Multivariate pattern analysis of anatomic, physiologic, and metabolic imaging data for improved management of patients with gliomas

by

Alexandra Elena Constantin

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Computer Science

and the Designated Emphasis

in

Communication, Computation, and Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ruzena Bajcsy, Chair
Professor Sarah J. Nelson
Professor Michael I. Jordan

Spring 2012

Multivariate pattern analysis of anatomic, physiologic, and metabolic imaging data for improved management of patients with gliomas

Abstract

Multivariate pattern analysis of anatomic, physiologic, and metabolic imaging data for improved management of patients with gliomas

by

Alexandra Elena Constantin

Doctor of Philosophy in Computer Science
Designated Emphasis in Communication, Computation, and Statistics

University of California, Berkeley

Professor Ruzena Bajcsy, Chair

The characterization of brain tumors involves the analysis of multiple heterogeneous data sets that include various types of medical images and spectroscopy, clinical and histopathology data, treatment history, and patient outcome. Of particular interest for such analyses is the utilization of recent advances in brain tumor imaging research, which have given rise to novel techniques for exploiting a number of different biological properties of normal and tumor tissue. The systematic analysis of these data could lead to better disease understanding, diagnosis, prognosis, and treatment.

The main focus of this thesis was the development of computer-assisted support for glioma understanding, diagnosis, and prognosis in clinical environments. This was achieved by analyzing heterogeneous biomedical data using multivariate pattern recognition methods. The tools developed in this thesis were used to characterize biological changes predictive of malignant transformations and treatment effects in gliomas, and for the early detection of disease progression. They were crucial in finding links between *in vivo* and *ex vivo* data that could give insight into the biology of brain cancer and help determine the right course of treatment for individual patients. The methods that are described in this thesis can contribute to clinical practice by improving the selection of biopsy sites and the targeting of treatment. The models that were learned in this thesis produced results with high classification accuracy, interpretability by means of clinical knowledge, and capacity to generalize the performance to new samples. The technical aspects covered in this thesis included the feature selection and modeling of biomedical data, the inference and evaluation of predictive models, and the use of models for clinical applications.

Acknowledgments

I am extremely thankful to my research advisors, Professors Ruzena Bajcsy and Sarah Nelson, whose encouragement, advice, and support during my graduate studies have made this thesis possible.

I would also like to thank Professor Michael Jordan for being part of my qualifying exam and dissertation committees and for his amazing courses on graphical models and Bayesian statistics; Doctor Annette Molinaro for her advice on statistics; Doctors Susan Chang, Soonmee Cha, and Tracy McKnight for helping me understand the clinical aspects of brain tumor research; Professor Mark D'Esposito for being on my qualifying exam committee, and for his course on fMRI methods. I am grateful to Doctor Joanna Phillips for her analysis of histological data and for graciously providing the histological images presented in this thesis.

I am also indebted to everyone in the Surbeck Laboratory for Advanced Imaging at University of California, San Francisco for their support, collaboration, discussion, and data acquisition. Particularly, I would like to thank Adam Elkhaled, Trey Jalbert, Jason Crane, Yan Li, Radhika Srinivasan, Emma Essock-Burns, Janine Lupo, Gabriela Bourne, Angela Jakary, Eugene Ozhinsky, and Chris Williams.

Generous financial support made my doctoral studies possible. I would like to thank Professor Sarah Nelson for providing funding for most of my studies through NIH grants CA097257, CA118816, and CA127612. I would also like to thank the Electrical Engineering and Computer Science Department at University of California, Berkeley for the privilege of serving as a graduate student instructor to David Wagner and Davis Tse and for providing my first year fellowship.

I am indebted to my undergraduate advisor, Professor Andrea Danyluk, who introduced me to machine learning and to research, and who continues to be a role model and an inspiration.

I would like to thank my family for their love and encouragement. I would like to express my deepest gratitude to my husband, Matt Piven, for his unconditional love, support, and understanding, and for editing most of my work. Lastly, I would like to thank my mom, Adriana Constantin, for raising me with a love for mathematics and computer science from a very early age, for her continuous involvement in my education, and for supporting me in all of my pursuits.

Contents

Contents **ii**

List of figures **vi**

List of tables **viii**

Acronyms **x**

1 Introduction **1**

2 The clinical assessment of patients with gliomas **3**
- 2.1 Nuclear magnetic resonance . . . 3
  - T1 and T2 relaxation . . . 4
    - Fluid-attenuated inversion recovery . . . 4
    - Contrast enhanced images . . . 5
  - Diffusion MRI . . . 6
  - Perfusion MRI . . . 6
  - Spectroscopy . . . 7
    - *In Vivo* MRS . . . 8
    - HRMAS spectroscopy . . . 10
    - Major metabolites in the brain . . . 11
- 2.2 Histopathology of gliomas . . . 13
  - Tumor cellularity . . . 14
  - MIB-1 index . . . 14
  - SMI-31 . . . 15
  - CA9 . . . 15
  - Microvascular morphology . . . 15
  - Necrosis . . . 16
  - WHO classification . . . 18
- 2.3 Important problems in the diagnosis and management of gliomas . . . 19

3 Multivariate pattern analysis of biomedical signals for patients with gliomas **22**
- 3.1 Preprocessing . . . 24

Preprocessing of MR images . . . 24
Preprocessing of spectra . . . 25
3.2 Feature selection . . . 25
Chi-squared feature selection . . . 27
Gain-Ratio feature selection . . . 27
Two-way conditional probability feature selection . . . 29
Wrapper-based feature selection . . . 30
3.3 Classification methods . . . 31
Linear discriminant analysis . . . 31
Logistic regression with ridge estimator . . . 32
Functional trees . . . 33
Support vector machines . . . 35
Decision stump boosting . . . 37
Mixed effects models . . . 38
3.4 Multivariate survival analysis . . . 39
Kaplan-Meier estimator of survival . . . 40
Cox proportional hazards model . . . 40
Brier score . . . 41
3.5 Evaluating the accuracy of a model . . . 41
Cross-validation . . . 41
Bootstrapping . . . 42
3.6 EM imputation of missing values . . . 43
3.7 Applications in brain tumor diagnosis, characterization, and treatment . . . 46

4 Identification of malignant transformations in gliomas using HRMAS spectra from image guided tissue samples 48
4.1 Data acquisition . . . 49
4.2 Analysis . . . 51
Preprocessing . . . 51
Feature selection . . . 51
Classification . . . 52
4.3 Validation . . . 53
4.4 Results . . . 53
Classification . . . 53
Feature selection results . . . 55
4.5 Discussion . . . 60
4.6 Conclusion . . . 61

5 Metabolic profiles of primary and secondary glioblastomas 62
5.1 Data acquisition, preprocessing, feature selection, and analysis . . . 63
5.2 Results . . . 63
5.3 Discussion . . . 63

6 Effects of treatment on the metabolic profile of grade 2 and grade 3 gliomas 65
6.1 Data acquisition 65
6.2 Analysis 66
6.3 Results 66
6.4 Discussion 69

7 Metabolic characterization of the histopathological properties of glioblastoma tissue 70
7.1 Data acquisition and preprocessing 71
7.2 Analysis 72
7.3 Results 72
7.4 Discussion 72

8 Identifying recurrent low grade gliomas that transformed to a higher grade based on *in vivo* magnetic resonance and spectroscopy 75
8.1 Data acquisition 76
8.2 Preprocessing 76
8.3 Analysis 76
8.4 Results 77
8.5 Discussion 78

9 Linking *ex vivo* and *in vivo* glioma biomarkers 82
9.1 Data acquisition 83
9.2 Preprocessing and analysis 84
9.3 Results 84
9.4 Discussion 86

10 Modeling tumor cellularity by linking MR imaging and spectroscopy with histology [21] 87
10.1 Data acquisition and preprocessing 87
10.2 Feature selection 88
10.3 Models 88
10.4 Results 89
10.5 Discussion 90

11 Multivariate serial analysis of glioblastomas 92
11.1 Data acquisition 92
11.2 Preprocessing 93
11.3 Analysis 93
11.4 Discussion 97

12 Automatic segmentation of brain tissue in patients with newly diagnosed gliomas 104
12.1 Data acquisition and preprocessing 104
12.2 Tumor segmentation 105

12.3 Healthy tissue segmentation . . . 106
12.4 Results . . . 107
12.5 Discussion . . . 107

13 Summary 110

Bibliography 113

List of figures

2.1 Anatomical MR images obtained from a patient diagnosed with GBM 5
2.2 Regions of interest 6
2.3 Diffusion maps obtained from a patient diagnosed with GBM 7
2.4 Perfusion imaging data obtained from a patient diagnosed with GBM 8
2.5 *In vivo* MRS data obtained from a patient diagnosed with GBM 9
2.6 *In vivo* lactate-edited MRS data obtained from a patient diagnosed with GBM 10
2.7 HRMAS and corresponding MRS obtained from a patient diagnosed with GBM 11
2.8 Multivariate imaging data and examples of normal and abnormal perfusion and spectroscopy 13
2.9 Example of biopsy site selection and corresponding histological appearance 14
2.10 Two H&E stained slides with different tumor scores 15
2.11 Two H&E stained slides with different MIB-1 scores 16
2.12 SMI-31 staining with different scores 17
2.13 CA9 staining 17
2.14 Factor VIII microvascular pattern and density 17
2.15 H&E slide showing necrosis and vascular hyperplasia 18
2.16 Histological appearance of tissue samples from gliomas of grades 2, 3, and 4 18
2.17 Heterogeneity of histological features in GBM tissue samples, by patient 19
2.18 T1-GAD and MRS of two seemingly similar biopsy sites with different histological properties 20

3.1 Supervised learning pipeline 23
3.2 Machine learning challenge: low and high tumor cellularity samples, plotted against two predictive features 24
3.3 Optimal number of features in a model as a function of the training set size and the classifier used 26
3.4 LDA example 32
3.5 A decision tree for determining whether a lesion transformed to a higher grade 34
3.6 SVM separating hyperplane 35
3.7 SVM slack variables 36
3.8 Example of SVM kernel application 37
3.9 Boosting 38

3.10 The training set and test set error rates converge to an asymptotic error 42

4.1 HRMAS spectrum for a recurrent grade 2 glioma 50
4.2 HRMAS spectrum for a recurrent LGG that transformed to grade 3 50
4.3 HRMAS spectrum for a recurrent LGG that transformed to grade 4 51
4.4 Plot of sensitivity and specificity of logistic ridge regression model for determining whether a tumor transformed to a higher grade 57
4.5 ROC curves comparing five classification methods used to distinguish between gliomas that transformed to a higher grade and those that did not 58

6.1 Metabolite levels in recurrent and newly diagnosed tumors, by grade 67

8.1 ROC curve of decision stump boosting model for non-invasively identifying whether a tumor transformed to a higher grade or not 80

10.1 Tumor cellularity predictions superimposed on a T1-GAD image 91

11.1 Event chart for 64 GBM patients 95
11.2 Kaplan-Meier overall survival curve 96
11.3 Nonlinearity test: Martingale-residual plots for the covariates in the Cox regression model 98
11.4 Nonlinearity test: component-plus-residual plots for the covariates in the Cox regression model 99
11.5 Influential observations: index plots of dfbeta for the Cox regression model 100
11.6 Proportional hazards assumption: plots of scaled Schoenfeld residuals against transformed time for each covariate in the Cox regression model 101
11.7 Prediction error curves for survival models, obtained using Brier score over time 102
11.8 Bootstrapping 0.632 prediction error curves for survival models with different covariates, obtained using Brier score over time 103

12.1 Segmentation results 109

List of tables

4.1 Characteristics of recurrent LGG patients enrolled in the study and treatment received prior to recurrence, tabulated according to the tumor grade at recurrence 49
4.2 Accuracy for distinguishing between recurrent LGGs which transformed to a higher grade and those that did not using five classifiers trained on HRMAS features selected using three feature selection methods 54
4.3 95% confidence intervals for accuracy results (presented in Table 4.2) for distinguishing between recurrent LGGs which transformed to a higher grade and those that did not using five classifiers trained on HRMAS features selected using three feature selection methods 55
4.4 Accuracy results for distinguishing between recurrent LGGs which upgraded to grade 3 and those that remained grade 2 using five classifiers trained on HRMAS features selected using three feature selection methods 56
4.5 Accuracy results for distinguishing between recurrent LGGs which upgraded to grade 4 and those that upgraded to grade 3 using five classifiers trained on HRMAS features selected using three feature selection methods 56
4.6 Percentage of times each feature was selected in one of the models for distinguishing between low grade tumors that recurred at different grades 59

5.1 Accuracy results for distinguishing between primary and secondary GBMs using five classifiers trained on HRMAS features selected using three feature selection methods . 63
5.2 95% confidence intervals for accuracy results (presented in Table 5.1) for distinguishing between primary and secondary GBMs using five classifiers trained on HRMAS features selected using three feature selection methods 64
5.3 Percentage of time each feature was selected in the logistic ridge regression model for distinguishing between primary and secondary GBMs 64

6.1 Number of treated and untreated patients with grade 2 and 3 gliomas 66
6.2 Bootstrapping accuracy of logistic ridge regression models for distinguishing between treated and untreated glioma tissue samples 68
6.3 Percentage of times each feature was selected as part of the model for distinguishing between newly diagnosed and recurrent grade 2 and 3 gliomas 68

6.4 Percentage of times each feature was selected as part of the model for distinguishing between newly diagnosed grade 3 gliomas and grade 2 gliomas that recurred at grade 3 68
6.5 Percentage of times each feature was selected as part of the model for distinguishing between newly diagnosed grade 2 gliomas and recurrent grade 2 gliomas 69

7.1 Number of tissue samples with low and high values of histopathological tissue properties 71
7.2 Accuracy results of multivariate decision stump boosting models for distinguishing between low and high values of histopathological tissue properties 73
7.3 Feature subsets capable of distinguishing between low and high values of histopathological tissue properties 73

8.1 LOOC and bootstrapping accuracy of additive logistic regression models for identifying recurrent LGGs that progressed to a higher grade 78
8.2 Feature subsets selected in the final decision stump boosting models which identify recurrent LGGs that progressed to a higher grade 78
8.3 Significance value of univariate parameters for logistic regression model which identifies recurrent LGGs that progressed to a higher grade 79

9.1 Number of tissue samples with low and high values of histopathological tissue properties. 83
9.2 Globally predictive features for distinguishing between low and high values of the histopathological parameters of GBM tissue 85
9.3 Accuracy results of decision stump boosting models for distinguishing between low and high values of the histopathological parameters of GBM tissue 86

10.1 Training and LOOC accuracy results for binary and three-category tumor cellularity models using different sets of parameters 89
10.2 Accuracy of binary cellularity model with EM estimated histology 90
10.3 Confusion matrix for full data three-category cellularity model 90

11.1 Significant univariate predictors of overall survival 94
11.2 Final multivariate model with significant interaction 97

12.1 Segmentation results 108

Acronyms

△R2* change in the relaxation rate.

2HG 2-hydroxyglutarate.

Acc accuracy.

Ace acetate.

ADC apparent diffusion coefficient.

ADP adenosine diphosphate.

Ala alanine.

AML acute myeloid leukemia.

Asc ascorbate.

Asp aspartate.

ATP adenosine triphosphate.

AUC area under the curve.

CBV cerebral blood volume.

CEL contrast enhancing lesion.

Cho choline.

CNI Cho to NAA ratio.

CPMG Carr-Purcell-Meiboom-Gill.

Cre creatine.

CSF cerebrospinal fluid.

DSC dynamic susceptibility contrast.

DTI diffusion tensor imaging.

DWI diffusion-weighted imaging.

EGFR epidermal growth factor receptor.

EM expectation maximization.

EPI echo planar imaging.

ERETIC electronic reference to access.

Eth ethanol.

FA fractional anisotropy.

FFT Fast Fourier Transform.

FID free induction decay.

FLAIR fluid-attenduated inversion recovery.

FSE fast spin echo.

FT functional tree.

GABA g-aminobutyric acid.

GBM glioblastoma multiforme.

Gd gadolinium.

Gd-DTPA gadolinium diethyltriamine pentaacetic acid.

GHMRF Gaussian Hidden Markov Random Field.

Glc glucose.

Gln glutamine.

Glu glutamate.

Glx glutamine and glutamate.

Gly glycine.

GM grey matter.

GPC glycerophosphocholine.

GSH glutathione.

HLSVD Hankel-Lanczos singular value decomposition.

HRMAS high resolution magic angle spinning.

Hyp-Tau hypo-taurine.

IDH1 isocitrate dehydrogenase 1.

IHC immunohistochemistry.

Lac lactate.

LDA linear discriminant analysis.

Leu leucine.

LGG low grade glioma.

Lip lipid.

LOOC leave-one-out cross-validation.

Lys lysine.

MCI myoinositol to total choline ratio.

ML maximum likelihood.

MR magnetic resonance.

MRF Markov random field.

MRI magnetic resonance imaging.

MRS magnetic resonance spectroscopy.

Myo-I myoinositol.

NAA N-acetylaspartate.

NAWM normal-appearing white matter.

NEC necrotic region.

NEL non-enhancing lesion.

NMR nuclear magnetic resonance.

OS overall survival.

PC phosphocholine.

PCA principal component analysis.

PCr phosphocreatine.

PE phosphoethanolamine.

PH peak height.

PTEN phosphatase and tensin homolog.

PWI perfusion-weighted imaging.

QP quadratic programming.

RECOV percent recovery to baseline.

RF recirculation factor.

RF pulse radio frequency pulse.

ROC receiver operating characteristic.

ROI region of interest.

SI scyllo-inositol.

SMO sequential minimal optimization.

Suc succinate.

SVM support vector machine.

T1-GAD T1-weighted sequence obtained after the injection of gadolinium contrast agent.

T2all FLAIR abnormality region.

Tau taurine.

TE echo time.

Thr threonine.

TR repetition time.

TSP 3-trimethylsilyl propionic acid.

Val valine.

WHO World Health Organization.

WM white matter.

Chapter 1

Introduction

Gliomas are primary brain tumors that are thought to be of glial origin. In adults, they account for 80% of such tumors, and result in more years of life lost than any other tumors [111]. Every year, approximately 14,000 adults in the United States are diagnosed with glioma [111] and approximately 10,000 adults die from this disease [111].

Gliomas have a complex evolution process which is characterized by a high degree of biological and clinical diversity. They include a variety of subtypes with a wide range of histopathology, molecular and genetic profiles, prognosis, and outcome. Despite major advances over the last two decades, the prognosis for patients with gliomas remains poor. Significant progress in the diagnosis, treatment, and prevention of these tumors requires both the timely implementation of new technology in the clinic and a continuing effort at increasing our understanding of brain tumor biology.

Neuroimaging plays a critical role in the diagnosis and preoperative planning of brain tumors and also serves as a means for evaluation of response to therapy. Recent advances in technology have given rise to novel data acquisition techniques that are capable of exploiting different biological properties of tissue. These methods include diffusion-weighted magnetic resonance imaging (MRI), contrast-enhanced perfusion MRI, and proton magnetic resonance spectroscopy (MRS). The characterization of brain tumors involves the integration of heterogeneous data sets that include various types of medical images, spectroscopy, clinical and histopathology data, treatment history, and patient outcome. Traditionally, these different types of biological signals are analyzed by radiologists, pathologists, and clinicians, who combine their experience in analyzing previous data with prior knowledge of the normal values of these biological signals in order to formulate a diagnosis and treatment plan. However, the increasing number of data modalities available make it likely that valuable but complex patterns within the data may remain undiscovered. The systematic analysis of such biomedical data could lead to better disease understanding, diagnosis, prognosis, treatment, and drug discovery.

The objective of this thesis was to develop multivariate pattern recognition models for integrating advanced magnetic resonance (MR) data and histology, in order to obtain algorithms that could assist in the clinical management and biological understanding of gliomas. The tools that were developed in this thesis were applied to characterize biological changes that were predictive of malignant transformations and treatment effects in gliomas, and to detect disease progression at an early stage. These tools were crucial in finding links between *in vivo* and *ex vivo* data that could give insight into the biology of brain cancer. The methods described in this thesis can contribute to clinical practice by improving the selection of biopsy sites and the targeting of treatment. The aim of this work was to produce results with high accuracy in classification, interpretability by means of clinical knowledge, and capacity to generalize the performance to new samples which are subsequently obtained. The technical aspects covered in this thesis included the feature selection and modeling of biomedical data, the inference and evaluation of predictive models, and the use of the models for clinical applications.

The rest of this document is structured as follows.

Chapter 2 describes the diagnosis and treatment of gliomas and the types of data acquisition technologies used to this effect. It also identifies major clinical problems that could benefit from the use of pattern analysis.

Chapter 3 gives a brief overview of pattern recognition methods and their application to the analysis of biomedical signals for patients with brain tumors.

Chapters 4-7 describe projects that utilize high resolution magic angle spinning (HRMAS) data for the characterization of glioma, while chapters 8-12 make use of *in vivo* MR imaging and spectroscopy.

Chapter 4 describes the use of pattern recognition methods for identifying malignant transformations in recurrent low grade gliomas (LGGs) based on HRMAS data.

Chapter 5 describes how the same pattern recognition tools can be used to distinguish between primary and secondary glioblastomas multiformae (GBMs) based on their metabolic profiles.

Chapter 6 describes the differences in the metabolic profiles of treated and untreated grade 2 and 3 gliomas.

Chapter 7 describes supervised learning models for the metabolic characterization of histopathological properties of GBM tissue.

Chapter 8 describes a machine learning model for non-invasively distinguishing between recurrent LGGs that transformed to a higher grade versus those that remained low grade.

Chapter 9 presents classification models for distinguishing between low and high values of histological parameters based on *in vivo* data.

Chapter 10 further refines the tumor cellularity prediction model by estimating the remaining histological parameters using the expectation maximization (EM) algorithm.

Chapter 11 describes a multivariate survival analysis model for GBMs. Most of the research conducted for this thesis relies on the definition of regions of interest (ROIs) in the imaging data. For this reason, Chapter 12 presents an automatic segmentation method that could aid in the definition of ROIs.

Finally, Chapter 13 provides a summary of the methods and studies used in this thesis, and describes the importance of the results.

Chapter 2

The clinical assessment of patients with gliomas

Gliomas are the most aggressive type of brain tumors, associated with a high degree of malignancy and discouraging survival rates [133]. Prognosis largely depends on the histological sub-type and grade of malignancy of the lesion, with the median survival ranging from 7 to 10 years for grade 2 tumors, 2 to 5 years for grade 3 tumors, and 15 to 18 months for grade 4 tumors [8].

Neuroimaging plays a crucial role in the diagnosis, therapy planning, evaluation of therapeutic effects, and early detection of recurrence of brain tumors. It is currently evolving from a morphology-driven discipline to incorporate biochemistry and physiologic parameters into the assessment of brain tumors. Although clinical and neuroimaging features of malignant gliomas are very useful in identifying abnormalities in the brain, the gold standard for diagnosis is based on the histologic examination of appropriately sampled tissue. Image-guided tissue biopsies are of interest to ensuring the most accurate diagnosis by reducing sampling errors.

2.1 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) is a complex modality which is being increasingly used in the diagnosis, characterization, and treatment of brain tumors. MR imaging of the protons in water is a powerful technique that allows the visualization of internal anatomy in a safe and noninvasive manner. MRI is used for evaluating brain tumors for several reasons: it provides good soft tissue contrast, it can perform multi-planar imaging, and it uses no ionizing radiation. MRI has been expanded in recent years to include special techniques such as diffusion-weighted imaging (DWI), perfusion-weighted imaging (PWI), and *in vivo* and *ex vivo* spectroscopy [7]. The contrast mechanism used to generate these data relies upon different physical and physiological properties, and provides complementary information that can be used in the analysis of brain tumors and surrounding tissue.

MR signals come from protons or other nuclei inside our body. In the absence of a magnetic field, these nuclei have a property known as spin, and the orientation of the spin vector is random

[11]. When the subject is placed inside a strong magnetic field, the spins of the nuclei align either parallel or anti-parallel with the field. Then a radio frequency pulse (RF pulse) is introduced, which excites these spinning protons, causing them to move out of alignment with the magnetic field. When the RF excitation stops, the nuclei return to their resting state within the magnetic field, giving off the RF energy acquired during the RF pulsation [59]. The signal is spatially localized by the rapid turning on and off of spatially varying magnetic field gradients [59]. The signal is detected by coils within the scanner and used to generate an image. The behavior of each nucleus reflects its chemical environment. The differences in the behavior of the nuclei result in different contrast mechanisms [7]. As the water content of tissue is on the order of 80%, the most abundant and widely used MR signals come from the protons in water.

There are thus three key components of an MRI scanner: a very strong magnet that creates the main magnetic field, which creates the parallel and anti-parallel energy levels; a set of gradient coils, which impose a gradient on that magnetic field, which provides the spatial encoding; and radio-frequency coils that either transmit the radio frequency pulses to excite the spin sample or detect the resulting signal [7].

T1 and T2 relaxation

MR signal is dependent on the number of mobile hydrogen protons, the time needed for the protons within the tissue to return to their original state of magnetization (T1 relaxation time), and the time required for the protons perturbed into coherent oscillation to lose their coherence (T2 relaxation time) [59]. The timing of the RF pulse and the reading of the radiated RF energy change the appearance of T1 and T2 images. The repetition time (TR) describes the time between successive repetitions of the RF pulse sequences [59]. The echo time (TE) describes the delay before the RF energy radiated by the tissue is measured [59]. The pulse sequences, described by the TR and TE, can be changed to maximize the differences in T1 and T2. This gives rise to T1 and T2 weighted images [59].

The standard MRI pulse sequence for anatomic and pathological detail is a spin echo sequence. The T1-weighted images (short TR; short TE) provide better anatomic detail, while the T2-weighted images (long TR; long TE) provide more detail about the pathology [59].

Fluid-attenuated inversion recovery

The Fluid-attenduated inversion recovery (FLAIR) uses a pulse sequence that removes the effects of fluid from the resulting image. Lesions that are normally covered by bright fluid signals using conventional T2-weighted images are made visible by this darkening of the fluid in this technique, making FLAIR very important for the differentiation of brain and spine lesions. Figure 2.1 shows a FLAIR image acquired from a patient diagnosed with GBM.

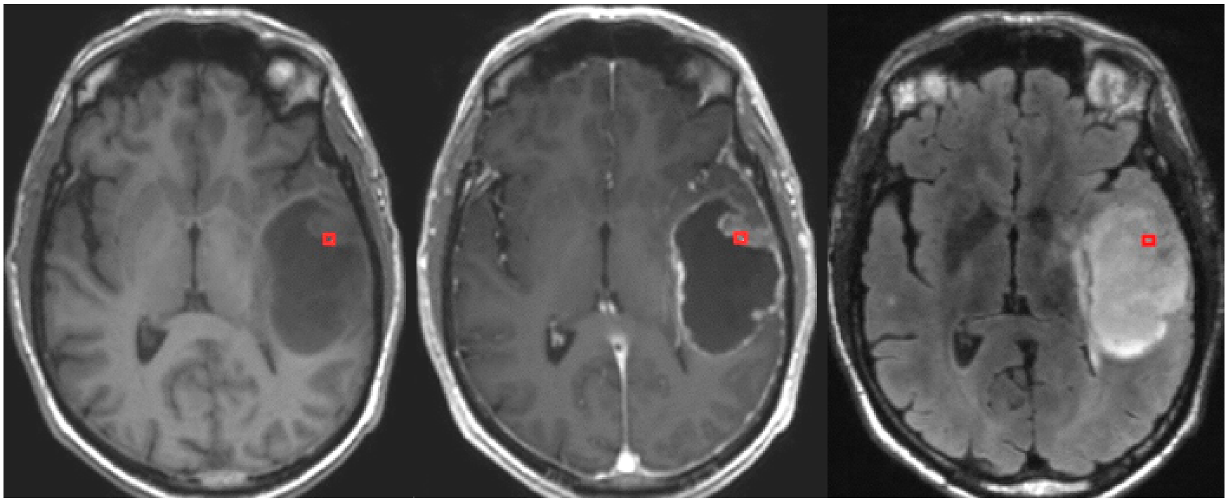


Figure 2.1: Anatomical MR images obtained from a patient diagnosed with GBM; (Left): a pre-contrast T1 image; (Middle): a post-contrast T1 image (T1-GAD); (Right): a FLAIR image; the region marked with red was later selected as a biopsy location

Contrast enhanced images

Neuroimaging techniques are designed to highlight differences between lesions and normal structures. For some lesions, mass effect is sufficient for diagnosis. For others, contrast agents have to be used to aid in diagnosis. Contrast agents are compounds that do not cross the intact blood-brain barrier but do cross abnormally permeable blood-brain barrier regions [7].

Gadolinium (Gd) is one of the primary contrast agents used in MRI. It causes marked signal change whenever there is breakdown of the blood-brain barrier, allowing high specificity tumor diagnosis based on imaging alone [7]. Administering gadolinium allows for the acquisition of two types of T1-weighted images: pre-contrast T1 and post-contrast T1 (T1-GAD). Figure 2.1 illustrates a pre-contrast and a post-contrast T1 image of a patient diagnosed with GBM. The region marked with red in this figure was later selected as a biopsy location. This region is highlighted in all of the imaging modalities presented in this chapter.

While a FLAIR image highlights the extent of abnormal tissue, a T1-weighted sequence obtained after the injection of gadolinium contrast agent (T1-GAD) is believed to highlight the most aggressive areas of the lesion, as well as necrotic tissue. Three ROIs can be defined based on these images: the FLAIR abnormality region (T2all), the contrast enhancing lesion (CEL), and the necrotic region (NEC), which are illustrated in Figure 2.2. The non-enhancing lesion (NEL) can also be defined as the difference between the T2all and the CEL and NEC regions.

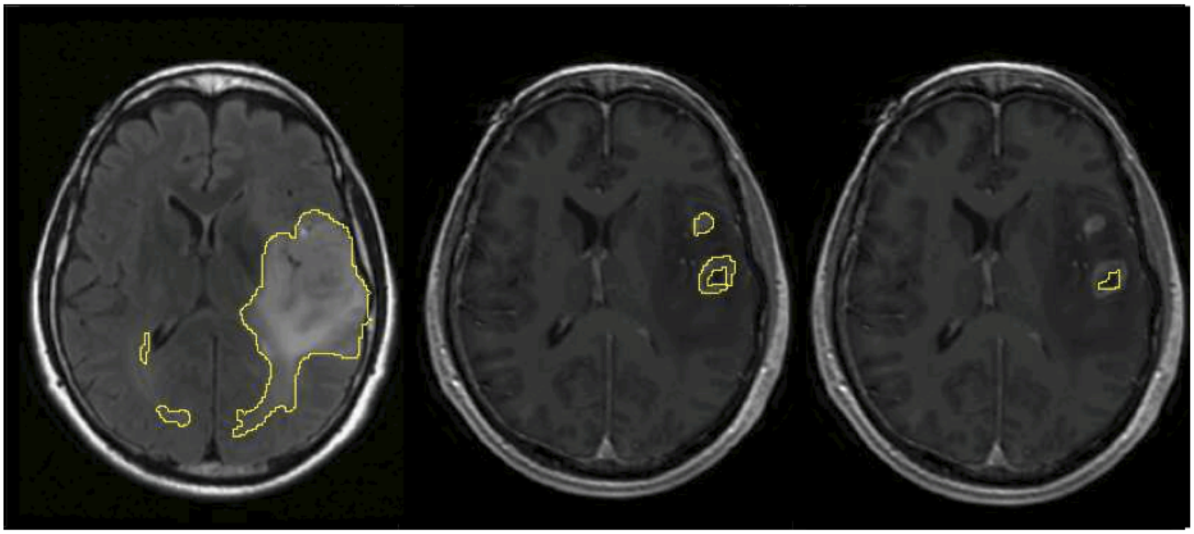


Figure 2.2: Regions of interest: FLAIR abnormality regions (T2all), contrast enhancing lesion (CEL), and necrotic region (NEC)

Diffusion MRI

Diffusion MRI allows the non-invasive evaluation of the diffusion of water molecules. Molecular diffusion in tissues reflects interactions of the molecules with obstacles such as macromolecules, fibers, and membranes. As a result this method can be used to reveal microscopic details about tissue architecture. DWI inserts an additional pair of gradient pulses into the imaging pulse sequence that measures the reduction in signal intensity caused by the motion of individual water protons. This results in a greater loss of signal intensity for protons with high diffusion rates compared with protons that have slow diffusion rates. It is possible to disentangle diffusion and relaxation effects by creating two types of maps: apparent diffusion coefficient (ADC) and fractional anisotropy (FA). FA describes the degree of anisotropy at each voxel, on a scale of zero for isotropic to one for fully anisotropic [7]. A low value for FA indicates that the cortical white matter (WM) tracts are isotropic, while a high value for FA indicates they are anisotropic with a preferred direction for diffusion [7]. ADC maps measure the magnitude of diffusion of water molecules within cerebral tissue. Figure 2.3 shows ADC and FA maps for the GBM patient used as an example throughout this chapter.

Perfusion MRI

The changes in signal intensity that are observed in T2-weighted MR images during the passage of a bolus of gadolinium through the microvasculature may be used to provide insight into cerebral blood volume, tissue microvasculature, angiogenesis, and vessel permeability [66]. The advantages

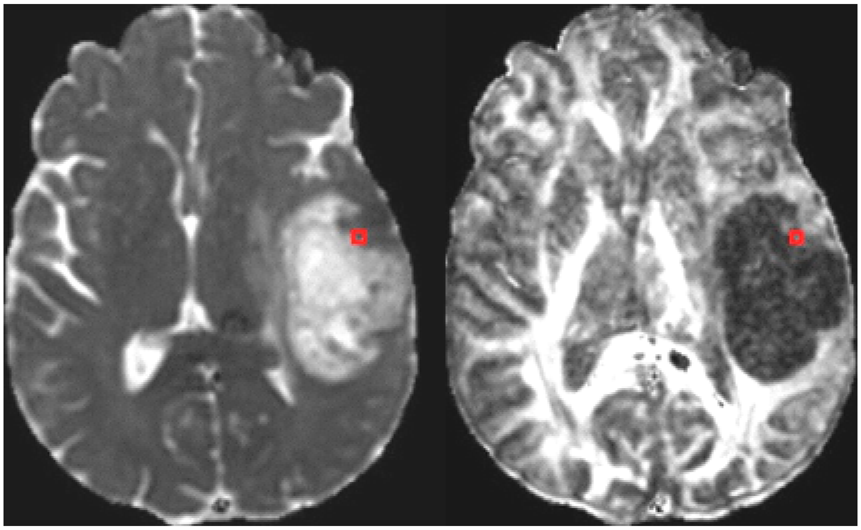


Figure 2.3: Diffusion maps obtained from a patient diagnosed with GBM; (Left): ADC map; (Right): FA map; the region marked with red corresponds to a biopsy site

of this approach are that it is relatively easy to implement as part of a conventional MR examination because patients with brain tumors are typically getting an injection of gadolinium for diagnostic purposes. Quantification of the changes in signal intensity uses the approximately linear relationship between concentration of the contrast agent and the changes in relaxivity to transform the signal-intensity curve to a change in the relaxation rate (ΔR2*) curve which increases during the bolus, followed by a recovery period [66]. Four parameters can be derived from this curve. The parameter most widely reported is the cerebral blood volume (CBV). This relates the area under the curve during the bolus from an ROI to the area obtained from a region of normal-appearing white matter (NAWM). Other parameters include the peak height (PH) in the ΔR2* curve, the percent recovery to baseline (RECOV), which reflects the leakiness of the tumor induced vessels, and the recirculation factor (RF) [66]. Figure 2.4 shows a T2-weighted image during the passage of a bolus of gadolinium. The area marked with red corresponds to a site where tissue was acquired. The figure also shows a ΔR2* curve for that site.

Spectroscopy

MRS uses the differences in chemical shift caused by alterations in the chemical environment of individual nuclei to evaluate the presence and levels of different compounds. It is based on the concept that certain nuclei have inherent spin properties. Examples include ^{1}H, ^{19}F, ^{31}P, and ^{13}C. These nuclei acquire small amounts of energy while in the presence of a static magnetic field. However, the introduction of electromagnetic radiation of the correct frequency, at right angles to the static magnetic field, causes the nuclei to jump to spin states of higher energy levels [59]. The nuclei subsequently drop to their ground spin states by emitting radiomagnetic radiation at a rate

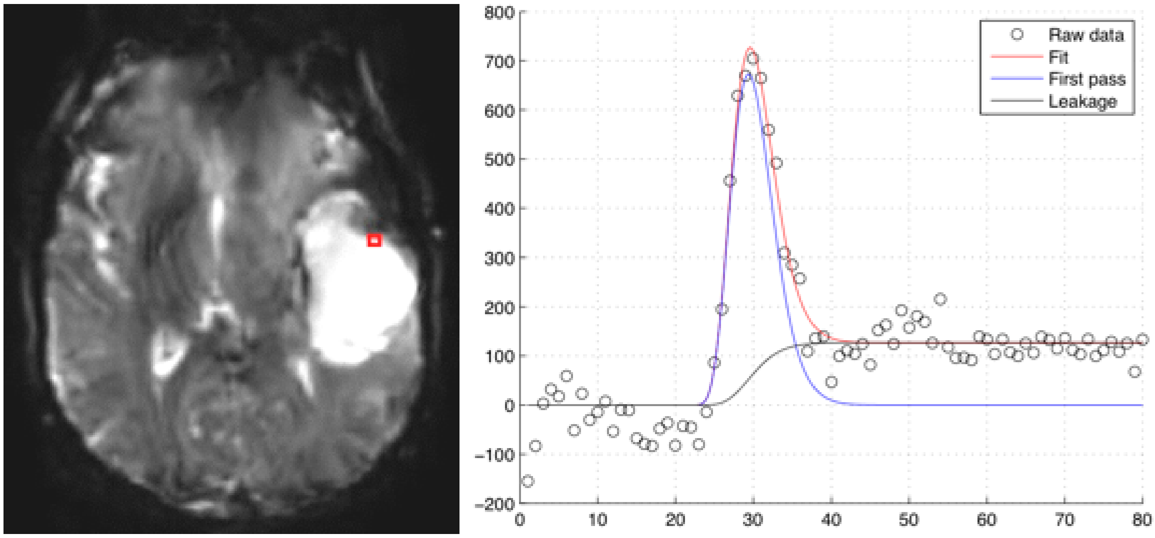


Figure 2.4: Perfusion imaging data obtained from a patient diagnosed with GBM; (Left): a T2-weighted image of a GBM patient during the passage of a bolus of contrast agent; the area marked with red corresponds to a biopsy site; (Right): $\Delta R2*$ curves at the biopsy site

determined by the T1 and T2 relaxation times. The emitted radiation is detected by a receiver coil, and the range of frequencies and their intensities comprise the MR spectrum.

Recent oncology research has shown that the evaluation of cellular metabolism can be very helpful for the diagnosis and assessment of treatment effects for patients with brain tumors. There is a growing body of evidence that MRS may contribute to the clinical evaluation of a number of pathologies and therapeutically induced changes in tumor biochemistry [2, 3, 9, 15, 23, 26, 33, 38, 41, 44, 45, 54, 63, 70, 77, 86–88, 97, 102, 104, 106–108, 117, 119, 129, 132]. It can be performed both *in vivo* and *ex vivo*.

In Vivo MRS

In vivo MRS uses similar spatial localization principles as MRI to measure the levels of different metabolites within regions of tissue. Arrays of spectra make it possible to both observe the heterogeneity within the lesion and also to examine surrounding tissue that may appear normal on MRI [87]. This provides a reference for comparing metabolite levels in the tumor and makes it possible to identify regions of abnormal metabolism outside the morphological lesion [87]. Figure 2.5 shows an example of an array of MRS data obtained from a GBM patient. Saturation bands were used to diminish the distortions produced by fat tissue. Figure 2.6 shows a lactate-edited MRS data array from a similar position. Both figures include the location of a future biopsy site. The MRS data from this biopsy site is compared to *ex vivo* HRMAS spectroscopy in Figure 2.7.

The dominant compounds observed using the *in vivo* MRS technique are levels of choline

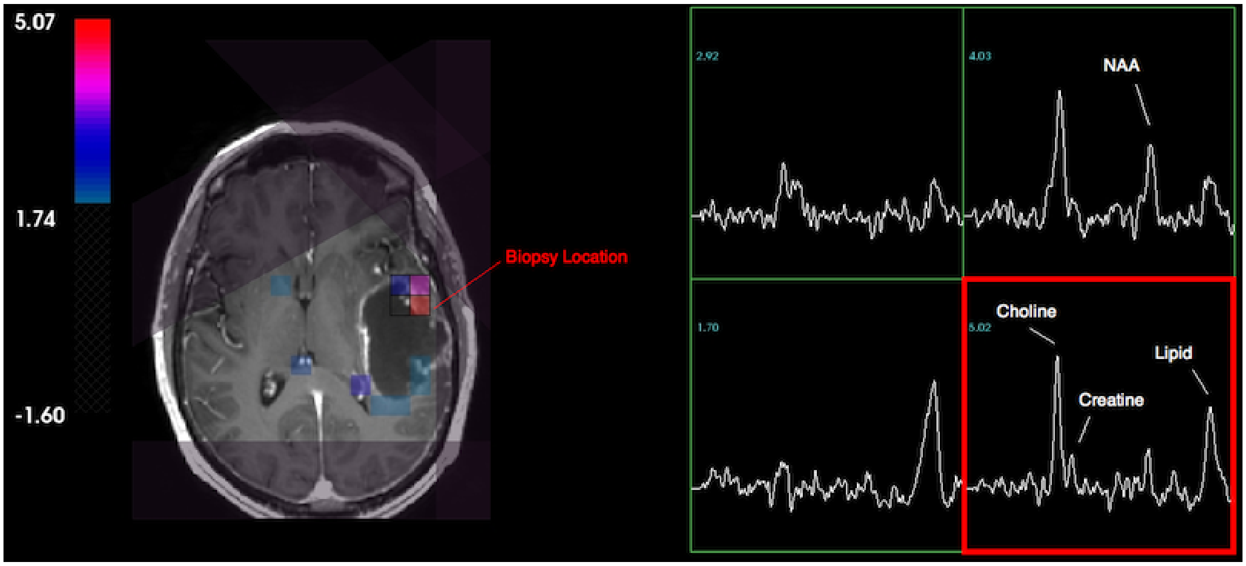


Figure 2.5: *In vivo* magnetic resonance spectroscopy data obtained from a patient diagnosed with GBM; saturation bands were used to diminish distortions due to fat tissue; the colored voxels overlayed on the T1-GAD image are abnormal; the overlay corresponds to the choline to NAA ratio (CNI); the voxel labeled in red was later selected as a biopsy location

(Cho), creatine (Cre), N-acetylaspartate (NAA), lactate (Lac), and lipid (Lip). The intensity of the Cho signal reflects changes in membrane synthesis and turnover associated with cell proliferation and remodeling. The Cre peak includes both creatine and phosphocreatine and is indicative of cellular bioenergetic processes. NAA is a marker of normal brain tissue that is typically assumed to correspond to the presence of actively functioning neurons. Lac is an end product of anaerobic metabolism and may therefore reflect ischemia, hypoxia, or both. Lip that is observed within the brain tissue is thought to reflect the presence of necrosis [87].

By comparing the relative concentration of metabolites from MRS data, it is possible to draw inferences about biological features such as neuronal viability, the presence of neurotoxins, and the process of membrane turnover within the volume of interest [54]. These can be used to formulate hypotheses about the likely underlying pathology [30, 54, 131]. The clinical impact of MRS in medicine was reviewed by Hollingworth et al. [54], Nelson [86, 87], and Smith et al. [117]. MRS shows promise as a method to complement routine diagnostic investigation. Significant research over the last decade has investigated the role of MRS imaging biomarkers in characterizing brain cancer [84, 131]. The combination of MRS with high-resolution anatomical images provided by conventional MRI provides unique information on brain tumor chemistry in inoperable tumors, and might complement neuropathology, guide or even replace biopsies, and help monitor therapy for operable brain tumors.

While a great tool for characterizing brain tumor metabolism noninvasively, *in vivo* spectroscopy suffers from some internal and external limitations that may strongly influence the re-

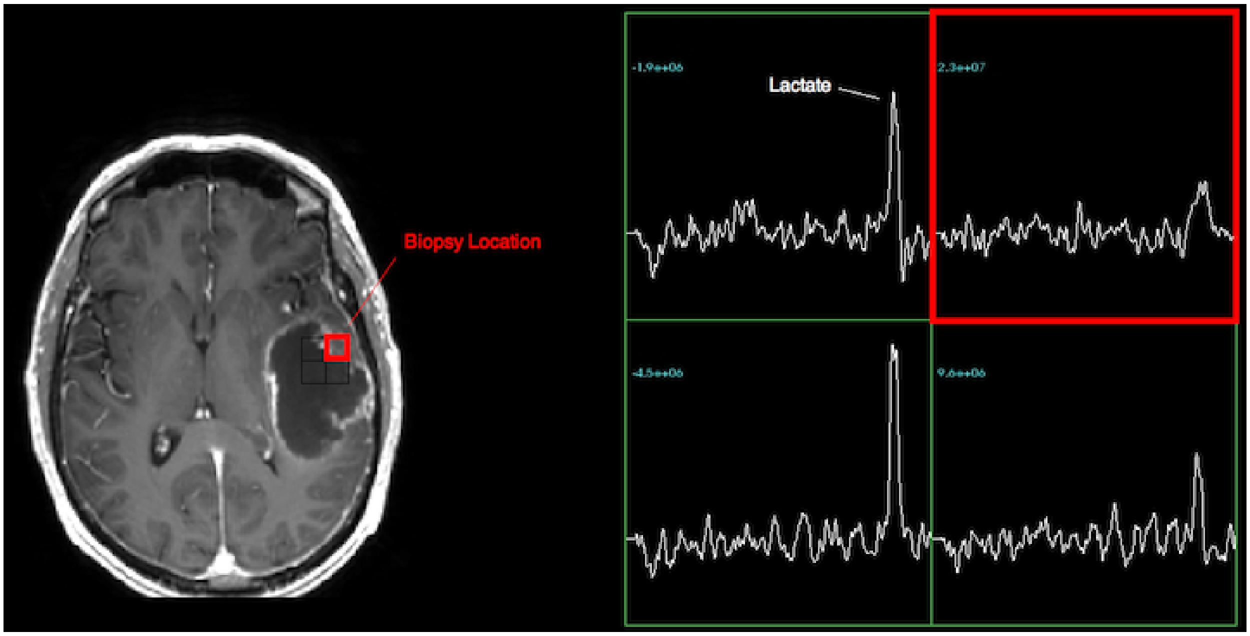


Figure 2.6: *In vivo* lactate-edited MRS data obtained from a patient diagnosed with GBM; the biopsy site marked with red is the same site marked in Figure 2.5

sults. In particular the limited *in vivo* spectral resolution restricts the number of metabolites that can be evaluated. Even among metabolites that are detectable with *in vivo* MRS, subtle differences between normal and pathological tissue can go unnoticed.

HRMAS spectroscopy

The development of HRMAS NMR spectroscopy and its recent application to *ex vivo* tissue samples made it possible to examine the biochemical composition of tissue by spectroscopic analysis [15]. HRMAS is a non-destructive technique for obtaining spectra with narrow peaks, improved signal to noise ratio, and higher spectral resolution over standard methods. Figure 2.7 shows an HRMAS spectrum and the corresponding MRS spectrum obtained from a patient with a GBM. Figure 4.3 better illustrates the higher resolution of the same HRMAS spectrum, and the much higher number of metabolites that it reveals.

The development of HRMAS is based on the observation that the spectral broadening that typically occurs in tissue has an angular dependence that can be overcome by placing the sample at a so-called magic angle with respect to the main magnetic field of the magnet. By spinning the sample, one can mimic the high molecular mobility of liquids that results in a well-averaged MR signal. An added bonus of HRMAS is that it can be used on unprocessed biopsy or surgical specimens, allowing much of the structure and biochemistry of the specimen to remain intact. Because HRMAS is a non-destructive technique, it can be applied to tissue prior to immunohistochemistry

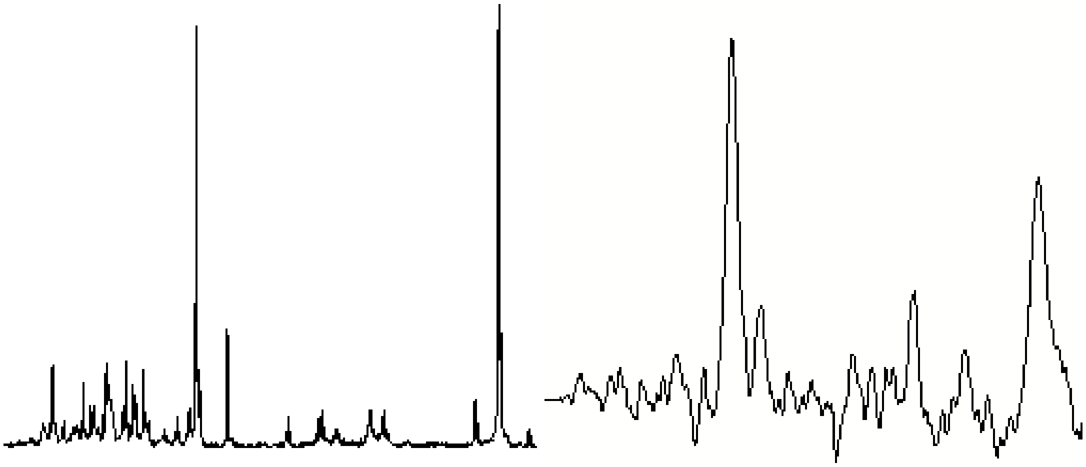


Figure 2.7: HRMAS and corresponding MRS obtained from a patient diagnosed with GBM; (Left): HRMAS spectrum obtained from a GBM patient's tissue sample; (Right): MR spectrum obtained from the biopsy location of the same GBM patient; HMRAS has narrower peaks and higher spectral resolution than MRS

(IHC) analysis, thus providing a link between metabolite concentrations, pathology, and clinical MRI and MRS. To this end, a large body of research examines the identification of biomarkers for brain tumor characterization and typing using HRMAS [46, 75, 88, 91, 104, 135]. *Ex vivo* magnetic resonance spectroscopy has also been used to link *ex vivo* metabolic profiles to *in vivo* MRS and hence to determine which metabolites are of interest for characterizing different pathologies [106–108].

Major metabolites in the brain

Choline (Cho) is a building block for cell membrane chemicals. Changes in Cho levels reflect changes in membrane synthesis and turnover associated with proliferation and remodeling [87]. The importance of this metabolite is highlighted by the fact that Cho deficiencies affect key genes in cell proliferation, differentiation, and apoptosis. Cho has a functional group that has a very strong magnetic signal, with nine equivalent protons, and so even at lower concentrations, it is visible in spectroscopic studies. It can be converted to and between various forms including **Glycerophosphocholine (GPC)** and **Phosphocholine (PC)**, which are also highly visible in proton spectroscopy. When cells divide quickly or there is substantial remodeling of membrane components, the Cho-containing compounds in solution increase. Thus, many brain tumors have elevated Cho peaks, thought to be associated with their increased cellularity and compression of surrounding brain tissue.

Recent studies on LGGs have established a link between mutations in the isocitrate dehydrogenase 1 (IDH1) gene and excessive production of **2-hydroxyglutarate (2HG)** [23, 32]. This is an important finding because overwhelming clinical evidence now suggests that glioma patients harboring IDH1 mutations carry a significant survival advantage, irrespective of treatment therapies employed for disease management [49, 56, 92, 137]. Although the function of both IDH-mutant enzymes and 2HG in gliomagenesis remains unclear, the improved prognosis associated with IDH1 mutations suggests that the presence of 2HG may be of prognostic value as a surrogate for favorable genotypes.

Myoinositol (Myo-I) is involved primarily in hormone-sensitive neuroreception. It is found mainly in astrocytes and helps to regulate cell volume. Elevated levels of Myo-I are seen in glial cell proliferation, such as gliosis. Reduced levels of Myo-I are seen in processes that cause glial cell destruction, such as tumor formation.

Lactate (Lac) levels in the brain are normally very low or absent. When oxygen supply is depleted, the brain switches to anaerobic respiration for which one end product is lactate. Therefore, elevated lactate levels are a sign of hypoxic tissue. Low oxygen supply can result from decreased oxygen supply or increased oxygen requirement. The former may be seen in vascular insults, or hypoventilation, and the latter may be seen in neoplastic tissue.

Lipids (Lips) are often present in the form of triglycerides, phospholipids, and fatty acids. These substances are incorporated into cell membranes and myelin. Within the brain, Lip peaks are only seen in the presence of destructive processes, such as formation of necrosis, inflammation, or infection [87].

N-acetylaspartate (NAA) is a marker of normal brain tissue that is typically assumed to correspond to the presence of actively functioning neurons. It is present in both gray and white matter. Its concentration has been reported to decrease with brain infection, ischemic injury, neoplasia, and demyelination.

Creatine (Cre) is indicative of cellular bioenergetic processes. The Cre peak receives contributions from both Cre and **phosphocreatine (PCr)**. PCr supplies phosphate to adenosine diphosphate (ADP) to form adenosine triphosphate (ATP) with the release of creatine. The overall level of Cre in normal brain is fairly constant. Reduced Cre levels may be seen in pathologies such as neoplasm, ischemic injury, infection, or some systemic diseases.

Glutamine and glutamate (Glx) resonate closely together. **Glutamate (Glu)** is an excitatory neurotransmitter in mitochondrial metabolism, and it is also a key component of the inhibitory compound neurotransmitter **g-aminobutyric acid (GABA)**. **Glutamine (Gln)** is a precursor to neuronal Glu and a potential fuel, which, like glucose, supports neuronal energy requirements.

Taurine (Tau) has been suggested to act as inhibitory neurotransmitter or modulator in the central nervous system. The immediate precursor of Tau, **Hypo-taurine (Hyp-Tau)**, has also been thought to have a synaptic role of its own or to act as a false neurotransmitter for GABA or Tau. Both amino acids depress neuronal firing with similar potency.

Reduced **Glutathione (GSH)** levels play an important role in the protection of cells against damage from free radicals and other electrophils and also influence cellular radiosensitivity, cellular response to hyperthermia, and cytotoxicity to some kinds of chemotherapeutic agents.

Alanine (Ala) is a non-essential amino acid with a non-reactive side group. It is the second most highly represented amino acid, to **leucine (Leu)**, in the primary structure of a sample of 1000 key proteins. Also, while it is not a reactive residue in proteins, it can play an important role in protein substrate recognition.

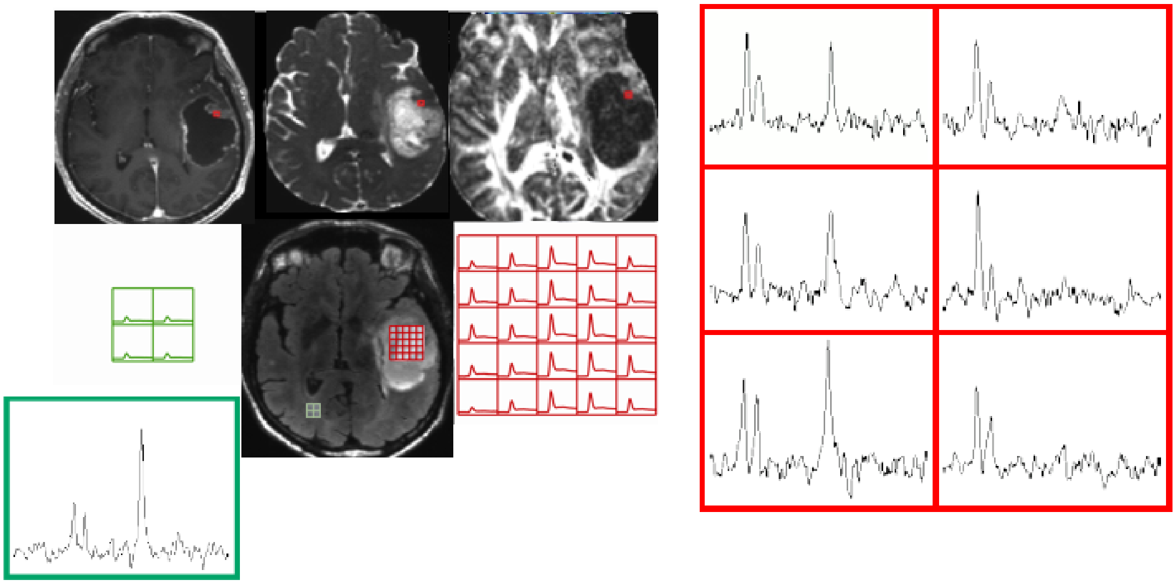


Figure 2.8: Multivariate imaging data and examples of normal and abnormal perfusion and spectroscopy

Under normal physiological conditions, the mature brain derives almost all of its energy from the aerobic oxidation of **glucose (Glc)**. Brain tumors, like most malignant tumors, depend heavily on Glc for their metabolic energy.

2.2 Histopathology of gliomas

While MRI is an important step in evaluating patients with suspected brain tumors, histopathological analysis of tissue samples from the lesion is always required to confirm a suspected diagnosis. Histology refers to the study of the microscopic anatomy of cells and tissue. A pre-operative MR exam, combining multiple images such as the ones illustrated in Figure 2.8, is used to guide the selection of biopsy sites by identifying regions suspected of containing viable tumor. Tissue samples may be obtained as part of a stereotactic procedure or during the tumor resection. A stereotactic biopsy procedure can be performed by using a computer to localize a target lesion on a three-dimensional image, as illustrated in Figure 2.9, and then extracting the tissue from that location. The tissue samples are fixed and then slices are prepared for examination under the microscope by a pathologist. The ability to identify microscopic structures is enhanced through the use of histological stains, which highlight features of interest. Figures 2.10, 2.11, 2.12, 2.13, and 2.14 show low and high values of some of these histological features of interest.

The histological analysis of brain tumor tissue identifies parameters describing the number of cancer cells and the invasive capacity of the tumor [72]. A short description of the parameters used

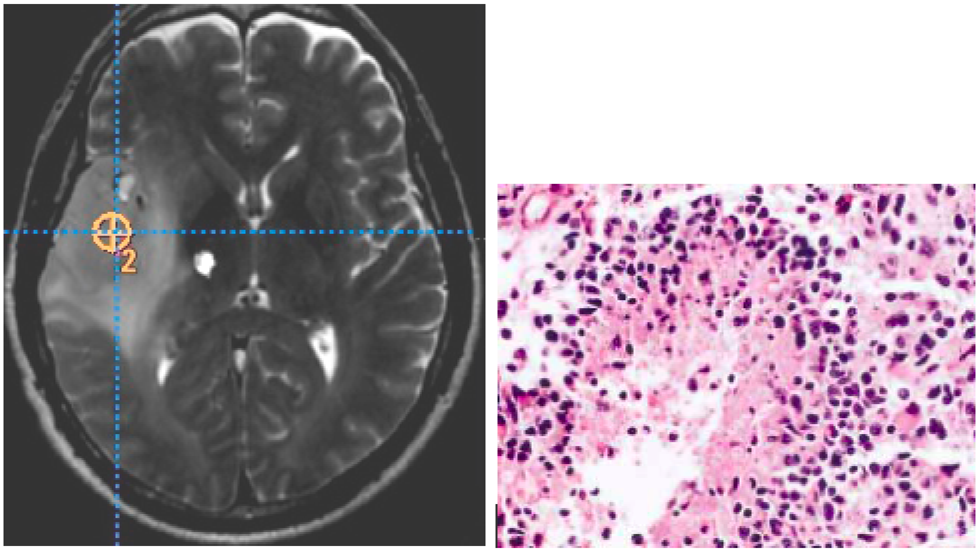


Figure 2.9: Example of biopsy site selection and corresponding histological appearance

for the samples evaluated in the studies described in this thesis is provided next.

Tumor cellularity

Tissue slides stained with H&E were analyzed for tumor cellularity. This is a measure of the density of cancer cells in a specimen. Slides were given a tumor cellularity score on the basis of the contribution of tumor cellularity to total cellularity. A score of 0 denoted neuropil without tumor, 1 indicated an infiltrating tumor margin containing detectable but not abundant numbers of tumor cells, 2 denoted a more cellular-infiltrated zone, and 3 denoted highly cellular tumor with few non-neoplastic cells. Figure 2.10 illustrates slides showing low and high tumor cellularity.

MIB-1 index

Cell cycle-associated nuclear proteins were utilized to estimate the growth potential of gliomas. The Ki-67 monoclonal antibody recognized epitopes associated with a specific category of cell cycle-associated proteins which appear to be necessary for the maintenance of the proliferative state [72]. The monoclonal antibody MIB-1 reacts with the Ki-67 epitope in routinely processed, paraffin-embedded tissue. For MIB-1-stained slides, a labeling index [(MIB-1-positive nuclei per total tumor cells counted per $200\times$ field)$\times$ 100%] was calculated based on the evaluation of at least three fields and more than 1000 cells. Slides showing low and high MIB-1 proliferation indices are illustrated in Figure 2.11.

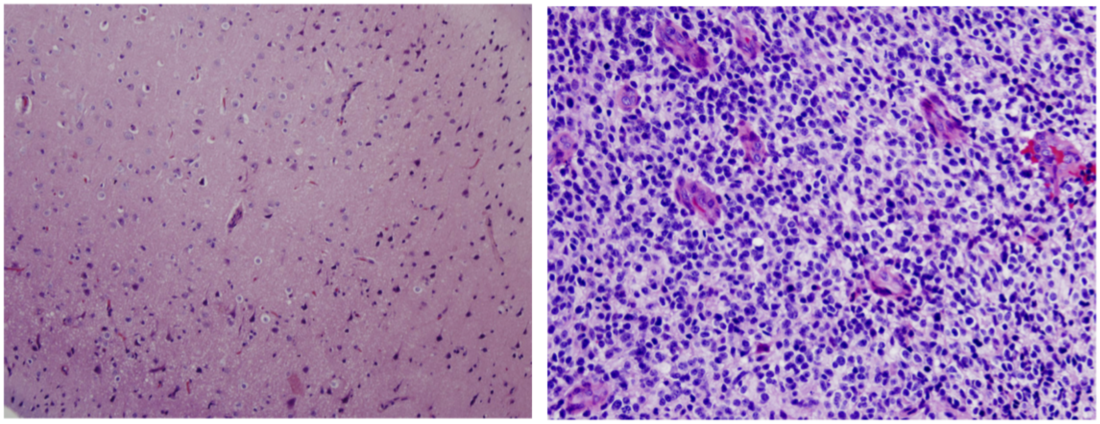


Figure 2.10: Two H&E stained slides with: (Left) a tumor score of 0, cellularity of 85 cells/field, and (Right) a tumor score of 3, cellularity of 921 cells/field (courtesy of Dr. Joanna Phillips)

SMI-31

Axonal integrity was measured based on a composite axon score for SMI-31 staining [72]. A score of axonal integrity was assigned to each SMI-31 stained slide on the basis of the extent of disruption of the normal axonal architecture. A score of 0 denoted no disruption, 1 denoted minimal disruption, 2 denoted mild disruption, and 3 denoted severe disruption. SMI-31 scores are illustrated in Figure 2.12.

CA9

The CA9 gene expression reflects hypoxia [72], a deficiency in the amount of oxygen that reaches the tissue. A hypoxia score of 0 denoted no positive CA9 staining, a score of 1 denoted that 10% of the tissue area was stained positive for CA9, a score of 2 denoted that 10-25% of the tissue area was stained positive for CA9, a score of 3 denoted that more than 25% of the tissue area was stained positive for CA9 . Illustrations of CA9 staining are provided in Figure 2.13.

Microvascular morphology

GBMs frequently present with regions of abnormal angiogenesis, which is the outgrowth of new blood vessels from preexisting vasculature [72]. Hyperplasia, the proliferation of cells which is not neoplastic, may signal the body's normal reaction to an imbalance or another stimulus [72]. On the basis of H&E staining and Factor VIII IHC, the microvascular morphology was graded as delicate (resembling normal cerebral vessels), simple vascular hyperplasia (circumferential hyperplasia with definitive lumen), or complex microvascular hyperplasia (glomeruloid-type vessels). These vascular parameters were measured on a four-tier ordinal scale (0, no contribution; 1, minimal; 2, prevalent; and 3, extensive). An overall hyperplasia score was assigned as: 0, signifying

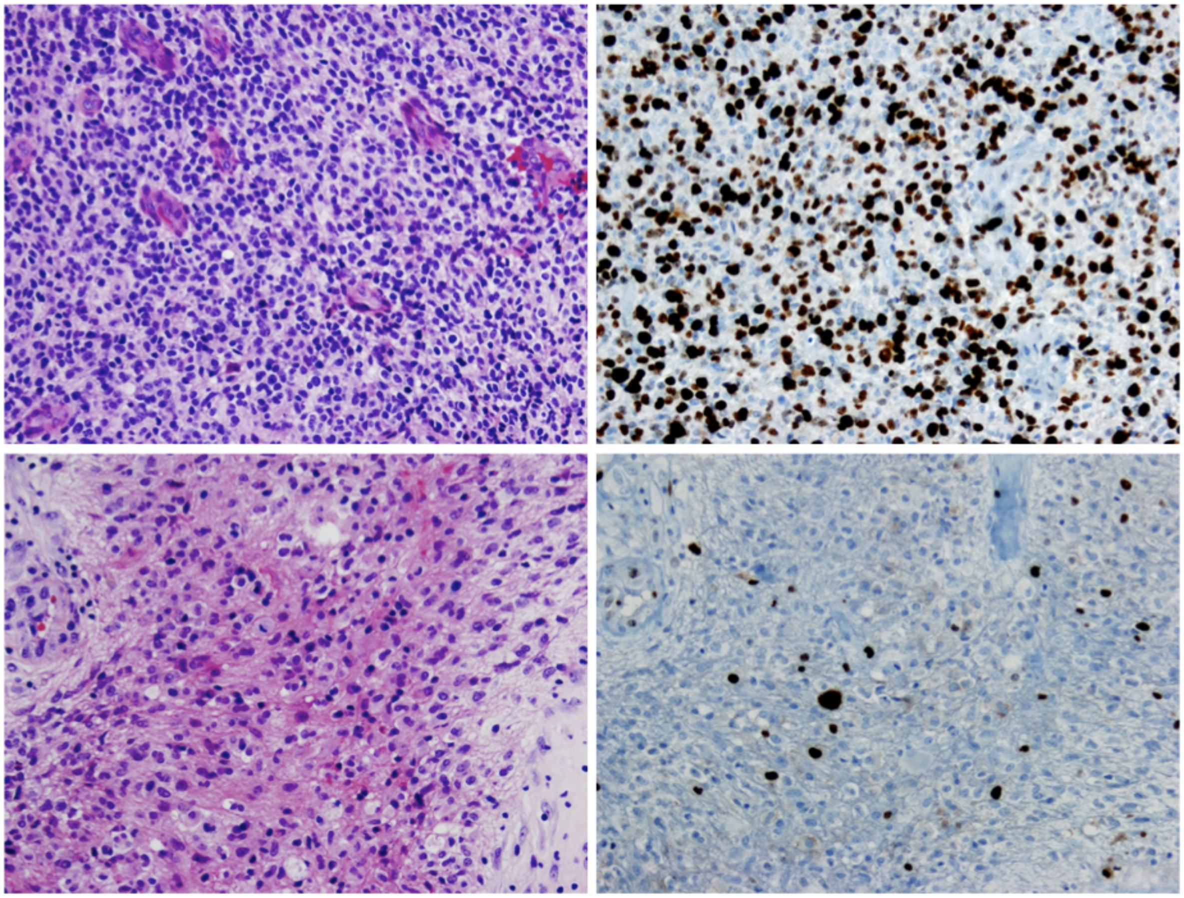


Figure 2.11: (Left): Two H&E stained slides and (Right): the corresponding MIB-1 stained slides; (Top): tumor score 3, cellularity 921 cells/field, MIB-1 of 35.5%; (Bottom): tumor score 3, cellularity 27.4 cells/field, MIB-1 of 7.89% (courtesy of Dr. Joanna Phillips)

normal vasculature, 1 signifying simple hyperplasia, and 2 signifying complex hyperplasia. Examples of the vascular properties that can be observed in gliomas are illustrated in Figure 2.14.

Necrosis

GBMs are often made up of a necrotic center, with a surrounding rim of hypercellular tissue. Figure 2.15 shows an H&E stained tissue sample containing necrosis. A necrosis score of 0 signified absent, 1 signified focal necrosis, and 2 signified extensive necrosis.

Figure 2.12: SMI-31 staining; (Left): SMI-31 score 0, no axonal disruption; (Middle): SMI-31 score 2, mild axonal disruption; (Right): SMI-31 score 3, severe axonal disruption (courtesy of Dr. Joanna Phillips)

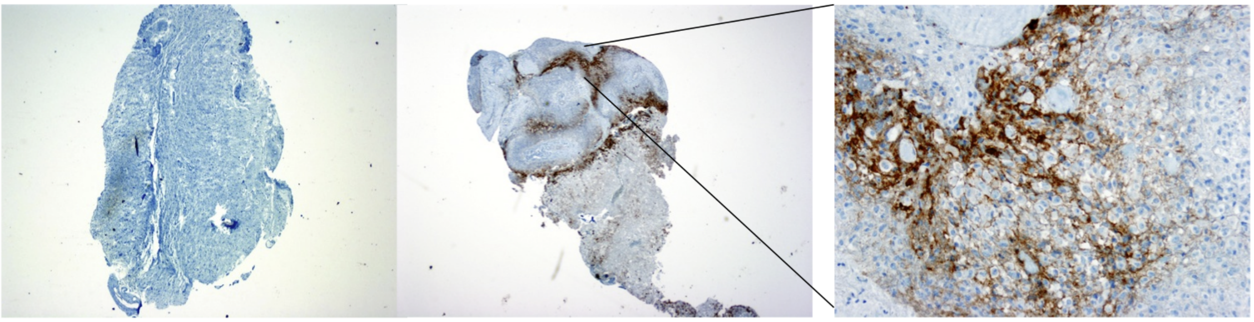


Figure 2.13: CA9 staining (courtesy of Dr. Joanna Phillips)

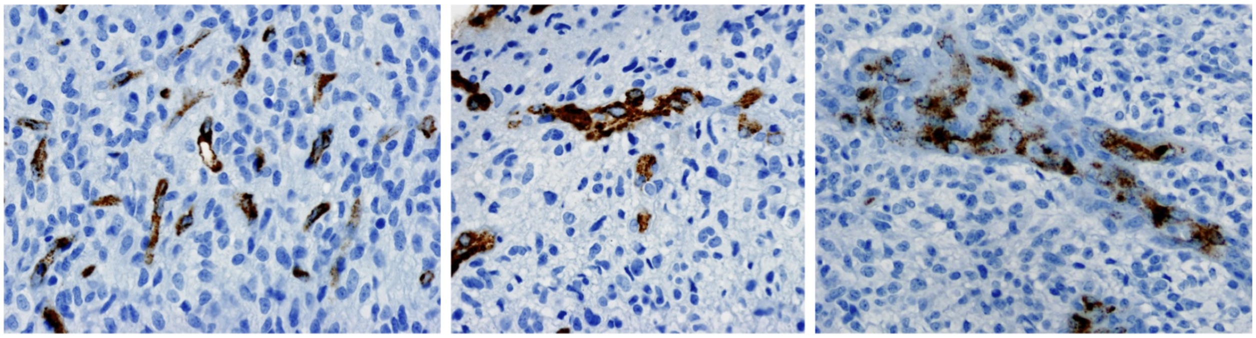


Figure 2.14: Factor VIII microvascular pattern and density; overall microvascular hyperplasia score: (Left) score 0, delicate hyperplasia, (Middle) score 1, simple hyperplasia, (Right) score 2, complex hyperplasia (courtesy of Dr. Joanna Phillips)

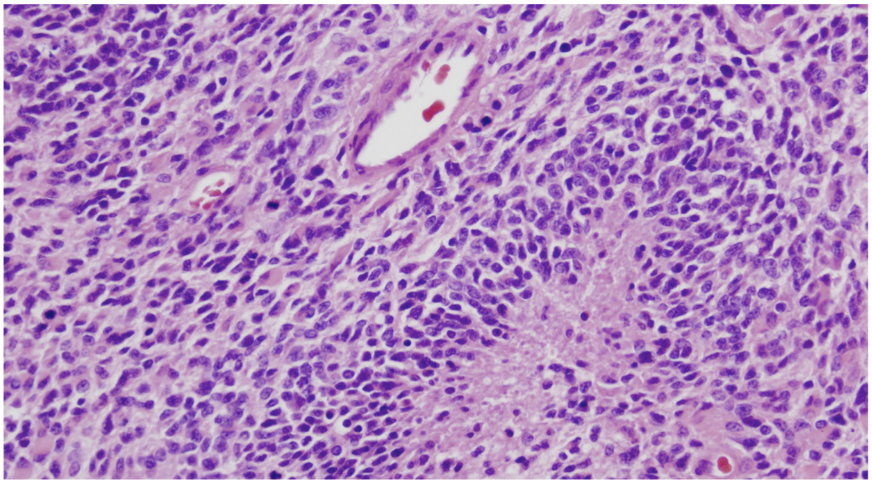


Figure 2.15: H&E slide showing necrosis and vascular hyperplasia (courtesy of Dr. Joanna Phillips)

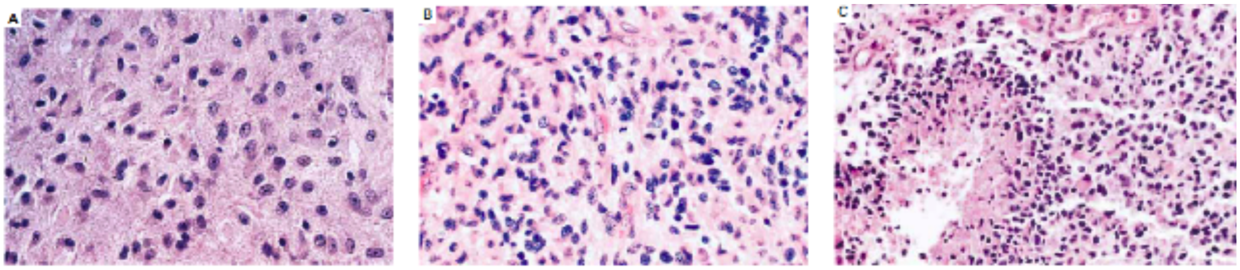
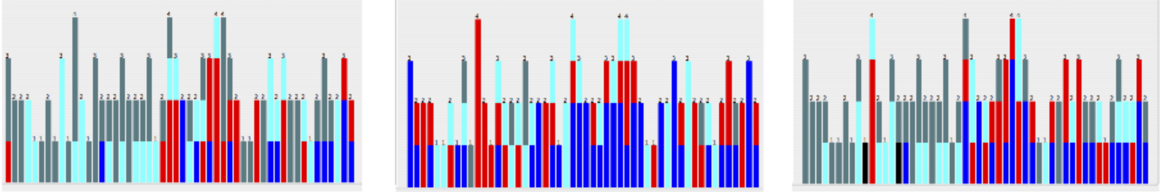


Figure 2.16: Histological appearance of tissue samples from (A) grade 2, (B) grade 3, (C) grade 4 gliomas

WHO classification

The World Health Organization (WHO) classifies gliomas on the basis of histologic features into four prognostic grades: grade 1 2, 3, and 4. Examples of histological stains from grade 2, 3, and 4 gliomas are illustrated in Figure 2.16. Grade 4 gliomas are also referred to as GBMs [80]. Grade 3 tumors are characterized by increased cellularity, nuclear atypia, and mitotic activity [133]. In addition, GBMs contain areas of microvascular proliferation, necrosis, or both [133]. These tumors may contain a delicate network of branching blood vessels [133]. GBMs can be separated into two main subtypes, primary and secondary [8]. Primary GBMs present themselves without clinical or histopathologic evidence of a preexisting less-malignant precursor lesion. Secondary GBMs typically develop in younger patients through malignant progression from a lower grade glioma.

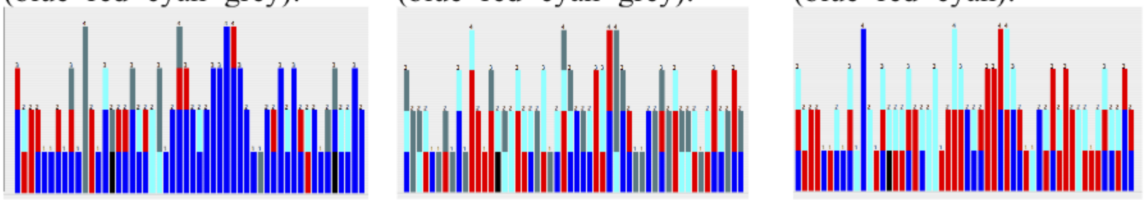
Gliomas are very heterogeneous and invasive, as illustrated in Figure 2.17. They typically contain both neoplastic and stromal tissue, which contribute to their histologic heterogeneity and



Tumor Score (blue<red<cyan<grey):

MIB-1 (blue<cyan<red<grey):

SMI31 (blue<red<cyan<grey):



Complex (blue<red<cyan<grey):

Simple (blue<red<cyan<grey):

Hyperplasia (blue<red<cyan):

Figure 2.17: Heterogeneity of histological features in GBM tissue samples, by patient

variable outcome [133]. The ability of histopathology to characterize and predict the behavior of brain lesions is thus limited by the sampling error due to the heterogeneity in the lesion as well as the inability to perform samplings at the time of presumed progression [130]. Molecular and metabolic studies can potentially allow for better classification of these tumors and for separation into different prognostic groups. Figure 2.18 shows the *in vivo* spectra corresponding to two biopsy locations that seem similar on a T1-GAD image. This figure shows that multivariate MR data provides complementary information, and that the integration of this data with histology has great potential.

2.3 Important problems in the diagnosis and management of gliomas

Brain cancer research is advancing at an accelerated pace which is being driven by the development of all the new data modalities and the demand for personalized medical treatment. These recent advances are providing ample opportunity for using machine learning methods for personalized medicine. Clinical endpoints that could benefit from a multivariate pattern analysis of multimodality data include tumor grading, early detection of malignant transformations, improved ways of characterizing progression and response to therapy, and personalized survival prognosis.

Traditionally, multivariate MR data are analyzed by radiologists, who combine their experience in analyzing previous images with prior knowledge of the contrast mechanisms of each image to formulate an opinion about the tumor, its type and degree of malignancy, and its borders. Visual interpretation is becoming more complex due to the increasing number of data modalities available.

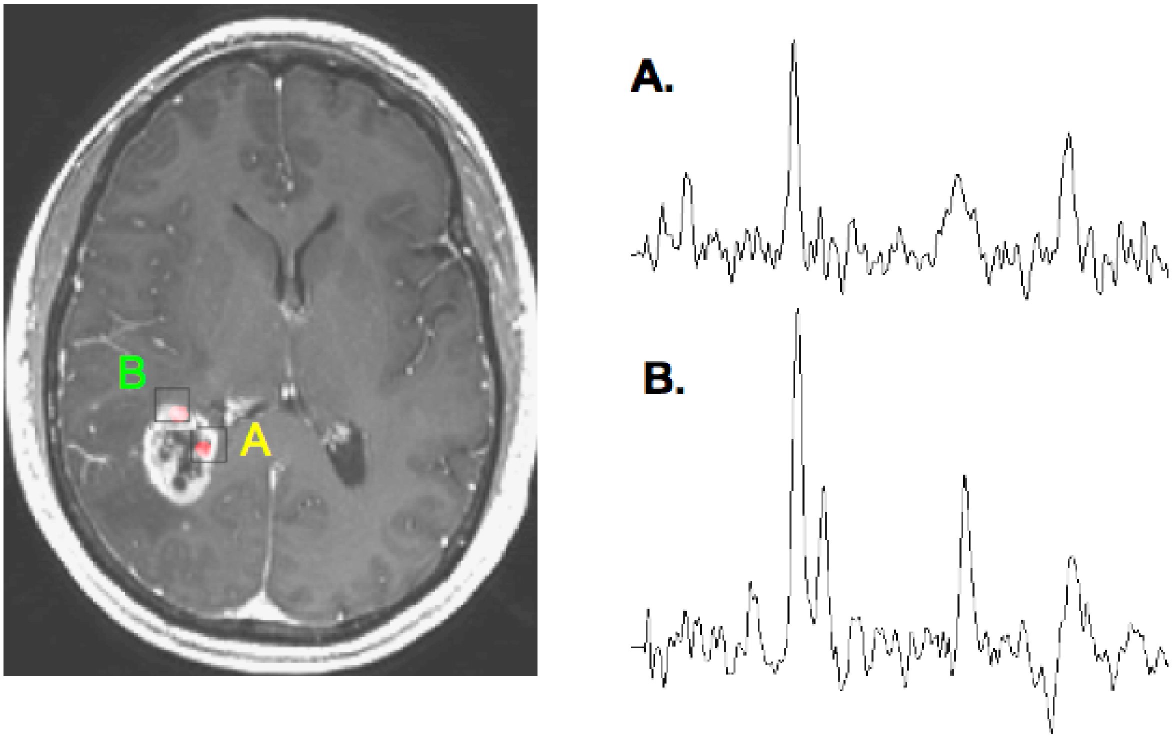


Figure 2.18: (Left): T1-GAD image of a GBM and two seemingly similar biopsy sites; (Right): Corresponding *in vivo* spectra at the two biopsy sites; (A) tumor score 1, hyperplasia score 1, complex vessel score 0, SMI-31 score 0, MIB-1 score 1.9%; (B) tumor score 2, hyperplasia score 2, complex vessel score 3, SMI-31 score 1, MIB-1 score 20%

Thus, it is likely that valuable patterns in the data may remain undiscovered. These issues have led many researchers to pursue computational methods for combining information from multivariate NMR data [2, 3, 9, 12, 15, 26, 27, 33, 38–41, 44, 45, 60, 63, 69, 70, 73, 74, 77, 79, 81, 82, 88, 90, 93, 96–98, 102, 106–108, 113, 114, 119, 129, 132, 139]. Not many of these studies are biologically interpretable nor do they explore the relationship between *in vivo* MR and histological features other than grade and tumor type.

One of the most significant limitations of *in vivo* MR imaging and spectroscopy is the underlying uncertainty about the characteristics of cells in an MRI voxel. While MRI and MRS data can be used to detect abnormalities and provide good spatial information about different tissue properties, the true composition of the tissues can only be determined histologically. However, samples obtained by biopsy or during surgical resection represent a relatively small amount of tissue and may therefore not be fully representative of the characteristics of the tumor. This means that the successful integration of histology and *in vivo* data through the use of multivariate pattern analysis can lead to quantitative non-invasive markers of biological behavior, which are extremely important for characterizing tumor heterogeneity, for defining the most appropriate region for obtaining image guided samples for histological analysis, and for evaluating patients for whom surgical resection is not appropriate.

Thus, this thesis identifies the following open and interesting research problems in the field of brain oncology that could benefit from the application and development of pattern analysis methods:

- assessment of tumor grade
- detection of malignant transformations in gliomas
- development of new quantitative *in vivo* biomarkers to guide evaluation of progression and response to therapy
- personalized survival estimates
- improved selection of biopsy sites
- importance of data features in determining clinical endpoints and in understanding tumor biology
- characterization of tumor heterogeneity
- finding connections between *ex vivo* and *in vivo* biomedical signals.

Chapter 3

Multivariate pattern analysis of biomedical signals for patients with gliomas

Pattern analysis is the process of finding general relations in a set of data. One of the most important pattern analysis tasks is supervised learning, the process of learning a model from supervised training data. The training data consist of a set of training examples, which are pairs of input objects and corresponding output values. Supervised learning algorithms analyze the training data and produce a classification function (if the output is categorical) or a regression function (if the output is continuous) that best describes the relationship in these data. The classification function should be able to predict the correct output value for any valid input object. This requires the learning algorithm to generalize from the training data to unseen situations in a reasonable way.

The development of pattern recognition systems involves the following steps, which will be described in more detail next: preprocessing the data, reducing data dimensionality and selecting relevant features, constructing a classifier, and measuring its performance on previously unseen data. The supervised learning pipeline is illustrated in Figure 3.1.

One of the challenges best addressed by machine learning is illustrated in Figure 3.2. The figure shows biopsy samples with high or low tumor cellularity scores, plotted according to two predictive features. Standard statistical analysis methods reveal that the two tumor cellularity groups are significantly different in terms of their excess Cho to Cre levels and the T1-GAD intensity values at the biopsy locations. However, there is a large amount of overlap between the two groups, which makes it hard to predict whether a biopsy sample has low or high tumor cellularity based on the two illustrated features. Supervised learning methods are able to identify and combine biomarkers that are sensitive and specific for individuals, rather than groups. These methods can identify a small set of predictive features that, when combined, can distinguish between low and high tumor score samples with good accuracy.

In order for supervised learning to be applicable to a specific problem there are some assumptions that need to be met. The data objects need to be represented as feature vectors of fixed length. They also need to be members of mutually exclusive classes. The training set needs to be a good representation of what will be encountered in a practical application. The data need to come from a fixed distribution that does not, for example, vary over time. Also, a loss function has to be defined

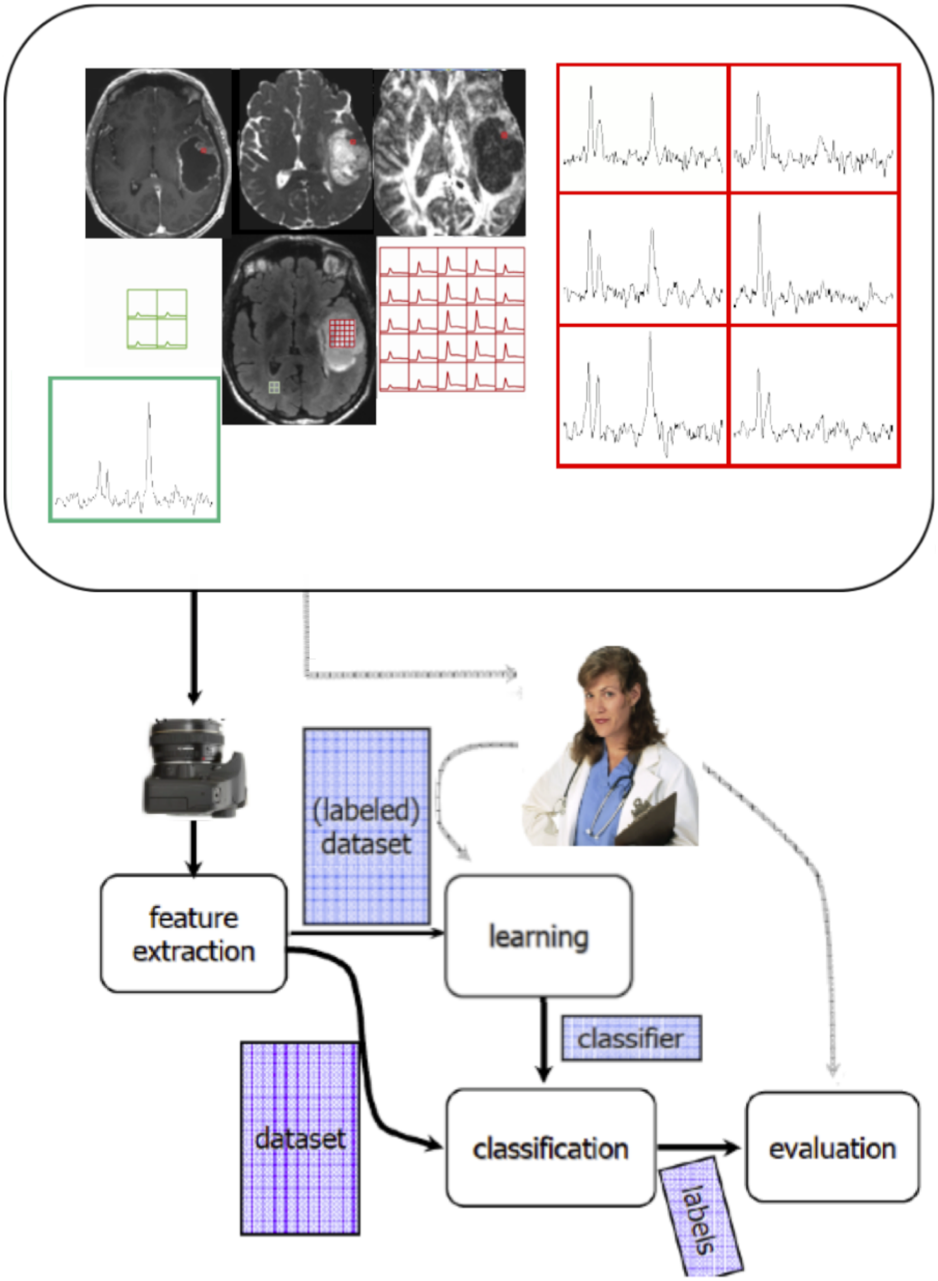


Figure 3.1: Supervised learning pipeline.

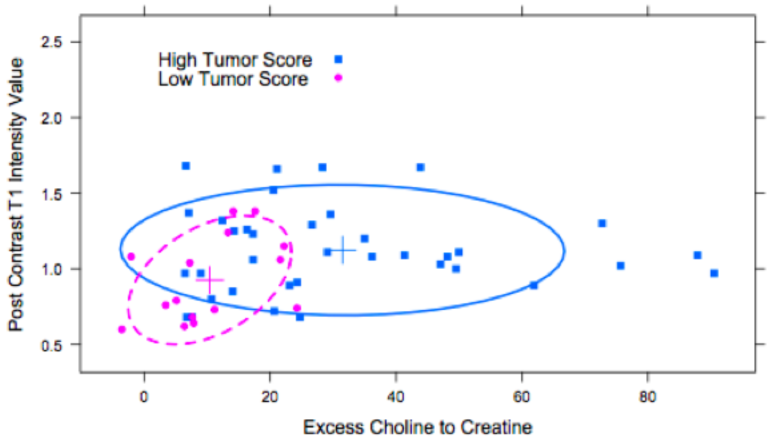


Figure 3.2: Low and high tumor cellularity samples, plotted against two predictive features; standard statistics discovers that the two tumor cellularity groups are significantly different, with high confidence; however, the two features cannot be used to distinguish between high and low tumor samples with high accuracy

and estimated using empirical loss.

Once these assumptions are met, a supervised learning model can be built and evaluated as described in the following subsections.

3.1 Preprocessing

Preprocessing aims to reduce the effects of noise and to transform raw data into a format that will make it easy to analyze. Furthermore, preprocessing steps standardize the parameters of images and spectral data in order to make them comparable across different samples, patients, and populations.

Preprocessing of MR images

For this thesis, the preprocessing of MR images includes noise reduction [42], intensity inhomogeneity reduction [116], normalization by the median intensity in NAWM, and registration [4, 5, 24, 47]. Image registration is the process of aligning images so that corresponding features can easily be related [47]. In this thesis, anatomical images are registered using rigid transformations [47], and perfusion and diffusion images are registered using nonrigid registration based on splines [47].

Preprocessing of spectra

Preprocessing of spectroscopic data involves transforming the raw free induction decay (FID) signal into the frequency domain using the Fast Fourier Transform (FFT).

MRS data also requires spatial transformations, phasing, correction, and water signal suppression. The spectra were normalized according to the integral of each spectrum.

In the case of HRMAS data, the frequency domain signal is shifted using 3-trimethylsilyl propionic acid (TSP) as a reference, and phased using zero-order phase correction. Residual water signal is removed using Hankel-Lanczos singular value decomposition (HLSVD). Each data sample in the frequency domain is normalized using the electronic reference to access (ERETIC) method [76] and the tissue weight of each sample.

Another optional step in the preprocessing of spectral data is the quantification of known metabolites. The most basic quantification technique is peak integration. In this technique, the region of the spectrum that contributes to the peak is marked on the frequency axis. The area of the NMR curve between the two marks is then calculated, and this integral is used to quantify the peak. More complex quantification methods such as LCModel [89], QUEST[100], and AQSES [115] quantify previously known chemicals in NMR spectra by incorporating prior knowledge about the set of frequencies that resonate for each chemical, the various resonance peaks, the relative area of these resonance peaks, and the shape of chemicals' spectra. These methods could potentially overlook important information in the spectrum, such as the presence of previously unquantified metabolites. In addition, methods that search for metabolites independently based on their shape, such as QUEST [100], cannot distinguish between low levels of a chemical or bad fits.

3.2 Feature selection

Dimensionality reduction is a crucial step when analyzing complex data with many parameters and relatively few data samples. This step reduces the complexity of the problem, makes the problem tractable, and improves the quality of the solution. If a model uses too few features, it will not have enough information to produce highly accurate classifications. On the other hand, if a model uses too many features, not enough data is available to accurately estimate all the parameters of the model. This leads to biased results that do not generalize well to unseen examples. The optimal number of features in a model depends on the training set size and on the classification method used, as illustrated in Figure 3.3.

Dimensionality reduction methods can be supervised or unsupervised. Unsupervised methods such as principal component analysis (PCA) try to transform the data in order to obtain a smaller set of variables that retains as much of the information in the data as possible. Supervised dimensionality reduction methods attempt to extract a set of features that retain the most characteristic information for a given problem.

In the medical field, a vast amount of data can be recorded for each patient. However, a large amount of this data is not relevant when trying to answer a specific question. This poses a great challenge for unsupervised feature selection methods. In cases such as this, supervised methods are

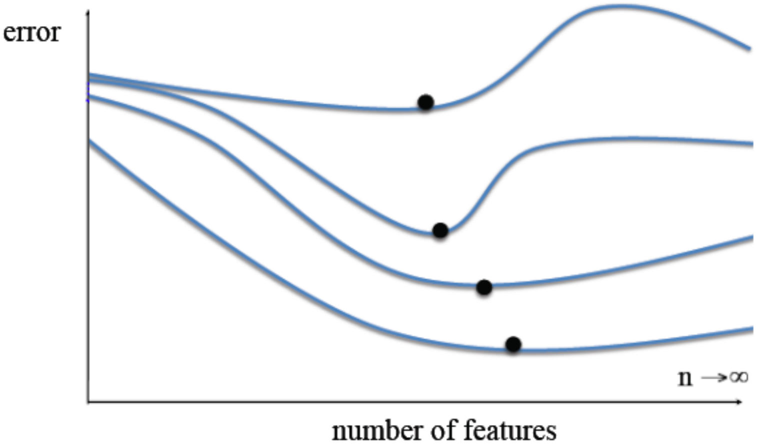


Figure 3.3: Optimal number of features in a model as a function of the training set size and the classifier used

more appropriate. For this reason, this thesis only explores the use of supervised dimensionality reduction methods.

Supervised feature selection techniques typically fall into two categories: feature ranking and subset selection. Feature ranking methods independently rank each feature by a metric and eliminate all the features that do not achieve an adequate score. Subset selection methods search the set of possible solutions for the optimal subset of features. This thesis uses both feature selection techniques. Feature ranking methods are usually very efficient and can be used to drastically reduce the dimensionality of the data. Once the data are of manageable size, optimal subset selection methods that take into account the interaction between features can be used to obtain a final feature set. The pseudocode for these methods is provided in Algorithms 1 and 2.

Because one of the goals of this thesis was to obtain parsimonious models that can be easily interpreted biologically, methods that transform the original features into a smaller set of mathematically related features that combine information from all the available data were not employed. Instead, three feature ranking metrics that lead to easily interpretable models were compared.

The three measures of association between features and classification output that were compared are based on the value of the chi-squared statistic with respect to the class, the information gain ratio with respect to the class, and a conditional probability-based technique that measures the mutual association between class decisions and feature values based on conditional probabilities.

Given a data set (X, Y) of data examples and associated classes, let n be the number of samples in the data set. X is an $n \times m$ matrix, containing n instances of m features. Y is the target attribute, and it can take on c different values.

Algorithm 1 Fliter(S')

```
Require: $S = \{(x_1, y_1), \ldots (x_K, y_K)\}$
Require: $length(x_i) = F$
for $i = 1$ to $F$ do
    Rank(feature $i$, $S$)
end for
$S' = (HighestRankedFeatures(X, Y))$
return $S'$
```

Chi-squared feature selection

The Pearson Chi-squared statistic can be used to measure the association between the target class feature and the input features [71]. This statistic assesses whether paired observations of two variables, expressed in a contingency table, are independent of each other. Consider a feature F that takes on f values and target feature Y which takes on c values. Each data observation is allocated to one cell of an $f \times c$ - dimensional contingency table according to its outcome. Let $O_{i,j}$ denote the frequency for a particular combination of values of the two variables. The null hypothesis is that the occurrence of these outcomes is statistically independent. Under the assumption of independence, the expected frequency in a cell is:

$$E_{i,j} = \frac{\sum_{k=1}^{f} O_{k,j} \sum_{k=1}^{c} O_{i,k}}{n}. \tag{3.1}$$

The value of the chi-squared test statistic is:

$$\chi^2 = \sum_{i=1}^{f} \sum_{j=1}^{c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}. \tag{3.2}$$

Continuous features are discretized into intervals according to a method described in [71]. At first, the values of the attribute F are ordered, and each value forms its own interval. The chi-squared test is performed as described above. Then, pairs of adjacent intervals with the lowest χ^2 values are merged until all of the intervals have χ^2 values exceeding a significance level. This process is repeated with a decreased significance level for each attribute, until an inconsistency rate is exceeded in the discretized data. An inconsistency occurs when two patterns are the same but classified into different categories. An inconsistency level of 5% is allowed.

Gain-Ratio feature selection

Gain-Ratio measures an attribute's usefulness in classification using information theory concepts. It is related to a measure commonly used in information theory, called entropy. This measure characterizes the impurity of a collection of examples. More formally, given a collection X, the entropy of X relative to the c-wise classification described above is:

$$Entropy(X) = \sum_{i=1}^{c} -p_i \log_2 p_i \tag{3.3}$$

where p_i is the proportion of X belonging to class i [83]. The interpretation of entropy is that it specifies the minimum number of bits of information needed to encode the classification of an arbitrary member of X.

Given entropy as a measure of the impurity of a collection of training examples, information gain can be defined as a measure of the effectiveness of an attribute in classifying the training data. Information gain is defined as the expected reduction in entropy caused by partitioning the examples according to this attribute. More formally, the information gain $Gain(X, F)$ of a feature F, relative to a collection of examples X, is defined as:

$$Gain(X, F) = Entropy(X) - \sum_{v \in Values(F)} \frac{|X_v|}{|S|} Entropy(X_v) \tag{3.4}$$

where $Values(F)$ is the set of all possible values for feature F and X_v is the subset of X for which attribute F has value v [83]. Intuitively, $Gain(X, F)$ is the information provided about the target function value, given the value of some other feature F. $Gain(X, F)$ is the number of bits saved when encoding the target value of an arbitrary member of X, by knowing the value of feature F.

The information gain measure favors attributes with many values over those with few attributes. One measure that avoids this problem is the gain ratio. This measure penalizes attributes with many values by incorporating a term called split information that is sensitive to how broadly and uniformly the feature splits the data:

$$SplitInformation(X, F) = -\sum_{i=1}^{c} \frac{|X_i|}{|X|} \log_2 \frac{|X_i|}{|X|} \tag{3.5}$$

where $X_1 \ldots X_c$ are the subsets of examples resulting from partitioning X by the c-valued target attribute Y [83]. Note that $SplitInformation$ is actually the entropy of X with respect to the values of feature F.

The gain ratio measure is then defined as follows [83]:

$$GainRatio(X, F) = \frac{Gain(X, F)}{SplitInformation(X, A)}. \tag{3.6}$$

In order to assess the value of continuous-valued features, new discrete-valued features are dynamically defined by partitioning the continuous attribute values into a discrete set of intervals. For a feature F, a new boolean feature F_t is created. F_t is true if $F < t$ and false otherwise. In order to pick the threshold t that produces the greatest gain ratio, the examples are sorted according to the continuous attribute F, and adjacent examples that differ in their target classification are identified. A set of candidate thresholds can be generated midway between the corresponding values of F. These candidate thresholds can then be evaluated by computing the gain ratio associated with each.

Two-way conditional probability feature selection

Probability based methods are another set of approaches for characterizing the association between features and a target class. A feature selection method proposed by Ahmad and Dey [1] measures the mutual association between class decisions and feature values using conditional probabilities. This method calculates the significance of an attribute using a two-way association function: a feature-to-class association and a class-to-feature association. The feature-to-class association is based on the observation that a change in the feature value should cause a change in the class decision, while the class-to-feature value association is based on the observation that a change in the class decision should be accompanied by a change in the feature's value.

Consider a contingency table O for a feature F and class Y, as described in Section 3.2. Conditional probabilities of classes and feature values can be calculated based on the frequencies in the contingency table:

$$\begin{aligned}
Pr[Y_i|F_j] &= \frac{O_{i,j}}{\sum\limits_{k=1}^{f} O_{k,j}} \\
Pr[Y_i|F_j^{\complement}] &= \frac{\sum\limits_{k=1}^{c} O_{i,k} - O_{i,j}}{n - \sum\limits_{k=1}^{f} O_{k,j}} \\
Pr[F_i|Y_j] &= \frac{O_{i,j}}{\sum\limits_{k=1}^{c} O_{i,k}} \\
Pr[F_i|Y_j^{\complement}] &= \frac{\sum\limits_{k=1}^{f} O_{k,j} - O_{i,j}}{n - \sum\limits_{k=1}^{c} O_{i,k}}
\end{aligned}$$

where $A^{\complement}$ is the complement of set A.

The feature-to-class association is a function of the discriminating power of the values of the feature:

$$DiscriminationPower(F) = \frac{\sum\limits_{i=1}^{c}\sum\limits_{j=1}^{f} \max(Pr[Y_i|F_j], Pr[Y_i|F_j^{\complement}])}{f} - 1. \tag{3.7}$$

The class-to-feature association is a function of the separability of the decision class with respect to the feature:

$$Separability(F) = \frac{\sum_{i=1}^{f}\sum_{j=1}^{c} \max(Pr[F_i|Y_j], Pr[F_i|Y_j^{\complement})}{c} - 1. \qquad (3.8)$$

The overall significance of an attribute is calculated as the average of the feature-to-class association and the class-to-feature association.

Wrapper-based feature selection

In order to obtain a stable, parsimonious model that only uses informative features, a second feature selection step can be performed by evaluating the suitability of subsets of features as a group, not just individually. This step is computationally intense, and for the problems explored in this thesis it cannot be performed without some prior dimensionality reduction. One way to evaluate the relative merit of subsets of features is via a wrapper method. Wrapper algorithms search through the space of possible feature sets and evaluate each subset of features by determining its performance when used in conjunction with a classification model. The search can be performed by utilizing a genetic algorithm for determining which subsets of features to evaluate.

Genetic algorithms represent the solution domain as a binary m-dimensional vector v, where $v_i = 1$ if feature i is selected as part of the feature subset, and $v_i = 0$ otherwise. The fitness function that measures the quality of a solution v is the leave-one-out cross-validation (LOOC) accuracy of a classifier based on the features selected for solution v.

At first, many individual solutions are randomly generated to form an initial population. During each successive generation, a proportion of the existing population is selected to breed a new generation of solutions. Individual solutions are selected through a fitness-based process, where fitter solutions are more likely to be selected. The next step is to use these selected individuals to generate a second generation population of solutions through genetic operators such as crossover and mutation.

For each new solution to be produced, a pair of parent solutions is selected for breeding. The child is produced by combining the two parents using crossover and mutation. Crossover involves comparing individual bits between the two parents. Single-point crossover selects one crossover position $k \in \{1, \ldots, m\}$. Two new offsprings are produced: one containing the first k bits of the first parent and the last $m - k$ bits from the second parent, and another offspring containing the first k bits of the second parent and the last $m - k$ bits from the first parent. Multi-point cross-over selects multiple such cross-over points. Uniform cross-over, which is used in this thesis, allows each position to be a possible cross-over point with a specific probability. Each of the bits of an offspring can also go through a mutation, which involves flipping the value of a bit with a certain probability.

New generations of populations are produced and evaluated until a termination criterion is met. At this point, the highest ranking solution is selected. The pseudocode for this procedure is provided in Algorithm 2.

Algorithm 2 WrapperSearch(S^*)

```
Require: $S = \{(x_1, y_1), \ldots (x_K, y_K)\}$
$i \leftarrow 0$
while Stopping criterion not met do
    $S_i = GenerateFeatureSubset(S)$
    $accuracy_i = Evaluate(S_i)$
end while
return $argmax_{S_i=\{(x'_1,y_1)\ldots(x'_K,y_K)\}}\{accuracy_i\}$
```

3.3 Classification methods

Classification methods use a supplied set of training data containing inputs and their associated labels to learn a model of the relationship between the input values and the labels. This model should be able to accurately describe the training data, and to make predictions about which category previously unseen data inputs belong to. The pseudocode for a learning a classifier is provided in Algorithm 3. This section reviews some classification methods whose results are fairly easy to interpret, such as logistic regression and linear discriminant analysis, as well as more complex non-linear methods such as functional trees (FTs), support vector machines (SVMs), and decision stump boosting.

Algorithm 3 BuildModel(S)

```
Require: $S = \{(x_1, y_1), \ldots (x_K, y_K)\}$
$S' \leftarrow$Filter($S$)
$S^* \leftarrow$WrapperSearch($S'$).
$M \leftarrow$ TrainClassifier($S^*$)
return $M$
```

Linear discriminant analysis

Linear discriminant analysis (LDA) [29] is closely related to regression. It attempts to express the target class variable Y as a linear combination of the predictors X. Suppose that in our data set (X, Y), the target class is binary. Then LDA tries to find a linear separator w such that [29]:

$$y = w^t x \tag{3.9}$$

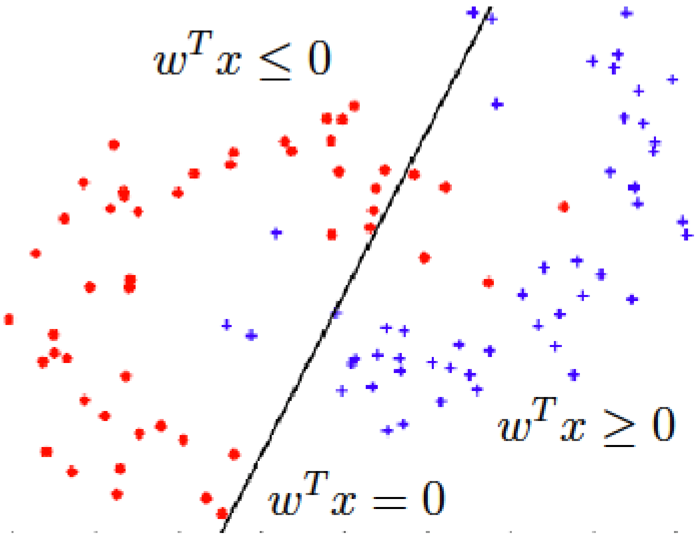


Figure 3.4: LDA example

provides the best separation between the two classes. This is illustrated in Figure 3.4. Fisher's linear discriminant w maximizes the ratio of between-class scatter to within-class scatter [29]. When the conditional densities $p(s|c = 0)$ and $p(x|y = 1)$ are multivariate normal with equal covariances Σ and means μ_0 and μ_1 then the optimal decision boundary has the equation [29]

$$w = \Sigma^{-1}(\mu_1 - \mu_0). \tag{3.10}$$

Logistic regression with ridge estimator

Logistic regression models the probability of occurrence of an event by fitting data to a logit function. Suppose that in our data set (X, Y), the target class Y is binary. The probability that an observation follows a logistic regression model is [13]:

$$p(X_i) = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}, \tag{3.11}$$

where β is an m-dimensional parameter vector. The log-likelihood l of the data under this model is [13]:

$$l(\beta) = \sum_{i=1}^{n} \{Y_i \log p(X_i) + (1 - X_i) \log(1 - p(X_i))\}. \tag{3.12}$$

Maximizing $l(\beta)$ yields the maximum likelihood estimator for β. This estimator maximizes the likelihood of the observed data given the model.

The maximization of the log-likelihood function can be done with a penalty on the norm of β [13]:

$$l^{\lambda}(\beta) = l(\beta) - \lambda||\beta||^2, \tag{3.13}$$

where $||\beta|| = (\sum_{j=1}^{m} \beta_j^2)^{1/2}$ is the norm of the parameter β. Let $\hat{\beta}^{\lambda}$ be the value that maximizes this equation. λ is called a ridge parameter and it controls the amount of shrinkage of the norm of β. When the number of predictors m is large compared to the number of samples n, this parameter can circumvent the problem of predictor collinearity. The new model parameter, $\hat{\beta}^{\lambda}$, is somewhat biased but has small variance. This parameter can be obtained by the Newton-Raphson maximization procedure, as described by Cessie and Houwelingen [13].

Functional trees

Tree induction methods [99] and regression models are popular techniques for supervised learning tasks, both for the prediction of nominal classes and numerical values. Regression methods fit a simple linear or logistic model to the data, resulting in low variance but potentially high bias estimates. If irrelevant attributes are included in a regression model, this usually results in a smaller error on the training data, but will not, however, increase the predictive power over unseen cases, and can sometimes significantly reduce accuracy. Decision trees, such as the one illustrated in Figure 3.5, classify data into categories based on a series of questions or rules about attributes of the class. A path in a decision tree can be seen as a conjunction of boolean expressions that explain how to classify the data. Decision tree classifiers are able to capture nonlinear patterns in the data, but are less stable and more prone to overfitting. These two schemes can be combined into FTs, which are able to use decision nodes with multivariate tests and leaf nodes that make predictions using linear or logistic functions. FTs are efficient to construct and easy to interpret, they can be used for either classification or numerical prediction, and they are formed through a constructive induction process that selects relevant attributes automatically, without changing the representation of the data or projecting it into a different space. This property of FTs is important in the medical imaging field, where it is necessary to be able to interpret the role of different attributes in the prediction process. This knowledge can provide new insights about the disease and treatment options.

FTs were first introduced by Gama [37]. Next, a short description of the algorithm is provided. The works of Gama [37] and Landwehr et al. [64] offer more implementation details. A FT is built by starting at the root. The existing set of attributes is extended using a constructor function that fits regression functions using the LogitBoost algorithm [64]. This algorithm starts out with a simple linear regression model based on the most predictive attribute. In every iteration, it computes response variables that encode the error of the currently fit model on the training examples, and then tries to improve the model by adding another simple linear regression function fit by least squared error. Because every multiple linear regression function can be expressed as a sum of simple linear regression functions, the general model does not change whether multiple or simple linear regression functions are used. LogitBoost is guaranteed to converge to the maximum likelihood estimate.

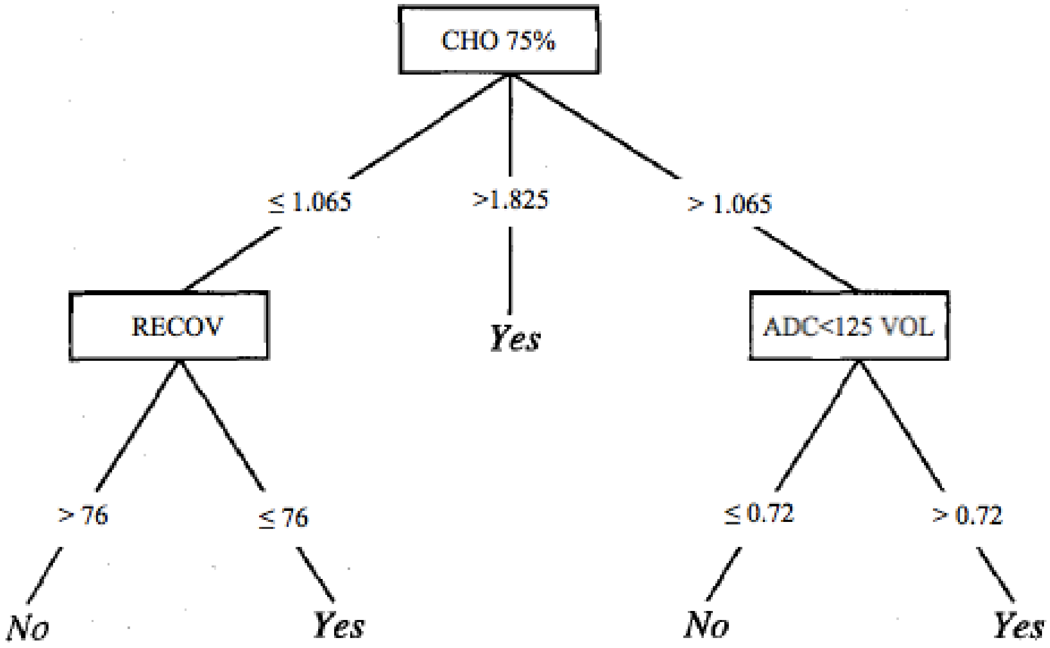


Figure 3.5: A decision tree for determining whether a lesion transformed to a higher grade

However, if the method is stopped before convergence, this will result in the automatic selection of the most relevant attributes. Therefore, LogitBoost is stopped based on a cross-validation method: more iterations are performed (and therefore more attributes are included) only if this improves prediction accuracy over unseen instances. Once the set of attributes is extended, the attribute that maximizes the information gain ratio [99] is selected as a splitting criterion. The child nodes are split recursively, by incrementally refining the regression models already fit at higher levels in the tree, thus taking into account the attributes that are only predictive locally. Tree growing stops if a node contains less than fifteen examples or if a particular split results in two subsets, one of which contains less than two examples. A linear model is only built at a node if that node contains at least five examples. Otherwise, a leaf with the majority class of the node is returned. Once a tree has been grown, it is pruned back using a bottom-up procedure. At each non-leaf node three possibilities are considered: performing no pruning, replacing the node with a leaf that predicts a constant, or replacing it with a leaf that predicts the value of the constructor function that was learned at the node during tree construction. The option that leads to the smallest error on a pruning data set is selected. Predicting a test instance using a FT is accomplished by traversing the tree in a bottom-down fashion. At each decision node the local constructor function is used to extend the set of attributes, and the decision test determines the path that the instance will follow. Once a leaf is reached, the instance is classified using either the constant or the constructor function at that leaf.

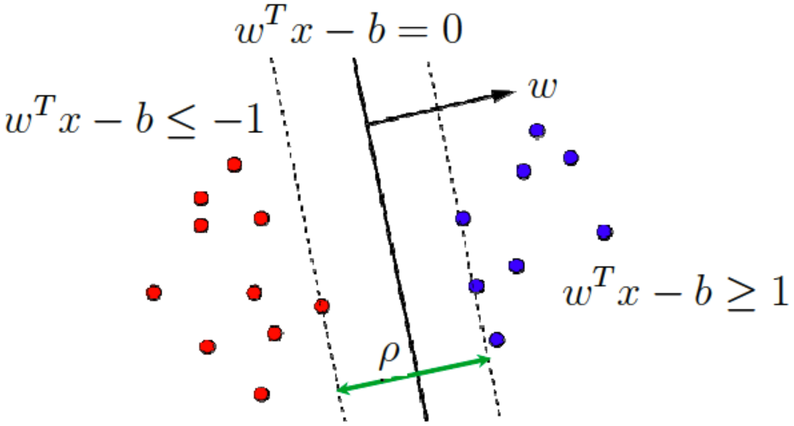


Figure 3.6: SVM separating hyperplane

Support vector machines

SVMs are a popular and versatile classification technique. In the simplest linear form, an SVM is a hyperplane that separates a set of positive examples from a set of negative examples with maximum margin. The margin is defined by the distance of the hyperplane to the nearest positive and negative examples. The separating hyperplane is

$$u = w^T \cdot x - b, \tag{3.14}$$

where w is the normal vector to the hyperplane, and x is the input vector [95]. The separating hyperplane is the plane $u = 0$, and the nearest points lie on the planes $u = \pm 1$. The margin ρ is thus $\rho = 1/||w||^2$. An example is illustrated in Figure 3.6.

Maximizing the margin can be expressed as an optimization problem:

$$\min_{w,b} \frac{1}{2}||w||^2 \quad s.t. \quad y_i(w^T \cdot x_i - b) \geq 1, \forall i, \tag{3.15}$$

where x_i is the ith training example and y_i is its associated label, which can take on the values ± 1 [95].

Using a Lagrangian, this optimization can be converted into a dual form which is a quadratic programming problem where the objective function L is solely dependent on a set of Lagrange multipliers α_i [95]:

$$L(\alpha) = \min_{\alpha} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_j (x_i \cdot x_j) \alpha_i \alpha_j - \sum_{i=1}^{n} \alpha_j$$

$$s.t. \quad \sum_{i=1}^{n} y_i \alpha_i = 0$$

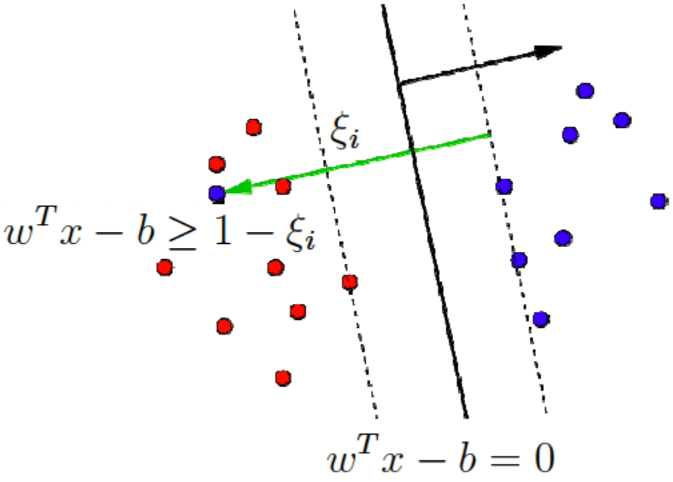


Figure 3.7: SVM slack variables

Once the Lagrange multipliers are determined, the normal w and the threshold b can be derived as:

$$w = \sum_{i=1}^{n} y_i \alpha_i x_i$$
$$b = w^T \cdot x_k - y_k \quad for \quad \alpha_k > 0.$$

Not all data sets are linearly separable. Slack variables, illustrated in Figure 3.7, can be introduced to penalize the failure of an example to reach the correct margin:

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C \sum_{i=1}^{n} \xi_i$$
$$s.t. \quad y_i(w^T \cdot x_i - b) \geq 1 - \xi_i, \forall i,$$

where ξ_i are slack variables the permit the margin failure and C is a parameter which trades off wide margin with small number of margin failures [95].

SVMs can be even further generalized to non-linear classifiers using kernel functions:

$$u = \sum_{k=1}^{n} y_j \alpha_j K(x_j, x) - b, \tag{3.16}$$

where K is a kernel function that measures the similarity or distance between the input vector x and the stored training vector x_j [95]. An example of how a kernel can transform a nonlinearly separable data set into a linearly separable one can be seen in Figure 3.8. Polynomial kernels are used in this thesis.

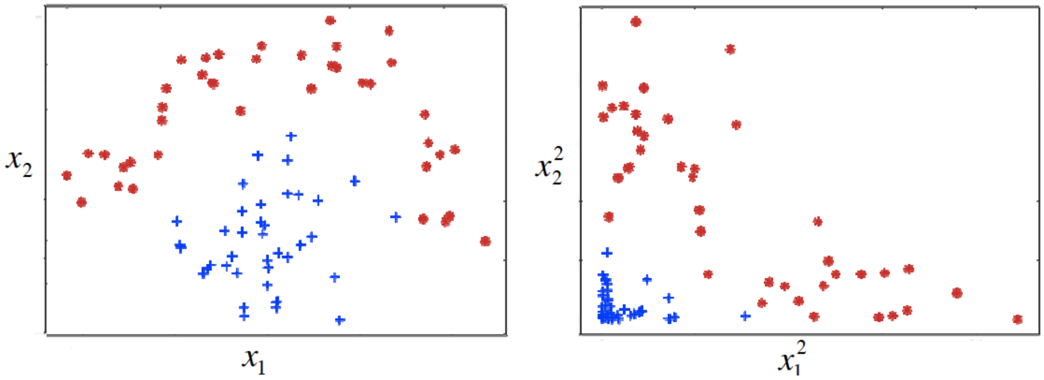


Figure 3.8: Example of SVM kernel application

Due to its large size, the quadratic programming problem that arises from SVMs cannot be easily solved via standard quadratic programming (QP) techniques. Platt introduced sequential minimal optimization (SMO), a simple algorithm that can quickly solve the SVM QP problem without any extra matrix storage and without using numerical QP optimization steps [94]. SMO decomposes the overall QP into sub-problems and solves the smallest possible optimization problem at every step. For the standard SVM QP problem, the smallest possible optimization involves two Lagrange multipliers. At every step, SMO chooses two Lagrange multipliers to jointly optimize, finds the optimal values for these multipliers, and updates the SVM to reflect the new optimal values [94].

Decision stump boosting

Boosting is an important recent development in classification methodology. It works by sequentially applying a classification algorithm to re-weight versions of the training data and then taking a weighted majority vote of the sequence of classifiers thus produced [36], as illustrated in Figure 3.9. Given the training data (X, Y), with $Y_i = -1$ or 1, the classification function is defined as:

$$F(x) = \sum_{i=1}^{W} w_i f_i(x), \tag{3.17}$$

where each $f_i(x)$ is a simple classifier, and w_i are constants. The corresponding prediction is $sign(F(x))$. The classifiers $f_i(x)$ are trained on W weighted versions of the training sample. At the beginning, all samples have equal weights. Then, in subsequent iterations, currently missclassified examples are given higher weight. The final classifier is defined as a linear combination of the classifiers from each stage. Let err_i be the weighted classification error at iteration i. Then $w_i = \log((1 - err_m)/err_m)$. The training examples that are missclassified are given new weights

$$v_j \leftarrow v_j \exp(w_i \mathbb{1}_{C_i \neq f_i(S_j)}) \tag{3.18}$$

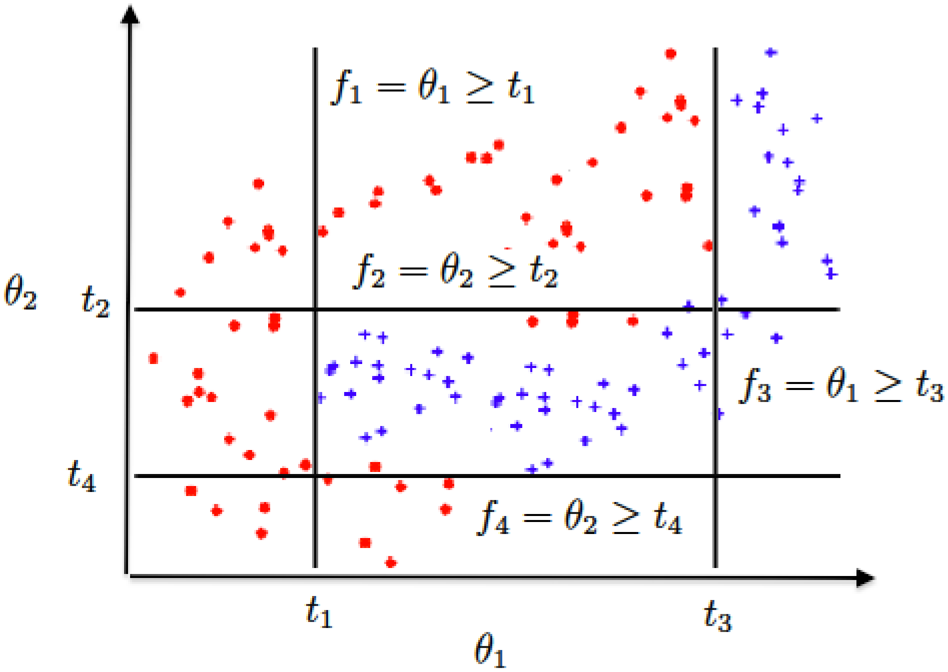


Figure 3.9: Boosting

after which the weights are re-normalized to add up to 1.

Decision stumps are used for the weak learners f_i. Decision stumps are single-split decision trees with only two terminal nodes. For numerical data, these are classifiers of the type $x \leq t$, where t is a threshold automatically chosen to maximize classification accuracy.

The success of this seemingly ad hoc technique can be explained in terms of the well-known statistical principles of additive modeling and maximum likelihood [36].

Mixed effects models

A mixed effects model is an extension of a regression model for cross-sectional data that introducing random effects in the model to account for variations between individuals and correlations within individuals. The random effects in a mixed effects model not only incorporate heterogeneity in the data but also incorporate correlation between the multiple measurements within each individual or cluster. This is because the data within the same individuals or the same clusters share the same random effects or similar characteristics, which leads to correlation in the data.

Let $y_{i,j}$ be the j^{th} response value for individual or cluster i, where $i = 1, 2, \ldots n$ and $j = 1, 2, \ldots n_i$. Let $y_i = (y_{i,1}, \ldots y_{i,n_i})$ be the n_i repeated observations within individual or cluster i. A general linear mixed effects model can be written as:

$$\begin{aligned} y_i &= X_i\beta + Z_i b_i + e_i \\ b_i &\sim N(0, D) \\ e_i &\sim N(0, R_i), \end{aligned}$$

where β are the population parameters or fixed effects, b_i are the random effects, X_i is a matrix containing the covariates of individual i, e_i are random errors of within-individual measurements, R_i is a variance-covariance matrix of within-individual observations, and D is the random covariance matrix of the random effects [136].

Logistic mixed effects models link the mean response of the data, μ, to the covariates, using the logit link:

$$g(\mu) = \log \frac{\mu}{1-\mu},$$

where $\mu = E(y) = P(y = 1)$. The logistic mixed effects model can be written as:

$$\begin{aligned} g(\mu_{i,j}) &= x_{i,j}^T\beta + z_{i,j}^T b_i + e_i \\ b_i &\sim N(0, D), \end{aligned}$$

where $\mu_i = E(y_{i,j}|\beta, b_i)$ is the conditional mean, $x_{i,j}$ and $z_{i,j}$ are vectors of covariates, β is a vector of mixed effects, and D is a covariance matrix [136].

Statistical inference for mixed effects models is typically based on the maximum likelihood method or the restricted maximum likelihood method [136].

Mixed effects can be used to model data in which several biopsy samples are acquired for each patient in the study. A simple logistic regression model with random intercept for a binary response $y_{i,j}$ taking only two possible values can be written as:

$$\begin{aligned} \log \frac{\mu_{i,j}}{1-\mu i, j} &= \beta_{0,i} + \beta_1 x_{i,j} = \beta_0 + b_i + \beta_1 x_{i,j} \\ b_i &\sim N(0, d^2), \end{aligned}$$

where $\mu_{i,j} = E(y_{i,j}) = P(y_{i,j} = 1)$ and $\beta_{0,i} = \beta_0 + b_i$.

3.4 Multivariate survival analysis

Survival analysis examines and models the time it takes for events to occur. A typical such event is death. However, applications of survival analysis are much broader. Survival analysis focuses on the distribution of survival times.

Kaplan-Meier estimator of survival

The Kaplan-Meier estimator estimates the survival function from life-time data. Let $S(t)$ be the probability that an item from a given population will have a lifetime exceeding t. For a sample from this population of size N, let the observed times until death of N sample members be $t_1, t_2, \ldots, t_N$. Corresponding to each t_i is n_i, the number at risk just prior to time t_i, and d_i, the number of deaths at time t_i.

The Kaplan-Meier estimator of the survival function at time t is [55]:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}. \tag{3.19}$$

When there is no censoring, n_i is just the number of survivors just prior to time t_i. With censoring, n_i is the number of survivors less the number of censored cases.

Cox proportional hazards model

The most interesting survival modeling examines the relationship between survival and one or more covariates. The Cox proportional hazards regression model is the most widely used such methods. One representation of the survival function, as noted in the previous subsection, is $S(t) = Pr(T > t)$. Another representation of the distribution of survival times is the hazard function, which estimates the risk of death at time t conditional on survival to that time [55]:

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr[(t \leq T < t + \Delta t) | T \geq t]}{\Delta t}.$$

The Cox proportional hazards model examines the relationship of the survival distribution to covariates, by specifying a model for the log hazard. For example, a parametric model based on the exponential distribution is written as [55]:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_k x_{i,k}, \tag{3.20}$$

where $\alpha(t) = \log h_0(t)$, is the baseline hazard function.

Consider two observations i and j. The hazard ratio for these two observations is [55]:

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t) \exp(\beta x_i)}{h_0(t) \exp(\beta x_j)} = \frac{\exp(\beta x_i)}{\exp(\beta x_j)}, \tag{3.21}$$

which is independent of the time t. Consequently, the Cox model is a proportional hazards model.

The Cox proportional hazards model can handle time-dependent covariates and lagged covariates.

In order to determine whether a fitted Cox regression model adequately describes the data, it is necessary to check for proportional hazards, nonlinearity, and influential observations [55].

Brier score

It is possible to measure the accuracy of survival models using the Brier score. This score is a function that measures the accuracy of a set of probability assessments. It measures the average squared deviation between predicted probabilities for a set of events and their outcomes.

The most common formulation of the Brier score is [55]:

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2, \tag{3.22}$$

where f_t is the probability that was forecast, and o_t is the actual outcome of the event at instance t.

3.5 Evaluating the accuracy of a model

For classification problems, it is natural to measure a classifier's performance in terms of the accuracy. The classifier predicts the class of each instance: if it is correct, it is counted as a success; if not, it is an error. The accuracy is just the proportion of correct classifications over a whole set of instances, and it measures the overall performance of the classifier. The classifier's likely future performance on new data is, of course, more interesting than its past performance on old data. To predict the performance of the classifier on new data, the classification accuracy needs to be assessed on a test set that played no role in the formation of the classifier. When a large amount of data is available, a model can be learned based on a large training set, and evaluated based on another large test set. The error rates of the model on the training and test sets converge to an asymptotic error rate, as the size of the data set increases. This is illustrated in Figure 3.10. The difference between the training set error and the test set error is due to overfitting. Once the accuracy has been determined, it is acceptable to produce a final classifier for actual use based on both the training and the test data [134].

When the amount of labeled data is limited, the question becomes how to make the most of the limited data set. From this dataset, a certain amount is held out for testing, and the remaining is used for training. The more data are used for the training, the better the classifier. The more data are used for testing, the better the accuracy estimate. Cross-validation and bootstrapping are techniques for dealing with this dilema. Among these, bootstrapping is the evaluation method of choice in most practical limited-data situations [134].

Cross-validation

Cross-validation reserves a certain amount of data for testing and uses the remainder for training, but in order to mitigate any bias caused by a particular set being chosen for holdout, the whole process is repeated, training and testing, several times, with different random samples. The initial data set is split into a fixed number of partitions, or folds. Each partition is in turn used for testing while the remainder of the data is used for training. The accuracy estimates obtained during the

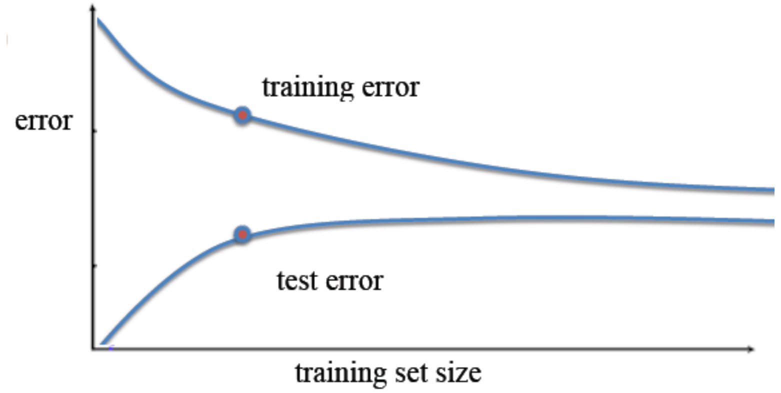


Figure 3.10: The training set and test set error rates converge to an asymptotic error

different iterations are averaged to yield the overall accuracy. LOOC is a special case of cross-validation, in which each fold contains only one data sample; thus one sample is in turn held out for testing, while the rest of the data is used for training. This procedure is attractive because it uses the greatest possible amount of data in each case, and no random sampling is involved. The pseudocode for this procedure is provided in Algorithm 4. Thus, leave-one-out offers the chance to make the most out of a small data set and to obtain as accurate an estimate of the classifier's performance on unseen data as possible [134].

Algorithm 4 Leave-One-Out Cross-Validation(D)

```
Require: $D = \{(x_1, y_1), \dots (x_N, y_N)\}$
for $i = 1$ to $N$ do
    $D_i \leftarrow D\{(x_i, y_i)\}$.
    $M_i \leftarrow$ BuildModel($D_i$)
    $error_i \leftarrow |M_i(x_i) - y_i|$
end for
return $accuracy = 1 - \frac{1}{N}\sum_{i=1}^{N} error_i$
```

Bootstrapping

Cross-validation methods provide unbiased, but high variance estimates of a classifier's prediction accuracy. This means that if the whole validation process were to be repeated, cross-validation estimates could vary significantly. Bootstrapping, on the other hand, provides nearly unbiased estimates of the prediction accuracy that are relatively low in variance, by repeating the validation process several times on different samples of the data, thus also avoiding overfitting. Bootstrapping is based on the statistical procedure of sampling with replacement. The data set is sampled with replacement to form a training set. For this, a data set with N instances is sampled N times, with replacement, to give another data set of N instances. Because some elements in this second data set will very likely be repeated, there must be some instances in the original data set that have not been picked - the test instances. For a reasonably large data set, the test set will contain about 36.8% of the instances, and the training set will contain about 63.2% of them [134], leading to the name 0.632 bootstrapping. Some instances will be repeated in the training set, bringing it up to a total size of N, which is the same as the original data set. The accuracy estimate obtained by training a classifier on the training set and calculating its accuracy on the test set will be a pessimistic estimate of the true prediction accuracy of a model on new data drawn from the sampling distribution, because even though the training set has size N, it only contains 63.8 % of the original data, which is far less than the amount used in LOOC. To compensate for this, the accuracy on the test sample is combined to that of the training sample, to give an unbiased estimate of the overall accuracy:

$$accuracy = 0.632 \times accuracy_{test} + 0.368 \times accuracy_{training}. \tag{3.23}$$

Then, the whole bootstrap procedure is repeated several times, with different replacement samples for the training set, and the results are averaged. The pseudocode for this procedure is provided in Algorithm 5.

The bootstrap procedure offers the best way to estimate prediction accuracy for very small data sets [134]. The estimate of the accuracy represents an unbiased estimate of the performance of the classifier on new data drawn from the sampling distribution [134]. However, before a model can be used with confidence in clinical practice, it is necessary to validate it on a completely independent data set, because the sampling distribution may not always accurately summarize the population distribution.

In this thesis, both LOOC and 0.632 bootstrapping with 200 repetitions [31] are used in order to evaluate the ability of the pattern recognition methods employed to generalize to unseen examples drawn from the sampling distribution.

The pseudocode for these methods is provided in Algorithms 4 and 5. Thus, the data set given as input to these algorithms contains N samples. The cross-validation algorithm calls Algorithm 3 using only $N - 1$ of the original samples each time. The bootstrapping algorithm calls Algorithm 3 using an N-sample data sets obtained by resampling the original data set with replacement. Algorithm 3 calls the wrapper-search algorithm (Algorithm 2) using the reduced or resampled data set. The wrapper-search algorithm, in turn, calls the cross-validation or bootstrapping algorithms respectively in the **Evaluate** step, thus ensuring that all of the prediction accuracy estimates are based on data that were left out during the model building.

3.6 EM imputation of missing values

Missing values often hinder the analyses of multivariate data. The most common practice is to remove data instances that contain missing values. However, in the medical field this practice is not acceptable, because even though the rate of missing values for each parameter is not very high,

Algorithm 5 Bootstrapping(D)

```
Require: $D = \{(x_1, y_1), \ldots (x_N, y_N)\}$
for $i = 1$ to $N$ do
    for $b = 1$ to $NFOLDS$ do
        $S_i \leftarrow$ Resample($D$)
        $error_b = 0$
        $total_b = 0$
        if $(x_i, y_i) \notin S_i$ then
            $M_{i,b} \leftarrow$ BuildModel($S_i$)
            $error_{i,b} \leftarrow |M_{i,b}(x_i) - y_i|$
            $error_b \leftarrow error_b + error_{i,b}$
            $total_b \leftarrow total_b + 1$
            $error_b \leftarrow error_b / total_b$
        end if
    end for
end for
$accuracy_b \leftarrow 1 - \frac{1}{N} \sum_{i=1}^{N} error_b$
$M \leftarrow$ BuildModel(D)
$accuracy_t \leftarrow 1 - \frac{1}{N} \sum_{i=1}^{N} |M(x_i) - y_i|$
return $accuracy \leftarrow 0.632 \times accuracy_b + 0.368 \times accuracy_t$
```

there are very few patients with complete data. In this thesis, the EM algorithm is used for filling in missing data with plausible values, as well as to estimate the histological parameters when desiring to predict tumor cellularity based on *in vivo* parameters alone.

Assuming a specific data model, it would be easy to estimate the model parameters if the missing values are known. Similarly, if the model parameters are known, it would be possible to use the model to estimate the missing values. The EM algorithm provides an iterative approach to the problem of maximum likelihood parameter estimation in statistical models with latent variables. In the expectation (E) step, the values of the missing variables are estimated by calculating the probability of the latent variables given the observed variables and the current values of the model parameters. In the maximization (M) step, the parameters are adjusted based on the current estimates of the missing values. These steps are repeated until the sequence of parameters converges to the maximum likelihood estimates that average over the distribution of missing values [58]. The EM algorithm for multiple imputations for Gaussian data is presented in detail by Schafer [109]. Below is a short description of the algorithm.

The dataset Y is assumed to be a matrix of n rows and m columns, with the rows corresponding to observations and the columns corresponding to variables. The complete dataset contains observed and missing portions: $Y = (Y_{obs}, Y_{mis})$. Let $y_{i,j}$, with $i \in 1 \ldots n$ and $j \in 1 \ldots m$, denote an individual element of Y and let y_i denote a row of the data matrix. The main model assumption is that $y_1 \ldots y_n$ are independent realizations of a multivariate normal distribution with mean vector μ and covariance matrix Σ.

When both μ and Σ are unknown, the conjugate prior distribution for the multivariate normal data model is the normal inverted-Whishart distribution [109]. Suppose that Σ is inverted-Wishart and that μ given Σ is assumed to be conditionally multivariate normal. Then the complete-data likelihood function is [109]:

$$L(\theta|Y) \propto |\Sigma|^{-\frac{n}{2}} exp\big\{\frac{n}{2} tr\Sigma^{-1}S\big\} exp\big\{-\frac{n}{2}(\bar{y}-\mu)^T\Sigma^{-1}(\bar{y}-\mu)\big\}. \tag{3.24}$$

The complete-data posterior is normal inverted-Wishart [109]:

$$\mu|\Sigma, Y \propto N(\mu_0', (\tau')^{-1}\Sigma), \Sigma|Y \propto W^{-1}(m', \Lambda'), \tag{3.25}$$

where the updated parameters are [109]:

$$\begin{aligned} \tau' &= \tau + n; m' = m + n; \mu_0' = \big(\frac{n}{\tau+n}\big)\bar{y} + \big(\frac{\tau}{\tau+n}\big)\mu_0; \\ \Lambda' &= \big[\Lambda^{-1} + nS + \frac{\tau}{\tau+n}(\bar{y}-\mu_0)(\bar{y}-\mu_0)^T\big]. \end{aligned} \tag{3.26}$$

The prior distribution of μ is assumed to be uniform over the m-dimensional real space. Under the improper prior, the complete-data posterior becomes [109]:

$$\mu|\Sigma, Y \propto N(\bar{y}, n^{-1}\Sigma); \Sigma|Y \propto W^{-1}(n-1, (nS)^{-1}). \tag{3.27}$$

If Y were fully observed, then the maximum likelihood (ML) estimation problem would amount to maximizing $log(L(\theta|Y)$. However, the ML estimates cannot be obtained in closed form when Y has missing values. The E-step of EM calculates the conditional expectations of the missing variables conditioned on the observed variables and fixed model parameters. This amounts to calculating the expected complete log likelihood, and thus to finding the complete-data sufficient statistics over $P(Y_{mis}|Y_{obs}, \theta)$ for assumed value of θ. The sufficient statistics are of the form $\sum_i y_{i,j}$ and $\sum_i y_{i,j}y_{i,k}$ [109]. Thus, it is necessary to find the expectations of $y_{i,j}$ and $y_{i,j}y_{i,k}$ over $P(Y_{mis}|Y_{obs}, \theta)$. Because the rows of Y are independent given θ, the following is true:

$$P(Y_{mis}|Y_{obs}, \theta) = \prod_{i=1}^{n} P(y_{i(mis)}|y_{i(obs)}, \theta). \tag{3.28}$$

The distribution $P(y_{i(mis)}|y_{i(obs)}, \theta)$ is a multivariate normal linear regression of $y_{i(mis)}$ and $y_{i(obs)}$ [109]. The parameters of this regression can be calculated by sweeping the θ-matrix on the positions corresponding to the variables in $y_{i(obs)}$, as described by Schafer [109]. The E-step consists of calculating and summing the expected values of y_{ij} and $y_{ij}y_{ik}$ for each j and k. Carrying out the M-step involves maximizing the expected complete log likelihood with respect to the parameters.

3.7 Applications in brain tumor diagnosis, characterization, and treatment

The potential of *ex vivo* and *in vivo* NMR has lately been extensively exploited in conjunction with multivariate pattern analysis methods in the separation between normal brain and brain tumors [60, 81, 104] as well as for the characterization of different types and degrees of malignancy in tumors [2, 15, 27, 33, 40, 44, 73, 77, 79, 88, 90, 97, 98, 102, 106–108, 113, 114, 128, 132]. Recent work using HRMAS for brain tumors has shown that tumor classification according to histological type [2, 26, 45, 73, 90, 97, 119, 129] and grade [2, 33, 40, 82, 90, 97, 101, 102, 113, 119, 132, 139] is possible with the use of multivariate methods such as LDA [97, 132], SVMs [2, 132], least square SVMs [2, 97, 132], logistic regression [101, 102], partial least square discriminant analysis [33], and multi-layer perceptrons [2]. Moreover, HRMAS multivariate studies have successfully revealed the status of tumor microheterogeneity [15, 106] and detected tumor metabolic alterations before they were morphologically detectable [2]. These studies combine dimensionality reduction methods such as PCA [26, 129, 132] and metabolite quantification [26, 97, 101, 102, 106, 107, 132, 135] with the robust classification methods listed above. Many dimensionality reduction methods and classifiers combine the data in such a way that the extracted components lack physical meaning. For example, in PCA, the entire spectra may participate in the principal components, and the linear combination may mix both positive and negative weights, which might partly cancel each other. On the other hand, some studies are biologically grounded through the use of concentrations of known metabolites that are quantified using methods such as LCModel [89], QUEST[100], and AQSES [115]. These methods quantify previously known chemicals in NMR spectra by incorporating prior knowledge about the set of frequencies that resonate for each chemical, the various resonance peaks, the relative area of the various resonance peaks, and the shape of chemicals' spectra. These methods could potentially overlook important information in the spectrum, such as the presence of previously unquantified metabolites. In addition, methods that search for metabolites independently based on their shape, such as QUEST [100], cannot distinguish between low levels of a chemical or bad fits. Thus, whole-spectrum analyses with statistical feature selection methods that maintain a link to the biological meaning of the features may provided additional benefits when compared to PCA or quantification based methods.

Most previous studies fall into two categories: biologically-grounded studies that limit themselves at describing independent metabolite behavior while ignoring the metabolite interconnectivity, and multivariate statistical studies that focus on brain tumor characterization and typing, without a link to the biological meaning behind the statistical analyses. In this thesis, robust feature selection and classification methods are combined to obtain accurate classifications with a parsimonious and interpretable set of features. The parsimonious sets of features can be used to gain insight into the biological processes of tumor cells, to inform *in vivo* MRS acquisition based on *ex vivo* results, and to better categorize tumor properties non-invasively. Compared to other multivariate studies, this work includes a relatively large number of data samples of each category. Another new aspect of our work is the detection of transformations in tumors that were previously of the same category, thus gaining insight into the malignant transformations that occur in LGGs that are undergoing treatment.

Chapter 4

Identification of malignant transformations in gliomas using HRMAS spectra from image guided tissue samples

The study presented in this chapter was published in the Journal of Artificial Intelligence in Medicine [18] and reproduced with permission.

Recent oncology research shows that the evaluation of cellular metabolism can be very helpful for the diagnosis and assessment of patients with brain tumors [18, 19]. HRMAS spectroscopy provides detailed metabolic data of whole biopsy samples for investigating tumor biology (see Figures 4.1, 4.2, and 4.3). Analysis of such data can lead to identification of metabolites that may be used as biomarkers for discriminating different types of cancer, and for grading and assessing the evolution of tumors. The identification of *ex vivo* metabolites can also inform the acquisition of *in vivo* MRS, which can lead to a non-invasive assessment of tumor biology.

LGGs include a diverse group of tumors, with distinct characteristics, patterns of occurrence, response to treatment, and survival timelines. The objective of this study is to determine whether quantitative metabolic parameters derived from HRMAS data are predictive of the biologic behavior of recurrent LGGs. This is an important clinical question because of the need to determine whether a lesion has transformed to a more malignant phenotype and to treat each patient with the therapy that is most likely to be effective for their particular lesion. Thus, the goal of this study was to explore multivariate pattern recognition methods to generate statistical models of the metabolic characteristics of recurrent LGGs that correlate with aggressive biology and poor clinical outcome. These models can be used for the early detection of malignant transformations in individual LGGs, and can lead to a timely change in treatment for each patient.

Generating a predictive model of the metabolic characteristics of recurrent gliomas using HRMAS begins with identifying regions in the HRMAS spectrum that can be used to accurately discriminate between different tumor grades in patients with recurrent or newly diagnosed gliomas, without making any prior assumptions about which metabolites are present. Initially, all the regions in the HRMAS spectrum that have a mutual association with disease stage or tumor grade are identified. Then, a small subset of these features is selected to build a parsimonious model that is capable

Table 4.1: Characteristics of recurrent LGG patients enrolled in the study and treatment received prior to recurrence, tabulated according to the tumor grade at recurrence

Characteristic	Grade 2	Grade 3	Grade 4	Total
Male	9	13	4	26
Female	13	11	3	27
Mean age (years)	37.6	49.5	46	44.4
Biopsy	22	24	7	53
Resection	17	18	5	40
Chemotherapy	10	14	1	25
Radiation	4	6	1	11
Total	22	24	7	53

of diagnosing new patients based on their HRMAS spectra. These features are then traced back to metabolites that are known to appear in the chemical shift range corresponding to the regions that were identified. These metabolites represent the best set of discriminatory features and would therefore also be of interest for acquiring *in vivo* data that would contribute to assessing glioma grade.

In this study, robust feature selection and classification methods are combined to obtain accurate classifications with a parsimonious and interpretable set of features. The parsimonious set of features can be used to inform *in vivo* MRS acquisition, and hence to better characterize of tumor properties non-invasively.

4.1 Data acquisition

This study involved 53 patients who had previously been diagnosed with WHO grade 2 gliomas and were presenting for surgical resection due to suspected disease recurrence. The patients had received prior standard-of-care treatment with surgical resection, radiation, or chemotherapy. Table 4.1 provides information about the patient baseline characteristics and treatment prior to recurrence, broken down by grade of recurrence. The differences in baseline characteristics and prior treatment between patients with gliomas that recurred at different grades were not statistically significant.

Pre-surgical *in vivo* MR examinations enabled the planning of targeted biopsies for sampling tissue from patient lesions. Imaging parameters derived from post-processed data helped guide the designation of small (5 mm^3), putative tumor regions. Regions of suspected tumor located in relatively homogeneous areas of the MR images were designated as targets for tissue sampling using surgical navigation software.

Tissue samples were divided into two parts. One part was flash-frozen in liquid nitrogen and the other was fixed using conventional pathological techniques. The fixed component was examined

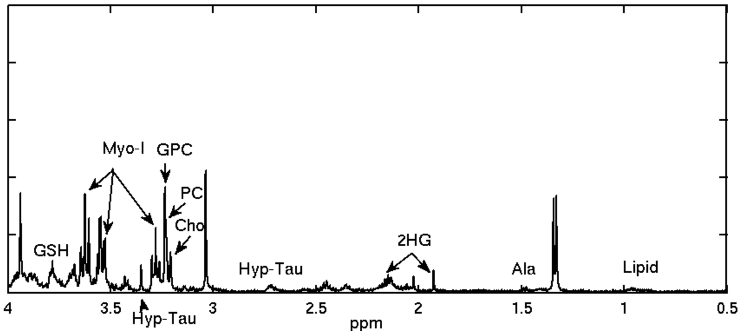


Figure 4.1: HRMAS spectrum for a recurrent grade 2 glioma

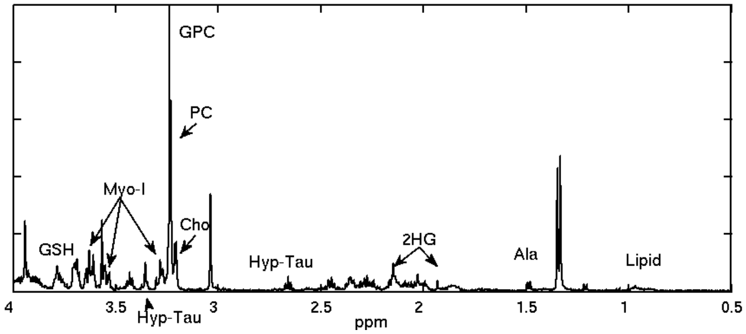


Figure 4.2: HRMAS spectrum for a recurrent LGG that transformed to grade 3

by a pathologist for histological features and for consistency with the tumor grade diagnosis. Only tissue samples that contained tumor cells and that were consistent with the diagnosis were included in the analysis. Histological analysis of tissue samples collected from the patients revealed that 7 tumors transformed to grade 4, 24 transformed to grade 3, and 22 remained grade 2.

The frozen part of the tissue sample was analyzed with *ex vivo* HRMAS. Tissue samples weighing between 0.78 and 28.14 mg (mean = 9.56 mg) were loaded into a 35-μl zirconium rotor (custom-designed by Varian) with 3 μl of 99.9% atom-D deuterium oxide containing 0.75 wt% TSP for chemical shift referencing. Data were acquired at 11.7 T, 1°C, 2250 Hz spin rate in a 4-mm gHX nanoprobe with a Varian INOVA 500 MHz multi-nuclear spectrometer. The nanoprobe gHX is an inverse probe, which was optimized for the direct detection of protons and the indirect detection of X-nuclei (e.g. 13C, 31P, 15N), and which was equipped with a magic-angle gradient coil. A rotor-synchronized 1D CPMG pulse sequence was run with TR/TE=4s/144 ms, 512 scans, 40,000 acquired points, and 20 KHz spectral width for a total time of 35 minutes. The 180° hard pulses were spaced 888 ms apart and synchronized to two turns of the rotor, resulting in 162 pulses.

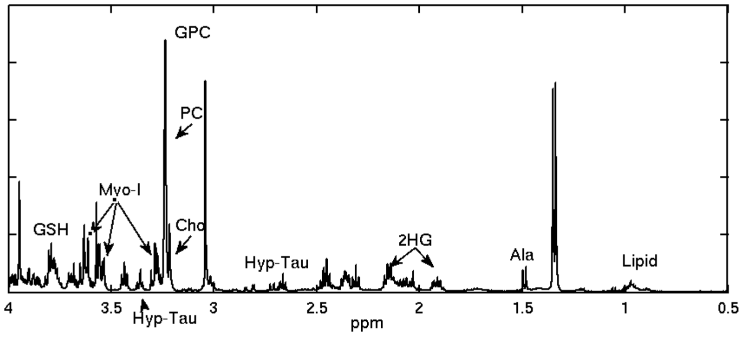


Figure 4.3: HRMAS spectrum for a recurrent LGG that transformed to grade 4

4.2 Analysis

Pattern analysis methods were used to build models capable of distinguishing between patients with recurrent LGGs that transformed to a higher grade and those that remained low grade.

The development of pattern recognition systems involves four major steps that can be identified as follows: preprocessing the data, reducing the data dimensionality, and selecting relevant features, constructing a classifier, and predicting its performance on previously unseen data.

Preprocessing

The raw free induction decay HRMAS signal was preprocessed using jMRUI [122]. The original time domain representation of the signal was transformed into the frequency domain using the fast Fourier transform. The frequency domain signal was shifted using TSP as a reference, and phased using zero-order phase correction. Residual water signal was then removed using Hankel-Lanczos singular value decomposition. Each data sample in the frequency domain was normalized using the ERETIC *in vivo* concentrations method [76] and the tissue weight of each sample.

The preprocessed data were grouped into frequency bins of widths 5, 10, 15, 20, and 25 samples to account for the fact that different metabolites have different linewidths. The value used for each bin was obtained using trapezoidal numerical integration for the 5, 10, 15, 20, or 25 samples corresponding to that bin. This resulted in an input vector with 18,266 dimensions. This method is equivalent to identifying areas of parts of the spectrum, and is therefore less susceptible to noise than using all the spectral data points.

Feature selection

Because one of the goals of this project was to obtain parsimonious models that can be easily interpreted biologically, non-linear dimensionality reduction methods or methods that transform the original features into a smaller set of mathematically related features that combine information from all the available data (such as PCA) were not employed. Instead, the dimensionality of the data was reduced by identifying regions in the HRMAS spectrum that have a mutual association with disease stage or tumor grade, using three feature ranking metrics that led to easily interpretable results. The three measures of association between features and classification output that we compared are based on the value of the chi-squared statistic with respect to the class [71], the information gain ratio with respect to the class [83], and a conditional probability-based technique that measures the mutual association between class decisions and feature values based on conditional probabilities [1]. The 40 highest-ranked features in the HRMAS spectra were obtained for each of the three association metrics, for each classification problem. Selecting 40 features was a heuristic loose enough to include most features with high and moderate levels of association, while keeping the second feature selection step computationally tractable.

A second feature selection step was performed in order to obtain a stable, parsimonious model capable of diagnosing new patients. A wrapper-based feature selection method was used to evaluate the suitability of subsets of features as a group, not just individually. The wrapper-based method [61] used genetic algorithms [43] to search through the space of possible feature sets and evaluate each subset of features by determining its performance when used in conjunction with a classification method. The genetic algorithm represented a solution as a binary vector that encoded whether a feature was included in the subset or not. LOOC accuracy was used to measure the fitness of a possible solution. At first, 50 individual solutions were randomly generated to form an initial population. During each successive iteration, a subset of the population was chosen to breed a new generation of solutions. Fitter solutions were more likely to be selected to breed. The new generation of solutions was obtained from pairs of parent solutions by combining the two parents using cross-over and mutation. Fifty iterations were performed, after which the highest ranked solution was selected.

Classification

The classification methods used to build diagnostic models were LDA [29] (which attempts to express the target class as a linear combination of predictors), logistic ridge regression [13] (which models the probability of occurrence of an event by fitting the data to a logit function), FTs [37] (which use decision nodes with multivariate tests and leaf nodes that make predictions using linear functions), SVMs with a polynomial kernel [95] (which project the data into a different space and separate positive and negative examples using a hyperplane with maximum margin), and decision stump boosting [36] (which sequentially applies decision stump classification to re-weight versions of the training data and then take a weighted majority vote of the sequence of classifiers thus produced). These methods were chosen because of their popularity and ease of interpretation.

Parsimonious diagnostic models based on very few relevant features were obtained after data preprocessing, dimensionality reduction, and model learning. The features chosen as part of the most discriminative subset for each model were then traced back to metabolites that are known to appear in the chemical shift range corresponding to the regions that were identified. These metabolites correspond to the best set of discriminatory features and their *in vivo* acquisition would be beneficial for assessing glioma grade non-invasively.

4.3 Validation

The models were validated using training accuracy, LOOC, bootstrapping, receiver operating characteristic (ROC) area under the curve (AUC), precision, recall, and F-measure. Algorithms 4 and 5 provide more details about the analysis pipeline and the number of data samples used at each step.

4.4 Results

Classification

In order to determine whether quantitative metabolic parameters derived from HRMAS data are predictive of malignant transformations in recurrent LGGs, the performance of diagnostic models for distinguishing between patients with recurrent gliomas that transformed to a higher grade and those that remained grade 2 were compared. The results of this analysis are illustrated in Tables 4.2 and 4.3. Logistic ridge regression and decision stump boosting models were able to distinguish between recurrent gliomas that transformed to a higher grade and those that did not with 100% training accuracy, 96% LOOC accuracy, and 96% bootstrapping accuracy. LDA, FTs, and SVMs were able to achieve LOOC accuracies above 90% and bootstrapping accuracies above 85%. The three feature ranking methods were comparable in performance. The pattern recognition methods for this classification task were further compared using a plot of precision and recall, illustrated in Figure 4.4, and ROC curves, illustrated in Figure 4.5. These curves show that the five classification methods are comparable, but the logistic ridge regression and decision stump boosting methods have a slight advantage. The difference in performance between logistic ridge regression and decision stump boosting and the other three methods is statistically significant, as can be seen from the non-overlapping 95% confidence intervals of the bootstrapping accuracy presented in Table 4.3.

An additional validation experiment was performed in order to address the concern that the high classification accuracies obtained may be due to overfitting caused by the large number of features explored. Labels of "transformed" and "not transformed" were randomly assigned to the HRMAS spectra, and bootstrapping accuracy results were obtained using the same methods that generated the best model in the previous analysis. Logistic ridge regression models were built based on the data with random labels. The features were filtered using information gain ratio. The best subset of features was then selected using a genetic wrapper-based search. The experiment was repeated 100 times, for different random generations of the labels. The average bootstrapping classification accuracy was 63% with a 95% confidence interval of [49-77%] (range 42-81%), significantly lower and more variable than the 96% bootstrapping accuracy with a 95% confidence interval of [95%-97%], obtained using the true labels. The random labels of the data sets that resulted in bootstrapping accuracies in the high end of the range actually had a significant number of labels that matched the true labels.

In order to assess the ability to distinguish between different degrees of disease malignancy based on HRMAS data, two other comparisons were performed. Statistical models were built

Table 4.2: Accuracy for distinguishing between recurrent LGGs which transformed to a higher grade and those that did not using five classifiers trained on HRMAS features selected using three feature selection methods

Accuracy	Features	LDA	Logistic	FT	SVM	Boosting
Training	Chi-square	0.93	1.00	0.93	0.94	1.00
	Gain ratio	0.85	1.00	0.93	0.94	1.00
	Conditional	0.94	1.00	0.92	0.94	0.98
LOOC	Chi-square	0.94	0.94	0.91	0.94	0.96
	Gain ratio	0.93	0.96	0.91	0.93	0.94
	Conditional	0.98	0.91	0.93	0.94	0.91
Bootstrap	Chi-square	0.87	0.96	0.89	0.85	0.96
	Gain ratio	0.84	0.96	0.89	0.83	0.94
	Conditional	0.89	0.95	0.89	0.83	0.94
ROC AUC	Chi-square	0.94	0.90	0.95	0.92	0.94
	Gain ratio	0.99	1.00	1.00	1.00	1.00
	Conditional	1.00	0.99	1.00	0.99	0.97
Precision	Chi-square	0.92	0.92	0.90	0.92	0.90
	Gain ratio	0.98	1.00	0.99	1.00	0.99
	Conditional	0.99	0.98	0.99	0.99	0.98
Recall	Chi-square	0.91	0.92	0.90	0.92	0.90
	Gain ratio	0.98	1.00	0.99	1.00	0.99
	Conditional	0.99	0.98	0.99	0.99	0.98
F Measure	Chi-square	0.91	0.92	0.90	0.92	0.90
	Gain ratio	0.98	1.00	0.99	1.00	0.99
	Conditional	0.99	0.98	0.99	0.99	0.98

to distinguish between recurrent LGGs that upgrade to grade 3 and those that remained grade 2 (100% training accuracy, 96% LOOC and bootstrapping accuracy), and between recurrent LGGs that transformed to grade 4 and those that transformed to grade 3 (100% training accuracy, 98% LOOC and bootstrapping accuracy). These comparisons also show a slight advantage when using logistic ridge regression and decision stump boosting.

Overall, the results suggest that metabolic parameters derived from HRMAS are predictive of malignant transformations in LGGs. Models that were built based on these parameters are specific and sensitive enough to be used for diagnosing individual patients, and do not merely reflect average differences between different patient groups. All classification and feature selection methods exhibit good performance, with logistic ridge regression and decision stump boosting slightly outperforming the other classification methods.

Table 4.3: 95% confidence intervals for accuracy results (presented in Table 4.2) for distinguishing between recurrent LGGs which transformed to a higher grade and those that did not using five classifiers trained on HRMAS features selected using three feature selection methods

Accuracy	Features	LDA	Logistic	FT	SVM	Boosting
Training	Chi-square	0.82-0.98	0.93-1.00	0.84-0.99	0.84-0.99	0.93-1.00
	Gain ratio	0.72-0.93	0.93-1.00	0.84-0.99	0.84-0.99	0.93-1.00
	Conditional	0.84-0.99	0.93-1.00	0.82-0.99	0.84-0.99	0.90-1.00
LOOC	Chi-square	0.84-0.99	0.84-0.99	0.79-0.97	0.84-0.99	0.87-1.00
	Gain ratio	0.82-0.98	0.87-1.00	0.79-0.97	0.82-0.98	0.84-0.99
	Conditional	0.90-1.00	0.79-0.97	0.82-0.98	0.84-0.99	0.79-0.97
Bootstrap	Chi-square	0.86-0.88	0.95-0.97	0.88-0.90	0.84-0.86	0.95-0.97
	Gain ratio	0.83-0.85	0.95-0.97	0.88-0.90	0.82-0.84	0.93-0.95
	Conditional	0.88-0.90	0.94-0.96	0.88-0.90	0.82-0.84	0.93-0.95
ROC AUC	Chi-square	0.88-1.00	0.82-0.98	0.89-1.00	0.85-0.99	0.88-1.00
	Gain ratio	0.96-1.00	1.00-1.00	1.00-1.00	1.00-1.00	1.00-1.00
	Conditional	1.00-1.00	0.96-1.00	1.00-1.00	0.96-1.00	0.93-1.00
Precision	Chi-square	0.78-0.98	0.78-0.98	0.74-0.98	0.78-0.98	0.74-0.98
	Gain ratio	0.89-1.00	0.89-1.00	0.89-1.00	0.89-1.00	0.89-1.00
	Conditional	0.89-1.00	0.89-1.00	0.89-1.00	0.89-1.00	0.89-1.00
Recall	Chi-square	0.71-0.99	0.77-0.99	0.71-0.97	0.77-0.99	0.71-0.97
	Gain ratio	0.85-1.00	0.85-1.00	0.85-1.00	0.85-1.00	0.85-1.00
	Conditional	0.85-1.00	0.85-1.00	0.85-1.00	0.85-1.00	0.85-1.00
F measure	Chi-square	0.75-0.99	0.78-0.99	0.72-0.98	0.78-0.99	0.72-0.98
	Gain ratio	0.87-1.00	0.87-1.00	0.87-1.00	0.87-1.00	0.87-1.00
	Conditional	0.87-1.00	0.87-1.00	0.87-1.00	0.87-1.00	0.87-1.00

Feature selection results

Although the direct identification of metabolites that are predictive of malignant transformations is not necessary to distinguish those patients who exhibit malignant transformations, analyzing the different metabolites that lead to high classification accuracy can lead to the detection of metabolic biomarkers and to the better understanding of tumor metabolism. Thus, one of the goals of this study was to model the metabolic transformations in gliomas in such a way that the results are easy to interpret biologically, and able to inform *in vivo* data acquisition. To this effect, feature selection results are presented. The features in this study represent normalized concentrations found in small chemical shift ranges. In order to interpret these features, their chemical shift ranges were linked to metabolites typically found in those spectral regions using the QUEST [100] quantification algorithm.

Table 4.4: Accuracy results for distinguishing between recurrent LGGs which upgraded to grade 3 and those that remained grade 2 using five classifiers trained on HRMAS features selected using three feature selection methods

Accuracy	Features	LDA	Logistic	FT	SVM	Boosting
Training	Chi-Square	0.91	1.00	0.89	0.96	1.00
	Gain Ratio	0.85	0.91	0.85	0.96	1.00
	Conditional	0.87	1.00	1.00	0.96	1.00
LOOC	Chi-Square	0.87	0.96	0.87	0.96	0.96
	Gain Ratio	0.91	0.87	0.72	0.94	0.96
	Conditional	0.89	0.96	0.94	0.91	0.94
Bootstrap	Chi-Square	0.85	0.95	0.86	0.83	0.95
	Gain Ratio	0.80	0.87	0.82	0.82	0.94
	Conditional	0.84	0.96	0.87	0.84	0.95

Table 4.5: Accuracy results for distinguishing between recurrent LGGs which upgraded to grade 4 and those that upgraded to grade 3 using five classifiers trained on HRMAS features selected using three feature selection methods

Accuracy	Features	LDA	Logistic	FT	SVM	Boosting
Training	Chi-Square	1.00	1.00	1.00	1.00	1.00
	Gain Ratio	1.00	1.00	0.79	0.97	1.00
	Conditional	1.00	1.00	1.00	0.97	1.00
LOOC	Chi-Square	1.00	1.00	1.00	1.00	0.96
	Gain Ratio	1.00	1.00	0.79	0.97	0.97
	Conditional	0.98	1.00	0.97	0.97	0.97
Bootstrap	Chi-Square	0.97	0.99	0.97	0.93	0.98
	Gain Ratio	0.96	0.97	0.79	0.90	0.96
	Conditional	0.97	0.98	0.79	0.90	0.97

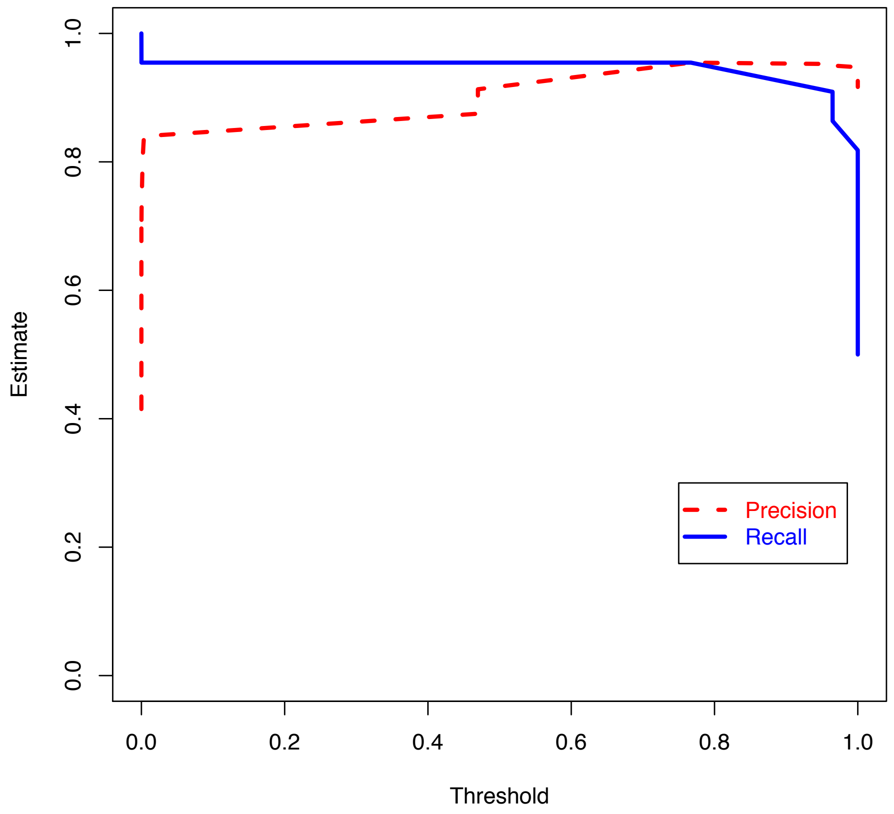


Figure 4.4: Plot of specificity and sensitivity for different thresholds of the logistic ridge regression model with features selected using the gain ratio filter and a genetic search wrapper-based method; the response variable is whether the tumor transformed to a higher grade

There was significant overlap in the 40 highest-ranking features selected using the chi-squared statistics, the information gain ratio, and the conditional probability-based association techniques. These features corresponded to nine metabolites. The highest-ranking of the 40 features were very similar across the three measures, even though the ranks differed slightly. Parameters corresponding to Myo-I, 2HG, Hyp-Tau, and Cho compounds were identified among the 40 highest-ranking features by all three association techniques, while parameters corresponding to GSH and Ala were identified only by some of the methods.

Table 4.6 shows the percentage of times each metabolite was identified by the logistic ridge regression genetic search wrapper-based feature subset selection algorithm as being part of the feature subset which was best at discriminating between tumor grades, during the bootstrapping process. Thus, for each comparison, up to 200 models were built on resampled versions of the original data set, and the percentages of times certain features were selected as part of these models are reported in Table 4.6, grouped by corresponding metabolite.

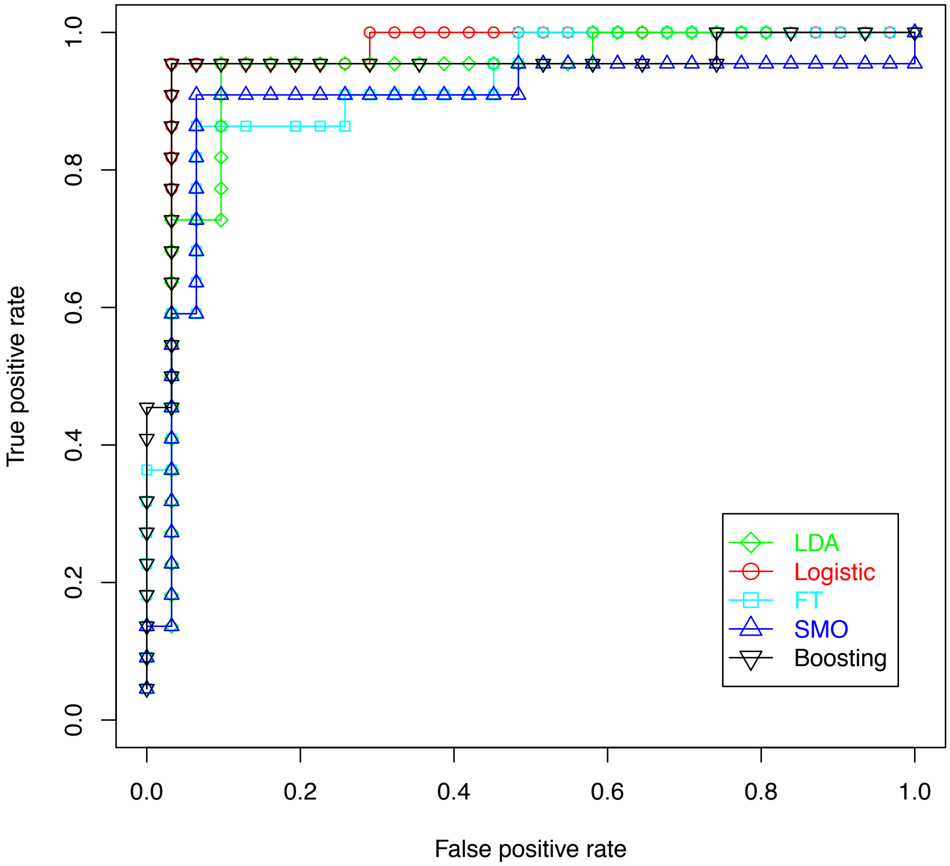


Figure 4.5: ROC curves comparing five classification methods used to distinguish between gliomas that transformed to a higher grade and those that did not, based on features selected using the gain ratio filter and a genetic search wrapper-based method

Table 4.6: Percentage of times each feature was selected in one of the models for distinguishing between low grade tumors that recurred at different grades

Metabolite		Transformed vs. not			Transf. to grade 3 vs. 2			Transf. to grade 4 vs. 3		
Name	Range	Gain ratio	Chi-sq.	Cond.	Gain Ratio	Chi-sq.	Cond.	Gain ratio	Chi-sq.	Cond.
Myo-I	3.51-3.54	100%	13%	78%	75%	48%	76%	33%	0%	0%
	3.61-3.68	96%	94%	85%	15%	0%	0%	0%	0%	0%
2HG	1.75-1.87	100%	78%	63%	26%	80%	54%	21%	64%	48%
	2.24-2.3	70%	92%	87%	67%	89%	28%	0%	0%	0%
Hyp-Tau	2.62-2.69	57%	99%	77%	75%	100%	100%	100%	94%	100%
	3.35-3.39	95%	99%	63%	95%	33%	45%	67%	73%	9%
Cho/PC/GPC	3.21-3.29	96%	100%	100%	78%	100%	100%	0%	75%	15%
	3.69-3.71	0%	53%	48%	26%	100%	50%	0%	0%	0%
Lip	0.96-1.00	100%	19%	100%	89%	0%	65%	0%	0%	0%
GSH	3.79-3.81	0%	97%	36%	100%	15%	100%	100%	78%	100%
Ala	1.4-1.5	0%	17%	0%	0%	57%	0%	88%	0%	3%

Distinguishing between recurrent LGGs that transformed to a higher grade and those that remained grade 2 was possible using areas of the *ex vivo* NMR spectrum corresponding to Myo-I, 2HG, Hyp-Tau, Choline, GPC, PC, GSH, and Lip. On average, the relative levels of Myo-I parameters were 56% lower in gliomas that transformed to a higher grade compared to those that remained grade 2. The gliomas that transformed to a higher grade also had 2HG levels in the $[2.24-2.3]$ chemical shift range that were 120% higher, 2HG levels in the $[1.75-1.87]$ range that were 30% higher, Hyp-Tau levels in the $[2.62-2.69]$ range that were 137% higher, Hyp-Tau levels in the $[3.35-3.39]$ range that were 57% higher, Cho compound levels in the $[3.21-3.29]$ range were 26% higher, Cho compound levels in the $[3.69-3.71]$ range were 83% higher, Lip levels that were 53% higher, GSH levels that were 39% higher, and Ala levels that were 25% higher.

Distinguishing between recurrent LGGs that transformed to grade 4 and those that recurred as grade 3 was possible using features corresponding to Hyp-Tau, GSH, Ala, and Cho. Another predictive feature shows increased activity in the 3.79 chemical shift range, where metabolites are very hard to quantify. Levels of Hyp-Tau, GSH, Cho, Myo-I, and 2HG were useful in distinguishing between gliomas that transformed to grade 3 and those that remained grade 2.

While some of the differences in metabolite levels between tumors that transformed to a higher grade and those that remained grade 2 are very large, individual metabolites are not sensitive or specific enough to distinguish between the two groups of lesions. Pattern recognition methods were able to objectively combine the information provided by all of these metabolites into a model that can accurately identify lesions that have undergone malignant transformations.

4.5 Discussion

The methods presented in this study were able to accurately detect malignant transformations in recurrent LGGs based on small sets of metabolites, without any prior knowledge. While the bootstrapping accuracies of the models created are very promising, the confidence in the models could be further strengthened by performing validation on an independent data set.

Distinguishing between LGGs that recurred at different grades was possible based on features corresponding to Myo-I, 2HG, Hyp-Tau, Cho compounds, GSH, Lip, and Ala. The fact that some features were selected in very different percentages of the models when using different filtering methods shows that that the features are highly correlated. Including all of these features in a model would be superfluous. In fact, the models created during the bootstrapping phase were based on 3 to 8 features each. This also explains why some metabolites identified as important biomarkers in the literature were not selected as features in these models. There are two reasons why a feature may not be selected by the algorithm: either because it is not well correlated with the outcome, or because it is highly correlated with some of the other features already in the model and it does not provide any additional information for the classification. Thus, the features identified by the pattern classification methods are a parsimonious set of metabolites that are specific and sensitive enough to lead to a quantitative formula that can be used to diagnose individual patients and can be applied to unseen data.

This analysis identified 2HG as a metabolite whose increased concentration is highly predictive of malignant transformations in recurrent LGGs, but whose concentration is generally low in newly diagnosed grade 4 gliomas. Recent studies have demonstrated that cancer-associated IDH1 mutations lead to the accumulation of 2HG [23]. These mutations have also been reported in acute myeloid leukemia (AML) [23], suggesting that a common mechanism may underpin both these AML and secondary grade 4 glioma malignancies. These studies offer novel opportunities to develop tumor directed therapeutic strategies.

Several of the metabolites selected by the pattern recognition methods for their discriminative power, such as Cho, Lip, and Myo-I, can be acquired using current *in vivo* MRS methods, implying that using pattern recognition methods in conjunction with *in vivo* spectroscopy may be successful in determining whether patients with recurrent LGGs have transformed to a more malignant phenotype without the need for a tissue diagnosis. Lip, Lac, Cho, NAA, and Cre are already being acquired in *in vivo* patient studies. Myo-I is one of the most important metabolites in terms of its discriminative power in all of the models explored. This metabolite can be acquired *in vivo*, and the results in this study suggest that it should be included in future *in vivo* acquisition protocols.

4.6 Conclusion

This study demonstrates the feasibility of using quantitative pattern recognition methods for the metabolic assessment of tissue samples obtained from brain tumor biopsies. The findings in this study enhance the knowledge obtained from previous HRMAS and MRS classification studies, because they suggest that it is possible to obtain high classification accuracy by using only a few spectral features obtained without any prior knowledge. The pattern recognition methods described in this paper identified biomarkers of importance in detecting malignant transformations in LGGs. Most of the metabolite parameters revealed by this method can be acquired *in vivo*. The use of MRS at high magnetic fields and with a robust classification approach should improve the characterization, typing, and prognostication of brain tumors. It can also be applied to assist in stratifying patients for appropriate therapeutic protocols and in the monitoring of new therapies.

Chapter 5

Metabolic profiles of primary and secondary glioblastomas

GBMs were first classified into primary and secondary tumors by Scherer [110]. Primary GBMs occur *de novo* without a recognizable precursor lesion. The clinical history of the disease is usually short, with a median survival of approximately one year following diagnosis. In contrast, secondary GBMs develop slowly and arise from preexisting, lower-grade astrocytomas [67, 110, 112].

The histological differentiation between primary and secondary GBMs is difficult. However, molecular genetic analyses suggest that there are at least two distinct pathways contributing to the tumorigenesis of GBM. Primary GBMs frequently show amplification and overexpression of the epidermal growth factor receptor (EGFR), phosphatase and tensin homolog (PTEN) mutations, and partial or complete loss of chromosome 10. Secondary GBMs generally feature mutations of the p53 or IDH1 genes, overexpression of platelet-derived growth factor receptors, and loss of heterozygosity at 17p, 19q, and 10q. In addition, they rarely overexpress EGFR [103].

Recent studies on LGGs have established a link between mutations in the IDH1 gene and excessive production of 2HG [23, 32]. This is an important finding because overwhelming clinical evidence now suggests that glioma patients harboring IDH1 mutations carry a significant survival advantage, irrespective of treatment therapies employed for disease management [49, 56, 92, 137].

Given the potential prognostic value of metabolites such as 2HG, the goal of this study was to characterize differences in the metabolic profiles between primary and secondary GBMs using HRMAS spectroscopy. Being able to accurately distinguish between primary and secondary GBMs based on their metabolic profiles can help identify the right course of treatment and predict the clinical behavior of these tumors [19].

Table 5.1: Accuracy results for distinguishing between primary and secondary GBMs using five classifiers trained on HRMAS features selected using three feature selection methods

Accuracy	Features	LDA	Logistic	FT	SVM	Boosting
Training	Chi-Square	0.98	1.00	0.95	0.98	1.00
	Gain Ratio	1.00	1.00	0.95	0.98	1.00
	Conditional	0.95	1.00	1.00	0.98	1.00
LOOC	Chi-Square	0.98	0.98	0.95	0.98	0.95
	Gain Ratio	1.00	1.00	0.95	1.00	0.98
	Conditional	1.00	0.98	0.98	0.98	0.95
Bootstrap	Chi-Square	0.96	0.98	0.95	0.98	0.98
	Gain Ratio	0.97	0.98	0.95	0.98	0.98
	Conditional	0.95	0.98	0.98	0.98	0.97

5.1 Data acquisition, preprocessing, feature selection, and analysis

This study included 36 newly diagnosed primary GBMs and 7 recurrent LGGs that progressed to grade 4. The secondary GBM patients had received prior standard-of-care treatment with surgical resection, radiation, or chemotherapy before the data acquisition.

The acquisition, preprocessing, feature selection, and analysis of the data were identical to those described in Chapter 4.

5.2 Results

Five classification models and three feature selection algorithms were used to distinguish between the metabolic profiles of primary and secondary GBMs. The results are summarized in Tables 5.1 and 5.2. The models were able to distinguish between primary and secondary GBMs with 98% bootstrapping accuracy, suggesting that the two types of tumors are indeed different entities. The features selected in the logistic ridge regression model are shown in Table 5.3. Myo-I, 2HG, Hyp-Tau, Cho compounds, and GSH provide enough information to accurately distinguish between primary and secondary GBMs.

5.3 Discussion

The ultimate goal of this work is to improve the clinical management of patients with gliomas. The data demonstrate the utility of HRMAS in distinguishing between primary and secondary GBMs. Myo-I, 2HG, Hyp-Tau, and GSH levels are sufficient to obtain good separation between the two

Table 5.2: 95% confidence intervals for accuracy results (presented in Table 5.1) for distinguishing between primary and secondary GBMs using five classifiers trained on HRMAS features selected using three feature selection methods

Accuracy	Features	LDA	Logistic	FT	SVM	Boosting
Training	Chi-square	0.91-1.00	0.95-1.00	0.88-0.99	0.91-1.00	0.95-1.00
	Gain ratio	0.95-1.00	0.95-1.00	0.88-0.99	0.91-1.00	0.95-1.00
	Conditional	0.88-0.99	0.95-1.00	0.95-1.00	0.91-1.00	0.95-1.00
LOOC	Chi-square	0.91-1.00	0.91-1.00	0.88-0.99	0.91-1.00	0.88-0.99
	Gain ratio	0.95-1.00	0.95-1.00	0.88-0.99	0.95-1.00	0.91-1.00
	Conditional	0.95-1.00	0.91-1.00	0.91-1.00	0.91-1.00	0.88-0.99
Bootstrap	Chi-square	0.95-0.97	0.97-0.99	0.94-0.96	0.97-0.99	0.97-0.99
	Gain ratio	0.96-0.98	0.97-0.99	0.94-0.96	0.97-0.99	0.97-0.99
	Conditional	0.94-0.96	0.97-0.99	0.97-0.99	0.97-0.99	0.96-0.98

Table 5.3: Percentage of time each feature was selected in the logistic ridge regression model for distinguishing between primary and secondary GBMs

Metabolite	Chemical Shift	Gain ratio	Chi-square	Conditional Probability
Myo-I	3.61-3.68	32%	32%	100%
2HG	2.24-2.30	67%	84%	33%
Hyp-Tau	3.35-3.39	82%	37%	10%
Cho/PC/GPC	3.21-3.29	100%	86%	38%
	3.69-3.71	40%	82%	72%
GSH	3.79-3.81	43%	43%	9%

glioma types. Since primary and secondary GBMs are indistinguishable histologically, this study also shows that HRMAS provides complementary information.

Chapter 6

Effects of treatment on the metabolic profile of grade 2 and grade 3 gliomas

LGGs include a diverse group of tumors, with distinct characteristics, patterns of occurrence, response to treatment, and survival. In clinical practice, magnetic resonance is used routinely for assessing the tumor status, therapy response, and progression free survival. However, differentiating between tumor progression and treatment effect using magnetic resonance imaging poses significant challenges [120]. For example, while contrast-enhanced MRI is often used to assess tumor progression and response to treatment, enhancement following treatment is not specific to tumor [120].

To address this issue, the objective of this study [17] is to investigate the differences in the metabolic profiles of newly diagnosed and recurrent grade 2 and 3 gliomas, using HRMAS spectroscopy. Identifying differences in metabolic parameters between these two groups is important in understanding how tumor progression affects the pathways of glioma growth and invasion, and in identifying biomarkers that may be associated with treatment effect. In order to assess whether treated and untreated gliomas of the same grade are different entities, and thus might have differential response to subsequent therapy, a multivariate pattern recognition method was used to discriminate between them based upon their metabolic profiles.

6.1 Data acquisition

This study included 137 patients: 86 were newly diagnosed non-enhancing grade 2 (n=55) and grade 3 (n=31) gliomas, and 51 were recurrent grade 2 gliomas that recurred as grade 2 (n=22) or grade 3 (n=29). The number of patients in each category is summarized in Table 6.1. A total of 349 image-guided biopsy samples were obtained from the 137 patients; the samples were then analyzed using HRMAS spectroscopy and histology. HRMAS data were obtained by scanning the biopsy tissues with a Varian 500 MHz spectrometer, equipped with a gHX gradient nanoprobe. Samples were evaluated at 1°C while the tissue was spun at 2250 Hz at the magic angle ($\theta = 54.7°$). The fully relaxed water presaturation sequence parameters were pulse width=7.8 mgrs, transients=128,

Table 6.1: Number of treated and untreated patients with grade 2 and 3 gliomas

Treated	Grade 2	Grade 3	Total
Yes	22	29	51
No	55	31	86
Total	77	60	137

sweep width=40 kHz, and 40,000 points. The ERETIC method was used to provide a constant reference for quantification.

6.2 Analysis

The *ex vivo* spectra were processed using jMRUI and a customized QUEST fitting algorithm to measure metabolite concentrations. A random effects model was used to compare metabolite concentrations in treated and untreated lesions, by grade. P-values were adjusted for multiple comparisons using Holm's method. A supervised learning algorithm was used to determine whether it is possible to accurately discriminate between the recurrent and newly diagnosed gliomas based on their metabolic profiles. Spectral parameters obtained from small chemical shift bins formed the input vectors, and the associated categories (newly diagnosed or recurrent) were the desired outputs or classes. The supervised learning method we used was multivariate logistic ridge regression with automatic wrapper-based feature selection, as described in Chapter 4. This method treated all biopsy samples as being independent. The supervised learning models were validated using 0.632 bootstrap accuracy estimates.

6.3 Results

A comparison of the metabolite levels in samples from patients with treated and newly diagnosed gliomas of grades 2 and 3 revealed significant differences in the myoinositol to total choline ratio (MCI) ($p =< 0.001$). The supervised learning model for distinguishing between these lesions had 90.25% accuracy (CI=$[89.69, 90.81\%]$) and was based on features of the spectra corresponding to Myo-I, Cho, aspartate (Asp), GSH, Cre, and Lac. Lower MCI levels ($p = 0.004$) and higher levels of threonine (Thr) ($p = 0.002$) and Asp ($p < 0.001$) were detected in recurrent grade 3 gliomas, compared to newly diagnosed grade 3 gliomas. The multivariate supervised learning model for distinguishing between recurrent and newly diagnosed grade 3 gliomas based on features of the spectra corresponding to Myo-I, Cho, GSH, NAA, and lactate had 92.45% accuracy (CI=$[89.75, 95.15\%]$). Lower MCI levels ($p = 0.003$) and higher GSH levels ($p = 0.014$) were detected in recurrent grade 2 gliomas compared to their newly diagnosed counterparts. The supervised learning model for distinguishing between recurrent and newly diagnosed grade 3 gliomas

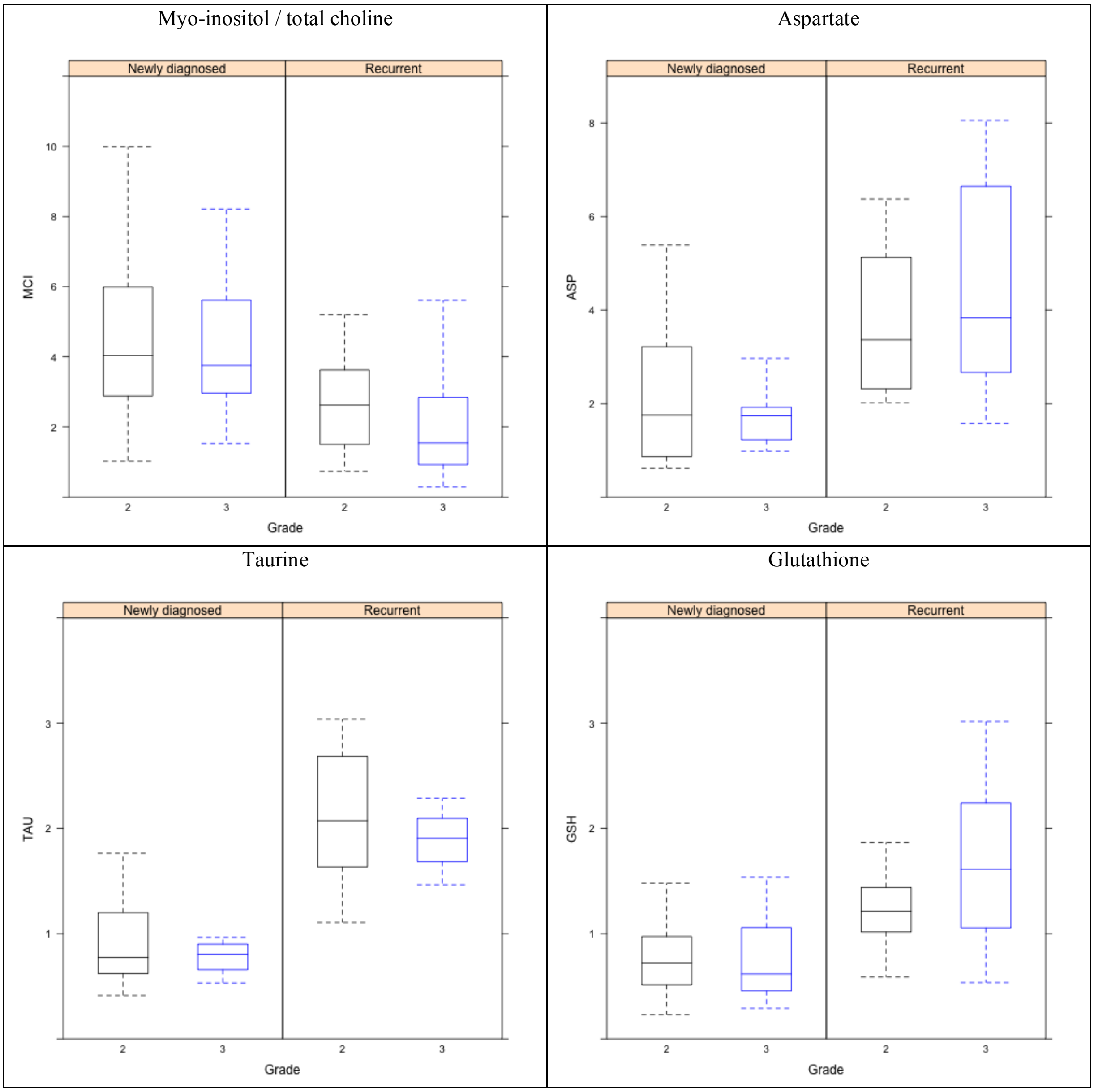


Figure 6.1: Metabolite levels in recurrent and newly diagnosed tumors, by grade

based on features of the spectra corresponding to Cho, GSH, Glc, Asp, 2HG, phosphoethanolamine (PE), and Lac had 94.44% accuracy (CI=$[92.2, 96.68\%]$).

A comparison of four metabolite levels in recurrent and newly diagnosed grade 2 and 3 gliomas is illustrated in Figure 6.1. Table 6.2 provides a summary of the bootstrapping accuracy of the logistic ridge regression models for distinguishing between treated and untreated glioma tissue

Table 6.2: Bootstrapping accuracy of logistic ridge regression models for distinguishing between treated and untreated glioma tissue samples

Grade	Boostrap Accuracy	Bootstrap 95% CI
2+3	0.90	0.89-0.91
3	0.92	0.90-0.95
2	0.94	0.92-0.97

Table 6.3: Percentage of times each feature was selected as part of the model for distinguishing between newly diagnosed and recurrent grade 2 and 3 gliomas

Metabolite	Chemical Shift	Percent Selected	Higher in Recurrent
Myo-I	3.55; 3.63	100%	No
PE	3.93-3.94	100%	No
Lac/Thr	1.32	100%	Yes
2HG	4.02	100%	Yes
	4.96	100%	Yes
Cre	3.04	7%	No

Table 6.4: Percentage of times each feature was selected as part of the model for distinguishing between newly diagnosed grade 3 gliomas and grade 2 gliomas that recurred at grade 3

Metabolite	Chemical Shift	Percent Selected	Higher in Recurrent
Myo-I	3.53-3.56; 3.62-3.64; 4.06-4.07	100%	No
PC/GPC/Cho	3.2-3.23	100%	Yes
Lac/Thr	1.31-1.32	100%	Yes
GSH	2.57	100%	No
Hyp-Tau	3.35	100%	No
NAA	4.53	83%	No

samples. Tables 6.3, 6.4, and 6.5 show the percentage of times each feature was selected as part of the final subset of features used for classification during the bootstrapping process.

Table 6.5: Percentage of times each feature was selected as part of the model for distinguishing between newly diagnosed grade 2 gliomas and recurrent grade 2 gliomas

Metabolite	Chemical Shift	Percent Selected	Higher in Recurrent
Myo-I	3.56	100%	No
PE	3.94-3.98	100%	No
2HG	2.23; 1.89; 4.02	94%	Yes
Glc	4.69	94%	Yes
Cho/PC/GPC	3.21-3.26	94%	Yes
Glx/GSH/Ala	3.77	93%	No
Lac	4.10-4.11	66%	Yes
Asp	2.78; 2.82	61%	No

6.4 Discussion

Grade 2 and 3 newly diagnosed gliomas had significantly different metabolic profiles than recurrent gliomas of the same grade. MCI, a biomarker associated with treatment effect [120], was significantly lower in recurrent lesions, regardless of grade. Multivariate supervised learning models were able to distinguish between treated and recurrent gliomas with more than 90% accuracy, suggesting that the metabolic profiles of these tumors are significantly different.

Chapter 7

Metabolic characterization of the histopathological properties of glioblastoma tissue

In vivo and *ex vivo* NMR spectroscopy provide useful information about the biochemical cellular processes that occur in the tissue under investigation. They offer an objective, clinically feasible metabolic assessment of disease. *In vivo* and *ex vivo* NMR spectroscopy have been used to separate normal brain from brain tumors [57, 104], as well as to characterize different degrees of malignancy [33, 77, 85, 88]. Although HRMAS is increasingly used to analyze tumor tissue biochemistry, histopathological analysis of surgically extracted tissue samples remains the gold standard for assigning the type and grade of a tumor. Because HRMAS spectra are obtained directly from biopsy samples, they can be used to identify correlations between metabolic profiles and histopathological properties of the same tissue. Combining these two types of analyses can provide further insights into the biology of brain tumors and improve the diagnosis process.

Most of the previous HRMAS studies focus on brain tumor characterization and typing, but do not describe the various histopathological tissue properties that play a role in the histological diagnosis and typing of brain tumors. However, there is an increasing interest in determining the correlation between metabolic features derived from *ex vivo* HRMAS and histopathological tissue properties [108], as HRMAS promises to provide an objective methodology for tumor classification and diagnosis. Previous studies have shown that HRMAS can help in predict the growth characteristics within different regions of a tumor [78] and can help distinguish between glioma grades [33]. Moreover, HRMAS can successfully reveal the status of tumor microheterogeneity [14] and can observe tumor metabolic alterations before they are morphologically detectable [2]. Nevertheless, these studies describe univariate metabolite behavior, while ignoring the interconnectivity of the metabolites.

In this study, multivariate pattern recognition methods were used to identify sets of features that provide the best sensitivity and specificity in distinguishing between low and high levels of histopathological features corresponding to tumor cellularity, proliferation, axonal integrity, and microvascular alterations. The purpose of the study was to determine whether supervised learn-

Table 7.1: Number of tissue samples with low and high values of histopathological tissue properties

Histology	Low	High
CA9	36	24
Complex	19	16
Hyperplasia	24	38
MIB-1	27	34
SMI-31	35	26
Tumor cellularity	27	35

ing methods based on HRMAS-derived parameters can successfully distinguish between different histopathological tissue properties. Additionally, this study aimed to extract representative metabolic tissue models for each histopathological tissue property and to analyze which of the considered input feature values are most sensitive in revealing whether a histological tissue property is present or absent.

This study analyzed the correlation between metabolic and histopathological tissue properties in newly diagnoses GBMs, because these gliomas show the highest degree of heterogeneity and the presence of the most malignant histopathological features.

7.1 Data acquisition and preprocessing

The study included 63 tissue samples obtained from 33 patients with newly diagnosed GBMs.

Pre-surgical *in vivo* MR examinations enabled the planning of targeted biopsies for sampling tissue from patient lesions. Imaging parameters derived from post-processed data helped guide the designation of small (5 mm^3), putative tumor regions. Regions of suspected tumor located in relatively homogenous areas of the MR images were designated as targets for tissue sampling using surgical navigation software.

Tissue samples were divided into two parts. One part was flash-frozen in liquid nitrogen and the other was fixed using conventional pathological techniques. The fixed component was examined by a pathologist for histological features such as tumor cellularity, MIB-1 proliferation index, SMI-31, CA9, hyperplasia, and the presence of simple, and complex vasculature. Only tissue samples that contained tumor cells were included in the analysis. Each histopathological tissue property was transformed into a binary variable describing either low or high values of that property. The number of samples in each category for each tissue property are listed in Table 7.1.

The frozen part of the tissue sample was analyzed with *ex vivo* HRMAS, as described in Chapter 4. The HRMAS data was also preprocessed according to the methods described in Chapter 4.

7.2 Analysis

Pattern analysis methods were used to build models capable of distinguishing between low and high levels of the various histopathological tissue properties.

The preprocessed data were grouped into frequency bins of widths 5, 10, 15, 20, and 25 samples to account for the fact that different metabolites have different linewidths. The value used for each bin was obtained using trapezoidal numerical integration for the 5, 10, 15, 20, or 25 samples corresponding to that bin. This resulted in an input vector with 18,266 dimensions. The dimensionality of the data was first reduced using the information gain ratio with respect to the class. A wrapper-based feature selection method using genetic algorithms for searching was used to evaluate the suitability of subsets of features as a group.

The classification method used to build metabolic models of the histopathological tissue properties was decision stump boosting. This method sequentially applies decision stump classification to re-weight versions of the training data and then take a weighted majority vote of the sequence of classifiers thus produced. It is an easy-to-interpret method that works well in the presence of distinct subgroups within each of the classes.

The models were validated using LOOC and bootstrapping.

7.3 Results

Metabolic models were built for each of the histopathological tissue properties. The accuracy results of these models are illustrated in Table 7.2, and the metabolic features selected in each model are listed in Table 7.3. The model for hyperplasia had 88% bootstrapping accuracy (95% CI=[87-89%]) and was based on features corresponding to Myo-I, Ala, lysine (Lys), ethanol (Eth), and Glc. The models for complex vasculature and tumor cellularity had 87% bootstrapping accuracy (95% CI=[86-88%]). The model for complex vasculature was based on metabolic features corresponding to Cre, PCr, PE, Myo-I, and Cho compounds, while that for tumor cellularity was based on GABA, Hyp-Tau, acetate (Ace), and Lys. High/low CA9 tissue samples were detected with 85% bootstrapping accuracy (95% CI=[84-86%]) and was based on Ala, Lac, Cho, Asp, Gln, and NAA. The model for low or high axonal disruption (SMI-31) had 81% bootstrapping accuracy (95% CI=[80-82%]), and was based on metabolic features corresponding to Hyp-Tau, Lac, Thr, GABA, and Ace. Lastly, tissue samples with low and high proliferation indices (MIB-1) were distinguished with 78% bootstrapping accuracy (95% CI=[77-79%]), based on features corresponding to Cre, Hyp-Tau, 2HG, Thr, GPC, and Myo-I.

7.4 Discussion

This study demonstrates the feasibility of using quantitative pattern recognition methods for the metabolic assessment of tissue samples obtained from brain tumor biopsies. The methods presented in this study were able to accurately classify tissue samples according to various histopathological properties. While the bootstrapping accuracies of the models created are very promising,

Table 7.2: Accuracy results of multivariate decision stump boosting models for distinguishing between low and high values of histopathological tissue properties

Histology	LOOC	LOOC 95% CI	Bootstrap	Bootstrap 95% CI
CA9	0.92	0.85-0.99	0.85	0.84-0.86
Complex	0.93	0.89-0.97	0.87	0.87-0.88
Hyperplasia	0.91	0.85-0.97	0.88	0.87-0.89
MIB-1	0.83	0.75-0.91	0.78	0.77-0.79
SMI-31	0.83	0.76-0.90	0.81	0.80-0.82
Tumor cellularity	0.93	0.88-0.98	0.87	0.86-0.88

Table 7.3: Feature subsets capable of distinguishing between low and high values of histopathological tissue properties

Histology	Metabolite	Chemical Shift	Percent Selected	Association
CA9	Ala	1.5	100%	negative
	Lac	4.1	95%	positive
	Asp	2.82	89%	positive
	Gln/NAA	2.48	90%	negative
Complex	Cre/PCr/PE	3.95	96%	negative
	My-I	3.29	67%	negative
	PC/GPC/Cho	3.23	97%	negative
Hyperplasia	My-I	3.64	76%	positive
	My-I	3.29	100%	negative
	Ala/Lys	1.56	96%	positive
	Eth/Glc	3.82	100%	positive
MIB-1	Cre	3.04	95%	negative
	hyp-Tau	2.64	77%	positive
	2HG	1.81	100%	positive
	Thr/GPC	4.3	99%	negative
	My-I	3.29	79%	negative
SMI-31	hyp-Tau/SI	3.36	96%	negative
	Lac/Thr	1.32	100%	positive
	GABA/Ace	1.92	94%	negative
TSCORE	GABA/Ace	1.91	100%	negative
	hyp-Tau	2.68	100%	negative
	Lys	1.62	93%	negative
	GABA	2.99	85%	positive

they do not fully explain the difference in histopathological properties. Metabolic features corresponding to Myo-I, Hyp-Tau, Cho compounds, Cre, Ala, Lac, and GABA are useful for distinguishing between tissue with low or high values of several histopathological parameters. These findings could aid in a better understanding of glial tumor tissue metabolism, and therefore could improve in diagnosis and prognosis of the tumor type and grade.

Chapter 8

Identifying recurrent low grade gliomas that transformed to a higher grade based on *in vivo* magnetic resonance and spectroscopy

Histopathological evaluation of tumor tissue remains the standard method for the diagnosis and classification of gliomas, but diagnostic challenges are well recognized, including subjective grading criteria, tissue sampling error, and lack of specific tumor markers. The objective of this study is to determine whether the quantitative parameters derived from metabolic and physiologic *in vivo* imaging characteristics are predictive of the biologic behavior of gliomas. This is an important clinical question because of the need to determine the malignancy of each tumor and to treat each patient with the therapy that is most likely to be effective for their particular lesion.

While some patients undergo biopsy or surgery to provide tissue for histological diagnosis, sampling error is a significant problem and having a non-invasive imaging technique to direct the surgeon to the most malignant region would be beneficial. For inoperable cases, imaging could provide an alternative method for predicting transformation to higher grade and establishing specific characteristics of the lesion.

The problem of determining whether a lesion has transformed to a more malignant phenotype based on metabolic and imaging parameters is of particular importance, because biopsies are not usually available at the time of progression. Having non-invasive biomarkers that are able to reliably define the spatial extent of recurrent tumor, predict whether the lesion has transformed to higher grade and provide an early indication of the success or failure of subsequent treatments would be extremely valuable for the management of these patients.

Although it is used for clinical evaluation of gliomas, standard MR provides ambiguous findings with respect to tumor grade and treatment effects. The work described in Chapter 4, using HRMAS to detect malignant transformations in recurrent LGGs, suggests that metabolic changes can be accurately used to predict transformation to a higher grade.

The objective of this study [20] was to build a multi-parametric diagnostic model that can predict whether gliomas originally diagnosed a being low grade have undergone malignant transformations to a higher grade, based on *in vivo* MR imaging parameters. This is an important clinical application because it will determine the therapeutic options offered to patients. In some cases this involves surgical resection to provide histological confirmation of tumor grade, but in other cases the preferred options are radiation, temozolomide, or standard chemotherapy. Having an objective, non-invasive model to predict tumor grade would allow patients to make an informed decision about the risks and benefits of different approaches and could have a major impact upon quality of life and survival.

8.1 Data acquisition

The study population included 61 patients with an original diagnosis of LGG, who were scheduled for surgical resection due to suspected recurrence. MR examinations were performed in order to provide data for planning image-guided surgery. The protocol included a FLAIR sequence, a T2 weighted fast spin echo (FSE) sequence, and a T1-GAD. These were used to define the spatial extent of the anatomic lesion. Physiological images that were acquired comprised of 6-directional diffusion tensor imaging (DTI) data with b=$1000mm^2$, and dynamic susceptibility contrast (DSC) images before, during, and after the injection of 0.1 mmol/kg body weight gadolinium diethyltriamine pentaacetic acid (Gd-DTPA) contrast agent at a rate of $3mL/sm$. Metabolic imaging data were obtained using lactate-edited three-dimensional MRS using point-resolved spectroscopy volume-selection with echo planar and phase encoding localization. Image-guided tissue samples were obtained during surgery and were used to confirm the diagnosis. Histological analysis indicated that 37 of the patients had lesions that progressed to a higher grade (31 to grade 3, and 6 to grade 4), and 24 patients had lesions that remained grade 2.

8.2 Preprocessing

The MR images were normalized using the median intensity value in the normal-appearing white matter and registered to the T1-gad. ADC andFA maps were obtained from the DWI sequence and CBV, PH, RF, and RECOV were estimated from the PWI susceptibility curves. The spectral amplitudes and line-widths of Cho, Cre, NAA, Lac, and Lip were estimated from the spectroscopy data. Four ROIs were created manually: the T2all, the CEL, the NEC, and the NEL. One hundred imaging parameters were extracted and used for analysis, including the volumes of the ROIs and of the regions with low ADC, high Cho to NAA ratio (CNI), and high CBV, as well as the median, 25th percentile, 75th percentile, and maximum intensity values of the imaging intensities and metabolite heights and areas inside the T2all.

8.3 Analysis

The problem of identifying lesions that progressed to a higher grade was formulated as a supervised learning problem. The input vector was comprised of the 100 imaging parameters, and the associated grade of recurrence (low grade or high grade) was the desired output or class. The goal was to build a supervised learning model to determine the probability that a lesion recurred at a higher grade for any valid input vector, after having seen a number of training examples, by generalizing from the presented training data to unseen situations in a reasonable way. The type of model we used was additive logistic regression, or boosting [36]. This method uses voting to combine the output of univariate regression models, which are built iteratively and in such a way that new models complement the previous ones by becoming experts for instances handled incorrectly by earlier models. Each models contribution is weighted by its performance. Suppose f_j is the j^{th} regression model in the ensemble and $f_j(a)$ is its prediction for instance a. Then, for a two-class problem, the probability estimate that a belongs to the first class is

$$p(1|a) = \frac{1}{1 + exp(\sum_j f_j(a))}[1]. \tag{8.1}$$

The variables used as input to the multivariate additive logistic regression model were selected using an automatic, wrapper-based feature selection method [134]. The final model was evaluated using LOOC and bias-corrected bootstrapping. These are the standard validation methods for supervised learning algorithms built on small data sets. LOOC repeatedly leaves out one data sample for testing, uses the rest of the samples for training the model, and then averages the accuracy results. It provides an accuracy estimate which is unbiased but has a high variance. Bootstrapping uses sampling with replacement to repeatedly split the data into a training set and a test set, and averages the accuracy results in order to obtain an overall accuracy estimate which is pessimistic, but has low variance.

8.4 Results

An additive logistic regression model based on automatically selected features was built to distinguish between tumors that recurred at a higher grade and those that remained grade two. This model had 93.02% LOOC accuracy with a 95% CI of $[87.84 - 98.20\%]$. The bias-corrected bootstrapping accuracy was 83.59% with a 95% CI of $[80.19 - 86.99\%]$. The model assigned a high probability of progression to lesions with the 75^{th} percentile of normalized choline height above 1.82, the 25^{th} percentile of percent recovery to baseline below 76, the 75^{th} percentile of CNI height above 1.25, and maximum choline height above 1.63. In order to compare the utility of different imaging sequences for identifying malignant transformations in recurrent LGGs, we built supervised learning models using sets of variables selected from spectral, perfusion, diffusion, and anatomical imaging parameters, each individually. The results are summarized in Tables 8.1 and 8.2. Selecting only spectral or perfusion parameters leads to models with lower accuracies than were obtained by selecting parameters from all imaging sequences, but these differences in accuracy are not statistically significant. Selecting only diffusion or anatomical parameters leads to models with significantly lower accuracy than the models using spectral and perfusion parameters. Table 8.1 also provides accuracy results for the univariate models that obtained more than 70%

Table 8.1: LOOC and bootstrapping accuracy of additive logistic regression models for identifying recurrent LGGs that progressed to a higher grade

Initial parameters	LOOC	LOOC 95% CI	Bootstrap Acc.	Bootstrap 95% CI
Model 1: All	0.93	0.88-0.98	0.84	0.80-0.87
Model 2: MRSI	0.86	0.77-0.95	0.79	0.75-0.82
Model 3: DSC	0.85	0.76-0.94	0.79	0.76-0.82
Model 4: DTI	0.76	0.66-0.86	0.69	0.67-0.71
Model 5: Anatomy	0.79	0.70-0.88	0.72	0.70-0.74
Model 6: Cho area 75%	0.67	0.56-0.79	0.76	0.73-0.79
Model 7: T1-gad 75%	0.62	0.50-0.73	0.75	0.73-0.77
Model 8: CNI height 75%	0.72	0.61-0.83	0.72	0.69-0.74
Model 9: T2all volume	0.58	0.47-0.69	0.71	0.69-0.73
Model 10: NEL volume	0.60	0.49-0.72	0.71	0.69-0.73
Model 11: CBV>2 Volume	0.79	0.70-0.88	0.70	0.68-0.73

Table 8.2: Feature subsets selected in the final decision stump boosting models which identify recurrent LGGs that progressed to a higher grade

Initial parameters	Parameters Selected
Model 1: All	RECOV 25%, Cho area 75%, CNI height 75%, Cho max height
Model 2: MRSI	CNI area >2 volume, CNI area median, Lac median, Cho area 75%
Model 3: DSC	RECOV 25%, PH 75%, CBV>3 volume, median PH
Model 4: DTI	FA median, ADC 75%
Model 5: Anatomy	T1-gad 25%, FLAIR 75%, T1-gad 75%

bootstrapping accuracy. These variables are the 75^{th} percentile Cho area, T1-gad intensity, and CNI height inside the T2all ROI, and the volume of the T2all, NEL, and CBV greater than two.

8.5 Discussion

Multivariate pattern recognition methods using parameters obtained from *in vivo* perfusion and spectroscopy data can identify recurrent LGGs that progressed to a higher grade with fairly good accuracy. High levels of choline and CNI, high intensity values in the contrast-enhancing lesion, low percent recovery to baseline, large lesion volumes, and large volumes of CBV greater than two are indicative of malignant transformations in recurrent LGGs. Choline-based parameters are associated with increased tumor cellularity, and a low percent recovery to baseline is indicative of

Table 8.3: Significance value of univariate parameters for logistic regression model which identifies recurrent LGGs that progressed to a higher grade

Parameter	P-value	Odds ratio (higher grade vs. grade 2)
T2all volume	0.006	1.05
NEL volume	0.008	1.05
T1-gad 75%	0.012	125.65
median T1-gad	0.038	45.83
ADC <125 volume	0.008	1.43
ADC <15 volume	0.011	1.14
median CBV > 2	0.013	1.29
CBV > 2 area	0.024	1.49
RECOV 25%	0.014	0.86
median RECOV	0.026	0.86
INVREC 75%	0.02	0.8
median INVREC	0.027	0.83
median CNI height	0.01	1.73
CNI height 75%	0.011	1.39
CNI area 75%	0.013	1.48
CNI area median	0.014	1.8
CNI area maximum	0.017	1.17
CNI height maximum	0.019	1.15
CNI height > 2 volume	0.02	1.08
CNI area > 3 volume	0.022	1.12
CNI area > 2 volume	0.025	1.08
CNI height > 3 volume	0.029	1.1
CNI area > 2 volume	0.041	1.1
Cho height 75%	0.017	5.08
Cho height maximum	0.017	2.11
Cho area maximum	0.019	1.97
Cho area 75%	0.025	4.57
median Cho height	0.028	7.76
median Cho area	0.042	6.89
median Lac area	0.037	2273.12
Cre height maximum	0.05	3.89

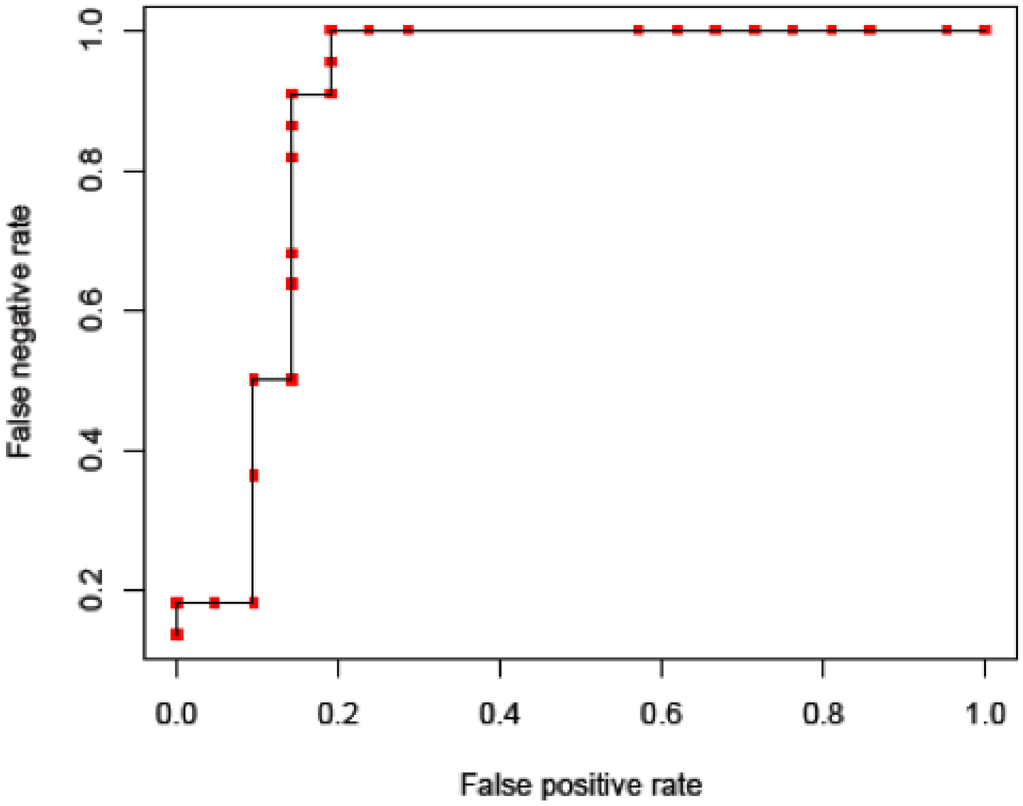


Figure 8.1: ROC curve of decision stump boosting model for non-invasively identifying whether a tumor transformed to a higher grade or not

leaky vasculature. The results of this study emphasize the importance of spectroscopy and perfusion data in detecting malignant transformations in LGGs. The multivariate additive logistic regression model built using these data can be used in clinical practice to identify patients suspected of recurrence at a higher grade before histological confirmation of the diagnosis is possible. While the *in vivo* parameters are influenced by the lesions histological subtypes, the parameters and cutoffs selected by the multivariate model are predictive of progression across all histological subtypes. An analysis of the model built with features selected from all of the parameters revealed that all astrocytomas and oligoastrocytomas with a 75^{th} percentile choline area above 1.065, as well as all oligodendrogliomas with a 75^{th} percentile choline area above 1.82, progressed to a higher grade. All lesions with a maximum choline height lower than 1.63, regardless of histological subtype, did not progress to a higher grade. For astrocytomas and oligoastrocytomas, a 25^{th} percentile recovery lower than 76 indicated that the lesions progressed to a higher grade. Progression to a higher grade was also indicated by 25^{th} percentile recovery lower than 76 in cases of oligodendromas with a 75^{th} percentile choline area below 1.82. Classification errors were most common among the oligodendroglioma subtype. A multivariate model built on the oligodendroglioma data alone revealed that a high volume of low ADC, low recovery to baseline, and low T1-GAD were predictive of malignant transformations.

Chapter 9

Linking *ex vivo* and *in vivo* glioma biomarkers

Tumor grading is a complex process that involves the search for histological features predictive of malignancy in biopsy specimens. Gliomas exhibit heterogeneous histopathological tissue properties, such as viable tumor cells, necrotic tissue or regions where the tumor infiltrates normal brain. Small image-guided biopsy samples allow for histological information to be related back to imaging parameters. However, due to the small size of the biopsies, it is very unlikely that one biopsy sample will contain all the histological features required for grading a heterogeneous brain tumor. Thus, being able to identify histological features of tissue using non-invasive MR data with high spatial resolution is instrumental in improving the biopsy selection process, in accurately assessing the malignancy of tumors, and in finding links between biological processes and non-invasive biomedical signals.

Many studies have been done to assess the correlation between *in vivo* MRI and MRS and various histological parameters [6, 25, 28, 50, 52, 65, 126, 127, 138]. These studies focus on linking pairs of *in vivo* and histological parameters, and do not attempt to obtain tumor profiles of histology estimates based on multiple data modalities. Other studies tried to asses the utility of *in vivo* magnetic resonance imaging and spectroscopy in determining the grade of a tumor [12, 27, 40, 44, 77, 82, 98, 113, 128, 139]. The main disadvantage of these studies is that they use features extracted from images of the whole brain in order classify tumors according to a subjectively obtained tumor grade, based on a specific biopsy location. Thus, they do not offer a direct link between images and histological features.

This study analyzed the relationship between parameters of brain tumors obtained through *in vivo* MRI, *in vivo* MRS, and *ex vivo* IHC. Multivariate pattern recognition methods were used to build models of histological features based on *in vivo* parameters. The successful integration of histology and *in vivo* data through the use of multivariate pattern analysis can lead to quantitative non-invasive markers of biological behavior, which are extremely important for characterizing tumor heterogeneity, for defining the most appropriate regions for obtaining image guided samples, and for evaluating patients for whom surgical resection is not appropriate. The successful prediction of histological features from imaging data could thus lead to a non-invasive "virtual biopsy"

Table 9.1: Number of tissue samples with low and high values of histopathological tissue properties.

Histology	Low	High
Tumor cellularity	37	81
Hyperplasia	30	87
Complex Vasculature	74	42
Simple Vasculature	56	61
MIB-1	80	38
SMI-31	49	67
CA9	57	47

with high spatial resolution.

9.1 Data acquisition

This study included 118 tissue samples from 52 patients with newly diagnosed GBMs.

MR examinations were performed in order to provide data for planning image-guided surgery. The protocol included a FLAIR sequence, a T2 weighted FSE sequence, and a T1-GAD sequence. These were used to define the spatial extent of the anatomic lesion. The physiological images that were acquired comprised of 6-directional DTI data with b=$1000mm^2$, and DSC images before, during, and after the injection of 0.1 mmol/kg body weight Gd-DTPA contrast agent at a rate of $3mL/sm$. Metabolic imaging data were obtained using lactate-edited three-dimensional MRS using point-resolved spectroscopy volume-selection with echo planar and phase encoding localization. Imaging parameters derived from post-processed data helped guide the designation of small (5 mm^3), putative tumor regions. Regions of suspected tumor located in relatively homogeneous areas of the MR images were designated as targets for tissue sampling using surgical navigation software.

Tissue samples were fixed using conventional pathological techniques and examined by a pathologist for histological features such as tumor cellularity, MIB-1 proliferation index, SMI31, CA9, hyperplasia, and the presence of simple, and complex vasculature. Only tissue samples that contained tumor cells were included in the analysis. Each histopathological tissue property was transformed into a binary variable describing either low or high values of that property. The number of samples in each category for each tissue property are listed in Table 9.1.

9.2 Preprocessing and analysis

The MR images were normalized using the median intensity value in the normal-appearing white matter and registered to the T1-GAD. ADC and FA maps were obtained from the DWI sequence and CBV, PH, RF, and RECOV were estimated from the PWI susceptibility curves. The spectral amplitudes and line-widths of Cho, Cre, NAA, Lac, and Lip were estimated from the spectroscopy data.

The problem of identifying image voxels corresponding to a high or low histopathological feature was formulated as a supervised learning problem. The input vector was comprised of the imaging parameters at the biopsy locations, and the associated histopathological feature (low or high) was the desired output or class. The goal was to build supervised learning models to determine the probability that a voxel corresponds to a low or high level of the histopathological parameter of interest, for any valid input vector, after having seen a number of training examples, by generalizing from the presented training data to unseen situations in a reasonable way. The type of model used was decision stump boosting. This method uses voting to combine the output of univariate regression models, which are built iteratively and in such a way that new models complement the previous ones by becoming experts for instances handled incorrectly by earlier models. The contribution of each model was weighted by its performance.

The variables used as input to the multivariate additive logistic regression model were selected using an automatic, wrapper-based feature selection method [134]. The final model was evaluated using LOOC and bias-corrected bootstrapping.

9.3 Results

The globally predictive features for distinguishing between low and high values of tumor cellularity, hyperplasia, complex and simple vasculature, MIB-1, SMI-31, and CA9 are listed in Table 9.2.

Decision stump boosting models based on automatically selected feature subsets were built to distinguish between low and high values of tumor cellularity, hyperplasia, complex and simple vasculature, MIB-1, SMI-31, and CA9. The accuracies of these models are illustrated in Table 9.3. It is possible to distinguish between low and high tumor cellularity samples and low and high complex vasculature samples with 86% bootstrapping accuracy (95% CI=[83-89%]). Tissue samples with microvascular hyperplasia can be distinguished *in vivo* with 84% bootstrapping accuracy (95% CI=[81-87%]). Tissue samples with high axonal disruption were identified with 83% bootstrapping accuracy (95% CI=[80-86%]). Samples with low or high MIB-1 proliferation index were classified with 79% bootstrapping accuracy (95% CI=[76-82%]). Low and high CA9 samples were distinguished with 75% bootstrapping accuracy (95% CI=[71-79%]).

Table 9.2: Globally predictive features for distinguishing between low and high values of the histopathological parameters of GBM tissue

Histopathology	Parameter	p-value	Association
Tumor cellularity	T1-GAD	0.03	positive
	CBV	0.02	positive
	Choline	0.01	positive
	Excess choline to NAA	0.009	positive
	Excess choline to Creatine	0.003	positive
	CNI	0.009	positive
Hyperplasia	Peak height	0.04	positive
	Lactate	0.03	positive
	RECOV	0.01	negative
	CBV	0.07	positive
Complex vasculature	RECOV	0.05	negative
	Peak height	0.06	positive
Simple vasculature	FA	0.01	negative
	CBV	0.05	positive
	Creatine	0.07	negative
MIB-1	T1-GAD	0.04	positive
	Peak height	0.02	positive
	CBV	0.0004	positive
	Excess choline to creatine	0.003	positive
	Lactate	0.09	positive
SMI31	RECOV	0.01	negative
	T1-GAD	0.02	positive
	Peak height	0.02	positive
	NAA	0.08	negative
CA9	RECOV	0.02	negative
	Lactate	0.03	positive
	T1-GAD	0.05	positive

Table 9.3: Accuracy results of decision stump boosting models for distinguishing between low and high values of the histopathological parameters of GBM tissue

Histopathology	LOOC	LOOC 95% CI	Bootstrap	Bootstrap 95% CI
Tumor cellularity	0.94	0.88-1.00	0.86	0.83-0.89
Hyperplasia	0.90	0.82-0.98	0.84	0.81-0.87
Complex Vasculature	0.86	0.79-0.93	0.86	0.83-0.89
Simple Vasculature	0.82	0.72-0.92	0.79	0.76-0.82
MIB-1	0.85	0.77-0.93	0.79	0.76-0.82
SMI31	0.84	0.74-0.94	0.83	0.80-0.86
CA9	0.81	0.71-0.91	0.75	0.71-0.79

9.4 Discussion

Multivariate pattern analysis models were able to distinguish between low and high histopathological tissue properties with good bootstrapping accuracies. Tumor cellularity, hyperplasia, complex and simple vasculature, and SMI31 were easier to classify than MIB-1 and CA9. While not accurate enough for diagnosis, these models could be used to guide biopsy site selection, by targeting areas of the tumor with high values of malignant histopathological features. This process could reduce sampling error and lead to more accurate histopathological diagnosis and grading.

Chapter 10

Modeling tumor cellularity by linking MR imaging and spectroscopy with histology [21]

The study presented in this chapter was published in the Proceedings of the 2010 MICCAI Workshop on Computational Imaging Biomarkers for Tumors [21].

This study extends the analyses performed in Chapter 9 by analyzing the relationship between parameters of brain tumors obtained through *in vivo* MRI and MRS, and *ex vivo* IHC, helped by the correlation between various histopathological features. The goal of this project was to provide a quantitative definition of tumor cellularity based on the *in vivo* parameters. Biopsy samples obtained from previously untreated patients with a diagnosis of GBM were used to find the link between imaging parameters at specific biopsy locations and IHC parameters from the corresponding tissue samples. A logistic ridge regression model of tumor cellularity was learned from the *in vivo* parameters and the remaining histological parameters. The tumor cellularity model was then tested on examples which contained only *in vivo* parameters, by first estimating the remaining IHC parameters by applying the EM algorithm, and then using the complete parameter vector for classification.

10.1 Data acquisition and preprocessing

Fifty-two patients received full MR examinations on a 3T MR scanner. The MR examination included a three-dimensional T1-GAD and a three-dimensional FLAIR sequence, acquired at a resolution of $1mm \times 1mm \times 3mm$. The exam also included 6-directional axial DWI with b=$1000s/mm2$ and dynamic PWI with an injection of 0.1 mmol/kg body weight Gd-DTPA contrast agent at 5mL/s and lactate-edited three-dimensional MRS data. Chemical shift imaging was performed using point-resolved spectroscopy volume-section techniques ($10mm \times 10mm \times 10mm$ nominal spatial resolution). Water suppression was achieved either through the use of spectral-spatial spin-echo pulses or CHESS and outer volume suppression was performed using PRESS

[77]. The spectral amplitudes and line-widths of Cho, Cre, NAA, Lac, and Lip were estimated. All images were rigidly registered to the T1-GAD image. The registration was performed through the maximization of normalized mutual information using a gradient ascent algorithm [123].

Tissue sample locations were selected based on surgically accessible areas with low ADC, elevated CNI, or elevated PH and low RECOV. Upon surgical excision, the tissue samples were flash-frozen in liquid nitrogen. For analysis, the tissue samples were fixed in paraffin and analyzed by a pathologist, who scored the sample for tumor cellularity (on a scale from 0 to 3), evaluated the MIB-1 index based on the number of cells stained with Ki-67 antibody relative to the total cell count, calculated the axonal integrity score (SMI-31), and noted the presence or absence of simple, complex, and delicate vasculature, hyperplasia, and CA9 gene expression.

The data set consists of 118 tissue samples obtained from 35 patients.

10.2 Feature selection

Each tissue sample had ten associated IHC parameters which were the imaging parameters at the biopsy locations obtained by creating a $5mm$ biopsy mask using the recorded biopsy location of each tissue sample. The median intensity values inside this mask in each of the MR and MRS images were used as the *in vivo* parameters. A total of seventeen *in vivo* parameters were used: two anatomical parameters, two diffusion parameters, six perfusion parameters, and seven spectral parameters. A low/high cellularity variable was created for classification, by including tumor scores of less than 2 into the low cellularity category, and the rest into the high cellularity category. A wrapper-based feature selection method with genetic search was used to select an optimal subset of features for the classification.

10.3 Models

The problem of classifying data samples into low and high tumor cellularity samples was formulated as a supervised learning problem. The *in vivo* and IHC parameters of each tissue sample were used to form a feature vector, or input. The associated tumor cellularity class was the desired output. The task of the supervised learner was to determine the tumor cellularity class for any valid input vector, after having seen a number of training examples, by generalizing from the presented data to unseen situations in a reasonable way. The EM algorithm for multiple imputations was first used to estimate the missing values in the training data. A logistic ridge regression classifier was then used to learn a model of tumor cellularity based on the *in vivo* MRI and MRS parameters, as well as the remaining histological parameters. The model was then used to predict the tumor cellularity of unseen examples, either using the full input vector, or only the *in vivo* parameters. In the latter case, the EM algorithm for multiple imputations was used to estimate the histological parameters, and the resulting input vector is used for classification.

Table 10.1: Training and LOOC accuracy results for binary and three-category tumor cellularity models using different sets of parameters

Features	Binary Cellularity		Three-Categ Cellularity	
	LOOC	Bootstrap	LOOC	Bootstrap
in vivo + hist	0.95	0.94	0.82	0.77
hist	0.92	0.90	0.73	0.69
MRS + hist	0.88	0.87	0.70	0.65
Perf + hist	0.89	0.87	0.71	0.68
Diffu + hist	0.90	0.86	0.77	0.70
Anat + hist	0.88	0.85	0.73	0.68
in vivo	0.94	0.86	0.43	0.34

10.4 Results

The data was first preprocessed and the missing values were estimated using the EM algorithm. The log likelihood threshold for convergence was set to 10^{-4}. The logistic ridge regression classifier was used to learn a model for classifying low and high tumor cellularity based on the remaining histological parameters and the *in vivo* parameters.

The model was first tested using complete input vectors. Table 10.1 shows that the full *in vivo* and histology model was able to distinguish between low and high tumor cellularity with 94% bootstrapping accuracy. This model outperformed the model based on *in vivo* parameters alone. The full model was based on seven parameters: SMI-31, the presence of complex vasculature, the excess of Cho to Cre, the RF and PH of the $\Delta R*$ perfusion curve, and the intensity values of the FLAIR and T1-GAD images at the biopsy locations. The histology parameters were, in general, the most important in predicting tumor cellularity. This was expected, because the imaging parameters came from an area around the estimated biopsy location, and might not have exactly corresponded to the tissue samples themselves.

Table 10.2 shows that the information obtained from the histology parameters can be used to improve tumor cellularity classification even when histological data is not available. Using the full model to predict tumor cellularity based on *in vivo* parameters and EM estimates of the "missing" histological parameters yields a significantly higher prediction accuracy compared to the *in vivo* model. Learning the structure of the model from histology and using the *in vivo* parameters to estimate the histological data gives better performance than trying to estimate tumor cellularity using the *in vivo* parameters directly.

The three-category tumor cellularity model built a regression function for each class. The accuracy of the three-category model was summarized in Table 10.1. The full model was able to predict tumor cellularity with 77% accuracy, given a three-category response. The confusion matrix for the full model was shown in Table 10.3. Most of the instances were classified correctly,

Table 10.2: Accuracy of binary cellularity model with EM estimated histology

Training Data	Test Data	Bootstrap Acc
in vivo + hist	*in vivo* + hist	0.94
in vivo + hist	*in vivo* + hist EM estimates	0.9
in vivo	*in vivo*	0.84

Table 10.3: Confusion matrix for full data three-category cellularity model

Class. 1	Class. 2	Class. 3	Actual
32	4	0	1
4	21	0	2
0	8	42	3

and the classification errors that occurred were always between two adjacent classes.

In addition to testing the accuracy of the model on examples with known classes, the model was also used to predict the cellularity of every voxel inside the manually defined T2all of an MR exam. The results are illustrated in Figure 10.1. The model classified a large part of the voxels in the center of the abnormal region as being high cellularity. After performing morphological operations on the resulting mask, we notice a center of high cellularity, with lower cellularity outwards. The amount of data available for estimating the joint distribution of *in vivo* and histological parameters was relatively small given the number of histological parameters that need to be estimated. Also, the data were not randomly selected and came from a relatively limited distribution of parameters. This explains some of the noise in the original high cellularity mask. Obtaining more data from random locations in the abnormality region is necessary in order to be able to more accurately extrapolate tumor cellularity values beyond the biopsy locations.

10.5 Discussion

This project provided a framework for learning a quantitative *in vivo* definition of tumor cellularity. The model proposed was able to accurately classify data instances into tumor cellularity categories using a set of *in vivo* and histological parameters, as well as the *in vivo* parameters and estimates of histological parameters. Using estimates of other histological parameters provided better accuracy results than directly predicting tumor cellularity based on only the *in vivo* parameters. This suggests that histological parameters are highly correlated, and that *in vivo* parameters provide information about all of the histological parameters. This information can be used to improve the accuracy of individual histological models. The pattern recognition framework described in this

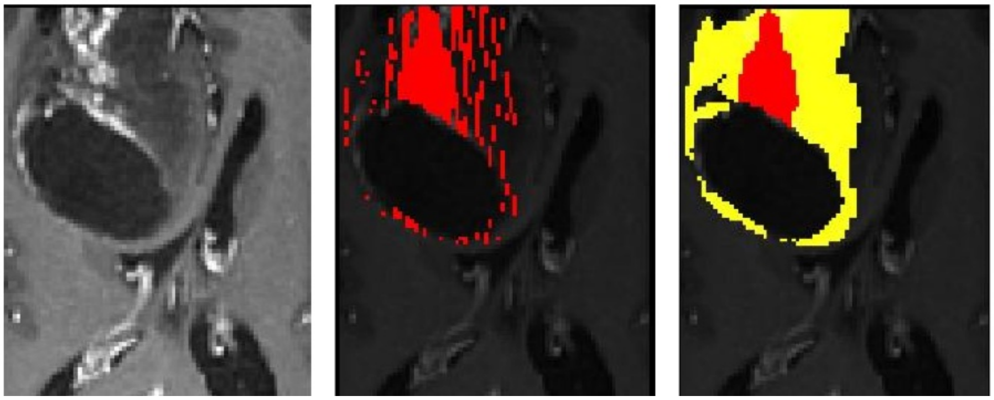


Figure 10.1: Tumor cellularity predictions superimposed on a T1-GAD image; (Left): T1-GAD images; (Middle): the high cellularity mask in red; (Right): tumor cellularity masks (red = high, yellow = low), corrected using morphological closing

project was able to identify *in vivo* parameters that were instrumental in predicting high tumor cellularity in all models, such as high Cho and low RECOV. This framework was used to create predicted maps of tumor cellularity at the spatial resolution of anatomical images provided that training data covering the entire distribution of parameters in the region of interest is available. Building such a model could lead to a better biopsy selection process or to the early prediction of tumor cellularity and grade before biopsies are acquired and processed.

Chapter 11

Multivariate serial analysis of glioblastomas

The median survival of patients with GBMs is approximately 15 months [125]. However, there is considerable variability among individuals, with some benefiting from different types of treatment. Previously published research revealed several factors that influence the prognosis of patients with GBMs [10, 34, 35, 51, 53, 62, 121, 124]. Preoperative and postoperative MRI characteristics have been reported as having predictive value in relation to survival [22, 48, 68, 105]. However, individual MRI characteristics are not sensitive and specific enough to provide an accurate prognosis estimate for individual patients. This study presents a multivariate model of survival based on serial MR parameters that could be used to predict individual patients' prognoses. A univariate analysis of the data used in this study was previously published by Yan et al. [68].

11.1 Data acquisition

Sixty-four patients with newly diagnosed primary GBMs participated in this study. They had a median age of 53 (range 27-77). Thirty-eight were male, and 26 were female. Thirty-six were on corticosteroid therapy, while 21 were not. Sixty-three patients underwent subtotal or gross total resection, while three patients only had biopsies. Patients had their baseline scans after surgical resection or biopsy, but prior to radiation and chemotherapy. Subsequent scans were acquired immediately after radiation and every two months afterwards, up to a maximum of one year. When tumor progression was suspected, patients received an additional scan at approximately one month, to help distinguish true progression from pseudoprogression.

MR data were acquired using either 1.5T or 3T GE scanners (GE Healthcare Technologies) between November 2002 and March 2007. Anatomic MR imaging included a T1-weighted image (repetition time [TR]/echo time [TE] = $54/2ms$), a T2-weighted FLAIR image (TR/TE/inversion time [TI] = $10002/127 - 157/2200ms$), and pre- and post-contrast T1-weighted images (TR/TE = $26/2 - 8ms$). DWIs were acquired with 3- or 6-directional axial diffusion echo planar imaging (EPI) sequences (TR/TE = $5000 - 10000/63 - 110ms$, b = $1000s/mm^2$). A bolus of $0.1\ mmol/kg$ of body weight Gd-DTPA was injected intravenously at the speed of $5mL/s$. PWIs were achieved by collecting a series of T2-weighted EPI (TR/TE/Flip = $1000 - 2000/54ms/30 - 358$, 60-80 timepoints) acquired before, during, and after the bolus injection.

11.2 Preprocessing

The anatomical images were aligned to the corresponding T1-GAD image. ADC maps were calculated from theDWI, resampled to the same resolution as the T1-GAD, and then rigidly aligned to it. Perfusion datasets were resampled to the same resolution as the T1-GAD and then nonrigidly aligned to it. CBV, RECOV, PH, and RF were calculated for each voxel using software developed in the Surbeck Laboratory for Advanced Imaging. CBV intensities were obtained by fitting the dynamic perfusion data by a modified gamma-variate function with a recirculation parameter. PH and RECOV values were estimated using a simple nonparametric procedure.

ROIs included the NAWM, the CEL, the T2all, and the NEL, which was defined as the T2all lesion subtracted from the CEL and NEC (Figure 2.2). NAWM was segmented automatically using a hidden Markov random field model with an EM algorithm on the pre-contrast T1-weighted image. The masks of tumor lesions were segmented using a region-growing segmentation tool. All ROIs excluded the resection cavity. Volumetric, diffusion, and perfusion parameters were analyzed within these abnormalities. Due to smaller acquisition coverage of DWI and PWI compared with the anatomic images, these parameters were limited to the region that overlapped. To emphasize any difference between patients, ADC, CBV, PH, and RF maps were normalized to the median value within NAWM. The value of RECOV is calculated as a percentage of the baseline and was therefore not normalized.

11.3 Analysis

Figure 11.1 shows the patient survival information and time to progression. The median overall survival (OS) was 588 days (95% CI = [468-708]), with 11 patients sensored. Figure 11.2 illustrates the Kaplan-Meier curve for overall survival.

Table 11.1 shows the significant univariate predictors of overall survival. These include the volumes of the T2all, the NEL, and the CEL ($p < 0.0001$), and parameters and extracted from the ADC, RF, and PH.

A multivariate model was built using stepwise selection of covariates, using the following steps:

S1: Fit model containing all variables significant at the 0.2 level in the univariate analysis.

S2: Sort p-values of coefficients and use them to delete insignificant variables one at a time.

S3: Re-add removed covariate if removal procedure causes a change of more than 30% in the coefficients.

S4: Try to add, one at a time, each variable not yet included in the multivariate model, and test if it is significant or an important confounder (it changes coefficients more than 30%).

Table 11.1: Significant univariate predictors of overall survival

VAR	COEF	EXP(COEF)	STANDARD ERROR		P-VALUE	
			NON-ROBUST	ROBUST	NON-ROBUST	ROBUST
t2all volume	0.01	1.01	0.003	0.002	0.0000102	0.00000008
nel volume	0.01	1.01	0.003	0.003	0.0000294	0.0000007
cel volume	0.04	1.04	0.01	0.01	0.000865	0.000303
median adc in cel	-1.86	0.16	0.74	0.581	0.0117	0.00142
90th adc in cel	-0.95	0.39	0.37	0.36	0.011	0.00854
10th adc in nel	2.95	19.15	1.13	1.16	0.00906	0.0112
10th adc in cel	-1.71	0.18	0.85	0.69	0.044	0.0135
10th adc in t2all	2.92	18.5	1.12	1.18	0.00913	0.0137
median adc in t2all	1.66	5.28	0.71	0.74	0.0184	0.0241
median adc in nel	1.51	4.54	0.68	0.72	0.0257	0.0347
median rf in t2all	1.15	3.17	0.65	0.42	0.0772	0.00642
90th rf in t2all	0.13	1.14	0.11	0.05	0.246	0.01
90th rf in nel	0.12	1.13	0.12	0.05	0.307	0.0121
median rf in nel	1.21	3.35	0.79	0.54	0.163	0.0263
10th rf in cel	1.07	2.93	0.53	0.51	0.0442	0.0347
90th ph in t2all	0.29	1.33	0.19	0.12	0.134	0.0145

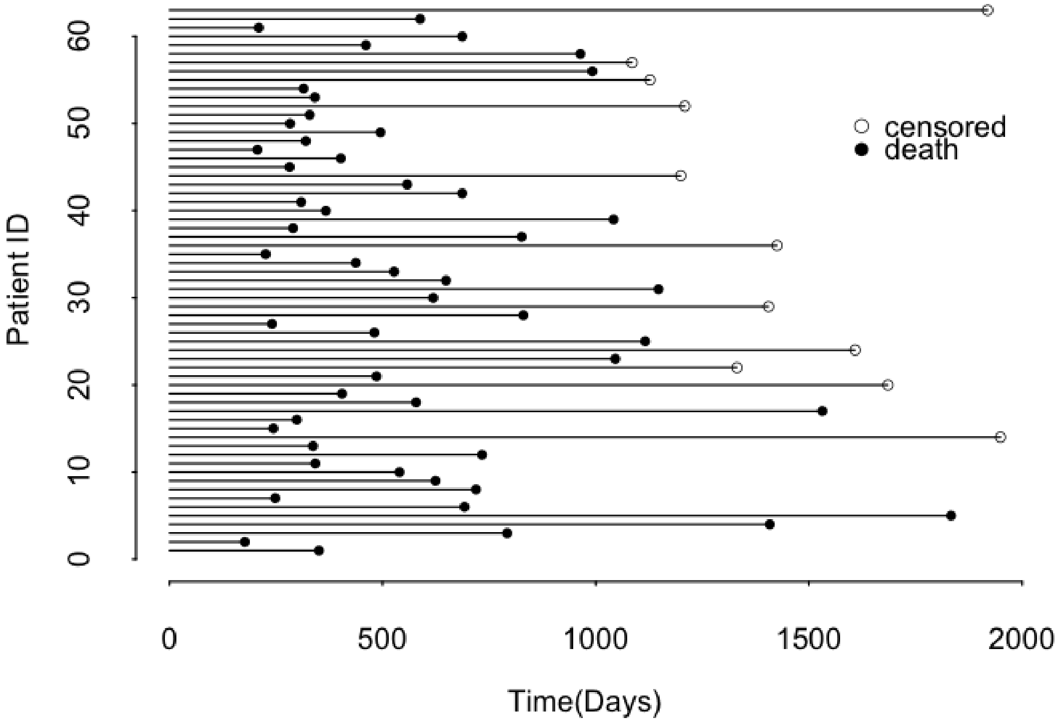


Figure 11.1: Event chart for 64 GBM patients

S5: Assess each interaction term individually.

S6: Evaluate model assumptions.

The final multivariate model is illustrated in Table 11.2. It includes nine covariates and an interaction term between median ADC in the CEL region and the 10th percentile of RECOV in the CEL region (partial likelihood test p-value = 0.00284).

Figures 11.3-11.6 evaluate the model assumption. Figure 11.3 shows the Martingale-residual plots for the covariates in the multivariate Cox regression model. This plot reveals the linear relationship between the covariates and the overall survival. This relationship is also illustrated in the component-plus-residual plots illustrated in Figure 11.4. The index plots of dfbetas for the Cox regression model of overall survival are illustrated in Figure 11.5. They reveal that none of the observations are outliers in all covariates. Figure 11.6 plots the scaled Schoenfeld residuals against transformed time for each covariate in the model fit in Table 11.2. The solid line is a smoothing-spline fit to the plot, with the broken lines representing a ± 2 standard-error band around the fit. These plots are a check of proportional hazards assumption for each covariate.

The final multivariate model in Table 11.2 was used to predict individual overall survival. Ten fold cross-validation and 0.632 bootstrapping were used to test the accuracy of the predictions.

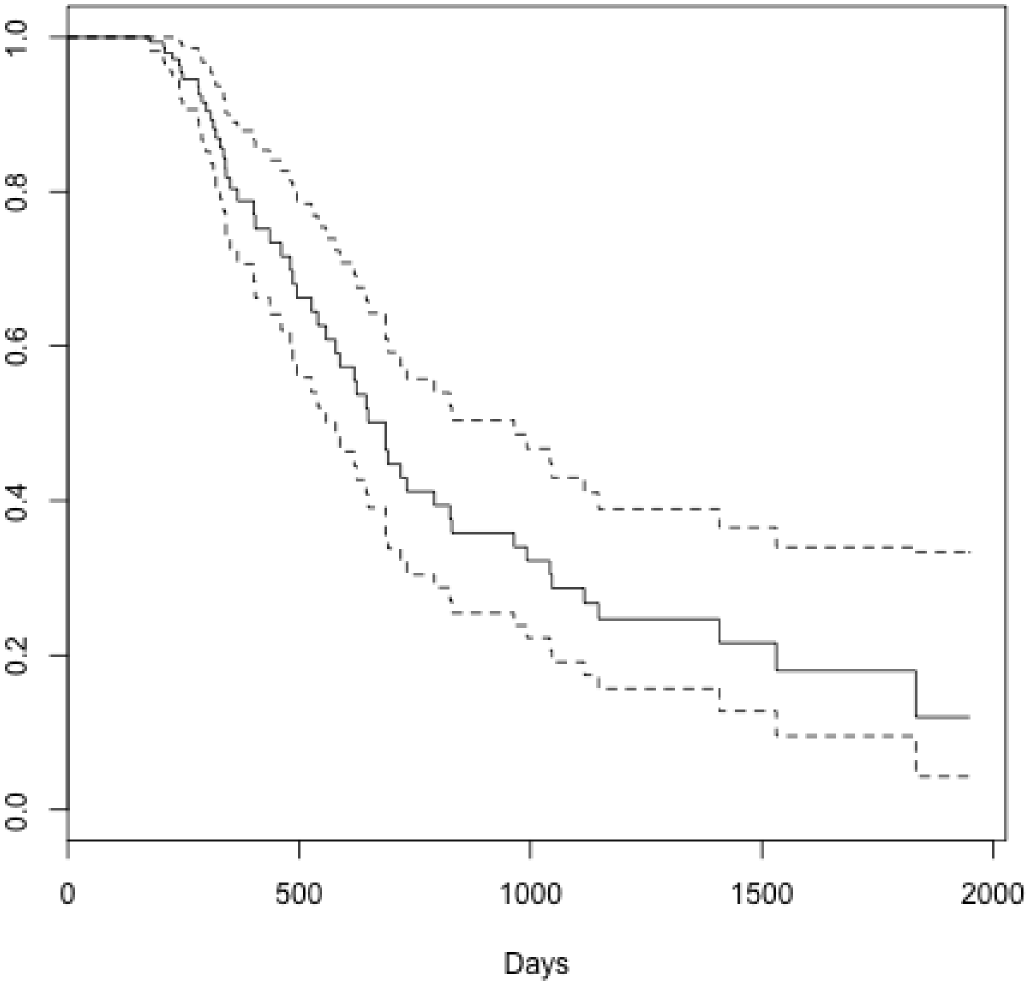


Figure 11.2: Kaplan-Meier overall survival curve (median OS was 588 days)

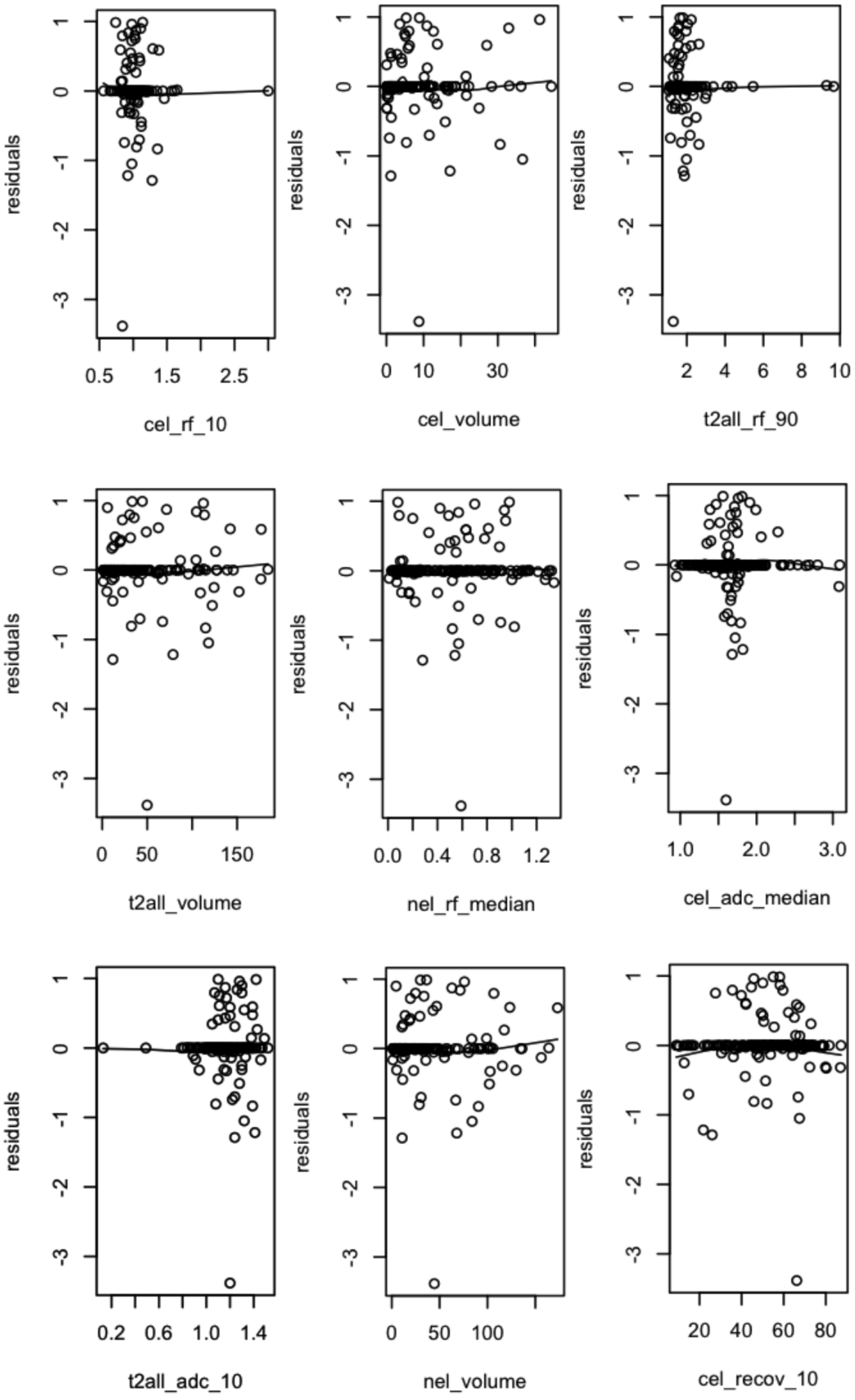


Figure 11.3: Nonlinearity test: Martingale-residual plots for the covariates in the Cox regression model

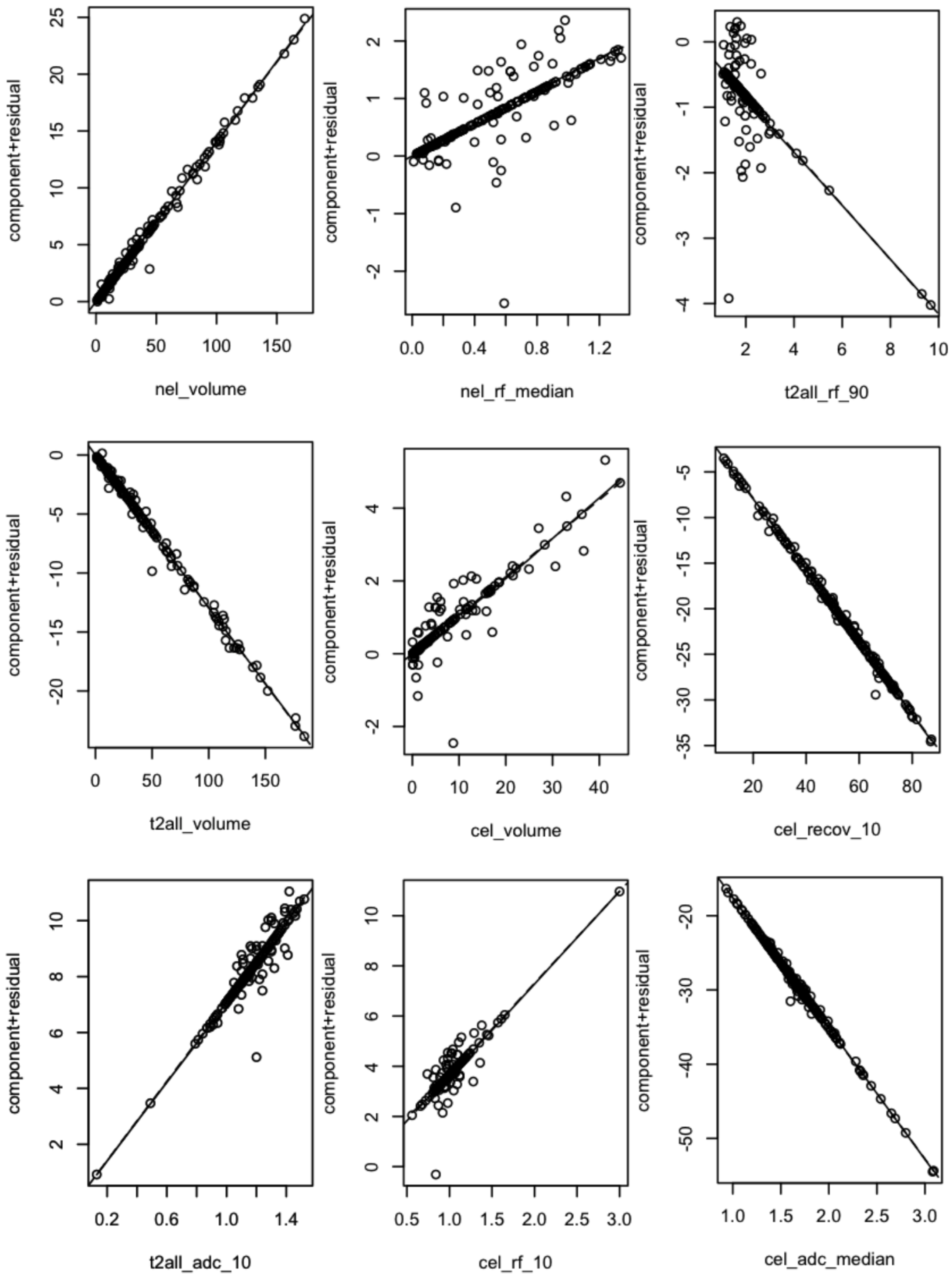


Figure 11.4: Nonlinearity test: component-plus-residual plots for the covariates in the Cox regression model

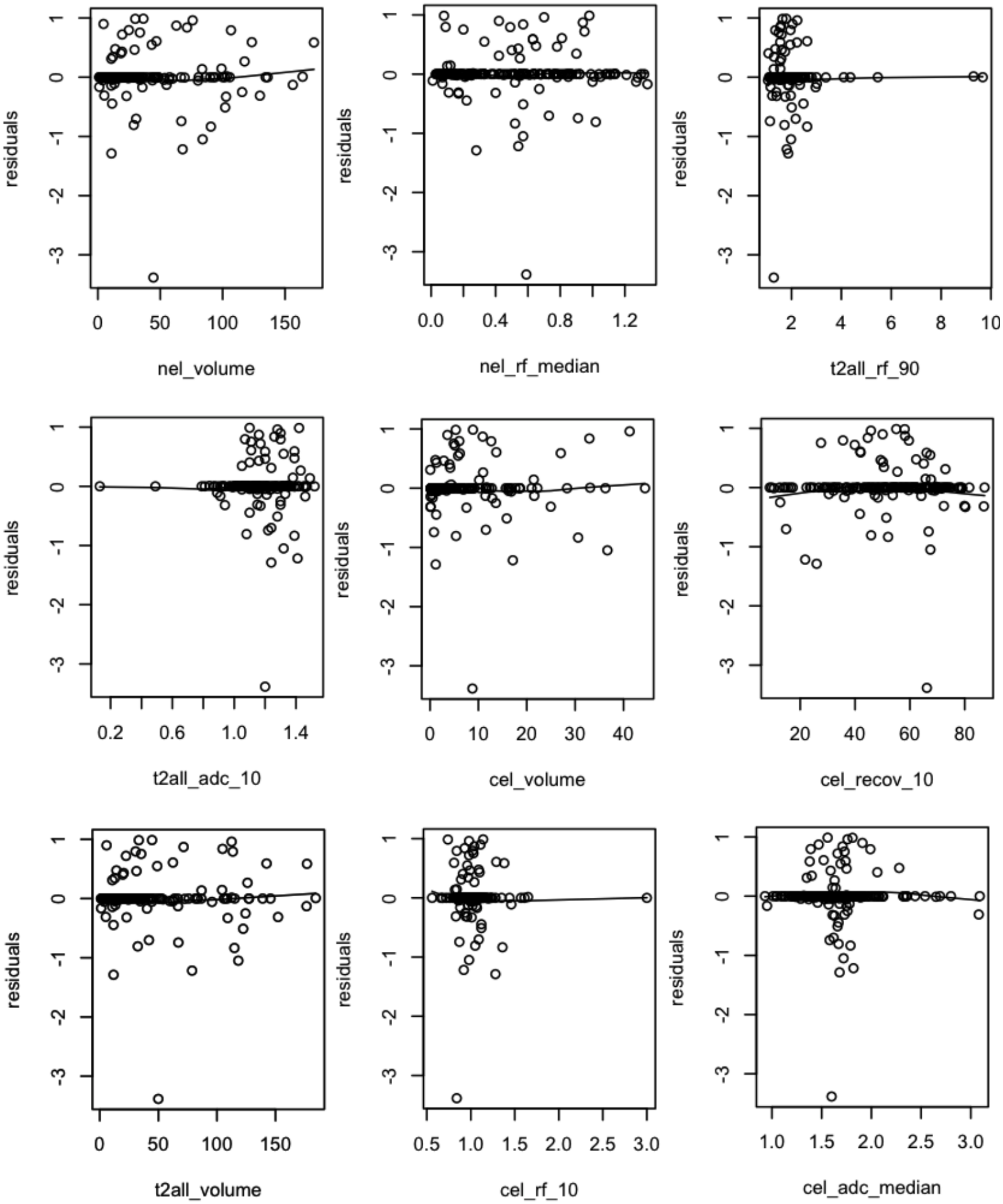


Figure 11.5: Influential observations: index plots of dfbeta for the Cox regression model

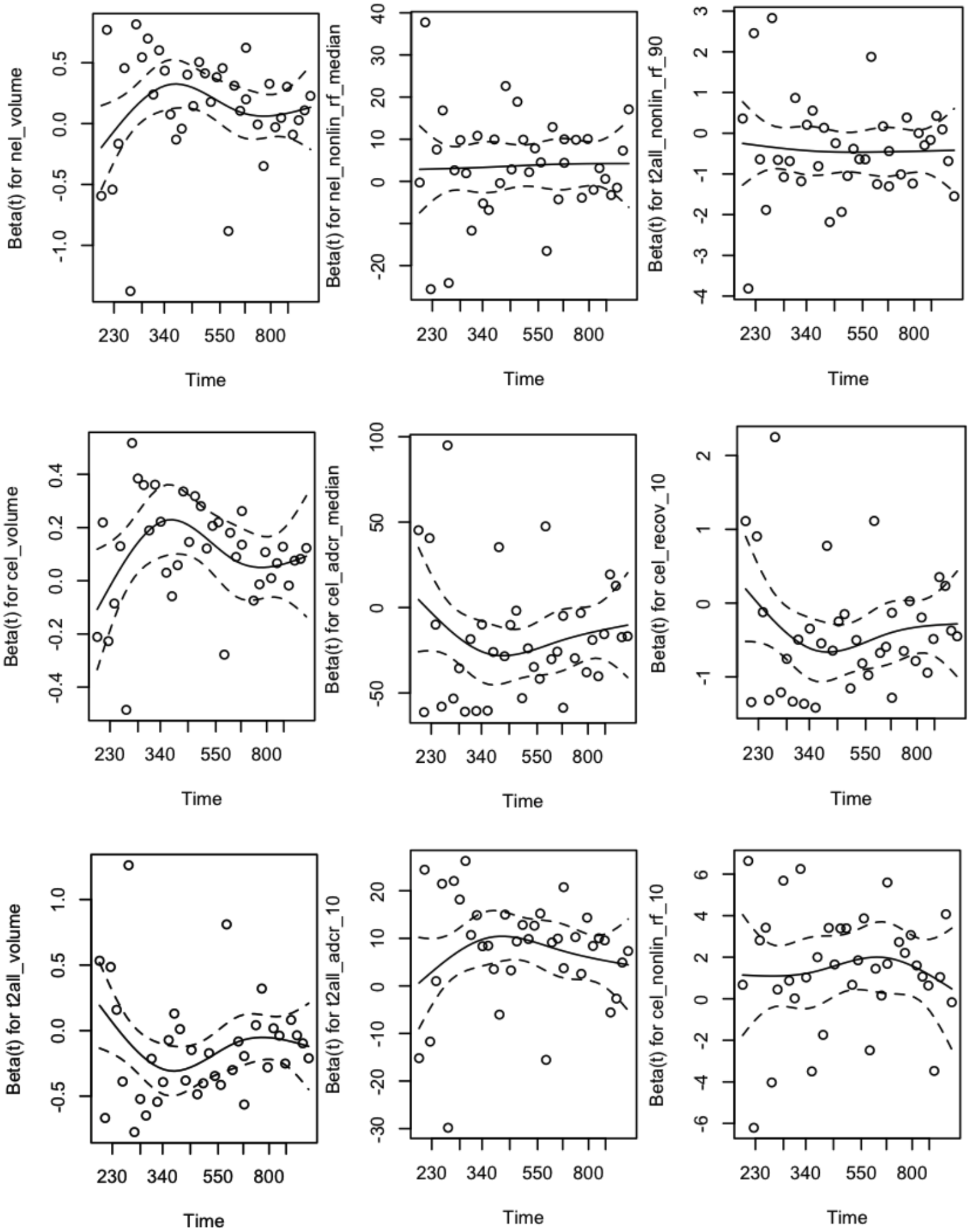


Figure 11.6: Proportional hazards assumption: plots of scaled Schoenfeld residuals against transformed time for each covariate in the Cox regression model

Table 11.2: Final multivariate model with significant interaction (partial likelihood test p-value=0.00284)

VAR	COEF	EXP(COEF)	S.E.	ROBUST S.E.	P-VALUE
t2all volume	-0.129	0.8786	0.06	0.05	0.010003
nel volume	0.14	1.151	0.06	0.05	0.008054
cel volume	0.105	1.111	0.06	0.03	0.002402
median adc in cel	-17.59	2.29E-08	5.15	4.69	0.000176
90th rf in t2all	-0.416	0.66	0.22	0.16	0.008168
10th adc in t2all	7.084	1193	1.93	1.47	0.00000151
median rf in nel	3.655	38.68	1.59	1.59	0.021281
10th rf in cel	1.4	4.054	0.59	0.45	0.001724
10th recov in cel	-3.936	0.6746	0.13	0.11	0.000326
median adc in cel X 10th recov in cel	0.247	1.28	0.07	0.07	0.000198

The prediction error curves obtained using the Brier score over time are illustrated in Figure 11.7. The dashed line represents the chance prediction error. The figure illustrates that the model can accurately predict the overall survival of patients with either relatively short or relatively long survival times. The model is less accurate for patients with median survival times.

Figure 11.8 compares the bootstrapping prediction error curves of the multivariate model with and without the interaction term, and of the univariate prediction models. Using the full model in Table 11.2 produces the lowest prediction errors. However, the model excluding the interaction term has very high errors for patients with median survivals. Among the univariate models, the volume parameters and the 10^{th} percentile ADC inside the T2all region produce the smallest prediction errors.

11.4 Discussion

MRI is the principal method being used to evaluate response to therapy and tumor progression in patients with GBM. The current study examined the predictive value of integrating anatomic, DWI, and PWI examinations that were obtained postsurgery, pretreatment, and at post-treatment follow-ups to determine whether the information obtained could accurately predict overall survival.

The median OS for the patients in this study with GBM was 84 weeks, but a large range of outcomes were observed. Previous univariate studies have found several variables that are associated with worse survival [68], but this multivariate study showed how these values can be combined to predict survival. A multivariate model with 9 significant parameters and one interaction suggests that different MR modalities provide complementary information about survival. Prediction errors

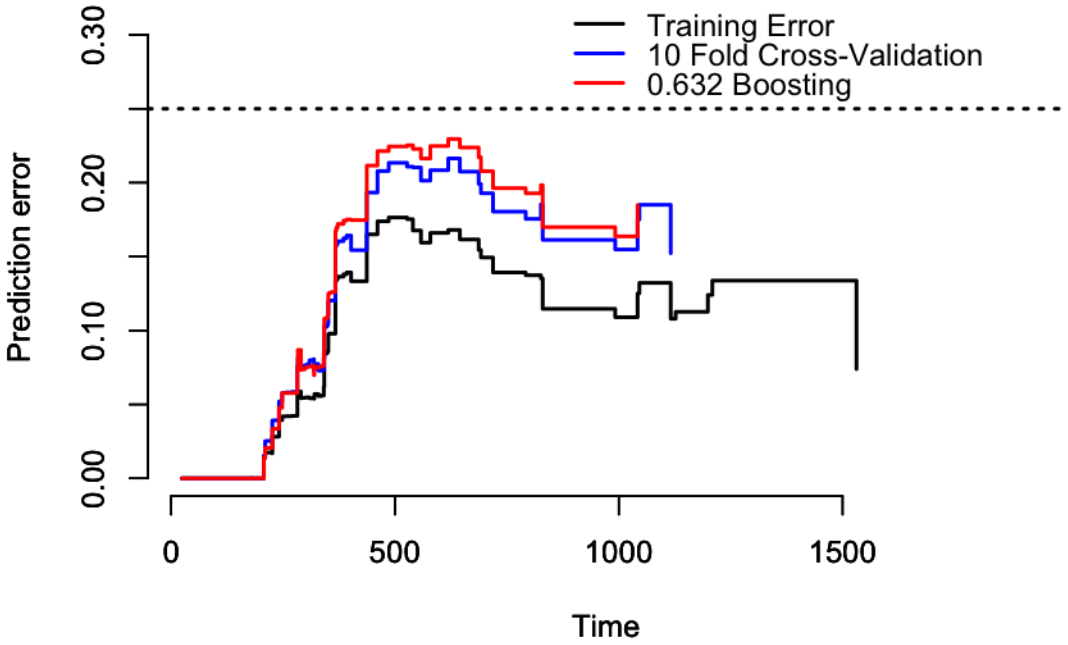


Figure 11.7: Prediction error curves for survival models, obtained using Brier score over time

curves suggest that the evolution of GBMs follows a complex process not fully explained by MR diffusion and perfusion parameters.

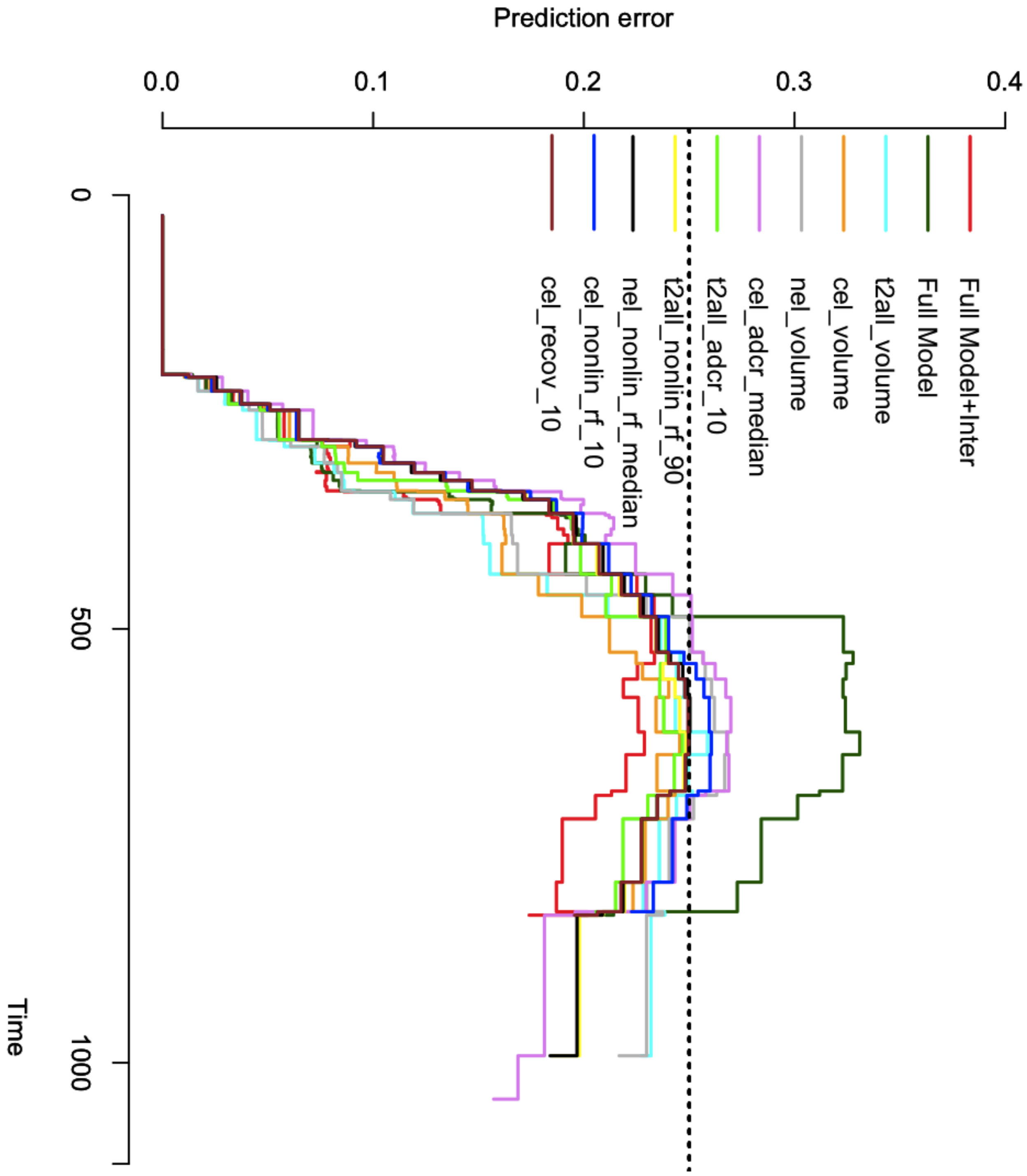


Figure 11.8: Bootstrapping 0.632 prediction error curves for survival models with different covariates, obtained using Brier score over time

Chapter 12

Automatic segmentation of brain tissue in patients with newly diagnosed gliomas

The study presented in this chapter was published in the Proceedings of the 2010 IEEE International Symposium on Biomedical Imaging [16] and reproduced with permission.

Effective brain tumor treatment depends on the accurate identification of boundaries between tumor, edema, and healthy tissue. Most of the studies previously described in this thesis rely on the identification of ROIs in the MR images that describe the tumor and abnormality extent. High grade tumors challenge our ability to make distinctions between abnormal brain tissue types, because they are diffuse, infiltrative, and heterogenous. They invade the surrounding healthy tissue and are comprised of enhancing and non-enhancing tumor tissue and edema, making the transition from tumor to healthy tissue gradual and hard to identify.

In this study, spectroscopy data were used for the coarse detection of the tumor region. Once the tumor area was identified, the FLAIR abnormality region was finely determined using thresholding with morphology and surface smoothing. Areas of contrast enhancement and necrosis were then identified by analyzing the abnormal FLAIR region in a T1-GAD. Once the tumor regions were removed, the remaining tissue was segmented into WM, grey matter (GM), and cerebrospinal fluid (CSF) using a hierarchical graphical model based on the intensity values of the pre-contrast T1-weighted image, the T2-weighted FSE image, and the T1-GAD image. The novel aspects of this algorithm include the use of spectroscopy for coarse tumor detection, which makes it possible to use fast thresholding methods with high accuracy, as well as the identification of ROIs within the tumor. Together, the pieces of this segmentation algorithm provide a framework for accurately segmenting MR images into healthy and abnormal ROIs. The method can be used with few corrections to replace the manual segmentation of such images used in the medical field.

12.1 Data acquisition and preprocessing

Sixty-nine patients with a diagnosis of high grade glioma received MRI examinations preceding surgery. Thirty-two were scanned on a 1.5T MR scanner, and 37 were scanned on a 3T MR scanner. The MR examination included a three-dimensional T1-weighted sequence, acquired both with and without gadolinium contrast agent, a three-dimensional T2-weighted sequence, and a three-dimensional FLAIR sequence. These images had a nominal resolution of $1mm \times 1mm \times 3mm$. Chemical shift imaging was performed using point-resolved spectroscopy volume-section techniques ($10mm \times 10mm \times 10mm$ nominal spatial resolution). For the spectroscopy, water suppression was achieved either through the use of spectral-spatial spin-echo pulses or CHESS, and outer volume suppression was performed using very selective suppression pulses [77]. The spectral amplitudes and line-widths of Cho, Cre, NAA, Lac, and Lip were estimated. The CNI index was calculated using a robust linear regression algorithm [77]. All images were rigidly registered to the T1-GAD. The registration was performed through the maximization of normalized mutual information using a gradient ascent algorithm [123]. The images were stripped of the skull using the Brain Extraction Tool (BET) [118].

12.2 Tumor segmentation

The patients considered in this study were diagnosed with GBM, which is the most common and most aggressive type of brain tumor [7]. Scans of GBM usually show a heterogeneous mass with a hypo-intense or necrotic center and a variable ring of enhancement surrounded by edema [7]. In this study, the ROIs within a GBM are the CEL and the NEC.

The tumor segmentation algorithm used the CNI index to detect the tumor. A thresholding algorithm with morphology was then used to find the boundaries of the tumor. Surface smoothing was then applied as a last step correction. The algorithm is described in more detail next, and the pseudocode is provided below. The algorithm used spectroscopy data for coarse detection of the tumor region. Spectroscopy data have a much coarser resolution than structural images, and therefore cannot be used for an accurate delimitation of the boundary of tissue abnormality. Furthermore, spectroscopy data were only available for part of the structural image. In order to obtain

Algorithm 6 Tumor segmentation algorithm

```
compute mask $CNI_{abnormal}$
fit Gaussian distribution through FLAIR intensity values
select threshold $t_0$ fitted based on Gaussian distribution
threshold FLAIR image:$I_{t_0} = I_{FLAIR} > t_0$
fill holes, erode, then dilate $I_{t_0}$
find disjoint regions in the binary image $I_{t_0}$
select regions of $I_{t_0}$ which contain abnormal CNI: $L_0$
define small windows around connected regions in $L_0$
optimize threshold $t$ on local window
select final abnormal region $L$ and smooth boundary
threshold CEL in region $L$ of T1-GAD image
threshold NEC in region $L$ of T1-GAD image
```

high quality spectroscopy data, lipid has to be avoided during the acquisition process, which means that part of the tumor might not have spectroscopic information. Even though spectroscopy data cannot be used to finely determine the boundaries of tumors, this data reflect metabolite information that is a lot more reliable in detecting abnormal tissue than the information available in the structural images. Therefore, using spectroscopy for tumor detection and structural images for boundary delineation can lead to more accurate results than using each modality on its own. An abnormality index was derived for the CNI as described in Section 12.1. Once the areas of abnormal CNI were identified, further processing was done on the FLAIR image in the neighborhood of the tumor to determine the FLAIR abnormality region. In FLAIR images of patients diagnosed with GBM, the abnormal region is typically brighter than the rest of the image. A thresholding algorithm was used to identify the bright FLAIR areas in the vicinity of abnormal CNI. A Gaussian distribution was fitted through to the FLAIR intensity values, and a threshold was selected based on this distribution. The rough area of the abnormal FLAIR region was estimated using the area of the abnormal CNI region. This area was used to select an initial thresholding value. The threshold was then optimized by maximizing the inter-region contrast between the area of high FLAIR intensity and the adjacent regions. The contrast was measured in terms of the average gray-level intensity of the ROI, I_{ROI}, and the average intensity of the local background, I_{BG}, and was computed as $C = \frac{I_{ROI} - I_{BG}}{I_{ROI} + I_{BG}}$. In order to obtain the intensity of the local background, the bounding box around the ROI was computed, and its edges were increased by 5% in each direction. This provided a local analysis window that was large enough to accurately recompute the tumor boundaries, yet small enough to keep the noise levels reduced. The pixels inside the extended box that did not belong to the ROI were considered local background pixels. After thresholding, three-dimensional morphological operations were used for spatial correction: the ROI was dilated and eroded and gaps were filled. The surface of the ROI was modeled by surface tessellation using connected triangles. This model was used to smooth the boundaries of the FLAIR abnormality. Areas of contrast enhancement and necrosis were then identified by analyzing the FLAIR abnormality region in a T1-GAD image. A thresholding algorithm was used to identify the contrast enhancing lesion as the brighter region in the abnormal tissue. Necrosis was identified through thresholding as the darker region within the abnormal tissue. The same segmentation accuracy metrics were used to determine appropriate thresholding values.

12.3 Healthy tissue segmentation

After the abnormal brain region was identified using the FLAIR image, the healthy tissue was further segmented into WM, GM, and CSF based on three images: the pre- and post-contrast T1-weighted images, and the T2-weighted image. A hierarchical graphical model was used to model the intensity values of the three healthy tissue types in the three types of MR images, and the parameters of the model were estimated using the EM algorithm.

The image classification problem involved assigning to each voxel a class label taking a value from the set $\mathcal{L}$. Each pixel in an image was characterized by an intensity value Y. The true and unknown labeling of the image was denoted as X^*, and $\hat{X}$ was an estimate of X^*, both of which were interpreted as particular realizations of a random field X, which was a Markov random field (MRF) with a specified distribution $P(X)$. The observable image was denoted by Y, which was a realization of a Gaussian Hidden Markov Random Field (GHMRF). The problem of classification was recovering X^* given the observed image Y.

The algorithm used for single channel segmentation was introduced in [140] and is briefly described next. This algorithm was then extended to multichannel segmentation. The EM algorithm was used to iteratively fit the parameters of the model. The strategy underlying the EM algorithm was the following: estimate $\hat{X}$ given the current θ estimate, then estimate the new parameters θ by maximizing the expectation of the complete-data log likelihood, $E[logP(X, Y|\theta)]$. The E-step of the EM algorithm calculated the conditional expectation $Q(\theta|\theta^t)$, while the M-step maximized it to obtain the next estimate.

Algorithm 7 EM algorithm

```
E-step: $Q(\theta|\theta^t) = E[\log P(X, Y|\theta)|Y, \theta^t]$
M-step: $\theta^{t+1} = argmax_\theta Q(\theta|\theta^t)$
```

Given several types of images of the same structure, a hierarchical model was formed that combined complementary brain tissue type information from several imaging modalities. The data from m imaging modalities were denoted as $Y = Y_1, \dots Y_m$. Each of the imaging modalities had its own parameters, $\Theta = \theta_1, \dots, \theta_m$. The remaining parameters of the model were the labels corresponding to the underlying tissue types, θ_L. Then, according to Bayes' rule, $p(\Theta, \theta_L|Y, M) \propto \{\prod_{i=1}^m p(Y_i|\theta_i, \theta_L)\} p(\theta_L|M)$. The marginal posterior for the common parameters was obtained by marginalizing over the parameters specific to each distribution:

$$p(\theta_L|Y, M) \propto \int_{\theta_1} \dots \int_{\theta_m} p(\Theta, \theta_L|Y, M) d\theta_1 \dots d\theta_m. \tag{12.1}$$

12.4 Results

The segmentation algorithm was run on the data described in Section 12.1. The segmentation results for the T2all, CEL, NEC, and WM were then compared to expert manual segmentation. These four types of tissue were the ROIs that were manually segmented for all the acquired scans. WM was of interest because it can be used for normalization in other studies. The precision (P), recall (R), and F-measure results of this comparison are summarized in Table 12.1, and some examples are illustrated in Figure 12.1. Table 12.1 also illustrates the benefits of using CNI for tumor detection, by contrasting the results with those of using FLAIR alone for the abnormality detection.

12.5 Discussion

The algorithm proposed in this study was able to obtain the FLAIR abnormality with high accuracy. This algorithm was fast, and did not require any training. The use of spectroscopy data

Table 12.1: Segmentation results

ROI	with CNI			without CNI		
	P	R	F	P	R	F
T2all	0.95	0.95	0.95	0.48	0.95	0.64
CEL	0.85	0.88	0.86	0.56	0.80	0.66
NEC	0.87	0.85	0.85	0.33	0.81	0.47
WM	0.98	0.90	0.94	0.98	0.77	0.86

for tumor detection eliminated the need for a slow, complicated algorithm for tumor detection or for correcting spatial segmentation errors. The possible drawback in the use of spectroscopic images was the extra acquisition time. However, spectroscopic images should be routinely acquired for glioma patients in order to assess tumor progression and treatment response. The accurate spectroscopy-based tumor detection made it possible to use a fast thresholding with morphology and surface smoothing to delineate the boundaries of the abnormal region, and to further detect the contrast enhancing lesion and necrosis. Accurately removing the abnormal brain tissue from the images made it possible to use a clustering Bayesian algorithm for healthy tissue segmentation. This algorithm could not cope with the heterogeneity in tumor appearance, but provided a good mathematical framework for integrating the appearance of common healthy structures from different types of images.

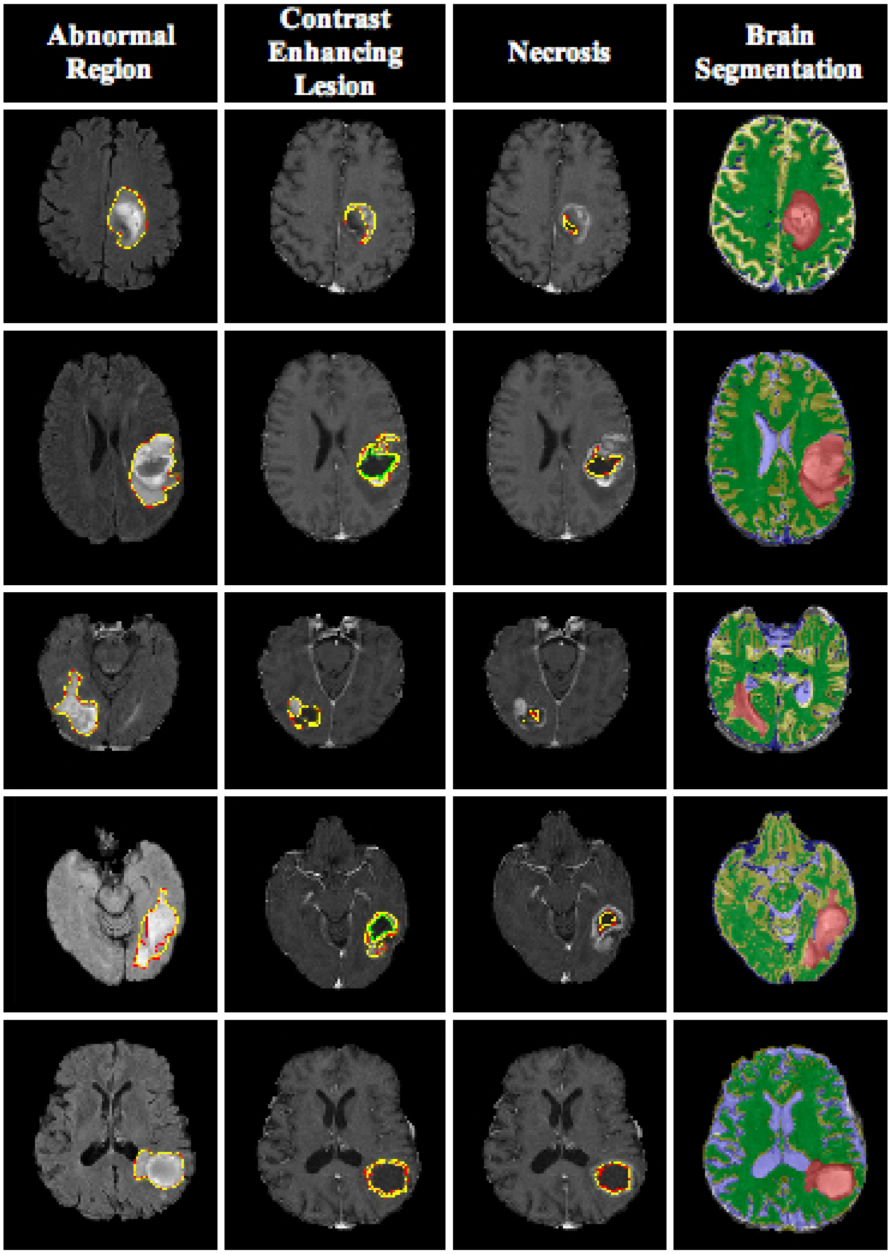


Figure 12.1: Segmentation results: each row corresponds to one patient; the columns, from left to right, illustrate the abnormal brain region on a FLAIR image, the CEL and NEC on a T1-GAD image, and an overlay of the abnormal, WM, GM, and CSF automated segmentation on top of a T1-weighted image; the ROIs segmented by our algorithm have red outer boundaries and green inner boundaries; the manually segmented ROIs have yellow outer boundaries and blue inner boundaries; the tumor is red, the CSF is blue, the WM is green, and the GM is yellow

Chapter 13

Summary

The objective of this thesis was the analysis of advanced NMR data in conjunction with multivariate pattern recognition methods to assist in the clinical management and biological understanding of gliomas, the most malignant type of brain tumors. The tools developed in this thesis were used to characterizing biological changes predictive of malignant transformations and treatment effects in gliomas, and for the early detection of disease progression. These tools were crucial in finding links between *in vivo* and *ex vivo* data that could give insight into the biology of brain cancer. The methods described in this thesis can be easily translated into clinical practice by improving the selection of biopsy sites and the targeting of treatment.

The main contribution of this thesis was the development of computer-assisted support for glioma understanding, diagnosis, and prognosis in clinical environments based on pattern recognition methods. The work described in this thesis produced results with high accuracy in classification, interpretability by means of clinical knowledge, and capacity to generalize the performance to new samples. The technical aspects covered in this thesis include the feature extraction and modeling of biomedical data, the inference and evaluation of predictive models, and the use of the models for clinical applications. These methods were described in Chapter 3.

Chapter 4 compared the use of several multivariate classification methods for identifying recurrent LGGs that transformed to a higher grade based on HRMAS data. The highest bootstrapping classification accuracy obtained by these classifiers was 96%. The use of interpretable feature selection methods lead to the identification of a parsimonious set of metabolites that are predictive of malignant transformations in recurrent LGGs. The identification of these metabolites can inform the acquisition of *in vivo* MRS, which could lead to better non-invasive identification of malignant transformations in LGGs.

The HRMAS analysis was extended in Chapter 5 to identify the differences between primary and secondary GBMs, with bootstrap classification accuracy of 98%. This analysis identified 2HG, a metabolite whose increased concentration is highly predictive of malignant transformations in recurrent LGGs, but whose concentration is generally low in newly diagnosed GBMs. This provided evidence supporting the hypothesis that primary and secondary GBMs are different disease entities, and that 2HG is useful in discriminating between them. Distinguishing between these two categories is important in administering the right treatment plan.

HRMAS data were also used to compare the metabolic profiles of treated and untreated grade 2 and 3 gliomas in Chapter 6. This work identified chemical changes that may be due to treatment effect, and that could thus help differentiation between tumor and treatment induced transformations in non-malignant tissue. This work could also help in the detection of early response to treatment. Pattern recognition methods determined that treated and untreated gliomas have very different metabolic profiles. This suggests that these tumors might benefit from different types of treatment.

Histological analysis is a relatively subjective process that varies across experts. The development of HRMAS led to an objective method for analyzing tissue. However, this data are significantly harder to interpret. Finding a link between HRMAS metabolites and histological parameters could lead to faster, more objective glioma grading and typing. This type of analysis could also provide new insights into the biology of brain tumors. For this reason, pattern classification methods for modeling histological features based on *ex vivo* metabolic data were described in Chapter 7. These models achieved bootstrapping accuracies of 78-88%, with tumor cellularity, hyperplasia, and the presence of complex vasculature being easier to model than CA9 or the MIB-1 proliferation index.

A natural continuation of the HRMAS work was to explore the use of *in vivo* spectroscopy and imaging to detect malignant transformations in LGGs. The model was developed in Chapter 8 and had 93% LOOC accuracy and 84% bootstrap accuracy. The previous HRMAS study indicates that this model could be improved by acquiring Myo-I *in vivo*. In the clinic, this model can lead to the early detection of tumor progression and to a faster change of treatment for those patients exhibiting malignant transformations.

Tumor grading is a complex process that involves the search for histological features predictive of malignancy either in a large biopsy sample or in several smaller biopsy specimens. In this work, the biopsy samples were small, in order to allow for their precise localization on the anatomical images. It is very unlikely that one biopsy sample contained all the histological features required for grading, because of tumor heterogeneity. In general, classification models linking *in vivo* and *ex vivo* data could help select biopsy sites that contain histological features of interest. The link between *in vivo* and histological glioma parameters was explored in Chapter 9. This chapter describes classification models for distinguishing between low and high values of histological parameters. These models achieved bootstrapping accuracies of 75-86%, with tumor cellularity, hyperplasia, and the presence of complex vasculature being easier to model than CA9 or the MIB-1 proliferation index. Chapter 10 describes a method for improving the tumor cellularity prediction by estimating the remaining histological parameters using the EM algorithm. These models can be used to create maps of tumor cellularity and other histological estimates at the spatial resolution of anatomical images. These maps can be used to improve the biopsy site selection process.

Multivariate analysis methods can also be used for predicting survival. An application of multivariate survival methods for GBM patients is presented in Chapter 11. This work describes the methodology for building a multivariate survival analysis model and using cross-validation and bootstrapping to estimate its accuracy.

Most of the projects involving *in vivo* data make use of manually defined ROIs. In order to help with the definition of the ROIs, an automated tissue segmentation technique for patients with newly diagnosed GBM was developed and described in Chapter 12. This method used multi-modality MRI and MRS data. The spectroscopy data were used for coarse detection of the tumor region. Once the tumor area was identified, further processing was done on a FLAIR image in the neighborhood of the tumor to determine the T2all abnormality region. Areas of contrast enhancement and necrosis were then identified by analyzing the T2all in a T1-GAD image. The healthy brain tissue is then segmented into WM, GM, and CSFusing a hierarchical graphical model whose parameters are estimated using the EM algorithm. This method segmented the T2all region with 95% accuracy compared to manual segmentation. The CEL and NEC regions were segmented with 85% accuracy.

This results in this thesis demonstrated that the systematic analysis of patients' biomedical data could lead to better disease understanding, diagnosis, prognosis, and treatment.

Bibliography

[1] A. Ahmad and L. Ley. “A feature selection technique for classificatory analysis”. In: *Pattern Recognition Letters* 26 (2005), pp. 43–56.

[2] O.C. Andronesi, K.D. Blekas, D. Mintzopoulos, L. Astrakas, P.M. Black, and A.A. Tzika. “Molecular classification of brain tumor biopsies using solid-state magic angle spinning proton magnetic resonance spectroscopy and robust classifiers”. In: *International Journal of Oncology* 33.5 (2008), pp. 1017–1025.

[3] C. Arus. “Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE”. In: *American Journal of Neuroradiology* 25 (2004), pp. 1696–1704.

[4] R. Bajcsy and S. Kovacic. “Multiresolution elastic matching”. In: *Journal of Computer Vision, Graphics, and Image Processing* 46.1 (1989), pp. 1–21.

[5] R. Bajcsy, R. Lieberson, and M. Reivich. “A computerized system for elastic matching of deformed radiographic images to idealized atlas images”. In: *Journal of Computer Assisted Tomography* 7.4 (1983), pp. 618–625.

[6] R.F. Barajas, J.G. Hodgson, J.S. Chang, S.R. Vandenberg, R.F. Yeh, A.T. Parsa, M.W. McDermott, M.S. Berger, W.P. Dillon, and S. Cha. “Glioblastoma multiforme regional genetic and cellular expression patterns: influence on anatomic and physiologic MR imaging.” In: *Radiology* 254.2 (2010), pp. 564–76.

[7] M. Bernstein and M.S. Berger. *Neuro-oncology: the essentials*. New York, NY: Thieme Medical Publishers, 2008.

[8] J.N. Bruce and B. Kennedy. “Glioblastoma Multiforme”. In: *MedScape* (2009). Accessed 25 October 2011. URL: `http://emedicine.medscape.com/article/283252-overview`.

[9] H. Bruhn, J. Frahm, M. Gyngell, K. Merboldt, W. Hanicke, R. Sauter, and C. Hamburger. “Noninvasive differentiation of tumors with use of localized 1H-MR spectroscopy in vivo:initial experience in patients with cerebral tumors”. In: *Radiology* 172 (1989), pp. 541–548.

[10] P.C. Burger and S.B. Green. “Patient age, histologic features, and length of survival in patients with glioblastoma multiforme.” In: *Cancer* 59.9 (1987), pp. 1617–1625.

[11] J.T. Bushberg, J.A. Seibert, Jr. E.M. Leidholdt, and J.M. Boone. *The essentials physics of medical imaging*. 2nd Edition. Philadelphia, PA: Lippincott Williams & Wilkins, 2002.

[12] F. Cartes-Zumelzu. “MRI for brain tumours: a multimodality approach”. In: *Magazine of European Medical Oncology* 2 (July 2009), pp. 15–19.

[13] S. Le Cessie and J.C. van Houwelingen. “Ridge estimators in logistic regression”. In: *Applied Statistics* 41.1 (1992), pp. 191–201.

[14] L.L. Cheng, D.C. Anthony, A.R. Comite, P.M. Black, A.A. Tzika, and R.G. Gonzalez. “Quantification of microheterogeneity in glioblastoma multiforme with ex vivo high-resolution magic-angle spinning (HR-MAS) proton magnetic resonance spectroscopy”. In: *Neuro-Oncology* 2 (2000), pp. 87–95.

[15] L.L. Cheng, M.J. Ma, L. Becerra, T. Ptak, I. Tracey, A. Lackner, and R.G. Gonzalez. “Quantitative neuropathology by high resolution magic angle spinning proton magnetic resonance spectroscopy”. In: *National Academy of Science USA* 94 (1997), pp. 6408–6413.

[16] A. Constantin, R. Bajcsy, and S.J. Nelson. “Unsupervised segmentation of brain tissue in multivariate MRI”. In: *Proceedings of the 2010 IEEE International Symposium on Biomedical Imaging: from Nano to Macro*. ©2010 IEEE. Reprinted with permission. Rotterdam, Netherlands: IEEE, 2010, pp. 89–92.

[17] A. Constantin, A. Elkhaled, L. Jalbert, K. Smith, T. McKnight, A. Molinaro, J. Phillips, S.M. Chang, R. Bajcsy, and S.J. Nelson. “Effects of treatment on the metabolic characteristics of grade 2 and grade 3 gliomas”. In: *Proceedings of the 20th Annual Meeting of the International Society for Magnetic Resonance in Medicine*. Melbourne, Australia: ISMRM, 2012, p. 5396.

[18] A. Constantin, A. Elkhaled, L. Jalbert, R. Srinivasan, S. Cha, S.M. Chang, R. Bajcsy, and S.J. Nelson. “Identifying malignant transformations in recurrent low grade gliomas using high resolution magic angle spinning spectroscopy”. In: *Artificial Intelligence in Medicine* 55 (2012). ©2012 Elsevier. Reprinted with permission., pp. 61–70.

[19] A. Constantin, A. Elkhaled, L. Jalbert, R. Srinivasan, S. Cha, S.M. Chang, R. Bajcsy, and S.J. Nelson. “Multivariate pattern analysis for identification of metabolites that are predictive of malignant transformations in gliomas using HRMAS spectra from image guided tissue samples”. In: *Proceedings of the 19th Annual Meeting of the International Society for Magnetic Resonance in Medicine*. Montreal, Canada: ISMRM, 2011, p. 724.

[20] A. Constantin, L. Jalbert, A. Elkhaled, R. Parvataneni, A. Molinaro, J. Phillips, S. Cha, S.M. Chang, R. Bajcsy, and S.J. Nelson. “Multivariate analysis of in vivo MR imaging parameters for detecting transformations to a higher grade in patients with recurrent low grade gliomas”. In: *Proceedings of the 20th Annual Meeting of the International Society for Magnetic Resonance in Medicine*. Melbourne, Australia: ISMRM, 2012, p. 2231.

[21] A. Constantin, S.J. Nelson, and R. Bajcsy. “Modeling tumor cellularity in newly diagnosed GBMs using MR imaging and spectroscopy”. In: *Proceedings of the 2010 MICCAI Workshop on Computational Imaging Biomarkers for Tumors: From Qualitative to Quantitative*. Beijing, China: MIDAS Journal, 2010.

[22] F.W. Crawford, I.S. Khayal, C. McGue, S. Saraswathy, A. Pirzkall, S. Cha, K.R. Lamborn, S.M. Chang, M.S. Berger, and S.J. Nelson. “Relationship of pre-surgery metabolic and physiological MR imaging parameters to survival for patients with untreated GBM”. In: *Neuro-Oncology* 91.3 (2009), pp. 337–351.

[23] L. Dang, D.W. White, S. Gross, B.D. Bennett, M.A. Bittinger, E.M. Driggers, V.R. Fantin, H.G. Jang, S. Jin, M.C. Keenan, K.M. Marks, R.M. Prins, P.S. Ward, K.E. Yen, L.M. Liau, J.D. Rabinowitz, L.C. Cantley, C.B. Thompson, M.G. Vander Heiden, and S.M. Su. “Cancer-associated IDH1 mutations produce 2-hydroxyglutarate”. In: *Nature* 462.7274 (2009), pp. 739–744.

[24] R. Dann, J. Hoford, S. Kovacic, M. Reivich, and R. Bajcsy. “Evaluation of elastic matching system for anatomic (CT, MR) and functional (PET) cerebral images”. In: *Journal of Computer Assisted Tomography* 13.4 (1989), pp. 603–611.

[25] M. Deckert, G. Reifenberger, and W. Wechsler. “Determination of the proliferative potential of human brain tumors using the monoclonal antibody Ki-67”. In: *Journal of Cancer Research and Clinical Oncology* 115.2 (1989), pp. 179–188.

[26] A. Devos. “Quantification and classification of magnetic resonance spectroscopy data and applications to brain tumour recognition”. PhD thesis. Katholieke Universiteit Leuven, 2005.

[27] A. Devos, A.W. Simonetti, M. van der Graaf, L. Lukas, J.A.K. Suykens, L. Vanhamme, L.M. Buydens, A. Heerschap, and S. Van Huffel. “The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification”. In: *Journal of Magnetic Resonance* 173.2 (2005), pp. 218 –228.

[28] A.K. Dinda, K. Kharbanda, C. Sarkar, S. Roy, M. Mathur, and A.K. Banerji. “In-vivo proliferative potential of primary human brain tumors; its correlation with histological classification and morphological features: I. Gliomas”. In: *Pathology* 25.1 (1993), pp. 4–9.

[29] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern classification*. New York, NY: Wiley, 2001.

[30] R.N. Stickland (editor). *Image processing techniques for tumor detection*. New York, NY: Marcel Dekker, 2002.

[31] B. Efron and R. Tibshirani. “Improvements on cross-validation: the .632+ bootstrap method”. In: *Journal of the American Statistical Association* 92.438 (1997), pp. 548–560.

[32] A. Elkhaled, L.E. Jalbert, J.J. Phillips, H.A.I. Yoshihara, R. Parvataneni, R. Srinivasan, G. Bourne, M.S. Berger, S.M. Chang, S. Cha, and S.J. Nelson. “Magnetic resonance of 2-Hydroxyglutarate in IDH1-mutated low-grade gliomas”. In: *Science Translational Medicine* 4.116 (2012), pp. 116–121.

[33] G. Erb, K. Elbayed, M. Piotto, J. Raya, A. Neuville, M. Mohr, D. Maitrot, P. Kehrli, and I.J. Namer. “Toward improved grading of malignancy in oligodendrogliomas using metabolomics”. In: *Magnetic Resonance in Medicine* 59 (2008), pp. 959–965.

[34] M.C. Etienne, J.L. Formento, C. Lebrun-Frenay, J. Gioanni, M. Chatel, P. Paquis, C. Bernard, A. Courdi, R.J. Bensadoun, J.P. Pignol, M. Francoual, P. Grellier, M. Frenay, and Milano. G. “Epidermal growth factor receptor and labeling index are independent prognostic factors in glial tumor outcome”. In: *Clinical Cancer Research* 4.10 (1998), pp. 2383–2390.

[35] G. Filippini, C. Falcone, A. Boiardi, G. Broggi, M.G. Bruzzone, D. Caldiroli, R. Farina, M. Farinotti, L. Fariselli, G. Finocchiaro, S. Giombini, B. Pollo, M. Savoiardo, C.L. Solero, and M.G. Valsecchi. “Prognostic factors for survival in 676 consecutive patients with newly diagnosed primary glioblastoma”. In: *Neuro-Oncology* 10.1 (2008), pp. 79–87.

[36] J. Friedman, T. Hastie, and R. Tibshirani. “Additive logistic regression: a statistical view of boosting”. In: *The Annals of Statistics* 28.2 (2000), pp. 337–407.

[37] J. Gama. “Functional trees”. In: *Machine Learning* 55.3 (2004), pp. 219–250.

[38] J. Garcia, S. Tortajada, C. Vidal, M. Julia-Sape, J. Luts, S. Van Huffel, C. Arus, and M. Robles. *On the use of long TE and short TE SV MR spectroscopy to improve the automatic brain tumor diagnosis*. Tech. rep. Katholieke University Leuven, 2007.

[39] J.M. Garcia-Gomez. “Pattern recognition approaches for biomedical data in computer-assisted cancer research”. PhD thesis. Universidad Politecnica de Valencia, 2008.

[40] J.M. Garcia-Gomez, S. Tortajada, J. Vicente, C. Saez, X. Castells, J. Luts, M. Julia-Sape, A. Juan-Ciscar, S. Van Huffel, A. Barcelo, J. Arino, C. Arus, and M. Robles. “Genomics and metabolomics research for brain tumour diagnosis based on machine learning”. In: *Proceedings of the 9th International Work Conference on Artificial Neural Networks*. Ed. by Francisco Sandoval Hernandez, Alberto Prieto, Joan Cabestany, and Manuel Grana. Berlin: Springer, 2007, pp. 1012–1019.

[41] J.M. Garcia-Gomez, S. Tortajada, C. Vidal, M. Julia-Sape, J. Luts, A. Moreno-Torres, S. Van Huffel, C. Arus, and M. Robles. “The influence of combining two echo times in automatic brain tumor classification by magnetic resonance spectroscopy”. In: *Nuclear Magnetic Resonance in Biomedicine* 21.10 (2008), pp. 1112–1125.

[42] D. Gering. “Recognizing deviations from normalcy for brain tumor segmentation”. PhD thesis. Cambridge, MA: Massachusetts Institute of Technology, 2003.

[43] D.E. Goldberg. *Genetic algorithms in search, optimization and machine learning*. Boston, MA: Addison-Wesley Longman Publishing, 1989.

[44] H. Gonzalez-Velez, M. Mier, M. Julia-Sape, T.N. Arvanitis, J.M. Garcia-Gomez, M. Robles, A. Peet, C. Arus, B. Celda, S. Van Huffel, P. Lewis, D. Dupplaw, and S. Dasmahapatra. “HealthAgents: distributed multi-agent brain tumor diagnosis and prognosis”. In: *Applied Intelligence* 30.3 (June 2009), pp. 191–202.

[45] F.F. Gonzslez-Navarro, L.A. Belanche Munoz, E. Romero, A. Vellido, M. Julia-Sape, and C. Arus. “Feature and model selection with discriminatory visualization for diagnostic classification of brain tumors”. In: *Neurocomputing* 73.4-6 (2010), pp. 622–632.

[46] V. Govindaraju, K. Young, and A.A. Maudsley. “Proton NMR chemical shifts and coupling constants for brain tumor metabolites”. In: *Nuclear Magnetic Resonance in Biomedicine* 13 (2000), pp. 129–153.

[47] J.V. Hajnal, D.J. Hawkes, and D.L.G. Hill. *Medical image registration*. Boca Raton, FL: CRC Press, 2001.

[48] M.A. Hammoud, R. Sawaya, W. Shi, P.F. Thall, and N.E. Leeds. “Prognostic significance of preoperative MRI scans in glioblastoma multiforme”. In: *Neuro-Oncology* 27.1 (1996), pp. 66–73.

[49] C. Hartmann, B. Hentschel, W. Wick, D. Capper, J. Felsberg, M. Simon, M. Westphal, G. Schackert, R. Meyermann, T. Pietsch, G. Reifenberger, M. Weller, M. Loeffler, and A. von Deimling. “Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas”. In: *Acta Neuropathology* 120 (2010), pp. 707–718.

[50] Y. Hayashida, T. Hirai, S. Morishita, M. Kitajima, R. Murakami, Y. Korogi, K. Makino, H. Nakamura, I. Ikushima, M. Yamura, M. Kochi, J.I. Kuratsu, and Y. Yamashita. “Diffusion-weighted imaging of metastatic brain tumors: comparison with histologic type and tumor cellularity”. In: *American Journal Neuroradiology* 27.7 (2006), pp. 1419–1425.

[51] M.E. Hegi, A.C. Diserens, T. Gorlia, M.F. Hamou, N. de Tribolet, M. Weller, J.M. Kros, J.A. Hainfellner, W. Mason, L. Mariani, J.E.C. Bromberg, P. Hau, R.O. Mirimanoff, J.G. Cairncross, R.C. Janzer, and R. Stupp. “MGMT gene silencing and benefit from temozolomide in glioblastoma”. In: *New England Journal of Medicine* 352.10 (2005), pp. 997–1003.

[52] S. Higano, X. Yun, T. Kumabe, M. Watanabe, S. Mugikura, A. Umetsu, A. Sato, T. Yamada, and S. Takahashi. “Malignant astrocytic tumors: clinical importance of apparent diffusion coefficient in prediction of grade and prognosis”. In: *Radiology* 241.3 (2006), pp. 839–846.

[53] M. Hohwieler Schloss, S.R. Freidberg, G.J. Heatley, and T.C. Lo. “Glucocorticoid dependency as a prognostic factor in radiotherapy for cerebral gliomas”. In: *Acta Oncology* 28.1 (1989), pp. 51–55.

[54] W. Hollingworth, L.S. Medina, R.E. Lenkinski, D.K. Shibata, B. Bernal, D. Zurakowski, B. Comstock, and J.G. Jarvik. “A systematic literature review of magnetic resonance spectroscopy for the characterization of brain tumors”. In: *American Journal of Neuroradiology* 27 (2006), pp. 1404–1411.

[55] D.W. Hosmer, S. Lemeshow, and S. May. *Applied survival analysis: regression modeling of time-to-event data*. Hoboken, NJ: Wiley Series in Probability and Statistics, 2008.

[56] C. Houillier, X. Wang, G. Kaloshi, K.Mokhtari, R. Guillevin, J. Laffaire, S. Paris, B. Boisselier, A. Idbaih, F. Laigle-Donadey, K. Hoang-Xuan, M. Sanson M, and J.Y. Delattre. "IDH1 or IDH2 mutations predict longer survival and response to temozolamide in low-grade gliomas". In: *Neurology* 75 (2010), pp. 1560–1566.

[57] F.A. Howe, S.J. Barton, S.A. Cudlip, M. Stubbs, D.E. Saunders, M. Murphy, P. Wilkins, K.S. Opstad, V.L. Doyle, M.A. McLean, B.A. Bell, and J.R. Griffiths. "Metabolic profiles of human brain tumours using quantitative in vivo 1H magnetic resonance spectroscopy". In: *Magnetic Resonance in Medicine* 49 (2003), pp. 223–232.

[58] M.I. Jordan. *An Introduction to probabilistic graphical models*. In preparation, 2007.

[59] A.H. Kaye and E.R. Laws Jr. *Brain tumors*. Edinburg, UK: Churchill Livingstone, 1995.

[60] B.M. Kelm, B.H. Menze, C.M. Zechmann, K.T. Baudendiestel, and F.A. Hamprecht. "Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: pattern recognition vs. quantification". In: *Magnetic Resonance in Medicine* 57 (2007), pp. 150–159.

[61] R. Kohavi and G.H. John. "Wrappers for feature subset selection". In: *Artificial Intelligence* 97.1 (1997), pp. 273–324.

[62] M. Lacroix, D. Abi-Said, D.R. Fourney, Z.L. Gokaslan, W. Shi, F. DeMonte, F.F. Lang, I.E. McCutcheon, S.J. Hassenbusch, E. Holland, K. Hess, C. Michael, D. Miller, and R. Sawaya. "A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival". In: *Journal of Neurosurgery* 95.2 (2001), pp. 190–198.

[63] C. Ladroue. "Pattern recognition techniques for the study of magnetic resonance spectra of brain tumours". PhD thesis. St. Georgeis Hospital Medical School, 2003.

[64] N. Landwehr, M. Hall, and E. Frank. "Logistic model trees". In: *Machine Learning* 59.1-2 (2005), pp. 161–205.

[65] M. Law, S. Yang, H. Wang, J.S. Babb, G. Johnson, S. Cha, E.A. Knopp, and D. Zagzag. "Glioma grading: sensitivity, specificity, and predictive values of perfusion MR imaging and proton MR spectroscopic imaging compared with conventional MR imaging". In: *American Journal of Neuroradiology* 24.10 (2003), pp. 1989–1998.

[66] M.C. Lee, S.M. Chang, S. Cha, and S.J. Nelson. "Dynamic susceptibility contrast perfusion imaging of radiation effects in normal appearing brain tissue: changes in the first-pass and recirculation phases". In: *Magnetic Resonance Imaging* 19.2 (2005), pp. 123–129.

[67] J.M. Legler, L.A. Ries, M.A. Smith, J.L. Warren, E.F. Heineman, R.S. Kaplan, and M.S. Linet. "Cancer surveillance series [corrected]: brain and other central nervous system cancers: recent trends in incidence and mortality". In: *Journal of National Cancer Institute* 91 (1999), pp. 1382–1390.

[68] Y. Li, J.M. Lupo, M.Y. Polley, J.C. Crane, W. Bian, S. Cha, S.M. Chang, and S.J. Nelson. "Serial analysis of imaging parameters in patients with newly diagnosed glioblastoma multiforme". In: *Neuro-Oncology* 13.5 (2011), pp. 546–557.

[69] J.C. Lindon, E. Holmes, and J.K. Nicholson. “Pattern recognition methods and applications in biomedical magnetic resonance”. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 39 (2001), pp. 1–40.

[70] P.J.G. Lisboa, I.O. Ellis, A.R. Green, F. Ambrogi, and M.B. Dias. “Cluster-based visualisation with scatter matrices”. In: *PatternRecognition Letters* 29.13 (2008), pp. 1814–1823.

[71] H. Liu and R. Setiono. “Chi2: feature selection and discretization of numeric attributes”. In: *Proceedings of the 7th International Conference on Tools with Artificial Intelligence*. Ed. by IEEE Computer Society. Los Alamitos, CA: IEEE Computer Society Press, 1995, pp. 388–391.

[72] D.N. Louis. *Classification of tumours: WHO classification of tumors of the central nervous system*. World Health Organization, 2007.

[73] L. Lukas, A. Devos, J.A. Suykens, L. Vanhamme, F.A. Howe, C. Majos, A. Moreno-Torres, M. Van der Graaf, A.R. Tate, C. Arus, and S. Van Huffel. “Brain tumor classification based on long echo proton MRS signals”. In: *Artificial Intelligence in Medicine* 31 (2004), pp. 73–89.

[74] J. Luts. “Classification of brain tumors based on magnetic resonance spectroscopy”. PhD thesis. Katholieke Universiteit Leuven, 2010.

[75] M.C. Martinez-Bisbal, L. Marti-Bonmati, J. Piquer, A. Revert, P. Ferrer, J.L. Llacer, M. Piotto, O. Assemat, and B. Celda. “1H and 13C HR-MAS spectroscopy of intact biopsy samples ex vivo and in vivo 1H MRS study of human high grade gliomas”. In: *Nuclear Magnetic Resonance in Biomedicine* 17 (2004), pp. 191–205.

[76] M.C. Martinez-Bisbal, D. Monleon, O. Assemat, M. Piotto, J. Piquer, J.L. Llacer, and B. Celda. “Determination of metabolite concentrations in human brain tumour biopsy samples using HR-MAS and ERETIC measurements”. In: *Nuclear Magnetic Resonance in Biomedicine* 22.2 (2009), pp. 199–206.

[77] T.R. McKnight. “Proton magnetic resonance spectroscopic evaluation of brain tumor metabolism”. In: *Seminars in Oncology* 31 (2004), pp. 605–617.

[78] T.R. McKnight, K.R. Lamborn, T.D. Love, M.S. Berger, S. Chang, W.P. Dillon, A. Bollen, and S.J. Nelson. “Correlation of magnetic resonance spectroscopic and growth characteristics within grades II and III gliomas”. In: *Journal of Neurosurgery* 106 (2007), pp. 660–666.

[79] T.R. McKnight, S.M. Noworolski, D.B. Vigneron, and S.J. Nelson. “An automated technique for the quantitative assessment of 3D-MRSI data from patients with glioma”. In: *Journal of Magnetic Resonance Imaging* 13.2 (2001), pp. 167–177.

[80] R.E. McLendon, D.D. Bigner, S.H. Bigner, and J.M. Provenzale. *Pathology of tumors of the central nervous system: a guide to histologic diagnosis*. New York, NY: Oxford University Press, 2000.

[81] B.H. Menze, M.P. Lichy, P. Bachert, B.M. Kelm, H.P. Schlemmer, and F.A. Hamprecht. “Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors”. In: *NMR in Biomedicine* 19 (2006), pp. 599–609.

[82] V. Metsis, H. Huang, F. Makedon, and A. Tzika. “Heterogeneous data fusion to type brain tumor biopsies”. In: *Proceedings of the 5th Conference on Artificial Intelligence Applications and Innovations*. Ed. by Iliadis, Maglogiann, Tsoumakasis, Vlahavas, and Bramer. IFIP Advances in Information and Communication Technology. Thessaloniki, Greece: Springer, Apr. 2009, pp. 233–240.

[83] T. Mitchell. *Machine learning*. Boston, MA: McGraw Hill, 1997.

[84] R. Nafe, S. Herminghaus, U. Pilatus, E. Hattingen, G. Marquardt, W. Schlote, H. Lanfermann, and F. Zanella. “Morphology of proliferating and non-proliferating tumor cell nuclei in glioblastomas correlates with preoperative data from proton-MR-spectroscopy”. In: *Neuropathology* 24.4 (Sept. 2004), pp. 172–182.

[85] W.G. Negendank, R. Sauter, T.R. Brown, J.L. Evelhoch, A. Falini, E.D. Gotsis, A. Heerschap, K. Kamada, B.C. Lee, M.M. Mengeot, E. Moser, K.A. Padavic-Shaller, J.A. Sanders, T.A. Spraggins, A.E. Stillman, B. Terwey, T.J. Vogl, K. Wicklow, and R.A. Zimmerman. “Proton magnetic resonance spectroscopy in patients with glial tumors: a multicenter study”. In: *Neurosurgery* 84 (1996), pp. 449–458.

[86] S.J. Nelson. “Analysis of volume MRI and MR spectroscopic imaging data for the evaluation of patients with brain tumors”. In: *Magnetic Resonance in Medicine* 46 (2001), pp. 228–239.

[87] S.J. Nelson. “Multivoxel magnetic resonance spectroscopy of brain tumors”. In: *Molecular Cancer Therapy* 2.5 (2003), pp. 497–507.

[88] K.S. Opstad, B.A. Bell, J.R. Griffiths, and F.A. Howe. “An investigation of human brain tumour lipids by high-resolution magic angle spinning 1H MRS and histological analysis”. In: *Nuclear Magnetic Resonance in Biomedicine* 21 (2008), pp. 677–685.

[89] K.S. Opstad, B.A. Bell, J.R. Griffiths, and F.A. Howe. “Toward accurate quantification of metabolites, lipids, and macromolecules in HRMAS spectra of human brain tumor biopsies using LCModel”. In: *Magnetic Resonance in Medicine* 60.5 (2008), pp. 1237–1242.

[90] K.S. Opstad, C. Ladroue, B.A. Bell, J.R. Griffiths, and F.A. Howe. “Linear discriminant analysis of brain tumour 1H MR spectra: a comparison of classification using whole spectra versus metabolite quantification”. In: *Nuclear Magnetic Resonance in Biomedicine* 20 (2007), pp. 763–770.

[91] K.S. Opstad, A.J. Wright, B.A. Bell, J.R. Griffiths, and F.A. Howe. “Correlations between in vivo 1H MRS and ex vivo 1H HRMAS metabolite measurements in adult human gliomas”. In: *Journal of Magnetic Resonance Imaging* 31 (2010), pp. 289–297.

[92] D.W. Parsons, S. Jones, X. Zhang, J.C. Lin, R.J. Leary, P. Angenendt, P. Mankoo, H. Carter, I.M. Siu, G.L. Gallia, A. Olivi, R. McLendon, B. A. Rasheed, S. Keir, T. Nikolskaya, Y. Nikolsky, D.A. Busam, H. Tekleab, L.A. Diaz Jr., J. Hartigan, D.R. Smith, R.L. Strausberg, S.K.N. Marie, S.M.O. Shinjo, H. Yan5, G.J. Riggins, D.D. Bigner, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V.E. Velculescu, and K.W. Kinzler. "An integrated genomic analysis of human glioblastoma multiforme". In: *Science* 321 (2008), pp. 1807–1812.

[93] P. Pels. "Analysis and Improvement of quantification algorithms for magnetic resonance spectroscopy". PhD thesis. Katholieke Universiteit Leuven, 2005.

[94] J.C. Platt. "Fast training of support vector machines using sequential minimal optimization". In: Cambridge, MA: MIT Press, 1999, pp. 185–208.

[95] J.C. Platt. *Sequential minimal optimization: a fast algorithm for training support vector machines*. Tech. rep. Microsoft Research, 1998.

[96] J.B. Poullet. "Quantification and classification of magnetic resonance spectroscopic data for brain tumor diagnosis". PhD thesis. Katholieke Universiteit Leuven, 2008.

[97] J.B. Poullet, M.C. Martinez-Bisbal, D. Valverde, D. Monleon, B. Celda, C. Arus, and S. Van Huffel. "Quantification and classification of high-resolution magic angle spinning data for brain tumor diagnosis". In: *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Ed. by IEEE Engineering in Medicine and Biology Society. Piscataway, NJ: IEEE, 2007, pp. 5407–5410.

[98] M.C. Preul, Z. Caramanos, R. Leblanc, J.G. Villemure, and D.L. Arnold. "Using pattern analysis of in vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors". In: *Nuclear Magnetic Resonance in Biomedicine* 11.4-5 (1998), pp. 192–200.

[99] J.R. Quinlan. "C4.5: Programs for Machine Learning". In: *Machine Learning* 16.3 (1994), pp. 235–240.

[100] H. Ratiney, M.J. Albers, H. Rabeson, and J. Kurhanewicz. "Semi-parametric time-domain quantification of HR-MAS data from prostate tissue". In: *Nuclear Magnetic Resonance in Biomedicine* 23 (2010), pp. 1146–1157.

[101] V. Righi. "Biochemistry in healthy and neoplastic human tissues: metabolic alteration revealed by HR-MAS nuclear magnetic resonance spectroscopy". PhD thesis. Universita di Bollogna, 2008.

[102] V. Righi, J.M. Roda, J. Paz, A. Mucci, V. Tugnoli, G. Rodriguez-Tarduchy, L. Barrios, L. Schenetti, S. Cerdan, and M.L. Garcia-Martin. "1H HR-MAS and genomic analysis of human tumor biopsies discriminate between high and low grade astrocytomas". In: *Nuclear Magnetic Resonance in Biomedicine* 22 (2009), pp. 629–637.

[103] Y. Ruano, T. Ribalta, A. Rodriguez de Lope, Y. Campos-Martin, C. Fiano, E. Perez-Magan, J.L. Hernandez-Moneo, M. Mollejo, and B. Melendez. “Worse outcome in primary glioblastoma multiforme with concurrent epidermal growth factor receptor and p53 alteration”. In: *American Journal of Clinical Pathology* 131 (2009), pp. 257–263.

[104] A. Rutter, H. Hugenholtz, J.K. Saunders, and I.C.P. Smith. “Classification of brain tumors by ex vivo 1H NMR spectroscopy”. In: *Journal of Neurochemistry* 64.4 (1995), pp. 1655–1661.

[105] S. Saraswathy, F.W. Crawford, K.R. Lamborn, A. Pirzkall, S. Chang, S. Cha, and S.J. Nelson. “Evaluation of MR markers that predict survival in patients with newly diagnosed GBM prior to adjuvant therapy”. In: *Neuro-Oncology* 91.1 (2009), pp. 69–81.

[106] A. Croitor Sava, T. Laudadio, J.B. Poullet, D. Monleon, M.C. Martinez-Bisbal, B. Celda, and S. Van Huffel. “Combining HR-MAS and in vivo MRI and MRSI information for robust brain tumor recognition”. In: *Proceedings of the 4th European Conference of the International Federation for Medical and Biological Engineering (ECIFMBE 2008)*. Ed. by Jos Vander Sloten et al. Heidelberg, Germany: Springer, 2008, pp. 340–343.

[107] A. Croitor Sava, T. Laudadio, J.B. Poullet, D. Monleon, M.C. Martinez-Bisbal, B. Celda, and S. Van Huffel. *Data fusion of HR-MAS and in-vivo Information with application in brain tumor recognition*. Tech. rep. ESAT-SISTA, Katholieke University Leuven, 2008.

[108] A. Croitor Sava, M.C. Martinez-Bisbal, S. Van Huffel, J.M. Cerda, D.M. Sima, and B. Celda. “Ex vivo high resolution magic angle spinning metabolic profiles describe intratumoral histopathological tissue properties in adult human gliomas”. In: *Magnetic Resonance in Medicine* 65.2 (2011), pp. 320–328.

[109] J.L. Schafer. *Analysis of incomplete multivariate data*. London, UK: Chapman and Hall, 1997.

[110] H.J. Scherer. “Cerebral astrocytomas and their derivatives”. In: *American Journal of Cancer* (1940), pp. 159–198.

[111] J.A. Schwartzbaum, J.L. Fisher, K.D. Aldape, and M. Wrensch. “Epidemiology and molecular pathology of glioma”. In: *Nature Clinical Practice Neurology* 2 (2006), pp. 494–503.

[112] J.N. Scott, N.B. Rewcastle, P.M. Brasher, D. Fulton, N.A. Hagen, J.A. MacKinnon, G. Sutherland, J.G. Cairncross, and P. Forsyth. “Long-term glioblastoma multiforme survivors: a population-based study”. In: *Canadian Journal of Neurological Sciences* 25 (1998), pp. 197–201.

[113] A.W. Simonetti, W.J. Melssen, F.S. de Edelenyi, J.J.A. van Asten, A. Heerschap, and L.M.C. Buydens. “Combination of feature-reduced MR spectroscopic and MR imaging data for improved brain tumor classification”. In: *Nuclear Magnetic Resonance in Biomedicine* 18.1 (2005), pp. 34–43.

[114] A.W. Simonetti, W.J. Melssen, M. van der Graaf, G.J. Postma, A. Heerschap, and L.M. C. Buydens. “A chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy”. In: *Analytical Chemistry* 75.20 (2003), pp. 5352–5361.

[115] A.W. Simonetti, J.B. Poullet, D.M. Sima, B. De Neuter, L. Vanhamme, P. Lemmerling, and S. Van Huffel. *An open source short echo time MR quantitation software solution: AQSES*. Tech. rep. 05-168. ESAT-SISTA, Katholieke University Leuven, 2006.

[116] J. Sled, A. Zijdenbos, and A. Evans. “A nonparametric method for automatic correction of intensity nonuniformity in MRI data”. In: *IEEE Transactions on Medical Imaging* 17 (1999), pp. 87–97.

[117] I.C.P. Smith and G.N. Chmurny. “NMR spectroscopy and cancer research: the present and the future”. In: *Analitical Chemistry* 62.15 (1990), 853A–890A.

[118] S.M. Smith. “Fast robust automated brain extraction”. In: *Human Bran Mapping* 17.3 (Nov. 2002), pp. 143–155.

[119] R.L. Somorjai, B. Dolenko, A.K. Nikulin, N. Pizzi, G. Scarth, P. Zhilkin, W. Halliday, D. Fewer, N. Hill, I. Ross, M. West, I.C. Smith, S.M. Donnelly, A.C. Kuesel, and K.M. Briere. “Classification of 1H MR spectra of human brain neoplasms: the influence of preprocessing and computerized consensus diagnosis on classification accuracy”. In: *Journal of Magnetic Resonance Imaging* 6.3 (1996), pp. 437–444.

[120] R. Srinivasan, J.J. Phillips, S.R. VandenBerg, M.Y.C. Polley, G. Bourne, A. Au, A. Pirzkall, S. Cha, S.M. Chang, and S.J. Nelson. “Ex vivo MR spectroscopic measure differentiates tumor from treatment effects in GBM”. In: *Neuro-Oncology* 12.11 (2010), pp. 1152–1161.

[121] A.M. Stark, J. Hedderich, Held-Feindt J., and H.M. Mehdorn. “Glioblastoma-the consequences of advanced patient age on treatment, survival”. In: *Neurosurgery Review* 30.1 (2007), pp. 56–62.

[122] D. Stefan, F. Di Cesare, A. Andrasescu, E. Popa, A. Lazariev, E. Vescovo, O. Strbak, S. Williams, Z. Starcuk, M. Cabanas, D. van Ormondt, and D. Graveron-Demilly. “Quantitation of magnetic resonance spectroscopy signals: the jMRUI software package”. In: *Measurement Science and Technology* 20.10 (2009), pp. 104035–104044.

[123] C. Studholme, D. L. Hill, and D. J. Hawkes. “An overlap invariant entropy measure of 3D medical image alignment”. In: *Pattern Recognition* 32.1 (Jan. 1999), pp. 71–86.

[124] W. Stummer, H.J. Reulen, T. Meinel, U. Pichlmeier, W. Schumacher, J.C. Tonn, V. Rohde, F. Oppel, B. Turowski, C. Woiciechowsky, K. Franz, and T. Pietsch. “Extent of resection and survival in glioblastoma multiforme: identification of and adjustment for bias”. In: *Neurosurgery* 62.3 (2008), pp. 564–576.

[125] R. Stupp, W.P. Mason, M.J. van den Bent, M. Weller, B. Fisher, M.J.B. Taphoorn, K. Belanger, A.A. Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R.C. Janzer, S.K. Ludwin, T. Gorlia, A. Allgeier, D. Lacombe, J.G. Cairncross, E. Eisenhauer, and R.O. Mirimanoff. “Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma”. In: *New England Journal of Medicine* 352.10 (2005), pp. 987–996.

[126] T. Sugahara, Y. Korogi, M. Kochi, I. Ikushima, T. Hirai, and Okuda et al. “Correlation of MR imaging-determined cerebral blood volume maps with histologic and angiographic determination of vascularity of gliomas”. In: *American Journal of Roentgenology* 171.6 (1998), pp. 1479–1486.

[127] T. Sugahara, Y. Korogi, M. Kochi, I. Ikushima, Y. Shigematu, T. Hirai, T. Okuda, L. Liang, Y. Ge, Y. Komohara, Y. Ushio, and M. Takahashi. “Usefulness of diffusion-weighted MRI with echo-planar technique in the evaluation of cellularity in gliomas”. In: *Journal of Magnetic Resonance Imaging* 9.1 (1999), pp. 53–60.

[128] A.R. Tate, J.R. Griffiths, I. Martinez-Perez, A. Moreno, I. Barba, M.E. Cabanas, D. Watson, J. Alonso, F. Bartumeus, F. Isamat, I. Ferrer, F. Vila, E. Ferrer, A. Capdevila, and C. Arus. “Towards a method for automated classification of 1H MRS spectra from brain tumours”. In: *Nuclear Magnetic Resonance in Biomedicine* 11 (1998), pp. 177–191.

[129] A.R. Tate, J. Underwood, D.M. Acosta, M. Julia-Sape, C. Majos, A.l. Moreno-Torres, F.A. Howe, M. van der Graaf, V. Lefournier, M.M. Murphy, A. Loosemore, C. Ladroue, P. Wesseling, J. Luc Bosson, M.E. Cabanas, A.W. Simonetti, W. Gajewicz, J. Calvar, A. Capdevila, P.R. Wilkins, B.A. Bell, C. Remy, A. Heerschap, D. Watson, J.R. Griffiths, and C. Arus. “Development of a decision support system for diagnosis and grading of brain tumours using in vivo magnetic resonance single voxel spectra”. In: *Nuclear Magnetic Resonance in Biomedicine* 19 (2006), pp. 411–434.

[130] T. Tihan, J. Barletta, I. Parney, K. Lamborn, Sneed P.K., and S. Chang. “Prognostic value of detecting recurrent glioblastoma multiforme in surgical specimens from patients after radiotherapy: should pathology evaluation alter treatment decisions”. In: *Human Pathology* 37.3 (2006), pp. 272–282.

[131] R. Tosi and V. Tugnoli (editors). *Nuclear magnetic resonance spectroscopy in the study of neoplastic tissue*. New York, NY: Nova Science Publishers, 2005.

[132] V. Van Belle, J.B. Poullet, D. Monleon, B. Celda, M. Martinez-Bisba, and S. Van Huffel. “Classification of brain tumors by means of high resolution magic angle spinning spectra”. In: *Proceedings of the Belgian Day on Biomedical Engineering of the IEEE Engineering in Medicine and Biology Society Symposium*. Piscataway, NJ: IEEE, 2006, pp. 199–202.

[133] P.Y. Wen and S. Kesari. “Malignant Gliomas in Adults”. In: *The New England Journal of Medicine* 359 (2008), pp. 492–507.

[134] I.H. Witten and E. Frank. *Data mining: practical machine learning tools and techniques*. 2nd. The Morgan Kaufmann Series in Data Management Systems. San Francisco, CA: Morgan Kaufmann Publishers, 2005.

[135] A.J. Wright, G. A. Fellows, J.R. Griffiths, M. Wilson, B.A. Bell, and F. A. Howe. “Ex-vivo HRMAS of adult brain tumours: metabolite quantification and assignment of tumour biomarkers”. In: *Molecular Cancer* 9 (2010), pp. 66–84.

[136] L. Wy. *Mixed effects models for complex data*. Boca Raton, FL: CRC Pres, Taylor and Francis Group, 2010.

[137] H. Yan, D.W. Parsons, G. Jin, R. McLendon, B.A. Rasheed, W. Yuan, I. Kos, I. Batinic-Haberle, S. Jones, G.J. Riggins, H. Friedman, A. Friedman, D. Reardon, J. Herndon, K.W. Kinzler, V.E. Velculescu, B. Vogelstein, and D.D. Bigner. “IDH1 and IDH2 mutations in gliomas”. In: *New England Journal of Medicine* 360 (2009), pp. 765–773.

[138] H. Yerli, A.M. Agildere, O. Ozen, B. Geyik E.and Atalay, and A.H. Elhan. “Evaluation of cerebral glioma grade by using normal side creatine as an internal reference in multi-voxel 1H-MR spectroscopy”. In: *Diagnostic and Interventional Radiology* 13.1 (Mar. 2007), pp. 3–9.

[139] E.I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E.R. Melhem, and C. Davatzikos. “Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme.” In: *Magnetic Resonance in Medicine* (2009).

[140] Y. Zhang, M. Brady, and S. Smith. “Segmentation of brain MR images through a Hidden Markov Random Field model and the Expectation-Maximization algorithm”. In: *IEEE Transactions on Medical Imaging* 20.1 (2001), pp. 45–57.