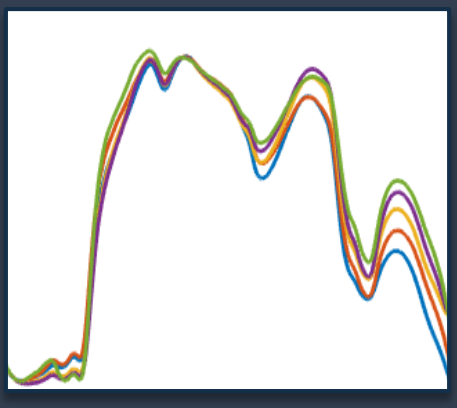
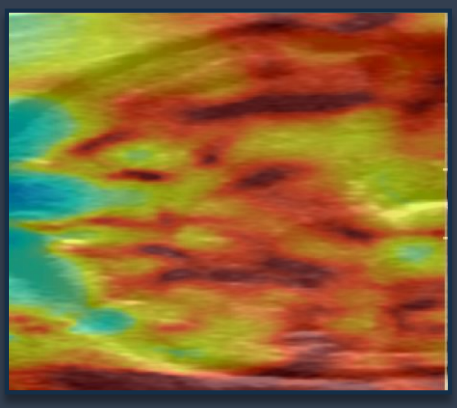
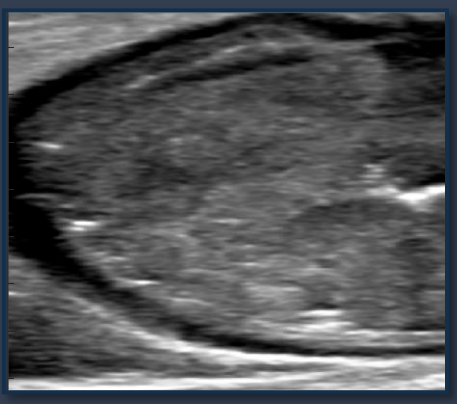
8-4-2021

Tissue type classification and resection margin assessment in colorectal cancer specimens using Ultrasound, Elastography, and Diffuse reflectance spectroscopy

Thesis

Master of Science

Technical Medicine

Medical Imaging and Interventions

University of Twente

C.A.W. van Egmond

ANTONI VAN LEEUWENHOEK
NEDERLANDS KANKER INSTITUUT

UNIVERSITY OF TWENTE.

Graduation Committee

Chairmen and Medical supervisor

Prof. Dr. T.J.M. Ruers

Nanobiophysics group
Faculty of Science and Technology
University of Twente
Enschede, The Netherlands

Department of Surgery
Netherlands Cancer Institute
Antoni van Leeuwenhoek
Amsterdam, The Netherlands

Technical supervisor

Dr.Ir. N. Bosschaart

Biomedical Photonic Imaging group
Faculty of Science and Technology
University of Twente
Enschede, The Netherlands

Process supervisor

E.M. Walter, MSc.

Faculty of Science and Technology
University of Twente
Enschede, The Netherlands

Technical/daily supervisor

F. Geldof, MSc.

Department of Surgery
Netherlands Cancer Institute
Antoni van Leeuwenhoek
Amsterdam, The Netherlands

Extra Technical supervisor (not in committee)

Dr. B. Dashtbozorg

Department of Surgery
Netherlands Cancer Institute
Antoni van Leeuwenhoek
Amsterdam, The Netherlands

External member

J.K. van Zandwijk, MSc.

Magnetic Detection and Imaging group
Faculty of Science and Technology
University of Twente
Enschede, The Netherlands

Preface

This thesis represents the research of my final year of the master Technical Medicine with the track Medical Imaging and Interventions. In first instance, I would like to thank the research group of Theo Ruers in the NKI-AvL for providing me an interesting graduation position in a team with great colleages from different disciplines. I would like to thank Behdad and Freija for their guidance during the last year. Thank you for your enthusiasm, support, critical questions, input, discussions and our good team work. Furthermore, I would like to thank Nienke for her support, critical questions, and feedback from the technical point of view. From the clinical point of view, I would like to thank Theo for his clinical input and critical questions. Also, I would like to thank Elyse for her support and positivity during my professional and personal development in the last two years. I would like to thank Jordi for being my external supervisor. Last, but definitely not least, I would like to thank Roan, my family and friends for their support, positive energy, and trust.

Abstract

Introduction: Colorectal cancer (CRC) ranks third in terms of incidence worldwide, and second in terms of mortality. An oncologic resection with negative resection margins is essential for overall survival and local recurrence. Achieving negative circumferential resection margins (CRM) in locally advanced rectal cancer is problematic through the effects of neoadjuvant radiotherapy and the technically challenging procedure. Although diffuse reflectance spectroscopy (DRS) can distinguish colorectal tumor tissue from healthy tissue, it is not able to distinguish tumor from the neoadjuvant effect called fibrosis. Furthermore, it is challenging to determine the resection margin with DRS. Ultrasound (US) B-mode may be of additional value since the healthy colorectal wall and fat are discriminative from tumor tissue. However, it is difficult to distinguish tumor tissue from fibrosis through the same ultrasonic appearance. Ultrasound elastography is of additional value to US B-mode in colorectal tumor grade assessment and response assessment. However, the current literature does not show whether US elastography can differentiate between tumor and fibrosis. Furthermore, it has not been used yet to estimate the resection margin. Therefore, an ex vivo study was performed to investigate the added value of US B-mode and shear wave elastography (SWE) in the classification of colorectal tumor tissue versus the healthy colorectal wall, fat, and fibrosis and in the estimation of the resection margin.

Tissue type classification: An ex vivo study was performed to investigate the ability of US B-mode and SWE to distinguish colorectal tumor tissue from healthy colorectal tissue, fat, and fibrosis.
Ultrasound B-mode images and SWE images were retrieved simultaneously from freshly excised CRC tissue with maximal three measurements per specimen. Superpixel segmentation was used to divide the US B-mode and SWE images into groups of pixels that share common characteristics. Subsequently, features of US B-mode and SWE images were extracted per superpixel. The superpixels were labeled with a tissue type using histopathology. The US B-mode and SWE features were used as input for machine learning-based classification algorithms. A classification algorithm was developed using nine selected features (five B-mode, and four SWE features) and a Bagged Trees algorithm. Furthermore, another classification algorithm was developed using a Fine Gaussian SVM (FG-SVM) algorithm.
The Bagged Trees algorithm resulted in a Matthews correlation coefficient (MCC) of 0.40 for the test set, an accuracy of 0.86, an area under the curve (AUC) value of 0.83, sensitivity 0.87, and specificity 0.74. The FG-SVM algorithm resulted in an MCC of 0.40 for the test set, an accuracies of 0.89, an AUC value of 0.81, sensitivity of 0.91, and specificity of 0.63.
The results demonstrate the capability of the proposed technique for the illustration of tumor tissue location during CRC surgery. In this, the B-mode and SWE images are automatically divided into superpixels which result in the detection of the tumor areas in the complete image. The tissue type classification using both imaging modalities is capable of distinguishing colorectal tumor tissue from neoadjuvant effects (fibrosis) and thereby avoiding positive CRM's and preserving healthy tissue.

Resection margin estimation: The recognition of tumor tissue is the first step, but the estimation of the resection margin is more clinically relevant. Therefore, a second ex vivo study was performed to investigate whether a combination of DRS, US B-mode, and SWE can be used to estimate the resection margin in ex vivo specimens.
The study population was a sub-study population of the previous study and consisted of 14 specimens with tumor tissue within 1 cm from the resection plane. Regression analysis was performed using DRS, B-mode, and SWE features to estimate the resection margin. The preliminary results of this study showed that features of the US B-mode and SWE images result in a significantly lower mean absolute error (1.88 mm) of the resection margin than DRS (2.10 mm) or DRS in combination with B-mode and SWE features (1.96 mm). However, the results of this study are preliminary since limited data was available with successful DRS, B-mode, and SWE measurements of specimens with tumor tissue within 1 cm of the resection plane.

Discussion and conclusion: In conclusion, the results of the ex vivo studies in this thesis gave new insight into how US B-mode and SWE can be used to distinguish tumor tissue from fibrosis, the healthy colorectal wall, and fat. Furthermore, it demonstrated how the US techniques can be combined with DRS in the estimation of the resection margin of colorectal tumors. However, optimization of the classification and regression algorithms and future research is needed to investigate whether these techniques can lead to less positive resection margins of rectal tumors while preserving more healthy tissues.

Table of Contents

List of Figures

List of Tables

Abbreviations

AUC	area under the curve
CRC	colorectal cancer
CRM	circumferential resection margin
CRT	chemoradiotherapy
CT	computed tomography
DRS	diffuse reflectance spectroscopy
ERUS	Endorectal ultrasound
FG-SVM	Fine Gaussian-support vector machine
IQR	interquartile range
kNN	k-nearest neighbor
LDA	linear discriminant analysis
MAE	mean absolute error
MCC	Matthews correlation coefficient
ML	machine learning
MRF	mesorectal fascia
MRI	magnetic resonance imaging
NIR	near-infrared
RMSE	root mean square error
ROC	receiver operating characteristic curve
SWE	shear wave elastography
TME	total mesorectal excision
TNM	tumor invasion, lymph node metastases, distant metastases
US B-mode	ultrasound brightness mode
VIS	Visual

1 Introduction

1.1 Context

Colorectal cancer (CRC) ranks third in terms of incidence worldwide, and second in terms of mortality.[1] In 29% of the colorectal cancer diagnoses, the tumor is located in the rectum.[2] The standard therapy for rectal cancer is surgery.[3] Patients with a locally advanced rectal tumor (T3 and T4) or lymph node involvement on imaging are also treated with neoadjuvant chemoradiotherapy.[3–5] The local advancement comprises tumor invasion into the mesorectal fascia. Neoadjuvant therapy aims to downstage tumors for a more effective oncological resection.[6] Furthermore, it results in improved prognosis and decreased local recurrence rates.[7] However, local therapy effects, such as peritumoral inflammation, edema, fibrosis, and necrosis, leading to a loss of surgical planes[8].

Since the surgical technique is based on this structure, the loss of surgical planes leads to a challenge. Furthermore, rectal surgery is challenging through limited space in the pelvic cavity.[4,6,9] Besides, tumor tissue can have the same palpable stiffness and appearance as fibrosis, which makes it hard for the surgeon to distinguish.[10] Several studies demonstrated that a higher clinical and pathological T stage of rectal cancer is related to a higher rate of positive circumferential resection margins (CRM).[11–14] The proportion of resected rectal tumors with a positive CRM was 5.4% in the Netherlands in 2011-2015 among 12,089 patients.[15] However, when dividing the tumors in cT1-3 and cT4 tumors, the occurrences of positive CRM were 4.7% and 12.7% respectively.

1.2 Problem description

The resection margin is a prognostic factor for recurrence rate and overall survival in surgically treated, locally advanced rectal cancers. Patients with a positive CRM (distance to the tumor <1mm) have a significantly higher chance of local recurrence and poorer survival than patients with tumor-free resection margins.[8,11,13,15–20]

To decrease the number of positive CRM's in colorectal cancer, in vivo and ex vivo studies have been performed with diffuse reflectance spectroscopy (DRS).[4,21,22] DRS is a non-invasive technique that shows the tissue properties of superficial tissue layers. DRS is useful in the discrimination between tumor and surrounding healthy tissue. However, the studies demonstrate that it was not possible yet to discriminate colorectal tumors from fibrosis tissue when both tissue types are present in a specimen.

Another technique that has been studied to discriminate between malignant tumor tissue and chemotherapy effects is ultrasound elastography.[23] This technique shows the stiffness of tissues as a colored map on top of an ultrasound B-mode image. Stiffness is a key property in characterizing malignant tumors since malignancies are generally stiffer than healthy tissue structures. Furthermore, the US B-mode image can contribute to discriminate colorectal tumor tissues from healthy tissues as well.[24,25] The B-mode features for malignancies are hypoechoic bowel wall thickening with irregular contour, the loss of stratification of the wall layers, and the absence of normal peristalses.[26,27] Since the B-mode image is always part of the ultrasound elastography, the B-mode features are often combined with elastography features.[28–33] It has been shown that ultrasound elastography was useful in breast cancer to discriminate malignancies from chemotherapy effects.[23] Furthermore, ultrasound elastography is useful for tumor staging, prediction of lymph node metastases, and prediction of therapy response in colorectal cancer.[29,32–36] However, the application of ultrasound elastography to differentiate colorectal tumors from therapy effects has not been studied yet. Additionally, the technique has not been studied to estimate the CRM intraoperatively or in ex vivo specimens.

The current study will focus on the possibility of ultrasound elastography to discriminate malignant tumors from fibrosis, the healthy colorectal wall, and fat in surgical specimens. Furthermore, the

additional value of ultrasound B-mode and elastography to DRS will be studied on the estimation of the circumferential resection margin.

1.3 Aims and research questions

The goal of this study is to investigate the benefit of ultrasound elastography in distinguishing tumor tissue from healthy colorectal tissue and fibrosis. Secondly, the goal is to combine US B-mode, elastography, and DRS to estimate the resection margin. These purposes can be described with the following research questions:

1. To what extend can US B-mode and elastography be used to distinguish colorectal malignant tumor tissue from healthy tissue and fibrosis in ex vivo surgical specimens?
2. What is the added value of US B-mode and elastography in the estimation of the resection margin to DRS?

To answer both research questions, two studies will be performed: 'Tissue type classification' and 'Resection margin estimation'. These studies contain the following steps:

Tissue type classification:

1. Ex vivo measurements on surgical specimens to acquire DRS, ultrasound B-mode, and elastography data.
2. Annotation of the different tissue types in the ultrasound B-mode images based on histopathological coupes of the surgical specimens.
3. Feature extraction including intensity features from ultrasound B-mode images, and elastography features from elastography data.
4. Comparison of extracted features for tumor, fibrosis, fat, and healthy colorectal tissue.
5. Feature selection and feature importance analysis.
6. Classification of tumor versus the other tissue types based on ultrasound B-mode and elastography features.

Resection margin estimation:

1. Data selection from ex vivo measurements in the tissue type classification study, from specimens with tumor tissue within 1 cm from the resection margin.
2. DRS data preprocessing and feature extraction.
3. Combination of the US B-mode and elastography features with DRS features.
4. Labeling of the data with the resection margin based on histopathological coupes of the surgical specimen.
5. Feature subset selection of the best representative features for resection margin assessment.
6. Regression analysis to estimate the resection margin based on US B-mode, elastography, and DRS features.

1.4 Outline

The thesis is outlined in the following chapters:

Chapter 2: A clinical background on colorectal cancer with the epidemiology, diagnostic methods, treatments, and surgical margin explanations.

Chapter 3: A technical background on DRS, US B-mode, and US elastography. This background includes the technical explanation, applications, advantages, and limitations per technique. Besides, the theoretical background on machine learning is explained.

Chapter 4: An ex vivo study on tissue type classification based on ultrasound B-mode and elastography. The data collection of ultrasound B-mode, elastography, and DRS data is explained. Secondly, the methodology of the data analysis of the ultrasound B-mode and elastography data and the correlation with pathology are described. Subsequently, the ultrasound B-mode and elastography features are

compared per tissue type. Thereafter, the classification results of tumor versus fat, healthy colorectal tissue, and fibrosis are shown. Lastly, discussion points and conclusions are drawn based on the results.

Chapter 5: An ex vivo study on the resection margin estimation based on DRS, US B-mode, and elastography. The combination of the techniques is described, and the correlation of the techniques to the pathology is shown. Furthermore, DRS data preprocessing and feature extraction are explained. Thereafter, the results of the regression analysis to estimate the resection margin based on US B-mode, elastography, and DRS features are shown. Lastly, the results will be discussed and a conclusion is drawn.

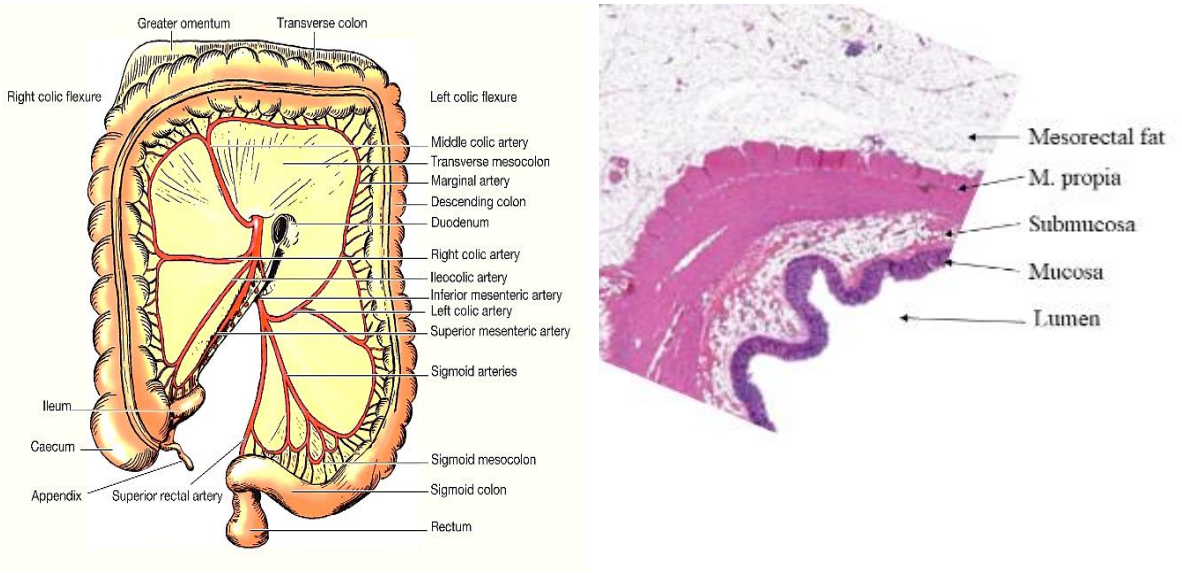
Chapter 6: Overall discussion, conclusion, and recommendations.

2 Clinical background

This clinical background provides more insight into the clinical problem of positive resection margins in rectal cancer. Since the tissue structure, tumor types, tumor classification, treatments, and margin assessment are comparable for colon and rectal cancer, the surgical specimens of tumors that are located in the rectum and colon will be included in the ex vivo study to collect more data. Therefore, this chapter explains the clinical background of both colon and rectal cancer.

2.1 Anatomy and physiology

The colon and rectum belong to the large intestines. The large intestines include the caecum, appendix, colon, rectum, and anal canal.[37] The colon comprises the colon ascendens, colon transversum, colon descendens, and the sigmoid colon. The anatomy is shown in Figure **2-1**A. The large intestines consist of several concentric layers, as shown in Figure **2-1**B. From the lumen to outwards, these layers are columnar epithelium, lamina propia, muscularis mucosae, submucosa, muscularis propia (m. propia) (comprising of an inner circular layer and an outer longitudinal layer of smooth muscle), and finally, the serosa. The principal physiological function of the large intestines is the absorption of water, sodium, and chloride from the liquid contents derived from the small intestines. Furthermore, it allows for the storage and excretion of the unabsorbed residue as feces.



A. B.

Figure 2-1: A. Anatomy of the large intestines.[37] B. Histological concentric layers of the colorectal wall.

2.2 Pathology

Most colorectal cancers arise from adenomatous polyps, which are benign neoplasms of the columnar epithelium.[38] Besides adenomatous polyps, other types of polypoid lesions include hyperplastic polyps, serrated adenomas, flat adenomas, hamartomatous polyps, and inflammatory polyps. The adenoma can be considered malignant when neoplastic cells pass through the m. mucosae and infiltrate the submucosa.[39] Most colorectal malignancies are located distal to the left colic flexure (60-70%), particularly in the rectum and sigmoid colon. The growing pattern of colorectal malignancies can be both longitudinally and in-depth, with the possibility to infiltrate surrounding organs and tissue. Local invasion is particularly evident for rectal tumors, due to the close relationship with organs such as the bladder, prostate, vagina, uterus, ureters, perineural muscles, and pelvic bones.

2.3 Colorectal cancer epidemiology

Colorectal cancer was diagnosed in 10.8% of all patients who were diagnosed with cancer in 2019 in the Netherlands.[40] Besides, 24.9% of these patients were diagnosed with rectal cancer. In this year, 4781 patients died of colorectal cancer, which comprises 14% of all patients who died of cancer that year.

The 5-year survival depends strongly on the tumor stadium, as shown in Figure **2-2**. The 5-years survival of stadium IV is 11% for colon cancer and 14% for rectal cancer.[40] Whereas the 5-years survival percentages for stadium I, II, and III of colorectal cancer are between 70 and 98%. In 2018, 27% of the colon cancer diagnoses belonged to stadium I, 27% stadium II, 26% stadium III, and 20% stadium IV. For rectal cancer, the distribution was 28%, 16%, 39%, and 17%, respectively. The description of each stadium is shown in Table **2-2**. The current study focuses mainly on patients diagnosed with colorectal cancer with stadium II, III, and IV.

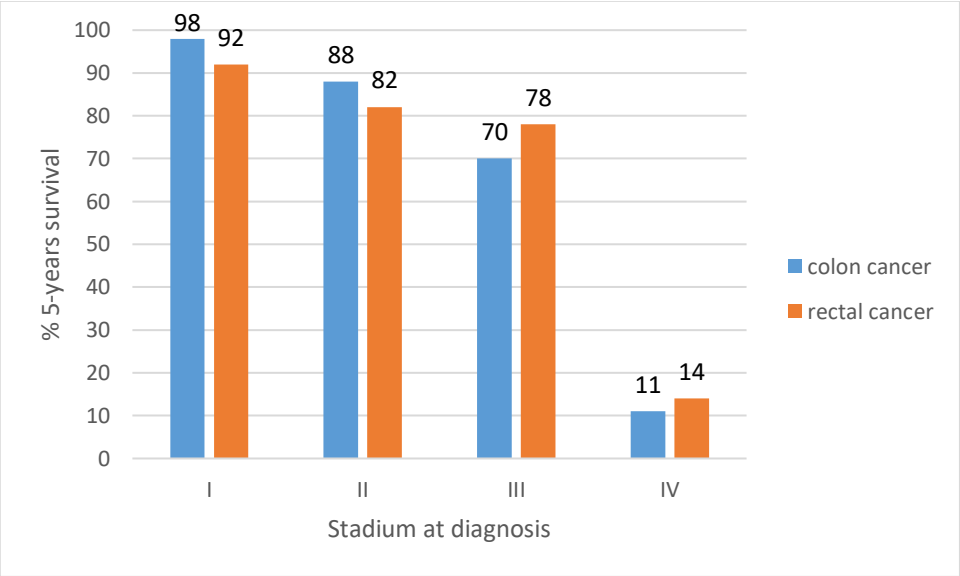


Figure 2-2: The percentage of patients that were diagnosed with colorectal cancer that survived 5-years since diagnosis in the Netherlands in 2018 [40]

2.4 Diagnosis and screening

Most colorectal tumors are found by screening or investigation of symptoms unrelated to the adenoma since most adenomas are silently present. Only large adenomas may cause gross bleeding, and anemia subsequently.[41] Besides, large rectal adenomas can result in symptoms such as urgency, mucus discharge, and tenesmus. Other symptoms are abdominal pain and a change in bowel habits.

Since 2014, the screening program for colorectal cancer has been introduced in the Netherlands. The number of screening-detected colorectal cancers (3733) was 28% of the total diagnosed colorectal cancers (9412 colons + 3711 rectums) in 2018.[40] The screening includes a biennial fecal immunochemical test that measures the amount of hemoglobin (Hb) in feces.[42] When the test is positive (Hb $\geq$ 15µg/g feces), the patient will be referred for a colonoscopy. The colonoscopy is used to prove the malignancy, preferably by taking biopsies.[3] During this procedure, the distal side of the tumor is marked with black tattoo ink. This tattoo is visible from outside the colon, which helps to localize the tumor during surgery. Subsequently, a computed tomography (CT) scan of the thorax- abdomen is performed to show the local tumor invasion, and to show or to exclude metastases. For rectal cancer, a magnetic resonance imaging (MRI) scan is made to stage rectal cancer concerning the T grade and N grade. With the use of these diagnostic modalities and histological assessment, the tumor is graded with a TNM grade and a stadium. The TNM grade describes the local tumor invasion (T), lymph node metastases (N), and metastases to distant organs (M). The TNM gradation for colorectal cancer is shown in Table **2-1**.[43] In the case of a T3 tumor, the distance to the mesorectal fascia (MRF) is also taken into

account. A distinction is made between MRF- and MRF+, where MRF+ means that the distance from the tumor to the MRF ≤1 mm, and MRF- means that the distance from the tumor >1 mm. When the TNM grading is based on clinical information, including physical examination and imaging modalities, the TNM grading is preceded with a 'c'. When the TNM grading is based on surgical and pathological information, the TNM grading is preceded with a 'p'. Lastly, when it concerns a TNM staging after neoadjuvant therapy, the cTNM or pTNM is identified by a 'y' prefix. The stadium of the tumor is assigned according to the TNM grading, as shown in Table **2-2**. The stadium is crucial for treatment possibilities and survival.

Table 2-1: TNM gradation colorectal cancer according to the Union for International Cancer Control 8th edition[43]

TNM grade	Description according to grade
T1	Tumor invades the lamina propia or submucosa, ≤ 2 cm
T2	Tumor invades the muscularis propia, or > 2 cm
T3	Tumor invades the serosa or the perirectal tissue
T4	Tumor perforates the serosa or invades adjacent structures
T4a	Tumor invades nearby structures (other parts of the colon or other organs)
T4b	Tumor perforates the serosa
N0	No lymph nodes containing tumor cells
N1	1-3 lymph nodes with tumor cells
N2	>3 lymph nodes with tumor cells
M0	No metastases to distant organs
M1	Metastases to distant organs
M1a	Metastases confined to one organ without peritoneal metastases
M1b	Metastasis in more than one organ
M1c	Metastasis to the peritoneum with or without other organ involvement

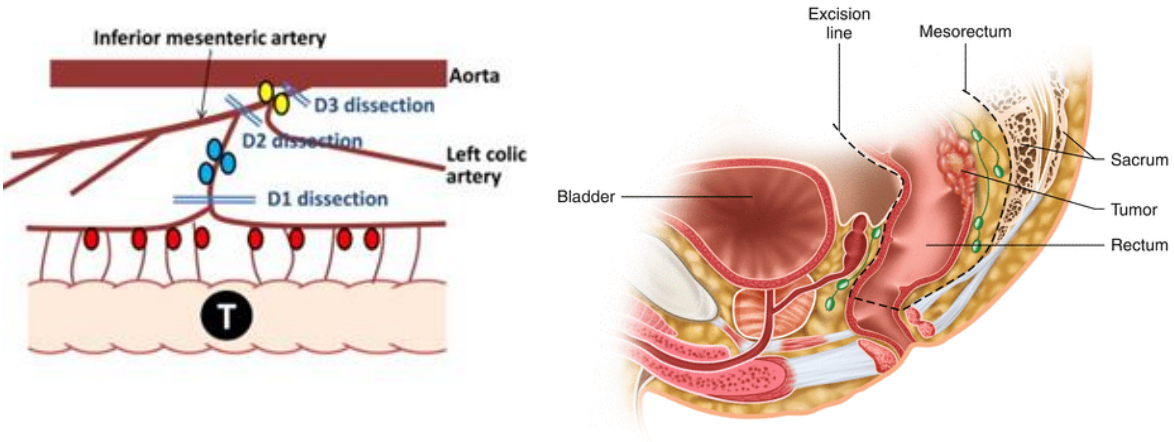
Table 2-2: Tumor stadium based on TNM [43]

Stadium	Combination TNM grade	Description
I	T1-2, N0, M0	Non-advanced tumor without metastases
II	T3-4, N0, M0	Locally advanced tumor without metastases
III	Tx, N1-2, M0	Tumor with lymph node metastases
IV	Tx Nx M1	Tumor with metastasis to distant organs

Tx and Nx involve all T- and N stages

2.5 Treatment

The standard therapy for colorectal cancer is surgery.[3] When the tumor is localized in the colon, the standard surgical procedure is complete mesocolic excision.[3,44] This surgical technique consists of three components. Firstly, resection of the primary tumor and the mesocolon as one package through a sharp dissection over the retroperitoneal fascia while preserving the visceral fascia. The dissection is based on the embryologic planes. Secondly, ligation of the segmental arteries with the pericolic and intermediary lymph nodes, as shown with D2 dissection in Figure **2-3**A.[3,45] These lymph nodes are the first and second lymph nodes for metastases of the colon tumor. Thirdly, resection of enough length of the colon, at least 5 cm proximally and distally from the tumor.



A. B.

Figure 2-3: Surgical technique of colorectal cancer. A. The dissection levels for a tumor (T) that is localized in the left-sided and/or recto-sigmoid colon. The D1 dissection includes the pericolic nodes (yellow), D2 dissection includes the intermediate nodes (blue) too, and D3 dissection includes the main nodes (red) as well.[45] D2 dissection is advised according to the Dutch guidelines on colorectal cancer.[3] B. The surgical technique of total mesorectal excision where the rectal tumor is excised including the perirectal and intermediate lymph nodes and mesorectal fat.[46]

The standard surgical technique for rectal cancer is a total mesorectal excision (TME).[3,47] TME consists of a sharp pelvic dissection of the avascular plane between the autonomic nerve plexus and the mesorectum. The surgical excision is shown schematically in Figure **2-3**B. With this technique, the rectum is excised including the mesorectum and local lymph nodes.[46,47] Furthermore, the autonomic nerve plexus can be preserved, which reduces the risk of postoperative urinary dysfunction, sexual dysfunction, or female infertility. The TME procedure is technically challenging because of the small working space in the pelvic area, the complicated anatomy, and multiple dissection planes.[48]

In case of reactive benign adhesions, or suspicion for tumor invasion in adjacent organs, the advice is to perform a multi-visceral resection.[3] This means that multiple tissues are resected in one surgical specimen. Since this procedure is more extensive, the risk of complications increases. The malignant infiltration is hard to distinguish from inflammatory adhesions during surgery. Therefore, the surgeon may interpret inflammatory adhesions as a macroscopic invasion. Nowadays, only histopathological analysis can provide precise differentiation. A recent study on multivisceral resections in Japan shows that malignant invasion was present in 60.9% of the patients that received a multivisceral resection.[49] Which means that for 39.1% of the patients it was not needed to perform such an extensive surgery.

In addition to surgery, patients can be treated with chemotherapy, radiotherapy, and/or immunotherapy. Neoadjuvant therapy involves preoperative treatment with (one of) these therapies. Neoadjuvant therapy aims to downstage the tumor for a more effective oncological resection.[6–8] The type of therapy depends on the localization of the tumor, lymph node metastases, tumor grade, and the distance of the tumor to the MRF. Other factors that influence the decision for neoadjuvant therapy are local recurrence, survival, toxicity, comorbidities, functional outcomes, and vulnerability. The use of neoadjuvant therapy is a trade-off in absolute risk reduction in local recurrence versus morbidity and functional outcome.

The Dutch guidelines for neoadjuvant therapy differ between colon and rectum tumors.[3] For colon tumors, neoadjuvant chemotherapy is advised for T4bN0-2M0 tumors. The commonly described chemotherapy includes 5-Fluorouracil (5-FU) and Oxaliplatin. For rectal tumors, a short scheme of radiotherapy (5x5 Gy) is advised for cT3c-dN0 and cT1-3N1 (MRF-) tumors. Furthermore, neoadjuvant chemoradiotherapy is advised for cT4 tumors, and/or MRF+, and/or N2 signs, and/or extramesorectal pathological lymph nodes. This treatment consists of radiotherapy (25x2 Gy or 28x1.8 Gy) and chemotherapy Capecetabine (825 mg/m^2). In the case of a cT4 tumor without other risk factors, surgery

can be considered without neoadjuvant therapy. In the case of a cN2 grade without other risk factors, a short scheme of radiotherapy (5x5 Gy) can be considered.

Neoadjuvant immunotherapy is only given to patients in a clinical trial in the Netherlands Cancer Insitute-Antoni van Leeuwenhoek (NKI-AvL). In this clinical trial, called the NICHE study, patients receive ipilimumab (1 mg g^{-1}), nivolumab (3 mg kg^{-1}), and celecoxib.[50]

2.6 Margin assessment

A positive resection margin is a strong unfavorable survival indicator for colorectal cancer patients. The resection margins in colorectal cancer surgery include the distal, proximal, and circumferential margin. The circumferential resection margin (CRM) is the most crucial one since the standard surgical technique already includes a distal and proximal margin of at least 5 cm. According to the Dutch guidelines, the CRM is positive, when the distance from the tumor to the resection surface is ≤1mm.[3] Two illustrations of a positive and negative CRM are shown in Figure **2-4**. Furthermore, two pathology coupes of a positive and negative CRM from two colon specimens are shown in Figure **2-5**. The pathology coupes are shown with distances from the resection plane to the tumor and from the beginning of the tumor to the lumen. Both coupes are from patients that did not receive neoadjuvant therapy. The pathology coupe with a positive CRM (0.51 mm) is from a T3 tumor that was localized in the colon, and the pathology coupe with a negative CRM (4.11 mm) is from a T2 tumor that was localized in the sigmoid colon.

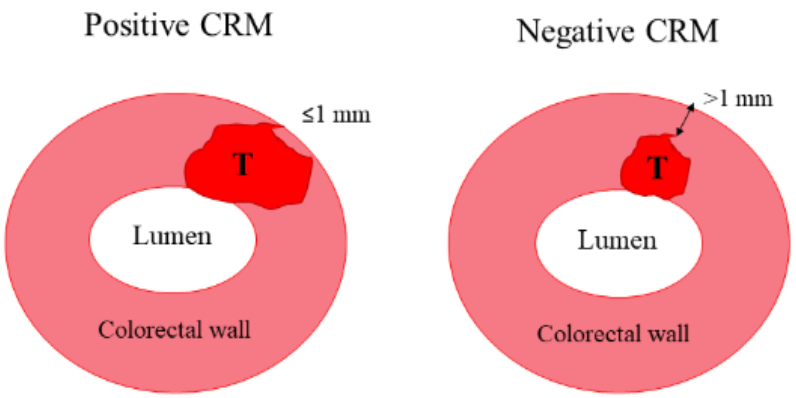
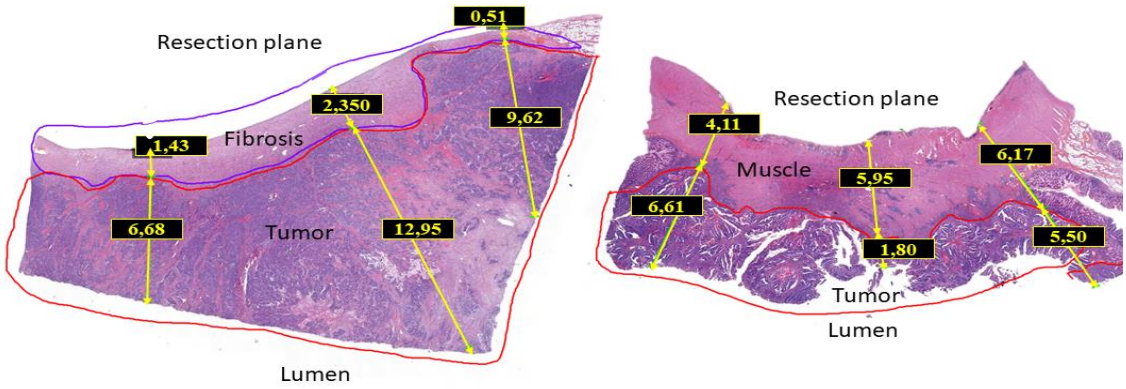


Figure 2-4: Illustration of a positive and negative circumferential margin (CRM). T=tumor. The CRM is positive when the distance between the resection margin and tumor ≤1 mm and the CRM is negative when this distance >1 mm.



A. B.

Figure 2-5: Histopathological coupes with H&E staining of two surgical specimens. Both coupes have the resection plane on top and the luminal side at the bottom. The tumor is delineated with red, and fibrosis is delineated with purple. The distances are shown in millimeters. A. Pathology coupe of a specimen from a T3 tumor that was localized in the colon, with the shortest

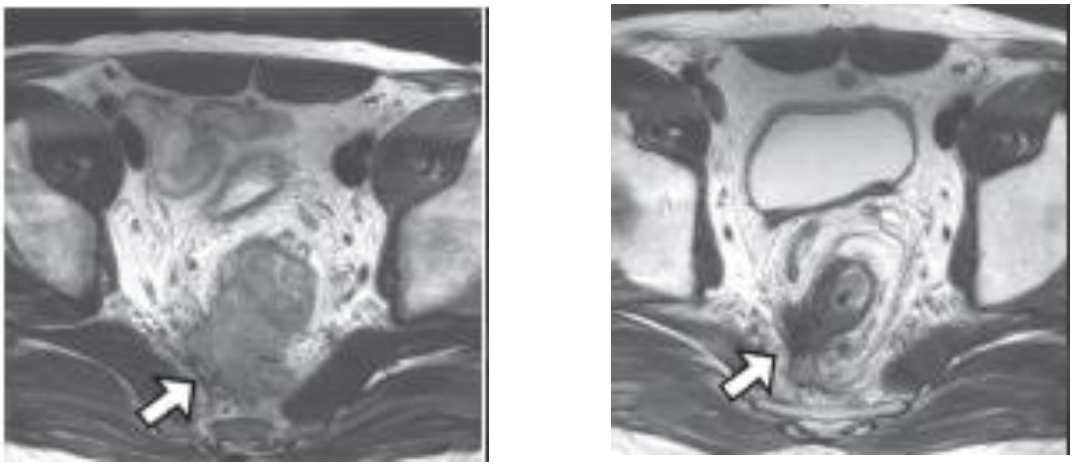
distance from the tumor to the surface of 0.51 mm, indicating a positive circumferential resection margin. B. Pathology coupe of a specimen from a T2 tumor localized in the sigmoid colon. The shortest distance from the tumor to the surface is 4.11 mm, indicating a negative circumferential resection margin.

A positive CRM is related to a higher local recurrence and poorer overall survival.[8,11,13,15–20] The problem of positive CRMs is mainly reported for rectal cancer, in contrast to colon cancer. The limited published data on CRM after colon surgery demonstrated positive CRM rates between 2% and 12% for locally advanced colon cancers without distant metastasis (Stage II-III).[18,51,52] Goffredo et al., a multi-institutional study in the US with 189,343 patients, demonstrated that a positive CRM was identified in 9% of stage II (T3-4N0M0), and 12% of III (TxN1-2M0) colon cancers.[18] A positive CRM after colon cancer was independently related with positive proximal and distal margins, an inadequate lymph node resection, an increasing number of positive nodes, tumor deposits, low facility volume, and a larger tumor size. Furthermore, a positive CRM was confirmed to be independently associated with a decreased overall survival at 5 years (54% vs 68%). A decrease in local recurrence has not been reported for colon cancer surgery with a positive CRM.

The proportion of resected rectal tumors with a positive CRM was 5.4% in the Netherlands in 2011-2015 among 12,089 patients.[15] However, when dividing the tumors in cT1-3 and cT4 tumors, the occurrences of positive CRMs were 4.7% and 12.7% respectively. Similarly, other studies demonstrated that a higher clinical and pathological T stage is related to a higher rate of CRM involvement.[11–14] Since patients with cT4 tumors are generally treated in specialized centers, including the NKI-AvL, the number of positive CRM margins can be higher in these hospitals.

Positive CRMs are especially a problem in rectal cancer surgery through the technically challenging procedure and the local effects caused by neoadjuvant radiotherapy.[9] The quality of the TME procedure directly influences the CRM, since an incomplete TME doubles the incidence of a positive CRM.[53] An incomplete TME concerns an irregular mesorectal surface with defects down to the muscularis propria. Independent prognostic factors for an incomplete TME are a pathologic body mass index (BMI), and the absence of downstaging after long-course chemoradiotherapy (the tumor grade remained ypT3/4 after neoadjuvant chemoradiotherapy).[53] A pathologic BMI influences the TME quality since obesity (BMI>30) makes exposure and other technical aspects of the surgery more difficult and in underweight patients (BMI<18.5) there may be minimal tissue between the tumor and the mesorectal envelope.[48] Furthermore, the absence of downstaging after chemoradiotherapy challenges a complete TME through a smaller distance between the tumor and the resection plane compared to a downgraded T1/T2 tumor. Besides, radiotherapy can result in local effects such as fibrosis. The surgical planes that are essential for a complete TME can become unclear through tumor invasion and fibrosis. Thereby, fibrosis and tumor are hard to discriminate by the surgeon.[4,10,49] When the surgeon suspects tumor invasion in other organs, while it is fibrosis, the chance of too extensive surgery (multivisceral resection) increases. In contrast, the chance of positive resection margins increases when the tumor tissue is interpreted as fibrotic tissue.

The chance of a positive CRM can be predicted based on preoperative imaging and thereby assessing the tumor involvement in the mesorectal fascia or assessing pathological response. However, an interpretation of MRI performed after chemoradiotherapy is problematic due to the same hypointensity of fibrotic tissue and tumor.[8,32,54] Other imaging modalities, such as CT and the endorectal US are also unreliable in restaging rectal cancer after neoadjuvant chemoradiotherapy.[54–56] An example of two MRI images before and after neoadjuvant chemoradiotherapy is shown in Figure **2-6**. It shows that the tumor is decreased after chemoradiotherapy, but the tumor involvement is still challenging to estimate. Therefore, tumor involvement in the MRF was still expected based on the MRI, whereas the histopathological assessment showed the pathologic response of the tumor. Therefore, the surgery might be performed too extensively, resulting in an increased chance of complications.



A. B.

Figure 2-6: Problematic response evaluation of chemoradiotherapy based on T2-weighted MRI images. A. Baseline T2-weighted image shows a bulky mass with mesorectal fascia (MRF) involvement (arrow). B. Post-CRT T2-weighted image, intermixed tumor, and fibrosis at the initial tumor site (arrow). MRF involvement was still suspected. However, pathologic response was confirmed after total mesorectal excision. Thus, there was a discrepancy between the preoperative MRI and pathology. [Adapted from Seo et al.[8]]

A positive CRM for rectal cancer surgery is related to a higher local recurrence. A retrospective analysis on rectal cancers resected in Sweden between 2005 and 2013, demonstrated a local recurrence of 8.8% for resections with a positive CRM and 3.3% for negative CRM within a period of 4,25 years.[57] Quircke et al., an international multicenter randomized controlled trial between 1998 and 2005, demonstrated a local recurrence rate within 3 years of 17% for patients with a positive CRM, and 6% for patients with a negative CRM after TME.[14] Furthermore, they showed that the 3-year disease-free survival was 50% for patients with a positive CRM, and 79% for patients with a negative CRM. This study included TNM stages I-III, and these patients received short-term neoadjuvant radiotherapy or postoperative chemoradiotherapy. Nagtegaal et al., a large multicenter trial in the Netherlands, included non-irradiated patients with rectal cancer between 1996 and 1999.[58] Local recurrences within 2 years was present in 16.4% of the patients with a positive CRM (≤1 mm), 14.9% of the patients with a CRM of 1.1-2.0 mm, and 2-4-10.3% of the patients with a CRM >2 mm. The local recurrence was significantly lower for margins >2 mm compared to margins ≤2 mm. Besides, the study showed that the 2-year survival was 69.7% for patients with a positive CRM (≤1 mm), and 90% for patients with a negative CRM. Furthermore, distant metastases occurred in 37.6% of patients with a positive CRM and 12.7% in patients with a negative CRM. Alberda et al., a retrospective study with patients that underwent surgery for locally recurrent rectal cancer between 1990-2013 in the Netherlands, reported similarly that a tumor-free margin of >2 mm is related to higher overall survival and significantly lower local recurrence than tumor-free margins between 0-2 mm.[16] In this study population, 29 of the 150 included patients who received neoadjuvant (chemo)radiotherapy. In summary, a positive CRM after rectal cancer surgery is related to a higher chance of local recurrence, distant metastases, and poorer overall survival.

With all the risks considered, it is important to avoid a positive CRM, which is challenging due to the neoadjuvant effects. Furthermore, the surrounding healthy colorectal tissue, with blood vessels, nerves, and ureters must be prevented as well.[4] Damage to these structures can lead to complications such as urogenital and anorectal dysfunction.[59,60] Intraoperative tissue recognition can be a helpful tool to avoid positive resection margins while preventing the surrounding healthy tissue.[10,61]

One of the techniques that have been used for intraoperative tissue recognition in colorectal cancer, is diffuse reflectance spectroscopy. Furthermore, B-mode ultrasound and ultrasound elastography are used for colorectal tumor assessment, to show the tumor grade, or to estimate response on neoadjuvant therapy. These three techniques will be explained and discussed in the following chapter.

3 Technical Background

This technical background explains the technical properties, applications in the field of (colorectal) cancer, advantages, and limitations for diffuse reflectance spectroscopy (DRS), ultrasound (US), and US elastography. Furthermore, the basics of machine learning are explained. This background serves as preparation for the ex vivo studies that are performed using these techniques. Firstly, the background of DRS is shown because this technique has already been used for distinguishing colorectal tumors from fat, the healthy colorectal wall, and fibrosis. Thereafter, the background of ultrasound is shown because this technique has been used to assess the tumor grade of colorectal tumors and may be useful to compensate for the limitations of DRS. Furthermore, it is needed to understand the ultrasound technology before explaining ultrasound elastography. Thirdly, the background of US elastography is shown because it has been shown that a combination of US B-mode and US elastography is superior to only US B-mode to assess colorectal tumors. Lastly, the background of machine learning is explained because one ex vivo study uses machine learning in the classification of tumor tissue and healthy colorectal tissue, and the other ex vivo study uses it in regression analysis to estimate the resection margin.

3.1 Diffuse reflectance spectroscopy

One of the techniques that can be used for intraoperative tissue recognition is DRS. In this section, the DRS technology will be introduced first and thereafter the four previous studies at the NKI on the application of DRS in margin assessment will be discussed.[4,10,21,22]

3.1.1 Physics of diffuse reflectance spectroscopy

DRS is an optical imaging technique that uses a broadband white light that is sent through an emitting fiber into the tissue. The light interacts with tissue through absorption and scattering. The reflected light can be measured by a spectrometer through receiving fibers. Based on the tissue interaction, a specific spectrum can be drawn from the reflected light, as shown in Figure **3-1**. The reflected spectrum shows characteristics of individual tissue types and can be used to distinguish tumor tissue from healthy tissue. DRS can be used for superficial measurements up to some millimeters in depth. The measurement depth of the DRS probe depends on the optical properties of the tissue, the wavelength of the light, and the distance between the emitting and receiving fibers.[21] As a rule of thumb, the sampling depth is similar to the distance between the emitting and receiving fibers.[62]

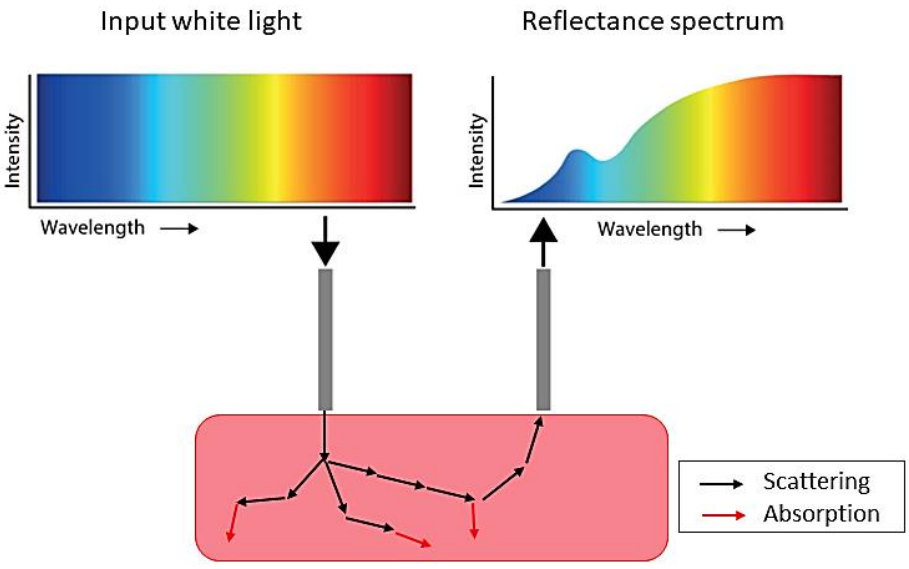


Figure 3-1: Basics of diffuse reflectance spectroscopy [Adapted from Baltussen et al.(2019)[4]]

The absorption depends on the absorption spectra of different chromophores in the probed tissue. Figure **3-2** shows the absorption spectra of some chromophores for visible (VIS) and near-infrared (NIR) wavelength ranges. It shows that deoxygenated hemoglobin (Hb), oxygenated hemoglobin (HbO_2) have mainly absorption in the VIS wavelength range. Although these chromophores are interesting to distinguish benign from malignant tissue, it is not practical to use the absorption of these chromophores during surgery since blood is present in the surgical area.[61,63] Since the absorption of blood is negligible for wavelengths in the NIR range, this range is more interesting for an intraoperative approach. Other chromophores that dominate the reflectance spectra and that can be used to distinguish benign from malignant tissue are lipid and water. Although lipid is mainly found in the fat tissue such as the perirectal fat, it is also present in other tissue since lipids are the main components of cell membranes. The main absorbance of lipid is shown in 1211 nm and secondly around 1393 and 1414 nm. The water absorption is mainly at 1453 nm. The water content is higher in malignant tissue compared to the surrounding healthy tissue.[61,63] However, this content is very dependent on the cancer type and healthy tissue type. Through the different concentrations of chromophores in tumor tissue and healthy colorectal tissue, the reflectance spectra of these tissue types show different characteristic peaks.

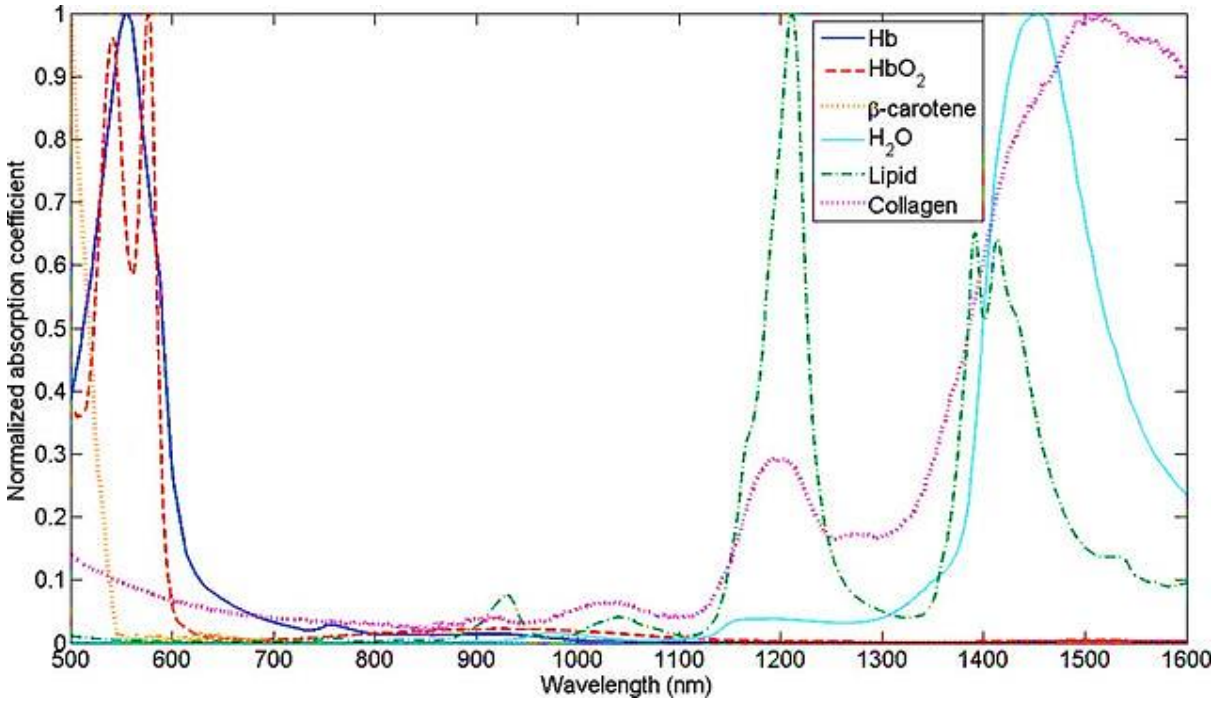


Figure 3-2: Absorption coefficient per chromophore for the visible (400 – 700 nm) and the near-infrared (800-1600) wavelength ranges [Adapted from: Nachabe et al.[64]]

Light scattering can be caused by Rayleigh and Lorentz-Mie.[65] Both processes describe scattering caused by small particles, bubbles, surface roughness, droplets, cells, fibers, density fluctuations, crystalline defects, and micro-organelles. The scattering process depends on the wavelength. Rayleigh scattering occurs when the particles are much smaller in diameter than the wavelengths of the electromagnetic radiation. On the other hand, Lorentz-Mie scattering occurs when the particle sizes are larger than the wavelength. Lorentz-Mie scattering is generally predominant for NIR spectroscopy since the larger particles that are present in the probed tissue dominate the light interactions relative to the small particles. Lorentz-Mie is anisotropic, dependent on the shapes of the scattering particles, and not strongly wavelength-dependent.

Next to the tissue interactions absorption and scattering, the resulting reflectance spectra depend on other factors that are not related to the tissue composition. Examples of these factors are the intensity of the emitted light, the distance between the light source and the tissue, the distance between the tissue and the spectrometer, the environment light, the temperature of the environment and the tissue, and the

sensitivity of the spectrometer. All these factors can result in unwanted effects in the spectra that are not needed to measure the tissue composition. Therefore, calibration is performed before DRS measurements to compensate for these effects. Generally, a white reflectance standard is used to measure the calibration spectrum, since a white standard reflects 100% of the emitted light.[66] Furthermore, the light intensity of the background is measured after the calibration and after each measurement. Although DRS measurements are performed in a dark room, there is always some background light. The resulting reflectance spectrum can be described as a function of wavelengths with Equation **3-1.**

$$R_{cal} = \frac{R_{meas} - R_{dark}}{R_{white} - R_{dark'}} \quad (3\text{-}1)$$

Where R_{cal} is the calibrated tissue spectrum, R_{meas} the measured tissue spectrum, R_{dark} the measured spectrum of the dark background, R_{white} the measured spectrum of the white reference sample, and R_{dark}' the measured spectrum of the dark background during calibration. Subsequently, the measured tissue spectrum can be used to extract optical properties such as the absorption coefficient and scattering coefficients per wavelength. The relation between reflectance, absorbance and scattering can be described using mathematical model.[67] However, these mathematical models are theoretically applicable for homogeneous tissue. Since the probed tissue of ex vivo specimens is heterogeneous, these mathematical models are not always working properly. Therefore, machine learning algorithms are currently used to distinguish tissue types based on DRS spectra.

The advantages of DRS are that it is non-destructive, does not require exogenous contrast agents, and has the potential to be performed in real-time. DRS technology has already been successfully evaluated in multiple oncological domains for discriminating tumor tissue from healthy tissue with classification accuracies of 0.86–1. [10,21,61,62,68–71] Baltussen et al. showed in several studies that the DRS can be used for margin assessment during rectal cancer surgery.[4,10,21,22]

3.1.1 Applications of diffuse reflectance spectroscopy in colorectal cancer

Baltussen et al. (2017) performed an ex vivo study on colorectal tumor specimens with DRS.[10] The DRS measurements were performed within one hour after excision. The probe that was used for these measurements contained three fibers, as shown in Figure **3-3**. The fibers consisted of one emitting fiber with a halogen broadband light source, one receiving fiber connected to a spectrometer 400 – 1100 nm (VIS range), and one receiving fiber connected to a spectrometer of 900-1700 nm (NIR range). The distance between the receiving and emitting fibers was 2 mm. Two classifiers were trained to classify tissue types based on DRS spectra. Firstly, a quadratic classifier was trained to distinguish fat from tumor tissue and healthy colorectal wall. Secondly, a linear support vector machine was trained to discriminate tumor tissue from the healthy colorectal wall.

Using both classifiers, it was possible to distinguish tumor tissue from fat and healthy colorectal tissue with an accuracy of 0.95. An example of DRS spectra of the three tissue types is shown in Figure **3-4**. The high accuracy of the classification can be due to the low number of included patients with a rectum tumor (2/38). The specimens of colon tumors are generally easier to interpret since these patients did not receive radiotherapy, in contrast to patients with rectal cancer. Through radiotherapy effects, the rectal cancer specimens generally contain fibrosis. Since fibrosis is a major problem during surgery and is hard to distinguish from tumor tissue, it is important to investigate whether DRS can help to discriminate tumor from fibrosis as well. Another reason for the high accuracy can be that the measurement samples that were located on the border between of two tissue types were excluded from this analysis.

The study did a separate analysis of these samples and achieved an accuracy of 80% for the tissue type that was most prominently present in the measured volume. The lower accuracy is logical since it is more challenging to recognize a measurement volume as healthy colorectal tissue, in case 40% of the

volume is tumor tissue. On the other hand, it would be more necessary to recognize all tumor cells, even if the tumor volume is in the minority.
Otherwise, small tumor deposits can be neglected which can result in positive margins. Another limitation is that the maximum measurement depth of this DRS system was ~2 mm. Despite a depth of 2 mm is enough to distinguish between a positive (≤1 mm) and negative (>1 mm) CRM, a larger sampling depth can give more information on the tissue type of deeper layers. This information is useful since it gives the surgeon more certainty for a negative margin when a larger tissue volume consists of healthy colorectal tissue instead of only 2 mm. Furthermore, when spectra of multiple fiber distances are compared, the differences between spectra can be used to observe tissue boundaries at a certain depth. Subsequently, this information can be used to estimate the CRM peri-operatively.

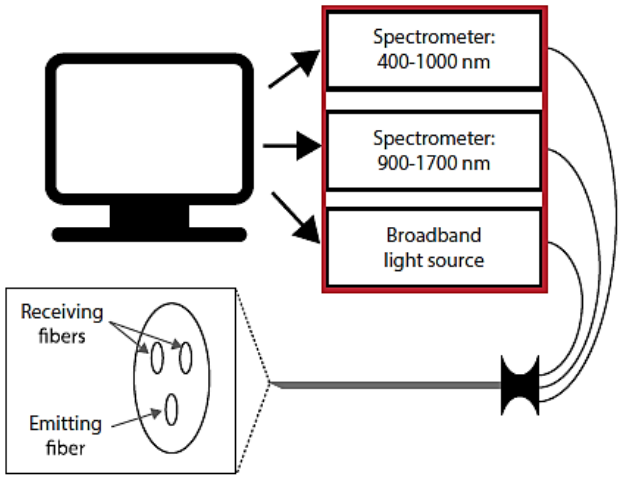


Figure 3-3: Schematic DRS measurement system which was used in the in vitro, ex vivo, and in vivo colorectal studies of the NKI-AvL.[4,10,21] The system consists of two spectrometers and a broadband light source, which were all controlled by a computer. Measurements are performed using a probe that includes three fibers. One emitting fiber transports the light from the broadband light source. The two receiving fibers transport the reflected light from the tissue to the two spectrometers. The fiber distance between the emitting and receiving fibers was 2 mm in the in vitro and ex vivo study. [10,21] In contrast, the fiber distance was 1.29 mm for the in vivo study.[4] [Adapted from Baltussen et al. (2019)][4]

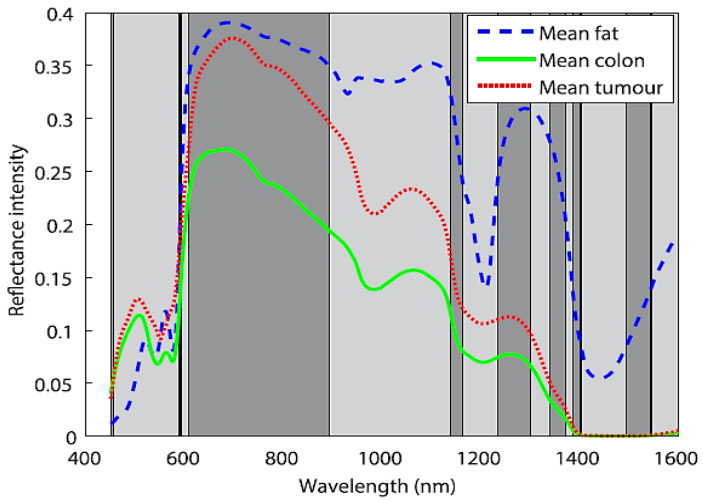


Figure 3-4: Averaged diffuse reflectance spectra of fat, healthy colorectal tissue, and colorectal tumor tissue. The mean spectrum of fat (blue) was based on 134 measurement locations on peri-colonic/peri-rectal fat. The mean spectrum of the colon (green) was based on 150 measurements on the healthy colorectal wall. The mean spectrum of tumor tissue (red) was based on 164 measurement locations at colorectal tumor tissue. The fat spectrum shows absorption at 1211 nm, which is logical through the high concentration of lipid compared to the other tissue types. The colon spectrum shows less reflectance and thus more absorption around 700 nm, that may be caused through a higher concentration of Hb than in tumor and fat. The tumor spectrum has the same shape as the colon spectrum but shows less absorption over in the range of 600-1400 nm. [Adapted from Baltussen et al. (2017)[10]]

Baltussen et al. (2019) performed a second ex vivo study on specimens from patients who received neoadjuvant therapy.[21] The purpose was to create a classifier to distinguish fibrosis from tumor tissue in rectal specimens with DRS. The same DRS probe as the previous study was used to collect DRS spectra. Two classification analyses were performed. Firstly, a linear SVM classifier was trained based on 96 pure fibrosis measurements, and 96 pure tumor measurements. The measurement sample volume was defined as pure when only one tissue type was present within a depth of 2 mm. Since only five pure tumor measurements were retrieved in this study, pure tumor measurements from the previously described study were also used for the first analysis.[10] This classification resulted in a mean accuracy of 0.88, a sensitivity of 0.91, a specificity of 0.86, and a Matthews Correlation Coefficient (MCC) of 0.76. These performance measures are further explained in Section **3.4.1**. Secondly, a linear SVM classifier was trained on tumor samples including mixtures with other tissue types. The fibrosis class still consisted of pure fibrosis samples. When tumor tissue was present within 3 mm from the resection surface, independently from the tissue type on top of the tumor, the measurement was classified as tumor tissue. In 68% of the mixed tumor samples, fibrosis was on top of tumor tissue. The second classification resulted in a lower performance with an accuracy of 0.61, a sensitivity of 0.51, a specificity of 0.66, and an MCC of 0.17. In conclusion, it may be possible to differentiate tumor from fibrosis as separate entities. However, when there is a healthy tissue layer on top of tumor tissue, which is mostly fibrosis, it becomes much harder.

Thirdly, an in vivo study was performed to assess the accuracy of DRS in the discrimination of colorectal tumor tissue from healthy tissue.[4] The accuracy of DRS was compared to the surgeon's judgment. The study population consisted of 32 patients with colorectal cancer with a cT3 or cT4 stage. The DRS measurements were performed using a probe with two receiving fibers and one emitting fiber with a fiber distance of 1.29 mm. The measurements were performed during colorectal surgery. The surgeons were asked to acquire DRS measurements of healthy fat, healthy colorectal wall, and tumor tissue. Furthermore, the surgeons were asked to indicate how certain they were that the tumor measurements contained tumor tissue. Patients were excluded when the surgeon was unable to localize superficial tumor tissue. The pathology coupes that corresponded to the measurement locations were used for the tissue type labeling with tumor tissue, fat, or healthy colorectal wall. A measurement was labeled as tumor when tumor tissue was present within 1.5 mm from the surface according to the pathology coupe. The classification of complete DRS spectra consisted of two steps. Firstly, an SVM classifier was used to distinguish fat from the healthy colorectal wall and tumor tissue. Subsequently, the samples that were classified as fat were excluded, and a second SVM classifier was used to distinguish the healthy colorectal wall from tumor tissue.
The study population consisted of 32 patients with colorectal cancer. Five patients had rectal cancer, four of them received neoadjuvant (chemo)radiotherapy, and one of them received neoadjuvant chemotherapy only. The classification of tumor resulted in an accuracy of 0.94, sensitivity of 0.90, specificity of 0.94, and an MCC of 0.73. The study also investigated the influence of the distance between the measurement surface to the tumor. They showed that the accuracy and MCC were optimal when the depth was 1-1.5 mm. Furthermore, the accuracy and MCC decreased when the depth was 2 mm or more. This result follows the rule of thumb that the measurement depth is approximately the same as the fiber distance, in this case, 1.29 mm.
Lastly, the classification results were compared to the surgeon's judgment. The comparison was made for the measurement locations of which the surgeon was uncertain whether tumor tissue was present or not. The surgeon indicated to be uncertain in 54 out of 270 locations. The threshold of the classification was adjusted such that no false-negative classifications were obtained on the locations of which the surgeon was uncertain. The classification resulted in 25% of the healthy tissue locations that were falsely classified as tumor tissue. Whereas the surgeon's judgment resulted in 69% of the healthy tissue locations that were falsely classified as tumor tissue.
Thus, classification based on DRS causes a large decrease in the number of false-positive locations compared to the surgeons' judgment. Therefore, DRS might result in less extensive surgery and fewer

complications after surgery. The tumor tissue was correctly classified by both the DRS classifier and the surgeon, resulting in a true-positive rate of 100%. However, there were only two measurements with tumor tissue, of all 54 measurements that were indicated as uncertain by the surgeon. Therefore, more data is needed to investigate whether DRS is better in detecting tumor involvement and prevents positive resection margins. Another study limitation was that only five of the 32 measured patients had rectal cancer, and four of these patients received neoadjuvant radiotherapy. Whereas especially the radiotherapy effects are hard to distinguish from tumor tissue. Therefore, these patients are particularly of interest in avoiding positive margins and preventing too extensive surgery. Lastly, the measurement depth was 1.5 mm. Although 1 mm is enough for a negative CRM, a tumor-free margin of 2 mm would be better in terms of survival and local recurrence. Furthermore, information about the tissue layers beneath the surface can provide useful information for the surgeon, as explained before.

The fourth study was performed to optimize the algorithm for tissue classification in colorectal cancer based on DRS.[22] Nine different DRS datasets were used for optimizing the algorithm. The datasets were from ex vivo studies and one in vivo study on colorectal cancer. The fiber distance differed per data set and ranged from 0.3 to 2.0 mm. Furthermore, some datasets were based on a reusable fiberoptic probe whereas other datasets were based on a disposable fiberoptic probe.
The algorithm that needed to be optimized consisted of three steps: normalization of the DRS spectra, feature extraction, and classification between healthy colorectal muscle and tumor tissue. Several techniques in all three steps were compared to determine the optimal classification strategy. Normalization techniques can be used to compensate for intensity differences that are not caused by tissue properties. For example, a variating tissue thickness or the coupling between the fibers and spectrometer might cause a variation in the detection efficiency of the signal. The normalization techniques that were applied were standard normal variate (SNV), normalization at 800 nm, and using the area under the curve (AUC). With the SNV method, each spectrum is transformed independently to remove slope variations on an individual basis. Each spectrum is normalized based on its mean and standard deviation. The normalization at 800 nm was used since it was assumed that no absorption was present at this wavelength. Therefore, the intensity of all reflectance spectra should be 1 at 800 nm. The AUC normalization technique sets all AUCs to 1.
Thereafter, four feature extraction techniques were evaluated. The first technique was the extraction of the intensity values, locations, and the right side gradient from the seven most prominent peaks. The second technique was the selection of 23 features that describe the coarse shape of the spectra while emphasizing the differences between different tissue classes. The third technique consisted of spectral band selection using k-means clustering. The final feature extraction technique consisted of 12 optical features that describe the composition of the measured tissue.
Subsequently, five different supervised classification models were used to distinguish colon wall from tumor tissue. The classifiers were k-nearest neighbor (kNN), a decision tree classifier, linear discriminant analysis (LDA), a linear SVM, and a neural network.
The results showed that normalization does not seem to have added value to non-normalization in the case of using the same DRS probe within one dataset. When multiple DRS probes are used, the normalization at 800 nm and SNV normalization improved the classification for some datasets. The best combination of a normalization technique, a feature extraction method, and a classifier that resulted in the highest MCC and accuracy was different per dataset. Generally, the best normalization techniques were normalization at 800 nm and SNV normalization. The best feature extraction methods were mainly spectral band selection and tissue optical features. Lastly, the best classifiers were SVM, LDA, and neural networks.
The results show that a classification algorithm must be optimized per application. For example, when another DRS probe is used than in these datasets, the optimal combination might be different. However, the optimal combinations that are found in this study can be used as starting point for further data analysis of DRS data in colorectal ex vivo studies.

In summary, several previous studies on the application of DRS on the margin assessment during colorectal surgery have been performed, which showed that DRS is a promising technique to distinguish colorectal tumor tissue and healthy tissue and to help the surgeon with decision making intraoperatively. However, the system is not ready to be applied yet due to several limitations:

- It is difficult to collect a sufficient large in vivo data set as the correct selection of superficial tumor locations is limited, since the surgeon is not always able to select those.
- A limited number of measurements was performed on specimens from patients who received neoadjuvant radiotherapy, while surgery with a negative CRM is especially challenging for these patients.
- DRS has not been proven to distinguish pure fibrotic from tumor tissue with a layer of fibrosis or healthy tissue on top. Whereas this is one of the challenges for surgeons in patients who received radiotherapy.
- The maximum measurement depth was 1-2 mm, whereas a depth of 5 mm provides more certainty on the resection margin. On the other hand, when the measurement volume of DRS increases, it becomes harder to estimate the tissue boundaries within the volume, which is needed to discriminate between a positive and negative resection margin.

To solve these shortcomings, an additional technology is needed that can:

- Select optimum measurement sites.
- Better differentiate fibrosis from tumor tissue.
- Differentiate multiple tissue layers within a measurement volume.
- Acquire information of deeper tissue layers than 2 mm.
- Be used intraoperatively during colorectal surgery.

3.2 Ultrasound

Ultrasound (US) can be used to visualize the tumor and the surrounding tissue layers on the locations of the DRS measurements. Furthermore, US can be used to acquire information of tissue layers up to ~4 cm in-depth, which is sufficient to visualize a complete cross-section of the rectum or colon. Ultrasound is of benefit to other imaging techniques because it gives direct visual feedback, it is low-cost, it is accessible in a surgical environment, and it is non-invasive. The current section includes the physical background of ultrasound, applications in the field of colorectal cancer, and the strengths and limitations of the technique.

3.2.1 Physics of ultrasound

Ultrasound is an imaging technique that uses high-frequency sound waves for medical imaging (typically 2-18 MHz).[72] The piezo elements in an ultrasound transducer convert electrical signals to acoustic signals, which produces pulses of ultrasound.[73] These pulses undergo tissue interactions such as absorption, refraction, reflection, and scattering. The piezo elements convert the reflected sound waves into electrical pulses. These pulses are interpreted into a two-dimensional greyscale image of the anatomy, known as the brightness mode (B-mode). The more energy in the returning echoes, the brighter the image. Reflection occurs at tissue boundaries due to the difference in acoustic impedance. The larger the difference in acoustic impedance, the higher the energy of a reflected pulse, and the brighter the image. The acoustic impedance depends on the density and the speed of sound, which are tissue-specific properties. The depth of the tissue boundary can be determined based on the speed of sound in soft tissues (1540 ms^{-1}) and the time that is taken for the ultrasound pulse to travel the distance to the interface and for the reflected wave to return to the transducer.

The depth and resolution of ultrasound imaging depend on the frequency. The axial resolution is the depth for which two tissues can be discriminated, this depth is ~4 cm for transducers with 15 MHz and ~13 cm for transducers with a frequency of 5 MHz.[72] Furthermore, the lateral resolution is the distance at which two structures along the beam width can be distinguished. This distance is ~0.3-0.8 mm for 15-18 MHz.[74] High-frequency transducers (5-15 MHz) are used for imaging superficial structures with a higher resolution, and lower frequency transducers (2-5 MHz) are used for deeper structures. Since the current study focuses on imaging superficial structures, a high-frequency transducer is preferred.[73,75]

3.2.2 Applications of ultrasound in colorectal cancer

Ultrasound is not used as a standard in the diagnostics of colorectal cancer. Only endorectal ultrasound (ERUS) is advised in the Dutch guidelines in addition to MRI to discriminate cT1 from cT2 rectal tumors.[3] Ultrasound is helpful to distinguish the different tissue layers of the colon and rectum, and can therefore discriminate different tumor grades. ERUS is the most reported US approach in the assessment of rectal tumors.[9,25,26,28–33,76,77] Furthermore, some studies with transabdominal and intraoperative approaches have been reported.[27,78]

3.2.2.1 Endorectal ultrasound

The ERUS approach and an US B-mode image of the healthy rectal wall are shown in Figure **3-5**A. The ERUS approach shows that a transducer is inserted inside the rectum. Generally, a radial transducer is used with a 360º field of view and a frequency range of 6-16 MHz.[77] The US image shows that the tissue layers of the rectal wall can be distinguished based on alternating echogenicity. From the lumen towards the perirectal fat, the following layers can be distinguished: the interface between the lumen and mucosa (hyperechoic), the mucosa (hypoechoic), submucosa (hyperechoic), m. propia (hypoechoic), and serosa/perirectal fat (hyperechoic). During ERUS the lumen is generally empty through preparation with a laxative. To obtain an optimal contact surface between the transducer and tissue, a balloon surrounding the transducer is filled with water or ultrasound gel. The ultrasonic features for malignancy are a hypoechoic bowel wall thickening with irregular contour, the loss of stratification of the wall layers, and the absence of normal peristalses.[26,27] The tumor stage can be determined based on the level of invasion through the rectal wall, as described in Table **3-1**.[79] Endorectal ultrasound provides an accurate assessment of rectal tumors, with an accuracy between 64 to 95%.[26] The accuracy varies since it depends strongly on the tumor stage and neoadjuvant therapy.

Oien et al. used ERUS for pretreatment staging of rectal cancer.[76] They found that ERUS had an accuracy of 88% for the discrimination between benign adenomas from early rectal cancer. Whereas the accuracies were lower for distinguish T1 or T2 tumors, with accuracies of 0.66 and 0.57. T1 tumors were mostly overstaged as T2 tumors, and T2 tumors were mostly overstaged as T3 tumors. Pomerri et al. used ERUS for post-treatment staging.[24] They showed that the specificity and negative predictive value (NPV) were 92% and 95% for ERUS in discriminating ≤T3 tumors from T4 tumors. However, the number of patients with a ypT4 tumor was only 7, which was 8% of the total study population. Furthermore, they demonstrated that ERU was inaccurate in predicting the exact T-stage after chemoradiotherapy. The inaccuracy is mainly caused by the effects of radiotherapy. These tissue changes such as peritumoral inflammation, edema, fibrosis, and necrosis have a similar hypoechoic appearance as tumor lesions.[26,54,56] This may result in overstaging of lower T-stages (T0-T2).[25,30] The overstaging with ERUS was the highest for T0 tumors, with 66% overstaging.[24] Several studies also compared ERUS with ultrasound elastography in colorectal cancer approaches, these studies will be discussed in Section 3.3.2.[28,29,31–33,80]

Table 3-1: Tumor grade description at endorectal ultrasound[79]

Tumor grade	The description on B-mode Ultrasound
uT0	The tumor does not infiltrate the submucosa
uT1	The tumor infiltrates the submucosa
uT2	The tumor infiltrates the Muscularis propia
uT3	The tumor infiltrates the serosa or perirectal fat
uT4	The tumor involves adjacent organs or structures

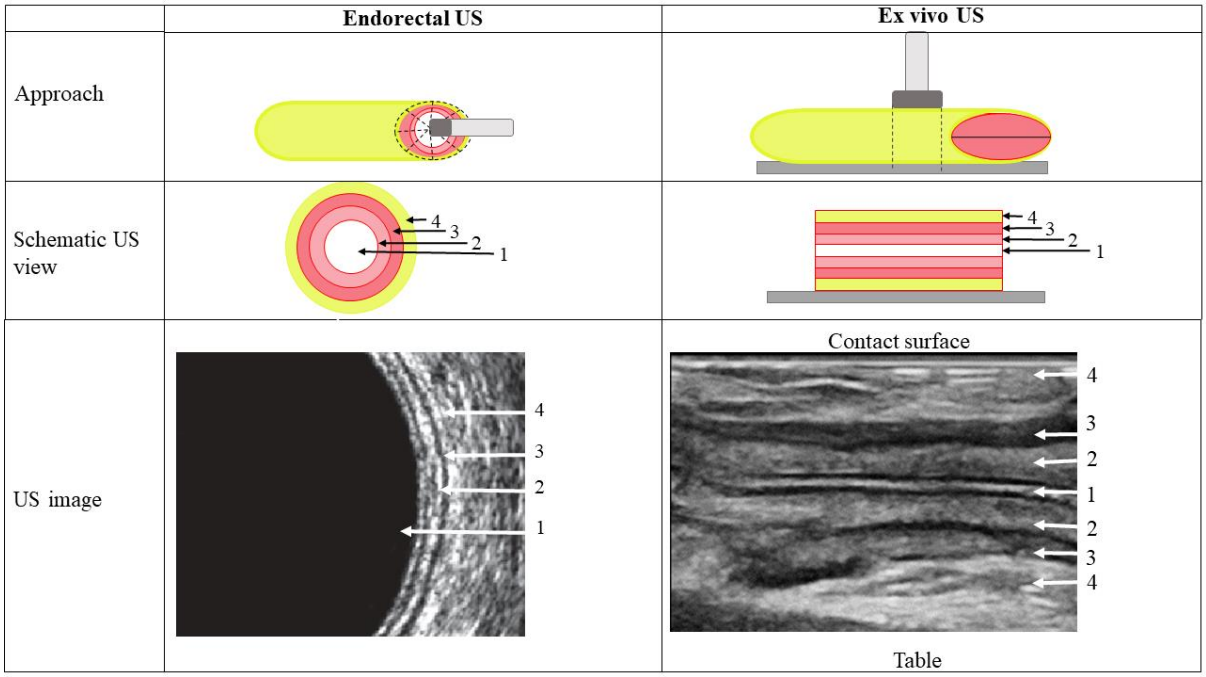


Figure 3-5: Endorectal ultrasound versus ex vivo ultrasound on a rectum specimen. The illustrations show the following layers: white and nr. 1 = lumen, light pink and nr. 2 = submucosa, dark pink and nr. 3 = Muscularis propia, yellow and nr. 4 = perirectal fat.

With ERUS, the transducer is placed endoluminal inside the rectum. The piezo elements inside the transducer can rotate, therefore, US images from multiple angles (dashed lines) can be combined into a 360º field of view. Since the tissue layers are thin relative to the lumen, a subpart of an ERUS image is enlarged to show the different layers. The lumen is shown as hypoechoic since the lumen contains water or gel. Thereafter, the inner muscle layer is shown as a hyperechoic layer, and the external muscle layer is shown as a hypoechoic layer. Subsequently, the perirectal fat is hyperechoic. [ERUS image adapted from Miro et al.[77]]

The ex vivo US is performed on a rectal specimen. The rectal specimen is enveloped in the perirectal fat (yellow), and the rectal wall (pink) is enclosed with sutures (black line) at the proximal and distal side. In the ex vivo US approach, the transducer is placed on top of the rectum specimen. The ultrasound image shows the tissue layers between the transducer and the table beneath the specimen. From top to bottom, the first layer is the perirectal fat (hyperechoic), the second layer is the outer muscle (hypoechoic), and the third layer is the inner muscle (hyperechoic). Thereafter, the lumen (hypo/hyperechoic) is shown. The lumen can be shown as hyperechoic when feces is present, or hypoechoic when air is present. Beneath the lumen, the tissue layers are repeated in the reversed order.

3.2.2.2 Transabdominal ultrasound

Tomizawa et al. used transabdominal US to diagnose colorectal cancer.[27] The study used the colorectal wall thickness to discriminate between the healthy colorectal wall and colorectal cancer. They found that the average wall thickness of healthy colorectal tissue was 2.8±0.4 mm and 12.7±5.2 mm in colorectal cancer. A threshold value of >4.3 mm was suspected for colorectal cancer.

3.2.2.3 Intraoperative ultrasound

Greif et al. is the only study on intraoperative US to localize colorectal tumors during surgery.[78] The study focused on the intraoperative approach of US in detecting colorectal adenomas and T1 tumors. It showed that intraoperative US was successful in localizing colorectal tumors, visualizing the bowel wall and its penetration by malignant tumors in 12/12 T1 lesions of colorectal cancer. Furthermore, they used US to determine the distal resection margin for rectal tumors. They managed to determine the distal resection margin for all six patients with rectal cancer. This can help the surgeon to determine whether a sphincter sparing surgery with a tumor-free margin of 2 cm is possible.

3.2.2.4 Ex vivo ultrasound

The current study will use an ex vivo approach of ultrasound. Since the approach differs from the standard imaging method ERUS, another US B-mode image of an ex vivo setting is shown in Figure 3-5B. The US approach shows that the transducer is directly placed on the surgical specimen of a colon or rectum tumor. From top to bottom, the following layers can be distinguished until the lumen: pericolic/perirectal fat (hyperechoic), the submucosa (hyperechoic), the muscularis propia (hypoechoic), and lastly the lumen (hyper/hypoechoic). The measurement depth is generally 4 cm, since the specimen is mostly about 1-2 cm thick, the table beneath the specimen is also shown on the US image. The table is usually shown as a bright line. The lumen looks different compared to the ERUS approach. First of all, it is flattened instead of round, and it can be hypoechoic in case of air, or hyperechoic in case of feces inside the lumen. Since the interface between the lumen and the colorectal wall differs from ERUS, the first two layers, the interface between the lumen and mucosa, and mucosa are not recognizable on the ex vivo US image. The accuracy of ex vivo US might differ from ERUS due to another transducer position. When the transducer is positioned directly on the outer wall of the bowel, instead of in the lumen, the penetration of T3 and T4 tumors are more superficial to the transducer. With a high-frequency probe, these superficial tissue layers are shown with a higher resolution than the deeper layers. Thus, the accuracy for a correct grading of T3 tumors with the ex vivo US may be higher in comparison to ERUS. The same ultrasonic features for malignancies that were mentioned for ERUS can be used for ex vivo US as well: hypoechoic bowel wall thickening with irregular contour, and the loss of stratification of the wall layers. The US images of specimens with rectal tumors from three different patients are shown in Figure **3-6**. All three patients received neoadjuvant chemoradiotherapy. The tumor grades are based on the pathological report on the specimen. Figure **3-6**A shows a T2 tumor with fibrosis as a radiotherapy effect. The tumor is recognizable at the location where the m. propia is interrupted. However, the tumor grade is hard to determine on US B-mode through the radiotherapy effects. According to the T2 stage, the tumor does not invade the serosa and perirectal fat, but on the US image, it might be interpreted as a T3 or T4 tumor. Figure **3-6**B shows a T3 tumor, therefore, the tumor invades the serosa or perirectal fat but does not invade adjacent tissues or organs. Figure **3-6**C shows a T4 tumor, which means that the tumor invades other organs/tissues. The stratification of the wall layers is completely lost at the side of the tumor. At the left side of the image, the normal structure of the healthy rectum wall is shown.

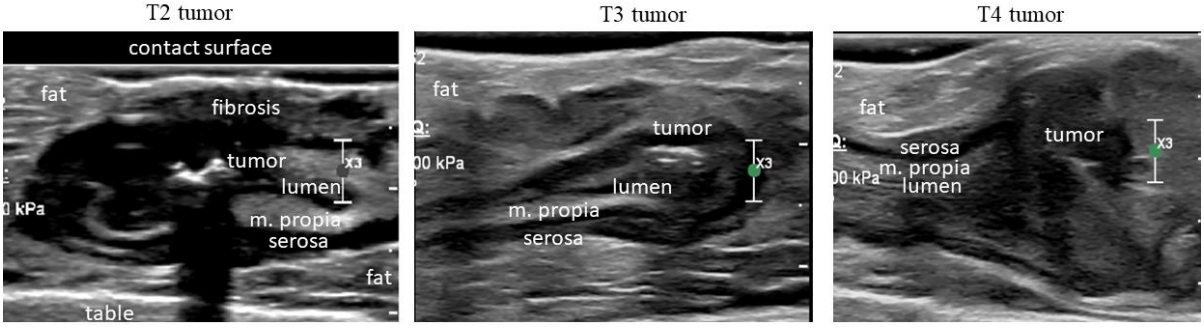


Figure 3-6: Ex vivo ultrasound of three rectum tumors. The tumor grade is according to the pathological report of the surgical specimens. The contact surface of the transducer is on top of the images, and the table is at the bottom of the image. A. US image of a specimen with a T2 tumor, where fibrotic effects are shown in the perirectal fat and serosa. B. US image of a specimen with a T3 tumor, the fat layer can be discriminated from the hypoechoic serosa. However, the Muscularis propia and serosa are interrupted by the tumor. C. US image of a specimen with a T4 tumor. The tissue wall stratification is completely lost at the right side of the image, and the tumor reaches into the perirectal fat and adjacent tissues/organs.

In summary, several studies show that US B-mode can be used to find the tumor location. This property of US can be useful to determine the measurement location of DRS. The additional values of US are that a larger measurement area can be assessed in one image, with high-frequency transducers upon ~4 cm in depth. Furthermore, the different layers of the colorectal wall can be clearly distinguished on US B-mode images. Especially the differences between healthy colorectal tissue or fat and malignant tumors are obvious. However, it is still challenging to distinguish between tumor and fibrosis. Therefore, an additional technique might be useful to improve this discrimination.

3.3 Ultrasound elastography

Ultrasound elastography can be a useful addition to DRS and the US B-mode since the technique adds information about tissue stiffness. Elastography images can be retrieved simultaneously with US B-mode images. US elastography is of additional value to ERUS in tumor grading and in discriminating between benign and malignant rectal tumors. Furthermore, US elastography is of additional value to B-mode US in therapy response assessment in breast cancer. The current section will describe the history and physical background of ultrasound elastography. Thereafter, the two US elastography techniques strain elastography and shear wave elastography will be explained and compared. Lastly, an overview of studies with USE will be shown, and relevant studies will be discussed.

3.3.1 History

Soft-tissue elasticity has been used for years as a diagnostic tool, as stiffness is a key property of abnormal tissue and organs.[81] Soft-tissue elasticity can principally be estimated by palpation. Although palpation is still an important part of physical examination, soft-tissue imaging methods are needed to characterize the masses and to detect lesions that are not palpable. Generally used imaging techniques such as MRI, CT, and US, assess the shape, size, and location of masses. However, it is not always possible to distinguish various types of masses and distinguish benign from malignant lesions. Other advanced imaging techniques such as nuclear imaging and diffusion-weighted imaging have shown their added value in distinguishing malignant from benign lesions by the addition of physiological information. However, none of these methods have been able to distinguish various types of masses based on stiffness. The addition of both qualitative and quantitative images of tissue stiffness offers a new way to help distinguish masses and diffuse growing diseases from one another. Ultrasound B-mode images have already been used to judge masses based on compressibility. The relative stiffness of a lesion can be visualized by measuring the diameter of a lesion before and during compression with the US transducer. Another US technique, Fremitus, uses vibrations induced by the patients humming to observe vibrations in the tissue using color and power Doppler. As stiff lesions vibrate less than softer tissues, these lesions become visible as a dark area surrounded by color. However, this method is inconsistent in the display of lesions due to variability in vibrations. The start of ultrasound elastography was in 1980. It started with the measurement of tissue motion based on the m-mode, color Doppler and vibrational Doppler. However, the image had a relatively low resolution and required an inconvenient external vibratory device to induce tissue motion. The currently used ultrasound elastography technologies are strain and shear wave elastography.

3.3.2 Physics of ultrasound elastography

Both strain and shear wave elastography are based on the physics of elasticity. Elasticity is generally described as the correlation between stress and strain., where stress is the applied force divided by the unit area and strain describes the tissue displacement to the original length. Three different elastic moduli describe the relationship between stress and strain depending on the direction of deformation, as shown in Figure **3-7**. The Young's modulus (E) describes the relationship between stress and strain in the case of longitudinal displacement, parallel to the direction of the applied force. The shear modulus (G) describes the elasticity based on tangential displacement to the surface. The bulk modulus (K) is defined as when a normal inward force or pressure produces a volume change. The higher the elastic modulus, the more material tends to resist deformation, which can be interpreted as increased stiffness.

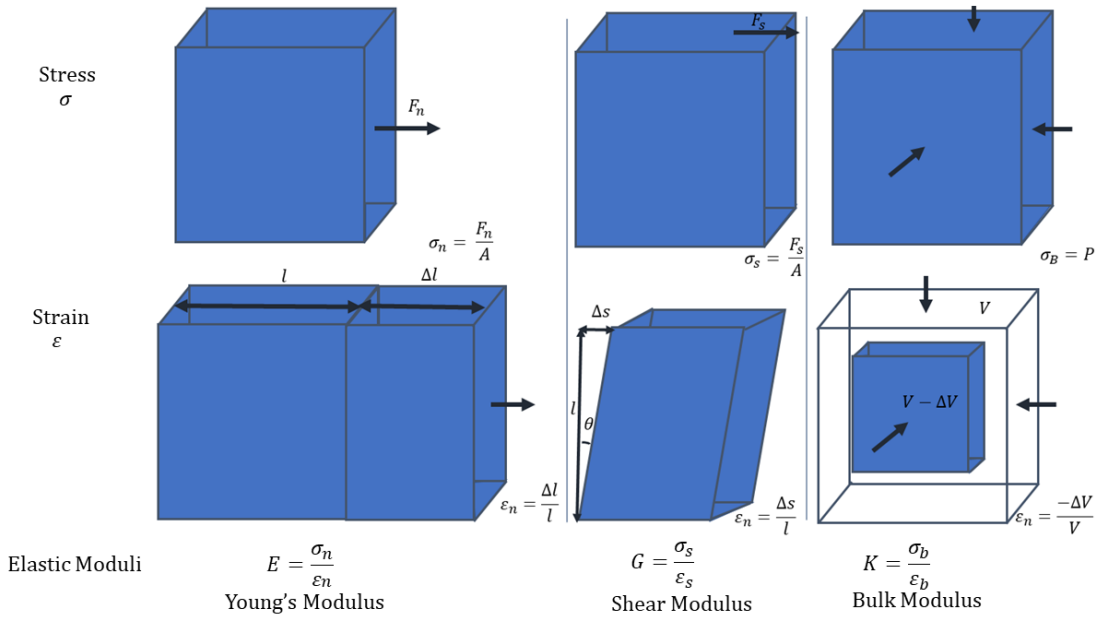


Figure 3-7: Tissue deformation described by stress **σ** (force per unit area), strain **ε** (expansion per unit length), and elastic moduli (stress divided by strain). The type of elastic modulus depends on the direction of the applied force and tissue displacement. The Young's modulus E is the relation between the normal stress $\boldsymbol{\sigma_n}$and normal strain $\boldsymbol{\varepsilon_n}$, where normal is perpendicular to the surface. The Shear modulus G depends on the shear stress $\boldsymbol{\sigma_s}$ which produces a shear strain $\boldsymbol{\varepsilon_s}$, where shear is tangential to the surface. The Bulk modulus K is defined when a normal inward force or pressure P that is equal to the bulk stress $\boldsymbol{\sigma_b}$, produces a volume change $\boldsymbol{\varepsilon_n}$.[Adapted from Sigris et al.[82]]

The elastic modulus also describes the propagation speed of waves through a material with a certain density. In ultrasound, two different wave speeds can be described. A schematic view of both waves, consisting of longitudinal waves and shear waves is shown in Figure **3-8**.[81] The longitudinal wave speed describes the wave speed in the propagation direction of the emitted waves by the US transducer. This speed in soft tissue is approximately 1540 m/s. The reflections of longitudinal waves are used to make B-mode images. The wave speed in the perpendicular direction to the propagation describes the shear wave speed. The shear wave speed is in the order of 1-10 m/s. However, the shear wave speed is not directly interpretable as tissue elasticity. Therefore, the shear wave speed can be converted to Young's modulus in kilo Pascal (kPa). The Young's modulus is directly interpretable for soft tissue elasticity, since a low Young's modulus describes a soft tissue, whereas a high Young's modulus describes a stiff tissue. The shear wave speed(c_s) can be converted to Young's Modulus with Equation **3-2**. The density(ρ) in kg/m^3 is assumed to be a constant of 1000 kg/m^3.[83]

$$E = 3G = 3\rho c_s^2 \qquad (3\text{-}2)$$

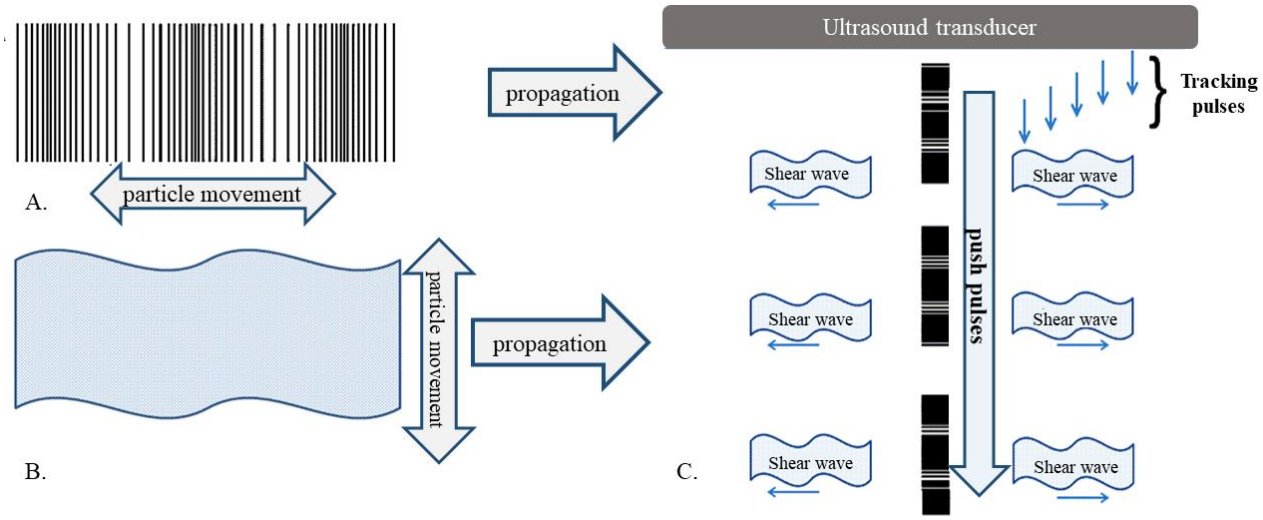


Figure 3-8: Schematic view of the longitudinal waves in ultrasound and the particle movement. A. Longitudinal compression waves with the wave speed. The particles making up the material move along the same axis as the direction of propagation. The longitudinal waves are used to make B-mode Ultrasound images. B. Shear waves with the wave speed. The particles making the material move along the perpendicular direction to the direction of propagation. C. Creation and tracking of the shear waves. The ultrasound transducer emits a beam consisting of a series of compressional wave pulses, which push the tissue away from the transducer and causes shear waves to be generated at the margins of the push beam. The shear waves travel away from the push pulse beam and are tracked by lower power tracking pulses, which are emitted by the transducer as well. In this way, the speed of the shear waves can be estimated [Adapted from Garra[81]]

3.3.3 Strain elastography

Strain elastography is a qualitative method to show tissue elasticity. The relative Young's modulus is estimated based on the tissue displacement in the direction of an applied force. In Figure **3-7** this can be seen as the normal strain ε_n, which is used in the calculation of Young's modulus. Since the stress applied is unknown, only a relative Young's modulus is shown. The applied stress can consist of manual compression on the tissue with the ultrasound transducer, but, the natural tremor of the operators' hand is also enough to produce a strain. Furthermore, tissue displacement as a result of internal physiologic motion, such as cardiorespiratory movements, may also be used for strain elastography. A strain elastography color map is exposed over a B-mode image, as shown in Figure **3-9**. In this strain map, the stiffer areas (compared to the green areas) are shown in red, and relative softer areas (compared to the green areas) are shown in blue. However, it is more typical in other studies that stiff tissue is displayed in blue, and soft tissue in red.[82] The green bar below the color bar shows whether enough stress is applied to produce a strain. When this bar is colored green instead of grey, the pressure is enough.

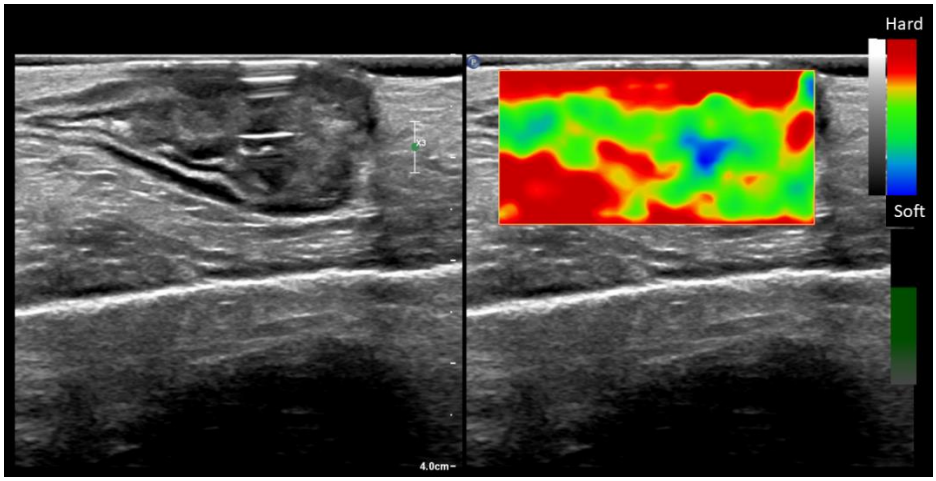


Figure 3-9: Example of a strain elastography ultrasound image of a surgical specimen from a sigmoid tumor. The stiffer areas relative to green areas are colored red, and relative softer areas are displayed in blue. The bar beneath the color bar is colored green when enough pressure is applied to provide a strain, when the pressure is insufficient, this bar is colored gray.

Parameters to use strain elastography as a semi-quantitative method, are strain ratio and the 5-point elasticity score.[28,31,80,84–90] The strain ratio gives the ratio between the mean strain in a reference tissue and the mean strain of the tumor. In colorectal approaches, the healthy colorectal wall or perirectal tissue can be used as reference tissue. The elasticity score is also known as the Ueno score or Tsukuba score and is illustrated in Figure **3-10**. The colors in this figure are adapted to the color bar in Figure **3-9**. The scoring system can be used to characterize suspicious hypoechoic lesions on the US B-mode image. The higher the score, the higher the chance that the lesion is malignant. A score of 1 corresponds to green in the entire hypoechoic lesion; A score of 2 corresponds to a mosaic pattern of green and red and deformity of the lesion; A score of 3 corresponds to red in the central part and green in the peripheral part of the tumor; A score of 4 corresponds to red in the entire hypoechoic tumor; Lastly, a score of 5 corresponds to red in the entire hypoechoic tumor and its surrounding area. This scoring system is based on breast tumors and not yet applied to colorectal tumors.

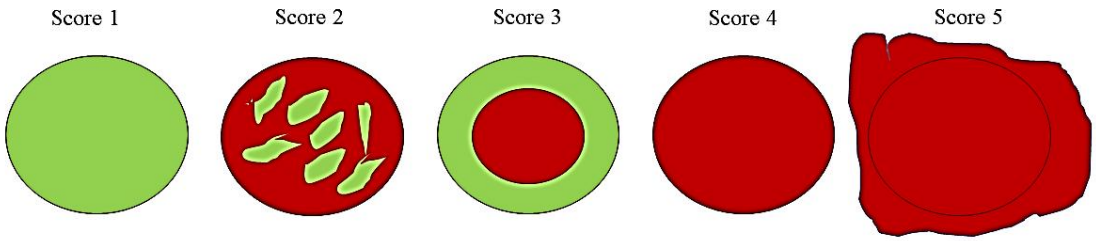


Figure 3-10: Ueno score based on strain elastography colors of suspicious hypoechoic lesions on US B-mode. The colors are adapted to the color bar that was used in the current study. Score 1 for an entire green area, showing the same stiffness in comparison to the surrounding area. Score 2 for a mosaic pattern of green and red, indicating a mixture of stiffness. Score 3 for red in the central part and green in the peripheral part of the tumor. Score 4 for red in the entire tumor, indicating that the tumor is stiffer than the surrounding area. Score 5 for red in the entire tumor and the surrounding area. [Adapted from Pardal et al. [91]]

3.3.1 Shear wave elastography

Shear wave elastography (SWE) uses dynamic stress, consisting of compressional wave pulses, to produce shear waves. The wave pulses result in tissue displacement of 1-2 µm which causes shear waves in a perpendicular direction, as shown in Figure **3-8**C. The speed of the shear waves (c_s) can be measured directly with ultrasound and converted into Young's modulus. The conversion to Young's modulus is made since Young's modulus says directly something about the stiffness of a tissue. Therefore, the pixels with a high Young's modulus, colored as red, represent stiff tissues, whereas pixels with a low Young's modulus, colored as blue, represent soft tissues. Subsequently, stiff tissues have a higher probability to be malignant. An SWE image of a surgical specimen from a sigmoid tumor is shown in Figure **3-11**. The SWE image is shown as a color map, superimposed on a B-mode image. Each pixel represents Young's modulus in kPa. The color bar ranges from 0 to 200 kPa. Since each pixel has a stiffness value, multiple statistic metrics can be retrieved from a region of interest, such as the mean, maximum, minimum, interquartile range (IQR), and standard deviation.

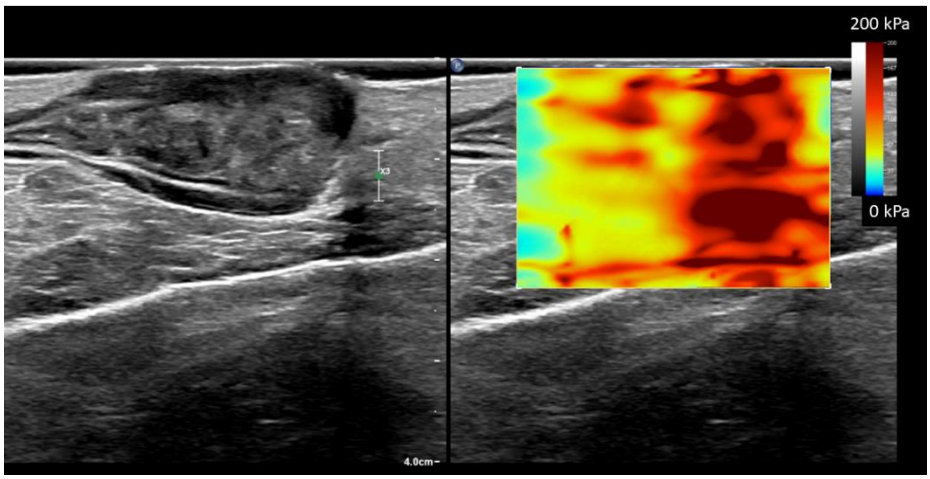


Figure 3-11: Ultrasound shear wave elastography image of a surgical specimen from a sigmoid tumor. The color bar ranges from 0 – 200 kPa, with blue representing soft tissues red representing stiff tissues.

3.3.1 Comparison of strain and shear wave elastography

Strain elastography and shear wave elastography are both methods that can be applied in clinical practice. The two techniques differ in multiple aspects, which makes one technique more suitable for certain applications than the other one. The most important differences are shown in Table **3-2**. For the application of ultrasound elastography as an intraoperative approach for colorectal cancer, the maximum depth is not a limiting factor since the ultrasound transducer can be placed directly on the bowel wall. The quantification of shear wave elastography is an advantage since the stiffness values and variables of the stiffness in a region of interest can be used for an objective judgment. The ability of shear wave elastography to assess diffuse lesions better than focal lesions is an advantage as well, as two of the characteristics for colorectal tumors are the irregular contour and the loss of stratification of the wall layers.

Table 3-2: Comparison between strain- and shear wave elastography [Adapted from Garra [81]]

	Strain elastography	**Shear wave elastography**
Image quality	Excellent if properly performed	Good, when the operator is trained in US assessment
Real-time imaging	Yes	No, 0,6-2,5 sec delay due to low frame rate (Philips: 0.4-1.6 Hz[5] for SWE and 20-30 Hz for B-mode[92])
Maximum depth	Good. Depends on applied force.	Limited. Typically 6 cm or less for good quality
Quantification	Qualitative/Semi-quantitative with strain ratio	Excellent. Quantification of stiffness from a region of interest or per pixel and color display is calibrated.
Diffuse/focal disease	Better for focal than diffuse lesions. Since the image is not calibrated, diffuse stiffness may appear the same on an image as diffuse softness	Better for diffuse than focal lesions. Algorithms and computations are designed for diffuse disease. It may give erroneous values in focal lesions
Operator dependence	Considerable. Hands-on training usually required	Limited if the operator already knows how to use ultrasound

3.3.2 Applications of ultrasound elastography

Ultrasound elastography has been used in the oncological field to determine the tumor grade[31,32,35,80], to distinguish between benign and malignant tumors[28,34,87,89,90,93], to predict lymph node metastases[33], to assess therapy response[88,94], or to predict tumor deposits[29]. The additional value of ultrasound elastography has been studied as well for distinguishing between fibrosis stages of liver fibrosis.[95–98] An overview of recent studies with strain elastography is shown in Table **3-3**, and an overview of studies with shear wave elastography is shown in Table **3-4**. The endpoint of each study shows for which distinction the elastography was used. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve is shown as a performance metric. In case the AUC was not available, the sensitivity, specificity, and accuracy are shown by selecting a cutoff value that was reported. It shows that the AUC varies from 0.623 to 0.943 for strain elastography. The performance for shear wave elastography ranges from 0.857 to 0.981. The studies which are of interest for the current project will be discussed more extensively in the following subsections.

3.3.2.1 Ultrasound elastography versus Brightness-mode ultrasound

Multiple studies showed that a combination of B-mode and elastography variables increases the AUC in comparison to one of the two techniques.[32,35,80,87,89,94,99]

Waage et al. showed that strain elastography with a strain ratio threshold of 1.25 was superior to ERUS and MRI in the differentiation of benign rectal adenomas and malignant adenocarcinomas.[28] The sensitivity, specificity, and accuracy were as follows: ERUS: 0.96, 0.62, and 0.90; Strain ratio with cut-off of 1.25: 0.96, 0.86, and 0.94, and MRI: 0.99, 0.07, and 0.87. The study showed as well that the strain ratio did not differ for patients who received chemoradiotherapy and those who did not. The histopathological label ‘adenocarcinoma’ was assigned based on resection specimens. However, in the case of patients who received neoadjuvant therapy, the specimen was verified as adenocarcinoma based on biopsy before radiation therapy. Since the ERUS, strain elastography and MRI were retrieved between the neoadjuvant therapy and surgery, a complete pathologic response to chemoradiotherapy was not taken into account.

Fan et al. and Chen et al. used SWE for preoperative tumor staging in rectal cancer.[32,35] Fan et al. compared SWE to ERUS and MRI in the staging of rectal cancer before surgery. They demonstrated that a maximum elasticity modulus of ≥90.7 kPa was distinctive for locally advanced tumors (T3-4) compared to T2 and T1. The sensitivity, specificity, and accuracy were 83%, 92%, and 89%, respectively. To discriminate T1 from T2, the best cutoff value for the maximum elasticity was <65.0 kPa. The study did not report a cutoff value to distinguish T3 and T4, since the number of tumors with

T4 was small (3 cases). When using <65.0 kPa for T1, 65.0-90.7 kPa for T2, and ≥90.7 kPa for T3 and T4 tumors, the overall accuracy was 85.5% (47/55). For ERUS the accuracy was 78.2% (43/55) and for MRI it was 74.6% (41/55). Therefore, SWE had a higher accuracy than ERUS and MRI, although, not statistically significant. Chen et al. compared SWE to ERUS to stage rectal adenomas and adenocarcinomas (T1-T3) before surgery. They reported cut-off values of the mean elasticity modulus for T1, T2, and T3 cancers, 26.9 kPa, 70.3 kPa, and 112.0 kPa respectively. With the cutoff value of 26.9 kPa, all adenomas could be discriminated from adenocarcinomas. Furthermore, the cutoff value of >112 kPa resulted in the discrimination of T3 from adenomas and T1-2 tumors with a sensitivity, specificity of 98% and 100%. The overall accuracy for correlation with the pathology was 95.7% (67/70) for SWE and 75.7% (53/70) for ERUS. As a result, the diagnostic performance was significantly higher for SWE compared to ERUS. Furthermore, they demonstrated that SWE may reduce observer dependency compared to ERUS since the concordance rate between two observers was 78.6% for ERUS and 85.7% for SWE. The limitation of both studies was that patients who received neoadjuvant CRT were excluded, while the preoperative staging is especially challenging in these patients.

The combination of radiomic features from SWE, ERUS, and CT has been studied by another study of Chen et al.[33] They developed a multiparametric radiomics model of a rectal tumor for preoperative prediction of lymph node (LN) metastasis. The radiomic features of the three modalities were extracted from three regions of interest: one at the tumor, one surrounding the largest pelvic LN, and one in the peri-tumoral fat (perirectal fat <1cm from the tumor). The results showed that the following elasticity features of SWE were the best risk predictors for LN metastasis: the maximum, minimum, and range of the rectal tumor, contrast/intensity variability of lymph nodes and, contrast/sum average of peri-tumoral fat. The risk predictors of the ERUS were: continuity, LN number, and maximum of LN. The CT risk predictors were: range/intensity variability of LN, contrast/variance of peritumoral fat, and sum average/contrast of the rectal tumor. The radiomics model based on the combination of SWE, ERUS, and CT features, had a concordance index of 0.872 for the training cohort, and 0.857 for the validation cohort.

3.3.2.2 Ultrasound elastography for pathologic response assessment

The use of shear wave elastography in pathologic response assessment has been studied comprehensively in the field of breast cancer. Jing et al. showed that a relative change in tumor stiffness after two cycles of neoadjuvant chemotherapy (NAC) correlated significantly with pathological responses of breast cancer specimens.[100] Response was defined when the loss of tumor cells was >90%. The mean stiffness at baseline, the mean stiffness after two cycles of NAC, and the relative change in mean stiffness were significantly higher for responders compared to non-responders. With a cutoff value of -36.1% for the relative change in mean stiffness, responders could be distinguished from non-responders with a sensitivity of 72.9% and a specificity of 85.7%. Lee et al. investigated the accuracy of SWE in detecting residual breast cancer after NAC.[94] Furthermore, they examined the addition of SWE to the conventional B-mode US in diagnostic performance. The B-mode and SWE images were retrieved 1 day before surgery, after completion of NAC. The maximum elasticity from SWE was used to distinguish between pathologic complete response and residual cancer. Pathologic complete response was defined as the absence of residual invasive cancer and ductal carcinoma in situ. They showed that the maximum elasticity was significantly higher for women with residual cancer compared to women with a pathologic complete response. A cutoff value for maximum elasticity of >30kPa had the highest diagnostic accuracy to indicate residual cancer, with a sensitivity, specificity, and accuracy of 83.6%, 80.0%, and 83.1%. The diagnostic performance of SWE (AUC = 0.880), and SWE combined with B-mode (0.877), were significantly higher than the B-mode US only (0.702).

The application of ultrasound elastography in response assessment has been investigated for colorectal cancer as well. Xiao et al. examined patients with ERUS in combination with strain elastography, and contrast-enhanced ultrasonography, to provide preoperative staging of rectal cancer after CRT.[80] The tumor staging was based on the local invasion of the tumor into the tissue layers based on the criteria by Hildebrandt et al, as shown in Table **3-1**. The strain elastography map was used to identify the boundary of the lesion by the color difference after compression. This boundary was used to determine the local tumor invasion T-stage. The T-stage according to the US B-mode was compared to the pathological T-stage postoperatively. The accuracy of using a combination of B-mode, strain elastography, and contrast-enhanced ultrasonography in staging rectal cancer after neoadjuvant chemo-radiotherapy was 84.9% (45/53). Most of the patients that were misdiagnosed were overstaged (5/8). The overstaging was mainly related to the peri-tumoral tissue reaction after neoadjuvant therapy.

Rafaelsen et al. performed a study towards the therapeutic response on CRT in patients with locally advanced (T3-4) rectal cancer with shear wave elastography.[34] This study is not shown in the literature overview, since the study does not report a performance metric of ultrasound elastography. The shear wave speed of the tumor and surrounding fat was measured with ultrasound elastography at the baseline, after two weeks from the start of CRT, and after six weeks. This shear wave speed was compared to the pathological stage after surgery. The results demonstrated that the tumors that were downgraded had almost an equal shear wave speed than the non-responders at baseline (3.14 m/s and 3.13 m/s). The shear wave speed was significantly lower for the responders than the non-responders after two weeks from the beginning of CRT (1.95 m/s vs 2.47 m/s). However, when comparing the shear wave speed at six weeks, there is a smaller difference between the responders and non-responders (2.05 vs 2.20 m/s). It is not shown whether this difference is significant. Furthermore, they showed that the perirectal fat of the non-responders had an increased shear wave speed from 1.56 to 2.41 m/s in six weeks.

3.3.2.3 Ultrasound elastography in surgical approaches

Intraoperative USE has not been studied yet in the field of colorectal cancer. However, one ex vivo study used strain elastography to discriminate Crohn's lesions from colorectal adenomas and adenocarcinomas. Furthermore, two studies used shear wave elastography intraoperatively to assess pancreatic lesions or to assess lymph node metastases.

Havre et al. performed an ex-vivo study with strain elastography on freshly excised surgical specimens of colorectal cancer.[86] The purpose of this study was to investigate whether strain elastography could discriminate between colorectal adenocarcinomas and stenotic inflammatory Crohn's lesions. The study showed that strain ratio measurements and visual evaluation of strain differences could not differentiate stenotic Crohn's disease from adenocarcinoma. However, another finding was that the SR was higher for 18 sections with adenocarcinomas compared to four sections from adenomas.

Recently, some studies were published on the intraoperative application of shear wave elastography.[101–103] Bae et al. performed a study for an intraoperative evaluation of axillary lymph node (LN) metastasis using shear wave elastography and nodal size in breast cancer.[101] The study developed a nomogram for the calculation of the probability of LN metastasis among excised LNs. The nomogram involved the nodal size, mean stiffness, and stiffness ratio. The stiffness ratio was the stiffness of the LN region in comparison to perirectal fat. The AUCs of the nomogram performance were 0.856 and 0.791 for training and validation set respectively in the discrimination between metastatic and non-metastatic lymph nodes. The study showed that the shear wave elastography has a significantly higher mean stiffness in metastatic lymph nodes compared to non-metastatic lymph nodes (23.54 vs. 10.41 kPa). Furthermore, they showed a significant difference in the elastic ratio for the lymph node versus the surrounding fat, the ratio was 3.24 for metastatic and 1.49 for non-metastatic nodes. Silva et al. used shear wave elastography and contrast-enhanced ultrasound to assess pancreatic lesions.[102] With a cutoff value of 28.7 kPa for the mean stiffness, the sensitivity, specificity, and accuracy were 0.744, 0.467, and 0.667 respectively.

The assessment with the contrast-enhanced US had a better performance with a sensitivity, specificity, and accuracy of 1.00, 0.40, and 0.833. However, the disadvantage is that the contrast-enhanced US is an invasive method, and it takes time to inject contrast fluid.

Table 3-3: Studies with strain elastography in the field of cancer

Study	Organ	End-point	Variables	Performance	Statistic metric
Waage (2015)[31]	Colorectal	Adenomas vs T1/T2 tumors	SR (cut-off value >1.25); SR+ERUS	0.82,0.86,0.84; 1.00, 0.88, 0.95	Sens, spec, acc
Waage (2015)[28]	Colorectal	Adenomas vs adenocarcinomas	SR (cut-off value >1.25)	0.96, 0.86, 0.94	Sens, spec, acc
Xiao (2018)[80]	Colorectal	Tumor stages T0-T4	Visual assessment of ERUS, SE, CRUS	0.849	Accuracy
Chang (2013) [89]	Breast	Benign vs malignant	5-point score; 5-point score +B-mode	0.943; 0.965	AUC
Fernandes (2019) [88]	Breast	Response vs no response[a]	SR	0.81	AUC
Mohey (2014) [87]	Breast	Benign vs malignant	5-point score (cut-off > 3); 5-point score + B-mode	0.697,0.951,0.817; 0.909,0.951,0.938	Sens, spec, acc
Zhang (2017) [85]	Pancreas	Benign vs malignant	5-point score; SR	0.91; 0.93	AUC
Hahn (2017) [90]	Soft-tissue tumors	Benign vs malignant	SR; 5-point score	0.700; 0.623	AUC

*SR = strain ratio; 5-point score is score as shown in Figure **3-10**; B-mode = brightness mode ultrasound; ERUS = endorectal ultrasound; SE = strain elastography; CRUS = contrast enhanced ultrasound; sens, spec, acc = sensitivity, specificity, accuracy; AUC = area under curve.*

[a]Response defined as the absence of residual invasive cancer confirmed with pathology;

3.3.1 Summary

In summary, several studies showed that ultrasound elastography is a valuable technique to distinguish malignant from benign tissue in several types of cancer, including colorectal cancer. Both strain and shear wave elastography have been proven to be of benefit to just B-mode US in distinguishing adenomas from adenocarcinomas. The application of ultrasound elastography to assess the pathological response of chemoradiotherapy has mainly been studied for breast cancer. The studies that focused on colorectal approaches are not convincing yet. Namely, Xiao used visual assessment of strain elastography to estimate the tumor grade instead of a quantitative measure.[80] Therefore, the study did not propose a reproducible method to estimate tumor grade objectively. Rafaelsen showed that tumors that were downgraded had a larger decrease in shear wave speed compared to tumors that did not.[34] However, the study only shows relative values compared to baseline and did not show a performance of shear wave elastography. Surgical approaches of USE to estimate colorectal tumor invasion have not been studied yet. Havre was the only study that used strain elastography in an ex vivo setting for colorectal adenocarcinomas.[86] Despite it showed that there was a difference in strain ratio between adenomas and adenocarcinomas, the number of cases was small. Furthermore, it was not mentioned whether these patients received neoadjuvant therapy. Another shortcoming of the current literature is that none of the studies focused on showing an automatic estimation to the surgeon about which tissue is malignant and which is not. Thereby, none of the studies focused on the use of ultrasound elastography in estimating the circumferential resection margin.

To compensate for the shortcomings mentioned above, the current study is needed to investigate whether ultrasound elastography can distinguish between tumor and surrounding tissues in colorectal

approaches for patients who received neoadjuvant therapy in an intraoperative setting. Furthermore, the current study will examine which features of ultrasound elastography are useful for this purpose. Besides, the study will show whether ultrasound elastography is of additional value to diffuse reflectance spectroscopy in estimating the circumferential tumor margin.

Table 3-4: Studies with shear wave elastography in the field of cancer and fibrosis

Study	**Organ**	**End-point**	**Variables**	**Performance**	**Statistic metric**
Chen (2017)[35]	Colorectal	Adenomas vs pT1-pT3; T0/pT1/pT2 vs pT3	Mean stiffness	0.981; 1.000	AUC
Chen (2020) [29]	Colorectal	Tumor deposits vs no tumor deposits	Radiomic features shear wave + B-mode	0.916	AUC
Fan (2019) [32]	Colorectal	pT1 vs pT2-4; pT1-2 vs pT3-4	Max stiffness	0.959; 0.9354	AUC
Chang(2013)[89]	Breast	Benign vs malignant lesions	Mean stiffness; mean stiffness +B-mode	0.928; 0.964	AUC
Jing (2016)[100]	Breast	Response vs no response [a]	Δ Mean stiffness [d]	0.802	AUC
Lee (2015) [94]	Breast	Response vs no response [b]	Maximum stiffness + B-mode	0.877	AUC
Cassinotto (2014)[95]	Liver	F0 vs ≥ F1; F0-1 vs ≥ F2; F0-2 vs ≥F3 [c]	Mean stiffness	0.89; 0.88; 0.93	AUC
Villani (2020) [96]	Liver	F0-1 vs ≥ F2; F0-2 vs ≥F3; F0-3 vs F4	Median stiffness	0.899; 0.900; 0.899	AUC
Silva (2020)[102]	Pancreas	Malignant vs benign focal pancreatic lesions	Stiffness (cutoff >28.7 kPa)	0.744,0.467,0.667	Sens, spec, acc
Chen (2018) [33]	LN colorectal	Lymph nodes metastases vs lymph nodes without metastases	Multiple (CT, ERUS, and SWE features)	0.857	C-index
Bae (2020) [101]	LN breast	Lymph nodes metastases vs lymph nodes without metastases	Mean stiffness + elasticity ratio	0.856; 0.791 [e]	AUC

LN = Lymph nodes;
ERUS = endorectal ultrasound; CT = computed tomography; SWE = shear wave elastography;
AUC = area under curve; sens, spec, acc = sensitivity, specificity, accuracy; C-index = concordance index

[a] *Response on chemotherapy was defined when the loss of tumor cells was >90%*

[b] *Response was defined as the absence of residual invasive cancer and ductal carcinoma in situ.*

[c] *Liver fibrosis stages: F0 = no fibrosis, F1=mild fibrosis, F2 =significant fibrosis, F3=severe fibrosis;F4=liver cirrhosis*

[d] *AUC of training+ validation set and AUC of the validation set*

[e] *Relative change in mean stiffness after two weeks of chemotherapy compared to baseline, and the cutoff value is shown for a response on chemotherapy*

3.4 Introduction to machine learning

In the current study, machine learning (ML) techniques are used to train a classification algorithm that can classify tissue as tumor or non-tumor based on B-mode and elastography features. Furthermore, ML was used to develop a regression algorithm to estimate the CRM based on B-mode, elastography, and DRS features. This section will explain some background information on ML.

ML is an interdisciplinary field that aims to construct algorithms that can learn from and make predictions on data.[104] ML algorithms in medical imaging classification are especially useful for classification problems that are not obvious and vulnerable to human errors. Machine learning algorithms have been used in analyzing SWE as well. For example, using quantified SWE features to classify malignant and benign breast lesions, predict malignancy in thyroid nodules, and determine liver fibrosis stages.[105–110] The ML algorithms in these studies are based on statistical features from SWE images and the application of a classifier. These studies show that multiple statistical features were useful in the classification between malignant tumors and healthy tissues. ML techniques can be used to aim for the best predictive possibility by using multiple features and multiple classifications or regression algorithms.

Machine learning algorithms can be divided into categories of supervised learning and unsupervised learning. Supervised learning means that a classifier or regression model is trained with samples that are labeled with the golden standard. Once a classifier is trained it can be used to classify previous unseen test samples. Unsupervised learning involves training based on clusters or similarities in data with no labels provided. In the current study, supervised learning was used to train classifiers and regression models. The training set consisted of DRS features as well as textural and statistical features from a certain region of interest from US B-mode images and elastography images that are labeled with a tissue type.

Input features (variables) are fed to the algorithm which will be used to learn which combination of variables belong to a certain class in the case of classification. In the current study, textural and statistical features of US B-mode and elastography images will be used for training algorithms. Textural features allow a quantitative and objective assessment of tissue heterogeneity by evaluating the distribution and relationship of pixel or superpixel grey levels in the image. Marcon et al. reported that textural analysis in combination with machine learning is a useful diagnostic tool in distinguishing malignant from benign lesions and lesions from normal tissue.[111] Statistical features are generally used in machine learning algorithms with SWE. These features involve the mean, variance, interquartile ranges, and distribution of stiffness values of regions of interest.[110]

It is important to constrain the classifier in such a way that it does not overfit the training data because overfitting results in model errors that do not generalize beyond the training set to new data. Due to overfitting, the model performs perfectly on a training set, while fitting poorly on a test set.[112] The poor performance on the test set can be induced through differences in features in the test set for a certain tissue type than the training set. Generally, overfitting can be caused by noisy training data. Noisy training data consist of feature sets retrieved from artifacts or wrong ground truth labels. A too large amount of noisy training samples compared to reliable training samples can result in a trained algorithm that is adjusted to classify based on noisy data instead of true signals. Overfitting can be reduced by selecting training samples that have a truly correct tissue label and which are not from regions of interest with image artifacts. Furthermore, the number of features must be adapted to the number of samples to avoid the curse of dimensionality. This refers to the phenomena when the number of features is very large relative to the number of samples causing problems in training algorithms. The optimal number of features depends strongly on the classifier and the feature-label distribution.[113] Generally, the optimal number of features can be determined as the square root from the number of samples.

The input data for a machine learning algorithm is usually split into a train-, validation, and test set. The training set is used to learn the machine learning algorithm in which feature values belong to a certain outcome. Thereafter, the validation set is used to optimize the parameters of machine learning algorithms. Through the validation, a classifier that is trained on the training set can be directly evaluated with 'new' data and it can be tested on overfitting. For smaller datasets, k-fold cross-validation is often used instead of creating a separate validation set. In the case of cross-validation, the validation set is part of the training set. Cross-validation can be performed up to a maximum of N times for a database of N-samples. For example, in the case of 10 training samples, 10-fold cross-validation with leave-one-out testing means that 9/10 samples are used for training and 1/10 is used for validation. The test set is held aside during the training and optimization of the algorithm, to evaluate the performance of the algorithm on unseen data. A dataset is generally divided into 80% training and 20% test set.

3.4.1 Classification

Machine learning algorithms can be used for classification and regression tasks. Classification algorithms aim to predict the class/label based on the input data. Classification is used in tissue classification, to distinguish tumor tissue from other tissue types. The performance of a classifier can be evaluated based on the confusion matrix. The confusion matrix is shown in Table **3-5**. In the current study, tumor tissue belongs to the positive class and other tissue types are the negative class.

Table 3-5: Confusion matrix result of classification models

	Predicted Positive	*Predicted Negative*
Actual positive	TP	FN
Actual negative	FP	TN

TP: Number of true-positive cases that are correctly identified as positive; FP: Number of false-positive cases that are misclassified as negative cases; FN: number of false-negative cases that are misclassified as positive cases; TN: number of true-negative cases that are correctly classified as negative.

The confusion matrix can be used to calculate several evaluation metrics that assess the effectiveness of the algorithm. The commonly used evaluation metrics and the formulas are shown in Table **3-6**. The accuracy assesses the overall effectiveness of the algorithm by estimating the probability of the true value of the negative and positive classes. The error rate shows the estimation of misclassification probability. The sensitivity shows the probability of the positive class being classified as positive. The specificity approximates the probability that a negative class is classified as negative. Precision shows the probability that a predicted positive class is truly positive.

Table 3-6: Evaluation metrics with the corresponding formulas based on the confusion matrix [114]

Measure	**Formula**
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Error rate	$\frac{FP + FN}{TP + TN + FP + FN}$
Sensitivity/Recall	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
Precision	$\frac{TP}{TP + FP}$

The distribution in the confusion matrix depends on the cutoff point. Since the output of a classifier is a score between 0 and 1, the cutoff value determines whether the prediction for a sample belongs to the positive class (class 1) or the negative class (class 0). The optimal cutoff point can be determined based on a ROC curve. This curve plots the sensitivity against the specificity for different thresholds between

0 and 1. The area under the ROC curve is a common metric to show how well a classifier performs. The AUC ranges from 0 to 1, where an AUC of 1 means that the classifier works perfectly and classifies each positive sample as positive and each negative sample as negative, and an AUC of 0.5 means that the classifier is performing completely random, as it classifies 50% of the data as negative and 50% of the data as positive independently from the input class. Although the AUC is a commonly used metric, it is sensitive to imbalanced data. When a dataset is unbalanced, the classifier often classifies instances into the majority class, which leads to poor performance in classifying the minority class. Rebalancing class sizes can improve the AUC, where the largest improvement in AUC can be achieved when two classes are fully rebalanced to be of equal size.[115] The Matthews Correlation Coefficient (MCC) is a reliable statistical performance measure in the case of imbalanced data.[116] The MCC is less influenced by imbalanced test sets since it considers mutual accuracies and error rates on both classes and involves all values of the confusion matrix.[114] The MCC ranges from 1 for a perfect prediction to -1 for the worst possible prediction and 0 for random prediction.[116] The MCC produces a high score only if the prediction obtained good results in all of the four confusion matrix categories (true positives, false negatives, true negatives, and false negatives), proportionally both to the size of positive elements and the size of negative elements in the dataset. The equation to calculate the MCC is shown in Equation **3-3**. The MCC can be used to determine the cutoff value of the predicted scores as well, by determining the cutoff value which results in the highest MCC.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP) \cdot (TP+FN) \cdot (TN+FP) \cdot (TN+FN)}} \qquad (3\text{-}3)$$

Additionally, the McNemar test can be used to compare the predictions of two classifiers.[117] The McNemar test is based on a 2x2 contingency table, as shown in Table **3-7**. The McNemar test can be performed by the calculation of the chi-square test as shown in Equation **3-4**. The p-value of this chi-square test can be used to show the significance of the difference in the predictions from the two models. The p-value is based on one degree of freedom, and when the p-value is below 0.05, the performances of the two models are assumed to be different.

Table 3-7: Contingency table

	Model 2, correct predictions	*Model 2, wrong predictions*
Model 1, correct predictions	A	B
Model 1, wrong predictions	C	D

$$\chi^2 = \frac{(B-C)^2}{B+C} \qquad (3\text{-}4)$$

3.4.1 Regression

Regression refers to the method of studying the relationship between independent variables and dependent variables.[118] The analysis involves estimating continuous values as opposed to discrete classes of data. A regression analysis was used to estimate the circumferential margin based on US B-mode, elastography, and DRS features.

Regression models can be compared using the root mean square error (RMSE) and mean absolute error (MAE). The MAE gives the same weight to all errors, while the RMSE personalizes variances as it gives errors with larger absolute values more weight than errors with smaller absolute values. The RMSE is more appropriate to represent model performance than the MAE when the error distribution is expected to be Gaussian.[119] Whereas the MAE is directly interpretable as the error of the regression model in the prediction of the CRM. The MAE and RMSE are calculated with Equations **3-5** and **3-6**. Where e_i is the error between the expected and predicted outcome and n is the number of samples.

$$MAE = \frac{1}{n}\sum_{i-1}^{n}|e_i| \qquad (3\text{-}5)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n} e_i^2} \qquad (3\text{-}6)$$

The described technologies, DRS, US B-mode, and US elastography are applied in ex vivo studies that are shown in Chapter **4** and **5**. It will be shown how these three techniques are combined in ex vivo measurements of colorectal specimens. Furthermore, the ex vivo study in Chapter **4** uses machine learning to classify tumor versus healthy colorectal tissue, fat, and fibrosis using US B-mode and US elastography. Thereafter, the ex vivo study in Chapter **5** uses machine learning in regression algorithms to estimate the resection margin in colorectal specimens based on DRS, US B-mode, and US elastography.

4 Tissue type classification

4.1 Introduction

Rectal cancer surgeries are especially challenging in patients who received neoadjuvant therapy since the discrimination between the malignant tumor and fibrosis is difficult. Whereas DRS is helpful to discriminate colorectal tumor tissue from healthy tissue, it is not able yet to discriminate tumor from fibrosis.[10,21,22] This discrimination is especially difficult when a thin layer of fibrosis or healthy tissue is on top of the tumor.[21] Ultrasound elastography may be useful to improve the discrimination of colorectal tumor tissue and fibrosis, since it adds information of the anatomy with the B-mode image and information of the stiffness with the elastography image. US elastography is useful in pathologic response assessment in breast cancer patients after neoadjuvant chemotherapy.[88,94,100] Furthermore, US elastography has been proven to be helpful in colorectal tumor grade estimation, distinguishing adenomas from adenocarcinomas, predicting tumor deposits in the perirectal fat, or lymph node metastases.[31–35,80,120] However, current studies with US elastography and colorectal cancer are not convincing yet for an intraoperative assessment, and for patients who received neoadjuvant therapy. Furthermore, most studies used only one elastography feature to discriminate malignancies from healthy tissue. Thereby, the US B-mode was mainly used with a visual assessment to recognize malignancies, which requires staff members that are specialized in US. Therefore, the current study was performed to fulfill these shortcomings and to investigate whether a combination of US B-mode and elastography features is useful to classify automatically between colorectal tumor tissue, the healthy colorectal wall, fat, and fibrosis.

An ex vivo setup was used with freshly excised specimens to investigate the properties of US B-mode and elastography. The data acquisition of DRS data is also described in this chapter since the B-mode and elastography data was retrieved simultaneously with DRS data. Further data analysis with DRS is demonstrated in Chapter **5**. This chapter describes the feature extraction of the B-mode and elastography data. Subsequently, these B-mode and elastography features were correlated to a tissue type (tumor tissue, healthy colorectal wall tissue, fat, or fibrosis) based on the histopathological assessment of specimens. Thereafter, feature selection was performed to select a set of features that resulted in the best classification between tumor and healthy tissue or fibrosis. Finally, the tissue type labels and selected B-mode and elastography features were used to train, validate, and test classification algorithms to automatically discriminate colorectal tumor tissue from the healthy colorectal wall, fat, and fibrosis.

4.2 Data acquisition

4.2.1 Study population

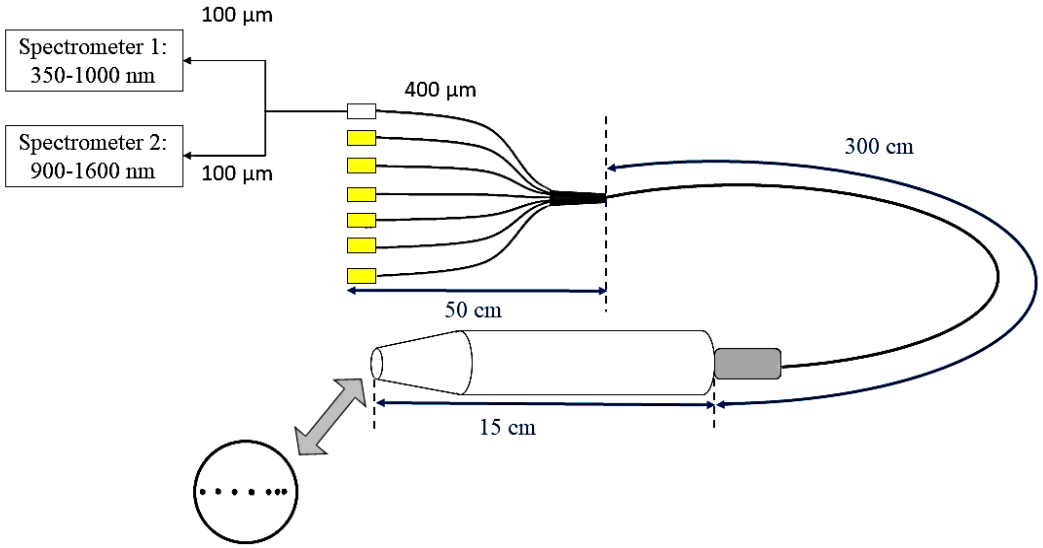
The study population consisted of patients with locally advanced colorectal cancer, who underwent colorectal surgery in the NKI-AvL. Despite the clinical relevance of this study would be for patients with rectal cancer, patients with colon cancer were included as well to increase the number of patients. Furthermore, the rectum and colon have the same bowel layer structure. The inclusion periods were November 2019 and May-October 2020. Patients were included when the tumor had a T3 or T4 grade preoperatively and when the tumor was located in the colon or rectum. Patients with T2 tumors were also included in case the tumor was located within 1 cm from the mesorectal fascia, which makes it possible to perform superficial measurements of the tumor. Patients were excluded from the study when the tumor responded to neoadjuvant therapy. The response was described as downgrading to T0 or T1 according to the radiological report. The reason for this is that it would not be possible to perform superficial measurements of the tumors due to the envelopment of the mesorectal fascia.

4.2.2 Measurement protocol and data acquisition

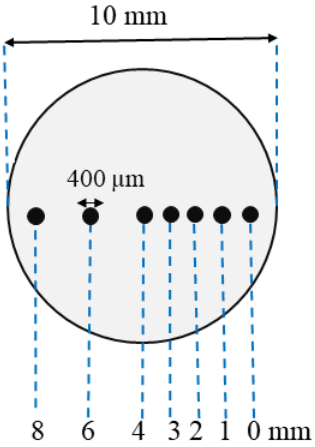
The ex vivo measurements were performed with three modalities: DRS, superficial US, and US elastography. The US elastography measurements were integrated into a protocol that was used for another study in the NKI-AvL. This study included the same specimens and has the goal to discriminate tumor from healthy tissue and fibrosis using superficial US B-mode and DRS simultaneously. When the current study shows that US elastography is a useful technique, the US elastography data can be directly compared to the superficial US B-mode and DRS. The superficial US B-mode and US elastography measurements require both a different US transducer. The superficial US B-mode measurements were performed with the Philips L15-7 broadband compact linear array transducer since it is suitable for intraoperative measurements through the small format. Besides, this transducer has high pulse frequencies, between 15 and 7 MHz, and is therefore suitable for superficial measurements. The L15-7 transducer is shown in Figure **4-1**A. However, with this US transducer, it is not possible to perform elastography measurements. Therefore, the Philips EPIQ 7 Ultrasound system eL18-4 transducer was used, as shown in Figure **4-1**B . This transducer has pulse frequencies between 22 and 2 MHz. The DRS system developed in the NKI-AvL is used for the DRS measurements. Figure **4-2** shows an illustration of the DRS setup, the contact surface of the probe, and the probe itself. The DRS console consists of two spectrometers; one for the 350-1100 nm range (Avantes, AVASPEC-HS2048XL-EVO) and one for the 900-1600 nm range (Avantes, AVASPEC-NIR256-1.7-RS). A cutoff point of 1000 nm was chosen to combine the spectra from both spectrometers. A 400-µm core fiber is illuminated with a halogen lamp (Avantes, AVALIGHT-HAL-S-MINI) and serves as an illuminating fiber. The linear DRS probe contains six emitting fibers and one receiving fiber, with different fiber distances between 1-8 mm.

A. B.

Figure 4-1: Hardware used for the ex vivo measurements. A. The L15-7io Philips Ultrasound transducer for superficial ultrasound measurements. B. The eL18-4 Philips Ultrasound transducer for ultrasound elastography measurements.



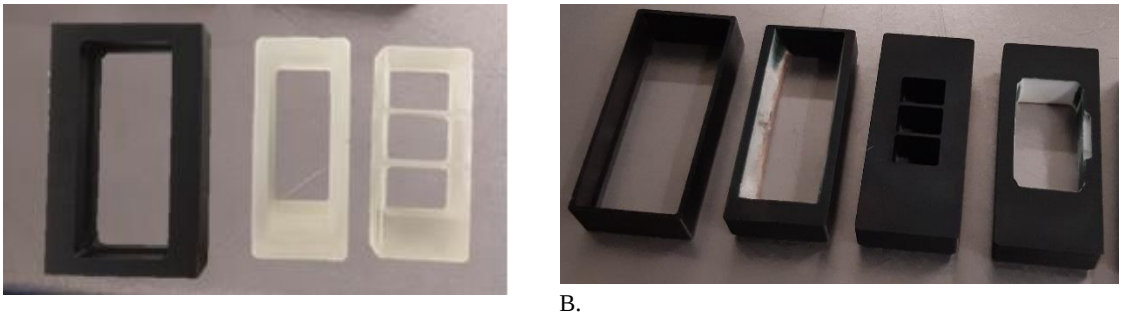
A.



B. C.

Figure 4-2: Illustration of DRS setup. A. The DRS system consists of six halogen lamps that are connected with the probe via six 400 µm emitting fibers. The 400 µm receiving fiber is split into two fibers of 100 µm that are connected to two different spectrometers. B. Contact surface of the DRS probe that shows the six emitting fibers and one receiving fiber. Each emitting fiber has a different distance to the receiving fiber, varying from 1 to 8 mm. C. Linear DRS probe developed in the NKI-AvL that was used for the ex-vivo measurements.

To combine the three modalities, molds were used in such a way that the central DRS measurement was at the same location as the center of the L15-7 transducer, and the eL18-4 transducer. However, for the first 13 specimens, there were only molds available for the L15-7 transducer and the DRS probe. For the last 17 specimens, new molds were introduced which fitted both US transducers and the DRS probe. The old and new molds are shown in Figure **4-3**. The outer molds were designed in such a way that all inner molds would fit and the measurement location does not change when changing from modality. Besides, the inner molds were designed that they fit in only one direction in the outer mold. Since the protocol slightly differed between the old molds and the new molds, the protocol with the old molds is called 'Protocol I', and the protocol with the new molds is called 'Protocol II'. The central location of the measurements with the eL1805 transducer in Protocol I might be deviating from the central DRS point through the missing mold. However, both B-mode images from the two US transducers could be compared to see whether the eL18-4 transducer was shifted relative to the L15-7 superficial transducer. Furthermore, there are no differences expected between the measurements using these two protocols since the US transducers, the DRS probe, and the measurement locations were the same.



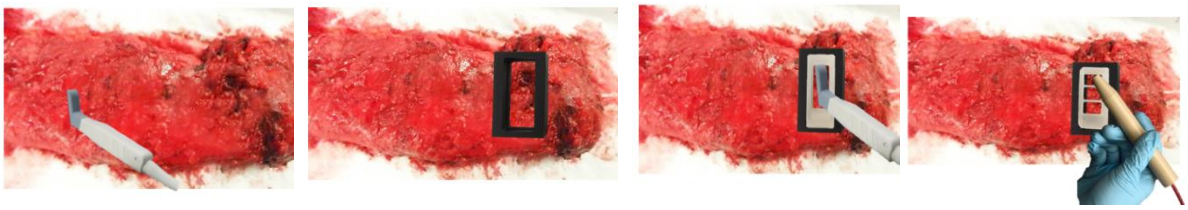
A. B.

Figure 4-3: A. Molds that were used for DRS and ultrasound measurements in Protocol I, for the first 13 measurements. From left to right: outer mold, inner mold for the L15-7 transducer, and the inner mold for the DRS probe. The DRS probe can be placed in the three holes. B. Molds used for measurement in Protocol II, for the last 17 measurements. From left to right: outer mold, inner mold for the eL18-4 ultrasound elastography transducer, inner mold for DRS probe, and the inner mold for L15-7 ultrasound transducer.

Before starting the measurements, the DRS system was calibrated using a white Spectralon® sample. The DRS probe was put into a custom-made probe holder to hold the probe tightly and keep the surface of the optical fibers at the distal end parallel to the reflectance standard at a fixed distance. The white sample reflects the light uniformly over the probe surface. The resulting calibrated spectrum was used as the system response to compensate for the spectral shape of light emitted by the lamp and the wavelength-dependent sensitivity of the detector as well as any wavelength-dependent sensitivity in the optics and gratings of the system.[66] Subsequently, a background measurement was performed to measure the ambient light, dark current, and electric offsets of the spectrometers. The calibration was performed as soon as the spectrometer temperature was stable and the light output was stable. Thereafter, Protocol I or II was performed, as shown in Figure **4-4** and Figure **4-5** respectively.

The measurement setup was adjusted based on the results of a phantom study that was performed to prepare the ultrasound elastography measurements. This study is described in Appendix **A**. The main adjustment was that surgical pads were placed beneath the specimen to minimize the deterioration of the metal table on the shear wave elastography measurements. Since the measurement depth of the eL18-4 transducer was minimal 4 cm and the specimens were mostly ~1-2 cm thick, surgical pads were placed beneath the specimen such that the distance from the transducer to the table was >4 cm.

The first four steps were the same for both protocols. The first step was to find the tumor location in the freshly excised specimen. The localization starts with determining the distal and proximal sides of the colon. This is important for recognizing the measurement locations during pathologic assessment, which will be explained in Section **4.2.3**. The distal side can be determined based on the black tattoo, which is placed distally from the tumor during colonoscopy. Furthermore, the proximal and distal sides can be determined based on anatomical interpretation. Subsequently, the specimen was inspected by visual assessment and palpation. The location of the malignant tumor could be roughly estimated based on preoperative imaging, the muscle-retracting sign (muscle layer being pulled toward a neoplastic tumor), and palpation.[121] Thereafter, the specimen was inspected by the L15-7 US transducer to find the most superficial location of the tumor. The superficiality is important for the measurement location since the DRS measurements are mainly useful for assessment up to ~8mm in depth since the maximum fiber distance is 8 mm.[21] When the measurement location was determined, the second step was performed by placing the outer mold at this location. Subsequently, the third step was to place the inner mold of the L15-7 transducer into the outer mold and to retrieve a superficial US B-mode image. The direction of the transducer should be perpendicular to the direction of the colon. This position is important for the combination with histopathology coupes. Thereafter, the fourth step was to replace the inner mold with the inner mold for the DRS probe and to retrieve three DRS measurements from top to bottom.



1. Tumor localization with the superficial L15-7 transducer.
2. Placing outer mold.
3. Placing inner mold for superficial transducer and perform US measurements.
4. Replacing inner mold for and perform three DRS measurements.

5. Place ink marks, with two black marks on top, and one purple ink mark at the bottom.
6. Place eL18-4 transducer in the same direction as the L15-7 transducer and with center placed on central ink mark (yellow circle) and perform US elastography measurement.
7. Repeat for maximal three rows and surround ink marks with yellow ink.

Figure 4-4: Step-by-step description of Protocol I for first 13 ex vivo measurements at colorectal specimens with DRS, superficial US, and US elastography.

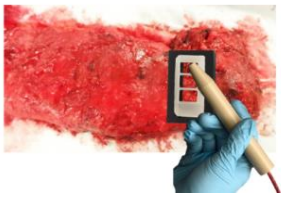
The fifth step in Protocol I was to mark the three DRS measurement locations with two black spots on top and one purple spot at the bottom. Since the US transducer was oriented along these lines, the purple marker location corresponds to the right part of the B-mode image and the black markers to the left and central part of the image. These markers are essential in correlation with histopathology coupes. Thereafter, the sixth step was to place the eL18-4 transducer with its center at the central measurement point. The transducer was placed in such a direction that the US images from both transducers had the same left-right orientation. Subsequently, a B-mode US image and the corresponding shear wave elastography image were retrieved. Shear wave elastography was chosen instead of strain elastography because quantitative stiffness values could be retrieved per pixel. The seventh step was to repeat Protocol I three times, with measurement locations next to each other. This resulted in three superficial US B-mode images from the L15-7 transducer, nine DRS measurements, three US B-mode images from the eL18-4 transducer, and three SWE images. After the ex vivo measurements, the surgical specimen was brought to the pathology department for pathological examination.

The fifth step in Protocol II was to replace the inner mold for the DRS measurements with the inner mold for the eL18-4 transducer. Therefore, this transducer could be placed exactly at the central DRS measurement as well. Subsequently, an US B-mode and SWE images were retrieved just as in Protocol I. Thereafter, the sixth step was to replace the inner mold of the eL18-4 transducer with the DRS inner mold, and the seventh step was to apply the three markers with two black ink marks on top and one purple ink mark at the bottom. Protocol II was repeated three times just as Protocol I.

1. Tumor localization with the superficial L15-7 transducer

2. Placing outer mold

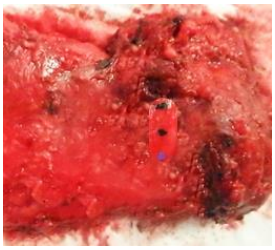
3. Placing inner mold for superficial transducer and perform US measurements



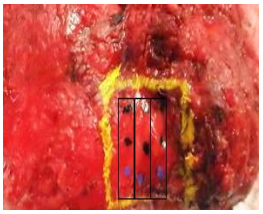
4. Replacing inner mold and perform three DRS measurements

5. Replace inner mold with mold for the eL18-4 transducer and perform USE measurements

6. Replace the inner mold with the inner mold of the DRS probe, to apply three ink marks on a row.



7. Place ink marks, with two black marks on top, and one purple ink mark at the bottom.



8. Repeat for maximal three rows and surround ink marks with yellow ink

Figure 4-5: Step-by-step description of Protocol II for the last 17 ex vivo measurements at colorectal specimens with DRS, superficial US, and US elastography.

For the US elastography measurements, several adjustments were made to the default settings in the EPIQ7 module. The transparency of the shear wave elastography color map was set to zero. Using this setting, the colors that represent the stiffness based on the color bar are clearly shown. This is important for the data extraction and will be further explained in Section **4.3.1**. Furthermore, the level of the confidence map was adjusted. The confidence map gives insight into the quality of SWE measurements. The color map was superimposed on the US B-mode image next to the SWE map with the colors red, yellow, and green. The red color was shown when the system was not able to detect shear waves, yellow when the system detected a few shear waves, and green when the system detected enough shear waves. The confidence map was set at 35% for the first 15 ex vivo measurements. With the level of 35%, just yellow and green colored regions were shown in the elastography image. The regions where the transducer could not detect shear waves were not shown in the elastography map. The disadvantage of this method was that elastography data is thrown away. Since the confidence map was optimized for measurements in the human body, the ex vivo setup could influence the confidence map mistakenly. Therefore, the confidence level was turned down in consultation with Philips for the last 15 ex vivo measurements. Since then, the shear wave maps were retrieved with a confidence level of 0%. To investigate the influence of the confidence level on the ex vivo elastography measurements afterward, the confidence maps were collected as well.

In summary, each ex vivo measurement resulted in nine DRS measurements, three US B-mode images from the L15-7 transducer, three US B-mode images from the eL18-4 transducer, three shear wave elastography images, and three confidence maps. An example of each modality is shown in Figure **4-6**.

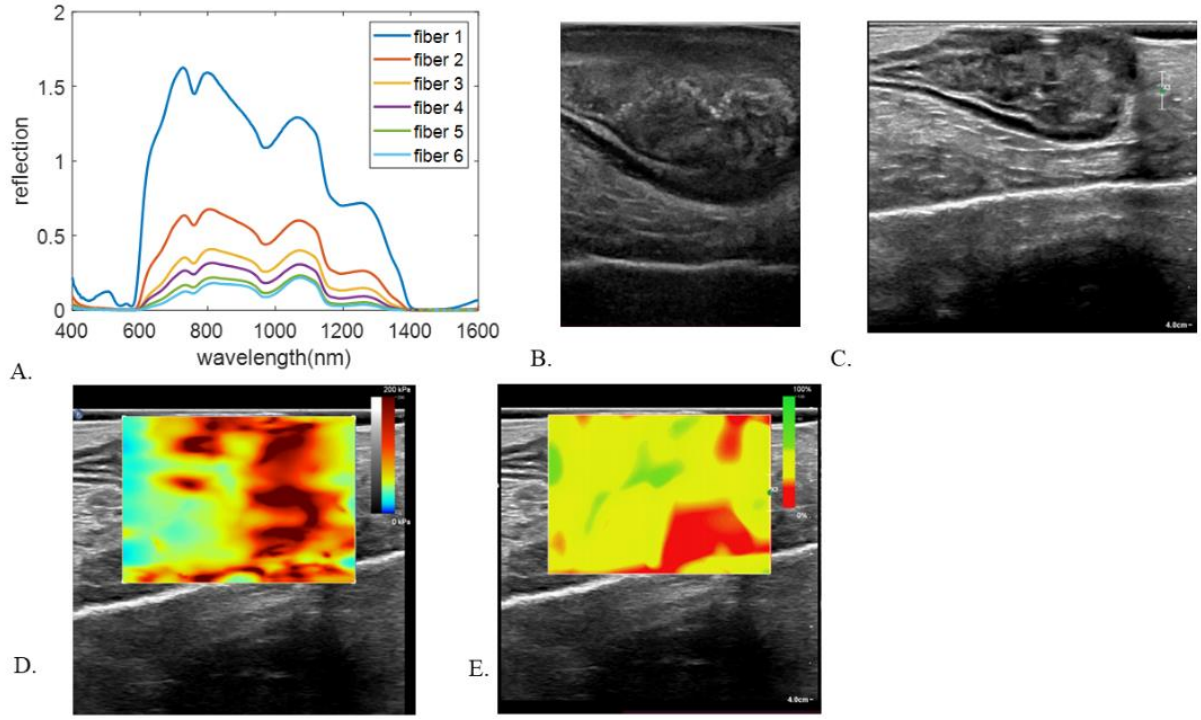


Figure 4-6: Resulting images from one ex vivo measurement of a T2 sigmoid tumor. A. DRS spectra with one spectrum per fiber distance. B. US B-mode image from the L15-7 superficial US transducer. C. US B-mode image from the eL18-4 US transducer. D. Shear wave elastography image. The colors represent stiffness values in the range of 0-200 kPa. E. Confidence map of shear wave elastography quality. Green represents areas where the transducer was able to detect all shear waves, yellow represents areas where the system could detect a few shear waves, and red represents areas where the system could not detect shear waves.

4.2.3 Correlation with histopathology results

Two days after surgery, the specimen was macroscopically inspected by the pathologist. The specimen was cut into slices of ~3 mm, in a perpendicular direction to the colon, as shown in Figure **4-7**A. Each slice was divided into subparts to be able to fit in numbered cassettes of 2 cm in length and 1 cm in width. During this division, the pathologist tried to place the three marker points together into one cassette, as shown in Figure **4-7**B. When this was not possible, the sub-slice was divided into two parts. In that case, one of the parts should contain the top black marker, and the other part should contain the purple and middle black marker, to be able to recognize the top and bottom of a measurement row in the final histopathological coupe. Thereafter, the cassettes were put into formalin to fixate the tissue to preserve proteins and other structures. The formalin fixation process takes ~36h for colorectal tissue cassettes.[122] After formalin fixation, the cassettes were embedded in paraffin to create a solid block of each cassette, which can easily be sliced into thin coupes. These coupes can be used for H&E staining and microscopic evaluation. Finally, an experienced pathologist delineates the tumor cells with a red line and the fibrotic cells with a purple line on the digital pathology slides. An example of a histological coupe with outlines is shown in Figure **4-7**C.

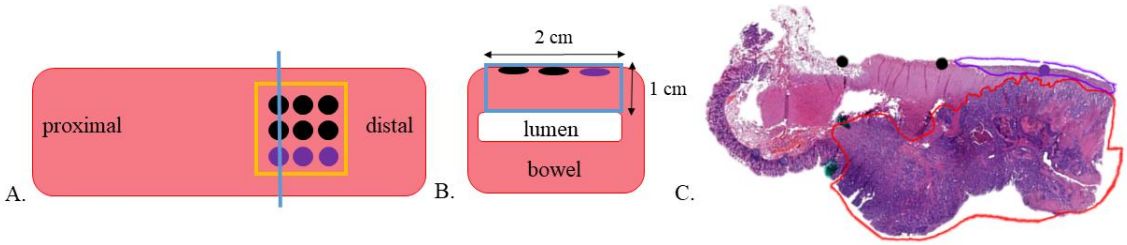


Figure 4-7: Pathology process for colorectal surgical specimen A. The specimen is localized with the distal side to the right and the proximal side to the left. The pathologist cuts a 3 mm slice from the bowel along the blue line, through the center of the markers. The result of this cut is shown in B. From this slice, a subpart is cut of max. 2 x 1 cm. Afterward, this block is put into formalin and embedded paraffin. Thereafter, the block is cut into thin coupes and is stained with H&E, as shown in C. The tumor region(s) are delineated with red, and fibrotic region(s) with purple.

The next step was to find the marker locations back in the histological coupes in Aperio ImageScope v. 12.4, Leica Biosystems. The center of an inked marker is used as a reference and an arrow was placed pointing to this reference, as shown in Figure **4-8**A. Based on the ink marker color, the purple marker could be distinguished from the black markers, as shown in Figure **4-8**B and C. Since this color difference is not visible without a high zoom factor, black and purple dots are placed at these locations, as shown in Figure **4-8**D. Since the distance between the markers should be equal, due to marker placement within a mold, the distance between the markers was used to check whether the locations of the markers are correct. When it was not possible to find the third marker, the location of this marker was based on this distance between the other two. An example is shown in Figure **4-8**D, where the distance between two markers is 7.5 mm.

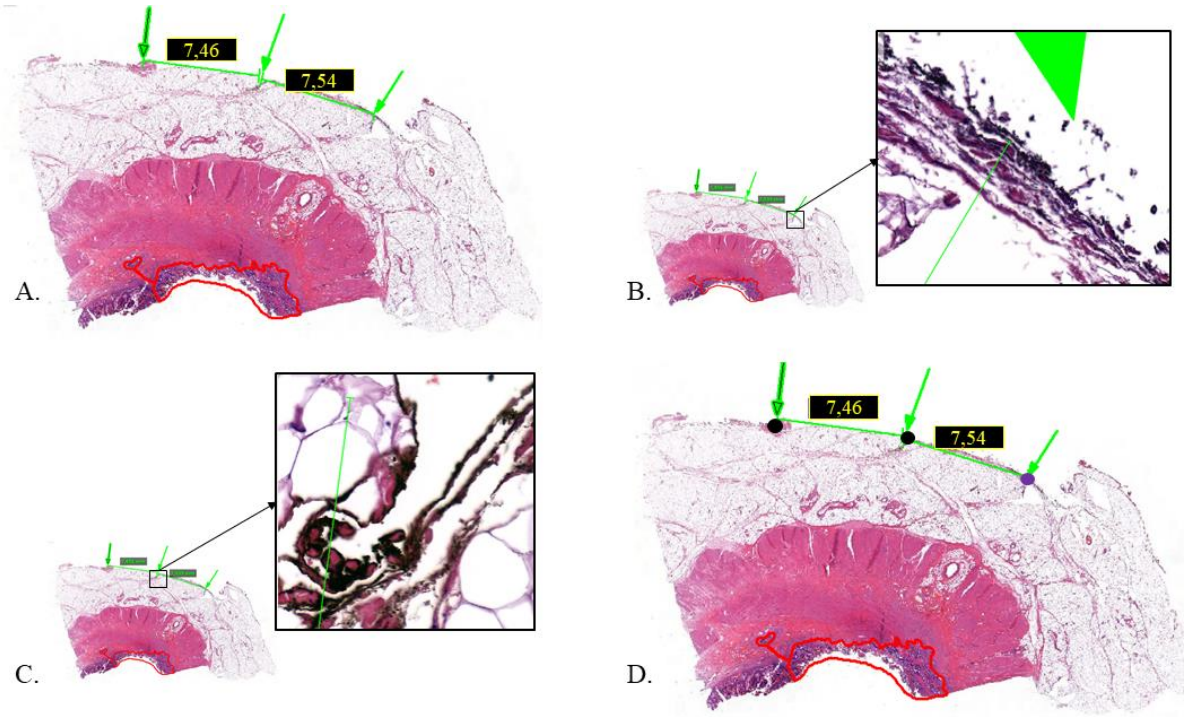


Figure 4-8: Overview of pathology coupes and markers. A. Pathology coupe of a T2 rectum tumor. B. Zoomed in at the purple ink mark. C. Zoomed in at the central black ink mark. D. Pathology coupe including black and purple dots to recognize the marker locations in the overview image.

4.3 Data analysis

After all ex vivo measurements and histopathological assessment of the specimens, the data analysis was performed. The data analysis was performed on the US B-mode images and SWE images from the eL18-4 transducer which is shown in Figure **4-6**C and D. The data analysis of the DRS spectra is described in Chapter **5**. The US B-mode image from the L15-7 transducer was used to check whether the eL18-4 transducer was shifted concerning the center of the other US transducer. This was needed for data from Protocol I because there was no mold available for the eL18-4 transducer.

All data analysis was executed with MATLAB R2020a (MathWorks Inc., Nathick, Massachusetts, US). The data analysis consists of several steps, which are shown in Figure **4-9**. Firstly, the shear wave elastography data was quantified with the conversion from red, green, and blue (RGB) color maps to stiffness values. Thereafter, superpixel segmentation was used to segment multiple regions from the US B-mode of the eL18-4 transducer and the corresponding elastography image. Thirdly, B-mode and elastography features were extracted from these regions. Fourthly, the regions were labeled with a tissue type using histopathology coupes. Subsequently, feature subset selection was performed. Thereafter, the selected feature set and tissue type labels were used to train, validate, and test several classification

algorithms. Lastly, a performance assessment was done to see how well the classifier could distinguish colorectal tumor from healthy tissue and fibrosis.

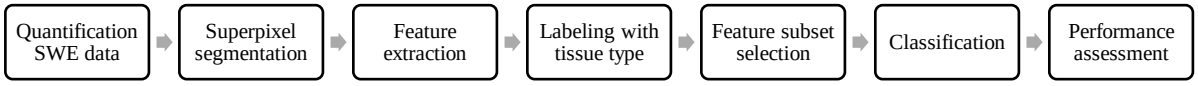


Figure 4-9: Overview of the data analysis part of the tissue type classification. Each subpart will be explained in one of the following subsections. SWE = shear wave elastography

4.3.1 Extracting elastography maps

The data from the ultrasound elastography measurements with the eL18-4 transducer consisted of DICOM files, with an US B-mode image, and an SWE image. Each pixel in the US B-mode image had a value between 0 and 1 that represents the grey intensity. Each pixel in the SWE image contained three values between 0 and 1, that represent the red-, green- and blue color channels (RGB value). The combination of these three values determines the color. The color bar next to the elastography image shows which RGB color corresponds to which stiffness, in a range between 0 and 200 kPa. It was not possible to export the stiffness values from the elastography image directly, so the stiffness values were reconstructed based on the RGB-maps and the color bar. The reconstruction from RGB map to stiffness map, is shown in Figure **4-10**. Firstly, the color bar was extracted based on its x- and y-coordinates, resulting in a matrix with RGB-values per row in the color bar. Thereafter, an array was created with values from 0 to 200 kPa, equally divided over the number of rows in the color bar. Therefore, a stiffness value could be assigned to each row in the color bar. The stiffness values were combined with the RGB-map in such a way that the lowest blue-colored pixel corresponded to 0 kPa, and the highest red-colored pixel to 200 kPa.

The colored region was automatically extracted from the original elastography image. The extraction was based on the characteristic outline, that had the same color in each elastography image. Each pixel in the RGB image was correlated to a row in the color bar. The corresponding rows in the color bar were determined by calculating the Euclidean distance. The Euclidean distance $d(p, q)$ was calculated with Equation **4-1**. The p and q are two pixels in the colored map and color bar. The $R, G,$ and B correspond to the red, green, and blue values respectively. The calculation of the Euclidean distance was repeated for each row in the color bar per pixel in the colored part. The row that resulted in the shortest distance with the selected pixel was assigned to that pixel. Each pixel could be assigned to a stiffness value by repeating the distance calculation.

$$d(p,q) = \sqrt{(q_R - p_R)^2 + (q_G - p_G)^2 + (q_B - p_B)^2} \quad (4\text{-}1)$$

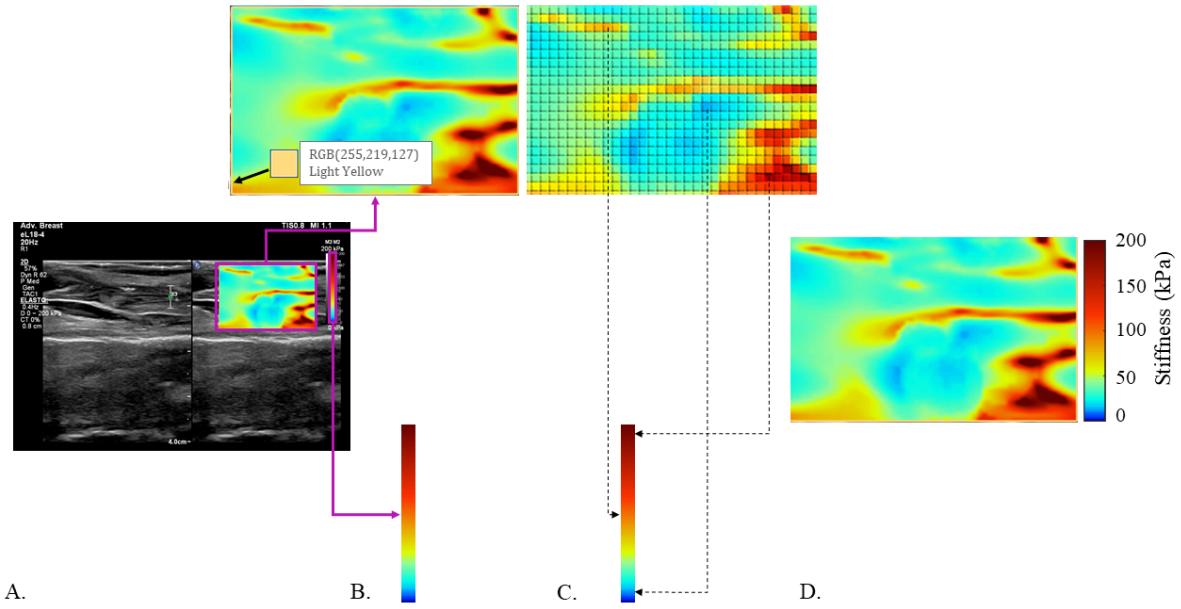
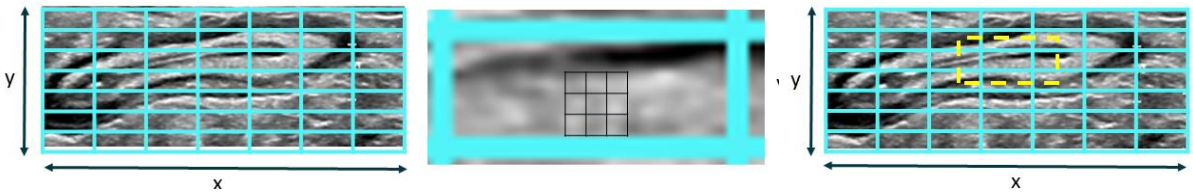


Figure 4-10: The method to turn colored pixels into a stiffness value (kPa). A→B:The color bar was extracted from the original image based on the x- and y-coordinates. Furthermore, the colored part in the image was segmented based on the characteristic delineation with a stable RGB value for each elastography image. C. The Euclidean distance was calculated with the RGB value per pixel in the colored part in comparison with each row in the color bar. The pixel in the colored part was assigned with a stiffness value based on the shortest Euclidean distance with a pixel in the color bar. D. The resulting image where each pixel had a stiffness value, as shown in the new color bar.

4.3.2 Superpixel segmentation

Thereafter, the US B-mode images were subdivided into superpixels. A superpixel is a set of connected pixels with a similar intensity. The use of superpixels was a tradeoff between feeding a complete ultrasound elastography image to an algorithm or feeding each pixel separately to an algorithm. Complete images could be used to train a neural network to predict the tumor area. However, training a neural network requires a huge amount of labeled data.[107] The advantage of superpixels to pixels is that a superpixel is less sensitive for noisy pixels while the local information of a pixel is also taken into account to extract representative features. A 2-D superpixel segmentation technique was used to automatically obtain groups of pixels that share common characteristics. Superpixel algorithms over-segment an image into several homogeneous subregions. One of the superpixel algorithms is simple linear iterative clustering (SLIC). This method is superior to other algorithms in showing image boundaries, it is faster, memory efficient, easy to use, flexible in the compactness and number of the superpixels it generates.[123,124] The SLIC algorithm has been used for tumor segmentation from B-mode ultrasound images before.[124,125]

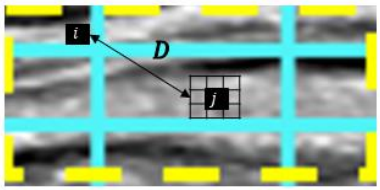
The SLIC algorithm developed by Achanta et al. is shown in Figure **4-11**.[123] The SLIC algorithm starts by dividing an image into a number (k) of clusters (C_k). The number k can be adjusted as an input parameter. Each cluster consists of S number of pixels, where $S = \frac{N}{k}$ and N is the total number of pixels in the image. Secondly, the center of the cluster is determined as the 3x3 neighborhood with the lowest gradient. This means that the center is determined at an area of 3x3 pixels where the change of intensities is the lowest relative to the rest of the cluster. This center movement prevents the superpixel from localizing at a boundary or noisy pixels. Thirdly, the assignment step was performed iteratively per cluster. The assignment step starts with defining the searching area that contains twice as many pixels as the cluster.

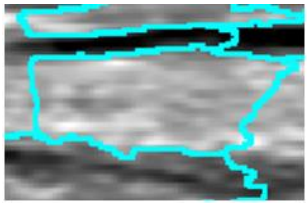


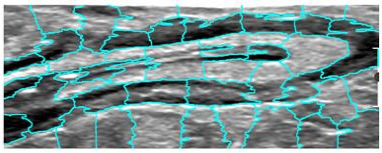
1. Division image in clusters C_k with S number of pixels per cluster. S = N/k, with N for the number of pixels in the image and k for the number of superpixels wanted.

2. Zoomed in at one cluster. The cluster center moves to the 3x3 neighborhood with the lowest gradient, which is equal to the lowest difference in grey intensities within a 3x3 matrix of pixels.

3. The distance D is calculated per cluster C_k, for each pixel within a searching area of 2S pixels, so twice as many pixels as the original cluster (delineated with yellow).







4. The distance D is the distance between pixel 'i' and cluster center 'j'. D concerns the distance in position and grey value. When distance D is lower for cluster k, than for another cluster, the cluster number k is assigned to pixel i.

5. After repeating this process for each pixel in the searching area, the contour is adjusted to pixels with corresponding grey values.

6. Result of superpixel over-segmentation when distance D is calculated iteratively for each cluster and pixel in the searching area. The resulting superpixels consist of grouped pixels with about the same grey value.

Figure 4-11: Simple linear iterative clustering algorithm to extract superpixels[123]

Fourthly, distance D was calculated iteratively for each pixel in the searching area. Distance D takes into account the position and intensity of the pixel to the cluster center. The calculation of distance D is shown in Equations **4-2** to **4-4**. Where d_g is the absolute difference between the grey value of the cluster center g_j and grey value of the pixel in the searching area g_i. Furthermore, d_{xy} is the distance between the cluster center j and the pixel in the searching area i. The positions of i and j are described with x- and y-coordinates. Through dividing $d_{xy}$by S, the distance is converted to a relative distance. Furthermore, factor m was introduced to control the compactness of a superpixel. When m is large, distance D depends more on the difference in grey intensities. When m is small, d_{xy} has more influence, and the superpixels will be more compact. The standard range of m is in the range of 1- 40.

$$d_g = |g_j - g_i| \quad (4\text{-}2)$$

$$d_{xy} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (4\text{-}3)$$

$$D = d_g + \frac{m}{S} d_{xy} \quad (4\text{-}4)$$

The searching area $2S$ overlaps with the searching areas for other clusters, thus the distance D was calculated to multiple clusters per pixel. The smallest distance D to a cluster determined the cluster number k that was assigned to that pixel. When a new pixel is connected to a cluster, the cluster center is adjusted to the 3x3 neighborhood with the lowest gradient again. The displacement of the cluster center is the residual error. Computing the new cluster centers and residual errors are known as the update steps. The residual error can be minimalized by repeating the assignment and update steps through the iteration parameter. The entire algorithm is described in Algorithm **4-1**.

Algorithm 4-1: Simple linear iterative clustering superpixel segmentation for grey intensity images[123]

```
//Initialization
//Initialize cluster centers C_k = [g_k, x_k, y_k]^T by sampling pixels at regular grid step S.
Move cluster centers to the lowest gradient position in a 3x3 neighborhood
Set label l(i) = −1 for each pixel i
Set distance d(i) = ∞ for each pixel i
Repeat
    // Assignment
    for each cluster center C_k do
        for each pixel i in a 2S x 2S region around C_k do
        Compute the distance D between Ck and i.
                if D < d(i) then
                        set d(i) = D
                        set l(i) = k
                end if
        end for
    end for
    // Update
    Compute new cluster centers
    Compute residual error E
```

The superpixel segmentation was applied to the region in the US B-mode image, which corresponds to the available elastography map, as shown in Figure **4-12**A-B. The number of clusters was set to 40 since this number was enough to select regions with the same intensity in the majority of the US B-mode images. The number of clusters is a trade-off between selecting small enough regions of just the same tissue type while selecting a large enough area to extract features that contain useful information about an area. Furthermore, a higher number of superpixels would increase the workload, since each superpixel must be labeled with a tissue type. After the 2D-superpixel segmentation, each superpixel was turned into a mask image, as shown in Figure **4-12**C-D. Erosion was applied to the mask image to remove the boundary pixels. The eroded superpixel is shown in Figure **4-12**E. This eroded mask was superimposed on the elastography image, as shown in Figure **4-12**F. Thereafter, B-mode and elastography features were extracted per superpixel.

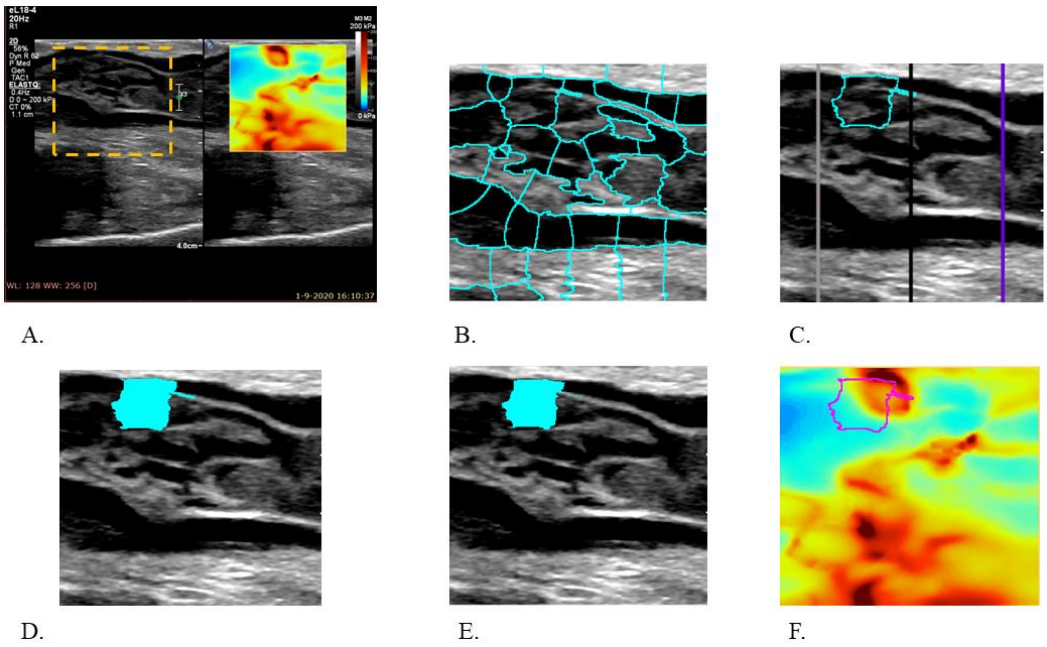


Figure 4-12: An overview of the superpixel selection method. A. The region of the colored part in the elastography image was extracted from the B-mode image. B. The automatic superpixel selection was applied to this region. C. Each superpixel was selected one by one. D. Mask image of the pixels within the superpixel. E. The mask of the eroded superpixel with the boundary pixels. F. The location of the superpixel was applied to the elastography map to calculate elastography variables per superpixel.

4.3.3 Feature extraction

The shear wave elastography and B-mode features were extracted per superpixel. The elastography features include the mean, standard deviation, minimum, and maximum stiffness. Furthermore, histogram features were extracted, which show the number of stiffness values within a range of 10 kPa relative to the superpixel size. Since the total range was 0-200 kPa, 20 histogram features were extracted. An example of a histogram from one superpixel is shown in Figure **4-13**A. Besides, the interquartile range (IQR) of the histogram was extracted. The IQR is the middle 50% of the relative histogram values. The B-mode features were retrieved from the same superpixels. The US B-mode data consist of grey values between 0 and 1, with an intensity of 0 for black pixels and 1 for white pixels. Preprocessing contrast adjustment was applied on the B-mode images to emphasize contrast differences, and to normalize the different US B-mode images. Thereafter, similar parameters were retrieved from the grey intensity values as the elastography features, thus the mean, standard deviation, maximum, and minimum intensity. Furthermore, relative histogram values of grey intensities from 0 to 1 with a bin width of 0.1 were extracted, an example is shown in Figure **4-13**B. Besides, the IQR of the histogram was extracted.

In total, it resulted in 40 features, with 25 SWE features and 15 US B-mode features per superpixel. After repeating the feature extraction for all superpixels, the features were combined into one table with one row per superpixel, and its 40 features. Furthermore, the superpixel number, image number, and specimen number were added to each superpixel. Therefore, it could be tracked to which superpixel, image, and specimen a data sample belongs when the features are combined into one dataset from all B-mode and SWE images of the ex vivo study.

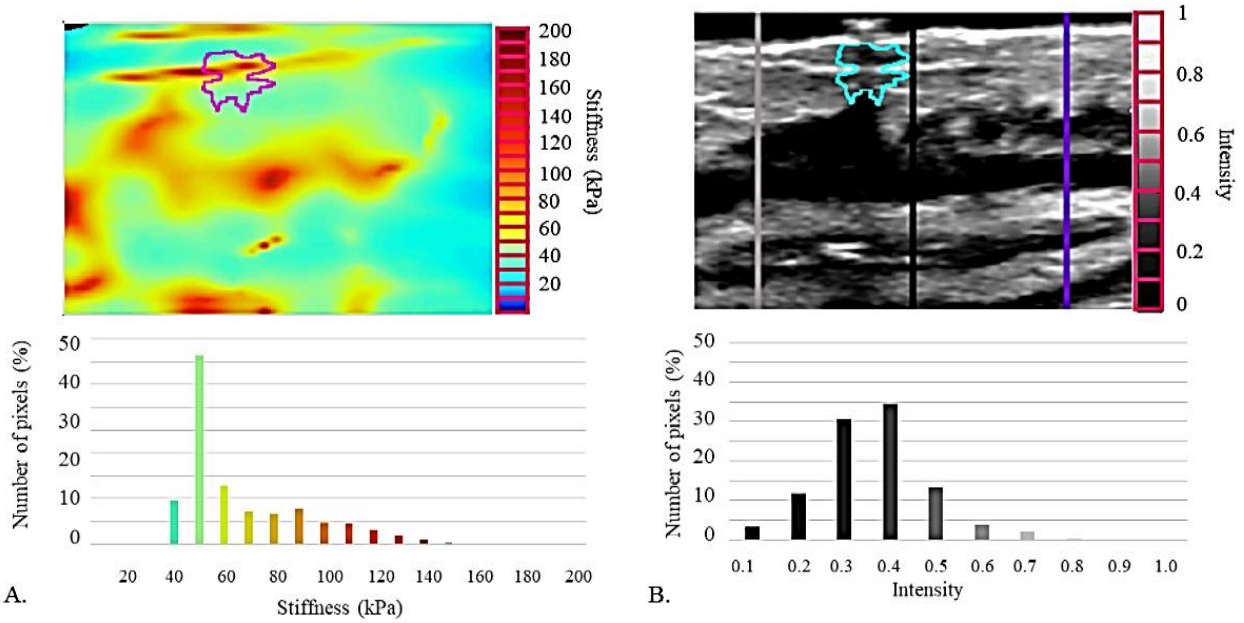


Figure 4-13: Explanation of histogram features. These features are expressed in a percentage that presents the ratio of pixels with a certain stiffness/intensity value relative to the total number of pixels in a superpixel. A. The elastography histogram features were determined per stiffness range with a bin width of 10 kPa, where bin 1 consists of pixels with a value between 1-10 kPa, bin 2 for 11-20 kPa until bin 20 for 190-200 kPa. This histogram shows for example that 45% of the pixels in this superpixel has a stiffness between 41 and 50 kPa. B. The US B-mode histogram features were determined per intensity range with a bin width of 0.1. Bin 1 consisted of pixels with an intensity between 0 - 0.1, bin 2 0.11 - 0.2 until bin 10 for 0.9-1.0. The histogram shows that most of the pixels have an intensity value between 0.31 and 0.4.

4.3.4 Tissue type labeling

Each superpixel was labeled with a tissue type based on the histopathology results. The labeling took place blindly for the US B-mode and SWE features for the specific superpixels. The histopathology coupes and US images were correlated through the marker locations. In the pathology coupe, vertical lines were drawn perpendicularly from the black and purple markers on the surface to the lumen, as shown in Figure **4-14**A.

The marker location lines in the US-B-mode image were determined based on the molds that were used to apply the markers during the ex vivo measurements. Through the perfect alignment of the molds, the central marker location at the specimen was at the midline of both US B-mode images. The location of the purple line was at the right side in both US B-mode images through the orientation of the transducer during the ex vivo measurements. Both the US B-mode images from the L15-7 and eL18-4 US transducers were used for the correlation with histopathology. By comparing anatomical landmarks on both US B-mode images, it could be noticed when the eL18-4 transducer was slightly replaced during the ex vivo measurements concerning the markers. This was mainly important for the measurements acquired with Protocol I since the eL18-4 transducer could be shifted easily without the mold. The distance between the central marker and the two markers aside was 8 mm during the ex vivo measurement (Figure **4-14**B). This distance was turned into a number of pixels based on the pixel size (Figure **4-14**C-E). In the US B-mode image from the superficial L15-7 transducer, the pixel size was 0.066 mm, so the distance between the central marker and the markers aside was 8/0.066 = 120 pixels. The pixel size in the US B-mode image of the eL18-4 transducer was 0.071 mm, so the distance between the central marker and the markers aside was 8/0.071 = 113 pixels. Since the superpixel image is a part of the US B-mode image from the eL18-4 transducer, there were also 113 pixels between the marker locations in the superpixel image. However, the central marker is not automatically at the central pixel in the superpixel image. Therefore, the marker locations in the superpixel image were determined by subtraction of the pixel column numbers of the markers in the US B-mode image with the first column number of the superpixel region in the US B-mode image (blue delineation in Figure **4-14**D).

Another important landmark to correlate the pathology coupe with the ultrasound images was the lumen. The bottom of the histopathology coupe corresponds to the mucosa, which is the tissue layer adjacent to the bowel lumen. In the ultrasound B-mode images, the lumen is recognizable as a hypoechoic or hyperechoic area between the colorectal tissue layers as described in Section 3.2.2. The lumen is shown as the yellow dashed line in Figure **4-14**C-D.

The location of a superpixel was correlated to a location in the histopathology coupe based on the marker lines and the lumen. The tissue type at that location in the coupe was assigned as the golden standard label. As shown in Section **4.2.3**, the tumor cells were delineated with red, and fibrotic tissue was delineated with purple. Furthermore, fat could be recognized as white cells, and muscle as the pink cells outside the delineations. Each superpixel in the US image was labeled with 'tumor', 'fibrosis', 'muscle', ' fat', or 'nothing'. When there was a mixture of tissue types in a superpixel, a different label was assigned describing the two tissue types, for example, 'mix tumor fibrosis'. The 'nothing' label was assigned to the gel layer between the US transducer and the specimen, the table beneath the specimen, and locations with ultrasonic artifacts. Furthermore, the lumen and the tissue layers beneath the lumen were assigned as 'nothing', since the tissue type was unknown for these tissue layers. Next to the tissue type label, a certainty score between 1 and 10 was assigned to each superpixel with a 1 for unclear labels, and a 10 for labels we were very certain of.

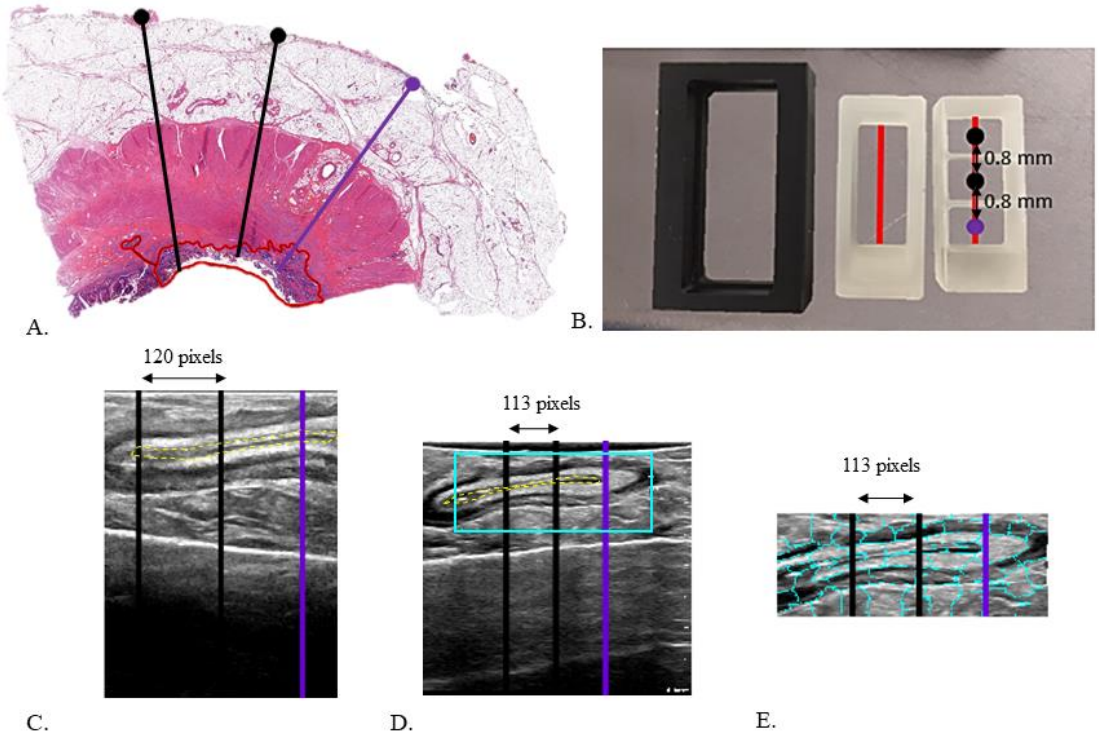


Figure 4-14: Method to correlate superpixel locations with histopathology coupes. The black and purple vertical lines show the locations beneath the black and purple markers at the surface of the specimen. A. H&E stained histopathological coupe of a rectum specimen, with vertical lines perpendicular to the surface and lumen at the three marker locations. The tumor is delineated with red. B. The molds that were used for marker placement at the specimen, showing that the distance between markers was 0.8 mm. C. US B-mode image from the superficial L15-7 US transducer, where the lumen is delineated with a yellow dashed line. The central black line is located exactly in the middle of the image and the locations of the lines aside are determined based on the 8 mm distance between de markers and a pixel size of 0.066 mm. D. US B-mode image from the eL18-4 ultrasound transducer, where the blue box shows the location of the superpixel map. The central black line is located exactly in the middle of the image, and the lines aside are determined based on the 8 mm distance and pixel size of 0.071 mm. E. Superpixel map as a part of the US B-mode image from the eL18-4 transducer.

4.3.5 Data visualization

The features per tissue type were visualized to get an insight into how individual features differed per tissue type. The features were shown from the superpixels that were labeled with one tissue type, and a certainty score of 10. The two-sample t-test was applied on the mean, maximum, minimum, and standard deviation to see whether these features were significantly different for tumor tissue than fibrosis, fat, or muscle. A significance level of 5% was used.

4.3.6 Feature ranking

Feature ranking was performed to find which individual features had the highest correlation with discrimination between tumor tissue, fibrosis, and healthy tissue. Feature ranking gives an insight into how important individual features are. Several techniques can be used for feature importance analysis. The choice for a certain method depends on whether the statistic problem concerns a classification or regression problem. Furthermore, it depends on the feature types which are divided into categorical and/or continuous variables. The current statistical problem is a classification problem between malignant and healthy tissue, and the variables concern continuous variables. Three common feature selection methods that are suitable based on these two characteristics are Maximum Relevance Minimum Redundancy (mRMR), ReliefF, and the Chi-square test. mRMR is a method that looks for the feature set with the highest dependency for the outcome but with minimal mutual information between features in the set.[126] ReliefF is a feature weight-based algorithm that operates by randomly picking instances and then finds their 10 nearest neighbors from the same class and other classes. After that, the features of these samples are compared to the randomly selected samples and then used in updating the rank of each feature. It gives higher ranks to the features that have the same value in the same class but are different in the opposite class.[127] The Chi-square test assesses the usefulness of features by measuring the value of the Chi-square statistic p-value concerning its relevance with classes. This test only works with discrete variables.

All three techniques were implemented in the Statistics and Machine Learning Toolbox in Matlab R2020a. Therefore, these techniques could be fastly applied to the extracted features and tissue type labels. The tissue type labels were turned into binary labels with tumor and mixtures with tumor as 1, and muscle, fat, fibrosis, and mixtures of these tissue types as 0. The ten most important features were determined per feature selection method and compared to each other.

4.3.7 Machine learning algorithm

The binary tissue type labels (healthy and tumor) and 40 extracted features from the US B-mode and SWE superpixels were used as input for a machine learning classification algorithm. A flowchart of the machine learning model is demonstrated in Figure **4-15**. The model starts by dividing the complete dataset into a training set and test set. Thereafter, the binary labels and features were used as input for the feature subset selection. The feature subset selection was performed to reduce the number of features. The selected feature set was used for the training and testing of classification models. The predicted scores resulting from the classification models were scores between 0 and 1. By comparing the predicted scores to the input binary labels, a performance assessment was performed, resulting in an area under the receiver operating curve (AUC) and a Matthews correlation coefficient (MCC). These performance measures were used to compare different classification models. The different steps in the machine learning model will be explained in more detail in the following sections.

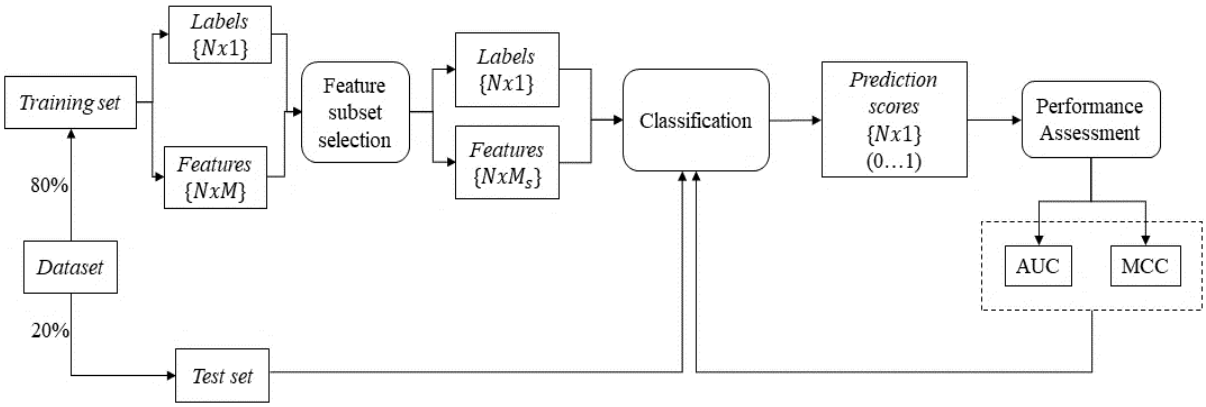


Figure 4-15: Overview of the machine learning classification model. It starts with dividing the complete dataset into a train and test set. The features and labels of the training set are used as input for the feature subset selection. The features are fitted in an NxM matrix, with N for the number of superpixels, and M for the number of features. The binary labels with N number of superpixels contain values of 1 for tumor and 0 for muscle/fat/fibrosis. Thereafter, feature subset selection was performed to reduce the number of features. The resulting feature subset and the binary labels were used as input for a classification algorithm. The classification algorithm resulted in a prediction score between 0 and 1 for each sample. Performance assessment was done by comparing the prediction score to the golden standard input label. From the performance assessment, the area under the receiver operating curve (AUC) and Matthews correlation coefficient (MCC) were determined. These performance measures were used to compare several classification algorithms.

4.3.8 Training and test set

The total dataset was subdivided into a train and test set. The total dataset consists of superpixels from all ex vivo measurements with ultrasound elastography. Each row in the dataset represents a superpixel with the corresponding specimen number, image number, superpixel number, B-mode features, elastography features, the tissue type label, and a certainty score of the tissue type label. The dataset was randomly divided into a training and test set based on the specimen number. As a result, 80% of the specimens belonged to the training set, and 20% of the specimens belonged to the test set. The division was based on the specimen number since multiple superpixels were retrieved from one specimen. The superpixels of one specimen may be highly correlated to each other, and when superpixels of one specimen would be present in both the training and test set, the classification algorithm is trained and tested on similar data. Since the goal of testing is to determine the performance of the classification algorithm to unseen data from new specimens, the test set should contain data from other specimens than the training set. The training set was used for feature selection, training and validating the classification algorithm. Afterward, the test set was used to assess the performance of the classification algorithm.

4.3.9 Feature subset selection

Feature selection was performed to prevent overfitting. Furthermore, a too large feature set involves more processing time and memory.[113] Sequential forward feature selection (SFFS) was used for the features selection.[128] Since the individual feature importance can differ from the importance of multiple features together, the feature ranking methods described in Section **4.3.6** are not sufficient to select the best feature set for the classification problem. Feature subset selection aims to obtain a subset of features to achieve good classification performance and high computational efficiency. SFFS is a bottom-up search procedure that starts from an empty set and gradually adds features selected by an evaluation criterion.[129] SFFS is an iterating process in which new features are included that produce a minimum classification error compared with the addition of any other feature.

The input data for the SFFS algorithm consisted of samples that were labeled with a certainty score of 10. The SFFS algorithm uses the loss criterion to select features and to determine when to stop. The loss measures the misclassification rate for the validation set. When adding new features to the feature subset, the loss should be decreasing. When the loss increases, the SFFS stops with feature selection. The feature set with the lowest misclassification rate can be determined by calculating the loss-function

for different candidate feature subsets. 10-Fold holdout cross-validation was used to select different candidate feature sets. The selected feature set contained both US B-mode and SWE features. Subsequently, this feature set was used to train and classification algorithms.

4.3.10 Classification

Tissue type classification was used to predict whether a superpixel contained tumor tissue or not, based on the B-mode and SWE features. Firstly, some initial experiments with the training data were performed to see which classifiers were suitable for this classification problem. The four classifiers with the best performance were selected.

Firstly, the classifiers were trained with samples that had a certainty score of 10. Secondly, the training set consisted of samples with a certainty score of 10 and 9, and so on until the fifth time with certainty scores from 6 to 10. Through this process, the advantage of using more training data could be compared to the increasing chance of including data with a wrong golden standard tissue type label. Subsequently, the training set was divided into training and validation sets using 10-fold hold-out cross-validation. The MCC was used to compare the performances of the four classifiers since the MCC is a good metric to use in unbalanced datasets.

Afterward, the classification algorithms were tested to see how the classification algorithm would perform on unseen data. The classification algorithms were also tested with five different test sets, with the same certainty score distribution as the training sets. Furthermore, Mcnemar's test was performed to assess whether the classifiers differed in predictions.

4.4 Results

The results show de patient characteristics from the specimens that were included in the data analysis. Secondly, the tissue types that were present in the histopathology coupes are shown. Thereafter, the US B-mode and SWE features are shown per tissue type. Thereafter, the results of the feature ranking are shown. Subsequently, the resulting feature subset from the SFFS feature selection is presented. Lastly, the results of the classification algorithms are shown.

4.4.1 Study population

Ex vivo data was retrieved from specimens of 30 patients. One specimen, number ex-24, was excluded from the study since the pathological coupes could not be retrieved. Therefore, the number of included patients is 29 instead of 30. The patient characteristics of the included specimens are shown in Table **4-1**. The distribution of tumor locations was equally divided over the colon, sigmoid, and rectum. Approximately half of the patients received neoadjuvant therapy (45%). The types of therapy differed among patients. Two patients received chemoradiotherapy in combination with immunotherapy, and six patients received chemoradiotherapy. Four patients received either chemotherapy, radiotherapy, or immunotherapy, and one patient received radio- and immunotherapy. The majority of the tumors were locally advanced and had a T3 or T4 grade. The maximum diameter of the tumors was on average 35 mm and ranged from 2-85 mm. Two specimens had a positive circumferential resection margin of 0 and 0.5 mm.

Table 4-1: Patient characteristics total study population

Characteristics	**Total study population (count (%))**
Total nr. of patients	29
Tumor location	
Colon	9 (31)
Sigmoid	10 (34)
Rectum	10 (34)
Neoadjuvant therapy	
Yes	**13 (45)**
Chemo- + radio- + immunotherapy	2 (15)[3]
Chemo- + radiotherapy	6 (46)
Radio- + immunotherapy	1 (8)
Chemotherapy	1 (8)
Radiotherapy	1 (8)
Immunotherapy	2 (15)
No	**16 (55)**
Pathological Tumor grade[1]	
T0	2 (7)
T1	1 (3)
T2	7 (24)
T3	12 (41)
T4	7 (24)
Pathological report	
Max. tumor diameter (mm)[2]	35 (2-85)[4]
CRM (mm)	19 (0-95)
CRM positive (<1 mm)	2 (7)

All values are shown as the count with the percentage relative to the total number of patients in the brackets.
CRM = circumferential resection margin according to pathological report
[1]tumor grade according to the pathological report of the resection
[2]maximum diameter of resected tumor according to pathological report
[3]percentage of patients who received neoadjuvant therapy
[4]the range with the minimum and maximum value in the train or test set

The total study population is split into a training and test set for the data analysis. The patient characteristics of the specimens in the training and test set are shown in Table **4-2**. The training set consists of 23 specimens and the test set consisted of six specimens. The specimens that belong to the test set are Ex-1, Ex-4, Ex-5, Ex-11, Ex-17, and Ex 20. The majority of the specimens in the test set is from tumors that were localized in the rectum (67%), while the majority in the training set is from the sigmoid (42%). The number of patients who received neoadjuvant therapy was in the minority in the training set (42%) but in the majority in the test set (83%). All patients who received neoadjuvant therapy in the test set received chemotherapy, four of the five received radiotherapy as well, and two patients received also immunotherapy. The tumor grade was mainly T3 or T4 in the training set while the test set had most tumors with grade T2. The maximum diameter of the tumor was generally smaller in the test set than the training set (21 vs. 39 mm). The training set contained two specimens with a positive margin, although the averaged CRM was higher in the training set than in the test set (21 vs. 13 mm).

Table 4-2: Patient characteristics of specimens in the train and test set

Characteristics	**Training set (count (%))**	**Test set (count (%))**
Total nr of patients	23	6
Tumor location		
Colon	7 (30)	2 (33)
Sigmoid	10 (43)	0
Rectum	6 (26)	4 (67)
Neoadjuvant therapy		
Yes	**8 (35)**	**5 (83)**
Chemo- + radio- + immunotherapy	0	2 (40)
Chemo- + radiotherapy	5 (63) [3]	2 (40)
Radio- + immunotherapy	1 (13)	0
Chemotherapy	0	1 (20)
Radiotherapy	1 (13)	0
Immunotherapy	2 (25)	0
No	**15 (65)**	**1 (17)**
Pathological Tumor grade [1]		
T0	2 (9)	0
T1	1 (4)	0
T2	4 (17)	3 (50)
T3	10 (43)	2 (33)
T4	6 (26)	1 (17)
Pathological report		
Max. diameter (mm) [2]	39 (2-85) [4]	21 (15-30)
CRM (mm)	21 (0,5-45)	13 (4-25)
CRM positive (<1 mm)	2	0

All values are shown as the count with the percentage relative to the total number of patients in the brackets.
CRM = circumferential resection margin according to pathological report
[1]tumor grade according to the pathological report of the resection
[2]maximum diameter of resected tumor according to pathological report
[3]percentage relative to the total number of patients who received neoadjuvant therapy
[4]the range with the minimum and maximum value in the train or test set

With 29 included ex vivo specimens and a maximum of three ultrasound B-mode and SWE measurements per surgical specimen, the total number of measurements amounted to 71. Six of these measurements were not usable due to unclear histopathology correlation. Therefore, 65 US B-mode images and corresponding SWE images from 29 surgical specimens are included in the data analysis.

4.4.2 Pathology

Each US B-mode and elastography measurement was correlated to a pathology coupe. Figure **4-16** shows the distribution of tumor and fibrosis in the coupes. It shows that 27 pathology coupes contained tumor tissue and fibrosis (41%). Furthermore, 25 coupes contained only tumor (38%), and 4 coupes contained only fibrosis (6%). The rest of the coupes did not contain tumor or fibrosis (15%). Figure **4-17** shows some pathology coupes of specimens from patients who received neoadjuvant therapy. Thereby, the corresponding US B-mode, and SWE images are shown. Furthermore, Figure **4-18** shows pathology coupes of specimens from patients who did not receive neoadjuvant therapy. It is shown that the amount of tumor tissue and fibrosis varies strongly per pathology coupe. Furthermore, fibrosis is more often present in (one of) the pathology coupes of specimens from patients who received neoadjuvant therapy than patients who did not (69% vs. 56%).

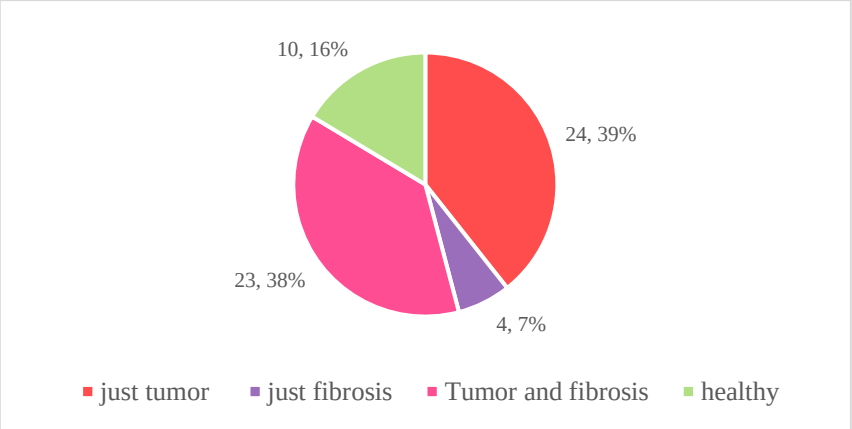
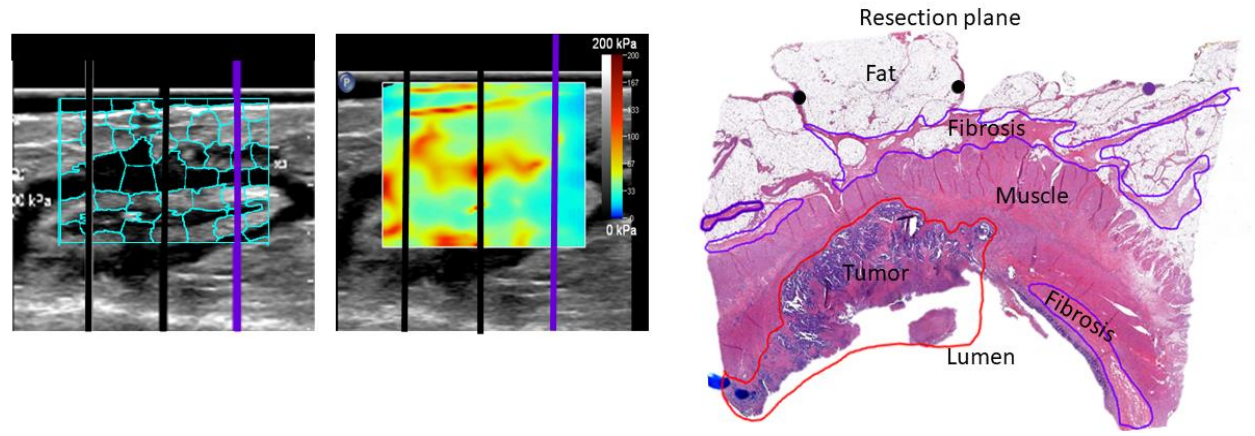
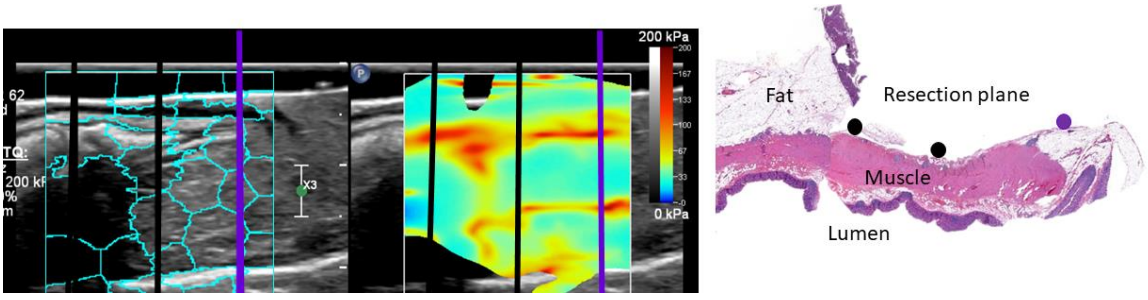


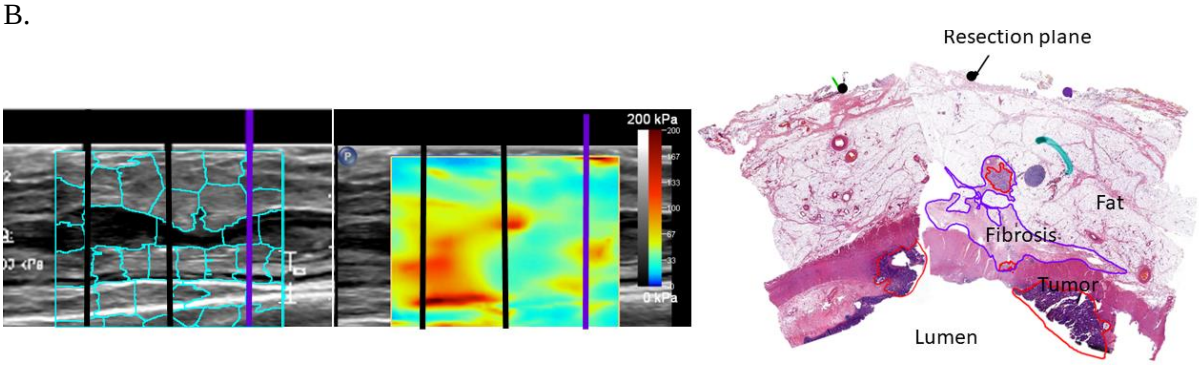
Figure 4-16: Distribution of tumor and fibrosis in the pathology coupes that correspond to the ultrasound and shear wave elastography measurements.



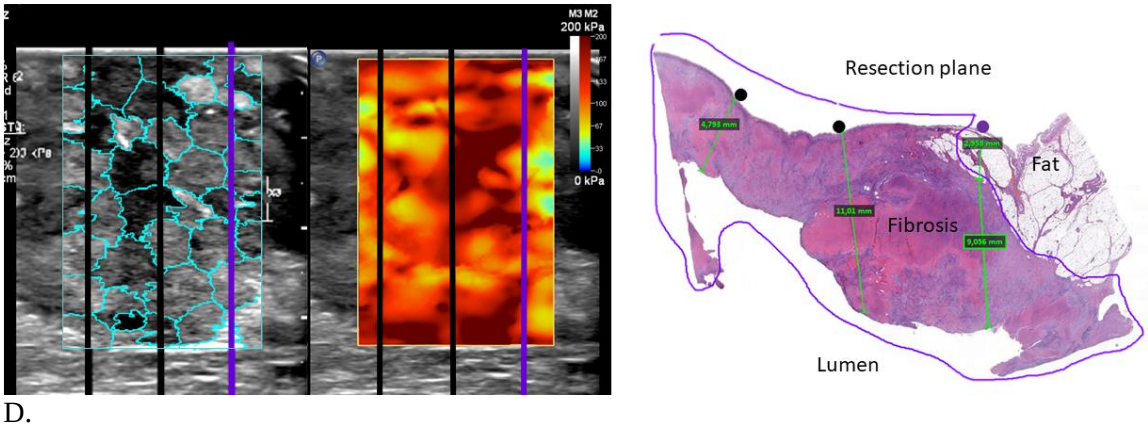
A.



B.

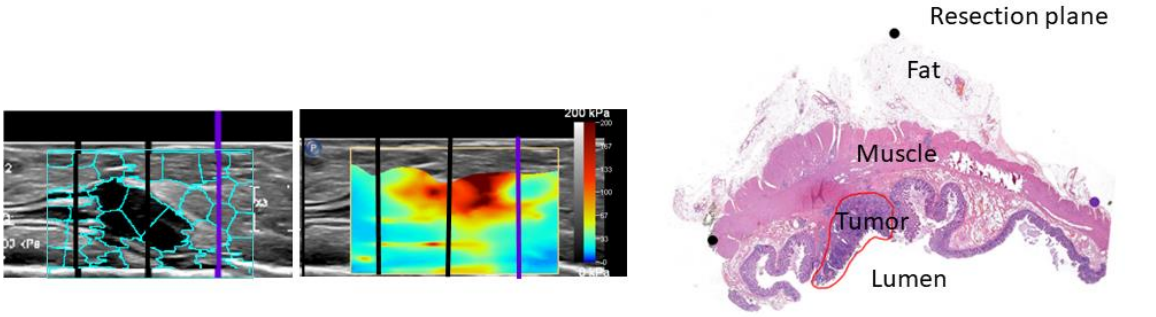


C.

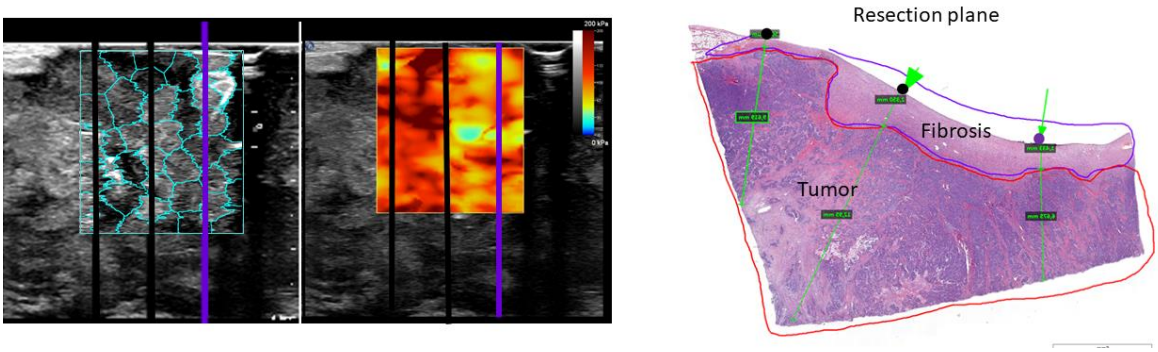


D.

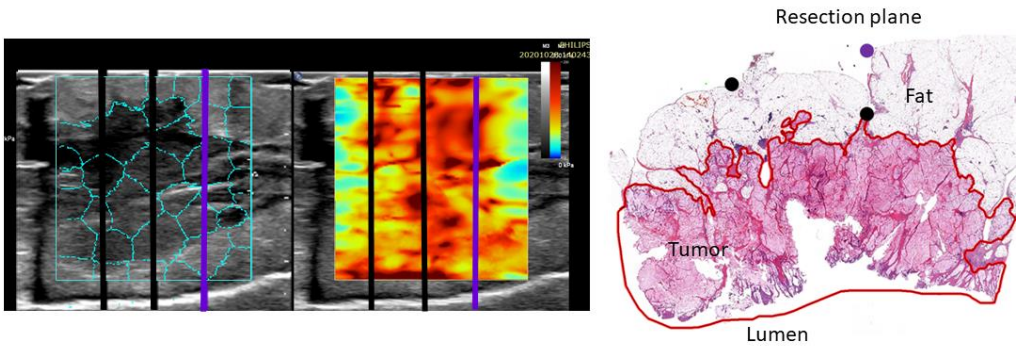
Figure 4-17: Results of ex vivo measurements from patients who received neoadjuvant therapy. The ultrasound B-mode image, shear wave elastography image, and pathology coupe are shown per specimen. A. Results from Ex-1, a rectum specimen with a T2 tumor. The pathology coupe contains tumor with muscle, fibrosis, and fat on top. B. Results from Ex-3, a colon specimen with no tumor left. The pathology coupe contains just muscle and fat. C. Results from Ex-17, a rectum specimen with a T3 tumor. The pathology coupe contains several areas with tumor cells and an area with fibrosis. D. Results from Ex-21, a colon specimen with no tumor left. The pathology coupe contains mainly fibrosis and a small area of fat.



A.



B.



C.

Figure 4-18: Results of ex vivo measurements from patients who did not receive neoadjuvant therapy. The ultrasound B-mode image, shear wave elastography image, and pathology coupe are shown per specimen. A. Results from Ex-8, a sigmoid specimen with a T1 tumor. The coupe contains a small area with tumor cells, and for the rest healthy muscle and fat. B. Results from Ex-22, a colon specimen with a T3 tumor. The pathology coupe contains a small layer of fibrosis on top of the tumor. C. Results from Ex-29, a sigmoid specimen with a T3 tumor. The pathology coupe contains mainly tumor tissue with a layer of fat on top.

Since the pathology coupes were used to label the US B-mode and SWE images per superpixel, the distribution of tissue types per superpixel is important as well. After excluding unlabeled superpixels and the specimen-wise training and test set division, the training set consisted of 2752 superpixels and the test set consisted of 423 superpixels. Figure **4-19**A-B shows the distribution of tissue types among the superpixels in the train and test set. The tissue types are grouped into tumor tissue, fibrosis, and healthy tissue. The tumor group contains superpixels that are labeled with tumor tissue, and mixtures of tumor tissue with other tissue types (fibrosis, fat, or muscle). The fibrosis group consists of superpixels that are labeled with fibrosis, and mixtures of fibrosis with fat or muscle. The healthy group contains superpixels that are labeled with fat, muscle, or mixtures of fat and muscle. There were more superpixels labeled with one tissue type than with mixtures of tissue types. The distribution was 92% vs 8% in the training set, and 84% vs 16% in the test set. Furthermore, the distribution per tissue type group differs for the train and test set. The tumor group includes 41% of the training set and 21% of the test set. Thereby, the healthy group comprises 45% in the training set and 71% in the test set.

Figure **4-19**C-D shows the distribution of tissue types among the superpixels with a certainty score of 10. The percentage of superpixels with a certainty score of 10 was 79% in the training set and 68% in the test set. The tissue type group distribution was almost the same for the training set, but not for the test set. The tumor group decreased from 21% to 6%, and the healthy group increased from 71% to 87%. Both the training and test set contained fewer superpixels that were labeled with mixtures of tissue types in this subset with certainty scores of 10 compared to the total training and test sets, 7%, and 6% respectively.

Figure 4-19: Distribution of pathology labels in the superpixels per tissue type group. The tumor group includes superpixels with tumor tissue and mixtures of tumor tissue. The fibrosis group includes superpixels with fibrosis and mixtures of fibrosis with fat or muscle. The healthy labels include superpixels with muscle, fat, or a mixture of muscle and fat. A. Distribution of tissue types in the total training set. B. Distribution of tissue types in the total test set. C. Distribution of tissue types in the superpixels that have a certainty score of 10 in the train set. D. Distribution of tissue types in the superpixels that have a certainty score of 10 in the test set.

4.4.3 Features and tissue types

In this section, the resulting US B-mode and shear wave elastography images and features are demonstrated. Figure **4-17** showed in the previous section some US B-mode and SWE images of specimens from patients who received neoadjuvant therapy. Furthermore, Figure **4-18** showed in the previous section B-mode and SWE images of specimens from patients who did not receive neoadjuvant therapy. Some of the B-mode images show (partly) the structure of a healthy bowel wall with alternating echogenicities (Figure **4-17** A-C and Figure **4-18**A). The B-mode images of tumor and fibrosis show hypoechoic areas that interrupt the normal bowel wall structure. The hypoechoic areas are corresponding to the location of tumor or fibrosis in the pathology coupes. Despite fibrosis is mostly less hypoechoic than tumor tissue, a clear boundary between the two tissue types is lacking (Figure **4-17** A, C and Figure **4-18**B). Furthermore, the B-mode image from the healthy colon shows a hypoechoic area but that area is beneath the lumen (Figure **4-17** B). Since the pathology coupe just shows the upper tissue layers until the lumen, the pathological information of this hypoechoic region is unknown. The SWE images from pathology coupes with tumor and fibrosis show red areas that represent stiff tissues. These red areas are mostly corresponding with the locations of tumor or fibrosis. However, there are also some stiff areas shown at locations without tumor tissue. These stiff areas are mainly located at boundaries between the mucosa and the lumen or near the contact surface of the US transducer. (Figure **4-17**A-B) The SWE images neither show a clear boundary between tumor and fibrosis, muscle, and fat.

All features presented in this section are retrieved from the total dataset (training + test set), from superpixels with a certainty score of 10, and labeled with one tissue type. The mean, standard deviation, maximum, and minimum intensities of the US B-mode images are shown per tissue type in Figure **4-20**. Table **4-3** shows the mean and 95% CI of these features and of the IQR. All features differ significantly for tumor tissue in comparison to the other three tissue types (p-value <0.01). The largest difference is observed for the maximum intensity and the mean intensity. All four intensity features are on average the lowest for tumor tissue, meaning that tumors are mostly shown more hypoechoic compared to the other tissue types. The average IQR of all superpixels with tumor tissue was lower than the average IQR for the other three tissue types. This means that tumor has less variety in intensity values within one superpixel than other tissue types.

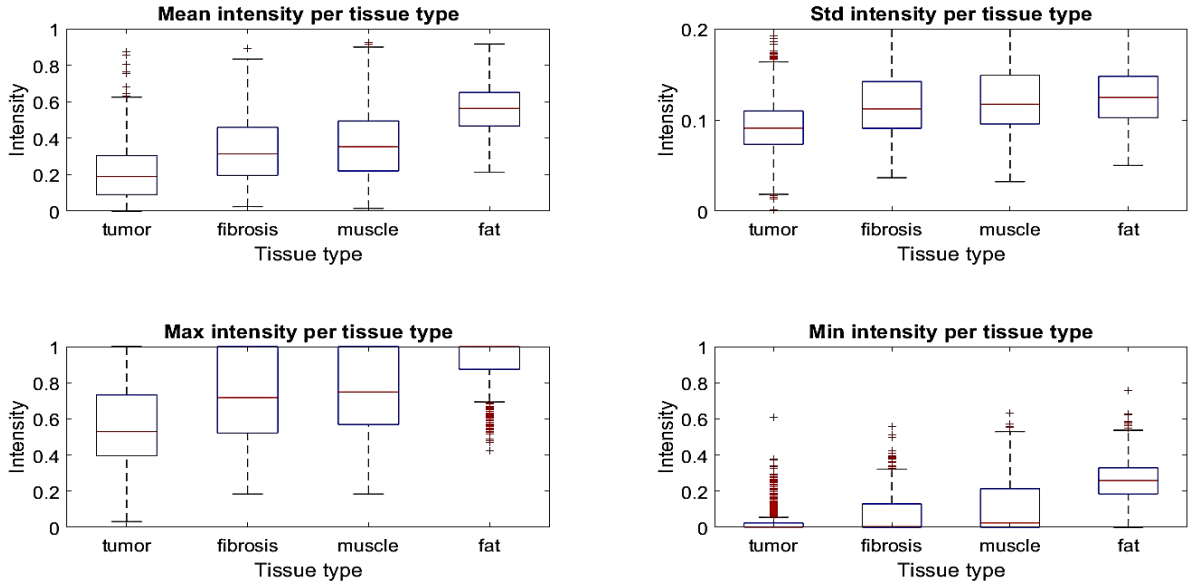


Figure 4-20: Boxplots of the mean, standard deviation, maximum, and minimum intensity on US B-mode images per tissue type. The red horizontal line shows the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles respectively. The whiskers extend to 1.5 times the interquartile range, and the outliers are plotted as the red plus signs. The feature values are retrieved from all superpixels in the total dataset with a certainty score of 10. It shows that the median of tumor tissue is lower than the median for the other tissue types for the mean, standard deviation, and maximum. However, the boxplots of tumor tissue are overlapping with the boxes of the other tissue types.

Table 4-3: Intensity features from the US B-mode images per tissue type

Tissue type	Mean	Standard deviation	Maximum	Minimum	IQR(%)
Tumor	0.20 (0.19-0.21)	0.09 (0.09-0.10)	0.54 (0.52-0.56)	0.03 (0.03-0.04)	12.6 (12.1-13.2)
Fibrosis	0.34 (0.32-0.36)	0.12 (0.12-0.12)	0.71 (0.69-0.74)	0.07 (0.06-0.09)	16.2 (15.7-16.8)
Muscle	0.36 (0.35-0.38)	0.13 (0.12-0.13)	0.75 (0.73-0.76)	0.10 (0.09-0.12)	14.9 (14.6-15.3)
Fat	0.56 (0.55-0.57)	0.13 (0.13-0.13)	0.91 (0.90-0.92)	0.26 (0.25-0.26)	15.6 (15.3-15.9)

*The mean of each feature is shown in intensity values with the 95% confidence interval in the brackets. The first four features are shown as intensity values of US B-mode images. Only the IQR is shown in percentages, that are corresponding to the y-axis of the histograms in Figure **4-21**. The averaged values are retrieved from all superpixels in the total dataset with a certainty score of 10. Superpixels with a mixture of tissue types are not taken into account.*

Figure **4-21** shows the averaged histogram values of the US B-mode images per tissue type. The histograms of the four tissue types differ from each other. The fat histogram peaks in the center, with the largest amount of pixels with an intensity between 0.4 – 0.6. The histograms of muscle and fibrosis have both mainly pixels with an intensity between 0.2 – 0.5 and these histograms are flatter compared to the histograms of fat and tumor. The histogram of tumor tissue has the most pixels with an intensity value of 0.2 and 0.3 and the histogram is shifted to the left.

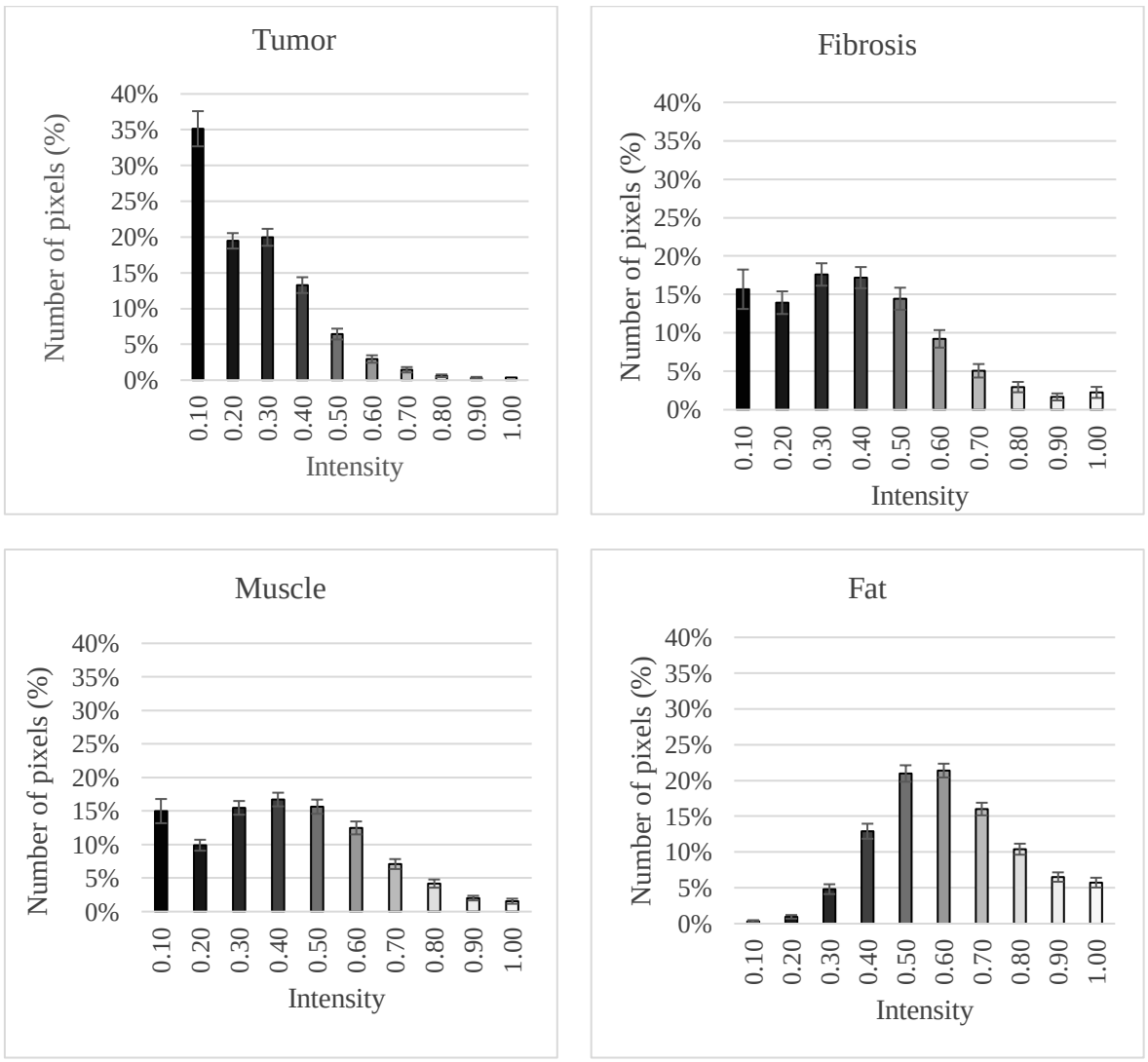


Figure 4-21: Averaged histograms of B-mode ultrasound intensities relative to the superpixel size. The intensity values are shown from the superpixels in the US B-mode images that are labeled with one of the four tissue types. The error bars show the 95% confidence intervals. The histogram values are retrieved from all superpixels in the total dataset with a certainty score of 10. Superpixels with a mixture of tissue types are not taken into account.

The mean, standard deviation, maximum, and minimum of the stiffness (kPa) from the SWE images are shown in Figure **4-22**. Table **4-4** shows the mean and 95% CI of these features and the IQR of the histogram values. The four SWE features have the highest values for fibrosis and the second highest value for tumor tissue. All four SWE features and the IQR are significantly higher for tumor tissue compared to muscle and fat (p-value <0.01). The mean stiffness and min stiffness are on average significantly higher for fibrosis compared to tumor tissue (p-value <0.01). The standard deviation in stiffness and the maximum stiffness are not significantly different for tumor and fibrosis (p-value >0.05). The boxplots show that the mean and maximum stiffness have the largest difference between tumor and both muscle and fat. Furthermore, it is shown that the mean, maximum, and minimum stiffness values spread over the complete range of 0-200 kPa for all tissue types. The IQR varies per tissue type, where the IQR of tumor tissue was lower than fibrosis but larger than muscle and fat. Therefore, tumor tissue has a larger variety in stiffness within one superpixel than muscle and fat, but a lower variety in stiffness than fibrosis.

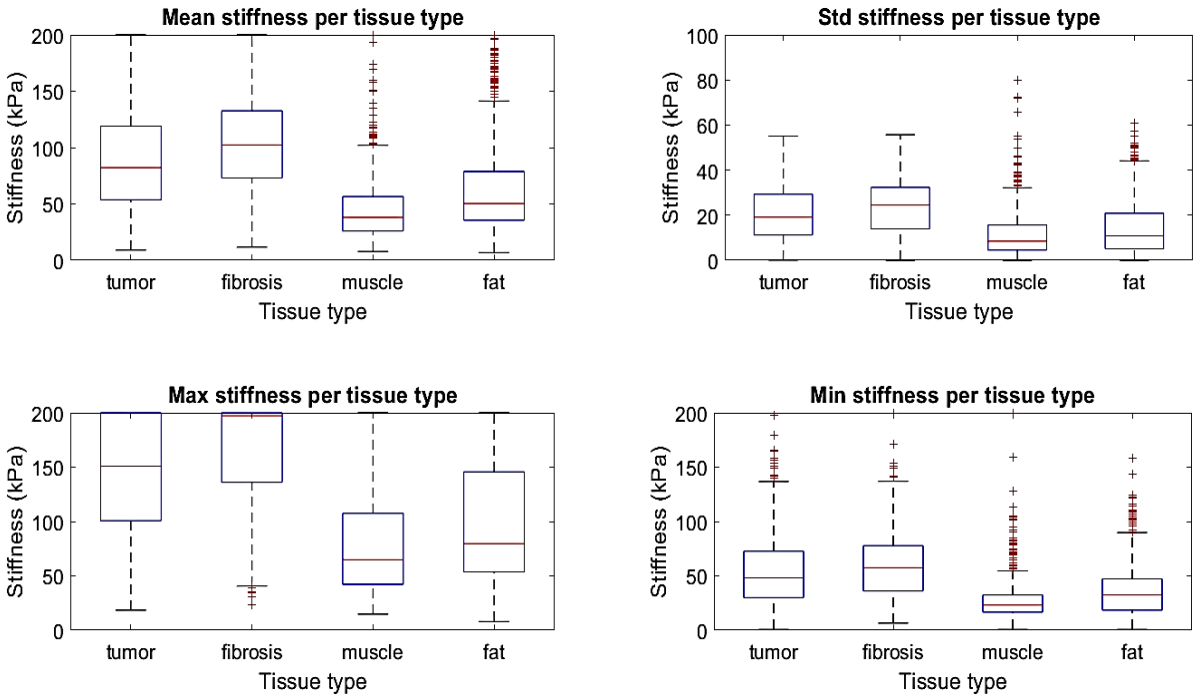


Figure 4-22: Shear wave elastography features per tissue type. The red horizontal line shows the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles respectively. The whiskers extend to 1.5 times the interquartile range, and the outliers are plotted as the red plus signs. All four features are significantly higher for tumor versus muscle and fat. The mean and minimum stiffness are significantly higher for fibrosis compared to tumor. The standard deviation and maximum stiffness are not significantly different for tumor tissue compared to fibrosis.

Tissue type	Mean (kPa)	Standard deviation (kPa)	Maximum (kPa)	Minimum (kPa)	IQR (%)
Tumor	90 (86-93)	21 (20-22)	144 (140-148)	54 (52-56)	5.3 (5.1-5.6)
Fibrosis	104 (99-109)	24 (23-25)	163 (158-169)	60 (57-64)	6.0 (5.6-6.3)
Muscle	46 (44-49)	12 (11-13)	82 (78-85)	29 (27-30)	2.5 (2.3-2.7)
Fat	63 (59-66)	15 (14-15)	100 (96-104)	37 (35-39)	3.2 (2.9-3.4)

The mean of each feature is shown as the stiffness in kilopascal with the 95% confidence interval in the brackets. Only the IQR is shown in percentages, that are corresponding to the y-axis of the histograms. The averaged features are retrieved from all superpixels in the total dataset with a certainty score of 10.

Table 4-4: Shear wave elastography histogram features per tissue type

The SWE histogram features are shown per tissue type in Figure **4-23**. The histograms show the relative number of pixels with stiffness values within ranges of 10 kPa. The first bar shows the percentage with a stiffness value from 1 up to and including 10 kPa, the second from 11 kPa up to and including 20 kPa, and so on. The histogram of the tumor is slightly shifted to the left. Superpixels with tumor tissue have on average the highest amount of pixels with stiffness between 40 and 60 kPa. The histogram of fibrosis shows that the majority of stiffness values is between 80 and 110 kPa. The histograms of tumor and fibrosis show both a larger spread in stiffness values than the muscle and fat. The histograms of muscle and fat are shifted to the left, with the highest amount of stiffness values between 20 and 40 kPa. Each histogram has a high peak at 200 kPa that deviates from the rest of the histogram.

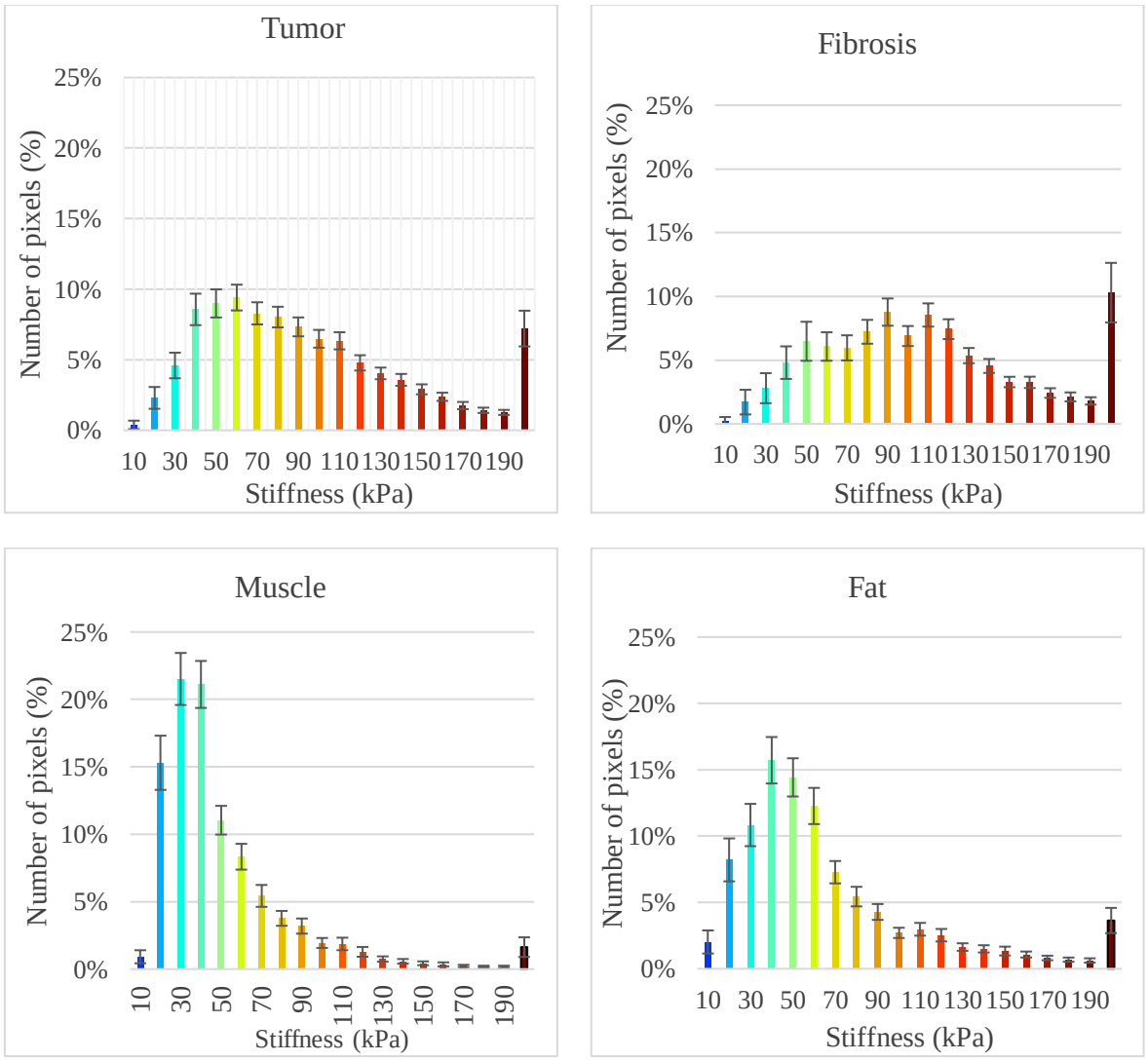


Figure 4-23: Averaged histograms of stiffness values in kilopascal relative to the superpixel sizes. The stiffness values are shown from superpixels in the shear wave elastography images, which are labeled with one of the four tissue types. The error bars show the 95% confidence intervals.

4.4.4 Feature ranking

The ten features with the highest rank according to mRMR, ReliefF, and Chi-square test are shown in Table **4-5**. It shows that each feature ranking method results in different features that have the highest correlation with the distinction between tumor and muscle, fat, or fibrosis. The features selected by mRMR and the Chi-square test are mainly US B-mode histogram features and have seven features in common. The ReliefF algorithm shows both US B-mode features and SWE features in the top ten. The histogram feature 'US B-mode bin 2' is selected by all three methods. The majority of selected features by the three methods are histogram features, that give information of a relative amount of a grey intensity value range or stiffness range. All US B-mode histogram features are selected by one or multiple algorithms. The distinctive elastography ranges are between 80-90 kPa, 110-120 kPa, 140-160 kPa, and 180-190 kPa according to the ReliefF.

Table 4-5: Ten features with the highest correlation score according to mRMR, ReliefF, and the Chi-square test

Feature importance	mRMR	ReliefF	Chi-square test
1 (most important)	Mean B-mode	B-mode bin 3 (0.21-0.30)	Mean B-mode
2	B-mode IQR histogram	Max. B-mode	B-mode bin 6 (0.51-0.60)
3	Mean Stiffness	B-mode bin 2 (0.11-0.20)	B-mode bin 7 (0.61-0.70)
4	B-mode bin10 (0.91-10)[a]	B-mode bin 4 (0.31-0.40)	Max B-mode
5	Std. B-mode	Elastography bin 19 (181-190)[b]	B-mode bin 2 (0.11-0.20)
6	Min. B-mode	B-mode IQR histogram	B-mode bin 8 (0.71-0.80)
7	B-mode bin 8 (0.71-0.80)	Elastography bin 9 (81-90)	B-mode bin 9 (0.81-0.90)
8	B-mode bin 6 (0.51-0.60)	Elastography bin 15 (141-150)	B-mode bin 1 (0.01-0.10)
9	B-mode bin 9 (0.81-0.90)	Elastography bin 12 (111-120)	Min. B-mode
10	B-mode bin 2 (0.11-0.20)	Elastography bin 16 (151-160)	Std. B-mode

[a]*The B-mode histogram features are shown with the bin number that represents a bar in the histogram of a certainty intensity range. For example, bin 2 is the 2nd bar with intensities between 0.1-0.2*

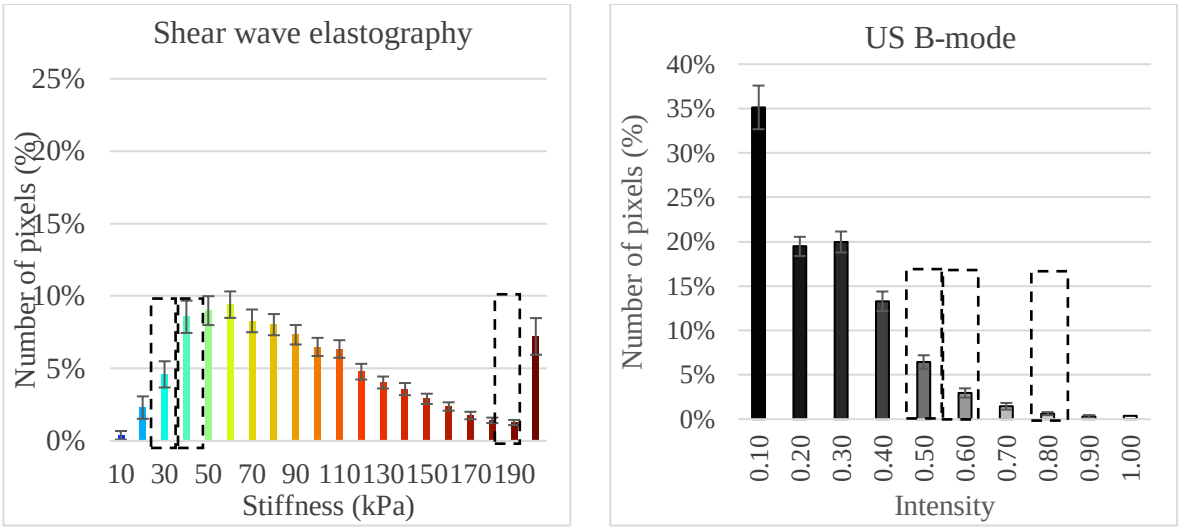
[b] *The elastography histogram features are shown with the bin number as well, which represents a bar with a certain stiffness range in kilopascal. For example, bin 9 is the 9th bar in the histogram with stiffness values between 80 and 90 kPa*

4.4.5 Feature subset selection

The feature subset selection resulted in nine features that are shown in Table **4-6**. The features were selected in order from the first row to the last row. In total, five US B-mode features and four elastography features were selected. Six of the nine features were histogram features, the histogram bars that were selected are shown in Figure **4-24**. Furthermore, the mean and standard deviation of B-mode intensity values and the IQR of the elastography histogram were selected. The nine selected features were used for the classification.

Table 4-6: Selected features from sequential forward feature selection

Feature importance	Feature name	Explanation
1 (most important)	Mean B-mode	The mean intensity of grey values
2	B-mode histogram bin 6	The relative amount of pixels with a grey value between 0.51 and 0.60
3	Elastography histogram bin 2	The relative amount of pixels with a stiffness value between 11-20 kPa
4	Std. B-mode	The standard deviation of grey values
5	B-mode histogram bin 5	The relative amount of pixels with a grey value between 0.41 and 0.50
6	Elastography IQR histogram	Interquartile range of histogram
7	Elastography histogram bin 19	The relative amount of pixels with a stiffness value between 181-190 kPa
8	B-mode histogram bin 8	The relative amount of pixels with a stiffness value between 0.71-0.80
9	Elastography histogram bin 3	The relative amount of pixels with a stiffness value between 21-30 kPa



A. B.

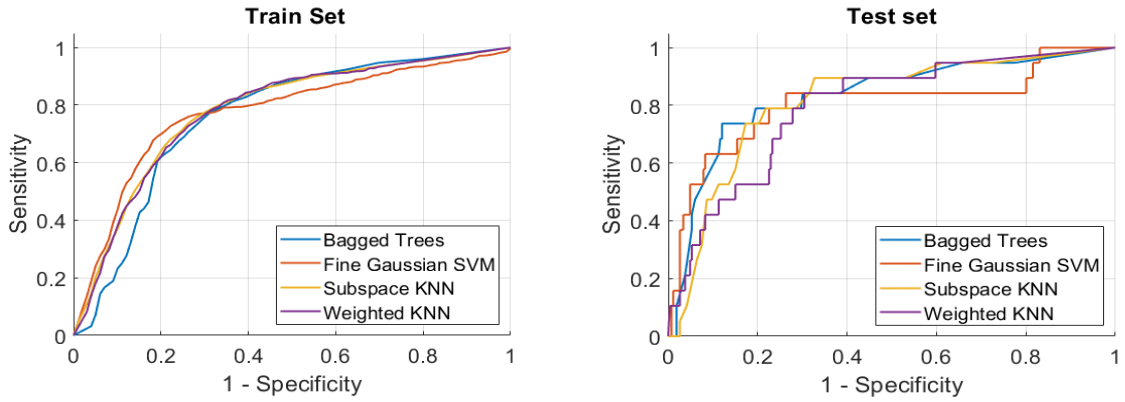
Figure 4-24: Selected histogram features shown in the color bars A. Elastography color bar with three selected stiffness regions. B. Ultrasound B-mode color bar with three selected regions.

4.4.1 Classification

The selected classifiers were Bagged Trees, Fine Gaussian SVM, Subspace KNN, and Weighted KNN. Table **4-7** shows the MCC values of these four classifiers after specimen-wise cross-validation which were trained with subsets of the training set based on certainty scores in labels. Five subsets were compared that differ from each other in the included certainty scores in the validation set. The first subset includes samples with a certainty score of 10, the second dataset included samples with a score of 10 and 9, until scores from 10 to 6. Therefore, the sample size differed for the datasets, nevertheless, the tumor class size was equal for all datasets. The second dataset results in the highest MCC for the Bagged Trees classifier. Whereas the first dataset results in the highest MCC for the Fine Gaussian SVM. The MCC's do not differ significantly between the different classifiers and datasets. The ROC curves of the classification results from the first dataset are shown in Figure **4-25**A. The ROC curves show that the Bagged Trees have a relatively low sensitivity in case of a specificity >0.8 compared to the other classifiers. The Fine Gaussian SVM curve shows that the sensitivity is relatively low with a specificity >0.7 compared to the other classifiers.

Table 4-7: MCC of the training set from four different classifiers and datasets with multiple certainty score ranges

Certainty scores	Sample size (#)	Tumor class (%)	Bagged Trees	Fine Gaussian SVM	Subspace KNN	Weighted KNN
Score 10	2186	40%	0.43	0.49	0.45	0.43
Score 9-10	2218	40%	0.49	0.48	0.44	0.43
Score 8-10	2349	41%	0.42	0.47	0.44	0.40
Score 7-10	2562	41%	0.44	0.44	0.40	0.39
Score 6-10	2702	41%	0.42	0.43	0.37	0.39



A. B.

Figure 4-25: Receiver operating characteristic curves of train and test set. A. The ROC curves of four different classifiers from the training set with a certainty score of 10. B. The ROC curves of four different classifiers from the test subset with a certainty score of 10 in the test set.

The classification results of the test set are shown in Table **4-8** and Figure **4-25**B. The classification results are compared for different input subsets, with the same criteria as the five subsets in the training classification. The relative tumor class size differs for the various certainty score ranges, the class size increased with the larger range of certainty scores. The average MCC of the Bagged Trees is not significantly different between the train and test results (p-value = 0.390). Whereas the average MCCs of the other three classifiers are significantly different for the train and test set (p-values < 0.05). The Bagged Trees classifier results in the highest MCC for the subset with a certainty score range between 6 and 10. The average MCC of the Bagged Trees is not significantly different from the average MCC of the Fine Gaussian SVM (p-value 0.06). However, both classifiers have significantly higher MCC values than the Subspace KNN and Weighted KNN (p-value < 0.05). The ROC curves in Figure **4-25**B are less smooth than the train ROC curves. The curve of the Fine Gaussian SVM from the test set shows the same pattern as in the training set, with lower sensitivities compared to the other classifiers for specificities <0.6. The Weighted KNN classifiers show lower sensitivities relative to the other classifiers for specificities >0.7.

Table 4-8: MCC of the test set, for four different classifiers and five different datasets with various certainty score ranges

Certainty scores	Sample size (#)	Tumor class (%)	Bagged Trees	Fine Gaussian SVM	Subspace KNN	Weighted KNN
Score 10	285	7%	0.40	0.40	0.27	0.22
Score 9-10	293	9%	0.40	0.41	0.29	0.25
Score 8-10	328	12%	0.44	0.34	0.28	0.23
Score 7-10	366	15%	0.44	0.38	0.34	0.29
Score 6-10	397	18%	0.45	0.41	0.36	0.33

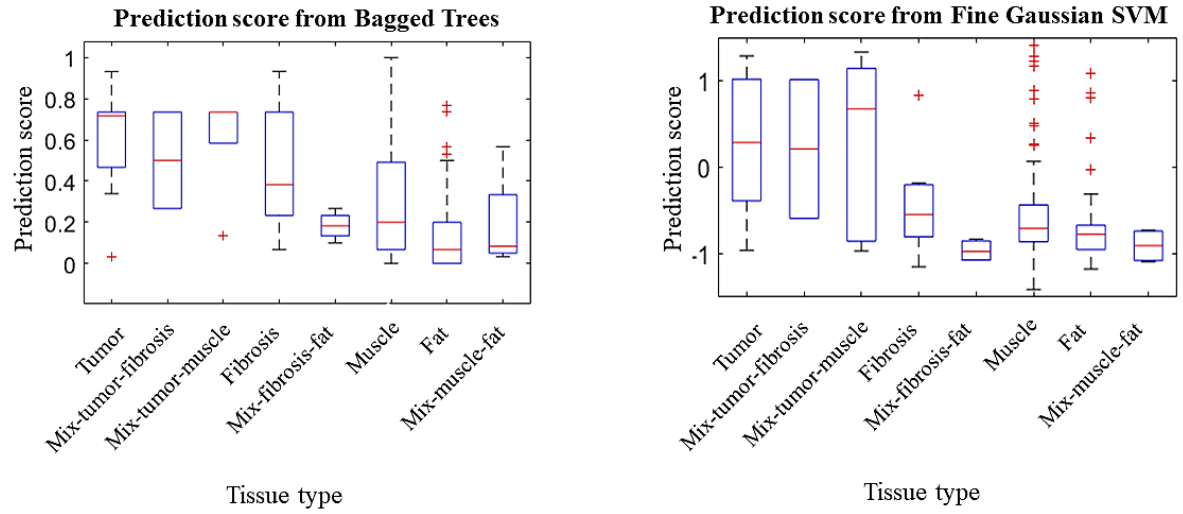
The performance metrics of the Bagged Trees and Fine Gaussian SVM are shown for the train and test set in Table **4-9**. The performance is based on the subsets with a certainty score of 10. The metrics of the Bagged Trees are based on a cutoff value of 0.5 for the prediction scores resulting from the classifier. Thus the samples with a prediction score >0.5 were classified as tumor tissue, and samples with a prediction score of ≤ 0.5 were classified as healthy. The cutoff value for the Fine Gaussian SVM was 0 since the prediction values vary between -1.5 and 1.5. The samples with a prediction value >0 were classified as tumor tissue and samples with a prediction score of ≤0 were classified as healthy. The sensitivity is higher than the specificity in both train and test sets, suggesting that the classifiers are superior in detecting tumors than detecting healthy tissue.

Table 4-9: Classification performance metrics of the Bagged Trees and Fine Gaussian SVM classifiers for the training and test set. Both sets just include samples with a certainty score of 10.

	Bagged Trees		Fine Gaussian SVM	
Performance metric	**Training set**	**Test set**	**Training set**	**Test set**
Sensitivity	0.76	0.87	0.83	0.91
Specificity	0.67	0.74	0.66	0.63
ACC	0.73	0.86	0.75	0.89
AUC	0.76	0.83	0.78	0.81
MCC	0.43	0.40	0.49	0.40

The predictions of the Bagged Trees classifier to the Fine Gaussian SVM for the first subset of the test set are compared with McNemar's test. This statistic test results in a Chi-squared of 0.18 and a p-value of 0.67. Since the p-value is >0.05, the performances of both models are not significantly different according to McNemar's test.

Figure **4-26**A shows the resulting prediction scores (probability of being tumor) of the Bagged Trees classifier per tissue type. It shows that tumor tissue and mixtures of tumor tissue and fibrosis mainly have high prediction scores. Since the prediction scores of tumor and fibrosis are mainly overlapping, it shows that the classifier has difficulties with distinguishing tumor and fibrosis. The mixture of tumor and muscle shows the largest spread in prediction scores, with 50% of the scores in the range of 0.18 and 0.83. The muscle, fat, and mixtures with muscle and fat have generally a lower prediction score than 0.5. Figure **4-26**B shows the prediction scores of the FG-SVM classifier. It shows that tumor tissue and mixtures with tumor tissue have a large variation in prediction scores. Especially the mixture with tumor and muscle have a large variation had has the most overlap with the prediction scores of the other tissue types. The other tissue types have low prediction scores, below 0 except for some outliers.

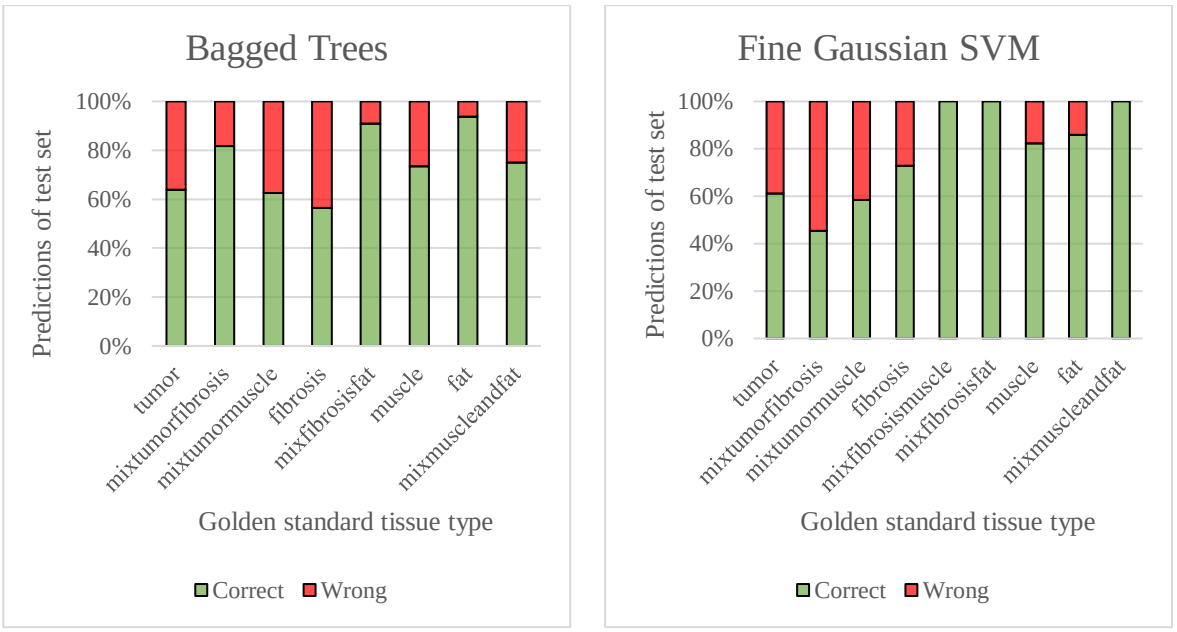


A. B.

Figure 4-26: Prediction scores (probability of being tumor) per tissue type of the test set including just samples with a certainty score of 10. The red horizontal line shows the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles respectively. The whiskers extend to 1.5 times the interquartile range, and the outliers are plotted as the red plus signs. The number of samples varies per tissue type: tumor (12), mix-tumor-fibrosis (2), mix-tumor-muscle (5), fibrosis (15), mix-fibrosis-fat (4),muscle (131), fat (113), and mix-muscle-fat(4) A. Prediction scores of the Bagged Trees classifier range from 0-1. B. Prediction scores of the Fine Gaussian SVM classifier range from -1.5 to 1.5.

Figure **4-27** shows the correct and wrong predictions based on a cutoff value of 0.5 per tissue type. Wrong means that tumor tissue or mixtures with tumor tissue were classified as healthy, or that the remaining tissue types were classified as tumor tissue. Fibrosis is more often wrongly classified as tumor tissue than the other healthy tissue types. However, the percentage of misclassifications of fibrosis is larger for the Bagged Trees than the FG-SVM (43% vs 27%). On the other hand, the Fine Gaussian SVM has a higher misclassification rate for tumor tissue, and mixtures with tumor tissue, than the Bagged Trees. Especially the mixture with tumor and fibrosis is often misclassified by the FG-SVM (55% FG-SVM and 18% Bagged Trees).

Therefore, these results show that the FG-SVM is more suitable to classify healthy tissue correctly as healthy, whereas the Bagged Trees classifier is better in classifying tumor or mixtures with tumor correctly as tumor tissue.



A. B.

Figure 4-27: Misclassification rate per tissue type in the test set with superpixels with a certainty score of 10, when using a cutoff value of 0.5 for the prediction scores from the Bagged Trees (>0.5 = tumor, ≤0.5 = healthy) and a cut-off value of 0 for the Fine Gaussian SVM (>0 = tumor, ≤0 = healthy). A. The Bagged Trees has the highest misclassification rate for fibrosis, resulting in a wrong classification of fibrosis as tumor tissue. Secondly, the tumor and mixture with tumor and muscle have a high misclassification rate. The fat and mixtures with fibrosis and fat are more often correctly classified as healthy than the other healthy tissue types. B. The Fine Gaussian SVM has low misclassification rates for the healthy tissues, but high misclassification rates for tumor tissue and mixtures with tumor tissue. The mixtures with tumor and fibrosis are more often misclassified as healthy than correctly classified as tumor tissue.

Figure **4-28** shows the misclassifications and correct classifications relative to the certainty scores of the tissue type labels. Both figures show that the misclassifications have more often a low certainty score for the tissue type labels than the correct classifications.

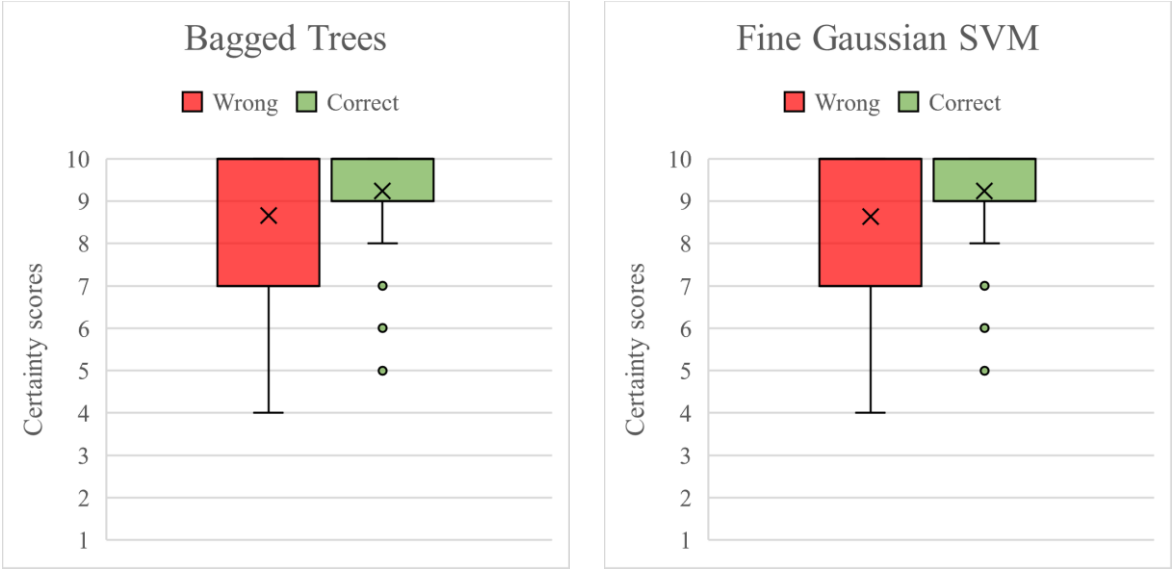


Figure 4-28: Misclassifications and correct classifications in the test set relative to the certainty scores of the tissue type label. The boxplots of both classifiers show that the superpixels with lower certainty scores (<9) are mostly misclassified.

Some examples of the test set classification results with the Bagged Trees and FG-SVM classifiers are shown in Figure **4-29**, Figure **4-30**, and Figure **4-31**. Examples from three different measurements are shown with the original US B-mode and elastography images. Furthermore, the golden standard tissue types and the certainty scores of the labels are shown per superpixel. The golden standard tissue types include mixtures of tissue types as well. The mixtures with tumor are shown as tumor tissue, the mixtures of muscle or fat with fibrosis are shown as fibrosis, and the mixture of muscle and fat is shown as muscle.

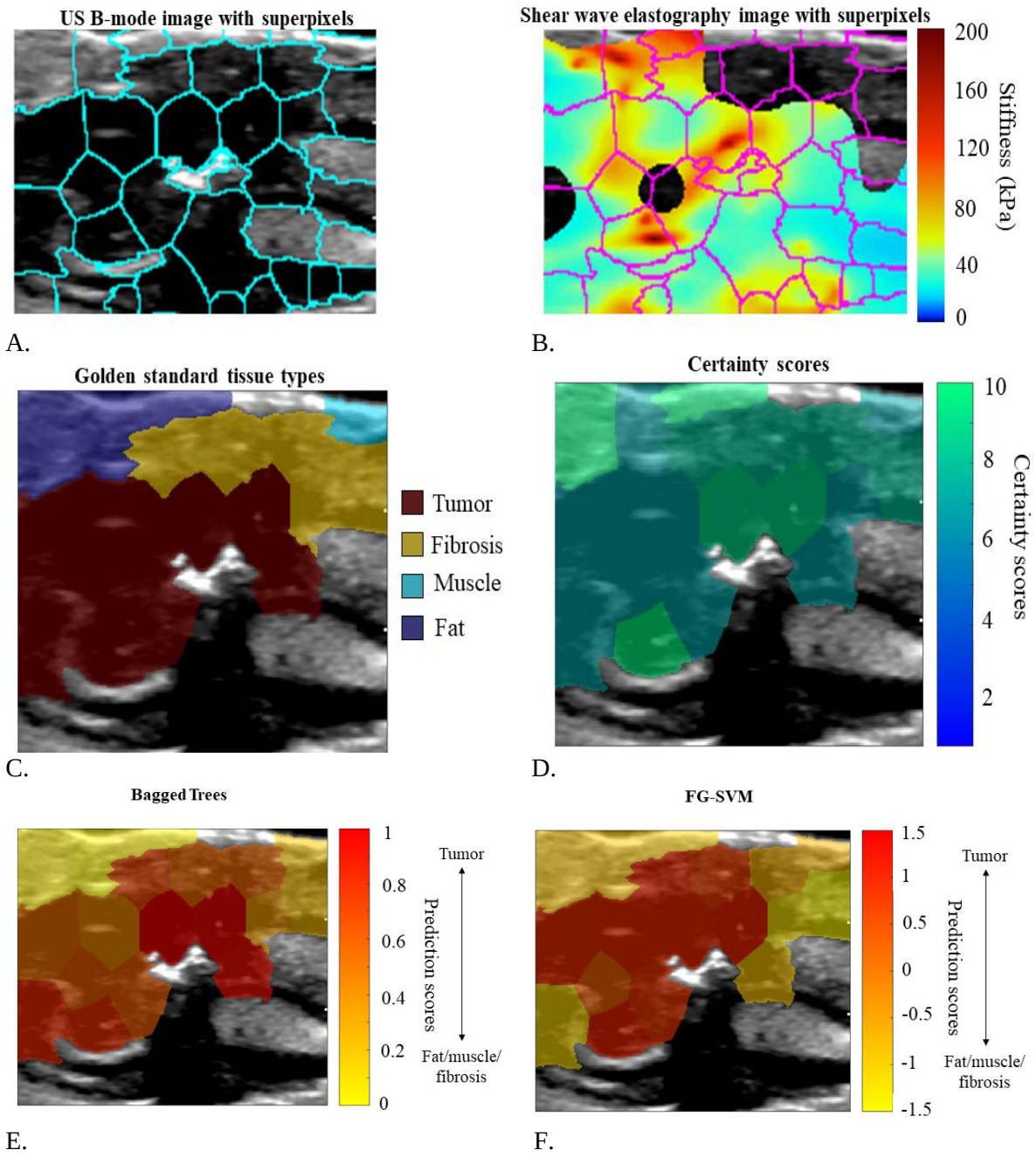


Figure 4-29: Results of Ex-1, measurement 1, one of the average predictions. A. Original ultrasound B-mode image with superpixels. B Shear wave elastography image with the same superpixels as the B-mode image. The superpixels that are not colored, had a confidence score <35% C. Golden standard tissue types based on the correlation with pathology. This image contains mainly tumor tissue, but also fibrosis, an area of fat, and a small area of muscle. D. Certainty scores of the tissue type labels, the scores were mainly between 7 and 8. E. The prediction scores as a result of the Bagged Trees classifier. The high scores are mainly at the location of tumor tissue, and some superpixels in the fibrosis area also have a score of 0.6-0.7. The healthy areas, muscle, and fat have low prediction scores. F. The prediction scores as a result of the FG-SVM classifier. Almost the complete tumor area has high prediction scores, however, the boundary between tumor and fibrosis is missing since some superpixels with fibrosis also have a high prediction score. The fat and muscle have low prediction scores.

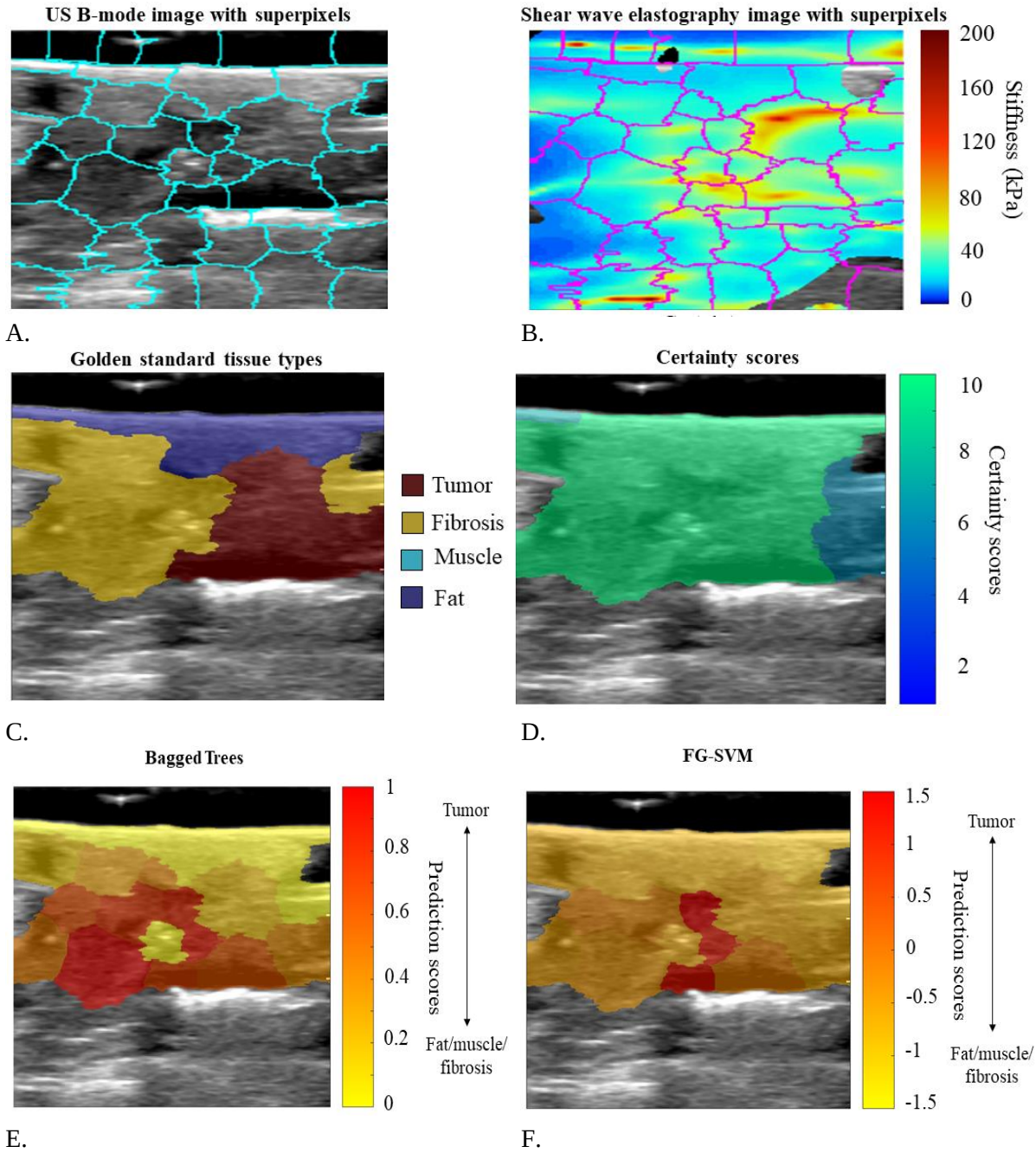


Figure 4-30: Results of Ex-5, measurement 1, one of the worst predictions. A. Original ultrasound B-mode image with superpixels. B Shear wave elastography image with the same superpixels as the B-mode image. C. Golden standard tissue types based on the correlation with pathology. This image contains fibrosis, tumor, and fat. D. Certainty scores of the tissue type labels, the scores were mainly 10, except for three superpixels aside. E. The prediction scores as a result of the Bagged Trees classifier. The high prediction scores are localized both at fibrosis and tumor. However, not the complete tumor area has high prediction scores. The fat area just contains low prediction scores. F. The prediction scores as a result of the FG-SVM classifier. There are only a few superpixels with a high prediction score, and two of them are in the tumor tissue area, and one of them is in the fibrotic area. Most of the tumor has low prediction scores.

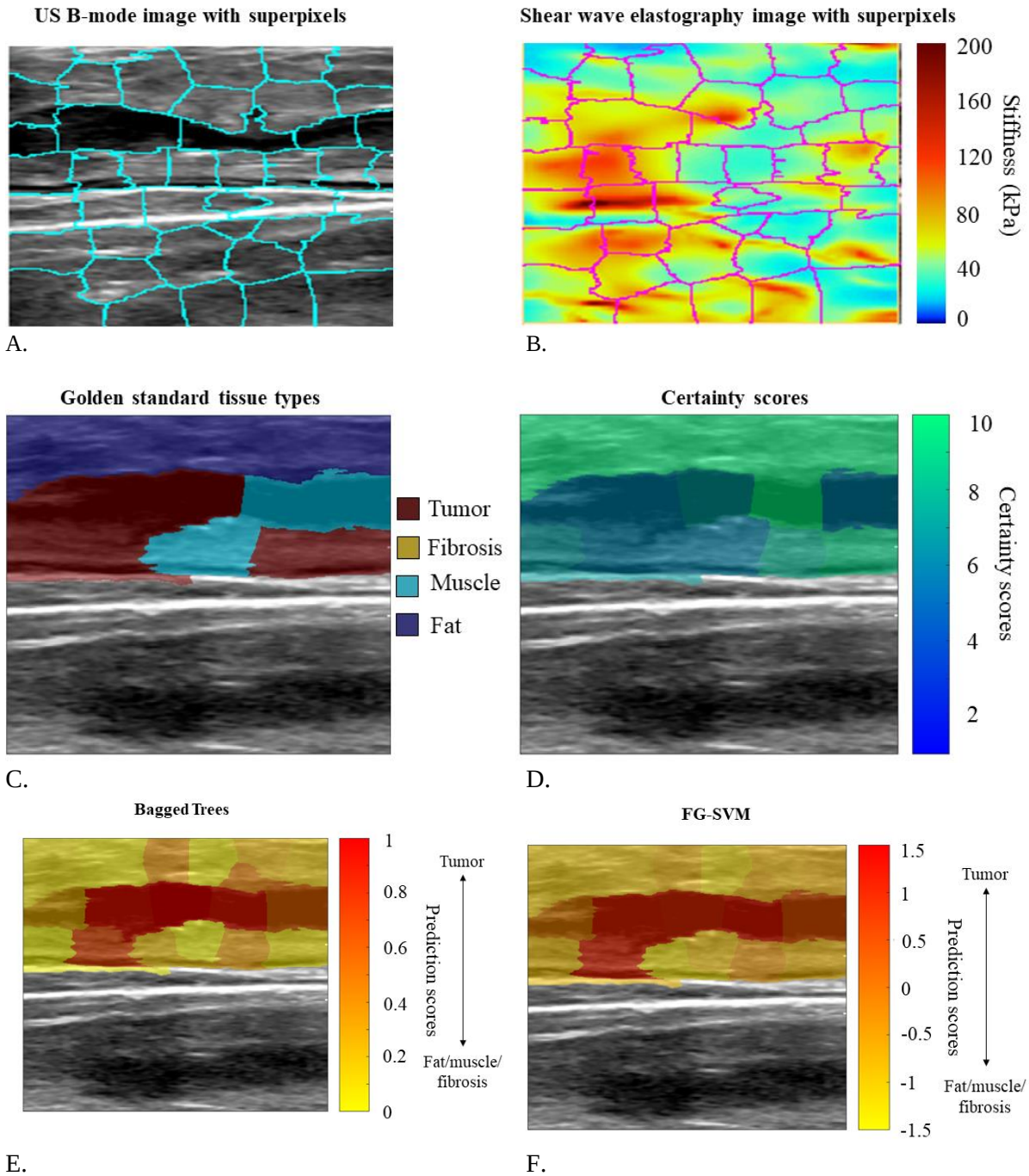


Figure 4-31: Results of Ex-17, measurement 2 A. Original ultrasound B-mode image with superpixels. B Shear wave elastography image with the same superpixels as the B-mode image. C. Golden standard tissue types based on the correlation with pathology. This image contains tumor, muscle, and fat. D. Certainty scores of the tissue type labels, the scores were mainly 10 at the fat area, but lower in the tumor and muscle areas. E. The prediction scores as a result of the Bagged Trees classifier. The high prediction scores are localized both at tumor and muscle. Furthermore, there are also low prediction scores in the tumor areas. The fat area has low prediction scores. F. The prediction scores as a result of the FG-SVM classifier. The prediction score map has almost the same distribution as the prediction scores of the Bagged Trees.

4.5 Discussion

This ex vivo study was performed to investigate whereas ultrasound B-mode and shear wave elastography can be used to distinguish colorectal tumor tissue from healthy colorectal muscle, fat, and fibrosis. The results showed that both ultrasound B-mode and shear wave elastography features are variable for the four tissue types. However, one of the features is not enough to distinguish tumor tissue from the other tissue types. Therefore, feature subset selection was performed which resulted in a combination of nine B-mode and elastography features. Furthermore, the classification results of the training and test sets show that the Bagged Trees and Fine Gaussian SVM classifier have the best performance based on the MCC.

The strengths and limitations of this study will be discussed per subject. Firstly, the correlation of the ultrasound images with the histopathology will be reviewed. Thereafter, the resulting B-mode and elastography features are correlated to literature, and deviating features will be clarified. Subsequently, the nine selected features are compared to the features in the top ten's from the three feature ranking methods. Afterward, the classification results are discussed and compared to the literature. Lastly, the clinical relevance and recommendations are shown.

4.5.1 Pathology correlation

The histopathology coupes were used as the golden standard for the tissue types of the measurement locations. Whereas using the markers on the coupe for correlation with the US images is an improvement compared to other studies with ultrasound elastography, it gave also some challenges. Other studies that used ultrasound elastography for classification between malignant and benign lesions, or the discrimination between tumor stages, used the B-mode image to determine the possible tumor location.[29,32,76,80,94,100] In the case of colorectal tumors, the measurement location was based on ultrasonic appearances according to Hildebrandt et al. or Beynon et al.[79,130] The ultrasonic characteristics of rectal tumors that were used included: hypoechogenicity, interruption of the healthy bowel structure, and a lost tissue layer stratification.[26,27,79,130] However, the current study shows, just as other studies, that tumors can have the same ultrasonic appearance as neoadjuvant therapy effects, such as fibrosis.[26,54,56] Therefore, the tumor location may be wrongly determined when fibrosis is present in the area of the tumor. To solve this problem, the markers on the histopathology coupes can be used to determine the location of the tumor in the B-mode and elastography image. Furthermore, the use of markers is more objective than the ultrasonic interpretation by a radiologist. Another strength of this method that it can be used to correlate multiple regions in one image to different tissue types. Whereas other studies used one region of interest per tumor to estimate the tumor grade, the current method can be used to show the exact location of tumor invasion. This is relevant for the surgical procedure to prevent positive resection margins. However, the correlation of histopathology with ultrasound also gave some challenges.

There were some cases that the ink marks at the pathology coupes were not visible anymore or were spread out. In these cases, the measurement surface colorectal wall was too slippery for a good attachment of the ink. Another reason was that a piece of tissue was ripped or replaced relative to the situation during the ex vivo measurement. When two or three markers in one coupe were lost, the pathology coupe was excluded from the data analysis. In the case that one marker was lost, the other two markers were used for correlation. It was assumed that the pathology coupe deformed with the same stretch in the horizontal direction. Therefore, the distance between two markers could be used to determine the location of the third marker. For example, when two black markers were found, and the distance between them was 7 mm, the distance to the third purple marker should be 7 mm as well. Furthermore, anatomical landmarks that were recognizable in the pathology coupe and the B-mode images were also taken into account to see whether the marker locations are logical. These landmarks included the fat layer, or the typical healthy colorectal tissue layers, or the tumor.

Therefore, the correlation of ultrasound to the pathology is still partly subjective, but less subjective than using only ultrasound characteristics. Furthermore, the pathology coupes that were difficult to interpret, were analyzed by two researchers.

4.5.2 Ultrasound B-mode and elastography features

The study gives insight into quantitative B-mode and elastography features per tissue type (tumor, fibrosis, muscle, and fat). The use of quantitative B-mode features is a different approach than other studies on ultrasound elastography. Other studies use B-mode images for visual assessment of colorectal tumors. [28,31,32,35] The quantitative features show how exactly the different tissue types differ on US B-mode images. Furthermore, the elastography features give new insights into other features that are relevant for tissue type discrimination, rather than the standardly used mean and maximum.[32,35,89,94,100,102] The results show that histogram features for example are also different per tissue type.

The US B-mode features show distinctive features for tumor tissue compared to the other three tissue types. The intensity features have lower values for tumor tissue than for the other tissue types, which is as expected since one of the tumor characteristics is that it appears hypoechoic, resulting in a black or dark grey color on the B-mode image. Although the averaged features of tumor tissue are distinctive from the other tissue types, the interquartile ranges of each feature are overlapping. Therefore, a cutoff value of one feature would always result in misclassifications. Furthermore, the histograms of each tissue type show that the tumor has a deviating peak of 0-0.1 compared to the other tissue types. Therefore, when more than 20% of a superpixel has intensities between 0-0.1, the chance is high that a superpixel is localized at tumor tissue. However, the histogram of the muscle may have higher intensity values between 0 and 0.1 as well, when dividing muscle into the submucosa and the m. propia. Since the submucosa generally appears hyperechoic and the m. propia hypoechoic, the current histogram may be averaging out and is not representative for individual superpixels of the two different tissue layers. When showing different histograms for the m. propia and the submucosa, the histogram of the m. propia may have a similar appearance as tumor tissue since most of the m. propia has low intensities as well. Furthermore, the histograms of tumor, fibrosis, and muscle show overlap between 0.11 and 0.40. Therefore, it is harder to distinguish the tissue types based on these intensity values. Lastly, all B-mode features of fibrosis are closer to the features of tumor tissue than muscle and fat. Therefore, it can be confirmed that fibrosis is harder to distinguish from tumor tissue.

The resulting elastography features show that tumor tissue is stiffer than the healthy colorectal wall and fat tissue, which corresponds to the expectation. However, the stiffness values of fat and the healthy colorectal wall, are stiffer in the current study than other studies that used shear wave elastography. Fan et al. show a mean stiffness between 5.0-5.8 kPa for the normal rectal wall, and maximum stiffness of 8.4-8.8 kPa.[32] Chen et al. showed a mean stiffness of 15.4 kPa for the normal rectal wall.[35] In the current study, the mean stiffness for the healthy colorectal wall is 47 kPa and the maximum stiffness is 82 kPa. Furthermore, Chen et al. showed that perirectal fat has a mean stiffness of 17.6 kPa and peritumoral fat had a mean stiffness of 31.3 kPa. Whereas the current study shows that fat has an average mean stiffness of 63 kPa and an average maximum stiffness of 100 kPa. The differences may be caused by a different measurement location of the healthy rectal wall and fat, the effects of neoadjuvant therapy, or higher tumor grades. Another cause may be that both Chen et al. and Fan et al. used another ultrasound system from a different brand to retrieve the ultrasound B-mode and elastography images than the Philips EPIQ7.
Firstly, Fan et al. retrieved the stiffness values of the normal rectal wall from another location than the tumor, whereas the current study used the stiffness values of the normal colorectal wall within the area of the tumor.[32] Chen et al. also retrieved stiffness values of the normal rectal wall and perirectal fat at a distance of 2-5 cm from the tumor.[35] Furthermore, Chen et al. compared the mean stiffness of the distant perirectal fat to the peritumoral fat and showed that the mean stiffness of the peritumoral fat was significantly higher than the mean stiffness of distant perirectal fat in the case of T3 tumors. Although

this comparison has not been made for the healthy rectal wall, the rectal wall next to the tumor is presumably stiffer than the rectal wall at a distance. This difference in stiffness can be caused by the effects of tumor tissue on the environment.
Secondly, both Fan et al. and Chen et al. excluded patients who received neoadjuvant therapy, whereas 45% of the specimens in the current study were retrieved from patients who received neoadjuvant therapy. [32,35] Rafaelsen et al. showed that the shear wave speed of perirectal fat increased after two weeks and six weeks since the start of chemoradiotherapy.[34] Since shear wave speed and stiffness are positively correlated, it can be assumed that the stiffness increased as well after two and six weeks of CRT. This clarifies why the stiffness values in the current study are higher than other studies that excluded patients who received neoadjuvant CRT. The current study did not compare the stiffness of the healthy colorectal wall and fat between patients who received neoadjuvant therapy and those who did not. However, further data analysis on the data retrieved in this study can be performed to investigate this.
Thirdly, Chen et al. showed that peritumoral fat has a significantly higher stiffness for T3 tumors (mean stiffness 86.5 kPa) than T2 tumors (24.5 kPa).[26] The stiffness for peritumoral fat around T4 tumors is not shown. However, the stiffness it is very likely that the stiffness of peritumoral fat is higher for T4 tumors than T3 tumors since T4 tumors perforate the peritumoral fat and T3 tumors invade the peritumoral fat. Since 24% of the specimens in the current study are from T4 tumors, whereas the other studies did not contain T4 tumors, the higher stiffness of healthy fat can be explained by the higher number of T4 tumors.
In summary, the healthy muscle and fat tissues have a higher stiffness than reported in other studies, but the higher values can be explained through the measurement location close to the tumor instead of at a distance from the tumor, the effects of neoadjuvant therapy, and the higher tumor grades.

Furthermore, the elastography features of fibrosis have overlap with the features of tumor tissue, and fibrosis is generally stiffer than tumor tissue. It is expected that tumor tissue and fibrosis are harder to distinguish each other than fat and muscle since this is also challenging for the surgeon to distinguish. Since the current study was the first that analyzed the stiffness of fibrosis surrounding colorectal tumors, the stiffness values of fibrosis cannot be compared to the literature. The mean stiffness of the colorectal tumors is within the same range compared to the results of Chen et al. They reported a mean stiffness of 119 kPa for rectal tumors, whereas the current study shows a mean stiffness of 90 kPa. A lower stiffness of tumor tissue may be a result of neoadjuvant CRT. Namely, Rafaelsen et al. demonstrated that the shear wave speed of tumors decreased after two and six weeks from the start of CRT. Therefore, the mean stiffness of tumors from patients who received CRT may be lower compared to patients who did not.

The mean and maximum stiffness a higher interquartile range for tumor tissue compared to the other tissue types. The large range of mean stiffness values of tumor tissue can be explained by the multiple tumor grades that were included in the study. For instance, both Chen et al. and Fan et al. show that a higher tumor grade correlates to higher mean and maximum stiffness values. Fan et al. reported maximum stiffness values of 76.7 kPa and 104.8 kPa for T2 and T3/T4 tumors respectively. Chen et al showed that mean stiffness values of 87.3 kPa and 157.3 kPa for T2 and T3 tumors respectively. The current study did not investigate the stiffness values per tumor grade since it was out of the scope of the study. Furthermore, the groups per tumor grade are quite small to conclude from. However, when using ultrasound elastography in the next step to estimate the resection margin, the tumor grade may be interesting since it influences the local invasion. Thereby, the discrimination between T3 and T4 tumors is especially interesting, since T4 tumors have a higher chance of a positive resection margin due to invasion into the serosa or adjacent structures. The previous studies did not report discriminative stiffness values between T3 and T4. Therefore, further data acquisition and analysis are needed to investigate the differences between T3 and T4 tumors.

Another remarkable result from the elastography histograms is that each tissue type has a deviating peak between 190 and 200 kPa. The high percentage of 190-200 kPa in the histogram of fibrosis is mainly caused by measurements from ex-21. Furthermore, the high percentage of 190-200 kPa in the histogram of tumor tissue is mainly caused by measurements from ex-22. The pathology coupes, B-mode, and shear wave elastography images of these specimens are shown in Figure **4-17**D and Figure **4-18**B. Ex-21 consists almost completely of fibrosis, whereas other specimens mostly have a small area of fibrosis. Furthermore, ex-22 has a much larger tumor than other specimens with tumor tissue. The large fibrosis area and large tumor result in a larger distance between the contact surface and the lumen for these specimens relative to other specimens. Therefore, relatively more superpixels with pathology were retrieved from these measurements. Since the superpixels from these elastography images also have high stiffness values, the measurements from only these two specimens resulted in high percentage values for the range of 190-200 kPa. Furthermore, the high percentages may be caused by the confidence level of the shear wave measurements. For the first fifteen ex vivo measurements, the level was set at 35%, therefore, stiffness values were missing at locations where the transducer could not retrieve enough shear waves. For the last fifteen measurements, the level was turned down to 0% since it was unknown whether the confidence maps were falsely low through the ex vivo setting. When the confidence level of 35% was used, it resulted in superpixels that were not completely filled with stiffness values. For example, some superpixels contained only 44 pixels with a stiffness value of 1251 in total, whereas the average number of pixels with a stiffness value is 2747 from the 3300 pixels in total. Since the percentages of the histograms are calculated by dividing the number of pixels within a certain stiffness range through the total number of pixels with a stiffness value in that superpixel, the influence of pixels with high stiffness values is higher when the total number of pixels in a superpixel is relatively low. On the other hand, when the confidence level was turned down to 0%, the pixels had more often a high stiffness value than before. The locations of superpixels with a histogram value >40% for the range of 190-200 kPa were compared to the confidence map, as shown in Figure **4-32**. Most of these superpixels were located in a red or yellow region. The reasons for a red or yellow color can be caused by the deterioration of shear waves. Therefore, several actions were taken to reduce this deterioration: the US transducer was kept stable during the elastography measurement, a thick layer of ultrasound gel (20 ml) was used to obtain an optimal contact surface between the transducer and the specimen, and a surgical pad was placed beneath the specimen to avoid deteriorating effects from the metal table. Another option may be to cast the specimen in gelatin to simulate a tissue-like environment. However, this is not possible in the current measurement protocol, where the ex vivo measurement is performed on freshly excised specimens within a limited amount of time. The influence of the confidence map on the elastography features can be investigated by assigning a confidence label (red, yellow, or green) to each superpixel based on the retrieved confidence maps. Subsequently, the superpixels with a confidence color of red can be excluded from the classification analysis, just as done with superpixels with a low certainty score for labeling.

Since both B-mode and SWE features show overlap for tumor, fibrosis, and muscle, it can be concluded that cutoff values for the stiffness and intensity would not be sufficient to discriminate tumor from healthy tissue and fibrosis.

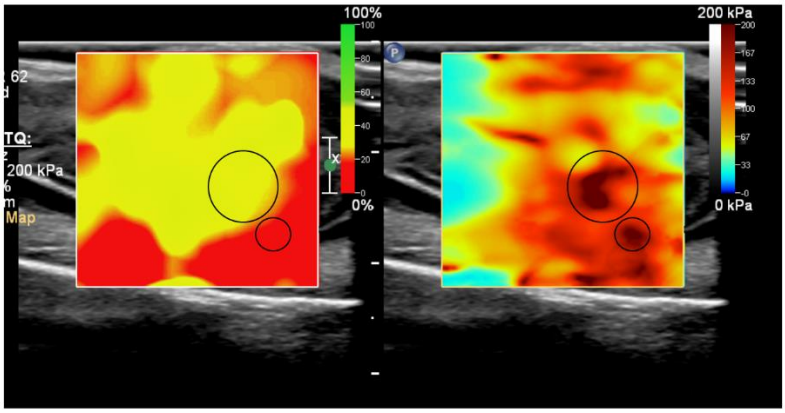


Figure 4-32: The confidence map and shear elastography map of ex_26 with black circles that surround two areas with stiffness values in the range of 190-200 kPa. The largest area corresponds to a yellow area in the confidence map, the smallest area corresponds to a red area.

4.5.3 Feature selection

The feature subset selection resulted in a combination of five B-mode and four shear wave elastography features. Since sequential forward feature selection aims to select the best feature set with the lowest misclassification rate, and 15 B-mode and 25 elastography features were used as input, it shows that a combination of the two techniques is better instead of using features of one of the techniques. This is also shown in other studies that compared a combination of B-mode and elastography features to only B-mode.[28,29,32,33]

The features that were selected through the subset feature selection are corresponding with the features that show the largest difference when assessing the boxplots and histograms. Furthermore, five of the nine selected features were also ranked in the top ten by at least one of the three feature ranking methods. The first feature is the mean intensity of the B-mode image, which has also the highest rank for mRMR and the Chi-square test. Since the hypoechoic appearance of tumor tissue is characteristic on ultrasound, this is a logically selected feature. Furthermore, in total six histogram features were selected, of which three B-mode and three elastography features. This shows that the distribution of stiffness values in a superpixel also contributes to the discrimination of tumor tissue from the other three tissue types. The nine selected features were used as input for the classification models.

4.5.4 Classification

The classification results show that the Bagged Trees and Fine Gaussian SVM classifier are the best classifiers to discriminate tumor tissue from the other three tissue types. Both classifiers have an MCC of 0.40 for the test set, but different sensitivities (0.87 and 0.91) and specificities (0.74 and 0.63). The MCC of the FG-SVM is variable for the five datasets with different certainty scores for the labeling, whereas the MCC of the Bagged Threes is almost the same for each dataset. When assessing the sensitivity of both classifiers, the sensitivity increases for the test set compared to the training set. Furthermore, the sensitivities of the FG-SVM classifier are higher than the Bagged Trees for the training set and the test set. A higher sensitivity is desirable in the context of preventing positive resection margins. It means that the classifier is better in detecting tumors than detecting healthy tissue, thus the chance of missed tumor (false negatives) is lower. However, the disadvantage is that healthy tissue can be classified as tumor tissue (false positives). When redundant healthy tissue is resected, the chance of complication increases.

The classification results were determined for multiple datasets since the test set had only 19 superpixels with tumor tissue that had a certainty score of 10, which was 7% of the total number of superpixels with a 10 score in the test set. Although it is desirable to assess the performance based on samples with a certain golden standard label, a too small class size is not representative for unseen data. Generally, the class sizes should be equal between the train and test set. Since the train-test distribution was just based

on the specimen numbers, the class size was not taken into account. The small class size was partly caused by a different number of measurements since the test set included coincidentally more specimens that had one or two measurements instead of three measurements relative to the training set. As a result, the test set contained 10 measurements from 6 specimens compared to 55 measurements from 23 specimens in the training set. Thus the actual test set is 15% of the total dataset instead of 20%, which was the ratio of the train-test distribution. Furthermore, the majority of the test measurements were performed at locations that contain both tumor and fibrosis. When both tissue types are present, the certainty score of the labels can be lower since the correlation of pathology to the US images is more challenging since fibrosis and tumor are harder to distinguish than tumor and fat or muscle. This explains the reduction of the tumor class size for the datasets with certainty scores of 9 and 10 relative to the lower certainty scores. The train and test set distribution would be perfect when the positive class size is the same in both datasets. Furthermore, the patient characteristics should be equal, such as the tumor grade distribution, and the number of patients who received neoadjuvant therapy. When all these aspects are taken into account, it is not possible to distribute the train and test set automatically, certainly with only 30 specimens included. Therefore, more specimens need to be included in further analyses to increase the chance of a more equally divided train and test set.

To get more insight into the performance of the classifier, the predictions on the complete test set are compared to the certainty scores and the ground truth tissue types. The results showed that the superpixels with a certainty score <9 were almost always misclassified by both classifiers. This can be caused by wrong pathology labels that were used as the golden standard since the superpixels with a lower certainty score have a higher probability to be labeled with a wrong pathology label. This observation emphasizes the need for reliable pathology labels. Firstly, the reliability of the pathology labels depends on the accuracy of the delineations in the pathology coupes. Although the pathology coupes are the golden standard, and surgeons rely on this, there is a high chance that the pathologist does not exactly delineate each cell with tumor or fibrosis separately. Since the delineations of tumor and fibrosis are region-based instead of cell-based, it may be that some fibrotic cells are within an area that is delineated as tumor tissue. This could of course decreases the accuracy of golden standard labels. Furthermore, the correlation of the pathology to the ultrasound images was done by one researcher. Thus, the reliability of the pathology labels and certainty scores is strongly dependent on the quality of labeling from this researcher. To increase the validity of the golden standard labels, another independent researcher need to assign golden standard labels and certainty scores to the superpixels as well. Thereafter, the inter-observer concordance rate can be determined. The higher the concordance rate, the more reliable the golden standard labels are.

Furthermore, the misclassification rate was variable per tissue type and per classifier. Whereas the Bagged Trees has higher misclassification rates for fibrosis than other healthy tissue types, the FG-SVM has higher misclassification rates for tumor tissue and mixtures with tumor tissue. However, the misclassifications for muscle and fat were lower for the FG-SVM than the Bagged Trees, therefore, the SVM is better in detecting truly healthy muscle and fat. These results are in contrast to the high sensitivity of the FG-SVM (0.91) and low specificity (0.63) of the test set. This difference is caused by another cutoff value for the prediction scores. The sensitivity and specificity were calculated at the point on the ROC curve where the MCC was the highest.

The performance of the current study cannot be compared to other ultrasound elastography studies since it is the first study that discriminated colorectal tumor tissue from fibrosis, the healthy colorectal wall, and fat.

4.5.5 Clinical relevance

The ultrasound B-mode images can help the surgeon to see whether the colon or rectum has a healthy appearance. However, when a tumor is present, or fibrosis, or a combination of both tissue types, it gets harder to discriminate tumor from the healthy tissues. In these cases, the classification algorithm can be used to automatically show where tumor tissue is present and where not. Through showing the prediction scores as a color map, the distinction can be made between areas that have a high probability of being tumor tissue, areas that have a high probability of being healthy, and areas that are doubtful. However, the main clinical purpose was to prevent positive resection margins. Thus when this technique can distinguish tumor tissue from other tissue types, it is also important that the distance to the tumor can be estimated. Therefore, another ex vivo study was performed to estimate the resection margin with the use of ultrasound B-mode and SWE.

4.5.6 Conclusion and recommendations

The goal of this ex vivo study was to investigate to what extend US B-mode and SWE can be used to distinguish colorectal tumor tissue from healthy colorectal muscle, fat, and fibrosis. The study gives a broad overview of which B-mode features and shear wave elastography features are discriminative per tissue type. It shows that a single feature is not enough to distinguish tumor tissue from the other tissue types. Furthermore, it shows that a combination of B-mode and SWE features are most suitable for the classification problem in this study. The tissue type classification of the test set with the Bagged Trees classifier shows that colorectal tumor tissue can be distinguished from fat, muscle, and fibrosis with an MCC of 0.40, an AUC of 0.83, an accuracy of 0.86, sensitivity of 0.87, and a specificity of 0.74. Furthermore, the tissue type classification of the test set with the FG-SVM shows that colorectal tumor tissue can be distinguished from fat, muscle, and fibrosis with an MCC of 0.40, an AUC of 0.81, an accuracy of 0.89, sensitivity of 0.91, and a specificity of 0.63. Furthermore, the classification results show that the discrimination of tumor tissue while preserving healthy tissues is still challenging. However, there are multiple points of improvement that can be done to improve the classification performance.

The recommendations to improve the classification algorithm are as follows:

Firstly, the labeling of the superpixels with the tissue type and certainty scores need to be performed by a second researcher to increase the reliability of the golden standard. Secondly, the influence of the confidence maps on the shear wave elastography measurements needs to be investigated. However, more data is needed to compare classification results with just superpixels from green areas relative to superpixels from red and yellow areas. Thirdly, the train and test sets need to be similar in patient characteristics, tumor characteristics, and tissue type class sizes. Fourthly, it may be investigated whether adding a second classifier would increase the total performance. Since muscle and fat are more likely to each other than fibrosis, it may be better to first classify tumor + fibrosis vs muscle + fat, and secondly to classify tumor vs fibrosis. In that case, the feature set can be better adapted to the separate classifications. To realize this, the feature subset selection needs to be done for the two classification problems. Furthermore, it needs to be investigated which classifiers are most suitable per classification problem. Another point of improvement is that other features may be useful such as the stiffness ratio of tumor tissue relative to the fat or muscle. The stiffness ratio has been used by Bae et al. to detect metastatic axillary lymph nodes.[101] Finally, since achieving negative resection margins is mainly challenging in patients who received neoadjuvant therapy, it needs to be investigated how this classifier specifically works for these patients.

The following chapter will show another ex vivo study that is performed to investigate the combination of ultrasound B-mode and shear wave elastography with DRS in the estimation of the resection margin. The resulting prediction scores (probability of being tumor) of the current ex vivo study are used as features in the following ex vivo study.

5 Resection margin estimation

5.1 Introduction

Negative resection margins are essential to achieve an adequate oncologic resection of colorectal tumors. The main problem in colorectal surgery is a positive circumferential resection margin. Achieving a negative CRM (>1 mm) is mainly challenging in patients with locally advanced rectal cancer because they received neoadjuvant radiotherapy. Therefore, intraoperative techniques may be helpful to guide the surgeon. DRS can distinguish colorectal tumors from healthy tissue.[4,10,21,22] Furthermore, the previous ex vivo study showed that ultrasound B-mode and elastography can be used as well to distinguish colorectal tumor tissue from healthy tissue and fibrosis. Each technique has its added value and limitations. DRS shows tissue properties based on characteristic reflectance spectra, however, it is mainly helpful to assess superficial tissues and does not directly show different tissue layers within a measurement volume. B-mode images clearly show the tissue layer structure, and tumors are clearly distinctive from the healthy colorectal wall and fat, but not directly from fibrosis. Lastly, shear wave elastography shows the stiffness of tissues and improves the classification of tumor tissue versus other tissue types than B-mode alone. Although the three techniques are shown to be useful in tissue type classification, it has not been studied yet whether these three techniques can be used to estimate the resection margin. Therefore, the current ex vivo study is performed. The purpose of this study is to investigate whether DRS, US B-mode, and shear wave elastography can be used to estimate the resection margin in ex vivo specimens. Furthermore, it will be investigated whether a combination of these techniques is beneficial in the estimation of the circumferential resection margin compared to one of the two techniques.

Although a classification analysis may be more suitable to show the surgeon whether the resection margin is positive or negative, for now, it is chosen to perform a regression analysis since limited data is available of positive resection margins. Therefore, the current study can be used to give a first idea of whether it is possible to estimate the resection margin based on the three techniques. Furthermore, a methodology will be shown on how the three techniques can be combined. It will also give insight into which feature set results in the best estimation with the lowest error.

The DRS, B-mode, and elastography measurements retrieved in the previous ex vivo study are used for the current study as well. The three techniques can be combined using the marker locations. The nine B-mode and elastography features that were selected with the feature subset selection are used in this study. Furthermore, the prediction scores resulting from the Bagged Trees and Fine Gaussian SVM are used as features. The preprocessing of the DRS data and feature extraction from the reflectance spectra is described. Subsequently, the correlation of the measurement data with the resection margins on the histopathology will be shown. Finally, different feature sets will be used as input for regression analysis to estimate the resection margin in colorectal specimens.

5.2 Data acquisition

The data acquisition of the DRS, ultrasound B-mode images and shear wave elastography images has been described in Section **4.2.2.** The resulting DRS data consisted of six spectra from six different fibers per measurement location. The B-mode images were retrieved from the eL18-4 US transducer, just as the shear wave elastography images. Furthermore, the ex vivo measurements resulted in three DRS measurements per US B-mode and elastography image. All ex vivo measurement locations where tumor tissue was present within one centimeter from the resection plane, were included in this study. The resection margin was based on the histopathological coupes.

5.3 Data analysis

A regression analysis was performed to estimate the resection margin based on the three techniques. The steps of the data analysis are shown in Figure **5-1**. The superpixels from the US B-mode and SWE images that were retrieved in the previous ex vivo study were used for the current study as well. The data analysis starts with describing the combination of these superpixels with the DRS measurements. Thereafter, the correlation with the histopathology is described. Subsequently, the B-mode and SWE features used in this ex vivo study will be explained. Furthermore, the pre-processing and feature extraction of the DRS spectra is shown. Finally, the feature subset selection and regression analysis are shown.

Figure 5-1: Overview of the data analysis. Each step will be explained in one of the following subsections.

5.3.1 Combining DRS, US B-mode, and elastography

In the ex vivo study of Chapter **5** is described that the B-mode and SWE images were divided into superpixels for feature extraction. The same superpixels were used in the current ex vivo study. Figure **5-2** shows an example of a B-mode image with superpixels and three vertical lines that correspond to the three DRS measurement locations. Per DRS measurement, three superpixels along the vertical line were selected until a maximum depth of 1 cm. When only two superpixels were fitting within 1 cm, the second superpixel was used as the third superpixel too, since each sample need the same number of features for the regression analysis. Figure **5-2** shows an example of which superpixels were selected per DRS measurement. It shows that 1a, 2a, and 3a correspond to the left DRS measurement, 1b, 2b, and 3b to the central DRS measurement, and 1c, 2c, and 3c to the right DRS measurement. The maximum depth of 1 cm was used because the maximum measurement depth of the DRS probe was ~8 mm. The depth of 1 cm can be easily recognized in the US B-mode image, through the white centimeter lines that are standardly shown in each B-mode image. Furthermore, a depth of 1 cm is sufficient for a clinically relevant estimation of the resection margin since a negative resection margin is defined as >1 mm distance to the tumor.

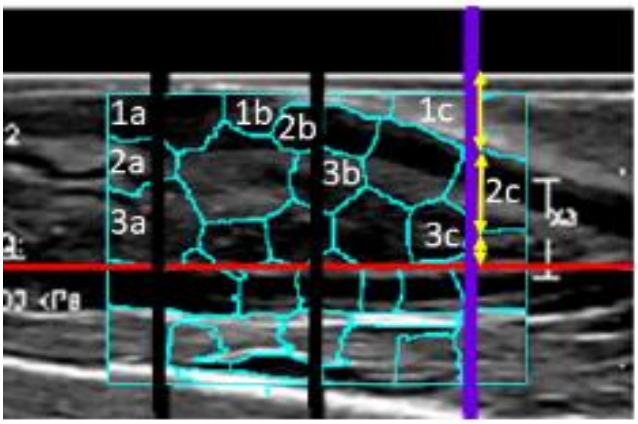


Figure 5-2: Ultrasound B-mode image with superpixel overlay. The vertical black and purple lines show the areas beneath the three DRS measurement locations and the horizontal red line represents 1 cm in depth. The superpixels with a number in it are selected for the regression analysis. The superpixels with 'a' correspond to the DRS measurement that was located at the left black line, etc. The numbers show the order of superpixels from the contact surface. The yellow arrows show the location of the measured superpixel heights that were used as features.

5.3.2 Correlation to the histopathology

The ground truth resection margins were based on the CRMs that were measured in the histopathological coupes at the DRS measurement locations, as shown in Figure **5-3.** The CRM was defined as the distance in millimeters between the measurement surface and the tumor, along the line perpendicular to the measurement surface and lumen.

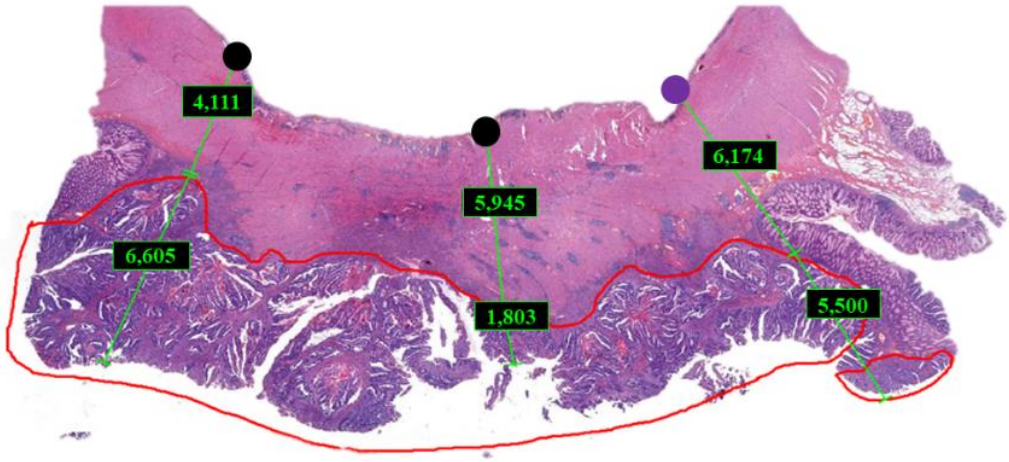


Figure 5-3: Example of circumferential margin measurement in the histopathological coupe. The CRM was measured from the DRS measurement location to the tumor, along the line perpendicular to the measurement surface and lumen. The CRMs in this example are 4.111, 5.945, and 6.174 mm respectively.

5.3.3 US B-mode and elastography

The pre-processing and feature extraction of these B-mode and shear wave elastography images have been described in Section **4.3.** The B-mode and SWE features that were used in the current study included the nine selected features, plus the height per superpixel measured on B-mode images, and the prediction scores (probability of being tumor). The nine selected features consisted of five B-mode features, and four elastography features, as shown in Table **4-6**. The B-mode features included three histogram features, the mean intensity, and the standard deviation of the intensity. The elastography features consisted of three histogram features and the interquartile range of the histogram.

The height of each superpixel was measured to estimate the thicknesses of tissue layers with different intensities on the B-mode image. Therefore, the height in centimeter was measured along the overlap of the superpixel with the vertical measurement line, as shown with the yellow arrows in Figure **5-2**. There is a high probability that different superpixels show different tissue types because the superpixels are based on echogenic intensity. As shown in the previous study, 91% of the total number of superpixels in the total dataset was labeled with only one tissue type. The layer thickness is a valuable feature to estimate the CRM since both are in the same order of magnitude. Furthermore, the tissue layer thickness may be of additional value to the DRS data, since the depth information is missing in DRS.

The prediction scores (probability of being tumor) that were retrieved from the classification in the previous study were used as a feature in the current study as well. Since the predictions of both the Bagged Trees or FG-SVM classifier were the same according to McNemar's test, the prediction scores from both classifiers were used as features.

In summary, the five B-mode features, four SWE features, superpixel height, and two classification scores were used per superpixel. In total, 12 features were retrieved per superpixel. Since three superpixels were selected per DRS measurement, 36 features were selected per measurement location.

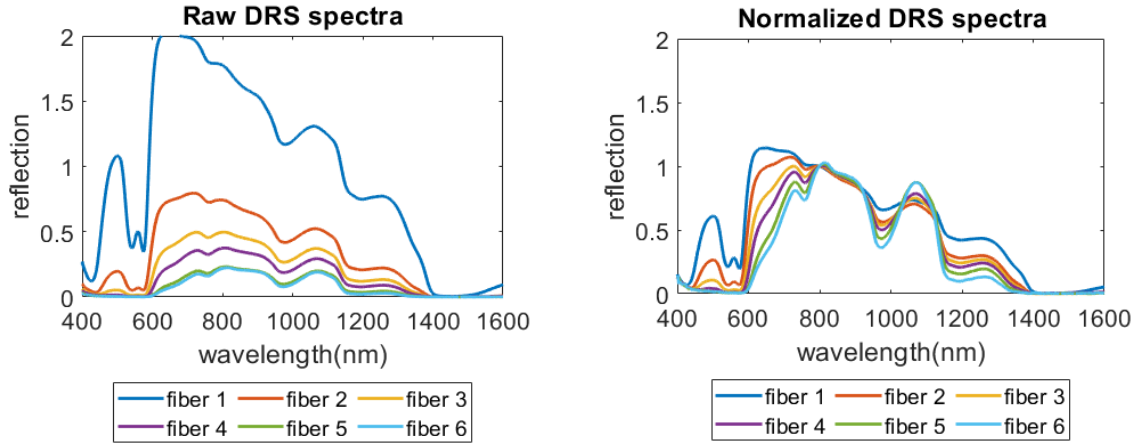
5.3.4 DRS spectra preprocessing

The DRS data included six spectra per measurement, with one spectrum per fiber distance. The six fibers have a fiber distance between the emitting and receiving fiber of 1, 2, 3, 4, 6, and 8 mm respectively. Each spectrum included 1200 wavelengths, in the range of 400-1600 nm.

The DRS spectra were normalized to compensate for undesired scatter effects. These undesired systemic variations are primarily caused by light scattering and differences in the effective path length.[65] A variation in the effective path length can be caused by the measured tissue thickness which deviates over de sample.[22] Light scattering can be caused by Rayleigh and Lorentz-Mie.[65] Both processes describe scattering caused by small particles, bubbles, surface roughness, droplets, cells, fibers, density fluctuations, crystalline defects, and micro-organelles. When the particle sizes are larger than the wavelength, which is generally the case for NIR spectroscopy, Lorentz-Mie scattering is predominant. Lorentz-Mie is anisotropic, dependent on the shapes of the scattering particles, and not strongly wavelength-dependent. Moreover, the coupling between the fibers and the spectrometer might cause a variation in the detection efficiency of the signal.[22] To eliminate these unwanted effects from the desirable absorbance spectra, the spectra can be normalized. The spectra were normalized at 800 because the minimum absorption was assumed to be present at this wavelength.[131] So, if the spectra intensity is higher than 1 at 800 nm, this was assumed to be unwanted scattering. The spectral data (X) is normalized based on the intensity at 800 nm (I_{800nm}), as shown in Equation **5-1**.

$$X_{800nm} = \frac{X}{I_{800nm}} \qquad (5\text{-}1)$$

An example of non-normalized spectra and normalized spectra are shown in Figure **5-4**.



A. B.

Figure 5-4: DRS spectra from fibers with six different fiber distances. Fiber 1 = 1 mm, fiber 2 = 2 mm, fiber 3 = 3 mm, fiber 4 = 4 mm, fiber 5 = 6 mm, fiber 6 = 8 mm. A. Raw DRS spectra from one measurement. B. Normalized DRS spectra at 800 nm.

5.3.5 Feature extraction DRS spectra

Feature extraction from the DRS spectra was performed to avoid overfitting due to high-dimensional data. Three peaks per fiber were extracted as features. The peaks were extracted in the areas of 935 nm, 985 nm, and 1211 nm. These peaks were chosen based on phantom studies and animal studies with DRS that showed that these peak heights were distinctive for different tissue types.[132] The first peak was the maximum value of the spectrum between 923 and 965 nm. Some absorption of lipid is present in this wavelength range.[63] The second peak was the minimum value of the spectrum between 950 and 1010 nm. In this wavelength range, there is some β-carotene and water absorption. The third peak, was extracted by calculating the difference between the maximal value in the range of 1225-1325 nm and the minimum value in the range of 1175 and 1275 nm. Since the main fat absorption peak is present at 1211 nm, and collagen absorption is present in both wavelength ranges, this peak gives insight into the concentration of lipid relative to collagen in the probed tissue.[64] The wavelength ranges of the three peaks are shown in Figure **5-5**. The three peaks were extracted from each of the six spectra individually, resulting in 18 DRS features per measurement.

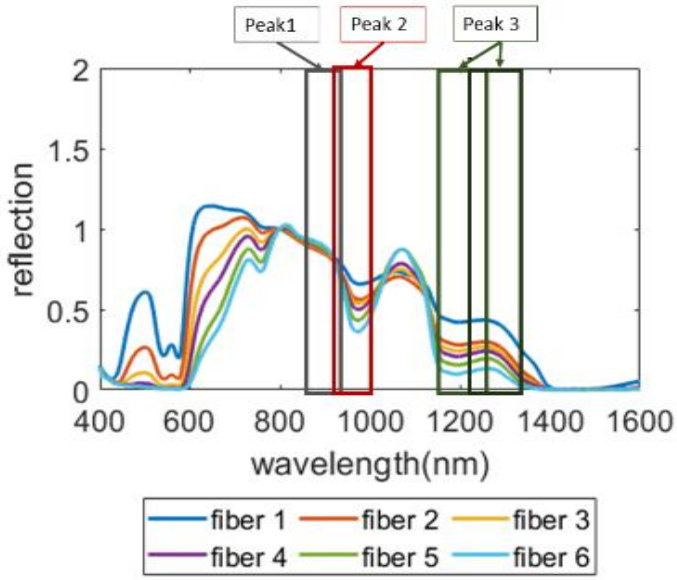


Figure 5-5: DRS spectra with the regions aligned for the peak extraction. Peak 1 is the maximum in the range of 925-965 nm, Peak 2 is the minimum peak in the range of 950-1010 nm, and Peak 3 is the maximum in the range of 1225-1325 nm subtracted with the minimum in the range of 1175-1275 nm. The three peaks are extracted per spectrum.

5.3.6 Feature subset selection

When combining the 36 features of the US B-mode and elastography data with the 18 features of the DRS data, there are 54 features in total. Since only 47 measurements were included, there was a high chance of overfitting. Therefore, feature subset selection needs to be performed. In the current study, the performance of the regression analysis was compared for different feature subsets. An overview of all feature sets is shown in Figure **5-6**. These different subsets were made to answer the following questions:

1. What is the additional value of normalization on the resection margin estimation?
2. What is the performance of the resection margin estimation with only DRS features?
3. What is the performance of the resection margin estimation with only ultrasound B-mode and shear elastography features?
4. What is the performance of the resection margin estimation with DRS features in combination with the ultrasound B-mode and shear wave elastography features?

The first question was answered through the comparison of a feature set with peak features retrieved from the raw DRS spectra to a feature set with peaks of the normalized DRS spectra. Both feature sets

included 18 peak features, since there were three peaks extracted per spectrum, and each measurement contained six spectra from different fibers. The performance of the regression analysis was compared between the two feature set to see whether normalization improves the performance. It is expected that normalization improves the performance in case the DRS measurements suffered from distortion. The peak features from the normalized spectra were used for further analyses.

The performance of the DRS features was tested with different subsets from the DRS features since 18 peak features can result still in overfitting. The DRS peak features were divided into two sets of peak features from three alternating fibers. Thus one set with nine peak features from fiber 1, 3, and 5, and another set with nine peak features from fiber 2, 4, 6. This division was made to include information from different depths in each feature set. The performance of the regression analysis was compared between the two feature sets. There is no substantial difference expected since the measurement volumes of the DRS fibers are mainly overlapping. In the case that one of the feature sets resulted in a better performance of the regression analysis than the other feature set, the best feature set was used for the combination with the ultrasound B-mode and elastography features. In the case that both features sets resulted in the same performance, one of the two sets was used randomly.

The US B-mode and elastography features were also divided into subsets to decrease the number of features and to see what the influence of different features was on the performance of the regression analysis. The nine selected B-mode and elastography features were compared to the classification scores and heights. Each measurement had features of three superpixels, thus 3 x 9 B-mode and elastography features, 3 Bagged Trees prediction scores, 3 Fine Gaussian SVM prediction scores, and 3 heights. Since only the number of B-mode and elastography features was already too much relative to the number of samples in the dataset, only the features of the first superpixel were used to compare the different feature sets. Subsequently, the features of the first superpixel were divided into three sets. One set with the nine selected B-mode and elastography features, a second set with the prediction score of the Bagged Trees classifier and the height of the superpixel, and a third set with the prediction score of the Fine Gaussian SVM classifier and the height of the superpixel as well. The performance of the regression analysis was compared between these three feature sets. Although it would be preferred to use the second or third dataset in the combination with the DRS features, it must be investigated whether the prediction scores are sufficient in combination with the height. It is expected that using the second and third feature set results in a better performance than the first feature set because the prediction score says something about the chance of a superpixel being tumor or not, and the height directly says something about the tissue layer thickness.

Subsequently, another regression analysis was performed using the prediction scores and heights of all three superpixels. In this analysis, a feature set with three Bagged Trees prediction scores and heights was compared to another feature set with three Fine Gaussian SVM prediction scores and heights. Since the features of the other two superpixels add more information about the tissue layers beneath the first layer, it is expected that the estimation of the resection margin improves and therewith the performance of the regression analysis.

Afterward, the Bagged Trees prediction scores and heights of the three superpixels were combined with the peak features of three alternating DRS fibers in one feature set. Furthermore, the SVM prediction scores and heights were combined with the peak features of three alternating DRS fibers in another feature set. The performance of the regression analysis was compared for both feature sets.

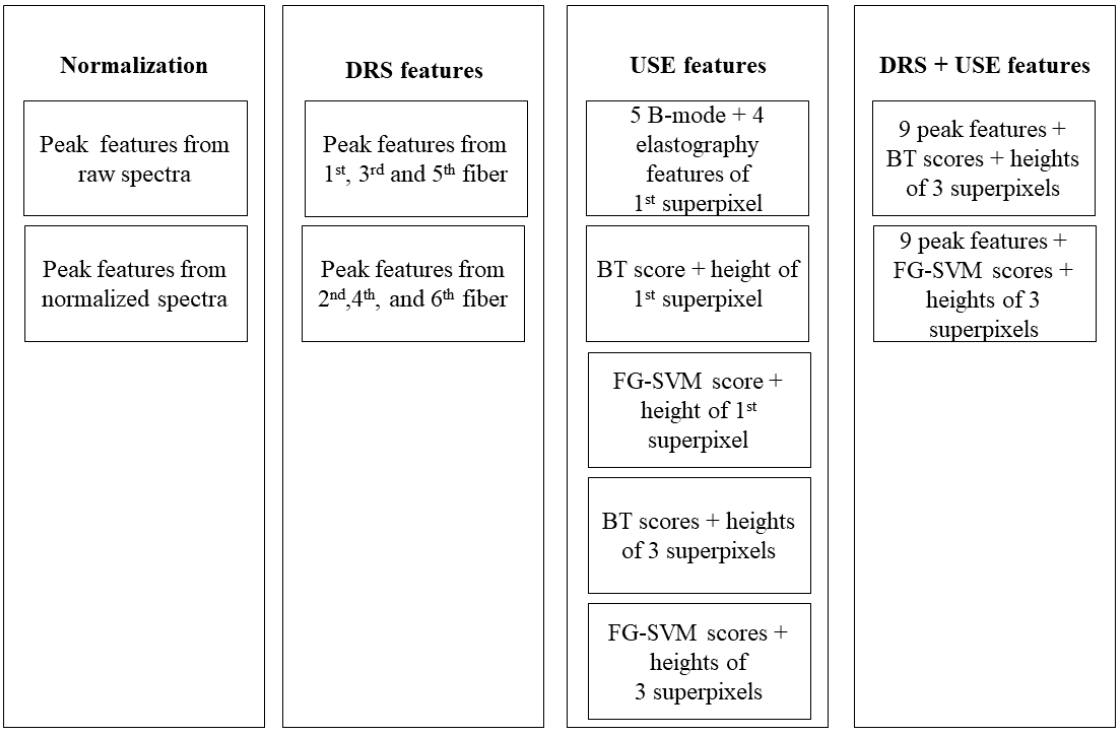


Figure 5-6: Overview of feature subsets. Firstly, two feature sets were compared with DRS features from normalized spectra and raw spectra. Secondly, DRS features of three alternating fibers were compared to features of the other three alternating fibers. Thirdly, the direct B-mode and elastography features were compared to the prediction scores and heights, separately for the Bagged Trees (BT) prediction score, and the Fine Gaussian SVM (FG-SVM) prediction score. Lastly, the prediction scores and heights were combined to the DRS peak features of three alternating fibers.

5.3.7 Regression analysis

Based on some primary experiments, two standard regression models with the highest performance were chosen to compare the different feature subsets. The regression models that were used were Medium Gaussian SVM and Bagged Trees. The validation method was 5-fold cross-validation without taken into account the specimen numbers. The RMSE and MAE of the regression analysis were used as performance measures. Furthermore, the total dataset was used as the training set, since too little data was available to split the data into a test, and training set. Therefore, it needs to be taken into account that the results are preliminary and will only give an idea of whether the combination of these features to estimate the resection margin is promising.

5.4 Results

5.4.1 Study population

In total, the specimens of 14 patients were included in this ex vivo study. These were the patients with tumor tissue within 1 cm from the resection surface. The patient characteristics of the included surgical specimens are shown in Table **5-1**. The sub-study population consisted of three tumors that were localized in the rectum, five in the sigmoid, and six in the colon. Only four patients received neoadjuvant therapy. Three of them received both chemo- and radiotherapy, and one patient received chemotherapy only. The majority of the tumors were locally advanced and had a T3 or T4 grade. The average maximum tumor size was 41 mm and ranged from 15-85 mm. One specimen had a positive CRM (0.5 mm).

Table 5-1: Patient characteristics of the specimens that were included in this study

Characteristics	**Study population (count (%))**
Total nr. of patients	14
Tumor location	
Colon	6 (43)
Sigmoid	5 (36)
Rectum	3 (21)
Neoadjuvant therapy	
Yes	**4 (29)**
Chemotherapy + radiotherapy	3 (75)[3]
Chemotherapy	1 (25)
No	**10 (71)**
Pathological Tumor grade[1]	
T0	0
T1	0
T2	2 (14)
T3	6 (43)
T4	6 (43)
Pathological report	
Max. tumor diameter (mm)[2]	41 (15-85)[4]
CRM (mm)	14 (0,5-45)
CRM positive (<1 mm)	1 (7)

All values are shown as the count with the percentage relative to the total number of patients in the brackets.
CRM = circumferential resection margin according to pathological report
[1]tumor grade according to the pathological report of the resection
[2]maximum diameter of resected tumor according to pathological report
[3]percentage relative to the total number of patients who received neoadjuvant therapy
[4]the range with the minimum and maximum value in the train or test set

5.4.1 Pathology

The regression analysis was performed using 47 measurement samples from 14 surgical specimens. Despite there were 158 measurements with DRS, ultrasound B-mode, and elastography data, measurements were excluded due to four different reasons: unreliable DRS data (36), missing elastography data in the first superpixel (18), no resection margin due to lack of tumor cells at the measurement location (48), resection margin >10 mm (9). The resection margin distribution of the included measurements is shown in Figure **5-7**. Most of the samples have a resection margin between 4 and 6 mm. Furthermore, seven samples have a resection margin of less than 2 mm.

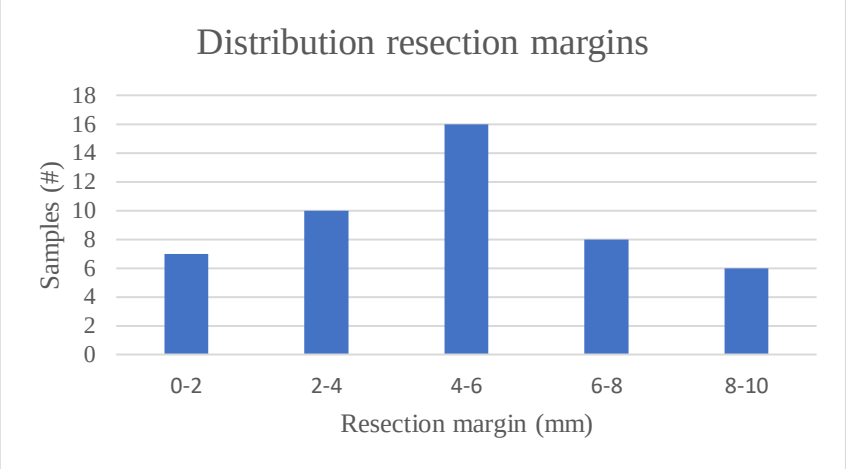


Figure 5-7: Distribution of resection margins for the included samples. Most samples had a resection margin between 4 and 6 mm.

5.4.2 Resulting DRS spectra

Figure **5-8** shows the mean DRS spectra from fiber 1, 3, and 5 per resection margin range. The spectra of fiber 3 and 5 show distortion at wavelengths higher than 1400 nm. The spectra from fiber 1, with a fiber distance of 1 mm, show mainly overlap of all spectra. Only the spectra of margins between 8 and 10 mm show a higher reflectance intensity in the range of 1275-1375 nm. The mean spectra of fiber 3, with a fiber distance of 3 mm, show more differences in absorbance between the resection margin ranges. The minimum between 950 and 1010 nm shows that margins <2 mm have more absorption in this wavelength region than margins ≥2 mm. Furthermore, the maximum in the range of 1275-1375 nm is distinctive per resection margin range, which is the lowest for resection margins between 0-2 mm, and the highest for resection margins between 8-10 mm. The averaged spectra of fiber 5, with a fiber distance of 6 mm, show variations between the margin ranges as well. The spectrum of 0-2 mm is overlapping with the spectrum of 2-4 mm. These two spectra show less reflectance in the range of 1000-1100 nm, and in the range of 1275-1375 nm.

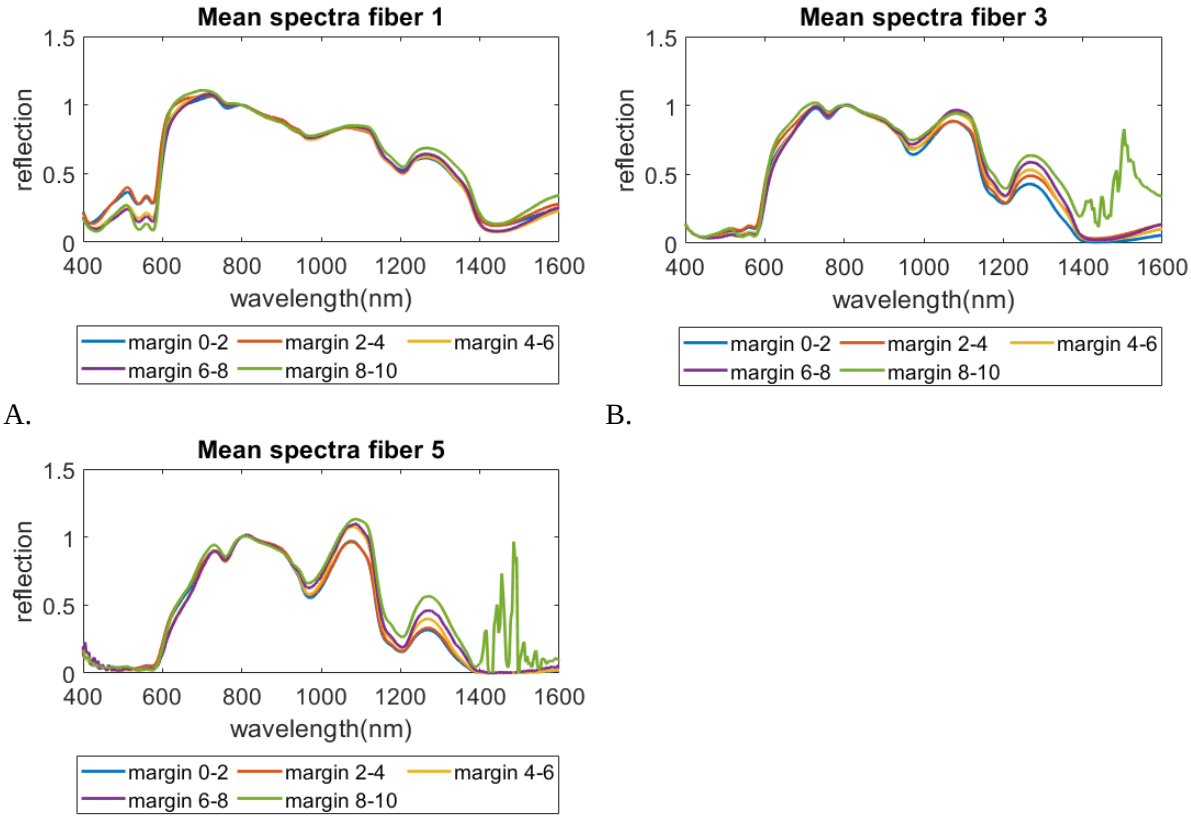


Figure 5-8: Resulting DRS spectra from fiber 1, 3, 5. The spectra are averaged per resection margin range in millimeters. A. Spectra from fiber 1 with a fiber distance of 1 mm. The spectra of the different resection margins are mainly overlapping. Only the spectrum from the 8-10 mm range has a higher peak after the lipid absorption peak of 1211 nm. B. Spectra from fiber 3 with a fiber distance of 3 mm. The minimum between 950 and 1010 nm shows that margins <2 mm have more absorption in this wavelength region than the other margins. Furthermore, the maximum between 1275-1375 nm is distinctive per resection margin range, which is the lowest for resection margins between 0-2 mm, and the highest for resection margins between 8-10 mm. C. Spectra from fiber 5 with a fiber distance of 6 mm. The minimum between 950 and 1010 nm shows that margins <4 mm have more absorption in this wavelength region than the other margins. The maximum between 1000 and 1100 nm shows more absorption for margins <4 mm relative to margins of ≥ 4 mm. Lastly, the maxima between 1275-1375 nm show differences in reflectance intensity per margin range.

5.4.3 Regression analysis

The results of the regression analysis are shown per feature subset and per subject. Therefore, first the results of raw DRS spectra are compared to normalized spectra. Thereafter, the results are shown of different DRS feature sets. Subsequently, the results are presented of different feature sets of the ultrasound B-mode and elastography data. Afterward, the results of the combination of DRS, US B-mode, and shear wave elastography are demonstrated. Lastly, the MAE is shown per resection margin range.

The results in Table **5-2** show the effect of DRS normalization on the regression performance. Both feature sets consisted of 18 features since three peak features from all six fibers were included. The MAE is lower for peak features retrieved from normalized DRS spectra compared to the features from raw DRS spectra. The RMSE of the Bagged Trees is lower for the features from normalized spectra, however, the RMSE is lower for the SVM regression model for the features from the raw spectra. The normalized features are used for further analysis because of the lower MAE.

Table 5-2: Effect of DRS spectra normalization on regression analysis

Feature set	Features (#)	Medium Gaussian SVM		Bagged Trees	
		RMSE (mm)	MAE (mm)	RMSE (mm)	MAE (mm)
Peak features from raw spectra	18	2.64	2.25	2.69	2.22
Peak features from normalized spectra	18	2.74	2.17	2.61	2.15

Table **5-3** shows the regression results of three different feature sets of DRS peak features. The feature sets included the peak features from all six fibers, the 1st, 3rd, and 5th fiber, or the 2nd, 4th, and 6th fiber. The performance of the regression analysis is better for the feature sets that included peak features from alternating fibers compared to the feature set with peaks from all six fibers. Furthermore, using the peak features of fiber 1, 2, and 3 results in the lowest RMSE and MAE for the Bagged Trees regression model. However, the RMSE and MAE of the Medium Gaussian SVM model are lower for the other feature set, with fibers 2, 3, and 6. The feature set with the lowest MAE is chosen for further analysis, which includes the peak features of fibers 1, 3, and 5.

Table 5-3: Regression results when using normalized DRS features of different fibers

Feature set	Features (#)	Medium Gaussian SVM		Bagged Trees	
		RMSE (mm)	MAE (mm)	RMSE (mm)	MAE (mm)
Peak features from all 6 fibers	18	2.74	2.17	2.61	2.15
Peak features from 1st, 3rd, 5th fiber	9	2.73	2.15	2.60	2.10
Peak features from 2nd, 4th, 6th fiber	9	2.66	2.14	2.61	2.16

The feature set with peak features of fiber 1, 3, and 5 is also used for further research to determine the MAE per resection margin range. These results are shown in Figure **5-9**. The MAE of both regression models is the lowest for samples with a resection margin between 4 and 6 mm. The MAE increases for the resection margins range between 0 and 2 mm and between 8 and 10 mm. The MAE of the Medium Gaussian SVM is not significantly different compared to the MAE of the Bagged Trees regression model (p-value>0.05).

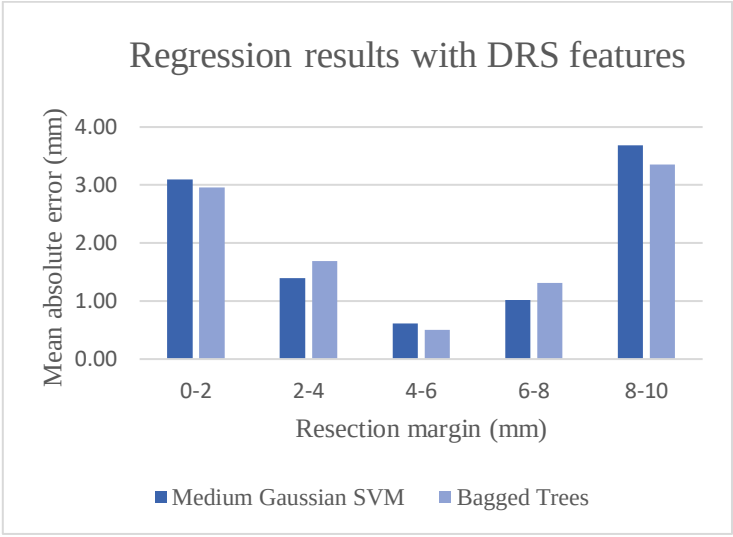


Figure 5-9: Regression results per resection margin range with the peak features of DRS fiber 1, 3, and 5 as feature set.

The results in Table **5-4** show the effect of using different feature sets consisting of ultrasound and elastography features, prediction scores of two classifiers, and the height of a superpixel in mm. The first five rows in Table **6-14** contain the results from features of just one superpixel per measurement. It shows that the Medium Gaussian SVM regression results in a better performance for using the prediction score and height as features instead of the US B-mode and elastography features. On the other hand, the Bagged Trees regression results in a lower MAE for the US B-mode and elastography features. Combining the prediction score and height with the B-mode and elastography features does not improve the regression performance. Lastly, the regression performance improves when using the prediction scores and heights of three superpixels compared to only one superpixel. Furthermore, the prediction scores of the Fine Gaussian SVM classifier in combination with the height result in a lower MAE and RMSE than using the scores of the Bagged Trees classifier.

Table 5-4: Regression analysis results when using different sets of ultrasound B-mode and elastography features.

Feature set	**Features (#)**	**Medium Gaussian SVM**		**Bagged Trees**	
		RMSE (mm)	**MAE (mm)**	**RMSE (mm)**	**MAE (mm)**
US B-mode and elastography features	9	2.75	2.23	2.71	2.18
BT prediction score + height	2	2.60	2.20	2.73	2.28
FG-SVM prediction score + height	2	2.55	2.18	2.64	2.25
US features + BT score + height	11	2.81	2.34	2.76	2.33
US features + FG-SVM score + height	11	2.79	2.33	2.81	2.31
BT score + height from 3 superpixels	6	2.48	2.05	2.46	1.92
FG-SVM score + height from 3 superpixels	6	2.31	1.88	2.45	1.91

BT = Bagged Trees classifier; FG-SVM = Fine Gaussian SVM classifier; height = superpixel height in cm at DRS measurement location. The US features include the nine ultrasound B-mode and elastography features. The first five rows include features of just the first superpixel, and the last two rows include the features of three superpixels per measurement.

The feature set with the Fine Gaussian SVM prediction scores and heights from three superpixels was used to determine the MAE per resection margin range. These results are shown in Figure **5-10**. The MAE of both regression models is the lowest for the resection margin range between 4 and 6 mm. The MAE is increasing towards smaller resection margins, however, the MAE is increasing more for resection margins between 8-10 mm. The MAE values are not significantly different between the two regression models (p-value>0.05).

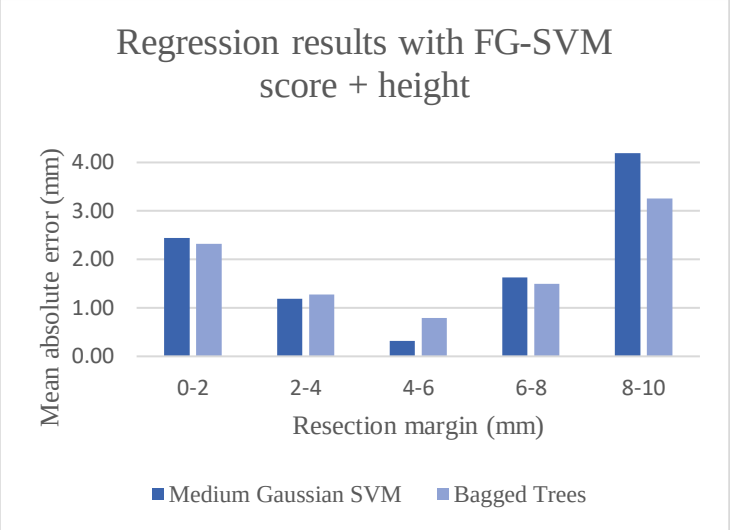


Figure 5-10: Regression results per resection margin range with the prediction scores from the Fine Gaussian SVM classifier and heights of three superpixels per measurement as the feature set.

Table **5-5** shows the regression performance of the combined feature set with the prediction scores, heights, and DRS peak features of fiber 1, 3, and 5. The combination of the Fine Gaussian SVM prediction scores, heights, and DRS features result in a better performance compared to the feature set with the Bagged Trees scores. Furthermore, the performances of the combined feature sets are better than the performances of just using DRS features that were shown in. However, the performances of the combined feature set slightly decreased compared to the performances of just the prediction scores and heights of superpixels that were shown in Table **5-4**. Lastly, the Medium Gaussian SVM regression model has a better performance than the Bagged Trees regression model based on the MAE.

Table 5-5: Effect of using prediction scores, heights, and DRS peak features of multiple superpixels in regression analysis

Feature set	**Features (#)**	**Medium Gaussian SVM**		**Bagged Trees**	
		RMSE (mm)	**MAE (mm)**	**RMSE (mm)**	**MAE (mm)**
BT score + height+ DRS features	15	2.48	2.08	2.69	2.13
FG-SVM score + height + DRS features	15	2.58	1.96	2.09	2.35

BT = Bagged Trees classifier; FG-SVM = Fine Gaussian SVM classifier; height = superpixel height in mm at DRS measurement location. The scores and heights are included from three superpixels per DRS measurement. The DRS features are the peak features of the 1st, 3rd, and 5th fibers.

Thirdly, the feature set with DRS features fiber 1, 3, and 5 in combination with the prediction scores and heights of three superpixels is used to show the MAE per resection margin range. These results are shown in Figure **5-11.** The MAE is not significantly different between the two regression models (p-value>0.05). However, the Medium Gaussian SVM has the lowest MAE for all ranges. The MAE is the lowest for the margin range of 4-6 mm. Furthermore, the MAE is lower for the resection margin range of 0-2 mm compared to the two other feature sets. Lastly, the MAE of both regression models decreased compared to the MAE of the regression analysis with the DRS features, but this decrease was not significant. However, the MAE values of the analysis with only SVM prediction scores and height, are significantly lower than the MAE of only DRS features and of the feature set with both DRS features, prediction scores, and heights (p-value<0.01).

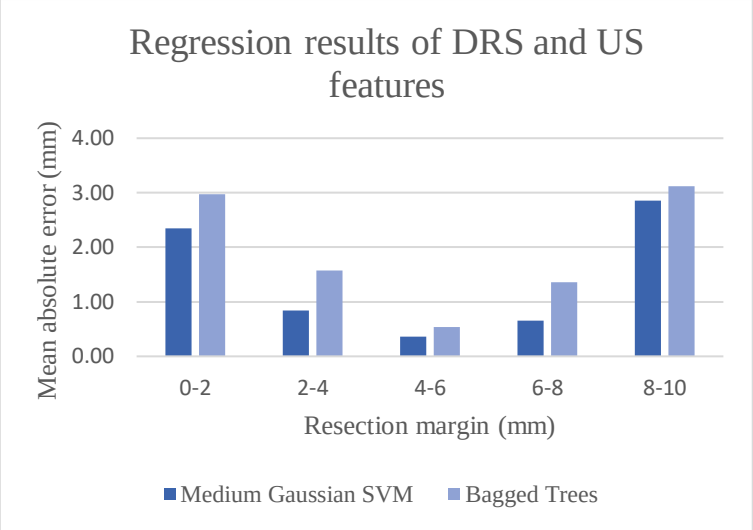


Figure 5-11: Regression results per resection margin range with classification FG-SVM prediction scores, superpixel heights, and DRS features in the feature set.

5.5 Discussion

The current ex vivo study was performed to show whether DRS, US B-mode, and shear wave elastography could be used to estimate the resection margin of colorectal specimens. The results show that a combination of the three techniques results in a better estimation of the resection margin than using only DRS or only US B-mode and elastography. However, it needs to be taken into account that these results are preliminary and more research is needed to draw conclusions. The reason for this is that there is no test set available to demonstrate how the regression analysis would perform on new data. Therefore, the MAE and RMSE may be optimistic. The error values shown in the tables are retrieved after regression analysis with 5-fold cross-validation. However, the cross-validation was random and did not take the specimen numbers into account. Since a maximum of nine measurement locations was used per specimen, there is a big chance that measurements from one specimen were present in both the train and validation set. It would be better to have different specimens in the train and validation set.

5.5.1 DRS

The DRS spectra show differences between the resection margin ranges in several peaks. The peaks with the largest difference are the minimum between 950-1010 nm, the maximum between 1000-1100 nm, the maximum between 1275-1375 nm. Generally, the spectra show more absorbance for shorter resection margins, thus tumor tissue has on average more absorbance than other tissue layers that are between the tumor and the resection plane, which are fat, muscle, or fibrosis. Furthermore, the wavelength ranges with the differences in reflection show that tumor may have relatively more collagen than the other tissue types, through more absorption in the ranges of 1275-1375nm and 1000-1100 for the shorter resection margins than the larger ones. Furthermore, the differences in reflection may be caused through a lower lipid concentration in tumor relative to the other tissue types. Distortion was shown in the spectra from resection margins 6-8 and 8-10 after 1400 nm that was not removed by the normalization. The cause for this distortion is unknown. Since the distortion was not shown in the wavelength range of the peak extraction, the distortion does not influence the features that were used in this study.

The DRS features consisted of three peaks per fiber. Although these peaks were determined beforehand, the second peak (min. 950-1010 nm) and third peak (the difference between the maximum in 1225-1325 nm and the minimum in 1175-1275) are discriminative for the different resection margins, especially for fiber 3 and 5. This is as expected since these spectra include information of a larger depth. However, the spectra of fiber 1 would be useful to tell something of the CRM, since it may show differences between a positive and a negative margin. Since the current study included only four measurements with a positive margin (<1 mm), more data is firstly needed to show whether the spectra of positive margins are deviating compared to negative margins.

The MAE of the analysis with peaks from the normalized spectra was lower than the analysis with peaks from raw spectra. However, the RMSE was lower for the analysis with peaks from the raw spectra. Baltussen et al. also demonstrated that normalization is of additional value when different DRS probes are used for measurements within one data set.[22] Since all measurements were performed with the same measurement protocol and DRS probe, the results are following the expectations. However, the normalized spectra are used for feature extraction. The spectra can be easily compared to the results of other studies, and in the future, the current dataset will be expanded with data from another DRS probe. Therefore, normalization is needed to use all spectra in one dataset.

The performance of the regression analysis was not different for the two feature sets of alternating fibers. This is also expected since measurement volumes of the fibers are overlapping. Furthermore, the performance of the three alternating fibers was the same or better than the complete DRS feature set.

In this work, three peak features were extracted per fiber, but other DRS features may be useful as well. Baltussen et al. investigated different feature extraction techniques of DRS for classification between tumor tissue and the healthy colorectal wall.[22] They demonstrated that shape-based features, spectral bands, and tissue optical features are useful for this classification problem.

5.5.2 US B-mode and elastography features

The feature set with direct B-mode and SWE features from the first superpixel and the feature set with the classification score and the height of the superpixel did not result in a different prediction of the resection margin. Both feature sets have their advantages and limitations. Whereas the first feature set directly says something of the B-mode and elastography features, the number of features is too high relative to the number of samples when including features of three superpixels per measurement. In that case, the feature set consists of 3 x 9 features with a sample size of 47. Therefore, regression analysis with this feature set would be sensitive for overfitting. The other feature set with the prediction score and height directly says something about the thickness of the first tissue layer and prediction of a tissue type. However, if the first superpixel does not contain tumor tissue, there is only information known until the depth of that superpixel. Therefore, it is logical that the MAE increases for samples of which the resection margins are larger than the depth of the first superpixel. Furthermore, a feature set of only two features is sensitive for underfitting, which happens when too little information is known about the outcome. Therefore, it is expected that the addition of the prediction scores and heights from the second and third superpixels results in a better performance of the regression analysis. Despite these features are not direct B-mode and shear wave elastography features, these two features contain (indirect) information of both techniques. A low number of features per superpixel is preferred since these two features are retrieved from three superpixels per measurement. Furthermore, the features will be combined with nine DRS features. To avoid overfitting, it is important to select the minimal number of the most relevant features. For now, it is assumed that these two features, the prediction score, and height, are relevant because of the correlation with the estimation of the resection margin. However, SFFS feature subset selection must be performed to examine whether these features are the most important ones for resection margin estimation and whether these features are the best to be combined with DRS. This feature subset selection was already performed for the tissue type classification, as described in Section **4.3.9**. The feature subset selection in this study is a preliminary study and firstly more measurement data is needed for a proper feature selection.

The use of superpixels is a practical method to distinguish tissue layers based on the B-mode intensity automatically. However, as shown in Chapter **5**, there are still superpixels that include mixtures of tissue type. In these cases, the boundary between the tumor and another tissue type may be within a superpixel, and the height does not say anything of the distance to the tumor. This problem may be reduced when more superpixels are used instead of ~40 superpixels per image, as shown in Section **4.3.3**. The next step is to apply deep learning on complete B-mode and shear wave elastography images and to make a prediction score image with a score per pixel. Thereafter, the distance can be measured based on the gradient between pixels. A gradient represents the boundary between healthy tissue and tumor because the prediction score changes. With this method, the estimation of the margin can be performed up to 0.071 mm, since the pixel size of the eL18-4 transducer is 0.071 mm. However, far more measurement data is needed to train such a deep learning algorithm. For example, Wang et al. applied deep learning with radiomics features of shear wave elastography to assess liver fibrosis stages and used 1990 images from 398 patients.[133] Another study, by Zhang et al., used deep learning to distinguish benign from malignant breast tumors and included 227 shear wave elastography images.[107]

5.5.3 Combining DRS, US B-mode, and elastography

The performance of the regression analysis significantly improved (p-value<0.05) after the combination of the DRS peaks with the classification scores and heights of three superpixels relative to using only of these feature sets. This is following the expectations since each feature contributes information that is useful for the resection margin prediction. Whereas the score and heights say something of the tissue layer thickness and an estimation of the tissue type in the order of mm, the DRS shows the tissue composition of a volume in the order of mm^3.

The MAE is the lowest for resection margins between 4-6 mm, this is logical since this group had the most number of measurements and therewith more training data. The MAE per resection margin range shows that the errors are negatively correlated with the number of included measurements. Therefore, the lower MAE is caused by the sample size. Accordingly, more data needs to be included, such that each resection margin range contains at least about 10 measurements. Furthermore, it is logical that the error increases for larger margins since the DRS measurement volume only reaches ~8 mm with fiber 6, and ~6 mm with fiber 5. Furthermore, when absorption of light occurs in the superficial tissue layer, less reflectance information can be retrieved from this absorption wavelength of the tissue layers beneath. This was also demonstrated by a study by Baltussen et al., which showed that it was harder to distinguish tumor tissue from healthy tissue types when a layer of healthy colorectal tissue, fat, or fibrosis was on top of the tumor.[21] For the larger resection margins it is also less relevant to estimate the margin precisely per mm since it does not influence the oncologic outcome. For example, when the true margin is 8 mm, and the estimation is 5 mm, the resection margin is negative either way. The reason that these margins are included is to enlarge the dataset. Furthermore, it may be useful to recognize rectal tumor invasion in the distal direction which can be the difference between preserving the anorectal sphincter or not.[134] With an MAE of more than 2 mm for resection margins <2 mm, this regression model is not yet clinically relevant. The results would be clinically relevant when it can distinguish positive margins (≤1 mm) from negative margins. Therefore, the error needs to be <1 mm. When this error is achieved, it must be investigated whether the use of these techniques decreases the number of positive resection margins while preserving more healthy tissues.

5.6 Conclusion

The goal of this study was to investigate whether US B-mode and elastography are of additional value to DRS in the estimation of the resection margin. This study is the first one that combined these three techniques. Therefore, it demonstrates a new method to estimate the resection margin of colorectal specimens. The preliminary results show that the combination of the three techniques results in a better estimation of the resection margin than using only DRS or only US B-mode and elastography. This study is a promising starting point for further research of using these three techniques in the resection margin estimation. The recommendations for this study are the following:

- Expand the data collection to have more ex vivo measurements with DRS, US B-mode, and shear wave elastography data and preferably with an equal sample size distribution.
- Improve the classification algorithm in Chapter **5** with the given recommendations to get better prediction scores.
- Expand feature extraction of DRS with features that were used for colorectal tumor classification by Baltussen et al. and perform feature subset selection afterward to select the most valuable features for resection margin prediction. [22]
- Divide the dataset beforehand into a train and test set to estimate the performance on new data samples.

6 Discussion and conclusion

The clinical problem behind these studies was that avoiding positive circumferential margins in locally advanced rectal carcinomas is problematic through the effects of neoadjuvant therapy and the technically challenging procedure.[4,9,10,49,53] Furthermore, it is important to preserve healthy structures, such as the healthy colorectal wall, blood vessels, nerves, and ureters.[4] Damage to these structures can lead to complications such as urogenital and anorectal dysfunction.[59,60] This clinical problem leads to the question of whether intraoperative tissue recognition can be helpful to avoid positive resection margins while preventing the surrounding healthy tissue.[10,61]

The optical imaging technique DRS has been used in ex vivo studies and an in vivo study to investigate whether it can be used to discriminate colorectal tumors from the healthy colorectal, perirectal/colonic fat, and fibrosis.[4,10,21,22] Although the studies demonstrated that DRS is useful to discriminate tumor tissue from the healthy colorectal wall and fat, there are some limitations of DRS that may be solved with additional imaging techniques. One limitation was that it was difficult to select measurement points at the locations where the tumor is most superficial. Since an intraoperative tool needs to help in the estimation of the circumferential margin, the measurements must be performed where the tumor is the closest to the resection plane. Furthermore, data samples with tumor tissue are needed to train algorithms in the classification of tumor tissue and fibrosis. Another limitation was that DRS has not been proven to distinguish pure fibrotic tissue from tumor tissue when a thin layer of healthy colorectal tissue or fibrosis is between the tumor and the resection plane. Subsequently, a limitation of the studies with DRS was that the maximum measurement depth was 1-2 mm, whereas the tissue layers deeper than 2 mm can help in the classification of the superficial tissue type, and thus in the estimation of the resection margin. However, it becomes harder to estimate tissue boundaries within a DRS measurement volume when the volume increases. Another limitation was that limited DRS data was available of patients who received neoadjuvant radiotherapy.

To solve these limitations, ultrasound could be added to DRS since it clearly shows the tissue layers of the healthy colorectal wall. Furthermore, colorectal tumors can clearly be distinguished from healthy tissue. However, when a patient received neoadjuvant therapy, which can result in local effects such as fibrosis, it is difficult to distinguish tumor from fibrosis through the same ultrasonic appearance, and in both cases, the healthy colorectal wall structure is lost.[26,54,56] Another technique, ultrasound elastography, is of additional value to US B-mode in colorectal tumor grade assessment.[32,33,35] Furthermore, US elastography has been used for response assessment in breast cancer and locally advanced rectal cancer.[34,94,100] However, the current literature does not show whether US elastography can differentiate between tumor and fibrosis. Furthermore, it has not been used yet to estimate the resection margin. Therefore, an ex vivo was performed to investigate whether US B-mode and shear wave elastography can be used to discriminate tumor from fibrosis, the healthy colorectal wall, and fat. The study population consisted of surgical specimens from patients with locally advanced colorectal carcinoma. The results of this study showed that the B-mode intensity of tumor is on average significantly lower than fibrosis, muscle, and fat (p-value<0.01). However, the tissue types cannot be distinguished purely based on a B-mode intensity. The stiffness values based on shear wave elastography are overlapping for both tumor and fibrosis. A classification algorithm was trained with five B-mode features and four elastography features to distinguish colorectal tumors from the healthy colorectal wall, fat, and fibrosis. A Bagged Trees classifier and FG-SVM classifier resulted in the best performance. The Bagged Trees classifier had an MCC of 0.43 for the training set, and 0.40 for the test set, accuracies of 0.73 and 0.86 and AUC values of 0.76 and 0.83, sensitivities of 0.76 and 0.87, and specificities of 0.67 and 0.74. The FG-SVM classifier had an MCC of 0.49 for the training set, and 0.40 for the test set, accuracies of 0.75 and 0.89, AUC values of 0.78 and 0.81, sensitivities of 0.83 and 0.91, and specificities of 0.66 and 0.63. The Bagged Trees resulted in overlapping prediction scores for tumor and fibrosis, therefore it shows that the Bagged Trees has difficulties with distinguishing these tissue types. The FG-SVM had low prediction scores for the healthy tissue types but resulted in a wide range

of prediction scores for tumor tissue. Therefore, when a cutoff value is chosen to detect all tumor tissues (high sensitivity), most of the healthy tissues are falsely classified as tumor tissue (low specificity). From these results can be concluded that the classification algorithms are not yet able to distinguish tumor while preserving healthy tissues. However, the ex vivo study resulted in new contributions to the literature. Firstly, it gave new insights into US B-mode and shear wave elastography features of colorectal tumor tissue, fibrosis, the colorectal wall, and fat. Furthermore, it provides a new method of how superpixels can be used to automatically extract features from regions with the same B-mode intensity from ultrasound B-mode and elastography images. Thirdly, it demonstrated how markers were used to correlate different tissue types on the ultrasound images to the pathology. Lastly, the results of this ex vivo study are a promising starting point for further research into tissue type classification of colorectal tumors and healthy tissues. Also, the classification algorithms have room for improvement. Firstly, the labeling of the superpixels with the tissue type needs to be done by a second observer. Secondly, the influence of the confidence maps on the shear wave elastography measurements can be investigated. Thirdly, the data set must be expanded to obtain equally divided train and test sets. Fourthly, it may be investigated whether the classification of tissue types can be divided into a two-step classification, with first classifying tumor and fibrosis versus the healthy colorectal wall and fat, and secondly classifying tumor from fibrosis. Fifthly, other features can be used to distinguish the tissue types, such as the stiffness ratio between two tissue types. Furthermore, the next step to make the classification based on ultrasound B-mode and elastography more clinically relevant, the classification predictions can be used to estimate the resection margin.

Therefore, another ex vivo study was performed to investigate how the ultrasound B-mode and elastography could contribute to DRS in the estimation of the resection margin. Subsequently, it was investigated whether a combination of these techniques is beneficial in the estimation of the circumferential resection margin compared to only DRS or only ultrasound B-mode and shear wave elastography. This study proposed a method to correlate the three techniques to the resection margin based on histopathology. Regression analysis was performed using DRS, B-mode, shear wave elastography, and the classification results of the first ex vivo study. The results showed that a combination of the three techniques resulted in the lowest MAE, compared to using only DRS or using US B-mode and elastography. However, the MAE is still too large for clinical relevance. When the regression analysis can help in distinguishing positive margins from negative margins it can be helpful for the surgeon to avoid positive margins while preserving healthy tissues. Since these results are preliminary, improvements of the resection margin estimation are expected when the data set is expanded such that the data can be divided into a train and test set, when other DRS features are extracted, when automated feature subset selection is performed, and when the classification algorithms from the first ex vivo study are improved.

In conclusion, the results of the ex vivo studies in this thesis gave new insight into how ultrasound B-mode, ultrasound shear wave elastography can be used to distinguish tumor from fibrosis, the healthy colorectal wall, and fat. Furthermore, it demonstrated how the ultrasound techniques can be combined with DRS in the estimation of the resection margin of colorectal tumors. However, more research is needed to investigate whether these techniques can lead to less positive resection margins of rectal tumors while preserving more healthy tissues.

7 Bibliography

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).

2. Siegel, R.L., Miller, K.D., Goding Sauer, A., Fedewa, S.A., Butterly, L.F., Anderson, J.C., Cercek, A., Smith, R.A., and Jemal, A. Colorectal cancer statistics, 2020. *CA. Cancer J. Clin.* (2020) doi:10.3322/caac.21601.

3. Richtlijn Colorectaal Carcinoom. *Nederlandse Vereniging voor Heelkunde.* (2019).

4. Baltussen, E.J.M., Brouwer De Koning, S.G., Sanders, J., Aalbers, A.G.J., Kok, N.F.M., Beets, G.L., Hendriks, B.H.W., Sterenborg, H.J.C.M., Kuhlmann, K.F.D., and Ruers, T.J.M. Tissue diagnosis during colorectal cancer surgery using optical sensing: An in vivo study. *J. Transl. Med.* **17**, 1–10 (2019).

5. Glynne-Jones, R., Wyrwicz, L., Tiret, E., Brown, G., Rödel, C., Cervantes, A., and Arnold, D. Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **28**, iv22–iv40 (2017).

6. Feeney, G., Sehgal, R., Sheehan, M., Hogan, A., Regan, M., Joyce, M., and Kerin, M. Neoadjuvant radiotherapy for rectal cancer management. *World J. Gastroenterol.* **25**, 4850–4869 (2019).

7. Li, Y., Wang, J., Ma, X., Li, T., Yan, Y., Xue, C., Hui, B., Liu, R., Ma, H., and Ren, J. A review of neoadjuvant chemoradiotherapy for locally advanced rectal cancer. *International Journal of Biological Sciences* vol. 12 1022–1031 (2016).

8. Seo, N., Kim, H., Cho, M.S., and Lim, J.S. Response assessment with MRI after chemoradiotherapy in rectal cancer: Current evidences. *Korean J. Radiol.* **20**, 1003–1018 (2019).

9. Granero-Castro, P., Muñoz, E., Frasson, M., García-Granero, A., Esclapez, P., Campos, S., Flor-Lorente, B., and Garcia-Granero, E. Evaluation of mesorectal fascia in mid and low anterior rectal cancer using endorectal ultrasound is feasible and reliable: A comparison with mri findings. *Dis. Colon Rectum* **57**, 709–714 (2014).

10. Baltussen, E.J.M., Snaebjornsson, P., de Koning, S.G.B., Sterenborg, H.J.C.M., Aalbers, A.G.J., Kok, N., and Beets, G.L. Diffuse reflectance spectroscopy as a tool for real-time tissue assessment during colorectal cancer surgery. *J. Biomed. Opt.* **22**, 1 (2017).

11. Warrier, S.K., Kong, J.C., Guerra, G.R., Chittleborough, T.J., Naik, A., Ramsay, R.G., Lynch, A.C., and Heriot, A.G. Risk factors associated with circumferential resection margin positivity in rectal cancer: A binational registry study. *Dis. Colon Rectum* **61**, 433–440 (2018).

12. Homan, J., Bökkerink, G.M., Aarts, M.J., Lemmens, V.E., Van Lijnschoten, G., Rutten, H.J., Wijsman, J.H., Nagtegaal, I.D., and De Wilt, J.H.W. Variation in circumferential resection margin: Reporting and involvement in the South-Netherlands. *Eur. J. Surg. Oncol.* **41**, 1485–1492 (2015).

13. Rickles, A.S., Dietz, D.W., Chang, G.J., Wexner, S.D., Berho, M.E., Remzi, F.H., Greene, F.L., James, W., Abbas, M.A., Peters, W., Noyes, K., and John, R.T. High rate of positive circumferential resection margins following rectal cancer surgery. **262**, 891–898 (2017).

14. Quirke, P., Steele, R., Monson, J., Grieve, R., Khanna, S., Couture, J., O'Callaghan, C., Myint, A.S., Bessell, E., Thompson, L.C., Parmar, M., Stephens, R.J., and Sebag-Montefiore, D. Effect of the plane of surgery achieved on local recurrence in patients with operable rectal cancer: a prospective study using data from the MRC CR07 and NCIC-CTG CO16 randomised clinical trial. *Lancet* **373**, 821–828 (2009).

15. Detering, R., Saraste, D., de Neree tot Babberich, M.P.M., Dekker, J.W.T., Wouters, M.W.J.M., van Geloven, A.A.W., Bemelman, W.A., Tanis, P.J., Martling, A., Westerterp, M., Aalbers, A., Beets-Tan, R., Boer, F. den, Breukink, S., Coene, P.P., Doornebosch, P., Gelderblom, H., … Westreenen, E. van. International evaluation of circumferential resection margins after rectal cancer resection: insights from the Swedish and Dutch audits. *Color. Dis.* **22**, 416–429 (2020).

16. Alberda, W.J., Verhoef, C., Schipper, M.E.I., Nuyttens, J.J., Rothbarth, J., De Wilt, J.H.W., and Burger, J.W.A. The importance of a minimal tumor-free resection margin in locally recurrent rectal cancer. *Dis. Colon Rectum* **58**, 677–685 (2015).

17. DeCaria, K., Rahal, R., Niu, J., Lockwood, G., and Bryant, H. Rectal cancer resection and circumferential margin rates in Canada: A population-based study. *Curr. Oncol.* **22**, 60–63 (2015).

18. Goffredo, P., Zhou, P., Ginader, T., Hrabe, J., Gribovskaja-Rupp, I., Kapadia, M., You, Y.N., and Hassan, I. Positive circumferential resection margins following locally advanced colon cancer surgery: Risk factors and survival impact. *J. Surg. Oncol.* **121**, 538–546 (2020).

19. Sauer, R., Liersch, T., Merkel, S., Fietkau, R., Hohenberger, W., Hess, C., Becker, H., Raab, H.R., Villanueva, M.T., Witzigmann, H., Wittekind, C., Beissbarth, T., and Rödel, C. Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: Results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years. *J. Clin. Oncol.* **30**, 1926–1933 (2012).

20. Hav, M., Libbrecht, L., Ferdinande, L., Geboes, K., Pattyn, P., and Cuvelier, C.A. Pathologic Assessment of Rectal Carcinoma after Neoadjuvant Radio(chemo)therapy: Prognostic Implications. *BioMed Research International* vol. 2015 (2015).

21. Baltussen, E.J.M., Brouwer de Koning, S.G., Sanders, J., Aalbers, A.G.J., Kok, N.F.M., Beets, G.L., Hendriks, B.H.W., Sterenborg, H.J.C.M., Kuhlmann, K.F.D., and Ruers, T.J.M. Using Diffuse Reflectance Spectroscopy to Distinguish Tumor Tissue From Fibrosis in Rectal Cancer Patients as a Guide to Surgery. *Lasers Surg. Med.* (2019) doi:10.1002/lsm.23196.

22. Baltussen, E.J.M., Sterenborg, H.J.C.M., Ruers, T.J.M., and Dashtbozorg, B. Optimizing algorithm development for tissue classification in colorectal cancer based on diffuse reflectance spectra. *Biomed. Opt. Express* **10**, 6096 (2019).

23. Hayashi, M., Yamamoto, Y., and Iwase, H. Clinical imaging for the prediction of neoadjuvant chemotherapy 2304-3873. *Chinese Clin. Oncol.* **9**, (2020).

24. Pomerri, F., Pucciarelli, S., Maretto, I., Zandon, M., Del Bianco, P., Amadio, L., Rugge, M., Nitti, D., and Muzzio, P.C. Prospective assessment of imaging after preoperative chemoradiotherapy for rectal cancer. *Surgery* **149**, 56–64 (2011).

25. Memon, S., Lynch, A.C., Bressel, M., Wise, A.G., and Heriot, A.G. Systematic review and meta-analysis of the accuracy of MRI and endorectal ultrasound in the restaging and response assessment of rectal cancer following neoadjuvant therapy. *Color. Dis.* **17**, 748–761 (2015).

26. Bor, R., Fábián, A., and Szepes, Z. Role of ultrasound in colorectal diseases. *World J. Gastroenterol.* **22**, 9477–9487 (2016).

27. Tomizawa, M., Shinozaki, F., Hasegawa, R., Fugo, K., Shirai, Y., Motoyoshi, Y., Sugiyama, T., Yamamoto, S., Kishimoto, T., and Ishige, N. Characteristics of colorectal cancer diagnosed with screening abdominal ultrasonography. *Mol Clin Oncol* **5**, 64–68 (2016).

28. Waage, J.E.R., Leh, S., Røsler, C., Pfeffer, F., Bach, S.P., Havre, R.F., Haldorsen, I.S., Ødegaard, S., and Baatrup, G. Endorectal ultrasonography, strain elastography and MRI differentiation of rectal adenomas and adenocarcinomas. *Color. Dis.* **17**, 124–131 (2015).

29. Chen, L. Da, Li, W., Xian, M.F., Zheng, X., Lin, Y., Liu, B.X., Lin, M.X., Li, X., Zheng, Y.L., Xie, X.Y., Lu, M. De, Kuang, M., Xu, J.B., and Wang, W. Preoperative prediction of tumour deposits in rectal cancer by an artificial neural network–based US radiomics model. *Eur. Radiol.* **30**, 1969–1979 (2020).

30. Ashraf, S., Hompes, R., Slater, A., Lindsey, I., Bach, S., Mortensen, N.J., and Cunningham, C. A critical appraisal of endorectal ultrasound and transanal endoscopic microsurgery and decision-making in early rectal cancer. *Color. Dis.* **14**, 821–826 (2012).

31. Waage, J.E.R., Bach, S.P., Pfeffer, F., Leh, S., Havre, R.F., Ødegaard, S., and Baatrup, G. Combined endorectal ultrasonography and strain elastography for the staging of early rectal cancer. *Color. Dis.* **17**, 50–56 (2015).

32. Fan, Z., Cong, Y., Zhang, Z., Li, R., Wang, S., and Yan, K. Shear Wave Elastography in Rectal Cancer Staging, Compared with Endorectal Ultrasonography and Magnetic Resonance Imaging. *Ultrasound Med. Biol.* **45**, 1586–1593 (2019).

33. Chen, L. Da, Liang, J.Y., Wu, H., Wang, Z., Li, S.R., Li, W., Zhang, X.H., Chen, J.H., Ye, J.N., Li, X.,

Xie, X.Y., Lu, M. De, Kuang, M., Xu, J.B., and Wang, W. Multiparametric radiomics improve prediction of lymph node metastasis of rectal cancer compared with conventional radiomics. *Life Sci.* **208**, 55–63 (2018).

34. Rafaelsen, S.R., Vagn-Hansen, C., Sørensen, T., Lindebjerg, J., Pløen, J., and Jakobsen, A. Ultrasound elastography in patients with rectal cancer treated with chemoradiation. *Eur. J. Radiol.* **82**, 913–917 (2013).

35. Chen, L. Da, Wang, W., Xu, J.B., Chen, J.H., Zhang, X.H., Wu, H., Ye, J.N., Liu, J.Y., Nie, Z.Q., Lu, M. De, and Xie, X.Y. Assessment of rectal tumors with shear-wave elastography before surgery: Comparison with endorectal US. *Radiology* **285**, 279–292 (2017).

36. Palmeri, M.L., Frinkley, K.D., Zhai, L., Gottfried, M., Bentley, R.C., Ludwig, K., and Nightingale, K.R. Acoustic Radiation Force Impulse (ARFI) Imaging of the Gastrointestinal Tract. *Ultrasound imaging* **27**, 75–88 (2005).

37. Mahadevan, V. Anatomy of the caecum, appendix and colon. *Surg. (United Kingdom)* **38**, 1–6 (2020).

38. Wolff, B.G., Fleshman, J.W., Beck, D.E., Pemberton, J.H., and Wexner, S.D. *The ASCRS Textbook of Colon and Rectal Surgery*. (Springer, 2007).

39. Ponz de Leon, M. and Di Gregorio, C. Pathology of colorectal cancer. *Dig. Liver Dis.* **33**, 372–388 (2001).

40. Nederlandse Kankerregistratie. cijfers over kanker. https://www.cijfersoverkanker.nl/selecties/Dataset_3/img5c877f8263442 (2021). [Assessed on 16-1-2021]

41. Ballinger, A.B. and Anggiansah, C. Colorectal cancer. *BMJ* **335**, 715–718 (2007).

42. Toes-Zoutendijk, E., van Leerdam, M.E., Dekker, E., van Hees, F., Penning, C., Nagtegaal, I., van der Meulen, M.P., van Vuuren, A.J., Kuipers, E.J., Bonfrer, J.M.G., Biermann, K., Thomeer, M.G.J., van Veldhuizen, H., Kroep, S., van Ballegooijen, M., Meijer, G.A., de Koning, H.J., … Blankenstein, M.A. Real-Time Monitoring of Results During First Year of Dutch Colorectal Cancer Screening Program and Optimization by Altering Fecal Immunochemical Test Cut-Off Levels. *Gastroenterology* **152**, 767-775.e2 (2017).

43. Sobin, L.H., Gospodarowicz, M.K., and Wittekind, C. *Classification of Malignant Tumours*. http://www.inen.sld.pe/portal/documentos/pdf/educacion/13072015_TNM Classification.pdf (2009).

44. Kim, N.K., Kim, Y.W., Han, Y.D., Cho, M.S., Hur, H., Min, B.S., and Lee, K.Y. Complete mesocolic excision and central vascular ligation for colon cancer: Principle, anatomy, surgical technique, and outcomes. *Surg. Oncol.* **25**, 252–262 (2016).

45. Ishiguro, M., Higashi, T., Watanabe, T., and Sugihara, K. Changes in colorectal cancer care in Japan before and after guideline publication: A nationwide survey about d3 lymph node dissection and adjuvant chemotherapy. *J. Am. Coll. Surg.* **218**, 969-977.e1 (2014).

46. Hakiman, H., Boostrom, S., and Fleshman, J. Total Mesorectal Excision with Autonomic Nerve Preservation: "Optimized Surgery" BT - Modern Management of Cancer of the Rectum. in (eds. Longo, W. E., Reddy, V. & Audisio, R. A.) 173–186 (Springer London, 2015). doi:10.1007/978-1-4471-6609-2_12.

47. Heald, R.. and Ryall, R.D.. Recurrence and survival after total mesorectal excision for rectal cancer. *Lancet* **327**, 1479–1482 (1986).

48. Campa-Thompson, M., Weir, R., Calcetera, N., Quirke, P., and Carmack, S. Pathologic processing of the total mesorectal excision. *Clin. Colon Rectal Surg.* **28**, 43–52 (2015).

49. Nishikawa, T., Ishihara, S., Emoto, S., Kaneko, M., Murono, K., Sasaki, K., Otani, K., Tanaka, T., Kiyomatsu, T., Hata, K., Kawai, K., Nozawa, H., and Watanabe, T. Multivisceral resections for locally advanced colorectal cancer after preoperative treatment. *Mol. Clin. Oncol.* 493–498 (2018) doi:10.3892/mco.2018.1559.

50. Chalabi, M., Fanchi, L.F., Dijkstra, K.K., Van den Berg, J.G., Aalbers, A.G., Sikorska, K., Lopez-Yurda, M., Grootscholten, C., Beets, G.L., Snaebjornsson, P., Maas, M., Mertz, M., Veninga, V., Bounova, G.,

Broeks, A., Beets-Tan, R.G., de Wijkerslooth, T.R., … Haanen, J.B. Neoadjuvant immunotherapy leads to pathological responses in MMR-proficient and MMR-deficient early-stage colon cancers. *Nat. Med.* (2020) doi:10.1038/s41591-020-0805-8.

51. Amri, R., Bordeianou, L.G., Sylla, P., and Berger, D.L. Association of Radial Margin Positivity With Colon Cancer. *JAMA Surg.* **150**, 890–898 (2015).

52. Khan, M.A.S., Hakeem, A.R., Scott, N., and Saunders, R.N. Significance of R1 resection margin in colon cancer resections in the modern era. *Color. Dis. Off. J. Assoc. Coloproctology Gt. Britain Irel.* **17**, 943–953 (2015).

53. Leonard, D., Penninckx, F., Fieuws, S., Jouret-Mourin, A., Sempoux, C., Jehaes, C., and Van Eycken, E. Factors predicting the quality of total mesorectal excision for rectal cancer. *Ann. Surg.* **252**, 982–988 (2010).

54. Kye, B.H., Kim, H.J., Kim, G., Kim, J.G., and Cho, H.M. Multimodal assessments are needed for restaging after neoadjunvant chemoradiation therapy in rectal cancer patients. *Cancer Res. Treat.* **48**, 561–566 (2016).

55. Xu, D., Ju, H.-X., Qian, C.-W., and Jiang, F. The value of TRUS in the staging of rectal carcinoma before and after radiotherapy and comparison with the staging postoperative pathology. *Clin. Radiol.* **69**, 481–484 (2014).

56. Gersak, M.M., Badea, R., Graur, F., Al Hajjar, N., Furcea, L., and Dudea, S.M. Endoscopic ultrasound for the characterization and staging of rectal cancer. Current state of the method. Technological advances and perspectives. *Med. Ultrason.* **17**, 227–234 (2015).

57. Agger, E.A., Jörgren, F.H., Lydrup, M.L.A., and Buchwald, P.L. Risk of local recurrence of rectal cancer and circumferential resection margin: population-based cohort study. *Br. J. Surg.* **107**, 580–585 (2020).

58. Nagtegaal, I.D., Marijnen, C.A.M., Kranenbarg, E.K., Van De Velde, C.J.H., and Van Krieken, J.H.J.M. Circumferential margin involvement is still an important predictor of local recurrence in rectal carcinoma: Not one millimeter but two millimeters is the limit. *Am. J. Surg. Pathol.* **26**, 350–357 (2002).

59. Burks, F.N. and Santucci, R.A. Management of iatrogenic ureteral injury. *Ther. Adv. Urol.* **6**, 115–124 (2014).

60. Giglia, M.D. and Stein, S.L. Overlooked Long-Term Complications of Colorectal Surgery. *Clin. Colon Rectal Surg.* **32**, 204–211 (2019).

61. de Boer, L.L., Molenkamp, B.G., Bydlon, T.M., Hendriks, B.H.W., Wesseling, J., Sterenborg, H.J.C.M., and Ruers, T.J.M. Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries. *Breast Cancer Res. Treat.* **152**, 509–518 (2015).

62. Gomes, A.J. and Backman, V. Algorithm for automated selection of application-specific fiber-optic reflectance probes. *J. Biomed. Opt.* **18**, 027012 (2013).

63. Bydlon, T.M., Nachabé, R., Ramanujam, N., Sterenborg, H.J.C.M., and Hendriks, B.H.W. Chromophore based analyses of steady-state diffuse reflectance spectroscopy: Current status and perspectives for clinical adoption. *J. Biophotonics* **8**, 9–24 (2015).

64. Nachabé, R., Evers, D.J., Hendriks, B.H.W., Lucassen, G.W., van der Voort, M., Rutgers, E.J., Peeters, M.-J.V., Van der Hage, J.A., Oldenburg, H.S., Wesseling, J., and Ruers, T.J.M. Diagnosis of breast cancer using diffuse optical spectroscopy from 500 to 1600 nm: comparison of classification methods. *J. Biomed. Opt.* **16**, 087010 (2011).

65. Rinnan, Å., Berg, F. van den, and Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends Anal. Chem.* **28**, 1201–1222 (2009).

66. Nachabé, R., Hendriks, B.H.W., Desjardins, A.E., van der Voort, M., van der Mark, M.B., and Sterenborg, H.J.C.M. Estimation of lipid and water concentrations in scattering media with diffuse optical spectroscopy from 900 to 1600 nm. *J. Biomed. Opt.* **15**, 037015 (2010).

67. Farrell, T.J., Patterson, M.S., and Wilson, B. A Diffusion Theory Model of Spatially Resolved, Steady-State Diffuse Reflectance for the Noninvasive Determination of Tissue Optical Properties in Vivo. *Med.*

Phys. **19**, 879–888 (1992).

68. Evers, D.J., Hendriks, B.H.W., Lucassen, G.W., and Ruers, T.J.M. Optical spectroscopy: Current advances and future applications in cancer diagnostics and therapy. *Future Oncology* vol. 8 307–320 (2012).

69. Kho, E., de Boer, L.L., Post, A.L., Van de Vijver, K.K., Jóźwiak, K., Sterenborg, H.J.C.M., and Ruers, T.J.M. Imaging depth variations in hyperspectral imaging: Development of a method to detect tumor up to the required tumor-free margin width. *J. Biophotonics* **12**, (2019).

70. Jermyn, M., Mercier, J., Aubertin, K., Desroches, J., Urmey, K., Karamchandiani, J., Marple, E., Guiot, M.C., Leblond, F., and Petrecca, K. Highly accurate detection of cancer in situ with intraoperative, label-free, multimodal optical spectroscopy. *Cancer Res.* **77**, 3942–3950 (2017).

71. De Boer, L.L., Bydlon, T.M., Van Duijnhoven, F., Vranken Peeters, M.J.T.F.D., Loo, C.E., Winter-Warnars, G.A.O., Sanders, J., Sterenborg, H.J.C.M., Hendriks, B.H.W., and Ruers, T.J.M. Towards the use of diffuse reflectance spectroscopy for real-time in vivo detection of breast cancer during surgery. *J. Transl. Med.* **16**, 1–14 (2018).

72. O'Brien, W.D. and O'Brien Jr, W.D. Ultrasound-biophysics mechanisms. *Prog. Biophys. Mol. Biol.* **93**, 212–255 (2007).

73. Powles, A.E.J., Martin, D.J., Wells, I.T.P., and Goodwin, C.R. Physics of ultrasound. *Anaesth. Intensive Care Med.* **19**, 202–205 (2018).

74. Goedee, H.S. Basisprincipes Echografie. *Nederlandse Vereniging voor Klinische Neurofysiology* https://nvknf.nl/neuromusculair/basis-principes/ (2018).

75. Prince, J.L. and Links, J.M. Medical imaging signals and systems. in *Pearson Education* 335–362 (Pearson Education, 2015).

76. Oien, K., Mjørud Forsmo, H., Rösler, C., Nylund, K., Waage, J.E., and Pfeffer, F. Endorectal ultrasound and magnetic resonance imaging for staging of early rectal cancers: how well does it work in practice? *Acta Oncol. (Madr).* **58**, S49–S54 (2019).

77. Miro, A.G.F., Grobler, S., and Santoro, G.A. Preoperative Staging of Rectal Cancer: Role of Endorectal Ultrasound. *Rectal Cancer - A Multidiscip. Approach to Manag.* (2011) doi:10.5772/38855.

78. Greif, F., Aranovich, D., Hananel, N., Knizhnik, M., and Belenky, A. Intraoperative ultrasound in colorectal surgery. *J. Clin. Ultrasound* **37**, 375–379 (2009).

79. Hildebrandt, U. and Feifel, G. Preoperative staging of rectal cancer by intrarectal ultrasound. *Dis. Colon Rectum* **28**, 42–46 (1985).

80. Xiao, Y., Xu, D., Ju, H., Yang, C., Wang, L., Wang, J., Hazle, J.D., and Wang, D. Application value of biplane transrectal ultrasonography plus ultrasonic elastosonography and contrast-enhanced ultrasonography in preoperative T staging after neoadjuvant chemoradiotherapy for rectal cancer. *Eur. J. Radiol.* **104**, 20–25 (2018).

81. Garra, B.S. Elastography: history, principles, and technique comparison. *Abdom. Imaging* **40**, 680–697 (2015).

82. Sigrist, R.M.S., Liau, J., Kaffas, A. El, Chammas, M.C., and Willmann, J.K. Ultrasound elastography: Review of techniques and clinical applications. *Theranostics* **7**, 1303–1329 (2017).

83. Mo, J., Xu, H., Qiang, B., Giambini, H., Kinnick, R., An, K.N., Chen, S., and Luo, Z. *2016 Mo - Bias of shear wave elasticity measurements in thin layer samples and a simple correction strategy.pdf. SpringerPlus* vol. 5 (Springer International Publishing, 2016).

84. Mutala, T.M., Ndaiga, P., and Aywak, A. Comparison of qualitative and semiquantitative strain elastography in breast lesions for diagnostic accuracy. *Cancer Imaging* vol. 16 (2016).

85. Zhang, B., Zhu, F., Li, P., Yu, S., Zhao, Y., and Li, M. Endoscopic ultrasound elastography in the diagnosis of pancreatic masses: A meta-analysis. *Pancreatology* **18**, 833–840 (2018).

86. Havre, R.F., Leh, S., Gilja, O.H., Odegaard, S., Waage, J.E., Baatrup, G., and Nesje, L.B. Strain

assessment in surgically resected inflammatory and neoplastic bowel lesions. *Ultraschall der Medizin* **35**, 149–158 (2014).

87. Mohey, N. and Hassan, T.A. Value of mammography and combined grey scale ultrasound and ultrasound elastography in the differentiation of solid breast lesions. *Egypt. J. Radiol. Nucl. Med.* **45**, 253–261 (2014).

88. Fernandes, J., Sannachi, L., Tran, W.T., Koven, A., Watkins, E., Hadizad, F., Gandhi, S., Wright, F., Curpen, B., El Kaffas, A., Faltyn, J., Sadeghi-Naini, A., and Czarnota, G. Monitoring Breast Cancer Response to Neoadjuvant Chemotherapy Using Ultrasound Strain Elastography. *Transl. Oncol.* **12**, 1177–1184 (2019).

89. Chang, J., Won, J., Lee, J., Park, I.A., Yi, A., and Moon, W.K. Comparison of Shear-Wave and Strain Ultrasound Elastography in the Differentiation of Benign and Malignant Breast Lesions. *AJR. Am. J. Roentgenol.* **201**, W347-56 (2013).

90. Hahn, S., Lee, Y.H., Lee, S.H., and Suh, J.S. Value of the strain ratio on ultrasonic elastography for differentiation of benign and malignant soft tissue tumors. *J. Ultrasound Med.* **36**, 121–127 (2017).

91. Pardal, R.C., Abrantes, A.F.L., Ribeiro, L.P.V., Almeida, R.P.P., Azevedo, K.B., Figueiredo, T.L., and Rodrigues, S.I. Screening of breast lesions: a comparative study between mammography, B-mode ultrasonography, sonoelastography and histological results. *Radiol. Bras.* **46**, 214–220 (2013).

92. Shamdasani, V. and Xia, H. *Philips ElastQ Imaging*. (2018).

93. Zhang, T.-T., Wang, L., Liu, H.-H., Zhang, C.-Y., Li, X.-M., Lu, J.-P., and Wang, D.-B. Differentiation of pancreatic carcinoma and mass-forming focal pancreatitis: qualitative and quantitative assessment by dynamic contrast-enhanced MRI combined with diffusion-weighted imaging. *Oncotarget* **8**, 1744–1759 (2017).

94. Lee, S.H., Chang, J.M., Han, W., Moon, H.G., Koo, H.R., Gweon, H.M., Kim, W.H., Noh, D.Y., and Moon, W.K. Shear-Wave Elastography for the Detection of Residual Breast Cancer After Neoadjuvant Chemotherapy. *Ann. Surg. Oncol.* **22**, 376–384 (2015).

95. Cassinotto, C., Lapuyade, B., Mouries, A., Hiriart, J.B., Vergniol, J., Gaye, D., Castain, C., Le Bail, B., Chermak, F., Foucher, J., Laurent, F., Montaudon, M., and De Ledinghen, V. Non-invasive assessment of liver fibrosis with impulse elastography: Comparison of Supersonic Shear Imaging with ARFI and FibroScan®. *J. Hepatol.* **61**, 550–557 (2014).

96. Villani, R., Cavallone, F., Romano, A.D., Bellanti, F., and Serviddio, G. Two-dimensional shear wave elastography versus transient elastography: A non-invasive comparison for the assessment of liver fibrosis in patients with chronic hepatitis C. *Diagnostics* **10**, (2020).

97. Chon, Y.E., Choi, E.H., Song, K.J., Park, J.Y., Kim, D.Y., Han, K.H., Chon, C.Y., Ahn, S.H., and Kim, S.U. Performance of Transient Elastography for the Staging of Liver Fibrosis in Patients with Chronic Hepatitis B: A Meta-Analysis. *PLoS One* **7**, 1–7 (2012).

98. Wong, V.W.S., Vergniol, J., Wong, G.L.H., Foucher, J., Chan, H.L.Y., Le Bail, B., Choi, P.C.L., Kowo, M., Chan, A.W.H., Merrouche, W., Sung, J.J.Y., and De Ledinghen, V. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology* **51**, 454–462 (2010).

99. Liu, Y., Chen, X., Wang, Z., Wang, Z.J., Ward, R.K., and Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **42**, 158–173 (2018).

100. Jing, H., Cheng, W., Li, Z.Y., Ying, L., Wang, Q.C., Wu, T., and Tian, J.W. Early evaluation of relative changes in tumor stiffness by shear wave elastography predicts the response to neoadjuvant chemotherapy in patients with breast cancer. *J. Ultrasound Med.* **35**, 1619–1627 (2016).

101. Bae, S.J., Youk, J.H., Yoon, C.I., Park, S., Cha, C.H., Lee, H.W., Ahn, S.G., Lee, S.A., Son, E.J., and Jeong, J. A nomogram constructed using intraoperative ex vivo shear-wave elastography precisely predicts metastasis of sentinel lymph nodes in breast cancer. *Eur. Radiol.* **30**, 789–797 (2020).

102. Platz Batista da Silva, N., Engeßer, M., Hackl, C., Brunner, S., Hornung, M., Schlitt, H.J., Evert, K., Stroszczynski, C., and Jung, E.M. sIntraoperative Characterization of Pancreatic Tumors Using Contrast-Enhanced Ultrasound and Shear Wave Elastography for Optimization of Surgical Strategies. *J. Ultrasound Med.* 1–13 (2020) doi:10.1002/jum.15545.

103. Kawabata, Y., Okada, T., Iijima, H., Yoshida, M., Iwama, H., Xu, J., Hatano, E., Fujimoto, J., and Suzumura, K. Intraoperative Ultrasound Elastography Is Useful for Determining the Pancreatic Texture and Predicting Pancreatic Fistula after Pancreaticoduodenectomy. *Pancreas* **49**, 799–805 (2020).

104. Brattain, L.J., Telfer, B.A., Dhyani, M., Grajo, J.R., and Samir, A.E. Machine learning for medical ultrasound: status, methods, and future opportunities. *Abdom. Radiol.* **43**, 786–799 (2018).

105. Zhang, Q., Xiao, Y., Chen, S., Wang, C., and Zheng, H. Quantification of Elastic Heterogeneity Using Contourlet-Based Texture Analysis in Shear-Wave Elastography for Breast Tumor Classification. *Ultrasound Med. Biol.* **41**, 588–600 (2015).

106. Gatos, I., Tsantis, S., Spiliopoulos, S., Karnabatidis, D., Theotokas, I., Zoumpoulis, P., Loupas, T., Hazle, J.D., and Kagadis, G.C. A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography. *Ultrasound Med. Biol.* **43**, 1797–1810 (2017).

107. Zhang, Q., Xiao, Y., Dai, W., Suo, J., Wang, C., Shi, J., and Zheng, H. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics* **72**, 150–157 (2016).

108. Bhatia, K.S.S., Lam, A.C.L., Pang, S.W.A., Wang, D., and Ahuja, A.T. Feasibility study of texture analysis using ultrasound shear wave elastography to predict malignancy in thyroid nodules. *Ultrasound Med. Biol.* **42**, 1671–1680 (2016).

109. Wang, J., Guo, L., Shi, X., Pan, W., Bai, Y., and Ai, H. Real-time elastography with a novel quantitative technology for assessment of liver fibrosis in chronic hepatitis B. *Eur. J. Radiol.* **81**, e31–e36 (2012).

110. Xiao, Y., Zeng, J., Niu, L., Zeng, Q., Wu, T., Wang, C., Zheng, R., and Zheng, H. Computer-aided diagnosis based on quantitative elastographic features with supersonic shear wave imaging. *Ultrasound Med. Biol.* **40**, 275–286 (2014).

111. Marcon, M., Ciritsis, A., Rossi, C., Becker, A.S., Berger, N., Wurnig, M.C., Wagner, M.W., Frauenfelder, T., and Boss, A. Diagnostic performance of machine learning applied to texture analysis-derived features for breast lesion characterisation at automated breast ultrasound: a pilot study. *Eur. Radiol. Exp.* **3**, (2019).

112. Ying, X. An Overview of Overfitting and its Solutions. *J. Phys. Conf. Ser.* **1168**, (2019).

113. Hua, J., Xiong, Z., Lowey, J., Suh, E., and Dougherty, E.R. Optimal number of features as a function of sample size for various classification rules. *Bioinformatics* **21**, 1509–1515 (2004).

114. Bekkar, M., Djemaa, H.K., and Alitouche, T.A. Evaluation Measures for Models Assessment over Imbalanced Data Sets. *J. Inf. Eng. Appl.* **3**, 27–38 (2013).

115. Xue, J.H. and Hall, P. Why does rebalancing class-unbalanced data improve AUC for linear discriminant analysis? *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1109–1112 (2015).

116. Chicco, D. and Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 1–13 (2020).

117. Raschka, S. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv* (2018).

118. Rong, S. and Bao-wen, Z. The research of regression model in machine learning field. *MATEC Web Conf.* **176**, 1033 (2018).

119. Chai, T. and Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **7**, 1247–1250 (2014).

120. Tang, X.Y., Huang, M.X., Han, S.Q., Chang, Y., Li, Z.P., Kao, X.M., Chen, Y.Y., Liu, C., Huang, Y. Di, Chen, Y.T., Lei, Z.J., and Chu, X.Y. The Circumferential Resection Margin Is a Prognostic Predictor in Colon Cancer. *Front. Oncol.* **10**, 1–10 (2020).

121. Toyonaga, T., Tanaka, S., Man-I, M., East, J., Ono, W., Nishino, E., Ishida, T., Hoshi, N., Morita, Y., and Azuma, T. Clinical significance of the muscle-retracting sign during colorectal endoscopic submucosal dissection. *Endosc. Int. open* **3**, E246–E251 (2015).

122. Lam, D., Kaneko, Y., Scarlett, A., D'Souza, B., Norris, R., and Woods, R. The Effect of Formalin Fixation on Resection Margins in Colorectal Cancer. *Int. J. Surg. Pathol.* **27**, 700–705 (2019).

123. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 2274–2282 (2012).

124. Zhang, D., Xu, M., Quan, L., Yang, Y., Qin, Q., and Zhu, W. Segmentation of tumor ultrasound image in HIFU therapy based on texture and boundary encoding. *Phys. Med. Biol.* **60**, 1807–1830 (2015).

125. Daoud, M.I., Atallah, A.A., Awwad, F., Al-Najjar, M., and Alazrai, R. Automatic superpixel-based segmentation method for breast ultrasound images. *Expert Syst. Appl.* **121**, 78–96 (2019).

126. Ramírez-Gallego, S., Lastra, I., Martínez-Rego, D., Bolón-Canedo, V., Benítez, J.M., Herrera, F., and Alonso-Betanzos, A. Fast-mRMR: Fast Minimum Redundancy Maximum Relevance Algorithm for High-Dimensional Big Data. *Int. J. Intell. Syst.* **32**, 134–152 (2017).

127. Sharifai, G.A. and Zainol, Z. Feature selection for high-dimensional and imbalanced biomedical data based on robust correlation based redundancy and binary grasshopper optimization algorithm. *Genes (Basel)*. **11**, 1–26 (2020).

128. The MathWorks, I. Sequential Forward Feature Selection. in *Statistics and Machine Learning Toolbox USer's Guide* 15–61 (2020).

129. Marcano-Cedeño, A., Quintanilla-Domínguez, J., Cortina-Januchs, M.G., and Andina, D. Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network. *IECON Proc. (Industrial Electron. Conf.* 2845–2850 (2010) doi:10.1109/IECON.2010.5675075.

130. Beynon, J., Foy, D.M., Roe, A.M., Temple, L.N., and Mortensen, N.J. Endoluminal ultrasound in the assessment of local invasion in rectal cancer. *Br. J. Surg.* **73**, 474–477 (1986).

131. Tanis, E., Spliethoff, J.W., Evers, D.J., Langhout, G.C., Snaebjornsson, P., Prevoo, W., Hendriks, B.H.W., and Ruers, T.J.M. Real-time in vivo assessment of radiofrequency ablation of human colorectal liver metastases using diffuse reflectance spectroscopy. *Eur. J. Surg. Oncol.* **42**, 251–259 (2016).

132. Geldof, F., Dashtbozorg, Behdad, Hendriks, B.H.W., and Sterenborg, Henricus J.C.M. Ruers, T.J.M. Combining diffuse reflectance spectroscopy and ultrasound for the assessment of multilayer tissue structures. (2021).[In preparation]

133. Wang, K., Lu, X., Zhou, H., Gao, Y., Zheng, J., Tong, M., Wu, C., Liu, C., Huang, L., Jiang, T., Meng, F., Lu, Y., Ai, H., Xie, X.Y., Yin, L.P., Liang, P., Tian, J., and Zheng, R. Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: A prospective multicentre study. *Gut* **68**, 729–741 (2019).

134. Sato, S., Kato, T., and Tanaka, J.I. Defining the distal margin of rectal cancer for surgical planning. *J. Gastrointest. Oncol.* **8**, 194–198 (2017).

135. Pavan, T.Z., Madsen, E.L., Frank, G.R., O. Carneiro, A.A., and Hall, T.J. Nonlinear elastic behavior of phantom materials for elastography. *Phys Med Biol.* **55**, 2679–2692 (2010).

136. Cao, R., Huang, Z., Varghese, T., and Nabi, G. Tissue mimicking materials for the detection of prostate cancer using shear wave elastography: A validation study. *Med. Phys.* **40**, 1–9 (2013).

137. Hall, T.J., Bilgen, M., Insana, M.F., and Krouskop, T.A. Phantom materials for elastography. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **44**, 1355–1365 (1997).

A. Appendix A: Phantom study ultrasound elastography

A.1 Introduction

The phantom study has been performed to acquire more insight into the practical applications of ultrasound elastography in an ex vivo setting. This study serves as a preparation for an ex vivo study with ultrasound elastography. The goal of this ex vivo study is to investigate whether ultrasound elastography helps to distinguish colorectal cancer from healthy colorectal tissue and fibrosis in surgical specimens. USE can be used to visualize the stiffness of soft tissues as a colored map, and to show the anatomy in a B-mode US image. The two techniques of USE are shear wave elastography and strain elastography. For the current approach, shear wave elastography will be used since it generates quantitative stiffness values. Therefore, features can be extracted that can be used as input for a classification algorithm to distinguish cancer from healthy tissue and fibrosis. The strain elastography just shows the relative stiffness of different tissue layers within one image. However, strain elastography may be used to visualize the boundaries between different tissue types. Despite the ex vivo study will focus on shear wave elastography, this phantom study will also look to the additional value of strain elastography.

The ultrasound system that is used to perform the USE measurements is developed for measurements on the human body. Therefore, the ex vivo setup may influence the quality of the ultrasound elastography images. The goal of this study is to investigate which measurement setup adjustments need to be made to obtain good quality ultrasound elastography measurements. Furthermore, the phantom study was performed to get used to ultrasound elastography measurements for the researchers.

To simulate a colorectal specimen, a phantom was made that consisted of multiple layers with different stiffness. Other studies that performed phantom studies to assess shear wave elastography, showed that phantoms of mixtures of agar and gelatin gels were suitable.[135–137] Through varying the agar concentration, phantom layers with a different stiffness can be obtained. The gelatin concentration should be the same in each layer to minimize osmotic pressure differences, and it helps to bond the inclusions to the background.[135]

A.2 Data acquisition

Two different phantoms were prepared, phantom 1 and 2. Phantom 1 consisted of two layers consisting of gelatin- and agar-solution mixtures of which the concentrations and estimated stiffness are shown in Table **A-1**. Phantom 2 consisted of four layers with gelatin- and agar-solution mixtures of which the concentrations are shown in Table **A-2**. Both agar and gelatin solutions were prepared separately, following the method described by Pavan (2010).[135] The preparation of the agar/water solutions started with mixing agar and water at room temperature. Thereafter, the mixture was heated in a microwave. The mixture was stirred when bubbles came up. This process was repeated until it became transparent. The 7.3% gelatin/water solution was mixed at room temperature. Thereafter, the mixture was heated in a pan until it became transparent. The molten agar and gelatin solutions were mixed at a temperature of about 90 degrees. The proportion was 40% gelatin solution and 60% agar solution. The mixture was cooled for some seconds at room temperature, and 0.047g formalin was added for each dry weight of gelatin (formalin concentration 1.017 g/cm^3). Thereafter, the mixture was poured into a box with a size of 13 x 13 x 15 cm (length x width x height). The box was placed in the fridge at 6 degrees Celsius. The box was placed at an angle after adding the second (in phantom 1 and 2) and fourth layers (in phantom 2) to create diagonal layers. The preparation process was repeated per layer, so each layer could be solidified before the next layer was added. Furthermore, different pigments were added per layer, to make the individual layers visually distinguishable. The resulting phantoms are shown in Figure **A-1**. Thereafter, ultrasound elastography measurements were performed 12-20h after preparation.

Table A-1: The two layers of phantom 1 with an estimated Youngs modulus, gelatin-, and agar concentrations

Layer number	*Estimated Young's modulus (kPa)* [135]	*Volume* [a] *(ml)*	*Gelatin mixture concentration (%)*	*Agar mixture concentration (%)*
1 (bottom layer)	79.6	400	7.3	2.27
2 (top layer	20	200	7.3	1.15

[a] *Volume is the total volume of one layer, after the combination of the gelatin and agar concentration*

Table A-2: The four layers of phantom 2 with an estimated Young's modulus, gelatin-, and agar concentrations

Layer number	**Estimated Young's modulus (kPa)** [135]	**Volume (ml)**	**Gelatin mixture concentration (%)**	**Agar mixture concentration (%)**
1(bottom layer)	79.6	400	7.3	2.27
2	47.5	200	7.3	1.71
3	30.2	200	7.3	Not included
4(top layer)	20	200	7.3	1.15

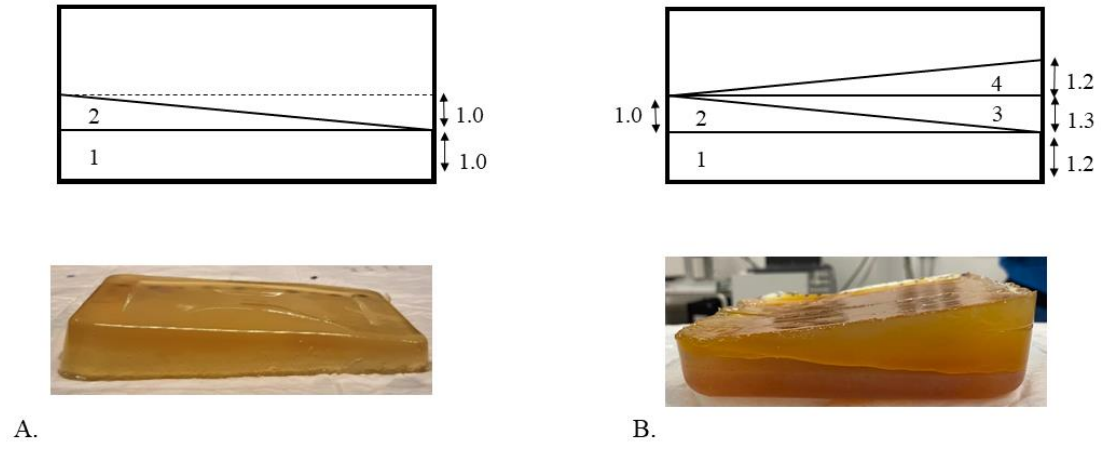
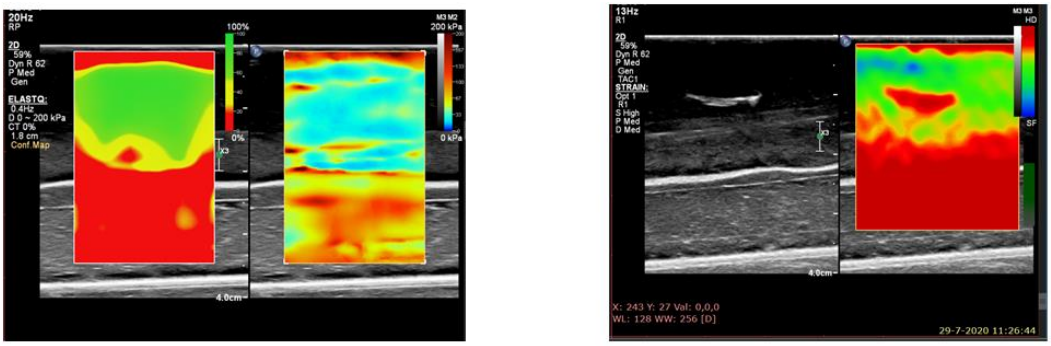


Figure A-1: Illustrations and photos of phantom 1 (A) and 2 (B). The thickness of each layer is shown in centimeters.

A.3 Ultrasound system

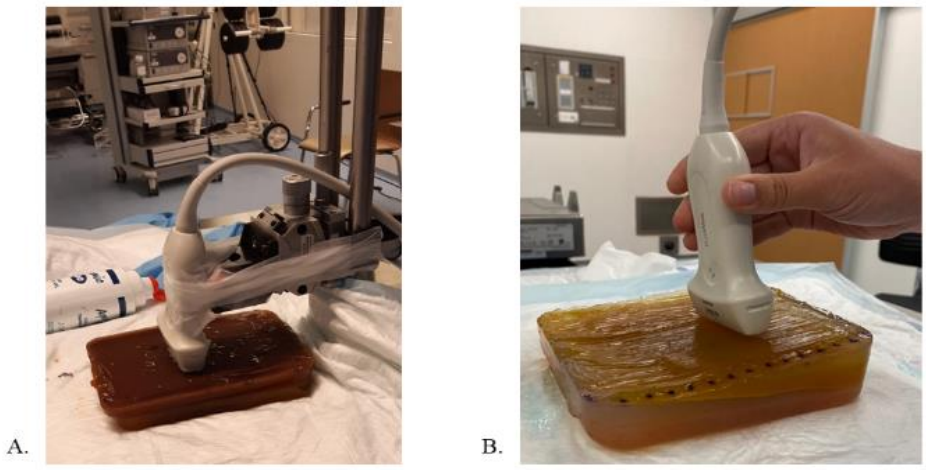
The Philips EPIQ 7 Ultrasound system with the eL18-4 US transducer was used to obtain shear wave elastography and strain elastography measurements. This transducer has pulse frequencies between 22 and 2 MHz. The shear wave elastography images were retrieved with the confidence map, as shown in Figure **A-2** A. The confidence map gives insight into the quality of shear wave measurements. The color map was superimposed on the US B-mode image next to the shear wave elastography map with the colors red, yellow, and green. The red color was shown when the system was not able to detect shear waves, yellow when the system detected a few shear waves, and green when the system detected enough shear waves. The confidence map was used to show the influence of measurement setup adjustments on the quality of shear wave elastography images. Furthermore, strain elastography images were retrieved from the same locations as the shear wave elastography images, as shown in Figure **A-2** B.



A. B.

Figure A-2: Resulting images from the phantom studies. A. Two similar B-mode ultrasound images with different color maps as an overlay. The confidence map is shown on the left, with green areas showing that the US system was able to detect all shear waves, with yellow for the areas from which the US system could detect a few shear waves and the red area where the system could not detect shear waves. The shear wave elastography image is shown on the right, and the colors represent stiffness values that are shown in the color bar. B. The same B-mode image as shown in A. but now with strain elastography. The red areas represent areas that are stiffer relative to the green-colored areas, and the blue area represents tissue that is softer than the green-colored tissues.

Multiple measurement setups have been used to examine their influence on elastography measurements. Firstly, the transducer that was fixated in a laboratory stand was compared to holding the transducer manually. This was investigated since the transducer needed to be held stable for some seconds to obtain good quality shear wave measurements. Despite the manual method is more flexible, the method with the laboratory stand may be better for stability. This setup is shown in Figure **A-3**. Secondly, the measurements of both phantoms were compared to investigate the influence of the phantom thickness on the shear wave elastography measurements. Phantom 1 had a thickness of 2 cm, and phantom 2 had a thickness of 4.7 cm. The reason for this experiment was that most of the ex vivo specimens were 1.5-2 cm thick, whereas the measurement depth of the ultrasound elastography system was fixated on 4 cm. Therefore, the question was whether the absence of tissue within the measurement volume would influence the quality of the shear wave elastography. Thirdly, the underground of a metal table was compared to the underground of a plastic table. Despite metal tables are generally used for ex vivo measurements, metal tables may deteriorate the shear waves. Therefore, a plastic table was compared to a metal table to investigate its influence on the shear wave elastography images.



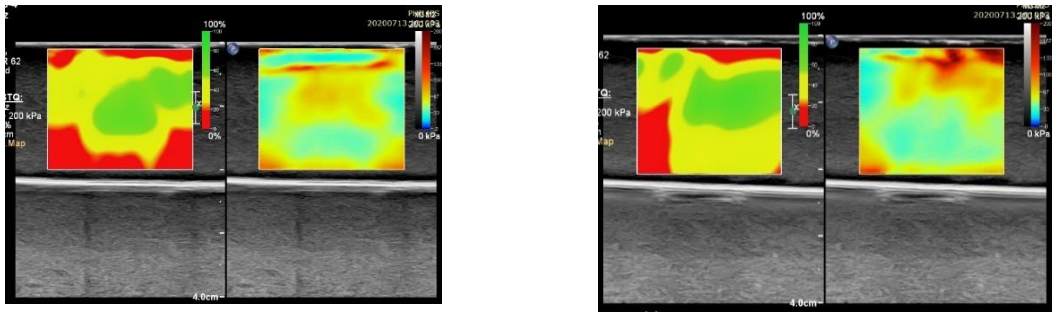
A. B.

Figure A-3: eL18-4 US transducer fixated in a laboratory stand versus manual.

A.4 Results

The resulting shear wave elastography images with the confidence maps are shown per measurement setup adjustment. Thereafter, the shear wave elastography images are compared to the strain elastography images.

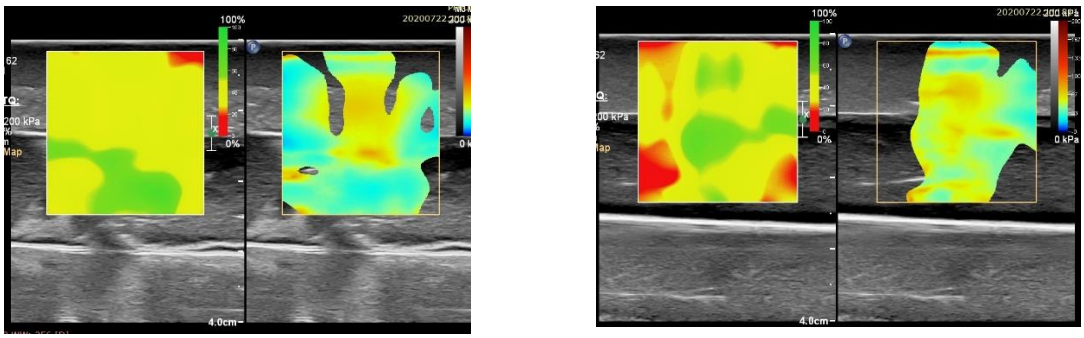
The influence of holding the transducer manually or in a laboratory stand was compared with the confidence maps, as shown in Figure **A-4**. The confidence map shows larger unconfident (red-colored) regions for the transducer in a holder than the transducer that is held manually. This was the case for multiple images at different locations on the phantom.



A. B.

Figure A-4: Influence of the transducer stability by hand or by a standard on the image quality at the same location on phantom 1. A. The US transducer is placed in a standard. The confidence map shows unconfident regions at the bottom and the top of the colored part. B. The US transducer is held manually. The confidence map shows an unconfident region at the left bottom that is largely following the confidence map in image A. furthermore, an uncertain region is shown at the top of the confidence map, although the region is smaller than the unconfident region in image A.

The metal table was in the region of shear wave measurements negatively influenced the quality of the ultrasound elastography measurement in phantom 1. The image with the plastic table showed more signal than the metal table with a confidence level of 35%, as shown in Figure **A-5**. However, the effect of the metal table was not present during measurements with phantom 2.



A. B.

Figure A-5: Influence of the table beneath the phantom. A. plastic table, the confidence map is mainly yellow and green, and the elastography map is almost full with a minimum confidence level of 35%. B. metal table beneath the phantom. There are more red spots, showing the noisy areas where the shear waves cannot be measured. The elastography map shows fewer signals with the same confidence level.

Generally, the strain elastography showed clear layers that were corresponding with the layers on the B-mode image, in contrast to shear wave elastography. One remarkable thing was that a lesion was detected in both US B-mode and strain elastography images, as shown in Figure **A-6**. It was shown as a white stripe in the top layer in the B-mode, and the red area in the green area in the strain map. When the phantom was sliced, this lesion turned out to be a stiffer piece in the phantom. The shear wave image does not show another color in the lesion area.

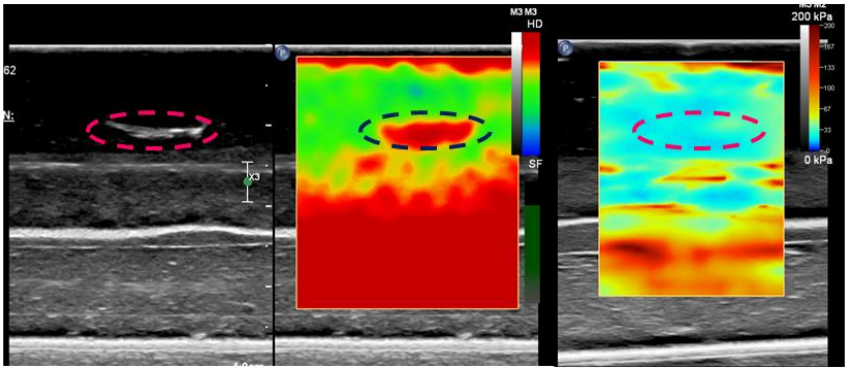


Figure A-6: Lesion (dashed delineation) detected with ultrasound B-mode and strain elastography, but not with shear wave elastography

A.5 Discussion

The goal of this study is to investigate which measurement setup adjustments need to be made to obtain good quality ultrasound elastography measurements.

The phantom study showed that the table within the measurement depth of 4 cm, influenced the elastography measurements. This can be caused by the boundary reflections that disturb the shear waves and therefore the movement of the shear waves cannot be tracked by the US transducer in this area. The problem was solved when using a plastic table beneath the phantom or using a phantom with at least a thickness of 4 cm (phantom 2). Furthermore, the quality of the ex vivo measurements improved as well when some pads were put beneath the specimen in case the specimen was less than 4 cm in thickness. This observation also highlights a limitation of the ultrasound system, as the measurement depth could not be decreased to less than 4 cm.

Another observation from the phantom study was that when the transducer was placed in a laboratory stand, it did not improve the elastography quality compared to when the transducer was held manually. One reason for this is that it was harder to ensure an optimal contact surface of the transducer with the phantom when the transducer was placed in the laboratory stand. Furthermore, the manual method was far more flexible in replacing the transducer. However, it is important to keep the transducer stable during the elastography measurements, since it takes ~2 seconds to acquire a shear wave elastography map. When the transducer is slightly moved within this period, the quality of the shear wave measurement decreases. Therefore, manually holding the transducer during shear wave elastography measurements is preferred when someone can hold it stable.

The strain elastography showed clear layers, in contrast to shear wave elastography. However, since strain elastography images contain only three colors, only three layers with a different stiffness can be discriminated. Furthermore, the strain elastography gives only qualitative information, therefore, it is unknown how ‘stiff’ a stiffer red region is. It is possible to obtain semi-quantitative features, such as the 5-score system and the strain ratio. However, the 5-score system is developed for breast cancer tumors and is not investigated for colorectal cancer yet. The strain ratio may be used, but then it is needed to determine the region of interest during the strain elastography measurements. In that case, the region needs to be determined based on suspicious areas on the B-mode image. This is a disadvantage relative to shear wave elastography. For these images, quantitative stiffness values can be obtained per pixel. Therefore, the interesting regions can be determined afterward based on histopathology results. Another observation was that a stiff lesion was detected by the US B-mode and the strain elastography. This is following the theory that strain elastography is more suitable to detect focal lesions than shear wave elastography.[81]

In summary, the following conclusions can be drawn from the phantom study

- Performing elastography measurements by holding the transducer manually is preferable to stabilize the transducer in a laboratory standard
- A plastic table is in favor of a metal table for shear wave measurements, since a plastic table gives less deterioration than a metal table. When a plastic table is not available, a surgical pad beneath the phantom, such that the metal table is out of the measurement range, is also sufficient.
- The strain elastography shows clear boundaries between layers of different stiffness, where the shear wave elastography does not. However, the shear wave elastography results in quantitative stiffness values per pixel.

A.6 Conclusion

The phantom study was performed to prepare the measurement of the ex vivo setting with ultrasound elastography. It showed that the metal table deteriorates the shear wave measurements. Furthermore, holding the transducer manually is preferred to fixation of the transducer into a stand. Lastly, the strain elastography shows clearer boundaries between tissue layers with different stiffness, but only three different layers can be distinguished. The phantom study leads to the following adjustments of the measurement protocol of the ex vivo study.

- The metal table needs to be out of range for the shear waves, therefore, a surgical pad is placed beneath the specimen in case the specimen was thinner than the measurement depth of 4 cm.
- One researcher with a stable hand performs the US elastography measurements.
- Shear wave elastography is used to assess the results of the ex vivo study since it gives quantitative values, but strain elastography data may be collected as well since it gives added value in showing the boundaries between tumor and other tissue types.

The next chapter demonstrates an ex vivo study is performed to investigate whether ultrasound B-mode and shear wave elastography can distinguish colorectal tumor from fat tissue, the colorectal wall, and fibrosis. The adjustments of the measurement protocol from this phantom study are taken into account in the ex vivo study.