Automated Classification of Cerebral Gliomas by Means of Quantitative Emission Tomography and Multimodal Imaging

Automatische Klassifizierung von Gliomen des menschlichen Gehirns auf Basis quantitativer Emissionstomographie und multimodaler Bildgebung

Der Technischen Fakultät
der Friedrich-Alexander-Universität
Erlangen-Nürnberg
zur
Erlangung des Doktorgrades

DOKTOR-INGENIEUR

vorgelegt von

Philipp M. Ritt
aus Duisburg

Als Dissertation genehmigt
von der Technischen Fakultät der
Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung:	13.12.2013
Vorsitzende des Promotionsorgans:	Prof. Dr.-Ing. habil. M. Merklein
Gutachter:	Prof. Dr.-Ing. J. Hornegger
	Prof. Dr. rer. nat. G. Anton

Abstract

Cerebral gliomas represent a common type of cancer of the human brain with many tumor grades which express a huge diversity in growth characteristics and have a highly varying malignancy. The optimal treatment for a cerebral glioma is only ensured if the underlying tumor grade is known. One very common grading scheme is the *World Health Organization (WHO) Classification of tumors of the central nervous system*, which differentiates four grades. The de facto standard of grading a glioma is based on bioptic samples which are obtained in invasive interventions. These interventions pose significant risks for the patients and add more time delays between an initial evidence of the tumor, usually found by X-ray computed tomography (CT) or magnetic resonance imaging (MRI) and the initiation of a treatment.
On the other side, versatile imaging modalities like CT, MRI and from the field of nuclear medicine, positron emission tomography (PET) cover various aspects of the morphology and physiology of a tumor. The information gained from medical imaging thus can indicate the grade of a cerebral glioma without any invasive intervention. The multimodal imaging often results in a high complexity that makes if difficult to diagnose the malignancy solely based on the visual interpretation of medical images.

In this thesis, we present approaches for an extensive pattern recognition pipeline for the grading of cerebral gliomas based on tomographic datasets from MRI, CT, and PET. More specifically, we use gadolinium contrast-enhanced T_1-weighted MRI, T_2-weighted fluid attenuated inversion recovery MRI, diffusion-weighted MRI, non contrast-enhanced low-dose X-ray CT, and dynamic (multiple acquired time frames) [18F]-Fluor-Ethyl-Tyrosine (FET) PET. Our setup includes image preprocessing, feature extraction and calculation, feature normalization, and finally fully automatic classification.
We propose the imaging modalities and the classifiers which performed best for our patient population and show that inter-dataset normalization as a preprocessing step helps to improve the classification rate for cerebral gliomas. As the PET is acquired over a lengthy time period which can lead to substantial patient motion, we present a retrospective motion correction technique based on image registration, which improves the image quality of the PET data.
The presented approaches underline that diagnostic statements can be gained from highly complex, multimodal image data in an automated fashion. We can differentiate not only low- and high-grade tumors, but also aid in distinguishing between the four WHO grades within some limitations.

Kurzfassung

Gliome repräsentieren eine häufige Krebserkrankung des menschlichen Gehirns. Es werden mehrere Gliomgrade unterschieden, die wiederum große Variabilität bezüglich ihres Wachstumsverhalten und ihrer Malignität aufweisen. Die optimale Behandlung eines Glioms ist nur sichergestellt, wenn der zugrundeliegende Tumorgrad bekannt ist. Ein verbreitetes Klassifikationsschema auf diesem Gebiet ist die *Klassifikation von Tumoren des zentralen Nervensystems der Weltgesundheitsorganisation (WHO)*, welches vier Tumorgrade differenziert. Der Standard bezüglich der Bestimmung des Tumorgrades ist die histopathologische Aufarbeitung von bioptischen Proben, die in invasiven Verfahren gewonnen werden. Diese Eingriffe stellen jedoch ein Risiko für den Patienten dar. Darüber hinaus tragen sie zu einem Zeitverzug zwischen dem initialen Hinweis auf einen Tumor, häufig gewonnen durch die medizinische Bildgebung, und der Einleitung einer Behandlung bei.
Im Gegensatz zu invasiven Methoden existieren verschiedene Bildgebungsverfahren wie die Röntgen-Computertomographie (CT), die Magnetresonanztomographie (MRT) und, aus dem Bereich der Nuklearmedizin, die Positronenemissionstomographie (PET). Diese Bildgebungsmodalitäten sind in der Lage, umfassende Aspekte der Tumorphysiologie und -morphologie darzustellen. Folglich können mithilfe jener Verfahren Indizien für den zugrundeliegenden Grad des Tumors gewonnen werden. Die multimodale Bildgebung generiert jedoch eine hohe Komplexität, die eine auf reiner Bildbetrachtung basierende Malignitätsdiagnostik erschwert.
In der vorliegenden Arbeit stellen wir Ansätze zur Beurteilung des Tumorgrads unter Anwendung von Mustererkennungsverfahren auf Bilddaten der MRT, der PET und der CT vor. Im Detail verwenden wir gadoliniumkontrastiertes, T_1-gewichtetes MRT, T_2-FLAIR MRT, diffusionsgewichtetes MRT, natives Niedrigdosis-CT und dynamisches PET der mit F-18-Ethyltyrosin (FET) dargestellten zerebralen Aminosäureaufnahme. Die vorgestellten Methoden umfassen die Vorverarbeitung der medizinischen Bilder, die Merkmalsextraktion und -berechnung, die Normierung der Merkmale und die vollautomatische Klassifizierung.
Wir ermitteln die besten Modalitäten und Klassifikatoren auf Basis unserer Patientenpopulation und zeigen, dass eine Normalisierung der Datensätze im Zuge der Datenvorverarbeitung die Klassifikationsrate erhöhen kann. Für die PET, die aufgrund ihrer ausgedehnten Aufnahmedauer potentiell in besonderem Maße von Bewegungsartefakten betroffen ist, stellen wir eine Bewegungskorrekturmethode vor, welche auf starrer und retrospektiver Bildregistrierung basiert. Diese Korrektur verbessert die Bildqualität der PET signifikant.
Weiterhin zeigen wir, dass mit unserem automatisierten Ansatz mit hoher Genauigkeit diagnostische Aussagen aus hochkomplexen, multimodalen Bilddaten gewonnen werden können. Nicht nur die Unterscheidung von niedrig- und hochgradigen Tumoren, sondern darüber hinaus die Abgrenzung der vier WHO-Grade kann in gewissen Grenzen realisiert werden.

Acknowledgment

First and foremost I would like to express my gratitude to Prof. Dr. Joachim Hornegger, who agreed to supervise my PhD after already co-supervising my diploma thesis. He always had an open ear for my problems, even despite his numerous obligations at the Pattern Recognition Lab and as Vice-President of the University of Erlangen-Nuremberg.
I am very thankful to Prof. Dr. Gisela Anton, as she readily agreed to co-supervise and review my thesis and always supported the interdisciplinary research between the Faculty of Engineering and the Faculty of Sciences.
I am indebted to Prof. Dr. Torsten Kuwert for employing me as medical physicist at the Clinic for Nuclear Medicine, giving me the opportunity to do research on medical imaging and for providing me advice that was often beyond job-related matters.
Many thanks also go to Dr. Stefan Steidl, Michal Cachovan, and Bharath Navalpakkam for the proof-reading of this thesis and for their support. For helping me on scientific questions regarding pattern recognition and magnetic resonance imaging, I'd like to thank Prof. Dr. Harald Quick, Prof. Dr. Björn Eskofier, and Florian Hönig.
At the Clinic of Nuclear Medicine, I would like to express my deepest gratitude to PD Dr. Rainer Linke, PD Dr. Daniela Schmidt, Dr. Torsten Fritscher, Dr. Carsten Hocke, Gerson Schütze, Michael Beck, Dr. Carolin Seiffert, Dr. Julia Bachschmidt, Evelyn Walz, and the technicians Anja Hocke, Valentina Totev, Marlene Winkler, Daniela Schultes, Waltraud Forster, Willi Amann, and Wolfgang Sprengart for their help and advice and for making the time at the clinic nice and special.
Last but not least, I would like to thank my parents, my sister Katharina and my brother Korbinian, all of my friends, and my girlfriend Katja for their moral support, their encouragement, and their patience with me.

Philipp M. Ritt

Contents

Chapter 1

Introduction

1.1 Motivation and Goal

The classification of pathological changes in the structure and function of the human body are daily tasks in medicine. In many cases, the decision for a class is based on the physician's knowledge and experience and is therefore rather subjective. Motivated by this fact, nowadays more and more guidelines and regulations for a classification based on hard metrics and quantitative measures are starting to prevail. This methodology ensures more objective and reproducible choices with less intra-, and inter-rater variability. Very lately, the fully automatic processing of medical data by methods of pattern recognition started to emerge. The examples of fully automated classification are versatile, starting from the well-established processing of blood samples and identification of blood cell types with neural networks [Ceel 07, Swol 03], automated analysis of electrocardiograms (ECG) (e. g. QT-time interval [Hnat 06]), automated seizure detection based on electroencephalograms (EEG) [Sack 11, Osor 11], automatic glaucoma detection on retinal fundus images of the human eye [Bock 10, Bern 11], classification of diseases of the human speech system (e. g. automatic evaluation of dysarthric speech [Noth 11] or the classification of the emotional state in children's speech [Stei 09]), and many more.
(Semi-)Automatic classification on the basis of three-dimensional medical image datasets seems to be very challenging and only few established applications exist, e. g. the computer aided diagnostic (CAD) systems for the detection of lesions in breast-MRI [Dorr 11] or in mammographic images [Oliv 10]. Marten et al. [Mart 04] describe another commercial available CAD system for the detection of nodules in thoracic CT, which is based on a variety of pattern recognition methods (e. g. thresholding, histograms, principal component analysis, and many more). Feulner et al. developed an automatic lymph node detection and segmentation in chest CT [Feul 12].
Additionally, it is a well-knownfact that data stemming from multiple modalities can improve the diagnostic performance of manual as well as of machine classification schemes, e. g. multimodal (ultrasound and mammography) classification of breast lesions [Druk 05, Sahi 09].
Cerebral gliomas are a common type of cancer with an incidence of 6 per 100 000 individuals per year in Germany [Well 04]. The most common subtypes of gliomas are glioblastoma (54%) and other astrocytomas (22%). Another subtype are oligo-

dendrogliomas. The different subtypes are subject to different treatments in clinical routine which is reflected in the guidelines for the diagnosis and treatment of this tumor entity [Well 04]. The differentiation of these subtypes, namely the classification of the tumor on basis of medical images and bioptic samples, is an important and challenging task. The gold standard in terms of diagnostic confidence is the grading of the tumor based on histologic analysis of invasively gained bioptic samples. The bioptic grading allows an accurate classification in over 90% of all cases [Well 04]. However, the process of taking the bioptic samples, whether being stereotactic, incisional, or excisional biopsies, in general results in an increased risk for complications during the surgery. This is reflected by a reported morbidity ranging from 5-9% [Sawi 98, McGi 05] for stereotactic biopsies, which evolves in the desire for accurate non-invasive grading.
It is of general consent that MRI of the human brain is the standard for an initial diagnosis of glial tumors. A variety of different MRI sequences are acquired to cover different aspects of the tumor. Nevertheless, other imaging modalities such as e. g. X-ray CT and emission computed tomography (ECT) using radiopharmaceuticals can add beneficial information for the task of tumor grading/classification. In the end, it is common that for each patient five to ten 3-D medical datasets from different modalities and different time points exist, each covering the same anatomical structures. This results in an increased complexity for the physician to diagnose and classify the multitude of images.
Machine classification approaches promise the ability to overcome the limitations of human beings, by providing classification on high-dimensional data based on objective measures/features.

The aim of this work is to develop and evaluate a machine classification approach based on objective measures for multimodal datasets of patients with gliomas of the brain with the help of established pattern recognition techniques.

1.2 State of the Art

In order to put the classification performance of our automatic approach based on multimodal image data into context, we compare our results with the results in the literature. As studies with comparable pathology, patient population, and image data are rare, we outline the state of the art for manual and single-modality classification as well as for automatic approaches. As gliomas are a pathology of the soft-tissue compartment of the human brain, MRI due to its high contrast on that field, is considered as the standard imaging modality. This is reflected by a high number of studies available in the literature: For example, Law et al. [Law 03] report a sensitivity and specificity for manual discrimination between low and high-grade gliomas on 160 patients solely on conventional MRI images (namely contrasted T_1 and T_2-FLAIR) of 72.5% and 65%. When features based on dynamic contrast-enhanced perfusion MRI (DCE-MRI) and MR spectroscopy (MRS) are taken into account, the overall classification performance could be increased to 93.3% sensitivity and 60% and specificity. More recent studies refine their classification methods based on MRS and DCE-MRI and achieve sensitivities and specificities of 91.5% and 100% respectively,

on 74 patients [Serv 11] and 94.7% and 93.7% on 51 patients [Arvi 09], again for the differentiation between low-grade and high-grade tumors. Machine classification approaches on MRI images are scarce, Zöllner et al. [Zoll 12, Zoll 10] report a sensitivity and specificity of 89% and 84% with their automatic approach based on DCE-MRI features, using a support vector machine (SVM) based classifier and feature reduction with principal component analysis (PCA). Zacharaki et al. [Zach 09] report a sensitivity and specificity of 85% and 96% for the differentiation of low- and high-grade gliomas using an SVM-based classifier as well. They extract shape, intensity, and texture features from T_1-, contrasted T_1-, T_2-, T_2-FLAIR-, and DCE-MRI out of manually defined regions of interest of 74 patients and perform SVM recursive feature elimination. In a later report, Zacharaki et al. [Zach 11] reach a classification accuracy of 94.5% for differentiation of low- and high-grade gliomas with a k-nearest neighbor classifier and 76.2% classification accuracy for a four-class problem (metastasis of other tumors, WHO2, WHO3, WHO4 gliomas). Besides DCE-MRI, which can be used to obtain perfusion parameters as cerebral blood flow and volume (CBF/CBV) or mean transit time (MTT), dynamic X-ray CT allows the calculation of CBV, CBF and MTT as well. For example, Ellika et al. [Elli 07] find a sensitivity and specificity of 92.9% and 100% on 19 patients with a manual classification based on CBF, CBV and MTT features for the differentiation of high- and low-grade gliomas. In a more recent study, Beppu et al. [Bepp 11] confirm these results and achieve a sensitivity and specificity of 90.9% and 83.3% for the same classification task (manual, low- vs. high-grade).
Although CT and MRI have a high sensitivity when it comes to finding intracranial suspicious formations [Lang 05, Sant 12], with PET or PET/CT, nuclear medicine offers a diagnostic device which is able to detect and grade gliomas as well. Various tracers are successfully applied for this task. Among them, the most commonly used are radioactive labeled sugars (FDG) and amino acids (FET, MET, FLT). FDG offers a sensitivity and specificity of 93% and 85% [Padm 03]. It is a general consent that amino-acids offer a better diagnostic performance for the differentiation of the subtypes of gliomas [Tsuc 08] and the differentiation from other diseases [Paul 06, Lau 10]. Pöpperl et al. [Popp 07, Popp 06] even report 94% sensitivity and 100% specificity for the manual classification of high- and low-grade gliomas on dynamic PET. However, in a recent study by Rapp et al. [Rapp 13] lower sensitivity and specificity of 80% and 65% are reported for the same task. Various studies indicate an increased manual classification accuracy when PET and MRI are combined, compared to the single modalities [Floe 05, Paul 05].

1.3 Contribution to the Progress of Research

The aim of this work was to implement a software setup, which covers all important steps for an automated classification for the application on multimodal image data of brain tumor patients and to evaluate the possibilities and limitations of this setup. The key parts of our solution are data registration, data normalization, feature extraction, feature preprocessing and classification. We achieved the following goals, which, to the best of our knowledge, have never been reported on that field:

- We were the first group that successfully classified gliomas from heterogeneous and multimodal datasets from clinical routine [Ritt 12, Ritt 13a], opposed to existing experiments that were carried out on data gained from single devices and single modalities. With our automated methods, we consistently achieved 10-15 % higher absolute classification rates for the patient population when compared to those extracted from the written medical reports of the patients. Additionally, we could show that deriving features from multimodal data greatly improves classification accuracy.

- For the derivation of features from multiple modalities, we developed a software framework which extracts data from three-dimensional, multimodal, static and dynamic image data (e. g. PET, SPECT, MRI, CT) and which, besides other pre-processing algorithms, incorporates a powerful rigid registration method. The accuracy of this registration of FET-PET images and MRI images has never been evaluated so far. From literature, only results for other PET tracers or artificial datasets were reported. To our best knowledge, we conducted the first clinical study for the accuracy of an automatic registration in this field. We found the accuracy to be excellent [Ritt 10, Kief 11] and sufficient for the multimodal feature extraction.

- Especially MRI and PET images show large intensity variations when acquired on differing machines, using differing imaging protocols. To overcome these limitations, we applied MRI normalization and quantitative accurate emission tomography in the field of the machine classification of gliomas [Ritt 11b, Ritt 13b], which has not been reported from literature yet. Additionally, we developed a motion correction technique in order to compensate for the potentially confounding influences of patient movement in dynamic PET.

- Besides the two-class problem of differentiating low and high tumor grades, more refined classification tasks are of huge interest. Ultimately, the differentiation of all of the four individual WHO grades for gliomas are the goal. However, this is a challenging task and thus has never been reported so far. With aforementioned techniques, we were able to differentiate the four WHO glioma grades to a large extent.

1.4 Structure of this Work

In the next chapter (Chapter 2) we provide the medical background for our classification methods (Chapter 5), namely, details and functions of the medical imaging devices (covering SPECT/PET, MRI, and X-ray CT) and the underlying pathophysiology of cancer and especially gliomas.
Chapter 3 provides the inclusion criteria for patients in our study and a description of the patient population. Additionally, the medical imaging devices and the imaging parameters that were used are given.
The necessary preprocessing that we apply, such as the definition of the volume of interest (VOI), inter-dataset normalization, image registration, and feature extraction/normalization and selection are explained in Chapter 4. Our dynamic PET

motion correction method is presented in the same chapter.
The different machine classifiers that we use on the preprocessed and extracted features are characterized in Chapter 5. We also provide the mathematical background of the classifiers and show examples for their decision boundaries in the same chapter. Additionally the procedure for a conventional receiver operating characteristic (ROC) of the gliomas is explained and measures for inter-rater concordance are introduced.
Chapter 6 presents the achieved classification results.
These results and the limitations of our study are discussed and compared to the state of the art in Chapter 7.
Chapter 8 provides a short summary of our work, followed by an outlook, motivation, and suggestions for future enhancement on the topic of machine classification in gliomas of the brain.

Chapter 2

Medical Imaging and Cerebral Gliomas

This chapter aims at providing some background on the medical imaging technology which is used for our study on classification of glioma grades. Herein, ECT and MRI are discussed more comprehensively compared to CT, as the majority of the experiments was carried out on those modalities. At the end of this chapter, a section about cerebral gliomas explains the foundations of this tumor class.

2.1 Quantitative Emission Tomography

Excerpts of the text in the current section were published in the *European Journal of Nuclear Medicine and Molecular Imaging* [Ritt 11b] and as a chapter in the book *Molecular Imaging in Oncology* [Ritt 13b]. Although these excerpts mainly cover aspects of SPECT, many methods are applicable for PET as well.

2.1.1 Introduction

Single-photon emission computed tomography (SPECT) and PET allow the visualization of the distribution of radioactivity within the human body. Both modalities are widely used for clinical purposes (for reviews, see e. g. Bockisch et al. [Bock 09], Schulthess et al. [Schu 06]). SPECT and PET also promise to exactly quantify the concentration of radioactivity within a given volume of tissue in absolute units, e. g. as kilobecquerels per cubic centimeter. This process is compromised by photon scatter [Jasz 84, Kora 88, Frey 94], photon attenuation [LaCr 94, Blan 96, Rome 06, El F 99] and partial volume artifacts [Kess 84, Gewo 00]. In PET, options to correct for these confounding variables have been developed and validated several decades ago (for a review, see e. g. Schelbert et al. [Sche 98], Boellard et al. [Boel 09]), owing to the technical advantages offered by positron decay and coincidence detection. In SPECT, technical progress in that regard has been considerably slower with various reports emphasizing the lack of quantitative abilities [Germ 01, Lewi 06, Sido 06]. For orientation, a simplified diagram of the image formation chain for quantitative PET and SPECT is given (Figure 2.1). We will start this overview with the reconstruc-

tion that nowadays incorporates many correction methods mentioned in the Sections 2.1.3-2.1.5. We will then focus on attenuation, scatter, and partial volume effects: In each of the paragraphs, a short explanation of the underlying effect and examples of correction techniques are given. In addition, we will briefly outline a calibration technique. We will conclude with a discussion of the potential of SPECT and PET quantification for clinical applications and present some validation studies. For simplification, we assume in the following that the kinetics of the activity distribution that is of interest are slow with respect to the imaging time. If this is not the case, significant quantification errors will occur and other approaches like dynamic SPECT and dynamic PET (as in our study) are beneficial [Gull 10].

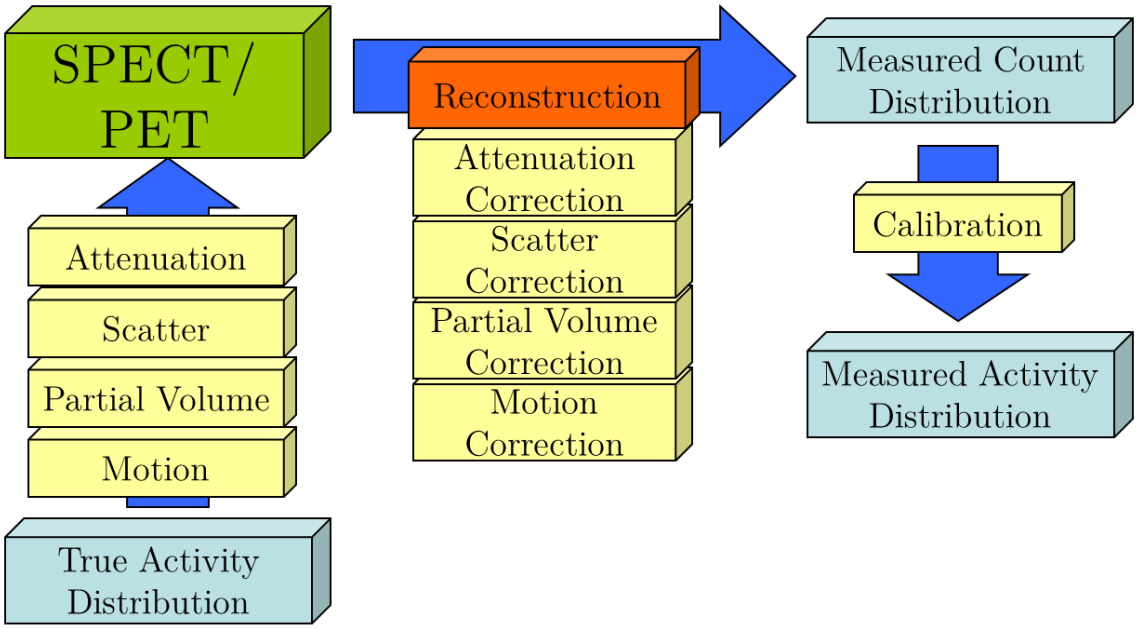


Figure 2.1: Illustration of a simplified image formation chain in ECT. The image of the true activity distribution is confounded by several effects including attenuation, scatter, partial volume and motion. The reconstruction, along with corrections for the mentioned effects, delivers a measured, three-dimensional count distribution. With a calibration step, the count distribution is translated into a measured activity concentration.

2.1.2 Quantitative Reconstruction

In general, two main families of reconstruction techniques are commonly used in clinical emission computed tomography: non-iterative (e. g. filtered back-projection, FBP) and iterative methods. Despite its higher demands on computation, iterative reconstruction seems to be superior with regard to quantification when compared to non-iterative methods. This is, in principle, mainly due to the possibility to implement corrections and system modeling methods more easily into iterative reconstruction than into non-iterative methods. Consequently, several publications report higher quantitative accuracy of iterative reconstructions compared to non-iterative

methods [Gill 91, Rose 95, Tsui 94].
For SPECT, in the following the most important parts of the system modeling will be explained (Figure 2.2). In general, the application of collimator-detector response models in the reconstruction leads to improved system resolution and quantitative accuracy, as can be seen in the literature listed in the following paragraphs.
As an example, one assumption of the FBP algorithm is that the sensitive volume of one collimator hole is of cylindrical shape. However, in reality, the sensitive volume is more cone-shaped. As a result, the system resolution, defined as full width at half maximum (FWHM) of a point source depends approximately linearly on the distance between source and detector for a gamma camera that employs absorptive parallel hole collimation. This effect, known as geometric response function, can easily be incorporated as a mathematical model into an iterative reconstruction – in contrast to FBP. The intrinsic effects of the detector are characterized by the intrinsic response function. This function mainly describes the effect of scatter in the crystal itself and the uncertainty in the position estimation of a detected photon.
In SPECT imaging, a compromise between collimator efficiency and image quality has to be made. Thicker collimator septa reduce the amount of septal penetration, however, they also reduce the efficiency of the collimator by covering sensitive area of the detector. As a result of the trade-off, a certain amount of septal penetration is allowed (e. g. 5%). In general, the probability for photons penetrating the septa of the collimator is described by a septal penetration function and can be incorporated to correct for the effect. Another possible interaction between the photons and collimator septa is scatter. The probability that photons are scattered by the septa is modeled by the septal scatter function and is in general more important for medium and high energy nuclides.
The combination of the four parts of the response function (see Figure 2.3) is known as the collimator-detector response function. It is used to correct for the aforementioned effects in the reconstruction step and consequently helps to improve system resolution and quantitative accuracy.
Analytical corrections for the response functions that can be implemented in non-iterative techniques are possible. However, there are several reports that show that a superior resolution can be achieved when corrections are implemented using iterative reconstruction techniques [Kohl 98b] and quantitative measures [Kohl 98a, Pret 98]. For example, Römer et al. [Rome 06] report the use of a three-dimensional depth dependent blur modeling (OSEM 3-D reconstruction) in a clinical environment. A more detailed review on the modeling of the collimator-detector response function can be found in [Zaid 06b].
For clinical SPECT quantification, iterative reconstruction is state of the art and in general recommended. Corrections for the geometric response and the intrinsic response are considered to be more important than the modeling of septal scatter and septal penetration, at least for low energy imaging. They should be applied if available, as the improved spatial resolution will significantly decrease the confounding effect of partial volume and thus result in higher quantitative accuracy of small structures.

For PET, due to radically different imaging principles compared to SPECT, the system modeling is different as well. Because of its coincidence detection, a so called

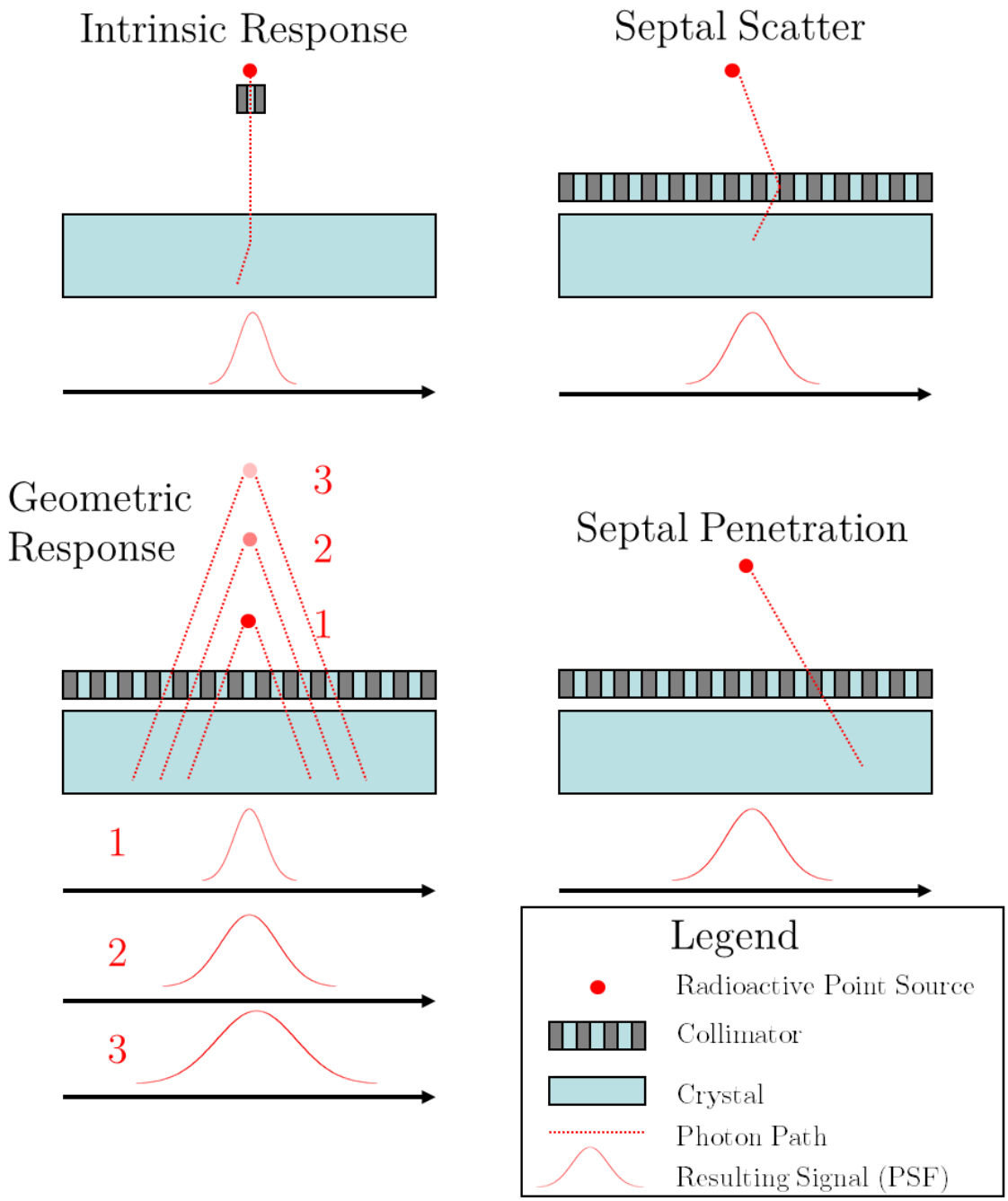


Figure 2.2: Simplified illustration of some components that form the collimator-detector response function. The geometric response function (lower left) models the source-to-collimator distance effects. The measured response in the crystal varies with source-to-collimator distance. The intrinsic response function (upper left) describes the effects of interactions in the detector crystal itself (the point source is collimated to form a pencil beam). The septal scatter function (upper right) and septal penetration function (lower right) model the interactions between gamma radiation and the collimator.

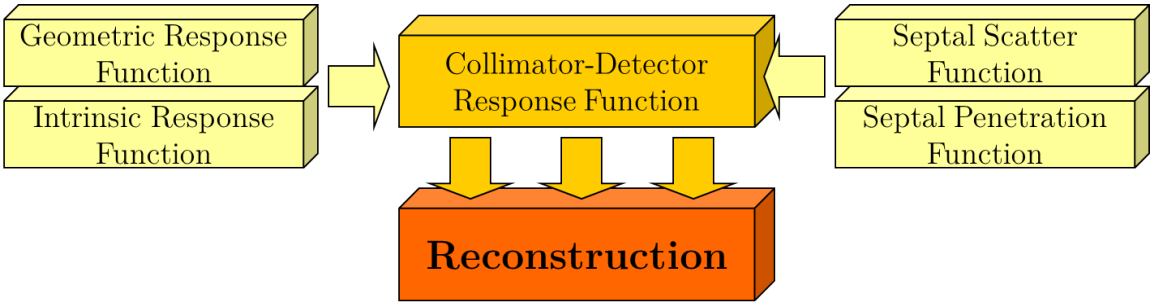
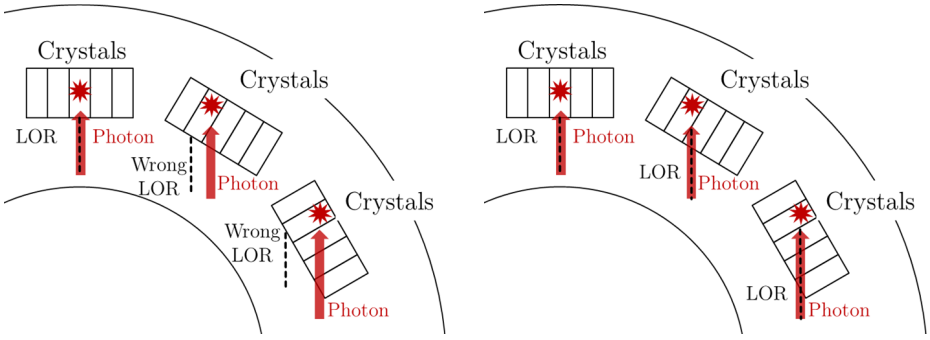


Figure 2.3: The combination of models for the intrinsic response, the geometric response, the septal scatter and the septal penetration is called the collimator-detector response function. It can be integrated into the (iterative) reconstruction step.

electronic collimation is applied. This omits the necessity of mechanical collimation in modern devices, compared to older devices with multiple detector rings that used collimators in order to prevent coincidence between different rings. However, problems similar to SPECT, like the effects of depth of interaction and crystal scatter exist. These effects can be modeled in an intrinsic response function and can be incorporated in the iterative reconstruction, as it is the case for SPECT. An example for this is Siemens TrueX reconstruction technique. It incorporates corrections for the depth of interaction effect. This effect otherwise would lead to the assumption of an erroneous line of response (LOR). The correction achieves an increased spatial resolution and increased quantitative accuracy, especially in the peripheral parts of the camera field of view [Ande 08]. For an illustration of this effect, see Figure 2.4



(a) Without TrueX (b) Siemens TrueX

Figure 2.4: Illustration of the depth of interaction effect for PET. Without corrections (a), the reconstruction assumes that the decay took place on a wrong line of response (LOR). The TrueX (b) iterative reconstruction technique corrects for the wrong LOR.

2.1.3 Attenuation Correction

ECT images are affected by attenuation artifacts. In the case of SPECT imaging, the probability P_{Det} that a gamma quantum emitted at position $\boldsymbol{d} \in \mathbb{R}^3$ (see Figure 2.5)

reaches the detector at position $\boldsymbol{D} \in \mathbb{R}^3$ (assumed that it is emitted in the proper direction) is calculated according to Equation 2.1 :

$$P_{\text{Det}} \propto \int_C -\mu(x,y,z)\,\mathrm{d}s \tag{2.1}$$

with the two endpoints of the path C given by $\boldsymbol{D}$ and $\boldsymbol{d}$.

$$P_{\text{Det}} \propto \int_C -\mu(x,y,z)\,\mathrm{d}s \tag{2.2}$$

with the two endpoints of the path C given by $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$.

The integral covers essentially the path C (by approximation a straight line) of the radiation from its origin through the object, to the location of detection. The probability for SPECT consequently depends on the (unknown) location of the decay $\boldsymbol{d}$ and on the linear attenuation coefficients $\mu(x,y,z)$ of the object. In contrary, in PET imaging the probability only depends on the linear attenuation coefficients along the line of response (LOR) $\boldsymbol{D_2} - \boldsymbol{D_1}$ where the decay happened, but not on the exact location in this LOR (Equation 2.2).
For the attenuation correction in the reconstruction step, the spatial distribution of the attenuation coefficients of the examined object for the photon energy of the radionuclide used needs to be known. Several methods for obtaining attenuation maps have been employed so far:
The maps can be estimated, if the contours of the object (e. g. via rough segmentation of the SPECT or PET image) and the attenuation coefficients are known (e. g. attenuation coefficient of water). The object can be assumed to be homogeneous with regard to this coefficient (Chang's correction [Chan 78]). This method is still applied very successfully to SPECT imaging of the brain, where only one class is predominant (soft brain tissue). However, it is not very accurate for ECT imaging of the thorax or pelvis, where large amount of other tissues (e. g. lung and bone) are present.
Another way of generating the attenuation maps is through a simple transformation of a transmission scan. The transmission images need to be converted to attenuation factors at the effective energy of the emission scan (140 keV for Tc-99m, 511 keV for PET) and corrected for the spatial registration between the emission and transmission images. The resulting attenuation map can be easily integrated into common iterative reconstruction techniques.
Before the introduction of hybrid SPECT/CT and PET/CT devices, radionuclide (source-based) transmission measurements (e. g. Gd-153, Tc-99m, Ba-133, Cs-137, Ge-68/Ga-68) were commonly employed. These source-based methods have the advantage that the same detectors could be used for both the emission and the transmission scans, which makes the method very cost-efficient. On the other hand, if using the same detectors severe disadvantages arise. This often leads to poor image quality due to a poor signal-to-noise ratio and due to the limited spatial resolution of the gamma camera detectors. Furthermore, because of radiation safety considerations, only relatively weak transmission sources are used, which results in lengthy transmission scans ($>$ 10 min). Additionally, the relatively high photon energies of the transmission sources (100-511 keV) lead to poor tissue contrast compared to X-ray

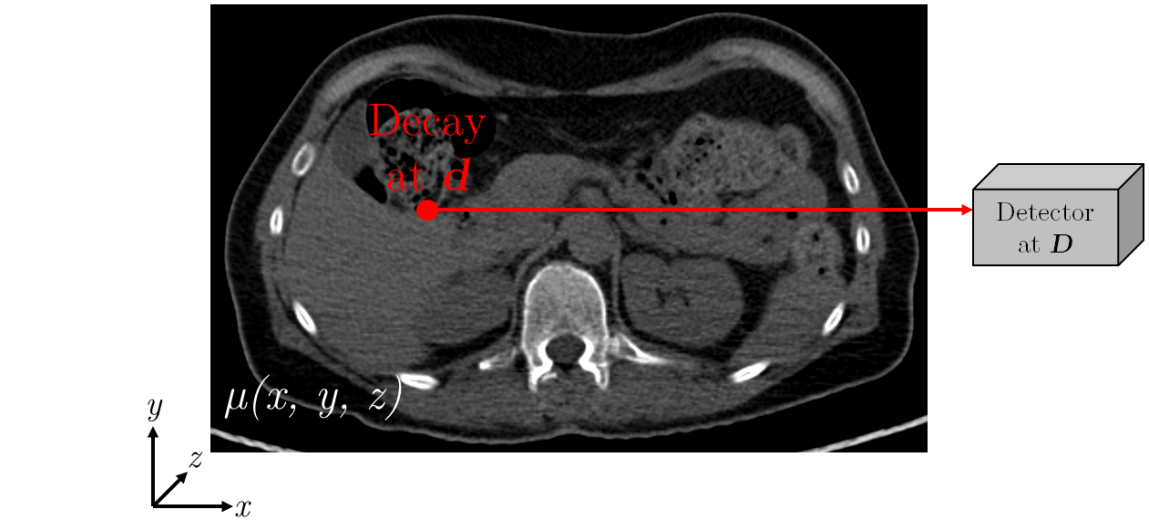
CT with effective energies of 40-90 keV. This also limits the quality of image fusion between emission and transmission images.
With the advent of hybrid devices and thus the availability of highly correlated high quality X-ray CT transmission scans, it is now standard to use these scans not only for high quality fused viewing but also for attenuation correction. The CT scans, usually in Hounsfield units, have to be converted to linear attenuation coefficients at the respective photon energy. It is important to note that the transformation of the CT transmission image to attenuation factors at the effective energy of the emission scan can introduce errors [Zaid 03]: First the transformation is specific for different acceleration voltages and beam filters of the CT. Second the polychromaticity of the X-ray beam also introduces artifacts, mainly caused by beam hardening. Patient motion that occurs between the emission and transmission acquisitions in a hybrid system can lead to artifacts, which in turn could lead to false readings of the SPECT images. A manual post-registration can help in most cases to prevent such artifacts [Chen 06].
Attenuation correction is mandatory for accurate ECT quantification. Methods that rely on the segmentation of contours in SPECT and PET are sufficient for quantification in relatively easy anatomies like the brain. Nevertheless, attenuation correction by CT is the clinical standard and highly recommended for quantification.

2.1.4 Scatter Correction

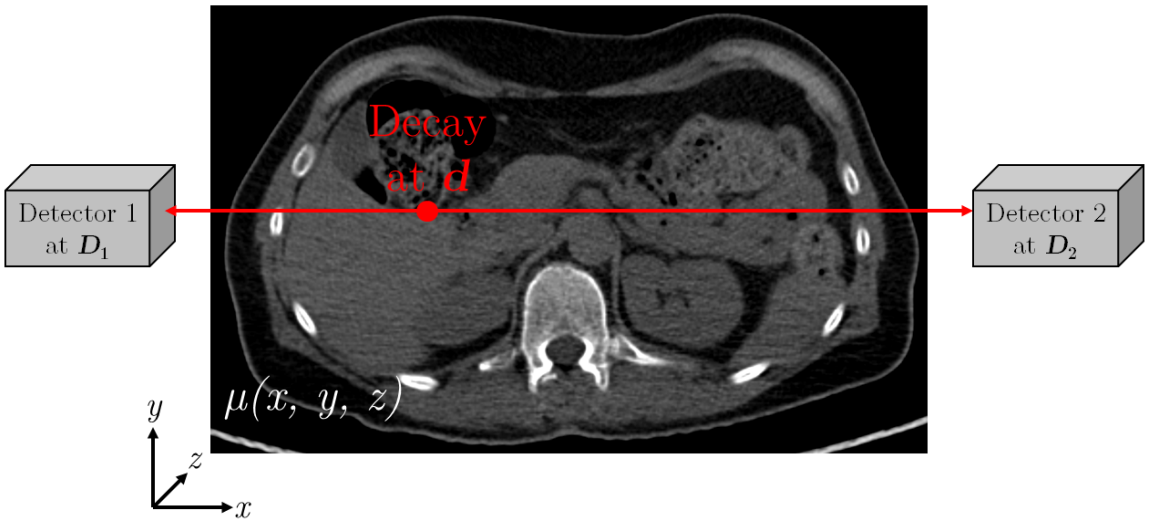
The simple reconstruction techniques in earlier methods neglected the cross talk between the individual lines of response. This assumption fails if photon scatter occurs. Scatter correction is another important requirement for (quantitative) ECT imaging. Scattered radiation is produced when gamma quanta emitted from decaying nuclei interact with surrounding atoms. Compton scattering is the prevalent scatter process in the energy range of clinically-utilized radio-tracers.
The energy E_{S} of the scattered photon depends only on the scattering angle ϕ and is given by Equation 2.3, where E_0 is the energy of the photon before the scattering and $m_e c^2$ the invariant mass of the electron. The energy transfer thus does not depend on the density or atomic number of the absorbing material. However, the total probability that a photon is scattered by this effect depends heavily on the properties of the absorbing material, most importantly the electron density.
As seen in Equation 2.3, the gamma quanta loose energy and change their momentum and direction in the scatter process. Because of the intrinsic energy resolution of the detector, the system cannot discriminate between unscattered quanta and quanta that have lost a small amount of energy in the scatter process. As a consequence some scatter is allowed in the image formation.

$$E_{\mathrm{S}} = \frac{E_0}{1 + \frac{E_o}{m_e c^2}(1 - \cos(\phi))} \tag{2.3}$$

In simple FBP, it is assumed that the decay took place exactly perpendicular to the detection plane and the detection location (for SPECT, under assumption of parallel-hole collimation). For scattered quanta, because of the change in direction



(a) SPECT



(b) PET

Figure 2.5: Simplified illustration for a decay event detected with SPECT (a) and PET (b). For SPECT, the signal of the decay at position $\boldsymbol{d}$ is decreased by attenuation effects. The amount of attenuation depends on the (spatially varying) linear attenuation coefficients $\mu(x, y, z)$ and on the distance between the detector and the decay, namely $|\boldsymbol{D} - \boldsymbol{d}|$. For PET, the amount of attenuation depends on $\mu(x, y, z)$ and on the distance between the two detector blocks $|\boldsymbol{D_1}\text{-}\boldsymbol{D_2}|$, but not on the exact position of the decay on the line of response.

in the scatter process, not only the distance of the radionuclei along the LOR is unknown, but also the correct positions of the LOR to the radionuclei. However, not all information about the originating nuclei is lost. Scattered radiation is therefore often understood as anisotropic noise that reduces the image quality of an ECT image.
There exist a variety of methods to correct for scattered radiation. Some of them rely on "passive" methods: For example, the camera's energy window could be narrowed or the lower discriminator cut-off of the window could be increased in order to avoid accepting scattered photons. Koral et al. [Kora 86] reported improved quantitative accuracy with this technique. A significant drawback of this method is that unscattered photons are also rejected due to the limited energy resolution of the gamma camera. Even with a relatively small energy window of ±5 keV for Tc-99m (140 keV), on basis of Equation 2.3, photons with scatter angles of up to 30 degrees are still accepted.
More common approaches utilize dual- [Jasz 84], triple- [Ogaw 91] or even multi-energy [Kora 88] windows. The additional scatter energy windows are placed below or above the photopeak energy window; the scatter images are acquired simultaneously with the photopeak image. For each pixel of the projection image, the amount of scattered radiation in the photo peak window image is estimated from the scatter window images. Subsequently, this amount can be subtracted from the projections or incorporated in the iterative reconstruction. Besides the simple multi-energy window approach, a multitude of other approaches exist for estimating and correcting for scatter radiation (for an overview see [Zaid 06a]). Despite the diversity in the methods employed, many reports stress the importance of an accurate scatter correction for quantitative ECT. Shcherbinin et al. [Shch 08] and Vandervoort et al. [Vand 07] incorporate a method based on the work of Wells et al. [Well 98] in their iterative reconstruction forward projection step that utilizes the Klein-Nishina formula and report an improved quantitative accuracy in phantoms measurements. Monte-Carlo methods seem to promise even more accurate results [Floy 84, Ljun 90, Frey 90]. However, its application in clinical practice is still limited due to the high computational costs and the patient specificity.
Although many scatter correction techniques are available, only very few are applicable in a clinical environment. Despite their simplicity, dual- or multi-energy window approaches can be recommended due to their ability for the correction for scattered radiation and the ease of application. For PET, due to the limited energy resolution of its detectors, model based scatter estimation techniques are more common. These methods utilize the emission and transmission images and mathematical models to estimate the extent of scattered radiation, individually for every patient [Olli 96]. The results from the scientific literature (see Section 2.1.7) for in-vivo quantification support this recommendation. The more sophisticated techniques still have to prove their superiority and applicability in daily clinical usage.

2.1.5 Partial Volume Correction

Partial volume effects are caused by the limited spatial resolution of emission tomography devices. Region of interest (ROI) in structures with heterogeneous activity distribution below approx. two-times the FWHM of the spatial resolution are de-

graded. The activity is either under- or over-estimated, depending on the combination of "spill-in" and "spill-out" effects. Spill-in refers to the effect that activity from outside the ROI or structure due to the limited spatial resolution is integrated into the ROI: The activity inside the ROI is increased. Spill-out is understood as the activity of the ROI/structure which is distributed over the borders (again due to the limited spatial resolution) and therefore "lost" for the quantification of that structure: the activity inside the ROI is decreased.
The degree of the partial volume effect depends on the (spatially varying) system resolution of the imaging system, the patient (e. g. motion), and the true distribution of radioactivity in the image.
In SPECT systems, the spatial resolution (which, in the following, is understood as the FWHM of a point source) is limited mainly by the collimator performance. Unlike PET, SPECT utilizes absorptive collimation to identify the direction of the photon LOR. Only a small fraction of the gamma quanta that hit the collimator surface pass through it. This leads to a heavily limited detection efficiency when compared to PET systems (see e. g. Cherry et al. [Cher 03, p. 340]). Since there is a trade-off between spatial resolution and detection efficiency, SPECT collimators are typically designed with the maximum allowable resolution in order to partially compensate the limited detection efficiency.
Besides the collimator design and geometry, the achievable spatial resolution is also influenced by the detector intrinsic resolution (the spatial resolution of the detector itself, without a collimator). Today, most SPECT detectors are made of a single crystal plate of NaI that illuminates an array of photomultipliers. The intrinsic resolution of the detector is influenced by the photo-peak energy of the imaged radionuclide and the crystal thickness. Higher gamma quantum energy leads to better intrinsic resolution, due to a higher scintillation light output. A larger crystal thickness increases intrinsic resolution, due to the broader spread of the scintillation light before it exits the crystal.
Clinical SPECT detectors typically possess an intrinsic spatial resolution in the range of 3-5 mm for Tc-99m. However, the image resolution for the SPECT system depends highly on the collimator design and the source-to-collimator distance. For parallel-hole collimation of Tc-99m and typical source-to-collimator distances it commonly ranges from 7 to 15 mm FWHM, which is considerably higher than that seen in PET (2-5 mm FWHM). By applying other collimator geometries, e. g. (multi-)pinhole, even lower spatial resolution than in PET can be achieved [Schr 03]. Branderhorst et al. [Bran 10] report sub-millimeter (FWHM) resolution for their small animal SPECT camera using Tc-99m. However those collimator geometries are more frequently for small animal studies than in clinical practice. A more detailed description on collimator geometries can be found in [Cher 03].
For clinical PET, the achievable spatial resolution mainly depends on the size of the individual crystal elements and the detector block geometry (e. g. size of the gantry bore). Ultimately, the PET resolution is limited by the positron's free path length in between the initial decay and the annihilation (which is what is detected in PET). Another factor is the acollinearity of the two coincident gamma quanta: The remainder of the kinetic energy of the positron at the point of annihilation causes the two emitted gamma quanta to not be exactly emitted in a perpendicular direction (180°)

but in smaller angles. These effects sum up to physical limits of ≈ 2 mm for a clinical PET system.
In the following, the ratio of apparent activity concentration to true activity concentration is called recovery coefficient. The approaches for partial volume correction can be divided into two groups. Approaches that need additional information (e. g. CT, MR) on the structures that are imaged and methods that work solely on the emission images.
A simple to implement and thus common post-reconstruction approach in the latter group was founded on experiments with physical phantoms and simulation studies or theoretical derivations. Based on the approaches of Hoffman et al. [Hoff 79] and Kessler et al. [Kess 84] recovery coefficients for simple geometries (e. g. spheres, discs, cylinders) are estimated and used for calculating the true amount of radioactivity in those structures. Several groups report an improved quantification accuracy of this approach [Blan 96, Gewo 00, Chen 98]. A known limitation of this method is the sole applicability on simple geometries; in general the distribution of the radioactivity might not follow this assumption. Seo et al. [Seo 09] report that with their implementation of a deconvolution based partial volume correction an accuracy of 10% for In-111 in lesions with a volume of down to 8 ml can be achieved.
Every approach that finally leads to an improved spatial resolution of the imaging system helps to avoid partial volume effects. Thus, e. g. the methods described in Section 2.1.2 which incorporate information about the systems (collimator-detector) response function could, to some extent, be understood as partial volume correction technique. For example, Hutton et al. [Hutt 98] implement their version of detector-response modeling into the Maximum Likelihood Expectation Maximization (MLEM) and Ordered Subset Expectation Maximization (OSEM) reconstruction for a SPECT system and report improved accuracy for the simulated MCAT phantom. The other group of methods incorporate structural information in the form of segmented MR or CT images in the partial volume correction step. The segmentation can be done automatically or manually by the definition of ROIs. Pretorius et al. [Pret 09] apply a method based on the work of Da Silva et al. [Da S 01] and Tang et al. [Tang 96] that incorporates multiple two-class segmentations (regions with activity and regions without activity) of co-registered myocardial CT images on the MCAT phantom. They report improved visual characteristics as well as a higher quantitative accuracy of the corrected SPECT data.
An evolution of the aforementioned approach is the geometric transfer matrix (GTM) method (Equation 2.4 first applied by Rousset et al. [Rous 93, Rous 98] on brain PET studies: it allows an almost arbitrary number (n) of regions with homogeneous activity distribution. The observed activity t_i of a certain tissue class i is assumed to be a linear combination of the true activities T_j of all other tissues classes $j = 1 \ldots n$. The spatial definition of the regions is commonly done on segmented CT or MRI images. The ω_{ij} represent the regional transfer coefficients: The diagonal terms represent the spill-out of every region; the off-diagonal terms define the spill-in of other regions. The ω_{ij} can be computed from the defined regions and the point spread function of the imaging system. In the end, one get a full-rank transfer matrix: the true activities

T_j can be computed by solving the linear equation defined in Equation 2.4.

$$\begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix} = \begin{pmatrix} \omega_{11} & \omega_{21} & \cdots & \omega_{n1} \\ \omega_{12} & \ddots & & \omega_{n2} \\ \vdots & & \ddots & \vdots \\ \omega_{1n} & \omega_{2n} & \cdots & \omega_{nn} \end{pmatrix} \times \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{pmatrix} \tag{2.4}$$

Du et al. [Du 05] compare variants of the GTM approach to uncorrected images in a physical brain phantom filled with Tc-99m and find a strong reduction of the bias induced by the partial volume effect. Soret et al. [Sore 03] studied the effect of the GTM approach on I-123 again using a physical brain phantom and segmented CT acquisitions. For uncorrected images and small structures (e. g. putamen) they find an underestimation of up to 50% when compared to the true activity concentration. The application of the partial volume correction leads to an over-estimation of about 10%, which could be considered to be a significant improvement.

2.1.6 Calibration

A calibration of the SPECT imaging system volume sensitivity S_{Vol} (Equation 2.5) is the final requirement for absolute quantitative imaging. This is typically obtained by a correlation of the results to a calibrated well counter. The principle is briefly outlined in this section, details can be found for example in the NEMA protocols [Perf 07]. In order to avoid partial volume effects, a large cylindrical phantom with known activity concentration c_{Vol} is scanned. Corrections for attenuated and scattered photons are applied in the reconstruction. A large VOI with volume V_{VOI} is placed on the reconstructed image. T_0 is the start time, T_{acq} the duration of the acquisition. $T_{\frac{1}{2}}$ is the half-life of the used radionuclide and T_{cal} the time of the activity calibration. R represents the counting rate measured in the VOI. Finally, according to Equation 2.5, a calibration factor from the detected counts per second to Becquerel is derived. In the equation, the measured count rate R is normalized by the volume V_{VOI} and the activity concentration c_{Vol}. The other factors of the equation account for the differences between aforementioned time points.

$$S_{\mathrm{Vol}} = \frac{R}{V_{\mathrm{VOI}} \cdot c_{\mathrm{Vol}}} \cdot \exp\left(\frac{T_0 - T_{\mathrm{cal}}}{T_{\frac{1}{2}}} \cdot \ln 2\right) \cdot \left(\frac{T_{\mathrm{acq}}}{T_{\frac{1}{2}}} \cdot \ln 2\right) \cdot \left(1 - \exp\left(-\frac{T_{\mathrm{acq}}}{T_{\frac{1}{2}}} \cdot \ln 2\right)\right)^{-1} \tag{2.5}$$

The calibration factor is specific for every radio nuclide as to different intrinsic detector sensitivities and collimators used. Due to non-linearities of the detector at different count rates and dead time effects at high activities, count rate-dependent calibration factors for the same radionuclide should be applied as well. Most notably those effects are more pronounced for high energy radionuclides. For example, Dewaraja et al. [Dewa 08] report on the dead time and pulse pile-up effects for the SPECT quantification of therapeutic activities 2-6 GBq I-131. For PET imaging,

this step is often automated and included in the daily quality control procedure. A homogenous Ge-68/Ga-68 phantom with known activity is placed in the PET field of view and data is acquired over a certain period of time. Subsequently, a calibration factor is calculated, stored, and automatically applied to the reconstructed PET data. However, to achieve more accurate values, a cross calibration between the well counter and the system, analogously to the above listed SPECT procedure is recommended [Gewo 02].

2.1.7 Validation Studies

Many reports show that accurate absolute quantification is possible. The majority of the methods are evaluated in phantom or simulation studies. Recently, Shcherbinin et al. [Shch 08] report errors between 3% and 5% for a study on a torso phantom for the isotopes Tc-99m, I-123, I-131, and In-111. Du et al. [Du 06] achieve a 2% error for I-123 in a brain phantom. Vandervoort et al. [Vand 07] performed a simulation study on the MCAT cardiac-torso phantom and on a Tc-99m filled torso phantom and achieve 8% error in the simulation and about 4% for the phantom study. Da Silva et al. [Da S 99] reach an error of 8% in an anthropomorphic phantom with cardiac insert for Tc-99m. For PET, Jentzen et al. [Jent 08] compare I-124 and F-18 recovery coefficients in phantoms and find that with application of a recovery correction coefficient specific for the size of the lesion, accurate quantification is possible, at least for structures larger than $\approx 1.5 \times$ FWHM. Apostolova et al. [Apos 10] reach 10% quantitative accuracy with F-18 in a phantom study for solitary pulmonary nodules, applying their version of motion and partial volume correction. Willowson et al. [Will 12] find PET to be accurate within 8% for larger structures and threshold based volumes of interest in a body phantom for Y-90.
Unfortunately only very few reports about in-vivo absolute ECT quantification exist: Our group [Zein 10] evaluated the activity of Tc-99m-DPD (a tracer for bone imaging) in the bladder of 16 patients. The reference activity is determined by the measurement of the activity concentration in the urine right after SPECT imaging. We find an average deviation of 6.8% between the activity concentration obtained in SPECT and a well counter measurement, with the corrections for physical effects provided by the camera manufacturer. Another in-vivo study by Da Silva et al. [Da S 01] evaluates the accuracy of absolute quantification of Tc-99m-Sestamibi (tracer for cardiac imaging) in the myocardium of eight pigs. The comparison of SPECT quantification (with their version of partial volume and attenuation correction) and the ex-vivo activity concentration of the excised myocardia revealed a deviation of 10%. Quantification in humans was shown by Willowson et al. [Will 08]. They studied Tc-99m-macro aggregated albumin in lung perfusion for 12 patients and found an average error of 2.6% (ranging from -7% to $+4\%$) with corrections for scatter, attenuation and partial volume. The macro aggregated albumin is trapped almost entirely in the capillaries of the lung, thus the total activity in the lung can be compared to the injected activity. Almeida et al. [Alme 99] evaluated the quantitative accuracy of the uptake of I-123 labelled epidepride in a certain brain region (striatum) of monkeys (papio anubis baboons). They validated their results using PET acquisitions of C-11 epidepride of the same subjects and found a deviation for defined regions of interest

of less than 10% between both modalities, using corrections for attenuation, scatter, and partial volume.

2.1.8 Conclusion

A survey of the current literature as well as our own work show that ECT can be quantitative with errors below 10% even in a clinical environment (Figure 2.6). To achieve this, a careful setup and calibration of a state of the art SPECT/CT or PET/CT system is required. Equally important is an iterative reconstruction software which is able to model the imaging physics and to correct for degrading factors. Due to the higher spatial resolution of the clinically PET systems and the reduced partial volume effects, PET can reach a better quantitative accuracy for smaller structures than SPECT.
However, there are also some limitations on the applicability in clinical routine. Many methods rely on complicated manual procedures. At this point more work is needed to enable intelligent automation. In order to further reduce the quantification errors, the incorporation of better imaging models within the reconstruction is mandatory. The improved models should allow a better compensation of physical effects and patient induced artifacts (e. g. motion).
In general, more sophisticated phantom and animal experiments are needed. Yet the goal of routine application of absolute quantification in clinical ECT imaging is in reachable range.

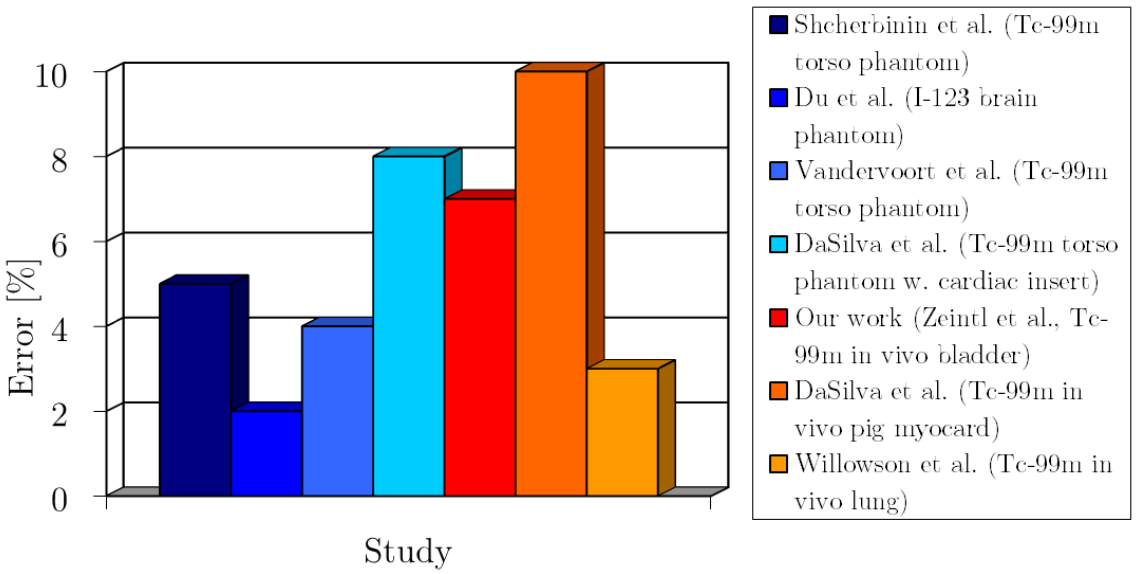


Figure 2.6: Overview of the accuracy of ECT quantification from studies reported in the literature (blue bars phantom studies, red/orange bars in-vivo studies).

2.2 Magnetic Resonance Imaging

2.2.1 History and Basics

MRI as an imaging modality was originally developed by Lauterbach et al. and Mansfield et al. in 1973. It is an evolution of Nuclear Magnetic Resonance Spectroscopy (NMR), which did not offer any spatially resolved measurements at that time. In the following section, we explain the basics that are necessary for our work. For further explanations please refer to e. g. [Reis 08].
In medicine, most MRI applications are based on physical effects of hydrogen atoms, more precisely, of the H-1 nuclei. In the semi-classical description, the H-1 nucleus can be understood as a rotating, positively charged sphere. The rotating positive charge results in a magnetic momentum which points to the direction of the angular momentum of the atom. In matter, the magnetic momentum of the single atoms points into random directions. Effectively, no aggregate magnetization of the matter can be measured. In the presence of an external magnetic field, a force results that tends to align the magnetic moment along this external field. Due to the conservation of the magnetic momentum, this alignment is not allowed, the magnetic moment starts to precess around the direction (parallel) of the external field. The frequency of this precession is called Lamor frequency. An anti-parallel precession is also possible. This state corresponds to a higher energetic level and has a lower probability for the atom to exist in anti-parallel than to exist in parallel precession. Within the context of scientific literature, the term magnetic moment is also known as spin. For particles with the charge q and the mass m, the spin $\boldsymbol{s}$ is in relation to the magnetic moment $\boldsymbol{m}$ according to the following Equation 2.6. The proportional factor g is known as Lande factor.

$$\boldsymbol{m} = g\frac{q}{2m}\boldsymbol{s} \tag{2.6}$$

2.2.2 Signal and Contrast Mechanism

As opposed to X-ray CT. the contrast mechanism of MRI is not based on electron densities but on multiple parameters. As MRI uses the signal from hydrogen atoms/ protons for image formation in most medical applications, the density of the protons contained in the imaged volume influences the signal intensity, i. e. , the more spins in the volume that are flipped, the more radio-frequency (RF) energy is deposited in the volume and hence the higher the emitted signal is when those spins flip back to their lower energy level.
In matter, the protons are packed within a lattice with the surrounding atoms and their rotating spins can interact with each other. Consequently, the energy from the higher energy state can not only be emitted as a photon (at RF) but can also be dissipated in the form of proton movement and vibrations. The average lifetime of the nuclei in the higher energetic state is called spin-lattice relaxation time or T_1-time.
Another parameter is the T_2-time or spin-spin relaxation time. As previously described, when flipped, the spins precess around the direction of the external magnetic field. An RF signal can be measured as long as there are enough spins in a voxel that precess synchronously (precess in phase). The local magnetic field for the spins in a voxel varies due to several reasons. These are intrinsic, e. g. a spatially varying

Tissue	T_2[1] [ms]	T_1 (1.5 T)[1] [ms]	T_1 (3 T)[1] [ms]	Attenuation[2] [HU]	FET-Uptake[3] [SUV]
White Matter	69	884	1 084	30	1.1
Gray Matter	99	1 124	1 820	33	0.8
Cerebrospinal Fluid	2 100[4]	4 210[3]	-	15	-
Blood	275	1 441	1 932	40	-

[1] all [Stan 05] except CSF [2] all [Cala 81] [3] all [Webe 00] [4] [Melh 97]

Table 2.1: Typical values for MRI, CT, and PET contrast parameters.

magnetic moment of molecules and extrinsic, e. g. a non-homogeneous external magnetic field. The result of this inhomogeneity is that each spin precesses at a different frequency and the spins which were in sync after the RF pulse loose their synchronicity. The emitted signals of the spin-systems interfere with each other in a destructive manner and the measured signal diminishes. The time that describes how fast this de-synchronization takes place, is called T_2-time. To be exact, two different T_2-times exist: the effective (measured) time is called T_2^*-time, it includes the intrinsic and extrinsic inhomogeneities. With special MRI acquisition techniques, one can correct for the extrinsic inhomogeneities and subsequently measure the T_2-time, which characterizes only the intrinsic field inhomogeneities.
Proton density, T_1- and T_2-time are patient and tissue specific. The following relation is true for most tissue types and applications.

$$T_2^* < T_2 < T_1 \tag{2.7}$$

Table 2.1 lists typical values of the relaxation times for tissue types encountered in MR imaging of the brain. For comparison, also CT-Hounsfield values and FET-PET standardized uptake values (SUV) are shown.

2.2.3 Spatial Encoding and Image Reconstruction

In order to represent a spatial distribution of the signal which is emitted from the precessing spins, additional steps are necessary. As opposed to CT and emission tomography where the projection concept prevails, which means obtaining 2-D images under different viewing angles, MRI uses a different reconstruction scheme. The projection concept was only used in the early inception of MRI.
All the sequences in our study apply slice selective excitation, only the spins in a slice of certain thickness are excited by the RF pulse and subsequently only these spins emit a signal and contribute to the image. For a three-dimensional imaging, multiple slices need to be acquired separately and have to be stacked in order to form a volume. Slice selective excitation is done by applying a gradient field perpendicular to the desired slice, which defines the gradient encoding direction. The gradient changes the strength of the magnetic field in this direction and leads to a position dependent Lamor frequency. By emitting the exciting RF pulse at a specific frequency, only spins with the appropriate Lamor frequency are flipped.
The two-dimensional encoding in the slice itself is achieved by other methods, namely

phase- and frequency encoding. A short gradient pulse in a direction perpendicular to the gradient encoding direction is applied at the time between the RF-excitation and recording of the echo, for phase encoding. The gradient pulse results in a spatial dependent Lamor frequency. Again, the spins precess at different speeds, depending on their spatial location along the phase encoding axis. After the short gradient pulse, all spins in the slice precess with the same frequency again. The intermittent gradient causes a phase shift of the spins and the amount of the phase shift is proportional on the spins position along the phase encoding axis. In order to resolve N voxels along a phase encoding axis, the process needs to be repeated N times. Depending on the sequence, this is done by repeating the same gradient (e. g. for echo planar imaging) or by increasing the strength of the gradient for every repetition (regular spin-echo and gradient-echo sequences).
In frequency encoding another gradient is switched during the acquisition of the echo. The gradient is along an axis which is perpendicular to both, phase- and slice encoding direction. This causes variations in the Lamor frequencies along the frequency encoding axis. The echoes are emitted and acquired at frequencies that depend on the position of the spins along this axis.
The previously described spatial encoding fills a two-dimensional k-space matrix line-by-line, with one separate matrix for every slice. It is a 2-D Fourier method wherein the spatial distribution of the image intensities subsequently is reconstructed by an inverse Fourier transformation of the k-space matrix for every slice. The reconstructed slices are stacked in order to form a final 3-D image.

2.2.4 Relevant Sequences

Due to the retrospective nature of our work, the used sequences show a certain heterogeneity. In general, we use three contrast mechanisms in our study, the contrast-enhanced T_1-weighted images, the fluid attenuated inversion recovery T_2-weighted images, and images representing the apparent diffusion coefficients, measured by diffusion-weighted MRI.

Enhanced T_1-MRI

For T_1-weighted imaging, the sequences use either spin-echo (SE) or gradient-echo (GRE) techniques. In the following is a schematic outline of SE (Figure 2.7) and GRE (Figure 2.8) sequences. The signal intensities I_{SE}, I_{GRE} of a certain voxel for spin-echo and gradient-echo sequences depend on time-to-echo T_{E}, time-to-repeat T_{R} and on T_1 and T_2 time of the tissue in the respective voxel according to the Equations 2.8 and 2.9.

$$I_{\mathrm{SE}} \propto \rho_{\mathrm{P}} \cdot \left(1 - e^{-T_{\mathrm{R}}/T_1}\right) \cdot e^{-T_{\mathrm{E}}/T_2} \tag{2.8}$$

$$I_{\mathrm{GRE}} \propto \rho_{\mathrm{P}} \cdot \frac{\left(1 - e^{-T_{\mathrm{R}}/T_1}\right) \sin\alpha}{1 - e^{-T_{\mathrm{R}}/T_1} \cos\alpha} \cdot e^{-T_{\mathrm{E}}/T_2} \tag{2.9}$$

As endovascular contrast agent, e. g. Gadobutrol can be used. It contains Gadolinium (Gd) which would be toxic in its free form and therefore is bound in highly stable complexes. Gd has seven unpaired electrons and is thus ferromagnetic. It changes

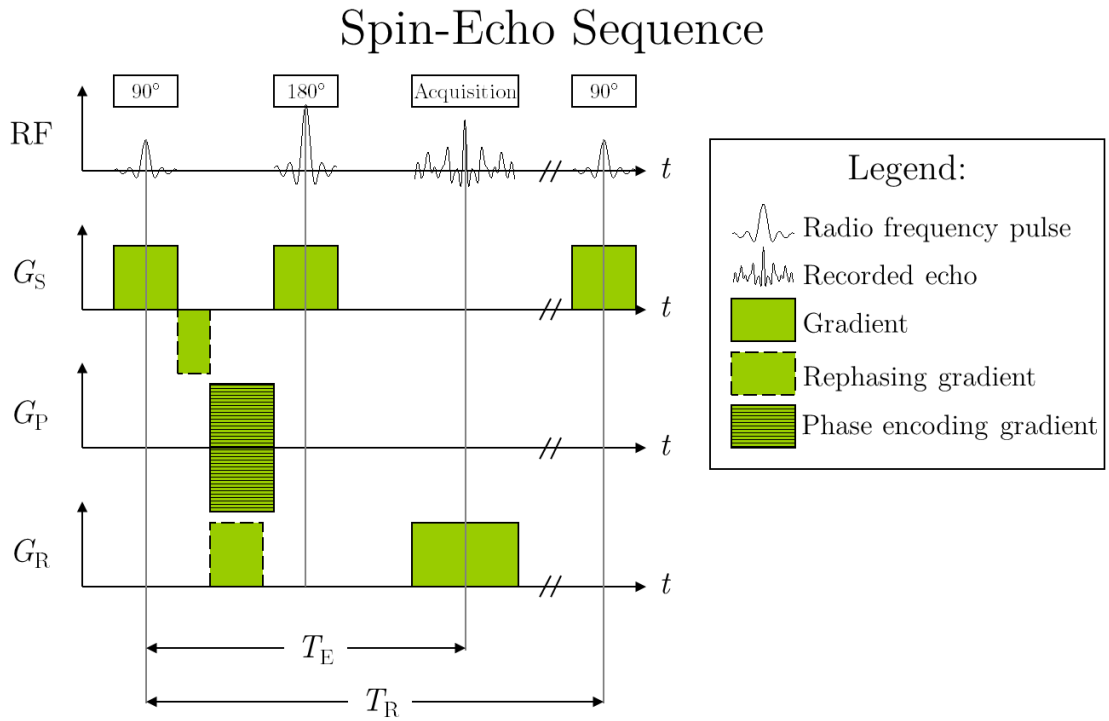


Figure 2.7: Schematic outline of a spin-echo MRI sequence: The echo is formed by a 180° RF pulse, following the initial excitation by a 90° pulse. G_S is the slice selective gradient, G_P the phase encoding gradient and G_R the read out gradient that is switched during the acquisition of the RF signal. T_E and T_R represent the time-to-echo and time-to-repeat.

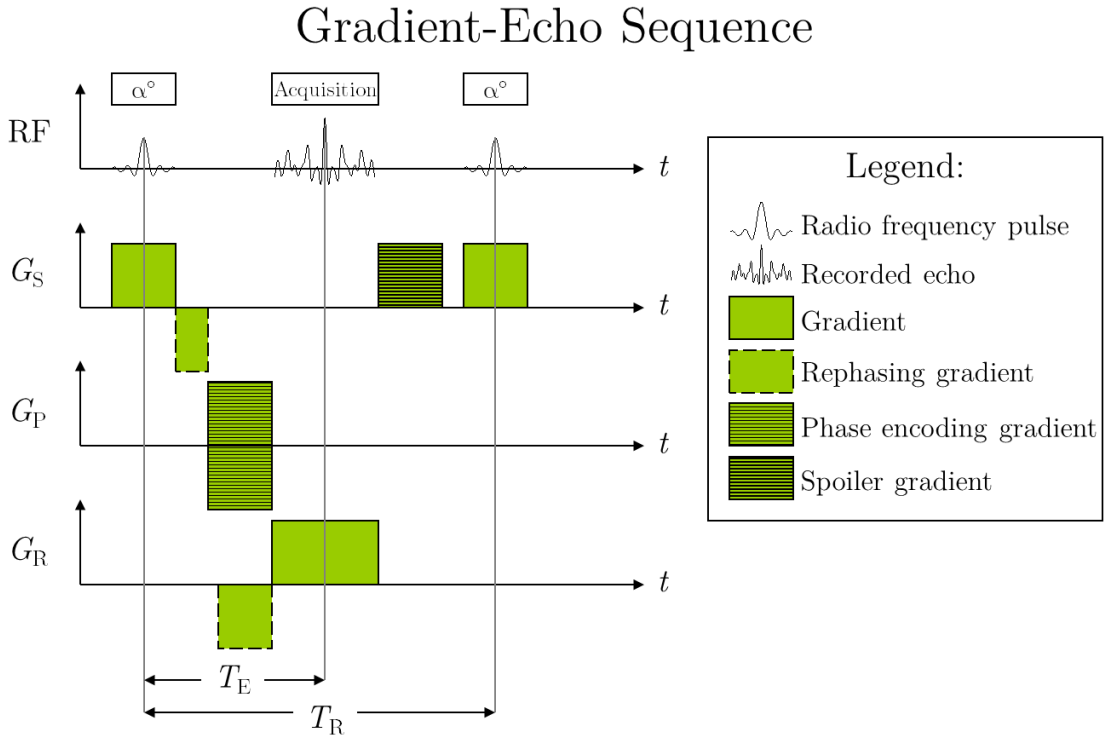


Figure 2.8: Schematic outline of a gradient-echo MRI sequence. The echo is formed by a rephasing signal of the read out gradient G_R, following the initial excitation by a RF pulse that flips the spins by a flexible angle of α degrees. The remaining magnetization is destroyed by a spoiler gradient. G_S is the slice selective gradient and G_P the phase encoding gradient. T_E and T_R represent the time-to-echo and time-to-repeat.

the relaxation times of adjacent protons. The surrounding tissue shows a shortened T_1-time and has an increased signal intensity in T_1-weighted sequences. The injection of Gadobutrol is widely used for identifying pathologies of the brain. For example, a by a tumor disrupted blood-brain-barrier results in an extravasation of the contrast agent and consequently leads to an increased signal in that area.

Fluid Attenuated T_2-MRI

In fluid attenuated inversion recovery (FLAIR) MRI, the signal of cerebrospinal fluid (CSF) is suppressed using an inversion recovery sequence. For an unsuppressed sequence, the CSF would appear as hyper-intense (brighter than surrounding normal tissue) and potentially mask tumors which often appear hyper-intense as well. A diagram of an inversion recovery sequence is given in Figure 2.9. The signal intensity

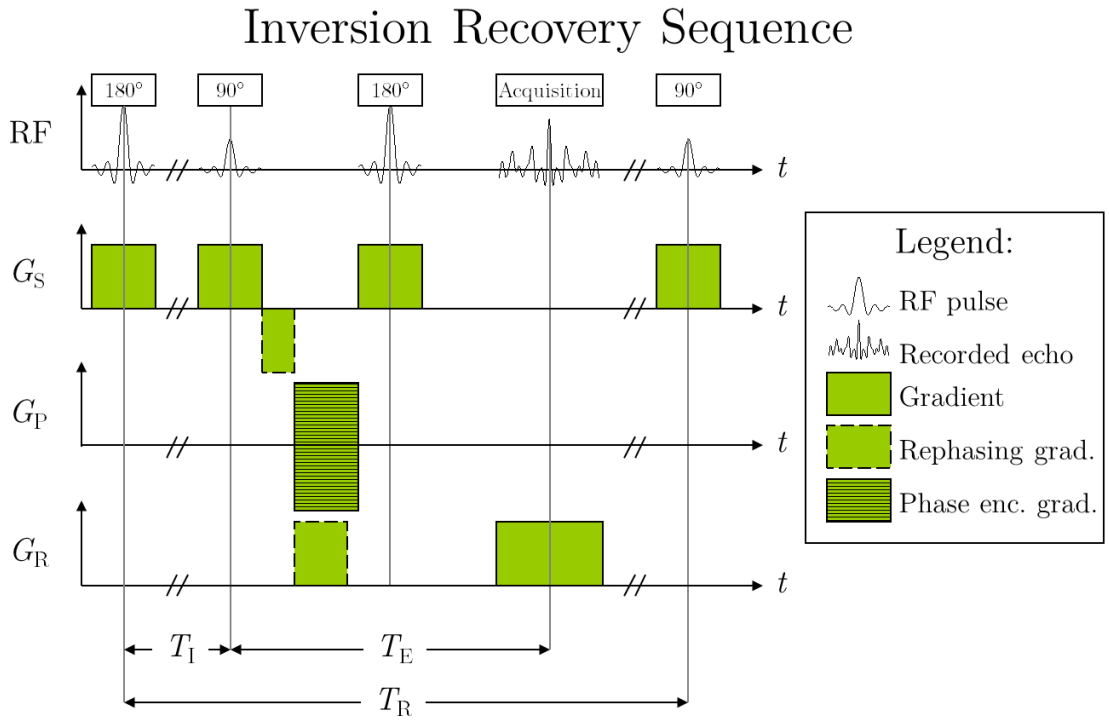


Figure 2.9: Schematic diagram of an inversion recovery sequence. The spins are inverted by the 180° pulse at the beginning. After the inversion time T_I, an ordinary spin-echo sequence is carried out. G_S is the slice selective gradient, G_P the phase encoding gradient and G_R the read out gradient that is switched during the acquisition of the RF signal. T_E and T_R represent the time-to-echo and time-to-repeat.

of a voxel for the inversion recovery sequence depends on inversion time T_I, time-to-repeat T_R, time-to-echo T_E and T_1, T_2 time of the tissue in the respective voxel according to Equation 2.10.

$$I_{IR} \propto \rho_P \cdot \left(1 - 2e^{-T_I/T_1} + e^{-T_R/T_1}\right) \cdot e^{-T_E/T_2} \tag{2.10}$$

Diffusion MRI

In our study, all diffusion MRI data were acquired using an echo-planar imaging technique which is characterized in the sequence diagram 2.10.

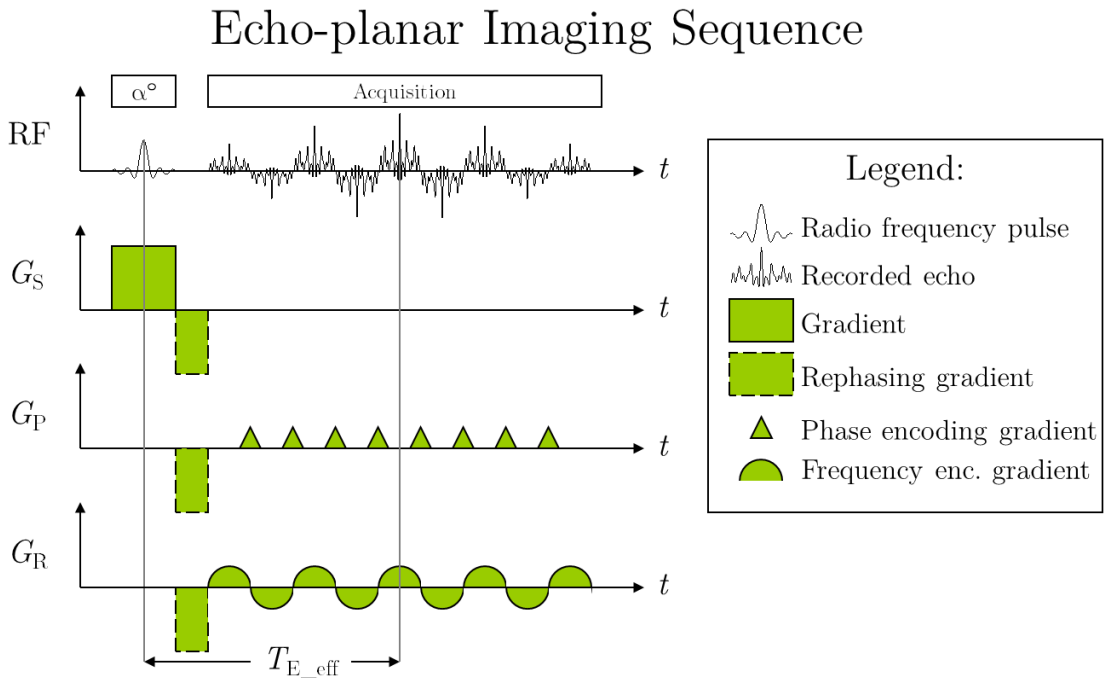


Figure 2.10: Schematic diagram of an echo-planar imaging MR sequence. Only one excitation RF pulse that flips the spins by a flexible angle of α degrees. An echo train is generated using alternating signals of the frequency-encoding/read-out gradient G_R. The phase encoding is done by short gradient pulses G_P during the ramp up of G_R. G_S is the slice selective gradient, T_{E_eff} and T_R represent the effective time-to-echo and time-to-repeat.

2.3 Transmission Tomography

A German version of excerpts of the current section was published in *Der Nuklearmediziner* [Ritt 11a]. The original article was recently included in the German guidelines for skeletal scintigraphy [DGN 13].
X-ray computed tomography (CT) is a well established method for the diagnoses of a multitude of medical problems. Most clinical CTs work according to the rotate-rotate principle: The main components consist of an X-ray tube and a detector that rotates around the object that should be imaged.
The X-ray tube is an evacuated unit consisting of the cathode and anode. By the acceleration of electrons which are emitted by the cathode and slowed down at the impact on the anode, X-rays are produced. A fan- or cone-shaped X-ray beam, which is formed by mechanical and electrical collimation and filtering, is sent through the imaged object and detected by an array of detector elements on the opposing side of the gantry. In the following, transmission images of the object under different viewing angles are acquired. Subsequent electronics and computers allow the calculation of the spatial attenuation map for X-rays of the object (reconstruction). Multiple rotations of the X-ray tube are carried out in order to achieve a larger field of view (FOV) in the axial direction. In between single rotations, the patient table is moved stepwise (sequential technique, "step and shoot") or continuously during the rotations ("helical CT").
For image reconstruction, computed tomography mostly applies filtered back projection, which is the inversion of the Radon transform, with adaptions for the fan- or cone-beam geometry. Since recently, iterative techniques have shown promise in providing increased image quality at lower radiation doses [Will 13]. In general, images can be reconstructed in different sharpness and with varying slice thickness and resolution.
For a speed-up in acquisition time, most modern CTs have multiple detector lines. Consequently, during one rotation, a larger axial field of view can be covered. Short scan times are especially important for acquisitions of organs that suffer from movement artifacts, such as heart and thorax/abdomen and are in general convenient for the patient.
In the following, some basic terms of CT are explained:

- **Focus:** The size of the impacting electron beam on the anode, projected onto the detector. Ideally, a point-like focus should be chosen. Certain technical limitations, among them mainly thermal constraints, restrict the minimum size. Since a small focus would exceed the thermal capacity of the anode at long CT examinations, it is necessary to have a larger focus in these cases.

- **kV:** The kV value reflects the acceleration voltage that is used on the X-ray tube. In general, kV values for abdominal and cranial CT range between 110 and 140 kV. X-rays with higher energy (high kV value) penetrate the patient easier and lead to an increased signal-to-noise ratio at the same tube current.

- **Collimation:** Modern CT devices operate with multiple detector lines. These lines are placed next to each other and have differing slice widths, at least for

some devices. It is common that the detector lines are pooled electronically or mechanically. A collimation of 32×0.6 mm means, that 32 detector lines with a width of 0.6 mm each are read out. The same detector could be collimated with 16×1.2 mm as well, consequently every two detector lines are pooled. The total collimation is for both examples $32 \cdot 0.6$, resp. $16 \cdot 1.2$ mm = 19.2 mm. A smaller collimation, together with adapted reconstruction and tube current, increases the spatial resolution of the CT acquisition.

- **Pitch:** The pitch number is defined as the ratio of table feed per rotation to the total collimation, measured at the center of rotation. A scan with a table feed of 10 mm per rotation and a total collimation of 10 mm consequently has a pitch of 1. At 20 mm table feed and constant collimation of 10 mm increases the pitch to 2. In general, with decreasing pitch the redundancy in the tomographic projections is increased: The same axial position is scanned in multiple rotations. With the appropriate reconstruction technique, this redundancy can be used (e. g. by averaging) to increase the signal-to-noise ratio and with it the image quality. As drawback of a decreased pitch, the patient dose is increased and the scanned field-of-view per time is lowered.

- **mAs:** For single slice computed tomography, the mAs-value corresponds to the tube current multiplied by the rotation time (current-time product). The higher the mAs value is, the more X-ray photons are available per rotation. Consequently, the signal-to-noise ratio and the image quality is in general increased. The mAs value is used as an indicator for image quality. It should be noted that an increased mAs value is directly related to an increased patient dose. As mentioned before, the pitch influences the acquisition quality as well. For this reason, a pitch independent indicator for image quality is defined by calculating $\mathrm{mAs}_{\mathrm{eff.}} = \mathrm{mAs} \,/\, \mathrm{Pitch}$. The $\mathrm{mAs}_{\mathrm{eff.}}$ value follows the same rules as the mAs value: Higher $\mathrm{mAs}_{\mathrm{eff.}}$ value $\rightarrow$ higher SNR and improved image quality but also higher patient dose. Lower $\mathrm{mAs}_{\mathrm{eff.}}$ value $\rightarrow$ lower SNR, lower image quality, but also decreased patient dose. The individual requirements of the examination usually define the optimal $\mathrm{mAs}_{\mathrm{eff.}}$ value. Especially for attenuation correction and anatomical co-registration like in SPECT/CT and PET/CT, low values are sufficient.

- **$\mathrm{CTDI}_{\mathrm{Vol}}$:** Is an acronym for "Computed Tomography Dose Index". The addition "Vol" is an acronym for volume and helps to differentiate a multitude of other CTDI values. $\mathrm{CTDI}_{\mathrm{Vol}}$ is the effective (= pitch corrected), weighted (= dose-depth profile corrected) dose value in a slice of the thickness of the effective collimation.

- **Dose-Length-Product (DLP):** It is defined as the product of the $\mathrm{CTDI}_{\mathrm{Vol}}$ value and the length of the CT scan. It is proportional to the effective dose of the CT acquisition and for this reason one of the most important parameters for estimating the radiation burden of the examined patient. From literature or simulation studies, factors for obtaining the effective dose from the DLP can be gained [Cher 03]. For fusion and attenuation correction of PET and SPECT,

CT images with limited field of view and low mAs values are sufficient ("low-dose CT"). The radiation of these low-dose scans is in the range of some mSv effective dose [Gilm 06].

- **Reconstruction kernel:** In FBP reconstruction, a mathematical operation for inverting the Radon transform (folding with a kernel) is applied. This approach is able to reduce the inherent "smearing" of the back projection. In clinical practice, a multitude of kernels is established, some are more blurring ("soft"), some are neutral, and some are sharpening ("bone"). Depending on the goal of the examination and on the type of the examined structures (bones, soft tissues), the appropriate kernel for the image reconstruction is chosen.

- **Hounsfield Unit (HU):** In CT, slice images of the spatial distribution of the linear attenuation coefficients for the used X-rays are attained. For generating device independent images which are not influenced by properties of the X-ray beam, such as photon energy spectrum, Hounsfield et al. chose the linear attenuation coefficients of water and air as base of a new scale. Every CT for medical applications is calibrated on these attenuation coefficients by acquiring special phantoms on the device. Air subsequently corresponds to a HU of -1000, water to a HU of 0.

- **Windowing:** The human perception is only able to differentiate up to $80-100$ gray level values. On the contrary, the dynamic range of CT system often offers > 4000 different gray level values. A so called windowing technique is applied for the reading of CT images and exploiting all the information which is represented by this multitude of gray levels. The window levels are typically defined by a center C and width W value in HU. The C value in HU represents a median-gray on the reading device (film/monitor). The width defines which HU value on the reading device is represented as black ($C - 0.5 * W$ HU) and white ($C + 0.5 * W$ HU).

2.4 Cerebral Gliomas

Gliomas are tumors stemming from glial cells and are the most common class of tumors in the brain. Glial cells are cells in the brain that support and protect the neurons and form the myelin, which plays an important role in the signal transmission of the brain. Tumors are understood as an uncontrolled and abnormal growth of somatic cells. The abnormal cells often, but not necessarily (e. g. leukemia), form aforementioned tumors, which are accumulations of the cancer cells. A tumor itself can be benign, pre-malignant or malignant, depending on its behavior in regard to other surrounding tissues. The behavior is defined as malignant if the tumor invades surrounding tissues or forms distant colonies (which are called metastases). Tumors that do not posses invasive and metastatic properties are called benign. In general the disease is called cancer if a tumor shows malignant behavior. Most theories assume, that tumors are monoclonal growths, which means that they descend from an alteration of a single normal cell, which consequently starts to duplicate in an uncontrolled manner.

Tumors can be grouped in four major groups, according to their origin: epithelial, mesenchymal, hematopoeitic, and neuroectodermal. The most common tumors originate from epithelial cells and are called carcinomas. The carcinomas itself can be divided in two subgroups, the scamous cell carcinomas, originating from cells that form protective layers (e. g. oesophagial, dermal cells) and adenocarcinomas, origination from cells that form secretory layers (stomach, lung). Non-epithelial tumors are called sarcomas (stemming from mesenchymal tissue), hematopoeitic cancer (stemming from blood cells) and neuroectodermal tumors (stemming from cells of the nervous system.
As all tumors related to our work are gliomas, a neuroepithelial tumor, we will focus on these in the following sections.

2.4.1 Causes for Gliomas

Epidemologic studies have shown that gliomas and cancer in general are hugely influenced by environmental factors. These influences can be divided into chemical (e. g. tobacco) and physical (e. g. ionizing radiation) agents. It is well known that even viruses can induce cancer. Quite often carcinogens (cancer causing agents) are also mutagens (change the DNA). However, there exist also non-genetic mechanisms that promote tumor genesis.

2.4.2 WHO Classification of Tumors of the Central Nervous System

As stated in the *WHO Classification of Tumors of the Central Nervous System* [Loui 07], based on their histopathological resemblance in bioptic samples, tumors can be classified into certain classes, groups, and subgroups. One of the first reliable and strong concepts for the grading of cerebral tumors was introduced in 1926 by P. Bailey and H. Cushing [Bail 26]. In the following years, more research was carried out on this topic, which finally resulted in the publication of the volume *Histological Typing of Tumors of the Central Nervous System*. Its first edition was published in 1979 by the World Health Organization (WHO), the most recent edition was published in 2007 [Loui 07]. Most of the groups and subgroups in the WHO scheme have an associated *WHO grade*. This grade reflects the prognosis for a patient suffering from a tumor of the respective group and therefore is an important indicator for the optimal treatment of the disease. The grades are WHO grade 1 (in short WHO 1), WHO 2, WHO 3 and WHO 4. A higher grade reflects a higher malignancy. E. g. a WHO 1 tumor can be completely curable with very little lifetime restrictions when treated adequately, whereas patients suffering from a WHO 4 tumor only have a median time of survival between one and two years when treated (three months untreated).

2.4.3 Entity of Gliomas

As already stated, gliomas stem from glila cells, which themselves are a neuroepithelial tissue. The WHO classification scheme lists nine groups in the class of neuroepithelial tumors. The patients that were included in our study cover three of these

groups: These are *astrocytic tumors*, *oligodendroglial tumors*, and *oligoastrocytic tumors* (a mixed-type tumor). In the following sections more details about the groups can be found.

Astrocytic Tumors

A large proportion of the glial tumors is of astrocytic origin. Astrocytes are a special kind of glial cells. They are star shaped and have numerous functions: They provide structural support for the brain and play an important role for the blood-brain-barrier. Additionally they are able to release and re-uptake several neurotransmitters (e. g. ATP, GABA and glutamate) and therefore influence synaptic transmissions. However, the understanding of the role of the astrocytes is far from complete and considered to be at an initial state. The group of astrocytomas (astrocytic tumors) itself can be divided in further subgroups according to their histopathology. The subgroups encountered in our patient collective are:

- Pilocytic astrocytoma (WHO 1)
- Diffuse astrocytoma (WHO 2)
- Anaplastic astrocytoma (WHO 3)
- Glioblastoma (WHO 4)

Oligodendroglial Tumors

Oligodendroglial tumors (oligodendrogliomas) are believed to stem from oligodendroglia, another type of glial cells. One of their main roles is to provide electrical insulation of neurons by wrapping their axons with a myelin sheath. In order to do so, they extend their branches to several axons which leads to a typical appearance (small nucleus and branch like extensions of the membrane). Oligodendrogliomas are divided into two subgroups:

- Oligodendrogliomas (WHO 2)
- Anaplastic oligodendrogliomas (WHO 3)

Oligoastrocytic Tumors

Oligoastrocytic tumors are mixed type tumors of the aforementioned groups. They contain astrocytic and oligodendroglial parts. Two subgroups with a differing grading exist, these are:

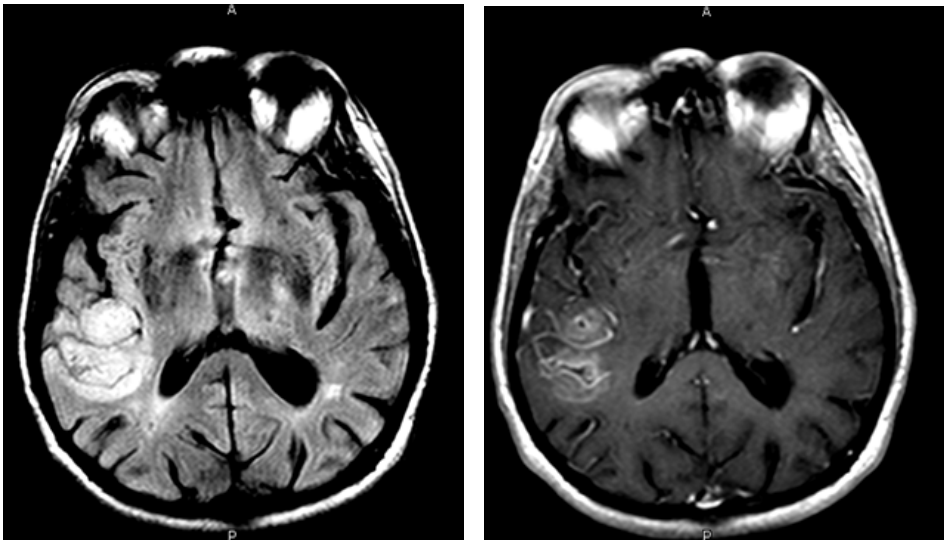
1. Oligoastrocytomas (WHO 2)
2. Anaplastic oligoastrocytomas (WHO 3)

2.4.4 Grading of Cerebral Gliomas

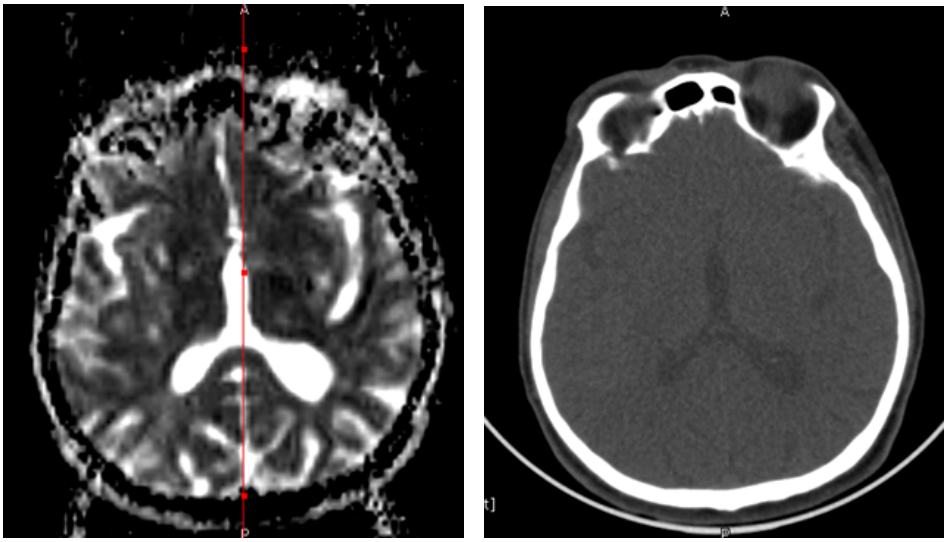
The gold-standard in terms of diagnostic confidence is the classification and grading of glial tumors based on histologic analysis of invasively gained bioptic samples. The bioptic grading allows for an accurate classification in over 90% of all cases [Well 04]. One has to keep in mind that the WHO classification scheme is mainly founded on the histopathologic appearance of those tumors.
It is of general consent that MRI of the human brain is the standard for the initial diagnosis of glial tumors. As mentioned in the previous sections, a variety of MRI sequences are acquired in order to cover different aspects of the glioma.
Metrics that can be obtained from contrasted T_1 and T_2 images and which are thought to indicate the class and the grade of a suspected glioma to a certain degree [Riem 02, Upad 11, Asar 94], are:

- Tumor heterogeneity
- Formation of a cyst or necrosis
- Hemorrhage or bleeding
- Crossing of the mid line
- Extent and degree of an edema
- Degree and heterogeneity of the contrast enhancement
- Shift of other regions of the brain (called mass effect)

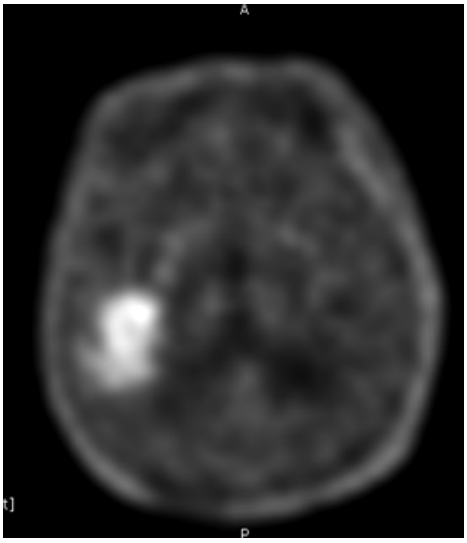
Other imaging modalities like X-ray CT, PET, and SPECT can add beneficial information for the task of tumor grading/classification: The extent and the kinetics of the amino acid uptake (obtained by FET-PET) and existence and shape of calcification (obtained by X-ray CT) of a glial tumor could give indications of its underlying class and grade [Rapp 13, Popp 07]. Figure 2.11 provides an example that shows the multimodal aspects of a glial tumor.



(a) T_2-FLAIR MRI (b) Contrast-enhanced T_1-MRI



(c) Diffusion-weighted MRI (d) Low-dose X-ray C



(e) FET-PET

Figure 2.11: Multimodal images of a 68 year old, male patient, diagnosed with Glioblastoma (WHO4) by biopsy.

Chapter 3

Patient Population and Image Acquisition

3.1 Patient Population

An initial screening of patients examined with FET-PET/CT in the Clinic of Nuclear Medicine in Erlangen yielded 232 patients between 05th of June 2007 and 21st of March 2011. After applying our inclusion criteria (listed below), we had 16 patients left. As part of a cooperation, we received another 32 patients from the Institute of Neuroscience and Medicine at Forschungszentrum Jülich, Germany, which met our inclusion criteria. Table A.1 provides additional details like age, weight, injected dose, time between the acquisition and several other parameters of the patient population. The **inclusion criteria** were:

- Dynamic PET acquisition using the amino acid FET, optional: CT images
- T_2-weighted MRI, contrast-enhanced T_1-weighted MRI, optional: diffusion-weighted MRI
- No prior treatment (esp. surgery, chemotherapy, radiotherapy)
- After imaging: tumor classification by histology on bioptic samples

As we have basically four different classification tasks, we create different (overlapping) subsets of our patient population for those tasks. The number of patients in each class is balanced by a random downsampling. The subsets are defined in Table 3.1. Table A.2 lists which patient is included in each subset.

3.2 Image Acquisition

The image acquisition consists of two main procedures for all patients in our collective: one is the PET(/CT) (Section 3.2.1), the other the MRI (Section 3.2.2). The average time distance between the separate PET(/CT) and MRI acquisition was ~ 18 $(0 - 77)$ days. The PET(/CT) and MRI sections themselves have instrumentation and acquisition parts: The medical devices and their specifications are given in the

Subset Name	Contained Modalities	Classes	Patients in Class
SUB44	T_1-MRI, T_2-MRI, PET	LG/HG	22/22
SUB32	T_1-MRI, T_2-MRI, PET	WHO1/2/3/4	8/8/8/8
SUB22	ADC-MRI	LG/HG	11/11
SUB14	X-ray CT	LG/HG	7/7

Table 3.1: The modalities, tumor classes, and number of patients contained in the subsets for the different classification tasks. LG: Low-grade, HG: High-grade.

instrumentation part, patient preparation, data acquisition and data reconstruction routines are described in the acquisition part.

3.2.1 PET/CT

The PET and CT data for 16 patients were acquired using a hybrid PET/CT system (TruePoint Biograph 64, Siemens Healthcare MI, Knoxville) (Figure 3.1). The PET and the CT are integrated into a common gantry in sequential spatial order. The PET data for the remainder of the patients (32) were acquired on a ECAT EXACT HR+ (Siemens Healthcare MI, Knoxville) 3.1.

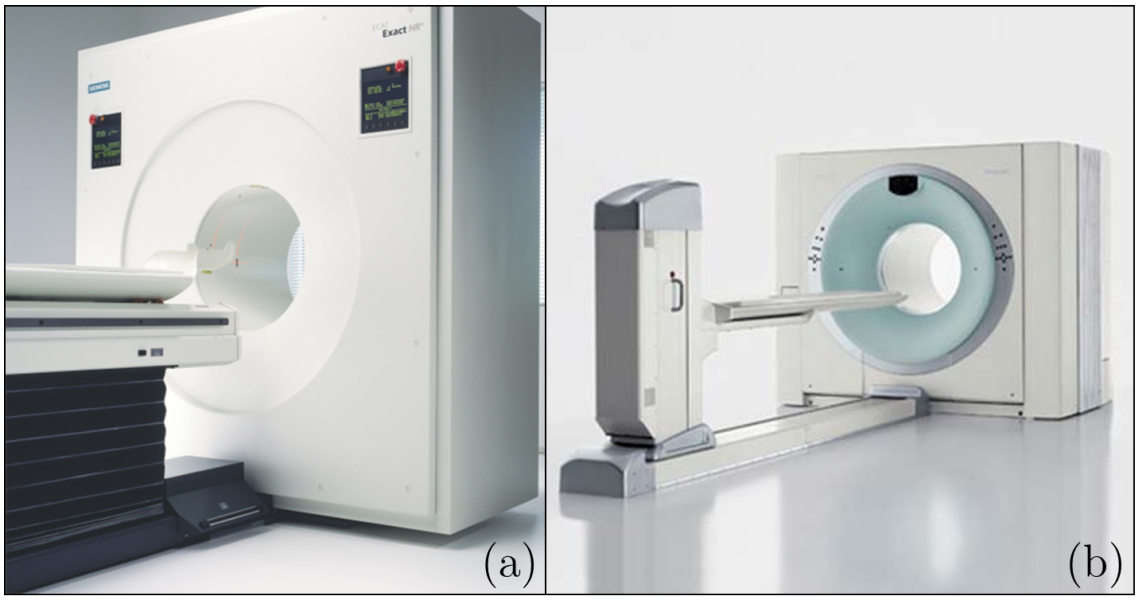


Figure 3.1: Product images of the PET system used in our studies: (a) Siemens Ecat Exact HR+, (b) Siemens Biograph 64 TruePoint PET/CT.

CT Instrumentation

The CT part consists of a 64 slice device with 40 detector rows. The detector consists of 26 880 ceramic elements in total. The transaxial FOV is 50 cm.

CT Acquisition

After the patient positioning, a low-dose CT scan was acquired. The following parameters were chosen: Size of the focal spot: 1.2 mm; Rotation time: 1.0 s; Tube voltage: 120 kV. The actual mean current for our patient collective was about 50 mAs. For each patient, we reconstructed two datasets using the manufacturer supplied FBP methods, one with a softer kernel (B08s) for attenuation correction purposes, one with a sharper kernel (B41s) for automatic classification and reading by the physicians. Both datasets have a matrix size of $512 \times 512 \times 111$ ($0.49 \times 0.49 \times 2.00$ mm). The reconstructed images have an intensity resolution of 16 bits. The image intensities are in Hounsfield Units (HU). After the CT procedure, the patient was transferred to the PET part of the gantry by an automatic table movement in axial direction.

PET Instrumentation

The TruePoint Biograph 64 system is equipped with Lutetium Oxyorthosilicate (LSO) as detector material. Each detector block consists of $13 \times 13 = 169$ individual detector elements, each with the size $4.0 \times 4.0 \times 20$ mm. Our scanner has four detector block rings, each consisting of 48 detector blocks. This results in a total number of 32 448 detector elements. The transaxial FOV is 605 mm, the axial 216 mm. The sensitivity of the system is 8.0 cps $\cdot$ kBq^{-1} [Jako 11]. According to the manufacturer, the axial resolution is 5.7 mm, and the transaxial resolution is 4.8 mm at 10 cm distance from the center of the FOV, following the measurement procedures of the NEMA 2001 standard.
The other scanner used for our study was the ECAT EXACT HR+. It is equipped with Bismuth Germanate (BGO) as detector material and each detector block consists of $8 \times 8 = 64$ individual detector elements, each with the size $4.0 \times 4.4 \times 30$ mm. It has four detector block rings, each consisting of 72 detector blocks. The total number of detector elements subsequently is 18 432. The transaxial FOV is 155 mm, the axial FOV is 583 mm. The sensitivity of the scanner is 26.4 cps$\cdot$ $\mathrm{Bq}^{-1}\cdot$ ml^{-1} [Adam 97], the axial resolution is 5.3 mm, the transaxial resolution is 5.4 mm at 10 cm distance to the center of the FOV. The ECAT EXACT HR+ is no hybrid device, thus uses three Ge-68/Ga-68 rod sources to acquire transmission data for attenuation correction.

PET Acquisition

The PET acquisition started simultaneously with the injection of the radioactive tracer. The PET images were acquired over $40 - 50$ minutes, the raw PET detector events were recorded together with a time stamp (listmode acquisition). This allowed a retrospective reconstruction into any desired combination of time bins. For the Biograph 64, we choose five short (1 minute each) bins at the beginning of the acquisition and subsequently seven long (5 minutes each) time bins throughout the end of the acquisition (40 minutes total). The time bins for the ECAT EXACT HR+ varied slightly: five 1 minute frames followed by five 3 minute frames, followed by four or six 5 minute frames (40/50 minutes total). The raw data of each bin were reconstructed into (nearly) isotropic 3-D datasets with a size of $168 \times 168 \times 109$ ($2.03 \times 2.03 \times 2.02$ mm), using the iterative OSEM algorithm with six iterations

and eight subsets and CT based attenuation correction for the Biograph 64 and $128 \times 128 \times 63$ ($2.00 \times 2.00 \times 2.42$ mm) with either Filtered Back Projection (Shepp-Logan Filter FWHM 2.48 mm) or iterative OSEM with six iterations and 16 subsets and Ge-68/Ga-68 transmission based attenuation correction for the ECAT EXACT HR+. Corrections for decay, scatter and random coincidences were applied for both scanners according to the implementation of the manufacturer. As post-processing step at the Biograph 64, smoothing with a Gauss filter (Kernel width 5 mm) was used. Consequently this yielded in total twelve 3-D datasets for the Biograph 64 and 14 or 16 3-D datasets for the ECAT scanner. The intensity resolution of these files was 16 bits. As both PET scanners were calibrated in order to allow absolute quantification, the image intensities were converted to Becquerel per Milliliter ($\mathrm{Bq} \cdot \mathrm{ml}^{-1}$) with the help of the calibration factors. These factors are individual for every acquisition and were stored in the DICOM tags for rescale slope (tag address: $0028, 1053$) and rescale intercept (tag address: $0028, 1052$) of the reconstructed datasets. An overview of the PET acquisition parameters can be found in Table A.8, the assignment of the acquisition protocols to the individual patients can be found in Table A.2.

3.2.2 MRI

MRI Instrumentation

The MRI acquisitions were done on various systems (for images, see Figure 3.2). One Philips 1.0 T system (Gyroscan NT), three Siemens 1.5 T systems (Magnetom Avanto, Magnetom Sonata, Magnetom Symphony, Siemens Healthcare, Erlangen), one Philips 1.5 T system (Intera, Philips Electronics N.V.) and one 3 T system (Magnetom Trio, Siemens Healthcare). The following section lists the acquisition parameters for the MRI data used in our study. In general, even the parameters within one weighting differ, which reflects the fact that the data originate from clinical routine.

MRI Acquisition

T_1-MRI
For the T_1-weighted MRIs an intravenous injection of 0.1 mmol/kg body weight Gadobutrol (Gadovist, Bayer Schering Pharma) was used in most cases as contrast agent (alternatively Gadodiamide, Omniscan, GE Healthcare; Gadopentetate dimeglumine, Magnevist, Bayer Schering Pharma). The images were acquired with a TE ranging from 1.7 to 17 ms and TR ranging from 145 to 690 ms. The typical slice thickness was $3-6$ mm with in-plane pixel size of $0.4-0.9$ mm. However, the parameters for the (almost) isotropic T_1-MRI (MPRAGE) differed: The TE was $2.2-3.93$ ms, TI was $900-1100$ ms, TR was $1950-2200$ ms, in plane pixel size was $0.5-1.1$ mm with $1.0-1.5$ mm slice thickness.

T_2-Flair MRI
The anisotropic T_2-Flair MRIs were acquired using an inversion time of $1800-2500$ ms, TE ranged $79-150$ ms, TR $5000-10000$ ms. The in-plane pixel size was $0.4-1.0$ mm with a slice thickness of $3-6$ mm. Again, the parameters for the isotropic datasets differed: The TI was 1800 ms, TE was 389 ms, TR was 5000 ms.

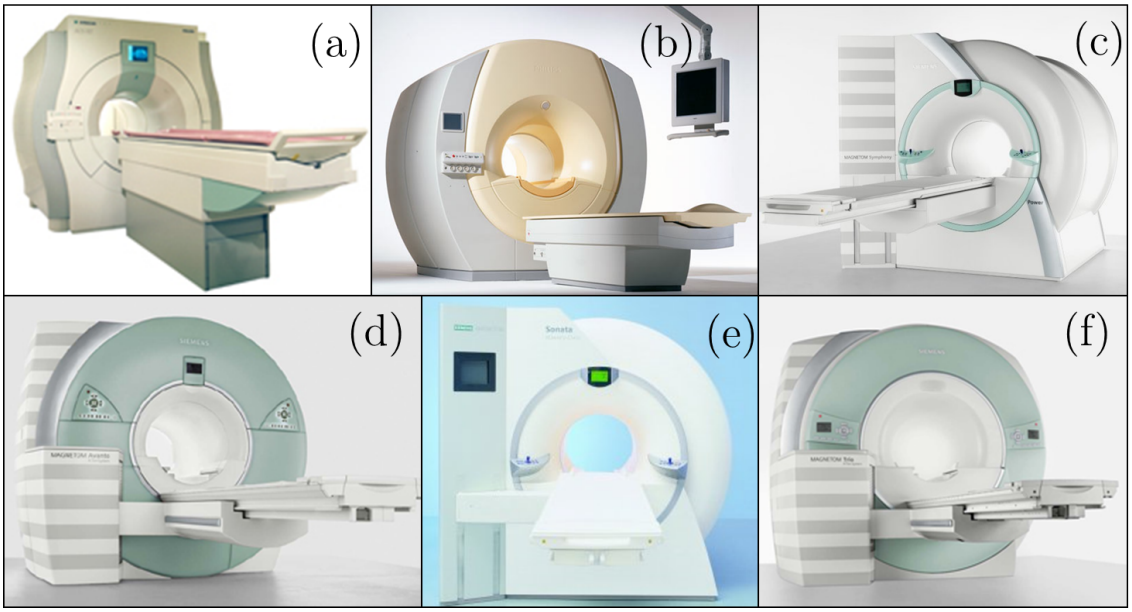


Figure 3.2: Product images of the MRI systems used in our study. (a) Philips Gyroscan NT, (b) Philips Intera, (c) Siemens Magnetom Symphony, (d) Siemens Magnetom Avanto, (e) Siemens Magnetom Sonata, (f) Siemens Magnetom Trio.

The in-plane pixel size was 0.5 mm and the slice thickness was 1 mm.

ADC MRI
For the diffusion weighted MRIs, the TE was $91 - 101$ ms, TR was $3200 - 3900$ ms. The in-plane pixel size was $0.6 - 1.8$ mm and the slice thickness ranged $5 - 6$ mm.

An overview of these parameters for the different MRI sequences can be found in the appendix, Table A.7. The assignment of the sequences to the individual patients in our study is listed in Table A.2.

Chapter 4

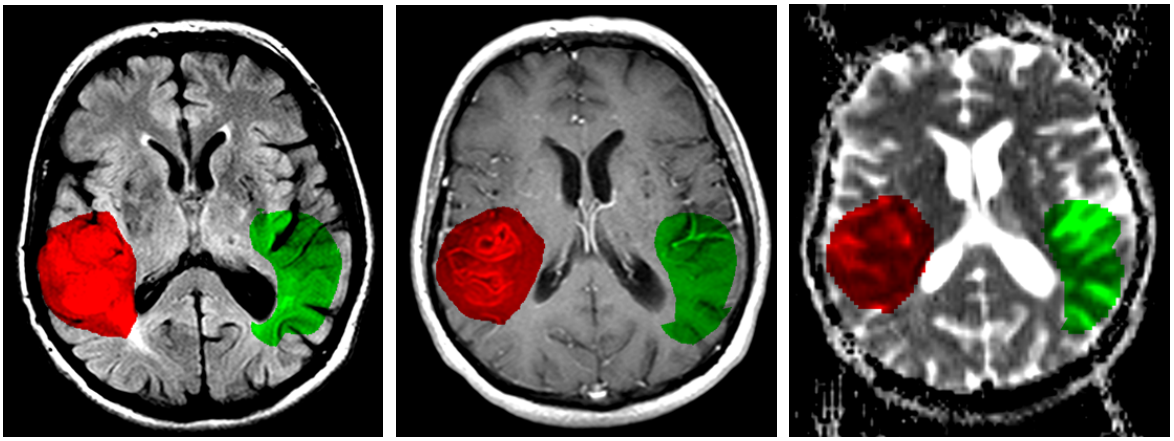
Preprocessing and Feature Handling

Proper preprocessing is crucial for a high classification performance. For this reason, we applied multiple methods to our patient datasets. These preprocessing methods are described in the current chapter. We wrote a software framework for the purpose of feature extraction and for the incorporation of the various preprocessing steps: This framework is briefly introduced in Section 4.2. After that, the criteria for the definition of the volumes of interests (VOI) from which the features are extracted are outlined (Section 4.1). We then propose methods for inter-device and inter-patient normalization of medical images (Section 4.3). We outline a motion correction technique for dynamic PET as well (Section 4.5). The automated multimodal image registration is a necessary prerequisite in many of our preprocessing steps. For this reason, we evaluated the accuracy of the multimodal registration (Section 4.4). Thereafter, we provide a detailed description of the extracted features by specifying the equations to calculate them (Section 4.6). Feature normalization is carried out in two different methods (Sections 4.7.1 and 4.7.2). In order to ensure proper inter-device calibration and as we calculated the quantitative values of PET and CT (namely kBq/ml and HU) in our own framework, we confirmed the accuracy of these calculations by a phantom experiment (Section 4.2.1). In our framework, the single features were collected into feature sets. These feature sets are subsequently used in the classification stage. The feature selection process and feature transformations are described in Section 4.8.

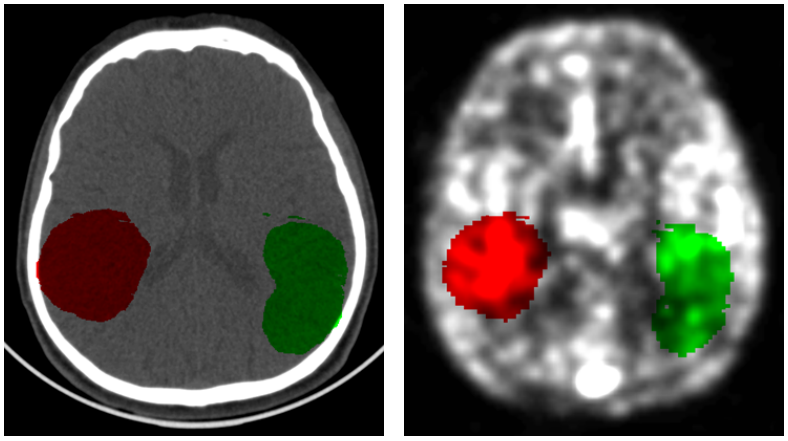
4.1 Volume of Interest Definition

In general, it is possible to define a VOI in several ways: there exist fully automatic (e. g. by automatic segmentation), semi-automatic (e. g. by segmentation using manually defined seed points) or manual methods. We decided to use a manual delineation of the VOI on the basis of the images from T_2-weighted MRI. This is motivated by the fact that an implementation and evaluation of the automated methods was beyond the scope of this work. The manual segmentation is known to be robust, therefore we could focus on the classification task itself. In the future, automated methods should be implemented as a viable enhancement on the way towards further automation.
The T_2-weighted MRI itself is considered as the standard sequence for the tumor localization in the brain. It has a superior soft-tissue contrast when compared to

X-ray CT. T_2-MRI also promises increased sensitivity and specificity over FET-PET for detecting intracranial lesions. This is mainly caused by the lower spatial resolution of PET when compared to MRI and due to non-specific FET uptake of healthy structures in the brain, which is especially given for tumor that are small and show low FET uptake.
The borders of the tumors were delineated slice-by-slice with help of the ITK-Snap tool [Yush 06] by a board-certified radiologist. In addition to the tumor VOI, one contra-lateral, healthy reference VOI with approximately the same volume as the tumor VOI was drawn. An example of this procedure is given in Figure 4.1. Due to the heterogeneity of MRI sequences, the image windowing for drawing the VOIs had to be adjusted manually to ensure proper tumor visualization.
The average volume for the tumor and reference VOI were 42.6 cm^3, ranging from 2.8 cm^3 to 156 cm^3. The VOIs were subsequently saved as DICOM datasets and imported into the InSpace volume renderer. Further computations such as co-registration, inter-dataset normalization and feature extraction and calculation were carried out in that program. InSpace is a standalone version of the commercially available Syngo InSpace application (Siemens Healthcare).



(a) T_2-FLAIR MRI (b) Contrast-enh. T_1-MRI (c) Diffusion-weighted MRI



(d) Low-dose X-ray C (e) FET-PET

Figure 4.1: Example for a volume of interest in multimodal images of a 68 year old, male patient, diagnosed with Glioblastoma (WHO4) by biopsy. The red VOI covers the tumor, the green VOI a contralateral, healthy reference region. In the current case, the ventricles of the left and right hemisphere are asymmetric due to the compression of the ventricle by the tumor. As the VOIs should cover the soft-tissue parts of the brain and not areas filled with cerebrospinal fluid, the VOIs needed to be adjusted and consequently became asymmetric.

4.2 Software Framework: Multi-Modality Work Bench

We developed a software tool which incorporates several of the methods presented in the following sections in a single application. We called the software "Multi-Modality Work Bench" (MMWB). It is entirely written in C++ and incorporates methods of the Insight Toolkit (ITK, [Iban 05]). MMWB is realized as a plug-in to the InSpace volume rendering software. It handles static as well as dynamic 3-D images (multiple 3-D datasets ordered as consecutive time frames). So far, it was successfully tested with MRI, CT, PET and Ultrasound (US) datasets. It extracts the features in fully automated fashion and outputs a result file with the extracted features of a certain patient. These result files are later on used in the classification step of our experiment. The main features of MMWB are:

- Concurrent handling of up to six static and three dynamic 3-D datasets for feature extraction.
- Importing of VOI files (MetaImage file format), generated by ITK-Snap (Section 4.1), containing up to 5 VOIs.
- Motion correction for dynamic datasets by image registration (Section 4.5).
- MR image intensity normalization by histogram matching (Section 4.3.3).
- Feature extraction from the multimodal datasets:
  - Co-registration (Section 4.4) of all images to the reference dataset (dataset where VOIs were defined).
  - Transfer of VOIs to all dynamic and static datasets.
  - Feature extraction from all dynamic and static datasets (list of features, Section 4.6).
  - Calculation of quantitative ECT features (Section 4.3.1).
- Output of a file which contains all extracted features for a specific patient.
- Intuitive graphical user interface.
- Visualization of multi-modal features in the InSpace application (e. g. see Figure 4.2).
- Configurable via config file.

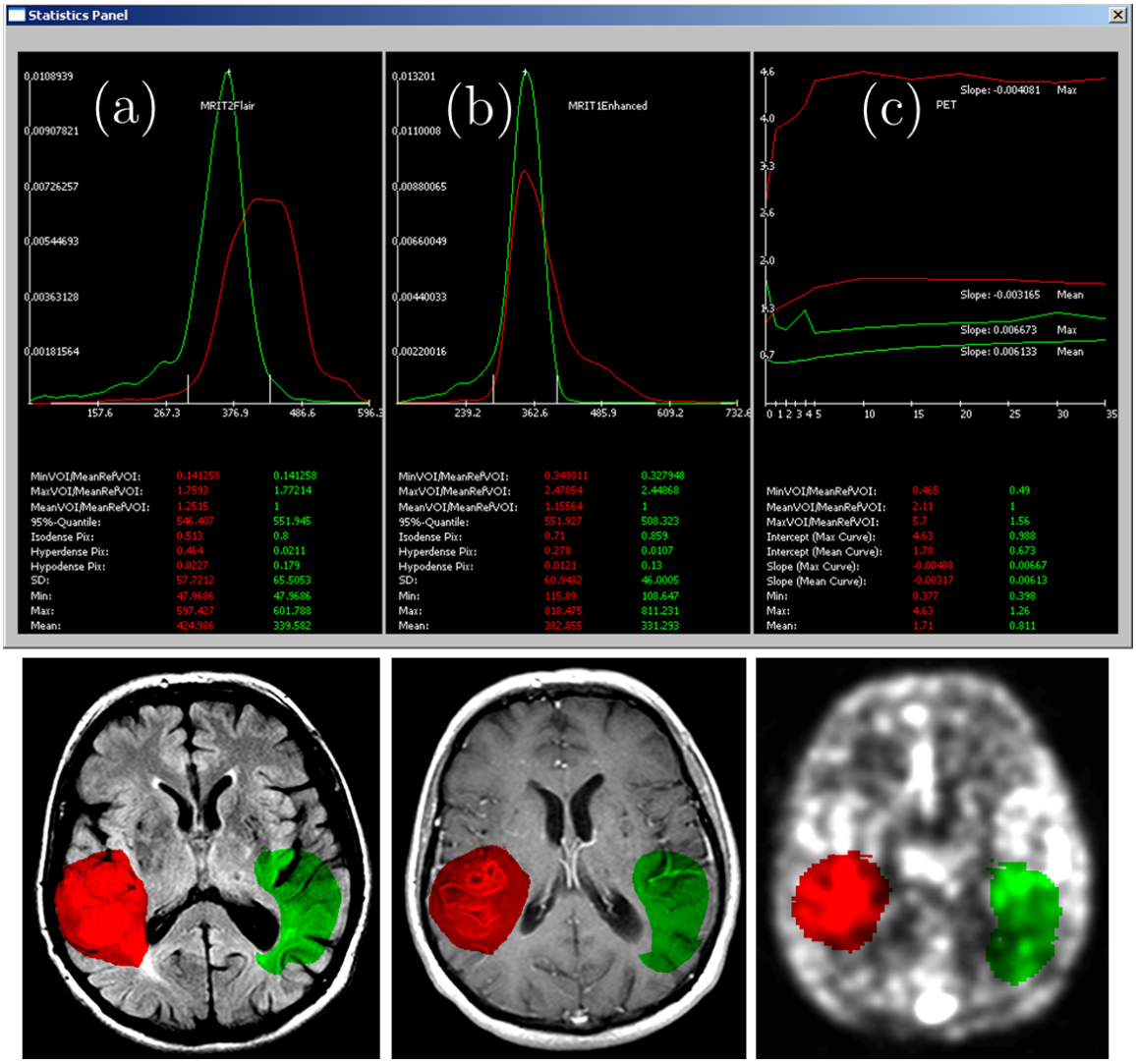


Figure 4.2: Example of the MMWB display of the extracted features of a 68 year old, male patient, diagnosed with Glioblastoma (WHO4) by biopsy. For the static modalities ((a): T_2-FLAIR MRI, (b): contrast-enhanced T_1-MRI), the distribution of image intensities in certain VOIs (red: tumor VOI, green: contralateral, healthy VOI) are displayed in the upper part of the window. For the dynamic modalities ((c): FET-PET), the upper part of the window displays the time-activity-curves of mean and max gray levels in the VOIs. Additionally, the slope of the curves between 10-40 minutes are shown as well. Below these curves, the values for several extracted features are listed. One can see, that the T_2-MRI of the patient shows increased intensity values when compared to the reference region, which represents an increased protein accumulation in the tumor. In the T_1-MRI, one finds an extravasation of the intravenous Gadolinium contrast agent, which indicates a disruption of the blood brain barrier and leads to hyper-intense image intensities when compared to the reference region. This could potentially indicate a higher grade tumor. The PET shows a distinct increase in FET-uptake quotient between tumor and reference VOI, which, together with the negative slopes of the time-activity-curves, points towards a higher grade tumor as well.

Description	Expected Value	Measured Value	Deviation
Mean activity concentration	8.05 kBq/ml	8.10 kBq/ml	$\approx 1\%$
Mean SUV	1.00	1.01	$\approx 1\%$
Intercept of dyn. curve	8.05 kBq/ml	8.11	$\approx 1\%$
Slope of dyn. curve	0 kBq/ml/min	0.00	$\approx 0\%$
Quotient to reference VOI	1	1	$\approx 0\%$
Mean HU value	0 HU	−1.8 HU	$\approx 2\%$
Quotient to reference VOI	1	0.99	$\approx 1\%$

Table 4.1: Comparison of the expected and measured values for PET phantom measurement. The measurement confirmed our ECT calibration: Only minor deviations were found.

4.2.1 Validation by Phantom Measurements

We conducted several phantom measurements in order to validate our PET system calibration and to ensure that our software extracts the correct voxel intensity values from PET and CT.
Firstly, the quantitative accuracy in PET was evaluated: We filled a standard quality control phantom (Flangeless Esser PET Phantom, DataSpectrum Corp.) with 73.1 MBq of F-18 (measured in well counter) suspended in water. The active volume of the phantom was measured by weighting the phantom in filled and empty state, calculating the difference and dividing by the density of water (0.998 g/cm^3 at 21°C). We found the volume to be 5.69 l, thus the activity concentration was 12.8 kBq/ml at preparation time and 7.6 kBq/ml at acquisition time. The cross calibration factor between the well counter and the PET device was 1.06, which leads to a corrected activity concentration of 8.06 kBq/ml at acquisition time. The phantom was dynamically measured for 40 minutes and reconstructed in 12 volumes of $200 \times 200 \times 109$ voxels ($4.07 \times 4.07 \times 2.03$ cm^3) each, using the OSEM algorithm with 6 iterations, 8 subsets, 5.0 mm Gauss filter, scatter correction, and CT-based attenuation correction. The 12 volumes consisted of data from five 1-minute frames and seven 5-minute frames. These parameters resemble the parameters of the PET acquisitions for the patients (Section 3.2.1). The phantom was evaluated in the same way as the patient data: Three VOIs were manually drawn and the intensities in the VOIs were evaluated using our tool (Section 4.1).
Table 4.1 shows a comparison of expected values and measured values for the analysis of the phantom. We always used the expected value as normalizing constant for the calculation of the deviation, except for CT where we chose 100 HU. This is motivated by the fact, that the soft tissue in brain varies around the same extent (see Table 2.1).

In general, the deviations are in the low single-digit percentage range. The deviations that we found were not caused by our software but by inaccuracies in the respective scanning process. It is known that absolute quantification in emission tomography devices is only possible within certain errors (Section 2.1). The values for the features *Mean activity concentration*, *Mean SUV* and *Intercept of dyn. curve*

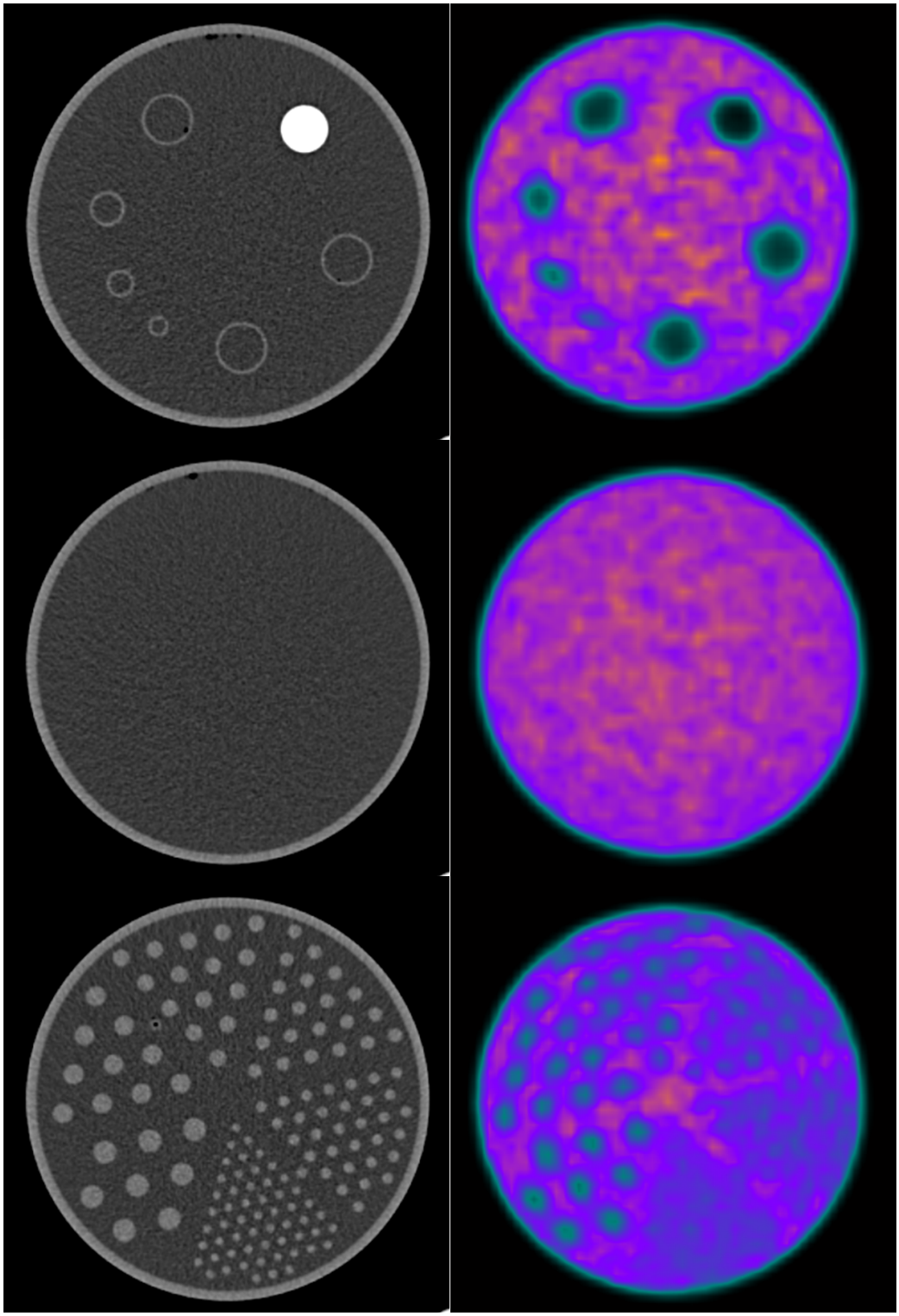


Figure 4.3: CT (*left*) and PET (*right*) images of an ECT quality control phantom with three different sections. *First row*: Cold cylinders in hot background. *Second row*: Homogeneous section of hot background. *Third row*: Cold bars in hot background.

therefore are considered to be extracted in an accurate way. The *Mean HU value* of the phantom slightly differs from the expected 0 HU due to beam hardening artifacts of the CT scan. However, the deviation of 2 HU is negligible for our purposes as the inter-patient variation is considerably higher.

4.3 Inter-Dataset Normalization

Inter-dataset normalization for medical images refers to a process with the goal to ensure that identical anatomical or physiological structures have constant intensity values, regardless of their imaging system and sequence variant (inter-device) or patient (inter-patient). For this, it is desirable to measure absolute physical quantities in order to achieve inter-device comparability. These can be HU in CT, absolute activity concentration in ECT or relaxation times in MRI. Very little need for additional inter-device standardization is given for medical imaging devices that rely on these physical properties. Nevertheless, a small amount of inter-patient variation still exists: The HU values of bones in CT, as well as soft tissues in MRI are known to vary across different patients. Additionally, the image intensities vary if external contrast agents are used (e. g. iodine for CT and gadolinium for MRI) and depend on the application protocol (amount, time point, bolus pressure,...).
Emission tomography consequently suffers from the same inherent problem as it is based on radioactive tracers which are external contrast agents as well. Thus, the image intensities depend on the amount and the distribution volume of the injected radioactive tracer, even if the ECT device itself is calibrated to absolute activity concentrations. There exist several approaches to overcome these limitations:

4.3.1 PET Normalization

In emission tomography like PET and SPECT, the concept of the calculation of standardized uptake values (SUV) is widely used for improved diagnostic results [Habe 91]. In our case, we use the following method to calculate the SUV (Equation 4.1).

$$\mathrm{SUV} = \frac{C_{\mathrm{measured}}}{A_{\mathrm{injected}}/M_{\mathrm{patient}}} \tag{4.1}$$

where C_{measured} is the activity concentration in kBq/ml, measured in the PET image and A_{injected} is the injected activity in MBq at the time point of the image acquisition and M_{patient} is the body weight of the patient in kg.
We also used other methods for the inter-patient standardization. We calculated quotients of tumor and reference region (see Section 4.1). This concept is established in nuclear medicine and is expected to deliver more reliable diagnostic results when compared to other methods. It is recommended in the *German guidelines for brain tumor imaging by PET and SPECT using labeled amino acids* [Lang 11].

4.3.2 CT Normalization

In X-ray CT, inter-patient normalization is not commonly applied. In practice, the scanners are calibrated to the attenuation values of air and water at the tube voltages of the device with the help of body mimicking phantoms. However, some patient variations remain, which is mainly due to beam hardening. One can partially overcome these effects if CT devices with two X-ray sources are used. The two sources operate at different tube voltages [Heis 09]. We did not take into account dual-source CT in

our study, as we did not have this kind of data available. Still, we do not consider this to be a major drawback as we found the inter-patient variation to be very low as only the soft tissue parts of the brain were covered. This can be observed from the HU values in the reference VOI, which represent normal brain tissue: The standard deviation of ≈ 8 HU (at ≈ 39 HU mean) is low compared to the range of general soft tissue in CT, which is at least 100 HU for the soft tissues of the brain. Nevertheless, we applied the quotient normalization method known from ECT since it does not rely on sophisticated acquisition techniques.

4.3.3 MRI Normalization

The biggest range of soft tissue intensities is observed in the MRI sequences. This is mainly caused by the fact that the images of the patients stem from different MRI devices and were obtained using different sequences. This variation is expressed by a large range of image intensities for normal brain tissue acquired by similar MRI sequences and can be seen in Figure 4.5. The inter-dataset variance of $\pm \approx 47\%$ for T_1-, $\pm \approx 36\%$ T_2-weighted MRI, and $\pm \approx 109\%$ for diffusion-weighted MRI is much higher than the variation seen in CT.
In order to reduce the amount of inter-patient and inter-device variation, we applied a histogram matching approach which is based on the work of Nyul et al. [Nyul 99]. It employs a piecewise-linear mapping of a target histogram to a reference histogram, based on certain calculated landmarks of these histograms. The approach is reported to be robust and sufficient for "easy" anatomies like cranial MRI. Still, more advanced methods which are even suitable for the normalization of whole-body MRI exist: Jäger et al. [Jage 10] describe an approach that relies on non-rigid registration of the target and the reference histogram. The approach consists of two important steps, step one employs an affine alignment of the histograms, step two incorporates a regularized, non-rigid registration. The regularizer term incorporates the first derivative of the deformation field and ensures that neighboring intensities are mapped as similar as possible. Both, affine as well as the non-rigid registration use either the Sum of Squared Differences or the Jeffrey Divergence as measure for the goodness of fit between the reference and the deformed target histogram. The method has the advantage that it allows for intensity transformations that are not piecewise-linear but non-linear, which is probably the case in reality.
As we aim for the normalization of cranial MRI, we decided to apply the well evaluated approach by Nyul et al. for its ease of usage. Still, we believe more sophisticated normalization methods should be tested in future research. Our method worked as follows:
Let $\mathcal{V}$ be an image consisting of an array of voxels v_i with certain intensities. Without further restrictions we assume we have integer intensities and that $v \geq 0 \quad \forall v \in \mathcal{V}$, namely $v \in \mathbb{N}_0^+$. $\mathcal{V}_j, \quad j \in \{1, \ldots, N\}$, is the set of all N images in the study, each image has m_j voxels. The maximum, minimum and mean intensities are defined as $\overline{m}_j = \max\{v\} \quad \forall v \in \mathcal{V}_j$, $\underline{m}_j = \min\{v\} \quad \forall v \in \mathcal{V}_j$, $\mu_j = \frac{1}{m_j} \sum_{i=1}^{m_j} v_i \quad \forall v \in \mathcal{V}_j$. The mean intensity of the MR image is considered as a good approximation for dividing the image into the foreground $\mathcal{F} = \{v_i \text{ with } v_i > \mu\}$ and the background $\mathcal{B} = \{v_i \text{ with } v_i \leq \mu\}$. This approach has been proven to work for a multitude of

MRI images, including cranial MRI [Nyul 99].
Now let $\mathcal{H}_j$ be the histogram of $\mathcal{F}_j$ for image $\mathcal{V}_j$ and $d_{k,j}$, $k \in \{10, 20, \ldots, 90\}$, be the k-th percentile of $\mathcal{H}_j$ (deciles). The image containing the transformed and corrected intensities is called target image and denoted by $\mathcal{V}^*$. One fixed image $\mathcal{V}_0$ is the reference image: In our setup, the reference image is an MRI image of a healthy patient which is not part of the datasets that are classified, in order to avoid a potential bias of the intensity normalization towards a certain tumor class.
Finally, we define a piecewise linear transformation $f(x)$ for the image intensities:

$$f(x) = \begin{cases} \dfrac{\underline{m}_0 - d_{10,0}}{\underline{m}_j - d_{10,j}}(x - \underline{m}_j) + \underline{m}_0, & \text{if } \underline{m}_j \leq x < d_{10,j} \\ \dfrac{d_{10,0} - d_{20,0}}{d_{10,j} - d_{20,j}}(x - d_{10,j}) + d_{10,0}, & \text{if } d_{10,j} \leq x < d_{20,j} \\ \vdots & \\ \dfrac{d_{90,0} - \overline{m}_0}{d_{90,j} - \overline{m}_j}(x - d_{90,j}) + d_{90,0}, & \text{if } d_{90,j} \leq x \leq \overline{m}_j \end{cases}$$

Using the above definitions, the corrected images $\mathcal{V}_j^*$ are calculated using Algorithm 1.

input : Reference image $\mathcal{V}_0$, target images $\mathcal{V}_j, j = 1, 2, \ldots, N$
ouput: Transformed images $\mathcal{V}_j^*, j = 1, 2, \ldots, N$
begin
  calculate $\underline{m}_0$, $\overline{m}_0$ and μ_0 for the reference image;
  calculate the histogram $\mathcal{H}_0$ of the foreground $\mathcal{F}_0$;
  calculate $d_{k,0} \quad \forall k \in \{10, 20, \ldots, 90\}$ from $\mathcal{H}_0$;
  forall the $j \in \{1, \ldots, N\}$ **do**
    calculate $\underline{m}_j$, $\overline{m}_j$ and μ_j for $\mathcal{V}_j$;
    calculate the histogram $\mathcal{H}_j$ of the foreground $\mathcal{F}_j$;
    calculate $d_{k,j} \quad \forall k \in \{10, 20, \ldots, 90\}$ from $\mathcal{H}_j$;
    map the intensity values of all $v \in \mathcal{V}_j$ according to function $v^* = f(v)$;
    save the mapped intensity values in the transformed image $\mathcal{V}^*$;
  end
end

Algorithm 1: Algorithm for the MRI intensity normalization

Figure 4.4 illustrates the piecewise linear rescaling of the image intensities. The parameters were set to 100 levels of the histogram $\mathcal{H}$ and eleven matching points (nine deciles and two extrema) for the linear transformation. As mentioned above, the background of the images was not considered for the histogram generation and matching. This choice of parameters is reported to successfully reduce the inter-dataset variation on a huge amount of different MRI datasets [Nyul 00].
The actual implementation uses the itk::HistogramMatchingFilter of the Insight Toolkit (ITK, [Iban 05]).

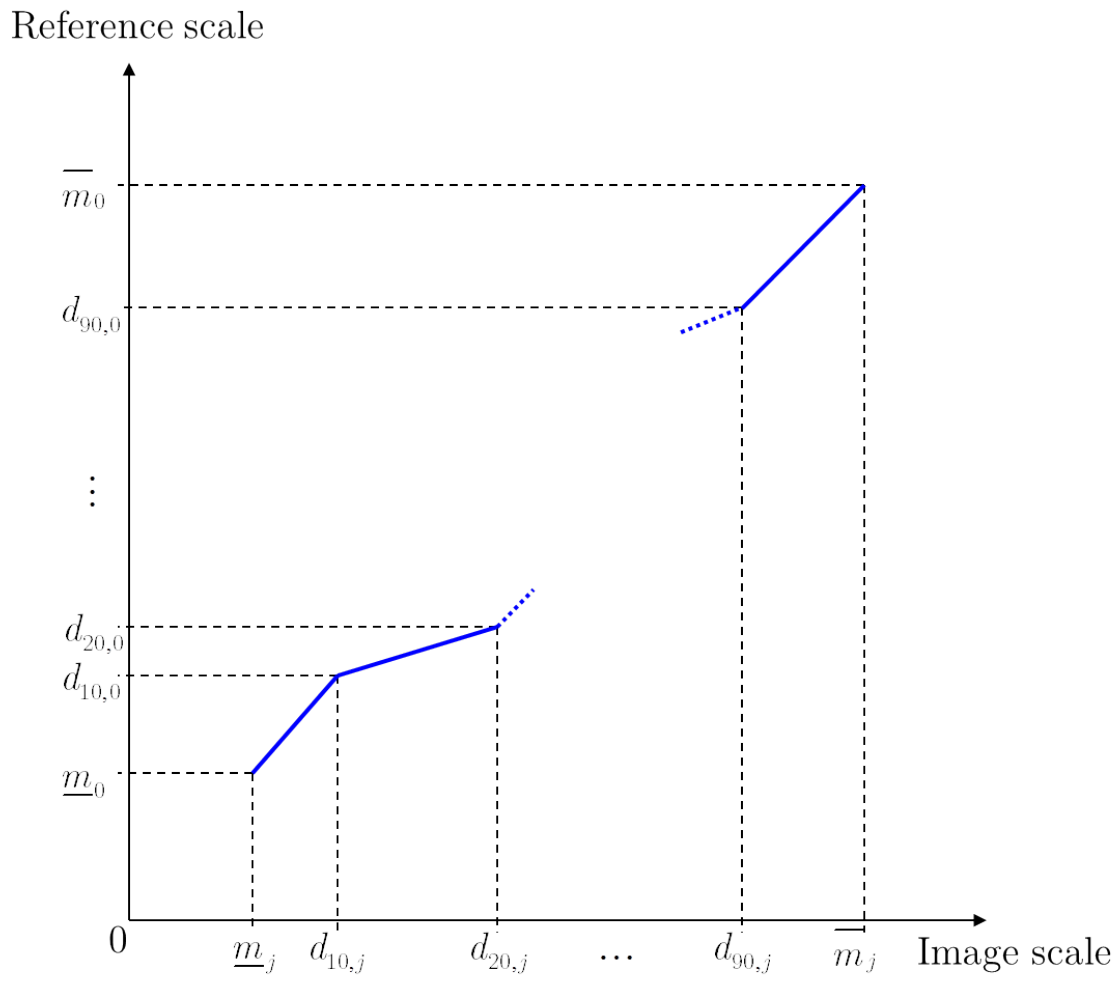
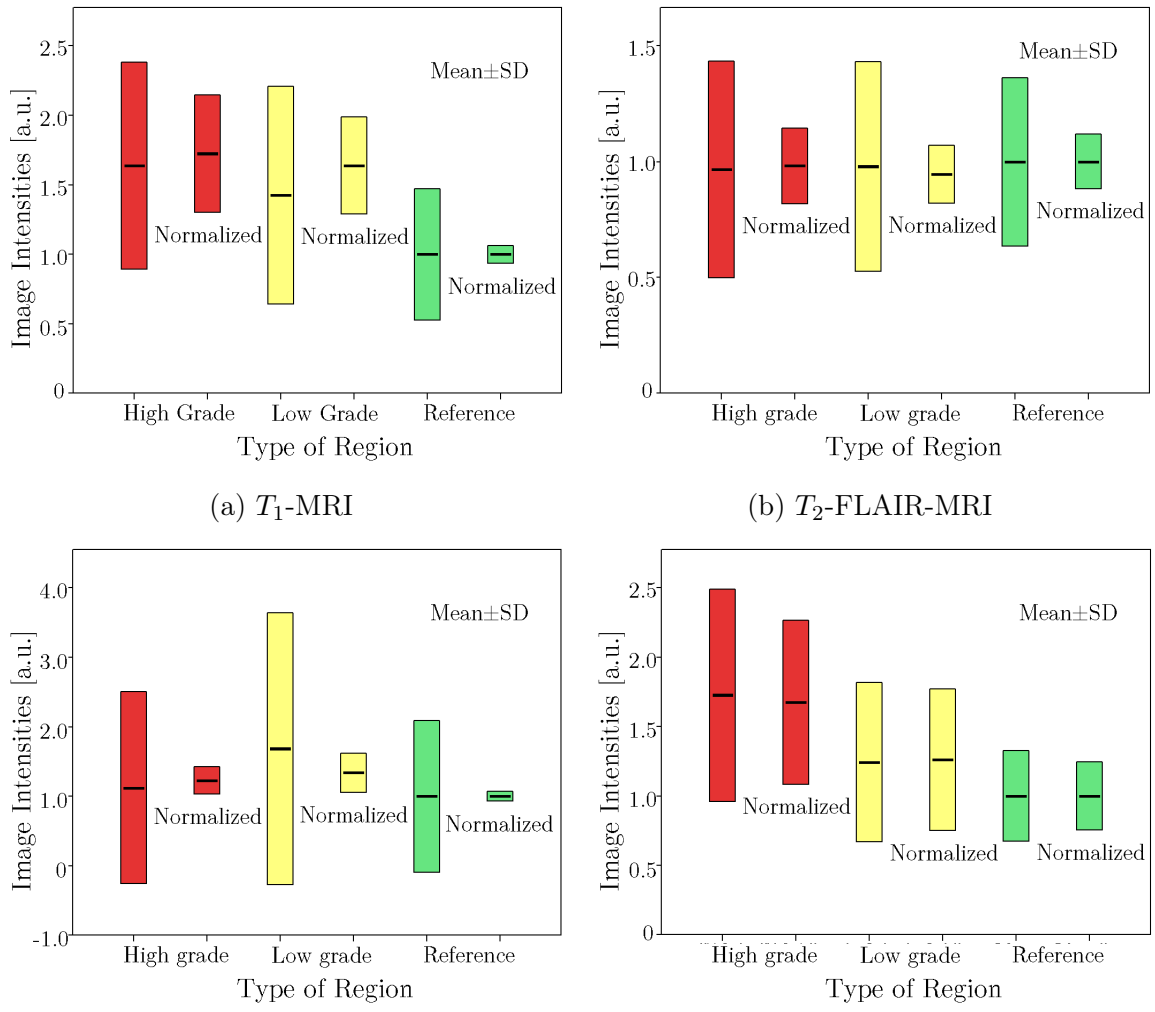


Figure 4.4: The intensity x of the target image $\mathcal{V}_j$ is piecewise linearly scaled on basis of points derived from the reference image $\mathcal{V}_0$ and $\mathcal{V}_j$. Together with the extrema $\overline{m}$ and $\underline{m}$, we have chosen the deciles d_k, $k \in \{10, 20, \ldots, 90\}$ as landmark points.

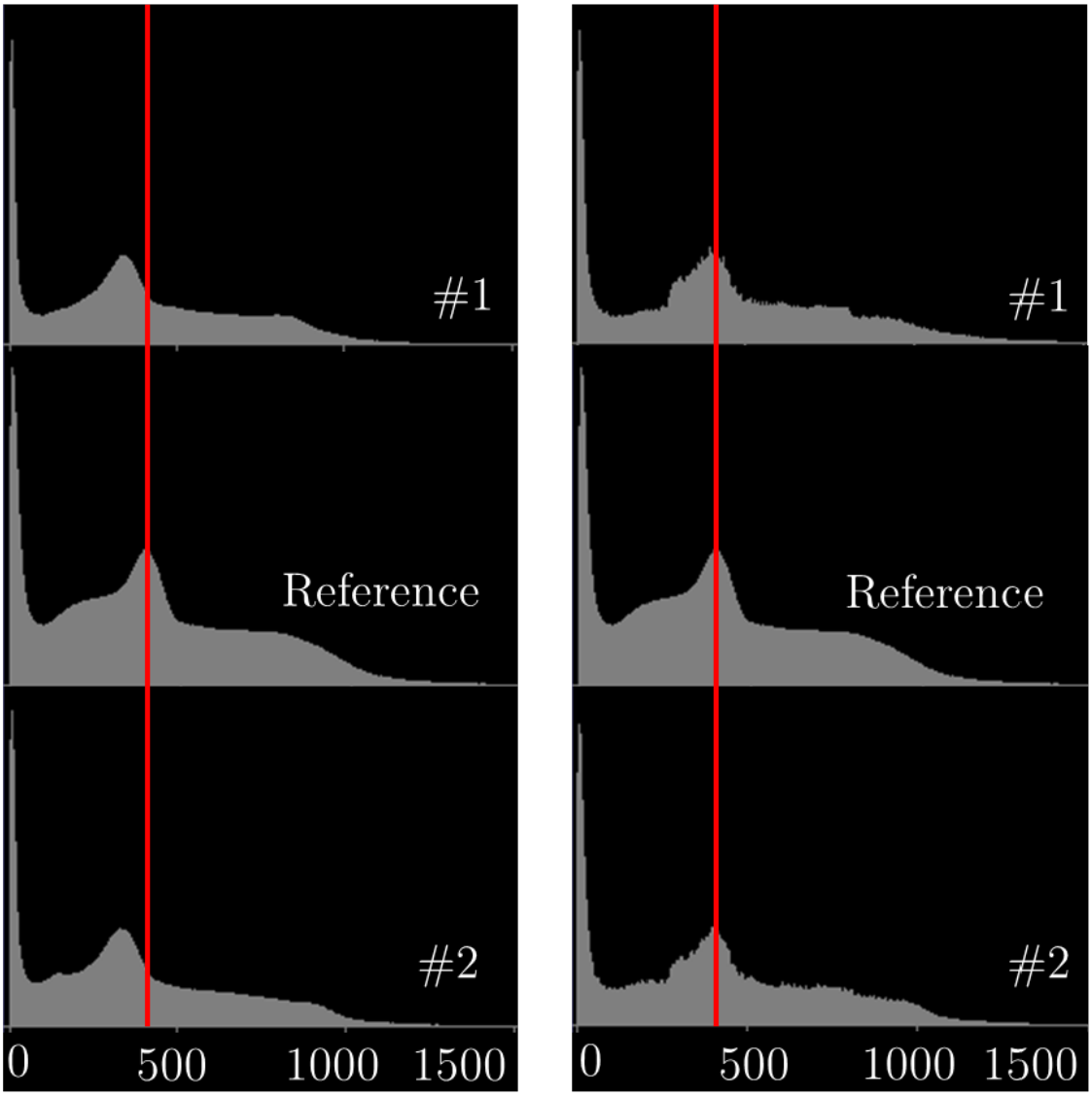
By applying this method, we successfully reduced the inter-dataset variation for our patient collective, not only in the healthy reference regions but also in the different tumor regions. The effect is visually perceptible (see Figure 4.7), aligns the image's histograms (Figure 4.6) and reduces the inter-dataset variation (Figure 4.5).
There exist multiple reports on the benefit of MRI intensity normalization for other segmentation and classification tasks, e. g. [Weis 04, Coll 04]. However, to the best of our knowledge we were the first group that applied MRI normalization for the automatic classification of cerebral gliomas. The amount of the increased classification accuracy is presented in Section 6.3 and discussed in Section 7.3.



(a) T_1-MRI (b) T_2-FLAIR-MRI

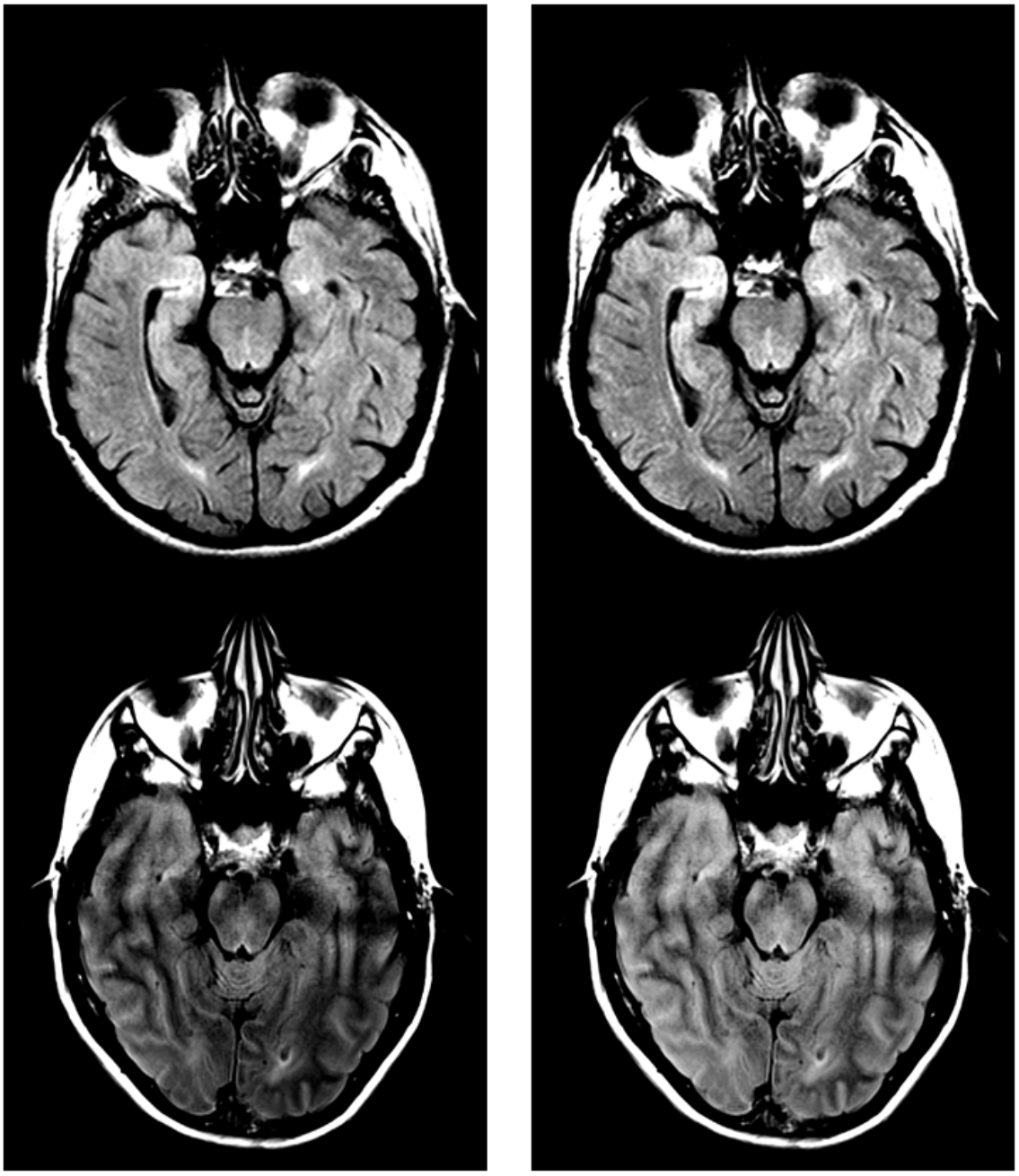
(c) ADC-MRI (d) PET

Figure 4.5: The range of image intensities, separated by type of VOI and normalized/non-normalized datasets are shown. The range of the image intensities was calculated by the mean and one standard deviation for all patients in subset SUB44. *Red:* high-grade tumors, *yellow:* low-grade tumors, *green*: healthy reference region. The left bar of each colored pair reflects non-normalized values, the right bar inter-dataset normalized values. All intensities and standard deviations are displayed as multiples of the mean intensity of the reference region. One can see that the inter-dataset variation is significantly reduced for the normalized datasets.



(a) (b)

Figure 4.6: Column (a) compares the histograms of two patients to a reference histogram. Column (b): After matching to the reference histogram, the example histograms show an improved alignment especially in the intensity regions of the soft tissue in the brain. The red line indicates the peak intensity of the brain tissue in the reference histogram.



(a) Non-Normalized (b) Normalized

Figure 4.7: Column (a) displays non-normalized MRI images of two patients, column (b) the normalized images. As all images are shown using the same windows settings of center 350 and width 250, one can see that the differences of the intensity values of the brain are successfully reduced by the presented approach.

4.4 Image Registration

Parts of the following section are based on the two publications [Ritt 10, Kief 11].
Image registration is a crucial part of our workflow. The medical images from different modalities, acquired at different points in time usually do not have identical intrinsic orientation. This means that a voxel with certain coordinates does not necessarily contain the same anatomical structure of a voxel with the same coordinates in another dataset. The coordinate transformation from the voxel in one dataset to the corresponding voxel in another dataset is in our case obtained by application of a fully automated rigid registration approach, introduced by Hahn et al. [Hahn 10]. Some algorithmic details are outlined in the Section 4.4.1. As studies about the accuracy of retrospective registration between PET and MRI of the brain were still lacking at the point of above mentioned publication, we present an evaluation for this topic in Section 4.4.2.

4.4.1 Implementation and Algorithmic Details

The registration of the images was carried out using an InSpace plugin. The dataset which was used to define the VOIs (in our case the T_2-weighted MRI) was considered as reference volume. Subsequently, all other datasets were registered to this dataset, using callbacks to a plugin which is based on the algorithms of Hahn et al. [Hahn 10]. The registration itself is realized as a rigid registration and therefore has 6 degrees of freedom (three translations, three rotations). It is based on the pixel intensities of the images and the normalized mutual information is used in its objective function. The objective function is iteratively optimized by a hill climbing algorithm. The algorithm features a multi-resolution approach in order to speed up computational time. For further details refer to [Hahn 10].

4.4.2 Validation of Registration Accuracy

To ensure that the VOIs are transferred in an anatomical correct way, we evaluated the registration accuracy for independently acquired cranial FET-PET and T_1-weighted MRI datasets of 49 patients. The mean time between the imaging was 80 days, ranging from 0 to 366 days. The images were registered using the InSpace platform and the previously described method (Section 4.4.1). The distances between the centers of gravity of the manually delineated skull contours of PET and MRI images were measured on multi-planar reconstructed views in axial, coronal and sagittal views (see Figure 4.8 for an example). The accuracy and reproducibility of the measurement method itself was evaluated by performing the procedure on 20 identical MRI and 20 identical PET images, which inherently were perfectly aligned. The isotropic voxel sizes of the PET images was ≈ 2 mm and ≈ 1 mm for the MRI images.

The accuracy of the measurement method itself was 1.21 ± 0.61 mm (95% confidence interval [1.01 mm; 1.40 mm] for the root-mean-squared error and 0.59 ± 0.51 mm for the error in a single dimension. The mean distance for the centers of gravity of the PET-MR patient data was 2.31 ± 1.08 mm (95% confidence interval [2.00 mm; 2.63 mm] with mean distances of 0.66 ± 0.56 mm in right-left (R→L)

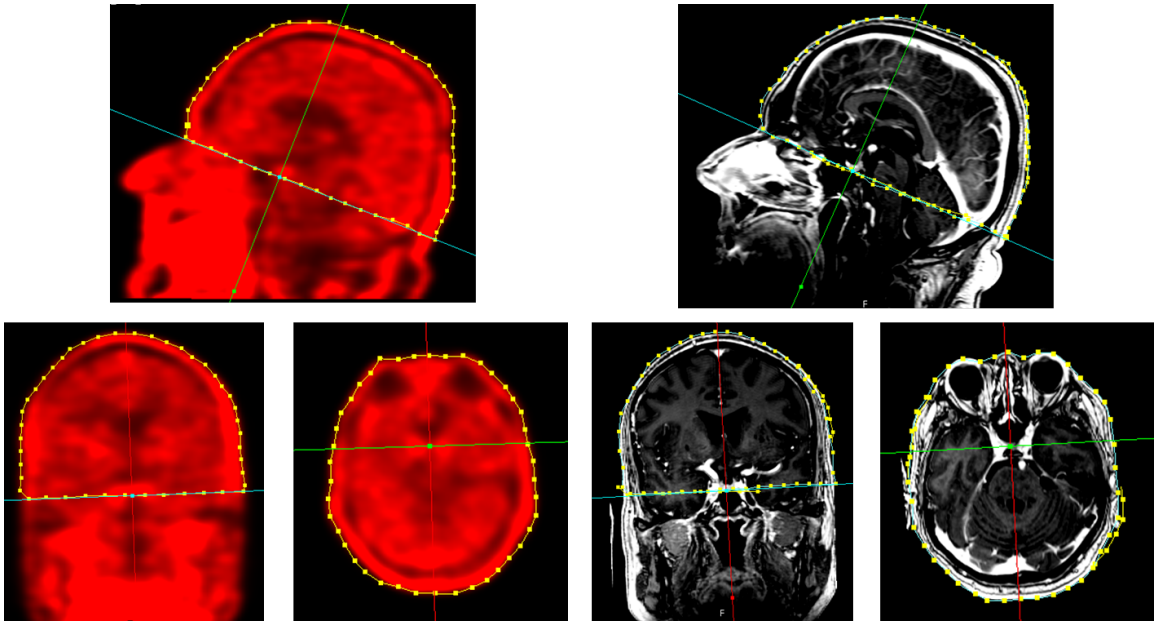


Figure 4.8: MPR views of PET (red colors) and MR (gray colors) images of the skull with delineated contours.

direction, of 1.36 ± 0.99 mm in anterior-posterior (A→P) direction, and of 1.36 ± 1.06 mm in foot-head (F→H) direction (Figure 4.9). The distances in R→L direction were found to be significantly smaller ($p < 0.05$) than for the other two directions using a Friedman test [Shes 03, p. 845].

We concluded, with the confidence intervals in mind, that the measured mean distances for the centers of gravity are significantly higher than the accuracy of the measurement method itself. This is presumably caused by a limited accuracy of the rigid registration. Possible causes are differences in the position of the brain relative to facial and cervical soft-tissues in the independently acquired datasets. This assumption could explain the significantly higher misalignment in A→P and F→H directions as well. An example image is given in Figure 4.10. It can be seen that the automatic registration achieved better accuracy in the axial slice than in sagittal slices (green circles). Still, we found that a mean misalignment in the range of the size of one voxel is sufficient and will not influence the feature extraction and classification results in a negative manner.

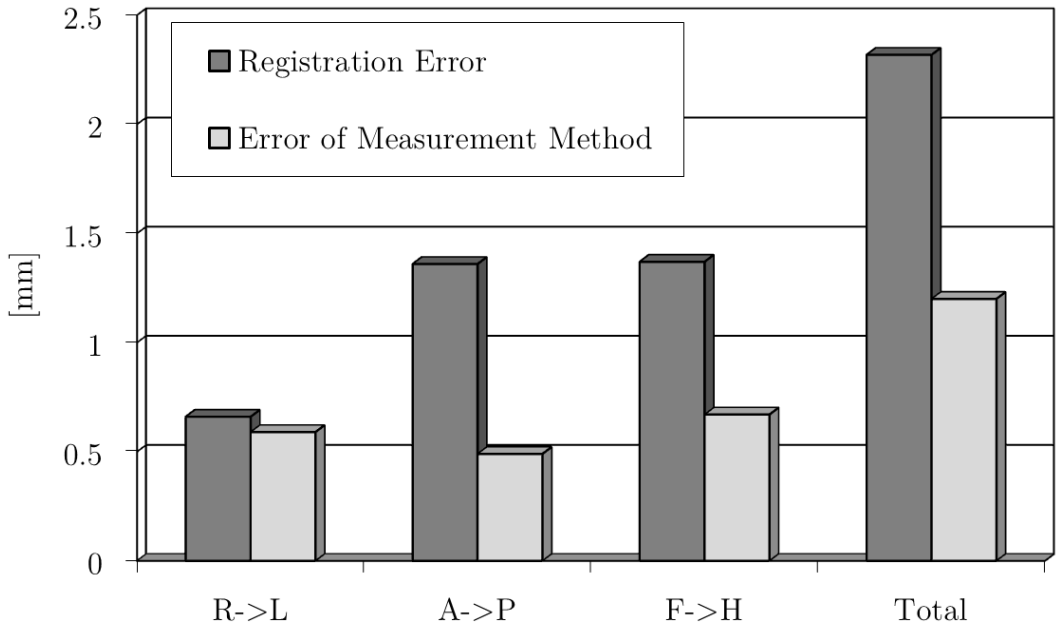


Figure 4.9: Misalignment of separately acquired and retrospectively registered PET and MR images of skulls. The misalignment is separated by spatial direction. A→P: Anterior-Posterior; F→H: Foot-Head; R→L: Right-Left.

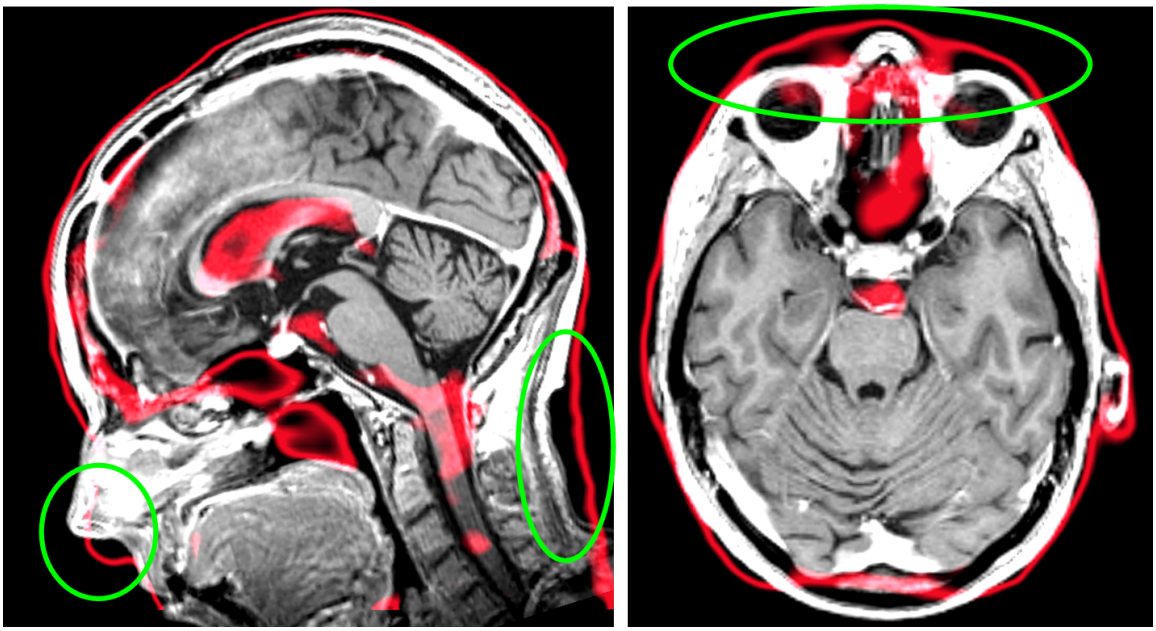


Figure 4.10: Misalignment between PET (red contours) and MR (gray colors) images that were retrospectively registered. The misalignment probably results from non-rigid motion in the cervical region.

4.5 PET Motion Correction

As our PET images were acquired dynamically (see Section 3.2.1), multiple 3-D datasets exist, each covering different time periods of the complete acquisition. These datasets are called frames. Due to the lengthy (up to 50 minutes) acquisition and the fact that for PET, in opposite to e. g. radiotherapy, the heads of the patients are not completely fixed, the occurrence of patient motion is probable. In fact, patient motion was noticeable in $\approx 90\%$ of our datasets (for an example, see Figure 4.11. In

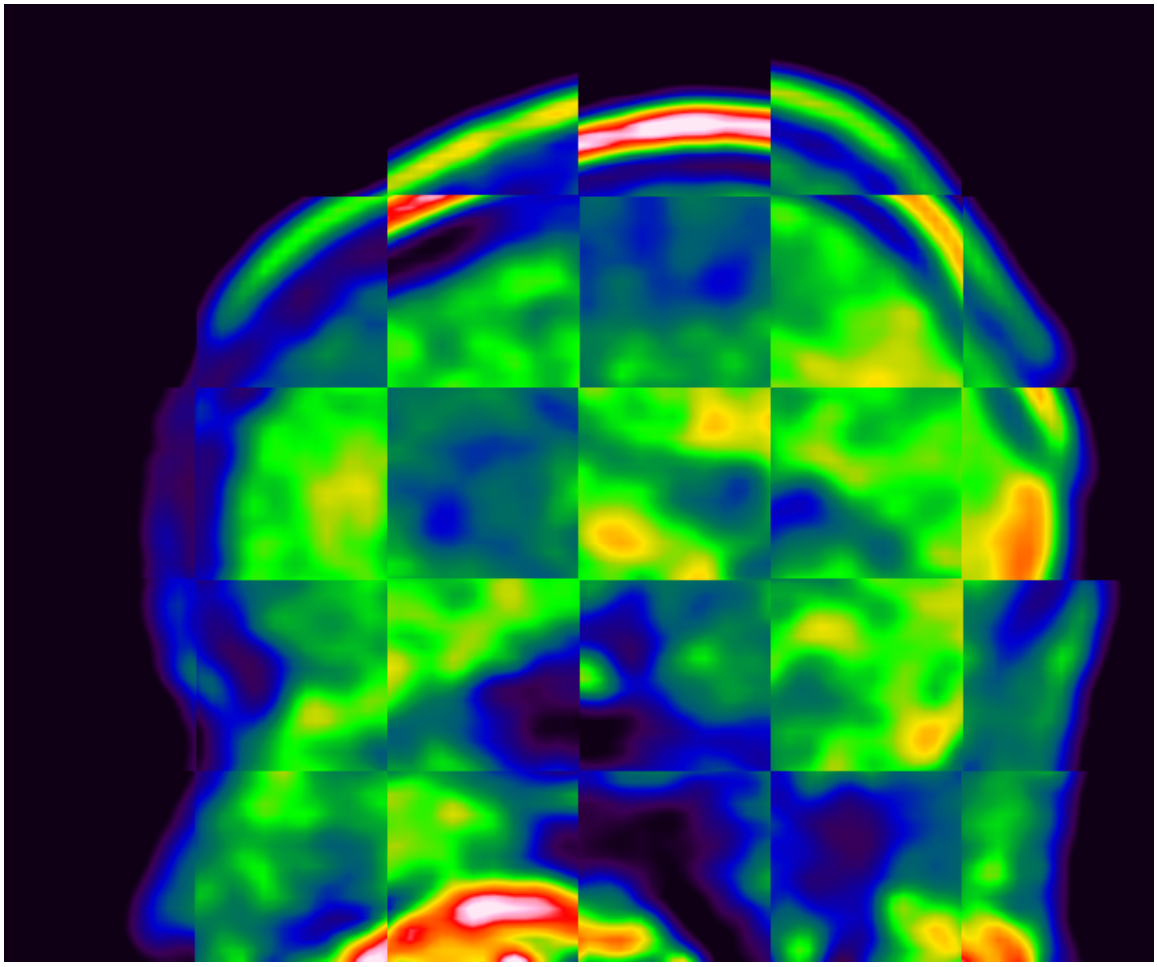


Figure 4.11: Checkerboard of 8th and 12th frame of a dynamic PET dataset of one patient. The noticeable misalignment is the result of motion that occurred in between the respective frames. Please note that the difference in intensities is partially due to the kinetics of the radioactive tracer and not due the patient motion.

order to partially compensate the motion, we applied the previously described and validated registration technique (Section 4.4). The reconstructed 3-D PET datasets were registered as a preprocessing step for our feature extraction. As before, the registration was of rigid manner and relied on the mutual information as a measure for goodness of fit. In a first step we evaluated two different strategies for the motion correction.
The first approach (Figure 4.12) employed a reference dataset which was kept con-

stant: In our setting, the last frame of the N frames was picked as the reference frame. All other $N-1$ frames were then registered to this reference frame.

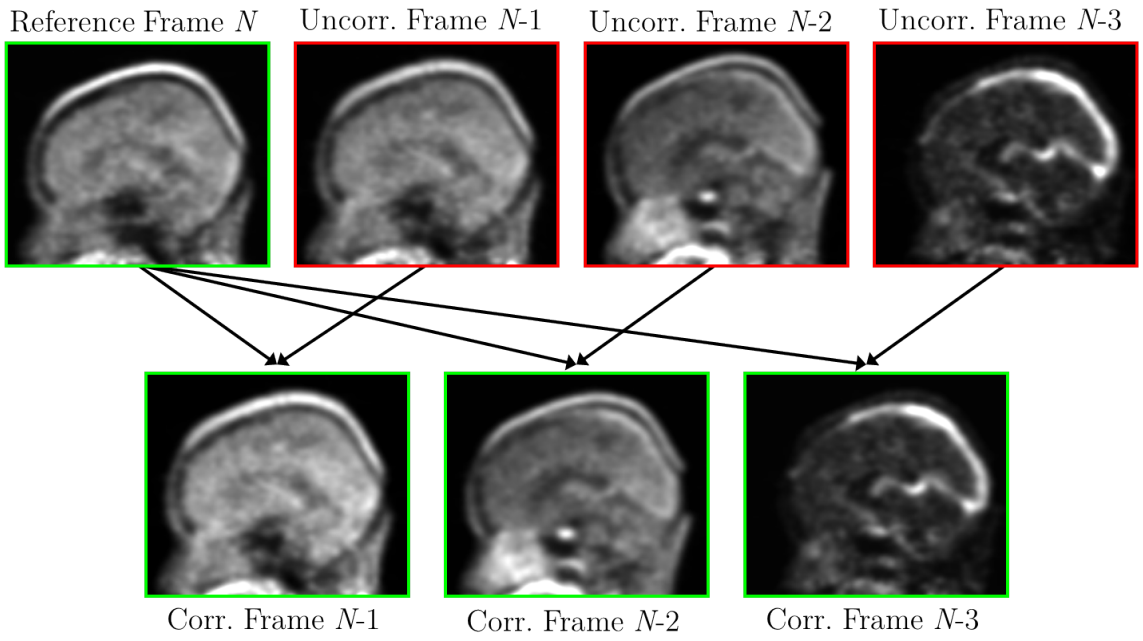


Figure 4.12: Illustration of motion correction incorporating a fixed reference frame.

The second approach employed a recursive technique. Its schematic can be found in Figure 4.13. First, frame N was picked as reference and frame $(N-1)$ was registered to it. By resampling with the registration matrix, a new corrected frame $(N-1)$ is calculated. The frame $(N-1)$ was the new reference to which the frame $(N-2)$ was then registered. This was repeated until all frames were corrected.

The recursive method provided better visual results (see Figure 4.14). This might be motivated by the fact that the intensity distribution changed over the acquired time period, even without any patient motion. Thus neighboring frames were more similar to each other than other arbitrary combinations like in the approach with fixed reference frame.

Therefore, we chose the recursive reference method for all of the further motion corrections. An example of the achieved correction is given in Figure 4.15, where one finds a significantly better match of 8th and 12th frame of the dynamic PET when compared to Figure 4.11.

For evaluating the efficacy of our motion correction method, we chose a random subset of 27 patients out of our patient collective. In a blinded experiment, one experienced nuclear medicine reader was presented with either an uncorrected or a corrected dynamic PET image. The reader rated the amount of patient motion in the image on a subjective scale with grades ranging from 0 to 3. Grade 0 implied no visible motion, grade 3 heavy motion. Table A.9 contains the detailed result of this analysis. A Wilcoxon signed-rank test (two-sided) [Shes 03, p. 189] revealed significant differences between the grading of corrected and uncorrected images ($p < 0.05$). The corrected images received lower ranks, thus the rated amount of motion is lower in those images. In another experiment, the image pairs were analyzed retrospectively

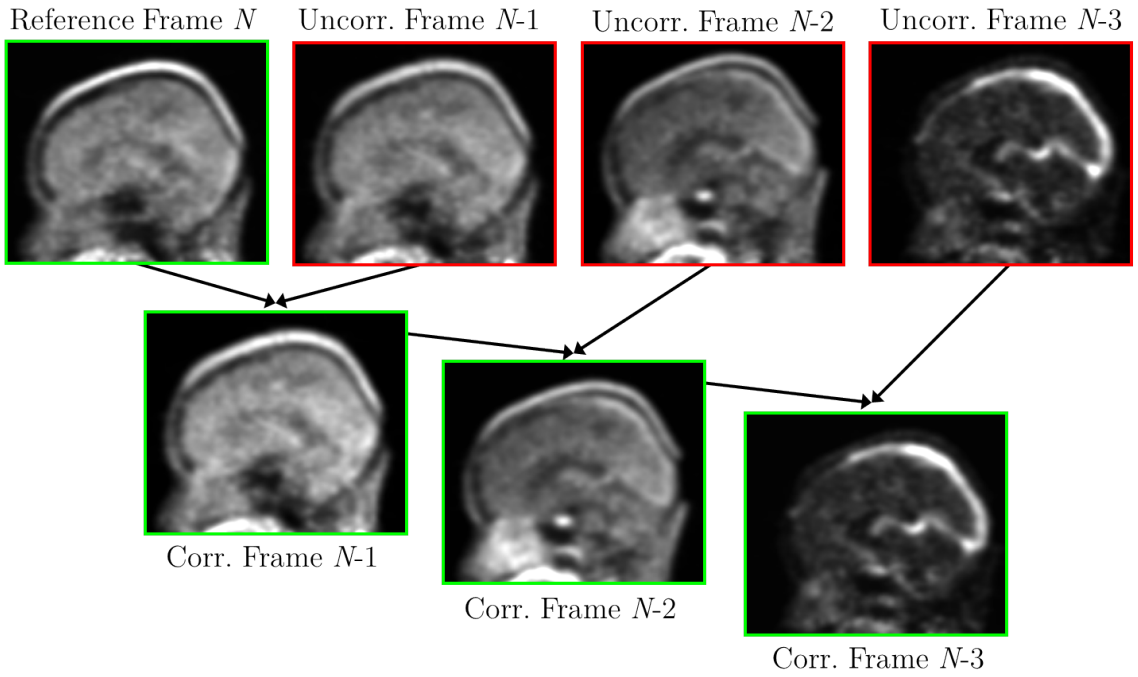
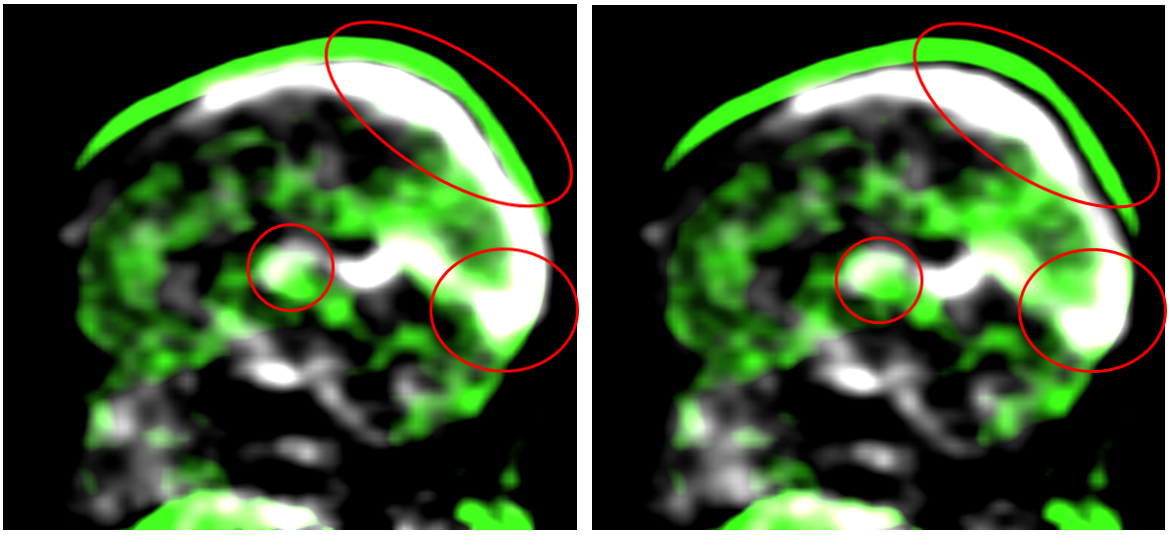


Figure 4.13: Illustration of PET motion correction using a recursive reference frame. The reference frame changes after each single registration.

in side-by-side view in order to determine, if a) the motion correction reduces the amount of motion, b) does not change the amount of motion or c) even introduces more motion. In 59% (16 cases), the amount of motion was reduced; in 22% (6 cases) the amount of motion stayed the same; however, in 19% (5 cases) the motion correction introduced artifacts that increased the subjective amount of motion (Figure 4.16).
The influence of the motion correction on the accuracy of the automatic classification is described in Section 6.4.



(a) Fixed reference frame (b) Recursive reference frame

Figure 4.14: Side by side comparison of fused images (first and last frame) of motion correction with (a) fixed reference frame and (b) recursive reference frame for the same patient. White depicts the first frame, green the last frame. The amount of matching is better for the recursive method in the areas indicated by the small red ellipses. In the area indicated by the big red ellipse, two different structures can be seen: In white, the cranial sinus, a venous blood vessel is the prevailing structure. In green, the scalp on top of the skull is the most prominent structure. These two structures are separated among others by the skull, the meninges (membrane of the brain), and liquor. The method with the fixed reference frame aligns the two structures too closely together, the recursive method yields results which are anatomically correct.

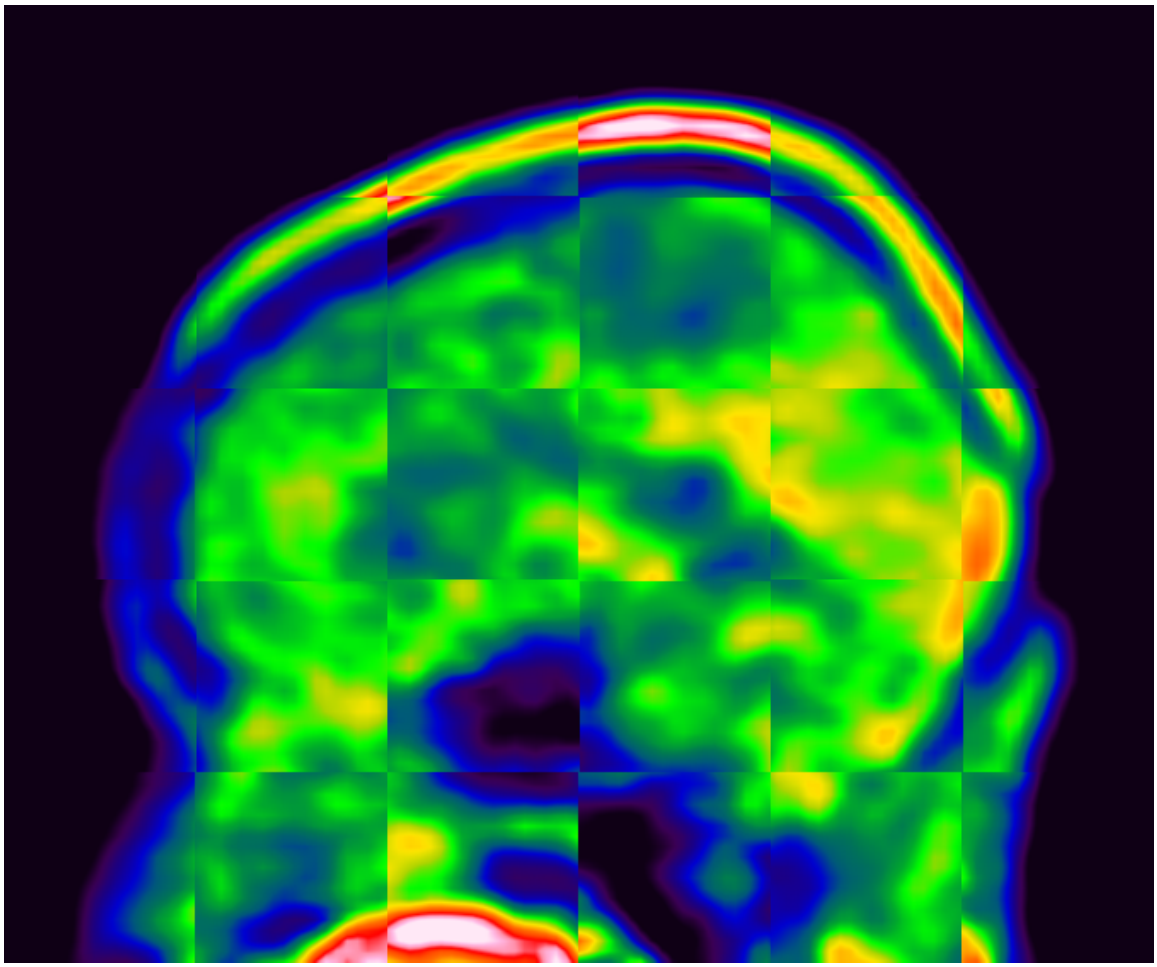


Figure 4.15: Checkerboard of 8th and 12th frame of a dynamic PET dataset after registration. When compared to the uncorrected Figure 4.11, the amount of misalignment is significantly reduced.

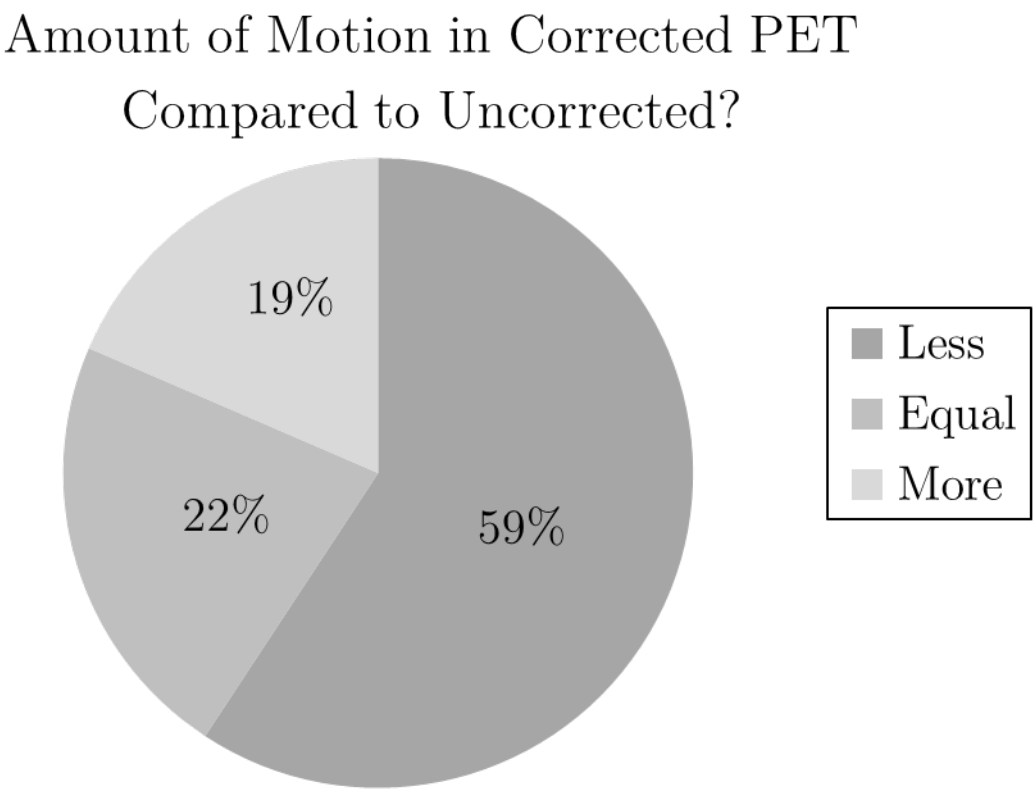


Figure 4.16: Pie chart for the results of the PET motion correction. We found that motion correction significantly reduces the amount of motion in the PET images. However, in some cases, weak "wiggling" is introduced by the correction and the amount of perceived motion is slightly increased.

4.6 Extracted Features

Depending on the modalities included in the classification task, the number of extracted features varied. The nature of the features can be described as statistical, contextual or textural. Statistical means in this context that the feature is calculated on the basis of the intensity value distribution of a single VOI. Contextual means that the feature is calculated by combining features from more than one VOI. In our case this always refers to a combination of the tumor VOI and the reference VOI of a patient dataset. In our implementation, based on the work of Haralick et al. [Hara 73], textural features incorporate information about the co-occurrence of intensity values in a VOI.
In general, the features were extracted from different medical datasets. We distinguished static and dynamic datasets. The static datasets were: T_2-weighted MRI, T_1-weighted MRI, diffusion-weighted (ADC) MRI, and X-ray CT. PET was the only dynamic modality in our experimental setup. Nevertheless, the feature extraction framework can handle any dynamic DICOM data, regardless of its origin. For a detailed description of the datasets see Chapter 3. Tables A.3 and A.4, which provide an overview on all features, are found in the appendix.
All features were chosen to represent the key metrics for tumor grading. These key metrics were introduced in 2.4.4. Those are:

- Tumor heterogeneity
- Formation of a cyst or necrosis
- Hemorrhage or bleeding
- Crossing of the mid line
- Extent and degree of an edema
- Degree and heterogeneity of the contrast enhancement
- Shift of other regions of the brain (called mass effect)
- Extent, degree and kinetic of amino acid uptake
- Extent and degree of calcification

In the following, we reference for all features which metrics they cover.

4.6.1 Statistical Features

The intensity values of the N voxels in a volume of interest (VOI) are $x_i \in \mathbb{R}$.

Mean Intensity Value
The mean voxel value μ of the tumor VOI was calculated by the arithmetic mean. A lower arithmetic mean in T_1-MRI indicates the *formation of a cyst or necrosis* or

hemorrhage or bleeding. A high mean in ADC-MRI indicates an *edema*.

Minimum Intensity Value
The minimum voxel value $x_{\min}$ of the VOI. Again, a low minimum intensity value for T_1-MRI represents the *formation of a cyst or necrosis*.

Maximum Intensity Value
The maximum voxel value $x_{\max}$ of the VOI. A high maximum intensity value in contrasted T_1-MRI hints to a distinct *contrast enhancement*, which is an important indicator of tumor grade. In CT, a high maximum intensity values is very likely a *calcification*.

95%-Quantile of Intensity
First, the probability density function (PDF) $\rho(x) \in \mathbb{R}$ of the intensity distribution was estimated by randomly choosing a number M of voxels (10% of the total voxel number N in the VOI, in order to reduce computational requirements) and calculating 4.2.

$$\rho(x) = \frac{1}{\sigma M\sqrt{2\pi}} \sum_i \exp\left\{-\frac{(x-x_i)^2}{2\sigma^2}\right\} \tag{4.2}$$

with the bandwidth $\sigma = (x_{\max} - x_{\min})M^{-\frac{1}{2}}$. The 95%-Quantile $x_{95\%}$ of the PDF subsequently was calculated solving Equation 4.3 for x.

$$0.95 = \int_{-\infty}^{x_{95\%}} \rho(x)\,\mathrm{d}x \tag{4.3}$$

In our experiment, bandwidth and mean intensity in the VOI were always in a way that $\int_{-\infty}^{x_{95\%}} \approx \int_{0}^{x_{95\%}}$. The integration was numerically carried out using the trapezoidal rule.

$$\int_a^b f(x)\,\mathrm{d}x \approx 0.5 \cdot (b-a)(f(a)+f(b)) \tag{4.4}$$

The 95%-quantile is an indicator for the maximum of the intensities in the VOI and is not as prone to noise as the maximum intensity value. For this, it shares the same motivation as the maximum: a high value in T_1-MRI indicates a *contrast enhancement*.

Position of Peak in Intensity
The position of the peak x_{peak} of the tumor PDF was:

$$x_{\text{peak}} = \operatorname*{argmax}_x \rho(x) \tag{4.5}$$

The peak position of the PDF represents the gray-level values of the most predominant class in the VOI. It indicates a variety of metrics. For contrasted T_1-MRI the *extent of contrast enhancement* and the *degree of edema*, for ADC-MRI and CT the *degree of edema*.

Height of Peak in Intensity
Following Equation 4.5, the height ρ_{peak} of the peak in the tumor PDF was:

$$\rho_{\text{peak}} = \rho(x_{\text{peak}}) \tag{4.6}$$

Slope and Intercept of Mean Intensity Value Curve
For dynamic datasets with J frames, indexed by $j = 1, \ldots, J$, and time points t_j of the frame, the slope b of the mean intensity value was calculated by linear regression (based on least squares) of the mean value μ_j of those frames for the tumor VOI . The arithmetic mean values of μ_j and t_j over the J frames are denoted by $\bar{\mu}$ and $\bar{t}$.

$$b = \frac{\sum_j (\mu_j - \bar{\mu})(t_j - \bar{t})}{\sum_j (\mu_j - \bar{\mu})^2} \tag{4.7}$$

The intercept a was calculated by

$$a = \bar{t} - b\bar{\mu} \tag{4.8}$$

The slope and the intercept of the mean intensities for the dynamic PET represent the average *extent, degree and kinetic of amino acid uptake*. A low amino acid uptake points toward a lower tumor grade, whereas a negative slope is an indicator of a higher tumor grade.

Slope and Intercept of Maximum Intensity Value Curve
Similarly to the calculation of *Slope of Mean Intensity Value Curve*, the slope for the maximum intensity value curve $b_{\max}$ was calculated on the basis of the maximum value of all voxels $x_{\max,j}$ in the tumor VOI of the j-th time frame. $\bar{x}_{\max}$ denotes the arithmetic mean of the $x_{\max,j}$ over J frames.

$$b_{\max} = \frac{\sum_j (x_{\max,j} - \bar{x}_{\max})(t_j - \bar{t}_j)}{\sum_j (x_{\max,j} - \bar{x}_{\max})^2} \tag{4.9}$$

And the intercept:

$$a_{\max} = \bar{t} - b_{\max}\bar{x}_{\max} \tag{4.10}$$

It was shown for PET, that in some cases the kinetics of the extremal values can offer better tumor grading [Popp 07]. In general, the same aspects apply as for the kinetics of the mean intensity.

Slope and Intercept of Minimum Intensity Value Curve
Again, similarly to the features *Slope of Mean Intensity Value Curve*, the slope for the minimum intensity value curve $b_{\min}$ was calculated on the basis of the minimum value of all voxels $x_{\min,j}$ in the VOI of the j-th time frame. Again, $\bar{x}_{\min}$ denotes the arithmetic mean of the $x_{\min,j}$ over J frames.

$$b_{\min} = \frac{\sum_j (x_{\min,j} - \bar{x}_{\min})(t_j - \bar{t}_j)}{\sum_j (x_{\min,j} - \bar{x}_{\min})^2} \tag{4.11}$$

And the intercept:

$$a_{\min} = \bar{t} - b_{\min}\bar{x}_{\min} \tag{4.12}$$

4.6.2 Contextual Features

Proportion of Iso-intense Voxels
Iso-intensity was defined to lie within one standard deviations σ around the peak intensity value x_{peak} of the reference VOI: all values below were hypo-intense, all values above were hyper-intense. The proportion of iso-intense voxels in the tumor VOI is denoted by P_{iso}. The total number of voxels in the tumor VOI is N, the number of iso-intense voxels is N_{iso}. $\mathcal{X}$ is a the set of pixel intensities that fulfills the specified condition.

$$N_{\text{iso}} = |\mathcal{X}| \qquad \mathcal{X} = \{x_i | x_{\text{peak}} - \sigma \leq x_i < x_{\text{peak}} + \sigma\} \tag{4.13}$$

$$P_{\text{iso}} = \frac{N_{\text{iso}}}{N} \tag{4.14}$$

The proportion of voxels in the tumor VOI reflects the *tumor heterogeneity*. A high proportion of iso-intense voxels indicates low heterogeneity, which could point towards a tumor of lower grade.

Proportion of Hypo-intense Voxels
Similarly to the feature *Proportion of Iso-intense Voxels*, the proportion of hypo-intense voxels P_{hypo} in the tumor VOI is

$$N_{\text{hypo}} = |\mathcal{X}| \qquad \mathcal{X} = \{x_i | x_i < x_{\text{peak}} - \sigma\} \tag{4.15}$$

$$P_{\text{hypo}} = \frac{N_{\text{hypo}}}{N} \tag{4.16}$$

with the number of hypo-intense voxels in the primary VOI being N_{hypo}.
The proportion of hypo-intense voxels reflects the extent of *edema* and *cyst or necrosis* as seen in T_1-MRI.

Proportion of Hyper-intense Voxels
Similarly to the calculation of *Proportion of Iso-intense Voxels*, the proportion of hyper-intense voxels P_{hyper} in the tumor VOI is

$$N_{\text{hyper}} = |\mathcal{X}| \qquad \mathcal{X} = \{x_i | x_i \geq x_{\text{peak}} + \sigma\} \tag{4.17}$$

$$P_{\text{hyper}} = \frac{N_{\text{hyper}}}{N} \tag{4.18}$$

with the number of hypo-intense voxels in the primary VOI being N_{hyper}.
The proportion of these voxels indicate the extent of contrast enhancement for T_1-MRI, the extent of an edema for ADC-MRI, and the degree of calcification for CT.

Kullback-Leibler (KL) divergence
The Kullback-Leibler divergence $D(\rho_{\text{tumor}} || \rho_{\text{reference}}) \in \mathbb{R}_0^+$ is a measure for the difference of two probability density functions. We computed the KL divergence for the PDFs of the reference ($\rho_{\text{reference}}$) and the tumor (ρ_{tumor}) VOI. For the classification,

we used the symmetric variant of the KL divergence D_{eff}. This variant is based on the left and right KL divergences and calculated by:

$$D(\rho_{\text{tumor}}||\rho_{\text{reference}}) = \int_{-\infty}^{\infty} \rho_{\text{tumor}}(x) \ln \frac{\rho_{\text{tumor}}(x)}{\rho_{\text{reference}}(x)} \, \mathrm{d}x \tag{4.19}$$

$$D(\rho_{\text{reference}}||\rho_{\text{tumor}}) = \int_{-\infty}^{\infty} \rho_{\text{reference}}(x) \ln \frac{\rho_{\text{reference}}(x)}{\rho_{\text{tumor}}(x)} \, \mathrm{d}x \tag{4.20}$$

$$D_{\text{eff}}(\rho_{\text{reference}}||\rho_{\text{tumor}}) = D(\rho_{\text{reference}}||\rho_{\text{tumor}}) + D(\rho_{\text{tumor}}||\rho_{\text{reference}}) \tag{4.21}$$

The idea for using the KL divergence is, that the difference between the tumor- and the reference PDF is related to the tumor grade. If both PDFs represent healthy tissue, D will be close to zero. With increasing D, the likelihood for a high-grade tumor increases as well.

Quotient Mean Tumor to Mean Reference VOI
The feature Q_{mean} was calculated by:

$$Q_{\text{mean}} = \frac{\mu_{\text{tumor}}}{\mu_{\text{reference}}} \tag{4.22}$$

where μ_{tumor} and $\mu_{\text{reference}}$ represent the mean intensity value in the tumor and in the reference VOI.
The motivation of quotients of parameters of tumor and reference VOI is to gain a certain stability against inter-patient and inter-dataset variation. In general, the same tumor metrics as for the mean intensity value are addressed. A $Q_{\text{mean}} < 1$ in T_1-MRI is a strong indicator of *formation of a cyst or necrosis* or *hemorrhage or bleeding*. A $Q_{\text{mean}} > 1$ in ADC-MRI indicates an *edema*.

Quotient Maximum Tumor to Mean Reference VOI
This feature Q_{max} was calculated by:

$$Q_{\text{max}} = \frac{x_{\text{max,tumor}}}{\mu_{\text{reference}}} \tag{4.23}$$

where $x_{\text{max,tumor}}$ is the maximum intensity value in the tumor VOI and $\mu_{\text{reference}}$ is the mean intensity value in the reference VOI.
Analogously as for the quotient of the means, here the metrics for the maximum intensity values are addressed. A $Q_{\text{max}} \gg 1$ in contrasted T_1-MRI shows a textitdisruption of the blood-brain-barrier. A *calcification* in the tumor leads to $Q_{\text{max}} \gg 1$ in CT.

Quotient Minimum Tumor to Mean Reference VOI
This feature Q_{min} was calculated by:

$$Q_{\text{min}} = \frac{x_{\text{min,tumor}}}{\mu_{\text{reference}}} \tag{4.24}$$

where $x_{\text{min,tumor}}$ is the minimum intensity value in the tumor VOI and $\mu_{\text{reference}}$ is the mean intensity value in the reference VOI.

Analogously as for the previous two features, when $Q_{\min} < 1$ in T_1-MRI, it is very likely that the tumor leaded to the *formation of a cyst or necrosis*.

Quotient Maximum Tumor to Mean Reference VOI (Dynamic)
In the case of dynamic modalities, the quotients were defined in a different way: $Q_{\max}$ was calculated by finding the maximum value $x_{\max}$ of all N voxels in the tumor VOI of all time frames J and dividing by the mean $\bar{\mu}$ of the reference VOI over all time frames. Namely

$$x_{\max} = \max_{i,j} \{x_{i,j}\} \tag{4.25}$$

$$\bar{\mu} = \frac{1}{NJ} \sum_j \sum_i x_{i,j} \tag{4.26}$$

$$Q_{\max} = \frac{x_{\max}}{\bar{\mu}} \tag{4.27}$$

This feature was only calculated for dynamic PET. Here the idea is that with increasing tumor grade $Q_{\max}$ increases as well.

Quotient Mean Tumor to Mean Reference VOI (Dynamic)
This feature is the quotient Q_{mean} of the mean over all frames of the reference VOI ($\bar{\mu}_{\text{reference}}$) and of the tumor VOI.

$$Q_{\text{mean}} = \frac{\bar{\mu}_{\text{primary}}}{\bar{\mu}_{\text{reference}}} \tag{4.28}$$

As for the previous feature, a $Q_{\max} > 1$ indicates a tumor. With increasing $Q_{\max}$, a higher grade tumor is likely.

Quotient Minimum Tumor to Mean Reference VOI (Dynamic)
$Q_{\min}$ was calculated by finding the minimum value $x_{\min}$ of all N voxels in the tumor VOI of all time frames J and dividing by the mean $\bar{\mu}$ of the reference VOI over all time frames.

$$x_{\min} = \min_{i,j} \{x_{i,j}\} \tag{4.29}$$

$$\bar{\mu} = \frac{1}{NJ} \sum_{i,j} x_{i,j} \tag{4.30}$$

$$Q_{\min} = \frac{x_{\min}}{\bar{\mu}} \tag{4.31}$$

4.6.3 Textural Features

Whereas most metrics of List 4.6 are covered by previously described features, especially the aspect of *tumor heterogeneity* is represented poorly so far. For this, we decided to include textural features in order to map this aspect in the set of features. For the textural features, so called *graytone spatial-dependence probability-distribution matrices* [Hara 73], also known as gray-level co-occurrence (GLCM) matrices are calculated. These matrices contain the probabilities that a VOI's voxel with intensity

value l has a neighboring voxel in a specific direction $\boldsymbol{r}$ with an intensity value m (see Figure 4.17). In 3-D, one voxel has 26 neighbors, as long as it is not at the border of the VOI. Consequently, the probabilities are calculated for 26 directions and one gets co-occurrence matrices $\boldsymbol{P_r}$ for every direction. Since $\boldsymbol{P_r} = \boldsymbol{P}^{\top}_{-\boldsymbol{r}}$ (symmetry for opposite directions), the number of necessary matrices can be reduced to 13. In order to achieve a certain directional invariance, out of the 13 co-occurrence matrices, one mean matrix $\boldsymbol{P}$ with elements $p_{l,m}$ is calculated, where the p are the aforementioned probabilities. Since we used 256 intensity levels, $\boldsymbol{P} \in \mathbb{R}^{256\times256}$. Voxels which are at the border of the VOI and do not have the neighbors in the specified directions are omitted in the calculations, but in general represent only a small amount of the total number of voxels as we had rather large VOIs. We calculated 6 different textural features according to Connors et al. [Conn 84]. The elements of the gray-level co-occurrence matrices $p_{l,m}$ were normalized, so that $\sum_{l,m} p_{l,m} = 1$. Despite its long history, the proposed textural features are still successfully applied for automatic classification of medical MRI [Zulp 12].

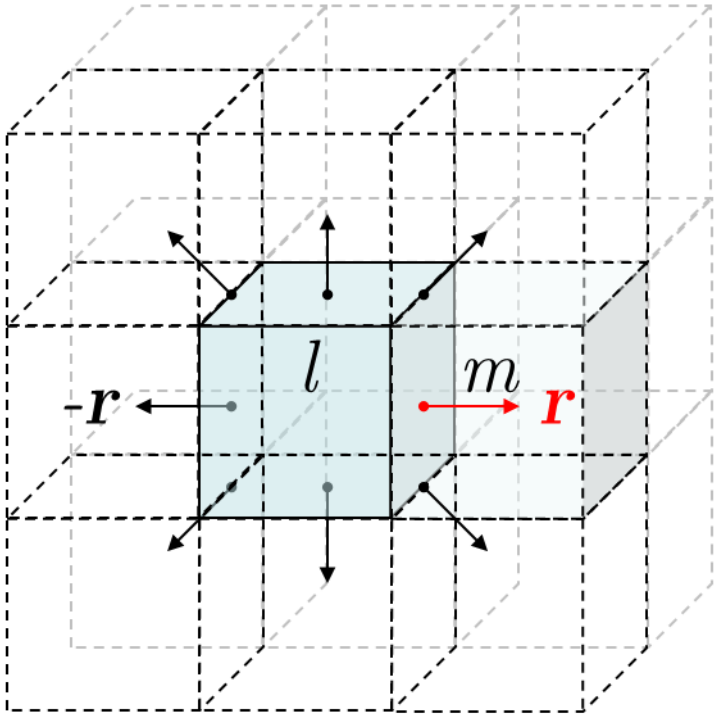


Figure 4.17: Schematic diagram of neighbors of a voxel with intensity value l. The GLCM matrices contain the probabilities that a VOI's voxel has a neighboring voxel in a specific direction $\boldsymbol{r}$ with an intensity value m. (In 3-D every voxel has 26 neighbors, consequently 26 such matrices exist. Due to the symmetry of the problem, it is sufficient to calculate the matrices for 13 directions, omitting opposite directions. For clearness, not all directions and neighbors are shown.)

Energy

$$E = \sum_{l,m} (p_{l,m})^2 \tag{4.32}$$

The energy is a measure for the homogeneity (similar gray-levels) and for textural uniformity (neighboring voxels having similar gray-level pairs) of the VOI. In the extremal case, when the whole VOI features the same image intensity or when all neighbors have the same gray-level pair, $p_{l,m}$ has only one non-zero entry. Due to the normalization of $p_{l,m}$, this entry is 1 and the energy consequently is $E = 1$ as well. Thus, the closer E approaches 1, the higher is the homogeneity and textural uniformity.

Entropy

$$H = -\sum_{l,m} p_{l,m} \log(p_{l,m}) \tag{4.33}$$

The entropy is a measure for the orderliness of the VOI. If the image has a low textural uniformity (white noise image, neighboring voxels having quasi-random gray-level pairs), many entries of $p_{l,m}$ will be close to 0 and the entropy H will approach high values. The entropy is inversely correlated to the energy.

Local Homogeneity

$$L = \sum_{l,m} \frac{1}{1 + (l-m)^2} \, p_{l,m} \tag{4.34}$$

The local homogeneity is especially sensitive to the homogeneity of the VOI, which means L will be small if strong edges (large differences in neighboring voxels) are present. On the opposite, L approaches its maximum if the difference between neighboring voxels is small (high near-diagonal entries in $p_{l,m}$). The local homogeneity is strongly inversely correlated to the inertia.

Inertia

$$I = \sum_{l,m} (l-m)^2 \, p_{l,m} \tag{4.35}$$

In inertia, the $p_{l,m}$ are weighted by the squared differences of their intensity values. Inertia is a measure for the contrast in the VOI. A VOI that features high differences in the gray-level of the neighboring voxels (meaning strong edges and high spatial image frequencies) leads to a high inertia. On the other side, if a lot of neighboring voxels have similar intensity values (soft edges, low spatial image frequencies), the $p_{l,m}$ will have only have high entries near its diagonal where the difference $(l - m)$ is small, thus the inertia becomes small.

Cluster Shade

$$A = \sum_{l,m} (l + m - \mu_l - \mu_m)^3 \, p_{l,m} \tag{4.36}$$

where $\mu_l = \sum_{l,m} l p_{l,m}$ and $\mu_m = \sum_{l,m} m p_{l,m}$.

Cluster Prominence

$$B = \sum_{l,m} (l + m - \mu_l - \mu_m)^4 \, p_{l,m} \tag{4.37}$$

with μ_l, μ_m as defined above.

4.7 Feature Normalization

The different imaging modalities have different image intensity levels, consequently the features in our experiment have differing numerical ranges. The numerical range of our features spans about ≈ 10 orders of magnitude, as seen in Figure 4.18. In order to ensure an equal weight of the features in the classification process, feature normalization is crucial. We applied two different feature normalization methods: linear scaling to a range and linear scaling to unit variance. We compared their influence for our classification task.
Not all classifiers are equally prone to a bias that is potentially introduced by the different scaling of the features. E. g. the nearest neighbor classifier using the Euclidean distance as measure is sensitive to scaling differences. Implicitly, more weight is assigned to the dimensions of the multidimensional feature vectors that dominate in terms of the magnitude of the features values. The nearest neighbor classifier with Euclidean distance is translation invariant but not scale invariant.
On the other hand, for adaptive boosting with decision stumps as weak classifier, the scaling of the feature vector dimensions is not of importance. The decision stumps classify on basis of the threshold of a univariate feature. For this, only the order of the instances is relevant. The decision stump is scaling and translation invariant.
Above examples are of course only valid if we can assume that all feature values can be uniquely coded using the implemented computer number format.

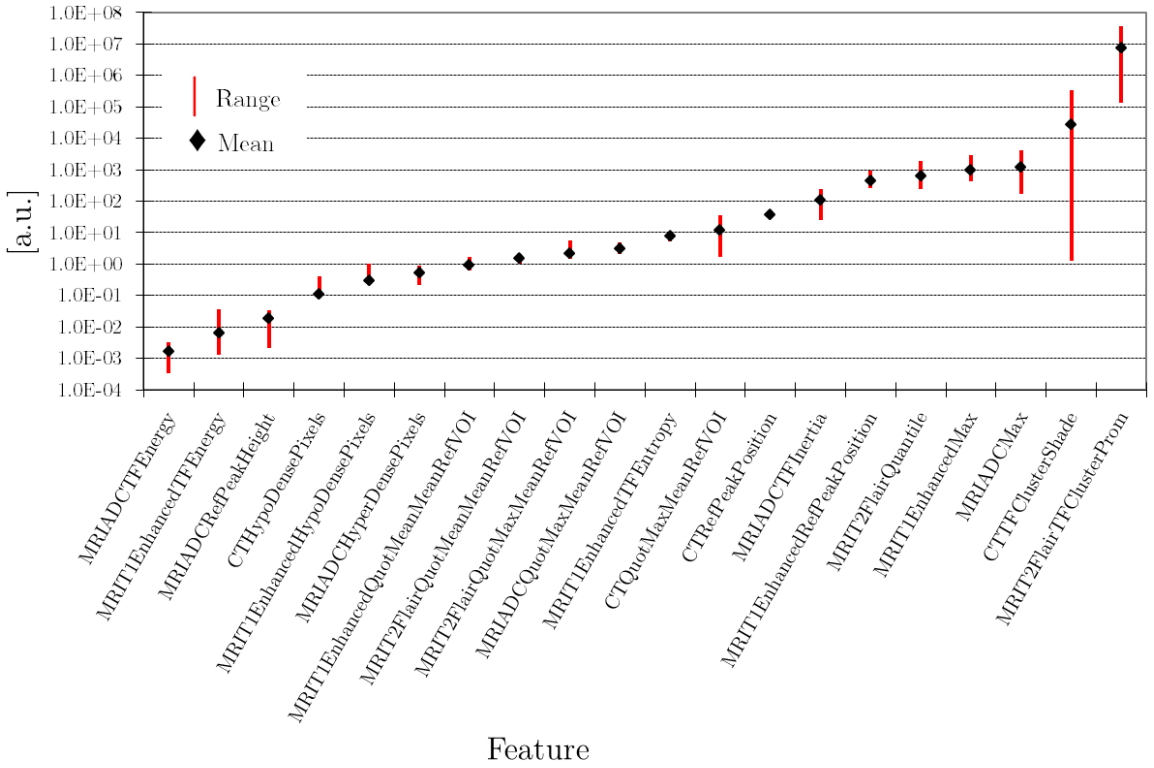


Figure 4.18: The mean values and numeric range of 20 selected features. The features have a large dynamic range which spans ≈ 10 orders of magnitude. The black marker indicates the mean value of each feature, the red bars indicate the range (from min to max).

4.7.1 Linear Scaling to Range (LSR)

One of the most frequently proposed methods for feature normalization is scaling the features to the interval $[-1;1]$. The training set $\mathcal{S}$ consists of N instances $\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\} \quad i = 1, \ldots, N$, with $\boldsymbol{x}_i \in \mathbb{R}^d$ and $y_i \in \mathbb{Z}$. $x_{i,j}$ denotes the j-th element of sample $\boldsymbol{x}_i$. By applying Equation (4.38) to the $x_{i,j}$ we obtained the normalized values $\tilde{x}_{i,j}$. $\max_j$ and $\min_j$ represent the maximum and minimum values of the j-th dimension in the feature set.

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \frac{1}{2}(\max_j + \min_j)}{\frac{1}{2}(\max_j - \min_j)} \qquad i = 1, \ldots, N,\ j = 1, \ldots, d \tag{4.38}$$

In the following, the term **normalized by LSR** always refers to the method mentioned above. This method reduces the numeric range over all dimensions of the feature vector. As we aimed to verify the numeric range of the feature vectors as seen by the classifier during classification of the unknown test samples, we used the leave-one-out cross-validation (LOO-CV) strategy. We divided the data into test- and training set, normalized the training set with appropriate choices of $\max_j$ and $\min_j$, normalized the test set using the same parameters, and stored the normalized test set. This procedure was repeated for all LOO-permutations. The resulting numeric range of all test data for some features can be seen in Figure 4.19. Naturally, due to the LOO-CV, this range exceeds the interval $[-1;1]$.

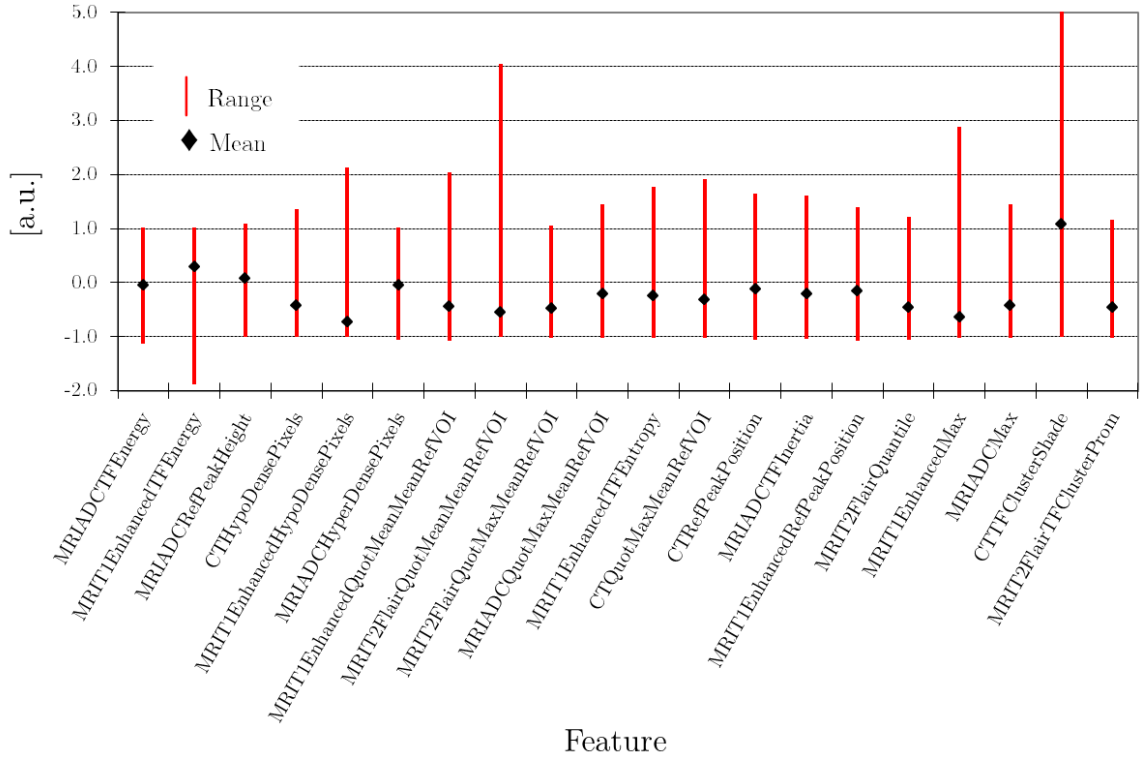


Figure 4.19: The mean values and the numeric range of 20 selected features after normalization by linear scaling to the range $[-1;1]$. The black marker indicates the mean value of each feature, the red bars indicate the range (from min to max).

4.7.2 Linear Scaling to Unit Variance (LSUV)

As an alternative to scaling to a specified range 4.7.1, scaling to unit variance can be used to ensure an equal weight of all features during the classification process. Scaling to unit variance does not rely on (potentially noisy) extremal values like, e. g. maxima, minima but on the mean μ_j and the standard deviation σ_j of j-th dimension. We chose to scale the feature values in such a way that for the transformed features $\tilde{\boldsymbol{x}}_i$, $\tilde{\mu}_j = 0$ and $\tilde{\sigma}_j^2 = 1$. This is achieved using Formula 4.39 on the values. As previously, $x_{i,j}$ denotes the j-th element of sample $\boldsymbol{x}_i$:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j} \qquad i = 1, \ldots, N,\ j = 1, \ldots, d \tag{4.39}$$

In the following, the term **normalized by LSUV** always refers to this method. Again, we find a similar range for all displayed features (Figure 4.20). These results were obtained analogously as for the numeric range of LSR as previously described. For this, $\tilde{\mu}_j$ and $\tilde{\sigma}_j$ differ slightly from their targeted values.
When we compare LSUV to LSR, we find that the former performs better in keeping the mean values close to zero and the latter prevents larger outliers.

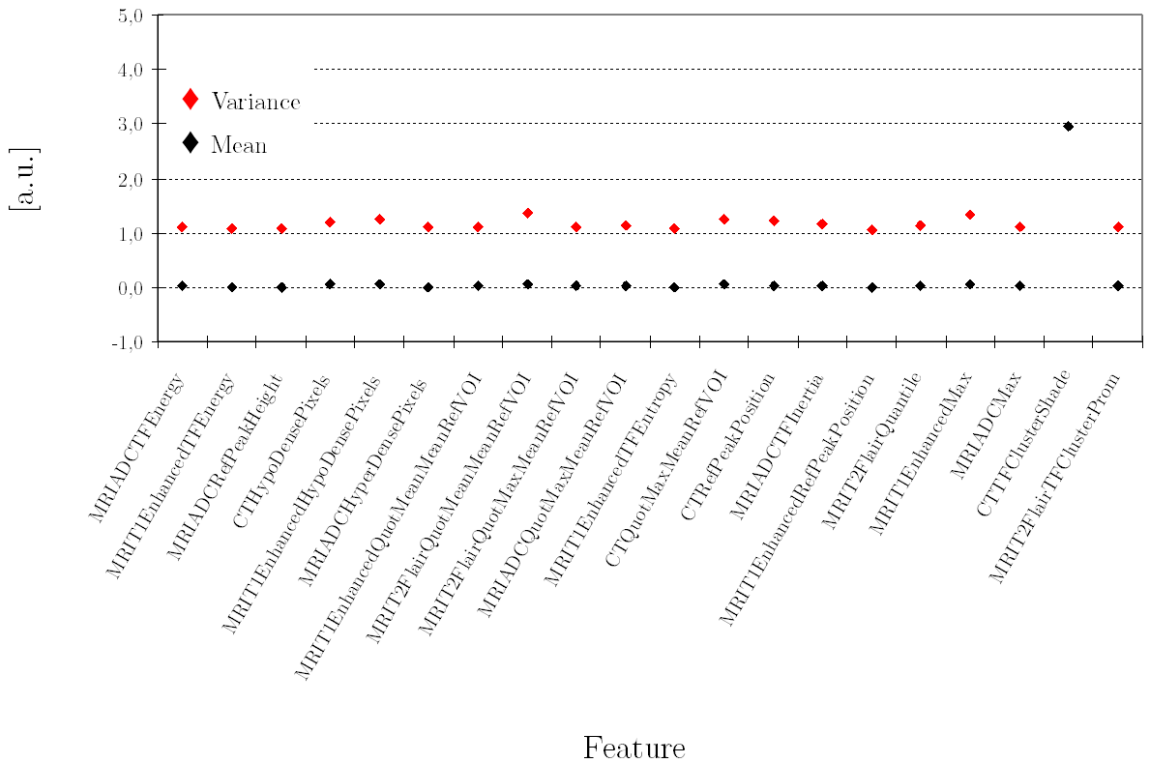


Figure 4.20: The mean and variance values of 20 selected features after normalization by linear scaling to unit variance and zero mean. The red marker indicates the mean value of each feature, the black markers the variance.

4.8 Feature Selection and Feature Transformation

4.8.1 Manual Feature Selection

As we aimed to examine how the specific types of features are suited for the automatic classification of cerebral gliomas, we chose a manual approach for selecting these sets over automatic approaches.
The features listed in Section 4.6 were one-dimensional. Each feature was extracted for different modalities, e. g. the *Mean Intensity Value* of the tumor VOI was extracted for T_1-MRI, T_2-MRI, ADC-MRI and CT. The univariate features are either directly used for classification or collected into feature sets, according to manually defined rules. E. g. the feature set with the index 56 consisted of the six texture features for each modality. In the case of patient subsets SUB44 and SUB32 (Chapter 3) with modalities T_1-MRI, T_2-MRI and FET-PET, the feature set 56 had the dimension 12, with 6 texture features from T_1-MRI and 6 texture features from T_2-MRI.
Table A.5 provides a list of the different feature sets for the classification of the subsets SUB44 and SUB32 and specifies which features are included in a certain feature set. We included the univariate features (features 1-50), as well as all features for a specific modality (feature sets 51-56). The feature sets 57-62 contained the transformed features: these were features consisting of a varying component number of the PCA transformation of all features. Details about the PCA transformation are found in Section 4.8.2.
Table A.6 provides the same information for the classification of the subsets SUB22 (ADC-MRI) and SUB14 (X-ray CT). In case of these modalities the number of different feature sets was 21.
In the classification stage, every feature set was classified. Consequently this resulted in 62 classification rates for SUB44 and SUB32 and 21 classification rates for SUB22 and SUB14, for each classifier.

4.8.2 Principal Component Transformation

We applied the principal component analysis (PCA) as feature transformation method in our experiments. It is based on the work of Karl Pearson in 1901 [Pear 01]. It transforms the data linearly into a new coordinate system. The transformation is characterized by the fact that the new coordinate axes lie in the direction of greatest data variances and are orthogonally to each other. Coordinates that were possibly correlated are transformed to new, uncorrelated coordinates.
Let $\boldsymbol{S} \in \mathbb{R}^{d \times N}$ the matrix containing as columns the feature vectors $\boldsymbol{x} \in \mathbb{R}^d$ of the training set $\mathcal{S} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)\}$ with classes $y \in \mathbb{Z}$. We assume that the $\boldsymbol{x}$ are centered, thus have zero mean, which can be achieved by subtracting their means. Furthermore, let $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ be the covariance matrix of the data. As $\boldsymbol{\Sigma}$ is estimated by the maximum likelihood method (Equation 4.40), it is positive

semi-definite and symmetric.

$$\begin{aligned} \boldsymbol{\Sigma} &= \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \\ &\text{with the mean } \boldsymbol{\mu} = \frac{1}{N}\sum_i \boldsymbol{x}_i \end{aligned} \tag{4.40}$$

Therefore, a transformation

$$\boldsymbol{D} = \boldsymbol{U}^\top \boldsymbol{\Sigma} \boldsymbol{U} \tag{4.41}$$

exists, where $\boldsymbol{D} \in \mathbb{R}^{d\times d}$ is a diagonal matrix with the Eigenvalues of $\boldsymbol{\Sigma}$ and $\boldsymbol{U} \in \mathbb{R}^{d\times d}$ is an orthogonal matrix with its column vectors being Eigenvectors of $\boldsymbol{\Sigma}$. For dimension reduction, the Eigenvectors are sorted by the size of their corresponding Eigenvalues $\lambda_i \in \mathbb{R}_0^+, \quad i = 1, \ldots, d$. A transformation matrix $\hat{\boldsymbol{U}} \in \mathbb{R}^{d\times k}, \quad k \leq d$ is formed by using the Eigenvectors belonging to the k largest Eigenvalues as columns in $\hat{\boldsymbol{U}}$. A data point $\boldsymbol{x}$ is then transformed by

$$\hat{\boldsymbol{x}} = \hat{\boldsymbol{U}}^\top \boldsymbol{x} \quad . \tag{4.42}$$

Analogously to Equation 4.41, the matrix $\hat{\boldsymbol{D}} \in \mathbb{R}^{k\times k}$ is the resulting covariance matrix after this transformation. The variance $v \in \mathbb{R}_0^+$ in the data, that is explained by those k largest components (PCA components) is accordingly:

$$v_k = \frac{\sum_{i=1}^{k}\lambda_i}{\sum_{j=1}^{d}\lambda_j} \tag{4.43}$$

The components of the Eigenvectors are known as loadings and describe how much each initial component of $\boldsymbol{x}$ contributes to the new coordinates $\hat{\boldsymbol{x}}$.
The PCA as described above decorrelates the data, the covariance matrix of the transformed data is a diagonal matrix with certain Eigenvalues λ_j on the diagonal, which usually are different from each other. Making these Eigenvalues the same is called "whitening" the data. The whitened data $\boldsymbol{w} \in \mathbb{R}^k$ can be obtained by transforming $\hat{\boldsymbol{x}}$ according to Equation 4.44.

$$\boldsymbol{w} = \hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{x}} = \hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{U}}^\top \boldsymbol{x} \tag{4.44}$$

The resulting covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ of the PCA is then:

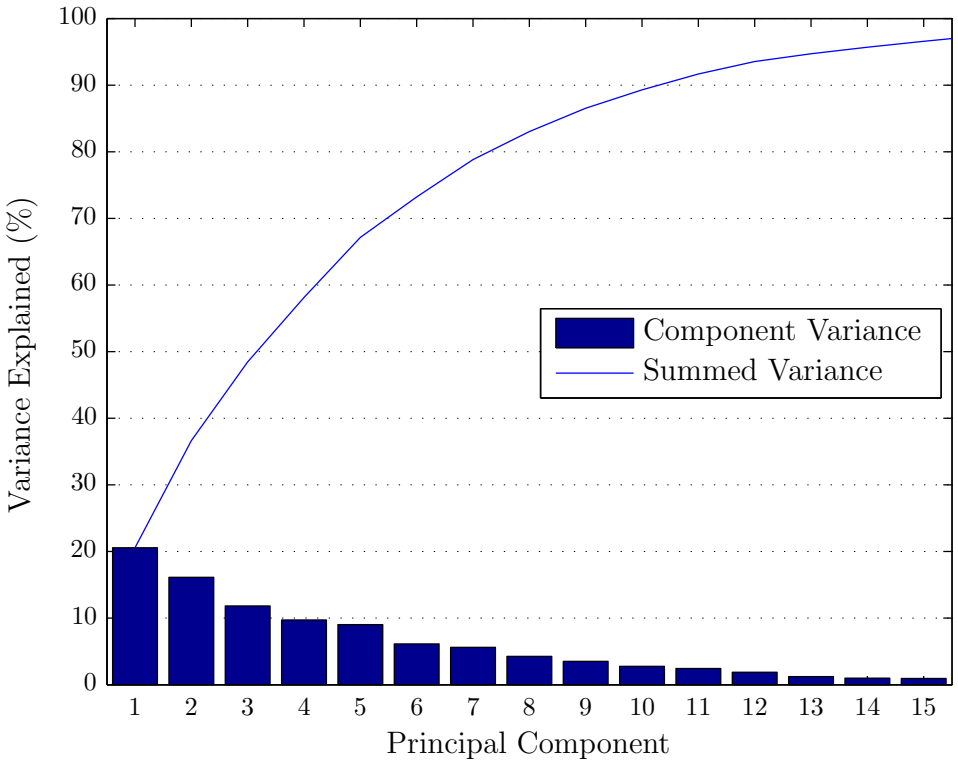
$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{w}} &= \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{w}_i - \boldsymbol{\mu}_{\boldsymbol{w}})(\boldsymbol{w}_i - \boldsymbol{\mu}_{\boldsymbol{w}})^\top \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{U}}^\top \boldsymbol{x}_i - \hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{U}}^\top \boldsymbol{\mu}\right)\left(\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{U}}^\top \boldsymbol{x}_i - \hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{U}}^\top \boldsymbol{\mu}\right)^\top \\ &= \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{U}}^\top(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^\top \hat{\boldsymbol{U}}\hat{\boldsymbol{D}}^{-\frac{1}{2}\top} \\ &= \hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{U}}^\top \boldsymbol{\Sigma}\hat{\boldsymbol{U}}\hat{\boldsymbol{D}}^{-\frac{1}{2}\top} \\ &= \hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{D}}\hat{\boldsymbol{D}}^{-\frac{1}{2}\top} = \boldsymbol{E} \ , \end{aligned} \tag{4.45}$$

with the identity matrix $\boldsymbol{E} \in \mathbb{R}^{k \times k}$ and the mean $\boldsymbol{\mu}_{\boldsymbol{w}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{w}_i$.

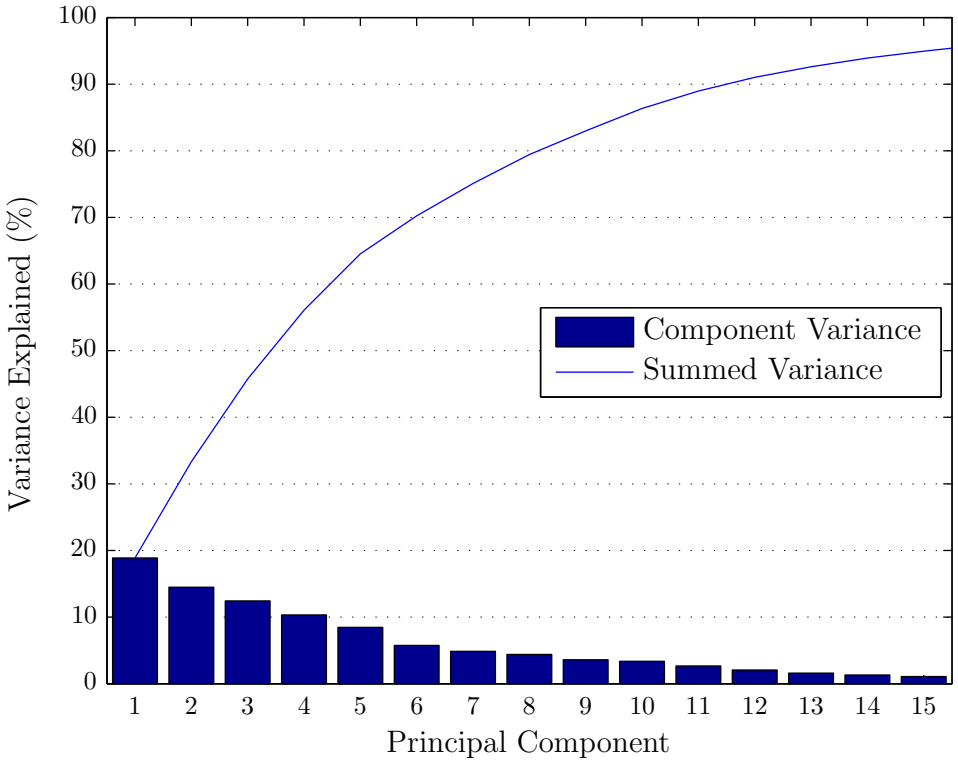
In our experiments, we used the transformation 4.42 to decorrelate the data and to reduce the dimension of the feature vectors to $k = \{3, 5, 7, 10, 15\}$ components. However, when applying the LSUV method (Section 4.7.2) to the PCA components, the result is whitened data as well.
Figure 4.21 shows the explained variance depending on the choice of k for patient subsets SUB44 and SUB32. One sees that $> 90\%$ of the variance are explained by the first 11 (SUB32) and 12 (SUB44) components. The 10 largest loadings of the first PCA component are listed in Table 4.2. We found that all different imaging modalities contribute to the transformed vectors to a significant amount, which makes the PCA features truly multimodal. The loadings for the second and third PCA components can be found in the Appendix in Tables A.10 and A.11.

SUB44 Component #	Loading	Feature Name
1	+0.17	MRIT2FlairHyperDensePixels
2	−0.17	MRIT2FlairIsoDensePixels
3	−0.16	PETQuotMinVOIMeanRef
4	+0.16	MRIT2FlairQuotMeanMeanRefVOI
5	−0.16	PETDynVOIMin
6	+0.15	MRIT2FlairEffectiveKLDPDF
7	−0.15	PETInterceptMin
8	−0.15	MRIT1EnhancedQuotMinMeanRefVOI
9	+0.14	MRIT1EnhancedHypoDensePixels
10	−0.13	MRIT1EnhancedIsoDensePixels
SUB32 Component #	Loading	Feature Name
1	+0.16	PETQuotMinVOIMeanRef
2	−0.16	MRIT2FlairHyperDensePixels
3	+0.16	MRIT2FlairIsoDensePixels
4	+0.16	PETDynVOIMin
5	−0.15	MRIT2FlairQuotMeanMeanRefVOI
6	−0.15	MRIT2FlairEffectiveKLDPDF
7	+0.15	PETInterceptMin
8	−0.15	MRIT1EnhancedHypoDensePixels
9	+0.14	MRIT1EnhancedIsoDensePixels
10	+0.13	MRIT1EnhancedQuotMinMeanRefVOI

Table 4.2: The ten largest loadings (coefficients) of the first Eigenvector of the principal component analysis for the features of patient subset SUB44 and SUB32. The transformed feature vectors are to a large degree formed by a linear combination of feature from T_2-MRI, T_1-MRI, and PET.



(a) SUB32



(b) SUB44

Figure 4.21: Pareto chart of the explained variance depending on the PCA component for patient subsets SUB32 (a) and SUB44 (b). The bars indicate the variance explained by the individual component, the line graph shows the summed variance that is explained by the union of those components.

Chapter 5

Classification

In this chapter, we aim to introduce the underlying principle of each classifier. For this, we derive basis equations and algorithms, decision functions, and decision boundaries. We also present our cross-validation setup in order to estimate a generalized classification accuracy and introduce measures for the classification performance.
The huge variety of different classification concepts that we used in this work was motivated by the results from literature. E. g. Caruana et al. [Caru 06] presented a comparison between ten supervised learning methods, SVMs, Neural Nets, Logistic Regression, Naïve Bayes, memory-based learning, Random Forests, decision trees, bagged trees, boosted trees, and boosted stumps and found that "although some methods clearly perform better or worse than other methods on average, there is significant variability across the problems and metrics. Even the best models sometimes perform poorly, and models with poor average performance occasionally perform exceptionally well.". Similar work was done earlier by King et al. [King 95]. They compared a variety of classifiers (e. g. Naïve Bayes, k-Nearest Neighbor, Linear and Quadratic Discriminant Analysis, and Neural Networks) on a set of different problems from the field of image analysis, medicine, engineering, and finance. Their "main conclusion is that there is no single best algorithm,...". The general consensus thus is, that it is not a priori clear which classifier will perform better on an unknown problem.
We applied several automatic approaches, as well as one semi-automatic classification approach to our data. All our automatic approaches are based on algorithms that rely on supervised learning. In general, we worked with balanced datasets, which means that all classes have equal occurrence probabilities.
We used two classifiers which base on Bayes' theorem and which assume Gaussian data: the Naïve Bayes classifier and the Linear Discriminant Analysis. The former assumes that the individual dimensions of the feature vector of the data are mutually independent. The latter transforms the data in a certain way in order to find an optimal decision boundary. Naïve Bayes yields quadratic decision boundaries, LDA linear decision boundaries, both assuming Gaussian probability densities. Classifiers that in theory can approximate arbitrary functions as decision boundaries in the feature space are: AdaBoost, which fits an additive model of weak classifiers (Decision Stumps). A Neural Network in the form of a single hidden layer perceptron and non-linear activation functions. A Support Vector Machine with soft margins and

polynomial or Gaussian kernels. Additionally, we used a k-Nearest Neighbor classifier as simple approach which can lead to highly non-linear decision boundaries as well.
For further information and more detailed derivations, please refer to e. g. Hastie et al. [Hast 09], which were used as orientation for writing these sections. Technically, the scripting and cross-validation was carried out using MathWorks MATLAB [MATL 10]. The classifiers were incorporated by Java calls to the WEKA framework (Version 3.6.0) [Hall 09].

Notation
If not specified otherwise, italics (e. g. z or Z) denote scalars, lowercase boldface ($\boldsymbol{z}$) vectors, uppercase boldface ($\boldsymbol{Z}$) matrices and uppercase calligraphic $\mathcal{Z}$ sets.
The norm $||.||_2$ is the L^2-norm: $||\boldsymbol{x}||_2 = \left(\sum_{i=1}^{d} x_i^2\right)^{\frac{1}{2}}$.
The sign function is denoted by $\operatorname{sgn}(x)$:

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{5.1}$$

The determinant of a matrix $\boldsymbol{Z}$ is denoted by $|\boldsymbol{Z}|$. If not noted otherwise, a single quotation mark as in $f'(x)$ denotes the derivative of the function $f(x)$ at point x.

5.1 Bayes' Classifier

We used two classifiers that found on Bayes' theorem, which states that the class conditional probability density can be expressed by the posterior probability, the prior probability for class $y \in \mathbb{Z}$ and the evidence of feature vector $\boldsymbol{x} \in \mathbb{R}^d$.
Let $p(\boldsymbol{x}|y)$ be the class conditional probability density for the feature vector $\boldsymbol{x}$ and the class $y \in \mathbb{Z}$, $p(y|\boldsymbol{x})$ be the corresponding posterior probability, $p(\boldsymbol{x}, y)$ be the joint probability, $p(y)$ be the prior probability for the class y and $p(\boldsymbol{x})$ the evidence of $\boldsymbol{x}$ in the feature space. Bayes' theorem is subsequently derived by the definition of the conditional probability:

$$p(\boldsymbol{x}|y) = \frac{p(\boldsymbol{x}, y)}{p(y)} \quad \text{if } p(y) > 0 \tag{5.2}$$

$$p(y|\boldsymbol{x}) = \frac{p(y, \boldsymbol{x})}{p(\boldsymbol{x})} \quad \text{if } p(\boldsymbol{x}) > 0 \tag{5.3}$$

using $p(y, \boldsymbol{x}) = p(\boldsymbol{x}, y)$ and combining 5.2 and 5.3 leads to Bayes' theorem

$$p(\boldsymbol{x}|y) \cdot p(y) = p(y|\boldsymbol{x}) \cdot p(\boldsymbol{x}) \tag{5.4}$$

For classification, we aim to assign class y^* to feature vector $\boldsymbol{x}$ that maximizes the posterior. The decision rule for a Bayes' classifier is subsequently:

$$\begin{aligned} y^* &= \underset{y}{\operatorname{argmax}}\, p(y|\boldsymbol{x}) \\ &\text{which can be rewritten with Bayes' rule} \\ &= \underset{y}{\operatorname{argmax}}\, \frac{p(\boldsymbol{x}|y)\cdot p(y)}{p(\boldsymbol{x})} \\ &= \underset{y}{\operatorname{argmax}}\, p(\boldsymbol{x}|y)\cdot p(y) \end{aligned} \tag{5.5}$$

As we do not know the class conditional probability density $p(\boldsymbol{x}|y)$ directly and have to estimate it from our training data, we model the class conditional as Gaussian distribution with class specific mean vector $\boldsymbol{\mu}_y \in \mathbb{R}^d$ and co-variance matrix $\boldsymbol{\Sigma}_y \in \mathbb{R}^{d\times d}$, namely $p(\boldsymbol{x}|y) \in \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$.

5.2 Gaussian Classifier

For the Gaussian classifier, the class conditional probability $p(\boldsymbol{x}|y)$ is modeled by a Gaussian.

$$p(\boldsymbol{x}|y) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_y|}} \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_y)\} \quad , \tag{5.6}$$

The decision boundary $F(\boldsymbol{x})$ for Gaussian classifiers can be calculated by equating the posterior probabilities for the two classes. Here, we arbitrarily assign $y \in \{-1, 1\}$.

$$\begin{aligned} p(y=1|\boldsymbol{x}) &= p(y=-1|\boldsymbol{x}) \\ \log\frac{p(y=1|\boldsymbol{x})}{p(y=-1|\boldsymbol{x})} &= 0 = F(\boldsymbol{x}) \\ \log\frac{p(\boldsymbol{x}|y=1)\,p(y=1)}{p(\boldsymbol{x}|y=-1)\,p(y=-1)} &= 0 \\ \log\frac{p(y=1)}{p(y=-1)} + \log\frac{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_{y=-1}|}}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_{y=1}|}} - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{y=1})^\top\boldsymbol{\Sigma}_{y=1}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{y=1}) + & \\ + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{y=-1})^\top\boldsymbol{\Sigma}_{y=-1}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{y=-1}) &= 0 \\ \frac{1}{2}\boldsymbol{x}^\top\left(\boldsymbol{\Sigma}_{y=-1}^{-1}-\boldsymbol{\Sigma}_{y=1}^{-1}\right)\boldsymbol{x} + \boldsymbol{x}^\top\left(\boldsymbol{\Sigma}_{y=1}^{-1}\boldsymbol{\mu}_{y=1} - \boldsymbol{\Sigma}_{y=-1}^{-1}\boldsymbol{\mu}_{y=-1}\right) + & \\ + \log\frac{p(y=1)}{p(y=-1)} + \log\frac{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_{y=-1}|}}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_{y=1}|}} + \frac{1}{2}\left(\boldsymbol{\mu}_{y=-1}^\top\boldsymbol{\Sigma}_{y=-1}^{-1}\boldsymbol{\mu}_{y=-1} - \boldsymbol{\mu}_{y=1}^\top\boldsymbol{\Sigma}_{y=1}^{-1}\boldsymbol{\mu}_{y=1}\right) &= 0 \ , \end{aligned} \tag{5.7}$$

which represents a quadratic decision boundary for $\boldsymbol{x}$.
From the training set $\mathcal{S} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_N, y_N)\}$ with N_y as the number

of samples in a certain class, the $\boldsymbol{\mu}_y$, $\boldsymbol{\Sigma}_y$, and $p(y)$, are estimated by the maximum likelihood method:

$$\boldsymbol{\mu}_y = \frac{1}{N_y} \sum_{i:y_i=y} \boldsymbol{x}_i \tag{5.8}$$

$$\boldsymbol{\Sigma}_y = \frac{1}{N_y} \sum_{i:y_i=y} (\boldsymbol{x}_i - \boldsymbol{\mu}_y)(\boldsymbol{x}_i - \boldsymbol{\mu}_y)^\top \tag{5.9}$$

$$p(y) = \frac{N_y}{N} \tag{5.10}$$

For a potentially $d = 62$ dimensional feature vector like in our case and assuming mutual dependency of the feature dimensions, one would have to estimate the d dimensions of $\boldsymbol{\mu}_y$ plus $d \cdot (d+1)/2$ dimensions of $\boldsymbol{\Sigma}_y$ plus the prior $p(y)$ which results in a total of 2016 parameters per class.

5.2.1 Naïve Bayes

We apply a simplification of the problem in order to reduce the number of those parameters. It is assumed that the dimensions of the feature vectors are mutually independent. The co-variance matrices $\boldsymbol{\Sigma}_y$ only have non-zero elements on the diagonal. With this, the number of parameters to be estimated is reduced to $2d + 1 = 125$ per class.
Thus the Naïve Bayes approach is:

$$p(\boldsymbol{x}|y) = \prod_{i=1}^{d} p(x_i|y) \tag{5.11}$$

The decision rule of the Naïve Bayes algorithm is then:

$$y^* = \operatorname*{argmax}_y p(y) \prod_{i=1}^{d} p(x_i|y) \tag{5.12}$$

Without further limitation, the decision boundary of the classifier is quadratic. The Naïve Bayes classifier inherits this property from the underlying Gaussian classifier. Figure 5.1 is an example for such a decision boundary for the patient subset SUB44 and a two-dimensional feature space. If we want to take into account the mutual dependency of the different feature vector dimensions, but aim only to find linear decision boundaries, the Linear Discriminant Analysis can be applied.

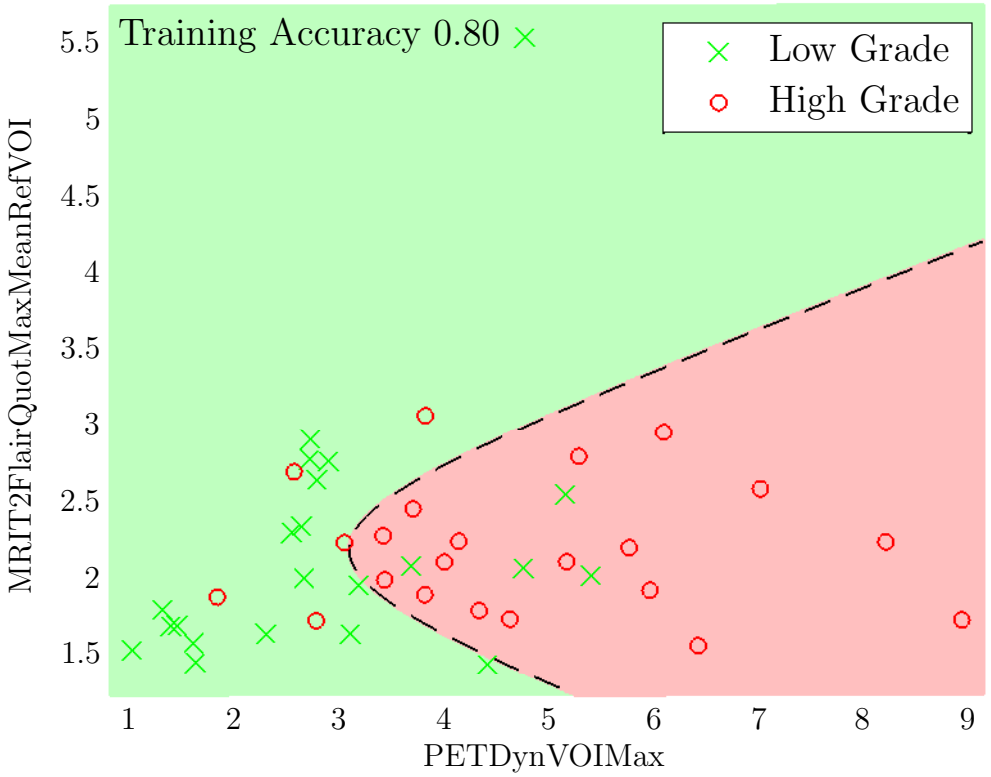


Figure 5.1: Decision boundary of the Naïve Bayes classifier for patient subset SUB44 and a two-dimensional feature space. The training accuracy is 80% in this example. The dashed line represents the decision boundary. The instances of SUB44 are colored in red if the underlying tumor was high-grade and colored in green if it was low-grade.

5.2.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a classifier that is based on the projection of samples, in order to yield decision boundaries that are linear in the feature space. Most basic work on this topic has been carried out by Fisher in 1936 [Fish 36]. However, we do not follow the original derivation but use the equivalency of LDA to a multivariate Gaussian classifier with a feature transformation which leads to identical covariance matrices for all classes [Frie 89]. The principle is as follows:
We assume that the class-conditional probability $p(\boldsymbol{x}|y)$ of a feature vector $\boldsymbol{x} \in \mathbb{R}^d$ with a class $y \in \{-1, 1\}$ is $p(\boldsymbol{x}|y) \in \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$. The means $\boldsymbol{\mu}_y \in \mathbb{R}^d$, the covariance matrices $\boldsymbol{\Sigma}_y \in \mathbb{R}^{d\times d}$, and the priors $p(y) \in [0, 1]$ are again maximum likelihood estimated by Equations 5.8 - 5.10.
As a consequence of this, the covariance matrices are positive semi-definite and symmetric, which means that they can be decomposed into orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{d\times d}$ and diagonal matrices $\boldsymbol{D} \in \mathbb{R}^{d\times d}$.

$$\begin{aligned}\boldsymbol{\Sigma}_y &= \boldsymbol{U}_y\boldsymbol{D}_y\boldsymbol{U}_y^\top = (\boldsymbol{U}_y\boldsymbol{D}_y^{\frac{1}{2}})(\boldsymbol{U}_y\boldsymbol{D}_y^{\frac{1}{2}})^\top \\ &= (\boldsymbol{U}_y\boldsymbol{D}_y^{\frac{1}{2}})\,\boldsymbol{E}\,(\boldsymbol{U}_y\boldsymbol{D}_y^{\frac{1}{2}})^\top\end{aligned} \tag{5.13}$$

where $\boldsymbol{E}$ is the identity matrix. Analogously we find for $\boldsymbol{\Sigma}_y^{-1}$:

$$\begin{aligned}\boldsymbol{\Sigma}_y^{-1} &= \left(\boldsymbol{U}_y\boldsymbol{D}_y\boldsymbol{U}_y^\top\right)^{-1} = \boldsymbol{U}_y^{\top-1}\boldsymbol{D}_y^{-1}\boldsymbol{U}_y^{-1} \\ (\text{since } \boldsymbol{U}^{-1} = \boldsymbol{U}^\top) \quad &= \boldsymbol{U}_y\boldsymbol{D}_y^{-1}\boldsymbol{U}_y^\top = (\boldsymbol{U}_y\boldsymbol{D}_y^{-\frac{1}{2}})\boldsymbol{E}(\boldsymbol{U}_y\boldsymbol{D}_y^{-\frac{1}{2}})^\top\end{aligned} \tag{5.14}$$

When we combine Equations 5.14 and 5.6, we get:

$$\begin{aligned}p(\boldsymbol{x}|y) &= \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_y|}}\exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_y)^\top(\boldsymbol{U}_y\boldsymbol{D}_y^{-\frac{1}{2}})\boldsymbol{E}(\boldsymbol{U}_y\boldsymbol{D}_y^{-\frac{1}{2}})^\top(\boldsymbol{x}-\boldsymbol{\mu}_y)\} \\ &= \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}_y|}}\exp\{-\frac{1}{2}(\boldsymbol{D}_y^{-\frac{1}{2}}\boldsymbol{U}_y^\top\boldsymbol{x} - \boldsymbol{D}_y^{-\frac{1}{2}}\boldsymbol{U}_y^\top\boldsymbol{\mu}_y)^\top\boldsymbol{E}(\boldsymbol{D}_y^{-\frac{1}{2}}\boldsymbol{U}_y^\top\boldsymbol{x} - \boldsymbol{D}_y^{-\frac{1}{2}}\boldsymbol{U}_y^\top\boldsymbol{\mu}_y)\}\end{aligned} \tag{5.15}$$

One finds that Equation 5.15 represents a Gaussian distributed random variable with transformed coordinates $\boldsymbol{x}' = \phi_y(\boldsymbol{x}) = \boldsymbol{D}_y^{-\frac{1}{2}}\boldsymbol{U}_y^\top\boldsymbol{x}$. The covariance matrix of this transformed variable is the identity matrix.
When looking at the decision boundary for Gaussian classifiers (Equation 5.7), for class-wise identical covariance matrices $\boldsymbol{\Sigma}_{y=1} = \boldsymbol{\Sigma}_{y=-1} = \boldsymbol{\Sigma} = \boldsymbol{E}$, which is the case for the transformed variables $\boldsymbol{x}'$, the quadratic terms cancel and we get a boundary that is linear in $\boldsymbol{x}'$.

$$\begin{aligned}\log\frac{p(y=1)}{p(y=-1)} - \frac{1}{2}\boldsymbol{\mu}_{y=1}'^\top\boldsymbol{E}\boldsymbol{\mu}_{y=1}' + \frac{1}{2}\boldsymbol{\mu}_{y=-1}'^\top\boldsymbol{E}\boldsymbol{\mu}_{y=-1}' + \boldsymbol{x}'^\top\boldsymbol{E}(\boldsymbol{\mu}_{y=1}' - \boldsymbol{\mu}_{y=-1}') = 0 \\ \log\frac{p(y=1)}{p(y=-1)} - \frac{1}{2}||\boldsymbol{\mu}_{y=1}'||_2^2 + \frac{1}{2}||\boldsymbol{\mu}_{y=-1}'||_2^2 + \boldsymbol{x}'^\top(\boldsymbol{\mu}_{y=1}' - \boldsymbol{\mu}_{y=-1}') = 0\ ,\end{aligned} \tag{5.16}$$

with $\boldsymbol{\mu}_y' = \phi_y(\boldsymbol{\mu}_y)$. See Figure 5.2 for an example of an LDA decision boundary.

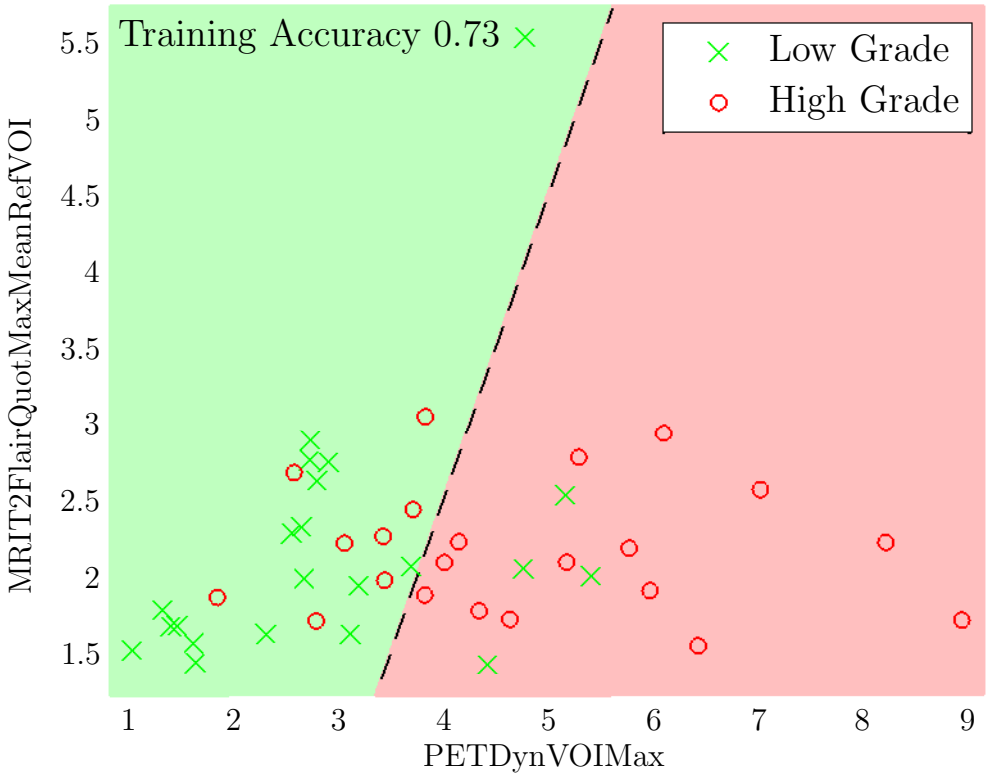


Figure 5.2: Decision boundary of the Linear Discriminant Analysis for patient subset SUB44 and a two-dimensional feature space. The training accuracy is 73% in this example. The dashed line represents the decision boundary. The instances of SUB44 are colored in red if the underlying tumor was high-grade and colored in green if it was low-grade. Please refer to Figure 5.1 for the same data classified with a Naïve Bayes classifier.

A disadvantage of this approach is, that the transformation is class specific. Also, the dimension of the initial feature vectors is not reduced, which means $\phi_y(\boldsymbol{x}) \in \mathbb{R}^d$. The decision rule of the LDA then is:

$$
\begin{aligned}
y^* &= \underset{y}{\operatorname{argmax}}\, p(y|\phi_y(\boldsymbol{x})) \\
&\overset{\text{(Bayes rule)}}{\Leftrightarrow} \quad = \underset{y}{\operatorname{argmax}}\, p(\phi_y(\boldsymbol{x})|y)\, p(y) \\
&\overset{\text{(Gaussian dist.)}}{\Leftrightarrow} \quad = \underset{y}{\operatorname{argmax}}\, p(y) \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_y|}} \exp\{-\frac{1}{2}(\phi_y(\boldsymbol{x}) - \phi_y(\mu_y))^\top \boldsymbol{E}\, (\phi_y(\boldsymbol{x}) - \phi_y(\boldsymbol{\mu}_y))\} \\
&\Leftrightarrow \quad = \underset{y}{\operatorname{argmax}} \left\{ \log(p(y)) - \frac{1}{2}(\phi_y(\boldsymbol{x}) - \phi_y(\boldsymbol{\mu}_y))^\top (\phi_y(\boldsymbol{x}) - \phi_y(\boldsymbol{\mu}_y)) \right\} \\
&\Leftrightarrow \quad = \underset{y}{\operatorname{argmin}} \left\{ \frac{1}{2} ||\phi_y(\boldsymbol{x}) - \phi_y(\boldsymbol{\mu}_y)||_2^2 - \log(p(y)) \right\}
\end{aligned}
\tag{5.17}
$$

Which means that for equal class occurrences, the $\boldsymbol{x}$ is classified as belonging to the class that minimizes the distance between $\phi_y(\boldsymbol{x})$ and the centroid $\phi_y(\boldsymbol{\mu}_y)$ of the class y.

5.2.3 Multi-Class for Naïve Bayes and LDA

Both classifiers are inherently able to be applied to class numbers $k > 2$. Therefore we did not use the one-versus-one classification scheme here.

5.3 AdaBoost

AdaBoost, an abbreviation for "Adaptive Boosting", is an algorithm that was originally published by Freund and Shapire in 1995 [Freu 95]. It is nowadays one of the most frequently applied boosting schemes. It is based on the intuitive idea that a combination of weak classifiers, which only perform slightly better than random prediction, can eventually lead to a strong classifier. What follows is the derivation of some key aspects of the AdaBoost algorithm.
Boosting can be understood as fitting an additive model (Equation 5.18) of a set of M basis functions $b(\boldsymbol{x}, \boldsymbol{\gamma}_m) \in \mathbb{R}$, with the individual weights $\beta_m \in \mathbb{R}$ and a specific set of parameters of the basis function, defined by a parameter vector $\boldsymbol{\gamma}_m$.

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} \beta_m b(\boldsymbol{x}; \boldsymbol{\gamma}_m) \tag{5.18}$$

For comparison, in three layer perceptrons (single hidden layer) $b(\boldsymbol{x}; \boldsymbol{\gamma}) = \mathrm{act}^H(\gamma_0 + \boldsymbol{\gamma}_1^\top \boldsymbol{x})$ is a (usually non-linear) activation function of the hidden layer and $\boldsymbol{\gamma}$ models a linear combination of the feature vector's elements (see Section 5.5). For AdaBoost, the $b(\boldsymbol{x}; \boldsymbol{\gamma})$ are weak classifiers.
Now let $\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\}, \quad i = 1, \ldots, N$ be the training set, with feature vectors $\boldsymbol{x}_i \in \mathbb{R}^d$ and assigned class labels $y_i \in \{-1, 1\}$. The additive models are fit by minimizing a certain loss function $L \in \mathbb{R}$ (Equation 5.19) which takes into account the summed error of all samples in the training set.

$$\underset{\{\beta_m, \boldsymbol{\gamma}_m\}_{m=1}^M}{\operatorname{argmin}} \sum_{i=1}^{N} L\left(y_i, \sum_{m=1}^{M} \beta_m b(\boldsymbol{x}_i; \boldsymbol{\gamma}_m)\right) \tag{5.19}$$

In case of AdaBoost, the following simplification and specifications to Equation 5.19 apply:

- The modeling is carried out in a stage-wise forward manner, which means that the basis functions are added sequentially, without changing the parameters and weights of the already added basis functions. $f_m(\boldsymbol{x}) = f_{m-1}(\boldsymbol{x}) + \beta_m b(\boldsymbol{x}; \boldsymbol{\gamma}_m)$
- The basis functions $b(\boldsymbol{x}; \boldsymbol{\gamma}_m)$ are weak classifiers $G_m(\boldsymbol{x}) \in \{-1, 1\}$.
- The loss function is the exponential loss $L(y, f(x)) = \exp\{-y\, f(x)\}$.

In consequence, at a step m, one must solve the problem:

$$(\beta_m, G_m) = \underset{\beta, G}{\operatorname{argmin}} \sum_{i=1}^{N} \exp\{-y_i(f_{m-1}(\boldsymbol{x}_i) + \beta G(\boldsymbol{x}_i))\} \tag{5.20}$$

One can re-write this with $w_i^{(m)} = \exp\{-y_i f_{m-1}(\boldsymbol{x}_i)\}$:

$$(\beta_m, G_m) = \underset{\beta, G}{\operatorname{argmin}} \sum_{i=1}^{N} w_i^{(m)} \exp\{-y_i \beta G(\boldsymbol{x}_i)\} \tag{5.21}$$

As $w_i^{(m)}$ does not depend on $G(\boldsymbol{x})$ or on β, it can be seen as weight factor that is updated after each iteration.
By splitting the sum in correctly and incorrectly classified terms and by using the indicator function I

$$I(y_i \neq G(\boldsymbol{x}_i)) = \begin{cases} 1 & \text{if } y_i \neq G(\boldsymbol{x}_i) \\ 0 & \text{if } y_i = G(\boldsymbol{x}_i) \end{cases} \tag{5.22}$$

Equation 5.21 can be reformulated to:

$$\begin{aligned} (\beta_m, G_m) &= \underset{\beta, G}{\operatorname{argmin}} \{e^{-\beta} \sum_{i:y_i = G(\boldsymbol{x}_i)} w_i^{(m)} + e^{\beta} \sum_{i:y_i \neq G(\boldsymbol{x}_i)} w_i^{(m)}\} \\ &= \underset{\beta, G}{\operatorname{argmin}} \{(e^{\beta} - e^{-\beta}) \sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(\boldsymbol{x}_i)) + e^{-\beta} \sum_{i=1}^{N} w_i^{(m)}\} \ . \end{aligned} \tag{5.23}$$

This optimization problem is solved in two steps: first, for every $\beta > 0$, some factors can be neglected as they are constants.

$$G_m = \underset{G}{\operatorname{argmin}} \sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G(\boldsymbol{x}_i)) \tag{5.24}$$

By using this solution for the G_m (Equation 5.24) in the objective function (Equation 5.20) and solving for β_m, one yields:

$$\beta_m = \frac{1}{2} \log \frac{1 - \text{err}_m}{\text{err}_m} \tag{5.25}$$

With $\text{err}_m = \frac{\sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G_m(\boldsymbol{x}_i))}{\sum_{i=1}^{N} w_i^{(m)}}$, which effectively is an error rate, where each misclassification is weighted by $w_i^{(m)}$.
As already defined, the update rule is of additive nature

$$f_m(\boldsymbol{x}) = f_{m-1}(\boldsymbol{x}) + \beta_m G_m(\boldsymbol{x}) \quad , \tag{5.26}$$

which leads to the update rule for the weights

$$w_i^{(m+1)} = w_i^{(m)} \cdot \exp\{-y_i \beta_m G_m(\boldsymbol{x}_i)\} \quad . \tag{5.27}$$

By using the indicator function I and $-y_i G_m(\boldsymbol{x}_i) = 2 \cdot I(y_i \neq G_m(\boldsymbol{x}_i)) - 1$, one can reformulate this to

$$w_i^{(m+1)} = w_i^{(m)} \cdot \exp\{2\beta_m \cdot I(y_i \neq G_m(\boldsymbol{x}_i))\} \cdot e^{-\beta_m} \quad . \tag{5.28}$$

Commonly, the term $2\beta_m$ is referred to as $\alpha_m = 2\beta_m$. Please note, that in its original version of the algorithm (Alg. 2), the factor $e^{-\beta_m}$ for updating the weights is omitted. However, this is still equivalent to Equation 5.28. The factor changes all weights in the same way and is therefore neither relevant in the training of the classifiers $G_m(\boldsymbol{x})$ on basis of the weighted samples, nor on the calculation of the weighted error rate err_m, since $e^{-\beta_m}$ multiplies numerator and denominator of the fraction.

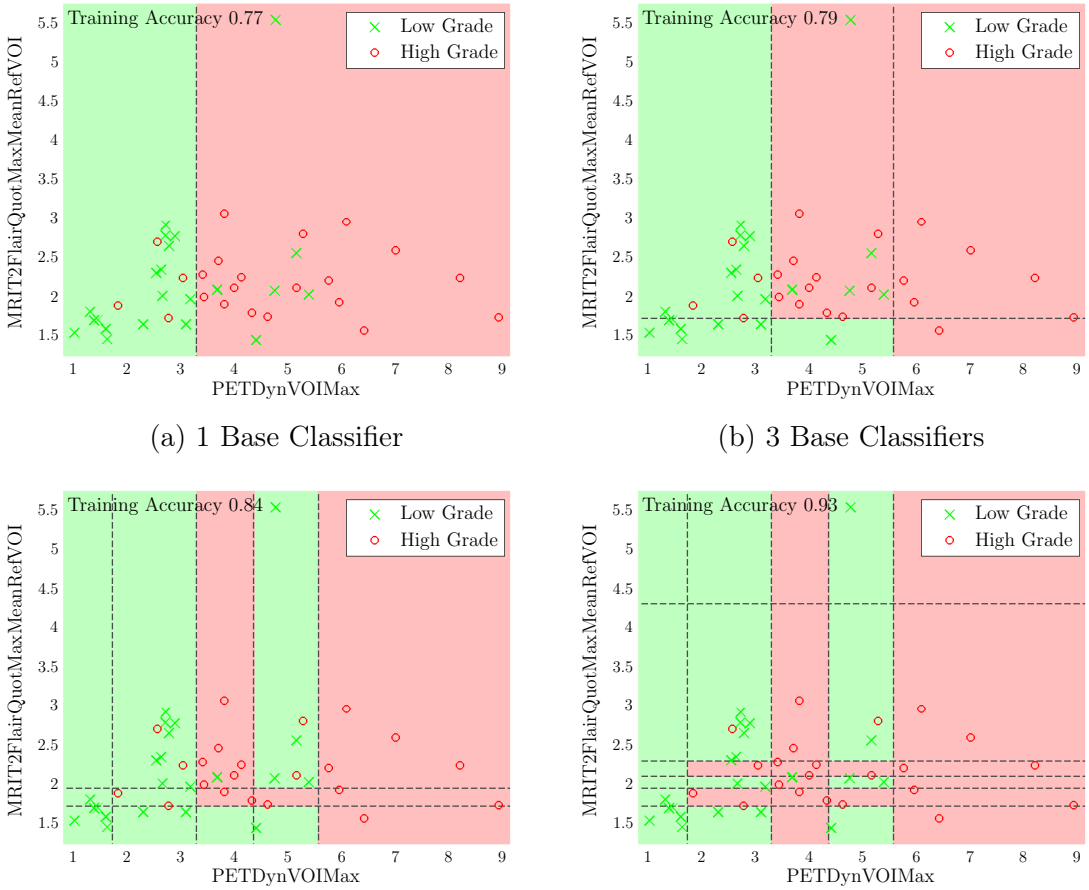
In conclusion, AdaBoost iteratively optimizes an exponential loss function. In each step of the iteration, a new weak classifier is fitted to re-weighted data. The weights of the data depend on the misclassification rate in the previous step. The weight is increased for misclassified samples since the argument of the exponential function in Equation 5.28 is positive in these cases. In an illustrative view, with increasing iteration number, the classifier focuses on the hard-to-classify samples. Figure 5.3 displays the decision boundaries of an AdaBoost scheme with decision stumps for a varying number of iterations. One can see how the weak classifiers are forced to focus more and more on the hard-to-classify instances. Different variants for the actual implementation of AdaBoost exist. We used for our experiments the *AdaBoost.M1* algorithm (Algorithm 2), which originally has been proposed by Freund et al. [Freu 96]. With this, the classification rule for an unknown sample $\boldsymbol{x}$ is simply

input : Training Set $\mathcal{S}$, weak classifier $G_m(\boldsymbol{x})$, integer M specifying the number of iterations
ouput: Final classifier $G^*(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m G_m(\boldsymbol{x})$
begin
  Initialize the weights $w_i = 1/N \quad \forall i$;
  for $m = 1$ **to** M **do**
    Fit the classifier $G_m(\boldsymbol{x})$ to the training data $\mathcal{S}$, using the weights w_i;
    Calculate the weighted error rate $\text{err}_m = \frac{\sum_{i=1}^{N} w_i^{(m)} I(y_i \neq G_m(\boldsymbol{x}_i))}{\sum_{i=1}^{N} w_i^{(m)}}$;
    If $\text{err}_m > 0.5$ set $M = m - 1$ and abort loop;
    Calculate $\alpha_m = \log\{\frac{(1-\text{err}_m)}{\text{err}_m}\}$;
    Update the weights $w_i^{m+1} = w_i^{(m)} \cdot \exp\{\alpha_m \cdot I(y_i \neq G_m(\boldsymbol{x}_i))\} \quad \forall i$;
  end
  Output the final classifier $G^*(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m G_m(\boldsymbol{x})$;
end

Algorithm 2: AdaBoost.M1

$y^* = G^*(\boldsymbol{x}) = \text{sgn}(\sum_{m=1}^{M} \alpha_m G_m(\boldsymbol{x}))$.
As mentioned, we used decision stumps for the weak classifiers. These are one-level decision trees [Ai 92]. They classify on basis of one single feature. For real-valued feature vectors, as in our case, a decision stump is fully defined by the feature it operates on and a threshold for this feature that separates the two classes.

Multi-Class Implementation
Since the AdaBoost.M1 algorithm needs the classification accuracy of the weak classifier to be at least $> \frac{1}{2}$, which is hard to achieve in the 4 class case using decision stumps, we applied a one-versus-one classification scheme in that cases. Consequently we trained $4 \cdot (4-1)/2$ binary classifiers and decided by majority vote. Ties (both classes have the same number of votes) were assigned by random decision.



(a) 1 Base Classifier

(b) 3 Base Classifiers

(c) 11 Base Classifiers

(d) 18 Base Classifiers

Figure 5.3: Decision boundaries of the weak classifiers for multiple iteration numbers (=number of base classifiers) for Adaptive Boosting with decision stumps as weak classifiers. a) 1 weak classifier, b) 3 weak classifiers, c) 11 weak classifiers, and d) 18 weak classifiers. Please note that for c) and d), not all decision boundaries of the weak classifiers are displayed as they lie too close to each other. The data is from patient subset SUB44 and a two-dimensional feature space. The training accuracy increases from a) 77% to b) 79% to c) 84% until d) 93%. The dashed lines represent the decision boundaries of the weak classifiers. The instances of SUB44 are colored in red if the underlying tumor was high-grade and colored in green if it was low-grade. Please also compare to Figures 5.1 and 5.2 for the same data classified with a Naïve Bayes classifier and Linear Discriminant Analysis.

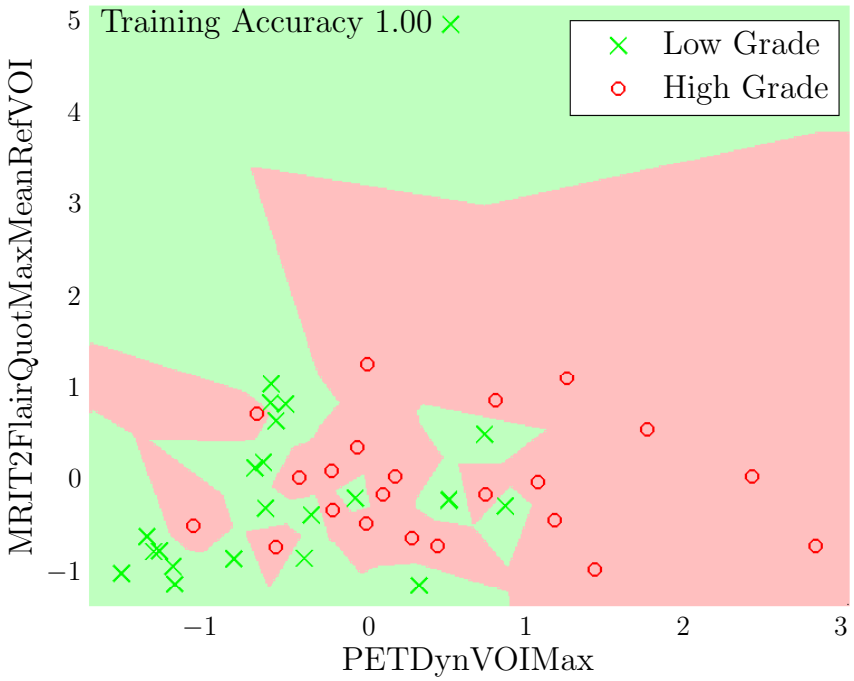
5.4 Nearest Neighbor

The k-Nearest Neighbor (kNN) classifier is a model free classifier – no specific properties of the underlying probability distributions of the features of distinct classes are assumed. It is thought to originate from ideas and formulations of Fix and Hodges in 1951. Later on several properties were proved by Cover et al. [Cove 67], e. g. that for 1-nearest neighbor classification and dense training sets (number of training samples $N \to \infty$) it approaches twice the error rate of the Bayes classifier and therefore can be used to estimate the best achievable training performance in some cases. kNN is a prototype method, which means that the training data are represented by a set of points in feature space – called prototypes. Furthermore it is a "lazy" classifier in the sense that all training samples need to be kept in the memory and all computation is done during classification of an unknown sample $\boldsymbol{x} \in \mathbb{R}^d$. The algorithm is given in pseudo-code in Algorithm 3:

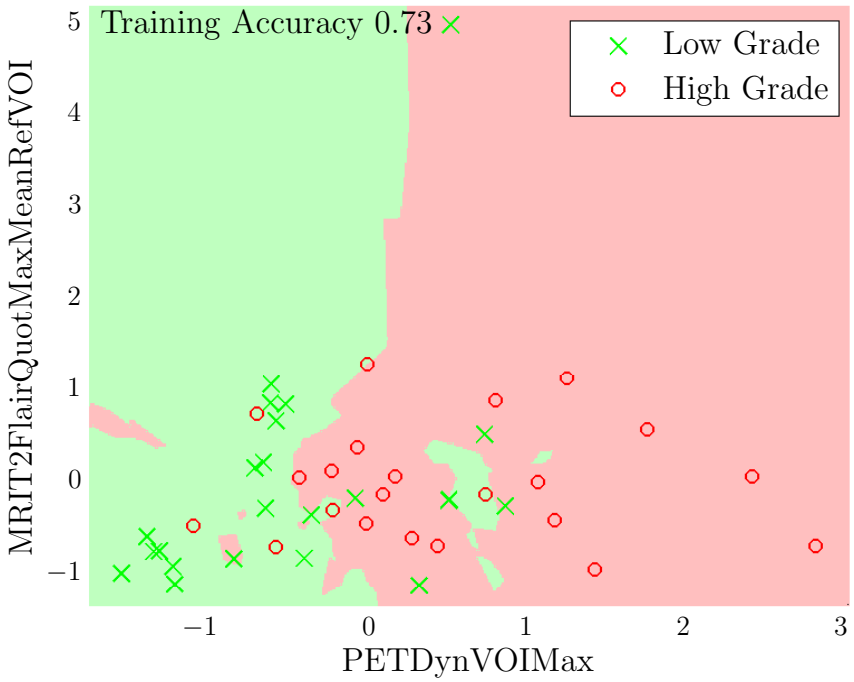
input : Test sample $\boldsymbol{x}$, training set $\mathcal{S}$, distance measure $d(\boldsymbol{x}, \boldsymbol{x}_i)$, integer k specifying the number of nearest neighbors to be evaluated
ouput: Class estimate y^* of sample $\boldsymbol{x}$, according to the classifier
begin
  for $i = 1$ **to** N **do**
    Calculate the distances $d_i(\boldsymbol{x}, \boldsymbol{x}_i)$;
  end
  Determine the k-closest training samples based on d_i;
  Determine the most frequent class label y^* among the k-closest training samples;
  Output the class estimate y^* of $\boldsymbol{x}$;
end

Algorithm 3: k-Nearest Neighbor Classifier

We used the Euclidean distance as distance metric $d(\boldsymbol{x}, \boldsymbol{x}_i) = ||\boldsymbol{x} - \boldsymbol{x}_i||_2$ and evaluated the classification accuracy for our problems for different numbers of neighbors k. The nearest neighbor method is inherently able to handle classification problems with a class number > 2, thus was not applied in a one-versus-one classification scheme.



(a) 1-Nearest Neighbor



(b) 3-Nearest Neighbors

Figure 5.4: Decision boundaries for the k-Nearest Neighbor classifier for a) 1-nearest neighbor and b) 3-nearest neighbors. The data is from patient subset SUB44 and a two-dimensional feature space. The training accuracy drops from a) 100% to b) 73%. The instances of SUB44 are colored in red if the underlying tumor was high-grade and colored in green if it was low-grade. The data were normalized to unit variance and zero mean (LSUV). Please compare to the other figures in this chapter for the same data classified with other classifiers.

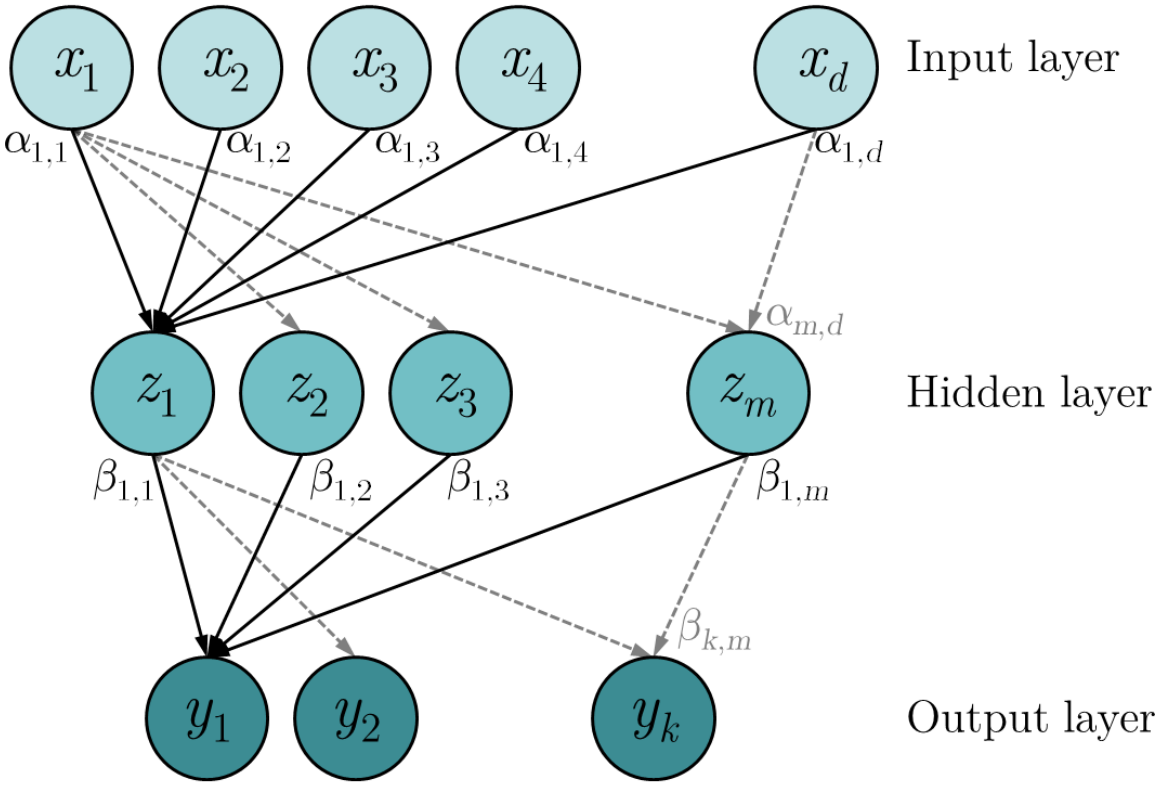


Figure 5.5: Schematic diagram of a perceptron featuring a single hidden layer. The input layer contains the elements of the feature vector $\boldsymbol{x} \in \mathbb{R}^d$. The output of the perceptron is a class vector $\boldsymbol{y}$. The output of the hidden layer is a vector $\boldsymbol{z} \in \mathbb{R}^m$. The layers are connected with weights $\alpha, \beta \in \mathbb{R}$. For clearness, not all connections and weights are shown.

5.5 Neural Networks

5.5.1 Definition of the Neural Network

We used a three layer perceptron (featuring a single hidden layer) as classifier of the neural networks family. The concept of the perceptron was first described in 1958 by Rosenblatt [Rose 58], which did not feature hidden layers in its original form. The network in our implementation is a feed-forward network without shortcuts. the input layer is only connected to the hidden layer and the hidden layer is only connected to the output layer. Every neuron of a certain layer is connected to all neurons of its succeeding layer (see also Figure 5.5). Now let $\boldsymbol{x} \in \mathbb{R}^d$ be a feature vector and $\boldsymbol{y}$ be its class, coded as binary vector, i. e. , for two classes $\boldsymbol{y} \in \left\{(0,1)^\top, (1,0)^\top\right\}$ and for four classes $\boldsymbol{y} \in \left\{(0,0,0,1)^\top, (0,0,1,0)^\top, (0,1,0,0)^\top, (1,0,0,0)^\top\right\}$. The single hidden layer perceptron and most other neural networks are inherently able to handle classification problems with a class number > 2. Thus we did not apply a one-versus-one classification scheme in case of the classification of the four individual WHO grades.

The input of the neurons of the input layer contains the elements $x_p, \quad p = 1, \ldots, d$, of the feature vector, the output of the neurons in the output layer is an estimated class

vector $\boldsymbol{y}^*$ as previously described. The number of input neurons equals the dimension of the feature vector d, the number of hidden neurons is m, the number of output neurons equals the class number k. The connections between the neurons of the input layer and the hidden layer are weighted by factors $\alpha_{o,p}, \quad o = 1, \ldots, m, \quad p = 1, \ldots, d$, the weights between the hidden layer and the output layer are $\beta_{j,o}, \quad j = 1, \ldots, k, \quad o = 1, \ldots, m$. We also allow for constant biases $\alpha_{o,0}$ and $\beta_{j,0}$. The input $\text{in}_o^{(H)}(\boldsymbol{x}) \in \mathbb{R}$ of the o-th neuron of the hidden layer under the input of a feature vector $\boldsymbol{x}$ is a weighted sum (Equation 5.29). For input and activation function, the superscript (H) indicates the hidden layer and (O) the output layer:

$$\text{in}_o^{(H)}(\boldsymbol{x}) = \alpha_{o,0} + \sum_{p=1}^{d} \alpha_{o,p} x_p \quad o = 1, \ldots, m \tag{5.29}$$

At the hidden layer, an activation function is applied to the input. This activation function should be a differentiable, non-linear function, in order to achieve a non-linear decision boundaries and to be able to update the weights by gradient descent steps. In our case, the activation function at the hidden layer is a sigmoid function $\text{act}^{(H)}(a) = \frac{1}{1+\exp\{-a\}} \in \mathbb{R}$. With this, the output of the o-th neuron of the hidden layer is:

$$z_o(\boldsymbol{x}) = \text{act}_o^{(H)}(\text{in}_o^{(H)}(\boldsymbol{x})) \quad o = 1, \ldots, m \tag{5.30}$$

As a consequence, the input of the j-th neuron of the output layer $\text{in}_j^{(O)}(\boldsymbol{x}) \in \mathbb{R}$ is the weighted sum of the outputs of the hidden layer:

$$\text{in}_j^{(O)}(\boldsymbol{x}) = \beta_{j,0} + \sum_{o=1}^{m} \beta_{j,o} z_o(\boldsymbol{x}) \quad j = 1, \ldots, k \tag{5.31}$$

Commonly, an activation function $\text{act}_j^{(O)}(\text{in}_j^{(O)}(\boldsymbol{x}))$ is also applied at the output layer. From literature, multiple possible functions are reported. To ensure the comparability of the output of the neural net and the binary coded classes $\boldsymbol{y}$, the functions should feature the properties $\sum_j \text{act}_j^{(O)}(\text{in}_j^{(O)}(\boldsymbol{x})) = 1$ and $\text{act}_j^{(O)}(\text{in}_j^{(O)}(\boldsymbol{x})) \in [0; 1]$. In the implementation that we used, the activation function of j-th neuron of the output layer was the "softmax" function, which fulfills the mentioned properties:

$$\text{act}_j^{(O)}(\text{in}_j^{(O)}(\boldsymbol{x})) = \frac{\exp\{\text{in}_j^{(O)}(\boldsymbol{x})\}}{\sum_{v=1}^{k} \exp\{\text{in}_v^{(O)}(\boldsymbol{x})\}} \quad j = 1, \ldots, k \tag{5.32}$$

5.5.2 Backpropagation of Errors

During training, the neural network learns by adjusting the weights α and β. In order to measure the goodness of fit of the classifier in its current state to the set $\mathcal{S}$ of the training data, we have to define an error function which is later on minimized. Several possibilities for an error function exist, however in most cases the sum-of-squared error

(SSE) function is used, but other functions are possible, e. g. cross-entropy. With SSE, the error is computed as follows:

$$\begin{aligned}\mathrm{SSE} &= \sum_{i=1}^{N}\sum_{j=1}^{k}(y_{i,j} - \mathrm{act}_j^{(O)}(\boldsymbol{x}_i))^2 \\ &= \sum_{i=1}^{N}\mathrm{SSE}_i\end{aligned} \tag{5.33}$$

Several techniques exist for optimizing the error function by adjusting the weights. The implementation that we used applied a gradient descent method for updating $\alpha^{(r+1)}, \beta^{(r+1)}$ from their previous versions $\alpha^{(r)}, \beta^{(r)}$. In this context, the superscripts indicate the iteration number r. Also, another free parameter γ is introduced which controls the step width of the gradient descent:

$$\alpha_{o,p}^{(r+1)} = \alpha_{o,p}^{(r)} - \gamma\sum_{i=1}^{N}\frac{\partial\,\mathrm{SSE}_i}{\partial\alpha_{o,p}^{(r)}} \quad o = 1,\ldots,m, \quad p = 1,\ldots,d \tag{5.34}$$

$$\beta_{j,o}^{(r+1)} = \beta_{j,o}^{(r)} - \gamma\sum_{i=1}^{N}\frac{\partial\,\mathrm{SSE}_i}{\partial\beta_{j,o}^{(r)}} \quad j = 1,\ldots,k, \quad o = 1,\ldots,m \tag{5.35}$$

The partial derivatives are simply calculated by subsequent application of the chain rule ($x_{i,j}$ is the j-th element of sample $\boldsymbol{x}_i$):

$$\begin{aligned}\frac{\partial\,\mathrm{SSE}_i}{\partial\alpha_{o,p}} &= \sum_{j=1}^{k}\frac{\partial\,\mathrm{SSE}_i}{\partial\,\mathrm{act}_j^{(O)}(\boldsymbol{x}_i)}\frac{\partial\,\mathrm{act}_j^{(O)}(\boldsymbol{x}_i)}{\partial\,\mathrm{in}_j^{(O)}(\boldsymbol{x}_i)}\frac{\partial\,\mathrm{in}_j^{(O)}(\boldsymbol{x}_i)}{\partial z_o(\boldsymbol{x}_i)}\frac{\partial z_o(\boldsymbol{x}_i)}{\partial\,\mathrm{in}_o^{(H)}(\boldsymbol{x}_i)}\frac{\partial\,\mathrm{in}_o^{(H)}(\boldsymbol{x}_i)}{\partial\alpha_{o,p}} \\ &= -2\sum_{j=1}^{k}(y_{i,j} - \mathrm{act}_j^{(O)}(\boldsymbol{x}_i))\cdot\mathrm{act}_j'^{(O)}(\boldsymbol{x}_i)\cdot\beta_{j,o}\cdot\mathrm{act}_o'^{(H)}(\mathrm{in}_o^{(H)}(\boldsymbol{x}))\cdot x_{i,p}\end{aligned} \tag{5.36}$$

$$\begin{aligned}\frac{\partial\,\mathrm{SSE}_i}{\partial\beta_{j,o}} &= \frac{\partial\,\mathrm{SSE}_i}{\partial\,\mathrm{act}_j^{(O)}(\boldsymbol{x}_i)}\frac{\partial\,\mathrm{act}_j^{(O)}(\boldsymbol{x}_i)}{\partial\,\mathrm{in}_j^{(O)}(\boldsymbol{x}_i)}\frac{\partial\,\mathrm{in}_j^{(O)}(\boldsymbol{x}_i)}{\partial\beta_{j,o}} \\ &= -2(y_{i,j} - \mathrm{act}_j^{(O)}(\boldsymbol{x}_i))\cdot\mathrm{act}_j'^{(O)}(\boldsymbol{x}_i)\cdot z_o(\boldsymbol{x}_i)\end{aligned} \tag{5.37}$$

With $\delta_{j,i} := -2(y_{i,j} - \mathrm{act}_j^{(O)}(\boldsymbol{x}_i))\cdot\mathrm{act}_j'^{(O)}(\boldsymbol{x}_i)$ and $s_{o,i} := \sum_{j=1}^{k}\beta_{j,o}\,\mathrm{act}_o'^{(H)}(\mathrm{in}_o^{(H)}(\boldsymbol{x}_i))\cdot\delta_{j,i}$ This can be re-written to:

$$\frac{\partial\,\mathrm{SSE}_i}{\partial\alpha_{o,p}} = s_{o,i}x_{i,p} \tag{5.38}$$

$$\frac{\partial\,\mathrm{SSE}_i}{\partial\beta_{j,o}} = \delta_{j,i}z_o(\boldsymbol{x}_i) \tag{5.39}$$

Where $\delta_{j,i}$ and $s_{o,i}$ are proportional to the difference of the output of the neural net and the actual class and can be understood as errors of the classification at a certain iteration. With these equations, the training of the neural network is as follows:

- The weights α, β are kept fixed and a feature vector $\boldsymbol{x}$ is propagated through the network (forward pass).
- The output of the net is compared to the actual class y_i and with it, the δ and subsequently the s are calculated (backward pass).
- Then, the δ and s are used to compute the gradients (Equations 5.36 and 5.37) and update the weights (Equations 5.34 and 5.35).

This kind of training algorithm is known as backpropagation of errors. This method ultimately goes back to the PhD-thesis of Werbos in 1974 [Werb 74] and is in depth described in Rumelhart et al. from 1986 [Rume 86]. The parameter γ influences the step width of every gradient descent step and is a free parameter of the classifier. It is commonly known as "learning rate". See Figure 5.6 for an example of the back-propagation of error process for an increasing number of iterations. In general, one does not aim to find the global minimum of the objective function of the neural net as it is very likely that the classifier overfits the problem at that point. In order to avoid overfitting, we optimize the number of iterations in a cross validation scheme and aim to find a number of iterations that terminates the process of learning before over-adaption occurs.

5.5.3 Speed-Up of Convergence

For a speed-up of the convergence, a momentum term [Rume 86] is added to α, β at training run $r+1$ (Equations 5.40 and 5.41). In an illustrative view, the momentum term enforces the convergence of the gradient descent at shallow areas of the objective function. The momentum term is weighted by the parameter κ, which itself is another free parameter that was optimized in our experimental setup.

$$\alpha_{o,p}^{(r+1)} = \alpha_{o,p}^{(r)} + \kappa \alpha_{o,p}^{(r)} - \gamma \sum_{i=1}^{N} \frac{\partial \, \mathrm{SSE}_i}{\partial \alpha_{o,p}^{(r)}} \quad o = 1, \ldots, m \quad p = 1, \ldots, d \tag{5.40}$$

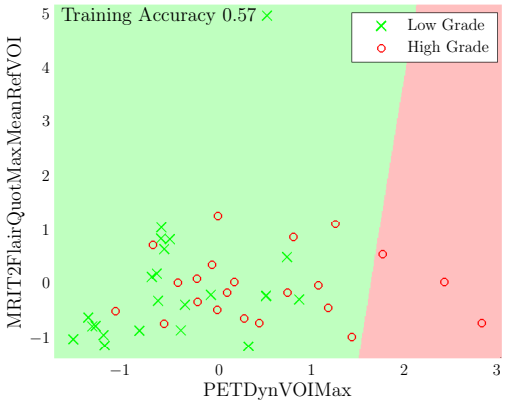
$$\beta_{j,o}^{(r+1)} = \beta_{j,o}^{(r)} + \kappa \beta_{j,o}^{(r)} - \gamma \sum_{i=1}^{N} \frac{\partial \, \mathrm{SSE}_i}{\partial \beta_{j,o}^{(r)}} \quad j = 1, \ldots, k \quad o = 1, \ldots, m \tag{5.41}$$

5.5.4 Training Algorithm and Decision Function

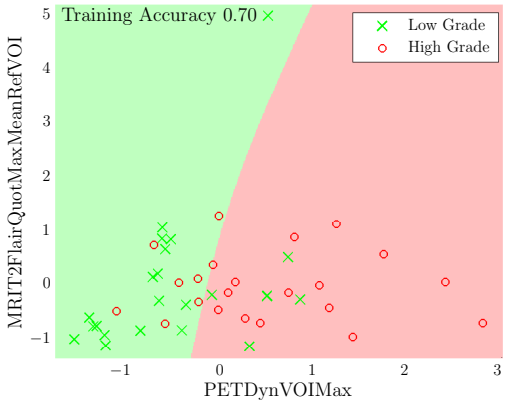
The algorithm for the training of the neural network using all training samples and backpropagation of error works as defined in Algorithm 4.
The decision function, given a feature vector $\boldsymbol{x}$ and a trained neural net is simply:

$$\boldsymbol{y}^* = \underset{\boldsymbol{y}}{\operatorname{argmin}} \sum_{j=1}^{k} \left(y_j - \mathrm{act}_j^{(O)}(\boldsymbol{x}) \right)^2 , \tag{5.42}$$

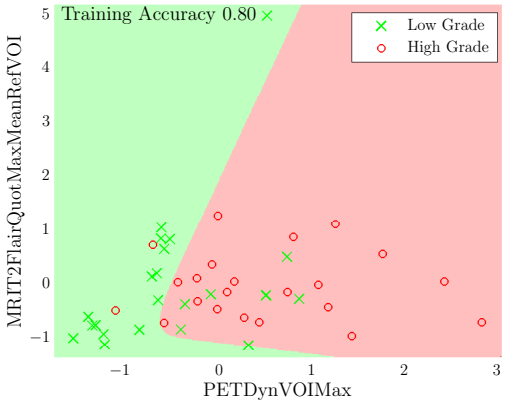
which effectively is deciding for the class $\boldsymbol{y}$ with the highest activation.



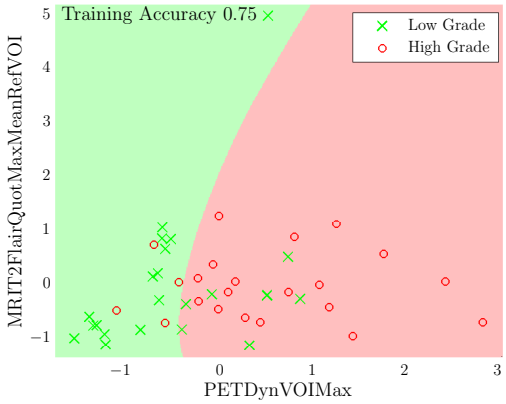
(a) $\gamma = 1.0 \quad \kappa = 0.1 \quad \text{Epoch } \# = 10$



(b) $\gamma = 1.0 \quad \kappa = 0.1 \quad \text{Epoch } \# = 100$



(c) $\gamma = 1.0 \quad \kappa = 0.1 \quad \text{Epoch } \# = 1000$



(d) $\gamma = 0.1 \quad \kappa = 0.1 \quad \text{Epoch } \# = 1000$

Figure 5.6: Decision boundaries of a single hidden layer perceptron for a different number of iterations (epochs) and two values of the learning rate γ. The momentum weight κ was held constant. One finds, with increasing iteration number, that the classifier adapts better to the training data. When comparing c) and d), we see that a lower step width delays the adaption of the classifier to the problem. However, large step widths could lead to the effect that the global minimum of the error function is missed. The training accuracy of this example was a) 57%, b) 70%, c) 80%, and d) 75%. The configuration of the network was: Two neurons in the input layer, two neurons in the hidden layer, and two neurons in the output layer. As in the examples for other classifiers, the data stem from patient subset SUB44 and a two-dimensional feature space and were normalized to unit variance and zero mean (LSUV). The instances of SUB44 are colored in red if the underlying tumor was high-grade and colored in green if it was low-grade. Please compare to the other figures in this chapter for the same data classified with other classifiers.

input : Training set $\mathcal{S}$, float γ specifying the learning rate, float κ specifying the momentum rate, integer R specifying the number of epochs
ouput: The trained neural network
begin
  Construct a neural net with a input layer with d neurons and one output layer with k neurons and a single hidden layer with $(d+k)/2$ neurons;
  Initialize the weights α, β to a random number $\in [-1, 1]$;
  for $r = 1$ **to** R **do**
    for $i = 1$ **to** N **do**
      Use feature vector x_i as input and calculate the output of the net;
      Update the weights $\alpha^{(r)}, \beta^{(r)}$ according to the backpropagation of error equations;
    end
  end
  Output the trained neural network;
end

Algorithm 4: Training algorithm for the single hidden layer perceptron

5.5.5 Initialization of the Neural Network

Neural networks are especially sensitive to the initially chosen weight values $\alpha_{o,p}^{(0)}, \beta_{j,o}^{(0)}$. E. g. if the starting weights are chosen to be equal, the input $\text{in}^{(H)}$ and the output z for all hidden neurons are equal as well, as are the inputs $\text{in}^{(O)}$ and output y. As a consequence, when updating the weights with the backpropagation of errors (Equation 5.36, 5.37) all weights are adjusted the same way and are equal again. If good adaption to the problem requires unequal weights, the problem can never be appropriately solved. As a solution for this, a random initialization of the weights was proposed by Rumelhart et al. [Rume 86]. Regarding the magnitude of the weights, initializing the starting weights to large values can result in very slow convergence of the algorithm. The reason is that the activation functions used in the hidden layer saturate for large inputs, which means that their derivatives become near zero. For good convergence and for avoiding saturation with sigmoidal activation functions, it is recommended for the inputs to be of order unity [Bish 95].
Let us assume that our input data are standardized, i. e. $\langle x_p \rangle = 0$ and $\langle x_p^2 \rangle = 1, p = 1, \ldots, d$ with $\langle . \rangle$ denoting the expected value. The expected average input $\langle \text{in}^{(H)} \rangle$ for a hidden unit is:

$$\begin{aligned} \langle \text{in}^{(H)} \rangle &= \sum_{p=1}^{d} \langle \alpha_p x_p \rangle \\ \overset{(\alpha_p, x_p \text{ uncorrelated})}{\Leftrightarrow} &= \sum_{p=1}^{d} \langle \alpha_p \rangle \langle x_p \rangle \\ \overset{\text{since } \langle x_p \rangle = 0}{\Leftrightarrow} &= 0 \end{aligned} \tag{5.43}$$

And the expected variance $\langle(\text{in}^{(H)})^2\rangle$ of the input is:

$$\begin{aligned}\langle(\text{in}^{(H)})^2\rangle &= \left\langle \left(\sum_{p=1}^{d} \alpha_p x_p\right) \left(\sum_{q=1}^{d} \alpha_q x_q\right) \right\rangle \\ \overset{(\alpha_p,\alpha_q \text{ uncorrelated})}{\Leftrightarrow} &= \sum_{p=1}^{d} \langle\alpha_p^2\rangle\langle x_p^2\rangle \\ &= \sigma^2 d\end{aligned} \tag{5.44}$$

with the variance σ^2 of the weights. Consequently, for large σ^2 the probability increases that some units of the hidden layer are already saturated from beginning, which could potentially lead to poor classification results [Hast 09].
For mentioned reasons, the starting weights in our implementation are chosen as random values near zero. E. g. for properly normalized input features, it is common to chose the starting weights $\alpha_{o,p}^{(0)}, \beta_{j,o}^{(0)} \in [-1, 1]$.

5.6 Support Vector Machines

Support Vector Machines (SVM) are based on the learning theory proposed by Vapnik et al. [Vapn 82] in 1982. SVMs are linear classifiers, the decision boundary that they fit is a hyperplane. As there exist an infinite number of separating hyperplanes (assuming linearly separable data), one can ask which of these hyperplanes is optimal in terms of generalization of the classifier. For that, SVM are constructed in such way that they fit a hyperplane which maximizes the distance (margin) to the closest training samples from both classes. It is thought that this helps in correctly classifying samples which have not been in the training set.

5.6.1 Hard Margin SVM

We start with a training set $\mathcal{S} = \{(\boldsymbol{x}_i, y_i)\}, \quad i = 1, \ldots, N$, with feature vectors $\boldsymbol{x}_i \in \mathbb{R}^d$ and class labels $y_i \in \{-1, 1\}$. Let

$$\left\{\boldsymbol{x} : \boldsymbol{\beta}^\top \boldsymbol{x} + \beta_0 = 0\right\} \tag{5.45}$$

be a hyperplane defined by $\boldsymbol{\beta}$ and β_0. For a point $\boldsymbol{x}_0$ on the plane one finds

$$\boldsymbol{\beta}^\top \boldsymbol{x}_0 = -\beta_0 \quad . \tag{5.46}$$

The signed distance $d(\boldsymbol{x}) \in \mathbb{R}$ of any point $\boldsymbol{x}$ to the plane is then (with $\hat{\boldsymbol{\beta}} = \frac{\boldsymbol{\beta}}{||\boldsymbol{\beta}||_2}$)

$$\begin{aligned} d(\boldsymbol{x}) &= \hat{\boldsymbol{\beta}}^\top (\boldsymbol{x} - \boldsymbol{x}_0) \\ &= \frac{\boldsymbol{\beta}^\top}{||\boldsymbol{\beta}||_2} \boldsymbol{x} + \frac{\beta_0}{||\boldsymbol{\beta}||_2}) \\ &= \frac{1}{||\boldsymbol{\beta}||_2} (\boldsymbol{\beta}^\top \boldsymbol{x} + \beta_0) \end{aligned} \tag{5.47}$$

Using Equation 5.47, for correct classification of all training data, one gets the condition

$$y_i \frac{1}{||\boldsymbol{\beta}||_2} (\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) \geq 0 \quad \forall i. \tag{5.48}$$

Now let us assume that all $\boldsymbol{x}_i$ should have a certain distance $M \in \mathbb{R}$ to the hyperplane:

$$y_i \frac{1}{||\boldsymbol{\beta}||_2} (\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) \geq M \quad \forall i \tag{5.49}$$

The fundamental principle of SVMs is, that the distance M is maximized, which leads to the following optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{\beta}, \beta_0}{\operatorname{argmax}}\, M \\ \text{subject to} \quad &y_i \frac{1}{||\boldsymbol{\beta}||_2} (\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) \geq M \quad \forall i \end{aligned} \tag{5.50}$$

Choosing $M = \frac{1}{||\boldsymbol{\beta}||_2}$ and reformulating as an equivalent convex optimization problem, one gets

$$\begin{aligned} &\underset{\boldsymbol{\beta},\beta_0}{\operatorname{argmin}} \frac{1}{2}||\boldsymbol{\beta}||_2^2 \\ &\text{subject to} \quad y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) \geq 1 \quad \forall i. \end{aligned} \tag{5.51}$$

The Lagrangian L_p in primal form, with multipliers λ_i (or $\boldsymbol{\lambda}$ as vector) is

$$\begin{aligned} L_p(\boldsymbol{\beta}, \beta_0, \boldsymbol{\lambda}) =& \frac{1}{2}||\boldsymbol{\beta}||_2^2 - \sum_{i=1}^{N} \lambda_i \left[y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) - 1 \right] \\ &\text{subject to} \quad \lambda_i \geq 0 \quad \forall i. \end{aligned} \tag{5.52}$$

The Karush-Kuhn-Tucker (KKT) conditions for optimality lead to further implications for the problem:

1. The primal constraints have to be fulfilled:

$$- y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) + 1 \leq 0 \quad \forall i \tag{5.53}$$

2. The dual constraints hold:

$$\lambda_i \geq 0 \quad \forall i \tag{5.54}$$

3. The gradient of the Lagrangian has to be zero:

$$\begin{aligned} \frac{\partial L_p(\boldsymbol{\beta}, \beta_0, \boldsymbol{\lambda})}{\partial \boldsymbol{\beta}} &\overset{!}{=} 0 \\ \rightarrow \boldsymbol{\beta} &= \sum_{i=1}^{N} \lambda_i y_i \boldsymbol{x}_i \end{aligned} \tag{5.55}$$

$$\begin{aligned} \frac{\partial L_p(\boldsymbol{\beta}, \beta_0, \boldsymbol{\lambda})}{\partial \beta_0} &\overset{!}{=} 0 \\ \rightarrow 0 &= \sum_{i=1}^{N} \lambda_i y_i \end{aligned} \tag{5.56}$$

4. The complementary slackness has to be fulfilled:

$$- \lambda_i \left[y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) - 1 \right] = 0 \quad \forall i \tag{5.57}$$

From the KKT conditions, some properties of the SVM are visible:

- If $\lambda_i > 0$, from Equation 5.57 follows:

$$y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) = 1 \tag{5.58}$$

  This means that the $\boldsymbol{x}_i$ is exactly at the boundary of the margin. These $\boldsymbol{x}_i$ are called *support vectors*, giving the classifier its name.

- Otherwise, if $y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) > 1$, the $\lambda_i = 0$. The points which are not on the boundary of the margin subsequently play no role in the definition of the hyperplane.

- From Equation 5.55 one can see that the parameters $\boldsymbol{\beta}$ of the optimally separating hyperplane is a linear combination of those training samples $\boldsymbol{x}_i$ for which $\lambda_i > 0$

Additionally, the conditions for the gradient of the Lagrangian (Equation 5.55 and 5.56) can be used to reformulate the optimization problem to its dual form L_d:

$$\begin{aligned} L_d = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j \\ \text{subject to} \quad \lambda_i, \lambda_j \geq 0 \quad \forall i, j \end{aligned} \tag{5.59}$$

The optimal values $\boldsymbol{\beta}^*, \beta_0^*$ that minimize L_d can be found by using quadratic programming techniques (e. g. Platt's Sequential Minimal Optimization [Plat 98]).
The decision function is then:

$$y^* = \operatorname{sgn}\left\{\boldsymbol{\beta}^{*\top} \boldsymbol{x} + \beta_0^*\right\} \tag{5.60}$$

5.6.2 Soft Margin SVM

Above formulations are to a large extent valid for linearly separable data only. However, in real-world scenarios, non-linearly separable problems are common. There exist two approaches in current implementations in order to make SVM usable even for those scenarios:
Cortes et al. [Cort 95] proposed the usage of soft margins. Consequently, a certain amount of misclassification is allowed. This can be achieved by introducing a slack variable $\boldsymbol{\xi}$ (with elements ξ_i) in Equation 5.51, which are weighted by a parameter C:

$$\begin{aligned} \underset{\boldsymbol{\beta}, \beta_0, \xi_i}{\operatorname{argmin}} \frac{1}{2} ||\boldsymbol{\beta}||_2^2 + C \sum_{i=1}^{N} \xi_i \\ \text{subject to} \quad y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) \geq 1 - \xi_i \\ \xi_i \geq 0 \quad \forall i. \end{aligned} \tag{5.61}$$

For soft margin SVMs, the parameter C controls the size of the margin. A higher value for C results in a smaller margin. For SVMs using mapping to higher dimensional spaces (Section 5.6.4), a higher value for C usually leads to decision boundaries which are better adapted (see Figure 5.7) but prone to overfit the training data. On the opposite, smaller values of C result in rather smooth, more general decision boundaries.

Analogously to the derivations for linearly separable data, the Lagrangian with multipliers $\boldsymbol{\lambda}, \boldsymbol{\mu}$ in primal form is:

$$\begin{aligned} L_p(\boldsymbol{\beta}, \beta_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \frac{1}{2}||\boldsymbol{\beta}||_2^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\lambda_i\left[y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) - (1-\xi_i)\right] - \sum_{i=1}^{N}\mu_i\xi_i \\ \text{subject to} \quad & \lambda_i \geq 0 \\ & \mu_i \geq 0 \\ & \xi_i \geq 0 \quad \forall i \end{aligned} \tag{5.62}$$

The KKT conditions for this Lagrangian are:

1. The primal constraints:

$$-y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) + (1-\xi_i) \leq 0 \quad \forall i \tag{5.63}$$

$$-\xi_i \leq 0 \quad \forall i \tag{5.64}$$

2. The dual constraints:

$$\lambda_i \geq 0 \quad \forall i \tag{5.65}$$

$$\mu_i \geq 0 \quad \forall i \tag{5.66}$$

3. Zero gradient of the Lagrangian:

$$\begin{aligned} \frac{\partial L_p(\boldsymbol{\beta}, \beta_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \boldsymbol{\beta}} &\stackrel{!}{=} 0 \\ \rightarrow \boldsymbol{\beta} &= \sum_{i=1}^{N}\lambda_i y_i \boldsymbol{x}_i \end{aligned} \tag{5.67}$$

$$\begin{aligned} \frac{\partial L_p(\boldsymbol{\beta}, \beta_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \beta_0} &\stackrel{!}{=} 0 \\ \rightarrow 0 &= \sum_{i=1}^{N}\lambda_i y_i \end{aligned} \tag{5.68}$$

$$\begin{aligned} \frac{\partial L_p(\boldsymbol{\beta}, \beta_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu})}{\partial \boldsymbol{\xi}} &\stackrel{!}{=} 0 \\ \rightarrow \lambda_i &= C - \mu_i \quad \forall i \end{aligned} \tag{5.69}$$

4. The complementary slackness:

$$-\lambda_i\left[y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0) - (1-\xi_i)\right] = 0 \quad \forall i \tag{5.70}$$

$$\mu_i\xi_i = 0 \quad \forall i \tag{5.71}$$

Some properties of the soft-margin SVM differ from the properties of the hard-margin version, such as:

- Among the support vectors ($\lambda_i > 0$), for points that lie exactly on the margin ($\xi_i = 0$), an upper bound for λ_i exists, as a consequence of Equations 5.71 and 5.69:

$$0 < \lambda_i < C \tag{5.72}$$

- On the contrary, for points with $\xi_i > 0$, from Equations 5.71 and 5.69 follows:

$$\lambda_i = C \tag{5.73}$$

Again, with help of the gradient conditions, L_p (Equation 5.62) can be reformulated to its dual form:

$$\begin{aligned} L_d &= \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j \\ \text{subject to} \quad & 0 \leq \lambda_i, \lambda_j \leq C \\ & \sum_{i=1}^{N} \lambda_i y_i = 0 \end{aligned} \tag{5.74}$$

The decision function is analogous to the hard margin case (Equation 5.60). One can reformulate this with the help of Equation 5.67 to:

$$y^* = \operatorname{sgn}(\sum_{i=1}^{N} \lambda_i y_i \boldsymbol{x}_i^\top \boldsymbol{x} + \beta_0) \tag{5.75}$$

We find, that the feature vectors only appear as scalar products in this function. Additionally, when comparing the dual forms (Equations 5.59 and 5.74) to their primal versions (Equations 5.52 and 5.62), one finds that in the dual form the feature vectors $\boldsymbol{x}$ only appear in scalar products as well. This is a favorable property to compute non-linear decision boundaries, which can be exploited, as shown Section 5.6.4.

5.6.3 Hinge Loss Function

From the complementary slackness for soft margin SVMs, defined by Equation 5.70, for any $\lambda_i > 0$ one finds

$$1 - y_i \underbrace{(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0)}_{= ||\boldsymbol{\beta}||_2 \cdot d(\boldsymbol{x}_i)} = \xi_i \quad , \tag{5.76}$$

where the term in brackets is $||\boldsymbol{\beta}||_2$-times the distance of the $\boldsymbol{x}_i$ to the separating hyperplane. Together with the primal constraints $\xi_i \geq 0$, this can be rewritten as hinge loss function $h(\boldsymbol{x}_i) \in \mathbb{R}_0^+$:

$$h(\boldsymbol{x}_i) = \max\left\{0,\, 1 - y_i(\boldsymbol{\beta}^\top \boldsymbol{x}_i + \beta_0)\right\} \tag{5.77}$$

The loss function is zero if the $\boldsymbol{x}_i$ is correctly classified ($d(\boldsymbol{x}_i)$ and y_i have the same sign) and the distance of the $\boldsymbol{x}_i$ to the hyperplane is greater $\frac{1}{||\boldsymbol{\beta}||_2}$. For all points that are closer to the separating hyperplane or which are classified incorrectly, the loss increases proportional to $d(\boldsymbol{x}_i)$.

5.6.4 Kernel Trick

In order to obtain linearly separable data, the features can be transformed to a higher dimensional space. This transformation of a feature vector $\boldsymbol{x} \in \mathbb{R}^d$ is denoted by $\phi(\boldsymbol{x}) \in \mathbb{R}^{d'}$ $(d' > d)$. The method is effective to fit highly non-linear decision boundaries in the feature space (see Figure 5.7).
Now we use the fact that the dual Lagrangian contains only scalar products of the feature vectors. These scalar products $\langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle$ for the transformed feature vectors can be represented by kernel functions $k(\boldsymbol{x}_i, \boldsymbol{x}_j) \in \mathbb{R}$:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \langle \phi(\boldsymbol{x}_i), \phi(\boldsymbol{x}_j) \rangle \tag{5.78}$$

With this, the Lagrangian in dual form 5.74 becomes

$$\begin{aligned} L_d &= \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j k(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ \text{subject to} \quad & 0 \leq \lambda_i, \lambda_j \leq C \\ & \sum_{i=1}^{N} \lambda_i y_i = 0 \end{aligned} \tag{5.79}$$

The decision function for a feature vector $\boldsymbol{x}$ is then

$$y^* = \operatorname{sgn}(\sum_{i=1}^{N} \lambda_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) + \beta_0) \tag{5.80}$$

We used two different kernel functions in our experiments, a Gaussian kernel and a polynomial kernel of degree two.

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\{-\gamma ||\boldsymbol{x} - \boldsymbol{x}'||_2^2\} \tag{5.81}$$

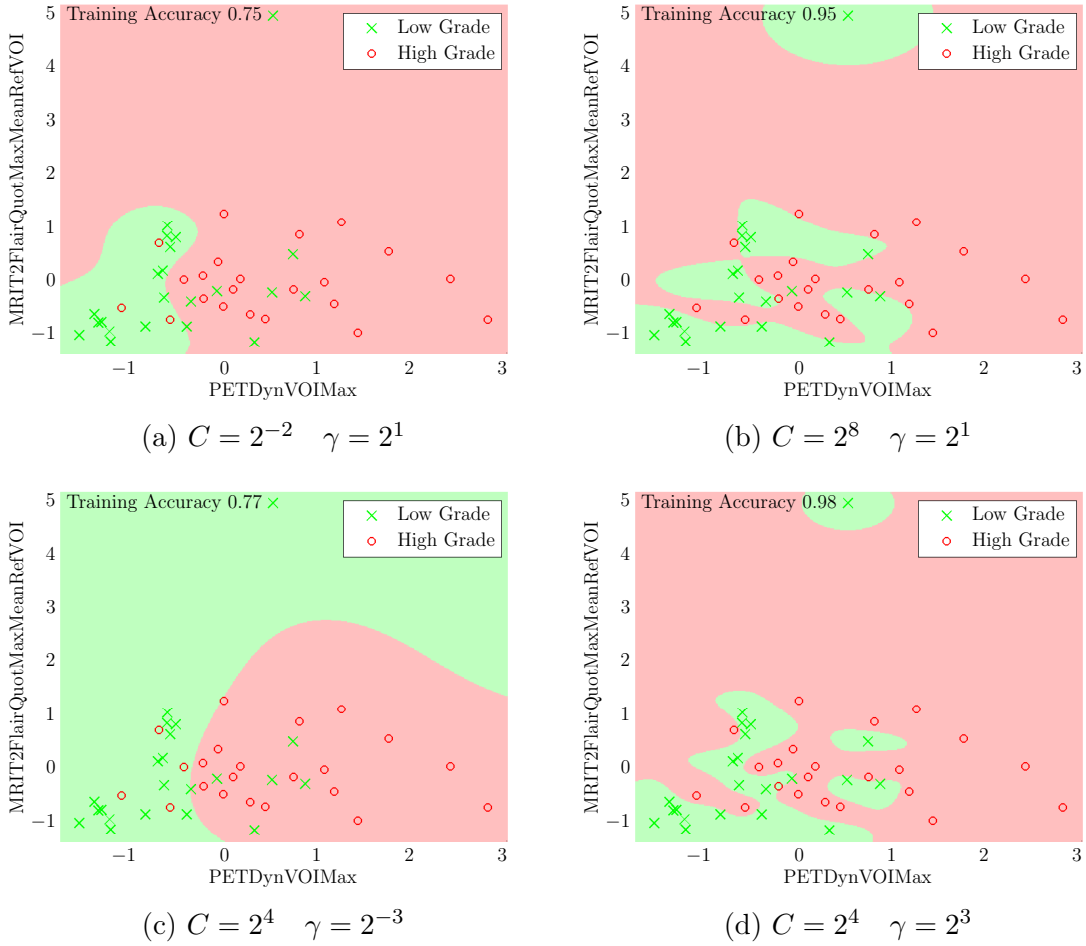
$$k(\boldsymbol{x}, \boldsymbol{x}') = (1 + \boldsymbol{x}^\top \boldsymbol{x}')^2 \tag{5.82}$$

5.6.5 Numerical Optimization

We used the libSVM software framework for the SVM classification [Chan 09]. There, the optimization problem posed in the Lagrangian dual (Equation 5.79) is solved by a sequential minimal optimization type decomposition: only a subset of the λ_i are modified in each iteration. The subset itself poses a smaller sub-problem that has to be minimized in each iteration. More details about this procedure can be found in [Chan 09]. How these subsets can be selected in an optimal way is described in Fan et al. [Fan 05].

5.6.6 Multi-Class SVM

The SVM as described above is able to handle two-class problems only. In the case of our 4-class classification, we used the one-versus-one classification scheme. We trained $4 \cdot (4-1)/2$ binary classifiers and decided by majority vote. Ties were assigned by random decision.



(a) $C = 2^{-2} \quad \gamma = 2^{1}$ (b) $C = 2^{8} \quad \gamma = 2^{1}$

(c) $C = 2^{4} \quad \gamma = 2^{-3}$ (d) $C = 2^{4} \quad \gamma = 2^{3}$

Figure 5.7: Decision boundaries of the C-Support Vector Classifier for different values of weight C and kernel width γ. One sees that C and γ influence the decision boundaries to a large extent. In general, larger values for C force the classifier to decision boundaries which are better adapted to the training data. However, the increased adaption might lead to over-fitting and to reduced generalizability for unknown test data. As in the previous examples for other classifiers, the data stem from patient subset SUB44 and a two-dimensional feature space and were normalized to unit variance and zero mean (LSUV). The training accuracy for this example was a) 75%, b) 95%, c) 77%, and d) 98%. The instances of SUB44 are colored in red if the underlying tumor was high-grade and colored in green if it was low-grade. Please compare to the other figures in this chapter for the same data classified with other classifiers.

5.7 Grid-Search for Best Parameters

For estimating the optimal values for the free parameters of the applied machine classifiers, we applied a grid search method. The cross validated accuracies (see Section 5.8) for all points on the grid were calculated. In the results section, usually only the highest accuracy corresponding to a single point on the grid is presented. Table 5.1 lists the free parameters of the different classifiers and the numerical range of each parameter.

Classifier	Parameter	Range	# Grid Points
AdaBoost	# of base classifiers M	$[5, 6, \ldots, 20]$	16
LDA	-	-	0
NaiveBayes	-	-	0
NearestNeighbor	# of neighbors k	$[1, 3, 5]$	3
NeuralNetworks	Learning rate γ	$[0.1, 0.3, \ldots, 0.9]$	
	Momentum rate κ	$[0.1, 0.3, \ldots, 0.9]$	
	# of epochs R	$[100, 200, \ldots, 1000]$	250
SVM (Poly)	Weight C	$[2^{-5}, 2^{-4}, \ldots, 2^{15}]$	21
SVM (RBF)	Weight C	$[2^{-5}, 2^{-4}, \ldots, 2^{15}]$	
	Kernel width γ	$[2^{-15}, 2^{-14}, \ldots, 2^{3}]$	399
		Total # grid points	689

Table 5.1: The numerical range for the free parameters of the machine classifiers. *Parameter* is the free parameter of the respective classifier. The *Range* specifies the numerical range, *# of Grid Points* is the total number of points on the grid for the respective free parameters.

5.8 Leave-One-Patient-Out Cross-Validation

In order to test the generalizability and to reduce a possible bias towards higher classification rates by over-fitting, all listed results of the machine classifiers were generated by performing a leave-one-patient-out cross-validation (LOPO-CV). The Figure 5.8 provides a diagram for our setup. The initial data was divided into training and test set. The training set was further divided into development and validation sets. The necessary statistics were calculated from the development set and the features from development, validation and test set were normalized with that information. Additionally, for the PCA features, the Eigenvectors were learned from the development set and the feature vectors from development, validation, and test set were transformed into the new coordinates. The optimal free parameters were then obtained be a grid search, in which the development set was used for training the classifier and the validation set for evaluating the classification accuracy. The grid search was five-fold cross-validated. The choice of parameters which obtained the best classification accuracy was then used for classifying the test set. The output of this algorithm is the LOPO-CV classification accuracy for every feature set. For

the classification of subsets SUB44 and SUB32 (those including T_1-MRI, T_2-MRI, FET-PET), we had 62 different feature sets. The algorithm consequently outputs one LOPO-CV classification accuracy for each of the 62 feature sets.

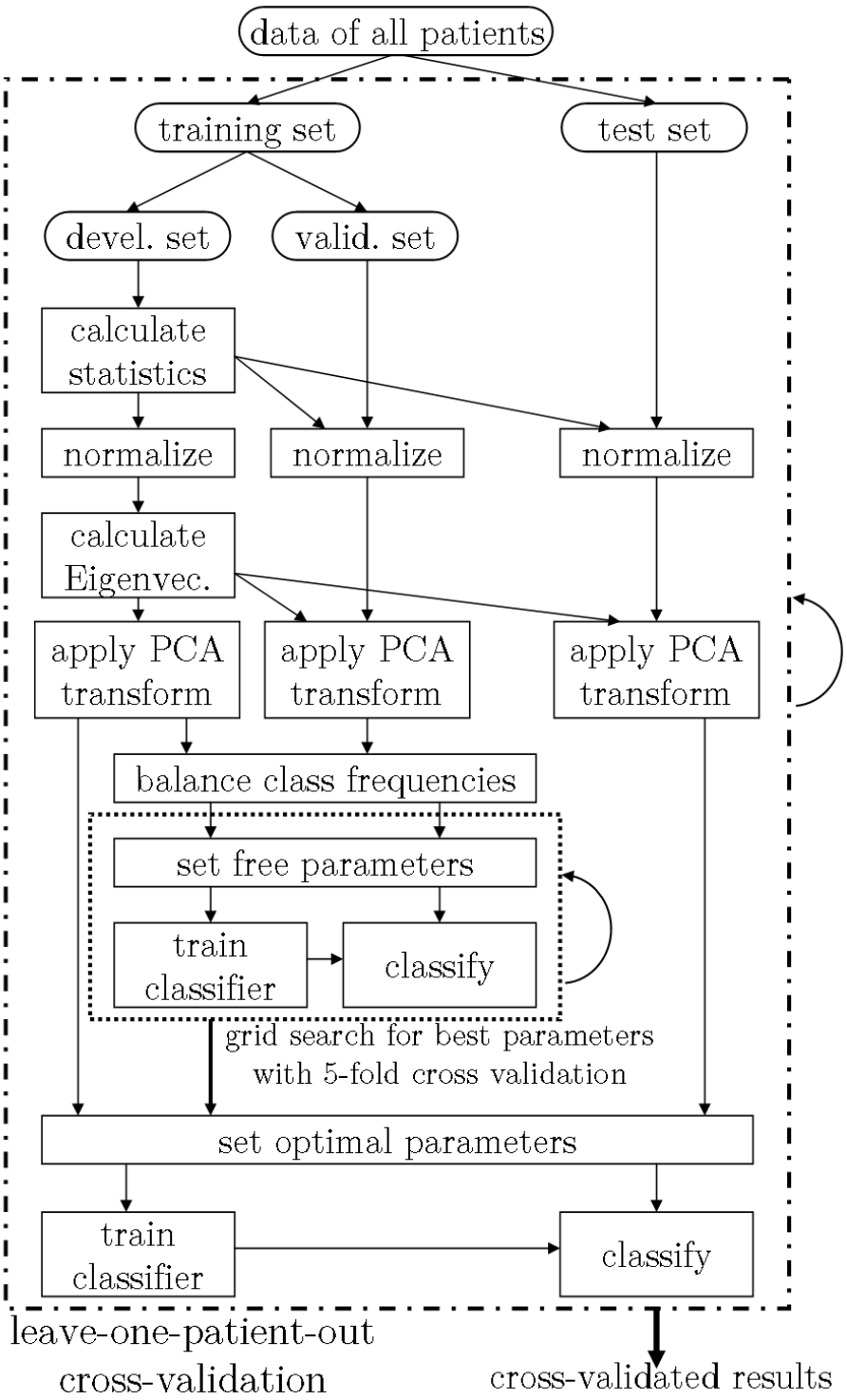


Figure 5.8: The setup of our experiments, incorporating a grid search for the best classifier parameters and a leave-one-patient-out cross validation. For the grid search, the training set was further subdivided into development and validation set. The parameters for the normalization and the PCA were learned from the development set only.

5.9 Receiver Operating Characteristic Analysis

In medicine, the receiver operating characteristic (ROC) curve is commonly employed to assess the significance of a feature for a binary classification task. In fact, many studies in the field of glioma grading involve ROC curves and the obtained parameter values are used in everyday medical routine [Popp 07, Popp 06, Calc 11]. In order to compare our methods to these results, we additionally carried out an ROC analysis for the task of differentiating between low-grade and high-grade gliomas.
The ROC analysis is applicable for multi-dimensional, scalar features and two-class classification. Commonly, it incorporates a diagram where sensitivity vs. $(1 -$ specificity) is plotted. The method or feature that results in the ROC curve with the largest area under the curve (AUC) is considered to be superior in terms of discriminative power when compared to the others. For reference, we always list the optimal parameter choice for our ROC curves. The optimal parameter value is the value with the highest Youden's index ($=$ sensitivity $+$ specificity -1). The Youden's index has its maximum when the accuracy is maximal. The accuracy is calculated by:

$$\text{accuracy} = \text{sensitivity} \cdot \text{prior probability} + \text{specificity} \cdot (1 - \text{prevalence}) \tag{5.83}$$

The prevalence for a two-class task with balanced groups, as in our case, is:

$$\text{prior probability} = 0.5$$

In opposite to the other machine classifiers in our study, the ROC analysis classifies retrospectively on all samples without any division into test and training sets.

5.10 Measures for Classification Performance

For evaluating the success of a specific classifier to classify samples, the results of the classification are often compared to a rating which is considered as ground truth (or gold standard). In our case this is the grading of the tumor based on the histological analysis of invasively gained bioptic samples. The output of the classifier is compared to this ground truth and multiple measures have been established for this comparison.

In order to present these measure in an illustrative way, we start with a confusion matrix for the classification, as seen in Figure 5.2. The elements $n_{j,i}$ in the matrix contain the absolute frequencies (number) of samples, which were classified as class i and have the underlying ground truth class j. The total number of classes is k. The total number of classified samples is $N = \sum_{j=1}^{k} \sum_{i=1}^{k} n_{j,i}$. The marginal frequencies are defined as $n_{j,*} = \sum_{i=1}^{k} n_{j,i}$ and $n_{*,i} = \sum_{j=1}^{k} n_{j,i}$.

5.10.1 Accuracy

We refer to classification accuracy ACC to the fraction of correctly classified examples, expressed as percentage. This measure is also known as total recognition rate. The

		Hypothesis				
		Class 1	Class i	...	Class k	
Reference	Class 1	$n_{1,1}$	$n_{1,i}$	...	$n_{1,k}$	$n_{1,*}$
	Class j	$n_{j,1}$	$n_{j,i}$	...	$n_{j,k}$	$n_{j,*}$
	⋮	⋮		⋱		⋮
	Class k	$n_{k,1}$	$n_{k,i}$	...	$n_{k,k}$	$n_{k,*}$
		$n_{*,1}$	$n_{*,i}$	...	$n_{*,k}$	N

Table 5.2: General confusion matrix of a multi-class classification with k different classes. The absolute frequencies $n_{j,i} \in \mathbb{N}$. The total number of samples is N. The marginal frequencies are calculated according to $n_{j,*} = \sum_{i=1}^{k} n_{j,i}$ and $n_{*,i} = \sum_{j=1}^{k} n_{j,i}$.

accuracy is the diagonal sum in the confusion matrix divided by the total number of samples:

$$\mathrm{ACC} = \frac{1}{N} \sum_{i=1}^{k} n_{i,i} \cdot 100\% \tag{5.84}$$

The accuracy as a measure to judge a classifier for suitability or to compare multiple classifiers has the limitation that it heavily depends on the prior probability of a certain class. A classifier that always decides for the class which has the largest prior probability $p_j = n_{j,*}/N$ can achieve high subjective classification rates for highly imbalanced classes. For this, other measures can be defined, e. g. the unweighted average recall UAR:

$$\mathrm{UAR} = \frac{1}{k} \sum_{j=1}^{k} \frac{n_{j,j}}{n_{j,*}} \cdot 100\% \tag{5.85}$$

The UAR does not depend on the prior probabilities and is equal to ACC when $n_{j,*} = \frac{N}{k} \quad \forall j$. However, as we deal with balanced or nearly balanced data, we do not expect this to be a significant source of bias for our experiments.

5.10.2 Cohen's Kappa

This measure was introduced by Cohen in 1960 [Cohe 60]. In its original context it was introduced for assessing the agreement of human raters. However, it can also be used for judging the classification performance of machine learning techniques. Cohen's κ is calculated on the basis of the observed agreement p_o of two raters and the by chance agreement p_c that is expected for statistically independent decisions of both raters:

$$\kappa = \frac{p_o - p_c}{1 - p_c} \tag{5.86}$$

The estimates for p_o is analogous to the accuracy:

$$p_o = \frac{1}{N} \sum_{i=1}^{k} n_{i,i} \tag{5.87}$$

p_c is estimated by an averaged product of the marginals:

$$p_c = \frac{1}{N^2} \sum_{i=1}^{k} n_{*,i}\, n_{i,*} \tag{5.88}$$

Cohen's Kappa features some useful properties. $\kappa = 1$ only if all off-diagonal elements in the confusion matrix are zero, which is induced by a perfect agreement between both raters. For other values of κ, several guidelines exist in literature. We follow the recommendation of Fleiss et al. [Flei 03], which propose the guidelines found in Table 5.3. In order to test if an obtained result yielding a certain κ-value is different from

κ	Strength of Agreement
$\kappa \leq 0.40$	Poor
$0.40 < \kappa \leq 0.75$	Fair to good
$0.75 < \kappa$	Excellent

Table 5.3: The strength of agreement of two independent raters depending on the magnitude of Cohen's Kappa according to Fleiss et al. [Flei 03].

the hypothesis that the two raters are independent ($\kappa = 0$) in a significant way, one needs to estimate the standard error $\mathrm{SE}(\kappa)$ of κ. Fleiss et al. [Flei 03, Flei 69] showed that this can be achieved by Equation 5.89:

$$\mathrm{SE}(\kappa) = \frac{1}{(1 - p_c)\sqrt{N}} \sqrt{p_c + p_c^2 - \frac{1}{N^3} \sum_{i=1}^{k} n_{*,i}\, n_{i,*} (n_{*,i} + n_{i,*})} \tag{5.89}$$

By assuming that κ's underlying distribution is normal, the according standard-score z is:

$$z = \frac{\kappa}{\mathrm{SE}(\kappa)} \tag{5.90}$$

For a significance level of $\alpha \leq 0.05$, which means that a classification result yielding a certain value κ has at most a 5% probability to be generated by chance, the critical value of z is $z \geq 1.645$ ($\alpha \leq 0.01$ for $z \geq 2.326$, $\alpha \leq 0.001$ for $z \geq 3.090$).
From this, we considered in our experiments a classification with an associated confusion table resulting in $\kappa \geq 0$ and a significance level $\alpha \leq 0.05$ to be significantly better than random prediction.

Chapter 6

Results

In this chapter, we aim to elaborate the differences in classification accuracy for feature normalization method, imaging modality, inter-dataset normalization method, PET motion correction and the different classifiers. All statistical analysis were carried out by help of the IBM SPSS Statistics [SPSS 10] package.

6.1 Feature Normalization Technique

The current section lists the classification results for the two different feature normalization techniques. These are linear scaling to range (LSR) and linear scaling to unit variance (LSUV) (for details, see Sections 4.7.1 and 4.7.2).
The Tables 6.1 (two-class) and 6.2 (four-class) show the maximum classification accuracies of the 62 feature sets for the different classifiers on patient subsets SUB44 (two-class) and SUB32 (four-class). All values in the tables reflect LOPO-CV results and were tested for statistical significance on the basis of Cohen's Kappa: The results that are not significantly better ($p < 0.05$) than random prediction are denoted in *italics*. The values of Cohen's Kappa are listed as well.
Additionally, the classification rates of the LSR and LSUV variants were compared by a two-sided Wilcoxon signed-rank test. In this context, one asterisk (*) indicates that the LSUV normalization method achieved significantly ($p < 0.05$) higher classification accuracy, two asterisks (**) indicate that the LSR method achieved significantly higher accuracy.
For reference, we also provide the results of the AdaBoost classifier, although, with decision stumps as weak classifier, it is insensitive to data scaling and translation. The last rows of the tables contain the mean, standard deviation and maximum of the respective column.

Automated Two-Class Classification

Classifier	LSR			LSUV		
	Max. Acc. [%]	Cohen's κ	Mod.	Max. Acc. [%]	Cohen's κ	Mod.
AdaBoost	79.6	0.59	PET	79.6	0.59	PET
LDA	72.7	0.46	MM	75.0	0.50	MM
NaiveBayes	72.7	0.46	MM	70.5	0.41	PET
NearestNeighbor	79.6*	0.59	PET	86.4*	0.73	MM
NeuralNetworks	79.6*	0.59	MM	86.4*	0.64	MM
SVM (Poly. Kernel)	81.8**	0.64	MM	75.0**	0.50	MM
SVM (RBF. Kernel)	84.1	0.68	PET	84.1	0.68	MM
Mean±SD (Max)	78.6 ± 4.3 (84.1)			79.6 ± 6.3 (86.4)		

Table 6.1: Maximum classification accuracies (*Max. Acc.*) for the two-class problem (low-grade/high-grade), separated by classifier and normalization technique. *Cohen's* κ was calculated according to Section 5.10.2. *Mod.* indicates the modality on which the feature set was calculated (*PET* = dynamic FET-PET with SUV inter-patient normalization, *MM* = Multimodality).

Automated Four-Class Classification

Classifier	LSR			LSUV		
	Max Acc. [%]	Cohen's κ	Mod.	Max Acc. [%]	Cohen's κ	Mod.
AdaBoost	65.6	0.54	T_2	65.6	0.54	T_2
LDA	56.3*	0.42	MM	62.5*	0.50	MM
NaiveBayes	56.3**	0.42	MM	50.0**	0.33	T_2
NearestNeighbor	59.4	0.46	MM	62.5	0.50	MM
NeuralNetworks	68.8	0.58	T_2	68.8	0.58	T_2
SVM (Poly. Kernel)	65.6	0.54	MM	62.5	0.50	MM
SVM (RBF. Kernel)	62.5	0.50	MM	62.5	0.50	MM
Mean±SD (Max)	62.0 ± 4.9 (68.8)			62.0 ± 5.8 (68.8)		

Table 6.2: Maximum classification accuracies (*Max. Acc.*) for the four-class problem (WHO1/WHO2/WHO3/WHO4), separated by classifier and normalization technique. *Cohen's* κ was calculated according to Section 5.10.2. *Mod.* indicates the modality on which the feature set was calculated (T_1 = contrasted T_1-MRI, T_2 = T_2-FLAIR-MRI, *MM* = Multimodality).

6.2 Imaging Modalities

This section focuses on the differentiation of the classification results for feature sets derived from the different modalities. Table 6.3 lists the results for the two-class problem and machine classifiers, Table 6.4 the results for the four-class problem and machine classifiers and Table 6.5 the results of the ROC analysis.
For all machine classifiers, the maximum classification accuracies of the 62 available features sets are listed, separated by imaging modality. As in the previous section, all values reflect LOPO-CV results and were tested for statistical significance with Cohen's Kappa: *Italics* denote non-significant results. The values of Cohen's Kappa are listed as well.
The last rows of the tables contain the mean, standard deviation and maximum of the respective column.
The classifications for feature sets from T_1-, T_2-MRI, PET and Multimodality (combining the three previously named) were carried out on patient subsets SUB44 (two-class) and SUB32 (four-class). The classification for ADC-MRI (SUB22, 22 patients) and X-ray CT (SUB14, 14 patients) was only done for two-classes, due to the limited dataset number in these groups.
The table for the ROC analysis contains three features for every modality that achieved the highest area under the curve (AUC). For reference, the threshold value representing the highest Youden's index for the separation of low- and high-grade tumors, the accuracy, the sensitivity and specificity using this threshold are listed as well. Every ROC curve was tested for significant differences ($p < 0.05$) of their AUCs to the null hypothesis AUC = 0.5. Non-significant results are denoted in *italics*. For PET, four ROC curves are given for the comparison in later sections.
The ROC curves itself can be found in Figures A.1 (T_1-MRI), A.2 (T_2-MRI) and A.3 (PET) in the Appendix.
Please note that the results for the PET are based on the SUV normalized datasets since this technique (Section 4.3) is the standard in nuclear medicine. The results of all other modalities base on the non-normalized variants.

Automated Two-Class Classification

Classifier	T_1		T_2		PET (SUV)	
	Max. Acc. [%]	Cohen's κ	Max. Acc. [%]	Cohen's κ	Max. Acc. [%]	Cohen's κ
AdaBoost	75.0	0.50	75.0	0.50	79.6	0.59
LDA	*56.8*	0.14	*59.1*	0.18	70.5	0.41
NaiveBayes	*63.6*	0.27	63.6	0.27	70.5	0.41
NearestNeighbor	65.9	0.32	65.9	0.32	79.6	0.59
NeuralNetworks	65.9	0.32	77.3	0.55	77.3	0.55
SVM (Poly.)	*61.4*	0.23	70.5	0.41	72.7	0.46
SVM (RBF.)	70.5	0.41	75.0	0.50	84.1	0.68
Mean±SD (Max)	65.6 ± 5.9 (75.0)		69.5 ± 6.8 (77.3)		76.3 ± 5.2 (84.1)	

Classifier	Multimodality		ADC		CT	
	Max. Acc. [%]	Cohen's κ	Max. Acc. [%]	Cohen's κ	Max. Acc. [%]	Cohen's κ
AdaBoost	68.2	0.36	81.8	0.64	71.4	0.46
LDA	75.0	0.50	*54.6*	0.10	64.3	0.32
NaiveBayes	72.7	0.46	*63.6*	0.27	64.3	0.32
NearestNeighbor	86.4	0.73	72.7	0.46	71.4	0.46
NeuralNetworks	86.4	0.73	86.4	0.73	64.3	0.32
SVM (Poly.)	81.8	0.64	72.7	0.46	71.4	0.46
SVM (RBF.)	84.1	0.68	86.4	0.73	78.6	0.59
Mean±SD (Max)	79.2 ± 7.2 (86.4)		74.0 ± 12.0 (86.4)		69.4 ± 5.4 (78.6)	

Table 6.3: Maximum classification accuracies *Max. Acc.* for the two-class problem (low-grade/high-grade), separated by classifier and imaging modality. T_1 = contrasted T_1-MRI, T_2 = T_2-FLAIR-MRI, *PET* = dynamic FET-PET using SUV inter-patient normalization, *Multimodality* = multimodal feature sets, *ADC* = ADC-MRI, *CT* = X-ray CT. *Cohen's* κ was calculated according to Section 5.10.2.

Automated Four-Class Classification

Classifier	T_1		T_2	
	Max. Acc. [%]	Cohen's κ	Max. Acc. [%]	Cohen's κ
AdaBoost	59.4	0.46	65.6	0.54
LDA	*37.5*	0.17	*34.4*	0.13
NaiveBayes	43.8	0.25	50.0	0.33
NearestNeighbor	53.1	0.38	53.1	0.38
NeuralNetworks	62.5	0.50	68.8	0.58
SVM (Poly.)	46.8	0.29	53.1	0.38
SVM (RBF.)	62.5	0.50	59.4	0.46
Mean±SD (Max)	52.5 ± 9.8 (62.5)		54.9 ± 11.4 (68.8)	
Classifier	PET (SUV)		Multimodality	
	Max. Acc. [%]	Cohen's κ	Max. Acc. [%]	Cohen's κ
AdaBoost	53.1	0.38	53.1	0.38
LDA	43.8	0.25	62.5	0.50
NaiveBayes	46.9	0.29	56.3	0.42
NearestNeighbor	53.1	0.38	62.5	0.50
NeuralNetworks	59.4	0.46	65.6	0.54
SVM (Poly.)	*37.5*	0.17	65.6	0.54
SVM (RBF.)	56.3	0.42	62.5	0.50
Mean±SD (Max)	50.0 ± 7.7 (59.4)		61.2 ± 4.7 (65.6)	

Table 6.4: Maximum classification accuracy *Max. Acc.* for the four-class problem (WHO1/WHO2/WHO3/WHO4), separated by classifier and imaging modality: T_1 = contrasted T_1-MRI, T_2 = T_2-FLAIR-MRI, *PET* = dynamic FET-PET using SUV inter-patient normalization, *Multimodality* = multimodal feature sets. *Cohen's* κ was calculated according to Section 5.10.2.

ROC Analysis

Contrasted T_1-MRI					
Feature #	AUC	Optimal Threshold	Accuracy	Sens.	Spec.
10	*0.61*	1.23	63.6	59.1	68.2
12	*0.59*	1.94	65.9	72.7	59.1
16	*0.59*	0.45	61.4	45.5	77.3
T_2-FLAIR-MRI					
Feature #	AUC	Optimal Threshold	Accuracy	Sens.	Spec.
10	*0.66*	3.15	68.1	100	36.4
9	*0.62*	0.35	65.9	100	31.8
3	*0.62*	684	65.9	50.0	81.8
ADC-MRI					
Feature #	AUC	Optimal Threshold	Accuracy	Sens.	Spec.
14	*0.65*	9.47	77.3	90.9	63.6
8	*0.60*	2.3E − 4	68.1	90.9	45.5
18	*0.58*	1.16E4	72.7	100	45.5
X-ray CT					
Feature #	AUC	Optimal Threshold	Accuracy	Sens.	Spec.
3	*0.74*	509	71.4	71.4	71.4
13	*0.67*	−0.31	71.4	42.9	100
18	*0.67*	659	71.4	57.1	85.7
PET (SUV)					
Feature #	AUC	Optimal Threshold	Accuracy	Sens.	Spec.
5	0.82	2.72	77.3	90.9	63.6
8	0.80	3.30	77.3	81.8	72.7
10	0.80	4.16	75.0	63.6	86.4
4	0.74	1.19	75.0	77.3	72.7

Table 6.5: Maximum area under the curve *AUC* values for the two-class problem (low-grade/ high-grade), separated by imaging modality. The *Optimal Threshold* (achieving highest accuracy) is listed in units of the feature. *Sensitivity* and *Specificity* are in %. *Feature Index* denotes the index of feature set. The name and type of the feature set can be found in Tables A.3, A.4 and for indices 51-62 in Table A.5.

6.3 Effect of Inter-Dataset Normalization

The current section lists the change in classification results for the inter-dataset normalization techniques. Please refer to Section 4.3 for the normalization methods.
As in the previous sections, for machine classifiers, Table 6.6 lists the maximum accuracies for the different feature sets, separated by modality. All results reflect LOPO-CV values and were tested for statistical significance ($p < 0.05$) with Cohen's Kappa: *Italics* denote non-significant results. The values of Cohen's Kappa are listed as well.
Additionally, the accuracy differences to the non-normalized variants are given. These differences were tested for significance ($p < 0.05$) with the help of a two-sided Wilcoxon signed-rank test. One asterisk (*) denotes results for which the normalized variant obtained significantly higher classification rates.
The last rows of the tables contain the mean, standard deviation and maximum of the respective column.
The classifications for feature sets from T_1-, T_2-MRI, PET and Multimodality were carried out on patient subsets SUB44 and SUB22 for ADC-MRI.
For the ROC analysis, as in the previous sections, Table 6.7 contains three features for every modality that achieved the highest AUC. For reference, the threshold value representing the highest Youden's index for the separation of low- and high-grade tumors, the accuracy, the sensitivity and specificity using this threshold are listed as well. Every ROC curve was tested for significant differences ($p < 0.05$) of their AUCs to the null hypothesis AUC = 0.5. Non-significant results are denoted in *italics*.
Additionally, the difference to the AUC of non inter-dataset normalized results is given.
Please note that the PET results are based on datasets without SUV inter-patient normalization. For the comparison to normalized PET, only feature sets were taken into account which do not rely on quotients to the reference region, as normalized and non-normalized variants would achieve the same results due to the underlying methodology of the SUV.

Automated Two-Class Classification

Classifier	T_1-MRI Normalized			T_2-FLAIR-MRI Normalized		
	Max. Acc. [%]	Cohen's κ	Diff. [%]	Max. Acc. [%]	Cohen's κ	Diff. [%]
AdaBoost	77.3	0.55	+2.3	90.9	0.82	+15.9*
LDA	*54.6*	0.10	−2.2	*61.4*	0.23	+2.3
NaiveBayes	63.6	0.28	±0.0	65.9	0.32	+2.3
NearestNeighbor	70.5	0.41	+4.6	77.3	0.55	+11.4*
NeuralNetworks	65.9	0.32	±0.0	65.9	0.32	−11.4
SVM (Poly.)	70.5	0.41	+9.1*	65.9	0.32	−4.6
SVM (RBF.)	72.7	0.46	+2.2	79.5	0.59	+4.6
Mean±SD (Max)	67.9 ± 7.4 (77.3)		+2.3	72.4 ± 10.5 (90.9)		+2.9
Classifier	FET-PET			ADC-MRI Normalized		
	Max. Acc. [%]	Cohen's κ	Diff. [%]	Max. Acc. [%]	Cohen's κ	Diff. [%]
AdaBoost	77.3	0.55	−2.3	72.7	0.46	−9.1
LDA	77.3*	0.55	+6.8*	63.6	0.27	+9.1*
NaiveBayes	75.0	0.50	+4.5	63.6	0.27	±0.0
NearestNeighbor	77.3	0.55	±0.0	72.7	0.46	±0.0
NeuralNetworks	79.6	0.59	+2.3	81.8	0.64	−4.6
SVM (Poly.)	72.7	0.46	±0.0	72.7	0.36	±0.0
SVM (RBF.)	77.3	0.55	−6.8	81.8	0.64	−4.6
Mean±SD (Max)	76.6 ± 2.1 (79.5)		+0.3	72.7 ± 7.4 (81.8)		−1.3
Classifier	Multimodality Normalized					
	Max. Acc. [%]	Cohen's κ	Diff. [%]			
AdaBoost	79.6	0.59	+11.4*			
LDA	77.3	0.55	+2.3			
NaiveBayes	79.6	0.59	+6.9*			
NearestNeighbor	77.3	0.55	−9.1			
NeuralNetworks	84.1	0.68	−2.3			
SVM (Poly.)	84.1	0.68	2.3			
SVM (RBF.)	86.4	0.73	2.3			
Mean±SD (Max)	81.2 ± 3.6 (86.4)		+2.0			

Table 6.6: Maximum classification accuracies for the two-class problem using inter-dataset normalization. The values for PET reflect non-normalized values. *Diff.*: Accuracy differences of inter-dataset normalized and non-normalized results. *Cohen's* κ was calculated according to Section 5.10.2.

ROC Analysis

T_1-MRI Normalized						
Feature	AUC	Diff.	Optimal Threshold	Accuracy	Sens.	Spec.
9	*0.65*	+0.04	0.18	63.6	72.7	54.5
10	*0.62*	+0.03	1.32	63.6	59.1	68.2
14	*0.62*	+0.03	8.85E − 3	63.6	72.7	54.5
T_2-FLAIR-MRI Normalized						
Feature	AUC	Diff.	Optimal Threshold	Accuracy	Sens.	Spec.
10	0.69	+0.03	3.12	65.9	100	31.8
11	*0.62*	±0.0	1.26	61.4	90.9	31.8
9	*0.58*	−0.04	0.87	61.4	63.6	22.7
ADC-MRI Normalized						
Feature	AUC	Diff.	Optimal Threshold	Accuracy	Sens.	Spec.
16	*0.65*	±0.0	1.32E − 1	63.6	81.8	45.5
15	*0.59*	−0.01	9.15	68.1	90.9	45.5
18	*0.56*	−0.02	7.91E3	68.1	100	36.4
FET-PET						
Feature	AUC	Diff.	Optimal Threshold	Accuracy	Sens.	Spec.
5	0.81	−0.01	8.95E3	79.5	77.3	81.8
8	0.79	−0.01	1.03E4	79.5	77.3	81.8
4	0.74	±0.0	3.43E3	77.3	77.3	54.5

Table 6.7: Maximum *AUC* values for the two-class problem, separated by imaging modality, using inter-dataset normalization. The values for PET reflect non-normalized values. *Diff.*: Accuracy differences of inter-dataset normalized and non-normalized results. The *Optimal Threshold* (achieving highest accuracy) is listed in units of the feature. *Sensitivity* and *Specificity* are in %. *Feature Index* denotes the index of feature set, for reference see Tables A.3 and A.4.

6.4 Effect of PET Motion Correction by Rigid Registration

The current section focuses on the influence of the PET motion correction on the classification results. Please refer to Section 4.5 for the motion correction method itself.
As in the previous sections, Table 6.8 lists the the maximum classification accuracy of all available feature sets for the motion corrected and SUV normalized PET data (patient subset SUB44). Again, all values were LOPO cross-validated and tested for statistical significance ($p < 0.05$) on basis of Cohen's Kappa: All results were significantly better than random prediction. The values of Cohen's Kappa are listed as well.
Additionally, the accuracy differences to the non-motion corrected variant are given. These differences were tested for significance ($p < 0.05$) with the help of a two-sided Wilcoxon signed-rank test. We did not find any significant differences between non-motion and motion corrected classification results.
For the ROC analysis, as in the previous sections, Table 6.9 contains the three features that achieved the highest (AUC) for motion corrected PET. For reference, the threshold value representing the highest Youden's index for the separation of low- and high-grade tumors, the accuracy, the sensitivity and specificity using this threshold are listed as well. Every ROC curve was tested for significant differences ($p < 0.05$) of their AUCs to the null hypothesis AUC = 0.5. All AUCs differed significantly from the null hypothesis.
Additionally, the difference to the AUC of the non-motion corrected results is given.

Automated Two-Class Classification

Classifier	PET Motion Corr. (SUV)		
	Max. Acc. [%]	Cohen's κ	Diff. [%]
AdaBoost	79.6	0.59	±0.0
LDA	75.0	0.50	+4.5
NaiveBayes	77.3	0.55	+6.8
NearestNeighbor	75.0	0.50	−4.6
NeuralNetworks	77.3	0.55	±0.0
SVM (Poly.)	72.7	0.46	±0.0
SVM (RBF.)	77.3	0.55	−6.8
Mean±SD (Max)	76.3 ± 2.2 (79.6)		±0.0

Table 6.8: Maximum classification accuracy *Max. Acc.* for the two-class problem for motion corrected and SUV normalized PET, separated by classifier. The column *Diff.* lists the differences in classification accuracy of motion corrected and non-motion corrected PET. *Cohen's* κ was calculated according to Section 5.10.2.

ROC Analysis

SUV Normalized, Motion Corrected PET						
Feature	AUC	Diff.	Optimal Threshold	Accuracy	Sens.	Spec.
5	0.84	+0.02	2.36	77.3	90.9	63.6
8	0.80	±0.0	3.13	77.3	81.8	72.7
10	0.80	±0.0	4.19	77.3	59.1	95.5

Table 6.9: Maximum *AUC* values for the two-class problem separated by classifier, for motion corrected and SUV inter-dataset normalized PET. The *Optimal Threshold* (achieving highest accuracy) is listed in units of the feature. *Sensitivity* and *Specificity* are in %. *Diff.* contains the AUC difference of motion corrected and non-motion corrected PET. *Feature* denotes the index of the feature set. The name and type of the feature set can be found in Tables A.3, A.4 and for indices 51-62 in Table A.5.

6.5 Choice of Classifier

The current section aims to elaborate the suitability of the different machine classifiers for the task of glioma classification.
For this reason, Table 6.10 lists the overview of the classifiers' performance for the patient subset SUB44 and two-class classification. Table 6.11 contains the results for patient subsets SUB32 and four-class classification.
In both tables, the classifiers are sorted by the mean rank of a Friedman test that was carried out on the classification rates of all 62 feature sets. One asterisk (*) indicates that the classifier achieved significantly higher ($p < 0.05$) classification rates compared to its directly preceding, lower-ranked classifier in a two-sided Wilcoxon signed-rank test. Additionally, the maximum accuracies, the mean accuracies and the standard deviations of the mean accuracies of classification accuracies of the 62 feature sets for the respective classifier values are listed.

Two-Class Classification

Classifier	Max. Acc. [%]	Mean Acc. ± SD. [%]	Mean Rank
LDA	75.0	42.3 ± 22.3	2.19
NaiveBayes	72.7	49.6 ± 15.8	2.95*
SVM (Poly.)	81.8	41.7 ± 26.5	2.96*
NearestNeighbor	86.4	57.2 ± 9.8	4.10*
AdaBoost	79.6	59.4 ± 9.2	4.81*
NeuralNetworks	86.4	59.8 ± 10.0	4.98
SVM (RBF.)	84.1	63.4 ± 7.4	6.00*

Table 6.10: Maximum (*Max. Acc.*), mean (*Mean Acc.*) and standard deviation (*SD.*) of classification accuracies of all 62 feature sets for two classes, separated by classifier. *Mean Rank* denotes the rank of the respective classifier in a Friedman test. A higher rank reflects higher classification accuracies. An asterisk (*) indicates that the classifier achieved significantly higher ($p < 0.05$) classification rates compared to its directly preceding, lower-ranked classifier, using a two-sided Wilcoxon signed-rank test.

Four-Class Classification

Classifier	Max. Acc. [%]	Mean Acc. ± SD. [%]	Mean Rank
LDA	62.5	22.1 ± 15.1	2.13
SVM (Poly.)	65.6	19.7 ± 17.6	2.15
NaiveBayes	56.3	29.4 ± 11.1	3.13*
NearestNeighbor	62.5	37.6 ± 10.6	4.37*
AdaBoost	65.6	40.6 ± 10.9	4.99*
SVM (RBF.)	62.5	41.1 ± 12.1	5.16
NeuralNetworks	68.8	45.2 ± 11.1	6.07*

Table 6.11: Maximum (*Max. Acc.*), mean (*Mean Acc.*) and standard deviation (*SD.*) of classification accuracies of all 62 feature sets for four-classes, separated by classifier. *Mean Rank* denotes the rank of the respective classifier in a Friedman test. A higher rank reflects higher classification accuracies. An asterisk (*) indicates that the classifier achieved significantly higher ($p < 0.05$) classification rates compared to its directly preceding, lower-ranked classifier, using a two-sided Wilcoxon signed-rank test.

6.6 Classification Accuracy of Medical Reports

For the majority of the patients in our experiments, several medical reports were available. In the ideal case, three reports would be available, one for the MRI, one for the PET and the third for the biopsy/surgery. The latter provides the ground truth diagnosis in the form of the results of the histology. Consequently, the diagnosis of the MRI and PET reports can be compared to the histology and by this, one can find an approximate classification accuracy of the medical reports. The MRI and PET reports were created earlier than the histological reports and therefore are not biased towards the ground truth. The written reports were available to us in 16 of 48 (33%) patients in the case of MRI and for 48 of 48 (100%) patients in the case of PET. The limited number of MRI reports was due to the fact that the primary purpose of the MRI images was the co-registration of the PET or simply that the report was not available as the MRIs were performed at some remote clinic practice. Additionally, we found that it was very uncommon for the medical reports to contain the naming of the exact glioma grade (WHO1-WHO4). Still, a differentiation between low- and high-grade gliomas was made in 13 (27%) of the MRI reports and 44 (94%) of the PET reports. Due to the low case number for MRI, we do not report a classification accuracy for this modality.
Accordingly, we calculated the confusion matrix for SUB44 (two-class problem) and PET (see Figure 6.12), based on the medical reports: **We found a classification accuracy of 75.6%**. Please note that for three low-grade patients of SUB44, no evidence of a glioma was reported in PET, although later a glioma was evidenced by histology. We decided to exclude these patients from the results in this section, which explains the imbalance between low- and high- tumor grade in the confusion matrix.

Classified as ->	Low-grade	High-grade
Low-grade	26.8	19.5
High-grade	4.9	48.8

Table 6.12: Confusion matrix for the glioma classification based on the PET reports. Ground truth classification was provided by histology of the tumors. This corresponds to a correct classification rate of 75.6%. All values in the confusion matrix are in %.

Chapter 7

Discussion

7.1 Influence of Feature Normalization Technique

When comparing the mean performance of the two feature normalization methods, linear scaling to range and linear scaling to unit variance, we found only minor differences. The mean classification accuracies of all classifiers on the two-class problem were 78.6% for LSR and 79.6% for LSUV. Even in the four-class problem, the means from the two scaling approaches did not differ (58.5%).
Significant differences arise when testing the classification results of all feature sets for the two scaling methods: For two classes, the SVM with polynomial kernel classifies significantly better with LSR and the neural networks classify significantly better with LSUV. These minor differences lead to a higher maximum classification accuracy for LSUV at two classes (86.4% vs. 84.1%). For four classes, the LSR normalization results in significantly better classification accuracies for the Naive Bayes classifier only. The maximum achieved classification rates for four classes are equal for LSR and LSUV (both 68.8%).
Besides the presented two approaches, other feature normalization methods exist: These are e. g. fitting to an appropriate distribution. For this method, a sufficient number of subjects in the patient population is needed as the underlying feature distributions have to be analyzed. Our patient population currently is limited in this regard, however, if a higher subject number is available, one should reconsider this method. Also, we did not test normalization methods which distribute the features evenly over a certain range, like rank normalization. In fact, Aksoy et al. [Akso 00] report that they found better classification results with rank normalization, compared to LSR and LSUV.
In general, when comparing the numeric range of the feature normalization techniques (Figures 4.19 and 4.19), we found only little differences between LSR and LSUV. As a consequence and in line with our expectations, both scaling methods performed equally well and we found only little differences in the classification accuracy. Nevertheless, for some classifiers differences existed and we decided to include both normalization methods in our further analyses.

7.2 Accuracy of Classification for Different Imaging Modalities

It was indicated from the literature that the modalities that we used provide different amounts of information for revealing the underlying tumor grade. For analyzing this, we evaluated the modality wise classification rates. We found distinct differences in the classification rates of the imaging modalities, for two as well as for four tumor classes.

Two Classes
Table 6.3 lists the means and the maxima of the achieved classification accuracies over all classifiers: We found the multimodal approach (86.4%) to be superior to PET (84.1%) and T_2- (77.3%), and T_1-MRI (75%). In the following, our results are compared to the state of the art from the literature:
Our MRI only based classification rates are higher than those reported by Haegler et al. [Haeg 12]. They performed a manual grading by consensus of two medical experts and achieved an accuracy of 64.9% on 37 patients with pre- and post-contrast T_1-, T_2-FLAIR-, and proton density MRI. In a study of Riemann et al. [Riem 02], a relatively high classification accuracy of 88% is reported for a visual assessment on pre- and post-contrast T_1- and T_2-MRI images . Their high accuracy for separating low and high tumor grades might be explained by the lack of WHO3 tumors in their patient population (24 WHO1+WHO2, 0 WHO3, 24 WHO4). In a large study (160 patients) Law et al. [Law 03] report an accuracy of 70.6% on grading by consensus of two medical experts on pre- and post-contrast T_1- and T_2-FLAIR-MRI. In a study with machine classifiers on MRI data, Zacharaki et al. achieved 87.8% accuracy for 98 patients [Zach 09] and reported 94.5% accuracy for the same two-class problem in a later publication [Zach 11]. There were some differences in comparison to our studies: First of all, our patient population is more heterogeneous with regard to the imaging parameters, e. g. one MRI system and three sequence variants vs. six systems and 28 sequence variants in our setup. Additionally, their approach relied on four different manually defined VOIs covering various regions of the same tumor (enhancing, non-enhancing, necrotic, edematous), as opposed to only 1 VOI covering the whole tumor in our case. This might introduce a significant amount of prior knowledge when compared to our approach. Li et al. [Li 06] report on two classes an accuracy of 89% with an SVM classifier for 154 patients. On the contrary to our study, their features were manually extracted, as medical experts had to rate e. g. the amount and heterogeneity of contrast agent, the amount of hemorrhage or the amount of necrosis.
For PET, various reports for the accuracy of differentiation between low- and high-grade gliomas exist: Pöpperl et al. report a rather high accuracy of 96% for a ROC analysis (AUC 0.967) based on dynamic PET of 54 patients [Popp 06]. So far, other groups were unable to reproduce these results with their patient populations even if using the same methods. Our lower classification results might also be due to the fact that the PET images for our patient population stem from two different PET scanners using three different image reconstruction methods, compared to 1 system and 1 reconstruction method. Even though the calibration methods that we applied

should have prevented significant inter-device deviations, these methods might not be sufficient. A recent study on a large patient population (n=143) of Rapp et al. [Rapp 13] achieved 74% accuracy for a ROC analysis (AUC 0.77) on dynamic PET data, which is slightly less compared to our results.
When comparing the machine classifiers to the accuracy of the medical records, we achieved a absolute classification rate that is 10-15% higher for the machine classifiers.
Even though the heterogeneity of study setups and patient populations make it hard to compare the achieved classification rates, the results from our own methods and those from the literature emphasize that PET is superior to MRI modalities in differentiating low-grade and high-grade tumors.
For the multimodal approach, we find that a combination of features from PET and MRI leads to higher classification rates than features from single modalities. To the best of our knowledge, no systematic study in literature exists on the classification accuracy of the exactly same multimodal features. On a similar note, Floeth et al. [Floe 05] assess increased diagnostic abilities of the multimodal (PET+MRI) approach over single modalities for the differentiation of brain tumors and non-neoplastic lesions. Similar results were reported from Pauleit et al. [Paul 05], which confirm an increased accuracy of the multimodal approach.

Four Classes
Compared to the two-class problem, the order for the achieved mean classification rates of the four-class task differed: T_1-MRI: 52.5%, T_2-MRI: 54.9%, PET: 50.0%, Multimodality: 61.2%. PET was no longer superior compared to the MRI modalities. The classification rates of the single modalities were very similar to each other. Still, we found multimodal features to achieve better classification rates. Our highest classification rate for differentiating the four WHO grades was 68.8%.
To the best of our knowledge, results in literature for differentiating each of the four WHO grades on basis of medical imaging are still lacking. For reference, we report results on similar topics: Zacharaki et al. [Zach 09] achieved an accuracy of 63% for differentiating WHO2, WHO3, WHO4, and metastases of other tumors on 98 patients. They derived features from four manually defined VOIs from MRI and classified using support vector machines. In a later publication of the same group [Zach 11], they were able to improve their classification rates on the same task to 76.3% by feature selection with a wrapper and the best-first search strategy. When comparing to our study, some differences arise: they used unbalanced classes with 22 WHO2, 18 WHO3, 34 WHO4 and 24 cases of metastases of other tumors, whereas we had a smaller patient population with balanced classes. Additionally, the previously listed (see *Two Classes*) differences still apply: They had a more homogenous study setup in terms of imaging modalities and probably introduced more prior knowledge by defining manual VOIs covering multiple tumor aspects.

ADC-MRI and CT
Our results for diffusion-based (ADC) MRI and X-ray CT were derived from smaller subsets of our patient collective (ADC-MRI: 22 patients, CT: 14 patients) when compared to T_1-MRI, T_2-MRI and PET (44 patients). ADC-MRI achieved a highest

classification rate of 86.4%, CT of 78.6%. Due to the limited patient number, results for four class classification were not calculated and the accuracies should not be compared to the classification rates of the other modalities.
Nevertheless, our results indicate that the classification rates for these modalities were significantly better than random prediction. An information of the tumor grade can be derived from diffusion MRI and X-ray CT. Especially the high classification rates for ADC-MRI indicate the potential of this modality.
Our findings are supported by several studies from literature. For diffusion MRI, Arvinda et al. [Arvi 09] report a ROC analysis with an accuracy of 88.2% on 51 patients with low- and high-grade gliomas. They achieve even better results (94% accuracy) for the classification based on the relative cerebral blood volume (rCBV), which was not available for our patients. On the same topic, Hilario et al. [Hila 12] reached an accuracy of 88.8% for a ROC analysis of ADC features and 85.6% for rCBV features of 162 patients.

ROC Analysis
The main purpose of the ROC analysis was to provide numbers for reference and for estimating how "good" our patient population can be classified when compared to other populations from literature, where results are often reported in the form of ROC.
For all modalities, the linear decision boundary (threshold) of the ROC is not sufficient to achieve a significant classification rate. The only exceptions from this are features derived from PET. All of the listed PET features with the highest AUC have a classification performance that is significantly better than random prediction.
In order to agree with the results from literature, we used the ROC analysis on one dimensional features only, which might explain why the classification rates of the ROC are lower than those of the machine classifiers.
In general, our results for ROC analyses are poor, especially when taking into account the substantial ROC accuracies reported from literature. As most of these studies feature very homogenous study setups and imaging data, this indicates that our more realistic, yet more heterogeneous patient database inhibited higher classification accuracies.

7.3 Effect of Inter-Dataset Normalization on Classification Accuracy and Class Overlap

Physicians deal with very heterogeneous image data which often stem from a variety of MRI, PET and CT scanners. Especially in MRI, an inter-device and inter-patient normalization is still not an established method. For this reason, we examined the influence of simple normalization methods like histogram matching and SUV technique on the distribution of pixel intensities. As seen in Figure 4.5, the variances of the mean gray-level values are reduced for all modalities and across all subgroups (low-grade, high-grade, reference region). This indicates that the applied methods successfully reduced the intra-class and inter-patient variance. However, it was not clear how this affected the achieved classification results for the differentiation of low-

and high-grade tumors. The changes in classification accuracy were as follows:
We achieved a modest increase in the mean accuracy over all classifiers for T_1- (+2.3%) and T_2-MRI (+2.9%). This leaded to a maximum accuracy of 90.9% for inter-dataset normalized T_2-MRI features. The mean classification accuracy for multimodal features from MRI and PET was increased as well by +2.0%. It was superior to the single modalities, as in the case of non-normalized datasets.
Not all modalities showed increased classification performance. For ADC-MRI, the histogram matching decreased the accuracy by -1.3% on average, which might be due to the very limited number of patients in that subgroup.
In general, we found inconsistencies in the changes of classification accuracy when comparing results of normalized and non-normalized data for some classifiers. This might indicate that the patient population was to low for achieving more stable results.
The results for the ROC analysis on normalized data pointed in the same direction as the machine classifiers. The area under the curve was increased for T_1- and T_2-MRI and decreased for ADC-MRI when compared to their non-normalized counterparts. However, this difference was not significant based on the 95% confidence intervals of the AUC.
In PET, normalization techniques like scanner-calibration to radioactive reference sources (inter-device normalization) and the SUV technique (inter-patient normalization) are well established and commonly applied. For our study, these methods did not change the classification rates significantly (mean accuracy normalized PET/ PET, 76.3%/76.6%) for all but one classifier. For the LDA classifier, the SUV normalized results showed a significantly lower maximum classification rate (70.5%) than the non-normalized version (77.3%).
The comparison of the ROC analysis leaded to ambivalent results. The AUC for the SUV-normalized PET was slightly higher than for the non-normalized PET, but the accuracy at the optimal separating point of the ROC-curve was lower for the SUV-normalized PET. Again, those difference did not fulfill the required significance level.

In conclusion, the normalization methods that we applied yielded an improvement in the classification performance. The high variance in the achieved improvements by inter-dataset normalization might indicate an insufficient amount of patient data and probably prevents more significant results. Furthermore, it is not clear how the more complex features like PCA features, texture features or contextual features are influenced by that normalization techniques. More work should be carried out on this topic, e. g. by implementing other normalization techniques like non-rigid histogram registration [Jage 09] and examining the effects of normalization on other features.

7.4 Effect of PET Motion Correction

A noticeable part of the dynamic PET datasets of our study was affected by patient motion ($\sim 93\%$). As already shown in Section 4.5, we were able to reduce the extent of motion in those datasets in significant manner by a rigid registration of the different time frame images. However, it was unclear if this reduction in patient motion translates to better classification results.
We did not find any significant differences for the classification rates for our two-

class problem. The mean classification rates of all machine classifiers was 76.3% for motion-corrected as well as for uncorrected PET images. Similarly, the ROC analysis did not reveal significant differences.
Consequently, the features that we used for the classification in PET might be stable to a certain amount of motion, especially considering the rather large size of the tumor VOIs (mean value $\sim$ 43 ml). Additionally, since it is based on reconstructed images, we are not able to correct for intra-frame motion with our method. The heads of the patients in our study were fixed with the help of proper cushioning. Other setups that do not apply proper immobilization might suffer from an increased amount of motion. For these studies, the presented correction method could lead to improvements that affect classification accuracy as well.
Our results showed that the motion correction helped to increase the visual appearance of dynamic PET images significantly. To our knowledge, studies employing motion correction for cerebral gliomas are lacking so far. On a similar note, Ikari et al. [Ikar 12] report on the influence of their motion correction method on FDG-PET of 172 patients suspected of Alzheimer's disease. They found very little differences for larger ($>$ 20 ml) VOIs and increased influence for small VOIs. In a small study (phantom experiment and three patients), Zanotti-Fregonara et al. [Zano 12] elaborate that the effect of motion and motion correction for the kinetic modeling in C-11-(R)-rolipram was below 10%, which they consider clinically irrelevant for this application. In contrast, Mourik et al. [Mour 09] conclude from simulations and data of 6 patients that motion correction "...provides major improvements in accuracy of pharmacokinetic analyses over non-motion-corrected data...".

7.5 Optimal Choice of the Machine Classifier

Depending upon the calculated metric (mean, max, rank) the ranking of the classifiers changes. When ranked according to the Friedman test, the SVM classifier with RBF kernel (mean acc. 63.4%), the neural network (mean acc. 59.8%) and adaptive boosting (mean acc. 59.4%) achieve the best classification results for the two-class problem. The other machine classifiers perform worse. The performance of the LDA classifier resembled those of the ROC analysis.
For four classes, when ranked according to the Friedman test, the neural network (mean acc. 45.2%), the SVM with RBF kernel (mean acc. 41.1%) and the AdaBoost classifier (mean acc. 40.6%) perform best.
Ranking the classifiers by the maximum classification rates changes their sequence. The neural networks (max acc. 86.4%), the nearest neighbor classifier (max acc. 86.4%) and SVM with RBF kernel (max acc. 84.1%) achieve the highest classification rates for two classes.
For four classes, the neural networks (max acc. 68.8%), the AdaBoost classifier (max acc. 65.6%) and the SVM with polynomial kernel (max acc. 65.6%) achieve the highest rates.
In general, we found that classifiers which are able to learn more complex decision boundaries perform better than classifiers that provide only simple boundaries. This indicates that similar tumor grades could lie rather in disconnected clusters in the

feature space.
Similar results are reported from literature: Zacharaki et al. [Zach 09] compared LDA, nearest neighbor, and SVM (RBF) classifiers in a similar study than ours and concluded that the SVM with radial basis function is superior in terms of maximum classification accuracy. In their later study [Zach 11], they found that the nearest neighbor classifier was superior for two classes and the voting feature intervals classifier was superior for four classes. Those results are in line with ours if assessing the classifiers by the maximum classification accuracies. The nearest neighbor classifier is ranked on top for two classes and falls back for four classes. In a report by Devos et al. [Devo 05], the SVM with RBF kernel was superior (AUC: 0.87 vs. 0.74) compared to LDA for differentiating the low- and high-grade tumors of 21 patients.
From our results and the literature, we recommend the application of either neural networks or SVM with RBF kernel for classification tasks similar to our problem. Especially SVM classifiers are easily applied and promise high classification results.

Classified as ->	WHO1	WHO2	WHO3	WHO4
WHO1	21.9	0.0	3.1	0.0
WHO2	0.0	18.6	3.1	3.1
WHO3	0.0	6.3	15.6	3.1
WHO4	0.0	3.1	9.4	12.5

Table 7.1: Example confusion matrix for four-class problem. The matrix corresponds to a classification accuracy of 68.8% and was obtained by neural networks classifying on T_2-MRI features. All values are in %.

7.6 Limitations of Our Study

One limitation of our study was the lesser number of patient datasets which we had available for analysis. The requirements for this kind of study are high: No prior treatment is allowed, a bioptically confirmed tumor is mandatory and dynamic PET and various MRI sequences need to be acquired. For our patient population, gathering the data was a process of several years. Nevertheless, studies with larger patient populations in all subgroups would be beneficial. A higher patient number than ours is especially necessary for the differentiation of the four WHO grades and the evaluation of the potential benefit of diffusion weighted MRI and X-ray CT.
The achieved maximum classification performance of 90.9% for the differentiation of low and high tumor grades and 68.8% for the four WHO grades was limited. We found a considerable amount of misclassification between the groups WHO2/WHO3 and WHO3/WHO4 (Figure 7.1).
Furthermore, our simple inter-dataset normalization and feature selection and feature reduction methods might have led to suboptimal results. Consequently, more studies should be carried out in order to test more advanced normalization and feature selection methods and to investigate the clinical relevance of the limited classification results.

When comparing the classification results based on medical image data to the tumor grades diagnosed by histology, one has to keep in mind that histological grades represent the tumor in the exact area where the biopsy was taken. It is well known that gliomas should be seen as heterogeneous formations which incorporate regions with varying tumor grade. Additionally, the inter-rater variability in the process of histological grading is not negligible. It is reported to be as high as 20% for glial tumors [Bent 10]. This variability could be reduced by a consensus agreement of independent histopathologic examinations of multiple experts.

Chapter 8

Summary and Outlook

8.1 Summary

We presented methodologies and techniques that show that a machine classification of cerebral gliomas is feasible. In general, our classification results were in line with those provided in literature and $10 - 15\%$ higher than those of the written medical reports, which is noteworthy when one takes into account our heterogeneous study setup which reflects true clinical applicability.
As expected, the two-class differentiation of low and high tumor grades achieved better classification results compared to the differentiation of the four individual WHO grades. Nevertheless, our results underline that the differentiation of the four WHO grades is feasible within certain limitations. Especially the WHO grades 3/4 and 2/3 had a significant overlap in our study.
When it comes to the value of the single modalities in differentiating between low and high-grade tumors, PET using FET as a tracer offers the most valuable information among the available modalities. This benefit was lost for the four-class differentiation, where MRI offered better classification accuracies.
In general, the multimodal approach with combined features from MRI and PET offered the best overall classification results. This is comparable with the clinical routine where examinations from nuclear medicine (PET) and radiology (MRI) are performed separately but interpreted in conjunction.
Within the limitations of our study, namely a lower number of datasets, diffusion-weighted MRI and CT offer potentially beneficial information on the tumor grade.
We showed, that the presented inter-dataset normalization techniques improved the classification accuracy for most modalities. The highest improvements were achieved through MRI normalization. The significance of the PET calibration and SUV normalization for the classification were ambiguous.
In our patient population, the application of motion correction by registration of subsequent reconstructed frames did not lead to an improved classification accuracy. This might be due to the fact that the features are stable to the amount of motion that we encountered in our study. Nevertheless, the presented motion correction technique improved the visual appearance of the PET images in significant manner and might be of importance for studies with more patient motion.
In general, for two classes as well as for four classes, the C-SVM classifier with radial

basis function or neural networks are recommended. For complex problems in this field (class number > 2), the AdaBoost classifier is a promising alternative to the aforementioned methods. It features an easy application with only 1 free parameter and fast computation.

8.2 Future Work and Outlook

With the advent of combined PET and, MRI scanners [Jude 08], the incorporation of multimodal features for the differentiation between tumor types and grades will become easier. Not only is the number of examinations reduced, but also the synergistic effects of the system can be exploited. The motion correction of the PET acquisition by the help of MRI is seen as one of the hot topics in the field of MR-PET [Grav 13, Ouya 13]. Furthermore, additional work should be carried out for incorporating diffusion-, perfusion-weighted, and spectroscopic MRI into a multimodal classification scheme. The accuracy for the differentiation of tumor grades from our own work as well as from literature [Bepp 11] promise improved diagnostic abilities of such a setup, especially when using MRI spectroscopy. A combined MR-PET scanner could facilitate such acquisitions [Neun 12]. Dynamic FET-PET is acquired over a lengthy time period anyways, which leaves plenty of time for sophisticated MRI sequences beyond simple T_1-, or T_2- weighting.
We see a lot of potential in the field of the MRI inter-dataset normalization. It should be tested if other methods like the one proposed by Jäger et al. [Jage 09] can help to achieve better normalization results and eventually lead to higher classification results.
Results from literature indicate [Zach 09, Zach 11] that more sophisticated feature reduction and selection approaches, like recursive feature elimination, sequential forward selection, could lead to improved classification accuracy and clarify the varying importance of the individual features on the classification process.
When comparing to the literature, we noticed a high heterogeneity in study setups and patient databases. A freely-accessible glioma database featuring imaging data from PET and MRI should be established.
The field of computer aided diagnoses (CAD), which we consider this current work to be a part of, is emerging as it proves its value and abilities in a growing field of applications. CAD will not replace the need for well-educated physicians but support their work by gathering and summarizing information from multimodal imaging data, helping to prevent false diagnoses, and reducing time-consuming procedures.

Appendix A

Additional Tables and Figures

Patient #	Sex	Age [a]	Weight [kg]	Tumor WHO Grade	Tumor Entity	Time betw. PET & MRI [d]	Time betw. PET & Biopsy [d]	Inj. FET Activity [MBq]
1	M	65.7	90	4	GBM	2	8	207
2	F	63.4	74	3	AO	0	59	198
3	M	46.2	76	2	DA	1	31	345
4	M	22.3	71	1	GG	11	156	152
5	F	41.8	70	4	GBM	11	35	213
6	M	30.1	95	1	GG	49	142	190
7	F	35.9	69	3	AOA	49	17	274
8	M	18.1	85	1	GG	35	113	240
9	F	23.9	59	1	PA	1	2	178
10	F	54.6	73	1	GG	77	7	193
11	M	38.6	73	1	PGT	70	188	199
12	M	45.1	86	1	DNT	70	247	205
13	M	63.6	71	3	AOA	14	95	148
14	F	68.6	68	4	GBM	1	10	205
15	F	22.6	56	1	GG	1	67	167
16	F	72.5	100	4	GBM	38	2	275
17	F	29.9	72	2	DA	2	2	185
18	F	41.0	77	2	DA	0	6	200
19	M	24.3	65	2	DA	16	7	219
20	M	29.9	72	3	AA	0	19	215
21	M	44.3	82	4	GBM	0	1	239
22	M	12.4	45	3	AA	14	4	110
23	M	49.9	94	4	GBM	23	30	259
24	M	54.8	105	3	AA	0	5	278
25	M	54.2	80	3	AA	0	13	236
26	F	41.4	62	2	DA	0	13	179
27	M	33.6	87	3	AA	0	7	180
28	M	27.6	98	2	DA	0	140	238
29	F	46.9	91	2	DA	40	112	250
30	F	35.6	85	2	DA	69	31	175
31	F	66.1	55	2	O	77	36	209
32	F	48.8	68	2	DA	6	21	240

Patient #	Sex	Age [a]	Weight [kg]	Tumor WHO Grade	Tumor Entity	Time betw. PET & MRI [d]	Time betw. PET & Biopsy [d]	Inj. FET Activity [MBq]
33	M	25.9	75	2	DA	1	2	211
34	M	39.2	80	2	DA	0	7	220
35	M	41.6	93	2	O	41	61	250
36	F	49.1	59	3	AA	7	11	183
37	M	52.7	77	3	AA	7	74	212
38	M	44.6	102	2	DA	8	22	256
39	F	66.3	90	3	AO	0	52	256
40	F	23.1	60	2	DA	0	37	217
41	F	67.6	73	4	GBM	4	2	215
42	M	41.7	89	3	AO	0	14	243
43	F	67.7	70	2	DA	30	31	218
44	M	54.1	75	2	DA	0	6	235
45	F	43.4	59	3	AO	7	84	232
46	F	32.3	54	3	AO	7	14	214
47	M	45.9	85	4	GBM	0	13	238
48	M	55.4	77	3	A	56	51	243
	22 F 26 M	Mean: 43.8 Min: 12.4 Max: 72.5	Mean: 77 Min: 45 Max: 105	8/17/15/8 I/II/III/IV		Mean: 17.6 Min: 0 Max: 77	Mean: 43.9 Min: 1 Max: 247	Mean: 218 Min: 110 Max: 345

Table A.1: Overview on relevant parameters of the patient population in our study. Abbreviations: AA: Anaplastic astrocytoma; DA: Diffuse Astrocytoma; PA: Pilocytic Astrocytoma; O: Oligodendroglioma; AO: Anaplastic oligodendroglioma; AOA: Anaplastic oligoastrocytoma; GG: Ganglioglioma; GBM: Glioblastoma; PGT: Papillary glioneural tumor; DNT: Dysembryoplastic neuroepithelial tumor.

Patient #	T_2 Flair Seq. #	Contrast. T_1 Seq. #	ADC Seq. #	PET Protocol #	CT	SUB44	SUB32	SUB22	SUB14
1	14	3	24	1	√	√	√	√	√
2	14	3	24	1	√	-	-	√	√
3	14	3	24	1	√	√	√	√	√
4	14	3	24	1	√	√	√	√	√
5	15	2	23	1	√	√	√	√	√
6	14	3	24	1	√	√	√	√	√
7	14	3	24	1	√	√	√	√	√
8	14	3	24	1	√	√	√	√	√
9	15	2	23	1	√	√	√	√	√
10	14	3	24	1	√	√	√	√	√
11	14	3	-	1	√	√	√	-	√
12	15	2	-	1	-	√	√	-	-
13	14	3	24	1	√	√	√	√	√
14	14	1	24	1	√	√	√	√	√
15	12	2	-	1	√	√	√	-	√
16	14	3	24	1	-	√	√	√	-
17	16	6	25	3	-	-	-	√	-
18	17	5	-	3	-	√	-	-	-
19	16	6	25	2	-	√	-	√	-
20	17	5	-	3	-	√	-	-	-
21	17	5	-	3	-	√	√	-	-
22	16	6	25	2	-	√	√	√	-
23	18	7	-	3	-	√	√	-	-
24	19	8	-	3	-	√	√	-	-
25	17	5	-	3	-	√	√	-	-
26	17	5	-	3	-	√	√	-	-
27	17	5	-	3	-	√	√	-	-
28	17	5	-	3	-	√	√	-	-
29	20	9	-	3	-	√	√	-	-
30	20	9	-	3	-	√	√	-	-
31	17	5	26	3	-	-	-	√	-
32	13	4	27	3	-	√	√	√	-
33	16	6	-	3	-	√	√	-	-
34	17	5	-	3	-	√	√	-	-
35	20	9	28	3	-	-	-	√	-
36	21	10	-	2	-	√	√	-	-
37	20	9	-	3	-	√	√	-	-
38	22	11	-	3	-	√	-	-	-
39	17	5	-	3	-	√	-	-	-
40	17	5	-	3	-	√	-	-	-
41	16	6	25	4	-	√	√	√	-
42	17	5	-	4	-	√	-	-	-
43	16	6	-	4	-	√	-	-	-
44	17	5	-	4	-	√	-	-	-
45	13	4	27	4	-	√	-	√	-
46	20	9	-	4	-	√	-	-	-

Patient #	T_2 Flair Seq. #	Contrast. T_1 Seq. #	ADC Seq. #	PET Protocol #	CT	SUB44	SUB32	SUB22	SUB14
47	17	5	-	4	-	√	√	-	-
48	20	9	28	4	-	√	-	√	-

Table A.2: List of the assignment of various MRI, PET and CT sequences to the individuals in our patient population. The parameters of the according MRI sequences can be found in Table A.7. The parameters of the PET acquisition protocols can be found in Table A.8. "-" denotes that the dataset was not acquired/ not available or was not used in our study. "√" indicates that this patient underwent a CT examination with the parameters described in Section 3.2.1, which was used in the study.

#	Feature	Type of Feature
1	Mean Intensity Value	Statistical
2	Minimum Intensity Value	Statistical
3	Maximum Intensity Value	Statistical
4	95%-Quantile of Intensity (PDF)	Statistical
5	Position of Peak in Ref.VOI in Intensity (PDF)	Statistical
6	Height of Peak in Ref.VOI in Intensity (PDF)	Statistical
7	Proportion of Iso-intense Voxels	Contextual
8	Proportion of Hypo-intense Voxels	Contextual
9	Proportion of Hyper-intense Voxels	Contextual
10	Effective KL Divergence (PDF)	Contextual
11	Quotient Mean Prim. to Mean Ref. VOI	Contextual
12	Quotient Maximum Prim. to Mean Ref. VOI	Contextual
13	Quotient Minimum Prim. to Mean Ref. VOI	Contextual
14	Texture Energy	Texture
15	Texture Entropy	Texture
16	Texture Local Homogeneity	Texture
17	Texture Inertia	Texture
18	Texture Cluster Shade	Texture
19	Texture Cluster Prominence	Texture

Table A.3: List of the extracted features for every static (T_2 FLAIR MRI, T_1 MRI, ADC MRI, and CT) dataset. In Section 4.6 in detail information about the calculation of every feature can be found.

#	Feature	Type of Feature
1	Slope of Mean Intensity Value Curve	Statistical
2	Slope of Maximum Intensity Value Curve	Statistical
3	Slope of Minimum Intensity Value Curve	Statistical
4	Intercept of Mean Intensity Value Curve	Statistical
5	Intercept of Maximum Intensity Value Curve	Statistical
6	Intercept of Minimum Intensity Value Curve	Statistical
7	Mean Intensity Value (Dynamic)	Statistical
8	Maximum Intensity Value (Dyn)	Statistical
9	Minimum Intensity Value (Dyn)	Statistical
10	Quotient Maximum Prim. to Mean Ref. VOI (Dyn)	Contextual
11	Quotient Mean Prim. to Mean Ref. VOI (Dyn)	Contextual
12	Quotient Minimum Prim. to Mean Ref. VOI (Dyn)	Contextual

Table A.4: List of the extracted features for the dynamic PET dataset. In Section 4.6 in detail information about the calculation of every feature can be found.

Feature (-set) #	Contained features in the set
1-50	Individual features
51	All features of all modalities
52	All features obtained from the T_2-Flair MRI
53	All features obtained from the contrasted T_1 MRI
54	All features obtained from the dynamic PET
55	All features obtained from the MRI (T_1+T_2)
56	All texture features
57	Transformed feature of the first PCA component
58	Transformed features of the PCA components 1-3
59	Transformed features of the PCA components 1-5
60	Transformed features of the PCA components 1-7
61	Transformed features of the PCA components 1-10
62	Transformed features of the PCA components 1-15

Table A.5: List of the *Features* and *Feature sets* for T_1-MRI, T_2-MRI, and FET-PET. An explanation which features are included in each set is given as well. The features 1-50 correspond to the features listed in Table A.3 for T_2-FLAIR MRI and contrasted T_1-MRI (2×19 features) plus the features listed in Table A.4 for the dynamic PET (12 features).

Feature (-set) #	Contained features in the set
1-19	Individual features
20	All features of the modality
21	All texture features

Table A.6: List of the *Features* and *Feature sets* for ADC-MRI and CT. An explanation which features are included in each set is given as well. The features 1-19 correspond to the features listed in Table A.3 for ADC-MRI and X-ray CT.

#	Type	Sequence	Variant	Options	TR [ms]	TE [ms]	TI [ms]	FA [°]	System	Voxel Size [mm]
1	T_1 contrast	IR,GR	SP,MP	-	2090	3.93	1100	15	Siemens Sonata (1.5T)	$1.2 \times 1.1 \times 1.1$
2	T_1 contrast	GR	SP,OSP	-	300	2.42	-	76	Siemens TrioTim (3.0T)	$0.6 \times 0.6 \times 3.0$
3	T_1 contrast	SE	SP	-	690	17	-	70	Siemens Sonata (1.5T)	$0.5 \times 0.5 \times 5.0$
4	T_1 contrast	IR,GR	SP,MP	-	1900	4.38	1100	15	Siemens Sonata (1.5T)	$0.5 \times 0.5 \times 1.0$
5	T_1 contrast	IR,GR	SP,MP	-	2200	3.93	1200	15	Siemens Avanto (1.5T)	$0.9 \times 0.9 \times 1.5$
6	T_1 contrast	IR,GR	SP,MP,OSP	IR,PFP	1950	2.2	900	9	Siemens TrioTim (3.0T)	$0.5 \times 0.5 \times 1.0$
7	T_1 contrast	SE	SS	-	440	10	-	69	Philips Intera (1.5T)	$0.5 \times 0.5 \times 5.0$
8	T_1 contrast	SE	-	-	509	14	-	90	Philips Gyroscan NT (1.0T)	$0.9 \times 0.9 \times 6.0$
9	T_1 contrast	SE	SP	-	515	10	-	90	Siemens Symphony (1.5T)	$0.9 \times 6.0 \times 0.9$
10	T_1 contrast	GR	-	-	145	1.7	-	80	Philips Intera (1.5T)	$0.4 \times 0.4 \times 6.0$
11	T_1 contrast	SE	SS	-	509	14	-	90	Philips Intera (1.5T)	$0.4 \times 0.4 \times 6.0$
12	T_2 FLAIR	SE,IR	SK,SP,MP	IR,PFP	5000	389	1800	120	Siemens TrioTim (3.0T)	$1.0 \times 0.5 \times 0.5$
13	T_2 FLAIR	SE,IR	SK,SP,MP,OSP	IF,FS	9000	124	2500	150	Siemens Sonata (1.5T)	$0.4 \times 0.4 \times 5.0$
14	T_2 FLAIR	SE,IR	SK,SP,MP,OSP	-	10000	109	2500	150	Siemens Sonata (1.5T)	$0.5 \times 0.5 \times 5.0$
15	T_2 FLAIR	SE,IR	SK,SP,MP	IR,FS	9000	79	2500	150	Siemens TrioTim (3.0T)	$0.4 \times 0.4 \times 3.0$
16	T_2 FLAIR	SE,IR	SK,SP,MP,OSP	IR,FS	9000	130	2500	150	Siemens TrioTim (3.0T)	$0.9 \times 5.0 \times 0.9$
17	T_2 FLAIR	SE,IR	SK,SP,MP,OSP	-	9000	109	2500	150	Siemens Avanto (1.5T)	$0.4 \times 0.4 \times 5.0$
18	T_2 FLAIR	IR	SK	-	5000	100	1900	90	Philips Intera (1.5T)	$0.9 \times 0.9 \times 5.0$
19	T_2 FLAIR	IR	-	-	5000	100	1900	90	Philips Gyroscan NT (1.0T)	$1.0 \times 6.0 \times 1.0$
20	T_2 FLAIR	IR,SE	SK,SP,MP,OSP	-	8690	110	2400	170	Siemens Symphony (1.5T)	$0.9 \times 0.9 \times 6.0$
21	T_2 FLAIR	IR	-	-	6000	100	2000	90	Philips Intera (1.5T)	$0.9 \times 0.9 \times 6.0$
22	T_2 FLAIR	IR	SK	-	6000	150	1950	90	Philips Intera (1.5T)	$0.4 \times 0.4 \times 6.0$
23	ADC	EP	SK,SP	PFP,FS	3800	91	-	90	Siemens TrioTim (3.0T)	$1.2 \times 1.2 \times 5.0$
24	ADC	EP	SK,SP	PFP,FS	3600	98	-	90	Siemens Sonata (1.5T)	$1.8 \times 1.8 \times 6.0$
25	ADC	EP	SK,SP	PFP,FS	3200	92	-	90	Siemens TrioTim (3.0T)	$0.6 \times 0.6 \times 5.0$
26	ADC	EP	SK,SP	PFP,FS	3300	101	-	90	Siemens Avanto (1.5T)	$0.6 \times 0.6 \times 5.0$
27	ADC	SE,EP	SK,SP	-	3300	96	-	90	Siemens Sonata (1.5T)	$1.8 \times 1.8 \times 5.0$
28	ADC	SE,EP	SK,SP	-	3900	94	-	90	Siemens Symphony (1.5T)	$1.8 \times 1.8 \times 5.0$

Table A.7: Overview of the different MRI sequences that were used for the acquisitions of our patient population. Acronyms: FA: FLip Angle, FLAIR: Fluid-attenuation inversion-recovery, ADC: Apparent diffusion coefficient, IR: Inversion recovery, GR: Gradient echo, SE: Spin echo, EP: Echo planar imaging, SP: Spoiled , MP: Magnetization Preparation, OSP: Over sampling phase , MTC: Magnetization Transfer Contrast, SK: Segmented k-space , SS: Steady state , PFP: Partial Fourier space , FS: Fat saturation. In detail explanations of the aforementioned techniques can be found in [Reis 08].

#	Matrix [Voxel]	Voxel Size [mm]	Recon. Algorithm	I #	S #	Filter FWHM [mm]	Acq. Time	Frames	System (All Siemens)
1	$168 \times 168 \times 109$	$2.0 \times 2.0 \times 2.0$	OSEM2D	6	8	Gauss 5	40’	12 (5×1’,7×5’)	Biograph 64
2	$128 \times 128 \times 63$	$2.0 \times 2.0 \times 2.4$	FBP	6	8	Shepp-Logan 2.5	40’	14 (5×1’,5×3’,4×5’)	ECAT EXACT HR+
3	$128 \times 128 \times 63$	$2.0 \times 2.0 \times 2.4$	FBP	6	8	Shepp-Logan 2.5	50’	16 (5×1’,5×3’,6×5’)	ECAT EXACT HR+
4	$128 \times 128 \times 63$	$2.0 \times 2.0 \times 2.4$	OSEM3D	6	16	-	50’	16 (5×1’,5×3’,6×5’)	ECAT EXACT HR+

Table A.8: List of the parameters of the different PET acquisitions. Acronyms: I: Number of iterations; S: Number of subsets; OSEM2D: Ordered subset expectation maximization with 2-dimensional beam modeling; OSEM3D: as before, but with 3-dimensional beam modeling; FBP: filtered back projection with the filter specified in the field *Filter*.

Patient #	Uncorrected	Corrected	Corrected Better?
1	1	1	1
2	2	1	1
3	2	2	−1
4	2	1	1
5	0	0	0
6	0	0	0
7	1	1	1
8	2	2	−1
9	2	1	1
10	1	1	1
11	2	2	1
12	1	1	−1
13	2	1	1
14	2	0	1
15	2	1	1
16	3	1	1
17	1	1	−1
18	2	2	1
19	1	1	−1
20	2	2	1
21	2	2	0
22	2	2	1
23	1	1	0
24	1	1	0
25	1	1	1
26	1	1	0
27	2	1	1
	Mean: 1.52	Mean: 1.15	Better/Equal/Worse 16/6/5

Table A.9: Subjective rating of the amount of motion in PET images before and after the application of motion correction. In general, the study was carried out blinded, for the last column the images were shown side-by-side (unblinded). For the first two columns, amount of motion: 0 = not visible, 1 = small, 2 = distinct, 3 = substantial. Last column, larger amount of motion in: −1 = corrected image, 0 = both images equal, 1 = uncorrected image.

SUB44 Component #	Loading	Feature Name
1	+0.31	MRIT2FlairQuantile
2	+0.30	MRIT2FlairMax
3	+0.30	MRIT2FlairMean
4	+0.27	MRIT1EnhancedRefPeakPosition
5	+0.27	MRIT2FlairRefPeakPosition
6	+0.26	MRIT1EnhancedMean
7	+0.25	MRIT1EnhancedQuantile
8	+0.25	MRIT2FlairMin
9	+0.24	MRIT1EnhancedMax
10	−0.21	MRIT1EnhancedRefPeakHeight
SUB32 Component #	Loading	Feature Name
1	+0.28	MRIT1EnhancedMax
2	+0.28	MRIT1EnhancedQuantile
3	+0.26	MRIT1EnhancedMean
4	+0.25	MRIT2FlairMax
5	+0.24	MRIT2FlairQuantile
6	+0.23	MRIT1EnhancedRefPeakPosition
7	+0.23	MRIT2FlairMean
8	−0.22	MRIT1EnhancedRefPeakHeight
9	+0.20	MRIT2FlairRefPeakPosition
10	+0.18	PETDynVOIMax

Table A.10: The ten largest loadings (coefficients) of the second Eigenvector of the principal component analysis for the features of patient subset SUB44 and SUB32. The transformed feature vectors are to a large degree formed by a linear combination of feature from T_2-MRI, T_1-MRI, and PET.

SUB44 Component #	Loading	Feature Name
1	+0.32	PETInterceptMax
2	+0.32	PETDynVOIMax
3	+0.31	PETQuotMeanVOIMeanRef
4	+0.30	PETInterceptMean
5	+0.29	PETDynVOIMean
6	+0.28	PETQuotMaxVOIMeanRef
7	+0.22	MRIT1EnhancedHyperDensePixels
8	−0.19	MRIT2FlairTFLocalHomog
9	−0.19	MRIT2FlairTFEnergy
10	+0.17	MRIT1EnhancedQuotMeanMeanRefVOI
SUB32 Component #	Loading	Feature Name
1	+0.14	MRIT2FlairRefPeakPosition
2	+0.12	MRIT2FlairMean
3	+0.12	MRIT2FlairQuantile
4	+0.11	MRIT2FlairMax
5	−0.11	MRIT1EnhancedTFInertia
6	−0.10	MRIT1EnhancedQuotMaxMeanRefVOI
7	+0.10	MRIT2FlairMin
8	−0.10	MRIT1EnhancedTFClusterShade
9	−0.09	MRIT1EnhancedTFClusterProm
10	−0.09	MRIT2FlairRefPeakHeight

Table A.11: The ten largest loadings (coefficients) of the third Eigenvector of the principal component analysis for the features of patient subset SUB44 and SUB32. The transformed feature vectors are to a large degree formed by a linear combination of feature from T_2-MRI, T_1-MRI, and PET.

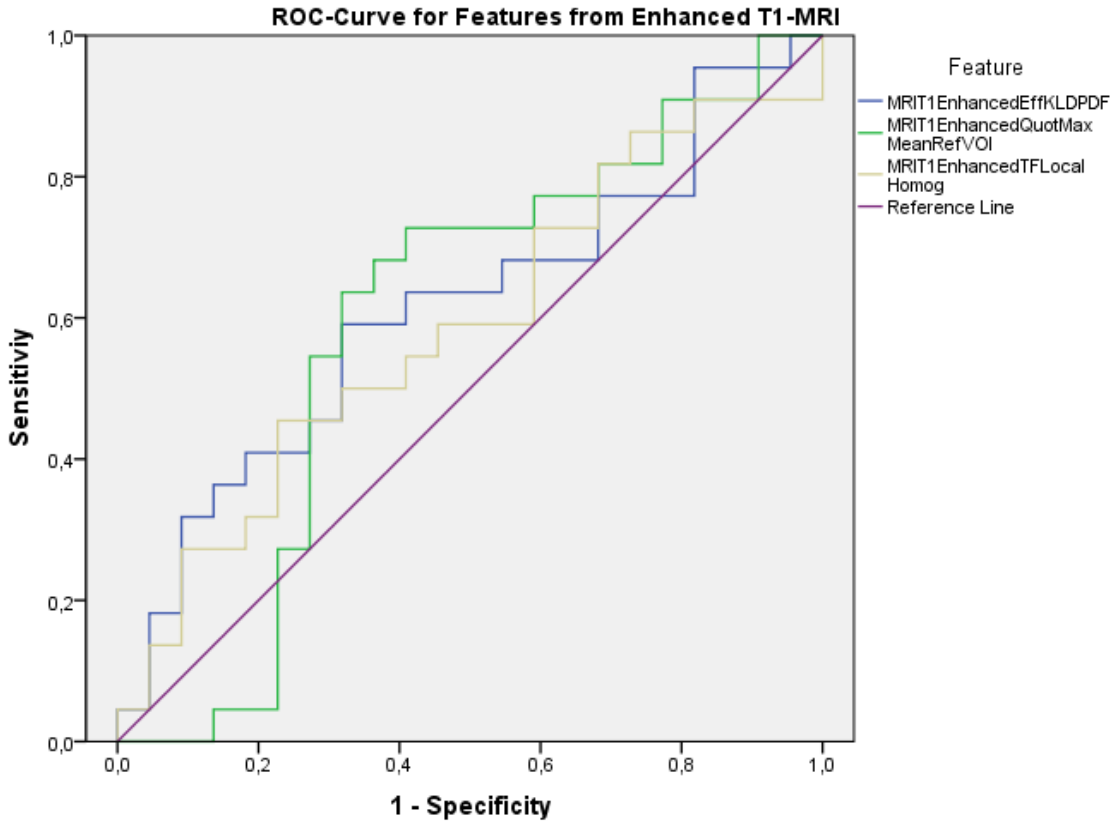


Figure A.1: ROC curve for non-normalized T_1-MRI. The 3 features with the highest area under the curve are plotted.

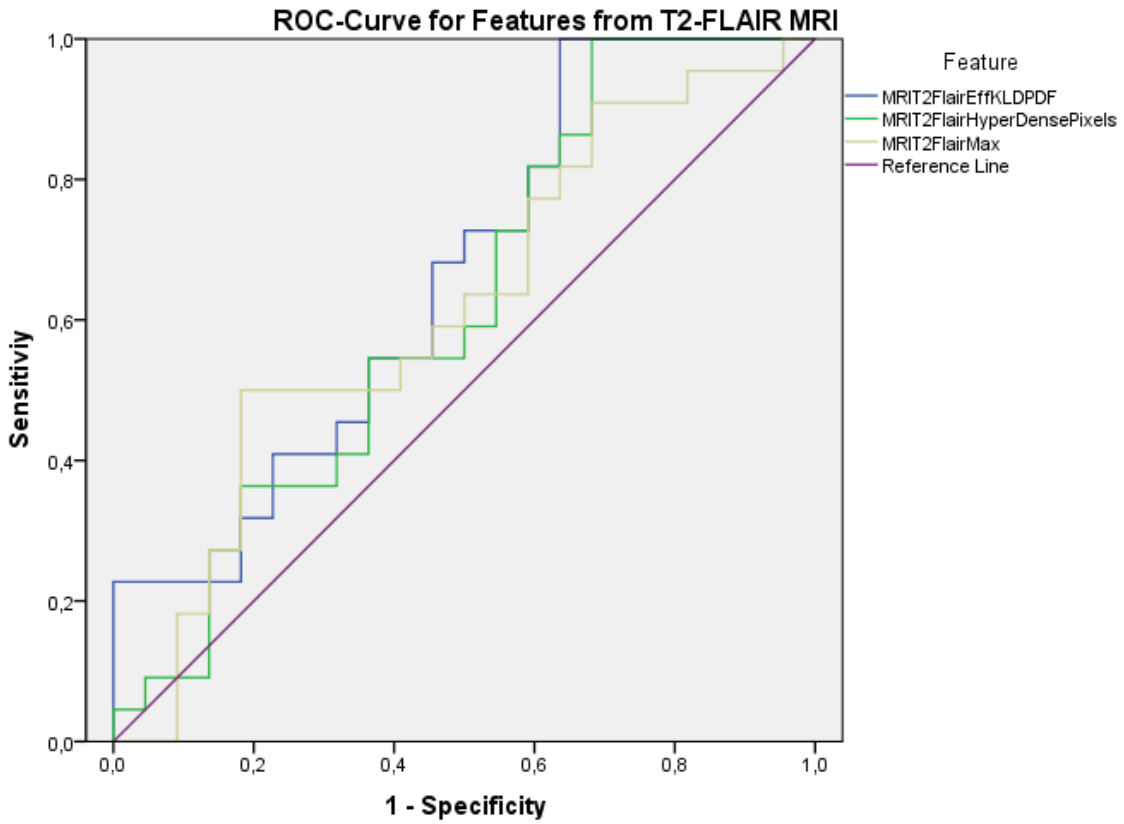


Figure A.2: ROC curve for non-normalized T_2-MRI. The 3 features with the highest area under the curve are plotted.

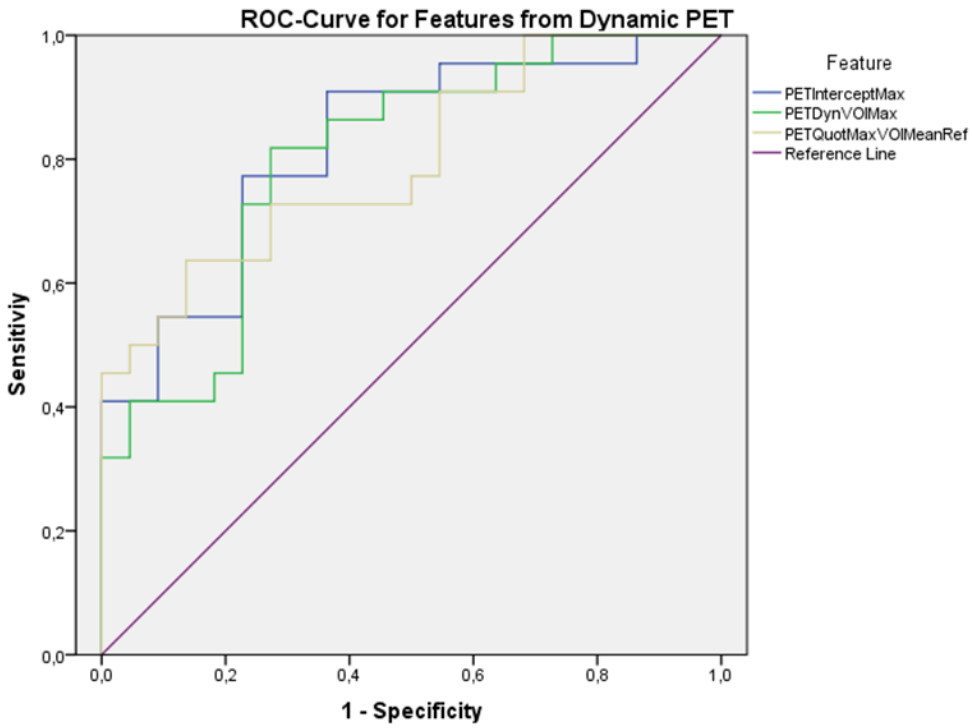


Figure A.3: ROC curve for SUV normalized PET. The 3 features with the highest area under the curve are plotted.

List of Acronyms

ADC	Apparent Diffucion Coefficient
A→P	Anterior-to-Posterior
ATP	Adenosine Triphosphate
AUC	Area Under The Curve
CBF	Cerebral Blood Flow
CBV	Cerebral Blood Volume
CT	X-Ray Computed Tomography
CTDI	Computed Tomography Dose Index
DCE-MRI	Dynamic Contrast-Enhanced MRI
DLP	Dose-Length-Product
DPD	TC-99m 2,3-dicarboxypropane-1, 1-diphosphonate
ECG	Electrocardiography
ECT	Emission Computed Tomography
EEG	Electroencephalography
FBP	Filtered Backprojection
FDG	[18F] Fluoro-D-Glucose
FET	[18F]-Fluor-Ethyl-Tyrosine
FLAIR	Fluid Attenuated Inversion Recovery
FLT	[18F] Fluorothymidine
FOV	Field Of View
FWHM	Full Width at Half Maximum
GABA	Gamma-Aminobutyric Acid
GRE	Gradient-Echo
H→F	Head-to-Feet
HU	Hounsfield Unit
ITK	Insight Toolkit
LOO-CV	Leave-One-Out Cross-Validation
LOPO-CV	Leave-One-Patient-Out Cross-Validation

LOR	Line Of Response
MET	Methyl-[11C]-Methionine
MLEM	Maximum Likelihood Expectation Maximization
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
MTT	Mean Time-to-Transit
OSEM	Ordered Subset Expectation Maximization
PCA	Principal Component Analysis
PDF	Probability Density Function
PET	Positron Emission Tomography
Poly	Polynomial
RBF	Radial Basis Function
RF	Radio Frequency
R→L	Right-to-Left
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SE	Spin-Echo
SPECT	Single Photon Emission Computed Tomography
SUV	Standardized Uptake Value
SVM	Support Vector Machine
TE	Time-to-Echo
TR	Time-to-Repeat
US	Ultrasound
VOI	Volume Of Interest
WHO	World Health Organization

List of Figures

List of Tables

Bibliography

[Adam 97] L.-E. Adam, J. Zaers, H. Ostertag, H. Trojan, M. Bellemann, and G. Brix. "Performance evaluation of the whole-body PET scanner ECAT EXACT HR + following the IEC standard". *IEEE Transactions on Nuclear Science*, Vol. 44, No. 3, pp. 1172 –1179, June 1997.

[Ai 92] W. I. Ai and P. Langley. "Induction of One-Level Decision Trees". In: *Proceedings of the Ninth International Conference on Machine Learning*, pp. 233–240, Morgan Kaufmann, 1992.

[Akso 00] S. Aksoy and R. M. Haralick. "Effects of Feature Normalization on Image Retrieval". 2000.

[Alme 99] P. Almeida, M. J. Ribeiro, M. Bottlaender, C. Loc'h, O. Langer, D. Strul, P. Hugonnard, P. Grangeat, B. Mazière, and B. Bendriem. "Absolute quantitation of iodine-123 epidepride kinetics using single-photon emission tomography: comparison with carbon-11 epidepride and positron emission tomography". *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 26, No. 12, pp. 1580–1588, December 1999.

[Ande 08] J. Anderson, D. Brandon, O. Oz, W. Erdman, and D. Mathews. "Impact of spatial resolution recovery on standard uptake values in clinical PET/CT". *Journal of Nuclear Medicine*, Vol. 49, No. Supp 1, p. 325, 2008.

[Apos 10] I. Apostolova, R. Wiemker, T. Paulus, S. Kabus, T. Dreilich, J. van den Hoff, M. Plotkin, J. Mester, W. Brenner, R. Buchert, and S. Klutmann. "Combined correction of recovery effect and motion blur for SUV quantification of solitary pulmonary nodules in FDG PET/CT". *European Radiology*, Vol. 20, No. 8, pp. 1868–1877, August 2010.

[Arvi 09] H. Arvinda, C. Kesavadas, P. Sarma, B. Thomas, V. Radhakrishnan, A. Gupta, T. Kapilamoorthy, and S. Nair. "Glioma grading: sensitivity, specificity, positive and negative predictive values of diffusion and perfusion imaging". *Journal of Neuro-Oncology*, Vol. 94, No. 1, pp. 87–96, August 2009.

[Asar 94] S. Asari, T. Makabe, S. Katayama, T. Itoh, S. Tsuchida, and T. Ohmoto. "Assessment of the pathological grade of astrocytic gliomas using an MRI score". *Neuroradiology*, Vol. 36, No. 4, pp. 308–310, May 1994.

[Bail 26] P. Bailey and H. Cushing. *A Classification of the Tumors of the Glioma Group on a Histogenetic Basis With a Correlated Study of Prognosis.* Lippincott, Philadelphia, 1926.

[Bent 10] M. J. Bent. "Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective". *Acta Neuropathologica*, Vol. 120, No. 3, pp. 297–304, September 2010.

[Bepp 11] T. Beppu, M. Sasaki, K. Kudo, A. Kurose, M. Takeda, H. Kashimura, A. Ogawa, and K. Ogasawara. "Prediction of malignancy grading using computed tomography perfusion imaging in nonenhancing supratentorial gliomas". *Journal of Neuro-Oncology*, Vol. 103, No. 3, pp. 619–627, July 2011.

[Bern 11] R. Bernardes, P. Serranho, and C. Lobo. "Digital Ocular Fundus Imaging: A Review". *Ophthalmologica*, Vol. 226, No. 4, pp. 161–181, October 2011.

[Bish 95] C. M. Bishop. *Neural networks for pattern recognition*. Oxford University Press, 1995.

[Blan 96] S. C. Blankespoor, X. Xu, K. Kaiki, J. K. Brown, H. R. Tang, C. E. Cann, and B. H. Hasegawa. "Attenuation correction of SPECT using X-ray CT on an emission-transmission CT system: myocardial perfusion assessment". *IEEE Transactions on Nuclear Science*, Vol. 43, No. 4, pp. 2263–2274, August 1996.

[Bock 09] A. Bockisch, L. S. Freudenberg, D. Schmidt, and T. Kuwert. "Hybrid Imaging by SPECT/CT and PET/CT: Proven Outcomes in Cancer Imaging". *Seminars in Nuclear Medicine*, Vol. 39, No. 4, pp. 276–289, July 2009.

[Bock 10] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, and G. Michelson. "Glaucoma Risk Index: Automated glaucoma detection from color fundus images". *Medical Image Analysis*, Vol. 14, No. 3, pp. 471–481, June 2010.

[Boel 09] R. Boellaard. "Standards for PET Image Acquisition and Quantitative Data Analysis". *Journal of Nuclear Medicine*, Vol. 50, No. Supplement 1, pp. 11–20, May 2009.

[Bran 10] W. Branderhorst, B. Vastenhouw, F. van der Have, E. Blezer, W. Bleeker, and F. Beekman. "Targeted multi-pinhole SPECT". *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 38, No. 3, pp. 552–561, March 2010.

[Cala 81] L. A. Cala, G. W. Thickbroom, J. L. Black, D. W. Collins, and F. L. Mastaglia. "Brain density and cerebrospinal fluid space size: CT of normal volunteers.". *American Journal of Neuroradiology*, Vol. 2, No. 1, pp. 41–47, January-February 1981.

[Calc 11] M. L. Calcagni, G. Galli, A. Giordano, S. Taralli, C. Anile, A. Niesen, and R. P. Baum. "Dynamic O-(2-[18F]fluoroethyl)-L-tyrosine (F-18 FET) PET for Glioma Grading: Assessment of Individual Probability of Malignancy". *Clinical Nuclear Medicine*, Vol. 36, No. 10, pp. 841–847, October 2011.

[Caru 06] R. Caruana and A. Niculescu-Mizil. "An empirical comparison of supervised learning algorithms". In: *Proceedings of the 23rd international conference on machine learning*, pp. 161–168, 2006.

[Ceel 07] H. Ceelie, R. B. Dinkelaar, and W. van Gelder. "Examination of peripheral blood films using automated microscopy; evaluation of Diffmaster Octavia and Cellavision DM96". *Journal of Clinical Pathology*, Vol. 60, No. 1, pp. 72–79, January 2007.

[Chan 09] C.-C. Chang and C.-J. Lin. "LIBSVM: a Library for Support Vector Machines". February 2009.

[Chan 78] L.-T. Chang. "A Method for Attenuation Correction in Radionuclide Computed Tomography". *IEEE Transactions on Nuclear Science*, Vol. 25, No. 1, pp. 638–643, February 1978.

[Chen 06] J. Chen, S. Caputlu-Wilson, H. Shi, J. Galt, T. Faber, and E. Garcia. "Automated quality control of emission-transmission misalignment for attenuation correction in myocardial perfusion imaging with SPECT-CT systems". *Journal of Nuclear Cardiology*, Vol. 13, No. 1, pp. 43–49, January-February 2006.

[Chen 98] C. H. Chen, R. F. J. Muzic, A. D. Nelson, and L. P. Adler. "A nonlinear spatially variant object-dependent system model for prediction of partial volume effects and scatter in PET". *IEEE Transactions on Medical Imaging*, Vol. 17, No. 2, pp. 214–227, April 1998.

[Cher 03] S. R. Cherry, J. A. Sorenson, and M. E. Phelps. *Physics in Nuclear Medicine*. Elsevier, Philadelphia, 3rd Ed., 2003.

[Cohe 60] J. Cohen. "A Coefficient of Agreement for Nominal Scales". *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37–46, April 1960.

[Coll 04] G. Collewet, M. Strzelecki, and F. Mariette. "Influence of MRI acquisition protocols and image intensity normalization methods on texture classification". *Magnetic Resonance Imaging*, Vol. 22, No. 1, pp. 81–91, January 2004.

[Conn 84] R. Conners, M. Trivedi, and C. Harlow. "Segmentation of a High-Resolution Urban Scene Using Texture Operators". *Computer Vision, Graphics, and Image Pprocessing*, Vol. 25, No. 3, pp. 273–310, March 1984.

[Cort 95] C. Cortes and V. Vapnik. "Support-vector networks". *Machine Learning*, Vol. 20, No. 3, pp. 273–297, September 1995.

[Cove 67] T. Cover and P. Hart. "Nearest neighbor pattern classification". *IEEE Transactions on Information Theory*, Vol. 13, No. 1, pp. 21–27, January 1967.

[Da S 01] A. J. Da Silva, H. R. Tang, K. H. Wong, M. C. Wu, M. W. Dae, and B. H. Hasegawa. "Absolute Quantification of Regional Myocardial Uptake of 99mTc-Sestamibi with SPECT: Experimental Validation in a Porcine Model". *Journal of Nuclear Medicine*, Vol. 42, No. 5, pp. 772–779, May 2001.

[Da S 99] A. J. Da Silva, H. R. Tang, M. C. Wu, and B. H. Hasegawa. "Absolute quantitation of myocardial activity in phantoms". *IEEE Transactions on Nuclear Science*, Vol. 46, No. 3, pp. 659–666, November 1999.

[Devo 05] A. Devos, A. Simonetti, M. van der Graaf, L. Lukas, J. Suykens, L. Vanhamme, L. Buydens, A. Heerschap, and S. V. Huffel. "The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification". *Journal of Magnetic Resonance*, Vol. 173, No. 2, pp. 218–228, April 2005.

[Dewa 08] Y. Dewaraja, M. Ljungberg, and K. Koral. "Effects of dead time and pile up on quantitative SPECT for I-131 dosimetric studies". *Journal of Nuclear Medicine*, Vol. 49, No. Supplement 1, p. 47, 2008.

[DGN 13] "DGN-Handlungsempfehlung (S1-Leitlinie) - Skelettszintigraphie". 2013.

[Dorr 11] M. Dorrius, M. der Weide, P. van Ooijen, R. Pijnappel, and M. Oudkerk. "Computer-aided detection in breast MRI: a systematic review and meta-analysis". *European Radiology*, Vol. 21, No. 8, pp. 1600–1608, August 2011.

[Druk 05] K. Drukker, K. Horsch, and M. L. Giger. "Multimodality Computerized Diagnosis of Breast Lesions Using Mammography and Sonography". *Academic Radiology*, Vol. 12, No. 8, pp. 970–979, August 2005.

[Du 05] Y. Du, B. M. W. Tsui, and E. C. Frey. "Partial volume effect compensation for quantitative brain SPECT imaging". *IEEE Transactions on Medical Imaging*, Vol. 24, No. 8, pp. 969–976, August 2005.

[Du 06] Y. Du and et al. "Model-based compensation for quantitative 123 I brain SPECT imaging". *Physics in Medicine and Biology*, Vol. 51, No. 5, p. 1269, April 2006.

[El F 99] G. N. El Fakhri, I. Buvat, M. Pélégrini, H. Benali, P. Almeida, B. Bendriem, A. Todd-Pokropek, and R. Di Paola. "Respective roles of scatter, attenuation, depth-dependent collimator response and finite spatial resolution in cardiac single-photon emission tomography quantitation: a Monte Carlo study". *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 26, No. 5, pp. 437–446, May 1999.

[Elli 07] S. Ellika, R. Jain, S. Patel, L. Scarpace, L. Schultz, J. Rock, and T. Mikkelsen. "Role of Perfusion CT in Glioma Grading and Comparison with Conventional MR Imaging Features". *American Journal of Neuroradiology*, Vol. 28, No. 10, pp. 1981–1987, November-December 2007.

[Fan 05] R. E. Fan, P. H. Chen, and C. J. Lin. "Working Set Selection Using Second Order Information for Training Support Vector Machines". *Journal of Machine Learning Research*, Vol. 6, pp. 1889–1918, December 2005.

[Feul 12] J. Feulner, S. K. Zhou, M. Hammon, J. Hornegger, and D. Comaniciu. "Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior". *Medical Image Analysis*, Vol. 17, No. 2, pp. 254–270, February 2012.

[Fish 36] R. A. Fisher. "The Use Of Multiple Measurements In Taxonomic Problems". *Annals of Human Genetics*, Vol. 7, No. 2, pp. 179–188, September 1936.

[Flei 03] J. Fleiss, B. Levin, and M. Paik. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, New Jersey, 3rd Ed., 2003.

[Flei 69] J. L. Fleiss, J. Cohen, and B. Everitt. "Large sample standard errors of kappa and weighted kappa.". *Psychological Bulletin*, Vol. 72, No. 5, p. 323, November 1969.

[Floe 05] F. W. Floeth, D. Pauleit, H.-J. Wittsack, K. J. Langen, G. Reifenberger, K. Hamacher, M. Messing-Jünger, K. Zilles, F. Weber, W. Stummer, H.-J. Steiger, G. Woebker, H.-W. Müller, H. Coenen, and M. Sabel. "Multimodal metabolic imaging of cerebral gliomas: positron emission tomography with [18F]fluoroethyl-l-tyrosine and magnetic resonance spectroscopy". *Journal of Neurosurgery*, Vol. 102, No. 2, pp. 318–327, February 2005.

[Floy 84] C. E. Floyd, R. J. Jaszczak, H. C. C, and C. R. E. "Energy and spatial distribution of multiple order Compton scatter in SPECT: a Monte Carlo investigation". *Physics in Medicine and Biology*, Vol. 29, No. 10, pp. 1217–1230, October 1984.

[Freu 95] Y. Freund and R. Schapire. "A desicion-theoretic generalization of on-line learning and an application to boosting". In: P. Vitányi, Ed., *Computational Learning Theory*, pp. 23–37, Springer Berlin / Heidelberg, 1995.

[Freu 96] Y. Freund and R. E. Schapire. "Experiments with a new boosting algorithm". In: *ICML*, pp. 148–156, 1996.

[Frey 90] E. C. Frey and B. M. W. Tsui. "Parameterization of the scatter response function in SPECT imaging using Monte Carlo simulation". *IEEE Transactions on Nuclear Science*, Vol. 37, No. 3, pp. 1308–1315, June 1990.

[Frey 94] E. C. Frey and B. M. W. Tsui. "Modeling the Scatter Response Function in Inhomogenous Scattering Media for SPECT". *IEEE Transactions on Nuclear Science*, Vol. 41, No. 4, pp. 1585–1593, August 1994.

[Frie 89] J. Friedman. "Regularized Discriminant Analysis". *Journal of the American Statistical Association*, Vol. 84, No. 405, pp. 165–175, March 1989.

[Germ 01] P. Germain, G. Roul, J. Baruthio, C. Jahn, P. M. Coulbois, B. Dumitresco, J. L. Dietemann, P. Bareiss, and A. Constantinesco. "Myocardial flow reserve parametric map, assessed by first-pass MRI compartmental analysis at the chronic stage of infarction". *Journal of Magnetic Resonance Imaging*, Vol. 13, No. 3, pp. 352–360, March 2001.

[Gewo 00] L. Geworski, B. O. Knoop, M. L. de Cabrejas, W. H. Knapp, and D. L. Munz. "Recovery correction for quantitation in emission tomography: a feasibility study". *European Journal of Nuclear Medicine*, Vol. 27, No. 2, pp. 161–169, February 2000.

[Gewo 02] L. Geworski, B. O. Knoop, M. de Wit, V. Ivanevi, R. Bares, and D. L. Munz. "Multicenter Comparison of Calibration and Cross Calibration of PET Scanners". *Journal of Nuclear Medicine*, Vol. 43, No. 5, pp. 635–639, May 2002.

[Gill 91] D. R. Gilland, R. J. Jaszczak, Z. Liang, K. L. Greer, and R. E. Coleman. "Quantitative SPECT brain imaging: effects of attenuation and detector responseat". In: *Nuclear Science Symposium and Medical Imaging Conference, 1991., Conference Record of the 1991 IEEE*, pp. 1723–1773, 1991.

[Gilm 06] M. D. Gilman, A. J. Fischman, V. Krishnasetty, E. F. Halpern, and S. L. Aquino. "Optimal CT Breathing Protocol for Combined Thoracic PET/CT". *American Journal of Roentgenology*, Vol. 187, No. 5, pp. 1357–1360, November 2006.

[Grav 13] P. Gravel, J. Verhaeghe, and A. J. Reader. "3D PET image reconstruction including both motion correction and registration directly into an MR or stereotaxic spatial atlas". *Physics in Medicine and Biology*, Vol. 58, No. 1, p. 105, January 2013.

[Gull 10] G. T. Gullberg, B. W. Reutter, A. Sitek, J. S. Maltz, and T. F. Budinger. "Dynamic single photon emission computed tomography - basic principles and cardiac applications". *Physics in Medicine and Biology*, Vol. 55, No. 20, pp. R111–R191, October 2010.

[Habe 91] U. Haberkorn, L. G. Strauss, A. Dimitrakopoulou, R. Engenhart, F. Oberdorfer, H. Ostertag, J. Romahn, and G. van Kaick. "PET Studies of Fluorodeoxyglucose Metabolism in Patients with Recurrent Colorectal Tumors Receiving Radiotherapy". *Journal of Nuclear Medicine*, Vol. 32, No. 8, pp. 1485–1490, August 1991.

[Haeg 12] K. Haegler, M. Wiesmann, C. Böm, J. Freiherr, O. Schnell, H. Brückmann, J.-C. Tonn, and J. Linn. "New similarity search based glioma grading". *Neuroradiology*, Vol. 54, No. 8, pp. 829–837, August 2012.

[Hahn 10] D. Hahn, V. Daum, and J. Hornegger. "Automatic Parameter Selection for Multimodal Image Registration". *Medical Imaging, IEEE Transactions on*, Vol. 29, No. 5, pp. 1140–1155, May 2010.

[Hall 09] M. Hall, E. Frank, G. Holmes, B. Pfahringer, and I. H. W. P. Reutemann. "The WEKA Data Mining Software: An Update". *SIGKDD Explorations*, Vol. 11, No. 1, pp. 10–18, June 2009.

[Hara 73] R. Haralick, K. Shanmugam, and I. Dinstein. "Textural Features for Image Classification". *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 6, pp. 610–621, November 1973.

[Hast 09] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2nd Ed., 2009.

[Heis 09] B. Heismann and M. Balda. "Quantitative image-based spectral reconstruction for computed tomography". *Medical Physics*, Vol. 36, No. 10, pp. 4471–4485, October 2009.

[Hila 12] A. Hilario, A. Ramos, A. Perez-Nunez, E. Salvador, J. Millan, A. Lagares, J. Sepulveda, P. Gonzalez-Leon, A. Hernandez-Lain, and J. Ricoy. "The Added Value of Apparent Diffusion Coefficient to Cerebral Blood Volume in the Preoperative Grading of Diffuse Gliomas". *American Journal of Neuroradiology*, Vol. 33, No. 4, pp. 701–707, April 2012.

[Hnat 06] K. Hnatkova, Y. Gang, V. N. Batchvarov, and M. Malik. "Precision of QT Interval Measurement by Advanced Electrocardiographic Equipment". *Pacing and Clinical Electrophysiology*, Vol. 29, No. 11, pp. 1277–1284, November 2006.

[Hoff 79] E. J. Hoffman, S.-C. Huang, and M. E. Phelps. "Quantitation in Positron Emission Computed Tomography: 1. Effect of Object Size". *Journal of Computer Assisted Tomography*, Vol. 3, No. 3, pp. 299–308, June 1979.

[Hutt 98] B. F. Hutton and Y. H. Lau. "Application of distance-dependent resolution compensation and post-reconstruction filtering for myocardial SPECT". *Physics in Medicine and Biology*, Vol. 43, No. 6, p. 1679, June 1998.

[Iban 05] L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide*. Kitware, Inc. ISBN 1-930934-15-7, http://www.itk.org/ItkSoftwareGuide.pdf, second Ed., 2005.

[Ikar 12] Y. Ikari, T. Nishio, Y. Makishi, Y. Miya, K. Ito, R. Koeppe, and M. Senda. "Head motion evaluation and correction for PET scans with 18F-FDG in the Japanese Alzheimer's disease neuroimaging initiative (J-ADNI) multi-center study". *Annals of Nuclear Medicine*, Vol. 26, No. 7, pp. 535–544, August 2012.

[Jage 09] F. Jäger and J. Hornegger. "Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging". *IEEE Transactions on Medical Imaging*, Vol. 28, No. 1, pp. 137–150, January 2009.

[Jage 10] F. Jäger. *Normalization of Magnetic Resonance Images and its Application to the Diagnosis of the Scoliotic Spine*. PhD thesis, Pattern Recognition Lab, University of Erlangen-Nuremberg, Martensstrasse 3, 91058 Erlangen, Germany, 2010.

[Jako 11] B. W. Jakoby, Y. Bercier, M. Conti, M. E. Casey, B. Bendriem, and D. W. Townsend. "Physical and clinical performance of the mCT time-of-flight PET/CT scanner". *Physics in Medicine and Biology*, Vol. 56, No. 8, p. 2375, April 2011.

[Jasz 84] R. J. Jaszczak, K. L. Greer, C. E. J. Floyd, C. C. Harris, and R. E. Coleman. "Improved SPECT Quantification Using Compensation for Scattered Photons". *Journal of Nuclear Medicine*, Vol. 25, No. 8, pp. 893–900, August 1984.

[Jent 08] W. Jentzen, R. Weise, J. Kupferschlager, L. Freudenberg, W. Brandau, R. Bares, W. Burchert, and A. Bockisch. "Iodine-124 PET dosimetry in differentiated thyroid cancer: recovery coefficient in 2D and 3D modes for PET(/CT) systems". *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 35, No. 3, pp. 611–623, April 2008.

[Jude 08] M. S. Judenhofer, H. F. Wehrl, D. F. Newport, C. Catana, S. B. Siegel, M. Becker, A. Thielscher, M. Kneilling, M. P. Lichy, M. Eichner, K. Klingel, G. Reischl, S. Widmaier, M. Rocken, R. E. Nutt, H. J. Machulla, K. Uludag, S. R. Cherry, C. D. Claussen, and B. J. Pichler. "Simultaneous PET-MRI: a new approach for functional and morphological imaging". *Nature Medicine*, Vol. 14, No. 4, pp. 459–465, April 2008.

[Kess 84] R. M. Kessler, J. R. J. Ellis, and M. Eden. "Analysis of Emission Tomographic Scan Data: Limitations Imposed by Resolution and Background". *Journal of Computer Assisted Tomography*, Vol. 8, No. 3, pp. 514–522, June 1984.

[Kief 11] A. Kiefer, T. Kuwert, D. Hahn, J. Hornegger, M. Uder, and P. Ritt. "Anatomical Accuracy of Abdominal Lesion Localization". *Nuklearmedizin*, Vol. 50, No. 4, pp. 141–173, April 2011.

[King 95] R. D. King, C. Feng, and A. Sutherland. "StatLog: Comparison of classification algorithms on large -world problems". *Applied Artificial Intelligence*, Vol. 9, No. 3, pp. 289–333, 1995.

[Kohl 98a] V. Kohli, M. A. King, P. Tin-Su, and S. J. Glick. "Compensation for distance-dependent resolution in cardiac-perfusion SPECT: impact on uniformity of wall counts and wall thickness". *IEEE Transactions on Nuclear Science*, Vol. 45, No. 3, pp. 1104–1110, June 1998.

[Kohl 98b] V. Kohli. "Comparison of frequency-distance relationship and Gaussian-diffusion-based methods of compensation for distance-dependent spatial resolution in SPECT imaging". *Physics in Medicine and Biology*, Vol. 43, No. 4, p. 1025, April 1998.

[Kora 86] K. F. Koral, N. H. Clinthorne, and W. Leslie Rogers. "Improving emission-computed-tomography quantification by Compton-scatter rejection through offset windows". *Nuclear Instruments and Methods in*

Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Vol. 242, No. 3, pp. 610–614, January 1986.

[Kora 88] K. F. Koral, X. Wang, W. L. Rogers, N. H. Clinthorne, and X. Wang. "SPECT Compton-Scattering Correction by Analysis of Energy Spectra". *Journal of Nuclear Medicine*, Vol. 29, No. 2, pp. 195–202, February 1988.

[LaCr 94] K. J. LaCroix, B. M. W. Tsui, B. H. Hasegawa, and J. K. Brown. "Investigation of the use of X-ray CT images for attenuation compensation in SPECT". *IEEE Transactions on Nuclear Science*, Vol. 41, No. 6, pp. 2793–2799, December 1994.

[Lang 05] K. Langen, K. Hamacher, D. Pauleit, F. Floeth, G. Stoffels, D. Bauer, G. Reifenberger, K. Zilles, and H. Coenen. "Evaluation of new 18F-labeled amino acids for brain PET". *Anatomy and Embryology*, Vol. 210, No. 5-6, pp. 455–461, December 2005.

[Lang 11] K. Langen, P. Bartenstein, H. Boecker, P. Brust, H. Coenen, A. Drzezga, F. Grünwald, B. Krause, T. Kuwert, O. Sabri, K. Tatsch, W. Weber, and M. Schreckenberger. "German guidelines for brain tumour imaging by PET and SPECT using labelled amino acids". *Nuklearmedizin*, Vol. 50, No. 4, pp. 167–173, April 2011.

[Lau 10] E. W. Lau, K. J. Drummond, R. E. Ware, E. Drummond, A. Hogg, G. Ryan, A. Grigg, J. Callahan, and R. J. Hicks. "Comparative PET study using F-18 FET and F-18 FDG for the evaluation of patients with suspected brain tumour". *Journal of Clinical Neuroscience*, Vol. 17, No. 1, pp. 43–49, January 2010.

[Law 03] 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". *American Journal of Neuroradiology*, Vol. 24, No. 10, pp. 1989–1998, November-December 2003.

[Lewi 06] D. H. Lewis, J. P. Bluestone, M. Savina, W. H. Zoller, E. B. Meshberg, and S. Minoshima. "Imaging Cerebral Activity in Recovery from Chronic Traumatic Brain Injury: A Preliminary Report". *Journal of Neuroimaging*, Vol. 16, No. 3, pp. 272–277, July 2006.

[Li 06] G.-Z. Li, J. Yang, C.-Z. Ye, and D.-Y. Geng. "Degree prediction of malignancy in brain glioma using support vector machines". *Computers in Biology and Medicine*, Vol. 36, No. 3, pp. 313–325, March 2006.

[Ljun 90] M. Ljungberg and S.-E. Strand. "Scatter and Attenuation Correction in SPECT Using Density Maps and Monte Carlo Simulated Scatter Functions". *Journal of Nuclear Medicine*, Vol. 31, No. 9, pp. 1560–1567, September 1990.

[Loui 07] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues. "The 2007 WHO Classification of Tumours of the Central Nervous System". *Acta Neuropathologica*, Vol. 114, No. 2, pp. 97–109, July 2007.

[Mart 04] K. Marten, T. Seyfarth, F. Auer, E. Wiener, A. Grillhösl, S. Obenauer, E. J. Rummeny, and C. Engelke. "Computer-assisted detection of pulmonary nodules: performance evaluation of an expert knowledge-based

detection system in consensus reading with experienced and inexperienced chest radiologists". *European Radiology*, Vol. 14, No. 10, pp. 1930–1938, October 2004.

[MATL 10] *MATLAB R2010b*. MathWorks, 3 Apple Hill Drive, Natick, Massachusetts 01760, USA, 2010.

[McGi 05] M. J. McGirt, G. F. Woodworth, A. L. Coon, J. M. Frazier, E. Amundson, I. Garonzik, A. Olivi, and J. D. Weingart. "Independent predictors of morbidity after image-guided stereotactic brain biopsy: a risk assessment of 270 cases". *Journal of Neurosurgery*, Vol. 102, No. 5, pp. 897–901, May 2005.

[Melh 97] E. R. Melhem, D. A. Israel, S. Eustace, and H. Jara. "MR of the spine with a fast T1-weighted fluid-attenuated inversion recovery sequence.". *American Journal of Neuroradiology*, Vol. 18, No. 3, pp. 447–454, March 1997.

[Mour 09] J. Mourik, M. Lubberink, F. Velden, A. Lammertsma, and R. Boellaard. "Off-line motion correction methods for multi-frame PET data". *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 36, No. 12, pp. 2002–2013, December 2009.

[Neun 12] I. Neuner, J. B. Kaffanke, K.-J. Langen, E. R. Kops, L. Tellmann, G. Stoffels, C. Weirich, C. Filss, J. Scheins, H. Herzog, and N. Shah. "Multimodal imaging utilising integrated MR-PET for human brain tumour assessment". *European Radiology*, Vol. 22, No. 12, pp. 2568–2580, December 2012.

[Noth 11] E. Nöth, A. Maier, A. Gebhard, T. Bocklet, W. Schupp, M. Schuster, and T. Haderlein. "Automatic Evaluation of Dysarthric Speech and Telemedical Use in the Therapy". *The Phonetician*, Vol. 103, No. 1, pp. 75–87, 2011.

[Nyul 00] L. Nyul, J. Udupa, and X. Zhang. "New variants of a method of MRI scale standardization". *IEEE Transactions on Medical Imaging*, Vol. 19, No. 2, pp. 143–150, February 2000.

[Nyul 99] L. Nyul and J. Udupa. "On standardizing the MR image intensity scale". *Magnetic Resonance in Medicine*, Vol. 42, No. 6, pp. 1072–1081, December 1999.

[Ogaw 91] K. Ogawa, Y. Harata, T. Ichihara, A. Kubo, and S. Hashimoto. "A practical method for position-dependent Compton-scatter correction in single photon emission CT". *IEEE Transactions on Medical Imaging*, Vol. 10, No. 3, pp. 408–412, September 1991.

[Oliv 10] A. Oliver, J. Freixenet, J. Marti, E. Perez, J. Pont, E. R. Denton, and R. Zwiggelaar. "A review of automatic mass detection and segmentation in mammographic images". *Medical Image Analysis*, Vol. 14, No. 2, pp. 87–110, April 2010.

[Olli 96] J. M. Ollinger. "Model-based scatter correction for fully 3D PET". *Physics in Medicine and Biology*, Vol. 41, No. 1, p. 153, January 1996.

[Osor 11] I. Osorio, A. Lyubushin, and D. Sornette. "Automated seizure detection: Unrecognized challenges, unexpected insights". *Epilepsy & Behavior*, Vol. 22, Supplement 1, No. 1, pp. S7–S17, December 2011.

[Ouya 13] J. Ouyang, Q. Li, and G. E. Fakhri. "Magnetic Resonance-Based Motion Correction for Positron Emission Tomography Imaging". *Seminars in Nuclear Medicine*, Vol. 43, No. 1, pp. 60–67, January 2013.

[Padm 03] M. Padma, S. Said, M. Jacobs, D. Hwang, K. Dunigan, M. Satter, B. Christian, J. Ruppert, T. Bernstein, G. Kraus, and J. Mantil. "Prediction of Pathology and Survival by FDG PET in Gliomas". *Journal of Neuro-Oncology*, Vol. 64, No. 3, pp. 227–237, September 2003.

[Paul 05] D. Pauleit, F. Floeth, K. Hamacher, M. J. Riemenschneider, G. Reifenberger, H.-W. MÃijller, K. Zilles, H. H. Coenen, and K.-J. Langen. "O-(2-[18F]fluoroethyl)-l-tyrosine PET combined with MRI improves the diagnostic assessment of cerebral gliomas". *Brain*, Vol. 128, No. 3, pp. 678–687, March 2005.

[Paul 06] D. Pauleit, A. Zimmermann, G. Stoffels, D. Bauer, J. Risse, M. O. Flüss, K. Hamacher, H. H. Coenen, and K.-J. Langen. "18F-FET PET Compared with 18F-FDG PET and CT in Patients with Head and Neck Cancer". *Journal of Nuclear Medicine*, Vol. 47, No. 2, pp. 256–261, February February 2006.

[Pear 01] K. Pearson. "On lines and planes of closest fit to systems of points in space". *Philosophical Magazine*, Vol. 2, No. 6, pp. 559–572, November 1901.

[Perf 07] "Performance Measurements of Gamma Cameras". Tech. Rep., National Electrical Manufacturers Association, 2007.

[Plat 98] J. Platt. "Sequential minimal optimization: A fast algorithm for training support vector machines". *Technical Report msr-tr-98-14*, April 1998.

[Popp 06] G. Pöpperl, F. W. Kreth, J. Herms, W. Koch, J. H. Mehrkens, F. J. Gildehaus, H. A. Kretzschmar, J. C. Tonn, and K. Tatsch. "Analysis of 18F-FET PET for Grading of Recurrent Gliomas: Is Evaluation of Uptake Kinetics Superior to Standard Methods?". *Journal of Nuclear Medicine*, Vol. 47, No. 3, pp. 393–403, March March 2006.

[Popp 07] G. Pöpperl, F. Kreth, J. Mehrkens, J. Herms, K. Seelos, W. Koch, F. Gildehaus, H. Kretzschmar, J. Tonn, and K. Tatsch. "FET PET for the evaluation of untreated gliomas: correlation of FET uptake and uptake kinetics with tumour grading". *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 34, No. 12, pp. 1933–1942, December 2007.

[Pret 09] P. H. Pretorius and M. A. King. "Diminishing the impact of the partial volume effect in cardiac SPECT perfusion imaging". *Medical Physics*, Vol. 36, No. 1, pp. 105–115, January 2009.

[Pret 98] P. H. Pretorius, M. A. King, T. S. Pan, D. J. de Vries, S. J. Glick, and C. L. Byrne. "Reducing the influence of the partial volume effect on SPECT activity quantitation with 3D modelling of spatial resolution in iterative reconstruction". *Physics in Medicine and Biology*, Vol. 43, No. 2, pp. 407–420, February 1998.

[Rapp 13] M. Rapp, A. Heinzel, N. Galldiks, G. Stoffels, J. Felsberg, C. Ewelt, M. Sabel, H. J. Steiger, G. Reifenberger, T. Beez, H. H. Coenen, F. W. Floeth, and K.-J. Langen. "Diagnostic Performance of 18F-FET PET in Newly Diagnosed Cerebral Lesions Suggestive of Glioma". *Journal of Nuclear Medicine*, Vol. 54, No. 2, pp. 229–235, February 2013.

[Reis 08] M. Reiser, W. Semmler, and H. Hricak. *Magnetic Resonance Tomography.* Springer Berlin Heidelberg, 2008.

[Riem 02] B. Riemann, K. Papke, N. Hoess, T. Kuwert, M. Weckesser, P. Matheja, H. Wassmann, W. Heindel, and O. Schober. "Noninvasive Grading of Untreated Gliomas: A Comparative Study of MR Imaging and 3-(Iodine 123)-L-methyltyrosine SPECT". *Radiology*, Vol. 225, No. 2, pp. 567–574, November 2002.

[Ritt 10] P. Ritt, D. Hahn, J. Hornegger, O. Ganslandt, A. Doerfler, and T. Kuwert. "Anatomische Genauigkeit der retrospektiven, automatischen und starren Bildregistrierung zwischen kranialer FET-PET und MRT". In: *Nuklearmedizin, Konferenzausgabe zur 48. Jahrestagung der Deutschen Gesellschaft für Nuklearmedizin, Leipzig*, Deutsche Gesellschaft für Nuklearmedizin, April 2010.

[Ritt 11a] P. Ritt, J. Hornegger, and T. Kuwert. "Technische und physikalische Aspekte der SPECT/CT". *Der Nuklearmediziner*, Vol. 34, No. 1, pp. 9–20, March 2011.

[Ritt 11b] P. Ritt, H. Vija, J. Hornegger, and T. Kuwert. "Absolute Quantification in SPECT". *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 38, Supplement 1, No. 1, pp. 69–77, May 2011.

[Ritt 12] P. Ritt, J. Hornegger, and T. Kuwert. "Automatische Klassifizierung von Gliomen des Gehirns auf Basis multimodaler Bilddaten aus PET, CT und MRT". In: *Nuklearmedizin, Konferensausgabe zur 50. Jahrestagung der Deutschen Gesellschaft für Nuklearmedizin, Bremen*, Deutsche Gesellschaft für Nuklearmedizin, April 2012.

[Ritt 13a] P. Ritt, J. Hornegger, T. Kuwert, and K. Langen. "Automated Classification of Cerebral Gliomas on Multimodal Imaging Data from PET and MR". In: *PSMR 2013 / 4th Jülich MR-PET Workshop, Aachen*, 2013.

[Ritt 13b] P. Ritt and T. Kuwert. "Quantitative SPECT/CT". In: O. Schober and B. Riemann, Eds., *Molecular Imaging in Oncology*, pp. 313–330, Springer Berlin Heidelberg, 2013.

[Rome 06] W. Römer, N. Reichel, H. A. Vija, I. Nickel, J. Hornegger, W. Bautz, and T. Kuwert. "Isotropic reconstruction of SPECT data using OSEM3D: Correlation with CT". *Academic Radiology*, Vol. 13, No. 4, pp. 496–502, April 2006.

[Rose 58] F. Rosenblatt. "The perceptron: a probabilistic model for information storage and organization in the brain". *Psychological Review*, Vol. 65, No. 6, pp. 386–408, November 1958.

[Rose 95] M. Rosenthal, J. Cullom, W. Hawkins, S. Moore, B. Tsui, and M. Yester. "Quantitative SPECT Imaging: A Review and Recommendations by the Focus Committee of the Society of Nuclear Medicine Computer and Instrumentation Council". *Journal of Nuclear Medicine*, Vol. 36, No. 8, pp. 1489–1513, August 1995.

[Rous 93] O. Rousset, Y. Ma, M. Kamber, and A. C. Evans. "3D simulations of radiotracer uptake in deep nuclei of human brain". *Computerized Medical Imaging and Graphics*, Vol. 17, No. 4-5, pp. 373–379, July-October 1993.

[Rous 98] G. Rousset, Y. Ma, and A. C. Evans. "Correction for Partial Volume Effects in PET: Principle and Validation". *Journal of Nuclear Medicine*, Vol. 39, No. 5, pp. 904–911, May 1998.

[Rume 86] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning internal representations by error propagation". In: D. E. Rumelhart and J. L. McClelland, Eds., *Parallel distributed processing: explorations in the microstructure of cognition*, pp. 318–362, MIT Press, Cambridge, MA, USA, 1986.

[Sack 11] J. C. Sackellares, D.-S. Shiau, J. J. Halford, S. M. LaRoche, and K. M. Kelly. "Quantitative EEG analysis for automated detection of nonconvulsive seizures in intensive care units". *Epilepsy & Behavior*, Vol. 22, Supplement 1, No. 1, pp. S69–S73, December 2011.

[Sahi 09] B. Sahiner, H.-P. Chan, L. M. Hadjiiski, M. A. Roubidoux, C. Paramagul, J. E. Bailey, A. V. Nees, C. E. Blane, D. D. Adler, S. K. Patterson, K. A. Klein, R. W. Pinsky, and M. A. Helvie. "Multi-modality CADx: ROC Study of the Effect on Radiologists' Accuracy in Characterizing Breast Masses on Mammograms and 3D Ultrasound Images". *Academic Radiology*, Vol. 16, No. 7, pp. 810–818, July 2009.

[Sant 12] A. Santra, R. Kumar, P. Sharma, C. Bal, A. Kumar, P. K. Julka, and A. Malhotra. "F-18 FDG PET-CT in patients with recurrent glioma: Comparison with contrast enhanced MRI". *European Journal of Radiology*, Vol. 81, No. 3, pp. 508–513, March 2012.

[Sawi 98] P. Sawin, P. Hitchon, K. Follett, and J. Torner. "Computed Imaging-Assisted Stereotactic Brain Biopsy: A Risk Analysis of 225 Consecutive Cases". *Surgical Neurology*, Vol. 49, No. 6, pp. 640–649, June 1998.

[Sche 98] H. R. Schelbert, C. K. Hoh, H. D. Royal, M. Brown, M. N. Dahlbom, F. Dehdashti, and R. L. Wahl. "Procedure Guideline for Tumor Imaging Using Fluorine-18-FDG". *Journal of Nuclear Medicine*, Vol. 39, No. 7, pp. 1302–1305, July 1998.

[Schr 03] N. U. Schramm, G. Ebel, U. Engeland, T. Schurrat, M. Behe, and T. M. Behr. "High-resolution SPECT using multipinhole collimation". *IEEE Transactions on Nuclear Science*, Vol. 50, No. 3, pp. 315–320, June 2003.

[Schu 06] G. K. von Schulthess, H. C. Steinert, and T. F. Hany. "Integrated PET/CT: Current Applications and Future Directions". *Radiology*, Vol. 238, No. 2, pp. 405–422, February 2006.

[Seo 09] Y. Seo, C. M. Aparici, M. R. Cooperberg, B. R. Konety, and R. A. Hawkins. "In Vivo Tumor Grading of Prostate Cancer Using Quantitative 111In-Capromab Pendetide SPECT/CT". *Journal of Nuclear Medicine*, Vol. 51, No. 1, pp. 31–36, January 2009.

[Serv 11] A. Server, B. Kulle, Øystein B. Gadmar, R. Josefsen, T. Kumar, and P. H. Nakstad. "Measurements of diagnostic examination performance using quantitative apparent diffusion coefficient and proton MR spectroscopic imaging in the preoperative evaluation of tumor grade in cerebral gliomas". *European Journal of Radiology*, Vol. 80, No. 2, pp. 462–470, November 2011.

[Shch 08] S. Shcherbinin, A. Celler, T. Belhocine, R. Vanderwerf, and A. Driedger. "Accuracy of quantitative reconstructions in SPECT/CT imaging". *Physics in Medicine and Biology*, Vol. 53, No. 17, p. 4595, September 2008.

[Shes 03] D. J. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. crc Press, 2003.

[Sido 06] C. Sidoti and U. Agrillo. "Chronic Cortical Stimulation for Amyotropic Lateral Sclerosis: A Report of Four Consecutive Operated Cases after a 2-Year Follow-up: Technical Case Report". *Neurosurgery*, Vol. 58, No. 2, p. E384, February 2006.

[Sore 03] M. Soret, P. M. Koulibaly, J. Darcourt, S. Hapdey, and I. Buvat. "Quantitative Accuracy of Dopaminergic Neurotransmission Imaging with 123I SPECT". *Journal of Nuclear Medicine*, Vol. 44, No. 7, pp. 1184–1193, July 2003.

[SPSS 10] *SPSS Statistics 19.0.0.1*. IBM, 2010.

[Stan 05] G. J. Stanisz, E. E. Odrobina, J. Pun, M. Escaravage, S. J. Graham, M. J. Bronskill, and R. M. Henkelman. "T1, T2 relaxation and magnetization transfer in tissue at 3T". *Magnetic Resonance in Medicine*, Vol. 54, No. 3, pp. 507–512, September 2005.

[Stei 09] S. Steidl. *Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech*. Logos Verlag, 1 Ed., 2009.

[Swol 03] B. Swolin, P. Simonsson, S. Backman, I. Löfqvist, I. Bredin, and M. Johnsson. "Differential counting of blood leukocytes using automated microscopy and a decision support system based on artificial neural networks - evaluation of DiffMasterTM Octavia". *Clinical and Laboratory Haematology*, Vol. 25, No. 3, pp. 139–147, June 2003.

[Tang 96] H. R. Tang, J. K. Brown, and B. H. Hasegawa. "Use of X-ray CT-defined regions of interest for the determination of SPECT recovery coefficients". In: *Nuclear Science Symposium, 1996. Conference Record., 1996 IEEE*, pp. 1840–1844, 1996.

[Tsuc 08] T. Tsuchida, H. Takeuchi, H. Okazawa, T. Tsujikawa, and Y. Fujibayashi. "Grading of brain glioma with 1-11C-acetate PET: comparison with 18F-FDG PET". *Nuclear Medicine and Biology*, Vol. 35, No. 2, pp. 171–176, February 2008.

[Tsui 94] B. M. Tsui, E. C. Frey, X. Zhao, D. S. Lalush, R. E. Johnston, and W. H. McCartney. "The importance and implementation of accurate 3D compensation methods for quantitative SPECT". *Physics in Medicine and Biology*, Vol. 39, No. 3, pp. 509–530, March 1994.

[Upad 11] N. Upadhyay and A. D. Waldman. "Conventional MRI evaluation of gliomas". *British Journal of Radiology*, Vol. 84, Special Issue 2, No. 2, pp. S107–S111, December 2011.

[Vand 07] E. Vandervoort and et al. "Implementation of an iterative scatter correction, the influence of attenuation map quality and their effect on absolute quantitation in SPECT". *Physics in Medicine and Biology*, Vol. 52, No. 5, p. 1527, March 2007.

[Vapn 82] V. Vapnik. *Estimation of Dependences Based on Empirical Data*. Springer Verlag, 1982.

[Webe 00] W. A. Weber, H.-J. Wester, A. L. Grosu, M. Herz, B. Dzewas, H.-J. Feldmann, M. Molls, G. Stöcklin, and M. Schwaiger. "O -(2-[18F] Fluoroethyl)-tyrosine and L -[methyl-11C] methionine uptake in brain tumours: initial results of a comparative study". *European Journal of Nuclear Medicine and Molecular Imaging*, Vol. 27, No. 5, pp. 542–549, May 2000.

[Weis 04] N. Weisenfeld and S. Warfteld. "Normalization of joint image-intensity statistics in MRI using the Kullback-Leibler divergence". In: *Biomedical Imaging: Nano to Macro, 2004. IEEE International Symposium on*, pp. 101–104 Vol. 1, 2004.

[Well 04] M. Weller. "Interdisziplinäre S 2 Leitlinie für die Diagnostik und Therapie der Gliome des Erwachsenenalters". Tech. Rep., Deutsche Krebsgesellschaft, 2004.

[Well 98] R. G. Wells, A. Celler, and R. Harrop. "Analytical calculation of photon distributions in SPECT projections". *IEEE Transactions on Nuclear Science*, Vol. 45, No. 6, pp. 3202–3214, December 1998.

[Werb 74] P. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, Cambridge, MA, 1974.

[Will 08] K. Willowson, D. Bailey, and C. Baldock. "Quantitative SPECT reconstruction using CT-derived corrections". *Physics in Medicine and Biology*, Vol. 53, No. 12, p. 3099, June 2008.

[Will 12] K. Willowson, N. Forwood, B. Jakoby, A. Smith, and D. Bailey. "Quantitative (90)Y image reconstruction in PET". *Medical Physics*, Vol. 39, No. 11, pp. 7153–7159, November 2012.

[Will 13] M. Willemink, T. Leiner, P. Jong, L. Heer, R. Nievelstein, A. Schilham, and R. Budde. "Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality". *European Radiology*, Vol. 23, No. 6, pp. 1632–1642, June 2013.

[Yush 06] P. A. Yushkevich, J. Piven, C. H. Hazlett, G. R. Smith, S. Ho, J. C. Gee, and G. Gerig. "User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability". *Neuroimage*, Vol. 31, No. 3, pp. 1116–1128, July 2006.

[Zach 09] 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". *Magnetic Resonance in Medicine*, Vol. 62, No. 6, pp. 1609–1618, December 2009.

[Zach 11] E. Zacharaki, V. Kanas, and C. Davatzikos. "Investigating machine learning techniques for MRI-based classification of brain neoplasms". *International Journal of Computer Assisted Radiology and Surgery*, Vol. 6, No. 6, pp. 821–828, November 2011.

[Zaid 03] H. Zaidi and B. Hasegawa. "Determination of the Attenuation Map in Emission Tomography". *Journal of Nuclear Medicine*, Vol. 44, No. 2, pp. 291–315, February 2003.

[Zaid 06a] H. Zaidi. *Quantitative Analysis in Nuclear Medicine Imaging*. Springer, Berlin, 1st Ed., 2006.

[Zaid 06b] H. Zaidi, E. C. Frey, and B. M. W. Tsui. *Collimator-Detector Response Compensation in SPECT*, pp. 141–166. Springer US, 2006.

[Zano 12] P. Zanotti-Fregonara, J.-S. Liow, C. Comtat, S. S. Zoghbi, Y. Zhang, V. W. Pike, M. Fujita, and R. B. Innis. "Image-derived input function in PET brain studies: blood-based methods are resistant to motion artifacts". *Nuclear Medicine Communications*, Vol. 33, No. 9, pp. 982–989, September 2012.

[Zein 10] J. Zeintl, A. H. Vija, A. Yahil, J. Hornegger, and T. Kuwert. "Quantitative Accuracy of Clinical 99mTc SPECT/CT Using Ordered-Subset Expectation Maximization with 3-Dimensional Resolution Recovery, Attenuation, and Scatter Correction". *Journal of Nuclear Medicine*, Vol. 51, No. 6, pp. 921–928, June 2010.

[Zoll 10] F. G. Zöllner, K. E. Emblem, and L. R. Schad. "Support vector machines in DSC-based glioma imaging: Suggestions for optimal characterization". *Magnetic Resonance in Medicine*, Vol. 64, No. 4, pp. 1230–1236, October 2010.

[Zoll 12] F. Zöllner, K. Emblem, and L. Schad. "SVM-based Glioma Grading: Optimization by Feature Reduction Analysis". *Zeitschrift für medizinische Physik*, Vol. 22, No. 3, pp. 205–214, September 2012.

[Zulp 12] N. Zulpe and V. Pawar. "GLCM Textural Features for Brain Tumor Classification". *IICSI International Journal of Computer Science issues*, Vol. 9, No. 3, pp. 354–359, May 2012.