levels in two-dimensional MRSI of heterogeneous brain lesions. *NMR in Biomedicine 24*, 7 (2011), 824–835.

[45] Cutter, G. R., Baier, M. L., Rudick, R. A., Cookfair, D. L., Fischer, J. S., Petkau, J., Syndulko, K., Weinshenker, B. G., Antel, J. P., Confavreux, C., et al. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain 122*, 5 (1999), 871–882.

[46] da Cruz, L. H., Rodriguez, I., Domingues, R., Gasparetto, E., and Sorensen, A. Pseudoprogression and pseudoresponse: imaging challenges in the assessment of posttreatment glioma. *American Journal of Neuroradiology 32*, 11 (2011), 1978–1985.

[47] Davie, C., Barker, G., Thompson, A., Tofts, P., McDonald, W., and Miller, D. 1H magnetic resonance spectroscopy of chronic cerebral white matter lesions and normal appearing white matter in multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry 63*, 6 (1997), 736–742.

[48] Davie, C., Hawkins, C., Barker, G., Brennan, A., Tofts, P., Miller, D., and McDonald, W. Serial proton magnetic resonance spectroscopy in acute multiple sclerosis lesions. *Brain 117*, 1 (1994), 49–58.

[49] De Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., De Brabanter, J., Pelckmans, K., De Moor, B., Vandewalle, J., and Suykens, J. LS-SVMlab toolbox user's guide. *ESAT-SISTA Technical Report* (2011), 10–146.

[50] De Stefano, N., and Filippi, M. MR spectroscopy in multiple sclerosis. *Journal of Neuroimaging 17*, s1 (2007), 31s–35s.

[51] De Vleeschouwer, S., Fieuws, S., Rutkowski, S., Van Calenbergh, F., Van Loon, J., Goffin, J., Sciot, R., Wilms, G., Demaerel, P., Warmuth-Metz, M., et al. Postoperative adjuvant dendritic cell–based immunotherapy in patients with relapsed glioblastoma multiforme. *Clinical Cancer Research 14*, 10 (2008), 3098–3104.

[52] De Vleeschouwer, S., Van Calenbergh, F., Demaerel, P., Flamen, P., Rutkowski, S., Kaempgen, E., Wolff, J. E., Plets, C., Sciot, R., and Van Gool, S. W. Transient local response and persistent tumor control in a child with recurrent malignant glioma: treatment with combination therapy including dendritic cell therapy: case report. *Journal of Neurosurgery: Pediatrics 100*, 5 (2004), 492–497.

[53] Dean, B. L., Drayer, B. P., Bird, C. R., Flom, R. A., Hodak, J. A., Coons, S. W., and Carey, R. G. Gliomas: classification with MR imaging. *Radiology 174*, 2 (1990), 411–415.

[54] Di Costanzo, A., Scarabino, T., Trojsi, F., Popolizio, T., Bonavita, S., de Cristofaro, M., Conforti, R., Cristofano, A., Colonnese, C., Salvolini, U., et al. Recurrent glioblastoma multiforme versus radiation injury: a multiparametric 3-t mr approach. *La radiologia medica 119*, 8 (2014), 616–624.

[55] Ding, C., and Peng, H. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology 3*, 02 (2005), 185–205.

[56] Dobes, M., Khurana, V. G., Shadbolt, B., Jain, S., Smith, S. F., Smee, R., Dexter, M., and Cook, R. Increasing incidence of glioblastoma multiforme and meningioma, and decreasing incidence of schwannoma (2000–2008): findings of a multicenter Australian study. *Surgical neurology international 2* (2011), 176.

[57] Doyle, T. J., Pathak, R., Wolinsky, J. S., and Narayana, P. A. Automated proton spectroscopic image processing. *Journal of Magnetic Resonance, Series B 106*, 1 (1995), 58–63.

[58] Droogan, A., Clark, C., Werring, D., Barker, G., McDonald, W., and Miller, D. Comparison of multiple sclerosis clinical subgroups using navigated spin echo diffusion-weighted imaging. *Magnetic resonance imaging 17*, 5 (1999), 653–661.

[59] Dunn, O. J. Multiple comparisons using rank sums. *Technometrics 6*, 3 (1964), 241–252.

[60] Earnest 4th, F., Kelly, P., Scheithauer, B., Kall, B., Cascino, T., Ehman, R., Forbes, G., and Axley, P. Cerebral astrocytomas: histopathologic correlation of MR and CT contrast enhancement with stereotactic biopsy. *Radiology 166*, 3 (1988), 823–827.

[61] Ellingson, B. M., Cloughesy, T. F., Lai, A., Mischel, P. S., Nghiemphu, P. L., Lalezari, S., Schmainda, K. M., and Pope, W. B. Graded functional diffusion map–defined characteristics of apparent diffusion coefficients predict overall survival in recurrent glioblastoma treated with bevacizumab. *Neuro-oncology 13*, 10 (2011), 1151–1161.

[62] Ellingson, B. M., Malkin, M. G., Rand, S. D., Connelly, J. M., Quinsey, C., LaViolette, P. S., Bedekar, D. P., and Schmainda, K. M. Validation of functional diffusion maps (fDMs) as a biomarker for human glioma cellularity. *Journal of Magnetic Resonance Imaging 31*, 3 (2010), 538–548.

[63] Ellingson, B. M., Malkin, M. G., Rand, S. D., LaViolette, P. S., Connelly, J. M., Mueller, W. M., and Schmainda, K. M. Volumetric analysis of functional diffusion maps is a predictive imaging biomarker for cytotoxic and anti-angiogenic treatments in malignant gliomas. *Journal of neuro-oncology 102*, 1 (2011), 95–103.

[64] Elson, A., Bovi, J., Siker, M., Schultz, C., and Paulson, E. Evaluation of absolute and normalized apparent diffusion coefficient (ADC) values within the post-operative T2/FLAIR volume as adverse prognostic indicators in glioblastoma. *Journal of neuro-oncology 122*, 3 (2015), 549–558.

[65] Enzinger, C., Barkhof, F., Ciccarelli, O., Filippi, M., Kappos, L., Rocca, M. A., Ropele, S., Rovira, À., Schneider, T., De Stefano, N., et al. Nonconventional mri and microstructural cerebral changes in multiple sclerosis. *Nature Reviews Neurology 11*, 12 (2015), 676–686.

[66] ESAT. `http://www.esat.kuleuven.be/sista/lssvmlab/`. [Online; accessed 09-August-2014].

[67] Filippi, M. Enhanced magnetic resonance imaging in multiple sclerosis. *Multiple Sclerosis Journal 6*, 5 (2000), 320–326.

[68] Filippi, M., Cercignani, M., Inglese, M., Horsfield, M., and Comi, G. Diffusion tensor magnetic resonance imaging in multiple sclerosis. *Neurology 56*, 3 (2001), 304–311.

[69] Filippi, M., Horsfield, M., Morrissey, S., MacManus, D., Rudge, P., McDonald, W., and Miller, D. Quantitative brain MRI lesion load predicts the course of clinically isolated syndromes suggestive of multiple sclerosis. *Neurology 44*, 4 (1994), 635–635.

[70] Filippi, M., Iannucci, G., Cercignani, M., Rocca, M. A., Pratesi, A., and Comi, G. A quantitative study of water diffusion in multiple sclerosis lesions and normal-appearing white matter using echo-planar imaging. *Archives of neurology 57*, 7 (2000), 1017–1021.

[71] Filippi, M., and Rocca, M. A. Magnetization transfer magnetic resonance imaging of the brain, spinal cord, and optic nerve. *Neurotherapeutics 4*, 3 (2007), 401–413.

[72] Filippi, M., Rocca, M. A., Barkhof, F., Brück, W., Chen, J. T., Comi, G., DeLuca, G., De Stefano, N., Erickson, B. J., Evangelou, N., et al. Association between pathological and mri findings in multiple sclerosis. *The Lancet Neurology 11*, 4 (2012), 349–360.

[73] Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of eugenics 7*, 2 (1936), 179–188.

[74] Fox, R. J., and Cohen, J. A. Multiple sclerosis: the importance of early recognition and treatment. *Cleveland Clinic journal of medicine 68*, 2 (2001), 157–171.

[75] Frahm, J., Merboldt, K.-D., and Hänicke, W. Localized proton spectroscopy using stimulated echoes. *Journal of Magnetic Resonance (1969) 72*, 3 (1987), 502–508.

[76] Freund, Y. A more robust boosting algorithm. *arXiv preprint arXiv:0905.2138* (2009).

[77] Freund, Y., and Schapire, R. E. A desicion-theoretic generalization of on-line learning and an application to boosting. In *Computational learning theory* (1995), Springer, pp. 23–37.

[78] Friedman, J., Hastie, T., and Tibshirani, R. *The elements of statistical learning*, vol. 1. Springer series in statistics Springer, Berlin, 2001.

[79] Friedman, J., Hastie, T., Tibshirani, R., et al. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics 28*, 2 (2000), 337–407.

[80] Friedman, J. H. Stochastic gradient boosting. *Computational Statistics & Data Analysis 38*, 4 (2002), 367–378.

[81] Fu, L., Matthews, P., De Stefano, N., Worsley, K., Narayanan, S., Francis, G., Antel, J., Wolfson, C., and Arnold, D. Imaging axonal damage of normal-appearing white matter in multiple sclerosis. *Brain 121*, 1 (1998), 103–113.

[82] Galbán, C. J., Chenevert, T. L., Meyer, C. R., Tsien, C., Lawrence, T. S., Hamstra, D. A., Junck, L., Sundgren, P. C., Johnson, T. D., Ross, D. J., et al. The parametric response map is an imaging biomarker for early cancer treatment outcome. *Nature medicine 15*, 5 (2009), 572–576.

[83] Galbán, C. J., Chenevert, T. L., Meyer, C. R., Tsien, C. I., Lawrence, T. S., Hamstra, D. A., Junck, L., Sundgren, P., Johnson, T. D., Galbán, S., et al. Prospective analysis of parametric MRI biomarkers: Identification of early and distinct glioma response patterns not predicted by standard radiographic assessment. *Clinical Cancer Research* (2011), clincanres–2098.

[84] Gantz, J., and Reinsel, D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the future 2007*, 2012 (2012), 1–16.

[85] Giovannoni, G., Cook, S., Rammohan, K., Rieckmann, P., Sørensen, P. S., Vermersch, P., Hamlett, A., Viglietta, V., Greenberg, S., et al. Sustained disease-activity-free status in patients with relapsing-remitting multiple sclerosis treated with cladribine tablets in the CLARITY study: a post-hoc and subgroup analysis. *The Lancet Neurology 10*, 4 (2011), 329–337.

[86] Gladwish, A., Koh, E.-S., Hoisak, J., Lockwood, G., Millar, B.-A., Mason, W., Yu, E., Laperriere, N. J., and Ménard, C. Evaluation of early imaging response criteria in glioblastoma multiforme. *Radiation Oncology 6*, 1 (2011), 121.

[87] Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[88] Hagan, M. T., and Menhaj, M. B. Training feedforward networks with the Marquardt algorithm. *Neural Networks, IEEE Transactions on 5*, 6 (1994), 989–993.

[89] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter 11*, 1 (2009), 10–18.

[90] Hamstra, D. A., Chenevert, T. L., Moffat, B. A., Johnson, T. D., Meyer, C. R., Mukherji, S. K., Quint, D. J., Gebarski, S. S., Fan, X., Tsien, C. I., et al. Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma. *Proceedings of the National Academy of Sciences of the United States of America 102*, 46 (2005), 16759–16764.

[91] Hamstra, D. A., Galbán, C. J., Meyer, C. R., Johnson, T. D., Sundgren, P. C., Tsien, C., Lawrence, T. S., Junck, L., Ross, D. J., Rehemtulla, A., et al. Functional diffusion map as an early imaging biomarker for high-grade glioma: correlation with conventional radiologic response and overall survival. *Journal of clinical oncology 26*, 20 (2008), 3387–3394.

[92] Hannoun, S., Bagory, M., Durand-Dubief, F., Ibarrola, D., Comte, J.-C., Confavreux, C., Cotton, F., and Sappey-Marinier, D. Correlation of diffusion and metabolic alterations in different clinical forms of multiple sclerosis. *PLoS One 7*, 3 (2012), e32525.

[93] Haralick, R. M., Shanmugam, K., and Dinstein, I. H. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 6 (1973), 610–621.

[94] Havrdova, E., Galetta, S., Hutchinson, M., Stefoski, D., Bates, D., Polman, C. H., O'Connor, P. W., Giovannoni, G., Phillips, J. T., Lublin, F. D., et al. Effect of natalizumab on clinical and radiological disease activity in multiple sclerosis: a retrospective analysis of the natalizumab safety and efficacy in relapsing-remitting multiple sclerosis (affirm) study. *The Lancet Neurology 8*, 3 (2009), 254–260.

[95] Hayton, T., Furby, J., Smith, K., Altmann, D., Brenner, R., Chataway, J., Hughes, R. A., Hunter, K., Tozer, D., Miller, D., et al. Grey matter magnetization transfer ratio independently correlates with neurological deficit in secondary progressive multiple sclerosis. *Journal of neurology 256*, 3 (2009), 427–435.

[96] He, J., Inglese, M., Li, B. S., Babb, J. S., Grossman, R. I., and Gonen, O. Relapsing-Remitting Multiple Sclerosis: Metabolic Abnormality in Nonenhancing Lesions and Normal-appearing White Matter at MR Imaging: Initial Experience 1. *Radiology 234*, 1 (2005), 211–217.

[97] He, K., Zhang, X., Ren, S., and Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (2015), pp. 1026–1034.

[98] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

[99] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine 29*, 6 (2012), 82–97.

[100] Hochmeister, S., Grundtner, R., Bauer, J., Engelhardt, B., Lyck, R., Gordon, G., Korosec, T., Kutzelnigg, A., Berger, J. J., Bradl, M., et al. Dysferlin is a new marker for leaky brain blood vessels in multiple sclerosis. *Journal of Neuropathology & Experimental Neurology 65*, 9 (2006), 855–865.

[101] Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. *Neural networks 2*, 5 (1989), 359–366.

[102] Hothorn, T., Bretz, F., and Westfall, P. Simultaneous inference in general parametric models. *Biometrical journal 50*, 3 (2008), 346–363.

[103] Hu, L., Baxter, L., Pinnaduwage, D., Paine, T., Karis, J., Feuerstein, B., Schmainda, K., Dueck, A., Debbins, J., Smith, K., et al. Optimized preload leakage-correction methods to improve the diagnostic accuracy of dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging in posttreatment gliomas. *American Journal of Neuroradiology 31*, 1 (2010), 40–48.

[104] Hu, L. S., Eschbacher, J. M., Heiserman, J. E., Dueck, A. C., Shapiro, W. R., Liu, S., Karis, J. P., Smith, K. A., Coons, S. W., Nakaji, P., et al. Reevaluating the imaging definition of tumor progression: perfusion MRI quantifies recurrent glioblastoma tumor fraction, pseudoprogression, and radiation necrosis to predict survival. *Neuro-oncology 14*, 7 (2012), 919–930.

[105] Hu, X., Wong, K. K., Young, G. S., Guo, L., and Wong, S. T. Support vector machine multiparametric MRI identification of pseudoprogression from tumor recurrence in patients with resected glioblastoma. *Journal of Magnetic Resonance Imaging 33*, 2 (2011), 296–305.

[106] Hua, J., Xiong, Z., Lowey, J., Suh, E., and Dougherty, E. R. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics 21*, 8 (2005), 1509–1515.

[107] Huang, J. Accelerating AI with GPUs: A New Computing Model. https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-artificial-intelligence-gpus/. Accessed: 2017-07-06.

[108] Hui, E. S., Cheung, M. M., Qi, L., and Wu, E. X. Towards better MR characterization of neural tissues using directional diffusion kurtosis analysis. *Neuroimage 42*, 1 (2008), 122–134.

[109] Husted, C., Goodin, D., Hugg, J., Maudsley, A. A., Tsuruda, J., De Bie, S., Fein, G., Matson, G., and Weiner, M. Biochemical alterations in multiple sclerosis lesions and normal-appearing white matter detected by in vivo 31P and 1H spectroscopic imaging. *Annals of neurology 36*, 2 (1994), 157–165.

[110] Indicators, O. Health at a glance 2011. *OECD Indicators, OECD Publishing, Paris 15* (2015), 2016.

[111] Inglese, M., and Bester, M. Diffusion imaging in multiple sclerosis: research and clinical implications. *NMR in Biomedicine 23*, 7 (2010), 865–872.

[112] Inglese, M., Li, B. S., Rusinek, H., Babb, J. S., Grossman, R. I., and Gonen, O. Diffusely elevated cerebral choline and creatine in relapsing-remitting multiple sclerosis. *Magnetic resonance in medicine 50*, 1 (2003), 190–195.

[113] Ioffe, S., and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning* (2015), pp. 448–456.

[114] Ion-Mărgineanu, A., Kocevar, G., Stamile, C., Sima, D. M., Durand-Dubief, F., Van Huffel, S., and Sappey-Marinier, D. Machine learning approach for classifying multiple sclerosis courses by combining clinical data with lesion loads and magnetic resonance metabolic features. *Frontiers in neuroscience 11* (2017).

[115] Ion-Mărgineanu, A., Van Cauter, S., Sima, D. M., Maes, F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Classifying glioblastoma multiforme follow-up progressive vs. responsive forms using multi-parametric mri features. *Frontiers in neuroscience 10* (2016).

[116] Ion-Mărgineanu, A., Van Cauter, S., Sima, D. M., Maes, F., Van Gool, S. W., Sunaert, S., Himmelreich, U., and Van Huffel, S. Tumour relapse prediction using multiparametric MR data recorded during follow-up of GBM patients. *BioMed research international 2015* (2015).

[117] Iranzo, A., Moreno, A., Pujol, J., Martí-Fàbregas, J., Domingo, P., Molet, J., Ris, J., and Cadafalch, J. Proton magnetic resonance spectroscopy pattern of progressive multifocal leukoencephalopathy in AIDS. *Journal of Neurology, Neurosurgery & Psychiatry 66*, 4 (1999), 520–523.

[118] Jain, S., Sima, D. M., Ribbens, A., Cambron, M., Maertens, A., Van Hecke, W., De Mey, J., Barkhof, F., Steenwijk, M. D., Daams, M., et al. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *NeuroImage: Clinical 8* (2015), 367–375.

[119] Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. Fsl. *Neuroimage 62*, 2 (2012), 782–790.

[120] Jensen, J. H., Helpern, J. A., Ramani, A., Lu, H., and Kaczynski, K. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magnetic Resonance in Medicine 53*, 6 (2005), 1432–1440.

[121] Kappos, L., De Stefano, N., Freedman, M. S., Cree, B. A., Radue, E.-W., Sprenger, T., Sormani, M. P., Smith, T., Häring, D. A., Piani Meier, D., et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity'(NEDA-4) in relapsing–remitting multiple sclerosis. *Multiple Sclerosis Journal 22*, 10 (2016), 1297–1305.

[122] Karpathy, A. CS231n convolutional neural networks for visual recognition.

[123] Khalifa, J., Tensaouti, F., Chaltiel, L., Lotterie, J.-A., Catalaa, I., Sunyach, M., Ibarrola, D., Noël, G., Truc, G., Walker, P., et al. Identification of a candidate biomarker from perfusion mri to anticipate glioblastoma progression after chemoradiation. *European radiology* (2016), 1–10.

[124] Kirov, I. I., Liu, S., Tal, A., Wu, W. E., Davitz, M. S., Babb, J. S., Rusinek, H., Herbert, J., and Gonen, O. Proton MR spectroscopy of lesion evolution in multiple sclerosis: Steady-state metabolism and its relationship to conventional imaging. *Human Brain Mapping 38*, 8 (2017), 4047–4063.

[125] Kleihues, P., and Ohgaki, H. Phenotype vs genotype in the evolution of astrocytic brain tumors. *Toxicologic pathology 28*, 1 (2000), 164–170.

[126] Kocevar, G. *Développement de Méthodes d'IRM Avancées pour l'Etude Longitudinale de la Sclérose en Plaques*. PhD thesis, Claude Bernard Lyon 1, 2017.

[127] Kocevar, G., Stamile, C., Hannoun, S., Cotton, F., Vukusic, S., Durand-Dubief, F., and Sappey-Marinier, D. Graph theory-based brain connectivity for automatic classification of multiple sclerosis clinical courses. *Frontiers in Neuroscience 10* (2016), 478.

[128] Kononenko, I., Šimec, E., and Robnik-Šikonja, M. Overcoming the myopia of inductive learning algorithms with RELIEFF. *Applied Intelligence 7*, 1 (1997), 39–55.

[129] Koshy, M., Villano, J. L., Dolecek, T. A., Howard, A., Mahmood, U., Chmura, S. J., Weichselbaum, R. R., and McCarthy, B. J. Improved survival time trends for glioblastoma using the SEER 17 population-based registries. *Journal of neuro-oncology 107*, 1 (2012), 207–212.

[130] Kreis, R. Issues of spectral quality in clinical ^{1}H-magnetic resonance spectroscopy and a gallery of artifacts. *NMR in Biomedicine 17*, 6 (2004), 361–381.

[131] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (2012), pp. 1097–1105.

[132] Kruskal, W. H., and Wallis, W. A. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association 47*, 260 (1952), 583–621.

[133] Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis an expanded disability status scale (EDSS). *Neurology 33*, 11 (1983), 1444–1444.

[134] Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. Package 'lmertest'. *R package version* (2015).

[135] Lamborn, K. R., Chang, S. M., and Prados, M. D. Prognostic factors for survival of patients with glioblastoma: recursive partitioning analysis. *Neuro-oncology 6*, 3 (2004), 227–235.

[136] LAMDA. `http://lamda.nju.edu.cn/Data.ashx`. [Online; accessed 09-August-2014].

[137] Larsson, H., Christiansen, P., Jensen, M., Frederiksen, J., Heltberg, A., Olesen, J., and Henriksen, O. Localized in vivo proton spectroscopy in the brain of patients with multiple sclerosis. *Magnetic resonance in medicine 22*, 1 (1991), 23–31.

[138] Lassmann, H., Van Horssen, J., and Mahad, D. Progressive multiple sclerosis: pathology and pathogenesis. *Nature Reviews Neurology 8*, 11 (2012), 647–656.

[139] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., and Laval-Jeantet, M. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. *Radiology 161*, 2 (1986), 401–407.

[140] Leary, S. M., Davie, C. A., Parker, G. J., Stevenson, V. L., Wang, L., Barker, G. J., Miller, D. H., and Thompson, A. 1H magnetic resonance spectroscopy of normal appearing white matter in primary progressive multiple sclerosis. *Journal of neurology 246*, 11 (1999), 1023–1026.

[141] Li, Y.-F., and Zhou, Z.-H. Towards making unlabeled data never hurt. *IEEE Transactions on Pattern Analysis and Machine Intelligence 37*, 1 (2015), 175–188.

[142] Liaw, A., and Wiener, M. Classification and regression by randomforest. *R News 2*, 3 (2002), 18–22.

[143] Lippmann, R. P. An introduction to computing with neural nets. *ASSP Magazine, IEEE 4*, 2 (1987), 4–22.

[144] Louis, D. N., Ohgaki, H., Wiestler, O. D., Cavenee, W. K., Burger, P. C., Jouvet, A., Scheithauer, B. W., and Kleihues, P. The 2007 WHO classification of tumours of the central nervous system. *Acta neuropathologica 114*, 2 (2007), 97–109.

[145] Louppe, G. Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502* (2014).

[146] Lublin, F. D., Reingold, S. C., et al. Defining the clinical course of multiple sclerosis results of an international survey. *Neurology 46*, 4 (1996), 907–911.

[147] Lund, R., Rand, S., Krouwer, H., Schultz, C., and Schmainda, K. Using rCBV to distinguish radiation necrosis from tumor recurrence in malignant gliomas. *International Journal of Radiation Oncology*Biology*Physics 63, Supplement 1*, 0 (2005), S65 – S66.

Proceedings of the American Society for Therapeutic Radiology and Oncology 47 th Annual Meeting American Society for Therapeutic Radiology and Oncology 47th Annual Meeting.

[148] Macyszyn, L., Akbari, H., Pisapia, J. M., Da, X., Attiah, M., Pigrish, V., Bi, Y., Pal, S., Davuluri, R. V., Roccograndi, L., et al. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. *Neuro-oncology 18*, 3 (2016), 417–425.

[149] McAlpine, D., and Compston, A. *McAlpine's multiple sclerosis*. Elsevier Health Sciences, 2005.

[150] McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the international Panel on the diagnosis of multiple sclerosis. *Annals of neurology 50*, 1 (2001), 121–127.

[151] Miller, D. H., Chard, D. T., and Ciccarelli, O. Clinically isolated syndromes. *The Lancet Neurology 11*, 2 (2012), 157–169.

[152] Moffat, B. A., Chenevert, T. L., Lawrence, T. S., Meyer, C. R., Johnson, T. D., Dong, Q., Tsien, C., Mukherji, S., Quint, D. J., Gebarski, S. S., et al. Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proceedings of the National Academy of Sciences of the United States of America 102*, 15 (2005), 5524–5529.

[153] Möller-Hartmann, W., Herminghaus, S., Krings, T., Marquardt, G., Lanfermann, H., Pilatus, U., and Zanella, F. Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. *Neuroradiology 44*, 5 (2002), 371–381.

[154] Mukherjee, P., Berman, J., Chung, S., Hess, C., and Henry, R. Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. *American journal of neuroradiology 29*, 4 (2008), 632–641.

[155] Muthuraman, M., Fleischer, V., Kolber, P., Luessi, F., Zipp, F., and Groppa, S. Structural brain network characteristics can differentiate CIS from early RRMS. *Frontiers in neuroscience 10* (2016).

[156] Nakahara, J., Maeda, M., Aiso, S., and Suzuki, N. Current concepts in multiple sclerosis: autoimmunity versus oligodendrogliopathy. *Clinical reviews in allergy & immunology 42*, 1 (2012), 26–34.

[157] Narayana, P. A. Magnetic resonance spectroscopy in the monitoring of multiple sclerosis. *Journal of Neuroimaging 15*, s4 (2005), 46S–57S.

[158] Narayana, P. A., Doyle, T. J., Lai, D., and Wolinsky, J. S. Serial proton magnetic resonance spectroscopic imaging, contrast-enhanced magnetic resonance imaging, and quantitative lesion volumetry in multiple sclerosis. *Annals of neurology 43*, 1 (1998), 56–71.

[159] Narayana, P. A., Wolinsky, J. S., Rao, S. B., He, R., Mehta, M., et al. Multicentre proton magnetic resonance spectroscopy imaging of primary progressive multiple sclerosis. *Multiple Sclerosis 10*, 3 suppl (2004), S73–S78.

[160] Narayanan, S., Fu, L., Pioro, E., De Stefano, N., Collins, D., Francis, G., Antel, J., Matthews, P., and Arnold, D. Imaging of axonal damage in multiple sclerosis: spatial distribution of magnetic resonance imaging lesions. *Annals of neurology 41*, 3 (1997), 385–391.

[161] Nasseri, M., Gahramanov, S., Netto, J. P., Fu, R., Muldoon, L. L., Varallyay, C., Hamilton, B. E., and Neuwelt, E. A. Evaluation of pseudoprogression in patients with glioblastoma multiforme using dynamic magnetic resonance imaging with ferumoxytol calls RANO criteria into question. *Neuro-oncology* (2014), 328.

[162] Nelson, S. J., and Cha, S. Imaging glioblastoma multiforme. *The Cancer Journal 9*, 2 (2003), 134–145.

[163] Nguyen, C., Wang, Y., and Nguyen, H. N. Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. *Journal of Biomedical Science and Engineering 6*, 05 (2013), 551.

[164] Nielsen, M. A. Neural networks and deep learning, 2015.

[165] Niu, G., Jitkrittum, W., Dai, B., Hachiya, H., and Sugiyama, M. Squared-loss mutual information regularization: A novel information-theoretic approach to semi-supervised learning. In *Proceedings of The 30th International Conference on Machine Learning* (2013), pp. 10–18.

[166] Ohgaki, H. Epidemiology of brain tumors. *Cancer Epidemiology: Modifiable Factors* (2009), 323–342.

[167] Ohgaki, H., and Kleihues, P. Genetic alterations and signaling pathways in the evolution of gliomas. *Cancer science 100*, 12 (2009), 2235–2241.

[168] Østergaard, L., Weisskoff, R. M., Chesler, D. A., Gyldensted, C., and Rosen, B. R. High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. part i: Mathematical approach and statistical analysis. *Magnetic resonance in medicine 36*, 5 (1996), 715–725.

[169] Ostrom, Q. T., Gittleman, H., Farah, P., Ondracek, A., Chen, Y., Wolinsky, Y., Stroup, N. E., Kruchko, C., and Barnholtz-Sloan, J. S. CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. *Neuro-oncology 15*, suppl 2 (2013), ii1–ii56.

[170] Ourselin, S., Stefanescu, R., and Pennec, X. Robust registration of multi-modal images: towards real-time clinical applications. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002* (2002), 140–147.

[171] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research 12*, Oct (2011), 2825–2830.

[172] Phuphanich, S., Wheeler, C., Rudnick, J., Hu, J., Mazer, M., Sanchez, C., Nuno, M., Chu, R., Black, K., and Yu, J. ATIM-25. TEN-YEAR FOLLOW UP WITH LONG TERM REMISSION IN PATIENTS WITH NEWLY DIAGNOSED GLIOBLASTOMA (GBM) TREATED WITH ICT-107 VACCINE (PHASE I). *Neuro-Oncology 18*, suppl 6 (2016), vi23–vi23.

[173] Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of neurology 69*, 2 (2011), 292–302.

[174] Polman, C. H., Reingold, S. C., Edan, G., Filippi, M., Hartung, H.-P., Kappos, L., Lublin, F. D., Metz, L. M., McFarland, H. F., O'Connor, P. W., et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Annals of neurology 58*, 6 (2005), 840–846.

[175] Poot, D. H., den Dekker, A. J., Achten, E., Verhoye, M., and Sijbers, J. Optimal experimental design for diffusion kurtosis imaging. *Medical Imaging, IEEE Transactions on 29*, 3 (2010), 819–829.

[176] Pope, W. B., Kim, H. J., Huo, J., Alger, J., Brown, M. S., Gjertson, D., Sai, V., Young, J. R., Tekchandani, L., Cloughesy, T., et al. Recurrent glioblastoma multiforme: ADC histogram analysis predicts response to bevacizumab treatment. *Radiology 252*, 1 (2009), 182–189.

[177] Posse, S., Otazo, R., Dager, S. R., and Alger, J. MR spectroscopic imaging: principles and recent advances. *Journal of Magnetic Resonance Imaging 37*, 6 (2013), 1301–1325.

[178] Poullet, J.-B. *Quantification and classification of magnetic resonance spectroscopic data for brain tumor diagnosis*. PhD thesis, KU Leuven, 2008.

[179] Poullet, J.-B., Sima, D., Luts, J., Garcia, M. O., Croitor, A., and Van Huffel, S. Manual: Simulation Package based on In vitro Databases (SPID). Accessed: 2015-02-10.

[180] Poullet, J.-B., Sima, D. M., Simonetti, A. W., De Neuter, B., Vanhamme, L., Lemmerling, P., and Van Huffel, S. An automated quantitation of short echo time MRS spectra in an open source software environment: AQSES. *NMR in Biomedicine 20*, 5 (2007), 493–504.

[181] Provencher, S. W. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. *Magnetic resonance in medicine 30*, 6 (1993), 672–679.

[182] Ranjeva, J.-P., Audoin, B., Duong, M. V. A., Ibarrola, D., Confort-Gouny, S., Malikova, I., Soulier, E., Viout, P., Ali-Chérif, A., Pelletier, J., et al. Local tissue damage assessed with statistical mapping analysis of brain magnetization transfer ratio: relationship with functional status of patients in the earliest stage of multiple sclerosis. *American Journal of Neuroradiology 26*, 1 (2005), 119–127.

[183] Ratiney, H., Sdika, M., Coenradie, Y., Cavassila, S., Ormondt, D. v., and Graveron-Demilly, D. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR in Biomedicine 18*, 1 (2005), 1–13.

[184] Rees, J. Advances in magnetic resonance imaging of brain tumours. *Current opinion in neurology 16*, 6 (2003), 643–650.

[185] Rinck, P. A. The basic textbook of the european magnetic resonance forum. `http://www.magnetic-resonance.org/ch/21-01.htm`. Accessed: 2017-04-05.

[186] Ropele, S., and Fazekas, F. Magnetization transfer MR imaging in multiple sclerosis. *Neuroimaging Clinics of North America 19*, 1 (2009), 27–36.

[187] Rosen, B. R., Belliveau, J. W., Vevea, J. M., and Brady, T. J. Perfusion imaging with NMR contrast agents. *Magnetic resonance in medicine 14*, 2 (1990), 249–265.

[188] Rosenblatt, F. *Principles of neurodynamics*. Spartan Book, 1962.

[189] Rovaris, M., Gass, A., Bammer, R., Hickman, S., Ciccarelli, O., Miller, D., and Filippi, M. Diffusion mri in multiple sclerosis. *Neurology 65*, 10 (2005), 1526–1532.

[190] Rovaris, M., Judica, E., Sastre-Garriga, J., Rovira, A., Pia Sormani, M., Benedetti, B., Korteweg, T., De Stefano, N., Khaleeli, Z., Montalban, X., et al. Large-scale, multicentre, quantitative MRI study of brain and cord damage in primary progressive multiple sclerosis. *Multiple Sclerosis Journal 14*, 4 (2008), 455–464.

[191] Rovira, À., Auger, C., and Alonso, J. Magnetic resonance monitoring of lesion evolution in multiple sclerosis. *Therapeutic advances in neurological disorders 6*, 5 (2013), 298–310.

[192] Roychowdhury, S., Maldjian, J. A., and Grossman, R. I. Multiple sclerosis: comparison of trace apparent diffusion coefficients with mr enhancement pattern of lesions. *American journal of neuroradiology 21*, 5 (2000), 869–874.

[193] Rulseh, A. M., Keller, J., Klener, J., Sroubek, J., Dbaly, V., Syrucek, M., Tovarys, F., and Vymazal, J. Long-term survival of patients suffering from glioblastoma multiforme treated with tumor-treating fields. *World Journal of Surgical Oncology 10*, 1 (2012), 220.

[194] Rutkowski, S., De Vleeschouwer, S., Kaempgen, E., Wolff, J., Kühl, J., Demaerel, P., Warmuth-Metz, M., Flamen, P., Van Calenbergh, F., Plets, C., et al. Surgery and adjuvant dendritic cell-based tumour vaccination for patients with relapsed malignant glioma, a feasibility study. *British journal of cancer 91*, 9 (2004), 1656–1662.

[195] Sajja, B. R., Wolinsky, J. S., and Narayana, P. A. Proton magnetic resonance spectroscopy in multiple sclerosis. *Neuroimaging clinics of North America 19*, 1 (2009), 45–58.

[196] Sarchielli, P., Presciutti, O., Pelliccioli, G., Tarducci, R., Gobbi, G., Chiarini, P., Alberti, A., Vicinanza, F., and Gallai,

V. Absolute quantification of brain metabolites by proton magnetic resonance spectroscopy in normal-appearing white matter of multiple sclerosis patients. *Brain 122*, 3 (1999), 513–521.

[197] Sauwen, N. *Unsupervised and semi-supervised Non-negative Matrix Factorization methods for brain tumor segmentation using multi-parametric MRI data.* PhD thesis, KU Leuven, 2016.

[198] Scalfari, A., Neuhaus, A., Degenhardt, A., Rice, G. P., Muraro, P. A., Daumer, M., and Ebers, G. C. The natural history of multiple sclerosis, a geographically based study 10: relapses and long-term disability. *Brain 133*, 7 (2010), 1914–1929.

[199] Schmainda, K. M., Prah, M., Connelly, J., Rand, S. D., Hoffman, R. G., Mueller, W., and Malkin, M. G. Dynamic-susceptibility contrast agent mri measures of relative cerebral blood volume predict response to bevacizumab in recurrent high-grade glioma. *Neuro-oncology 16*, 6 (2014), 880–888.

[200] Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. Nih image to imagej: 25 years of image analysis. *Nature methods 9*, 7 (2012), 671–675.

[201] Seber, G. A. *Multivariate observations*, vol. 252. John Wiley & Sons, 2009.

[202] Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., and Napolitano, A. Rusboost: Improving classification performance when training data is skewed. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (2008), IEEE, pp. 1–4.

[203] Shu, N., Liu, Y., Li, K., Duan, Y., Wang, J., Yu, C., Dong, H., Ye, J., and He, Y. Diffusion tensor tractography reveals disrupted topological efficiency in white matter structural networks in multiple sclerosis. *Cerebral Cortex 21*, 11 (2011), 2565–2577.

[204] Sicotte, N. L., Voskuhl, R. R., Bouvier, S., Klutch, R., Cohen, M. S., and Mazziotta, J. C. Comparison of multiple sclerosis lesions at 1.5 and 3.0 Tesla. *Investigative radiology 38*, 7 (2003), 423–427.

[205] Šidák, Z. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association 62*, 318 (1967), 626–633.

[206] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I.,

Panneershelvam, V., Lanctot, M., et al. Mastering the game of go with deep neural networks and tree search. *Nature 529*, 7587 (2016), 484–489.

[207] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[208] Smets, T., Lawson, T. M., Grandin, C., Jankovski, A., and Raftopoulos, C. Immediate post-operative MRI suggestive of the site and timing of glioblastoma recurrence after gross total resection: a retrospective longitudinal preliminary study. *European radiology 23*, 6 (2013), 1467–1477.

[209] Smith, S. M. Fast robust automated brain extraction. *Human brain mapping 17*, 3 (2002), 143–155.

[210] Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., et al. Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage 23* (2004), S208–S219.

[211] Soh, L.-K., and Tsatsoulis, C. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *Geoscience and Remote Sensing, IEEE Transactions on 37*, 2 (1999), 780–795.

[212] SPID. `http://homes.esat.kuleuven.be/~biomed/software.php#SpidGUI`. [Online; accessed 09-August-2014].

[213] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research 15*, 1 (2014), 1929–1958.

[214] Stejskal, E. O., and Tanner, J. E. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics 42*, 1 (1965), 288–292.

[215] Stewart, G. Researches on the circulation time in organs and on the influences which affect it: Parts i.—iii. *The Journal of Physiology 15*, 1-2 (1893), 1.

[216] Stupp, R., Mason, W. P., Van Den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J., Belanger, K., Brandes, A. A., Marosi, C., Bogdahn, U., et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *New England Journal of Medicine 352*, 10 (2005), 987–996.

[217] Sugahara, T., Korogi, Y., Tomiguchi, S., Shigematsu, Y., Ikushima, I., Kira, T., Liang, L., Ushio, Y., and Takahashi, M. Posttherapeutic intraaxial brain tumor: the value of perfusion-sensitive contrast-enhanced MR imaging for differentiating tumor recurrence from nonneoplastic contrast-enhancing tissue. *American Journal of Neuroradiology 21*, 5 (2000), 901–909.

[218] Sugiyama, M. `http://www.ms.k.u-tokyo.ac.jp/software.html`. [Online; accessed 09-August-2014].

[219] Suhy, J., Rooney, W., Goodkin, D., Capizzano, A., Soher, B., Maudsley, A. A., Waubant, E., Andersson, P., and Weiner, M. 1H MRSI comparison of white matter and lesions in primary progressive and relapsing-remitting MS. *Multiple sclerosis 6*, 3 (2000), 148–155.

[220] Sundin, T., Vanhamme, L., Van Hecke, P., Dologlou, I., and Van Huffel, S. Accurate quantification of 1 H spectra: From finite impulse response filter design for solvent suppression to parameter estimation. *Journal of Magnetic Resonance 139*, 2 (1999), 189–204.

[221] Suykens, J. A., Van Gestel, T., De Brabanter, J., De Moor, B., and Vandewalle, J. *Least squares support vector machines*. World Scientific, 2002.

[222] Suykens, J. A., and Vandewalle, J. Least squares support vector machine classifiers. *Neural processing letters 9*, 3 (1999), 293–300.

[223] Tartaglia, M., Narayanan, S., De Stefano, N., Arnaoutelis, R., Antel, S., Francis, S., Santos, A., Lapierre, Y., and Arnold, D. Choline is increased in pre-lesional normal appearing white matter in multiple sclerosis. *Journal of neurology 249*, 10 (2002), 1382–1390.

[224] Thakkar, J. P., Dolecek, T. A., Horbinski, C., Ostrom, Q. T., Lightner, D. D., Barnholtz-Sloan, J. S., and Villano, J. L. Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiology and Prevention Biomarkers 23*, 10 (2014), 1985–1996.

[225] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints abs/1605.02688* (May 2016).

[226] Tiberio, M., Chard, D., Altmann, D., Davies, G., Griffin, C., McLean, M., Rashid, W., Sastre-Garriga, J., Thompson, A., and Miller, D. Metabolite changes in early relapsing–remitting multiple sclerosis. *Journal of neurology 253*, 2 (2006), 224–230.

[227] Tourbah, A., Stievenart, J.-L., Abanou, A., Iba-Zizen, M.-T., Hamard, H., Lyon-Caen, O., Cabanis, E., and Stievenart, L. Normal-appearing white matter in optic neuritis and multiple sclerosis: a comparative proton spectroscopy study. *Neuroradiology 41*, 10 (1999), 738–743.

[228] Tran, B., and Rosenthal, M. Survival comparison between glioblastoma multiforme and other incurable cancers. *Journal of Clinical Neuroscience 17*, 4 (2010), 417–421.

[229] Tsien, C., Galbán, C. J., Chenevert, T. L., Johnson, T. D., Hamstra, D. A., Sundgren, P. C., Junck, L., Meyer, C. R., Rehemtulla, A., Lawrence, T., et al. Parametric response map as an imaging biomarker to distinguish progression from pseudoprogression in high-grade glioma. *Journal of Clinical Oncology 28*, 13 (2010), 2293–2299.

[230] Tsuruda, J. S., Kortman, K. E., Bradley, W. G., Wheeler, D. C., Van Dalsem, W., and Bradley, T. P. Radiation effects on cerebral white matter: MR evaluation. *American Journal of Roentgenology 149*, 1 (1987), 165–171.

[231] Van Cauter, S., De Keyzer, F., Sima, D., Croitor Sava, A., D'Arco, F., Veraart, J., Peeters, R. R., Leemans, A., Van Gool, S., Wilms, G., et al. Integrating diffusion kurtosis imaging, dynamic susceptibility-weighted MR imaging and short echo time chemical shift imaging for grading gliomas. *Neuro-oncology 16*, 7 (2014), 1010–1021.

[232] Van Cauter, S., Sima, D. M., Luts, J., ter Beek, L., Ribbens, A., Peeters, R. R., Osorio Garcia, M. I., Li, Y., Sunaert, S., Van Gool, S. W., et al. Reproducibility of rapid short echo time CSI at 3T for clinical applications. *Journal of Magnetic Resonance Imaging 37*, 2 (2013), 445–456.

[233] Van Cauter, S., Veraart, J., Sijbers, J., Peeters, R. R., Himmelreich, U., De Keyzer, F., Van Gool, S. W., Van Calenbergh, F., De Vleeschouwer, S., Van Hecke, W., et al. Gliomas: diffusion kurtosis MR imaging in grading. *Radiology 263*, 2 (2012), 492–501.

[234] Van Gool, S., Maes, W., Ardon, H., Verschuere, T., Van Cauter, S., and De Vleeschouwer, S. Dendritic cell therapy of high-grade gliomas. *Brain Pathology 19*, 4 (2009), 694–712.

[235] Vrabec, M., Van Cauter, S., Himmelreich, U., Van Gool, S. W., Sunaert, S., De Vleeschouwer, S., Šuput, D., and Demaerel,

P. MR perfusion and diffusion imaging in the follow-up of recurrent glioblastoma treated with dendritic cell immunotherapy: a pilot study. *Neuroradiology 53*, 10 (2011), 721–731.

[236] Wang, S., Martinez-Lage, M., Sakai, Y., Chawla, S., Kim, S., Alonso-Basanta, M., Lustig, R., Brem, S., Mohan, S., Wolf, R., et al. Differentiating tumor progression from pseudoprogression in patients with glioblastomas using diffusion tensor imaging and dynamic susceptibility contrast MRI. *American Journal of Neuroradiology 37*, 1 (2016), 28–36.

[237] Wattjes, M., Harzheim, M., Lutterbey, G., Klotz, L., Schild, H., and Träber, F. Axonal damage but no increased glial cell activity in the normal-appearing white matter of patients with clinically isolated syndromes suggestive of multiple sclerosis using high-field magnetic resonance spectroscopy. *American Journal of Neuroradiology 28*, 8 (2007), 1517–1522.

[238] Wattjes, M. P., Harzheim, M., Lutterbey, G. G., Bogdanow, M., Schild, H. H., and Träber, F. High field MR imaging and 1H-MR spectroscopy in clinically isolated syndromes suggestive of multiple sclerosis. *Journal of neurology 255*, 1 (2008), 56–63.

[239] Wen, P. Y., Macdonald, D. R., Reardon, D. A., Cloughesy, T. F., Sorensen, A. G., Galanis, E., DeGroot, J., Wick, W., Gilbert, M. R., Lassman, A. B., et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Journal of Clinical Oncology 28*, 11 (2010), 1963–1972.

[240] Wen, Q., Jalilian, L., Lupo, J. M., Molinaro, A. M., Chang, S. M., Clarke, J., Prados, M., and Nelson, S. J. Comparison of adc metrics and their association with outcome for patients with newly diagnosed glioblastoma being treated with radiation therapy, temozolomide, erlotinib and bevacizumab. *Journal of neuro-oncology 121*, 2 (2015), 331–339.

[241] Werring, D., Clark, C., Barker, G., Thompson, A., and Miller, D. Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. *Neurology 52*, 8 (1999), 1626–1626.

[242] Weybright, P., Sundgren, P. C., Maly, P., Hassan, D. G., Nan, B., Rohrer, S., and Junck, L. Differentiation between brain tumor recurrence and radiation injury using MR spectroscopy. *American Journal of Roentgenology 185*, 6 (2005), 1471–1476.

[243] Wolinsky, J. S., Narayana, P. A., and Fenstermacher, M. J. Proton magnetic resonance spectroscopy in multiple sclerosis. *Neurology 40*, 11 (1990), 1764–1764.

[244] Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. M. Bayesian analysis of neuroimaging data in fsl. *Neuroimage 45*, 1 (2009), S173–S186.

[245] Yamasaki, F., Sugiyama, K., Ohtaki, M., Takeshima, Y., Abe, N., Akiyama, Y., Takaba, J., Amatya, V. J., Saito, T., Kajiwara, Y., et al. Glioblastoma treated with postoperative radio-chemotherapy: prognostic value of apparent diffusion coefficient at MR imaging. *European journal of radiology 73*, 3 (2010), 532–537.

[246] Yang, Y., and Pedersen, J. O. A comparative study on feature selection in text categorization. In *ICML* (1997), vol. 97, pp. 412–420.

[247] Yoon, R. G., Kim, H. S., Kim, D. Y., Hong, G. S., and Kim, S. J. Apparent diffusion coefficient parametric response mapping MRI for follow-up of glioblastoma. *European radiology 26*, 4 (2016), 1037–1047.

[248] Zhang, J., Yu, H., Qian, X., Liu, K., Tan, H., Yang, T., Wang, M., Li, K. C., Chan, M. D., Debinski, W., et al. Pseudo progression identification of glioblastoma with dictionary learning. *Computers in Biology and Medicine 73* (2016), 94–101.

[249] Zhang, Y., Brady, M., and Smith, S. Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging 20*, 1 (2001), 45–57.

[250] Zhang, Y., Li, A., Peng, C., and Wang, M. Improve glioblastoma multiforme prognosis prediction by using feature selection and multiple kernel learning. *IEEE/ACM transactions on computational biology and bioinformatics 13*, 5 (2016), 825–835.

[251] Zhu, X., and Goldberg, A. B. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning 3*, 1 (2009), 1–130.

[252] Zierler, K. L. Equations for measuring blood flow by external monitoring of radioisotopes. *Circulation research 16*, 4 (1965), 309–321.

Curriculum Vitae

Adrian Ion-Mărgineanu was born on March 29, 1988, in Constanta, Romania. He received the Bachelor degree in the field of Telecommunications and Electrical Engineering from the Polytechnic University of Bucharest (Romania) in July 2010. He received the M.Sc. degree in the field of Information Technology and Computing in September 2012. For his master thesis he worked in collaboration with SensoMotoric Instruments in Berlin (Germany), on the subject of 3-D modelling of light tracking inside the eye.

Between October 2012 and March 2013 he was employed as a software engineer at S.C. Comsys SRL in Bucharest (Romania), where he worked in the analysis of financial data. Between March and August 2013 he worked as a scientific researcher at the Institute for Neuro- and Bio-Informatics in Lübeck (Germany), on adaptive sampling of the visual world. In October 2013 he joined the KU Leuven Electrical Engineering Department (ESAT-STADIUS) as a PhD student under the supervision of Prof. Sabine Van Huffel. His research has been part of TRANSACT, a European Union funded FP7-PEOPLE Marie Curie Initial Training Network (PITN-GA-2012-316679), with the aim of Transforming Magnetic Resonance Spectroscopy into a Clinical Tool.

List of publications

Papers in international journals

1. **Ion-Mărgineanu, A.**, Kocevar, G., Stamile, C., Sima, D.M., Durand-Dubief, F., Van Huffel, S., and Sappey-Marinier, D. Machine learning approach for classifying Multiple Sclerosis courses by combining clinical data with lesion loads and Magnetic Resonance metabolic features. Frontiers in Neuroscience: Neurodegeneration, 2017. (Accepted)

2. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Classifying Glioblastoma Multiforme Follow-Up Progressive vs. Responsive Forms Using Multi-Parametric MRI Features. Frontiers in Neuroscience: Brain Imaging Methods, vol. 10, page 615, 2017.

3. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, F., W. Van Gool, S., Sunaert, S., Himmelreich, U., and Van Huffel, S. Tumour Relapse Prediction Using Multiparametric MR Data Recorded during Follow-Up of GBM Patients. BioMed Research International, vol. 2015, Article ID 842923, 13 pages, 2015.

Papers in preparation for submission to international journals

1. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Classification of Recurrent Glioblastoma using modified Parametric Response Maps of contrast-enhanced T1-weighted MRI and Perfusion MRI.

Papers accepted in proceedings of international conferences

1. **Ion-Mărgineanu, A.**, Kocevar, G., Stamile, C., Sima, D.M., Durand-Dubief, F., Van Huffel, S., and Sappey-Marinier, D. A comparison of Machine

Learning approaches for classifying Multiple Sclerosis courses using MRSI and brain segmentations. The 26th International Conference on Artificial Neural Networks, ICANN2017, Alghero, Italy, September 2017.

2. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Comparison of manual and semi-manual delineations for classifying glioblastoma multiforme patients based on histogram and texture MRI features. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN2017, Bruges, Belgium, April 2017.

Abstracts in proceedings of international conferences

1. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Classifying GBM follow-up outcome using semi-manual delineations and multi-parametric MRI. 6th Dutch Bio-Medical Engineering Conference, Egmond aan Zee, January 2017.

2. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Impact of semi-automatic delineation of hotspots of contrast enhancing region in predicting the outcome of GBM patients after brain surgery. 24th Annual Conference of the International Society for Magnetic Resonance in Medicine (ISMRM 2016), Singapore, May 2016.

3. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Impact of semi-automatic delineation of hotspots of contrast enhancing region in predicting the outcome of GBM patients after brain surgery. 7th Annual meeting of the International Society for Magnetic Resonance in Medicine, Benelux Chapter (ISMRMB 2016), Eindhoven, The Netherlands, Jan. 2016.

4. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Tumour relapse prediction using multi-parametric MR data recorded during follow-up of GBM patients. 23rd Annual Conference of the International Society for Magnetic Resonance in Medicine (ISMRM 2015), Toronto, Canada, Jun. 2015.

5. **Ion-Mărgineanu, A.**, Van Cauter, S., Sima, D.M., Maes, W. Van Gool, S., F., Sunaert, S., Himmelreich, U., and Van Huffel, S. Tumour relapse prediction using multi-parametric MR data recorded during follow-up of GBM patients. 7th Annual meeting of the International Society for Magnetic Resonance in Medicine, Benelux Chapter (ISMRMB 2015), Ghent, Belgium, Jan. 2015.





FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
ESAT-STADIUS-BIOMED
Kasteelpark Arenberg 10, box 2446
B-3001 Leuven

