FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

In-vivo prostate cancer predictions in multi-parametric MRI using oversegmentation

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medizinische Informatik

eingereicht von

Dominik Hofer, BSc
Matrikelnummer 0925749

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig
Mitwirkung: Matthew D. DiFranco, PhD

Wien, 21. Juli 2017

Dominik Hofer	Robert Sablatnig

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

In-vivo prostate cancer predictions in multi-parametric MRI using oversegmentation

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Medical Informatics

by

Dominik Hofer, BSc
Registration Number 0925749

to the Faculty of Informatics

at the TU Wien

Advisor: Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig
Assistance: Matthew D. DiFranco, PhD

Vienna, 21st July, 2017

Dominik Hofer	Robert Sablatnig

Erklärung zur Verfassung der Arbeit

Dominik Hofer, BSc
Lerchenfelder Straße 69/29, 1070 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 21. Juli 2017

Dominik Hofer

Danksagung

Ich würde gerne meinem Diplomarbeits-Betreuer Prof. Robert Sablatnig für seine fachliche Unterstützung und Aufsicht danken sowie meinem zweiten Betreuer Matthew D. DiFranco meine Dankbarkeit aussprechen, der mich über die Jahre mit seinem Fachwissen und mit wertvollen Diskussionen unterstützt und begleitet hat. Des Weiteren möchte ich mich bei Alessio Montuoro für seine Hilfe bei diesem Projekt bedanken. Außerdem bedanke ich mich bei Peter Kuess und den Abteilungen für Radiologie und Radioonkologie von der Medizinischen Universität Wien für die Bereitstellung von Daten.

Meine tiefe Dankbarkeit gebührt meinen Eltern für ihre Liebe, Unterstützung und Ermutigung. Danken möchte ich außerdem meinem Partner und meinen Freunden, die mich unterstützt und immer wieder ermutigt haben. Diese Arbeit wäre ohne euch nie möglich gewesen. Danke.

Acknowledgements

I would like to thank my thesis advisor Prof. Robert Sablatnig for his supervision and guidance and my co-supervisor Matthew D. DiFranco for valuable discussions and continuous support over the years. Many thanks go to Alessio Montuoro for his help throughout this project. I would also like to thank Peter Kuess and the Departments of Radiology amd Radiation Oncology of the Medical University of Vienna for providing data.

I would like to express my very profound gratitude to my parents for their love, support and encouragement. I am also grateful to my partner and my friends for consistently supporting and encouraging me. This accomplishment would not have been possible without you. Thank you.

Kurzfassung

Obwohl Dank einer früheren Erkennung und einer besseren Behandlung die Mortalitätsrate gesenkt werden konnte, bleibt der Prostatakrebs die häufigste Krebsart bei Männern in den Industrieländern. Die multiparametrische Magnetresonanztomographie wird zunehmend im klinischen Bereich zur Erkennung von Prostatakrebs eingesetzt. In den letzten 15 Jahren kam die rechnerunterstützte automatische Prostatakrebserkennung in den Fokus der medizinischen Bildverarbeitung um Radiologen bei ihrer Entscheidungsfindung zu helfen. Diese Frameworks detektieren normalerweise die Position des Tumors anhand von berechneten Merkmalen oder geben eine Diagnose über eine Region in der Prostata. Diese Merkmale beruhen entweder auf pixel-basierten Berechnungen oder benötigen manuell annotierte Bereiche, auf deren Basis statistische Berechnungen durchgeführt werden. In dieser Arbeit wird ein Framework vorgestellt, das diese Bereiche mit Hilfe eines multi-modalen Superpixel-Algorithmus automatisch generiert und die Prostata in verschiedene Bereiche unterteilt. Diese Bereiche werden für die Berechnung von aussagekräftigeren, statistischen Merkmalen verwendet. Die Resultate des Frameworks werden anhand von zwei Datensätzen berechnet. Der erste Datensatz besteht aus multiparametrischen MRT-Scans von 20 Patienten, während der zweite Datensatz aus Scans von 25 Patienten besteht. Bei beiden Datensätzen wurden die Prostata, deren anatomische Zonen und der Prostatakrebs annotiert. Die Genauigkeit der Klassifizierung wurde mit der Receiver Operating Characteristic Kurve berechnet. Die durchschnittliche Area Under the Curve beim ersten Datensatz beträgt 0.87 mit einer Standardabweichung von 0.08. Beim zweiten Datensatz wurde ein Wert von 0.59 mit einer Standardabweichung von 0.11 gemessen. Das Framework zeigt bessere Resultate als vergleichbare Systeme und beweist dass Superpixel die Genauigkeit von rechnerunterstützten Systemen für die Erkennung von Prostatakrebs erhöhen können (von 0.85 zu 0.87 beim ersten und von 0.55 zu 0.59 beim zweiten Datensatz).

Abstract

Despite its steady reduction in mortality due to early detection and improved treatment, prostate cancer remains the most common cancer form in men in the developed countries. Multi-parametric magnetic resonance imaging is gaining clinical relevance and is increasingly used to diagnose prostate cancer. In the last 15 years, computer-aided detection systems that aid the radiologists in their clinical decision making have come into focus of medical image analysis. These frameworks normally detect cancer by computing pixel-based features or compute region-based features and give a diagnosis about a region of interest that was manually annotated. In this thesis, we propose a computer-aided detection system that automatically segments the prostate into specific regions of interest without the need for manual annotation. By incorporating a multi-modal, superpixel-based oversegmentation of the prostate into our framework, more accurate region-based features can be calculated. The system is evaluated on two datasets. The first dataset consists of multi-modal MRI scans of 20 patients of which 18 have biopsy-proven prostate cancer. The second dataset has multi-modal MRI scans of 25 patients. In both datasets, the prostate boundary, prostate zones and cancer lesions were annotated by experienced radiologists. Performance evaluation is based on receiver operating characteristic curve. The average area under the curve is 0.84 with a standard deviation of 0.08 for the first dataset. The second dataset shows an average area under the curve of 0.71 with a standard deviation of 0.11. The framework shows a better performance than comparable computer-aided detection systems in literature and proves that superpixels can improve the classification result for detecting prostate cancer (from 0.85 to 0.87 for the first and from 0.55 to 0.59 for the second dataset).

Contents

CHAPTER 1

Introduction

Prostate Cancer (CaP) accounts for 22% of all cancer diagnoses in Austrian men in 2012 [Sta15], and it is expected that 19% of all new cancer diagnoses for male US-Americans in 2017 will be prostate-related [SMJ17]. Despite its steady reduction in mortality due to early detection and improved treatment, CaP remains the most common cancer form for men in developed nations [FSB+10]. Current clinical practice for the screening of CaP involve Digital Rectal Examinations (DRE) and Prostate Specific Antigen (PSA) blood tests [YVM+12]. Definite diagnosis is usually obtained by analyzing histopathological samples from TransRectal UltraSound (TRUS) guided biopsies, providing the clinician with information on the Gleason Score (GS), a score that grades the malignancy of the cancer [YVM+12].

Multi-parametric Magnetic Resonance Imaging (mpMRI) of the prostate is used to diagnose CaP or assist the radiologists in guiding biopsies due to the complementary information obtained from T2-Weighted (T2W), Diffusion-Weighted (DWI), and Dynamic Contrast Enhanced (DCE) MR imaging acquired during a single diagnostic imaging session [PJY+13, YVM+12, HBH+11]. With the introduction of hybrid Positron-Emission-Tomography (PET)/MR systems, there is potential for including simultaneous molecular imaging sequences for further complementary disease information [LAT+16]. However, obtaining a diagnosis through this complex mpMRI data can be affected by inter- and intra-observer variability [LMF+15, WBT+14]. Therefore, Computer-Aided Detection (CAD) systems are developed to support clinicians in the detection and diagnosis of CaP.

The majority of CAD systems for CaP either rely on pixel-wise features for detection or on manual annotations for diagnosis [LMF+15]. The novel contribution of this thesis is a CAD system that predicts cancer regions of the prostate in mpMRI by computing Region Of Interest (ROI) based statistical features without the need of manual annotation of suspicious regions. This is achieved by implementing a superpixel algorithm into our framework that automatically oversegments the prostate into ROIs that are used for statistical feature calculation. The goal is to improve the state of the art in

mpMRI prostate CAD systems in terms of accuracy (specificity and/or sensitivity), speed, simplicity and generalizibility. The goal of a CAD system is to automate the process of finding CaP and to minimize the need for manual intervention. This thesis shows that oversegmentation eliminates the step of manual segmentation of suspicios regions without sacrificing classification accuracy and that superpixel-based statistical features improve the classification result. There are two other CAD systems that use automatic segmentation of ROIs to improve classification results. Litjens et. al. [LD$^+$14] use a fully-automated CAD system that first calculates a pixel-wise probability map for CaP. In a second step, they perform local maxima detection on the probability map and compute a ROI around each maximum. Finally, each ROI is classified as cancerous or healthy tissue. Vos et. al. [VBKH12] find ROIs by detecting blobs in the ADC map. In the next step, these ROIs are used for the Random Forest (RF) classifier. In comparison, our CAD system does not only find suspicious regions, but also segments the rest of the prostate gland into distinct ROIs. The results of this thesis can be used for other approaches of CAD systems that plan to automate the segmentation step and to implement statistical region-based features without the need of manual annotation of ROIs.

The remainder of this thesis is organized as follows:

- Chapter 2 provides a background on prostate anatomy, CaP, diagnosis techniques, mpMRI, and CAD systems and gives a literature review on related work regarding the computer-aided detection of CaP and superpixel algorithms.
- Chapter 3 explains the structure of the proposed CAD framework, the pre-processing steps, the algorithms that were used for oversegmentation, the image features and the classification model.
- Chapter 4 provides a validation of our CAD framework. The accuracy of the classifier is shown by computing the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. The importance of superpixel-based features in comparison with pixel-wise features and how it improves the CAD system is also explained. Furthermore, we compare the accuracies of different supervoxel parameter sets and examine how they influence the classification outcome.
- Chapter 5 gives a conclusion and summarizes the work and novel contributions of this thesis. Limitations of the CAD systems are discussed and ideas for future research are presented.

CHAPTER 2

Background

In this section, a literature review about the existing CAD approaches for CaP is presented. Generally, there can be distinguished between two types of CaP CAD systems: frameworks that predict the location of possible cancer lesions (CADe systems), and frameworks that give a diagnosis about manually defined ROIs (CADx systems). CADe systems automatically segment possible cancer regions and use these regions as inputs for a CADx. While the output of these two systems is different, the underlying framework is the same. Additionally, an overview of the most important superpixel algorithms is given in this chapter.

2.1 Prostate Anatomy

The prostate gland is a part of the male reproductive system [MTT12, SBL98]. It is located directly below the bladder and in front of the rectum and is slightly larger than a walnut [MTT12, SBL98]. The functions of the prostate include the production of a secretion that acts as a protective and liquid medium for the sperm cells in the ejaculatory fluid and the control of the flow of urine [MTT12, SBL98].

The prostate is divided into zones, as first defined by McNeal et. al. [McN81] (see Figure 2.1).

- The Central Zone (CZ) surrounds the transition zone and constitues approximately 25% of the prostate [McN81].
- The Transition Zone (TZ) surrounds the urethra as it passes through the prostate. It is located between the Peripheral Zone (PZ) and the CZ. With age, the TZ begins to enlarge until it becomes to largest zone of the prostate. The CZ and TZ are called Central Gland (CG) together because they are inseperable on MR images [YVM+12].

- The PZ is the largest part of the prostate (approximately 70% of the whole volume [McN81]). It is located closest to the rectum and can be felt during a Digital Rectal Examination (DRE). Approximately 70% of CaPs are located in the PZ [YVM+12].

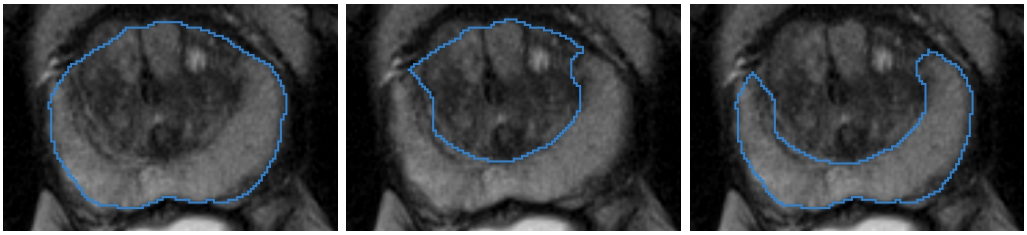


Figure 2.1: Axial slice from a T2-weighted prostate MRI of dataset 2 (see Chapter 3.1). Left: Annotated boundary of the whole prostate gland. Middle: Annotated central gland zone. Right: Annotated peripheral gland zone.

In men aged over 40 years, Benign Prostatic Hyperplasia (BPH) may occur. BPH refers to noncancerous increase of the prostate size due to an increase in the number of cells, mainly in the TZ. It is diagnosed histologically and may cause symptoms like obstruction of the bladder outlet [Tam97].

2.2 Prostate Cancer

CaP is one of the most common cancers in men worldwide. In Austria, it is the most common cancer form in men, where 22% of all new cancer diagnoses in 2012 were CaP and approximately one in ten cancer-related deaths were because of CaP [Sta15].

The diagnosis usually consists of a DRE along with PSA blood tests, followed by TRUS guided biopsies [HBH+11] to obtain histological samples for the final clinical diagnosis.

A DRE is an internal examination of the prostate in which a doctor inserts a lubricated finger into the rectum to palpate the prostate. CaP is detected by checking for an enlarged prostate gland, so smaller, lower-grade cancers may not be diagnosed. Furthermore, only the PZ is palpable with a DRE, so cancers that are located in other areas may be missed. Due to these disadvantages and an overall low sensitivity of 37% [SKR+98], imaging methods or blood tests are used to complement a DRE.

PSA blood tests measure the level of PSA in a patient's blood, which is often elevated in men with CaP [YVM+12]. PSA blood tests have a low specificity of 36% [HBH+11] and may lead to unnecessary biopsy [SKR+98]. When the PSA levels are low (0 to 4.0 ng./ml.), usually a prostate biopsy is not indicated. However, at this level a DRE performs poorly with a positive predictive value of only 9.7% and significant cancers may be missed [SvdCKdK+00].

For screenings, currenct clinical practice is to use a combination of DRE and PSA, although it is controversial that the benefits outweigh the risks of overdiagnosis and

overtreatment [GLT+81, LBN+14]. Followed by elevated PSA blood levels and/or if the doctor felt abnormalities during the DRE, usually an imaging-guided biopsy is done to obtain histopathological samples for analysis. Currently, TRUS is the standard approach for imaging-guided biopsies [YVM+12]. During this procedure, random systematic samples are taken from the prostate with ultrasound as a guidance to locate the prostate gland. With this technique, prostate samples are taken at different locations in the prostate according to a specific protocol [YVM+12]. TRUS is not used for a targeted biopsy of cancer regions because most prostate tumors are either not visualized with this technique or are undistuingishable from other tissue [YVM+12]. After the biopsy, the histopathological samples are analyzed to obtain the GS [HBH+11]. The disadvantages of this method are low sensitivity for clinically relevant cancers and overdiagnosis of small, low grade cancers. The reason of this is, due to its random sampling, there is a chance of missing relevant cancer leasons and detecting microfocal, clinically irrelevant cancers [YVM+12]. Therefore, research has focused on other imaging techniques to examine the usefulness for targeted biopsies [YVM+12]. MRI-directed prostate biopsy uses previously acquired MR imgages for an additional guidance for TRUS biopsies and yields better results than random systematic samples because it allows to target cancer lesions [YVM+12]. MRI-guided prostate biopsy uses real-time MR imaging as a guidance during the biopsy [YVM+12]. The third option is fused MR imaging- and TRUS-guided biopsy, where MR images are taken previously and are then co-registered with real-time TRUS images [YVM+12]. Targeted biopsies show improved results compared to non-targeted, random systematic samples [PEND+16]. PET/MRI is another promising method for guiding biopsies due to its high sensitivity and specificity [LAT+16] and improves the identification of CaP compared to mpMRI alone [PMK+16].

Because these diagnostic methods have their limitations, new imaging methods such as mpMRI gain clinical relevance and are increasingly used for localizing and detecting CaP [BRC+12].

2.3 Magnetic Resonance Imaging

MRI is a non-invasive medical imaging technique to obtain images of the inside of the body. The MRI scanner uses strong magnetic fields and receiver coils for imaging. When the patient is placed inside the scanner, the hydrogen protons within the patient's body align to the static magnetic field. In the next step, this alignment is disturbed by emitting a radio frequency pulse from the scanner, causing some hydrogen protons to get out of alignment with the static magnet field. As the nuclei return to their initial alignment, they emit the radio frequency energy which can be measured by the MRI scanner with receiver coils.

In the context of CaP, MR imaging provides functional tissue information along with anatomic information. It is the best currently available imaging method for CaP due to its high soft-tissue contrast and high resolution [FBV+15]. MpMRI, the combination of anatomic T2-weighted imaging with functional imaging such as DW or DCE imaging, is

used to obtain complementary information and to increase accuracy [HBH+11]. T2W, DWI and DCE MR images have shown to correlate with cancer lesions [WBT+14]. In a diagnostic meta-analysis by Rooji et. al. [dRHF+14], the specificity and sensitivity for the detection of CaP with mpMRI that uses a combination of T2W, DWI and DCE MR images are 0.74 and 0.88. MpMRI has such a high accuracy that it is able to detect clinically significant CaP and may therefore be used to target prostate biopsies [FBV+15], assist the clinicians in their decision-making and improve the diagnostic yield. It is recommended by the European Society of Urogenital Radiology (ESUR) that a combination of T2W MRI with at least two other modalities is used for the detection of CaP [dRHF+14].

2.3.1 T2-weighted MRI

T2W MRI relies on measuring the transverse (spin-spin) relaxation time of hydrogen protons in a patient's body. This modality has a high spatial resolution (0.676x0.676mm and 0.625x0.625mm per pixel for the datasets used in this thesis) and can therefore be used for differentiating between the zones of the prostate [BRC+12].

On T2W images, tumor tissue in the PZ can show as an area with lower intensity (hypointensity) compared to the values of non-tumor tissue. However, low-intensity ROIs can also represent benign abnormalities such as chronic prostatitis, atrophy, scars, postirradiation or hormonal treatment effects, hyperplasia or postbiopsy hemorrhage [HBH+11]. Furthermore, cancer in the PZ has different tissue characteristics than cancer that is located in the CG, which makes the detection of CaP in the CZ more difficult [BRC+12]. CaP in the PZ is easier to detect than cancer in the CG, because other tissue types in the CG may have similar intensities in T2W MRI [HBH+11].

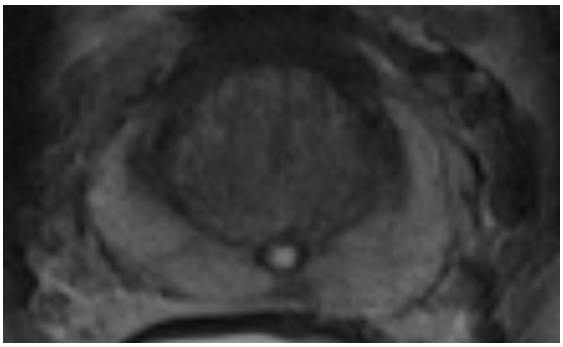


Figure 2.2: Axial slice of a T2-weighted MR image of dataset 2 (see Chapter 3.1). The anatomical regions (central gland, peripheral zone) can be distuingished.

2.3.2 Diffusion-weighted MRI

DW MRI uses the diffusion of water molecules to produce image contrast. Typically, images with varying contrast can be generated by applying different b-values (the degree

of applied diffusion weighting). These DW MR images can be used to calculate the Apparent Diffusion Coefficient (ADC) map, which describes carpillary perfusion and diffusion characteristics in the tissue [HBH+11]. The amount of water diffusion is shown as contrast in the ADC map, and different cell densities in the tissues are characterized by different contrast values. Tumor regions tend to have higher cell density which leads to lower water diffusion. Thus, lower ADC values are an image-based marker for possible tumor or other dense tissue and the ADC map can be used to detect cancer lesions because they show as a darker region compared to normal PZ tissue [HBH+11]. Furthermore, it has been shown that the ADC map may correlate with GS [LMF+15]. The advantages of DW MRI are that it is a readily available and non-invasive. However, it is sensitive to motion artifacts and has a lower spatial resolution than T2W MRI [HBH+11].

Figure 2.3: Example of a diffusion-weighted MR image of dataset 2 (see Chapter 3.1), in particular the ADC map.

2.3.3 Dynamic Contrast-Enhanced MRI

DCE MRI measures the vascular characteristics of tissue by exploiting the dynamic uptake and rapid washout of a contrast agent (usually gadolinium-based) [HBH+11]. It consists of a series of fast T1-weighted MR images that obtain pharmacokinetic features by measuring the prostate intensity values at certain timepoints before, during, and after the injection [HBH+11]. Subsequently, an intensity enhancement diagram over time can be generated and analyzed (see Figure 2.4). In this diagram, cancer lesions can be differentiated from normal tissue because they have an earlier, faster and greater enhancement and earlier wash-out than normal tissue [HBH+11]. More precisely, these pharmacokinetic parameters can be calculated qualitatively, semi-quantitatively and quantitatively [HBH+11]. Qualitative parameters are calculated by examining the shape of the time curve of the intensities of a voxel [HBH+11]. Semi-quantitative parameters characterize the shape of the diagram curve using mathematical modeling (e.g., the integral area of the time curve, maximum signal intensity, and time-to-peak enhancement, among others [HBH+11], Lemaitre20158). Quantitative parameters (e.g., kTrans) are used to compute maps of the distribution of the contrast agent. kTrans is the transfer constant from blood plasma into the extracellular extravascular space (see Figure 2.5).

DCE MRI is a functional MR imaging technique with a high sensitivity [HBH+11] for detecting CaP. It is an invasive technique as it requires the injection of a contrast agent. Another disadvantage is that tumor tissue may be hard to differentiate from prostatitis in the PZ and BPH in the CG [LMF+15].

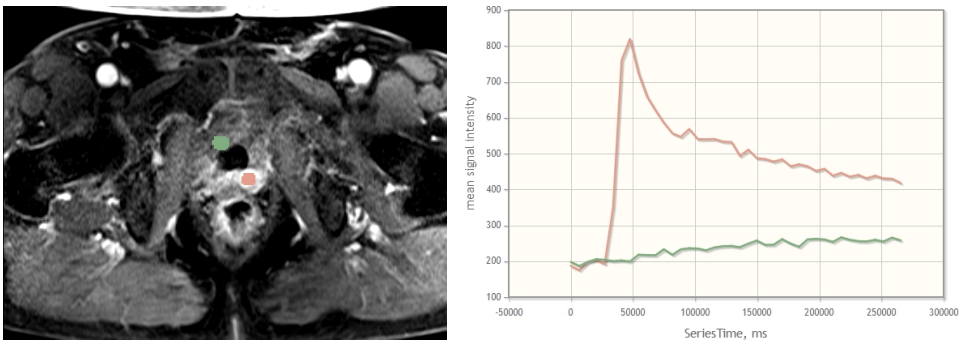


Figure 2.4: Example of an intensity enhancement diagram. The red line in the diagram (right) shows the intensity enhancement over time of the red annotated cancer region in the DCE MRI (left). The green line in the diagram shows the intensity enhancement for the green annotated normal tissue in the DCE MRI. The cancer region shows an earlier, faster and greater enhancement. Image taken from dataset 1 (see Chapter 3.1).

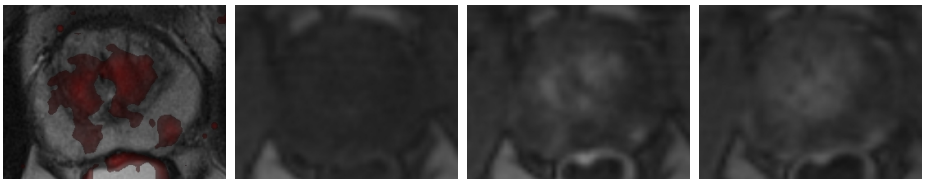


Figure 2.5: Examples of DCE MRI and pharmacokinetic parameter maps. From left to right: kTrans map overlayed on T2W MRI, DCE pre phase, DCE early phase, DCE late phase. Image taken from dataset 2 (see Chapter 3.1).

2.3.4 MR Spectroscopy

MR Spectroscopy Imaging (MRSI) is a Nuclear Magnetic Resonance (NMR) based technique that is used to examine the presence of various metabolite concentrations in a tissue [LMF+15]. It is helpful for detecting cancerous tissue in the prostate because CaP has other concentrations of metabolites than normal tissue [LMF+15]. Concretely, it has been shown that an increased concentration of choline and a a decreased concentration of citrate and spermine correlate with cancer lesions [LMF+15]. MRSI has a high sensitivity and high specificity and improves the detection of CaP when combined with other MRI modalities [LMF+15]. However, it has a low voxel size and has a high variability between patients [LMF+15].

2.3.5 PET/MRI

PET/MR imaging is a multimodal imaging technology that combines MR imaging with PET imaging in a single simultaneous acquisition. PET is a nuclear and functional imaging method that uses small amounts of radioactive material to provide information about metabolic processes in the body. In the context of CaP diagnosis, it is used to get biological information of the cancer lesions [LAT+16]. The advantages of PET is that it is capable of scanning the whole body for cancer because it is sensitive and highly specific for distant disease [LAT+16]. In hybrid PET/MR imaging, the advantages of both modalities are combined into a single scanner. By combining the two protocols, the detection of CaP may be improved because of the higher specificity of PET combined with the anatomic and functional information provided by MRI [LAT+16]. However, PET/MRI has several disadvantages including high cost, long acquisition protocols [LAT+16] small bore size of the scanner (limiting how big a patient can be) and the MR-based attenuation correction, which is inferior to CT-based attenuation correction.

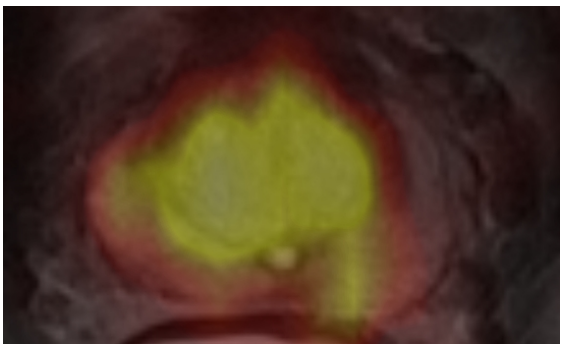


Figure 2.6: Example of a PET image overlayed on a T2W MRI. Image taken from dataset 2 (see Chapter 3.1)

2.4 Computer-Aided Detection Systems

Interpreting mpMRI images for the diagnosis of CaP requires expertise and the results have a low repeatability due to inter-observer variation [LMF+15, WBT+14]. With the research fields of computer vision and machine learning, the idea of developing software to aid the radiologists in their clinical decision making has come into focus of medical image analysis [LMF+15]. CAD systems help the radiologists in the detection and diagnosis of CaP lesions, reduce reading time and required expertise. There are two types of CAD systems: Computer-Aided Diagnosis (CADx) systems help the radiologists to diagnose CaP in manually defined ROIs and Computer-Aided Detection (CADe) systems highlight regions in the prostate that are suspected to have cancer. The CAD system in this thesis is a CADe system.

The typical workflow of a CAD system consists of loading and pre-processing the data, feature extraction and classification. If necessary, registration and/or segmentation can

be performed before the feature extraction. A detailed literature review of CAD systems can be found in Chapter 2.5.

2.5 Related Work

This section gives an overview about other approaches in literature that implement CAD systems for CaP and the most common superpixel algorithms.

2.5.1 CAD Systems

Chan et. al. [CWM+03] were the first to implement a CAD system for CaP. They use T2W, T2-mapping and line scan diffusion imaging and manually segmented tumor regions by a radiologist. Voxel-wise features include textural features that were extracted from a co-occurence matrix and discrete cosine transform and anatomical features using a cylindrical coordinate system. For classification they use a combination of maximum likelihood, Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) which result in a statistical map that indicated cancer probability in the PZ of the prostate. The authors present an average AUC of 0.84 for their best linear discriminant classifier and an average AUC of 0.76 for their best SVM classifier.

Madabhushi et. al. [MFM+05] propose a prediction framework for the detection of prostatic adenocarcinoma from high-resolution ex-vivo MRI. They extract a total of 35 3D features from the images, including first- and second-order statistical features, gradient based features, steerable Gabor filters and discrete cosine transform. The authors obtain a specificity of 98% and a sensitivity between 36% and 42%.

Langer et. al. [LvdKE+09] use T2W, DWI, T2-mapping and DCE images to build a model for prospectively identifying CaP in the PZ. Out of the DWI they calculate ADC maps and from the DCE images they calculate the k-trans and extravascular extracellular volume fraction, which are all used for feature calculation. They evaluate their model on a per-voxel basis and come to the conclusion that the best single parameter is the ADC map which results in an AUC of 0.689 and the best combination of parameters are the ADC map, T2W MRI and k-trans with an AUC of 0.706.

Liu et. al. [LY11] develop an automated CaP localization method by computing a new feature that incorporates the spatial information of the cancer. With this location map they avoid the need of manual PZ extraction, because 70%-80% of prostate tumors occur in the PZ of the prostate. They then use a SVM for classification. The algorithm is tested on 20 patients with biopsy-confirmed CaP and results show an AUC of 0.83 for distuingishing cancer from MR-positive benign lesions and an AUC of 0.89 for cancer vs MR-positive or MR-negative benign lesions.

Ozer et. al. [OHL+09] use pharmacokinetic parameters derived from DCE MRI along with T2W and DWI MRI. As classifiers they use Relevance Vector Machines (RVM) with a Bayesian framework and compared it to a SVM. They report a specificity of 0.78 and a

sensitivity of 0.74. Additionally to their work, Ozer et. al. [OLL+10] propose the use of fuzzy Markov Random Fields (MRF) as an unsupervised classifier and compare it to their previously described RVM and SVM. They report that the supervised methods perform better with an AUC of approximately 0.8 for the fuzzy MRF.

Tiwari et. al. [TVK+12] develop a voxel-wise classifier for MRI and MR spectroscopy. They use manual prostate segmentation and compute Haar wavelet and Gabor features and use a RF for classification. The algorithm is tested on 36 patients and results show an average AUC of 0.89.

Vos et. al. [VBKH12] implement a fully automated two-stage CAD system that uses an initial blob detection approach combined with an automatic ROI segmentation. In the first stage they use a Hessian blob detection at multiple scales for initial voxel classification. In the next step the ROIs are characterized with features that were extracted from histogram analysis. Subsequentially, a Linear Discriminant Analysis (LDA) classifier is trained which yields sensitivities of 0.41, 0.65 and 0.74 at false positive levels of 1, 3 and 5 per patient, respectively.

Niaf et. al. [NRML+12] develop a CAD system for cancer in the PZ of the prostate. The goal is to develop a framework that assists the radiologist by showing the probability of cancer in manually pre-defined ROIs in the PZ. Although the system does not approach the same problem (detection of CaP), the underlying algorithms are the same. The dataset consists of T2W, DWI and DCE images. The authors evaluate four different classifiers (nonlinear SVM, LDA, k-nearest neighbours and naive Bayes classifiers) and extract about 140 different features including grey-level, texture, gradient and functional features. To increase the performance of the classifier and to prevent overfitting, the number of features is reduced. The results show that the t-test feature selection approach combined with a SVM yielded the best result with an AUC of 0.89.

In a subsequent study, Niaf et. al. [NFR+14] propose a generalized version of a SVM, called probabilistic SVM (P-SVM) and show on mpMRI data that the P-SVM yields better classification results than the classic SVM.

Peng et. al. [PJY+13] evaluate the potential utility of a number of parameters from mpMRI consisting of T2W, DWI and DCE imaging. Their CADx system calculates features from manually annotated ROIs by radiologists and shows that a combination of the 10th percentile and the average of the ADC map and T2W skewness is effective for distinguishing CaP foci from normal PZ tissue. The framework yields an AUC of 0.95 (0.93 - 0.97). Furthermore, the authors show that the 10th percentile of the ADC map and kTrans moderately correlate with tumor GS.

A semi-automatic multi-image texture analysis framework is presented by Duda et. al. [DKM+14]. They extract 30 features from the images (T2W, DCE and DWI MRI), including autocorrelation, first order statistics, gradients, fractals, co-occurence matrices and run-length matrices. They report an accuracy of up to 99%, however they only validate one slice per modality.

Litjens et. al. [LDB+14] develop a fully-automated CaP CAD system that consists of two stages. In the first stage, they segment the prostate with a multi-atlas based segmentation approach and predict the probability of CaP for each voxel using intensity, pharmacokinetic, texture, blobness and anatomical voxel-wise features. In the second stage, they perform local maxima detection on the probability map. For each local maximum they segment the surrounding region to get ROIs. Finally, they classify each ROI using the original feature images and additional statistical, local contrast, symmetry and shape features. The result is a probability between 0 and 1 for each ROI, with 0 indicating that there is no CaP and 1 indicating definite CaP. At 0.1, 1 and 10 false positives per normal case the CAD system obtains a sensitivity of 0.42, 0.75 and 0.89. Furthermore, the authors show that the additional step of having a ROI-based classification step improves the result (sensitivity at 1 false positive per normal case improved from 0.55 to 0.75). Beside the framework by Vos et. al., the authors' CAD system is the only other one that incorporates automatically generated ROIs for improving cancer prediction.

Kwak et. al. [KXW+15] propose a CADx system for CaP that utilized two MRI sequences (T2-weighed and high-b-value diffusion weighted imaging) on a total of 244 patients. They extract voxel-wise features based on local binary patterns, local directional derivative patterns and variance measure operator and implement a three-stage feature selection method. They obtain an AUC of 0.83 (0.76 - 0.89) for distinguishing cancer from MR-positive benign tissue and an AUC of 0.89 (0.84 - 0.93) for distinguishing cancer from benign (MR-positive or MR-negative).

Khalvati et. al. [KWH15] present an automated radiomics-based approach for cancer detection using comprehensive texture feature models. In addition to T2W MRI and DWI, the model uses computed high-b DWI and correlated diffusion imaging. The feature set includes first-order statistical features like mean, standard deviation and skewness, second-order statistical features like correlation, variance, and entropy (computed on rectangular pixel neighbourhood), Gabor filters and Kirsch filters. The most important features are selected with feature selection analysis and then the best features for each modality are used for the model. In the next step, the best features for the combined modalities are selected. The authors achieve a maximum AUC of 0.91 at a sensitivity of 0.97 and a specificity of 0.87.

Chang et. al. [CKS+15] use the approach by Khalvati et. al. [KWH15] and introduce a radiomics-driven conditional random field framework for improving the classification result by not only studying the individual voxels, but also taking the inter-voxel spatial and feature relationships into account. Their approach enforces the compactness of the classified regions. The authors improve the result by Khalvati et. al. [KWH15] by 0.55%, leading to an AUC of approximately 0.92.

Rampun et. al. [RTZM16] use a texton-based approach to diagnose CaP in the PZ in T2W MRI. They extract square patches at random locations of the images. Then, they use k-means clustering to aggregate and cluster the textons. Out of all the patches, they build a texton dictionary which is used to generate a texton map for every peripheral

zone in the training set. They extract features out of the texton map and use it to train the classifiers (Bayesian network, random forests and k-NN). The best result is an AUC of 92.8 using the Bayesian network.

2.5.2 Superpixels

Images are represented digitally as a grid of pixels, and each pixel represents a sample of the original image at a certain location. However, a pixel grid does not represent a natural region of an image and, especially in the context of medical imaging, can be subject to noise (e.g., due to magnetic field inhomogeneity [LMF+15]). Superpixel algorithms (or in the three-dimensional case supervoxels) group together neighbouring pixels that share the same low-level features such as color, intensity or location (see Figure 2.7 for some examples). The results are homogeneous clusters of pixels that share the same properties. Superpixels are used in computer vision areas such as classical segmentation [RM03, TLZF16], where they serve as a basis for more sophisticated algorithms, and reduce computation time [TLZF16]. In the context of prostate MR imaging, they are used for the automatic segmentation of the prostate where superpixels instead of pixels are used as the basic processing unit for a graph cut algorithm, significantly reducing the computational cost [TLZF16].

In general, there can be distuingished between graph based methods and gradient ascent methods [ASS+10]. Graph-based methods consider each pixel as a node in a graph, and each edge has a certain weight which represents the similarity between two pixels. Typically, a cost function is defined over the graph and minimized to obtain the superpixels.

Felzenszwalb et. al. [FH04] implement a popular 2D image segmentation algorithm. Each pixel is represented as a node in a graph with edges that connect it to adjacent pixels. Each edge has a certain weight which is represented by the intensity similarities between the pixels. Subsequently, the pixels are grouped by merging pixels with low edge weights into the same region, while the boundaries between regions have high edge weights so that a superpixel represents a minimum spanning tree. The number of superpixels and their compactness cannot be controlled by the user.

Using the originally for classical segmentation intended Normalized Cuts algorithm [SM00], Ren et. al. [RM03] adopt it for another popular superpixel method that uses graph cuts to globally minimize a cost function using contour and texture cues.

Turbopixels by Levinsthein et. al. [AAN+09] is an algorithm based on level-set based geometric flow. Initially, superpixel seeds are placed on a regular grid. In the next step, the superpixels are grown based on an evolving contour. The resulting superpixels have the same size and compactness and adhere to boundaries.

Gradient-ascent-based algorithms start with an initial clustering, followed by iterations of optimizing and refining the superpixel centers and boundaries.

Mean-shift [CM02] is a non-parametric mode-seeking algorithm that finds local maxima in a density function. Each pixel is associated to a mode of the underlying density function. Pixels that are near the same mode in the color or intensity feature space are grouped into superpixels. The resulting superpixels have an irregular shape and a non-uniform size.

Quick-shift [VS08] is similar to Mean-shift as it also uses a mode-seeking algorithm. Each point in the feature space is moved towards a mode to increase the Parzen density estimate. The amount and size of the superpixel cannot be set by the user. The resulting segmentations are comparable to Mean-shift, at a lower computation time.

Simple Linear Iterative Clustering (SLIC) [ASS+12] uses local k-means clustering to obtain the superpixels. The number of superpixels, compactness and sigma can be set by the user. Furthermore it can be used in 2D and 3D. It is described in more detail in Chapter 3.3.1

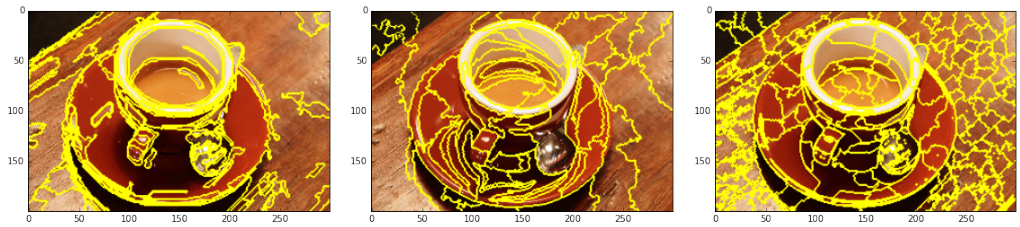


Figure 2.7: Comparison of three different superpixel algorithms. Left: Felzenszwalb's method [FH04]. Middle: SLIC [ASS+12]. Right: Quick-shift [VS08].

2.6 Summary

This chapter provided an overview of the topic and described the anatomy of the prostate, the current clinical practice for detecting CaP, mpMRI and the advantages of CAD systems. It was explained why mpMRI may be used to improve the diagnosis of CaP in comparison to DREs or PSA blood tests. Furthermore, the benefits of using CAD systems to aid the clinicians in their decision-making were stated and the goal of the thesis and how it was achieved were outlined.

CHAPTER 3

Methodology

In this section, the proposed method for the prediction of CaP lesions is presented. The CAD system consists of six iteratively executed steps (see Figure 3.1). Prior to the classification, the prostate is segmented into superpixels using the SLIC algorithm [ASS+12]. The SLIC parameters are optimized by calculating the Sørensen-Dice Coefficient (DICE) between overlaying superpixels and the cancer annotation (see Chapter 3.3.2) and choosing the parameter set that yields the highest DICE. The final classifier is generated by applying a superpixel oversegmentation with the calculated parameter set, and then optimizing the classification model with exhaustive grid search and cross validation. On the following pages, every step of the CAD system is described in detail. The framework is implemented in Python using packages like NumPy, SciPy [JOP+] and Scikit-Learn [PVG+11].

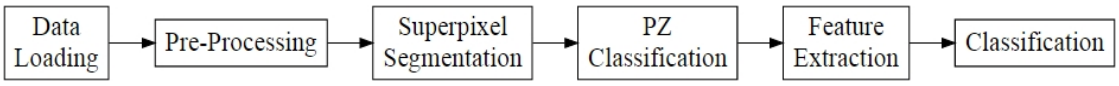


Figure 3.1: The workflow of the CAD system.

3.1 Data Description

In this thesis, two datasets are used for training and validating the classification. The goal is to examine how the oversegmentation and classification algorithms perform on different datasets. For the first dataset, publicly available data is used, whereas the second dataset is provided by the department of radiology and the department of radiation oncology of the Medical University of Vienna (MUV).

3.1.1 Dataset 1: Public Data

Lemaitre et. al. [LMF+15] provide a public dataset which is available at `http://i2cvb.github.io/`. It consists of T2W MRI, DCE MRI, DWI MRI and MRSI acquired from a 3.0 Tesla Siemens MRI system (see Figures 3.2 and 3.3). In total there are 20 patients of which 18 have biopsy proven CaP. Thirteen patients have CaP in the PZ, three in the CG, and two have CaP in both zones (see Table 3.1 for more details). The slices were annotated by an experienced radiologist and the annotations consist of the prostate boundary, the CG, the PZ and the cancer lesions. The different modalities were pre-registered, but have different voxel sizes and resolutions.

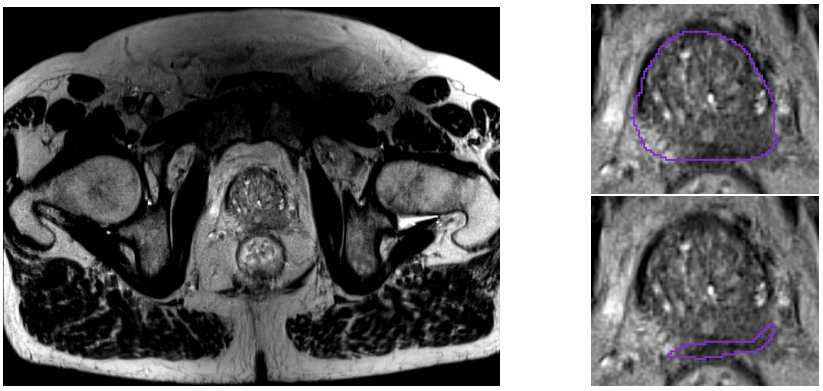


Figure 3.2: Left: Example of a whole T2W MRI slice of dataset 1 [LMF+15]. The volume has voxel dimensions of 320x320x20 and a voxel size of 0.625x0.625x3.600mm. Upper right: Cropped prostate with prostate boundary annotation. Lower right: Cropped prostate with cancer annotation. The cancer lesion has a lower intensity compared to the surrounding tissue.

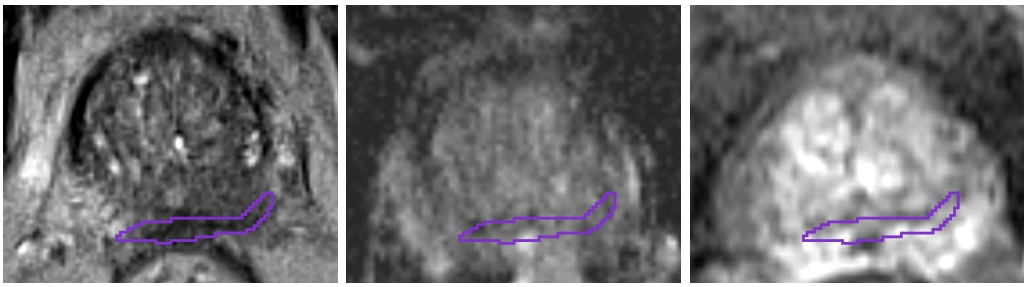


Figure 3.3: Examples of the different modalities of dataset 1 [LMF+15] with cancer annotation. From left to right: T2W, ADC, DCE early phase.

3.1.2 Dataset 2: Data from the Medical University of Vienna

Dataset 2 is provided by the department of radiology and the department of radiation oncology of the MUV and consists of T2W MRI, DCE volumes acquired at three

Patient Number	Modalities	CaP Location	CaP Size in % of the whole prostate volume
383	ADC, DCE, DWI, MRSI, T2W	PZ	6.82
384	ADC, DCE, DWI, MRSI, T2W	no cancer	
387	ADC, DCE, DWI, MRSI, T2W	CG	3.82
410	ADC, DCE, DWI, MRSI, T2W	PZ and CG	25.11
416	ADC, DCE, DWI, MRSI, T2W	PZ	32.99
430	ADC, DCE, DWI, MRSI, T2W	PZ	11.95
513	ADC, DCE, DWI, MRSI, T2W	CG	2.65
531	ADC, DCE, DWI, MRSI, T2W	no cancer	
634	ADC, DCE, DWI, MRSI, T2W	PZ	9.74
778	ADC, DCE, MRSI, T2W	PZ	4.93
782	ADC, DCE, MRSI, T2W	PZ	3.09
784	ADC, DCE, MRSI, T2W	PZ	9.66
799	ADC, DCE, MRSI, T2W	PZ	11.43
804	ADC, DCE, MRSI, T2W	PZ	1.65
836	ADC, DCE, DWI, MRSI, T2W	PZ and CG	54.97
870	ADC, DCE, MRSI, T2W	CG	6.15
996	ADC, DCE, MRSI, T2W	PZ	1.22
1036	ADC, DCE, MRSI, T2W	PZ	4.3
1041	ADC, DCE, MRSI, T2W	PZ	29.4

Table 3.1: Image modalities and the location of CaP for each patient of dataset 1.

different time points, the ADC map, Computed-Tomography (CT) images, PET images, and maps of quantitative functional features of the DCE MRI such as kTrans and iAUC (see Figures 3.4 and 3.5). Twenty-five patients were measured and annotated by an experienced radiologist using histological correlation as a gold standard. The annotations consist of the prostate boundary, the CG, the PZ, the cancer lesions and an anatomical division of the prostate into ROIs using the PI-RADS standard (see Figure 3.6), which was developed to standardize the interpretation of prostate mpMRI examinations [BRC+12, WBC+16, DAA+11]. One patient has CaP in the GC, eleven patients in the PZ and nine patients have CaP in both zones (see Table 3.2 for more details).

3.2 Pre-Processing

Before the prostate images can be used for machine learning, they are loaded and pre-processed to eliminate different resolutions and to normalize image intensities.

Patient Number	**Modalities**	**CaP Location**	**CaP Size in % of the whole prostate volume**
001	ADC, CT, DCE, iAUC, kTrans, PET, T2W	PZ	7.21
002	ADC, CT, DCE, iAUC, kTrans, PET, T2W	no cancer	
003	ADC, CT, DCE, iAUC, kTrans, PET, T2W	PZ	6.98
004	ADC	no cancer	
005	ADC, CT, DCE, iAUC, kTrans, PET, T2W	PZ and GC	12.4
006	ADC, CT, DCE, iAUC, kTrans, PET, T2W	PZ	7.71
007	ADC, CT, DCE, iAUC, kTrans, PET, T2W	PZ and CG	1.4
008	ADC, CT, DCE, iAUC, kTrans, PET, T2W	PZ	10.44
009	ADC, CT, DCE, iAUC, kTrans, PET, T2W	PZ	7.87
010	ADC, CT, DCE, iAUC, kTrans, PET, T2W	CG	14.35
011	ADC, CT, DCE, iAUC, kTrans, PET, T2W	PZ	5.37
012	ADC, DCE, iAUC, kTrans	no cancer	
013	DCE, iAUC, kTrans, T2W	no cancer	
014	ADC, DCE, iAUC, kTrans, T2W	PZ	3.67
015	ADC, DCE, iAUC, kTrans, T2W	PZ	5.25
016	ADC, DCE, iAUC, kTrans, T2W	PZ and CG	2.71
017	ADC, DCE, iAUC, kTrans, T2W	PZ and CG	3.49
018	ADC, DCE, iAUC, kTrans, T2W	PZ and CG	16.62
019	ADC, DCE, iAUC, kTrans, T2W	PZ and CG	6.88
020	ADC, DCE, iAUC, kTrans, T2W	PZ and CG	7.06
021	ADC, DCE, iAUC, kTrans, T2W	PZ	10.29
022	ADC, DCE, iAUC, kTrans, T2W	PZ	4.14
023	ADC, DCE, iAUC, kTrans, T2W	PZ and CG	2.33
024	ADC, DCE, iAUC, kTrans, T2W	PZ	6.62
025	ADC, DCE, iAUC, kTrans, T2W	PZ and CG	4.16

Table 3.2: Image modalities and the location of CaP for each patient of dataset 2.

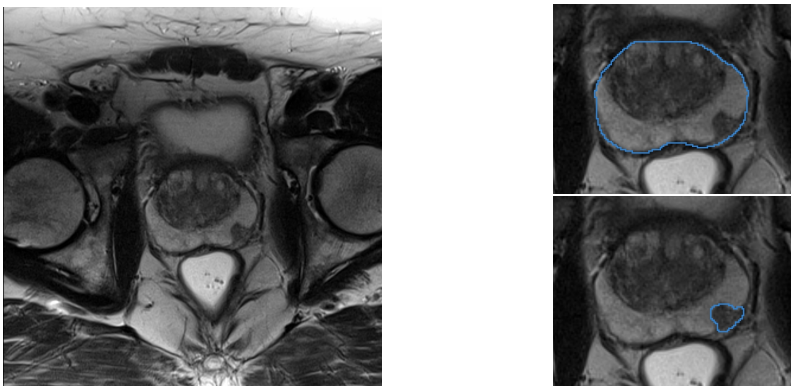


Figure 3.4: Left: Example of a whole T2W MRI slice of dataset 2. The volume has voxel dimensions of 320x320x20 and a voxel size of 0.625x0.625x3.600mm. Upper right: Cropped prostate with prostate boundary annotation. Lower right: Cropped prostate with cancer annotation. The cancer lesion has a lower intensity compared to the surrounding tissue.

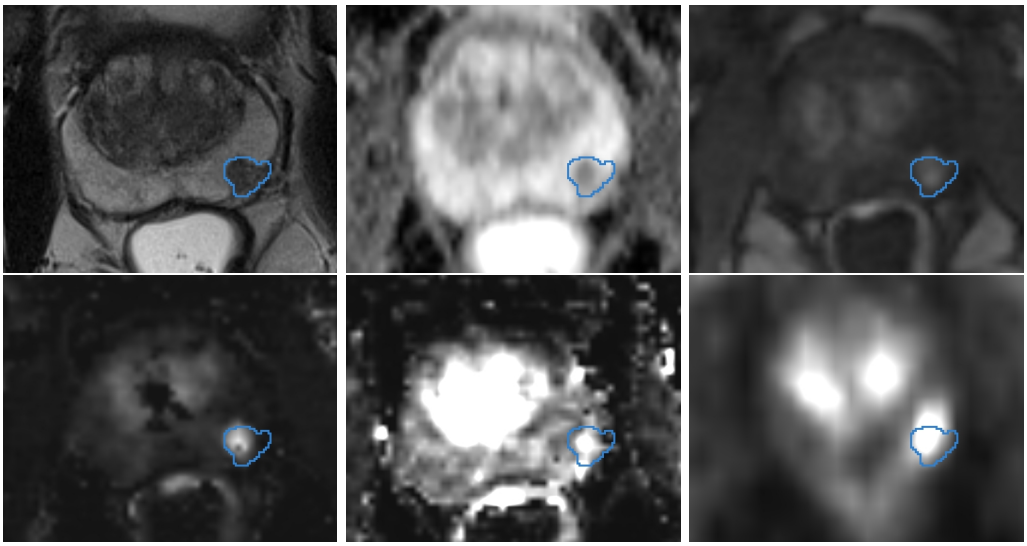


Figure 3.5: Examples of the different modalities of the MUV dataset with cancer annotation. From left to right: T2W, ADC, DCE early phase, kTrans, iAUC, PET.

3.2.1 Loading the Data

The three most used modalities in literature are T2W MRI, DWI and DCE MRI [LMF+15]. To make our CAD system acessible to the majority of datasets and comparable to other approaches, we use these three modalities as well. The DICOM files are loaded with PyDICOM. Because not every file has the same volume dimensions and voxel size, they are resampled with ITK [JMIC13] into the same format as the T2W volume. As

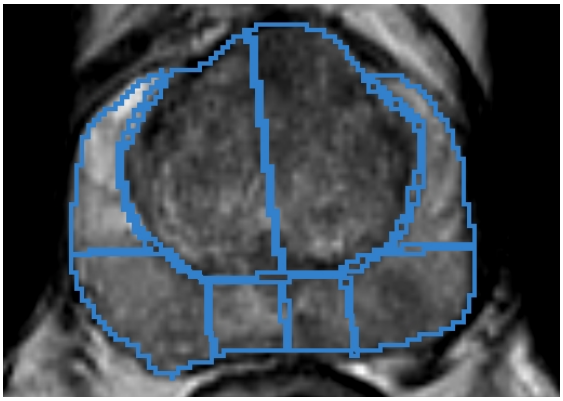


Figure 3.6: Axial slice of a T2W-MRI with annotations according to the standardized MRI prostate reporting scheme. The prostate is segmented into specific pre-defined regions. Image taken from dataset 2.

interpolation method, B-Spline interpolation is used because it incorporates information from more distant pixels and it is the preferable interpolation method for medical image processing [**?**]. After the resampling, a four dimensional array is generated in the format [z-coordinate, x-coordinate, y-coordinate, modality]. In addition to the T2W MR images and the ADC map, for dataset 1 the whole DCE series is saved in the array as well. For dataset 2, the three time points of the DCE series as well as the kTrans volume are saved. Furthermore, the annotations are also resampled into the same dimensions as the T2W volume and saved in a seperate array in the format [z-coordinate, x-coordinate, y-coordinate, annotationType]. For the MUV dataset the annotations are stored in the RTSTRUCT format which saves the boundary points of the segmentation in the physical space. The segmentations are extracted as a binary mask with 3D Slicer [FBKC$^+$12] and saved as NRRD-files. They are then loaded into Python with PyNRRD.

3.2.2 Image Normalization

The machine learning classifier uses texture features that are extracted from intensity values from the MR scans. However, MRI sequences like T1W and T2W produce variable intensities for the same tissue, even when using the same scanning protocol between patients. Hence, to improve the discriminative performance of the classifier, these variations have to be corrected. In literature, there are two approaches for normalization. The approach by Kwak et. al. [KXW$^+$15] uses the median and standard deviation of the voxels inside the prostate to normalize the intensities. The method by Ozer et. al. [OLL$^+$10] considers anatomical structures like blood vessels which have similar intensities across different MR scans to normalize the images. They transform the image to have zero mean and unit variance. However, they only use voxels of the PZ of the prostate.

In our framework, we use the approach by Kwak et. al. [KXW$^+$15] to normalize the images in order to reduce the need for manual interventions such as localizing blood

vessels or manual segmentation of the PZ in the MR scans.

After loading the data, the first step in our normalization method is histogram equalization to increase the contrast of the images. In the next step, Equation 3.1 is used for normalizing the image.

$$imageNormalized = \frac{imageRaw}{\sigma_p \ + \ 2 \ * \ \mu_p}, \tag{3.1}$$

where μ_p is the mean value and σ_p is the standard deviation of all voxels inside the prostate.

3.3 Superpixel-based ROI Detection

One of our CAD system's novel approaches is the use of oversegmentation for generating ROIs. To our knowledge, only two other papers use automatic ROI detection [VBKH12, LDB+14]. One major drawback of other CAD systems that use manual ROIs is that the user has to manually annotate ROIs in each new volume that needs to be classified, which results in lower automatization and therefore more work. Furthermore, the position of the cancer is not located automatically. Instead, the annotated ROI is classified to show the likelihood if the ROI is cancerous or not.

3.3.1 Superpixels

In this project, superpixels or supervoxels are used as a pre-processing step for feature detection. The goal is to improve the classification result by, in addition to low-level pixel- or voxel-wise features, computing statistical features that are defined over a ROI. In most other CaP CAD systems, the features are either only pixel- or voxel-wise features, or the ROIs are defined manually by an expert reader.

The oversegmentation algorithm that is used for our task is chosen based on the following requirements:

- The boundaries of the different tissues in the prostate should be preserved.
- The algorithm should be computationally efficient.
- The algorithm has to be available for 2D and 3D grayscale images.
- The parameters of the algorithm should be able to be controlled by the user.

To be able to set the parameters is especially important, as is it the goal to optimize the parameters of the oversegmentation algorithm to achieve a sufficient segmentation of the cancer lesions in the prostate (see Chapter 3.3.2).

Regarding the requirements for our oversegmentation algorithm, the SLIC method [ASS+12] is chosen for our CAD framework because it fulfills all specifications and is readily available for Python in Scikit-Learn [PVG+11]. In the next paragraphs, the algorithm is described in more detail.

SLIC

The SLIC algorithm [ASS+12] is based on a localized k-means clustering and consists of several stages. Here, the version of the algorithm for segmenting 3D grayscale volumes is explained. First, the clusters are initialized by setting k cluster centers with an equal distance throughout the volume, where k is the amount of supervoxels. The distance between the cluster centers is $S = \sqrt[3]{N/k}$, where N is the number of voxels in the volume. The next step is the assignment phase, where each voxel is associated with its nearest cluster center by computing the distance measure D. A cluster center has a volume of $S * S * S$ and the search region for the clustering is $2S * 2S * 2S$. Subsequently, the new cluster centers are updated by computing the mean of all voxels of the cluster. This step is repeated until convergence. The distance measure D computes the distance of a voxel to the cluster center. It is a combination of space and instensity distances and is written as

$$D = \sqrt{(\frac{d_c}{N_c})^2 + (\frac{d_s}{N_s})^2}, \tag{3.2}$$

where

$$d_c = \sqrt{(I_i - I_j)^2}, \tag{3.3}$$

is the intensity distance function for grayscale images or

$$d_c = \sqrt{(I_1 i - I_1 j)^2 + (I_2 i - I_2 j)^2} \tag{3.4}$$

is the multimodal distance function for 2 grayscale images and

$$d_s = \sqrt{(s_x * (x_i - x_j))^2 + (s_y * (y_i - y_j))^2 + (s_z * (z_i - z_j))^2} \tag{3.5}$$

is the spatial distance function. I is the intensity of a voxel and s_x, s_y and s_z are the voxel sizes in each dimension. This is done to consider anisotropic voxel dimensions. The distance measure as seen in Figure 3.6 normalizes the spatial and intensity distances using the normalization constants N_c and N_s to ensure that both distance measures are weighted equally. If the two distance measures are not equalized, the spatial distance outweights the intensity distance in large supervoxels. For small supervoxels, the opposite applies. $N_s = \sqrt[3]{N/k}$ is the maximum spatial distance. Normalizing the intensity is difficult because the intensities vary in every cluster [ASS+12]. Therefore, the authors replace the distance measure by a constant m, resulting in the distance equation

$$D = \sqrt{d_c^2 + (\frac{d_s}{S})^2 * m^2}, \tag{3.6}$$

where m is a constant that describes the importance of d_s relative to d_c. When m is large, the supervoxels are more compact. Finally, any voxels that are unconnected to their cluster center are assigned the label of their nearest cluster.

The clusters are computed in a multi-modal setting where T2W MRI and the ADC map are used as inputs for the SLIC algorithm, because these two modalities are available in both datasets (see Figure 3.7 for example segmentations). The DCE MRI is not used for computing the clusters because they are not saved in the same format in both datasets (dataset 1 has the whole time-series and dataset 2 has only three timepoints).

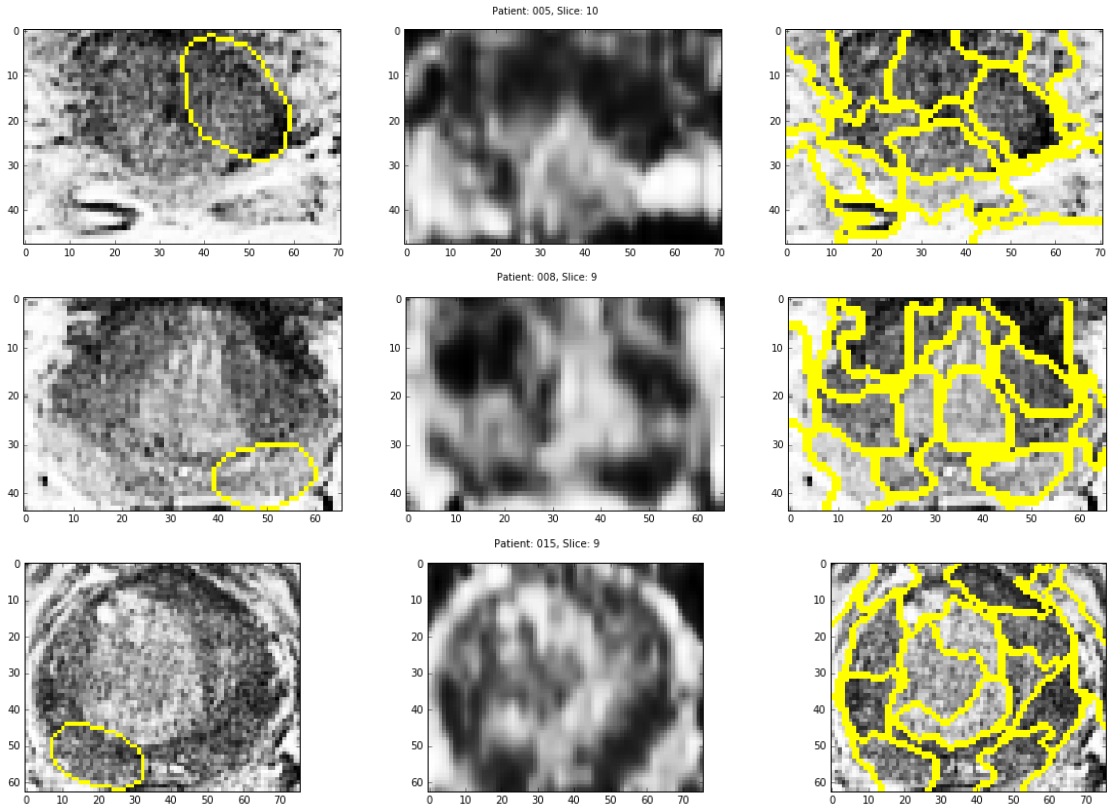


Figure 3.7: SLIC examples. Left: Axial slice of a T2W MRI with annotated cancer region. Middle: ADC map of the same slice. Right: SLIC segmentation. Images taken from dataset 2.

3.3.2 Optimizing the Superpixels

Before the features for the classification algorithm are calculated, the superpixels are optimized. The goal is to achieve the best accuracy for an automatic annotation of CaP

and to oversegment the prostate gland into distinct ROIs. The assumption is, that if a superpixel segments cancer lesions in the training data accurately, the same applies to test data.

Concretely, a parameter search is performed until the best overlap of the cancer annotations with superpixels is achieved (see Table 3.3 for details). The number of superpixels ranges from 50 to 150 for dataset 1 and 10 to 100 for dataset 2. As the oversegmentation is done on a per-slice basis for dataset 2, the overall number of superpixels for the whole volume is roughly the same as in dataset 1 in the lower parameter range. The minimum and maximum are set according to prior tests, in which a visual inspection of the oversegmentation accuracy for parameters that are lower or higher than these values finds no meaningful results. The compactness, which controls the shape of the superpixels, is set to three values on a log scale, as it is recommended in the documentation of the framework Scikit-Learn [PVG+11]. The sigma parameter, which controls the width of the Gaussian smoothing kernel that is applied prior to the segmentation, is set from 0.1 to 0.9 with a step size of 0.1. The threshold is the value of overlap from which a superpixel is considered a CaP (e.g., when the threshold is set to 0.9, a superpixel is considered part of the CaP lesion when their pixels overlap by 90% or more) and ranges from 0.1 to 0.9 with a step size of 0.1. While more parameter combinations may yield more precise results, a smaller step size is out of scope of this thesis due to its computational cost. For each iteration, the DICE of the overlap between the superpixels and CaP is computed and the parameter values with the best DICE are used in the CAD system.

Parameter	**Value Range (Step size)**
Number of Superpixels	50 - 150 (10), 10 - 100 (10)
Compactness	0.01, 0.1, 1
Sigma	0.1 - 0.9 (0.1)
Threshold	0.1 - 0.9 (0.1)

Table 3.3: The parameter range for the oversegmentation optimization.

The overlap is measured by computing the DICE of all superpixels that are considered part of a cancer region with the cancer annotations which is defined as:

$$DICE = \frac{2|A \cap B|}{|A| + |B|}, \tag{3.7}$$

where A is the set of pixels of the superpixel and B is the set of pixels of the annotated cancer region. A superpixel belongs to a cancer region if the area that overlaps with a cancer region exceeds a certain threshold. This threshold is set to values ranging from 0.1 to 0.9 with a step size of 0.1. For every value, the optimization algorithm is executed and the DICE for every parameter combination of the other parameters (number of segments, compactness, sigma) is computed. A disadvantage of DICE values are that they are not suitable for comparing segmentations on images with different sizes. However, since the oversegmented and annotated images are always the same, this is not a problem.

The final result of the superpixel accuracy is calculated by the following equation:

$$E = 1 - DICE, \tag{3.8}$$

which means that the result is better the bigger the DICE is.

An overwiew of the optimization procedure is given in Algorithm 3.1.

Algorithm 3.1: SLIC Optimization

```
1  for threshold = 0.1 to 0.9 do
2      for segments = min to end do
3          for compactness = (0.01, 0.1, 1) do
4              for sigma = 0.1 to 1 do
5                  for prostate in allProstates do
6                      slicSegments = calculateSlicSegments();
7                      prostateSegments = getOverlappingProstateSegments();
8                      //returns all SLIC segments that overlap with CaP annotation >
                        threshold
9                      dice = calculateDICE(cancerAnnotation, prostateSegments);
10                     score = 1 − dice;
11                     scoreList.add(score);
12                 end
13                 averageScore = sum(scoreList)/length(scoreList);
14                 averageScoreList.add(averageScore);
15             end
16         end
17         endResult = getLowestAverageScore(averageScoreList);
18     end
19 end
```

The cancer annotations in the two datasets show differences. While in the MUV dataset, the annotations are made on a per-slice basis, the annotated regions in the public dataset show a coherent, three-dimensional shape. Therefore, to achieve the best segmentation result, the SLIC algorithm is applied in either 3D for dataset 1 and in 2D for dataset 2.

3.4 Feature Detection

The machine learning algorithm needs features to build a classifier. In the proposed CAD system, three different types of features are used for training: pixel-based features, region-based features and functional features (see Table 3.4 for an overview). In the following sections, each type of feature is described and the most relevant features are explained. The features for the proposed CAD system in this thesis are chosen based on a literature review of CaP CAD systems by Lemaitre et. al. [LMF+15], where they summarize the most commonly used features in literature and on the features used in the work of Litjens et. al. [LD+14] and Vos et. al. [VBKH12], which are the only other two-stage CAD systems in literature and therefore the most similar CAD frameworks to the proposed approach.

3.4.1 Pixel-based Features

Pixel-based features are calculated based on the intensities of each pixel in the image. For T2W MRI and the ADC map, the intensity values and a Gaussian texture bank (Gaussian filters applied with kernel sizes of 0.5, 1 and 2) are used as features. Furthermore, different edge detection filters such as Sobel, Gabor, Canny and Prewitt are used and for each pixel the Gray Level Co-occurence Matrix (GLCM) is computed.

Intensity-based Features

The intensity values of the T2W MRI and the ADC map are used as intensity-based features in both datasets. Furthermore, Gaussian filters with kernel sizes of 0.5, 1 and 2 are used for the T2W MRI and ADC map.

Edge-based Features

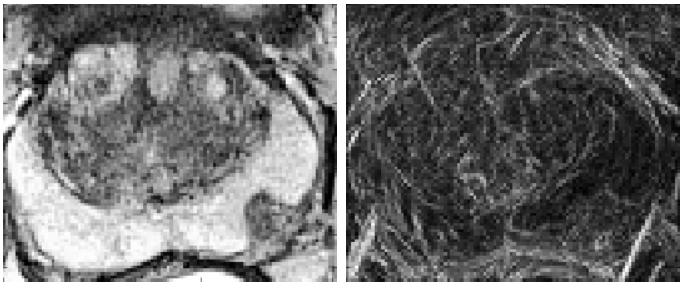
Edge detectors are commonly used features to detect CaP [LMF+15]. In this framework, Gabor, Sobel and Prewitt filters are used for the T2W MRI and the ADC map (see Figure 3.8). The Gabor filter is made rotation invariant by computing 16 different angles and choosing the maximum intensity for each angle.

Texture-based Features

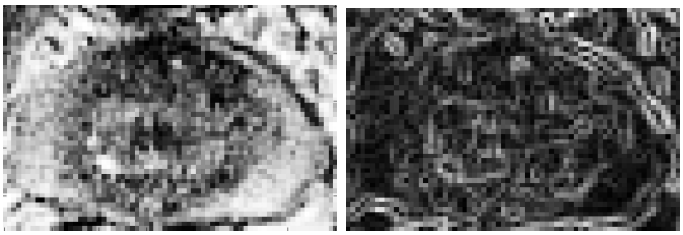
Texture-based features do not only consider individual pixels, but also the relationship between them. The GLCM computes the statistical distribution of pixel intensity values at specified positions relative to each other. GLCMs, or Haralick features as they are also called, extract second-order statistical features from the image [HSD73]. In this framework, the following features are calculated for each pixel: contrast, dissimilarity, homogeneity, energy, correlation and ASM. Because computing the GLCM features is computationally expensive, the features are calculated once and then saved.

Type	Name	Description
Pixel-based		
Intensity	T2W	Intensity values of the T2W MRI
Intensity	ADC	Intensity values of the ADC map
Filter	Gaussian	Multi-scale Gaussian filters
Filter	Gabor	Gabor filter
Filter	Canny	Canny filter
Filter	Prewitt	Prewitt filter
Texture	GLCM	GLCM filters
Anatomical	Distance map	Distance of each pixel to the boundary of the prostate
Region-based		
Statistical	Mean	Mean intensity value of each superpixel
Statistical	Variation	Variation of the intensity values of each superpixel
Statistical	Standard deviation	Standard deviation of all pixel intensities in a superpixel
Statistical	Skewness	Skewness of the histogram of all pixel intensities in a superpixel
Statistical	Kurtosis	Kurtosis of the histogram of all pixels in a superpixel
Statistical	Percentiles	10th, 25th and 75th percentiles of each superpixel
Anatomical	Superpixel properties	Size and shape of each superpixel
Anatomical	PZ probability	Map of probabilites for each pixel that it belongs to the PZ
Functional		
Pharmacokinetic	Variation	Variation of intensities over time for each pixel
Pharmacokinetic	Skewness	Skewness of histogram of intensities over time for each pixel
Pharmacokinetic	Kurtosis	Kurtosis of histogram of intensities over time for each pixel
Pharmacokinetic	Geometric mean	Geometric mean of intensities over time for each pixel
Pharmacokinetic	kTrans	Map of the kTrans value for each pixel

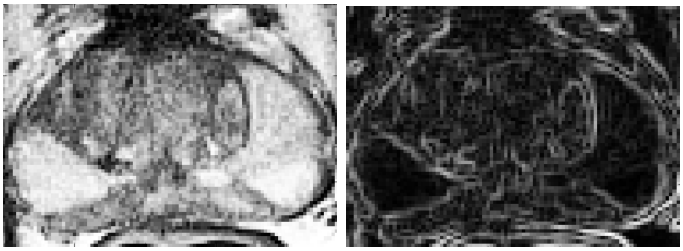
Table 3.4: Overview of the features that are used for the classifier.



(a) Gabor Edge filter.



(b) Sobel Edge filter.



(c) Prewitt Edge filter.

Figure 3.8: Examples of different edge filters.

Anatomical-based Features

Anatomical-based features take the anatomical position of each voxel into account. For this purpose, a distance map is computed that shows the distance of each voxel to the prostate boundary (see Figure 3.10). The distance map is computed from the prostate boundary annotation. It shows a high feature importance because approximately 70% of cancer lesions occur in the PZ of the prostate [YVM+12].

3.4.2 Region-based Features

ROIs consider a group of neighbouring pixels to allow the computation of region-based statistical features. The superpixels that were generated in a prior step are used as ROIs. The relevance and importance of the features correlate with the accuracy of the supervoxel segmentation. If a superpixel is not accurate enough and includes different tissues of the prostate, the features have less relevance.

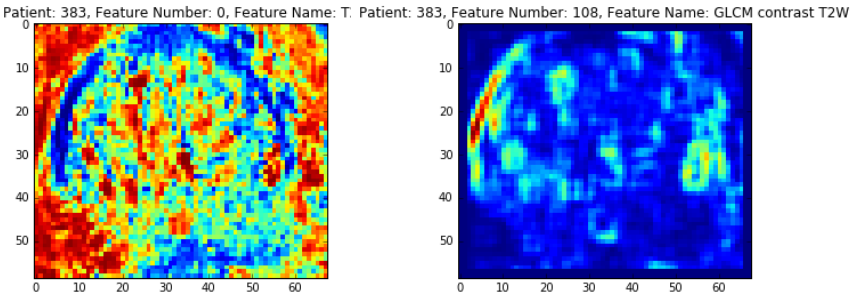


Figure 3.9: Examples of a GLCM. Left: Axial slice of a T2W MRI. Right: Calculated contrast GLCM from the same slice.

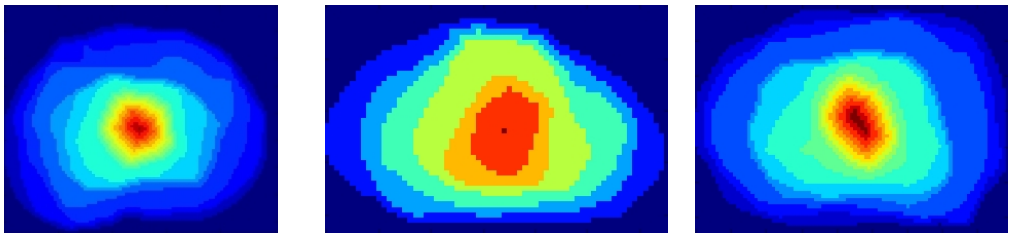


Figure 3.10: Examples of distance maps from three different prostates. Blue indicates a low distance to the prostate boundary and red a high distance.

Statistical Features

In the T2W MRI, statistical features that are used are the mean, variation, standard deviation, skewness and kurtosis of the pixel intensity histogram, and the 10th percentile. The 10th percentile is used because cancer lesions can show as a region with lower intensity compared to the surrounding tissue. The histogram skewness is used because it correlates with cancer lesions in T2W MRI [HBH$^+$11, PJY$^+$13]. In the ADC map, the mean, variation, standard deviation and the 10th percentile are used. The average and 10th percentile of the ADC map correlate with cancer lesions in the PZ of the prostate [PJY$^+$13]. In the DCE MRI, mean, variation, standard deviation, and the 75th percentile are used, because cancer lesions can show as bright spots in early phase DCE MRI. For all other images (edge filters, DCE pharmacokinetic parameters, distance map), the mean, minimum, maximum standard deviation, median, the 25th and 75th percentiles are used.

Anatomical-based Features

The superpixels segment CaP regions and share the same properties in terms of shape and size. Therefore, the location and size for each superpixel are calculated and used as features.

3.4.3 Functional Features

For the public dataset, features from the DCE series are calculated. For every pixel the intensity changes over time are considered and histogram-based calculations are made, because cancerous tissue has different pharmacokinetic parameters than normal tissue [HBH+11]. The variation, skewness, kurtosis and geometric mean for each pixel are calculated. In the MUV dataset only three DCE MR images for certain timepoints are available (pre, early and late), therefore computing histogram-based features is not feasible. However, the quantiative feature map kTrans that is computed by the MRI scanner is used.

3.4.4 Features for the PZ probability map

Approximately 70% of CaPs occur in the PZ of the prostate [YVM+12]. Furthermore, cancer in the PZ shows different characteristics in MRI than cancer in the CG, which makes it difficult to distuingish from normal tissue [BRC+12]. Because of this, some papers ([NRML+12, PCVH10, RCM+16]) only focus on the PZ for classification, which would require an expert reader to annotate the PZ or an automatic segmentation approach ([LY11, LDB+14]). In our CAD system, the PZ is automatically detected with pattern classification. The distance map features already focus the classification algorithm onto the peripheral parts of the prostate. In addition to this, a map that shows the probability that a pixel belongs to the PZ is computed and used as a feature. Prior to the main classification of CaP, a RF is trained with pixel-based and superpixel-based features that are chosen based on the work by Litjens et. al. [LDvdV+12], who implement a pattern recognition approach for the zonal segmentation of the prostate and on the literature review by Lemaitre et. al. [LMF+15], who provide an overview of the different zones of the prostate and their visual properties in mpMRI (see Table 3.5 for a detailed list of the features). In the next step, for new test data, the RF is used to generate the probability map. This map is then used as a feature for the CaP classifier. Figure 3.11 shows some examples of PZ probability maps.

3.5 Image Classification

In this section, the classification algorithm and how it is optimized is explained. According to the review by Lemaitre et. al. [LMF+15], the most commonly used classifier that is used in CAD systems for CaP is the SVM, followed by LDA and RFs. The two most similar approaches to the proposed CAD systems are using RFs [LD+14] and LDA [VBKH12]. The CaP prediction framework by Litjens et. al. [LD+14] yields the best result of the reviewed two-stage CAD systems (see Chapter 2.5) with a RF classifier, describing that it outperforms LDA and the Gentleboost classifier. Vos et. al. [VBKH12] choose a LDA classifier over a SVM because it yields the best accuracy of all tested classifiers. Hence, LDA, SVM and RFs are chosen as possible classifiers. For prior testing, the three classifiers are trained on sample data from both datasets after the superpixels

Type	Name	Description
Pixel-based		
Intensity	T2W	Intensity values of the T2W MRI
Intensity	ADC	Intensity values of the ADC map
Filter	Gaussian	Multi-scale Gaussian filters
Filter	Gabor	Gabor filter
Filter	Canny	Canny filter
Filter	Prewitt	Prewitt filter
Texture	GLCM	GLCM filters
Anatomical	Distance map	Distance of each pixel to the boundary of the prostate
Region-based		
Statistical	Mean	Mean intensity value of each superpixel
Statistical	Min	Minimum intensity value of each superpixel
Statistical	Max	Maximum intensity value of each superpixel
Statistical	Variation	Variation of the intensity values of each superpixel
Statistical	Standard deviation	Standard deviation of all pixel intensities in a superpixel
Statistical	Skewness	Skewness of the histogram of all pixel intensities in a superpixel
Statistical	Percentiles	10th percentile of each superpixel
Anatomical	Superpixel properties	Size and shape of each superpixel
Anatomical	PZ probability	Map of probabilites for each pixel that it belongs to the PZ

Table 3.5: The features for the PZ probability map.

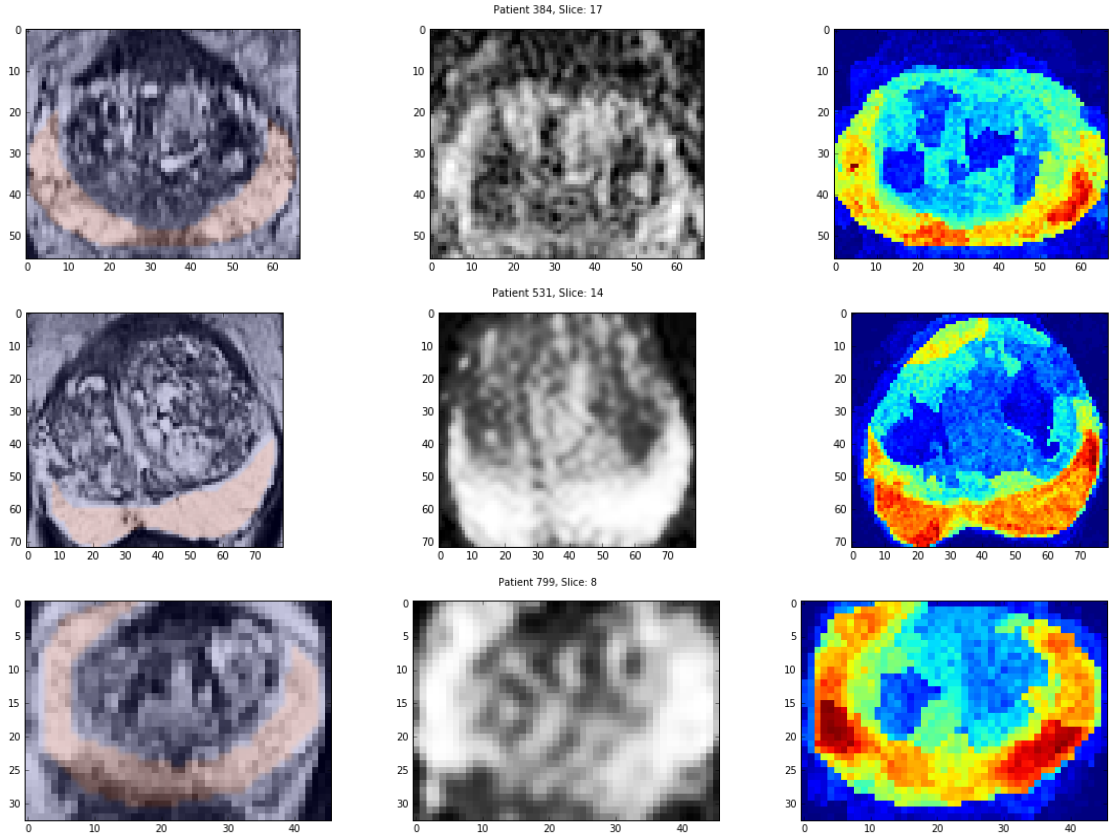


Figure 3.11: Examples of PZ probability maps. The left image shows a slice in the T2W MRI volume with the annotated PZ (red). The image in the center shows the corresponding slice in the ADC map. The image on the right shows the predicted PZ, with red indicating a high probability and blue indicating a low probability that the voxel is part of the PZ.

are optimized. RFs are chosen as a classifier for the proposed CAD system because they yield the best classification result.

3.5.1 Random Forest

The RF [Bre01] algorithm uses an ensemble of decision trees to make a prediction. Each tree is built using a random subset of the sample data (called bagging) and a subset of the features. When a new sample set needs to be classified, all trees make a prediction for the data. The final classification is the class that got the most predictions.

Each tree is grown to its full extent. For the nodes in each tree, a random subset of features is chosen for splitting them. For each node, the feature with the lowest impurity is chosen to split the sample subset, resulting in two child nodes. This process is repeated

until each node has a Gini index of 0 (or in other words, until each node contains only samples of the same class).

The advantages of RFs include its ease of use and robustness to overfitting due to the averaging of the trees. Furthermore, RFs can give estimates about the importance for each feature, making it possible to eliminate non-discriminative features and speed up the classification. In this thesis it is used for validating the accuracy of the superpixel algorithm. In literature it has been shown that certain features correlate with cancer lesions in MR images (e.g., the 10th percentile and average of the ADC map [PJY+13]). Therefore, the feature importance for statistical features that correlate with cancerous regions increases when the superpixels are more accurate in their segmentation.

The algorithm is implemented using Scikit-Learn [PVG+11].

3.5.2 Optimizing the Hyper-Parameters

To optimize the classification result, the best hyper-parameters (the parameters of a machine learning algorithm that are not directly learned) have to be found. In RFs, hyper-parameters that can be tuned are the number of trees, the number of features used for each tree and the maximum depth of the tree. The parameters are tuned using exhaustive grid search of a given parameter space. As scoring function, cross validation is used.

3.6 Summary

This section described the CAD framework and every step in detail. First, the data was loaded, resampled and normalized. Then, ROIs were automatically generated by applying a superpixel segmentation. In the next step, features for the classification algorithm were calculated. For more sophisticated, region-based features, the generated ROIs were used. Finally, the classification algorithm was applied. In addition to the workflow, complementary steps such as optimizing the superpixels, tuning the parameters of the classification model, and the algorithms that were used (SLIC, RFs) were explained and it was described how the CAD system automatically finds the best parameters for the oversegmentation and the classifier.

CHAPTER 4

Results

In this chapter, the CAD system is evaluated. The overall classification accuracy is assessed and compared to other frameworks in literature. Furthermore, it is examined if and how the superpixel-based features improve the classification results and which features and modalities are most important for the classification. It is shown which segmentation yields the best result and it is explained how the parameters affect the outcome of the CAD system. Because there are two datasets available, the results are shown for both. Generally, it is difficult to compare the different approaches in literature because the test datasets are only available for the specific studies or the evaluation approaches differ from each other. However, the public dataset is openly available and therefore the CAD system is directly comparable to other approaches that use the same dataset.

4.1 Classification Accuracy

The classification is evaluated by computing the AUC of the ROC curve, because the two most similar CAD systems in literature also use this form of validation [LD+14, VBKH12].

The ROC curve shows the classification result at different threshold rates. When shown in a diagram, the True Positive Rate (TPR, sensitivity) is plotted on the Y-axis and shows the amount of pixels that are correctly identified as cancerous, with respect to all cancerous pixels, while the False Positive Rate (FPR) is plotted on the X-axis and shows the proportion of pixels that are mistakenly considered as cancerous, with respect to all non-cancerous pixels. For plotting the ROC curve, the two metrics TPR and FPR are calculated for different threshold rates between zero and one (with zero being the value that predicts all pixels as cancerous and one being the value that predicts all pixels as non-cancerous). The AUC is calculated by computing the area under the resulting curve. The calculation of the AUC is pixel-based because the classification also uses features that are calculated on each pixels, and not only superpixel-based features.

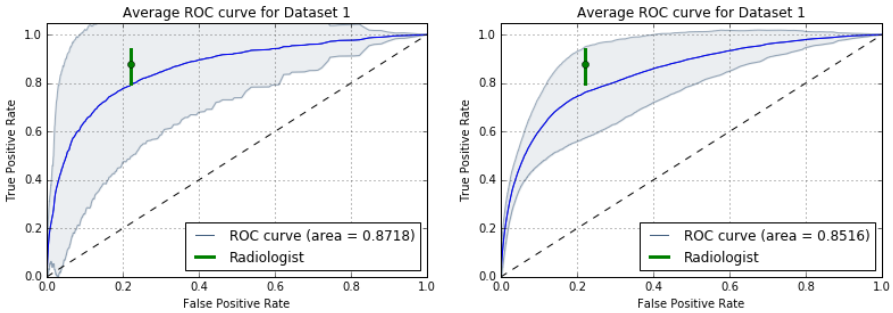


Figure 4.1: Averaged ROC curves of the classification performance for dataset 1. Left: The result when superpixels are added as features to the classifier. Right: The classification result with only pixel-based and no superpixel-based features. The blue area with lower opacity shows the 95% confidence intervall of the standard deviation of all ROC curves. The green line shows the performance of the radiologists as shown in the study by Rooji et. al. [dRHF+14].

To assess the overall accuracy for all datasets, leave-one-out cross validation is used [HTF09]. For this, one sample is used for testing while the remaining samples of the dataset are used for training. For each sample, this process is repeated and the AUC is computed. Finally, the classification accuracy is calculated by computing the mean of all AUCs.

4.1.1 Public Dataset

For the public dataset, the classifier yields an average AUC of 0.87 (see Figure 4.1 for the averaged ROC curves and Table 4.1 for all AUCs). The most similar work in literature that also uses automatic ROI detection is the CAD system by Litjens et. al. [LDB+14], which yields an AUC of 0.81. As Figure 4.1 shows, the performance of the classifier has a high variability with a standard deviation of 0.0779. When compared to a study by Rooji et. al. [dRHF+14] that reports a specificity and sensitivity of 0.74 and 0.88 when radiologists try to detect CaP with mpMRI (T2W, DWI and DCE MR images), the classifier shows a slighty worse averaged sensitivity at the given specificity, but the performance of the radiologists still is in the 95% confidence intervall of the ROC curve. However, this data is not based on the same dataset so a direct comparison is difficult. Figure 4.3 shows a visual comparison of the annotated cancer regions and the result of the CAD system. The probability map shows high values for areas with CaP, but also some false-positive classifications.

4.1.2 MUV Dataset

The average AUC for the MUV dataset is 0.59. As Table 4.2 shows, there is a high variability between the AUCs with patient 009 having an AUC of 0.9297 and patient 007 having an AUC of 0.4976, resulting in a standard deviation of 0.1085. Figure 4.2

Patient Number	AUC
383	0.9362
384	no cancer annotations
387	0.8014
410	0.7545
416	0.8335
430	0.8864
513	0.9114
531	no cancer annotations
634	0.9115
778	0.9328
782	0.8590
784	0.8760
799	0.7889
804	0.8058
836	0.8501
870	0.9125
996	0.8156
1036	0.6501
1041	0.7106
	Average: 0.8718 Standard Deviation: 0.0779

Table 4.1: The AUCs for all patients of the public dataset.

shows the average ROC curve for dataset 2. A possible reason for that variability could be that the CaP annotations are not accurate enough in some scans, resulting in wrong classifications. Figure 4.4 shows a comparison of the annotated cancer regions and the result of the CAD system.

4.2 Improvement of Classification by Adding Superpixels

In this section, it is shown that the oversegmentation improves the classification. For this purpose, the superpixel-based classification is compared to a classification that is computed with only pixel-based features (and has the superpixel-based features removed). The comparison is done visually by examining the classification maps for example patients and numerically by looking at the average ROC curves.

4.2.1 Public Dataset

When comparing the two average ROC curves (see Figure 4.1), the classifier shows a higher sensitivity especially at high specificity values, and the average AUC is slighty better (0.87 with superpixels and 0.85 without superpixels). However, the variability of

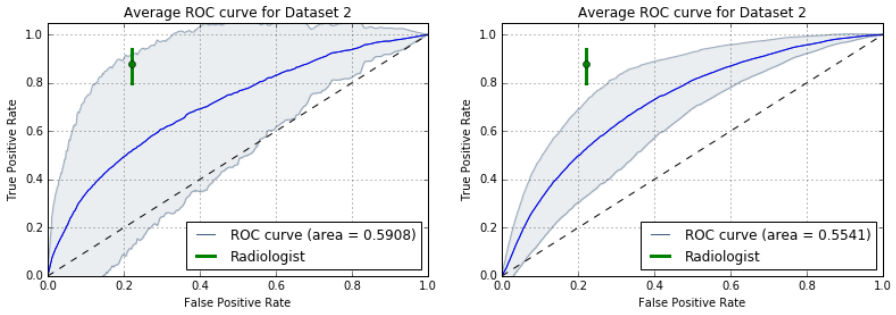


Figure 4.2: Averaged ROC curves of the classification performance for dataset 2. Left: The result with superpixel-based features added to the classifier. Right: The classification result with only pixel-based and without superpixel-based features. The blue area with lower opacity shows the 95% confidence intervall of the standard deviation of all ROC curves. The green line shows the performance of the radiologists as shown in the study by Rooji et. al. [dRHF+14].

Patient Number	**AUC**
001	0.7100
002	0.8654
003	0.6298
005	0.8551
006	0.7069
007	0.4976
008	0.8880
009	0.9297
010	0.7955
011	0.5764
014	0.6865
015	0.7241
016	0.7293
017	0.6316
018	0.6750
019	0.7189
020	0.6336
021	0.7027
022	0.7211
023	0.6152
024	0.7844
025	0.5425
	Average: 0.59 Standard Deviation: 0.1085

Table 4.2: The AUCs for all patients of the MUV dataset.

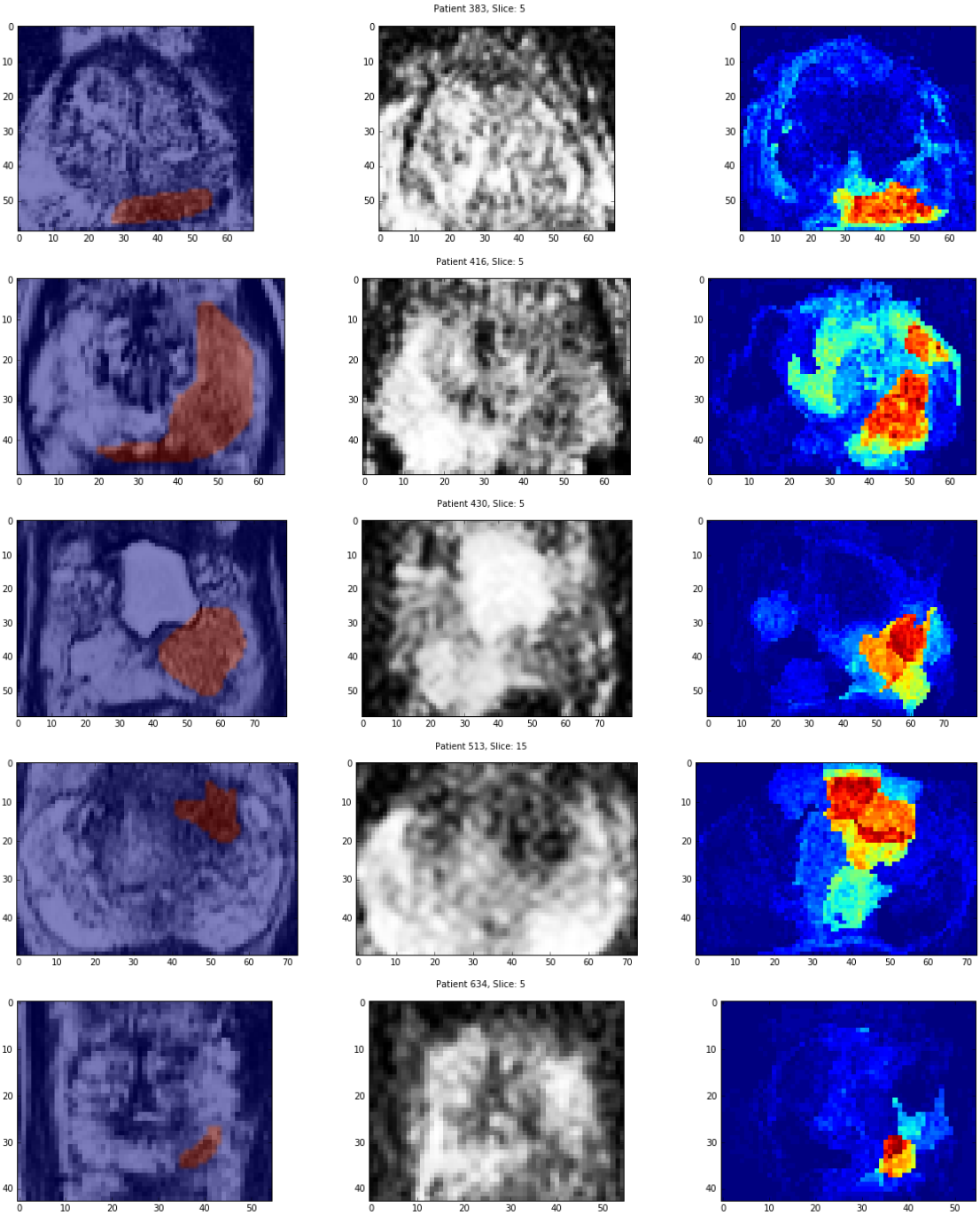


Figure 4.3: Visual evaluation of the classification accuracy for the public dataset. The left image shows a slice of a T2W MRI volume with annotated cancer regions (red). The image in the center shows the corresponding slice in the ADC map. The right image shows the predicted cancer regions, with red indicating a high probability and blue indicating a low probability for cancer.

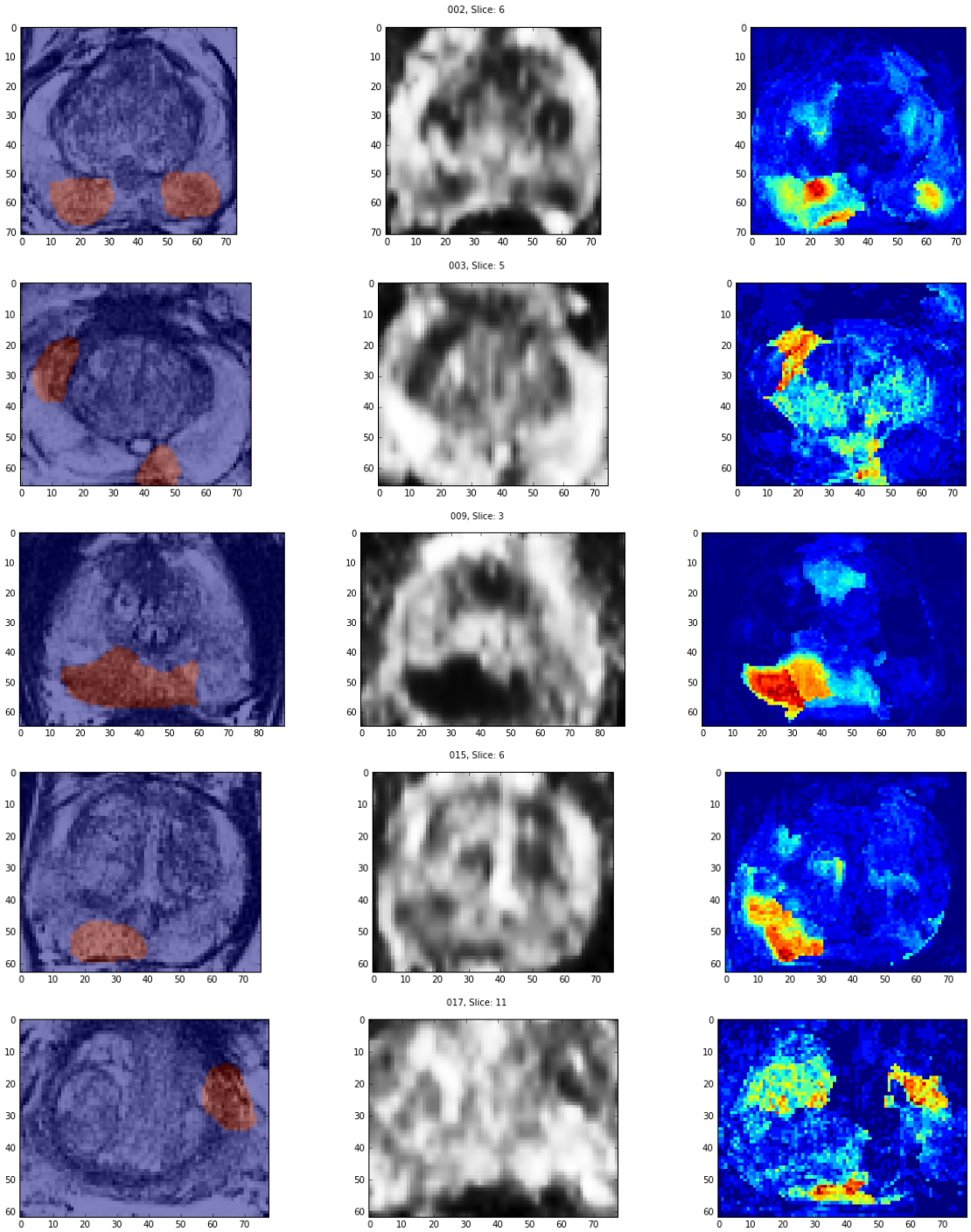


Figure 4.4: Visual comparison for the MUV dataset. The annotated cancer regions are overlayed on an axial T2W MRI slice (left), the ADC map (middle) and the result of the CAD system (right) are shown.

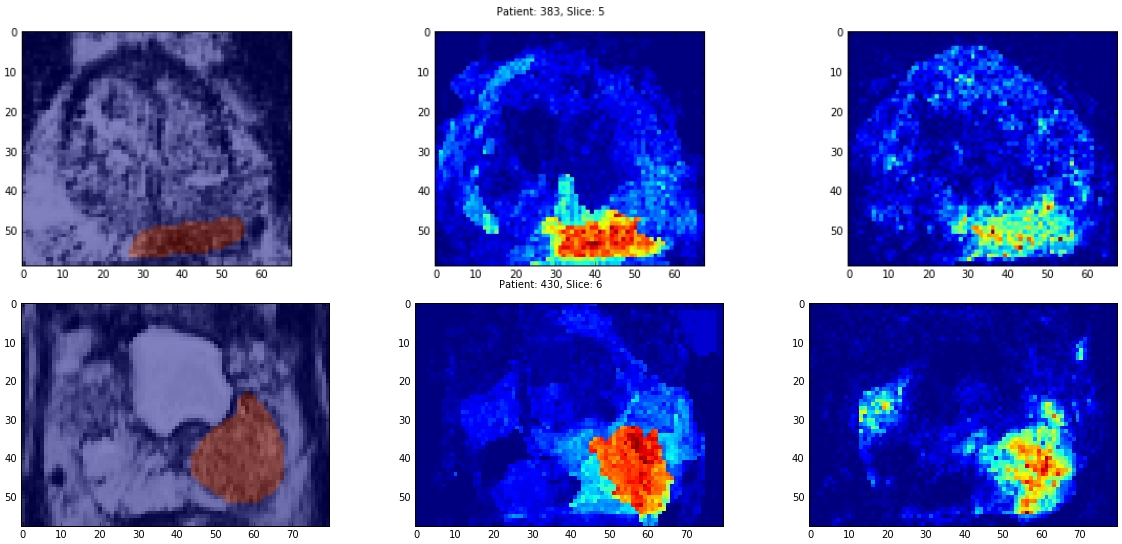


Figure 4.5: Visual Comparison how the classification changes when superpixel-based features are added. Left: Axial slice of a T2W MRI with annotated cancer regions (red). Center: Classification result with superpixel-based features. Right: Classification result with only pixel-based features. Red values indicate a high probability and blue values indicate a low probability for cancer.

the classification result is better when no superpixels are used. A possible explanation for this is that the result for the superpixel-based classification depends on the accuracy of the oversegmentation, which results in a higher performance when the superpixels are accurate, but also in worse performance when the superpixels do not accurately segment the CaP. However, when looking at a visual comparison (see Figure 4.5), it can be shown that the superpixel-based features increase the probability of a pixel that it is classified as cancerous at annotated regions, which leads to a higher sensitivity when the threshold is set to a higher value (resulting in a higher specificity). Thus, when adding superpixel-based features to the framework, the sensitivity is increased, especially at a high specificity.

4.2.2 MUV Dataset

Figure 4.2 shows that the superpixel features slightly increase classification performance from an AUC of 0.55 to 0.59, although the variability is higher. The higher variability may be due to superpixels being either accurate, resulting in an improved classification result, or not being accurate, resulting in a decreased classification result when compared to a classification without superpixels. The result highlights the importance of the oversegmentation being accurate to improve the accuracy of the classifier.

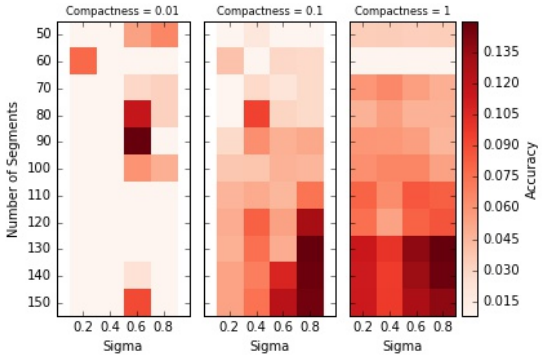


Figure 4.6: DICE of different parameter combinations. Left: Results with the compactness set to 0.01. Middle: Results with the compactness set to 0.1. Right: Results with the compactness set to 1. The red color indicates the accuracy of the superpixel segmentation, with dark red indicating a high accuracy and white indicating a low accuracy.

Number of Segments	**Compactness**	**Sigma**	**DICE**
150	1	0.2	0.5621
150	1	0.4	0.5602
130	1	0.2	0.5599
130	0.1	0.8	0.5593
150	1	0.8	0.5584
...			
20	0.01	0.4	0.1233

Table 4.3: The five parameter combinations with the best five and the worst DICE values for dataset 1. Additionally, the worst DICE value is shown.

4.3 Oversegmentation Accuracy

The superpixels are implemented with different combinations of parameter values (see Table 3.3 for the parameter space). In this section, the parameters that yields the best classification are presented.

4.3.1 Public Dataset

The CaP annotations in the public dataset have a 3D, coherent shape and span along several neighbouring slices. Therefore, SLIC is implemented in 3D to ensure that the superpixels have a similar shape and size in the 3D space. An implementation in 2D would result in superpixels that do not share the same properties as the cancer annotations. The best classification with an average AUC of 0.87 and an average DICE of 0.56 is achieved with the number of superpixels set to 150 per volume, the compactness set to 0.8 and sigma set to 0.8 (see Figure 4.6 and Table 4.3). The threshold for these parameter combinations was 0.1.

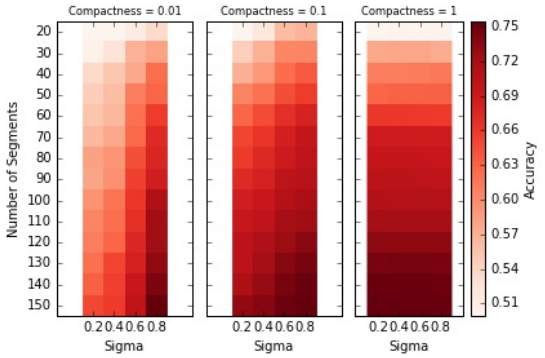


Figure 4.7: DICE of different parameter combinations. Left: Results with the compactness set to 0.01. Middle: Results with the compactness set to 0.1. Right: Results with the compactness set to 1. The red color indicates the accuracy of the superpixel segmentation, with dark red indicating a high accuracy and white indicating a low accuracy.

Number of Segments	**Compactness**	**Sigma**	**DICE**
150	1	0.2	0.7545
150	1	0.4	0.7540
150	1	0.6	0.7535
150	1	0.8	0.7529
140	1	0.2	0.7521
...			
50	0.01	0.2	0.2149

Table 4.4: The five parameter combinations with the best DICE values for dataset 2. Additionally, the worst DICE value is shown.

4.3.2 MUV Dataset

The SLIC algorithm in 2D yields the best classification result of an average AUC of 0.59 and a DICE of 0.75 for the MUV dataset. The number of superpixels is set to 140 per slice, the compactness to 1 and sigma to 0.2 (see Figure 4.7 and Table 4.4) at a threshold of 0.3. The reason that SLIC was implemented in 2D for this dataset is that the shape and location of the cancer annotations vary between the slices, and that they are not always located in adjacent slices, indicating that the annotations were made on a per-slice basis.

4.4 Feature Importance

In this section, it is examined which features contribute the most to the classification algorithm.

4.4.1 Public Dataset

As Figure 4.8 shows, the peripheral zone probability map is the most important feature, which indicates that the majority of CaPs are indeed located in this region of the prostate. The distance map is the second most important feature, which shows that the location of the CaP plays a significant role for the training of the classifier. The best four superpixel-based features are all from the DCE volume, followed by the feature that shows the mean value for each superpixel in the Gaussian-blurred T2W MRI. Generally, when looking at the 50 most important features, the first 30 features are primarily pixel-based features. However, three of ten most important features are superpixel-based, indicating that these features contribute to the accuracy of the classifier significantly.

Figure 4.9 shows that the superpixel-based features have a higher variability than the pixel-based features. The reasons for this may be that the superpixels are accurate in some mpMRI scans and inaccurate in others or that the CaP annotations are not accurate enough in some scans, showing that the superpixel-based features may be more sensible to wrong segmentations and annotations than pixel-based features.

4.4.2 MUV Dataset

Figure 4.10 shows that the original (T2W, ADC, DCE) and filtered (Gaussian, Median, Sobel, Prewitt) MRI volumes show a high feature importance. The functional feature kTrans and the GLCMs also play an important role in this dataset. Interestingly, the superpixel-based features do not show the same importance as in the public dataset, which, as mentioned in Chapter 4.2.2, may be due to the CaP annotations not being accurate enough. However, as in dataset 1, the anatomical features also show a high relevance, which further confirms that the location of CaP is a meaningful feature.

When looking at Figure 4.11, the features with the highest variability are mostly pixel-based ones (in contrast to dataset 1 where the features with the highest variability are mostly superpixel-based), which indicates that the CaP annotations may be inaccurate in some scans.

4.5 Discussion

A study by Rooji et. al. [dRHF+14] comes to a specificity and sensitivity of 0.74 and 0.88 when radiologists try to detect CaP with mpMRI at a combination of T2W, DWI and DCE MR images. When considering individual patients, the classifier performs better than the radiologists of the study [dRHF+14] and yields a sensitivity of 0.91 at a specificity of 0.74 for the public dataset and patient 383. However, the classifier performs worse than the radiologists for patient 002 of the MUV dataset at a specificity of 0.74 and a sensitivity of 0.79. Taking all patients into account and computing the average ROC curve, the radiologists perform slightly better, although their performance lies in the 95% confidence interval of the framework accuracy.

When looking at other CAD systems in the literature, our framework can be compared to the implementation by Litjens et. al. [LDB+14] because it also is a two-stage CAD system that has an automatic ROI detection. Having the best results in literature, Litjens et. al. report a mean AUC of 0.81 on a per-patient level which makes our implementation perform slighty better for the public dataset and slightly worse for the MUV dataset. However, these results are difficult to compare as the test data is completely different. Hence, we use a public dataset that is openly available to make our approach comparable for other implementations.

It is shown that the addition of superpixel-based features enhances the classificaton accuracy. Especially at values of high specificity (left parts of the ROC curves in Figures 4.1 and 4.2) the classifier shows a greater sensitivity when superpixels are added for generating features, which can also be confirmed by looking at the visual comparison of Figure 4.5. However, the higher variability of the superpixel-based features (see Figures 4.9 and 4.11) indicates that they may be more sensible to inaccurate segmentations or CaP annotations. This can also be shown when looking at the averaged ROC curves (see Figures 4.1 and 4.2), where the curves have a higher variability. Therefore, it is important to optimize the superpixel segmentations to achieve the best accuracy.

When comparing the predicted cancer regions and the annotated cancer regions (see Figures 4.3 and 4.4), it is noticeable that the classifier shows a high sensitivity for the detection of CaP. However, the specificity appears to be lower as regions that are normal are classified as cancerous. More precisely, when looking at individual ROC curves it is obvious that the sensitivity mostly reaches a value of 0.9 when the specificity is around 0.7.

Generally, as Figures 4.6 and 4.7 show, the superpixel algorithm yields a better CaP segmentation when the number of segments and sigma are in the higher end of the parameter range. The compactness does not play such a significant role.

4.6 Summary

This chapter presented the accuracy of the classification algorithm and the superpixels. It was shown that the superpixel-based features improved the classification result. In addition, it was explained how the superpixels were optimized and which parameter sets achieved the best accuracy and which features contributed the most to the classification model. To conclude, the results showed that the superpixel-based features improve the accuracy of the CAD system without the need for manual annotations.

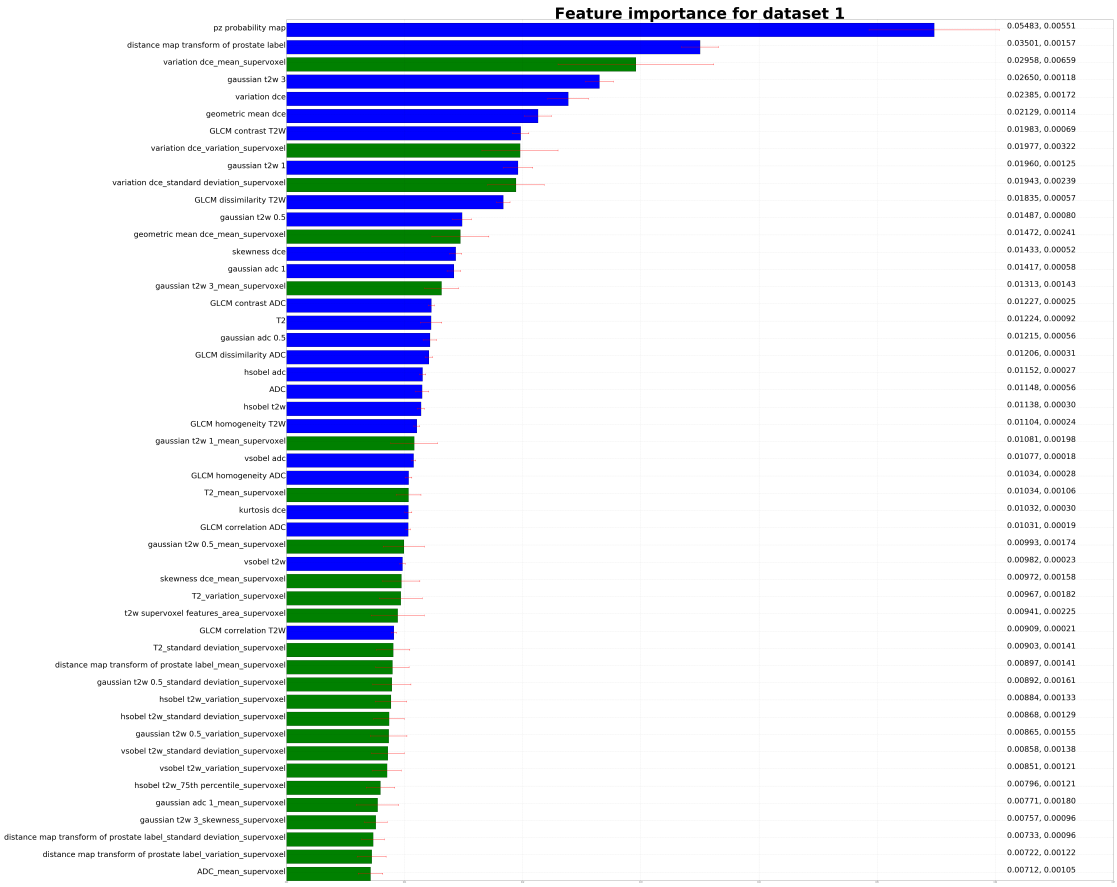


Figure 4.8: The most important features for the RF classifier for the public dataset. Blue indicates a pixel-based feature and green indicates a superpixel-based feature. The red lines on the right side of the feature bars show the variability and the numbers on the right show the importance and variability for each feature.

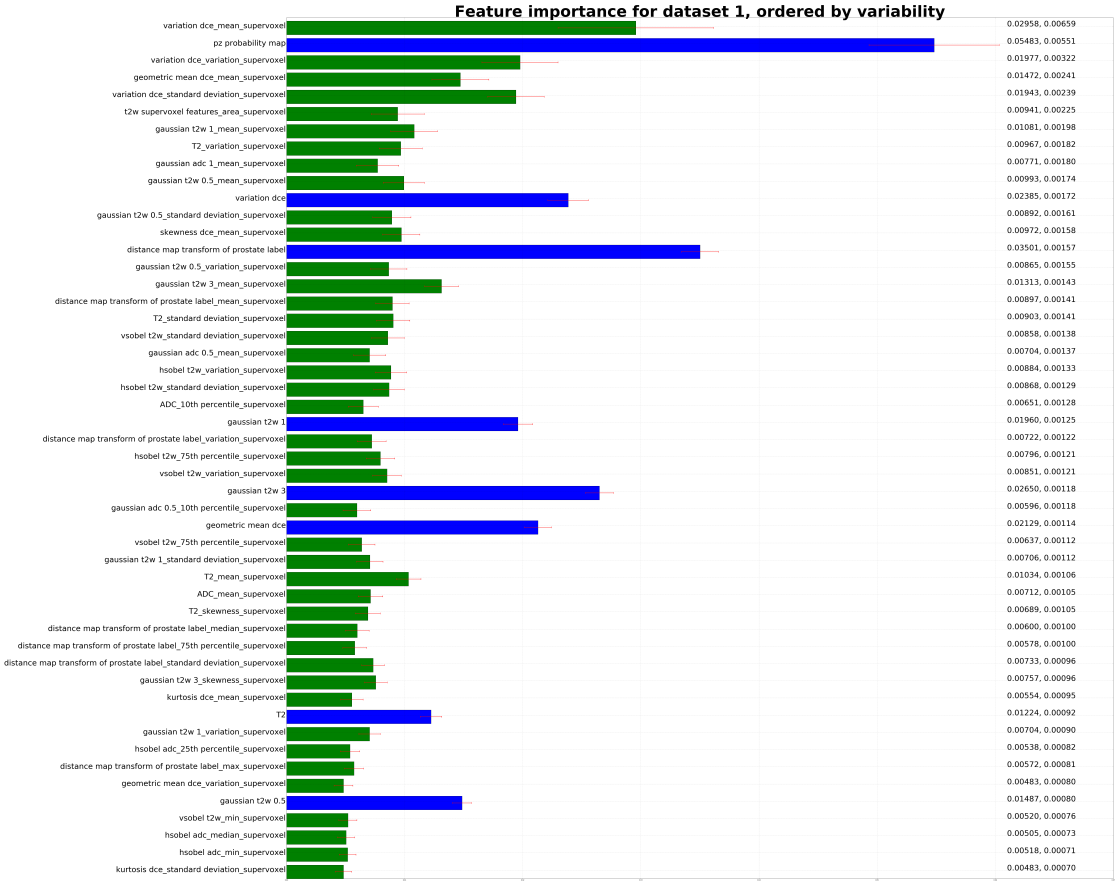


Figure 4.9: The features for the RF classifier for the public dataset with the highest variability. Blue indicates a pixel-based feature and green indicates a superpixel-based feature. The red lines on the right side of the feature bars show the variability and the numbers on the right show the importance and variability for each feature.

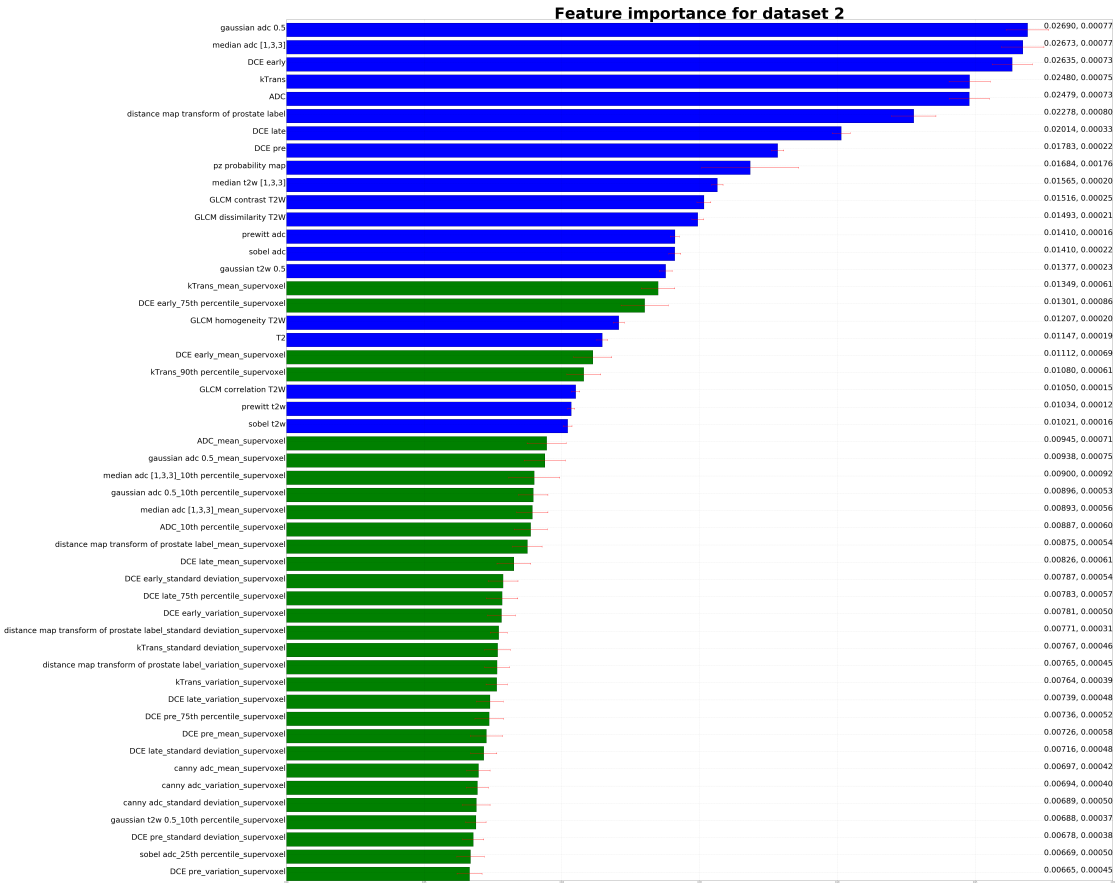


Figure 4.10: The most important features for the RF classifier for the MUV dataset. Blue indicates a pixel-based feature and green indicates a superpixel-based feature. The red lines on the right side of the feature bars show the variability and the numbers on the right show the importance and variability for each feature.

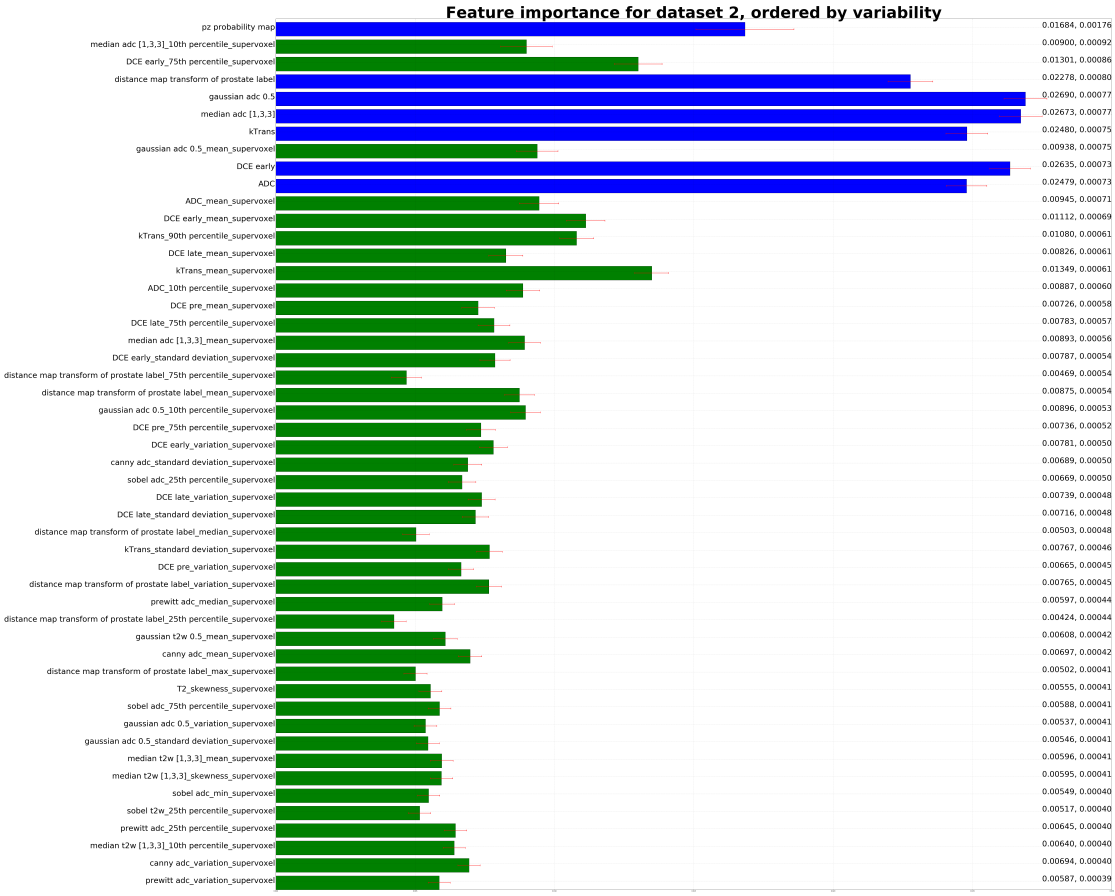


Figure 4.11: The features for the RF classifier for the public dataset with the highest variability. Blue indicates a pixel-based feature and green indicates a superpixel-based feature. The red lines on the right side of the feature bars show the variability and the numbers on the right show the importance and variability for each feature.

CHAPTER 5

Conclusion

This chapter gives an overview of this thesis, summarizes the results, contributions and limitations of this work, and proposes topics for future work.

5.1 Thesis Summary

This thesis proposes a CAD system that detects CaP based on pixel-wise and region-based statistical features. The ROIs for the features are generated automatically with an oversegmentation algorithm. Concretely, the superpixel algorithm SLIC [ASS+12] is used in a multi-modal configuration that uses mpMRI. Chapter 1 provides an introduction to the topic and describes the novel contributions of this work. Chapter 2 gives an overview of the topic and background information about MRI, CAD systems and superpixels. Furthermore, related work that is found in literature is discussed and compared. Chapter 3 describes the proposed CAD system and explains every step in detail. An overview about the frameworks used, the features that are computed and the superpixel algorithm SLIC [ASS+12] is given. Chapter 4 presents the results of the classification. The accuracy of the CAD system is compared to related work in literature and it is explained how the superpixel-based features improve the classification result. Additionally, it is shown how the superpixel parameters are optimized and which combination yields the best segmentation. Finally, the relevance of the features is shown for both datasets.

5.2 Contributions

The novel contribution of this work is the automatic ROI detection with superpixels. In general, there can be distuingished between two types of CAD systems. CADe systems detect the location of CaP on a pixel-wise basis. CADx systems give a diagnosis about manually annotated ROIs. This work is a combination of these two types and proposes the use of superpixels to automatically segment CaP and the rest of the prostate gland

into distinct ROIs. These ROIs are used to add region-based statistical features to the classifier to enhance the classification result. For the segmentation, SLIC [ASS+12] is used in either 2D or 3D. The parameters are chosen according to an optimization algorithm that tries different parameter values and chooses the combination that yields the best overlap of superpixels with CaP. The reason for this approach is that, due to the inter- and intra-patient variability in MRI scans [LMF+15, WBT+14] the parameters need to be adjusted for every new dataset to improve the classification result. The proposed approach can be used by other CAD systems that want to incorporate region-based statistical features into their framework without the need for manual annotation.

The following examinations are made regarding the oversegmentation and may be useful for other researchers that plan to implement superpixels into their CAD system:

- It is shown that superpixel-based features improve the classification result both in 2D and 3D.
- The superpixel-based features are more sensible to inaccurate segmentations or annotations than pixel-based features as they have a higher variability. Hence, it is important to optimize the parameters.
- The parameter optimization yields a more accurate oversegmentation when the superpixels have a more compact shape and the number of superpixels is high.

5.3 Conclusion

In this thesis, a CAD framework for the detection of CaP in mpMRI is presented. The novel contribution is an automatic oversegmentation of the prostate with a multi-modal superpixel algorithm to allow the computation of region-based statistical features without the need for manual annotation of CaP lesions. It is demonstrated that the superpixel-based features enhance the classification result and that the SLIC [ASS+12] algorithm can be applied in either 2D or 3D and in single- and multi-modal volumes. However, the higher variability in the AUC and the low feature importance of the superpixel-based features in dataset 2 show that if the superpixels are not accurate, the resulting classification may be worse, which may be the case for volumes that have inaccurate CaP annotations. The oversegmentation is done automatically by the framework, with an optizimation step to achieve the best segmentation result. The advantage of parameter optimization is that a better segmentation of the different tissues of the prostate greatly improves the importance of the features calculated from the superpixels, which results in a better classification. Our tests show that features that correlate with CaP lesions (e.g., the 10th percentile and average of the ADC map [PJY+13]) show a bigger feature importance when the accuracy is high. The output of the framework is a map that shows the probability of cancer for each pixel. The result of this thesis is an easy to use, flexible and efficient CAD system for the detection of CaP. It could be refined and used for automating the diagnosis of CaP or as additional information for guided biopsies.

When compared two releated work in literature, the CAD systems by Litjens et. al. [LDB+14] and Vos et. al. [VBKH12] are the most similar ones. The system by Litjens et. al. [LDB+14] first predicts the probability of CaP for each pixel. Secondly, they perform local maxima detection on the generated probability map and segment the area surrounding each maxima. The resulting ROIs are then used for computing region-based statistical features. The CAD systems performs with an average AUC of 0.81 on a per-patient level. The CAD system by Vos et. al. [VBKH12] uses blob detection to find maxima in the ADC map. Subsequently, the ROIs are segmented and classified using statistical region features. The authors obtain sensitivities of 0.15, 0.48 and 0.89 at false positive levels of 0.1, 1 and 10 per patient. However, the CAD system is outperformed by Litjens et. al. [LDB+14] with sensitivities of 0.42, 0.75 and 0.89. When comparing the AUC of the CAD system Litjens et. al. [LDB+14] with the proposed approach, our classifier yields a better result for dataset 1 with an average AUC of 0.87. However, for dataset 2 the AUC is 0.59. Generally, the CAD systems are difficult to compare because the datasets are not the same.

5.4 Limitations and Future Work

One of the limitations of the proposed CAD system is that the prostate has to be manually segmented, requiring an expert reader for new MR images that need to be classified. While the need for manual annotation by the user is reduced in this framework by using superpixels, further automating the workflow would be even more beneficial. An automatic segmentation of the prostate would require no manual input by the user at all, resulting in a fully automatic CAD system. While this step is out of scope for this thesis, several prostate segmentation algorithms can be found in the literature and could be added to this CAD system.

The framework has been tested on two distinct datasets. However, for each dataset the classifier has been trained seperately. Making predictions for new MRI data from other datasets without prior training on this data is not trivial because MR images have different intensity values for the same tissue, even if it is only a different patient. Future research could focus on how the classifier performs if it is applied to new test MR images without prior training on the new test dataset (or in other words, if the train and test datasets are different). Furthermore, as the datasets are relatively small (20 and 25 patients), it would be desirable to test the performance of the CAD system on a bigger dataset.

There are approaches in literature that try not to only detect, but also determine the GS of the cancer lesions. There are features that correlate with the GS, so it is feasible to try and stage the tumors with the proposed CAD framework.

The superpixels are calculated on intensity values of T2W MRI and the ADC map. It would be possible to experiment with including higher-order features in the superpixel calculation (e.g., Haralick features) and examine if it improves the segmentation accuracy.

This CAD system uses automatic superpixel oversegmentations for region-based features. They are optimized to find the best fit to pre-defined cancer regions in the training data. An idea for future work is to investigate the usefuleness of the framework for anatomical regions other than the prostate, because the workflow of the framework is similar as long as the region has annotated cancer lesions.

Bibliography

[AAN+09] Levinshtein A., Stere A., Kutulakos K. N., Fleet D. J., Dickinson S. J., and Siddiqi K. Turbopixels: Fast superpixels using geometric flows. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(12):2290–2297, December 2009.

[ASS+10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels. Technical report, EPFL, 2010.

[ASS+12] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, Nov 2012.

[BRC+12] J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V Logager, and Jurgen J. Fütterer. Esur prostate MR guidelines 2012. *European Radiology*, 22(4):746–757, 2012.

[Bre01] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[CKS+15] A. G. Chung, F. Khalvati, M. J. Shafiee, M. A. Haider, and A. Wong. Prostate cancer detection via a quantitative radiomics-driven conditional random field framework. *IEEE Access*, 3:2531–2541, 2015.

[CM02] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, May 2002.

[CWM+03] I. Chan, W. Wells, R. V. Mulkern, S. Haker, J. Zhang, K. H. Zou, S. E. Maier, and C. M. C. Tempany. Detection of prostate cancer by integration of line-scan diffusion, t2-mapping and t2-weighted magnetic resonance imaging; a multichannel statistical classifier. *Medical Physics*, 30(9):2390–2398, 2003.

[DAA+11] L. Dickinson, H. U. Ahmed, C. Allen, J. O. Barentsz, B. Carey, J. J. Futterer, S. W. Heijmink, P. J. Hoskin, A. Kirkham, A. R. Padhani, R. Persad, P. Puech, S. Punwani, A. S. Sohaib, B. Tombal, A. Villers,

J. van der Meulen, and M. Emberton. Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: Recommendations from a European consensus meeting. *European Urology*, 59(4):477 – 494, 2011.

[DKM+14] D. Duda, M. Kretowski, R. Mathieu, R. de Crevoisier, and J. Bezy-Wendling. *Multi-Image Texture Analysis in Classification of Prostatic Tissues from MRI. Preliminary Results*, pages 139–150. Springer International Publishing, Cham, 2014.

[dRHF+14] M. de Rooij, E. H. J. Hamoen, J. J. Fütterer, J. O. Barentsz, and M. M. Rovers. Accuracy of multiparametric MRI for prostate cancer detection: A meta-analysis. *American Journal of Roentgenology*, 202(2):343 – 351, 2014.

[FBKC+12] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. M. Fennessy, M. Sonka, J. Buatti, S. R. Aylward, J. V. Miller, S. Pieper, and R. Kikinis. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–41, 11 2012.

[FBV+15] J. J. Fütterer, A. Briganti, P. De Visschere, M. Emberton, G. Giannarini, A. Kirkham, S. S. Taneja, H. Thoeny, G. Villeirs, and A. Villers. Can clinically significant prostate cancer be detected with multiparametric magnetic resonance imaging? A systematic review of the literature. *European Urology*, 68(6):1045 – 1053, 2015.

[FH04] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[FSB+10] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer*, 127(12):2893–2917, 2010.

[GLT+81] G. L. Gomella, S. X. Liu, J. E. Trabulsi, K. Wm. Kelly, R. Myers, Showalter T., A. Dicker, and R. Wender. Screening for prostate cancer: the current evidence and guidelines controversy. *The Canadian Journal of Urology*, 18(5):5875–5883, 1981.

[HBH+11] C. M. A. Hoeks, J. O. Barentsz, T. Hambrock, D. Yakar, D. M. Somford, S. W. T. P. J. Heijmink, T. W. J. Scheenen, P. C. Vos, H. Huisman, I. M. van Oort, J. A. Witjes, A. Heerschap, and J. J. Fütterer. Prostate cancer: Multiparametric MR imaging for detection, localization, and staging. *Radiology*, 261(1):46–66, 2011.

[HSD73] R. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition, 2009.

[JMIC13] H. J. Johnson, M. McCormick, L. Ibáñez, and The Insight Software Consortium. *The ITK software guide*. Kitware, Inc., third edition, 2013.

[JOP+] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed 21.04.2017].

[KWH15] F. Khalvati, A. Wong, and M. A. Haider. Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. *BMC Medical Imaging*, 15(1):27, 2015.

[KXW+15] J. T. Kwak, S. Xu, B. J. Wood, B. Turkbey, P. L. Choyke, P. A. Pinto, S. Wang, and R. M. Summers. Automated prostate cancer detection using t2-weighted and high-b-value diffusion-weighted magnetic resonance imaging. *Medical Physics*, 42(5):2368–2378, 2015.

[LAT+16] L. Lindenberg, M. Ahlman, B. Turkbey, E. Mena, and P. Choyke. Evaluation of prostate cancer with PET/MRI. *Journal of Nuclear Medicine*, 57(Supplement 3):111S–116S, 2016.

[LBN+14] S. Loeb, M. A. Bjurlin, J. Nicholson, T. L. Tammela, D. F. Penson, H. Ballentine Carter, P. Carroll, and R. Etzioni. Overdiagnosis and overtreatment of prostate cancer. *European Urology*, 65(6):1046 – 1055, 2014.

[LD+14] G. Litjens, O. Debats, et al. Computer-aided detection of prostate cancer in MRI. *IEEE Transactions on Medical Imaging*, 33(5), May 2014.

[LDB+14] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman. Computer-aided detection of prostate cancer in MRI. *IEEE Transactions on Medical Imaging*, 33(5):1083–1092, May 2014.

[LDvdV+12] G. Litjens, O. Debats, W. van de Ven, N. Karssemeijer, and H. Huisman. *A Pattern Recognition Approach to Zonal Segmentation of the Prostate on MRI*, pages 413–420. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[LMF+15] G. Lemaitre, R. Martí, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review. *Computers in Biology and Medicine*, 60:8 – 31, 2015.

[LvdKE+09] D. L. Langer, T. H. van der Kwast, A. J. Evans, J. Trachtenberg, B. C. Wilson, and M. A. Haider. Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative t2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *Journal of Magnetic Resonance Imaging*, 30(2):327–334, 2009.

[LY11] X. Liu and I. S. Yetik. Automated prostate cancer localization without the need for peripheral zone extraction using multiparametric MRIa. *Medical Physics*, 38(6):2986–2994, 2011.

[McN81] J. E. McNeal. The zonal anatomy of the prostate. *The Prostate*, 2(1):35–49, 1981.

[MFM+05] A. Madabhushi, M. D. Feldman, D. N. Metaxas, J. Tomaszeweski, and D. Chute. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI. *IEEE Transactions on Medical Imaging*, 24(12):1611–1625, Dec 2005.

[MTT12] F. H. Martini, M. J. Timmons, and R. B. Tallitsch. *Human Anatomy*, volume 7. Pearson, 2012.

[NFR+14] E. Niaf, R. Flamary, O. Rouvière, C. Lartizien, and S. Canu. Kernel-based learning from both qualitative and quantitative labels: Application to prostate cancer diagnosis based on multiparametric MR imaging. *IEEE Transactions on Image Processing*, 23(3):979–991, March 2014.

[NRML+12] E. Niaf, O. Rouvière, F. Mège-Lechevallier, F. Bratan, and C. Lartizien. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Physics in Medicine and Biology*, 57(12):3833, 2012.

[OHL+09] S. Ozer, M. A. Haider, D. L. Langer, T. H. van der Kwast, A. J. Evans, M. N. Wernick, J. Trachtenberg, and I. S. Yetik. Prostate cancer localization with multispectral MRI based on relevance vector machines. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 73–76, June 2009.

[OLL+10] S. Ozer, D. L. Langer, X. Liu, M. A. Haider, T. H. van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick, and I. S. Yetik. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Medical Physics*, 37(4):1873–1883, 2010.

[PCVH10] J. O Barenstz P. C Vos, T. Hambrock and H. J Huisman. Computer-assisted analysis of peripheral zone prostate lesions using t2-weighted and dynamic contrast enhanced t1-weighted MRI. *Physics in Medicine and Biology*, 55(6):1719–1734, 2010.

[PEND+16] M. Piert, I. El Naqa, M. S. Davenport, E. Incerti, P. Mapelli, and M. Picchio. PET/MRI and prostate cancer. *Clinical and Translational Imaging*, 4(6):473–485, 2016.

[PJY+13] Y. Peng, Y. Jiang, C. Yang, J. B. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M. L. Giger, S. E. Eggener, and A. Oto. Quantitative analysis of multiparametric prostate MR images: Differentiation between prostate cancer and normal tissue and correlationwith gleason score-a computer-aided diagnosis development study. *Radiology*, 267(3):787–796, 2013. PMID: 23392430.

[PMK+16] M. Piert, J. Montgomery, L. P. Kunju, J. Siddiqui, V. Rogers, T. Rajendiran, T. D. Johnson, X. Shao, and M. S. Davenport. 18f-choline PET/MRI: The additional value of PET for MRI-guided transrectal prostate biopsies. *Journal of Nuclear Medicine*, 57(7):1065–1070, 2016.

[PVG+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[RCM+16] A. Rampun, Z. Chen, P. Malcolm, B. Tiddeman, and R. Zwiggelaar. Computer-aided diagnosis: detection and localization of prostate cancer within the peripheral zone. *International Journal for Numerical Methods in Biomedical Engineering*, 32(5):e02745–n/a, 2016. e02745 cnm.2745.

[RM03] X. Ren and J. Malik. Learning a classification model for segmentation. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 10–, Washington, DC, USA, 2003. IEEE Computer Society.

[RTZM16] A. Rampun, B. Tiddeman, R. Zwiggelaar, and P. Malcolm. Computer aided diagnosis of prostate cancer: A texton based approach. *Medical Physics*, 43(10):5412–5425, 2016.

[SBL98] D. Shier, J. Butler, and R. Lewis. *Hole's Essentials of Human Anatomy and Physiology*, volume 6. McGraw-Hill Science/Engineering/Math, 1998.

[SKR+98] F. H. Schröder, A. B. Kruger, J. Rietbergen, R. Kranse, P. van der Maas, P. Beemsterboer, and R. Hoedemaeker. Evaluation of the digital rectal examination as a screening test for prostate cancer. *Journal of the National Cancer Institute*, 90(23):1817–1823, 1998.

[SM00] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, August 2000.

[SMJ17] R. L. Siegel, K. D. Miller, and A. Jemal. Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, 67(1):7–30, 2017.

[Sta15] Statistik Austria. Österreichisches Krebsregister (Stand 02.10.2015)), 2015.

[SvdCKdK+00] F. H. Schröder, I. van der Cruijsen-Koeter, H. J. de Koning, A. N. Vis, R. F. Hoedemaeker, and R. Kranse. Prostate cancer detection at low prostate specific antigen. *The Journal of Urology*, 163(3):806 – 812, 2000.

[Tam97] T. Tammela. Benign prostatic hyperplasia. *Drugs & Aging*, 10(5):349–366, 1997.

[TLZF16] Z. Tian, L. Liu, Z. Zhang, and B. Fei. Superpixel-based segmentation for 3d prostate MR images. *IEEE Transactions on Medical Imaging*, 35(3):791–801, March 2016.

[TVK+12] P. Tiwari, S. Viswanath, J. Kurhanewicz, A. Sridhar, and A. Madabhushi. Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR in Biomedicine*, 25(4):607–619, 2012.

[VBKH12] P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman. Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis. *Physics in Medicine and Biology*, 57(6):1527, 2012.

[VS08] A. Vedaldi and S. Soatto. *Quick Shift and Kernel Methods for Mode Seeking*, pages 705–718. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[WBC+16] J. C. Weinreb, J. O. Barentsz, P. L. Choyke, F. Cornud, M. A. Haider, K. J. Macura, D. Margolis, M. D. Schnall, F. Shtern, C. M. Tempany, H. C. Thoeny, and S. Verma. PI-RADS prostate imaging – reporting and data system: 2015, version 2. *European Urology*, 69(1):16 – 40, 2016.

[WBT+14] S. Wang, K. Burtt, B. Turkbey, P. Choyke, and R. M. Summers. Computer aided-diagnosis of prostate cancer on multiparametric MRI: A technical review of current research. *BioMed Research International*, 2014.

[YVM+12] Joseph H. Y., Sadhna V., Jonathan S. M., Scott E., and Aytekin O. Imaging-guided prostate biopsy: Conventional and emerging techniques. *RadioGraphics*, 32(3):819–837, 2012. PMID: 22582361.