



DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF BREMEN

Doctoral Thesis

**Machine Learning and Multi-Modal Image
Analysis for Image-Guided Therapy and
Clinical Decision Support**

Jennifer Nitsch



AG Medical Image Computing
 Universität Bremen

 **Fraunhofer**
MEVIS

I confirm that this doctoral thesis is my own work and I have documented all sources and material used.

Bremen, 15.01.2020

Jennifer Nitsch

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
- Dr.-Ing. -

vorgelegt im Fachbereich 3
Mathematik und Informatik
der Universität Bremen
Januar 2020

Gutachter:

Prof. Dr. Ron Kikinis
Universität Bremen

Prof. Dr.-Ing. Bernhard Preim
Otto-von-Guericke-Universität Magdeburg

Datum des Promotionskolloquiums: 27. Februar 2020

Copyright © 2020 Jennifer Nitsch

Acknowledgments

The writing and completion of this thesis would not have been possible without assistance, support, and guidance of the following people and I would like to show my gratitude:

I first thank my advisor Ron Kikinis for his guidance, support, brainstorming, and motivation to help me finish this thesis. I am lucky to have worked intensively with him and I would not have made the progress, both personally and academically, without his guidance and mentoring during the projects that made up this thesis. Without him I would not have half of my anatomical and pathological knowledge. I enjoyed the countless radiological reviews to be able to distinguish anatomical structures in ultrasound and MR imaging and learning the correct medical terminology. Even during his busiest days and with the Atlantic Ocean in between, he responded instantly to emails and he always took time to discuss any issues on my mind.

Furthermore, I would like to thank Bernhard Preim for reviewing and critiquing this work and being member of the dissertation committee.

Much gratitude goes to Jan Klein, Hans Meine, and Jan Moltz, who supported the research of this dissertation at Fraunhofer MEVIS and at the University of Bremen. They have been honest mentors and I would not have made the same progress both, professionally and personally, without them. Thank you for your guidance and your support during the past years and for helping plan the steps to success. Thanks to Richard Rascher-Friesenhausen for sparking my interest and passion for medical image computing during the first years of my studies. My gratitude also goes to my many great colleagues at MEVIS and at the University of Bremen.

I am honored to have worked in partnerships with excellent collaborators. Special thanks go to Dorothea Miller, Ulrich Sure, Oliver Gembruch, Philipp Dammann, and Karsten Wrede from the Clinic of Neurosurgery, Essen University Hospital, who supported the research of Part I of this thesis. Also, the research of Part II of this thesis would not have been possible without Tina Kapur who, once again, connected the right people at

the right time to combine individual strengths in order to conduct excellent research collaboratively. Much gratitude goes to Mike Halle for his dedication and design of planning our joint collaboration during my one-year research stay at the Surgical Planning Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, USA. I thank him for his guidance, assistance, and support during my research year. He set up an excellent research team and it was a pleasure to work together with Jordan Sack, April Wall, and Anna Rutherford on the topics of Part II of this thesis. During my research stay I was grateful for the feedback on my research from Andriy Fedorov, Steve Pieper, Alex Golby, and Sarah Frisken. Many thanks to my former officemates Isaiah Norton and Simon Drouin for creating a friendly and inspiring atmosphere and who enjoyed the view over Fenway Park together with me. Special thanks to Katie Mastrogiacomo who made my time at 1249 Boylston Street especially delightful. I thank Katie and Gracia Kranz from the University of Bremen for their unconditional administrative support.

I also thank the DAAD (the German Academic Exchange Service) for the one-year research scholarship for doctoral students and the University of Bremen for the BremenIDEA travel stipend. Without this funding, the research stay in Boston would not have been possible.

To my close friends who have been around the last years, including Sarah Bartels, Stefan Bartels, Nora Breutigam, Nina Drücker, Sabrina Haase, Daniel Harms, and Janka Wilbig.

And finally, to my parents Regina and Joachim Nitsch who supported me in everything when needed for which I will always be grateful.

Abstract

This PhD thesis makes two different contributions to the field of medical image analysis.

The first is a novel approach for using segmentation-based image registration in the field of real-time image-guided glioma surgery. This was the first approach demonstrating that a segmentation-based registration of preoperative magnetic resonance imaging (MRI) with intraoperative ultrasound (iUS) imaging is more accurate, more robust, and faster than purely intensity-based registration. For the registration, central cerebral structures were segmented and served as a guiding-frame. Moreover, the intraoperative applicability of MRI-iUS registration within a strict time frame during glioma resection could be demonstrated with an overall computation time of about 91 seconds, which also included computing the segmentations.

The second contribution is a proof of concept of a novel automatic image-based severity assessment of liver cirrhosis for end-stage liver disease patients. Currently, the degree of severity is determined by a blood test from which the so-called MELD score is calculated. This score is also used for ranking patients on the liver transplant list. For an image-based assessment, quantitative hepato-splenic radiomic features are extracted and analyzed to determine their importance for classifying the severity of liver cirrhosis. Different clinically established metrics for assessing the disease severity were employed and compared with extracted radiomic features. Furthermore, a control group could be identified that was scanned with the exact same MR imaging protocol. This allowed a direct comparison of radiomic features extracted from cirrhotic patients with those from non-cirrhotic patients. Apart from automatic cirrhosis detection, important findings can be reported from our studies that are novel in this field. Some findings are also in line with perceptions of radiologists but were not quantified before. To the best of my knowledge, no other research group has previously investigated the specific radiomic features for risk-stratification of cirrhotic patients and for automatic cirrhosis classification.

Kurzfassung

Diese Doktorarbeit leistet zwei unterschiedliche Beiträge im Bereich der medizinischen Bildverarbeitung. Der erste ist ein neuartiger Ansatz zur segmentierungsbasierten Bildregistrierung im Bereich der bildgestützten Gliomchirurgie in Echtzeit. Zum Zeitpunkt der Veröffentlichung war dies der erste Ansatz, der gezeigt hat, dass eine segmentierungsbasierte Registrierung von Magnetresonanzbildern (MRT) mit intraoperativer Ultraschallbildgebung (iUS) in der Gliomchirurgie genauer, robuster und schneller ist als eine rein intensitätsbasierte Registrierung. Für die Registrierung wurden zentrale zerebrale Strukturen segmentiert und dienten als Stützstrukturen. Darüber hinaus konnte die intraoperative Anwendbarkeit der MRI-iUS-Registrierung in der Gliomchirurgie bewiesen werden, da durchschnittlich nur eine Rechenzeit von 91 Sekunden für die Registrierung benötigt wurde, die auch die Berechnung der Segmentierungen einschließt.

Der zweite Beitrag ist eine neuartige, automatische, bildbasierte Schweregradbeurteilung der Leberzirrhose für Patienten mit Lebererkrankungen im Endstadium. Derzeit wird der Schweregrad durch eine Blutuntersuchung ermittelt auf deren Grundlage der sogenannte MELD-Score berechnet wird. Dieser Score wird auch für die Einstufung von Patienten auf der Lebertransplantationsliste verwendet. Für eine bildbasierte Beurteilung werden quantitative hepato-splenische radiomische Merkmale extrahiert und in ihrer Bedeutung für die Einstufung des Schweregrades der Leberzirrhose analysiert. In dieser Arbeit wurden verschiedene, klinisch etablierte Metriken zur Beurteilung des Schweregrades der Erkrankung verwendet und mit extrahierten radiomischen Merkmalen verglichen. Darüber hinaus konnte eine Kontrollgruppe identifiziert werden, die mit dem exakt gleichen MR-Bildgebungsprotokoll gescannt wurde. Dies ermöglicht einen direkten Vergleich der radiomischen Merkmale von Zirrhose-Patienten mit denen von Nicht-Zirrhose-Patienten. Neben einer automatischen Zirrhose-Erkennung gibt es weitere wichtige neue Erkenntnisse. Einige Erkenntnisse stimmen auch mit der Wahrnehmung der Radiologen überein, wurden aber bisher nicht quantifiziert. Meines Wissens hat bisher keine andere Forschungsgruppe die spezifischen radiomischen Merkmale und ihre Bedeutung für die Risikostratifizierung von Zirrhosepatienten und die automatische Zirrhose-Klassifikation untersucht.

Contents

Acknowledgments	v
Abstract	vii
Kurzfassung	ix
1 Introduction and Background	1
1.1 About this thesis	1
1.1.1 Part I: Multi-Modal Image Segmentation of Cerebral Structures for Improving MRI-US Registration for Image-Guided Glioma Surgery	2
1.1.2 Part II: Clinical Decision Support for End-Stage Liver Disease	4
1.2 The Images	6
1.2.1 Magnetic Resonance Imaging	6
1.2.2 Ultrasound Imaging	7
1.2.3 Magnetic Resonance Imaging vs. Intraoperative Ultrasound Imaging	9
1.3 Pathologies and Anatomy Needed for this Thesis	11
1.3.1 Part I: Pathology of Gliomas	11
1.3.2 Part I: Anatomy of Central Cerebral Structures	12
1.3.3 Part II: Pathology and Known Anatomical Features of Liver Cirrhosis	15
1.4 Machine Learning Techniques	17
1.4.1 Classical Machine Learning	17
1.4.2 Deep Learning	19
1.4.3 Classical Machine Learning vs. Deep Learning	21
1.5 Object-Based Image Analysis (OBIA)	23
1.6 Imaging Informatics	26
2 PART I: Multi-Modal Image Segmentation of Cerebral Structures for Improv- ing MRI-US Registration for Image-Guided Glioma Surgery	29
2.1 Overview of Approaches Implemented for Part I of this Thesis	31

2.2	Introduction and Motivation	34
2.2.1	Registration for US-MRI fusion	35
2.2.2	Segmentation of US images	36
2.3	General Material and Methods of Part I	37
2.3.1	Data Sets for preMRI and iUS Image Fusion and Segmentation . . .	37
2.3.2	Evaluation Methods for Segmentation and Registration Results . . .	39
2.4	Approach 1: Segmentation of the Cerebral Falx and Adjacent Gyri	42
2.4.1	Segmenting Perifalcine Structures in Pre-Resection US	43
2.4.2	Results	46
2.4.3	Discussion	47
2.5	Approach 2: Segmentation of Perifalcine Structures Using OBIA	48
2.5.1	Segmenting Perifalcine Structures Utilizing OBIA	49
2.5.2	Results	50
2.5.3	Discussion	51
2.5.4	Conclusion	53
2.6	Approach 3: Corresponding MRI Segmentations of the Perifalcine Region .	54
2.6.1	Segmentation of Falx Cerebri and Tentorium Cerebelli in preMRI .	54
2.6.2	Results	57
2.6.3	Discussion	58
2.7	Approach 4: Automatic MRI-US Segmentations for Improving Registration	60
2.7.1	Segmentation-Based Registration for MRI-US Fusion	61
2.7.2	Results	68
2.7.3	Discussion	72
2.7.4	Conclusion	76
2.8	Approach 5: U-net-like Segmentation of Perifalcine Structures in iUS	77
2.8.1	Neural-Network-Based Segmentation in iUS	78
2.8.2	Results	81
2.8.3	Discussion	81
2.8.4	Conclusion	84
2.9	Conclusion and Outlook Part I	85
3	PART II: Clinical Decision Support for End-Stage Liver Disease	89
3.1	Introduction	92
3.2	Patient Selection and Image Data	94

3.3	MRI-based Radiomic Feature Analysis of End-Stage Liver Disease for Severity Stratification	97
3.3.1	Material and Methods	98
3.3.2	Results	99
3.3.3	Discussion	104
3.4	Prognostic Value of MRI-based Radiomic Features in End-Stage Liver Disease	107
3.4.1	Material and Methods	108
3.4.2	Results	109
3.4.3	Discussion	112
3.4.4	Conclusion	116
3.5	Important Radiomic Features for Healthier and Sicker Cirrhotic Patients . .	117
3.5.1	Material and Methods	118
3.5.2	Results	119
3.5.3	Discussion	120
3.6	Conclusion and Outlook Part II	126
 List of Figures		 133
 List of Tables		 135
 Bibliography		 137

1 Introduction and Background

This chapter gives a short introduction and overview of the topics covered in this dissertation. Furthermore, the fundamental background knowledge is provided, which is further completed in the Chapters of Part I and Part II that represent the two research projects this thesis is based on.

The background information given in this Chapter includes in Section 1.2 a short description of the used imaging modalities, in Section 1.3 an outline of the needed pathological and anatomical background, and in Section 1.4 an introduction to employed machine learning techniques. The subsequent Section 1.5 describes the basics of object-based image analysis as used in this thesis. Conclusively, Section 1.6 discusses the highly topical field of imaging informatics and its role for data acquisition, management, and preprocessing the in context of this thesis.

1.1 About this thesis

This PhD thesis fundamentally consists of two different topics within the field of medical image computing and analysis and implies a strong focus on clinically oriented research. These two topics are described in the two main chapters of the thesis:

- Part I: Multi-Modal Image Segmentation of Cerebral Structures for Improving MRI-US Registration for Image-Guided Glioma Surgery.
- Part II: Clinical Decision Support for End-Stage Liver Disease.

This thesis combines conventional medical image computing and machine learning techniques in order to develop very specialized approaches that fit the application scenarios. Furthermore, the advantages of modern deep learning techniques are used where appropriate.

Since the study of Part I was started with a limited patient data base that was gradually expanded, the subsequently developed approaches evolved in parallel. At first, a large amount of a priori knowledge was necessary in order to define specific image features

to segment the target regions. Later, with a larger amount of training data, these specific features could be automatically extracted from the respective image regions and a classifier learned the most important features of the target objects and regions. Finally, as conclusive work in Part I, a neural network is trained on the basis of the now considerably larger patient data base with carefully selected training data of the target regions. A combination of the attained results from the deep learning-based approach, together with classical image postprocessing techniques, demonstrated a significant additional boost in segmentation accuracy.

In Part II of this thesis, deep learning techniques were used for the tedious and time-consuming task of basically segmenting the target organs in magnetic resonance imaging (MRI). However, for the central medical image analyses and for the feature extraction task from segmented organs, it was not aimed to rely on deep learning techniques with difficult accessibility to the important features for classification. Rather, a well-known publicly available library with a large number of modifiable and descriptive radiomic features was employed as the basis to train classical machine learning classifiers for different experiments. As the first researchers to develop an image-based severity assessment for end-stage liver disease patients, the overall objective was proof of concept and the comparison and verification of the most important features from these experiments with clinical expectations of known anatomical and physiological changes during progressive stages of end-stage liver disease.

1.1.1 Part I: Multi-Modal Image Segmentation of Cerebral Structures for Improving MRI-US Registration for Image-Guided Glioma Surgery

The survival of low- or high-grade glioma (a type of brain tumor, see Section 1.3.1 for more information) patients correlates with the extent of tumor resection, whereas the patient's quality of life depends on the neuro-functional outcome after the surgical intervention [21, 17, 23, 165]. Knowing the location of the tumor and its relation to eloquent structures is vital for an optimal outcome after glioma surgery [108, 116].

In this application scenario, so-called neuronavigation systems are employed that allow an intraoperative, e.g., optical tracking of specific instruments and enable a visualization of, for instance, a needle tip within the preoperative planning data set on a central monitor in the operating room [34, 98, 78, 149].

While the tracking of instruments in relation to the patient's brain and structures of risk is very useful in identifying the tumor location and optimal surgical path towards it, the difference between the preoperative planning data set (which is usually a MRI of the

brain) and the the current intraoperative situation, increases with the amount of resected tumor volume [17, 98].

Likewise, the thereby induced tissue shift and deformations are increased, which also reduces the navigation accuracy [24].

Navigated intraoperative ultrasound (hereinafter also referred to as "intraoperative ultrasound" or "iUS") is a low-cost imaging modality that offers acquisition of ultrasound volumes in real-time when employed in conjunction with a neuronavigation system [34, 107, 109, 115].

An intraoperative image fusion of both modalities and, therefore, adapting the preoperative MRI (preMRI) with the deformations recorded in iUS imaging, can update and illustrate the respective changes in real-time for the medical team during the neurosurgical procedure [98]. However, accurate real-time multi-modal registration of preoperative MRI with intraoperatively acquired ultrasound imaging is not yet solved with a sufficient amount of accuracy [24, 34, 160, 50].

This research was conducted with clinical partners from the neurosurgical department at the University Hospital in Essen, Germany.

To the best of our knowledge, we were the first research group evaluating central cerebral structures for improving image registration of both modalities and we demonstrated applicability through superiority in speed, robustness, and accuracy over purely intensity-based registration.

Objectives of Part I: The goal of Part 1 was to implement a segmentation-based registration that uses central cerebral structures as reference frame. The idea is to overcome intensity-based differences of both imaging modalities and to increase the registration's robustness, speed, and accuracy by incorporating segmentations.

Overall, the registration should not exceed a maximum time window of three minutes: The recording of an ultrasound sweep via the navigation system occupies approximately two minutes. A 3D reconstruction of the thereby acquired 2D US slices into a volume and the computation of segmentations can be performed in parallel within this time window. In addition, the ultrasound device must be moved out of the immediate vicinity of the operating table again. This results in a total ideal computation time of three minutes, in order to guarantee that the surgical team can continue working without delay on the basis of the now updated preMRI-iUS fusion image.

The initial registration provided by the neuronavigation system is a landmark-based registration that takes facial fiducials of the patient's head into account to perform a

rigid registration with the preMRI [78]. This process is purely landmark-based and no image content is taken into account for the initial patient-to-navigation-system registration [78]. The iUS probe with its scanning width and depth is calibrated with a specific ultrasound phantom prior surgery with the navigation system [8]. The initial registration error of the rough registration of the navigation system is often very clearly visible on the monitor in the operating room when comparing the preMRI with the iUS. Ideally, this project should develop a registration method that allows a rigid registration with a mean Target Registration Error (mTRE) of less than 3.5 mm, which would be acceptable to our partners for a rigid registration. Furthermore, as a proof of concept, the segmentation-based registration method must demonstrate equal or superior registration accuracy than provided by conventional intensity-based registration.

User interactions should be kept to a minimum and should be realizable under real conditions in the operating room. The recording of the ultrasound volume already requires the use of the touch screen monitor of the neuronavigation system [8] and a minimum of interaction would be tolerated to perform/initiate the registration process.

1.1.2 Part II: Clinical Decision Support for End-Stage Liver Disease

The topics of Part II of this thesis were part of a one-year research stay at the Brigham and Women's Hospital (BWH), Harvard Medical School, Boston, USA. The research project was conducted together with collaborative partners from the Surgical Planning Laboratory (SPL) and the Division of Gastroenterology, Hepatology, and Endoscopy both affiliated with and located at the BWH.

The project's objective was to evaluate the feasibility of a severity assessment of liver cirrhosis of end-stage liver disease patients purely based on MR imaging of the abdomen. Cirrhosis is generally defined as the progressive transformation of normal liver tissue into non-functional scar tissue. The pathology of liver cirrhosis is described in detail in Section 1.3.3.

Currently, the severity of cirrhosis is assessed by a blood test that is calculated into the so-called MELD (Model for End-stage Liver Disease) score [97]. The MELD scoring system aims to identify high-risk patients by stratifying potential liver transplant recipients by their disease severity by estimating their 90-day mortality [53, 52].

The MELD was developed by the Organ Procurement and Transplantation Network (OPTN) and the United Network for Organ Sharing (UNOS) and is also used by Euro-transplant to risk-stratify patients on the liver transplantation list [97, 28]. The MELD score is determined to have a fair and objective score. However, being subject to technical

laboratory variability, it may well make a difference in which hospital or laboratory the blood parameters are extracted [68].

Before the MELD scoring system, the Child-Pugh score (also termed Child-Turcotte-Pugh-Score) was used to assess the prognosis and necessity for liver transplantation of patients with liver cirrhosis [112].

Besides blood parameters, this score also included symptoms such as ascites and hepatic encephalopathy [103, 111]. A major criticism of this score was that the severity of these symptoms was assessed solely by the treating physician and that the presence and degree of symptoms and severity were difficult to measure uniformly and objectively and, thus, subject to observer-dependent variability [103]. Therefore, this score was also highly criticized for its potential manipulability by physicians [103].

An MR image-based score may have the potential to better support the true urgency for a transplant for liver disease, e.g., by looking for (MRI) patterns of disease throughout the abdomen, by providing physicians with information that might allow them to defer transplants for people already too sick, and by supporting to re-include objectively measurable symptoms degrading quality of life if they could be assessed objectively by an image-based scoring system (re-include useful parameters from the Child-Pugh score).

Patients suffering from liver cirrhosis are usually regularly screened for hepatocellular carcinoma (HCC), a type of liver cancer for which cirrhotic patients have a significantly increased risk, which would provide the basis for an image-based screening and assessment for the severity of liver cirrhosis [155, 111].

Objectives of Part II: The aim of the research project of Part II was to utilize these MRI scans for a quantitative radiomic feature extraction in conjunction with known metrics for severity assessment of liver cirrhosis to evaluate the feasibility of a novel image-based severity assessment. The overall objective is to increase the accuracy, generalizability, and, therefore, fairness in risk stratification of end-stage liver disease patients.

1.2 The Images

This chapter briefly describes the medical imaging modalities used in this thesis. For Part I magnetic resonance imaging (MRI) scans and B-mode ultrasound scans of the brain were used as imaging techniques for prospective data acquisition. For the analyses in Part II, MRI scans from the abdomen were acquired retrospectively from the hospital data base.

1.2.1 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) scanners utilize strong magnetic fields, magnetic gradients, and radio waves to produce tomographic images of the human body [6, 105, 75]. A major advantage of this medical imaging technique is that no ionizing radiation or X-rays are involved during image acquisition [6, 88, 59]. A disadvantage of MRI scanners is that the acquisition of images takes longer than in computed tomography (CT) scanners that use X-rays for image generation: For instance, a high-resolution head scan takes a few seconds in modern CT scanners, whereas a high-resolution MRI scan can take up to ten minutes [6]. MR imaging is also much louder and requires the use of earplugs or noise-protection headphones during image acquisition [88]. However, MRI is the preferred imaging technique for the diagnostic visualization of soft tissue and allows to acquire different so-called MRI *sequences* within one session [39, 105]. MRI sequences are further explained in the next paragraph.

MR imaging makes use of the fact that hydrogen atoms, as they occur in the tissue of living organisms, can be induced while placed in an external magnetic field with a radio frequency pulse [59, 75]. This generates detectable radio-frequency signals that are measured by the built-in detectors of the MRI scanner [88]. The magnetic field gradients are used to localize the induced signals in space. MRI scanners basically map the location of water and fat in the body [39, 105]. Different contrasts are the result of different relaxation properties of anatomical structures which depends on the quantity of hydrogen atoms they contain [75]. The variation of parameters of pulse sequences leads to the acquisition of different MR sequences, e.g., T1- or T2-weighted images, FLAIR or DTI. [6, 39, 44, 59, 75, 88, 105]

MR Imaging Protocols Used in this Thesis

In Part I and in Part II of this thesis contrast-enhanced T1-weighted MR images were the basis for further analyses.

For Part I the basic imaging protocol consisted of a gadolinium-enhanced MRI of the brain acquired a few days before the planned glioma surgery. This represents a popular scan used as the basic preMRI during image-guided glioma surgery [98, 145, 160]. The MRI scans were scans from different manufacturers with different magnetic field strengths.

In Part II a state-of-the-art hepatocellular carcinoma (HCC) screening protocol was used for image acquisition which consisted of a series of multiple MRI scans [74]. Basically, for HCC surveillance a dynamic scanning series during contrast agent uptake is made and the following stages are differentiated: late arterial phase, portal venous phase, and equilibrium phase (also referred to as "delayed phase") [70, 74, 155].

The employed contrast agent was also gadolinium-based as in Part I. For Part II of this thesis the delayed phase acquired five minutes after contrast injection is considered the valuable imaging phase by radiologists in which the retention of contrast agent in nodular, fibrotic, necrotic, and, therefore, cirrhotic regions can be detected [70, 125]. More details on what kind of images were used for the studies in Part II are described in Section 3.2. Two example MR images of the delayed phase can be seen in Fig. 1.5 on page 16.

The remaining contrast agent leads to a heterogeneous enhancement of the liver tissue with higher intensities in diseased regions [74]. Therefore, utilizing the HCC screening protocol seemed an ideal starting point for an image-based severity assessment of liver cirrhosis.

HCC is ranked as the sixth most commonly diagnosed cancer worldwide and develops in up to 90 % of cases within the context of liver cirrhosis [5, 70]. Cirrhotic patients are under strict HCC surveillance including the above described series of dynamic MRI scans, ideally, every six months [136].

1.2.2 Ultrasound Imaging

Diagnostic ultrasound is a non-invasive, real-time medical imaging modality that uses pulsed sound waves of higher frequencies in order to create a diagnostic image. Different types of diagnostic ultrasound imaging exist and within this thesis the so-called B-mode ("B" for brightness) ultrasound is used and its functionality is briefly described in the following. [60, 95]

The ultrasound waves have higher frequencies than those audible by humans with frequencies above 20 kilohertz [95, 170]. The ultrasound probe uses the piezoelectric effect and alternates between creating and receiving the ultrasound waves (also referred to as "pulsed echo") that are sent through biological tissues [60]. The sent out sound waves are reflected differently by dissimilar tissue types dependent on their acoustic impedance

[170]. The ultrasound probe detects the speed and the strength of the reflected sound waves from diverse tissues from which a cross-sectional grayscale image can be calculated. Another commonly used diagnostic ultrasound technique would be Doppler ultrasound, representing a measurement of the blood flow displayed as colored overlay over a B-mode ultrasound image. [60, 95, 143, 170]

No ionizing radiation is induced and the transmitted sound waves are harmless [60]. For this reason, this real-time imaging technique is the preferred diagnostic imaging for the examination of pregnant women with their unborn children [14]. Another advantage is that ultrasound devices are relatively cheap imaging modalities when compared to the costs of MRI or CT scanners [30, 17].

An apparent disadvantage of ultrasound imaging is that the scanning width and depth is very limited when compared to CT or MRI scans where whole body scans can be acquired [17]. Through movement of the ultrasound probe the scanned body region can be increased but stays limited in its depth [170].

Furthermore, structures with high acoustic impedance, such as air or bone tissue, can absorb and reflect a great deal of the transmitted ultrasound waves, which severely impedes or even blocks the scanning of deeper, underlying anatomical structures [17, 170].

This makes ultrasound an artifact-rich imaging and explains the necessity of ultrasonic gels and fluids directly applied to the head of the ultrasound probe prior image acquisition [61]. Aside from that, so-called "speckles" are characteristic artifacts in ultrasound imaging that represent small scale structural inhomogeneities. These grainy structures are visible in Fig. 1.1 [61].

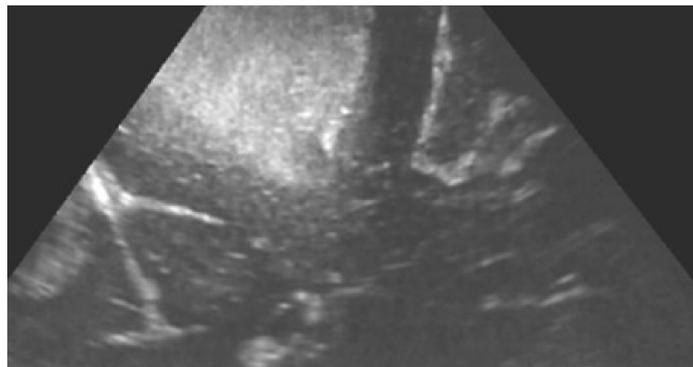


Figure 1.1: Speckles in ultrasound imaging as small granular accentuations in ultrasound imaging resembling small rice grains. Speckles are the most common artifacts and characteristic for ultrasound images [61].

Navigated Intraoperative Ultrasound

In Part I of this thesis navigated ultrasound is used in order to acquire ultrasound image volumes. The position of the ultrasound probe is tracked by a camera system that belongs to a neuronavigation system [50, 170]. By knowing the spatial location of each acquired 2D ultrasound image in real-time, the slice-wise acquired 2D ultrasound images can be reconstructed to 3D volumes by using the tracking information in form of transformation matrices [78, 8]. In the following it is referred to the navigated intraoperative ultrasound only in shortened form as "intraoperative ultrasound" (or abbreviated as "iUS") because of its sole context of application together with a neuronavigation system in this thesis. The detailed image acquisition of iUS volumes is further described in Section 2.3 on page 37 and the intraoperative setup is explained and illustrated in Fig. 2.1 on page 37 and Fig. 2.2 on page 38.

1.2.3 Magnetic Resonance Imaging vs. Intraoperative Ultrasound Imaging

In order to accomplish intraoperative image guidance on preMRI during glioma surgery, a neuronavigation system at our partner's hospital was used to acquire tracked 2D ultrasound images that could be reconstructed to 3D volumes by utilizing the transformation matrices received through optical tracking of the ultrasound probe [78, 8]. The intraoperative setup is further described in Section 2.2.

As described above, ultrasound and MR imaging are very different medical imaging techniques that considerably differ in their physical image generation, resulting in different contrasts and visibility of anatomical structures [44]. Therefore, a basic intensity-based registration does not provide enough spatial references to allow an accurate enough multi-modal image registration [17, 34, 50]. Another specificity of ultrasound images is the limited width and depth of scanned anatomical regions. These limitations are based on the small operating area available for the ultrasound head and are a result of the trade-off between an ultrasound frequency with maximum penetration depth and optimal resolution of the ultrasound image. The greater the frequency, the smaller the penetration depth of the ultrasound. However, the higher the frequency, the higher is the local resolution, i.e., the ability to distinguish between objects that are close to each other.

The MRI scanner can acquire a scan of the whole patient head whereas the ultrasound images and the resulting volumes only contain smaller regions of the patient's brain (see Fig. 1.2).

Alternative intraoperative imaging techniques for image-guided surgery that would

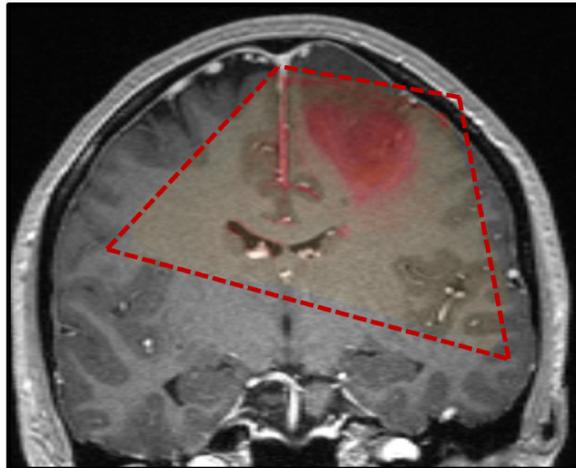


Figure 1.2: The image shows a coronal cross-section of the brain with a red overlay showing the limited image extent of the acquired ultrasound imaging in relation to the whole head scan of the preoperative MRI. The acquired iUS volume represents a volume with a cone-like shape.

provide complete head scans would be intraoperative CT or intraoperative MRI scanners that are far more costly modalities than an ultrasound device and, therefore, are not affordable for most hospitals in an intraoperative setup.

Another huge advantage of ultrasound imaging is the real-time acquisition during surgery with almost no delays for the medical team in the operating room. A neuronavigation system as well as an ultrasound device are basic equipment for neurosurgery and usually available in most neurosurgical departments and would not impose any additional costs in this application scenario.

1.3 Pathologies and Anatomy Needed for this Thesis

This chapter briefly describes the anatomy and pathologies that are important in context of this thesis. Subsequently, it is started with the needed background for image-guided glioma surgery of Part I. Afterwards, liver cirrhosis and related pathological and anatomical features are described in detail for Part II of this dissertation.

1.3.1 Part I: Pathology of Gliomas

Currently, gliomas are the most common brain tumors in adults and are categorized as primary central nervous system (CNS) tumors that infiltrate surrounding tissues [160, 45]. Prognostically, these tumors can be divided into low- and high-grade gliomas [129].

The World Health Organization (WHO) classifies these tumors histologically in grade I to IV. Low-grade gliomas (LGG) represent a grade I and II because their survival rate is higher because they are benign tumors and grow more slowly, whereas high-grade gliomas (HGG) are categorized as malignant tumors with grade III and IV [38]. Low-grade gliomas can progress to high-grade types. The incidence of these tumors is 6 per 100000 persons, which is not high, but they have very poor 10-year survival rates depending on the patient's age of 10 % to 62 % for LGGs and in high-grade tumors these can reduce to only a few months with maximum therapy [20, 38, 99, 140, 157].

Numerous studies have shown that the extent of tumor resection for LGG as well as for HGG has a decisive influence on the patient's prognosis [82, 108]. At the same time, the most radical possible tumor resection is a trade-off between overall survival and the risk of losing neurological functions [21, 116]. Postoperative preservation of independence is associated with longer survival and a better quality of life after surgery among glioma patients [71].

Several technical developments such as the use of neuronavigation systems for better image-guided resection by intraoperative imaging, fluorescence-assisted microsurgery, and electrophysiological monitoring have contributed in their combined usage to increasing the radicality of intracerebral tumor surgery in recent years, while simultaneously increasing patient safety [108, 17, 50].

LGGs and HGGs have very patient-specific manifestations. Besides varying in size and location, these tumors can differ immensely in their texture with both homogeneous and heterogeneous tissue regions. For example, these tumors can have calcifications, necrosis, adjacent cysts, and can be more or less perfused. All this also leads to very different tumor representations in contrast-enhanced MRI as used in Part I of this thesis. [64, 36, 35, 44]

1.3.2 Part I: Anatomy of Central Cerebral Structures

In Part I of this thesis central cerebral structures are segmented, that shall be briefly outlined in this section. Fig. 1.3 shows a comprehensive view of the here described anatomical structures.

Additionally, in Chapter 2.7 on page 61 the main anatomical structures are also illustrated in contrast-enhanced MRI and in ultrasound imaging on page 64.

The here described and displayed anatomical structures are summed up in accordance to *Netter's Atlas of Neuroscience* [31], which is a well known resource in the medical domain for anatomical knowledge combined with detailed illustrations.

Falx Cerebri and Tentorium Cerebelli

The falx cerebri and the tentorium cerebelli are central meningeal folds within the human brain. The falx separates the the left and right hemisphere, whereas the tentorium separates the cerebrum and the cerebellum.

In this thesis, the anatomical structures of the falx and the tentorium with adjacent sulci and gyri are also referred to as the perifalcine region or as the perifalcine structures. The falx and the tentorium might be thin folds but they tend to stay in place even when large

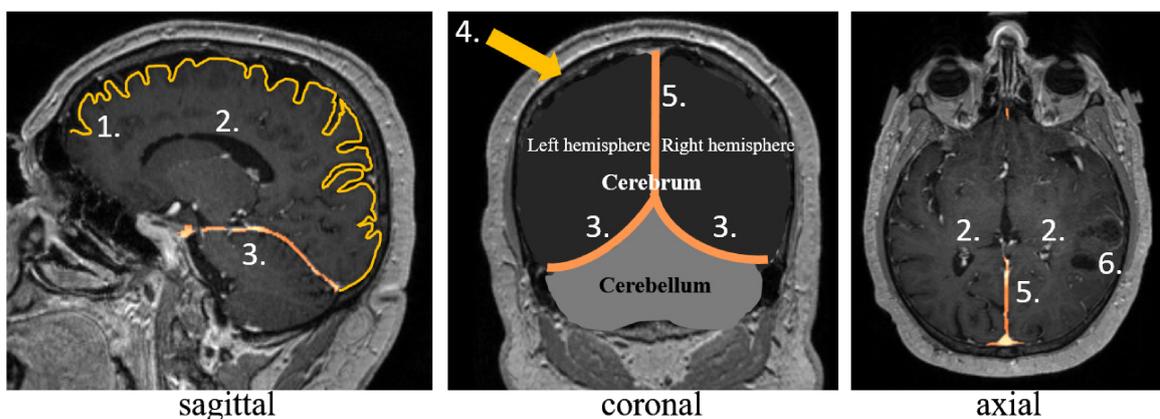


Figure 1.3: Anatomy of central cerebral structures in a sagittal, coronal, and axial view of the brain: 1. Illustrates as yellow outline the sulci and gyri of the cerebral cortex; 2. The black structure depicts an example of a lateral ventricle; 3. Shows the location of the tentorium cerebelli; 4. The arrow points towards the thin meninges containing as outer membrane the dura mater; 5. Shows the location of the falx cerebri; 6. Illustrates an example of a glioma (here: a cystic tumor).

tumors grow in their vicinity as can be seen in Fig 1.4. This feature makes these structures excellent support structures for multi-modal image registration.

Sulci and Gyri

The cerebral cortex is the outer layer of the human brain. Folds or ridges are termed as gyri (singular: gyrus). The contrary are grooves that are referred to as sulci (singular: sulcus). Gyri and sulci are depicted in Fig. 1.3.

Dura Mater

The dura mater is a thin membrane that surrounds the brain and the spinal cord. The dura is the outermost of a total of three layers called meninges and covers and protects the central nervous system. In Fig. 1.3 the dura is represented in the sagittal view of the patient head as the thin layer above the sulci and gyri of the cerebral cortex. In the coronal view the arrow points towards the dura.

Ventricles

The ventricles consist of four connected cavities in which the cerebrospinal fluid is produced: Two lateral ventricles (right and left) one for each hemisphere, the third ventricle, and the fourth ventricle. The ventricular system is a central anatomical structure in the

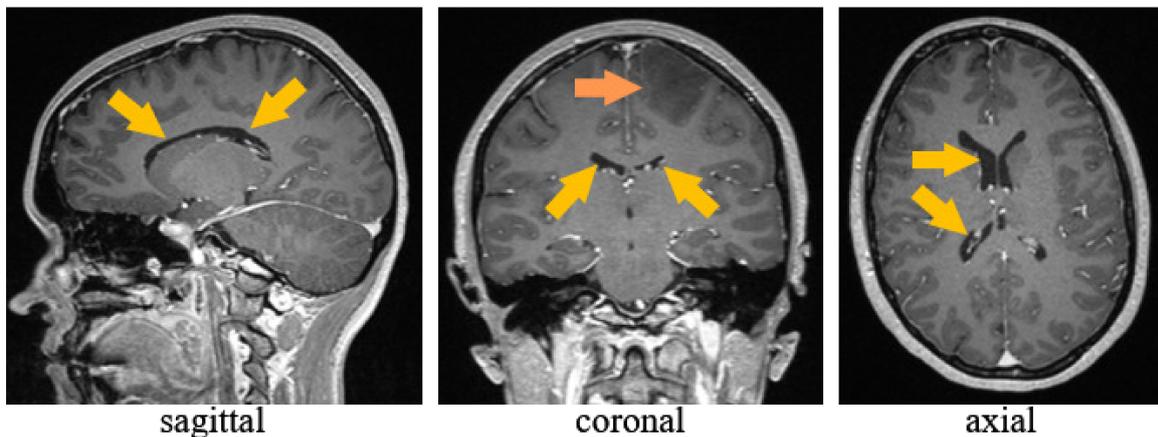


Figure 1.4: Anatomy of the ventricles in contrast-enhanced T1-weighted MRI. The yellow arrows point toward the ventricles in sagittal, coronal, and axial view. The orange arrow points towards a tumor that compresses slightly the right lateral ventricle in the right hemisphere.

brain. The ventricles have each a choroid plexus in which the cerebrospinal fluid (CSF) is produced.

This structure is specifically hyperechogenic in ultrasound imaging and is an example of a structure that is not represented in the same extent in MRI and iUS. This ultrasound hyperechogenicity can reduce the penetration depth of the ultrasound which should be avoided.

On T1-weighted MRI the CSF filled ventricles can be easily detected as large, central cavities with low gray value intensities. In Fig. 1.3 an example of a lateral ventricle can be seen in the sagittal view of the patient head. Fig. 1.4 illustrates the ventricles in an axial, sagittal, and coronal view. It can also be seen that the ventricles behave differently in the vicinity of a tumor than the falx and tentorium. Usually, the right and left lateral ventricles are very symmetric but Fig 1.4 demonstrates that tumors can easily compress and deform ventricles.

The ventricles in iUS imaging are illustrated on page 64 as the two black cavities with low gray value intensities below the falx in the top left and in the top right iUS image.

1.3.3 Part II: Pathology and Known Anatomical Features of Liver Cirrhosis

Cirrhosis is generally defined as the progressive transformation of normal liver tissue into non-functional scar tissue, which can have various etiological origins. The liver tissue transformation process usually takes years before the first symptoms appear.

The etiologies of liver cirrhosis as well as the severity assessment with the MELD scoring system are describes in detail in Chapter 3.1 on page 92.

The following Tab. 1.1 sums up the characteristic pathological changes of liver cirrhosis within the abdomen and, therefore, could be as well visible in abdominal MRI for radiomic feature extraction.

At the beginning of the research in Part II, it was not sure what kind of radiomic features would be detectable on the available image data. To be more precise, it was unknown whether these features would be predominantly visible for the naked eye or if latent image information would be discovered that bear the characteristics of liver cirrhosis. Therefore,

Feature	Comment
Liver surface nodularity	Reduced "smoothness" of the liver surface.
Increased heterogeneity of liver tissue	Visible in MRI imaging and due to bridging fibrosis and cirrhotic tissue in the liver. The unique ability of the liver to regenerate leads to intervening islands, "nodules", within fibrotic (and cirrhotic) tissue within the liver. Fibrosis is a preliminary stage of cirrhosis, but with the potential to regress and heal, while cirrhotic liver damage is irreversible. The impaired blood flow in encapsulated regions leads to an accumulation of contrast agent.
Liver atrophy	The liver shrinks which results in an overall smaller volume. In some areas the liver shrinks faster than in others, leading to pathological features as the decreased right to left lobe volume ratio.
Portal hypertension	Liver cirrhosis increases the resistance to blood flow resulting into higher pressure in the portal venous system and leading to differences in portal venous circulation. This results into visible, global morphological liver features as changes to the liver shape.
Varices	Esophageal varices due to higher blood pressure. Thus, vessels are enlarged and the risk of ruptures and severe internal bleeding is increased.
Splenomegaly	Enlargement of the spleen.
Expanded gallbladder fossa sign	This sign is known as a specific indicator of cirrhosis.
Ascites	Accumulation of fluid in the abdomen.
Sarcopenia	Muscle wasting.

Table 1.1: Pathological features of liver cirrhosis within the abdomen are summed up in this table. The table was attained by several publications with the focus of typical signs of liver cirrhosis [9, 48, 72, 125, 167].

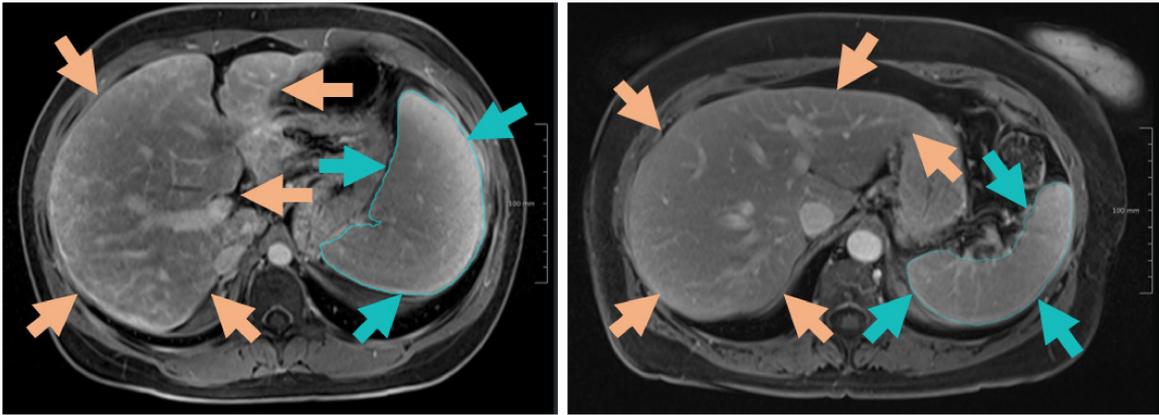


Figure 1.5: This figure depicts the abdomen of two patients in comparison on an axial slice of a contrast-enhanced MRI. The left image illustrates the abdomen of a cirrhotic patient, whereas the right image shows the abdomen of a non-cirrhotic patient. In both images the orange arrows point towards the liver surface and the blue arrows point towards the blue contours of the spleen. The liver is shown without contour in order to also depict the liver surface. The standard resolution of acquired MR images for HCC screening was not sensitive enough to capture the feature of liver surface nodularity. The cirrhotic liver, with a very heterogeneous contrast uptake within the whole liver tissue, can be clearly distinguished from the non-cirrhotic. The non-cirrhotic liver tissue is homogeneously enhanced and the contrast uptake is solely visible within the liver vasculature. Another feature of cirrhosis can be seen in comparison of these two MR images: The cirrhotic patient on the left also has an enlarged spleen (splenomegaly). The spleen of the non-cirrhotic patient demonstrates the otherwise normal size of the spleen within the abdomen.

a literature search was necessary to compare and classify the detected features and to validate if the detected changes are in line with expectations or if new correlations of image features and liver cirrhosis could be revealed. Fig. 1.5 shows the abdomen of a cirrhotic patient on contrast-enhanced MRI in comparison to a non-cirrhotic patient.

1.4 Machine Learning Techniques

This section gives a brief overview of classical machine learning and deep learning techniques and provides the background of techniques used in this thesis.

1.4.1 Classical Machine Learning

The core of classical machine learning solutions is pattern recognition by employing a selection of appropriate features, or by specifically tailored features, that best capture the characteristics of the particular application at hand [162]. This initial step is generally followed by selecting a suitable machine learning algorithm that best fits the classification task.

The especially engineered features, also referred to as handcrafted features, require a comprehensive domain knowledge to thereby generate satisfactory classification results [15, 162]. Consequently, the classification accuracy is highly dependent on the quality of problem-specific features. Selecting or engineering a large amount of features not only takes a considerable amount of time and effort but also does not necessarily lead to an increase in classification accuracy [162].

The contrary is often true, because large numbers of features or parameters can even impair the generalizability of the thereon attained classification models [15, 158]. This dilemma is often referred to as being the "curse of dimensionality" caused by large number of features in conjunction with relatively small sample sizes that can lead to an overfitting on the training data by learning very complex but unnecessary rules from many parameters [153, 158]. A subsequent application of the trained classification model to an independent test data set could reveal if the attained model generalizes well on the classification task [158].

Random Forest Classifier

Random forests are part of the *ensemble* learning methods that are commonly used for classification or regression tasks. The output of a random forest classifier is a class (classification) or a mean prediction (regression). [158, 10, 40]

A random forest classifier builds multiple decision trees and each tree is constructed from a sample of the training set (drawn with replacement). Each tree consists of nodes and each node splits at the best split derived from all features or a random subset of features from the training set. [10]

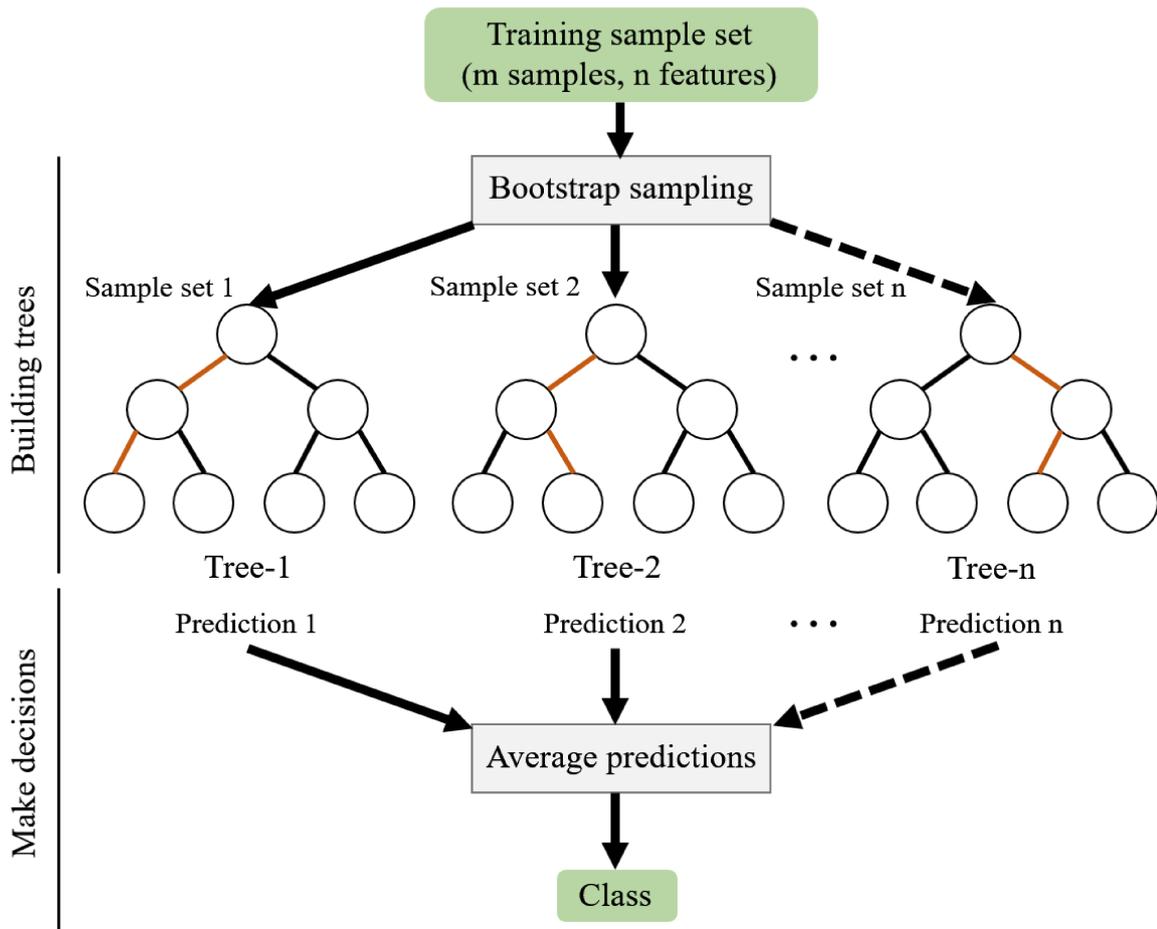


Figure 1.6: This figure illustrates the general functionality of a random forest classifier by building several decision trees and averaging the predictions from all trees.

The decision trees are built independently and in the end the predictions of all forest estimators are combined and averaged. The randomness in building decision trees and, thus, independent classifiers, reduces variance and overfitting and increases generalizability and robustness. [158]

The random forest classifier implementation used in this thesis is based on averaging probabilistic predictions, instead of a majority voting from the individual decision trees for a single class as used in the publication by Breiman [10]. Fig. 1.6 illustrates the here described random forest classifier.

1.4.2 Deep Learning

The most commonly used deep learning architectures in the medical imaging domain are convolutional neural networks often abbreviated as CNNs or Conv Nets [66, 142]. CNNs are deep neural networks and many different architectures exist that are especially engineered to model the specific learning task at hand, e.g. voxel classification/segmentation in this thesis. Unlike models of classic machine learning, CNNs do not require any prior knowledge or human effort to design handcrafted features, instead they provide end-to-end learning [66]. [142]

In order to grasp the success of CNNs, it is often referenced to the famous challenge in computer vision: The ImageNet competition [126]. The challenge involves the classification of more than 14 million images into more than thousand categories. The outstanding success of CNNs in this field transformed the landscape of used machine learning methods and is now the most active field.

The success of deep learning methods can also be seen by examining the submitted contributions to the "Grand Challenges in Biomedical Image Analysis" [25]. By browsing through the challenges on this well-known website in the field of biomedical imaging, it can be perceived that within organ segmentation challenges deep learning methods are leading in the high score lists and even represent the solely used machine learning technique in the respective challenge.

Neural networks consist of different components that are called feed-forward layers: Basically, an input and an output layer that are connected with a series of hidden layers in

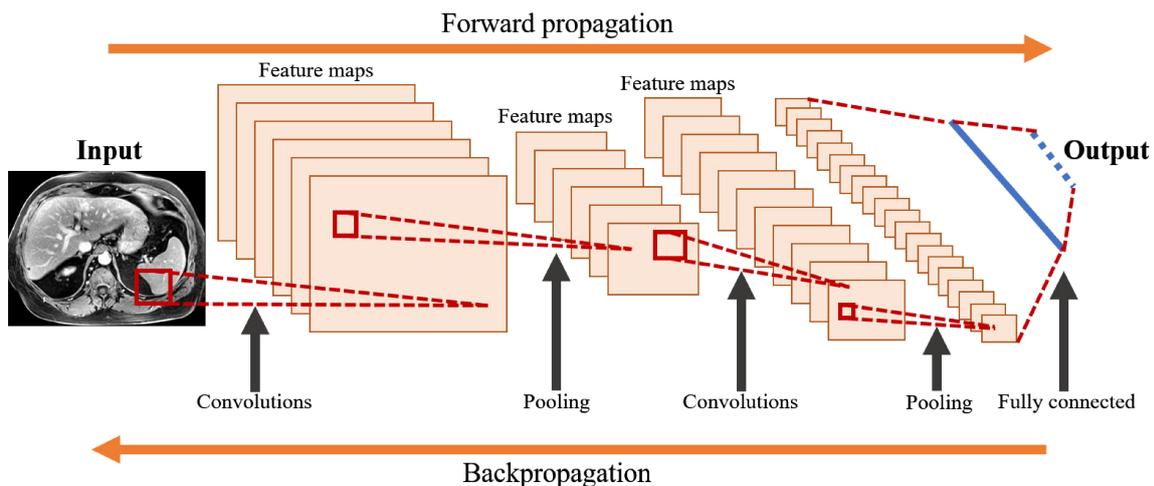


Figure 1.7: Structure of a CNN for image classification.

between. Hidden layers typically consist of a series of convolutional layers [66]. In each layer a certain number of neurons exists that are connected to the previous and successive layers [137]. CNNs in their basic form take an input image which is divided into image regions with fixed sizes and they produce as output a probability value for each class [66]. CNNs transfer the learning problem to an optimization problem of finding the optimal weights with a specific loss function in each neuron (also referred to as convolution) to correctly learn the desired features to generate the output.

With the connections of multiple neurons from multiple layers complex features can be derived and learned. Next to these convolutional layers that are organized in feature maps, a CNN consists of pooling layers [66]. Pooling layers reduce (unnecessary) feature dimensionality by reducing the number of neurons to the most active ones: For instance, it is not necessary to learn the "exact" position of a feature and a rough localization of, e.g., a specific edge feature is sufficient. In this way, it is possible to reduce memory requirements and to increase computational speed. This makes it possible to train even deeper networks for more complex tasks with accelerated speed. Likewise, the receptive field is increased in deeper convolutional layers which makes it possible to increase spatial context information.[66]

The specific features are learned from training data for which the desired output already exists as ground truth. The CNN learns through backpropagation of the layers in which the error is measured that was made to predict the output in order to adjust the weights accordingly to minimize the error and to "learn" from the recently passed example. Likewise, the process of predicting the output from the input image can be termed as forward propagation of the neural network. The more layers a CNN has, the deeper the network which explains the naming of this machine learning method. Fig. 1.7 illustrates a CNN with the aforementioned terms and definitions.

However, deeper series of feature extraction layers increase the number of parameters and the risk of overfitting on the training data remains. The obtained classification results must be evaluated and discussed just as critically as those from classical machine learning methods.

U-Net

A U-net is a type of fully convolutional networks that proved especially successful and award winning for medical image segmentation tasks [147, 124]. The U-net-like architecture is also used in this thesis to segment the liver and the spleen as organs from the abdomen on MR imaging in Part II. The great advantage of the network architecture is that it has

skip connections between the downsampling and upsampling path of the U-net. This makes it possible to provide local information to the global image information while upsampling. Therefore, the U-net can also learn features of specific image regions and combines localization with context information. These are both attributes needed to attain satisfactory segmentation results in 2D and 3D medical images. More details and the network architecture can be found in the original publication by Ronneberger et. al in [124]

1.4.3 Classical Machine Learning vs. Deep Learning

There are some pros and cons of classical machine learning techniques and deep learning methods which shall be briefly described in this section.

Generally, deep learning-based approaches are seen as computationally more expensive because training the networks can take a considerable amount of time. On the contrary, in classical machine learning the classification with a standard classification method is usually faster but this step does not take the longer preprocessing efforts of feature engineering into account. Besides, also the process of feature extraction before classification can be computationally expensive and time-consuming in classical machine learning. Generally, the necessary GPUs for deep learning are regarded as higher hardware costs [137].

Furthermore, larger amounts of training data are seen as necessary for neural networks to learn robust representations of the desired output. This is especially crucial when medical images with pathologies are processed, e.g., for a segmentation task. For instance, gliomas have patient-specific locations, volumes and shapes within the brain. Additionally, the tumor tissue itself can combine homogeneous and heterogeneous textures and also has patient specific characteristics.

It is comprehensible that a deep learning algorithm must first have seen a larger amount of training data in order to learn these variations to generate the desired classification output. Classical machine learning can benefit from expert knowledge about pathologies and the features can be built accordingly. Therefore, by benefiting from a priori knowledge in designing handcrafted features in classical machine learning, potentially smaller amounts of training data are necessary to attain satisfactory classification results.

The handcrafted features from classic machine learning approaches are also much easier to interpret. Tracing back learned features within the neurons of a deep learning network can be very challenging due to the complexity of the network itself and also due to the lack of interpretability of weights and learned convolutions in deep series of connected layers.

Nevertheless, when there is access to larger amounts of (medical imaging) data, deep learning is the method of choice today. Also transfer learning is an advantage of deep learning-based approaches, when a previously trained network could be used as basis for training a CNN for a new but related classification task. An example would be a deep learning-based liver segmentation previously trained on a different MRI sequence or different contrast agent. In this case, training times for learning these new adaptations can be potentially reduced and less training data for the related task might be necessary to obtain satisfying segmentation results.

The advantages and disadvantages summarized here have significantly influenced the choice of methods in this dissertation.

1.5 Object-Based Image Analysis (OBIA)

In Part I of this thesis a special image segmentation technique is used that is briefly described in this section.

Object-based image analysis (OBIA) is a concept of segmenting image objects instead of using pixels or voxels. The OBIA implementation used in this dissertation is based on the work by Schwier [133].

A prerequisite for OBIA is an appropriate partitioning of the image into objects, which is referred to as *over-segmentation* in the following. In this thesis the term *region* is used for a group of neighboring *objects*.

A typical criterion for forming adequate image objects for segmentation is heterogeneity. Different techniques exist to create over-segmentations. In this work an adaption of a watershed transformation [154] and the SLIC Superpixel method [1] are applied to

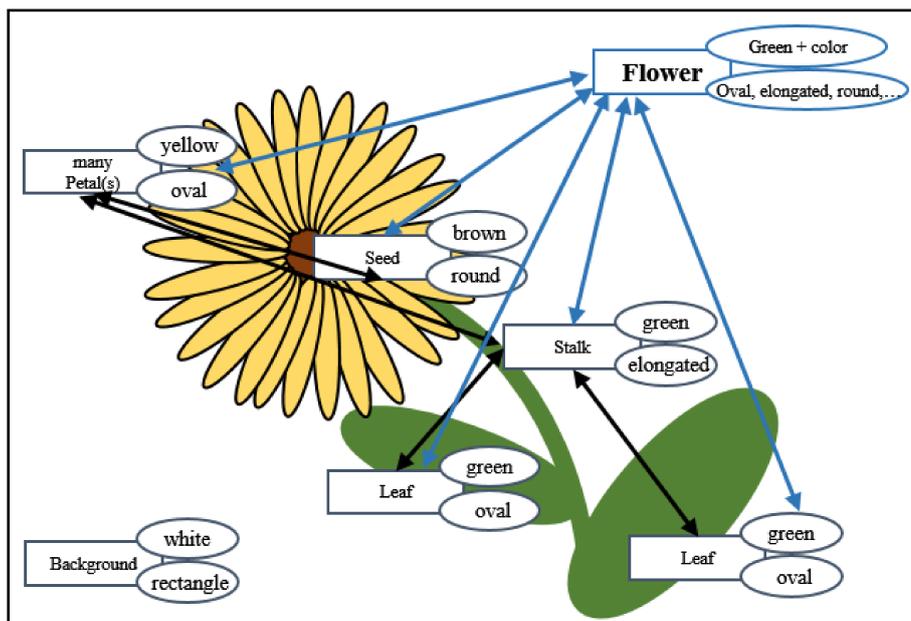


Figure 1.8: This figure shows the principle of an OBIA representation of objects (leaf, stalk, seed, petal) together with their attributes/features. Overlaid on the flower is the attributed relational graph as an example to classify objects using OBIA. For simplicity's sake and for the sake of a general overview, the labeling of each individual petal has been omitted in this graph. The individual objects form together the target structure *flower* that was aimed to be detected in the image.

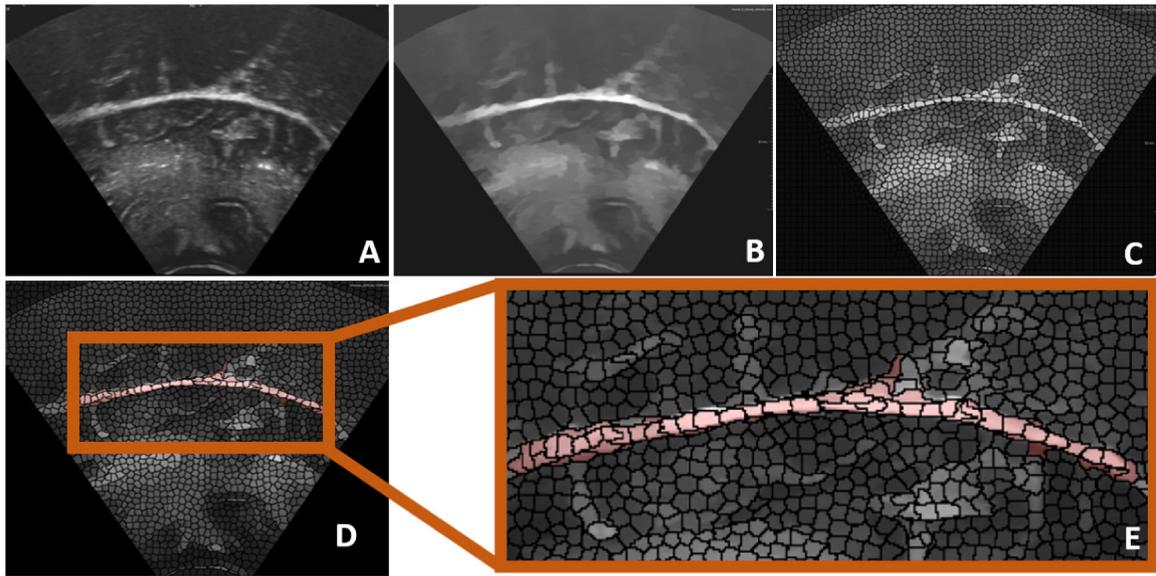


Figure 1.9: This figure gives a more explicit example of how the OBIA approach is used in this thesis. In **A** the original ultrasound image is illustrated. **B** shows the preprocessed (smoothed) iUS image, which is done in order to increase texture heterogeneity for the subsequent applied over-segmentation in **C**. In **D** and **E** objects are highlighted in red that represent the segmented target objects of the falx cerebri and tentorium cerebelli (see Section 1.3.2 for details on anatomy).

generate initial over-segmentations of potential target objects (see Sections 2.5 and 2.7). For an optimal segmentation result it is essential that the initial over-segmentation adequately captures the image objects that shall be segmented in order to generate accurate segmentations.

There is an explicit and an implicit approach of developing OBIA algorithms [133]. The explicit approach implies the formulation of (complex) rules that are needed to segment the target objects. Countless descriptive OBIA features (i.e. intensity, shape, and neighboring features) can be employed to describe relations of objects that can be further merged into regions. Fig. 1.8 shows a simple representation of how objects can be detected in an image by their individual characteristics (here: color, shape, neighboring relations) by using OBIA. In a subsequent step, these objects form the characteristics of a flower by relying on the neighboring features. This represents also a simple hierarchical segmentation approach by, first, identifying possible target objects and, subsequently, using the contextual information of merging these objects to form the target structure *flower*.

The example also illustrates the potential of the explicit OBIA approach to quickly

become very complex to be handled manually. Therefore, the implicit OBIA approach uses machine learning techniques to automatically classify objects and regions. For instance, in this dissertation in Section 2.7 object features are extracted using OBIA and a random forest classifier is employed that learns the specific features of the target objects from training samples. On the contrary, in Sections 2.4 and 2.5 the explicit OBIA approach is used. Fig. 1.9 illustrates an example closer to the actual application of the OBIA approach in this thesis.

1.6 Imaging Informatics

For both clinically oriented research projects in Part I and in Part II of this thesis the fundamental step was to build a patient data base as a prerequisite for all further analyses. This data base had to contain the specific patient cohort, related imaging data, clinical data (data of ongoing patient care from the electronic medical record (EMR), for instance, reports from previous visits, radiology reports, and laboratory parameters), and annotations either extracted from the EMR or especially created for further image analyses (e.g. comments about tumor type, tumor location, disease severity or segmentations).

The selection of the specific patient cohorts implies filtering patient data. In case of the project in Part I, it was a prospective research study and, therefore, patients that fulfil the study criteria could be actively included in the study as soon as they were diagnosed, decided for a surgical treatment in our partner's hospital, and agreed to sign the declaration of consent to participate in the study.

In case of the research project in Part II the specific patient cohort was selected retrospectively by searching the hospital's patient data base. An overview of the whole data acquisition process is given in Fig. 1.10. With the Research Patient Data Registry (RPDR), as centralized clinical data registry of all Partners Healthcare hospitals, it is possible for Partners researchers to gather clinical information from various hospital systems [86]. In preparation for my research stay at the Brigham and Women's Hospital, I had to undergo the proper training and certification in order to receive access to the data from the hospital systems.

An online query tool allows to explore available clinical data and can be used as self-service system. The query tool allows concatenation of several query criteria to identify the necessary patient cohorts. As a result, the query can be used to obtain patient identifiers and complete EMRs of the selected patients through the RPDR with IRB (institutional review board) approval. The IRB is consistent with the duties and functions of an ethical review committee.

For the research in Part II it was crucial to identify the patient cohort with ICD10 codes related to cirrhosis of the liver or to benign pancreatic cysts as control cohort (for more details in patient data selection see Chapter 3.2). Furthermore, these search criteria could be supplemented by the criterion of having had a MRI of the abdomen. As soon as the patient cohorts are identified, the RPDR online tool allows to display the number of selected patients that fulfil the criteria and the cohort's demographics. This is why this tool is especially useful to access the initial study feasibility. The results of the query can be

saved and the related patient data can be requested through the RPDR system. However, the access to radiology images of these patients needs to be requested separately. For this, the Medical Imaging Informatics Bench to Bedside (mi2b2) workbench had to be used as a secure download portal that serves as bridge between researchers and the Partners Healthcare PACS systems. The mi2b2 workbench allows to access stored medical images from several servers of the PACS system. Together with the already requested patient data from the query, the related radiology images can be requested for download.

The medical images are downloaded in DICOM (digital imaging and communication

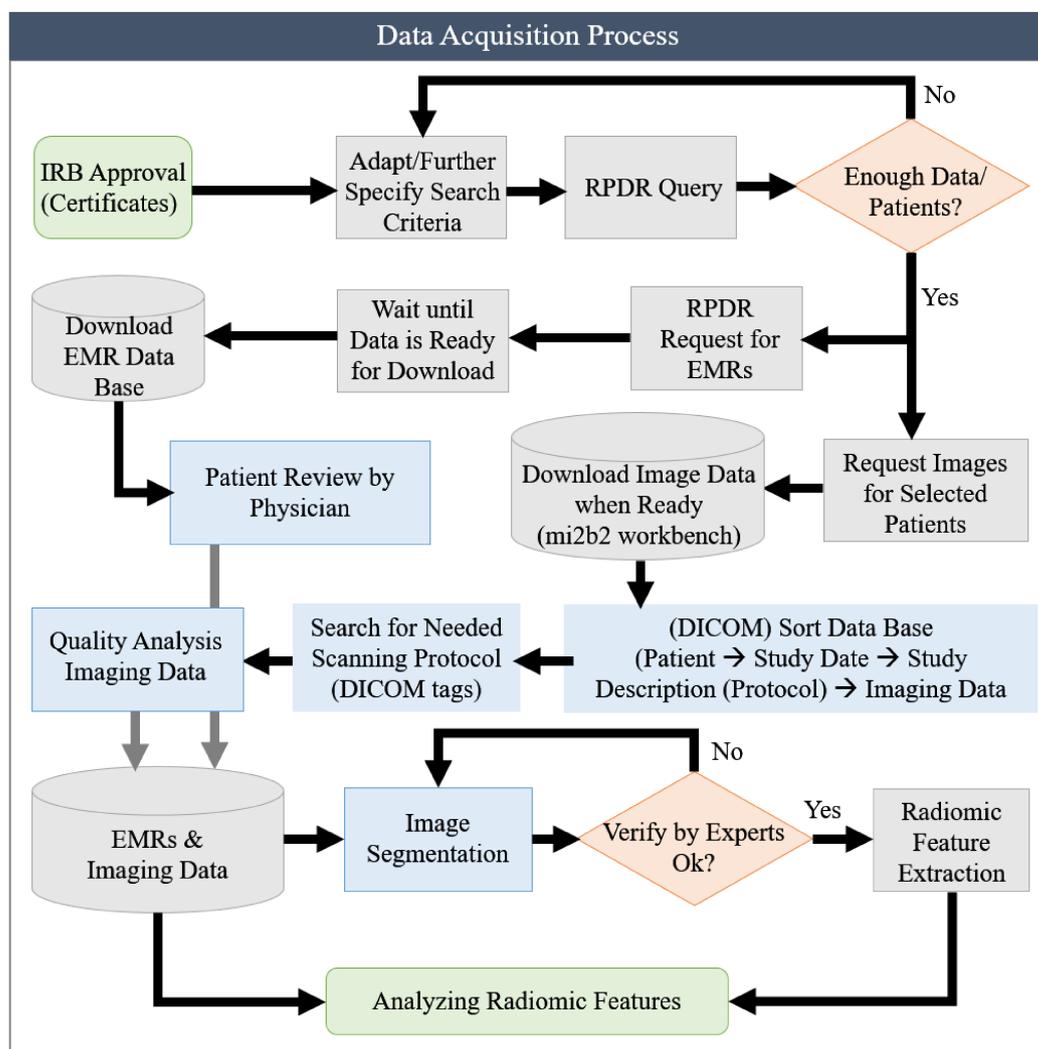


Figure 1.10: RPDR data acquisition process visualized in a flowchart. Start and end of the process are colored in green. Preprocessing steps are visualized as blue rectangles. Decisions are marked in orange.

standard in medicine) format from the workbench, which is a convenient standard for medical image computing [87]. Additionally, further processing of the radiology data was mandatory for further analyses in Part II. The downloaded imaging data seemed arbitrarily scattered and needed to be sorted and saved in the following folder structure for our data base for which the information from the DICOM tags could be used: Patient, study date, series description, specific radiology images. Afterwards, from this data base MR images are extracted for further analyses that follow a specific acquisition protocol in order to guarantee that the analyzed scans of the cirrhotic and the control cohort follow the exact same imaging protocol. Without a Partners hospital-wide standardization of scanning protocols and harmonization of their names this task would have been far more time-consuming.

The whole process of the above described data acquisition took nine months and also included patient data anonymization, patient chart review by a physician to confirm each patient's suitability for the studies of Part II, and attaining needed segmentations of the MRI data. During that time, a computer was always busy with almost no downtime with downloading, preprocessing, and automatically segmenting MR images.

This represents three quarters of the total project time of the research year in Boston and, thus, meets the general expectations of the amount of time needed for this process by data scientists. According to a survey conducted by CrowdFlower (since 2018 rebranded as Figure Eight) and commented in an article by Forbes data scientists spend 79 % of their time on collecting, cleaning, labelling, and organizing data [47, 32]. Analyzing the data with mining data for patterns, refining algorithms, and building training sets represent about 16 % of their time [47, 32], which is also in line with the time needed for these tasks in Part II of this thesis.

It was a great experience to have had access to these modern data base systems of the future for extracting patient data at the Brigham and Women's Hospital in Boston, USA. Without question, the studies in Part II of my thesis would not have been possible during the period of the research stay without a large part of the data acquisition being automated with them. This kind of systems for data mining EMRs and imaging data are powerful tools and will shape the future in collecting and analyzing large amounts of medical data. Especially, in the research field of clinical decision support systems and radiomic biomarker detection hospitals equipped with this technology will play a pioneering role.

2 PART I: Multi-Modal Image Segmentation of Cerebral Structures for Improving MRI-US Registration for Image-Guided Glioma Surgery

As mentioned in the first chapter of this dissertation this thesis focused on two main topics. Part I is described in this chapter and focused on multi-modal image segmentation techniques of cerebral structures in both intraoperative B-mode ultrasound as well as in gadolinium-enhanced T1-weighted MR images. The objective was to provide automatic and efficient MRI-US segmentations for improving intraoperative MRI-US image registration for image-guided glioma (brain tumor) resections. All patients and, therefore, all corresponding imaging data were acquired within this prospective study together with our clinical partners at the neurosurgical department at Essen University Hospital, Germany. With an increasing number of included patients, image data, and expert references (ground truths data for the evaluation of results) different methods for segmentation and the registration were developed, adapted, refined, and successively tested and trained on an increasing patient data base.

Brainlab, a company headquartered in Munich, Germany, provided us and our clinical partners with technical support in the challenging data acquisition process.

In the following an introduction to the topics of Part I is given in Section 2.2 together with a detailed motivation for improving image-guided surgery. This section complements the general motivation given in Chapter 1. Furthermore, this section describes the registration algorithms for MRI and ultrasound image fusion in the research area of image-guided glioma surgery and gives a general introduction of segmenting structures in ultrasound imaging.

The material and methods section provides an overview of the technical components used in the operating room and the thereby acquired image data (Section 2.3). The image

data is the basis for the implemented image processing approaches that are outlined in the following Sections 2.4-2.8. Each approach is prefaced with a short introduction, followed by a detailed description of the implementation, attained results, and concludes with a discussion.

The subsequent Section 2.1 gives an overview of the developed approaches for Part I of this thesis. A paper-like structure is followed for describing the methods in each section but the introduction and the general material and methods are combined for all sections in the beginning of this chapter in order to avoid redundancy. Part I closes with a conclusion and an outlook.

Contributions:

The overall objective was to implement a robust multi-modal image registration of preoperative MRI and intraoperative ultrasound imaging to improve intraoperative image guidance during low- and high-grade glioma resections. This goal was achieved by implementing a segmentation-based registration algorithm that performed significantly better than a pure multi-modal intensity-based registration. The segmentation-based registration even proved to be more robust and faster than the intensity-based registration. As a guiding frame for the registration, central cerebral structures were segmented in B-mode ultrasound imaging. We were the first research group to implement a segmentation-based US-MRI registration method and compared the results and benefits to a pure intensity-based registration for supporting intraoperative image guidance during glioma surgery.

2.1 Overview of Approaches Implemented for Part I of this Thesis for Improving MRI-US Registration

In Part I of this thesis five approaches are described that are implemented in order to improve an intraoperative registration of preoperative MRI scans with intraoperatively acquired ultrasound volumes for image-guided glioma surgery. The goal is to provide the surgical team in the operating room with an (almost) in real-time updated MRI-US fusion image on a monitor about tumor location, extent of tumor resection, and visually perceivable safety margin to risk structures.

Challenging here is to compensate the initial brain shift after craniotomy and dura opening, the deformations of brain tissue caused by surgically accessing the tumor's location within the brain, and the tissue shift and collapse by gradually resecting the tumor volume as mentioned in the introduction in Chapter 1 [49, 98].

Applying intraoperative image deformations derived from ultrasound imaging in almost real-time to the preoperative MRI would be an immense benefit for a save and complete tumor resection. Until today, this multi-modal registration problem is not solved and different approaches from other research groups exist that aim to target the specific image registration problems. In Section 2.2.1 these approaches are summarized and reflected in conjunction with our implemented solution for MRI-US image registration and fusion.

This is why the research of Part I in this thesis focused on segmenting central cerebral structures in ultrasound imaging that could be used as 3D reference structures for improving a registration with preoperative MR images. Consequently, the impact of other boundaries or artifacts visible in iUS are reduced and cannot confuse and hinder a correct registration of corresponding anatomical structures any more. The goal was to transform the MR image in the respective (partial) image region and visualize the registration result in a MRI-US fusion image as illustrated in Fig. 1.2 and in Fig. 2.16.

The five different approaches described in Part I of this thesis reflect the gradual development of the desired registration (see overview of attained results of these approaches in Tab. 2.1). As the focus was to use central cerebral structures as guiding frame, the perifalcine region together with the tentorium cerebelli were identified as reference structure for multi-modal image registration. As literally being the central structures of the brain the chance of their visibility in iUS during glioma surgery even with very different possible tumor locations is quite high (see Section 1.3 for the anatomical background). All preMRI and iUS images were acquired prospectively during our collaborative project together with the neurosurgical department of the University Hospital in Essen. In the beginning

the focus was to identify optimal ultrasound parameters and some initial experiments and exercises were necessary to optimize the acquisition of freehand B-mode sweeps in order to allow an optimal reconstruction of 3D ultrasound volumes for image registration. For instance, the speed of moving the ultrasound probe had an impact on the quality of the scanned iUS volume. Moreover, the pressure of the ultrasound probe on the brain tissue had to be adjusted and balanced between optimal contact to the brain tissue and ultrasound probe induced tissue deformation. After definition of a detailed ultrasound acquisition protocol, more data sets were acquired in the same manner.

Starting in approach 1 with an initial data set of pre-resection ultrasound images from four different patients, the focus was to evaluate an optimal preprocessing for the artifact-rich "speckled" ultrasound images (see Section 1.2.2) and to enhance the line-type structures of falx cerebri with adjacent sulci and gyri (anatomical structures described in detail in Section 1.3.2) prior segmenting these structures.

In approach 2 the segmentation method is further refined to an object-based image analyzing approach. A watershed transform is used as basis for further object-based analysis of the perifalcine target region. Target objects are classified with handcrafted features. The ultrasound data set for evaluation of the segmentation method could be extended to overall seven different patients.

Approach 3 describes a segmentation method for corresponding segmentations of the perifalcine region in contrast-enhanced preoperative MRI. The falx cerebri and the tentorium cerebelli represent sheet-like structures in 3D and are enhanced together with the vasculature by the contrast agent. The basic concept to segment these structures is to use the eigenvalues of the Hessian matrix to get a sheet probability for each voxel. A custom-built preprocessing of the MR image is implemented to reduce computation time and to enhance the target structures.

Approach 4 is the central publication of Part I that combines ultrasound and MRI segmentations and demonstrates that incorporating segmentations in the registration process could significantly improve the overall registration accuracy, robustness, and speed of preMRI and iUS image fusion. Eleven patient data sets could be used for this study with contrast-enhanced preMRI and corresponding iUS scans. An expert created reference segmentations and corresponding landmarks in preMRI and iUS imaging in order to evaluate the segmentations and the attained multi-modal registrations. For approach 4 the object-based image analyzing approach was further improved to a very sophisticated, hierarchical object- and region-based analyzing approach that uses SLIC Superpixel for an initial over-segmentation of iUS images instead of the watershed transform in approach 2

[1]. The SLIC Superpixel method allowed a more precise parcelling of the target structures for segmentation. A random forest classifier is trained on the extracted image features from the target objects and regions for an automatic classification of the objects representing the perifalcine region in iUS images. Besides, the inter-observer variability is discussed for segmenting the structures of the perifalcine region in iUS imaging which once again illustrates the complexity of the segmentation task by comparing human experts.

Approach 5 fundamentally uses the same eleven data sets as in approach 4 and allows direct comparison of attained segmentation results. Additionally, seven further iUS patient data sets could be acquired to train a U-net-like neuronal network to classify the falx, tentorium, and adjacent sulci and gyri. Furthermore, the trained U-net is applied to a publicly available ultrasound data set that was also acquired in the context of glioma surgery. Without any further training on the externally acquired ultrasound images very good segmentation results could be attained.

Approach	Section	Topic	Dice	Hausdorff	Jaccard
1	2.4	Automatic Segmentation of the Cerebral Falx and Adjacent Gyri in Pre-resection iUS	0.79	1.56 mm	0.64
2	2.5	Automatic Segmentation of Perifalcine Structures Using Object-Based Image Analysis	0.68	2.58 mm	0.53
3	2.6	Corresponding MRI Segmentations of the Perifalcine Region	Quality visually assessed by an expert as very good		
4	2.7	Automatic MRI-US Segmentations for Improving Registration in Image-Guided Neurosurgery	0.74	12.2 mm mTRE: 2.2 mm	
5	2.8	Neural-Network-Based Segmentation of Perifalcine Structures in multiple stages during glioma surgery	0.88	5.21 mm	

Table 2.1: Overview of approaches implemented for Part I of this thesis. The following evaluation metrics are abbreviated in the above table: Dice Coefficient as *Dice*; Hausdorff distance as *Hausdorff*; Jaccard index as *Jaccard*, mean Target Registration Error as *mTRE*. All evaluation methods used in this thesis are explained in Section 2.3.2. Different segmentation accuracies must also be seen in conjunction with what kind of perifalcine structures are segmented and in which phase during glioma surgery (see *Topic* column). The more stages during resection are included, the more difficult is the segmentation task and the accuracy might be decreased. The resection creates additional boundaries between which the algorithm also has to distinguish.

2.2 Introduction and Motivation

In glioma surgery residual tumor has a great impact on the overall survival rate but also on progression-free survival [21, 81, 104]. For the optimal outcome after surgery a balance has to be established between maximization of surgical cytoreduction and minimization of new permanent neurological deficits to preserve and prolong the patient's resulting quality of life [108, 116, 165].

Preoperative image data are acquired a few days before surgery. However, the difference between these preoperative scans and the current intraoperative situation increases with the extent of tissue removal caused by brain-shift after craniotomy and tissue deformation [107, 109, 115].

The preoperative MRI (preMRI) data set, usually shown on a separate screen in the operating room for the medical staff, is now merely a rough estimate. For this reason, intraoperative imaging is crucial to maximize the gross-total resection but with greatest accuracy, precision, and care of eloquent brain areas.

Therefore, intraoperative imaging, such as MRI or CT, have great importance during surgery but are not affordable or available in every neurosurgical department. On the contrary, intraoperative ultrasound (iUS) offers real-time information, it is widely available at low costs, and does not cause any radiation [107]. Moreover, tumor boundaries are even better distinguishable within iUS, which is an enormous benefit compared to other intraoperative imaging for safe gross-total resection [82, 135, 148].

However, several factors during resection degrade the quality of iUS. For example, additional structures or tissue boundaries are visible in iUS imaging but not in preMRI. These are induced during surgery through creating and extending the resection borders, by filling the resection cavities with saline water, by small air bubbles, and by blood clotting agents. Also, cerebral edemas are extremely common in surrounding brain tumors. Moreover, this increases reflection that diminishes the general ultrasound penetration depth. All the above-mentioned gravely impedes a solely intensity-based iUS and preMRI registration.

For this reason, several anatomical structures of the brain were segmented as part of this thesis that can act as guiding frame, i.e. as reference structures, in order to overcome the above mentioned multi-modal registration difficulties. These approaches are described in detail in the following Sections 2.4-2.8.

2.2.1 Registration for US-MRI fusion

Registration algorithms in this field of research of ultrasound and MRI fusion can be categorized in non-deformable as described in [22, 107, 110, 128] or deformable approaches as described in [29, 63, 115, 118, 119, 120, 121, 122] that further split into feature-based and intensity-based methods.

Feature-based methods find corresponding points or structures in both modalities and use correspondences to conclude the registration transformation [79].

Intensity-based methods rely on similarity measures and methods commonly used are: sum of squared differences (SSD), normalized mutual information (NMI), normalized cross-correlation (NCC), and normalized gradientfield (NGF) measures [7, 79, 80, 89].

Proposed solutions in the context of this specific registration problem are landmark-based methods that register points [37], lines or tubes, e.g. of vasculature [12, 106, 114] or surfaces [57], and methods that attempt to transform and simulate the intensities within preMRI to iUS in order to make both imaging modalities more comparable for a subsequent registration [62, 67, 156].

Rackerseder et al. [113] proposed coarse segmentations from both MRI and US imaging as an alternative to 3D landmark registration. The research focused on the registration aspect and regards the segmentations as given, i.e., either from propagated coarse annotations or from other already existing automatic methods as, for instance, implemented in *FreeSurfer* [117]. The segmentations are used to transfer to a (pseudo)-mono-modal registration problem comparable to the aforementioned methods. The coarse segmentations support a coarse registration for initializing the registration globally.

This refers to the general finding of the rough image region within the preMRI brain scan that corresponds to the only partially scanned brain volume acquired with each iUS sweep. This can also be seen in Fig. 2.16, where the iUS volumes are overlaid in red on the preMRI scans. In contrast to our research in Section 2.7, this step was not necessary because the neuronavigation system supported a rough initial registration of preMRI with intraoperatively acquired ultrasound imaging through an optical camera system that allowed tracking of the ultrasound probe in the operating room in relation to the patient's head (see Fig. 2.1 that illustrates the intraoperative setup) [78].

More exact, finer segmentations are used by Arbel et al. [3] and Mercier et al. [77] who registered preMRI to a probabilistic atlas. The atlas was used to segment the preMRI and they assigned different intensity transformations to different regions within the MRI to create a pseudo-US image.

The proposed combined segmentation and registration approach for US-MRI fusion

in Section 2.7 would fit into this registration category where segmentations are used to transfer additional information to the overall registration process.

2.2.2 Segmentation of US images

Most 2D ultrasound segmentation methods can be expanded to 3D segmentation approaches. Mozaffari and Lee [84] describe them as powerful as 3D methods. For slice-wise US acquisition 2D image segmentation relies on contrast and image resolution, whereas 3D approaches add the supplementary issue of reconstruction quality.

Even 3D ultrasound images that are acquired with a rotating ultrasound transducer face the problem of reconstruction errors by indirect smoothing of the image content when multiple images of the same region exist and an average voxel intensity is calculated for each voxel in the resulting ultrasound image.

This is why we chose 2D image segmentation within this PhD thesis in order to segment the slice-wise acquired iUS images in their highest, original resolution. This was especially beneficial for segmenting the thin and line-type structures of the perifalcine region as described in the following Sections 2.4-2.8.

Mozaffari and Lee in [84] distinguish none of the reviewed ultrasound segmentation strategies as gold standard or even trendsetting for every purpose in image processing. It is even more emphasized that this is an active field of research with numerous methods and segmentation strategies that are tailored for each scope of application.

2.3 General Material and Methods of Part I

This section gives an overview of the general material and methods used for the research project of Part I of this thesis. In the following, the intraoperative setup and image acquisition are described and illustrated. Additionally, the evaluation methods for segmentation and registration results that are applied in this thesis are outlined at the end of this section.

2.3.1 Data Sets for preMRI and iUS Image Fusion and Segmentation

In the context of this thesis we acquired corresponding preMRI and iUS images from overall 11 patients with low-grade and high-grade gliomas who underwent tumor resection in the neurosurgical department at Essen University Hospital, Germany. Each patient had at least one T1-weighted gadolinium-enhanced preMRI for surgical planning acquired

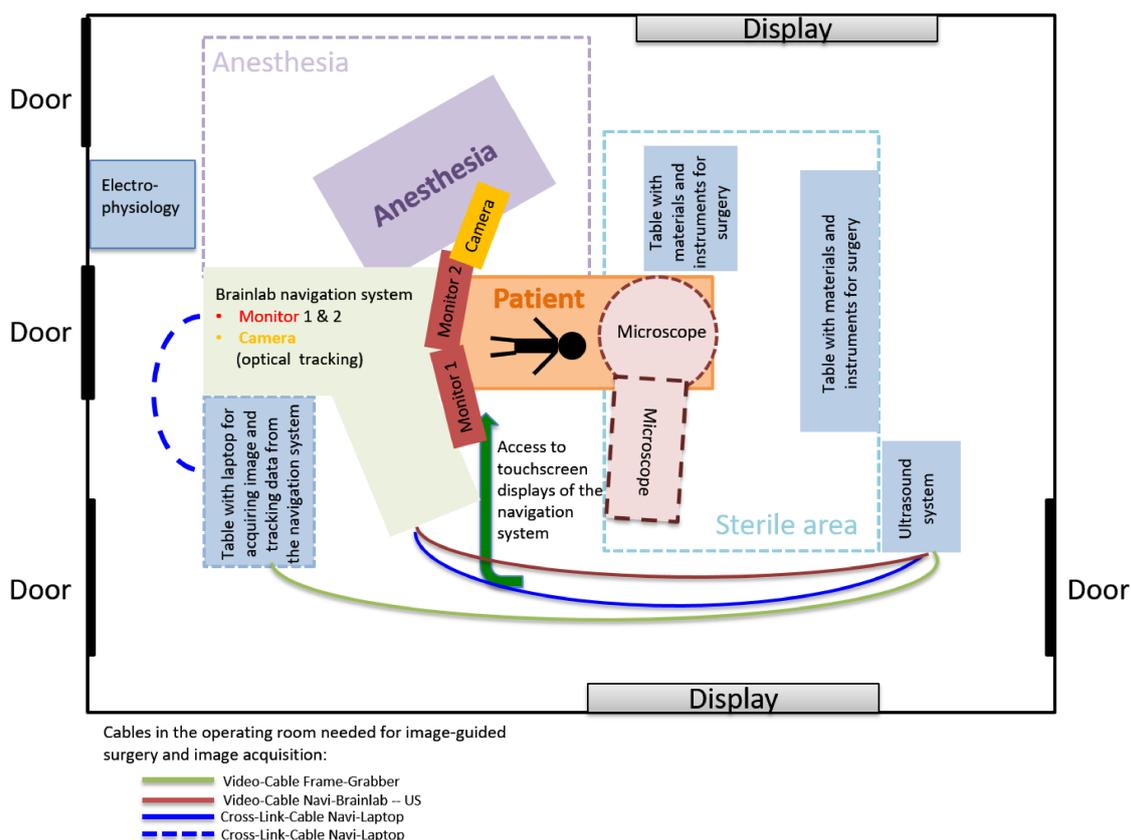


Figure 2.1: Intraoperative setup during image-guided glioma surgery. The illustration of the operating room shows an exemplary spatial arrangement of the devices and materials for an image-guided neurosurgical procedure.

a few days before tumor surgery. Apart from the aforementioned patient data sets that could be used for multi-modal image fusion, iUS scans from 18 different patients were later available in the project to further evaluate the segmentation algorithms.

Navigated and tracked freehand B-mode sweeps were made in a parallel, non-fan-like manner, before dura opening (right after craniotomy), during tumor surgery, and after tumor resection if possible.

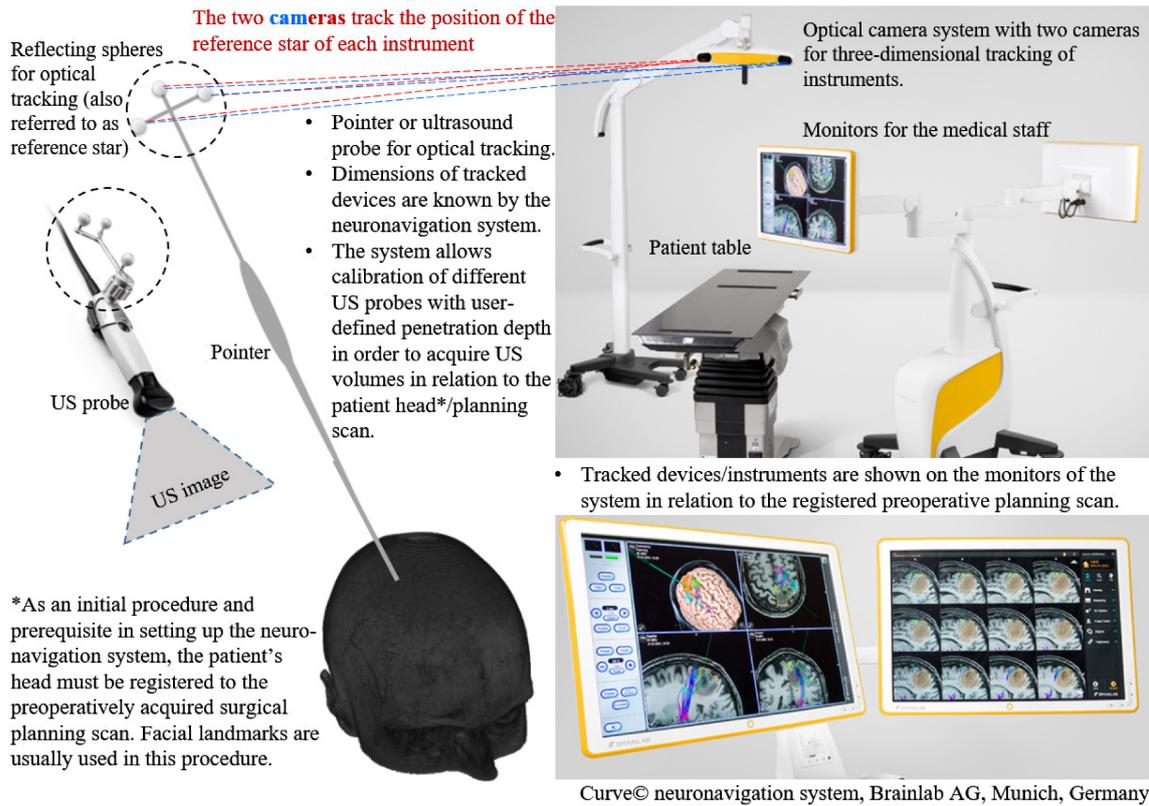


Figure 2.2: Image-guided surgery with a neuronavigation system. The figure illustrates the intraoperative setup and tracking of instruments in relation to the patient's head within the surgical planning data set. In our intraoperative setup the surgical planning data set was a contrast-enhanced MRI. In an initial step of setting up the neuronavigation system and the intraoperative tracking of instruments/devices, the patient head is registered to the surgical planning scan. For instance, facial landmarks of the eye or ear regions of the patient can be used for an initial landmark-based registration. A reference star (consisting of three or more spheres) is mounted with an adapter to the patient table close to the patient's head and functions as origin of the camera coordinate system.

The iUS scans were acquired using an US system (Alpha 10, Hitachi Aloka Medical, Japan) connected to a neuronavigation system (Curve, Brainlab AG, Munich, Germany) via a video cable. We used a micro-convex multifrequency probe with a mean frequency of 7.5 MHz and a fixed penetration depth for all patients of 8 cm.

For navigated freehand US acquisition an adapter was mounted onto the US probe for optical tracking. The navigation system saved 100-200 2D B-mode US images (slices) per sweep together with their tracking information.

The slices were reconstructed to a 3D volume for volumetric preMRI and iUS image registration (see Section 2.7) and the original 2D slices were used for iUS image segmentations in the subsequent chapters. Fig. 2.1 illustrates the intraoperative setup in the operating room with the neuronavigation system. Fig. 2.2 shows the used neuronavigation system and outlines the optical tracking of instruments, such as an US probe.

2.3.2 Evaluation Methods for Segmentation and Registration Results

In this section we briefly describe the methods for quality measurement of the segmentation and registration results that are applied in the following chapters.

Evaluation of Segmentation Results

The employed similarity criteria are here briefly described that are used to measure the quality of our results. The Dice coefficient, the Jaccard index, and the Hausdorff distance are established methods for the evaluation of 2D and 3D segmentation results.

The three methods provide numerical values that allow comparisons of different segmentation results as well as an immediate assessment of the achieved quality of the segmentation at hand in regard to a reference segmentation.

The following formula determines the calculation of the Dice coefficient, where P and Q represent the sets of pixels or voxels of two compared segmentation masks:

$$C_{Dice} = \frac{2|P \cap Q|}{|P| + |Q|} \quad (2.1)$$

According to the formula the Dice coefficient represents the size of the union of two sets. The coefficient is a value between 0 and 1, where 0 indicates no overlap of the sets of two compared segmentation masks and a value of 1 indicates an exact match between two sets.

The Jaccard index is a quite similar measure and is often used instead or in conjunction with the Dice coefficient and measures the similarity between sets and is defined as the

size of the intersection divided by the size of the union of two sets:

$$I_{Jaccard} = \frac{|P \cap Q|}{|P \cup Q|} \quad (2.2)$$

The Jaccard index also represents values between 0 and 1. In accordance to the Dice coefficient a Jaccard index of 1 would mean an exact union of P and Q and a 100 % overlap of two segmentation results.

The Hausdorff distance is the third quality measure used in this thesis to assess the accuracy of attained segmentation results and quantifies the maximum distance between two 2D or 3D segmentations and allows a comparison of contours or surfaces of the segmented object. Subsequently, the equation defines the Hausdorff distance, where S and T are the two different sets of surface pixels and voxels of the compared segmentations:

$$H(S, T) = \max\{d_{max}(S, T), d_{max}(T, S)\} \quad (2.3)$$

The maximum distance d_{max} is defined as:

$$d_{max} = \max_{s \in S} \min_{t \in T} ||s - t||. \quad (2.4)$$

According to the equation above the maximum calculated distances not symmetric because the following applies:

$$d_{max}(S, T) \neq d_{max}(T, S). \quad (2.5)$$

The Hausdorff distance is a symmetric measurement of both distances by considering the maximum of both distances. The distances are measured in mm and should be as small as possible by measuring the distance between segmentation result and the corresponding reference masks of an expert or other gold standard. [26, 43, 127, 141]

Evaluation of iUS and preMRI registration results

In order to quantitatively measure the performance of the registration approach described in Section 2.7 in comparison with the initial registration provided by the neuronavigation system and a solely intensity-based registration, the commonly applied mean target registration error (mTRE) is used as metric.

The mTRE measures the average distance between landmarks x (here: landmarks within the iUS imaging that represent the ground truth landmarks) and x' (here: landmarks

after multi-modal image registration within the deformed preMRI). The mTRE can be computed as follows for a number of landmarks n [43]:

$$mTRE = \frac{1}{n} \sum_{i=1}^n \|x - x'\|. \quad (2.6)$$

2.4 Approach 1: Automatic Segmentation of the Cerebral Falx and Adjacent Gyri in Pre-Resection 2D Ultrasound Images

Published as: J. Nitsch , J. Klein, D. Miller, U. Sure, H. Hahn (2015) *Automatic Segmentation of the Cerebral Falx and Adjacent Gyri in 2D Ultrasound Images*. In: H. Handels, T. Deserno, H. P. Meinzer, T. Tolxdorff (eds) *Bildverarbeitung für die Medizin 2015*. Informatik aktuell. Springer Vieweg, Berlin, Heidelberg. [93]

Contributions: Optimization of parameters for a diffusion stick filter as general, extremely useful preprocessing for iUS images in order to enhance line-type structures by simultaneously reducing speckles as characteristic artifacts in iUS images of the human brain. The Hessian filter is applied for further feature analysis of potential target structures. Both of these proposed filtering methods proved to be reliable in this context and represent the basic preprocessing methods used in subsequently developed US segmentation approaches.

Abstract. An automatic segmentation of the cerebral falx and adjacent gyri (perifalcine region) is presented for B-mode 2D ultrasound images. The movement of the brain tissue during neurosurgery reduces the accuracy of neuronavigation system which provide image guidance based on preMRI. Thus, the segmentation of the falx and its adjoining gyri in navigated iUS may be used to improve navigation within the preMRI scans by providing additional spatially updated image information of the patient's brain. The segmentation method was tested on 50 2D US images and achieved on average a Dice coefficient of 0.79, a Hausdorff distance of 1.56 mm, and a Jaccard index of 0.64.

Available data sets at the time of this study: Selection of 50 iUS images from four different patients acquired after craniotomy but before opening the dura mater of the brain. Per patient two up to three ultrasound volumes were acquired before dura opening.

Due to the generally small area of craniotomy and limited space for the US probe for acquiring ultrasound images, the target structures did not vary much when the volume was scanned from one angle within one sweep. Therefore, the US scanning angles of the target structure were slightly varied in each acquired iUS volume per patient.

The 50 iUS images used for evaluation in this study were carefully selected to show different representations from different viewpoints of the target structures to be segmented. This was done in order to have as many unique scans of the target region as possible to get a representative impression of the characteristic features for segmenting the perifalcine

region. The 50 selected iUS images always show the target structures for which an expert manually segmented the falx and adjacent gyri.

2.4.1 Segmenting Perifalcine Structures in Pre-Resection US

The first step within the segmentation method is the application of a diffusion stick filter as initially proposed in [159, 171]. Two diffusion stick filters are applied sequentially (kernel size 15×15). The advantage of the diffusion stick filter is that it offers edge enhancement of voxel or pixel values along lines and edges, while smoothing homogeneous regions. Therefore, the line-type structures of the falx and gyri with strong edges are enhanced while speckles within the 2D US images are reduced (see Fig. 2.3).

The kernel size was determined by an automatic test and was varied from 5 to 31 in 25 US images. The kernel size of 15×15 offers a good compromise between smoothing and edge enhancement. The kernel achieved one of the highest average Dice coefficients and one of the lowest average Hausdorff distances within the test as can be seen in Fig. 2.4. Followed by a histogram equalization to increase the global contrast between the falx and gyri (image content with strong edges with gray values with high intensities) and other structures (gray values with low intensities) within the image.

At this point the proposed method splits into two different parts. One part leads the resulting image, after speckle reduction and edge enhancement, into a Hessian filter. Whereas the other part is focused on defining the ROI, containing the image region of the falx and its adjacent gyri, and later used to mask the result image of the Hessian filter.

The Hessian filter is applied because it can detect the local strength and the direction of edges and lines. This has already proven to be advantageous to segment anatomical structures such as vessels, bronchi, and pulmonary fissures [65, 138] which are line-type or tube-like structures that have a similar image representation and characteristics as the perifalcine region. The Hessian filter is based on the eigendecomposition of the Hessian matrix that contains the second-order partial derivatives of the image intensities. The relation of the eigenvalues are used to describe the local image structure and enable an analysis of the strength and direction of lines and edges. The Hessian filter is applied on the whole preprocessed 2D US image and contains solely a preselection of possible candidates of lines and edges that may be structures of the falx and gyri. Thus, a ROI is determined, representing the image section that contains (ideally) solely the structures of the falx and its gyri. Due to the fact that US images are acquired to scan the perifalcine region, it is assumed that the strongest collection of edges represents the structures of the falx and gyri within the US image.

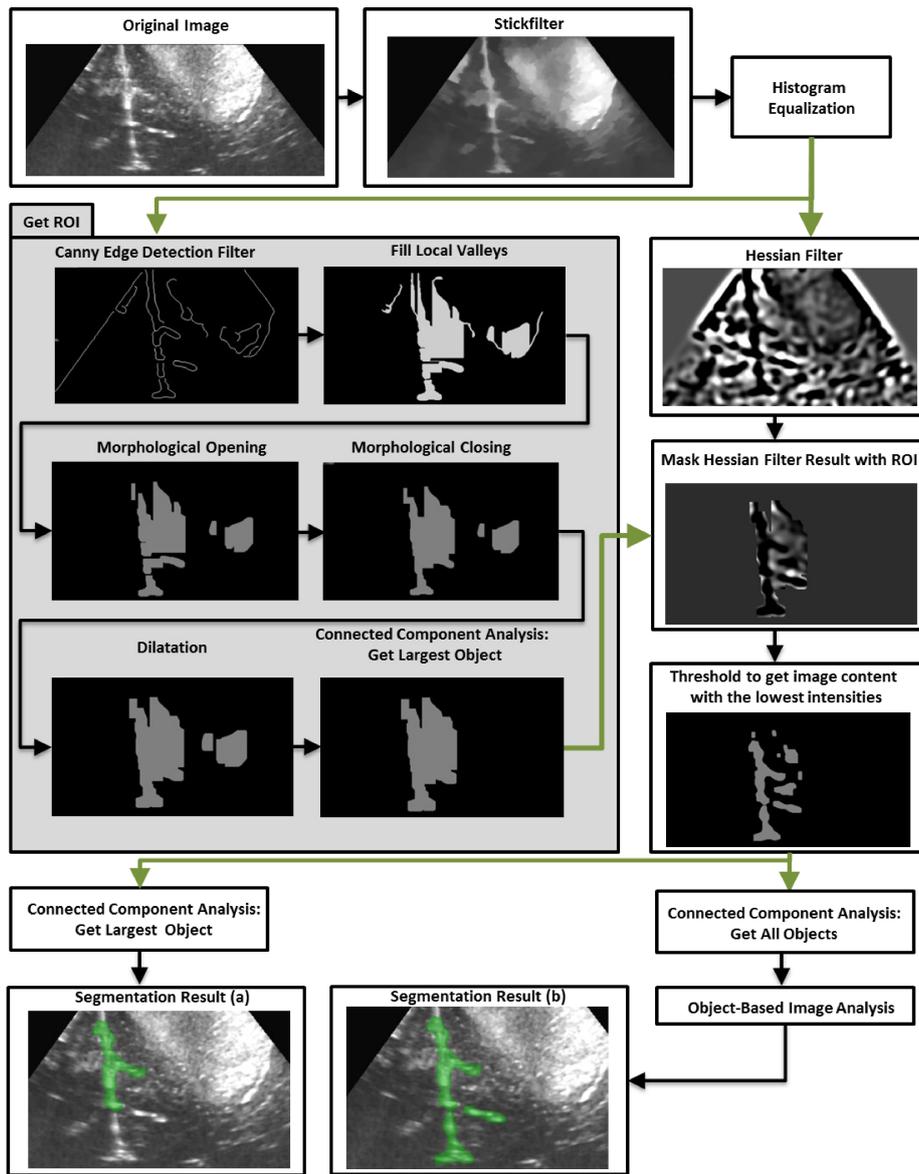


Figure 2.3: Image processing pipeline of the proposed segmentation method. The two segmentation results at the bottom represent the result from the connected component analysis (a) compared to the result of the OBIA approach (b).

In order to identify this image region, a canny edge detection filter is applied with a lower threshold of 0.8 and an upper threshold of 2.8 (Fig. 2.3). Note that the thresholds are determined in the same fashion as the kernel size mentioned above and they performed well on the 50 US images but might need further adaption and testing when applied to a larger amount of different patient data sets with more variance in tumor location and,

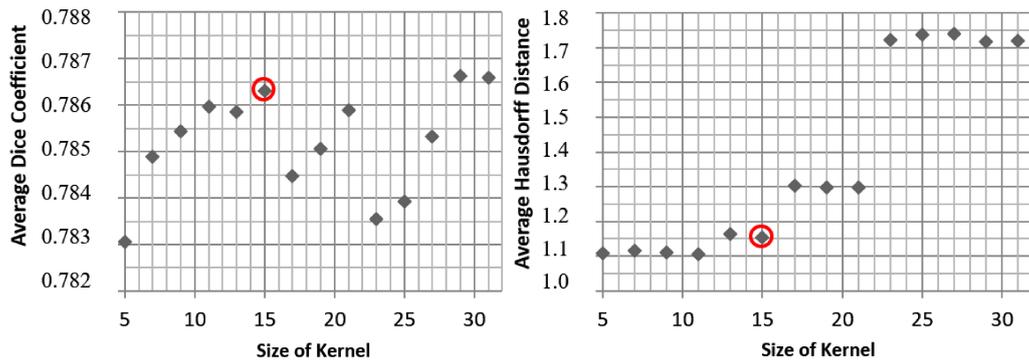


Figure 2.4: Finding the optimal parameters for the diffusion stick filters.

therefore, more diverse scanning angles of the target structures in iUS. Afterwards, local valleys, present within the resulting binary image, are filled (Fig. 2.3). Followed by a morphological opening, a closing and a dilation to cut-off thin and small connections between the filled image contents, to fill small gaps, and to slightly dilate the filtered image content to enable that the resulting binary image covers the image content as well as the boundaries of the objects that should be segmented (Fig. 2.3). Kernel sizes for these filters were evaluated on 20 images.

Finally, a connected component analysis is applied in order to get the largest image content from the binary image which is the region with the strongest collection of edges. This represents the ROI within the 2D US image that contains the structures of the falx and adjoining gyri.

The binary image of the ROI is then used to mask the result of the Hessian filter (Fig. 2.3). At first it was expected that a postprocessing by using solely a connected component analysis, in order to get the largest connected line from the image content, is sufficient. Nevertheless, after processing a larger amount of US images it turned out that the falx and the gyri do not necessarily represent one connected region within the 2D US slices. Thus a further analysis of the image content of the masked result of the Hessian filter is necessary to correctly extract the structures of interest.

To assure this, an object-based image analysing (OBIA) approach is applied to the masked result image in order to correctly extract the perifalcine region by the following image features: positional features (x and y position of presegmented structures), shape features (circularity, elongation and size) and intensity statistics (mean intensity).

A comparison of the segmentation results with and without using OBIA can be seen in Fig. 2.3. The result image of the Hessian filter proved to be especially suitable as basic image from which characteristic features could be derived. The used OBIA approach

Postprocessing	Dice coefficient	Hausdorff distance (mm)	Jaccard Index
With OBP			
Minimum	0.48	0.64	0.03
Maximum	0.89	7.33	0.81
Average \pm SD	0.79 ± 0.07	1.56 ± 1.40	0.64 ± 0.14
Without OBP			
Minimum	0.48	0.64	0.31
Maximum	0.84	15.42	0.73
Average \pm SD	0.68 ± 0.07	3.90 ± 3.41	0.60 ± 0.08

Table 2.2: Comparison of the overall achieved segmentation results: Minimum, maximum, and results on average with standard deviation (SD) for the 50 iUS images with and without using the OBIA approach in a preprocessing step.

has already shown to be a powerful method to extract and classify image features of US images [132].

2.4.2 Results

The quality of the segmentation was tested on 50 2D US images. The segmentation results are compared to reference segmentations of an expert. In 48 cases the falx and gyri could be correctly detected.

On average a Dice coefficient of 0.79, a Hausdorff distance of 1.56 mm, and a Jaccard index of 0.64 could be achieved. It could be shown that the postprocessing with an OBIA approach could considerably improve the segmentation results, concerning the fact that the falx and adjacent gyri do not necessarily represent a single connected structure as mentioned above.

A comparison of the results with and without the object-based postprocessing (OBP)

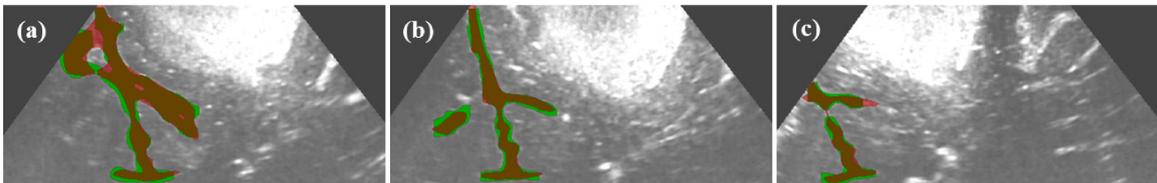


Figure 2.5: Visualization of additional segmentation results (a)-(c). The images show the reference segmentation (red) with an overlay of the segmentation results of the proposed method with OBIA (green).

can be seen in Tab. 2.2. Additional segmentation results for the perifalcine region are visualized in Fig. 2.5.

2.4.3 Discussion

The most challenging part in segmenting the cerebral falx and the adjacent gyri is that these structures do not provide a homogeneous representation within the US images. The falx itself is predominantly easily detected with its strong edges, whereas the adjacent gyri usually provide less contrast to the background and, therefore, less prominent edges. This momentarily minimizes the detection rate of the gyri and is the reason why the proposed method tends to undersegmentation when compared to the reference.

In two cases the contrast of falx and gyri to the background of the US image was too weak and a segmentation of these structures could not be performed when based on prominent edges. OBIA has proven to be a powerful approach to allow a feature-based image segmentation.

This is the reason why OBIA is used in further experiments of this dissertation as soon as more iUS images became available.

2.5 Approach 2: Automatic Segmentation of Perifalcine Structures Using Object-Based Image Analysis

Published as: J. Nitsch, J. Klein, P. Dammann, D. Miller, K. Wrede, U. Sure, H. Hahn (2015) *Automatic Segmentation of Perifalcine Structures in Two-Dimensional Ultrasound Images Using Object-Based Image Analysis*. In: Proceedings of the 14th Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery (CURAC 2015), pages 219-223, Bremen, Germany, September, 2015. [90]

Contributions: Further development of the segmentation approach by extending the segmentation from the previously described approach in Section 2.4 from the falx to the whole perifalcine region. A watershed transform is performed in order to generate an initial over-segmentation of the iUS image. An object-based image analyzing approach is now used as central segmentation approach instead of being used as postprocessing. Handcrafted features are defined to classify and segment the structures of the perifalcine region.

Abstract: In this paper we present an automatic segmentation method for structures of the perifalcine region in two-dimensional ultrasound (US) images. The segmentation method is implemented in context of improving image guidance during brain tumor (glioma) resection. Tissue removal and movement of brain tissue in glioma surgery considerably decrease the accuracy of neuronavigation systems which provide image guidance based on preoperative MRI (preMRI). The scanned central cerebral structures of the perifalcine region (such as the falx, the tentorium, the corpus callosum, and adjacent gyri) in intraoperative US (iUS) can be used as guiding frame to retrieve additional, spatially information and insights into local image and tissue deformation within the preMRI scan. Therefore, a segmentation of the perifalcine structures followed by a registration with the preMRI scan in future works may be used for an intraoperative update and even a partial fusion of the ROI in both modalities during different phases of the cerebral tumor resection. The method was evaluated on 50 US images and achieved on average a Dice coefficient of 0.68, a Hausdorff distance of 2.58 mm and a Jaccard index of 0.53.

Available data sets at the time of this study: Selection of 50 iUS images from seven different patients acquired before opening the dura mater and directly after opening the dura. As in the previously described approach the 50 iUS images were carefully selected

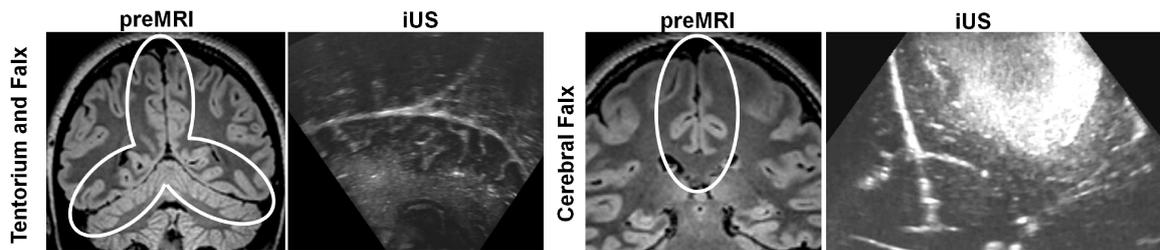


Figure 2.6: Perifalcine structures that can be equally depicted in preMRI and iUS and can help supporting image registration.

with the focus on uniqueness of the representation of target structures (perifalcine region) in scanned iUS volumes. An expert segmented structures of the perifalcine region in order to evaluate the segmentation method.

2.5.1 Segmenting Perifalcine Structures Utilizing OBIA

Fig. 2.6 shows the anatomical structures of the perifalcine region that are segmented in the here described approach. The figure shows the structure in both iUS and in preMRI. The proposed method starts with a preprocessing in which two diffusion stick filters are applied sequentially as described in [93] (see Section 2.4) and as proposed in [159, 171].

The advantage of the stick filter is that it offers edge enhancement of voxel values along lines and edges and thus, to emphasize the line-type structures of the perifalcine region while smoothing homogeneous regions of the US image with its characteristic speckles (Fig. 2.7). Subsequently, a median filter is applied with a 3×3 kernel to further suppress noise within the US image.

In the following a Hessian filter is employed in order to detect the local strength and direction of lines and edges of the perifalcine region as proposed in [93] (see Section 2.4). Within the Hessian filtered image the line-type structures are heightened and can be visualized as dark lines with the lowest intensities within the filtered image (Fig. 2.7).

This results into an image that delineates the lines and edges of the perifalcine region well enough to perform a watershed transform in order to create an initial over-segmentation. The over-segmentation is a prerequisite to allow an object-based image analysis (OBIA) as proposed in [46, 130]. The basic concept of the region-based analysis of OBIA is described in [46].

The over-segmentation separates the preprocessed US image in the initial objects that are further analyzed and characterized in order to segment the objects forming the perifalcine structures: the falx, the tentorium, the corpus callosum, and the adjacent gyri. The

initial objects should typically represent adherence to important region edges and borders, intensity and texture homogeneity, and compactness in order to delineate an adequate mosaic of the SOI within the US image as can be seen in Fig. 2.7.

For this we applied a watershed method as introduced in [42] that provides a parameter called ‘preflooding height’ p and that allows to define the size of the objects within the initial over-segmentation. The higher the chosen preflooding height, the larger are the object of the initial over-segmentation. In our case the lowest possible preflooding height $p = 0$ was chosen for the segmentation of the initial objects that are further analyzed using OBIA (Fig. 2.7).

Based on the preprocessed US image and the objects retrieved by the over-segmentation, the objects of the US image are classified using the following image features: intensity features (mean intensity, lower quartile, upper quartile), shape features (size, elongation, eccentricity, circularity), and neighbor-relation features (connectivity to the background of the US cone). The here computed image features for each object are described in more detail in [130].

The segmentation of the objects is performed in two major steps: First a rough preselection of possible objects belonging to the structures of interests (SOI) is made by using intensity features. Then neighboring possible candidates fulfilling the intensity criteria are merged to larger objects and the largest objects are further classified by the above mentioned shape features. The used OBIA approach was first presented in [46] and has already shown to be a powerful method for extracting and classifying objects by image features in US images as can be seen in [132].

2.5.2 Results

To evaluate the quality of our method we tested it on fifty 2D US images including US scans before and directly after dura opening and with different representations of the structures of the perifalcine region, varying in contrast, size, angle, and direction. The

Parameter	Dice coefficient	Hausdorff distance (mm)	Jaccard Index
Minimum	0.27	0.58	0.14
Maximum	0.84	11.00	0.73
Average±SD	0.68±0.11	2.58±2.42	0.53±0.11

Table 2.3: Comparison of the overall achieved segmentation results (SD: Standard Deviation).

segmentation results were compared to reference segmentations of a clinical expert with more than ten year's experience in image analysis of cerebral structures.

For detailed comparison of the proposed method with the references we chose the Dice coefficient, the Hausdorff distance, and the Jaccard index as quality measures (Tab. 2.3).

In 47 cases the SOI could be correctly detected, in which an average Dice coefficient of 0.68, a Hausdorff Distance of 2.58 mm, and a Jaccard index of 0.53 could be achieved. Some additional segmentation results are visualized in Fig. 2.8.

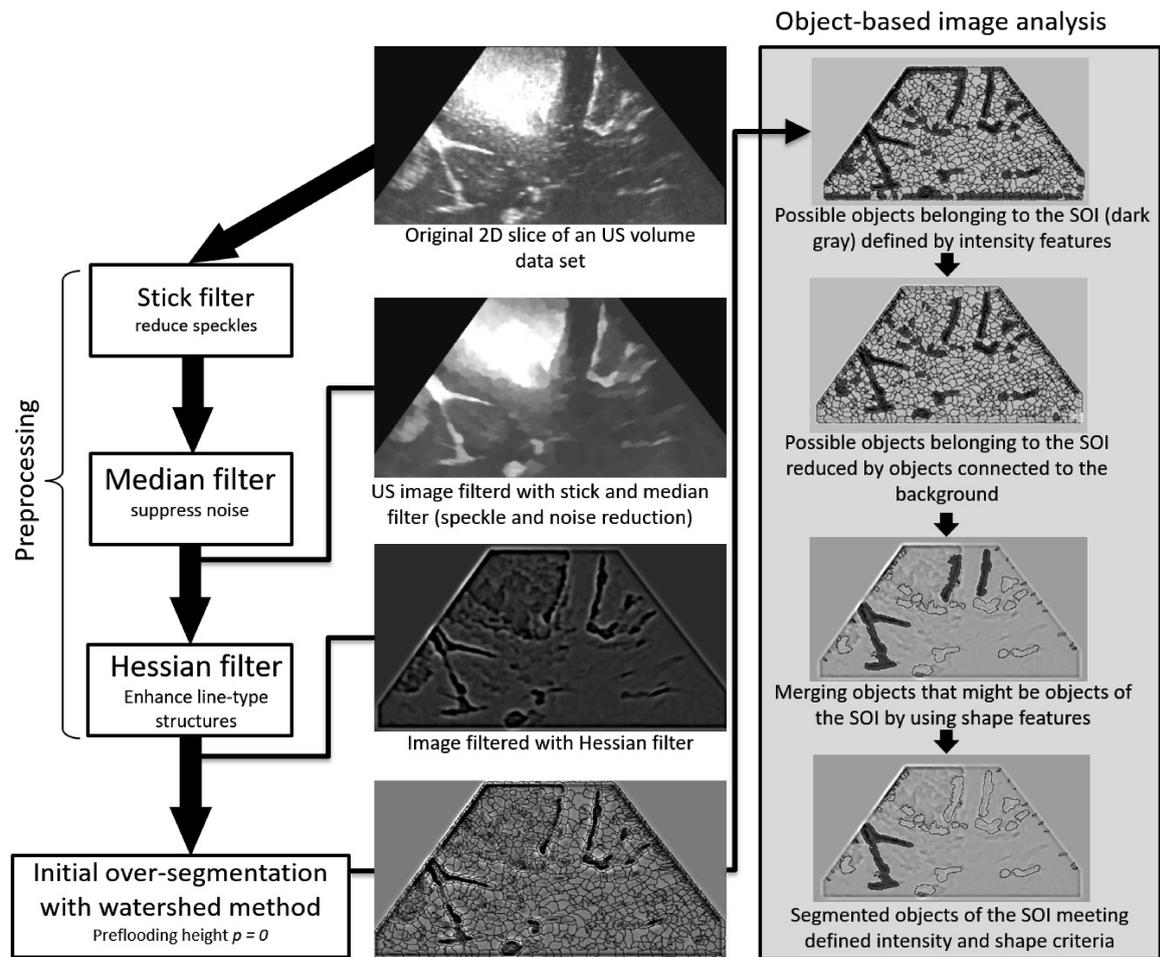


Figure 2.7: Image processing pipeline of the proposed segmentation approach.

2.5.3 Discussion

The segmentation of the perifalcine region is a challenging task due to variations of angle, elongation, size, and contrast of these line-type structures. The proposed method tends to undersegment the SOI which can be seen in Tab. 2.3 and in the screenshots of the

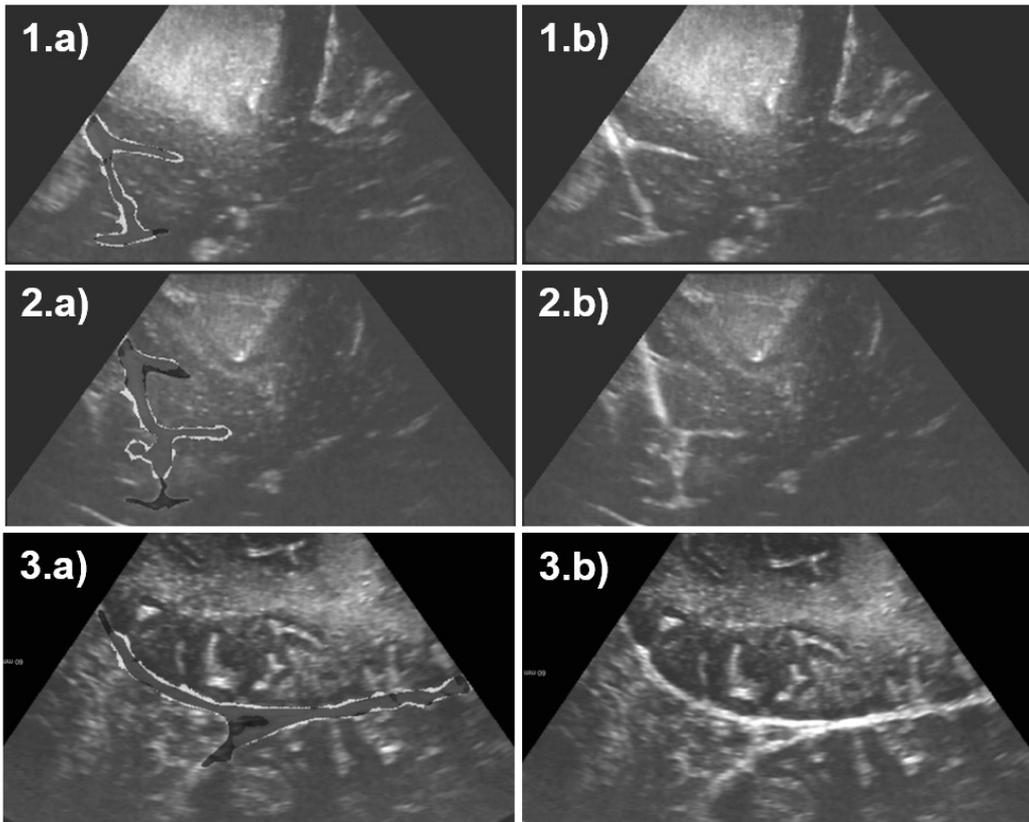


Figure 2.8: Visualization of additional segmentation results. The images show the reference segmentation (black) with an overlay of the segmentation results of the proposed method (white) in 1.a) - 3.a) compared to the unprocessed iUS images in 1.b) - 3.b).

segmentation results with references in Fig. 2.8. The reason for this is that the falx, the tentorium, and the corpus callosum provide good contrast to the image background and could, thus, be better detected and segmented than the adjacent gyri with considerably lower contrast in comparison to the previously mentioned structures. Sometimes there is just a thin connection between an adjacent gyrus and, i.e., the falx, which is difficult to identify by both the algorithm and the clinical expert.

This here proposed method segments all objects by the same image features chosen for the whole perifalcine region. In the subsequent approaches of this thesis the initial undersegmented structures/objects are further refined by identifying more specific image features for adjacent gyri that are connected to the initial detected object(s) (see hierarchical object-based image segmentation approach in Section 2.7).

2.5.4 Conclusion

An automatic segmentation method for the structures of the perifalcine region including the cerebral falx, the tentorium, the corpus callosum, and adjacent gyri was presented in this section. It must be pointed out that the segmentation of these structures is a challenging task due to shape and contrast variations of the line-type SOI.

In contrast to our previous studies [93, 92] (see Sections 2.4 and 2.5), it was tested within the here presented method if OBIA alone is an adequate approach to segment the perifalcine region and if appropriate enough image features could be determined that primarily describe the SOI. OBIA proved to be an approach with much potential for future works in segmenting the perifalcine region, verified by the promising initial results presented in this section. This is why the here proposed method was further refined in the subsequent approach of this thesis in Section 2.7.

2.6 Approach 3: Corresponding MRI Segmentations of the Perifalcine Region

The here described method was not published previously. The idea was to segment the falx cerebri and the tentorium cerebelli supplementary in preMRI in order to have segmentations of corresponding structures in iUS and preMRI for an initial alignment to start the registration process. But after being able of directly receiving the optical tracking information from the neuronavigation system together with the acquired iUS volumes during surgery a rough, image-based global initialization of the registration was not necessary anymore. Nevertheless, the segmentation could be useful to integrate other spatial information in the registration process by representing another reference structure. As aforementioned, the falx and tentorium are meningeal folds and, therefore, very thin fissures within the human brain (see Section 1.3.2). After contrast injection not only blood vessels within the brain are intensified, also the gaps between the right and left hemisphere (falx cerebri) and between the cerebrum and cerebellum are filled with contrast agent and highlighted as shown in Fig. 2.11.

Contributions: A segmentation algorithm that segments the falx cerebri and tentorium cerebelli in gadolinium-enhanced T1-weighted MRI. No other MR sequences are necessary to benefit from multiple contrasts in order to obtain satisfying segmentation results.

2.6.1 Segmentation of Falx Cerebri and Tentorium Cerebelli in preMRI

The target structures are sheet-like structures in 3D with relatively high gray value intensities in contrast-enhanced T1-weighted MRI as can be seen in Fig. 2.10 on page 59 and in Fig. 2.11 on page 61. The prerequisite for the here proposed segmentation method is a skull stripping in order to reduce other high intensity gray values and to reduce the overall computation time.

Skull stripping in contrast-enhanced T1-weighted MRI is more challenging than without contrast agent, especially, if the contrast-enhanced imaging is the only acquired scan before glioma surgery and when there is no option to utilize intensity differences in multiple sequences to segment anatomical structures by clever image subtractions. Additional boundaries within the brain appear through contrast enhancement of the vasculature or through contrast agent accumulation in or near the tumor.

After a non-uniformity intensity correction the same skull stripping is applied as described in detail in Section 2.7.1 in which context it was published. The first step of

the algorithm is now to enhance the sheet-like structures of the falx and tentorium. The basic concept to segment these structures is to use the eigenvalues of the Hessian matrix H $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$ to get a fissure probability for each voxel. The relation of the eigenvalues of the Hessian matrix H can be used to describe the local image structure. This basic concept is also used to enhance the line-type structures in iUS images in Section 2.4.1.

For example, a sheet-like structure has two small and one large eigenvalue. A tube- or a blob-like structure would be represented by two or three large eigenvalues. Before we calculate H using a derivative-of-Gaussian approach with $\sigma = 1$ mm, we further process the skull-stripped image. This is done in order to get an optimal input image to calculate H from which we want to extract the sheet-like structures.

We apply a morphological filter called constrained connection cost that fills all local valleys/cavities (areas with low gray value intensities) in the resulting image with a mean voxel intensity. The result can be seen in Fig. 2.10 in which the sheet-like target structures are predominantly structures with high intensities whereas the rest of the image has lower gray value intensities. We use the resulting image to calculate the Hessian matrix H .

As suggested in Lassen et al. [65] the feature image of the brain fulfills now similar criteria as the lung fissures between lobes and allows to use the there described filter for fissure enhancement (fissure similarity measure). The filter suppresses high intensity structures/voxels in the image with more than one large eigenvalue which represent typical features for tube-like blood vessels.

The resulting image has values within the interval $[0, 1]$ and after thresholding with an empirically defined value ≥ 0.1 the now obtained binary image contains possible fissure voxel candidates. A separate exclusion of vessel voxels in this binary candidate image is not necessary as proposed in Lassen et al. [65], because, in our case, the sheet-like structures of the falx and tentorium already represent structures with highest gray level intensities.

Afterwards, we also apply a vector-based connected component analysis which is based on the assumption that the largest eigenvalue of a sheet is perpendicular to the plane. The corresponding eigenvector of the largest eigenvalue represents the orientation of the structure/fissure. Neighboring fissure voxels have similar large eigenvalues. Therefore, a 3D connected components analysis with a 26-neighborhood can be applied to the candidate voxels from the binary image similar to van Rikxoort et al. [151]. The similarity is calculated by the inner product of the normalized eigenvalues which means a product of 1 for identical vectors. In our case we empirically defined an inner product of ≥ 0.99 to a connected component as sufficient enough for a good fissure segmentation. Fissure

components must have at least a volume of 0.1 ml in order to be considered part of the target fissure.

Subsequently, a morphological closing is applied with a kernel of $3 \times 3 \times 3$ voxels to close small gaps that are most likely due to the partial volume effect of the delicate fissures. Fig 2.10 shows an exemplary result of segmenting the fissures of the falx and tentorium in 3D and within 2D in the original contrast-enhanced MRI.

Alternative Preprocessing of the Feature Image

This alternative proved beneficial in four of five cases because otherwise structures from the brain surface were erroneously segmented as part of the fissures. In the remaining case

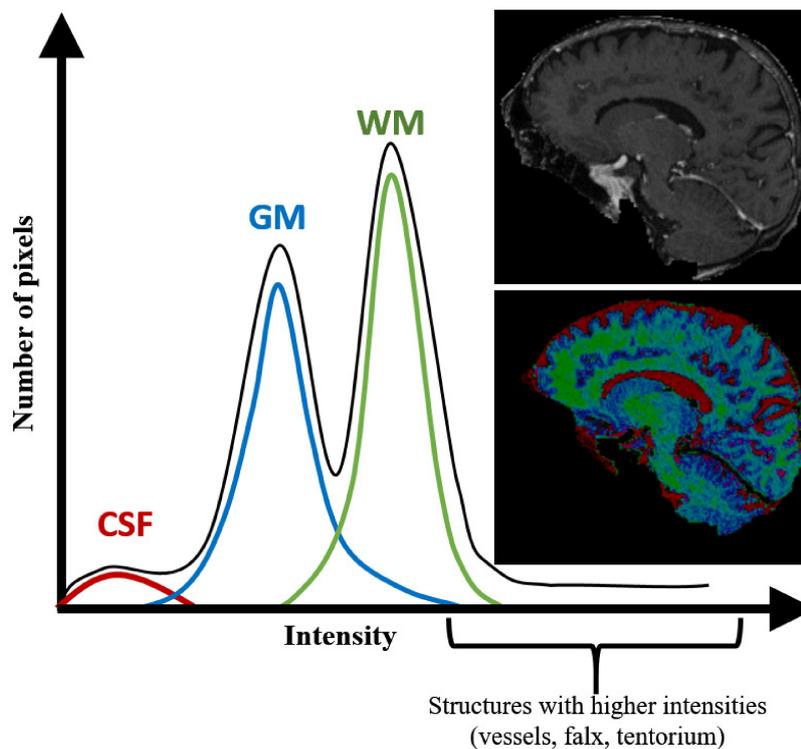


Figure 2.9: Smoothed model of a histogram of a non-uniformity corrected and contrast-enhanced T1-weighted MRI of the brain. The histogram shows the distribution of the three basic tissue classes (CSF, GM, WM) from a patient in our data base. The distribution can slightly differ from those of a healthy patient without a brain tumor. The images on the right show a colored sagittal brain image with CSF (red), GM (blue), and WM (green).

the alternative had neither a positive nor a negative impact on the attained segmentation result. For this the skull-stripped image is masked with its original values and a histogram analysis is performed. The histogram is smoothed and characteristic extreme values are extracted from the resulting curve. Each peak in the curve represents a tissue class (Gray matter (GM); white matter (WM), and cerebrospinal fluid (CSF)) that can be modeled by a Gaussian distribution.

This method is also termed as Gaussian mixture model (GMM). In our case we are only interested in deleting the structures WM and GM in our feature image. The fissures of the falx and tentorium are surrounded by CSF and excluding this fluid would also delete parts of the target structures/fissures due to the partial volume effect.

More specifically, GM and WM are extracted because we want to prevent extracting the surface of the brain (GM) or the surface of the WM as largest sheet-like structure. GM and WM represent the two largest turning points in the smoothed histogram curve that can be used as the mean (intensity) values to calculate each Gaussian distribution. The distance from these peaks to the local minimum between GM and WM in the histogram curve is used as standard deviation to calculate the Gaussian distributions (see Fig. 2.9).

The Gaussian distributions are used to threshold the image resulting in a reduced binary brain mask image. It can be noted from the histogram that there is an overlap between different tissue classes. Also, it can be seen that this overlap is higher between WM and GM than between GM and CSF. As we are only interested in a rough exclusion of WM and GM we neglect these overlaps. Through morphological filtering we aim to close smaller gaps in GM and WM by simultaneously preserving the target structure in the resulting binary mask.

2.6.2 Results

The segmented fissures of the falx and tentorium are very thin anatomical structures and sometimes discontinuous due to the partial volume effect. However, these structures with an occasional thickness of less than a couple of voxels have a considerable volume by representing a central anatomical structure of the brain. Therefore, creating manual reference masked by an expert was seen as too tedious and error prone. Instead, the results were visually inspected by an expert and evaluated as good and acceptable results considering the partial volume effect which lead to undersegmentation of the overall volume of the segmented structures. In Fig. 2.10 we show an example of a segmentation result in 2D (in sagittal, coronal, and axial view) and in 3D.

2.6.3 Discussion

We received very good segmentation results on five different contrast-enhanced MRI data sets. Nevertheless, the method can be imagined error prone in the presence of large tumors with increased vasculature and cysts with a considerable uptake of contrast agent. These structures would represent large clusters of higher gray value intensities. However, these structures could be easily segmented previously and the regions could be deleted from the skull-stripped image prior the here described falx and tentorium segmentation. A tumor segmentation is likely to be done anyway in the course of surgical planning of an image-guided intervention with a neuronavigation system.

Furthermore, the here segmented structures are different from those segmented in iUS in the previous chapters because only the fissures are segmented here and not the falx and tentorium with adjacent gyri and sulci that are filled with cerebrospinal fluid. Due to the partial volume effect and the thinness of the structure, it matters how the patient's falx and tentorium are positioned within the MRI scanner before image acquisition. Moreover, the scanning time after contrast injection that also correlates with the contrast uptake in the fissures has an impact on the segmentation quality. The displayed result in Fig. 2.10 shows a patient with good contrast uptake in the fissures and with less impact of the partial volume effect.

However, this method could represent a preprocessing of extracting the adjacent sulci and gyri by WM and GM segmentation only in the vicinity of the segmented fissures of the falx and tentorium. This spatial information could be used to fuse with the segmentations in iUS imaging of the perifalcine region and could be used to initialize and to speed-up the multi-modal image registration.

The segmentation of these fissures in contrast-enhanced MRI could also be helpful in other application scenarios, i.e. for segmenting the right and left hemisphere and for further anomaly detection of those "normally" mirroring compartments. Additional MRI sequences were not available for most of our data sets and contrast-enhanced T1-weighted MRI is the standard imaging acquired prior glioma surgery. Otherwise, MRI sequences with different contrast properties or with or without contrast agent could have improved or at least speeded up an intensity-based segmentation of the falx and tentorium in this case. Direct sequentially acquired pre- and post-contrast injection images could be subtracted from each other and a segmentation of blood vessels together with brain fissures could be obtained. Based thereon, the here described method could be used to differentiate the sheet-like structures of the falx and tentorium from the blob- and tube-like structures of blood vessels.

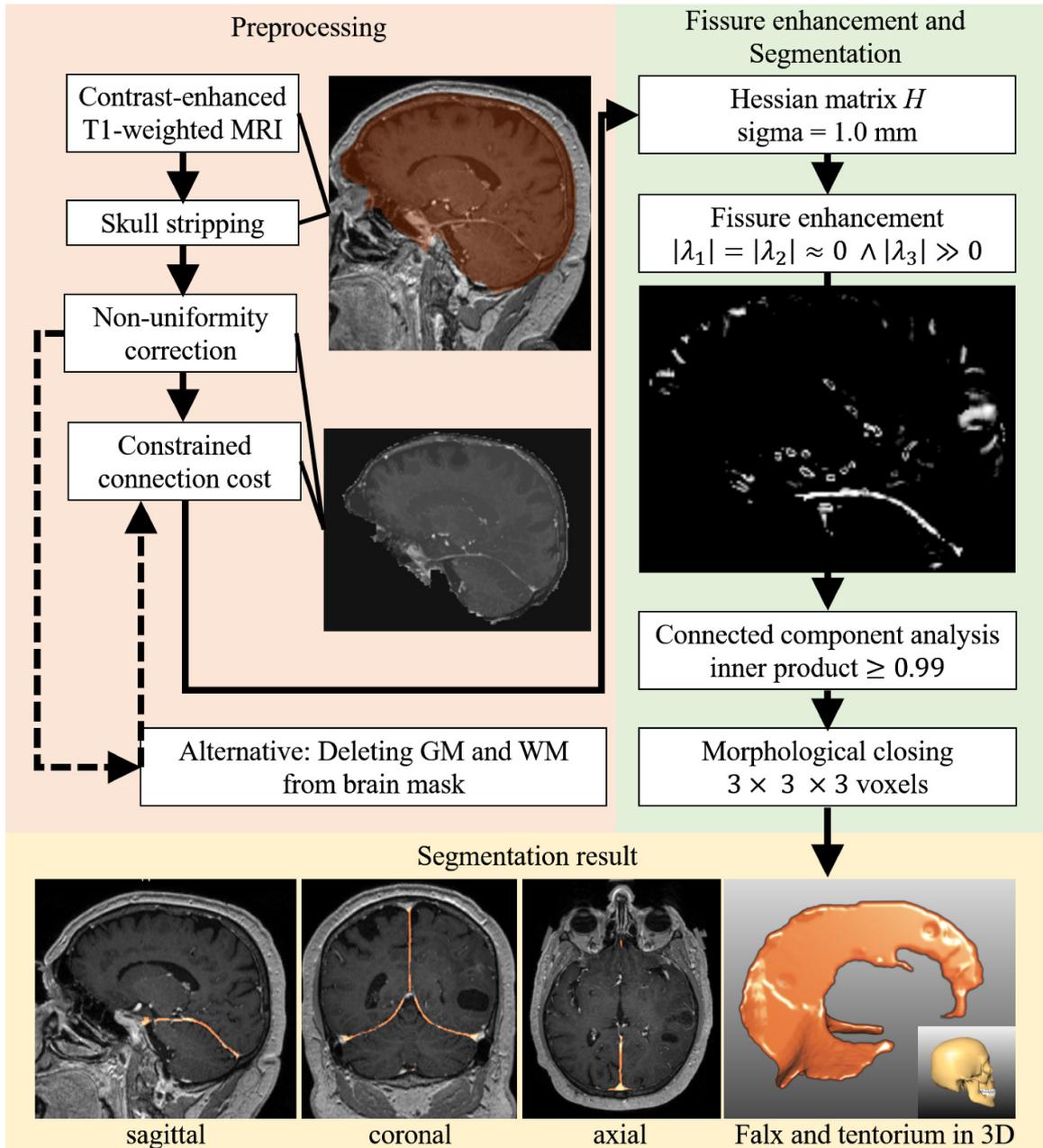


Figure 2.10: Overview of the falx and tentorium segmentation algorithm in preMRI with image preprocessing, fissure segmentation, and segmentation results.

2.7 Approach 4: Automatic MRI-US Segmentations for Improving Registration in Image-Guided Neurosurgery

Published as: J. Nitsch, J. Klein, P. Dammann, K. Wrede, O. Gembruch, J. H. Moltz, H. Meine, U. Sure, R. Kikinis, D. Miller (2019) *Automatic and Efficient MRI-US Segmentations for Improving Intraoperative Image Fusion in Image-Guided Neurosurgery*. In: *NeuroImage: Clinical*, 33, 101766. [91]

Contributions: Advanced, very sophisticated, and hierarchical OBIA approach for segmentation of falx cerebri and tentorium cerebelli with adjacent sulci and gyri. The approach differentiates between the perifalcine structures and other sulci and gyri in the vicinity (for anatomical information see Section 1.3.2). Furthermore, the initial registration by the neuronavigation system and a solely intensity-based registration are compared to a registration approach that incorporates the segmented structures of the perifalcine region in iUS for image fusion of preMRI with iUS imaging.

To our best knowledge our paper published the first iUS segmentation approach applicable within the process of glioma surgery in order to increase registration accuracy, speed, and robustness of preMRI and iUS images for improving intraoperative image guided therapy (IGT).

The second contribution of this paper is the introduction of an initial registration approach which demonstrates how valuable an incorporation of segmentations is in reducing the overall computation time and in increasing registration accuracy.

Abstract: Knowledge of the exact tumor location and structures at risk in its vicinity are crucial for neurosurgical interventions. Neuronavigation systems support navigation within the patient's brain, based on preoperative MRI (preMRI). However, increasing tissue deformation during the course of tumor resection reduces navigation accuracy based on preMRI. Intraoperative ultrasound (iUS) is therefore used as real-time intraoperative imaging. Registration of preMRI and iUS remains a challenge due to different or varying contrasts in iUS and preMRI. Here, we present an automatic and efficient segmentation of B-mode US images to support the registration process. The falx cerebri and the tentorium cerebelli were identified as examples for central cerebral structures and their segmentations can serve as guiding frame for multi-modal image registration. Segmentations of the falx and tentorium were performed with an average Dice coefficient of 0.74 and an average Hausdorff distance of 12.2 mm. The subsequent registration incorporates these

segmentations and increases accuracy, robustness and speed of the overall registration process compared to purely intensity-based registration. For validation an expert manually located corresponding landmarks. Our approach reduces the initial mean Target Registration Error from 16.9 mm to 3.8 mm using our intensity-based registration and to 2.2 mm with our combined segmentation and registration approach. The intensity-based registration reduced the maximum initial TRE from 19.4 mm to 5.6 mm, with the approach incorporating segmentations this is reduced to 3.0 mm. Mean volumetric intensity-based registration of preMRI and iUS took 40.5 s, including segmentations 12.0 s.

Available data sets at time of this study: Eleven patients with preMRI and corresponding iUS scans. An expert created reference segmentations for the iUS images as well as corresponding landmarks in iUS and preMRI in order to evaluate the multi-modal registration accuracy.

2.7.1 Segmentation-Based Registration for MRI-US Fusion

The segmentation-based registration approach can be divided into three different parts: 1.) Segmentation of perifalcine structures in iUS images, 2.) Skull stripping of preMRI scans as preprocessing prior registration, and 3.) Registration of preMRI with iUS images. These are described separately in the following three sections. In Fig. 2.12 an overview of the entire method for preMRI and iUS image fusion is given.

In this publication we proposed a novel approach that combines automatic segmenta-

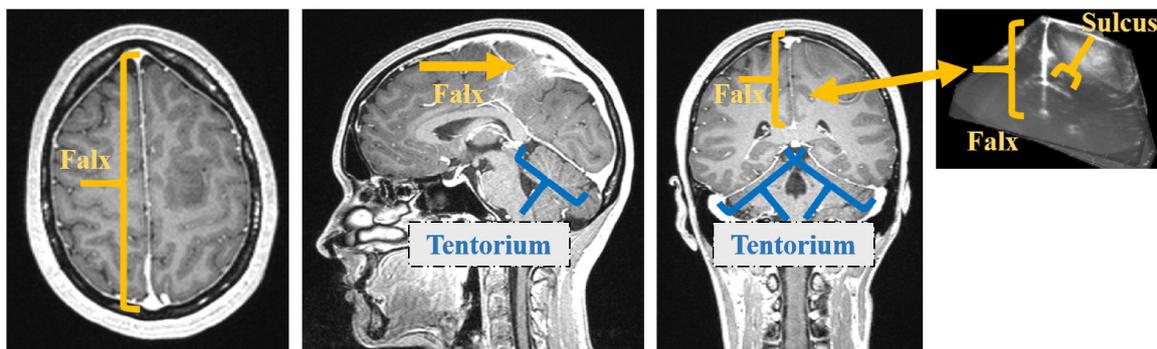


Figure 2.11: Anatomical structures of falx cerebri, tentorium cerebelli, and adjacent gyri (sg. gyrus) and sulci (sg. sulcus) in MRI and US imaging. This figure also gives an impression about the central position and thinness of these anatomical structures.

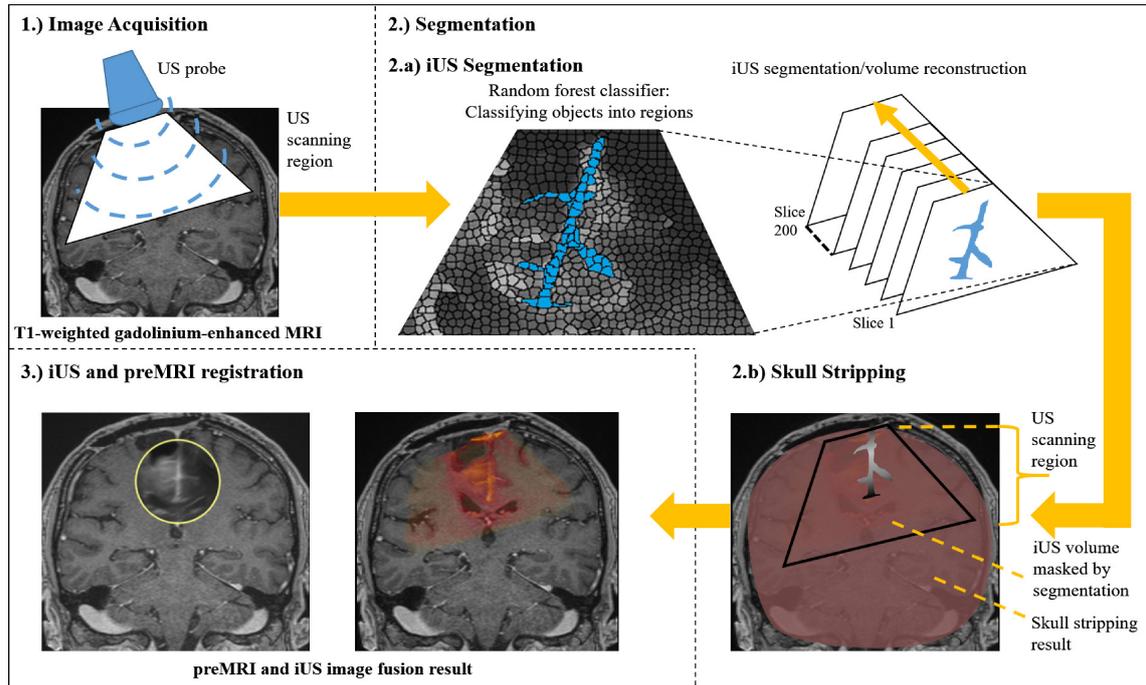


Figure 2.12: Overview of the entire segmentation and registration method for preMRI and iUS image fusion. Basically, the method consists of three steps: 1.) Image Acquisition (see Section 2.3): Acquisition of preoperative T1-weighted gadolinium-enhanced MRI and intraoperative B-mode ultrasound images. The iUS images are acquired freehand with a micro-convex US probe. 2.) Segmentation: 2.a) iUS segmentation (see Section 2.7.1.1): The falx cerebri and the tentorium cerebelli with their adjacent gyri and sulci are segmented slice-wise in 2D before the segmentation results are reconstructed to a 3D volume. Each US volume consists of 200 slices. For iUS segmentation an object-based image analysis approach is used that classifies objects into regions; 2.b) Skull stripping (see Section 2.7.1.2): The brain segmentation is essentially used to reduce complexity within the registration task in order to speed-up the computation and to prevent erroneous registrations. The skull stripping could be done prior to surgical intervention to avoid adding intraoperative computation time. 3.) iUS and preMRI registration (see Section 2.7.1.3): The masked preMRI brain image and the corresponding reconstructed iUS volume masked with the falx, tentorium, and adjacent gyri segmentations are used for registration in order to solely employ these anatomical structures for an initial alignment of both modalities.

tions of central anatomical structures that can act as a guiding-frame for a subsequent registration approach. The falx cerebri and tentorium cerebelli as prominent and central structures of the perifalcine region and for the brain in general are just examples of structures fulfilling the correspondence criteria in gadolinium-enhanced preMRI and in B-mode iUS images (see Fig. 2.11).

The segmentations are utilized as additional spatial information for the registration to enable an alignment of selected structures – and, respectively, for masking out possible disruptive structures – for supporting fast, automatic, and robust iUS/preMRI fusion in order to improve real-time image guidance during neurosurgery.

2.7.1.1 Automatic Segmentation of the Perifalcine Region in iUS Images

As mentioned in Section 2.3 Material and Methods, the input data for the segmentation method in iUS imaging are freehand and tracked series of 2D US scans which are segmented slice-wise.

The delicate anatomical structures of the falx, tentorium, and their adjacent gyri and sulci are segmented in 2D in order to provide a segmentation within images of maximal resolution and quality which would be reduced by a previously applied reconstruction. For each patient US scans with 200 slices are acquired and subsequently segmented prior to volume reconstruction. The image extent of each slice is 531×376 pixels.

Basically, the method described here is an object-based image analysis approach that classifies candidate regions within the US image series [46]. Compared to pixel-based classification it is advantageous that a region-based approach offers more expressive features describing the specific target structures or objects within the image. Especially if differences between target objects and background are subtle – as they are for falx, tentorium, sulci, and gyri – it is important to define sufficiently specific features to characterize the target objects.

This is the reason why we chose the object-based image analysis approach because the target structures here appear line-like for the human eye but are in fact discontinuous lines with enormous changes in their intensity along their extent even within the same iUS slice. Classic segmentation algorithms rely on homogeneous regions within and/or similar strong edges around the target structure. Both of these criteria are not given in this particular segmentation task.

The object-based image analysis approach is a powerful method to locate different, discontinuous regions of the target structure within an over-segmented image. Accordingly, the region-based segmentation requires a good initial over-segmentation of the whole

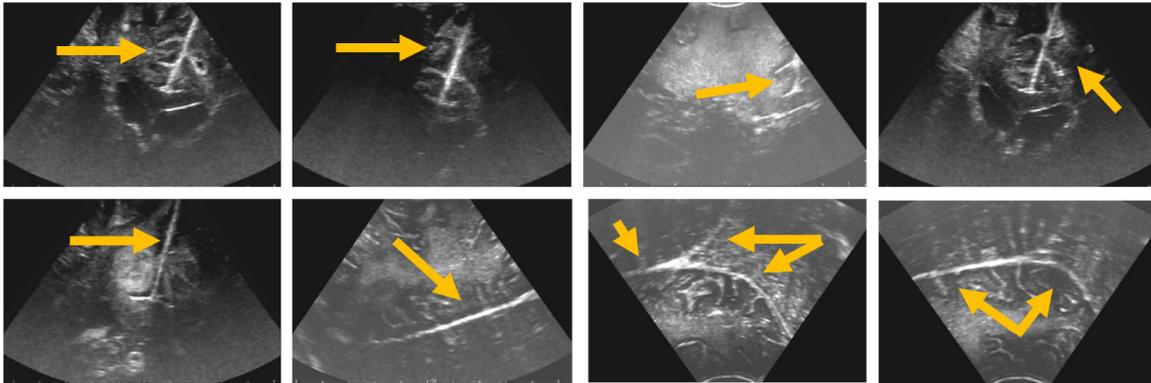


Figure 2.13: Falx and tentorium in iUS images showing their different representation as line-type structures scanned from different angles by the US probe.

image that captures the boundaries of the target structures. To facilitate a good initial clustering of image regions the US image slices are preprocessed with two subsequently applied stick filters by adopting [93] (see Section 2.4).

The stick filters sustain and strengthen line-type structures, while smoothing homogeneous regions and reducing characteristic speckles in US images. The falx and the tentorium may be scanned from different angles but the line-type character remains still present. Fig. 2.13 shows the falx and the tentorium and their different representation in iUS images.

In order to generate the over-segmentation, watersheds – as suggested by Schvier et al. [130] – and the SLIC Superpixel approach [1] are tested and compared, resulting in the conclusion that the SLIC Superpixel approach provides a more precise and robust partitioning of target objects of the falx, tentorium, gyri, sulci, and their boundaries. An example of an initial over-segmented US image can be seen in Fig. 2.14.

For the classification process a Random Forest classifier with 25 trees is used to classify the candidate regions. The classifier is trained with 686 positive samples and 824 negative samples from characteristic iUS slices of patient 1 that were excluded from the test data set. The training samples were selected manually by a human expert in order to guarantee that different examples with changing intensities as well as different perspectives of the target structures are covered by the training data.

The scanned iUS volume of patient 1 was especially acquired as training data set to fit these criteria and almost every slice within the US sweep represents an example of falx and tentorium with adjacent gyri and sulci. In the first step of the segmentation algorithm, all objects generated by the initial over-segmentation are classified.

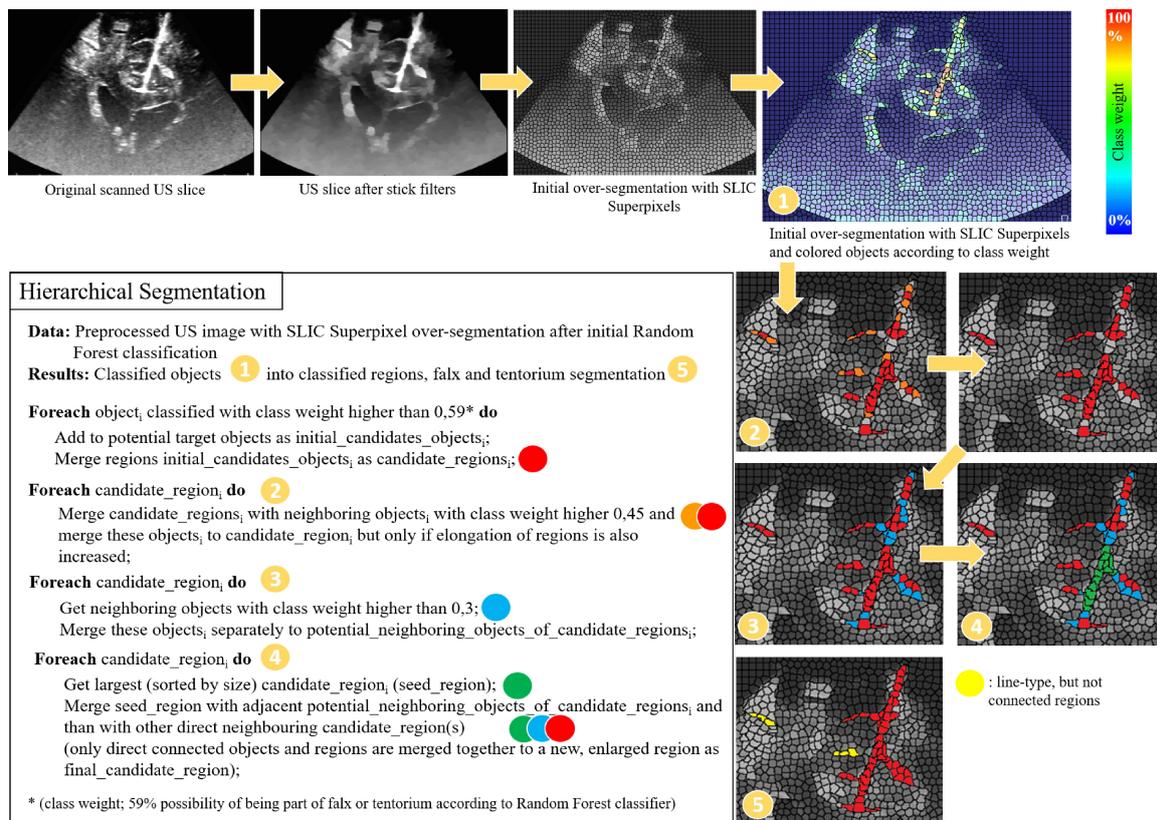


Figure 2.14: Pseudo code of the hierarchical classification and segmentation process. The segmentation approach first classifies objects and then objects into regions.

Hereafter, the region-based features are listed that are used for initial classification of these objects as originally proposed by Schwier et al. [130]: Intensity-based features (lower and upper quartile, mean, median, minimum, maximum, standard deviation), shape-based features (size, eccentricity, elongation, circularity), and direct neighbor relation features (connectivity to other objects and their features).

In addition to those, a subsequent classification process is applied that uses the initial classification result as basis to further analyze the connectivity of objects classified as potential candidate objects of the structures to be segmented and takes their neighbor relations to further analyze their probability of being part of the falx and tentorium in order to generate a more precise characterization of the overall anatomical structures of falx, tentorium, and their adjacent gyri and sulci.

This hierarchical segmentation approach first classifies objects and then merges objects into regions and is described in pseudo code and visualized in Fig. 2.14.

2.7.1.2 Skull Stripping of preMRI Scans as Preprocessing prior Registration

To prevent initial fitting to prominent pressure artifact or other artifacts of the US probe in iUS images onto the skull or brain surface in gadolinium-enhanced preMRI and in order to reduce the total computation time, a skull stripping is performed on the head and neck scans of the 11 patients prior to registration.

A profound review of methods on skull stripping on non-contrast-enhanced images is given by Kalavathi et al. in [51] in which methods are distinguished in five different categories: mathematical morphology-based, intensity-based, deformable surface-based, atlas-based, and hybrid methods. Kleesiek et al. [58] used Convolutional Neural Networks (CNNs) trained for skull stripping of 53 patient data sets consisting of T1-weighted images with and without contrast agent, T2-weighted and FLAIR images.

For our field of application of image guidance during glioma surgery, contrast-enhanced T1-weighted images are used as basic preMRI for navigation and were acquired a couple of days prior to surgery for planning and intraoperative image guidance. These contrast-enhanced images come with additional difficulties such as varying uptakes of contrast agents due to different time points the MRI scan was performed after exposition to the contrast agent and also varying contrast uptakes of pathological brain tissues, creating patient specific-boundaries within the brain to be skull-stripped.

According to Kalavathi et al. [51] the proposed segmentation method for brain extraction can basically be categorized as a hybrid method using mathematical morphology-based as well as intensity-based segmentation methods. As initial step a non-parametric method for automatic correction of intensity non-uniformity is used as described by Sled et al. [139].

Afterwards, a threshold is applied that focuses on filtering out lower intensities such as the cerebrospinal fluid and partly gray matter in order to increase the immediate gap between the center of the brain and the skull.

Subsequently, a morphological opening is employed that separates the thresholded image roughly in two different components: Brain and skull/neck. Successionally, a 3D connected components analysis is performed using a 3D-18-neighborhood.

Simultaneously, the center of gravity is calculated and the component that is closest to it is taken as a rough estimate of an initial coarse brain segmentation.

The previously mentioned threshold, that roughly separates the brain from the skull and neck, is adaptive, increasing continuously, starting from the experimentally determined value of 200 with a step size of 10 as long as there are less than two components after connected component analysis.

The non-uniformity intensity corrected image is then smoothed with a 3D Gauss

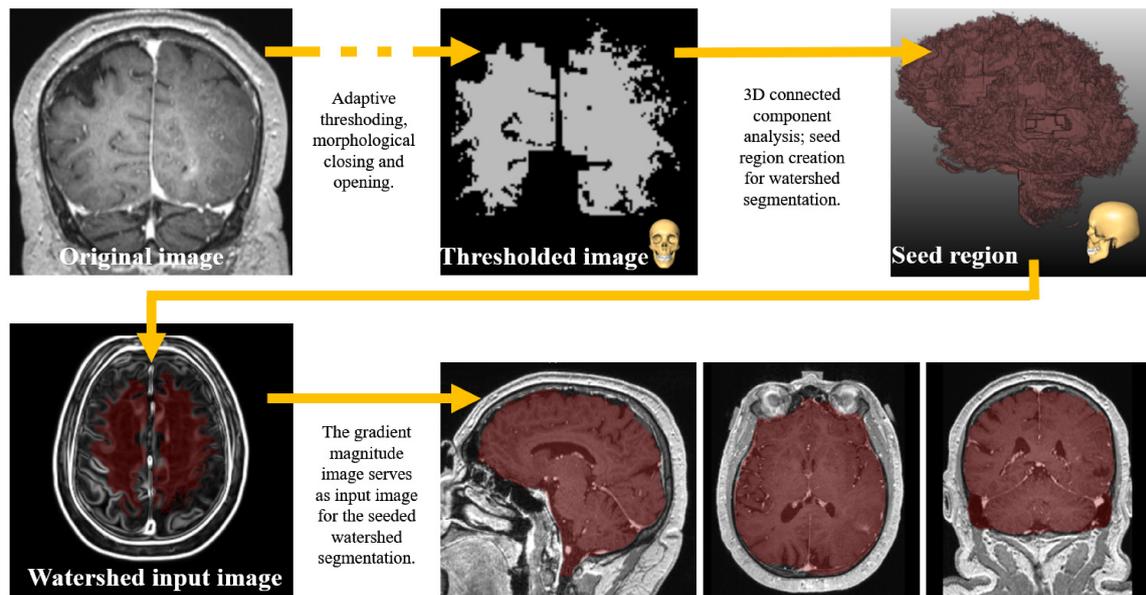


Figure 2.15: This figure shows the main image processing steps within the outlined skull stripping approach. At the end of the processing pipeline the resulting brain mask of patient 4 is shown and overlaid on the original preMRI in sagittal, axial, and coronal view.

smoothing using a sigma of 1 voxel on which the gradient magnitude is calculated with identical sigma before it serves as input image for a seeded watershed transformation. The applied watershed transformation method was proposed by Hahn et al. [41].

In our case, the previously attained, jagged segmentation of the brain serves as seed region. Remaining gaps within the brain mask are removed after watershed segmentation and are shut by morphological closing with a kernel of $3 \times 3 \times 3$.

Finally, a rolling ball closing with a radius of 5 mm is applied to remove residual holes and to smooth the overall surface of the segmented brain. Fig. 2.15. shows the main image processing steps of the skull stripping approach.

2.7.1.3 Registration of preMRI with iUS Images

Before both imaging modalities can be registered, the iUS scans of 200 slices per patient and the corresponding 2D segmentations must be reconstructed into a 3D volume. The reconstructed volumes are attained through accumulating the scanned or segmented 2D slices/volumes.

Besides trilinear interpolation, intersecting volumes are combined by means of weighted

averaging. For multi-modal image fusion we used the MERIT registration framework [7].

As a preprocessing step, the original gadolinium-enhanced T1-weighted image is resampled with a Lanczos filter to an isotropic voxel size of 1 mm [27, 13]. The corresponding reconstructed iUS volume is masked with the falx, tentorium and adjacent gyri segmentations to solely use these anatomical structures for an initial alignment of both imaging modalities.

Simultaneously, the head scan is masked by its skull stripping segmentation result to reduce the image content and computations needed for registration. Prior to registration all images masked by segmentations are resampled to the same isotropic voxel size of 1 mm.

The MERIT framework includes multiple significantly different image similarity measures. For our purpose we used the local cross-correlation (LCC) similarity measure that evaluates the local cross-correlation in a local window of $3 \times 3 \times 3$ around each voxel. LCC is robust to non-linear intensity relations which proved advantageous for our rigid iUS and preMRI registrations and has similar properties as the Normalized Mutual Information (NMI) similarity measure which is considered a standard similarity measure for robust multi-modal image registration.

Furthermore, we used a multi-level registration approach consisting of three levels which allows a downscaling of the image to be registered with the following scaling factors 0.63^2 , 0.63 and 1 . This allows the deformed image to be registered from coarse to finer resolutions.

2.7.2 Results

In the following we evaluate the previously described segmentation methods of the falx cerebri, the tentorium cerebelli, and adjacent gyri and sulci in 2D B-mode US images and the skull stripping method on gadolinium-enhanced T1-weighted preMRI.

To assess the quality of the segmentation, we use standard quality measurements by means of the Dice coefficient and Hausdorff distance.

Furthermore, we analyze our implemented registration approach for multi-modal image fusion of iUS and preMRI with distance measurements to landmarks placed by a human expert. We evaluate our registration results by comparing the initial registration from the navigation system with a solely intensity-based registration and with our segmentation-based registration approach that incorporates the prior attained segmentation results.

An overview of the whole segmentation and registration approach is given in Fig. 2.12. In the following we first evaluate the segmentation results before we measure the quality of

preMRI and iUS image fusion. This section concludes with an evaluation of inter-observer variability of falx and tentorium reference segmentations in iUS imaging.

Evaluation of segmentations in 2D B-mode images

Tab. 2.4 shows the 2D segmentation results of the 11 patients for segmenting the falx cerebri and the tentorium cerebelli in B-mode US images which is illustrated in Fig. 2.14.

The automatic segmentation results are compared to reference segmentations attained by segmenting the falx and the tentorium manually by an expert in 2D B-mode US slices. Consequently, the segmentations are also evaluated slice-wise. Average Dice coefficient and Hausdorff distances per US scan (set of 200 slices) or patient, respectively, are shown in Tab. 2.4.

The proposed method achieved an average Dice coefficient of 0.74 and an average Hausdorff distance of 12.2 mm over all 11 patients with an average computation time of 79 s for processing 200 US slices.

Evaluation of the Skull Stripping for Gadolinium-Enhanced T1-weighted MRI

Furthermore, we evaluated our skull stripping method (Fig. 2.15) and compared it to a reference segmentation by an expert. The subsequent Tab. 2.5 gives an overview of

Patient	Dice coefficient	Hausdorff distance (mm)	Computation time (s)
1	0.89	6.3	103
2	0.69	15.9	88
3	0.67	12.0	84
4	0.88	7.3	96
5	0.86	7.8	93
6	0.72	14.2	79
7	0.65	15.9	75
8	0.63	16.0	62
9	0.65	15.2	58
10	0.73	13.0	65
11	0.77	11.1	78
Mean	0.74	12.2	79

Table 2.4: Segmentation results of the implemented automatic 2D segmentation approach for falx cerebri and tentorium cerebelli segmentation in B-mode iUS images.

the segmentation results and the computation times for each patient. We focused our segmentation approach on a balance between speed and accuracy.

A manual correction can be done as post-processing step in which the resulting mask is interpolated between two indicated corrections. Corrections are only necessary if the segmentation is not accurate enough in the required region where the acquisition of B-mode US scans is planned during the intervention.

Automatic and manual correction times are also displayed in Tab. 2.5. The skull stripping method attained an average Dice coefficient of 0.82 and an average Hausdorff distance of 33.8 mm with an average computation time of 106 s.

Evaluation of the iUS and preMRI Registration

Results of the different registrations methods are compared in Tab. 2.6. For each patient ten to twelve corresponding landmarks between iUS and preMRI were defined by an expert. Moreover, we compare the computation time of the purely intensity-based method and the registration using the prior attained segmentations of the falx cerebri and the tentorium cerebelli in US images and the skull stripping from preMRI.

A comparison of computation times is given in Tab. 2.7. Registration results for iUS and preMRI image fusion are visualized in Fig. 2.16. All segmentations and registrations are performed on an Intel Core i7 PC with 2.60 GHz and a 16 GB RAM.

Patient	Dice coefficient	Hausdorff distance (mm)	Computation time (s)	Corrections
1	0.84	28.4	112	*
2	0.79	33.8	100	118
3	0.72	54.9	104	45
4	0.82	32.9	92	*
5	0.84	25.1	118	*
6	0.82	48.2	110	58
7	0.83	19.7	90	11
8	0.83	27.9	124	*
9	0.84	34.8	108	67
10	0.87	38.6	113	*
11	0.83	27.7	99	*
Mean	0.82	33.8	106	

Table 2.5: Segmentation results of the automatic 3D skull stripping method for gadolinium-enhanced T1-weighted MRI. * Indicates that no manual correction was needed.

Patient	Initial registration	Intensity-based registration	registration with segmentations
1	10.3(9.6-11.2)	3.3(2.5-5.6)	2.3(1.9-3.1)
2	5.8(3.0-8.0)	3.0(2.0-4.3)	1.7(1.0-2.3)
3	4.8(2.7-5.6)	3.2(2.3-4.0)	2.0(1.6-2.6)
4	6.2(4.0-8.3)	5.9(3.8-7.9)	3.2(2.2-4.3)
5	16.1(11.8-19.9)	3.0(2.4-4.3)	2.1(1.8-2.8)
6	24.3(22.0-25.7)	3.7(2.1-5.6)	2.1(1.7-2.7)
7	10.5(7.9-13.1)	3.6(2.6-5.5)	2.0(0.9-2.5)
8	31.1(30.1-34.0)	4.5(2.3-6.3)	2.2(0.9-3.0)
9	22.8(19.2-28.2)	3.0(1.3-4.8)	1.5(0.5-2.6)
10	36.1(31.9-40.1)	5.7(4.2-7.5)	3.1(2.0-4.3)
11	17.6(15.9-19.3)	3.6(1.7-5.7)	2.2(1.3-3.2)
Mean	16.9(14.4-19.4)	3.9(2.5-5.6)	2.2(1.5-3.0)

Table 2.6: mTRE values with initial registration, intensity-based registration, and with registration incorporating segmentations. For each patient ten to twelve landmarks were used to evaluate the registration accuracy. Within the brackets the lowest and the highest landmark distances within each patient are displayed.

Comparison of computation time (s)		
Patient	Intensity-based registration	registration with segmentations
1	32	9
2	31	9
3	45	19
4	30	7
5	28	11
6	30	12
7	84	20
8	30	13
9	45	15
10	48	8
11	42	9
Mean	41	12

Table 2.7: Comparison of computation time (s) needed in order to compute the preMRI and iUS registrations.

Inter-Observer Variability of Falx and Tentorium Reference Segmentations

Moreover, we analyzed the inter-observer variability of the combined falx, tentorium, sulci, and gyri reference segmentations in iUS images. For this we compared the reference segmentations from three different experts and used the Dice coefficient to assess similarities. We included this comparison in our evaluation in order to spotlight how complex the segmentation task is even for human experts for these central anatomical structures within iUS images. The Dice coefficient in Tab. 2.8 is calculated between our segmentation result and the three different references. In Tab. 2.9 the Dice coefficient is computed between the references of Expert 1 and Expert 2 as well as between Expert 1 and Expert 3 in order to determine the resemblance of the expert’s segmentations.

Inter-observer variability (dice coefficients)			
Patient	Expert 1	Expert 2	Expert 3
1	0.89	0.92	0.79
2	0.69	0.73	0.58
3	0.67	0.57	0.45
4	0.88	0.82	0.76
5	0.86	0.72	0.60
6	0.72	0.53	0.48
7	0.65	0.66	0.38
8	0.63	0.71	0.43
9	0.65	0.70	0.52
10	0.73	0.69	0.47
11	0.77	0.74	0.55
Mean	0.74	0.71	0.55

Table 2.8: Inter-Observer variability of the falx cerebri and tentorium cerebelli using three different expert’s segmentations to evaluate our iUS segmentation approach. Results are compared using the Dice coefficient. Segmentations from Expert 1 were used to assess the quality of the here proposed perifalcine region segmentation, see Tab. 2.4 for comparison.

2.7.3 Discussion

In general, the automatic segmentation of the perifalcine region is especially challenging due to the variance of tumor locations and neighboring anatomical structures, size, and image representation that correlates with the number of tumor patients. Nevertheless, we

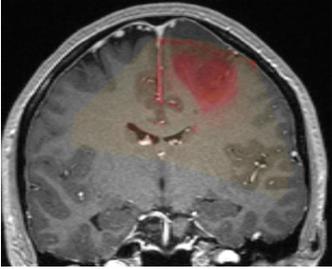
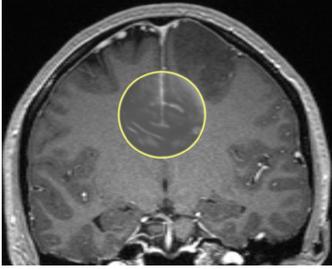
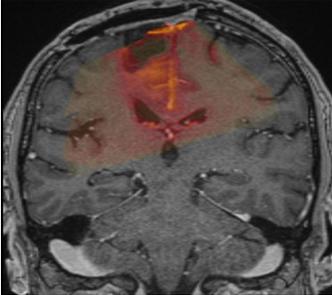
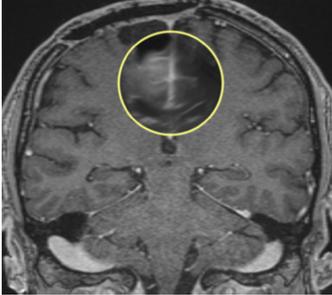
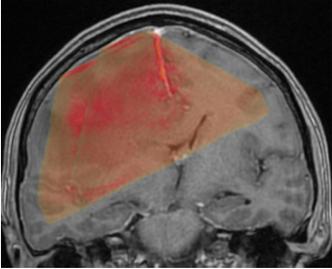
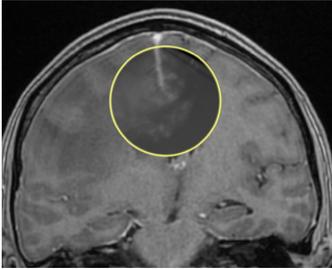
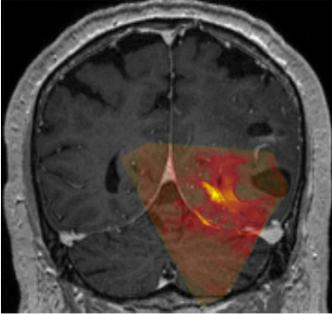
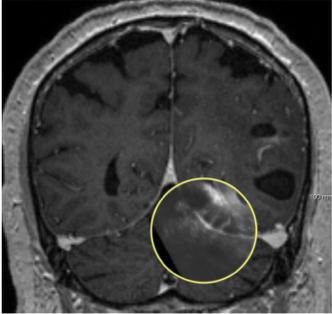
Patient 1			mTRE of 2.3(1.9-3.1)
Patient 2			mTRE of 1.7(1.0-2.3)
Patient 3			mTRE of 2.0(1.6-2.6)
Patient 4			mTRE of 3.2(2.2-4.3)

Figure 2.16: Registration results represented in a coronal view with a colored overlay (preMRI in gray; iUS in red) within the left column. Furthermore, an additional view for image fusion is proposed. Here the clinical staff can move the lens (yellow sphere) over the region of interest in order to inspect the registration result and image fusion. Patient 1 and patient 3 show average registration results. Patient 4 shows the registration with highest mTRE of which is due to immediate tissue deformation during the iUS scan after dura opening. This causes initial deformations that the here presented non-deformable registration approach cannot compensate. Patient 2 shows one of the best alignments of iUS and preMRI.

Inter-observer variability (dice coefficients)		
Patient	Expert 1 compared to Expert2	Expert 1 compared to Expert 3
1	0.93	0.70
2	0.84	0.42
3	0.96	0.53
4	0.82	0.75
5	0.78	0.65
6	0.72	0.44
7	0.83	0.31
8	0.85	0.41
9	0.74	0.49
10	0.72	0.45
11	0.94	0.54
Mean	0.83	0.52

Table 2.9: Inter-Observer variability in comparison to the chosen reference of Expert 1. The Dice coefficients represent the difference of the chosen reference to Expert 2 and 3.

proposed two segmentation methods that are used to support multi-modal image registration of preMRI and iUS images and showed that we can reduce the overall computation time needed for registration from 40.53 s to 18.38 s.

One could argue that the computation time for the segmentation of the scanned iUS images with on average 79 s must be added to the 18.38 s but the slice-wise segmentation can be performed in parallel to the volume reconstruction process of the 2D US images, which is essentially needed to register the scanned iUS volume to the corresponding region within the preMRI volume.

In our case, no additional computation time is needed to perform the segmentations that would prolong the overall registration process. The automatic skull stripping could be accomplished during surgical planning prior to surgery.

The overall methods achieved an average Dice score of 0.82, which leaves some areas (especially in regions adjacent to the eyes and areas nearing the brain stem and spinal cord) within the brain segmented incorrectly.

However, in half of our 11 cases no manual correction was needed because the segmentation errors are not in the same brain region as the iUS scans are expected later during the neurosurgical intervention and these errors could be neglected.

The segmentation of the iUS images revealed another important issue that seriously

influenced our evaluation and the here represented results: The inter-observer variability of the slice-wise segmentations of the perifalcine region. All experts have more than 10 years' experience in image analysis of cerebral structures. In Tab. 2.8 and in Tab. 2.9 the variabilities can be seen between the three different experts. The experts discussed their segmentation strategy before each of them generated reference segmentations in order to find rules to which extent they would segment the structures of the perifalcine region.

Expert 3 tends to segment more volume of the target structures than Expert 1 and 2 which includes regions Expert 1 and 2 were uncertain of belonging to the target structures. A slightly better correlation between the segmentations of Expert 1 and 2 can be observed. It can be seen that the results may differ based on which reference is used. To evaluate our segmentation results we used solely the references from Expert 1 who segmented the target structures, more reproducibly than the other experts, which can be observed by inspecting segmentations of adjacent slices within the overall scanned iUS volume.

Finding merging criteria to generate a ground truth from all three experts was discussed but it would lead to inconsistencies in how the structures would be segmented from slice to slice. This would lead to further conflicts in training the machine learning algorithm how to segment the structures of interest.

According to the presented variabilities between the expert segmentations, the overall segmentation results of the perifalcine region with a Dice coefficient of 0.74 and an average Hausdorff distance of 12.2 mm is a reasonable and acceptable segmentation result.

The segmentation of patient 1 is evaluated as the best segmentation result which is due to the fact that the training slices are drawn from this patient data set by using iUS data sets from during and after surgery in order to train the Random Forest classifier.

With our proposed approach we show a significant benefit by integrating segmentations into the registration process. The solely intensity-based registration decreases the mTRE comparing the initial registration from 16.9 mm to 3.9 mm which is further reduced by the segmentation-based registration to 2.2 mm. The maximum initial mTRE is reduced from 19.4 mm to 5.6 mm by the intensity-based registration and further decreased to 3.0 mm by the segmentation-based registration.

In Fig. 2.16 characteristic registration results of the segmentation-based registration are shown. Nonetheless, the mTREs of the proposed registration may seem high with 2.2 mm on average to the ground truth landmarks. But this is due to the fact that the ground truth registration is deformable whereas the here proposed algorithms for registration used a non-deformable registration approach to initially demonstrate the usefulness of segmentations within the registration process of this use case in glioma surgery.

2.7.4 Conclusion

IUS is an imaging modality of low cost and usually already present in many neurosurgical units, iUS holds great potential in improving intra-operative image guidance and would improve considerably the quality of treatments of a large number of patients, i.e. higher survival rate or improved patient's quality of life.

Being able to calculate the correct brain shift and tissue deformation from iUS and transferring this deformation to preMRI will support additional surgery planning. For instance, the deformation could also be applied to preoperative fMRI or DTI sequences and the shift of structures at risk within the operative field could be displayed on a screen within the operating room during glioma surgery.

The here described approach showed to increase robustness, accuracy, and speed of the registration process. It is demonstrated that the proposed segmentation-based registration can serve as initial registration right after craniotomy and before dura opening. The proposed approach could be used for further refinements with additional affine registrations to adapt and update the preMRI to local deformations within the operational field during the process of glioma resection (see Section 2.9 as detailed complement to this conclusion).

2.8 Approach 5: Neural-Network-Based Automatic Segmentation of Cerebral Ultrasound Images for Improving Image-Guided Neurosurgery

Published as: J.Nitsch, J. Klein, J. H. Moltz, D. Miller, U. Sure, R. Kikinis, H. Meine (2019). *Neural-Network-Based Automatic Segmentation of Cerebral Ultrasound Images for Improving Image-Guided Neurosurgery*. In: Proc of SPIE, 10951, 109511N. [94]

Contributions: Robust, automatic, neural-network-based segmentation method trained on iUS sweeps before, during, and after resection. Comparison with results from the approach in the previous section of this thesis [91] (see Section 2.7). The trained U-net is also applied to a publicly available data set and very good results were attained on even differently acquired B-mode iUS images for glioma surgery.

Available data sets at time of this study: Basically, the same data sets were used from the previously described approach in Section 2.7. Additionally, the number of iUS scans could be increased from 11 to 18 different patient scans for which an expert created reference segmentations of the perifalcine region.

Abstract: Segmentation of anatomical structures in intraoperative ultrasound (iUS) images during image-guided interventions is challenging. Anatomical variances and the uniqueness of each procedure impede robust automatic image analysis. In addition, ultrasound image acquisition itself, especially acquired freehand by multiple physicians, is subject to major variability. In this paper we present a robust and fully automatic neural-network-based segmentation of central structures of the brain on B-mode ultrasound images.

For our study we used iUS data sets from 18 patients, containing sweeps before, during, and after tumor resection, acquired at the University Hospital Essen, Germany. Different machine learning approaches are compared and discussed in order to achieve results of highest quality without overfitting. We evaluate our results on the same data sets as in a previous publication in which the segmentations were used to improve iUS and preoperative MRI registration.

Despite the smaller amount of data compared to other studies, we could efficiently train a U-net model for our purpose. Segmentations for this demanding task were performed with an average Dice coefficient of 0.88 and an average Hausdorff distance of 5.21 mm.

Compared with a prior method for which a Random Forest classifier was trained with handcrafted features, the Dice coefficient could be increased by 0.14 and the Hausdorff distance is reduced by 7 mm.

2.8.1 Neural-Network-Based Segmentation in iUS

In a previous study we demonstrated the potential of central anatomical structures of the brain as reference structures in supporting multi-modal image registration. Being able to approximate the brain shift and tissue deformation from iUS images and transferring this deformation to preMRI will support additional surgery planning.

In this paper, we compare different segmentation approaches for the central cerebral structures falx cerebri and tentorium cerebelli with connected sulci and gyri on B-mode iUS images. Both structures are meningeal folds that separate the left from the right hemisphere (falx cerebri) and the cerebrum from the cerebellum (tentorium cerebelli).

Data and Preprocessing

Before the iUS images are segmented with a U-net-like segmentation approach the iUS images are preprocessed. The seven additional iUS images in comparison to the previous study were acquired with the same iUS acquisition protocol as for all previous studies mentioned in this chapter. Each US volume/sweep contains 100-200 US slices per volume and were stored alongside with their tracking information on the navigation system during glioma resection. The 2D iUS images in each sweep have an average isotropic pixel size of 0.2 mm. The US images normalized within the interval $[0, 1]$ for processing with percentile mapping (P1=2%, P2= 98%).

Furthermore, we used the publicly available MNI BITE [76] database which contains multiple pre- and postresection B-mode ultrasound images from 14 brain tumor patients with either low-grade or high-grade glioma, similar to the acquired images at the neuro-surgical department at Essen. We use the public database in order to assess the general performance of our ultrasound segmentation method. The well-known database is utilized in numerous publications for comparisons in the field of multi-modal image registration of preoperative MRI with intraoperative ultrasound images. The images were acquired at the Montreal Neurological Institute in 2010. A prototype neuronavigation system, IBIS NeuroNav, was employed to record the US images. The ultrasound probe was an HDI 5000 (ATL/Philips, Bothell, WA) P7-4 MHz phased array transducer with an ultrasound depth setting of 6.5-8 cm which causes geometrical differences of the scanned ultrasound

cone when compared to the US images from the Essen University Hospital (see Fig. 2.17 and Fig. 2.18). Each freehand and tracked US sweep consists of 200-600 2D B-mode slices with a pixel size of $0.2 \times 0.2 \text{ mm}^2$. Before we could apply our neural-network-based segmentation, we rescaled and normalized the US images to the same interval as the US images from our study. Additionally, we had to apply a Gaussian smoothing ($\sigma = 1.5$ pixels) in order to increase the similarity of the US images in both studies.

Network Architecture and Training

We describe two approaches that vary in preprocessing. In a first step, in order to achieve a proof of concept in a reasonable time frame, the U-net [124], described in the following, is trained on US images with a downsampled isotropic pixel size of 1 mm. In the second experiment the US images were resampled to an isopixel size of 0.2 mm which is close to the original pixel size of all acquired US images.

For training and validation we used seven patient data sets and eleven data sets for testing. US slices for training were manually selected to guarantee that the segmentation method solely learns from segmentation examples that include adjacent sulci and gyri of the falx and tentorium to the same extent. The main distinction between testing, training and validation data is that the testing data sets are exclusively pre-resection iUS scans, while training and validation data sets include pre-, during-, and after-resection iUS sweeps in order to increase the robustness of the trained models by adding the complexity of differentiating also between the structure of interest (SOI) and visually similar resection boundaries and cavities. The perspective of the SOI varies in each iUS sweep due to miscellaneous tumor locations. The MNI BITE data sets were not included within the training data and were only used for further evaluation. In contrast to other research groups segmenting structures in US images, we preferred to apply the segmentation to 2D iUS images instead of using a 3D reconstruction. By this means, we sidestep reconstruction and other averaging errors which would have a considerable impact on the thin and line-type target structures.

For both aforementioned experiments a 2D U-net model was employed that works on three resolution levels, resulting in a receptive field of 44 pixels. The channel count is 64 on the first resolution level and is doubled (halved) along the downscaling (upscaling) path of the U-net, respectively. The here described neural network uses ReLu activations followed by a batch normalization and is trained using the Dice loss function and the Adam optimizer on a GeForce GTX Titan X.

For the first experiment the training was stopped due to convergence after about 23

RF Classifier)			U-net pixel size 1 mm				U-net pixel size 0.2 mm			
Patient	Dice	Hausdorff	Dice	Dice POST	Hausdorff	Hausdorff POST	Dice	Dice POST	Hausdorff	Hausdorff POST
1	0.89	6.3	0.79	0.84	31.3	8.2	0.76	0.92	36.7	4.1
2	0.69	15.9	0.72	0.74	15.6	10.7	0.70	0.85	23.5	6.9
3	0.67	12.0	0.75	0.76	22.9	9.9	0.73	0.87	31.0	4.6
4	0.88	7.3	0.69	0.88	23.0	6.9	0.75	0.90	27.1	6.0
5	0.86	7.8	0.80	0.89	20.5	9.2	0.78	0.91	30.7	5.3
6	0.72	14.2	0.76	0.83	19.4	7.6	0.73	0.89	26.5	3.3
7	0.65	15.9	0.80	0.84	28.6	9.5	0.74	0.87	36.3	5.1
8	0.63	16.0	0.69	0.72	28.6	9.5	0.65	0.82	37.0	5.5
9	0.65	15.2	0.70	0.73	26.8	8.2	0.68	0.84	33.6	5.1
10	0.73	13.0	0.73	0.77	30.2	9.3	0.70	0.88	38.2	6.4
11	0.77	11.1	0.67	0.81	42.3	6.8	0.63	0.90	48.4	5.0
Mean±SD	0.74±0.09	12.2±3.5	0.73±0.04	0.80±0.06	26.3±7.0	8.71±1.19	0.71±0.04	0.88±0.03	33.53±6.6	5.21±0.97

Table 2.10: The table compares previous segmentation results on 11 patient data sets that were achieved by using an object-based image analysis approach described in Nitsch et al. [94] (see Section 2.7) together with a Random Forest (RF) classifier trained on positive and negative examples of objects and image regions (see approach in Section 2.7). The table compares the results of our first and second U-net experiment using different image resolutions. Hausdorff distances are measured in mm in this table.

hours (batch size 15 slices, learning rate 0.005). We chose a coarser image resolution in a first experiment to validate the applicability of a neural-network-based segmentation while keeping computation times to a reasonable time frame. The second experiment applied to the original pixel size uses the same parameters but converged after 4 days and 18 hours.

Postprocessing

We apply a 0.5 argmax over the foreground class of the 2D U-net followed by a morphological closing with a kernel size of 3×3 pixels, before we extract the largest connected component. This is done because the falx and tentorium with their adjacent sulci and gyri could have very thin connections. Extracting just the largest component would lead to a considerable undersegmentation of these anatomical structures in 2D. Subsequently, a morphological opening with the same parameters is applied in order to prevent an artificial dilation of the segmented structures.

2.8.2 Results

As previously mentioned, we used seven patient data sets for training and validation and eleven data sets for testing. The eleven test data sets were chosen due to the fact that we wanted to compare our current results with a U-net segmentation to a conventional machine learning approach with adequately tailored preprocessing that uses a Random Forest classifier with handcrafted features as described in Nitsch et al. [94] Segmentation results are compared in Tab. 2.10 and visualized in Fig. 2.17.

There are no publicly available segmentations of the falx, tentorium, and adjacent sulci and gyri of the MNI BITE database. Therefore, we demonstrate the general applicability of our segmentation method by illustrating several visual results in Fig. 2.18. For this purpose we selected US slices in which other line-type structures or boundaries are visible within the scanned US cone and in which the target structure consists of multiple branches. As can be seen from the segmentation results in Fig. 2.18, our neural-network-based automatic segmentation is not overfitted on the training data of the University Hospital Essen. In fact, we could demonstrate that the method can be applied on data from another US device with even different US cone geometry.

2.8.3 Discussion

The initial results without postprocessing in Tab. 2.17 of both U-nets are less accurate than the segmentations from the Random Forest classifier. Both U-nets outperform the Random Forest classifier after postprocessing (except for patient data set 1 for the coarse resolution U-net).

The reason for this is that the U-nets learn to detect larger sulci and gyri in general, explaining also why the Hausdorff distances decrease drastically after applying the postprocessing which eliminates segmentations of other sulci and gyri in the vicinity.

Evidently, the Dice coefficients are lower than in other typical segmentation areas which is due to the thin and branching target structures. If one branch is left out, the coefficient drops distinctively. There are challenges of manually segmenting the falx and the tentorium in iUS images to a consistent degree and there are non-negligible intraobserver variances depictable in the training and testing data. Considering the above we would not expect higher Dice coefficients or smaller Hausdorff distances.

Furthermore, we segmented the falx and the tentorium only with their directly adjacent sulci and gyri that have a profound and direct connection to the aforementioned structures. The reason for this is that other or thinner sulci and gyri within the vicinity of the

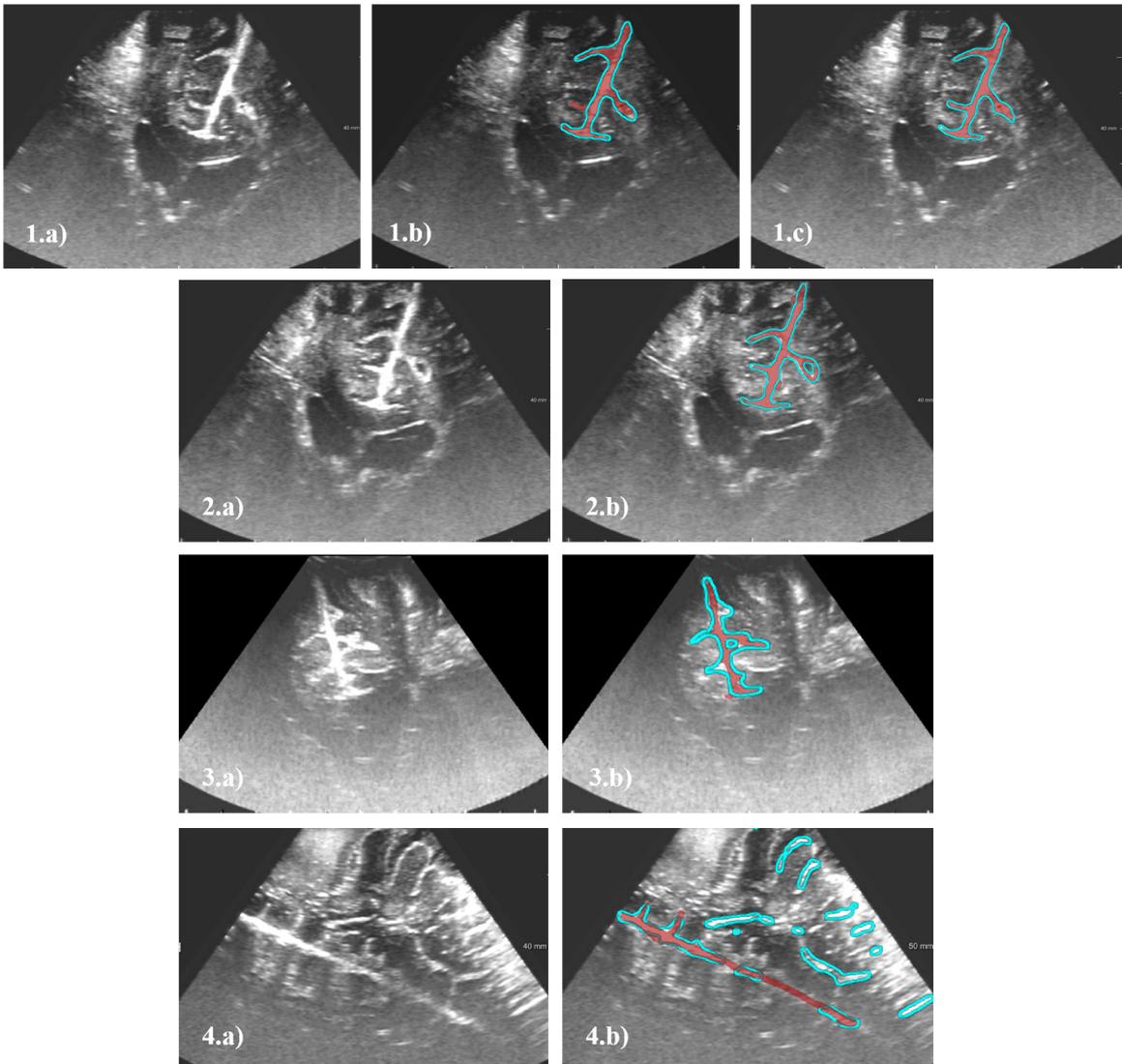


Figure 2.17: The images here show some of the results of the presented segmentation method (blue outline) in comparison with the expert segmentations (red overlay). The images 1.a)-c) show segmentation results from the first U-net experiment (1.b)) next to the result of the second experiment trained on a finer resolution (1.c)). The second experiment shows a more detailed segmentation and spreads also into thinner branches of the target structures. 2.a)-b) as well as 3.a)-b) represent more challenging segmentation tasks with bifurcations and inclusions, mastered by the trained model of the second U-net experiment. In 3.b) it can be even discussed that the algorithm might be more accurate/sensitive to borders than the expert segmentation in this case. 4.b) shows sulci and gyri in the vicinity that would be part of the segmented structures without connected component analysis on the U-net outputs.

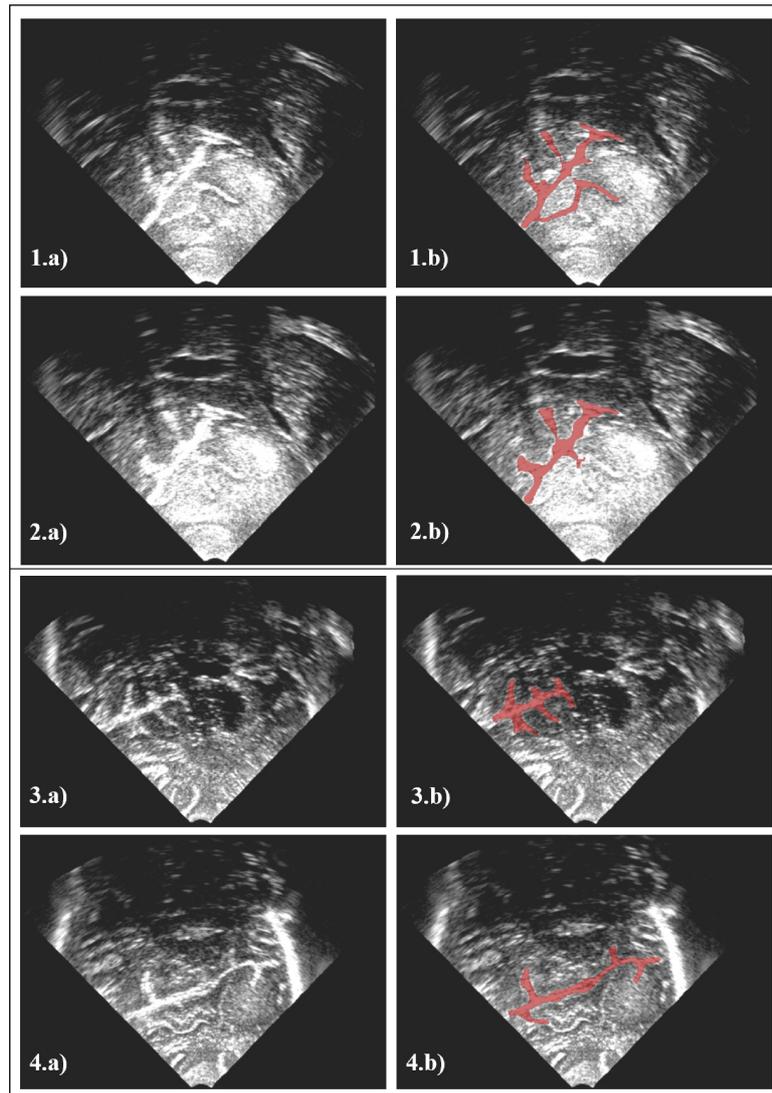


Figure 2.18: We applied our neural-network-based automatic segmentation to data sets from the publicly available MNI BITE database from the Montreal Neurological Institute. The first two pre-resection US sweeps (case 1 and 2) were used for comparison. Images a) always show the original image whereas the b) images show the original images with the overlaid segmentation result of our algorithm in red. The images 1.a) and b) are from case 1 slice 1 within the pre-resection data set. 2.a) and b) show slice 20 within the aforementioned patient data set. 3.a) and b) show a segmentation result of case 2 slice 100. Within the same data set 4.a) and 4.b) show the segmentation of slice 190. To demonstrate the applicability of our segmentation approach in a broader scope we selected challenging cases containing other line-type structures or boundaries within the scanned US slice and in which the target structures consist of multiple branches.

tumor underlie considerable changes due to brain shift and tissue removal during tumor surgery while the meningeal folds and close sulci and gyri in the perifalcine region stay predominantly in place.

Nevertheless, in some cases as illustrated in Fig. 2.17 4.b), the target structure might actually consist, or just be depicted, as multiple structures on an US slice. For future work we aim to combine the neural-network-based segmentation – due to the promising results of the trained classifier – with a Random Forest classifier as additional postprocessing to filter possible target structures after the U-net segmentation.

2.8.4 Conclusion

Our work demonstrates that automatic and robust segmentation of iUS is possible by using a U-net-like architecture, even on smaller data sets without overfitting by choosing different resection stages in training, validation, and test data sets.

Furthermore, we carefully selected the training examples instead of randomly splitting training, testing, and validation sets. This can motivate research for similar anatomical structures with lower variances in size and shape or in the segmentation of surgical devices in iUS images.

Generally, it is assumed that a U-net needs a lot more data before reasonable results can be expected, however, the here presented neural-network-based segmentation already outperformed our previously trained Random Forest classifier.

Nowadays, this is an important observation within an era of constantly growing data bases and an often neglected subject that even a task with a relatively small amount of data might be eligible for a deep learning approach. For the best of our knowledge, this observation has not yet been made in the here presented context (see Section 2.9 that complements this conclusion).

2.9 Conclusion and Outlook Part I

Overall the greatest challenge in developing the different approaches in Part I was the prospective acquisition and inclusion of patients. Patients had to be diagnosed with a low- or high-grade glioma in order to fit into the study. In addition, since we wanted to demonstrate the basic concept of better registration quality by using segmented anatomical structures as guiding frame, patients with larger or multilocular tumors had to be excluded from the study.

This had to be done because the scanned iUS volume would likely contain only the tumor region and no other structures would be visible during surgery that could support the registration. However, not many patients had to be excluded based on this criterion. Moreover, patients in which ultrasound image acquisition was challenging due to patient positioning on the table in the operating room or due to difficult accessibility of the tumor with the US probe had to be excluded. And most importantly, the patients had to agree to be included in the study. Furthermore, the annual case numbers of glioma resections performed can only be estimated. In reality these were significantly smaller than estimated before the start of the joint project, especially, after applying the aforementioned exclusion criteria.

The here described approaches in Part I represent a proof of concept of using central cerebral structures as reference structures for preMRI and iUS image registration. In the introduction of this thesis in Section 1.1.1, the requirements for a non-deformable segmentation-based registration were summed up as objectives of Part I. According to this, the following achievements were demonstrated with the implemented registration approach in this thesis:

- Fast and accurate segmentations of the peritumoral region as central cerebral structures could be computed automatically. A Dice coefficient of 0.88 and a Hausdorff distance of 5.21 mm could be attained. For segmenting 200 US slices less than 79 seconds were needed.
- The slice-wise segmentations could be computed in parallel with the reconstruction of 3D iUS volumes.
- It could be demonstrated that the segmentation-based registration performed faster, more robust, and more accurate than a conventional intensity-based registration by reducing mTRE from 3.8 mm to 2.2 mm. The segmentation-based registration and the intensity-based registration were both more accurate than the initial, rigid

registration by the neuronavigation system (mTRE 16,9 mm).

- The segmentations, the volume reconstruction, and the subsequent multi-modal registration took on average 91 seconds per scanned ultrasound volume (79 seconds for segmentation plus 12 seconds for registration, see Section 2.7). This fulfilled the time constraints of three minutes for this whole procedure and could still be considered a near real-time image fusion.
- All the previously described computational steps could be managed with one additional interaction of starting the segmentation-based registration process. This fulfilled the requirement to keep the user interactions for the registration to a minimum.

However, the here identified central cerebral structures for improving MRI-US registration are just examples and further segmentations of additional structures as guiding frame would lead to more independence of visible structures of the here proposed approach. Nevertheless, the acquired iUS images in this study were specifically recorded to capture the perifalcine region. Thus, for extending this approach to other structures more iUS volumes must be acquired to learn the characteristics of further anatomical structures.

In addition, by adding more structures it may also be possible to increase precision in estimating the non-uniform, three-dimensional tissue deformations that occur during tumor resection.

Additional structures could be blood vessels, such as smaller arteries or veins in the operational field. For this, an extension of the overall imaging protocol would be required for both preoperatively acquired MRI scans as well as for the ultrasound image acquisition. T1-weighted preMRI should be acquired pre- and post-contrast injection and an MR-Angiography should be considered as further MRI sequence. A pre- and post-contrast MRI would be beneficial in segmenting thicker blood vessels of the brain. MR-Angiography (MRA) would help in distinguishing smaller cerebral veins and arteries. Furthermore, Doppler ultrasound images must be acquired during surgery to get a corresponding representation of the vasculature. Besides, specific cerebral blood vessels could be detected, such as the *circulus willisii* and used for registration with the MRA image. This idea of using MRA images together with Doppler ultrasound for neurosurgical procedures in general, was initially proposed by Reinertsen et al. [114] in a small clinical validation study of a vessel-based registration technique with five patients.

Moreover, the integration of further dominant fissures into the registration process could be advantageous to increase flexibility of the registration method. This could be, for

instance, the *Sylvian fissure* as an example for an explicit sulcus or any other sulci from the cerebral cortex potentially visible during glioma surgery within iUS. To implement this, the FLAIR MRI sequence would be beneficial in which gray and white matter of the brain can be better distinguished as on contrast-enhanced T1-weighted MRI. From these images the sulci and gyri of the cerebral cortex could be segmented with a similar method as described in Section 2.6 by using a Gaussian mixture model.

The U-net proposed in Section 2.8 could be trained with these additional sulci and gyri from iUS images. By using the sulci and gyri segmentations from preMRI as well as the information about tumor location, a priori knowledge could be integrated to the registration algorithm. This could incorporate knowledge about visible anatomical structures and predict their location and would also make it possible to build a preoperative, patient specific atlas of segmented structures in the brain region that is potentially scanned during glioma surgery.

This a priori knowledge could support finding an optimal initialization for the registration algorithm and, potentially, further reduce the overall computation time for multi-modal image registration. This would also enable a global initialization of the registration algorithm without relying on an initial registration by the neuronavigation system, similar to what is proposed by Rackerseder et al. in [113].

Furthermore, good results are reported by Machado et al. in [69] for iUS feature extraction and subsequent uni-modal iUS-iUS feature-based registration during image-guided glioma surgery. In Machado et al. the 3D SIFT-Rank algorithm [146] is used to extract and match features from 3D iUS. The above proposed extension of providing a patient specific atlas of segmentations in both preMRI and iUS would extend this idea by Machado et al. to a multi-modal approach. However, instead of taking any strong hyperechogenic structures into account in iUS for registration, the new idea is here to solely rely on previously selected guiding-frame structures in iUS and preMRI.

With this, increased control could be obtained of what kind of anatomical structures should be used for image fusion/registration and the multi-modal registration could overcome different representation of structures and boundaries in both modalities. An example would be dominant resection boundaries and cavities in iUS that are not visible in preMRI. At the same time, this increases robustness to (hyperechogenic) artifacts in iUS images. This is the outstanding advantage that could be demonstrated with the segmentation-based registration approach of this thesis.

The research and proof of concept of Part I of this thesis - of a novel, near real-time, robust, and accurate non-deformable segmentation-based registration - has paved the

way for future research in this direction of incorporating specific anatomical structures as guiding-frame for multi-modal deformable image registration for image-guided glioma surgery.

In summary, it was an outstanding opportunity and experience to research with an interdisciplinary team on an utmost application-oriented project in very close cooperation with clinical partners. I was present at the surgeries myself and actively acquired the data sets on a laptop. The benefit of intraoperative image fusion became very clear during multiple visits and ultrasound imaging had an especially clear influence on the gross-total tumor resection. IUS represented an alternative modality for distinguishing the tumor boundaries and residual tumor in contrast to the visibility in preMRI or in contrast to visibility under the microscope with and without fluorescence assistance.

3 PART II: Clinical Decision Support for End-Stage Liver Disease

The research in Part II of this thesis was part of a one year research visit at the Surgical Planning Laboratory (SPL), Brigham and Women’s Hospital, Harvard Medical School, in Boston, USA. Together with our partners at the SPL and at the division of gastroenterology, hepatology, and endoscopy we researched in the field of MRI-based radiomic features for an automatic severity assessment of cirrhotic end-stage liver disease patients. The motivation and background will be outlined in detail in the following sections of this chapter.

This research can be situated in the area of clinical decision support by providing an image-based alternative or addition to a score based on a blood test that is currently used to risk-stratify cirrhosis patients [55]. As mentioned in the introductory chapter of this thesis, this is the so-called MELD (Model for End-Stage Liver Disease) score that is also used for transplant organ allocation in Europe and the USA [28, 97]. The overall goal is to improve a fair and objective prediction of the severity of liver cirrhosis for end-stage liver disease patients as explained in the introduction of Part II in Section 1.1.2.

Part II basically consists of two publications. A journal paper is already submitted (Section 3.3) and the other is in preparation (Section 3.4):

- *MRI-based Radiomic Feature Analysis of End-Stage Liver Disease for Severity Stratification*
- *Prognostic Value of MRI-based Radiomic Features in End-Stage Liver Disease*

Further experiments are described in Section 3.5 and are closely linked to the research and to the results of the two above mentioned sections. It is planned to publish the results of Section 3.5 in a clinical context.

The structure of Part II of this thesis is as follows: First, a combined, general introduction into the topic is given together with related work in Section 3.1. Afterwards, the material and methods section in 3.2 describes the utilized patient data within the research project

of Part II. Subsequently, the three different studies/planned publications are described separately. Each study consists of a short abstract, description of performed experiments, attained results, and a conclusive discussion section.

Contributions: A novel image-based severity assessment for liver cirrhosis and an automatic cirrhosis detection on contrast-enhanced MRI. To date, scientific research in this area has only concentrated on improving the MELD scoring system with other clinical parameters but without using available image data acquired for these patients, for instance, images from the HCC screening (see Section 1.2.1). In this work, radiomic features were analyzed in detail within the liver and spleen and various experiments were performed in order to predict the severity of liver cirrhosis. An additional objective was to determine the specific radiomic features that can serve as a radiomic biomarker to identify and risk-stratify end-stage liver disease patients.

Therefore, our goal was not just to collectively capture specific, progressive or even latent changes of liver cirrhosis in MR images, but also to understand the individual, underlying radiomic features that are important to describe these characteristics. In this way, it is also advantageous that the results of our initial studies allow detailed examination for plausibility and comparisons with clinically known manifestations of progressive cirrhosis (see Section 1.3.3). Furthermore, this supported to evaluate the feasibility of describing the disease with radiomic features accurate enough and to assess the significance of our results. Consequently, we chose to use a classic machine learning approach together with an established open-source library with descriptive radiomic features (PyRadiomics [150]) for a proof of concept of our initial and overall novel research in the field of providing image-based clinical decision support for the severity assessment of end-stage liver disease patients. This is why we decided against using a deep learning-based method as a first step for our research in order to verify our hypothesis to be able to risk-stratify cirrhotic patients based on radiomics in MR imaging.

Another uniqueness of our research is that we have a control patient cohort with healthy livers and spleens scanned with the exact same MRI protocol and MRI scanners (same manufacturer model). To the best of our knowledge we are the first research group in this field of research.

As stated above, this research was conducted together with our partners at the Brigham and Women's Hospital (BWH) located at the SPL and at the hepatology clinic. Valuable feedback was provided by all contributors to successfully accomplish the aims of our research.

I was responsible for the data collection through the Partner's HealthCare Hospital System in Boston (see Section 1.6) and I provided the initial list and data base of suitable patients and corresponding images for further chart review through our clinical partners. Patient exclusion criteria were managed by our clinical partners to obtain the final patient cohorts for our studies. Initial manual segmentations and feedback to obtain optimal, automatic liver and spleen segmentations were provided by a radiological expert from Fraunhofer MEVIS.

Furthermore, my task in this project was to contribute the design and implementation of the concept for automatic neural-network-based segmentations, accurate quantitative imaging feature extraction and its parametrization, machine learning analysis adapted to the available data, and evaluation and interpretation of resulting radiomic features.

Sections 3.1, 3.2, and 3.3 are in parts based on a manuscript that is already submitted to a journal. All co-authors mentioned in Section 3.3 have read and reviewed the manuscript.

3.1 Introduction

Nearly two million people worldwide die from complications of cirrhosis each year, making the disease the 11th most common cause of death globally [83, 4]. Cirrhosis is characterized by bridging fibrosis and regenerative nodules that disrupt the normal liver architecture. Cirrhosis is the final histologic pathway for chronic liver diseases caused by alcohol, viral hepatitis, non-alcoholic fatty liver disease, autoimmune disease, and metabolic disorders. The multiple etiologies of cirrhosis differ in their prevalence which can be summarized as follows: Hepatitis C (25-40 %), alcoholism (25-35 %), hepatitis B (15 %), nonalcoholic fatty liver disease (10 %), and autoimmune hepatitis (5 %) [125]. Cirrhosis is prognostically subdivided into compensated and decompensated cirrhosis, with the latter characterized by higher mortality and defined as the occurrence of at least one episode of variceal bleeding, ascites, or hepatic encephalopathy. The only current curative treatment for advanced cirrhosis is liver transplantation, which is limited by the small pool of available donor organs.

The Model for End-Stage Liver Disease (MELD) scoring system aims to stratify potential liver transplant recipients by estimating 90-day mortality. The MELD was developed by the Organ Procurement and Transplantation Network (OPTN) and the United Network for Organ Sharing (UNOS) and was implemented in 2002 for prioritizing organ allocation. The score is a formula based on the patient's serum creatinine (Cr), total serum bilirubin (TBIL), and Internal Normalized Ratio of prothrombin time (INR):

$$MELD_{(i)} = 9.57 \cdot \ln(Cr \text{ mg/dl}) + 3.78 \cdot \ln(TBIL \text{ mg/dl}) + 11.2 \cdot \ln(INR) + 6.43 \quad (3.1)$$

The formulation of MELD was modified by UNOS in January 2016 to include serum sodium (Na) [56, 97]:

$$MELD \text{ Score} = MELD_{(i)} + 1.32 \cdot (137 - Na) - 0.033 \cdot MELD_{(i)} \cdot (137 - Na). \quad (3.2)$$

MELD scores are rounded to the nearest integer and range from 6 to 40, with 6 being normal and 40 correlated with imminent death. UNOS organ allocation also provides mechanisms to expedite transplant for conditions such as HCC in the form of exception points that are added to a patient's MELD score. These exception points are handled on a case-by-case basis. For this reason, we do not take possible MELD exception points into account. A MELD score of 15 has been shown to be the inflection between the relative risk and benefit of transplant [144, 73].

Cirrhosis leads to successive morphological and textural tissue changes to the liver and surrounding vessels and organs. Evident characteristics of cirrhosis can include: portal

hypertension, liver surface nodularity, heterogeneous enhancement of the liver, varices, ascites, expanded gallbladder fossa, splenomegaly, and sarcopenia [125, 167].

Apart from these directly visible features, radiomic feature analysis - *radiomics* - has recently shown promising results in exploiting latent information in medical images. Radiomics has been used to identify biomarkers through quantitative image-based feature extraction and analysis [101, 100, 169]. Applications include correlating derived features with patient outcomes, such as survival and response to chemotherapy and radiation [168, 16]. Furthermore, radiomic parameters related to characteristic texture and morphological heterogeneity have shown the potential to yield excellent, non-invasive prognostic factors for patient outcome. Examples include tumor phenotype analysis for risk stratification of prostate cancer as well as lung lesion characterization and predicting treatment response [152, 150, 2, 161]. Radiomic features include different feature classes of quantitative features that capture different properties of a region of interest (ROI): first-order features (e.g. intensity features), second-order statistical features (e.g. texture features), and also morphological features to analyze shape ROI statistics.

MRI-based radiomic feature extraction also comes with additional challenges due to lack of signal normalization, MRI sequence standardization, and more common acquisition artifacts [131, 164]. This makes feature repeatability (also termed stability or reproducibility in this context) between different scanner types, models, or even different software versions on the same model, a research field on its own. As a consequence, MRI-based radiomic features can identify different MRI manufacturer models that are using the exact same acquisition protocol, which would confound a study such as ours focusing on disease.

For these reasons, we decided to focus our initial study exclusively on MRI data from a single centre, using the same manufacturer model, scanner software version, magnetic field strength, and the exact same MRI acquisition protocol. These restrictions allowed us to maximize control while assessing the value of radiomic features.

Related research has used contrast-enhanced T1-weighted MR images to automatically assess the stage of liver fibrosis. Yasaka et al. trained a deep convolutional neural network to learn characteristic image-based liver fibrosis features from contrast-enhanced T1-weighted MR images from 534 patient data sets, classifying fibrosis into the stages F0, F1, F2, F3, and F4, where a stage of F4 represents liver cirrhosis [163]. This is similar to the research by Choi et al. where liver fibrosis staging was performed on contrast-enhanced CT images [19]. Other researchers have attempted to improve the prognostic value of the MELD scoring system by evaluating a broader set of lab parameters, such as the MELD-Plus score, but do not include image-derived metrics [54].

3.2 Patient Selection and Image Data

Patient Data and Selection

This retrospective study of MRI scans and corresponding medical records was approved by the Institutional Review Board (IRB) of Partners HealthCare. The cirrhosis cohort and the control cohort are validated through chart review. In Tab. 3.1 we summarized the cohorts demographic information. MRI scans of both cohorts were acquired from June 1, 2015 to June 1, 2018 at the Brigham and Women’s Hospital (see Section 1.6 for more details on the data acquisition process).

Cirrhosis Cohort

MRI scans were acquired for hepatocellular carcinoma (HCC) surveillance in patients with cirrhosis. 417 patients with ICD10 codes of cirrhosis were initially identified using the Partners HealthCare Research Patient Data Registry (RPDR) that gathers clinical data from the Partners Hospital System [86]. Identification of patients with RPDR is further outlined in Section 1.6). Within this cohort we searched for patients that were scanned using a multiparametric, fat-suppressed T1-weighted MRI scanning series on a 3 Tesla scanner - a standard protocol used for HCC screening - including a five minutes delayed scan post contrast injection (Gadavist[®], Bayer HealthCare AG, Medical Care, NJ, USA; in Europe also known as Gadovist[®], Bayer AG, Germany). From this series, the five minutes postcontrast scan is used for radiomic feature extraction, representing a scan and contrast uptake phase where cirrhotic regions within the liver are enhanced (see Sections 1.2.1 and 1.3.3 for more details). This is a result from architectural cirrhotic tissue transformations that considerably impact perfusion within the liver parenchyma and delay the washout characteristically. In total, 191 MRI scans were acquired with the standardized protocol. Any scans were excluded that met the following criteria: if scans were not done on a

Cirrhosis Cohort (n=90)		
Age (years)	Mean \pm Standard Deviation	61 \pm 12
Sex	Male/Female	46/44
Control Cohort (n=77)		
Age (years)	Mean \pm Standard Deviation	63 \pm 11
Sex	Male/Female	14/63

Table 3.1: Cirrhosis and control cohort’s demographic information

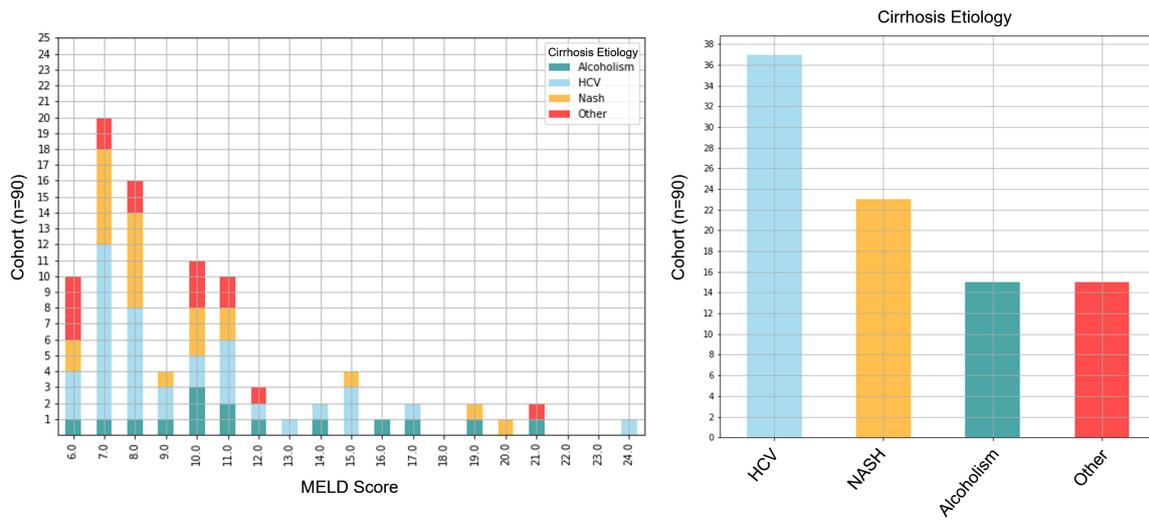


Figure 3.1: Distribution of MELD scores in liver cirrhosis cohort (n=90) (left) as well as an overview of different cirrhosis etiologies (right).

Siemens Verio MRI Scanner (Siemens Magnetom Verio, Siemens Medical Solutions, PA, USA) (n=31), if lab parameters were missing for MELD score calculation (n=9), if patients had a history of hepatic ablation (n=14), if cirrhosis could not be confirmed (n=10), if patients had a hepatic resection (n=1), if patients had a splenectomy (n=1). The final cohort consists of 90 different patients with 125 MRI scans. If a patient had multiple scans only the latest scan was used for feature analysis in order to prevent bias in our radiomic feature extraction and machine learning analysis. All included patients obtained their MELD labs on average within a period of ± 22 days from their MRI scan. The patient cohort was validated through chart review. In Tab. 3.1 we summarized the cohort's demographic information. In Fig. 3.1 we give an overview of cirrhosis etiologies in our patient cohort.

Control Cohort

The control cohort (or IPMN cohort in the following) are patients with a benign pancreas tumor - intraductal papillary mucinous neoplasia (IPMN) - that has generally no impact on neighboring anatomical structures within the abdomen [33]. The control cohort was also acquired using Partners RPDR for retrospective study data selection and acquisition. Patients were identified if they had ICD10 codes of pancreatic cysts/IPMN but with ICD10 codes of cirrhosis or HCC as basic exclusion criteria. Initially, 650 patients were selected. For a comparable study of radiomic features, MRI scans were selected matching the same MRI protocol, MRI scanner type, and contrast agent as described above for the cirrhotic

cohort. To our advantage, the exact same multiparametric MRI protocol could be identified through hospital-wide standardized protocol names. With these protocol and acquisition prerequisites 88 patients with corresponding MRI scans could be identified. Patients were excluded from the control cohort if they had multiple cysts larger than 10 mm within the liver parenchyma (n=3) (multiple larger cysts might suppress or corrupt the extracted healthy liver tissue features), if any technical problems occurred during scanning (n=4), if they had a splenectomy (n=1), and if they had diverse comorbidities at the same time (n=3) potentially impacting a normal liver or spleen anatomy. Collectively, 77 IPMN patients matched the previously mentioned criteria. Tab. 3.1 also summarizes the demographic information of the control cohort. Important note: The control cohort was not available (neither checked for exclusions, nor segmented and preprocessed) at the time when we conducted our initial study in Section 3.3.

Image Analysis

For quantitative radiomic feature extraction, we automatically segmented livers and spleens in our cohort using a U-net-like [124] architecture similar to Chlebus et al. [18]. The original image resolution of the fat-suppressed T1-weighted MR images acquired five minutes after contrast injection is $0.59 \text{ mm} \pm 0.05 \text{ mm}$ ranging from 0.5-0.86 mm with 3 mm in z-dimension and within the control group $0.56 \text{ mm} \pm 0.04 \text{ mm}$ ranging from 0.47-0.76 mm with also 3 mm in z-dimension. All images were resampled to 0.5 mm in x- and y-dimension. As preprocessing before segmentation a non-uniformity intensity correction was applied as described by Sled et al. [139] followed by a normalization to the interval $[0; 1]$. We started with 20 expert segmentations for training two individual neural networks for liver and spleen segmentation. Erroneous liver and spleen masks were successively corrected and the network was retrained. An expert with more than 10 years experience in abdominal radiology validated and corrected segmented contours if necessary. Feature extraction was performed using the PyRadiomics library (version 2.0.1) in Python. For our experiments we initially extracted features from liver and spleen segmentations using all available feature classes in the respective version for further analysis: first-order statistic features, shape-based 3D features, gray level co-occurrence matrix (GLCM) features, gray level size zone matrix (GLSZM) features, gray level run length matrix (GLRLM) features, neighboring gray tone difference matrix (NGTDM) features, and gray level dependence matrix (GLDM) features. Furthermore, we used LoG filters with sigma 1, 2, 3, 4, 5 mm. We also added the *liver-to-spleen volume ratio* as additional feature. In total 2577 radiomic features were extracted, 1288 each for liver and spleen.

3.3 MRI-based Radiomic Feature Analysis of End-Stage Liver Disease for Severity Stratification

Planned publication as: J.Nitsch, J. Sack, M.W. Halle, J.H. Moltz, A. Wall, A.E. Rutherford, R. Kikinis, H. Meine. *MRI-based Radiomic Feature Analysis of End-Stage Liver Disease for Severity Stratification*. (Submitted).

Abstract: We have developed a predictive model of disease severity for cirrhosis using MRI-derived radiomic features that correlates well with established measures of disease severity such as MELD score and presence of hepatic decompensation. In this retrospective study, the patient cohort (n=90) underwent a contrast-enhanced MR screening protocol for hepatocellular carcinoma (HCC). We use radiomic feature analysis to train four prediction models for assessing the patient's condition at time of scan: the patient's MELD score, if the patient's MELD score is ≥ 9 (median score of the cohort), if the patient's MELD ≥ 15 (the inflection between the risk and benefit of transplant), and if the patient is decompensated. Liver and spleen segmentations were used for feature extraction, followed by a random forest classifier. Our method was most predictive of the presence of decompensation. In the decompensation and median MELD experiments, a combination of liver and spleen features was found to have the best predictive value (AUC of 0.84 and 0.79, respectively), compared to liver features only (AUC of 0.80 and 0.71) and spleen features only (AUC of 0.80 and 0.77). All results are computed using cross-validation.

Introduction: The introduction here is a continuation of the general introduction in Section 3.1. This retrospective study aimed to determine if radiomic features derived from MRI scans of a cirrhotic patient cohort can predict the patients' disease severity as approximated by MELD score and presence of decompensation.

Furthermore, by focusing on severity assessment of end-stage liver disease we try to predict whether a patient already decompensated by applying the same extracted radiomic features. Compared to previous related work focused on fibrosis staging (see Section 3.1), we make no *a priori* assumptions about specific manifestations of disease in imaging beyond generally detecting them in the liver and spleen. Rather, we rely on objective image-derived radiomic features with established surrogates for liver disease severity.

For our radiomic feature analysis we focused our feature extraction on liver and spleen-derived features from contrast-enhanced T1-weighted MR images on 90 patients with liver cirrhosis (see Sections 1.3.3 and 3.2). As mentioned in Section 1.3.3 of this thesis, cirrhosis

is not only known for impacting the liver tissue but also leads to splenomegaly.

3.3.1 Material and Methods

The cirrhotic patient cohort that was used for this study as well as the image analysis and preprocessing steps were previously described in Section 3.2.

Machine Learning Analysis

We performed four different experiments using the resulting radiomic features to predict different surrogates of disease severity:

- Experiment 1 seeks to determine a direct, MELD-score-specific prediction model of MRI-derived features with our data set (meaning: radiomic features specific for every MELD score).
- Experiment 2 attempts to predict whether a patient has a MELD score above or below the cohort median. The cohort median MELD score is 8, resulting in an almost even split at MELD score ≥ 9 (46 patients with a lower MELD score and 44 with a higher MELD score). A MELD score of 9 has a clinical relevance as well, since a score of 10 has been suggested as a threshold at which transfer of care to a hepatologist should be considered [144, 85].
- Experiment 3 is similar to experiment 2, but attempts to predict whether a patient's MELD score is 15 or above (where 15 represents a value where the mortality risk of transplant and cirrhosis are approximately equal). Since our patient cohort includes more patients with less advanced cirrhosis, the population of the two classes is 77 patients with a MELD < 15 and 13 patients with a MELD ≥ 15 .
- Experiment 4 uses the same radiomic feature analysis, but instead uses liver decompensation (as determined by chart review) rather than MELD score as a surrogate for disease severity. Decompensation events (presence of ascites, variceal bleeding or hepatic encephalopathy) pose severe mortality risks and impact to patient quality of life. Decompensation can be directly assessed by review of the patient's clinical record, even in the absence of laboratory tests. The patient cohort consists of 62 compensated and 28 decompensated patients. Fig. 3.2 shows the number of compensated and decompensated patients in the cohort for each MELD score value.

We divided our cohort into stratified subsets and used cross-validation to define training and test data sets for our experiments. Additionally, we used the *fast correlation-based*

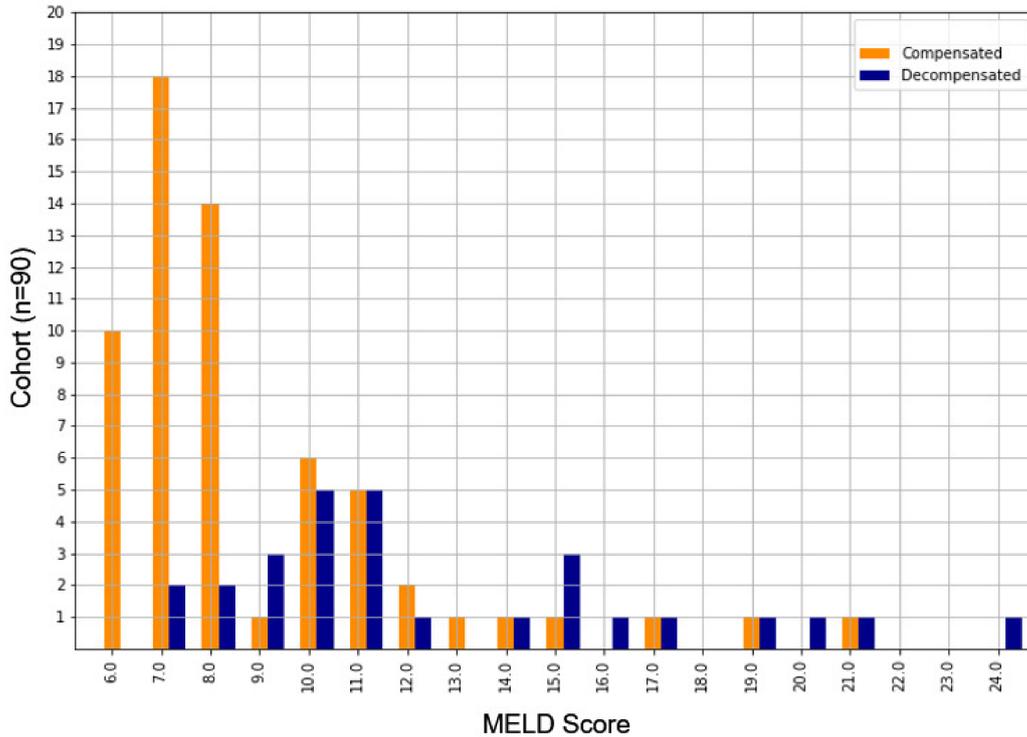


Figure 3.2: Distribution of MELD scores with an overview of compensated (n=62) and decompensated (n=28) patients.

filter, FCBF, for feature selection [166]. This filter allows to identify features with minimal redundancy and maximized relevancy due to pairwise analysis of correlations between features. The resulting features have the greatest prognostic power. Moreover, we used the random forest regressor or the random forest classifier for our regression and classification experiments, respectively. In related work random forests have shown to be a powerful classification method for machine learning analysis of radiomic features [101].

3.3.2 Results

In the following section, we describe the results of our experiments using radiomic feature analysis to predict different measures of cirrhosis severity. For each experiment, we used stratified k-fold cross-validation with $k = 5$ folds to split our cohort into training and test data sets. We repeated the cross-validation process 100 times to attain different stratified training and test groups in order to get representative average classification results. Fig. 3.3 gives an overview of the general feature extraction and classification approach. In Tab. 3.2 all results are summarized together with the ROC curves of the experiments in Fig. 3.4.

Experiment	Used radiomic features	AUC	P-value
2	Split at median MELD score		
	Liver and spleen features	0.79	< 0.01
	Liver features	0.71	< 0.01
	Spleen features	0.77	< 0.01
3	Split at MELD score ≥ 15		
	Liver and spleen features	0.68	< 0.02
	Liver features	0.71	< 0.01
	Spleen features	0.59	≈ 0.2
4	Decompensation		
	Liver and spleen features	0.84	< 0.01
	Liver features	0.80	< 0.01
	Spleen features	0.80	< 0.01
	MELD Score	0.79	< 0.001

Table 3.2: Classification results: Used radiomic features with respective area under ROC curve (AUC) and p-values for each classification task.

Experiment 1: Direct prediction of each MELD score with extracted radiomic features

For this experiment we tested different experimental settings and approaches. As the MELD score represents integer values within the interval $[6; 40]$ we employed the random forest regressor. We carried out different experiments by trying to detect a correlation with just liver-derived radiomic features, spleen-derived features and trying to correlate with the ensemble of both organ features. But even after reducing the feature space by applying the FCBF feature selection method on the training data and selecting the most important features (feature selection performed on liver features, spleen features, and liver and spleen features together) we could not verify a direct correlation of radiomic features in our remaining test data sets with specific MELD scores with an $R^2 = -0.0044$.

Experiment 2: Dividing cohort into two classes at median MELD score

Based on the observations in the previous experiment we modified the experiment by splitting the data into two classes at the median MELD score in our patient cohort with a median MELD score of 8 and a split at MELD score ≥ 9 . This experiment consequently

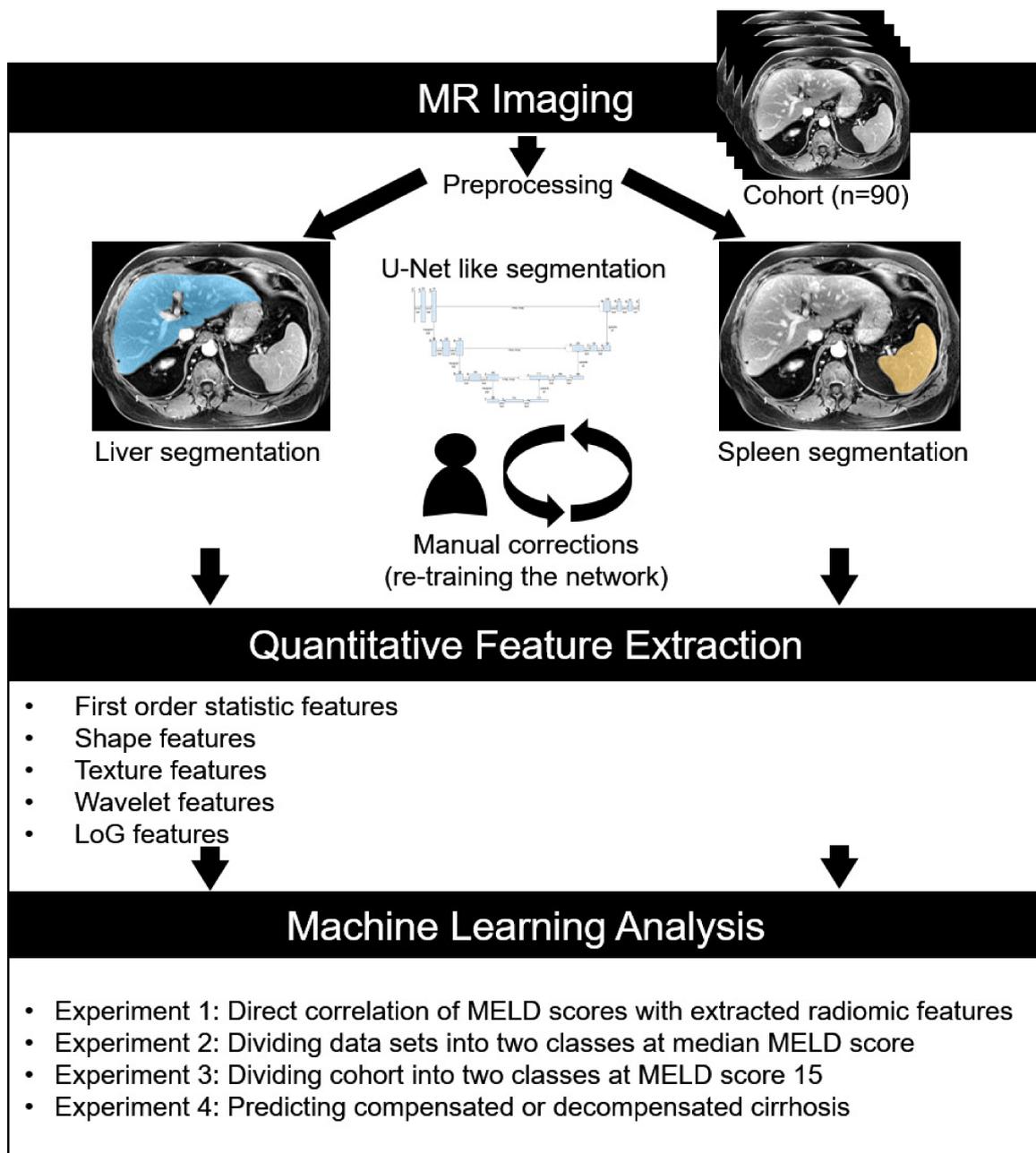


Figure 3.3: General Overview of the feature extraction process and performed experiments.

transferred a regression problem from the previous experiment to a classification problem and the random forest classifier was applied. The goal was to reduce the effects of class imbalances in our relatively small cohort (n=90) by defining a reasonable threshold to split our cohort into two categories: A lower and a higher cirrhosis disease stage. The importance of liver and spleen features was also evaluated for this classification task in

context of all conducted experiments. With a combination of liver and spleen features we achieved an AUC of 0.79, for liver features alone an AUC of 0.71, and for spleen features an AUC of 0.77. All results of the second experiment had a p-value of $p < 0.01$. All p-values in this study were calculated using a random permutation test to evaluate the significance of a cross-validated score. Table 3.2 gives an overview of the classification results of this experiment.

Experiment 3: Dividing cohort into two classes at MELD score 15

We used the same experimental setup as in the second experiment and also utilized a random forest classifier to train our prediction models. Likewise, we evaluated a radiomic feature correlation of liver-, spleen-, and liver- and spleen- derived features with the MELD scores under or equal or above score 15. We also used the same cross-validation strategy as in the previous experiment and an AUC of 0.69 with $p < 0.01$ could be attained for liver and spleen features, an AUC of 0.71 with $p < 0.01$ by using only liver features, and an AUC of 0.59 with $p \approx 0.2$ for using solely spleen features for this experiment (see Tab. 3.2).

Experiment 4: Predicting compensated or decompensated cirrhosis

The fourth experiment targets the status of liver decompensation as determined by a clinical hepatologist based on review of the electronic patient record. Utilizing both liver and spleen features for this classification task resulted in an AUC of 0.84, only using liver features led to an AUC of 0.80, and only using spleen features induced an AUC of 0.80. All results of the fourth experiment had a p-value of $p < 0.01$ as can be seen in Tab. 3.2. In comparison, MELD score of the same cohort has an AUC of 0.79 for predicting the status of liver decompensation (see Tab. 3.4).

Feature Importance

As aforementioned, the FCBF method was used for feature selection. Based on our cohort we cannot determine a generalized set of important radiomic features that is salient for every or even most of the training and test splits within the cross-validation process. We observe that a combination of liver and spleen features is important to obtain an optimal prediction result.

Furthermore, it is always a different mixture of LoG, texture-, and wavelet-based features from the feature classes GLCM, GLDM, GLSZM, and GLRLM that are ranked among the nine to 13 most important features. In addition, most of the important LoG features were computed with a sigma of 3 or 4 mm.

Shape features or the *liver-to-spleen ratio* were not ranked among the most important features. After feature selection we observed in our experiments that only nine to thirteen features seem to be important for the classification process and either liver and spleen features can represent the most important feature for classification.

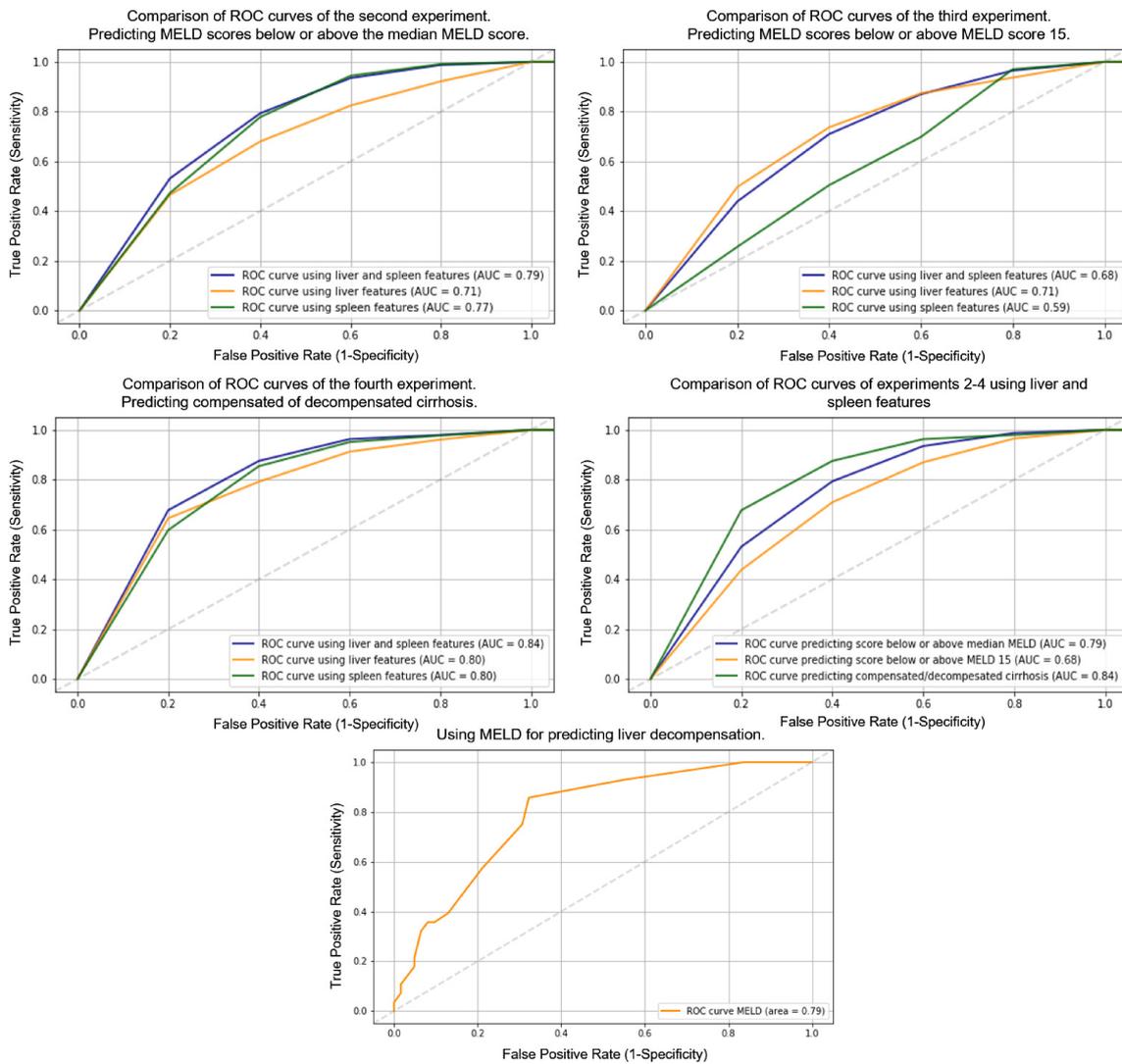


Figure 3.4: Overview of ROC curves of experiment 2 - 4, also comparing using liver and spleen radiomic features and solely using liver or spleen radiomic features for the classification task.

3.3.3 Discussion

In this exploratory study, we evaluated the potential for using radiomic features for severity assessment of patients with liver cirrhosis. Our hypothesis is that medical images of cirrhotic patients hold sufficient latent information to provide a more accurate or more clinically useful severity assessment compared to current clinical indicators such as MELD or decompensation.

Our experiment is complicated by the fact that no objective "ground truth" disease severity metric is available for us to model and evaluate against. Hence, we evaluated our predictive models against several existing clinical surrogates for disease severity: MELD score, MELD score ≥ 9 , MELD score ≥ 15 , and decompensation status. As we ultimately seek to improve on current practice, perfect alignment with existing metrics would be neither expected nor desired. Future clinical evaluation is required to fully assess the accuracy and utility of our method compared to (or in conjunction with) existing methods.

We evaluated the prognostic value of liver and spleen features together as well as solely using liver or spleen features in our experiments. Although we could not show a regression of MRI-derived radiomic features with each MELD score for severity assessment - which was an ambitious attempt from the start considering the heterogeneous patient cohort in its cirrhosis etiologies and partially very limited number of patients examples or even non-existing imaging examples for each MELD score in our data base (see Figure 3.1) - we have demonstrated that MRI-derived radiomic features have the potential to be used for severity stratification.

We received the best results for predicting a lower or higher severity in experiment 2 and 4 if a combination of liver and spleen features is used (exception: experiment 3, discussed below). For experiment 2 in which the median MELD score was used as threshold for a patient's classification into a lower or higher disease severity, it must be mentioned that spleen features alone (AUC 0.77) have shown a higher predictive value than solely using liver features (AUC 0.71). We believe this small but significant disparity represents a previously unreported discovery worthy of future study. In current clinical evaluation of liver disease, the spleen is typically considered only in passing as an impression of splenomegaly or a rough measure of size by a radiologist. We believe that a more detailed analysis of spleen features has previously overlooked value in the diagnostic assessment of cirrhosis.

In experiment 3, a MELD score 15 was used as threshold to define the two different classes for severity assessment, marking an important disease stage by considering if the respective patient should be listed for liver transplant. While broad conclusions are

difficult to make based on a relatively small cohort size of 90, the importance of spleen features decreased considerably in this experiment in contrast to experiment 2 where the lower MELD score was used as class threshold. This manifests not only in an AUC of 0.59 but is likewise indicated by a relatively high p-value of $p \approx 0.2$. In this context liver features alone claim the main contribution in this classification task. Liver and spleen features together even decrease the classification accuracy for this experiment. This could imply that for more advanced cirrhosis spleen features lose importance while liver features alone gain significance for severity assessment. However, spleen features remain important to distinguish disease stages for patients with a lower MELD score. One explanation for this result could be that physical and morphological changes to the spleen tend to occur in less acute cases and earlier in the progression of the disease, while the liver continues to undergo observable changes as MELD increases.

Experiment 4 achieved the best classification result in this work with an AUC of 0.84 for discriminating between compensated and decompensated cirrhosis, meeting or modestly exceeding MELD's predictive ability (AUC of 0.79). This result may imply that changes to the liver and spleen manifested in radiomic features align with a definition of severity defined by decompensation. While useful in clinical practice, decompensation is a crude binary measure of disease that does not provide detailed insight into the progression or severity of cirrhosis compared to, for instance, the MELD scoring system with a range for progressive severity assessment from 6 to 40. A larger patient cohort would allow a better understanding of how the liver and spleen change as patients approach and pass through the decompensation threshold. Furthermore, it must be mentioned that MELD score exception points are not handled in this study which might increase some patient MELD scores. Further studies are needed if this impacts the prognostic value of the MELD score for predicting liver decompensation.

Moreover, experiments 2-4 demonstrate that we can train reliable, predictive models for each classification task. Even with unbalanced data sets we demonstrate the significance of our cross-validated accuracy scores with random permutation tests. In accordance with this, the p-values of experiments 2-4 are always < 0.01 for either using liver or spleen features in the respective experiment (see corresponding p-values for each experiment in Tab. 2). Stability and robustness of the trained predictive models can also be seen in the ROC curves in Fig. 3.4 within the distance of each curve to the 50 % recall ratio. A separation of a training and test set was not feasible in the relatively small cirrhotic cohort containing a very heterogeneous distribution of disease severity. However, to increase the general robustness of our experiments and to find stable and reliable radiomic feature

for a radiomic signature a balanced data set would be desirable. Nevertheless, in our case an overall larger patient cohort would be an additional prerequisite. Generally, the classifier needs an "adequate" number of data sets - dependent on the complexity of the classification task - in order to learn to distinguish properly between two or more classes.

According to a study published by UNOS and OPTN the median MELD score at liver transplantation in the USA in 2018/2019 is a score of 35 [96]. For future studies we want to increase the span of MELD scores in our patient cohort from 6 to 40, which would mean including data from other research centers, requiring the alignment of factors such as scanner type, MRI protocols, lab test protocols, and documentation in patient records. Furthermore, an analysis of a larger patient cohort that includes a wider range of MELD scores (and more examples for each score) is more likely to yield a stable set of prognostic features.

Moreover, it has to be evaluated whether additional, objective surrogates for disease severity can be determined and included in future prediction models for the this study. For instance, the image feature analysis could be combined with other metrics derived from lab tests and the patient records (such as MELD, decompensation, and additional factors such as those used in MELD Plus [54]). Statistical analysis could then be used to weight the different components by relevance to form a more wholistic clinical decision support system.

Beyond expanding the patient cohort, several additional steps will be required in order to produce a fair, objective, transparent, and widely useful radiomics-based signature or biomarker for cirrhosis severity. In particular, the differences in imaging produced at different hospitals using MRI scanners made by different manufacturers must be accounted for. Fortunately, the HCC screening protocol used in our cohort corresponds to a widely-used standard in the field. In addition, our use of an open source library for radiomic feature extraction (PyRadiomics) [150] and the herewith open availability of our experimental setting should facilitate validation and extension of this study by the research community.

3.4 Prognostic Value of MRI-based Radiomic Features in End-Stage Liver Disease

Publication in preparation with planned title: *Prognostic Value of MRI-based Radiomic Features in End-Stage Liver Disease*

Abstract: In this study we assess the prognostic value of MRI-derived radiomic features for detecting liver cirrhosis. Cirrhotic patients (n=90) were scanned with hepatocellular carcinoma (HCC) MRI screening protocol. A non-cirrhotic (n=77) control cohort was scanned with exactly the same acquisition protocol but for intraductal papillary mucinous neoplasia (IPMN) that is known as having no impact on the organs (liver and spleen) from which we extract and analyze quantitative radiomic features. The goal is to analyze 1.) if specific radiomic features can be found in order to classify cirrhosis, 2.) the number of features needed for this classification, and 3.) to identify what kind of radiomic feature classes are generally important to classify cirrhosis. We aim to comprehend radiomic characteristics of cirrhosis to gather knowledge considerably impacting future studies for image-based severity assessment of end-stage liver disease. The fast correlation-based filter (FCBF) was used for feature reduction and selecting most relevant features. With a combination of liver and spleen features we identified wavelet features as most important, followed by first-order statistical features from which four radiomic features alone are sufficient to classify cirrhosis with an AUC of 0.94. All results are computed using the random forest classifier and cross-validation validated on an independent test set.

Introduction: The introduction here is a continuation of the general introduction in Section 3.1. The goal of this retrospective study is to analyze and understand the specific MRI-based radiomic features that are necessary to discriminate automatically between cirrhotic and healthy patients. In this study here, we hope to gain further insights on important features for risk stratification of cirrhosis with radiomic features.

To securely confirm cirrhosis, liver biopsies are still a state-of-the-art but they can be a risky procedure for patients whose liver might already have undergone major changes and lost a significant amount of its biological functionality [123]. Biopsies have an increased risk of complications and their use is controversial for smaller lesions due to possible sampling errors [134, 123]. On the contrary, MRI is part of clinical routine for diagnosis, therapy planning, and screening (i.e. in hepatocellular carcinoma (HCC) surveillance) and already complements laboratory tests in assessing the current patient's stage of cirrhosis [155].

Nevertheless, the acquired MR images are not yet used in clinical practice for an automatic image-based assessment, neither for detection nor for liver cirrhosis staging. Not even the ultrasound measures FibroSURE™ or FibroScan®, as non-invasive alternative to liver biopsies, give holistic insights on how compensated a cirrhotic patient's liver actually is [102].

To the best of our knowledge we are the first group investigating an image-based radiomic biomarker for cirrhosis detection. Our overall research's objective of this study is to produce a cirrhosis biomarker that can be used to support guidance in patient assessment and treatment or to supplement MELD (Model for End-Stage Liver Disease) to improve risk stratification and liver transplant prioritization.

As aforementioned, the uniqueness of this study is that besides the cirrhotic patient cohort a control cohort could be identified that was scanned following exactly the same MRI acquisition protocol, with the same contrast agent on the same type of MRI scanner. As a reminder: The control group are patients with a benign pancreas tumor, intraductal papillary mucinous neoplasia (IPMN), that has generally no impact on neighboring abdominal structures (see Section 3.2) [33].

3.4.1 Material and Methods

This section complements to Section 3.2 where the two patient cohorts are described in detail together with the applied preprocessing steps on the fat-suppressed and contrast-enhanced MR images. Basically, the same preprocessing steps are applied as to the cirrhotic data in the previous study for hepato-splenic feature extraction in Section 3.3. This also includes using the PyRadiomics library [150] for radiomic feature extraction.

Machine Learning Analysis

We used the cirrhosis cohort (n=90) and the control cohort (n=77) and defined a stratified, randomly assigned subset of (n=34) patients as test data set. The rest of the data (n=133) was used as training data set. This represents a split of approximately 80% training (used in cross-validation) and 20% independent test data.

The goal of the study was to gain insights on the most important radiomic features in order to classify the presence of cirrhosis. For this purpose, the training data is used for an initial feature selection using the fast correlation-based filter (FCBF) method to identify features with minimal redundancy and maximized relevancy due to pairwise analysis of correlations between features [166]. This is done in order to reduce the initial set of 2577 extracted features to those with the most prognostic power (see Section 3.2 for details on

feature extraction).

Afterwards, prediction models were trained based on these most important radiomic features on the training set by using a random forest classifier. As a final subsequent check, to evaluate whether a stable set of radiomic features could be found for this classification task, the trained classifier is applied to the as yet untouched test data set. During several experiments of analyzing feature importances (that are further described in the following section in conjunction with attained results) the set of used important radiomic features for classification is increased iteratively and the attained classification accuracy is analyzed. As a final validation, the subset of radiomic features obtaining the best prediction model for cirrhosis detection during training is evaluated on the test set.

Accordingly, different radiomic feature classes and their feature importances were analyzed individually to gain insights on their impact on predicting cirrhosis. All results are computed using 10-fold cross-validation. In addition, we applied the random permutation test to evaluate the significance of our scores.

3.4.2 Results

In this section we give an overview of our initial feature selection and reduction, overall feature importance and class-wise feature importance analyses, and obtained classification results for classifying cirrhosis. Fig. 3.5 gives an overview of our approach.

Radiomic Feature Importance for Classifying Cirrhosis

We perform an initial feature selection utilizing the established FCBF method on the whole training cohort. Basically, we wanted to reduce the 2577 radiomic features to a set of non-correlating features with the most prognostic power. Like other feature selection methods FCBF scores each feature and, thus, allows to rank the feature importance. In Tab. 3.3 the top 14 most important and non-correlating radiomic features are listed that remained after feature selection. We picked these ranked 14 features to analyze how many of them are actually necessary for an optimal classification of liver cirrhosis on the training data set.

In Tab. 3.4 we show the results of adding n features to the feature vector for training the classifier on the training data with respective AUCs and standard deviations. Features are added in accordance of their respective rank: starting from the most important feature $n = 1$ to the least important at rank 14.

To prevent an overfitting on the training data we introduce a delta to determine the minimum increase in accuracy a newly added feature to the feature vector must represent

in order to be counted as the new best set of features. We chose a $\delta = 0.005$ for selecting the best set of features in the training cohort.

As can be seen in Tab. 3.4 together with the constraint of a minimum increase in accuracy of $\delta = 0.005$, the four most important features are selected as having the most prognostic power for classifying cirrhosis on the training set. It can also be observed that the addition of further features does not necessarily lead to better classification accuracy. For this experiment we used stratified 10-fold cross-validation and repeated the cross-validation process 100 times to attain different stratified folds in each pass of cross-validation. This is done in order to obtain representative, averaged classification results. This subset of the first four most important features leads to an AUC of 0.99 ± 0.03 on the training set. Subsequently, the obtained classifier trained on the four most important features is applied to the test data which led to an AUC of 0.94.

Analyzing the Importance of Each Radiomic Feature Class for Classifying Cirrhosis

In this section we analyze the feature importances of each radiomic feature class, which we divide as follows: original features (first order, shape, GLCM, GLDM, GLRLM, GLSZM, NGTDM), wavelet features, and LoG features. These feature classes are the basic feature classes provided by the PyRadiomics library. Essentially, we repeat the previously described feature selection and reduction process that was used to identify the most important features for cirrhosis classification with the difference of only employing features from one feature class at the same time for classification.

Feature ranking and selection is performed as well and in the same manner as described above. Aside from that, it is evaluated how many features of the respective feature class are needed to obtain the optimal classification result on the training set. Tab. 3.5 summarizes the features from each feature class that are necessary for an optimal classification evaluated on the training set. The table also displays the classification results on the independent

Rank	Feature	Score	Rank	Feature	Score
1.	wavelet-HHH_glszm_SizeZoneNonUniformity_3DLiver	0.444	9.	log-sigma-5.0-mm-3D_ngtdm_Contrast_3DSpleen	0.187
2.	original_shape_Sphericity_3DLiver	0.441	10.	log-sigma-3.0-mm-3D_glszm_ZoneEntropy_3DLiver	0.128
3.	wavelet-LLL_ngtdm_Strength_3DSpleen	0.419	11.	wavelet-HLL_glcm_Correlation_3DLiver	0.107
4.	log-sigma-2.0-mm-3D_gldm_DependenceVariance_3DLiver	0.39	12.	original_glrIm_LowGrayLevelRunEmphasis_3DSpleen	0.085
5.	wavelet-HHL_glszm_SizeZoneNonUniformity_3DSpleen	0.325	13.	wavelet-HHL_firstorder_Range_3DLiver	0.071
6.	original_glcm_MaximumProbability_3DLiver	0.266	14.	log-sigma-4.0-mm-3D_glcm_ClusterShade_3DSpleen	0.060
7.	wavelet-LLH_glcm_Imc_3DSpleen	0.224			
8.	wavelet-LLH_glcm_Imc2_3DLiver	0.203

Table 3.3: Overview of the 14 most important liver and spleen radiomic features in the training data set. The FCBF method was used for feature scoring.

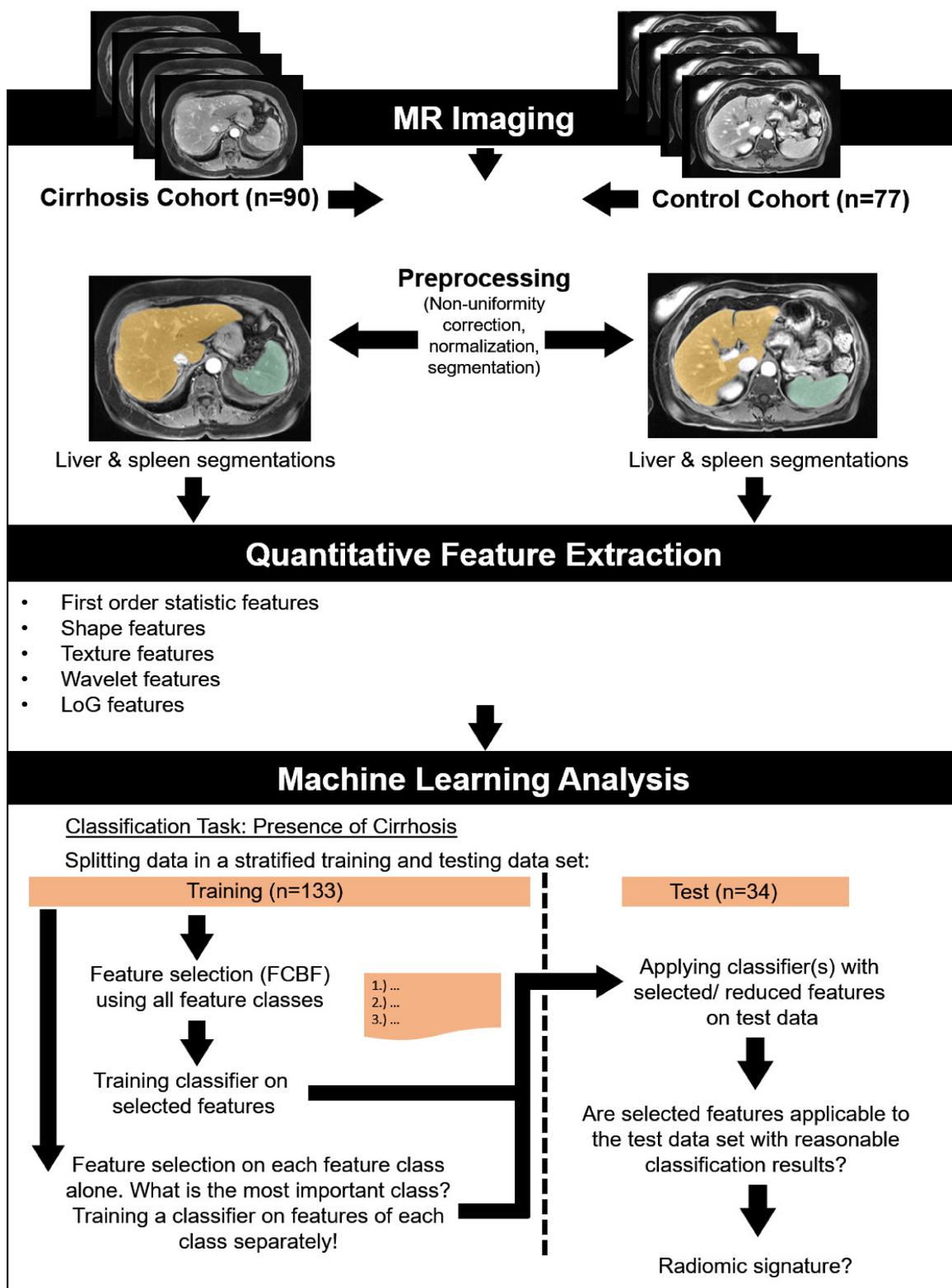


Figure 3.5: General overview of the feature extraction process, feature selection, and evaluation of feature importances in order to classify the presence of cirrhosis.

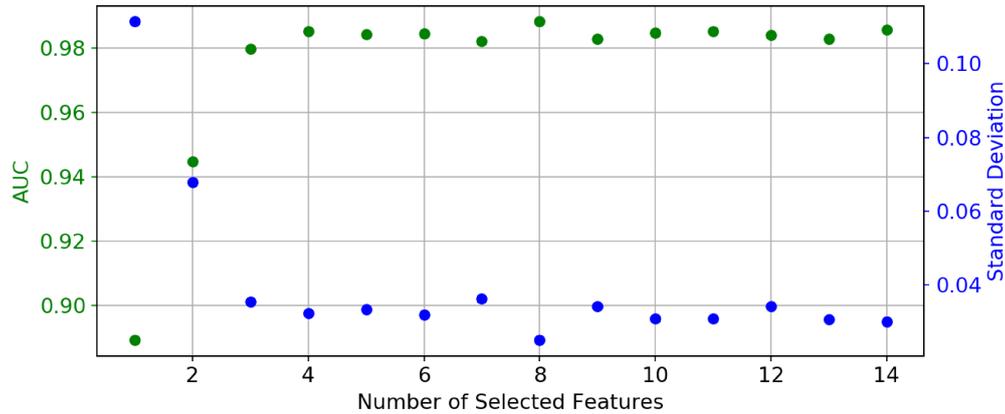


Table 3.4: AUCs and corresponding standard deviations of initial feature selection. Within this experiment the impact of the top 14 most important radiomic features on the classification result was analyzed. For this, the feature vector was extended successively from the feature with rank 1 to rank 14. In each iteration a classifier was trained with the selected subset of n features on the training data set. Subsequently, AUC and standard deviation were calculated for using the respective subset of the top n features.

test set by applying the best prediction model obtained from training the classifier on the training set to the test data.

Furthermore, we analyzed the importance of each feature class in the training set without prior feature selection and reduction. According to the aforementioned experiment, the classifier was trained separately on each feature class and the classification accuracies were computed respectively.

Again, we repeated the stratified 10-fold cross-validation process 100 times for a good averaged classification result. The *original* feature class is also further subdivided in its components. The results of this experiment are illustrated in Tab. 3.6 with corresponding AUCs and standard deviations.

3.4.3 Discussion

In this exploratory study we investigate whether we could identify MRI-based radiomic features that are characteristic for distinguishing between a cirrhotic and a non-cirrhotic cohort. The overall goal is to apply the gained knowledge from this research study to future studies for automatic, radiomic-based severity assessment of liver cirrhosis. In this

context this research represents another step towards understanding and decoding the specific, latent image information in MR imaging that could be used as radiomic biomarker to identify and risk-stratify end-stage liver disease patients.

A radiomic-based approach for identifying cirrhosis could represent a save alternative to the overall risky procedure of liver biopsy. Moreover, these results could add anatomical and texture knowledge of the liver and spleen parenchyma to the MELD score. The objective is to better detect potentially severe complications and increase the knowledge of pathogenesis through identifying specific cirrhotic tissue transformations. These characteristic transformations and architectural tissue changes could mark important transitions to more severe disease stages and could even mark the transition from compensated to decompensated cirrhosis.

Tab. 3.3 shows the ranked radiomic feature importances on the training data set.

Feature Class	Optimal Number of Features/ Most Predictive Features	AUC training data	AUC test data
Original features	1.) shape_Sphericity_3DLiver 2.) ngtdm_Strength_3DSpleen	0.96 ± 0.05	0.85 $p < 0.01$
Original features (except shape)	1.) ngtdm_Strength_3DSpleen 2.) glcm_MaximumProbability_3DLiver 3.) glcm_SumEntropy_3DSpleen	0.89 ± 0.09	0.85 $p < 0.01$
Shape features	1.) shape_Sphericity_3DLiver 2.) shape_VoxelVolume_3DSpleen	0.93 ± 0.07	0.82 $p < 0.01$
Wavelet features	1.) HHH_glszm_SizeZoneNonUniformity_3DLiver 2.) LLL_ngtdm_Strength_3DSpleen 3.) HHL_glszm_GrayLevelNonUniformity_3DSpleen 4.) LLH_glcm_Imc2_3DLiver	0.98 ± 0.03	0.91 $p < 0.01$
LoG features	1.) 2.0-mm-3D_gldm_DependenceVariance_3DLiver 2.) 2.0-mm-3D_ngtdm_Strength_3DSpleen 3.) 3.0-mm-3D_glrml_LongRunHighGrayLevelEmphasis_3DSpleen 4.) 2.0-mm-3D_glszm_ZonePercentage_3DLiver 5.) 5.0-mm-3D_firstorder_Skewness_3DLiver	0.93 ± 0.07	0.85 $p < 0.01$

Table 3.5: Feature classes and most important features after feature selection. After ranking the features with the FCBF method, it was evaluated on the training data set how many of the most important features from each class are needed to obtain the best classification result/best prediction model. The result is a prediction model trained on the training set (see *AUC training data*) that is then applied to classify the test set (see *AUC test data*).

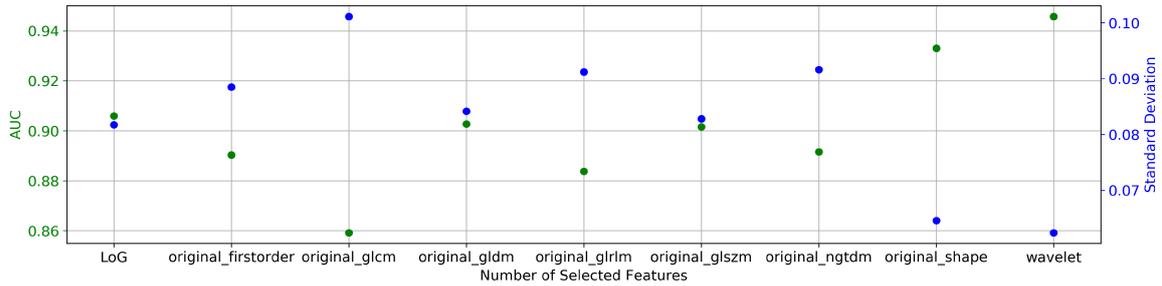


Table 3.6: AUCs and corresponding standard deviations of classification results using different feature classes on the training data set prior feature selection (reduction).

Noticeable is that from a total of 2577 extracted features only 14 non-correlating features remain as important features for classifying cirrhosis. It can also be seen that both liver (in total eight from 14 features) and spleen (in total six from 14 features) features are important for this classification task. The most dominant/predictive feature class are wavelet features (seven from 14 features) in this experiment, giving textual changes a high impact in predicting cirrhosis.

With our second experiment we determine how many of these 14 features are actually necessary to train an optimal prediction model for cirrhosis and we further narrowed down the 14 features in their importance. Tab. 3.4 shows that the first four features ranked in Tab. 3.3 are sufficient to attain a fairly high classification accuracy on the training data (AUC 0.99 ± 0.03). We verified this observation by applying the obtained prediction model on these four radiomic features to the independent test data set. This resulted in a comparably high classification accuracy with an AUC of 0.94.

Among these four most important radiomic features are two wavelet features that measure the homogeneity/heterogeneity of the gray value intensities within the liver (*glszm size zone non-uniformity* at rank 1) and the spleen (*ngtdm strength* at rank 3). The forth most important feature is also a texture feature that measures texture coarseness based on a Laplacian of Gaussian filtered image [150]. In this case a kernel with a *sigma* of 2 mm was applied. Besides these three texture features, a shape feature of the liver is ranked as second important radiomic feature: *shape sphericity*. This feature measures the roundness of the liver shape relative to a sphere.

All the aforementioned reported important features are unsurprising and conform with known cirrhotic tissue transformations (see Section 1.3.3 for comparison) and validate our findings from a clinical perspective. Nevertheless, these observations were not made quantitatively before in this specific field of research by using radiomics together with

the observation of more complex changes within the spleen tissue beyond splenomegaly (spleen enlargement).

Furthermore, we analyzed the importance of each feature class for predicting cirrhosis. The results can be seen in Tab. 3.5 and are conform with those from the previously described experiment. Consequently, the classifier trained explicitly on wavelet features achieved the highest classification accuracy by focusing on texture features within the liver and spleen. This resulted in an AUC of 0.98 ± 0.03 on the training data and an AUC of 0.91 on the test data by classifying cirrhosis with the selected subset of wavelet features.

The second most important feature classes based on the AUCs achieved on the test data set are: *original features* (historic first published set of PyRadiomic features) that also include shape features, *original features without shape features*, and *LoG features* that inherit features from other classes by filtering with different kernel sizes. All the aforementioned feature classes resulted in an AUC of 0.85 when the trained classifiers were applied on the test data. Shape features alone have the least impact with an AUC of 0.82 on the test data. However, it must be emphasized that the two most predictive features here are compliant with clinical observations concerning liver shape alterations and splenomegaly (see Section 1.3.3 for comparison). Nevertheless, the impact of each feature class alone is quite high. Another fact that deserves to be highlighted is that always liver and spleen features are among the most predictive features in each feature class. Conclusively, this experiment demonstrated that not a single feature class but a combination obtains best prediction results.

The last experiment was to illustrate the importance of each feature class, and sub-classes if appropriate, on the whole training set. On the contrary to previous experiments, this was done prior FCBF feature selection and, therefore, before reducing the whole set of 2577 features. The original features were further subdivided into sub-classes for further evaluation: original first order features, GLCM, GLDM, GLRLM, GLSZM, NGTDM, and shape features. Moreover, the feature importances of LoG and wavelet features were analyzed in this context. In Tab. 3.6 the importance of each feature class for predicting cirrhosis can be depicted together with achieved AUCs and standard deviations. The first three most important feature classes are wavelet features, shape features, and LoG features. The least important feature class for predicting cirrhosis are Gray Level Co-occurrence Matrix (GLCM) features. These results are compliant to those shown in Tab. 3.5.

3.4.4 Conclusion

With the exploratory results of this study we are one step closer to determine a radiomic signature for predicting cirrhosis. We demonstrated that wavelet features, or more general, features that measure heterogeneity, are the most important features to distinguish between *healthy* livers and spleens from those of a cirrhotic cohort. Nevertheless, shape and LoG features complement to the prediction accuracy. In total only four features out of 2577 features are sufficient to predict cirrhosis in the test data with a fairly high AUC of 0.94. We also evaluated that liver as well as spleen radiomic features are important for classifying cirrhosis.

To the best of our knowledge we are the first research group identifying specific MRI-based radiomic features for predicting cirrhosis. Our use of an open source library for radiomic feature extraction should facilitate validation and extension of our work by other researchers. However, the long-term goal of this study is not to predict cirrhosis but to characterize features that can also be used for severity stratification of cirrhotic patients. Future studies will show if the here gained knowledge about feature importances can complement severity predictions (see conclusion of Part II of this thesis in Section 3.6).

3.5 Important Radiomic Features for Healthier and Sicker Cirrhotic Patients

The results presented in this section were not published previously and are further investigations on severity assessment of cirrhosis using MRI-based radiomic features. The here presented findings round off the previous two studies in Sections 3.3 and 3.4 and complete the performed analyses on the topics of Part II of this thesis. It is planned to publish the results of the study in this section in a clinical journal. Accordingly, a paper-like structure is followed to describe the background, material and methods, and results in this context. The section closes with a discussion of attained findings.

Introduction: This section contains further experiments that connect and pursue the presented results from the studies described in Section 3.3 and in Section 3.4. As a reminder: In Section 3.3 a radiomic feature analysis is performed to train different prediction models for assessing the patient's condition at time of MRI scan: if the patient's MELD score is ≥ 9 , if the score is ≥ 15 , and if the patient is decompensated. Whereas in Section 3.3 several experiments were conducted to find the most important MRI-based radiomic features implying the most prognostic value to determine if a patient's scan reveals signs of cirrhosis.

In the following, the observations made in these two previous sections are combined by training prediction models on MRI-based radiomic features extracted from the liver and spleen and by analyzing the most important radiomic features for distinguishing between *healthier* (compensated and patients with a MELD score under 9) and *sicker* patients (decompensated and patients with a MELD score ≥ 9). The IPMN cohort was used as a healthy hepato-splenic baseline in these experiments in order to distinguish between:

1. IPMN patients (*healthy* cohort) (n=77) and cirrhotic-compensated patients (n=62);
2. IPMN patients and cirrhotic-decompensated patients (n=28);
3. IPMN patients and cirrhotic patients with a MELD score < 9 (n=46);
4. IPMN patients and cirrhotic patients with a MELD score ≥ 9 (n=44).

Unfortunately, the experiment to distinguish between healthier and sicker patients with a MELD score under or ≥ 15 only contains 13 patients above score 15. All other aforementioned experiments are performed using a separate, independent test set to evaluate the

findings. In this particular case it is not feasible to split into reasonable sets that allow sufficient training and subsequent testing.

The purpose of the study described in this section is to identify MRI-based radiomic features with the most prognostic power for the aforementioned four classification tasks. We expect this to provide us with further important insights into the characteristics of progressive cirrhosis and the assessment of severity using hepato-splenic radiomic features.

3.5.1 Material and Methods

The here used IPMN and cirrhotic patient cohorts were previously described in Section 3.2. The cirrhotic sub-cohorts were first mentioned in Section 3.3.1 on page 98 where different experiments are outlined in order to assess the severity of cirrhosis with MRI-based radiomic features. The same image preprocessing, liver and spleen segmentations, and the same parameters and public library for radiomic feature extraction were used as in the two previous studies as described in Section 3.2 on page 96.

Machine Learning Analysis

In order to produce comparable results to the previously described studies in this chapter, the same experimental setup is used here for dividing 80% training (used in cross-validation) and 20% test data splits as described in the previous study in Section 3.4.1. Likewise, patients were assigned randomly to the stratified training and test sets. This resulted in the following training and test splits within the aforementioned four experiments:

1. Experiment: The IPMN (n=77) and cirrhotic-compensated (n=62) cohort represent in total 139 patients: Of which 111 patients are in the training set and 28 are assigned to the test data set.
2. Experiment: The IPMN and cirrhotic-decompensated (n=28) cohort cover in total 105 patients: Of which 84 are in the training and 21 are in the testing data set.
3. Experiment: The IPMN and cirrhotic cohort with MELD scores < 9 (n=46) are in total 123 patients. Resulting into 98 patients in the training set and 25 in the test set.
4. Experiment: The IPMN and cirrhotic cohort with MELD scores ≥ 9 (n=44) are in total 121 patients: Of which 96 are assigned to the training set and the remaining 25 represent the test data set.

As in the previous studies we applied the fast-correlation based filter (FCBF) method for feature selection. Consequently, as the first step, the selection of most important features for each experiment was done once on the whole training data set for each experiment.

In accordance to the previous study in Section 3.4, the training set was not only used to determine the relevant features but also to evaluate the features that have the most prognostic power in the classification task of each experiment.

For identifying these features the same method was used as introduced Section 3.4: According to the feature importance rank of each feature, they were added iteratively to the feature vector used for classification. The set of best features that obtained the highest classification accuracy are considered as most powerful radiomic features for the classification task in the respective experiment. Aside from that, the same accuracy constraint of performing better than $\delta = 0.005$ as a previous set of best features must be satisfied to be a valid, better set of features (see Section 3.4.2 for details).

For these experiments stratified 5-fold cross-validation was employed within the training set and also repeated 100 times for each classification task and iteration for identifying the best set of features, respectively. The results are averaged classification accuracies. This was done in order to obtain comparable results to those represented in Sections 3.3 and 3.4. Moreover, for an additional validation of our findings, we also calculated the corresponding p-values of each experiment using a random permutation test to verify significance of resulting scores as in the previously described two studies within this chapter of this thesis.

3.5.2 Results

In this section we give an overview of the attained exploratory results of this study that are summarized in Tab. 3.7. The table shows the important non-correlating features after feature selection with the FCBF method for each experiment. The radiomic features that have the most predictive power and with which the best classification results could be achieved on the training data are highlighted. The prediction model trained on the training data set with this set of features, or this single radiomic feature, was then used to classify the independent test set. Corresponding AUCs are displayed in the last two columns of the aforementioned table.

Highest classification accuracies were attained in experiment 2 and 4 in which *sicker* cirrhotic patients were distinguished from the non-cirrhotic IPMN cohort. Experiment 1 in which the classifier had to identify IPMN patients and *healthier* cirrhotic-compensated patients marks the classification result with the lowest AUC of 0.75.

Experiment 1			Experiment 2		
	Cirrhosis	IPMN		Cirrhosis	IPMN
Cirrhosis	TP 15	FP 3	Cirrhosis	TP 5	FP 0
IPMN	FN 4	TN 6	IPMN	FN 0	TN 16

Experiment 3			Experiment 4		
	Cirrhosis	IPMN		Cirrhosis	IPMN
Cirrhosis	TP 5	FP 1	Cirrhosis	TP 8	FP 1
IPMN	FN 1	TN 18	IPMN	FN 0	TN 16

Figure 3.6: Confusion matrices that summarize the results of the attained prediction models on the test data of each experiment 1 - 4.

In Experiment 1 and 2 only one radiomic feature was necessary in order to distinguish cirrhotic-compensated and cirrhotic-decompensated patients from the IPMN cohort. In experiment 3 five features were needed to determine whether a *healthier* cirrhotic patient has a MELD score under 9 (median MELD in cirrhotic cohort) or belongs to the IPMN cohort. In experiment 4 three radiomic features were sufficient to distinguish between the non-cirrhotic IPMN cohort and the *sicker* cirrhotic patients with a MELD score equal or above 9.

Furthermore, a random permutation test was performed on the test set to verify the significance of the achieved classification accuracy. The resulting p-values are illustrated together with the AUCs of the test set in Tab. 3.7. Additionally, the confusion matrices are illustrated in Fig. 3.6 in order to summarize the results with respective type I and type II errors of the trained prediction models for each experiment.

3.5.3 Discussion

The aim of this study was to determine whether reliable predictive models can be trained to differentiate between a healthy baseline cohort (IPMN) and a healthier/mildly sick or an even sicker patient cohort. In addition, a statement should be made about which features

3.5 Important Radiomic Features for Healthier and Sicker Cirrhotic Patients

Exp.	Important Features After Feature Selection	AUC training	AUC test	
1	1.) <u>wavelet-HHH_glszm_SizeZoneNonUniformity_3DLiver</u> 2.) original_shape_Sphericity_3DLiver 3.) wavelet-LLL_ngtdm_Strength_3DSpleen 4.) log-sigma-2-0-mm-3D_glcm_InverseVariance_3DLiver 5.) wavelet-HHL_glszm_SizeZoneNonUniformity_3DSpleen 6.) wavelet-LLL_glcm_MaximumProbability_3DLiver 7.) log-sigma-2-0-mm-3D_glszm_ZonePercentage_3DLiver	8.) wavelet-LLH_glcm_Imc2_3DLiver 9.) wavelet-HLH_glszm_GrayLevelNonUniformity_3DLiver 10.) wavelet-HHH_gldm_DependenceEntropy_3DLiver 11.) log-sigma-5-0-mm-3D_glszm_HighGrayLevel...ZoneEmphasis_3DSpleen 12.) log-sigma-1-0-mm-3D_glcm_Imc2_3DLiver	0.90 ± 0.07	0.75 p < 0.01
2	1.) <u>log-sigma-5-0-mm-3D_glcm_Idm_3DSpleen</u> 2.) log-sigma-4-0-mm-3D_glszm_LargeAreaLowGray...LevelEmphasis_3DSpleen 3.) wavelet-LLH_glrlm_GrayLevelNonUniformity_3DSpleen 4.) original_shape_Sphericity_3DLiver 5.) log-sigma-2-0-mm-3D_glcm_DifferenceEntropy_3DSpleen 6.) wavelet-HLH_ngtdm_Busyness_3DSpleen 7.) log-sigma-2-0-mm-3D_firstorder_Kurtosis_3DSpleen 8.) log-sigma-5-0-mm-3D_firstorder_Uniformity_3DSpleen 9.) wavelet-HHH_glszm_GrayLevelNonUniformity_3DLiver 10.) log-sigma-2-0-mm-3D_gldm_Dependence...NonUniformityNormalized_3DLiver 11.) original_firstorder_Entropy_3DLiver	12.) log-sigma-4-0-mm-3D_glcm_Imc2_3DSpleen 13.) log-sigma-5-0-mm-3D_glcm_Inverse...Variance_3DSpleen 14.) wavelet-HHL_glszm_ZonePercentage_3DLiver 15.) wavelet-HHL_glszm_SizeZoneNonUniformity_3DSpleen 16.) log-sigma-1-0-mm-3D_firstorder_Skewness_3DLiver 17.) wavelet-LHH_glszm_ZonePercentage_3DLiver 18.) wavelet-HLL_glcm_ClusterShade_3DLiver 19.) original_glcm_InverseVariance_3DLiver 20.) log-sigma-5-0-mm-3D_firstorder_90Percentile_3DLiver 21.) wavelet-HHH_gldm_SmallDependenceLowGray...LevelEmphasis_3DLiver 22.) wavelet-LLL_firstorder_Uniformity_3DSpleen	0.99 ± 0.02	1.0 p < 0.01
3	1.) <u>wavelet-HHH_glszm_SizeZoneNonUniformity_3DLiver</u> 2.) <u>log-sigma-2-0-mm-3D_gldm_DependenceVariance_3DLiver</u> 3.) <u>original_shape_Sphericity_3DLiver</u> 4.) <u>wavelet-LLL_glcm_MaximumProbability_3DLiver</u> 5.) <u>original_ngtdm_Busyness_3DSpleen</u> 6.) wavelet-HHL_glszm_GrayLevelNonUniformity_3DSpleen 7.) wavelet-HHL_glszm_SizeZoneNonUniformity...Normalized_3DSpleen	8.) wavelet-LHH_glrlm_GrayLevelNonUniformity...Normalized_3DLiver 9.) log-sigma-4-0-mm-3D_glcm_ClusterShade_3DLiver 10.) log-sigma-3-0-mm-3D_glszm_GrayLevelNon...UniformityNormalized_3DSpleen 11.) original_glszm_ZoneEntropy_3DSpleen 12.) wavelet-HLH_firstorder_Maximum_3DSpleen	0.97 ± 0.03	0.92 p < 0.02
4	1.) <u>original_ngtdm_Busyness_3DSpleen</u> 2.) <u>log-sigma-3-0-mm-3D_glrlm_RunPercentage_3DSpleen</u> 3.) <u>log-sigma-2-0-mm-3D_gldm_DependenceVariance_3DLiver</u> 4.) log-sigma-1-0-mm-3D_glszm_GrayLevelNon...Uniformity_3DSpleen 5.) log-sigma-4-0-mm-3D_ngtdm_Strength_3DSpleen 6.) log-sigma-5-0-mm-3D_glrlm_LongRunHighGrayLevel...Emphasis_3DSpleen 7.) original_shape_Sphericity_3DLiver	8.) original_shape_Maximum2DDiameterColumn_3DSpleen 9.) log-sigma-4-0-mm-3D_firstorder_Entropy_3DSpleen 10.) wavelet-HHL_glszm_SizeZoneNonUniformity_3DSpleen 11.) wavelet-HLH_ngtdm_Busyness_3DSpleen 12.) wavelet-HHH_glrlm_LongRunEmphasis_3DLiver 13.) wavelet-LLL_firstorder_Entropy_3DLiver 14.) wavelet-HHH_glszm_SizeZoneNonUniformity_3DLiver 15.) wavelet-LLH_glcm_Imc2_3DLiver 16.) log-sigma-1-0-mm-3D_firstorder_Kurtosis_3DLiver	0.99 ± 0.01	0.96 p < 0.01

Table 3.7: Ranking of important radiomic features within each experiment (abbr. Exp.). Also, the most predictive features are determined on the training data that are needed for an optimal classification result (**bold** and underlined). The resulting classifier from the training data is then applied to the independent test data set. The corresponding AUCs are displayed in the last two columns of this table. Attained p-values are displayed together with the AUCs of the test set.

have the most predictive power in these classification tasks and how many features are actually necessary to distinguish between the IPMN cohort and the respective more or less diseased cohorts.

In the following the results of each experiment are discussed separately before a joint conclusion is made on the overall obtained results in this study:

Experiment 1: What stands out in this experiment is the similarity of the important features to those generally necessary to distinguish between the entire cirrhosis cohort and the IPMN cohort. Especially, when comparing the top three most important features (see Tab. 3.3 on page 110). This also marks texture features measuring heterogeneity in the images as the most important features for the classification in experiment 1 of this study. On the contrary, the performed analysis shows that in experiment 1 here only one feature is deemed necessary for the classification task at hand, whereas there are four features needed to generally distinguish between the IPMN and the cirrhosis cohort in Section 3.4. The respective classification accuracies also differ in these compared analyses: Experiment 1 here obtained an AUC of 0.75 on the test data, whereas the experiment from the previous study achieved an AUC of 0.85. Generally, it seems to be harder to distinguish between the healthy livers and spleens of the IPMN cohort and the healthier cirrhotic but still compensated cohort. The AUC of 0.75 is quite low compared to the other experiments. This can be a result of an overfitting on the training data.

Experiment 2: This experiment had the highest AUC with 1.0 on the test set. After ranking important features and analyzing how many of them are needed to distinguish between the IPMN cohort and the sicker cirrhotic-decompensated patients one feature alone achieved the highest classification accuracy on the training set. The high importance of this heterogeneity feature *log-sigma-5-0-mm-3D_glcm_Idm_3DSpleen* was confirmed after applying the thereby trained prediction model on the test set. Another interesting observation is the higher proportion of spleen features among the top ten ranked important features. This observation is also shared by experiment 4 where the classification task is as well to differentiate between the IPMN and a *sicker* cirrhotic patient cohort. This observation must be particularly emphasized when comparing the results of experiments in Section 3.3.2. Here spleen features were considered especially important for distinguishing between patients with lower and higher MELD scores. The difference in feature space seemed to be even more significant/higher when severely diseased cirrhotic patients are compared to a healthy hepato-splenic baseline cohort, confirming the results of the initial study in Section 3.3 on spleen feature importance.

Experiment 3: For this experiment five radiomic features turned out to be the most prognostic features to discriminate between the IPMN cohort and mildly diseased patients with a MELD score under 9. Four of these features are based on characteristic changes of the liver of which three measure the heterogeneity within the liver tissue and one

feature describes the liver shape. The remaining important feature is a spleen feature that measures heterogeneity within the spleen tissue. Overall an AUC of 0.92 could be achieved. Interesting here is that a patient that is also misclassified in the test set of experiment 1 for having cirrhosis is also misclassified here. Revisiting the electronic medical record of this patient revealed a borderline platelet count which can be from any number of causes such as portal hypertension, which is very common in cirrhotic patients. For future studies we will have to evaluate an exclusion of this peculiar patient from the IPMN cohort.

Experiment 4: Likewise, a fairly high classification accuracy could be obtained in this experiment with an AUC of 0.96. The purpose of this experiment was to determine the most important features for distinguishing between the non-cirrhotic IPMN and the severely diseased cirrhotic cohort with MELD scores equal or above 9. Three features were identified as being most predictive for this classification task and yielded to the highest classification accuracy on the training set. Two of these features are heterogeneous spleen features and measure texture coarseness. The third most predictive feature is a heterogeneity feature of the liver texture.

For experiment 1, 2, and 4 a p-value of < 0.01 could be attained and a p-value of < 0.02 was computed for experiment 3. The p-values were calculated with the random permutation test as in the previous studies of this chapter in order to evaluate the significance of the attained results in this study on the respective test sets.

Moreover, we re-examined the electronic medical records of misclassified patients again to see if there are any exceptional conditions.

An extreme example would be an IPMN patient that is later diagnosed with cirrhosis. Apart from the IPMN patient with the borderline platelet count that is already discussed above in experiment 3 one other patient showed irregularities. This patient belongs to the cirrhotic-compensated cohort and is as well misclassified in experiment 1. The patient was presumed NASH (Non-Alcoholic steatohepatitis)/alcohol induced cirrhosis and may have reversed hepatic fibrosis from ceasing alcohol use and weight loss. For future studies this must be taken under consideration if this is a typical patient for detecting cirrhosis. Otherwise, no other abnormalities could be detected neither in the imaging nor on the medical records that would be responsible for misclassification at the time.

To summarize the findings of our exploratory research we made the following general observations: Overall, the classification accuracy is slightly better for differentiating the IPMN cohort from sicker cirrhotic patients than discriminating between IPMN and mildly

diseased cirrhotic patients.

As in the results described in the previous two studies in this chapter we can also confirm the importance of radiomic spleen features in our experiments. Tendencies were already observed in Section 3.3 that spleen features are more important when distinguishing healthier (compensated, MELD score < 9 or 15) cirrhotic patients from sicker (decompensated, MELD score > 9 or 15) cirrhotic patients. The results of this study show an even stronger significance of the importance of spleen features when a healthy hepato-splenic IPMN cohort is compared to severely diseased cirrhotic cohorts.

Altogether, radiomic features are important that quantify the texture heterogeneity of the liver and/or spleen tissue. Another advantage of this study are larger patient cohorts that allowed dividing the data set into training and testing sets. Therefore, the here obtained results are unbiased and they confirm the findings and observations made in the initial study in Section 3.3.

Adding the gained knowledge here to the observations made in the studies in the previous sections, it is difficult to assume that only one radiomic signature would be sufficient to predict both the presence of cirrhosis and simultaneously allow a risk-stratification of cirrhotic patients. The experiments conducted here together with the experiments in the previous studies allow 1.) to basically distinguish between healthy livers and spleens from the IPMN cohort and those of the whole cirrhotic patient cohort and 2.) the experiments of the study in this section add an indirect way of a rough disease stratification to 1.) by distinguishing between the IPMN cohort and mildly and severely diseased patients. The next consecutive step in this context would be to verify the here obtained findings on a larger independent validation set and to expand the patient cohorts to allow a finer scale of severity assessment of cirrhotic patients. To attain this also more severely sick patients above a MELD score of 24 are absolutely necessary to extend the here acquired knowledge to a larger, more representative, range of disease severity. We hope to observe even more characteristic transitions on a larger scale that are identified by the respective most prognostic features. We assume that these features are possibly shifting in their individual importance for different end-stage liver disease stages.

Besides, integrating more MRI data in future studies (also from other centers that follow the same gold-standard MRI HCC screening protocol on a comparable MRI scanner) will allow an assessment of stability and reproducibility of MRI-based radiomic features in the here presented context.

With our research of this study we hope to inspire future studies to gain profound knowledge of the progressive characteristics of end-stage liver disease features on contrast-enhanced MR imaging to improve an early detection, risk stratification, and prediction of complications.

3.6 Conclusion and Outlook Part II

Part II of this thesis focused on clinical decision support of end-stage liver disease patients. Main objectives were a radiomic severity assessment of liver cirrhosis and a general detection of cirrhosis in five minutes delayed gadolinium-enhanced MR images (equilibrium phase). The MR images are part of a multiparametric scanning series of a gold standard hepatocellular carcinoma (HCC) screening protocol (see Section 1.2.1).

As being the first research group investigating in this specific field of research of a radiomic-based assessment of liver cirrhosis, we defined several subgoals for our research. These goals also split into three different studies that are part of Part II of this thesis:

First, we wanted to provide a proof of concept of the feasibility to risk-stratify cirrhosis patients using radiomics. This is why our first study of Part II of this thesis focused on different experiments using diverse surrogates, such as the MELD score or presence of decompensation, for assessing the severity of liver cirrhosis. With the initial study in Section 3.3 we could demonstrate that radiomic hepato-splenic features can be connected to disease severity. Observations show that we could even see a transition within feature importances from spleen to liver features with increasing severity. Unfortunately, our initial cirrhotic cohort of 90 patients was too small to significantly identify the individual importance of specific liver and spleen features in the first study of this chapter. Nevertheless, the first study supports our general hypothesis.

This leads to our second subgoal: We wanted to identify specific, radiomic features that are basically necessary to identify cirrhosis on MR images of the five minute delayed phase. Accordingly, the goal was to reduce the total set of 2577 radiomic features in order to determine the most prognostic features for cirrhosis screening. To implement a method that allows these analyses, we needed to integrate a baseline patient cohort with healthy livers and spleens on MRI. For this reason, patients with a benign pancreas tumor - intraductal papillary mucinous neoplasia (IPMN) - were included as additional patient cohort for our studies. The tumor is generally known as having no impact on other abdominal organs, especially, none on liver and spleen. As a prerequisite for our analysis, this cohort was scanned with the exact same type of MR scanner, MRI protocol, and contrast agent. This circumstance prevented facing and adding further challenges of extracting significant MRI-based radiomic features mainly due to the lack of MR signal

standardization [131]. Therefore, the IPMN cohort with 77 patients served as healthy liver and spleen baseline cohort in contrast to the cirrhotic cohort in our experiments. The results of the second study in Section 3.4 show that fundamentally four radiomic features are actually necessary to distinguish between non-cirrhotic IPMN and cirrhotic patients. A combination of liver and spleen heterogeneity as well as liver shape radiomic features turned out to be features with the most prognostic power for cirrhosis detection. Results were evaluated on unbiased, independent test sets.

The third study in Section 3.5 continues the investigations of the first and second study and connects and completes both research studies. Here, the radiomic features of the IPMN cohort function again as healthy baseline, whereas the classification task is now to distinguish between non-cirrhotic IPMN/healthy patients and mildly diseased or severely diseased cirrhotic patients. For this purpose, the same surrogates from the first study are used to define mildly and severe stages of disease: MELD score $<$ or ≥ 9 and compensated or decompensated cirrhosis. The results confirm findings and observations from the first study that could be verified now on an independent test set: radiomic spleen features seem particularly important for distinguishing between the groups of healthy/healthier patients and a cirrhotic cohort. This is an observation of the impact of splenic features that has not yet been made by the research community. Moreover, the prognostic power of radiomic features measuring texture heterogeneity within the liver and spleen from the second study can be confirmed. Together with the results of the first study the results represent a rough severity stratification of liver cirrhosis by identifying prognostic features to a.) generally distinguish between the baseline IPMN cohort and the overall cirrhosis cohort and b.) to further differentiate the features necessary to classify between IPMN and mildly diseased and c.) IPMN and severely diseased patients (see Fig. 3.7 for comparison).

Future work must target acquiring more data by either extending the study dates of the existing cohorts or by expanding from a single to a multi-center study. Preferably, both strategies should be realized concurrently. Advantageous for further data acquisition is the choice of using the widely accepted gold standard MRI protocol for HCC screening, especially, for integrating data from multiple centers. Simultaneously, the extended cohorts must contain a larger range of disease severity with a broader spectrum of MELD scores between 6 and 40 as first suggested in the conclusion of Section 3.3. Moreover, to actually evaluate radiomic-based severity assessment of cirrhosis in direct comparison to the established MELD scoring system (probability of 90 day observed mortality), the

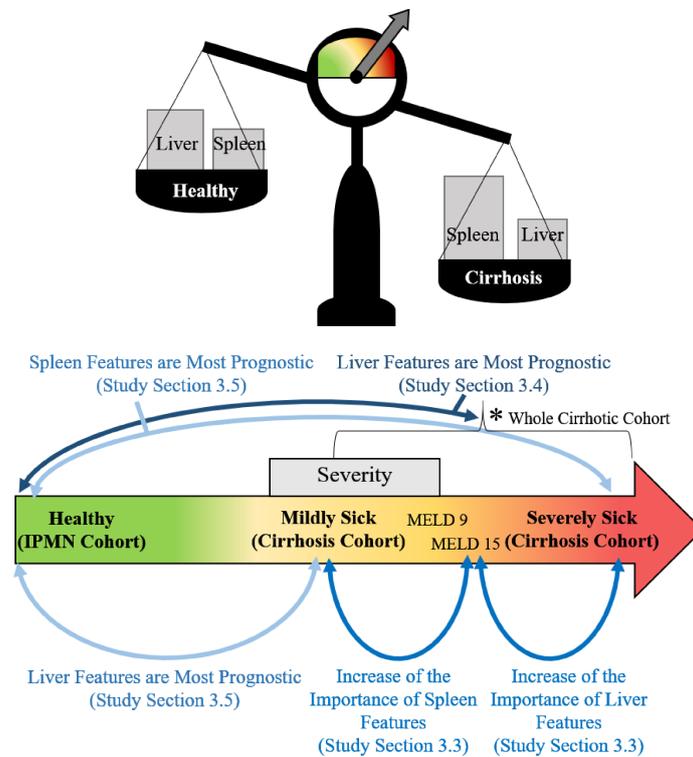


Figure 3.7: Summary of the results from the studies of Part II of this thesis. The arrows in this figure give an overview of exploratory findings of liver and spleen feature importances in the studies in Sections 3.3, 3.4, and 3.5. The arrows point towards the groups that are distinguished. The figure illustrates the importance of liver or spleen features evaluated in each study. In the first study in Section 3.3 it is generally distinguished between mildly diseased and severely diseased cirrhosis patients by using the MELD score and the status of liver decompensation to divide two classes for classification. In the second study in Section 3.4 the radiomic feature importances are further analyzed and it is generally distinguished between the *healthy* non-cirrhotic IPMN cohort (healthy livers and spleens) and the cirrhosis cohort. Further experiments of radiomic feature importances are conducted in Section 3.5 where the classification task was to distinguish between healthy livers and spleens and a.) mildly diseased and b.) severely diseased cirrhosis patients. The scales symbolizes the finding in our exploratory studies that liver and spleen features are both very important for an automatic image-based assessment of cirrhosis. The weights symbolize the general feature importances for this classification task.

radiomic features must also be correlated with patient outcomes, i.e., survival, as it is done in studies that evaluate new parameters for a more precise MELD score [54].

To maximize control in our initial studies, we deliberately chose to use single center data for extracting MRI-based radiomic features to prove the basic concept of our approach of MRI-based severity stratification and cirrhosis detection. Especially, in order to prevent dealing with the challenges of MR imaging of being a less standardized imaging modality. In future studies it has to be evaluated whether the here attained general observations of radiomic liver and spleen feature importances and their transition in progressive cirrhosis can be translated to other MRI scanners from other manufacturers. This implies a profound analysis of feature stability and reproducibility on various scanners with the constraint of using the same standardized HCC screening protocol and magnetic field strength, which seems a plausible and feasible task for the future.

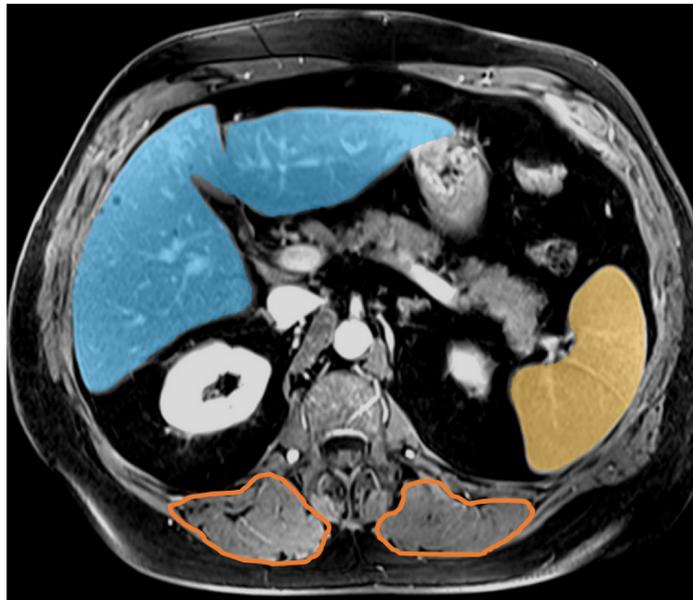


Figure 3.8: This figure shows the back muscles on a contrast-enhanced MRI (delayed phase) from a HCC scanning series. The back muscles are roughly segmented with orange contours. The liver is visualized with a blue and the spleen with a yellow overlay.

In addition to extending the data pool, it should be considered to analyze the impact of cirrhosis on other abdominal organs or structures. As mentioned in Section 1.3.3, other cirrhosis characteristics exist, for instance, sarcopenia, which is the characteristic receding of skeletal muscle mass and a severe complication for cirrhotic patients [167]. The patient basically consumes the body's own proteins and the process is also termed *muscle wasting*. This represents a common symptom for severe chronic diseases with metabolic disorders and/or decreased nutrient intake and uptake [11]. The *erector spinae muscles* are potentially very well suited to include these observations in further analyses. These muscles are clearly visible on the here used MR images and belong to the central skeletal muscles within the lumbar vertebra region (lower back). The muscles could be segmented automatically in future studies on MRI followed by a thickness measurement that could be repeated during follow-up examinations.

This would be similar to the approach suggested by Bridge et al. [11], where the body composition of cancer patients is analyzed on computed tomography images. An example of segmentable back muscles for the purpose in our study for cirrhosis patients can be seen in Fig. 3.8.

Additionally, liver cirrhosis is the most common cause of portal hypertension, which influences the blood flow and causes high blood pressure within the hepatic portal system. Consequently, this also influences the characteristic diameters of the central hepatic vasculature. For instance, the hepatic vein diameter reduction strongly correlates with cirrhosis and its severity, which is considered comparable to the MELD score [167].

Furthermore, it must be evaluated if the five minutes delayed phase is the only relevant sequence for cirrhosis detection and severity assessment within the series of multiparametric MRI scans for HCC screening. The gold standard protocol also suggests the acquisition of an arterial phase and a portal phase fat-suppressed T1-weighted MRI in order to detect disease related abnormalities during different stages of contrast uptake (as mentioned in Section 1.2.1). Therefore, it has to be evaluated whether the multiparametric scans allow a combined analysis to better distinguish between fibrosis, cirrhosis, benign tumors (e.g. hemangiomas), and HCC. Nevertheless, the equilibrium phase is clinically considered the most relevant phase to perceive cirrhotic changes [155].

Considering the acquired knowledge of our initial studies in Part II of this thesis, it is more likely that the specific set of important radiomic features changes during progressive cirrhosis (as can be visually assessed by the exploratory results in Fig. 3.7). Therefore, we assume that the radiomic signature for severity assessment must contain different components detecting and weighing various possible disease stages in order to evaluate

the most probable stage of disease. This would result in a radiomic signature with different components similar to different lab parameters from which the established MELD score is computed.

We believe that our pioneering research in this thesis opens new avenues of applying radiomics for disease stratification of end-stage liver disease patients. In particular, we hope to improve predicting and identifying critical disease stages, such as decompensation or optimal timing to benefit from organ transplantation. We believe that understanding the specific, underlying radiomic representation of the disease yields great potential for improving diagnosis, therapy, monitoring, and fair transplant organ allocation. A vision of the future would include providing a sophisticated, clinical decision support system that embeds radiomics for a standardized automatic assessment of cirrhosis for end-stage liver disease patients.

List of Figures

1.1	Speckles in Ultrasound Imaging	8
1.2	Scanned Ultrasound Image Region Displayed within Preoperative MRI . .	10
1.3	Anatomy of Central Cerebral Structures	12
1.4	Anatomy of Central Cerebral Structures - Ventricles	13
1.5	Comparison of a Cirrhotic and a Non-Cirrhotic Patient on MRI	16
1.6	Random Forest Classification	18
1.7	Structure of a CNN	19
1.8	Illustration of the Principle of an OBIA Representation of Objects	23
1.9	Example of Classifying Objects with OBIA in iUS	24
1.10	The RPDR Data Acquisition Process	27
2.1	Intraoperative Setup during Image-Guided Glioma Surgery.	37
2.2	Image-Guided Glioma Surgery with a Neuronavigation System.	38
2.3	Approach 1: Image Processing Pipeline	44
2.4	Approach 1: Diffusion Stickfilter	45
2.5	Approach 1: Visualization of Segmentation Results	46
2.6	Approach 2: Visualization of the Perifalcine Region	49
2.7	Approach 2: Image Processing Pipeline	51
2.8	Approach 2: Segmentation Results	52
2.9	Approach 5: Model of a Histogram of a T1-weighted MRI of the Brain . . .	56
2.10	Approach 5: Falx and Tentorium Segmentation in preMRI	59
2.11	Approach 3: Perifalcine Region in preMRI and iUS Imaging	61
2.12	Approach 3: Overview of the Segmentation-Based Registration Approach .	62
2.13	Approach 3: Perifalcine Region in iUS Imaging	64
2.14	Approach 3: Pseudo Code Hierarchical Segmentation	65
2.15	Approach 3: Skull Stripping prior Registration	67
2.16	Approach 3: Examples of preMRI and iUS Registration Results	73
2.17	Approach 4: Visualization of Segmentation Results	82
2.18	Approach 4: Visualization of Segmentation Results on MNI Bite Data Sets	83

3.1	Distribution of MELD Scores in Liver Cirrhosis Cohort.	95
3.2	Distribution of Compensated and Decompensated Patients	99
3.3	Overview of the Feature Extraction Process and Performed Experiments . .	101
3.4	Overview of ROC Curves of Experiment 2 - 4	103
3.5	General Overview of the Feature Extraction Process	111
3.6	Confusion Matrices of the Experiments 1 - 4	120
3.7	Summary of Experiments of Radiomic Feature Importances for Cirrhosis .	128
3.8	Back Muscles on Contrast-Enhanced MRI	129

List of Tables

1.1	Pathological Features of Liver Cirrhosis	15
2.1	Overview of Implemented Approaches in Part I	33
2.2	Approach 1: Results	46
2.3	Approach 2: Results	50
2.4	Approach 3: Segmentation Results in iUS Imaging	69
2.5	Approach 3: Skull Stripping Results in preMRI	70
2.6	Approach 3: preMRI and iUS Registration Results	71
2.7	Approach 3: Computation Time of preMRI and iUS Registration Results . .	71
2.8	Approach 3: Inter-Observer Variability of Segmentations in B-mode iUS .	72
2.9	Approach 3: Inter-Observer Variability Comparison to Expert 1	74
2.10	Approach 4: Comparing Segmentation Results	80
3.1	Cirrhosis and Control Cohort's Demographic Information	94
3.2	Classification Results with AUCs and P-values	100
3.3	Most Important Radiomic Features to Classify Cirrhosis	110
3.4	AUCs and Corresponding Standard Deviations of Initial Feature Selection	112
3.5	Importance of Different Feature Classes to Classify Cirrhosis	113
3.6	Classification Accuracy Using Different Radiomic Feature Classes	114
3.7	Importand Radiomic Features for "Healthier" and "Sicker" Patients	121

Bibliography

- [1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. “SLIC Superpixels Compared to State-of-the-Art Superpixel Methods”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.11 (2012), pp. 2274–2282. DOI: <https://doi.org/10.1109/TPAMI.2012.120>.
- [2] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, S. Cavalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. In: *Nature communications* 5 (2014), p. 4006. DOI: <https://doi.org/10.1038/ncomms5006>.
- [3] T. Arbel, X. Morandi, R. Comeau, and D. Collins. “Automatic non-linear MRI-ultrasound registration for the correction of intra-operative brain deformations”. In: *Computer Aided Surgery* 9.4 (2004), pp. 123–136. DOI: <https://doi.org/10.3109/10929080500079248>.
- [4] S. K. Asrani, H. Devarbhavi, J. Eaton, and P. S. Kamath. “Burden of liver diseases in the world”. In: *Journal of hepatology* 70.1 (2019), pp. 151–171. DOI: <https://doi.org/10.1016/j.jhep.2018.09.014>.
- [5] Attwa, M. H. and El-Etreby, S. A. “Guide for diagnosis and treatment of hepatocellular carcinoma”. In: *World Journal of Hepatology* 7.12 (2015), pp. 1632–1651. DOI: <https://doi.org/10.4254/wjh.v7.i12.1632>.
- [6] E. J. van Beek and E. A. Hoffman. “Functional Imaging: CT and MRI”. In: *Clinics in Chest Medicine* 29.1 (2008), pp. 195–216. DOI: <https://doi.org/10.1016/j.ccm.2007.12.003>.
- [7] T. Boehler, D. van Straaten, S. Wirtz, and H.-O. Peitgen. “A robust and extendible framework for medical image registration focused on rapid clinical application deployment”. In: *Computers in biology and medicine* 41.6 (2011), pp. 340–349. DOI: <https://doi.org/10.1016/j.compbiomed.2011.03.011>.

- [8] Brainlab AG Germany. *Software User Guide Rev. 1.0 Ultrasound Navigation Ver. 1.0*. URL: <https://www.brainlab.com/userguides/wp-content/uploads/userguides/1595/Ultrasound-Navigation-version.1.0-revision.1.0-en.pdf>. pp. 20, 35-38. Published: 2018, Accessed: 2020-01-04.
- [9] G. Brancatelli, M. P. Federle, R. Ambrosini, R. Lagalla, A. Carriero, M. Midiri, and V. Vilgrain. "Cirrhosis: CT and MR imaging evaluation". In: *European Journal of Radiology* 61.1 (2007), pp. 57–69. DOI: <https://doi.org/10.1016/j.ejrad.2006.11.003>.
- [10] L. Breiman. "Random Forests". In: *Machine Learning* 45.1 (2001), pp. 5–32. DOI: <https://doi.org/10.1023/A:1010933404324>.
- [11] C. P. Bridge, M. Rosenthal, B. Wright, G. Kotecha, F. Fintelmann, F. Troschel, N. Miskin, K. Desai, W. Wrobel, A. Babic, N. Khalaf, L. Brais, M. Welch, C. Zellers, N. Tenenholtz, M. Michalski, B. Wolpin, and K. Andriole. "Fully-Automated Analysis of Body Composition from CT in Cancer Patients Using Convolutional Neural Networks". In: *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Ed. by D. Stoyanovet et al. Cham: Springer International Publishing, 2018, pp. 204–213. ISBN: 978-3-030-01201-4. DOI: https://doi.org/10.1007/978-3-030-01201-4_22.
- [12] M. Bucki, O. Palombi, M. Baillet, and Y. Payan. "Doppler Ultrasound Driven Biomechanical Model of the Brain for Intraoperative Brain-Shift Compensation: A Proof of Concept in Clinical Conditions". In: *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*. Ed. by Y. Payan. Springer Berlin Heidelberg, 2012, pp. 135–165. ISBN: 978-3-642-29014-5. DOI: https://doi.org/10.1007/8415_2012_119.
- [13] W. Burger and M. J. Burge. *Digitale Bildverarbeitung : eine algorithmische Einführung mit Java*. 3rd. X.media.press. Berlin: Springer Vieweg, 2015. ISBN: 9783642046032. DOI: <https://doi.org/10.1007/978-3-642-04604-9>.
- [14] S. Campbell and A. Thoms. "Ultrasound Measurement of the Fetal Head to Abdomen Circumference Ratio in the Assessment of Growth Retardation". In: *BJOG: An International Journal of Obstetrics & Gynaecology* 84.3 (1977), pp. 165–174. DOI: <https://doi.org/10.1111/j.1471-0528.1977.tb12550.x>.
- [15] W. Che and Y. Zhang. "Deep Learning in Lexical Analysis and Parsing". In: *Deep Learning in Natural Language Processing*. Ed. by L. Deng and Y. Liu. Singapore: Springer Singapore, 2018, p. 94. ISBN: 978-981-10-5208-8. DOI: <https://doi.org/10.1007/978-981-10-5209-5>.

-
- [16] X. Chen, K. Oshima, D. Schott, H. Wu, W. Hall, Y. Song, Y. Tao, D. Li, C. Zheng, P. Knechtges, B. Erickson, and X. A. Li. "Assessment of treatment response during chemoradiation therapy for pancreatic cancer based on quantitative radiomic analysis of daily CTs: An exploratory study". In: *PloS one* 12.6 (2017), e0178961. DOI: <https://doi.org/10.1371/journal.pone.0178961>.
- [17] X. Chen. "Chapter 9 - Multimodal Image-Guided Brain Tumor Resection". In: *Image-Guided Neurosurgery*. Ed. by Alexandra J. Golby. Boston: Academic Press, 2015, pp. 213–214. ISBN: 978-0-12-800870-6. DOI: <https://doi.org/10.1016/B978-0-12-800870-6.00009-1>.
- [18] G. Chlebus, H. Meine, N. Abolmaali, and A. Schenk. "Automatic Liver and Tumor Segmentation in Late-Phase MRI Using Fully Convolutional Neural Networks". In: *Proceedings of CURAC* (2018), pp. 195–200.
- [19] K. J. Choi, J. K. Jang, S. S. Lee, Y. S. Sung, W. H. Shim, H. S. Kim, J. Yun, J.-Y. Choi, Y. Lee, B.-K. Kang, J. H. Kim, S. Y. Kim, and E. S. Yu. "Development and Validation of a Deep Learning System for Staging Liver Fibrosis by Using Contrast Agent-enhanced CT Images in the Liver". In: *Radiology* (2018), p. 180763. DOI: <https://doi.org/10.1148/radiol.2018180763>.
- [20] E. B. Claus and P. M. Black. "Survival rates and patterns of care for patients diagnosed with supratentorial low-grade gliomas". In: *Cancer* 106.6 (2006), pp. 1358–1363. DOI: <https://doi.org/10.1002/cncr.21733>.
- [21] J. Coburger, A. Merkel, M. Scherer, F. Schwartz, F. Gessler, C. Roder, A. Pala, R. König, L. Bullinger, G. Nagel, C. Jungk, S. Bisdas, A. Nabavi, O. Ganslandt, V. Seifert, M. Tatagiba, C. Senft, M. Mehdorn, A. W. Unterberg, K. Rössler, and C. R. Wirtz. "Low-grade Glioma Surgery in Intraoperative Magnetic Resonance Imaging: Results of a Multicenter Retrospective Assessment of the German Study Group for Intraoperative Magnetic Resonance Imaging". In: *Neurosurgery* 78.6 (2016), pp. 775–786. DOI: <https://doi.org/10.1227/NEU.0000000000001081>.
- [22] P. Coupe, P. Hellier, X. Morandi, and C. Barillot. "3D Rigid Registration of Intraoperative Ultrasound and Preoperative MR Brain Images Based on Hyperechogenic Structures". In: *International journal of biomedical imaging* 2012 (2012), p. 531319. DOI: <https://doi.org/10.1155/2012/531319>.
- [23] R. S. D'Amico, Z. K. Englander, P. Canoll, and J. N. Bruce. "Extent of Resection in Glioma—A Review of the Cutting Edge". In: *World Neurosurgery* 103 (2017), pp. 538–549. DOI: <https://doi.org/10.1016/j.wneu.2017.04.041>.
-

- [24] M. Del Bene, A. Perin, C. Casali, F. Legnani, A. Saladino, L. Mattei, I. G. Vetrano, M. Saini, F. DiMeco, and F. Prada. “Advanced Ultrasound Imaging in Glioma Surgery: Beyond Gray-Scale B-mode”. In: *Frontiers in Oncology* 8 (2018), p. 576. DOI: <https://doi.org/10.3389/fonc.2018.00576>.
- [25] DIAG Research Software Engineering, Diagnostic Image Analysis Group Radboud University Medical Center, Nijmegen, Netherlands, Fraunhofer MEVIS, Bremen, Germany. *Grand Challenges in Biomedical Image Analysis*. URL: <https://grand-challenge.org/>. Accessed: 2019-17-12.
- [26] L. R. Dice. “Measures of the Amount of Ecologic Association Between Species”. In: *Ecology* 26.3 (1945), pp. 297–302. DOI: <https://doi.org/10.2307/1932409>.
- [27] C. E. Duchon. “Lanczos Filtering in One and Two Dimensions”. In: *Journal of Applied Meteorology* 18.8 (1979), pp. 1016–1022. DOI: [https://doi.org/10.1175/1520-0450\(1979\)018<1016:LFI0AT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<1016:LFI0AT>2.0.CO;2).
- [28] Eurotransplant International Foundation. *ET Liver Allocation System (ELAS) - Eurotransplant Manual – version 4.0*; URL: http://www.eurotransplant.org/cms/mediaobject.php?file=chapter5_elas9.pdf. pp. 10-13. Published: 2013-03-08, Accessed: 2020-01-04.
- [29] P. Farnia, A. Ahmadian, T. Shabaniyan, N. D. Serej, and J. Alirezaie. “Brain-shift compensation by non-rigid registration of intra-operative ultrasound images with preoperative MR images based on residual complexity”. In: *International journal of computer assisted radiology and surgery* 10.5 (2015), pp. 555–562. DOI: <https://doi.org/10.1007/s11548-014-1098-5>.
- [30] A. Fedorov, P. Risholm, and W. Wells. “Deformable Registration for IGT”. In: *Intraoperative Imaging and Image-Guided Therapy*. Ed. by F. A. Jolesz. New York, NY: Springer New York, 2014, p. 219. DOI: https://doi.org/10.1007/978-1-4614-7657-3_14.
- [31] D. L. Felten, F. H. Netter, M. K. O’Banion, and M. S. Maida. *Netter’s Atlas of Neuroscience*. 3rd. Philadelphia: Elsevier, 2016, pp. 49, 61, 64, 86, 90–91. ISBN: 9780323265119.
- [32] Figure Eight Inc. *CrowdFlower Data Science Report 2016*. URL: https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf. Accessed: 2019-12-27.

-
- [33] S. Fritz, Küper-Steffen, R., K. Feilhauer, C. M. Sommer, G. M. Richter, R. Hennig, and J. Köninger. “Resection of benign side-branch intraductal papillary mucinous neoplasm of the pancreas-is long term follow-up indicated?: A case report and review of the literature”. In: *Medicine* 97.7 (2018), e9894. DOI: <https://doi.org/10.1097/MD.0000000000009894>.
- [34] R. L. Galloway. “Chapter 1 - Introduction and Historical Perspectives on Image-Guided Surgery”. In: *Image-Guided Neurosurgery*. Ed. by A. J. Golby. Boston: Academic Press, 2015, pp. 1, 10–11, 14, 16. ISBN: 978-0-12-800870-6. DOI: <https://doi.org/10.1016/B978-0-12-800870-6.00001-7>.
- [35] M. L. Gandía-González, S. Cerdán, L. Barrios, P. López-Larrubia, P. G. Feijoó, A. Palpan Jr., J. M. Roda, and J. Solivera. “Assessment of Overall Survival in Glioma Patients as Predicted by Metabolomic Criteria”. In: *Frontiers in Oncology* 9 (2019), p. 328. DOI: <https://doi.org/10.3389/fonc.2019.00328>.
- [36] M. Geurts and M. J. van den Bent. “On high-risk, low-grade glioma: What distinguishes high from low?” In: *Cancer* 125.2 (2019), pp. 174–176. DOI: <https://doi.org/10.1002/cncr.31834>.
- [37] D. G. Gobbi, R. M. Comeau, and T. M. Peters. “Ultrasound/MRI Overlay with Image Warping for Neurosurgery”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*. Ed. by S. L. Delp, A. M. DiGoia, and B. Jaramaz. Vol. 1935. Lecture notes in computer science. Springer Berlin Heidelberg, 2000, pp. 106–114. ISBN: 978-3-540-40899-4. DOI: https://doi.org/10.1007/978-3-540-40899-4_11.
- [38] R. Goldbrunner, M. Ruge, M. Kocher, C. W. Lucas, N. Galldiks, and S. Grau. “The Treatment of Gliomas in Adulthood”. In: *Dtsch Arztebl International* 115.20-21 (2018), pp. 356–364. DOI: <https://doi.org/10.3238/arztebl.2018.0356>.
- [39] V. P. Grover, J. M. Tognarelli, M. M. Crossey, and I. J. Cox. “Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians”. In: *Journal of Clinical and Experimental Hepatology* 5.3 (2015), pp. 246–255. DOI: <https://doi.org/10.1016/j.jceh.2015.08.001>.
- [40] G. Hackeling. *Mastering Machine Learning with scikit-learn*. Packt Publishing, 2014. ISBN: 9781783988372.

- [41] H. K. Hahn and H.-O. Peitgen. "The Skull Stripping Problem in MRI Solved by a Single 3D Watershed Transform". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*. Ed. by S. L. Delp, A. M. DiGioia, and B. Jaramaz. Vol. 1935. Lecture notes in computer science. Springer Berlin Heidelberg, 2000, pp. 134–143. ISBN: 978-3-540-40899-4. DOI: https://doi.org/10.1007/978-3-540-40899-4_14.
- [42] H. K. Hahn and H.-O. Peitgen. "IWT-interactive watershed transform: a hierarchical method for efficient interactive and automated segmentation of multidimensional grayscale images". In: *SPIE Medical Imaging, Proc in SPIE 5032* (2003). Ed. by M. Sonka and J. M. Fitzpatrick, pp. 643–653. DOI: <https://doi.org/10.1117/12.481097>.
- [43] H. Handels. *Medizinische Bildverarbeitung : Bildanalyse, Mustererkennung und Visualisierung für die computergestützte ärztliche Diagnostik und Therapie*. 2nd ed. Vieweg und Teubner Studium. Wiesbaden: Vieweg und Teubner, 2009, pp. 74–75, 94. ISBN: 978-3-8351-0077-0. DOI: <https://doi.org/10.1007/978-3-8348-9571-4>.
- [44] N. C. Himes and G. Young. "Chapter 2 - Background on Imaging Structural Imaging". In: *Image-Guided Neurosurgery*. Ed. by A. J. Golby. Boston: Academic Press, 2015, pp. 25–61, 31–36. ISBN: 978-0-12-800870-6. DOI: <https://doi.org/10.1016/B978-0-12-800870-6.00002-9>.
- [45] E. C. Holland. "Progenitor cells and glioma formation". In: *Current Opinion in Neurology* 14 (2001), pp. 683–688. DOI: <https://doi.org/10.1097/00019052-200112000-00002>.
- [46] A. Homeyer, M. Schwier, and H. K. Hahn. "A Generic Concept for Object-Based Image Analysis". In: *VISIGRAPP 2010, International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. 2010, pp. 530–533. ISBN: 978-989-674-028-3.
- [47] F. Inc. *Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says*. URL: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. Published: 2016-03-23, Accessed: 2019-12-23.
- [48] K. Ito, D. G. Mitchell, T. Gabata, and S. M. Hussain. "Expanded Gallbladder Fossa: Simple MR Imaging Sign of Cirrhosis". In: *Radiology* 211.3 (1999). DOI: <https://doi.org/10.1148/radiology.211.3.r99ma31723>.

-
- [49] F. A. Jolesz. "Intraoperative imaging in neurosurgery: where will the future take us?" In: *Acta neurochirurgica. Supplement* 109 (2011), pp. 21–25. DOI: https://doi.org/10.1007/978-3-211-99651-5_4.
- [50] F. A. Jolesz. "Foreword". In: *Image-Guided Neurosurgery*. Ed. by A. J. Golby. Boston: Academic Press, 2015, p. xiv. ISBN: 978-0-12-800870-6. DOI: <https://doi.org/10.1016/B978-0-12-800870-6.00028-5>.
- [51] P. Kalavathi and V. B. Prasath. "Methods on Skull Stripping of MRI Head Scan Images-a Review". In: *Journal of Digital Imaging* 29.3 (2016), pp. 365–379. DOI: <https://doi.org/10.1007/s10278-015-9847-8>.
- [52] P. S. Kamath and W. R. Kim. "The model for end-stage liver disease (MELD)". In: *Hepatology* 45.3 (2007), pp. 797–805. DOI: <https://doi.org/10.1002/hep.21563>.
- [53] P. S. Kamath, R. H. Wiesner, M. Malinchoc, W. Kremers, T. M. Therneau, C. L. Kosberg, G. D'Amico, E. R. Dickson, and W. R. Kim. "A model to predict survival in patients with end-stage liver disease". In: *Hepatology* 33.2 (2001), pp. 464–470. DOI: <https://doi.org/10.1053/jhep.2001.22172>.
- [54] U. Kartoun, K. E. Corey, T. G. Simon, H. Zheng, R. Aggarwal, K. Ng, and S. Y. Shaw. "The MELD-Plus: A generalizable prediction risk score in cirrhosis". In: *PLOS ONE* 12.10 (2017), e0186301. DOI: <https://doi.org/10.1371/journal.pone.0186301>.
- [55] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach. "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success". In: *British Medical Journal* 330.7494 (2005), p. 765. DOI: <https://doi.org/10.1136/bmj.38398.500764.8F>.
- [56] W. R. Kim, S. W. Biggins, W. K. Kremers, R. H. Wiesner, P. S. Kamath, J. T. Benson, E. Edwards, and T. M. Therneau. "Hyponatremia and mortality among patients on the liver-transplant waiting list". In: *The New England journal of medicine* 359.10 (2008), pp. 1018–1026. DOI: <https://doi.org/10.1056/NEJMoa0801209>.
- [57] A. P. King, J. M. Blackall, G. P. Penney, P. J. Edwards, D. L. G. Hill, and D. J. Hawkes. "Bayesian Estimation of Intra-operative Deformation for Image-Guided Surgery Using 3-D Ultrasound". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000*. Ed. by S. L. Delp, A. M. DiGoia, and B. Jaramaz. Vol. 1935. Lecture notes in computer science. Springer Berlin Heidelberg, 2000, pp. 588–597. ISBN: 978-3-540-40899-4. DOI: https://doi.org/10.1007/978-3-540-40899-4_60.
-

- [58] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller. "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping". In: *NeuroImage* 129 (2016), pp. 460–469. DOI: <https://doi.org/10.1016/j.neuroimage.2016.01.024>.
- [59] H. Kolem. "Magnetresonanztomographie (MRT)". In: *Medizintechnik: Verfahren Systeme Informationsverarbeitung*. Ed. by R. Kramme. Springer Berlin Heidelberg, 2007, pp. 271–274. ISBN: 978-3-540-34103-1. DOI: <https://doi.org/10.1007/978-3-540-34103-1>.
- [60] B. Köstering and H. Dudwiesus. "Ultraschalldiagnostik". In: *Medizintechnik: Verfahren Systeme Informationsverarbeitung*. Ed. by R. Kramme. Springer Berlin Heidelberg, 2007, pp. 311–315. ISBN: 978-3-540-34103-1. DOI: <https://doi.org/10.1007/978-3-540-34103-1>.
- [61] F. W. Kremkau and K. J. Taylor. "Artifacts in ultrasound imaging." In: *Journal of Ultrasound in Medicine* 5.4 (1986), pp. 227–237. DOI: <https://doi.org/10.7863/jum.1986.5.4.227>.
- [62] M. Kuklisova-Murgasova, A. Cifor, R. Napolitano, A. Papageorghiou, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, J. A. Noble, and J. A. Schnabel. "Registration of 3D fetal neurosonography and MRI". In: *Medical Image Analysis* 17.8 (2013), pp. 1137–1150. DOI: <https://doi.org/10.1016/j.media.2013.07.004>.
- [63] L. Mercier, D. Araujo, C. Haegelen, R.F. Del Maestro, K. Petrecca and D.L. Collins. "Registering Pre- and Postresection 3-Dimensional Ultrasound for Improved Visualization of Residual Brain Tumor". In: *Ultrasound in Medicine & Biology* 39 (2013), pp. 16–29. DOI: <https://doi.org/10.1016/j.ultrasmedbio.2012.08.004>.
- [64] J. Larsen, S. B. Wharton, F. McKeivitt, C. Romanowski, C. Bridgewater, H. Zaki, and N. Hoggard. "Low grade glioma: an update for radiologists". In: *The British Journal of Radiology* 90.1070 (2017), p. 20160600. DOI: <https://doi.org/10.1259/bjr.20160600>.
- [65] B. Lassen, J. Kuhnigk, O. Friman, S. Krass, and H. Peitgen. "Automatic segmentation of lung lobes in ct images based on fissures, vessels, and bronchi". In: *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2010, pp. 560–563. DOI: <https://doi.org/10.1109/ISBI.2010.5490284>.
- [66] Y. LeCun, Y. Bengio, and G. Hinton. "Deep Learning". In: *Nature* 521 (2015), pp. 436–444. DOI: <https://doi.org/10.1038/nature14539>.

- [67] M. M. J. Letteboer, P. W. A. Willems, M. A. Viergever, and W. J. Niessen. "Non-rigid Registration of 3D Ultrasound Images of Brain Tumours Acquired during Neurosurgery". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*. Ed. by R. E. Ellis and T. M. Peters. Lecture notes in computer science. Springer Berlin Heidelberg, 2003, pp. 408–415. ISBN: 978-3-540-39903-2. DOI: https://doi.org/10.1007/978-3-540-39903-2_50.
- [68] T. Lisman, Y. Van Leeuwen, J. Adelmeijer, I. T. A. Pereboom, E. B. Haagsma, A. P. Van Den Berg, and R. J. Porte. "Interlaboratory variability in assessment of the model of end-stage liver disease score". In: *Liver International* 28.10 (2008), pp. 1344–1351. DOI: <https://doi.org/10.1111/j.1478-3231.2008.01783.x>.
- [69] I. Machado, M. Toews, J. Luo, P. Unadkat, W. Essayed, E. George, P. Teodoro, H. Carvalho, J. Martins, P. Golland, S. Pieper, S. Frisken, A. Golby, and W. Wells. In: *International Journal of Computer Assisted Radiology and Surgery* 13.10 (2018), pp. 1525–1538. DOI: <https://doi.org/10.1007/s11548-018-1786-7>.
- [70] "Magnetic resonance imaging of the cirrhotic liver: Diagnosis of hepatocellular carcinoma and evaluation of response to treatment - Part 1". In: *Radiologia brasileira* 50.1 (2017), pp. 38–47. DOI: <https://doi.org/10.1590/0100-3984.2015.0132>.
- [71] A. Mainio, S. Tuunanen, H. Hakko, A. Niemelä, J. Koivukangas, and P. Räsänen. "Decreased quality of life and depression as predictors for shorter survival among patients with low-grade gliomas: a follow-up from 1990 to 2003". In: *European Archives of Psychiatry and Clinical Neuroscience* 256 (2006), pp. 516–521. DOI: <https://doi.org/10.1007/s00406-006-0674-2>.
- [72] G. Mamone and R. Miraglia. "The "expanded gallbladder fossa sign" in liver cirrhosis". In: *Abdominal Radiology* 44.3 (2019), pp. 1199–1200. DOI: <https://doi.org/10.1007/s00261-018-1878-4>.
- [73] P. Martin, A. DiMartini, S. Feng, R. Brown, and M. Fallon. "Evaluation for liver transplantation in adults: 2013 practice guideline by the American Association for the Study of Liver Diseases and the American Society of Transplantation". In: *Hepatology* 59.3 (2014), pp. 1144–1165. ISSN: 02709139. DOI: <https://doi.org/10.1002/hep.26972>.
- [74] A. P. Matos, F. Velloni, M. Ramalho, M. AlObaidy, A. Rajapaksha, and R. C. Semelka. "Focal liver lesions: Practical magnetic resonance imaging approach". In: *World Journal of Hepatology* 7.16 (2015), pp. 1987–2008. DOI: <https://doi.org/10.4254/wjh.v7.i16.1987>.

- [75] D. McRobbie, E. Moore, M. Graves, and M. Prince. *MRI from Picture to Proton*. Cambridge University Press, 2007, pp. 13–14, 30–41. ISBN: 978-1-139-45719-4.
- [76] L. Mercier, R. F. Del Maestro, K. Petrecca, D. Araujo, C. Haegelen, and D. L. Collins. “Online database of clinical MR and ultrasound images of brain tumors”. In: *Medical Physics* 39 (2012), pp. 3253–3261. DOI: <https://doi.org/10.1118/1.4709600>.
- [77] L. Mercier, V. Fonov, C. Haegelen, R. F. Del Maestro, K. Petrecca, and D. L. Collins. “Comparing two approaches to rigid registration of three-dimensional ultrasound and magnetic resonance images for neurosurgery”. In: *International journal of computer assisted radiology and surgery* 7.1 (2012), pp. 125–136. DOI: <https://doi.org/10.1007/s11548-011-0620-2>.
- [78] U. Mezger, C. Jendrewski, and M. Bartels. “Navigation in Surgery”. In: *Langenbeck's Archives of Surgery* 398 (2013), pp. 501–514. DOI: <https://doi.org/10.1007/s00423-013-1059-4>.
- [79] J. Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, 2003. ISBN: 9780198528418. DOI: <https://doi.org/10.1093/acprof:oso/9780198528418.001.0001>.
- [80] J. Modersitzki. *FAIR: Flexible algorithms for image registration*. Vol. FA06. Fundamentals of algorithms. Society for Industrial and Applied Mathematics (SIAM), 2009. ISBN: 978-0-89871-690-0.
- [81] A. V. Moiyadi. “Intraoperative Ultrasound Technology in Neuro-Oncology Practice-Current Role and Future Applications”. In: *World neurosurgery* 93 (2016), pp. 81–93. DOI: <https://doi.org/10.1016/j.wneu.2016.05.083>.
- [82] A. V. Moiyadi, P. Shetty, and A. Degaonkar. “Resection of Pediatric Brain Tumors: Intraoperative Ultrasound Revisited”. In: *Journal of pediatric neurosciences* 12.1 (2017), pp. 19–23. DOI: https://doi.org/10.4103/jpn.JPN_141_16.
- [83] A. A. Mokdad, A. D. Lopez, S. Shahraz, R. Lozano, A. H. Mokdad, J. Stanaway, C. J. L. Murray, and M. Naghavi. “Liver cirrhosis mortality in 187 countries between 1980 and 2010: A systematic analysis”. In: *BMC medicine* 12 (2014), p. 145. DOI: <https://doi.org/10.1186/s12916-014-0145-y>.
- [84] M. H. Mozaffari and W. Lee. *3D Ultrasound image segmentation: A Survey*. 2016. arXiv: 1611.09811 [cs.CV].

-
- [85] K. F. Murray and R. L. Carithers. "AASLD practice guidelines: Evaluation of the patient for liver transplantation". In: *Hepatology (Baltimore, Md.)* 41.6 (2005), pp. 1407–1432. DOI: <https://doi.org/10.1002/hep.20704>.
- [86] R. Nalichowski, D. Keogh, H. C. Chueh, and S. N. Murphy. "Calculating the benefits of a Research Patient Data Repository". In: *AMIA Annual Symposium proceedings. AMIA Symposium (2006)*, p. 1044.
- [87] National Electrical Manufacturers Association (NEMA). *Dicom Standard*. URL: <https://www.dicomstandard.org/>. Accessed: 2019-12-27.
- [88] National Institute of Biomedical Imaging and Bioengineering. *Magnetic Resonance Imaging (MRI)*. URL: <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri>. Accessed: 2020-01-04.
- [89] T. Netsch, P. Rosch, A. van Muiswinkel, and J. Weese. "Towards real-time multi-modality 3-D medical image registration". In: *Eighth IEEE International Conference on Computer Vision*. IEEE Computer Society, 2001, pp. 718–725. DOI: <https://doi.org/10.1109/ICCV.2001.937595>.
- [90] J. Nitsch, J. Klein, P. Dammann, K. Wrede, U. Sure, and H. Hahn. "Automatic Segmentation of Perifalcine Structures in Two-Dimensional Ultrasound Images Using Object-Based Image Analysis". In: *Proceedings of CURAC (2015)*, pp. 219–223.
- [91] J. Nitsch, J. Klein, P. Dammann, K. Wrede, O. Gembruch, J. H. Moltz, H. Meine, U. Sure, R. Kikinis, and D. Miller. "Automatic and efficient MRI-US segmentations for improving intraoperative image fusion in image-guided neurosurgery". In: *NeuroImage: Clinical* 22 (2019), p. 101766. DOI: <https://doi.org/10.1016/j.nicl.2019.101766>.
- [92] J. Nitsch, J. Klein, H. K. Hahn, K. Wrede, U. Sure, and D. Miller. "Automatic segmentation of the falx, tentorium and adjacent gyri in two-dimensional ultrasound images". In: *Deutsche Gesellschaft für Neurochirurgie. 66. Jahrestagung der Deutschen Gesellschaft für Neurochirurgie (DGNC)*. Düsseldorf: German Medical Science GMS Publishing House, 2015, DocP 163, p. 249. DOI: <https://doi.org/10.3205/15dgnc561>.
- [93] J. Nitsch, J. Klein, D. Miller, U. Sure, and H. K. Hahn. "Automatic Segmentation of the Cerebral Falx and Adjacent Gyri in 2D Ultrasound Images". In: *Proc BVM (2015)*. Ed. by T. Deserno, H. Handels, H.-P. Meinzer, and T. Tolxdorff, pp. 287–292. DOI: https://doi.org/10.1007/978-3-662-46224-9_50.
-

- [94] J. Nitsch, J. Klein, J. H. Moltz, D. Miller, U. Sure, R. Kikinis, and H. Meine. “Neural-network-based automatic segmentation of cerebral ultrasound images for improving image-guided neurosurgery”. In: *SPIE Medical Imaging, Proc in SPIE* 10951 (2019). DOI: <https://doi.org/10.1117/12.2512289>.
- [95] R. Novelline. *Squire’s Fundamentals of Radiology*. Vol. 5. Harvard University Press, 1997, pp. 34–35. ISBN: 978-0-674-83339-5.
- [96] Organ Procurement and Transplantation Network and United Network for Organ Sharing. *Median MELD at Transplant by Donor Service Area and Median PELD at Transplant Within the Nation*. URL: https://optn.transplant.hrsa.gov/media/2971/mts_dsa_distribution_20190311.pdf. Published: 2019-03-11, Accessed: 2020-01-04.
- [97] Organ Procurement and Transplantation Network and United Network for Organ Sharing. *Policy Note: Clerical Changes for Implementation of Adding Serum Sodium to the MELD Score*. URL: https://optn.transplant.hrsa.gov/media/1575/policynotice_20151101.pdf. pp. 4-5. Published: 2015-11-01, Accessed: 2020-01-04.
- [98] D. A. Orringer, A. Golby, and F. Jolesz. “Neuronavigation in the surgical management of brain tumors: current and future trends”. In: *Expert review of medical devices* 9.5 (2012), pp. 491–500. DOI: <https://doi.org/10.1586/erd.12.42>.
- [99] Q. T. Ostrom, G. Cioffi, H. Gittleman, N. Patil, K. Waite, C. Kruchko, and J. S. Barnholtz-Sloan. “CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016”. In: *Neuro-Oncology* 21.Supplement_5 (Nov. 2019), pp. v1–v100. DOI: <https://doi.org/10.1093/neuonc/noz150>.
- [100] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts. “Machine Learning methods for Quantitative Radiomic Biomarkers”. In: *Scientific Reports* 5 (2015), p. 13087. DOI: <https://doi.org/10.1038/srep13087>.
- [101] C. Parmar, P. Grossmann, D. Rietveld, M. M. Rietbergen, P. Lambin, and H. J. W. L. Aerts. “Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer”. In: *Frontiers in oncology* 5 (2015), p. 272. DOI: <https://doi.org/10.3389/fonc.2015.00272>.
- [102] K. Patel, M. Friedrich-Rust, Y. Lurie, M. Grigorescu, C. Stanciu, C.-M. Lee, E. R. Schiff, D. Häussinger, M. P. Manns, G. Gerken, I. Colle, M. Torbenson, E. Pulkstenis, G. M. Subramanian, J. G. McHutchison, and S. Zeuzem. “FibroSURE and FibroScan

- in relation to treatment response in chronic hepatitis C virus". In: *World Journal of Gastroenterology* 17.41 (2011), pp. 4581–4589. DOI: <https://doi.org/10.3748/wjg.v17.i41.4581>.
- [103] Y. Peng, X. Qi, and X. Guo. "Child-Pugh Versus MELD Score for the Assessment of Prognosis in Liver Cirrhosis: A Systematic Review and Meta-Analysis of Observational Studies". In: *Medicine* 95.8 (2016), e2877. DOI: <https://doi.org/10.1097/MD.0000000000002877>.
- [104] A. K. Petridis, M. Anokhin, J. Vavruska, M. Mahvash, and M. Scholz. "The value of intraoperative sonography in low grade glioma surgery". In: *Clinical neurology and neurosurgery* 131 (2015), pp. 64–68. DOI: <https://doi.org/10.1016/j.clineuro.2015.02.004>.
- [105] R. A. Pooley. "Fundamental Physics of MR Imaging". In: *RadioGraphics* 25.4 (2005), pp. 1087–1099. DOI: <https://doi.org/10.1148/rg.254055027>.
- [106] B. C. Porter, D. J. Rubens, J. G. Strang, J. Smith, S. Totterman, and K. J. Parker. "Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial markers". In: *IEEE transactions on medical imaging* 20.4 (2001), pp. 354–359. DOI: <https://doi.org/10.1109/42.921484>.
- [107] F. Prada, M. Del Bene, L. Mattei, L. Lodigiani, S. DeBene, V. Kolev, I. Vetrano, L. Solbiati, G. Sakas, and F. DiMeco. "Preoperative magnetic resonance and intraoperative ultrasound fusion imaging for real-time neuronavigation in brain tumor surgery". In: *Ultraschall in der Medizin (Stuttgart, Germany : 1980)* 36.2 (2015), pp. 174–186. DOI: <https://doi.org/10.1055/s-0034-1385347>.
- [108] F. Prada, M. D. Bene, R. Fornaro, I. G. Vetrano, A. Martegani, L. Aiani, L. M. Sconfienza, G. Mauri, L. Solbiati, B. Pollo, and F. DiMeco. "Identification of residual tumor with intraoperative contrast-enhanced ultrasound during glioblastoma resection". In: *Neurosurgical focus* 40.3 (2016), E7. DOI: <https://doi.org/10.3171/2015.11.FOCUS15573>.
- [109] F. Prada, M. Del Bene, A. Moiraghi, C. Casali, F. G. Legnani, A. Saladino, A. Perin, I. G. Vetrano, L. Mattei, C. Richetta, M. Saini, and F. DiMeco. "From Grey Scale B-Mode to Elastosonography: Multimodal Ultrasound Imaging in Meningioma Surgery-Pictorial Essay and Literature Review". In: *BioMed research international* 2015 (2015), p. 925729. DOI: <https://doi.org/10.1155/2015/925729>.

- [110] B. Presles, M. Fargier-Voiron, M.-C. Biston, R. Lynch, A. Munoz, H. Liebgott, P. Pommier, S. Rit, and D. Sarrut. "Semiautomatic registration of 3D transabdominal ultrasound images for patient repositioning during postprostatectomy radiotherapy". In: *Medical physics* 41.12 (2014), p. 122903. DOI: <https://doi.org/10.1118/1.4901642>.
- [111] R. N. H. Pugh, I. M. Murray-Lyon, J. L. Dawson, M. C. Pietroni, and R. Williams. "Transection of the oesophagus for bleeding oesophageal varices". In: *BJS (British Journal of Surgery)* 60.8 (1973), pp. 646–649. DOI: <https://doi.org/10.1002/bjs.1800600817>.
- [112] M. Quante, C. Benckert, A. Thelen, and S. Jonas. "Experience Since MELD Implementation: How Does the New System Deliver?" In: *International Journal of Hepatology* 2012.264015 (2012). DOI: <https://doi.org/10.1155/2012/264015>.
- [113] J. Rackerseder, M. Baust, R. Göbl, N. Navab, and C. Hennemersperger. "Initialize globally before acting locally: Enabling Landmark-free 3D US to MRI Registration". In: *21st International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI)* (2018).
- [114] I. Reinertsen, F. Lindseth, G. Unsgaard, and D. L. Collins. "Clinical validation of vessel-based registration for correction of brain-shift". In: *Medical Image Analysis* 11.6 (2007), pp. 673–684. DOI: <https://doi.org/10.1016/j.media.2007.06.008>.
- [115] I. Reinertsen, F. Lindseth, C. Askeland, D. H. Iversen, and G. Unsgard. "Intraoperative correction of brain-shift". In: *Acta neurochirurgica* 156.7 (2014), pp. 1301–1310. DOI: <https://doi.org/10.1007/s00701-014-2052-6>.
- [116] M. Renovanz, A.-K. Hickmann, C. Henkel, M. Nadji-Ohl, and N. J. Hopf. "Navigated versus non-navigated intraoperative ultrasound: Is there any impact on the extent of resection of high-grade gliomas? A retrospective clinical analysis". In: *Journal of neurological surgery. Part A, Central European neurosurgery* 75.3 (2014), pp. 224–230. DOI: <https://doi.org/10.1055/s-0033-1356486>.
- [117] M. Reuter, N. J. Schmansky, H. Rosas, and B. Fischl. "Within-subject template estimation for unbiased longitudinal image analysis". In: *Neuroimage* 61.4 (2012), pp. 1402–1418. DOI: <https://doi.org/10.1016/j.neuroimage.2012.02.084>. URL: <https://surfer.nmr.mgh.harvard.edu/>.
- [118] H. Rivaz, S. J.-S. Chen, and D. L. Collins. "Automatic deformable MR-ultrasound registration for image-guided neurosurgery". In: *IEEE transactions on medical imaging* 34.2 (2015), pp. 366–380. DOI: <https://doi.org/10.1109/TMI.2014.2354352>.

-
- [119] H. Rivaz and D. L. Collins. "Deformable registration of preoperative MR, pre-resection ultrasound, and post-resection ultrasound images of neurosurgery". In: *International journal of computer assisted radiology and surgery* 10.7 (2015), pp. 1017–1028. DOI: <https://doi.org/10.1007/s11548-014-1099-4>.
- [120] H. Rivaz and D. L. Collins. "Near real-time robust non-rigid registration of volumetric ultrasound images for neurosurgery". In: *Ultrasound in Medicine & Biology* 41.2 (2015), pp. 574–587. DOI: <https://doi.org/10.1016/j.ultrasmedbio.2014.08.013>.
- [121] H. Rivaz, Z. Karimaghloo, and D. L. Collins. "Self-similarity weighted mutual information: a new nonrigid image registration metric". In: *Medical image analysis* 18.2 (2014), pp. 343–358. DOI: <https://doi.org/10.1016/j.media.2013.12.003>.
- [122] H. Rivaz, Z. Karimaghloo, V. S. Fonov, and D. L. Collins. "Nonrigid registration of ultrasound and MRI using contextual conditioned mutual information". In: *IEEE transactions on medical imaging* 33.3 (2014), pp. 708–725. DOI: <https://doi.org/10.1109/TMI.2013.2294630>.
- [123] D. C. Rockey, S. H. Caldwell, Z. D. Goodman, R. C. Nelson, and A. D. Smith. "Liver biopsy". In: *Hepatology* 49.3 (2009), pp. 1017–1044. DOI: <https://doi.org/10.1002/hep.22742>.
- [124] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional networks for biomedical image segmentation". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015* (2015), pp. 234–241. DOI: https://doi.org/10.1007/978-3-319-24574-4_28.
- [125] C. G. Roth and S. Deshmukh. *Fundamentals of Body MRI*. 2nd Ed. Fundamentals of Radiology. Elsevier Health Sciences, 2016. ISBN: 9780323431415.
- [126] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision (IJCV)* 115.3 (2015), pp. 211–252. DOI: <https://doi.org/10.1007/s11263-015-0816-y>. URL: <http://www.image-net.org/>.
- [127] K. C. Santosh and R. S. Hegadi. *Recent Trends in Image Processing and Pattern Recognition : Second International Conference, RTIP2R 2018, Solapur, India, December 21–22, 2018, Revised Selected Papers, Part I*. 1st ed. Springer eBooks, Computer Science. Singapore: Springer, 2019, pp. 496–497. ISBN: 978-981-13-9181-1. URL: <https://doi.org/10.1007/978-981-13-9181-1>.
-

- [128] R. J. Schneider, D. P. Perrin, N. V. Vasilyev, G. R. Marx, P. J. Del Nido, and R. D. Howe. "Real-time image-based rigid registration of three-dimensional ultrasound". In: *Medical image analysis* 16.2 (2012), pp. 402–414. DOI: <https://doi.org/10.1016/j.media.2011.10.004>.
- [129] T. Schneider, C. Mawrin, C. Scherlach, M. Skalej, and R. Firsching. "Die Gliome des Erwachsenen". In: *Dtsch Arztebl International* 107.45 (2010), pp. 799–808. DOI: <https://doi.org/10.3238/arztebl.2010.0799>.
- [130] M. Schwier, T. Chitiboi, T. Hülnhagen, and H. Hahn. "Automated spine and vertebrae detection in CT images using object-based image analysis". In: *International Journal for Numerical Methods in Biomedical Engineering* 29.9 (2013), pp. 938–963. DOI: <https://doi.org/10.1002/cnm.2582>.
- [131] M. Schwier, J. van Griethuysen, M. Vangel, S. Pieper, P. S., C. Tempany, H. Aerts, F. Fennessy, and A. Fedorov. "Repeatability of Multiparametric Prostate MRI Radiomics Features". In: *Scientific Reports* 9.9441 (2019). DOI: <https://doi.org/10.1038/s41598-019-45766-z>.
- [132] M. Schwier and H. K. Hahn. "Segmentation hierarchies and border features for automatic pregnancy detection in porcine ultrasound". In: *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. 2014, pp. 931–934. DOI: <https://doi.org/10.1109/ISBI.2014.6868024>.
- [133] M. Schwier. "Object-based Image Analysis for Detection and Segmentation Tasks in Biomedical Imaging". PhD thesis. Jacobs University Bremen, 2016. URL: <http://urn-resolving.de/urn:nbn:de:gbv:579-opus-1006017>.
- [134] L. B. Seeff, G. T. Everson, T. R. Morgan, T. M. Curto, W. M. Lee, M. G. Ghany, M. L. Shiffman, R. J. Fontana, A. M. D. Bisceglie, H. L. Bonkovsky, and J. L. Dienstag. "Complication Rate of Percutaneous Liver Biopsies among Persons with Advanced Chronic Liver Disease in the HALT-C Trial". In: *Clinical Gastroenterology and Hepatology* 8.10 (2010), pp. 877–883. DOI: <https://doi.org/10.1016/j.cgh.2010.03.025>.
- [135] T. Selbekk, A. S. Jakola, O. Solheim, T. F. Johansen, F. Lindseth, I. Reinertsen, and G. Unsgård. "Ultrasound imaging in neurosurgery: Approaches to minimize surgically induced image artefacts for improved resection control". In: *Acta neurochirurgica* 155.6 (2013), pp. 973–980. DOI: <https://doi.org/10.1007/s00701-013-1647-7>.
- [136] El-Serag, H. B. and Davila, J. A. "Surveillance for hepatocellular carcinoma: in whom and how?" In: *Therapeutic Advances in Gastroenterology* 4.1 (2011), pp. 5–10. DOI: <https://doi.org/10.1177/1756283X10385964>.

-
- [137] D. Shen, G. Wu, and H.-I. Suk. "Deep Learning in Medical Image Analysis". In: *Annual Review of Biomedical Engineering* 19.1 (2017), pp. 221–248. DOI: <https://doi.org/10.1146/annurev-bioeng-071516-044442>.
- [138] H. Shikata, G. McLennan, E. A. Hoffman, and M. Sonka. "Segmentation of Pulmonary Vascular Trees from Thoracic 3D CT Images". In: *Journal of Biomedical Imaging* 2009.24 (2009), pp. 1–11. DOI: <https://doi.org/10.1155/2009/636240>.
- [139] J. G. Sled, A. P. Zijdenbos, and A. C. Evans. "A nonparametric method for automatic correction of intensity nonuniformity in MRI data". In: *IEEE Transactions on Medical Imaging* 17.1 (1998), pp. 87–97. DOI: <https://doi.org/10.1109/42.668698>.
- [140] N. R. Smoll, O. P. Gautschi, B. Schatlo, K. Schaller, and D. C. Weber. "Relative survival of patients with supratentorial low-grade gliomas". In: *Neuro-Oncology* 14.8 (2012), pp. 1062–1069. DOI: <https://doi.org/10.1093/neuonc/nos144>.
- [141] T. Sørensen. *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*. Vol. 5. 4. Det Kongelige Danske Videnskabernes Selskab, 1948, pp. 1–34.
- [142] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. Kruthiventi, and R. V. Babu. "Chapter 2 - An Introduction to Deep Convolutional Neural Nets for Computer Vision". In: *Deep Learning for Medical Image Analysis*. Ed. by S. K. Zhou, H. Greenspan, and D. Shen. Academic Press, 2017, pp. 26–30, 33. ISBN: 978-0-12-810408-8. DOI: <https://doi.org/10.1016/B978-0-12-810408-8.00003-1>.
- [143] J. Stoll. "Technology of Ultrasound-Guided Therapy". In: *Intraoperative Imaging and Image-Guided Therapy*. Ed. by F. A. Jolesz. New York, NY: Springer New York, 2014, pp. 155–162. ISBN: 978-1-4614-7656-6. DOI: https://doi.org/10.1007/978-1-4614-7657-3_10.
- [144] R. T. Stravitz. "Management of the Cirrhotic Patient Before Liver Transplantation: The Role of the Referring Gastroenterologist". In: *Journal of gastroenterology and hepatology* 2.5 (2006), pp. 346–354. ISSN: 1554-7914.
- [145] S. C. Thust, S. Heiland, A. Falini, H. R. Jäger, A. D. Waldman, P. C. Sundgren, C. Godi, V. K. Katsaros, A. Ramos, N. Bargallo, M. W. Vernooij, T. Yousry, M. Bendszus, and M. Smits. "Glioma imaging in Europe: A survey of 220 centres and recommendations for best clinical practice". In: *European Radiology* 28.8 (2018), pp. 3306–3317. DOI: <https://doi.org/10.1007/s00330-018-5314-5>.

- [146] M. Toews and W. Wells. In: *Medical Image Analysis* 17.3 (2013), pp. 271–282. DOI: <https://doi.org/10.1016/j.media.2012.11.002>.
- [147] Universität Freiburg, Department of Computer Vision - Pattern Recognition and Image Processing. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. URL: <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/>. Accessed: 2020-01-04.
- [148] G. Unsgård, O. Solheim, F. Lindseth, and T. Selbekk. “Intra-operative imaging with 3D ultrasound in neurosurgery”. In: *Acta neurochirurgica. Supplement* 109 (2011), pp. 181–186. ISSN: 0065-1419. DOI: https://doi.org/10.1007/978-3-211-99651-5_28.
- [149] G. Unsgård, O. Solheim, and T. Selbekk. “Intraoperative Ultrasound in Neurosurgery”. In: *Intraoperative Imaging and Image-Guided Therapy*. Ed. by F. A. Jolesz. New York: Springer New York, 2014, pp. 549–565. ISBN: 978-1-4614-7657-3. DOI: https://doi.org/10.1007/978-1-4614-7657-3_41.
- [150] J. J. M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. W. L. Aerts. “Computational Radiomics System to Decode the Radiographic Phenotype”. In: *Cancer research* 77.21 (2017), e104–e107. DOI: <https://doi.org/10.1158/0008-5472.CAN-17-0339>.
- [151] E. M. van Rikxoort, B. de Hoop, S. van de Vorst, M. Prokop, and B. van Ginneken. “Automatic Segmentation of Pulmonary Segments From Volumetric Chest CT Scans”. In: *IEEE Transactions on Medical Imaging* 28.4 (2009), pp. 621–630. DOI: <https://doi.org/10.1109/TMI.2008.2008968>.
- [152] B. Varghese, F. Chen, D. Hwang, S. L. Palmer, A. L. de Castro Abreu, O. Ukimura, M. Aron, M. Aron, I. Gill, V. Duddalwar, and G. Pandey. “Objective risk stratification of prostate cancer using machine learning and radiomics applied to multiparametric magnetic resonance images”. In: *Scientific reports* 9.1 (2019), p. 1570. DOI: <https://doi.org/10.1038/s41598-018-38381-x>.
- [153] M. Verleysen and D. François. “The Curse of Dimensionality in Data Mining and Time Series Prediction”. In: *Computational Intelligence and Bioinspired Systems*. Ed. by J. Cabestany, A. Prieto, and F. Sandoval. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 758–770. ISBN: 978-3-540-32106-4.

-
- [154] L. Vincent and P. Soille. "Watersheds in digital spaces: an efficient algorithm based on immersion simulations". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13.6 (1991), pp. 583–598. ISSN: 1939-3539. DOI: <https://doi.org/10.1109/34.87344>.
- [155] A. Watanabe, M. Ramalho, M. AlObaidy, H. J. Kim, F. G. Velloni, and R. C. Semelka. "Magnetic resonance imaging of the cirrhotic liver: An update". In: *World Journal of Hepatology* 7.3 (2015), pp. 468–487. DOI: <https://doi.org/10.4254/wjh.v7.i3.468>.
- [156] W. Wein, A. Ladikos, B. Fuerst, A. Shah, K. Sharma, and N. Navab. "Global registration of ultrasound to MRI using the LC2 metric for enabling neurosurgical guidance". In: *Medical image computing and computer-assisted intervention : MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention* 16.Pt 1 (2013), pp. 34–41.
- [157] W. Wick, C. Hartmann, C. Engel, M. Stoffels, J. Felsberg, F. Stockhammer, M. C. Sabel, S. Koeppen, R. Ketter, R. Meyermann, M. Rapp, C. Meisner, R. D. Kortmann, T. Pietsch, O. D. Wiestler, U. Ernemann, M. Bamberg, G. Reifenberger, A. von Deimling, and M. Weller. "NOA-04 Randomized Phase III Trial of Sequential Radiochemotherapy of Anaplastic Glioma With Procarbazine, Lomustine, and Vincristine or Temozolomide". In: *Journal of Clinical Oncology* 27.35 (2009), pp. 5874–5880. DOI: <https://doi.org/10.1200/JCO.2009.23.6497>.
- [158] I. H. Witten, E. Frank, and M. A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. 3rd. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011, pp. 308, 31, 357. ISBN: 978-0-12-374856-0.
- [159] C.-Y. Xiao, Z. Su, and Y.-z. Chen. "A diffusion stick method for speckle suppression in ultrasonic images". In: *Pattern Recognition Letters* 25.16 (2004), pp. 1867–1877. ISSN: 0167-8655. DOI: <https://doi.org/10.1016/j.patrec.2004.08.014>.
- [160] Y. Xiao, M. Fortin, G. U. H. Rivaz, and I. Reinertsen. "REtroSpective Evaluation of Cerebral Tumors (RESECT): A clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries". In: *Medical Physics* 44.7 (2017), pp. 3875–3882. DOI: <https://doi.org/10.1002/mp.12268>.
- [161] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, R. H. Mak, and H. J. Aerts. "Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging". In: *Clinical Cancer Research* (2019). DOI: <https://doi.org/10.1158/1078-0432.CCR-18-2495>.
-

- [162] G. Y., G. Y., and S. D. “Deformable MR Prostate Segmentation via Deep Feature Learning and Sparse Patch Matching”. In: *IEEE Transactions on Medical Imaging* 35.4 (2016), pp. 1077–1089. DOI: <https://doi.org/10.1109/TMI.2015.2508280>.
- [163] K. Yasaka, H. Akai, A. Kunimatsu, O. Abe, and S. Kiryu. “Liver Fibrosis: Deep Convolutional Neural Network for Staging by Using Gadoteric Acid-enhanced Hepatobiliary Phase MR Images”. In: *Radiology* 287.1 (2018), pp. 146–155. DOI: <https://doi.org/10.1148/radiol.2017171928>.
- [164] S. S. F. Yip and H. J. W. L. Aerts. “Applications and limitations of radiomics”. In: *Physics in Medicine and Biology* 61.13 (2016), R150–R166. DOI: <https://doi.org/10.1088/0031-9155/61/13/r150>.
- [165] R. L. Yong and R. R. Lonser. “Surgery for glioblastoma multiforme: Striking a balance”. In: *World neurosurgery* 76.6 (2011), pp. 528–530. DOI: <https://doi.org/10.1016/j.wneu.2011.06.053>.
- [166] L. Yu and H. Liu. “Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution”. English (US). In: *Proceedings, Twentieth International Conference on Machine Learning*. Ed. by T. Fawcett and N. Mishra. Vol. 2. 2003, pp. 856–863. ISBN: 1577351894.
- [167] Y. Zhang, X. M. Zhang, J. C. Prowda, H. L. Zhang, C. Sant’anna Henry, G. Shih, J. C. Emond, and M. R. Prince. “Changes in hepatic venous morphology with cirrhosis on MRI”. In: *Journal of magnetic resonance imaging : JMRI* 29.5 (2009), pp. 1085–1092. ISSN: 1053-1807. DOI: <https://doi.org/10.1002/jmri.21718>.
- [168] Y. Zhang, A. Oikonomou, A. Wong, M. A. Haider, and F. Khalvati. “Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer”. In: *Scientific reports* 7 (2017), p. 46349. DOI: <https://doi.org/10.1038/srep46349>.
- [169] L. Zhao, J. Gong, H. Yin, W. Qin, and M. Shi. “Multiparametric MRI Based Radiomics for the Prediction of Induction Chemotherapy Response and Survival in Locally Advanced Nasopharyngeal Carcinoma”. In: *International Journal of Radiation Oncology, Biology, Physics* 102.3S (2018), pp. 127–127. DOI: <https://doi.org/10.1016/j.ijrobp.2018.06.316>.
- [170] L. Zhao and F. A. Jolesz. “Navigation with the Integration of Device Tracking and Medical Imaging”. In: *Intraoperative Imaging and Image-Guided Therapy*. Ed. by F. A. Jolesz. New York, NY: Springer New York, 2014, pp. 259, 260, 261–262, 262. ISBN: 978-1-4614-7657-3. DOI: https://doi.org/10.1007/978-1-4614-7657-3_18.

- [171] F. Zoehrer, J. Drexel, and H. K. Hahn. “Speckle reduction for automated breast ultrasound comparison of three filtering methods”. In: *Proc BVM* (2010). Ed. by T. Deserno, H. Handels, H.-P. Meinzer, and T. Tolxdorff, pp. 390–94.