

Aus der
Radiologischen Universitätsklinik Tübingen
Abteilung Diagnostische und Interventionelle Radiologie

**Strukturierte Befundung der multiparametrischen
Magnetresonanztomographie der Prostata unter
Anwendung Künstlicher Intelligenz**

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Medizin

der Medizinischen Fakultät
der Eberhard Karls Universität
zu Tübingen

vorgelegt von
Reiß, Sebastian

2023

Dekan: Professor Dr. B. Pichler
1. Berichterstatter: Professor Dr. K. Nikolaou
2. Berichterstatter: Professor Dr. G. Bier

Tag der Disputation: 23.08.2023

Inhaltsverzeichnis

Inhaltsverzeichnis	II
1 Einleitung	1
1.1 Anatomie der Prostata	1
1.2 MRT-Bildgebung der Prostata: Methoden- und Indikationsbeschreibung	8
1.3 Künstliche Intelligenz in der radiologischen Bildgebung	13
2 Material und Methoden	22
2.1 Studiendesign	22
2.1.1 Einschlusskriterien und Zeitraum der Fallsammlung	23
2.1.2 Multiparametrische Magnetresonanztomographie	24
2.1.3 Verwendete Software	24
2.2 Studienablauf	25
2.2.1 mpMRT und Prostatektomien	25
2.2.2 Retrospektive Erhebung klinischer Daten	25
2.2.3 Segmentierung der Prostata	26
2.2.4 Trainieren und Anwenden der KI	29
2.3 Statistische Auswertung	31
2.3.1 Sørensen-Dice-Koeffizient	31
2.3.2 Äquivalenztest	32
2.3.3 Verallgemeinernde Schätzgleichung	33
2.3.4 Visuelle Analyse der KI-basierten Segmentierungen	35
3 Ergebnisse	41
3.1 Studienpopulation der Stichprobe Tübingen	41
3.2 Trainings- und Testdatensatz der Stichprobe Tübingen	55
3.3 Vergleich der Dice-Koeffizienten	61
3.4 Einflussfaktoren	63
3.5 Visuelle Untersuchung der automatisierten Segmentierungen	66
4 Diskussion	68
4.1 Zusammenfassung der Ergebnisse	68
4.1.1 Vergleich der Dice-Koeffizienten mit vergleichbaren Studien	68
4.1.2 Verallgemeinernde Schätzgleichung zur Analyse der Einflussfaktoren	69
4.1.3 Auswahl der Prädiktorvariablen und Interpretation der Regressionsparameter	69
4.1.4 Visuelle Analyse der Segmentierungen	71
4.1.5 Bizentrisches Studiendesign	72
4.1.6 Auswahl der Patienten, Stichprobengröße und Tumorstatus der Patienten	73

4.2	Limitationen	75
4.2.1	Durchführung der manuellen Segmentierung	75
4.2.2	Dice-Koeffizient als Maß der Performanz einer KI zur Bildsegmentierung	77
4.2.3	Berechnung der Dice-Koeffizienten nach Prostataabschnitt.....	77
4.2.4	Redundanz zweier Bilddatensätze.....	78
4.3	Künftige und an diese Arbeit anknüpfende Forschung.....	79
4.3.1	Untersuchen weiterer Prädiktorvariablen.....	79
4.3.2	Evaluierung der manuellen Segmentierung als Grundwahrheit.....	79
4.3.3	Segmentieren der Prostatazonen und suspekter Läsionen unter Berücksichtigung aller mpMRT-Sequenzen	80
5	Zusammenfassung.....	81
6	Literaturverzeichnis	83
7	Abbildungsverzeichnis	90
8	Tabellenverzeichnis	92
9	Erklärung zum Eigenanteil	93

Abkürzungsverzeichnis

2D	2-dimensional
3D	3-dimensional
BDR	<i>Bund Deutscher Radiologen</i>
CAD	<i>computer-aided diagnosis</i>
CNN	<i>Convolutional Neural Networks</i>
CT	<i>Computertomograph(ie)</i>
DCE	<i>dynamic contrast-enhanced</i>
DRG	<i>Deutsche Röntgengesellschaft</i>
DRU	<i>digitale rektale Untersuchung</i>
DWI	<i>diffusion weighted imaging</i>
ESUR	<i>European Society of Urogenital Radiology</i>
FoV	<i>Field of View</i>
FSE	<i>Fast spin echo</i>
GEE	<i>Generalized Estimating Equations</i>
GLM	<i>generalized linear model</i>
ILSVRC	<i>ImageNet Large Scale Visual Recognition Challenge</i>
IoU	<i>Intersection over union</i>
k.A.	<i>keine Angabe</i>
KI	<i>Künstliche Intelligenz</i>
MRT	<i>Magnetresonanztomograph(ie)</i>
PET	<i>Positronenemissionstomograph(ie)</i>
PI-RADS	<i>Prostate Imaging-Reporting and Data System</i>
PSA	<i>Prostata-spezifisches Antigen</i>
SESOI	<i>smallest effect size of interest</i>
T1w	<i>T1 weighted (T1-gewichtete Sequenz)</i>
T2w	<i>T2 weighted (T2-gewichtete Sequenz)</i>
TOST	<i>two one-sided tests</i>
TRUS	<i>transrektale Ultraschalluntersuchung</i>
TSE	<i>Turbo spin echo</i>
T-Zone	<i>Transitionszone</i>

1 Einleitung

Die folgende Arbeit beschäftigt sich mit der auf künstlicher Intelligenz basierten (KI-basierten) Segmentierung der Prostata in magnetresonanztomographisch aufgenommenen Bildsequenzen des kleinen Beckens und untersucht deren Qualität und Genauigkeit in Abhängigkeit von verschiedenen klinischen Parametern.

In Hinführung auf den Forschungsgegenstand wird zunächst Grundlegendes zur Anatomie sowie der Magnetresonanztomographie (MRT) der Prostata, hier speziell im Hinblick auf das Prostatakarzinom, erläutert. Außerdem soll ein Überblick zur KI-basierten Analyse radiologischer Bilddaten sowie zur automatisierten Segmentierung morphologischer Organstrukturen gegeben werden.

1.1 Anatomie der Prostata

Der folgende Abschnitt zur Anatomie der Prostata orientiert sich vorrangig an denjenigen anatomischen Strukturen und Kenngrößen, die im Verlauf der Arbeit Erwähnung finden und für das Verständnis von Bedeutung sind.

Neben den beiden paarigen Bläschen- (Glandula vesiculosa) und Bulbourethraldrüsen (Glandula bulbourethralis oder nach dem englischen Anatom William Cowper auch Cowpersche Drüsen) zählt die Prostata zu den akzessorischen Geschlechtsdrüsen des Mannes (Schulte, 2020). Ihre Größe entspricht in etwa der einer dicken Esskastanie (Sökeland et al., 2008) mit näherungsweise folgenden Kenngrößen: Sagittaldurchmesser: ca. 2 bis 3 cm; Breite: ca. 4 cm; Dicke: ca. 1 bis 2cm; Masse: ca. 20 g (Schünke et al., 2018). Berges et al. (2011) untersuchten in ihrer Studie unter anderem das altersspezifische Prostatavolumen bei 1240 Patienten über 50 Jahren. Das Volumen wurde dabei sonographisch erhoben und reichte von durchschnittlich 24 cm³ bei der Altersgruppe der 50- bis 54-Jährigen bis zu 38 cm³ bei den über 75-Jährigen (Berges & Oelke, 2011).

Die Prostata oder auch Vorsteherdrüse ist extra- bzw. subperitoneal verortet und lässt sich von kranial nach kaudal in die Abschnitte Prostatabasis, Drüsenmitte und Prostataspitze, den Apex, gliedern (McLaughlin et al., 2005). Abbildung 1 zeigt diese Höhenbezeichnungen schematisch an einem Sagittalschnitt der Prostata.

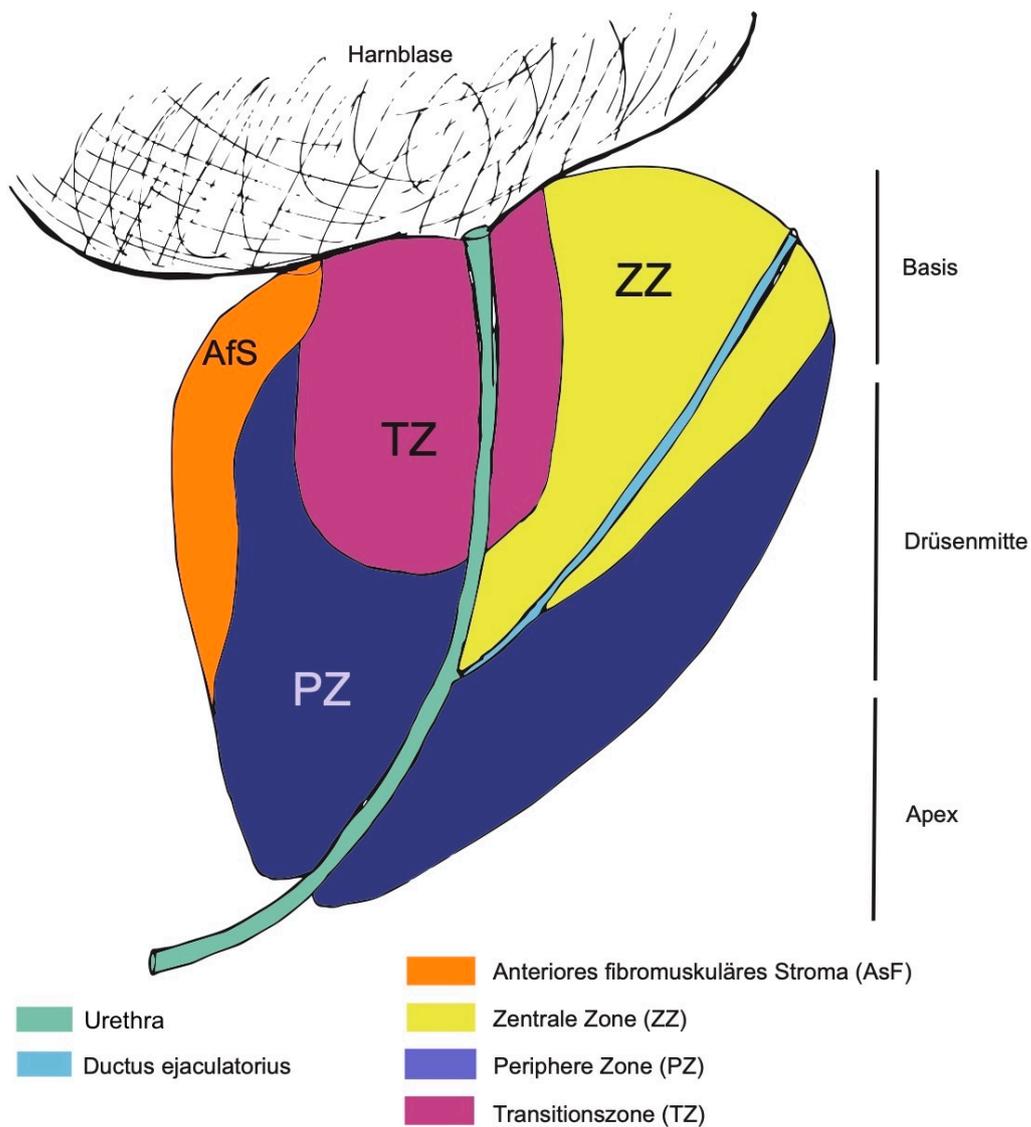


Abbildung 1. Schematischer Sagittalschnitt der Prostata mit Angabe der Prostataabschnitte und der zonalen Gliederung, abgezeichnet nach Jensen et al. (2019).

Die Prostatabasis liegt direkt unterhalb der Harnblase an deren Ausgang (Schulte, 2020). Ein Teil der Harnröhre, der Pars prostatica urethrae, wird von der Prostata umgeben (Schulte, 2020). Ausläufer der longitudinalen Muskelschicht der Harnblase bilden mit der longitudinalen Muskelschicht des präprostatistischen Sphinkters, einem Schließmechanismus, der eine retrograde Ejakulation verhindern und Urinkontinenz bewahren soll, die präprostatistische Harnröhre; der Begriff „präprostatistisch“ ist etwas irreführend, da auch dieser Abschnitt der Urethra von Prostatagewebe umschlossen ist (McLaughlin et al., 2005). Sie geht in den Samenhügel (Verumontanum) über, wo der paarig angelegte Ductus ejaculatorius, der gemeinsame Ausführungsgang von Samenbläschen und Samenleiter, mündet.

Die Prostata wird durch die vordere Rektumwand circa 4 cm oberhalb des Afters nach dorsal begrenzt (Schünke et al., 2018), was sie unter anderem einer digitalen rektalen Untersuchung (DRU) oder einer transrektalen Ultraschalluntersuchung (TRUS) zugänglich macht. Getrennt von der Rektumampulle wird die Prostata durch die sogenannte Denonvilliersche Faszie, kollagenem Bindegewebe, das 1836 nach seinem Erstbeschreiber Charles-Pierre Denonvillier benannt wurde und die posteriore Grenze der Prostata zum Rektum darstellt (Tzelves et al., 2022).

Die Prostata wird in zwei Seitenlappen (Lobus dexter und Lobus sinister) unterteilt, die nach dorsal durch den Lobus medius sowie nach ventral durch eine schmale Gewebsbrücke, den Isthmus prostaticae, miteinander in Verbindung treten (Schünke et al., 2018). Diese anatomische Einteilung nach Prostatalappen wurde nach und nach von einer zonalen Gliederung der Prostata abgelöst; vergleiche Abbildung 2 (McLaughlin et al., 2005).

So unterteilte bereits der Pathologe John E. McNeal (1988) die Prostata in ihrer Funktion als Drüse bezüglich der unterschiedlichen Histologie und Biologie in drei Hauptregionen: einer peripheren, einer zentralen und einer Transitionalzone.

Die periphere Zone der Prostata bildet mit circa 70 % Volumenanteil den größten Bereich des Organs (Bhavsar & Verma, 2014). Sie liegt dorsolateral sowie kaudal. Die zentrale Zone macht circa 25 % des Volumens der Prostata aus und

liegt einem mit der Spitze nach kaudal ausgerichteten Keil nachempfunden am kranialen Prostatapol (Aumüller et al., 2020; Bhavsar & Verma, 2014). Die Transitionszone („T-Zone“) stellt sich paarig als zwei Bereiche seitlich des proximalen Segments der Urethra dar und repräsentiert nur etwa 5 bis 10 % des Prostatagewebes. Die Periurethralzone als vierte Drüsenregion neben den drei oben genannten Hauptregionen umhüllt einem schmalen Saum gleich die proximale Urethra (Aumüller et al., 2020; McNeal, 1988; Schulte, 2020).

Das ventral gelegene anteriore fibromuskuläre Stroma als fünfte Organregion ist frei von Drüsen (Schulte, 2020).

Abbildungen 2 und 3 zeigen schematisch die zonale Gliederung der Prostata, deren Kenntnis in der MRT-Bildgebung speziell im Hinblick auf die Pathogenese und Lokalisation eines Prostatakarzinom von Bedeutung ist (McLaughlin et al., 2005; McNeal, 1981). In Abbildung 2 kommt der Samenhügel überproportional groß zur Darstellung, das Augenmerk soll vorwiegend auf die Einteilung der Prostata in Zonen gelegt werden. Die horizontalen Linien der beiden Zeichnungen auf der rechten Bildhälfte in Abbildung 3 stellen grob die Schnitte der Prostatabasis und des Apex, zu sehen auf der linken Bildhälfte, dar.

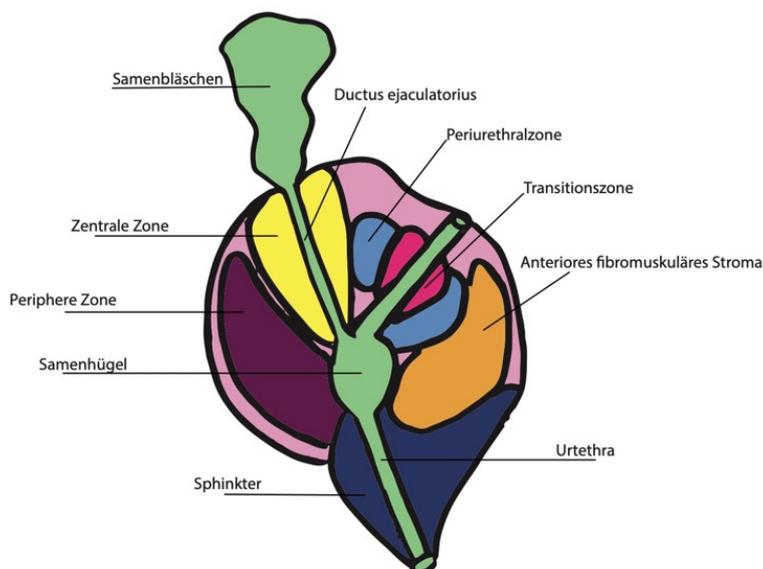


Abbildung 2. Schematischer Sagittalschnitt der Prostata mit zonaler Gliederung, abgezeichnet nach McLaughlin et al. (2005)

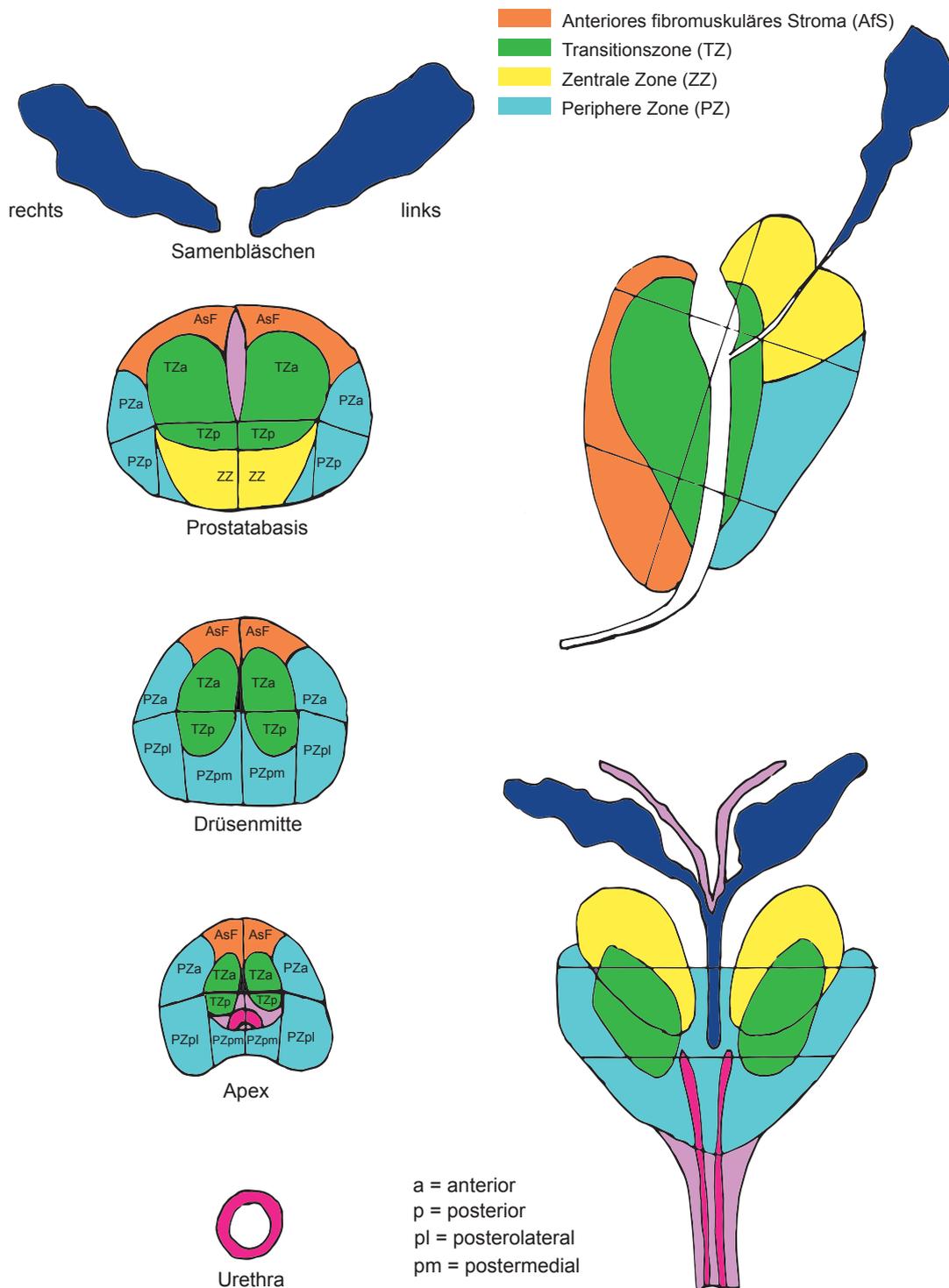


Abbildung 3. Zonale Gliederung der Prostata nach Weinreb et al. (2016)

Ein kräftiger Kapselapparat, bestehend aus fibrösem Bindegewebe sowie glatter Muskulatur, umhüllt die Prostata (Schulte, 2020). Der Begriff Prostatakapsel wird allerdings vor allem aus pathologischer Sicht kontrovers diskutiert, da auch die äußere Begrenzung der Prostata variabel von Drüsen durchzogen ist und Züge der glatten Muskulatur mit dem periprostatishen Gewebe verflochten sein können (Ayala et al., 1989; Walz et al., 2016). Aus onkologischen, bildmorphologischen und auch operativen Gesichtspunkten hat die Prostatakapsel jedoch eine große Bedeutsamkeit, denn gerade in der MRT-Bildgebung dient sie einer Abgrenzung gegenüber der Prostataumgebung (Villeirs & De Meerleer, 2007). In MRT-Befunden und Operationsberichten sind Formulierungen wie „kapselüberschreitendes Wachstum“, „Kapselinvasion“ oder „breitbasiger Kapselkontakt“ üblich, um die Ausbreitung eines Prostatakarzinoms zu verdeutlichen, was therapeutische und prognostische Relevanz für den jeweiligen Patienten haben kann.

Die Innervation durch das autonome Nervensystem hinsichtlich der Harnkontinenz, der Erektion des männlichen Gliedes und der Ejakulation erfolgt durch den Plexus hypogastricus inferior (Walz et al., 2016). Dieser liegt in vorwiegend sagittaler Ausrichtung eingeschlossen von einer dünnen Gewebsplatte zwischen Blase und Rektum (Walz et al., 2016) und erhält efferente parasympathische präganglionäre Nervenfasern des Sakralmarks S2 - S4 sowie sympathische Nervenfasern des thorakolumbalen Truncusanteils Th11 - L2 (Walsh et al., 1983; Yucel et al., 2005).

Walsh und Donker untersuchten bereits 1982 die Ursache für erektile Dysfunktion nach stattgefundenener radikaler Prostatektomie. Sie nannten vorrangig zwei Einflussfaktoren auf den Nicht-Erhalt der Potenz nach operativer Entfernung der Prostata: ein höheres Alter sowie den histologischen Status der Karzinomerkrankung, konkret ein kapselinfiltrierendes oder darüber hinauswachsendes Tumorgeschehen (Walsh & Donker, 1982). Lepor et al. (1985) untersuchten darauf aufbauend die exakte Verortung der Nerven und Gefäße, die unter anderem für die Erektion des männlichen Gliedes verantwortlich sind, das sogenannte Neurovaskuläre Bündel. Auch dank ihrer anatomischen Landkarte konnte gezeigt werden, dass ein sehr prominenter Teil

dieser neurovaskulären Versorgung posterolateral der Prostata verläuft (Lepor et al., 1985; Yucel et al., 2005). Abbildung 4 zeigt schematisch den Verlauf des posterolateralen Gefäßnervenbündels (rot).

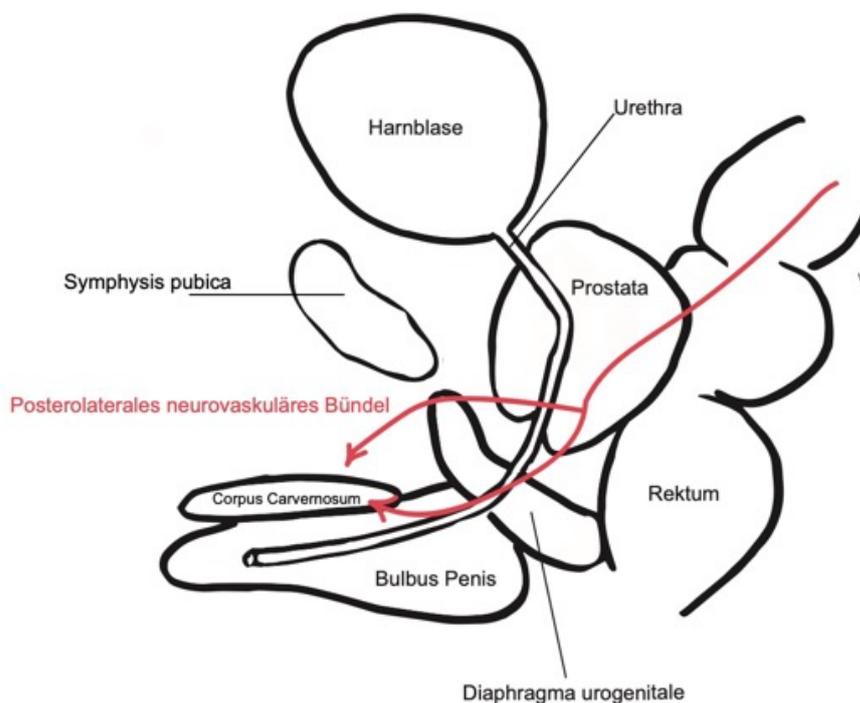


Abbildung 4. Verlauf des posterolateralen Gefäßnervenbündels (rot), abgezeichnet nach McLaughlin et al. (2005)

Die anatomischen Kenntnisse bezüglich der Prostata und ihrer begleitenden Strukturen sind nicht nur von Bedeutung für eine operative Versorgung beispielsweise in Form der radikalen Prostatektomie, sondern auch in deren Vorfeld im Hinblick auf die radiologische Diagnostik, um die Lokalisation suspekter Läsionen in der Prostata genau zu beschreiben oder ggf. mittels gezielter, MRT-gesteuerter Biopsie histologisch zu sichern. Im folgenden Abschnitt soll die Bildgebung der Prostata im klinischen Alltag näher skizziert werden.

1.2 MRT-Bildgebung der Prostata: Methoden- und Indikationsbeschreibung

Im Rahmen der Früherkennung des Prostatakarzinoms werden neben einer ausführlichen Anamnese und körperlichen Untersuchung initial die DRU sowie optional die Bestimmung des Prostata-spezifischen Antigens (PSA) empfohlen, bildgebende Verfahren sollen nicht zur Primärdiagnostik eingesetzt werden (Deutsche Krebsgesellschaft, 2021). Bei einem erstmaligen PSA-Wert von mindestens 4 ng/ml und in Zusammenschau mit anderen Einflussfaktoren, einer ansteigenden PSA-Dynamik (Grenzwert zwischen 0.35 ng/ml bis 0.75 ng/ml pro Jahr) sowie in Folge einer tastuspekten DRU kann nach aktuellen Leitlinien der Deutschen Gesellschaft für Urologie und dem Leitlinienprogramm Onkologie der Deutschen Krebsgesellschaft zum Prostatakarzinom eine Empfehlung zur bioptischen Abklärung erfolgen (Deutsche Krebsgesellschaft, 2021). Diese erfolgt in der Regel systematisch und unter Zuhilfenahme einer TRUS. Um eine möglichst hohe Detektionsrate zu erzielen, sollen bei der systematischen Biopsie mindestens 10 - 12 Stanzzyylinder entnommen werden (Das et al., 2019; Deutsche Krebsgesellschaft, 2021), außerdem sollen in der Bildgebung suspekt erscheinende Areale gezielt biopsiert werden (Deutsche Krebsgesellschaft, 2021).

Der MRT wird ein immer größerer diagnostischer Wert zugeschrieben (Weinreb et al., 2016); 2021 fand sie Eingang in die modifizierte Leitlinie zur Primärdiagnostik des Prostatakarzinoms (Deutsche Krebsgesellschaft, 2021). Die Bildgebung der Prostata mittels MRT soll nach der S3-Leitlinie der Deutschen Krebsgesellschaft zum Prostatakarzinom multiparametrisch erfolgen (Deutsche Krebsgesellschaft, 2021). Sie kann außerdem in Anbetracht flächendeckend verfügbarer 1.5 und 3 Tesla starker MRT-Scanner seit einigen Jahren und spätestens mit Aufnahme in die S3-Leitlinien „Prostatakarzinom“ der Deutschen Krebsgesellschaft 2018 (Deutsche Krebsgesellschaft, 2021) als Goldstandard in den Bereichen Stadieneinteilung (Staging) des Prostatakarzinoms (Caglic et al., 2019), Selektion von Patienten in Hinblick auf eine notwendige Gewebeentnahme (Ahmed et al., 2017), der gezielten Biopsie bei Biopsie-naiven Patienten (Kasisvisvanathan et al., 2018) sowie additiv in der Überwachung

weniger aggressiver Prostatakarzinome („Active Surveillance“) (Chiam et al., 2021; Klotz et al., 2020) angesehen werden (Manfredi et al., 2018).

Die multiparametrische MRT (mpMRT) verknüpft Sequenzen zur anatomischen Beurteilung mit funktionellen Sequenzen und hat dadurch eine große Aussagekraft hinsichtlich der Beurteilung suspekter Läsionen - und das als nicht-invasives diagnostisches Untersuchungsverfahren (Bardis et al., 2020). Zur Vorbereitung und Durchführung der mpMRT gibt es herausgegeben von der Deutschen Röntgengesellschaft (DRG) und dem Berufsverband der Deutschen Radiologen (BDR) eine Veröffentlichung mit klaren Empfehlungen (Franiel et al., 2021). Die zu akquirierenden MRT-Sequenzen mit stichpunktartigen Erläuterungen sind entsprechend dem Untersuchungsprotokoll dieser Empfehlungen in Tabelle 1 aufgeführt.

Tabelle 1. Bildsequenzen der mpMRT nach Empfehlung der DRG und des BDR

Sequenz	Erläuterung
T2w-TSE / FSE	<ul style="list-style-type: none">- 2-dimensional und mindestens biplanar- axiale Ebene obligatorischer Bestandteil- Prostata in allen Ebenen zentriert und vollständig abgebildet- Blasenboden und Urogenitaldiaphragma auf mindestens 2 Ebenen vollständig abgebildet- Samenbläschen und Lymphknoten im Foramen obturatorium vollständig auf mindestens einer Ebene abgebildet- axial, sagittal, koronar: Schichtdicke 3.5 mm (3 mm bevorzugt), Schichtinkrement 0 %, FoV in Phase \leq 200 mm, Phasenkodierrichtung in der Ebene \leq 0.8 mm, Frequenzcodierungsrichtung in der Ebene \leq 0.6 mm
DWI	<ul style="list-style-type: none">- obligatorischer Bestandteil- axiale Schnittbilder

- Schichtdicke 3.5 mm (3 mm bevorzugt), Schichtinkrement 0 %, FoV in Phase ≤ 200 mm, Phasenkodierrichtung in der Ebene ≤ 2.1 mm, Frequenzcodierungsrichtung in der Ebene ≤ 2.1 mm
 - Zur Berechnung der ADC-Karte sollten mindestens 2 verschiedene b-Werte gemessen werden
- DCE
- von zentraler Bedeutung für die MRT-Diagnostik von Prostatakrebs
 - axiale Schnittbilder
 - Schichtdicke 3.5 mm (3 mm bevorzugt), Schichtinkrement 0 %, FoV in Phase ≤ 200 mm, Phasenkodierrichtung in der Ebene ≤ 2.1 mm, Frequenzcodierungsrichtung in der Ebene ≤ 2.1 mm
 - Die Aufnahmedauer sollte 2 Minuten nicht unterschreiten (vorzugsweise 3 Minuten)
- T1w
- Zur Beurteilung des Knochens und der Lymphknoten sowie der Prostata bezüglich Blutungen
 - Abbildung des gesamten Beckens von der Aortenbifurkation bis zum Beckenboden
 - Schichtdicke ≤ 5 mm (2D) / ≤ 2 mm (3D), Schichtinkrement ≤ 10 %, Auflösung in der Ebene $\leq 2.0 \times 2.0$ mm
-

Abbildung 5 zeigt eine mpMRT der Prostata eines Patienten unserer Studie bei 3 Tesla (Siemens MAGNETOM Skyra) unter Verwendung einer Body-Array-Spule. Von links nach rechts sind jeweils in transversaler Schichtebene die T2-gewichtete (T2w), die diffusionsgewichtete (DWI) sowie die ADC-Karte dargestellt. Durch die roten Pfeile markiert ist jeweils eine im Durchmesser etwa 9 x 6 mm große Läsion, die sich in der T2w-Sequenz sowie in der ADC-Karte hypointens und in der DWI-Sequenz hyperintens zeigt. Es handelt sich um eine PI-RADS 4-Läsion in der medialen peripheren Zone rechts ohne organüberschreitendes Wachstum.

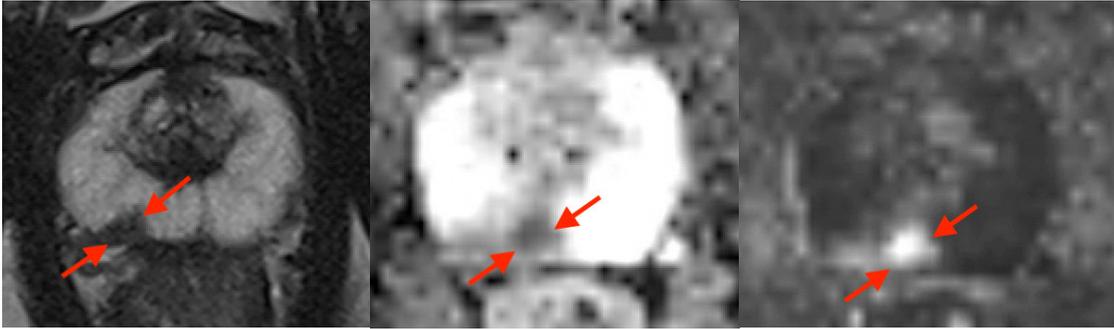


Abbildung 5. MRT-Sequenzen der Prostata, transversal. Von links nach rechts: T2w, ADC-Karte und DWI. Mit roten Pfeilen ist eine ca. 9 x 6 mm messende Läsion in der medialen peripheren Zone rechts markiert (PI-RADS 4).

2012 wurde mit Einführung des sogenannten Prostate Imaging-Reporting and Data System, kurz PI-RADS, durch die Europäische Gesellschaft für Uroradiologie (European Society of Urogenital Radiology, ESUR) (Barentsz et al., 2012) ein standardisiertes Vorgehen zur Erhebung und Auswertung von MRT-Bildern der Prostata beschrieben. 2015 in Form des PI-RADS 2.0 mit der Einführung maßgeblicher Sequenzen bezüglich der Beurteilung suspekter Läsionen in definierten Lokalisationen erstmals aktualisiert (Weinreb et al., 2016), liegt seit Anfang 2019 die Version 2.1 vor (Hotker & Donati, 2021), die Verbesserungen im Hinblick auf Unsicherheiten im Scoring sowie der Interreader-Variabilität anführt (Beyer et al., 2021).

Diesem Scoring-System, das in erster Linie die strukturierte Befundung der Prostata-MRT objektivierbar und damit nachvollziehbar und vergleichbar machen soll, liegt die radiologische Beurteilung dreier Sequenzen der multiparametrischen Magnetresonanztomographie (mpMRT) zugrunde: einer T2-gewichteten Sequenz (T2w) insbesondere zur Beurteilung der Transitionszone in der transversalen und sagittalen Ebene, einer diffusionsgewichteten Sequenz (diffusion weighted imaging, DWI) vorwiegend zur Beurteilung der peripheren Zone und schließlich einer dynamischen kontrastmittelgestützten Sequenz mit ihrem hohen räumlichen Auflösungsvermögen (Beyer et al., 2021; Hotker & Donati, 2021). Tabelle 2 zeigt

die fünf Stufen des PI-RADS-Scores in seiner aktuellen Version und deren Interpretation (Hotker & Donati, 2021).

Tabelle 2. PI-RADS-Score Version 2.1

PI-RADS V2.1	Wahrscheinlichkeit eines klinisch signifikanten Prostatakarzinoms
1	sehr gering
2	gering
3	intermediär
4	hoch
5	sehr hoch

Eine Metaanalyse mit 10 zugrunde liegenden Studien zeigte, dass 7.4 bis 58.5 % der durchgeführten Gewebeprobenentnahmen hätten vermieden werden können, wenn Biopsien lediglich bei mpMRT-auffälligen Patienten durchgeführt worden wären (Becerra et al., 2020).

Im Rahmen der sogenannten PROMIS-Studie beläuft sich der Anteil der Patienten, die durch eine vorgeschaltete mpMRT-Untersuchung auf eine Biopsie verzichten konnten, auf 27 %. Außerdem zeigen Ahmed et al. in ihrer Publikation, dass die mpMRT dazu beitragen kann, eine Überdiagnose klinisch nicht signifikanter Läsionen um 5 % gegenüber der primären Abklärung via TRUS-Biopsie zu verhindern und gleichzeitig klinisch signifikanten Prostatakrebs besser zu erkennen (Sensitivität der mpMRT: 93 %, die der TRUS-Biopsie: 48 %) (Ahmed et al., 2017).

In der multizentrisch angelegten Studie PRECISION belegen Kasivisvanathan et al. (2018) die Überlegenheit der MRT-gestützten Gewebeentnahme gegenüber der TRUS-gesteuerten bei Biopsie-nativen Patienten mit klinischem Anhalt für ein Prostatakarzinom: Klinisch relevante maligne Läsionen der Prostata konnten in 38 % der Fälle, in denen MRT-gesteuert biopsiert wurde, gefunden werden; dieser Anteil lag in der Ultraschall-gesteuerten Kohorte bei 26 %.

Auch Siddiqui et al. (2013) konnten zeigen, dass die MR/TRUS-Fusionsbiopsie der althergebrachten Biopsiemethode mit 12 systematisch entnommenen Stanzen hinsichtlich der finalen histologischen Bewertung eines Karzinoms mittels Gleason-Scores überlegen scheint (Siddiqui et al., 2013). Sowohl niedriggradige (hier sprechen die Daten dafür, dass die alleinige Gewebeentnahme mittels zielgerichteter MRT-gesteuerter Biopsie schlechtere Ergebnisse liefert als die systematische transperineale Biopsie) als auch klinisch signifikante und damit behandlungsbedürftige Prostatakarzinome können in hohem Maße durch eine Kombination aus systematischer transperinealer und zielgerichteter MR/TRUS-Fusionsbiopsie erkannt werden (Radtko et al., 2015).

Der aktuelle Stellenwert der mpMRT im Bereich des Prostatakarzinoms als hoch sensitive und gleichzeitig nicht-invasive Untersuchungsmethode ist unstrittig. Aus Gründen der Zeitersparnis und der Präzision ist die computerunterstützte und automatisierte Analyse und Segmentierung der Prostata in der MRT ein ausgesprochen zukunftsträchtiges Feld (Zhu et al., 2019). Das Schlagwort für die immense Datenmenge, die in der radiologischen Bildgebung tagtäglich gewonnen wird und zum größten Teil kaum detaillierte analytische Beachtung findet, lautet „Radiomics“.

1.3 Künstliche Intelligenz in der radiologischen Bildgebung

In den nachfolgenden Abschnitten und Kapiteln wurde beim Verfassen darauf geachtet, möglichst einheitlich deutsche Begrifflichkeiten zu verwenden und die englischen Namen mit anzugeben, auch wenn in vielen Fällen in der deutschen Literatur der englische Begriff gängig ist.

Unter dem Begriff „Radiomics“ versteht man in Anlehnung an die „Genomik“ oder „Proteomik“ die maschinenbasierte Analyse und Auswertung eines verhältnismäßig großen Datensatzes aus dem Bereich der tomographischen Bildgebung (MRT, Computertomographie (CT) oder Positronenemissionstomographie (PET)) (Cutaia et al., 2021). Die zugrunde

liegenden Algorithmen können auch aufgrund der heutigen technischen Voraussetzungen in Form sehr leistungsfähiger Computer - hier seien vor allem die Grafikprozessoren genannt - und immer leistungstärkerer Scanner auf enorme Bilddatenmengen zurückgreifen und diese höhergradig abstrahiert auf unterschiedlichste (pathologische) Muster untersuchen (Gillies et al., 2016). Mit Abstraktion ist hierbei die Transformation der dem Bild immanenten Informationen wie Farbe, Graustufen oder Kanten gemeint (LeCun et al., 2015).

„Multi-omik“, die Verbindung also verschiedenster Merkmale und deren Analyse und Auswertung, ist speziell im Bereich der onkologischen Medizin der wohl zukunftsreichste Ansatz, wenn es darum geht, mithilfe von KI datenbasiert Entscheidungshilfen zu generieren (Zanfardino et al., 2019).

Anders als bei der computerassistierten Detektion (*computer-aided diagnosis and detection*, CAD), die auf konkrete Fragestellungen Antworten liefern soll, kann die Radiomik durch das sogenannte „*Data-Mining*“ (das statistische Analysieren großer Datenmengen, das Fördern von Daten) selbst Hypothesen aufstellen und prüfen (Gillies et al., 2016). Der maßgebliche Begriff hinter diesem Analysesystem lautet Künstliche Intelligenz. Im Bereich der Bildanalyse liegen ihr grundsätzlich zwei unterschiedliche Architekturen zugrunde: maschinelles und tiefes Lernen (Castiglioni et al., 2021).

Tiefes Lernen (*deep learning*) stellt eine Subdomäne im Bereich des Maschinlernens oder maschinellen Lernens (*machine learning*) dar, beide Verfahren stehen unter dem Überbegriff „Künstliche Intelligenz“ (Bardis et al., 2020). Abbildung 6 soll diese Begrifflichkeiten speziell im Bereich der computerunterstützten Bildanalyse veranschaulichen.

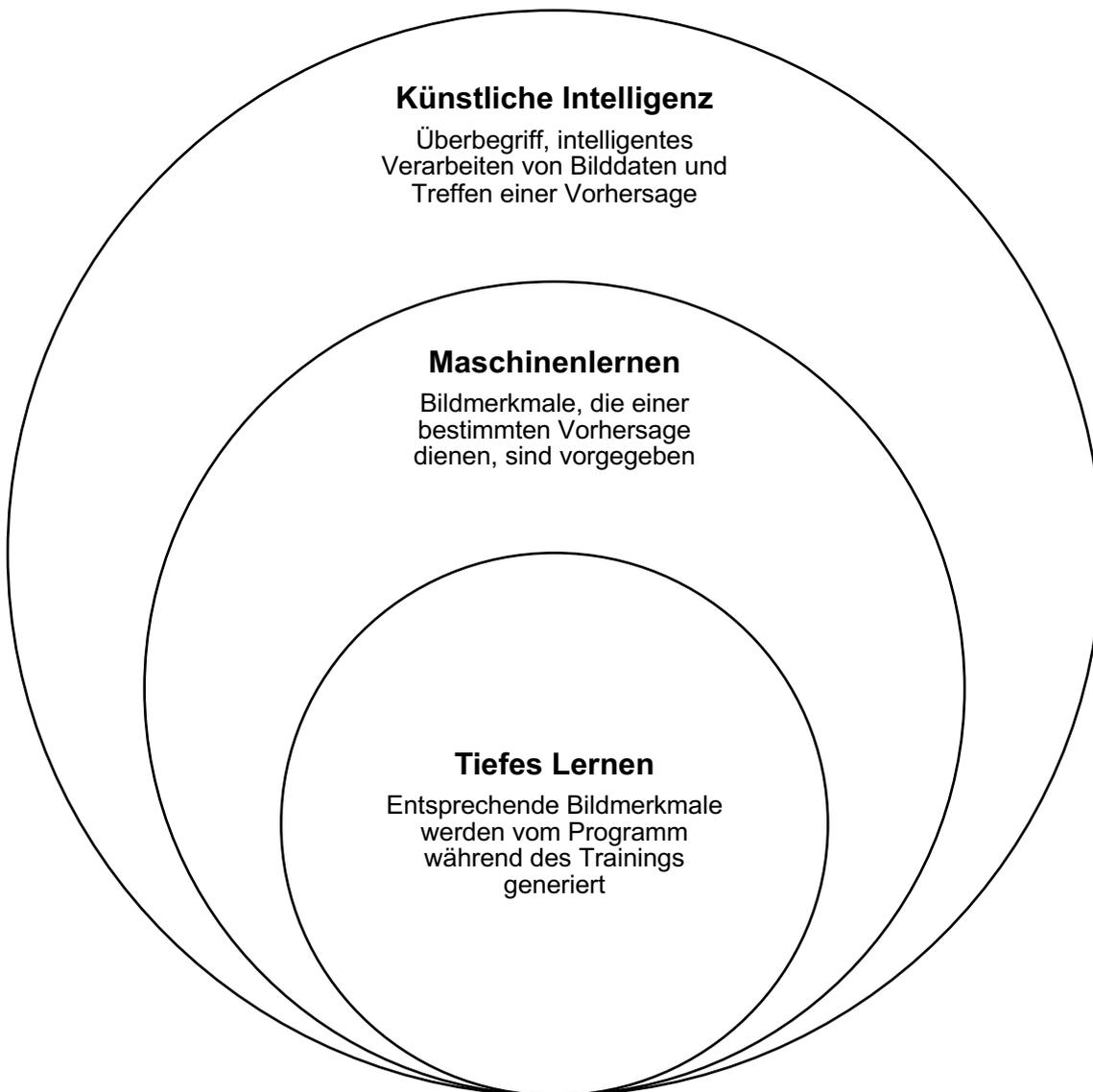


Abbildung 6. Zusammenhang von Künstlicher Intelligenz, Maschinenlernen und Tiefen Lernen, modifiziert nach Bardis et al (2020)

Künstliche Intelligenz soll vereinfacht ausgedrückt nach der Zufuhr (*input*) von Daten auf Grundlage unterschiedlich komplexer Algorithmen eine Aussage oder besser Vorhersage als Ausgabe (*output*) liefern. Mit Vorhersage ist in der Bildanalyse vorwiegend das Erkennen von Kategorien oder Klassen (*classification*) gemeint (Bardis et al., 2020). Die KI soll beispielsweise erkennen und vorhersagen, ob es sich bei umschriebenen Bereichen auf einem MRT-Bild

um Knochen, Gewebe, ein Organ oder Flüssigkeit handelt. Beiden, Maschinenlernen und tiefem Lernen, ist gemein, dass sie einen Datensatz zum Üben, den sogenannten Trainingsdatensatz, häufig einen Datensatz zum Validieren und schließlich einen Datensatz zum Anwenden, den Testdatensatz, benötigen (Castiglioni et al., 2021). Beim Maschinenlernen dienen (Bild-)Daten und eine vorgegebene Kategorisierung, Etikettierung oder Benennung dieser Daten als Zufuhr (Bardis et al., 2020). Die Vorhersage basiert auch auf bestimmten Erkennungsmerkmaltypen, die vor dem Training gewonnen wurden und der KI vorgegeben sind (Castiglioni et al., 2021). Sie validiert und evaluiert sich selbst und lernt durch ihre während des Trainings getroffenen Vorhersagen. *Deep Learning*-Systeme können sich im Gegensatz dazu Bildmerkmale und bestimmte Muster, die sie zur Bildanalyse und zum Treffen bestimmter Vorhersagen benötigen, selbst aus den Bilddaten erarbeiten (Castiglioni et al., 2021). Sie üben anhand des Trainingsdatensatzes und filtern dabei gleichzeitig Bildmerkmale, die für die zu treffende Vorhersage von Bedeutung sind, heraus; beim Maschinenlernen sind diese Bildmerkmale gleichbleibend und basieren auf vor dem Training von Hand festgelegten Mustern (Bardis et al., 2020).

Deep-Learning-Systeme können eigenständig Veränderungen in der Gewichtung ihrer auf unterschiedlichen Ebenen und in mehreren Schichten angeordneten Entscheidungsfunktionen vornehmen, um möglichst genaue Ergebnisse zu liefern. Diese Ebenen und Schichten sind einem biologischen neuronalen Netzwerk nachempfunden. Im Bereich der radiologischen Bilddatenanalyse sind dies wie oft in der Bildverarbeitung vorwiegend sogenannte „Gefaltete Neuronale Netzwerke“ (*Convolutional Neural Networks*, CNNs) (McBee et al., 2018). Ganz allgemein werden die erste und letzte Schicht (*layer*) dieser Netzwerke als Eingabe- (*input layer*) bzw. Ausgabe-Schicht (*output layer*) bezeichnet. Zwischen Ein- und Ausgabe-Schicht liegen sogenannte versteckte oder besser Zwischenschichten (*hidden layers*). Diese Zwischenschichten bilden eine zum Teil sehr große Anzahl von Knotenpunkten oder bildlich gesprochen Neuronen (Castiglioni et al., 2021).

Erste Ansätze zur automatisierten Bilderkennung auf Grundlage von mehrschichtigen neuronalen Netzwerken datieren aus den 1980er Jahren

(Krizhevsky et al., 2012), LeCun et al. konnten hierfür exemplarisch eine gut operierende KI zur Erkennung handschriftlich geschriebener Postleitzahlen kreieren (LeCun et al., 1989).

Mit dem Aufkommen und der Verfügbarkeit immenser Datenmengen sowie bahnbrechenden Fortschritten im Bereich der Computer- und Grafikprozessoren um die Jahrtausendwende (McBee et al., 2018) markierten Krizhevsky et al. in ihrem 2012 veröffentlichten Papier und ihrer im Vergleich zu anderen mit einer deutlich niedrigeren Fehlerrate arbeitenden KI einen Meilenstein im Forschungsfeld der automatisierten Bilderkennung: Krizhevsky et al. gewannen den *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC, umfassender visueller Erkennungs-Wettbewerb) erstmalig auf Grundlage eines CNN mit einer fast 10 % geringeren Fehlerrate. *ImageNet* verfügt über einen Datensatz von über 15 Millionen Bildern, eingeteilt in etwa 22000 Kategorien (Krizhevsky et al., 2012). Der ILSVRC, erstmalig abgehalten im Jahr 2010 und ab 2012 als Nachfolge des seit 2005 etablierten sogenannten *PASCAL VOC*-Wettbewerbes, nutzt für seine Aufgaben einen Teil dieser Bilder (etwa 1.2 Millionen Bilder unterteilt in 1000 Kategorien); sie teilen sich auf in einen Trainings- und einen Testdatensatz (Russakovsky et al., 2015). Die Kategorien, beispielhaft „Vogel“, „Hauskatze“, „Flugzeug“ oder „Monitor“, sind dabei bewusst einfach gehalten (Russakovsky et al., 2015). Eine Aufgabe des Wettbewerbs besteht darin, Objekte zu erkennen und dem präsentierten Bild im Anschluss eine entsprechende Kategorie zuzuordnen; jedes Bild ist dabei mit nur einer Kategorie als sogenannte Grundwahrheit (*ground truth*) versehen. Um dem Problem entgegenzuwirken, dass auf einem Bild mehrere Objekte zu sehen und zu erkennen sein können, ist es dem KI-Algorithmus erlaubt, bis zu 5 Kategorien pro Bild zu nennen (Russakovsky et al., 2015). Die Top-1-Fehlerrate (*top-1-error rate*) gibt an, zu wieviel Prozent der Algorithmus die korrekte Kategorie nicht an erster Stelle genannt hat, die Top-5-Fehlerrate (*top-5-error rate*), zu wieviel Prozent sie nicht unter den bis zu fünf genannten Kategorien zu finden ist. 2012 ist deshalb als Meilenstein im Rahmen des ILSVRC zu nennen, weil mit dem Algorithmus der *SuperVision*-Arbeitsgruppe erstmalig eine KI auf Grundlage eines CNN teilnahm - und den Wettbewerb mit Abstand gewann (Krizhevsky et

al., 2012; Russakovsky et al., 2015). Die der KI präsentierten Bilder wurden hinsichtlich des Erkennens der jeweiligen Bildkategorie zu 37.5 % an erster Stelle als nicht-richtig erkannt (Top-1-Fehlerrate). Zu 17 % wurde das präsentierte Bild in den von ihr genannten obersten fünf Möglichkeiten, gelistet nach absteigender Wahrscheinlichkeit, als nicht-richtig erkannt (Top-5-Fehlerrate) (Krizhevsky et al., 2012). Diese Zahlen mögen sehr hoch erscheinen, doch gilt es, sie in Relation zu den Algorithmen der anderen Teilnehmer zu setzen: Die Leistung der zweitplatzierten KI auf Basis der Bildanalyse unter anderem mittels des sogenannten Fisher-Vektors lag bei einer Top-5-Fehlerrate von 26.2 %, die der besten KI im Jahr 2011 bei 25.8 %, im Jahr 2010 bei 28.2 % (Russakovsky et al., 2015). Das CNN von *SuperVision* war bereits 2012 um mindestens 10 Prozentpunkte besser als jeder andere Algorithmus. 2015 überragte schließlich die Leistung einer KI auf Grundlage des *Deep-Learning*-Modells die eines Menschen (McBee et al., 2018).

Der Gedankensprung vom Erkennen von Containerschiffen, Kirschen, Leoparden und Pilzen (Krizhevsky et al., 2012) auf Fotobildern zur Analyse klinischer Bilder ist nicht weit. Ronneberger et al. haben 2015 mit ihrem sogenannten „fully convolutional network“ U-Net gezeigt, wie leistungsstark die automatisierte Segmentierung auf biomedizinischen Bildern wie etwa einer elektronenmikroskopischen Aufnahme von Nervengewebe mit neuronalen Strukturen sein kann. „U-Net“ beschreibt dabei die symmetrische Architektur des hier zugrundeliegenden CNN, dessen Schichtanordnung vergleichbar dem Buchstaben „U“ skizziert werden kann. Die Besonderheit dieser KI besteht unter anderem darin, dass nicht abertausende annotierter Trainingsdaten vonnöten sind. Durch das Augmentieren der Daten reichen nur wenige Beispielmuster als Trainingsset. Mit Augmentation, wörtlich dem „Unterstützen“ der wenigen Trainingsdaten, ist hier das wiederholte Präsentieren dieser Trainingsbilder, abgeändert in unterschiedlichen Variationen, gemeint: rotiert, horizontal gekippt, vergrößert, verzerrt und minimal in Höhe und Breite verschoben (Jensen et al., 2019; Sanford et al., 2020). Für die *Cell Tracking Challenge* (Zell-Detektion Wettbewerb) des *International Symposium on Biomedical Imaging (ISBI)* 2015, bei der eine Teilaufgabe das Segmentieren von Astrozyten eines Glioblastoms

darstellte, bestand das Trainingsdatenset aus nur 35 teilweise annotierten Mikroskopiebildern. Hier erreichte das U-Net einen - im weiteren Verlauf dieses Abschnitts näher erläuterten - Jaccard Index von 92 %; die zweitplatzierte KI erreichte lediglich 83 %. Bei einer weiteren Aufgabe innerhalb des Wettbewerbs reichten dem U-Net 20 Trainingsbilder, um einen Jaccard Index von 77.5 % und damit den ersten Platz zu erzielen. Hier lag der zweitbeste Algorithmus bei 46 % (Ronneberger et al., 2015). Der Jaccard Index, auch „*Intersection over union*“ (IoU) genannt, wird gebildet aus dem folgenden Quotienten: Der Flächeninhalt der Schnittfläche von Grundwahrheit und automatisiert vorgenommener Segmentierung geteilt durch den gesamten annotierten Flächeninhalt beider Segmentierungen. Er ist ein statistisches Mittel zum Vergleich zweier Bilder und stellt eine mathematisch sehr ähnliche Alternative zum Sørensen-Dice-Koeffizienten dar, auf den im Kapitel 2.3.1 näher eingegangen wird.

Die Anforderungen an eine KI im medizinischen Kontext müssen sehr hoch sein, da viele Entscheidungen, die auf deren Vorhersage beruhen, weitreichende Konsequenzen nach sich ziehen. Nimmt man zum Beispiel das Erkennen und Segmentieren von Organstrukturen auf MRT-Bildern, so können daraus Therapieentscheidungen, die Bestimmung der Lokalisation gezielter Biopsieentnahmen und Prognosen resultieren.

Die Segmentierung der Prostata ist nicht nur im onkologischen Bereich häufig der erste diagnostische Schritt in der Beurteilung eines MRT-Bildes. Geschulte Radiologinnen und Radiologen umfahren dabei das Organ als Ganzes, einzelne Zonen oder den Ausschnitt, der von Interesse ist, manuell auf jedem einzelnen Schnittbild der mpMRT-Bildserien (Clark et al., 2017). Eine KI an dieser Stelle des klinischen Alltages zu implementieren wäre allein schon aufgrund der Zeitersparnis ein großer Fortschritt (Zhu et al., 2019). Das Umschreiben der Prostata in der mpMRT in ihrer Gesamtheit ist notwendig, um in einem nächsten Schritt die periphere von der Transitionszone zu unterscheiden, was wiederum zur Klassifikation nach PI-RADS unerlässlich und schließlich von therapeutischer und prognostischer Konsequenz ist. Eine präzise Segmentation ist außerdem

Grundlage für ein möglichst genau berechnetes Prostatavolumen. Dieses wiederum ist von großer Bedeutung zur Abschätzung der Diagnosewahrscheinlichkeit etwa einer benignen Prostatahyperplasie oder eines Prostatakarzinoms, dessen Verlauf und Therapieansprechen sowie zur Einordnung des PSA-Wertes respektive der Berechnung der PSA-Dichte (Bardis et al., 2020).

In einer systematischen Übersichtsarbeit von Bardis et al. (2020) wurden Studien auf Grundlage Maschinenlernen-basierter Anwendungen zur Beurteilung einer mpMRT zusammengefasst. Das Hauptaugenmerk lag dabei auf der Segmentierung der Prostata und der Detektion suspekter Läsionen. Tabelle 3 gibt einen Überblick über diejenigen Studien, die auf Grundlage eines CNN automatisiert die Prostata segmentiert haben.

Tabelle 3. Studien zur CNN-basierten Prostatasegmentation nach Bardis et al. (2020)

Arbeitsgruppe	Jahr	CNN Architektur	Patienten (N)	mpMRT- Sequenz	Dice
Milletari et al.	2016	V-Net	80	T2w	0.87
Zhu, Q. et al.	2017	U-Net	81	T2w	0.89
Clark et al.	2017	U-Net	134	DWI	0.89
Zhu, Y. et al.	2018	U-Net	163	T2w, DWI	0.93
Karimi et al.	2018	3 Schichten	49	T2w	0.88
Tian et al.	2018	7 Schichten	140	T2w	0.85

Die in Tabelle 3 angeführten Dice-Koeffizienten (siehe Kapitel 2.3.1) von 0.85 bis 0.93 zeigen bereits akzeptabel bis gut funktionierende automatisierte Segmentierungen. Turkbey und Haider (2021) fassen erst kürzlich veröffentlichte Studien zur automatisierten Abgrenzung der Prostata, ihrer Zonen und möglicher Läsionen zusammen. Sie geben außerdem einen Überblick über bereits verfügbare klinische Anwendungen auf diesem Gebiet, von denen zwei auch in Europa zugelassen sind (Turkbey & Haider, 2022). Wang et al. erreichten mit ihrem CNN einen mittleren Dice-Koeffizienten von 0.86 (Wang et al., 2019),

Ushinsky et al. 0.898 (Ushinsky et al., 2021). Die bis dato höchsten publizierten Ergebnisse liefern Bardis et al. mit 0.940 (Bardis et al., 2021) und Sanford et al. mit einem Dice-Koeffizienten von 0.931 (Sanford et al., 2020).

Die in den letzten Jahren erreichten Überschneidungen, ausgedrückt in den Dice-Koeffizienten manueller mit automatisierter planmetrischer Abgrenzung der Prostata auf MRT-Bildern, lassen auf bereits gut bis sehr gut funktionierende KI-Algorithmen schließen. Die vorliegende Arbeit untersucht nun klinische Parameter als mögliche Einflussfaktoren auf die Güte der Organsegmentierung einer KI. Neben der Qualität und Anzahl der Trainingsdaten sowie der Präzision der manuell vorgenommenen Segmentierung als systemimmanenter Komponente besteht die Annahme, dass es externe Faktoren gibt, die sich auf die Performanz der KI auswirken. Das Wissen um solche Einflussfaktoren könnte die automatisierte Segmentierung noch weiter verbessern und im Einzelfall einen Hinweis für ein genaueres Betrachten der Sequenzen geben, ehe Entscheidungen auf Grundlage der Segmentierung abgeleitet werden.

2 Material und Methoden

2.1 Studiendesign

Der vorliegenden Arbeit liegt eine KI-basierte Bildauswertung sowie eine retrospektive Kohortenstudie der Abteilung für Diagnostische und Interventionelle Radiologie des Universitätsklinikums Tübingen (UKT) im Rahmen des Projektes Künstliche Intelligenz zur Detektion der Resektabilität eines Prostatakarzinoms (RESECT) zugrunde.

Zu den beteiligten Partnern des RESECT-Projektes gehören neben dem UKT die Klinik für Radiologie und Nuklearmedizin der Universitätsmedizin Mannheim (UMM) der medizinischen Fakultät der Universität Heidelberg sowie die Klinik für Urologie und Urochirurgie der UMM der medizinischen Fakultät der Universität Heidelberg, das Heinrich-Lanz Zentrum für Digitale Gesundheit, die Koordinierungsstelle Telemedizin Baden-Württemberg und das Fraunhofer-Institut für Digitale Medizin MEVIS, des Weiteren die Smart Reporting GmbH sowie die Deutsche Röntgen-Gesellschaft e.V. Neben der Ethik-Kommission der Universität Tübingen („Künstliche Intelligenz zur Detektion der Resektabilität eines Prostatakarzinoms“, Nummer des Ethikantrages: 754/2020BO2) wurde die Studie von der Ethik-Kommission II der Universität Heidelberg (Medizinische Fakultät Mannheim) genehmigt.

Übergeordnetes Ziel des RESECT-Projektes ist es, das behandelnde Kollektiv bei gesichert diagnostiziertem Prostatakarzinom bezüglich der optimalen Therapieentscheidung zu unterstützen. Es soll KI-basiert eine prätherapeutische Risikoabschätzung hinsichtlich des Verbleibs eines Resttumors, konkret also die Entscheidung bezüglich einer R0 (tumorfrei)- oder R1 (Schnittrand mit Tumorrest)-Situation nach einer operativen Intervention, getroffen werden. Dazu wurden multizentrisch an den beiden Standorten Tübingen und Mannheim klinische Daten und Bilddaten erhoben, aufgearbeitet und analysiert.

Ziel der vorliegenden Dissertationsschrift ist es, mögliche Einflussfaktoren auf die Performanz der KI hinsichtlich des Erkennens und der Segmentierung der Prostata zu untersuchen. Konkret sollen folgenden drei Fragestellungen erörtert werden:

- 1) Gibt es einen Unterschied hinsichtlich der Güte oder Performanz der Organsegmentierung eines KI-Algorithmus bei drei unterschiedlichen Trainingsdatensätzen?
- 2) Haben der Body-Mass-Index, der PI-RADS-Score, die Anzahl suspekter Läsionen im MRT, die bildmorphologisch beschriebene extraprostatatische Ausbreitung eines Tumorgeschehens, das MRT-diagnostisch bestimmte Prostatavolumen und die PSA-Dichte Einfluss auf die Performanz der KI?
- 3) Welche qualitativen Unterschiede gibt es zwischen der manuellen und der automatischen Segmentierung der Prostata?

2.1.1 Einschlusskriterien und Zeitraum der Fallsammlung

Das Projekt „RESECT“ ist multizentrisch angelegt. Der Standort Tübingen hat aus einer von der Universitätsklinik für Urologie zur Verfügung gestellten Liste mit Patienten, die konsekutiv im Zeitraum 02.01.2017 bis 30.03.2021 eine radikale Prostatektomie erhalten haben, 185 Patienten ausgewählt, bei denen im Vorfeld der Operation eine multiparametrische Magnetresonanztomographie (mpMRT) in domo durchgeführt wurde.

Einschlusskriterien waren demnach

- stattgehabte Radikale Prostatektomie und
- vorhandene und präoperativ durchgeführte mpMRT in domo.

Die Segmentierung der Prostatae sowie das Annotieren der neurovaskulären Bündel und der Sphinkter der Prostatae wurden nach Einweisung in die Software und einer Schulung hinsichtlich des Erkennens und Annotierens anatomischer Strukturen auf Prostata-MRT-Datensätzen vom Doktoranden im Zeitraum 17.11.2021 bis 27.01.2022 vorgenommen. Der Doktorand hatte zu diesem Zeitpunkt sein Medizinstudium erfolgreich abgeschlossen. Die fertigen Segmentierungen und Annotationen wurden anschließend von einer Assistenzärztin für Radiologie im 3. Weiterbildungsjahr kontrolliert und gegebenenfalls korrigiert.

2.1.2 Multiparametrische Magnetresonanztomographie

Alle 185 MRT-Untersuchungen wurden mit einer Body-Array-Spule durchgeführt. Allen Patienten wurde das Kontrastmittel Gadubotrol (Gadovist®, Bayer Healthcare) intravenös verabreicht.

Die akquirierten Prostata-MRT-Bildsequenzen entsprachen ausnahmslos den leitliniengerechten Vorgaben und Empfehlungen der DGR und des BDR zur Durchführung einer mpMRT-Untersuchung (Deutsche Krebsgesellschaft, 2021; Franiel et al., 2021):

- T2-gewichtete MR-Sequenzen (T2w): Koronare, sagittale und axiale T2 - TSE, axiale T1-TSE
- Diffusionsgewichtete MR-Sequenzen („Diffusion-weighted imaging“, DWI)
- Dynamische kontrastmittelangehobene MR-Sequenzen („dynamic contrast enhanced“-MRT, DCE) nach Kontrastmittelgabe

2.1.3 Verwendete Software

Tabelle 4 zeigt die in dieser Arbeit verwendete Software und deren Hersteller.

Tabelle 4. Verwendete Software

Produkt	Hersteller
SPSS Statistics	IBM
Satori RESECT AI Development Platform	Fraunhofer MEVIS
SAP IS-H	Affinis AG
Medavis RIS	Medavis GmbH
Smart Radiology	Smart Reporting GmbH

SPSS diente der Auswertung und Analyse der Daten, die retrospektiv mithilfe einer Excel Tabelle aufgearbeitet wurden. Grundlage der Erhebung waren Befunde und Berichte in der digitalen Patientenakte (SAP IS-H (Industrial Solution - Healthcare)) und dem Radiologieinformationssystem RIS (Medavis). Die SATORI RESECT AI Plattform von Fraunhofer MEVIS diente browserbasiert

der Segmentierung der Prostaten und suspekter Läsionen sowie der Annotation weiterer anatomischer Strukturen. Auf Smart Radiology wurden ebenfalls browserbasiert klinisch relevante Daten zusammengeführt und etwa das Prostatavolumen und die PSA-Dichte automatisiert errechnet.

2.2 Studienablauf

2.2.1 mpMRT und Prostatektomien

Im Rahmen ihrer Prostatakarzinomerkrankung wurde bei den 185 Probanden, die für die RESECT-Studie eingeschlossen werden konnten, im Zeitraum 16.09.2016 bis 30.03.2021 sowohl eine mpMRT als auch eine radikale Prostatektomie vorgenommen.

2.2.2 Retrospektive Erhebung klinischer Daten

Die eingeschlossenen 185 Patienten wurden anonymisiert (UKT_1 bis UKT_185). Es wurden retrospektiv anhand der Patientenakte (Arztbriefe, Pathologiebefunde, Laborbefunde, Radiologiebefunde, Aufklärungsbögen Anästhesie, Operationsberichte) folgende klinischen Parameter erhoben und in einer Excel-Tabelle notiert:

- Alter zum Zeitpunkt der Prostatektomie
- Körpergröße (laut Narkoseaufklärungsbogen zur Prostatektomie)
- Körpergewicht (laut Narkoseaufklärungsbogen zur Prostatektomie)
- Zuletzt vor Prostatektomie gemessener Prostata-spezifisches Antigen (PSA) - Wert
- Gleason Score der Stanzbiopsie (wenn vorhanden laut Arztbrief bei externer Biopsie bzw. laut Pathologiebefund bei in domo durchgeführter Probenentnahme)
- Gleason Score des Prostatektomiepräparates (Pathologiebefund)
- Größe und Gewicht des Prostatektomiepräparates (Pathologiebefund)
- TNM-Stadium (laut Pathologiebefund des Prostatektomiepräparates)
- Resektionsstatus (laut Pathologiebefund des Prostatektomiepräparates)

- PI-RADS-Score (laut mpMRT-Befund)
- Volumen der Prostata gemessen im MRT (bei Angaben aller Längen im Raum (apikal-basal, sinister-dexter, ventral-dorsal) Ausrechnung mithilfe der SMART-Radiology-tools; ansonsten geschätztes Volumen laut MRT-Befund)
- Anzahl der Prostataläsionen im MRT (laut MRT-Befund)
- Extraprostatische bzw. organüberschreitende Ausbreitung einer suspekten Läsion im MRT (kapselüberschreitendes bzw. -infiltrierendes Wachstum eines Tumors; laut MRT-Befund)
- Durchmesser der Indexläsion (laut MRT-Befund)
- Feldstärke des Scanners (laut MRT-Protokoll)

Anhand der oben genannten Daten wurden folgende Parameter berechnet:

- PSA-Dichte ($\frac{\text{erhobener PSA-Wert}}{\text{Prostatavolumen gemessen im MRT}}$)
- Body-Mass-Index (BMI; $\frac{\text{Körpergewicht}}{\text{Körpergröße}^2}$)

2.2.3 Segmentierung der Prostata

Die Bildserien der mpMRT der Prostata wurden anschließend anonymisiert auf die SATORI Plattform (Fraunhofer MEVIS) geladen. Die Prostaten wurden vom Doktoranden segmentiert. Das neurovaskuläre Bündel sowohl links als auch rechts wurde annotiert, ebenso der prostatiche Sphinkter. Die Segmentierung und die Annotationen dienen als Trainingsdaten für die Künstliche Intelligenz (KI). Abbildungen 6 und 7 zeigen die Anwendermaske der browserbasierten SATORI Plattform beim Einzeichnen der Prostatabegrenzung für zwei unterschiedliche Patienten. Darauf sind im linken, großen Arbeitsfenster (T2-gewichteter axialer Schnitt) die bereits vollständig umfahrene Prostata sowie die Annotation des neurovaskulären Bündels links und rechts (NVB right und NVB left) zu erkennen. Im rechten oberen Fenster ist der koronare Schnitt mit in Abbildung 6 zu sehender gelber Linie auf der dem Arbeitsfenster entsprechenden Höhe sichtbar. Das rechte untere Fenster stellt die ADC-Karte dar. Die rote Umrandung

entspricht der manuellen Segmentierung der Prostata. Der in Abbildung 7 und 8 in unterschiedlicher Größe zu sehende orangefarbene Punkt ist das Werkzeug, mit dem die Einzeichnung der Prostatagrenzen erfolgte.

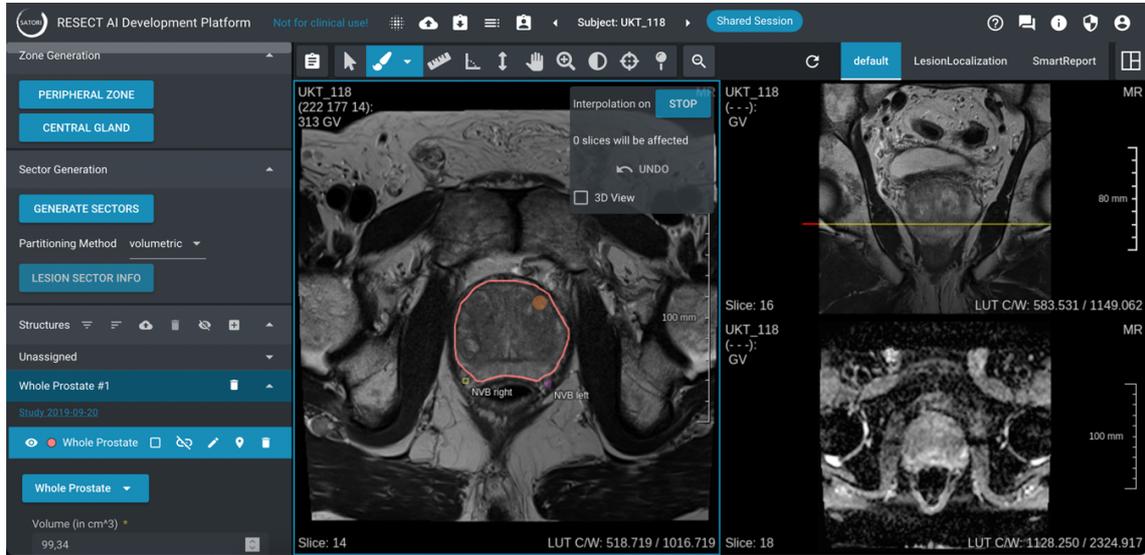


Abbildung 7. Segmentierung der Prostata auf der SATORI-Plattform, Beispiel 1

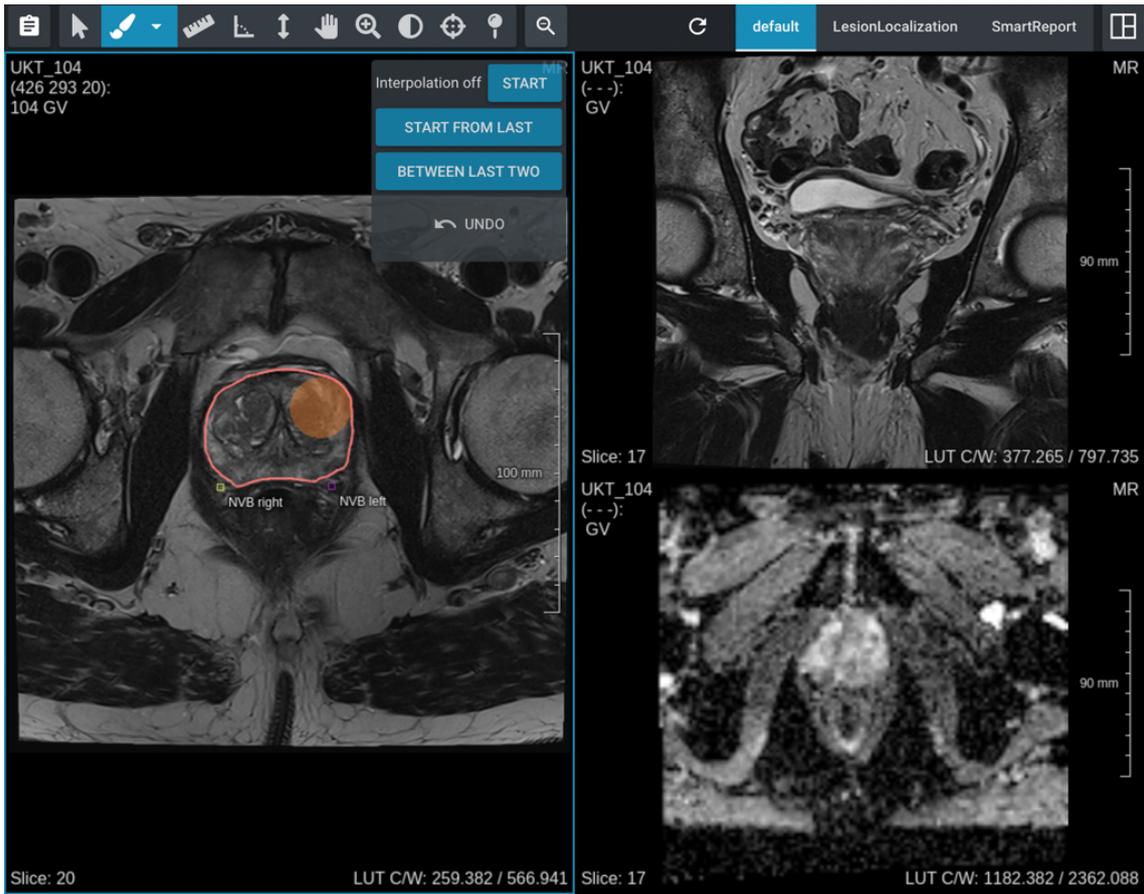


Abbildung 8. Segmentierung der Prostata auf der SATORI-Plattform, Beispiel 2

Abbildung 9 skizziert den Arbeitsfluss von der Auswahl der Patienten bis zum Segmentieren der Prostata.

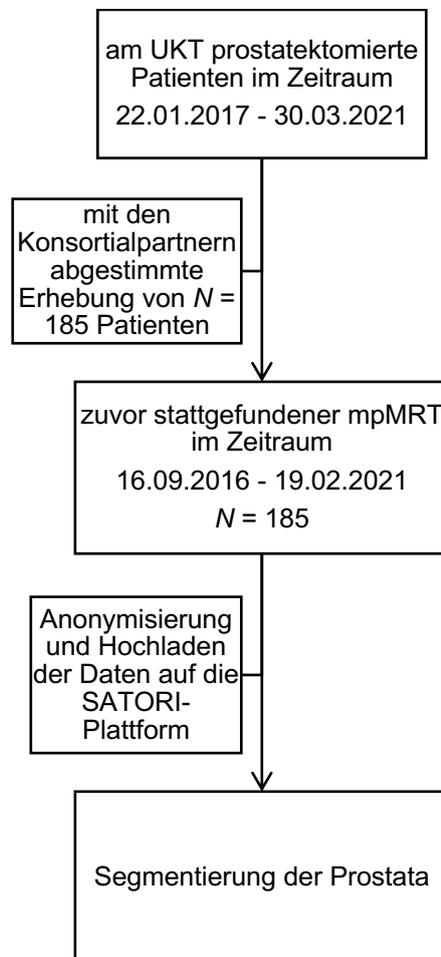


Abbildung 9. Arbeitsfluss bis zur Segmentierung der Prostata

2.2.4 Trainieren und Anwenden der KI

Anschließend wurden basierend auf folgenden Trainingsdatensätzen drei unterschiedliche KIs beübt:

- 1) 100 Probanden der Stichprobe Tübingen (UKT_1 bis UKT_100)
- 2) 100 Probanden der Stichprobe Mannheim (MA_retro_pos1 bis MA_retro_pos100)

- 3) 100 Probanden der Stichproben Tübingen und Mannheim (UKT_1 bis UKT_50 und MA_retro_pos1 bis MA_retro_pos50)

Die auf Basis dieser unterschiedlichen Datensätze trainierten KIs segmentierten anschließend jeweils 52 ihnen unbekannte mpMRT-Bildserien (UKT_134 bis UKT_185). Die automatisierte Organsegmentation wurde anhand des Sørensen-Dice-Koeffizienten mit der manuellen Segmentation verglichen.

Abbildung 10 zeigt das Trainieren und Anwenden der unterschiedlichen KI-Algorithmen.

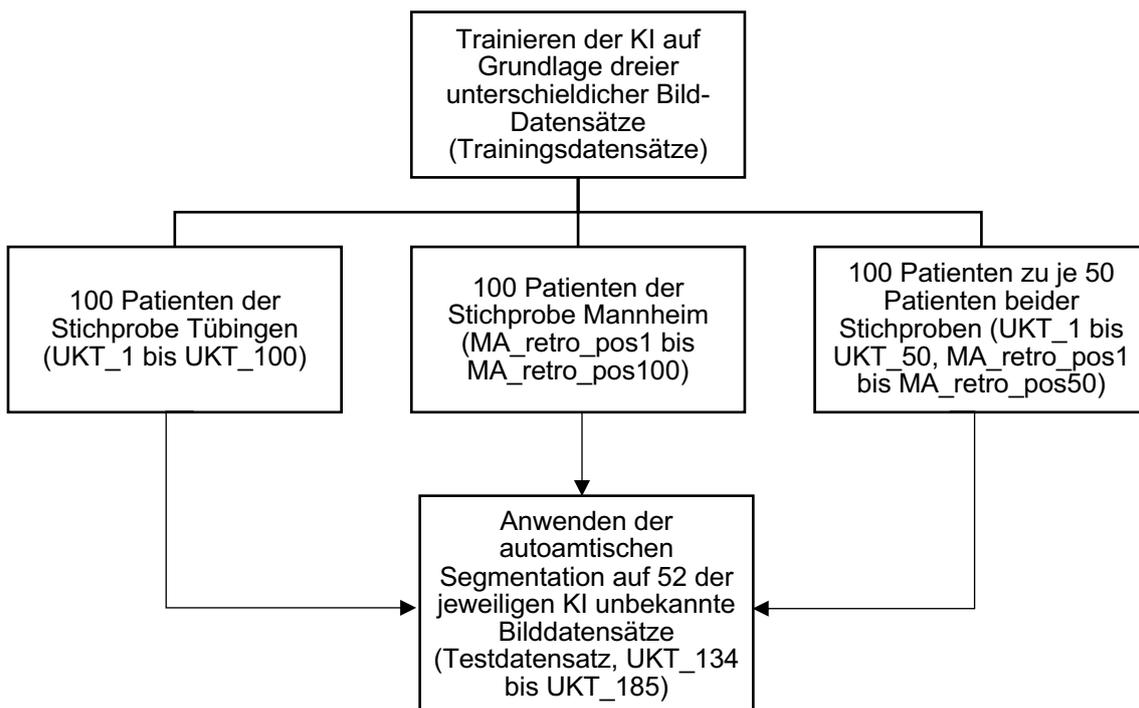


Abbildung 10. Zusammensetzung der Trainingsdatensätze

2.3 Statistische Auswertung

Die erhobenen Daten wurden mit Hilfe des Programms SPSS 22.0 (Statistical Package for Social Sciences) deskriptiv analysiert und nach Bravais-Pearson korreliert. Die drei KIs, die auf Grundlage unterschiedlicher Trainingsdatensätze entstanden sind, wurden auf Äquivalenz getestet und mögliche Einflussfaktoren auf die Performanz der KIs auf Grundlage einer verallgemeinernden Schätzgleichung ausgewertet. Zur Untersuchung der Einflussfaktoren wurden Sørensen-Dice-Koeffizienten berechnet.

2.3.1 Sørensen-Dice-Koeffizient

Der Sørensen-Dice-Koeffizienten oder kurz Dice-Koeffizient wird verwendet, um eine automatische Bildsegmentation mit der manuell vorgenommenen zu vergleichen. Er findet sehr häufig Anwendung auf dem Gebiet der medizinischen Bilddatenauswertung (Carass et al., 2020). Als Maß der 2-dimensionalen Überschneidung zweier Areale kann er Werte zwischen 0 (keine Überschneidung) und 1 (vollkommene Überschneidung) annehmen (Bardis et al., 2020).

Der Sørensen-Dice-Koeffizienten (Dice) zweier Bilddatensätze A und B kann wie folgt formuliert werden:

$$\text{Dice (A, B)} = \frac{2 \times |\text{Überschneidungen (A, B)}|}{(|A| + |B|)}$$

wobei $|A|$ die Kardinalität (Anzahl der Elemente dieser Menge, konkret: Anzahl der Pixel) des Bilddatensatzes A beziffert, $|B|$ die Kardinalität des Bilddatensatzes B.

Der Dice-Koeffizient kann außerdem mit den Begrifflichkeiten „Richtig-positiv segmentierte Pixel“ (*true positives*, (TP)), „Falsch-Positive“ (*false positives* (FP)) und „Falsch-negativ segmentierte Pixel“ (*false negatives* (FN)) wie folgt beschrieben werden:

$$\text{Dice (A, B)} = \frac{2 \times |TP|}{(2 \times TP + FP + FN)}$$

2.3.2 Äquivalenztest

Die Bestimmung der Dice-Koeffizienten dient einer quantitativen Beurteilung der Performanz der jeweiligen KI. Zum Testen auf Unterschiede zwischen den Dice-Koeffizienten der drei KIs wurde ein Äquivalenztest durchgeführt. Dabei wird untersucht, ob der Unterschied zwischen zwei Gruppen kleiner als die Differenz zuvor bestimmter Grenzen ist und die einzelnen Gruppen folglich als äquivalent zueinander bezeichnet werden können. Es soll damit also untersucht werden, ob der Unterschied so hinreichend klein und nahe bei Null ist, dass er keinen relevanten oder signifikanten Effekt hat. Der Äquivalenztest kann folglich als Variation des Signifikanztests angesehen werden (Lakens et al., 2018). Dieser ist zur Untersuchung der Frage nach Unterschiedlichkeit zwischen zwei Gruppen ungenügend, da man bei nicht-signifikantem Ergebnis zwar die Alternativhypothese („Es gibt einen Unterschied zwischen den beiden Gruppen.“) ablehnen muss respektive nicht annehmen kann, gleichzeitig aber nicht auf einen Beweis für das Annehmen der Nullhypothese („Es gibt keinen Unterschied zwischen den beiden Gruppen.“) schließen kann (Abwesenheitsnachweis, *proof of absence*, bei Signifikanz nicht gleichbedeutend mit Evidenznachweis, *proof of evidence*).

Die kleinste Effektgröße, die beim Testen auf Äquivalenz noch von Interesse ist (*smallest effect size of interest* (SESOI)), gibt die Grenzen an, in denen sich die Differenz zweier Gruppen bewegen soll, damit diese beiden noch als äquivalent bezeichnet werden können.

Es gibt zwei Möglichkeiten, statistische Äquivalenz zu testen. Zum einen kann man zwei einseitige t-Tests (*two one-sided tests*, TOST) gegen die zuvor bestimmte obere und untere Grenze als Nullhypothesen durchführen. Werden beide Tests statistisch verworfen, so liegt Äquivalenz vor. Legt man bei diesen beiden Signifikanztests ein Signifikanzniveau von 5 % (0.05) zu Grunde, so kann man zum anderen auch ein Prognoseintervall mit Sicherheitsniveau von 90 % für den Wert der Differenz bestimmen. Liegt dieses 90%-Konfidenzintervall innerhalb des Intervalls zwischen den zuvor festgelegten Grenzen, so kann man die Äquivalenz annehmen.

Von entscheidender Bedeutung für die Aussagekraft der Äquivalenz ist das gut begründete Setzen der oberen und unteren Grenze. Dazu geben Lakens et al. (2018) grundsätzlich und ganz allgemein gesprochen zwei Ansätze: eine objektive Bestimmung und eine subjektive. Die objektive beruht auf messbaren oder sichtbaren Effekten wie beispielsweise einem veränderten Blutspiegel oder einer Erweiterung der Pupillen nach Gabe einer bestimmten Substanzdosis. Subjektiv kann man sich den Grenzen zur Äquivalenztestung nähern, indem man auf ein Maß zur Bestimmung der Effektstärke zurückgreift, zum Beispiel Cohens d (definiert als Differenz der Stichprobenmittelwerte geteilt durch die gemeinsame Streuung bei gleichen Stichprobengrößen). Weiter kann mit Grenzwerten, die bereits in anderen Studien und Arbeiten etabliert sind, gerechnet werden. Außerdem ist ein auf eigener Erfahrung beruhender und aus den Daten ableitbarer Ansatz denkbar (Lakens et al., 2018).

Die Grenzen für den Nachweis von Äquivalenz wurde auf ± 0.01 festgelegt, das heißt die obere Grenze beträgt $+0.01$, die untere -0.01 . Das entspricht gerundet auf zwei Dezimalen 1 % der Werte zwischen 0 und 1, die der Dice-Koeffizient annehmen kann. Liegt das 90%-Konfidenzintervall (entspricht zwei einseitigen Signifikanztests mit $\alpha = 0.05$) der Differenzen der Dice-Koeffizienten innerhalb dieser unteren und oberen Grenze, so lässt sich daraus schließen, dass es keinen signifikanten Unterschied zwischen den jeweiligen KIs gibt oder anders ausgedrückt, dass es einen Effekt gibt, der äquivalent zur Null ist.

2.3.3 Verallgemeinernde Schätzgleichung

Zur Untersuchung des Einflusses der verschiedenen Prädiktorvariablen auf die Güte oder Performanz der drei KIs in Form der jeweiligen Dice-Koeffizienten wurde eine verallgemeinernde Schätzgleichung (*Generalized Estimating Equations*, GEE) durchgeführt (Liang & Zeger, 1986). Von Liang und Zeger (1986) Mitte der 80er Jahre eingeführt sind GEEs eine sehr beliebte Methode unter anderem im Bereich der Epidemiologie und anderen naturwissenschaftlichen Disziplinen (Ghisletta & Spini, 2004). Dabei werden verallgemeinerte lineare Modelle (*generalized linear model*, GLM) um

Messwiederholung bei longitudinalen und hierarchisch aufgebauten Studiendesigns oder aber in Clustern zusammengefasste Daten (Ghisletta & Spini, 2004). Ghisletta und Spini (2004) zeigen in ihrer Veröffentlichung die Vorteile der GEE gegenüber anderen Regressionsmodellen. So muss eine mögliche Korrelation der Variablen innerhalb der Cluster berücksichtigt werden, der geschätzte Standardfehler könnte sonst verzerrt sein. Dem wird die Analyse mittels GEE gerecht. Außerdem gehen traditionelle Regressionsmodelle von der Annahme aus, dass die Fehlerterme normalverteilt und die abhängigen Variablen stetig sind. GEE sind für die Analyse nicht-normalverteilter korrelierter Daten geeignet und liefern die Möglichkeit, sowohl stetige, ordinal verteilte, dichotom- oder polytome Daten simultan zu analysieren; sie lockern generell gesprochen einige Annahmen, die bei althergebrachten Regressionsmodellen getroffen werden müssen (Ghisletta & Spini, 2004).

In der vorliegenden Arbeit entsprechen die drei auf unterschiedlichen Trainingsdatensätzen geübten KIs (1 bis 3) den Clustern oder Subjektvariablen. Diese wurden weiter in Innersubjektvariablen (Patienten-ID 1 bis 52) unterteilt. Die Prädiktorvariable „extraprostatistische Ausbreitung“ ist dichotom bzw. binär konzipiert, der PI-RADS-Score sowie die Anzahl suspekter Läsionen ordinal und der BMI, das Prostatavolumen sowie der Durchmesser der Indexläsion metrisch. Die Dice-Koeffizienten der 52 Patienten des Testdatensatzes dienen als abhängige respektive Antwortvariablen. Bei der Auswahl der Prädiktorvariablen wurde darauf geachtet, dass sie sich inhaltlich nicht zu sehr ähneln oder überschneiden.

Als Interaktionsterm wurde das Produkt der jeweiligen Gruppe mit den einzelnen Prädiktorvariablen definiert, um entsprechende Interaktionseffekte hinsichtlich der Güte innerhalb der unterschiedlich trainierten KIs zu untersuchen. Anhand der Signifikanz kann so darüber geurteilt werden, ob eine der Prädiktorvariablen durch die Subjektvariable „Gruppe“ (1-3 entsprechend der drei unterschiedlich trainierten KIs) moderiert wird.

2.3.4 Visuelle Analyse der KI-basierten Segmentierungen

Die Dice-Koeffizienten sind ein mathematisch präzises und objektiv vergleichbares Maß für die Übereinstimmung zweier Segmentierungen. Für die praktische Anwendung automatisierter Segmentierungen der Prostata kann es aber beispielsweise vorkommen, dass eine KI auf MRT-Bildern die Prostata selbst sehr gut umrandet, gleichzeitig aber Strukturen einzeichnet, die außerhalb der Prostata liegen und damit die Höhe des Dice-Koeffizienten verringern. Eine Radiologin oder ein Radiologe, der diese Bilder beurteilt, würde die fälschlicherweise gezogenen Linien erkennen und gleichzeitig von einer präzise vorgenommenen Segmentierung der Prostata profitieren. Es lohnt sich deshalb, die automatisiert vorgenommenen Segmentierungen visuell zu analysieren und qualitativ zu bewerten.

Hierfür wurden die Segmentierungen der drei KIs für jeden einzelnen der 52 Patienten des Testdatensatzes von kranial nach kaudal gruppiert in die Organabschnitte „Prostatobasis“, „Drüsenmitte“ und „Apex“ gruppiert und nach der in Tabelle 5 gezeigten Kategorien beurteilt.

Tabelle 5. Kategorien (1 bis 4) zur qualitativen Analyse der Dice-Koeffizienten

1	Keine Abweichung im Vergleich zur manuellen Segmentierung
2	Geringe Abweichung im Vergleich zur manuellen Segmentierung
3	Starke Abweichung im Vergleich zur manuellen Segmentierung
4	Segmentierung von Strukturen außerhalb der Prostatagrenzen

Die nun folgenden Abbildungen 11 bis 15 sollen die zur visuellen Analyse definierten Kategorien aus Tabelle 5 veranschaulichen. In Gelb eingezeichnet sind jeweils die automatisch vorgenommenen Segmentierungen aller drei KIs, in Rot die manuell erstellte Segmentierung. Auf einigen Abbildungen sind alle vier Markierungen (drei Gelbe, eine Rote) zu sehen. Weitere Erläuterungen sind den Bildunterschriften zu entnehmen.

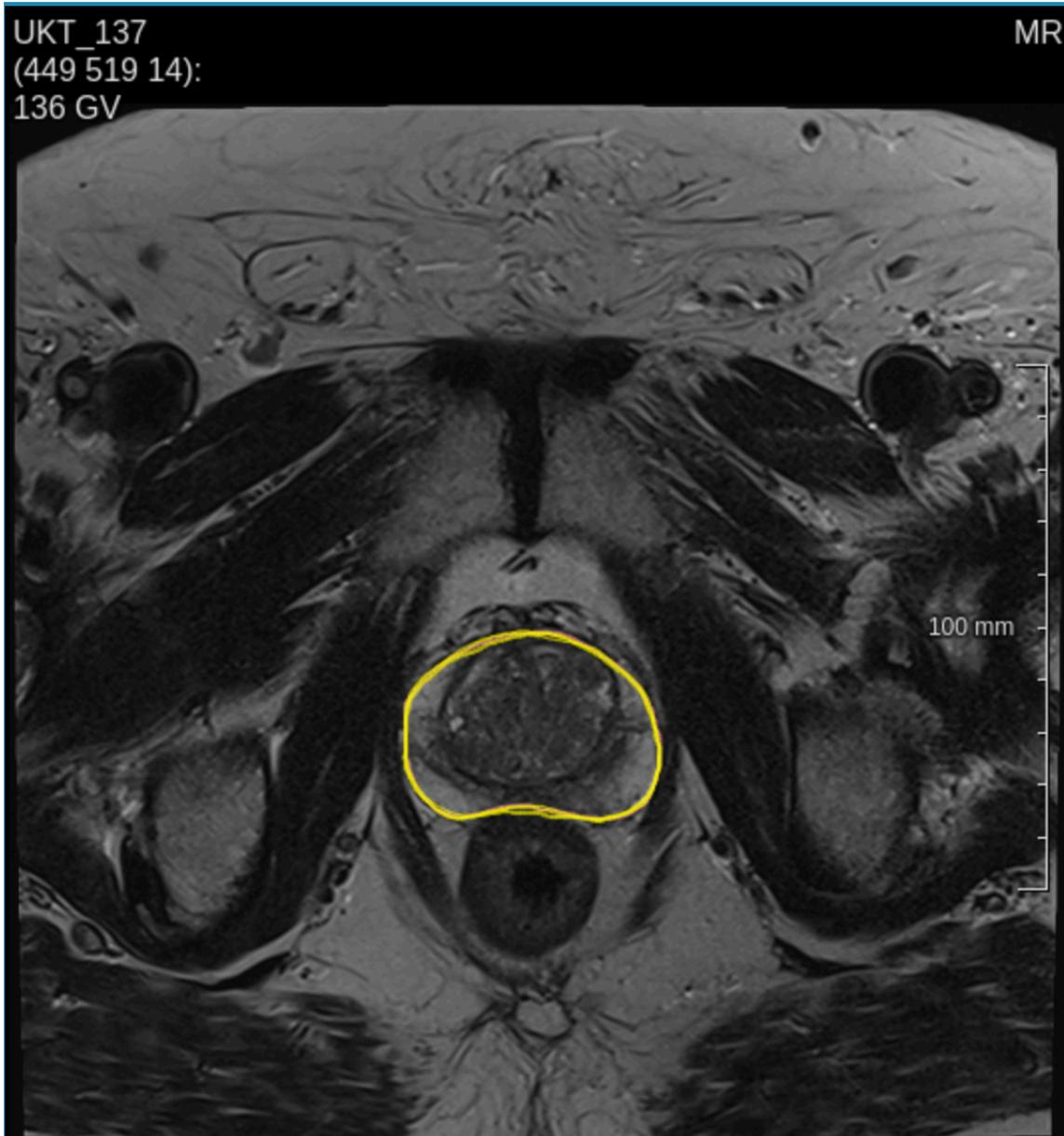


Abbildung 11. Beispiel für die Kategorie 1 im Bereich der Drüsenmitte: Nahezu vollständige Überlappung aller automatisch vorgenommenen Segmentierungen (Gelb) mit der manuellen Segmentierung (Rot).

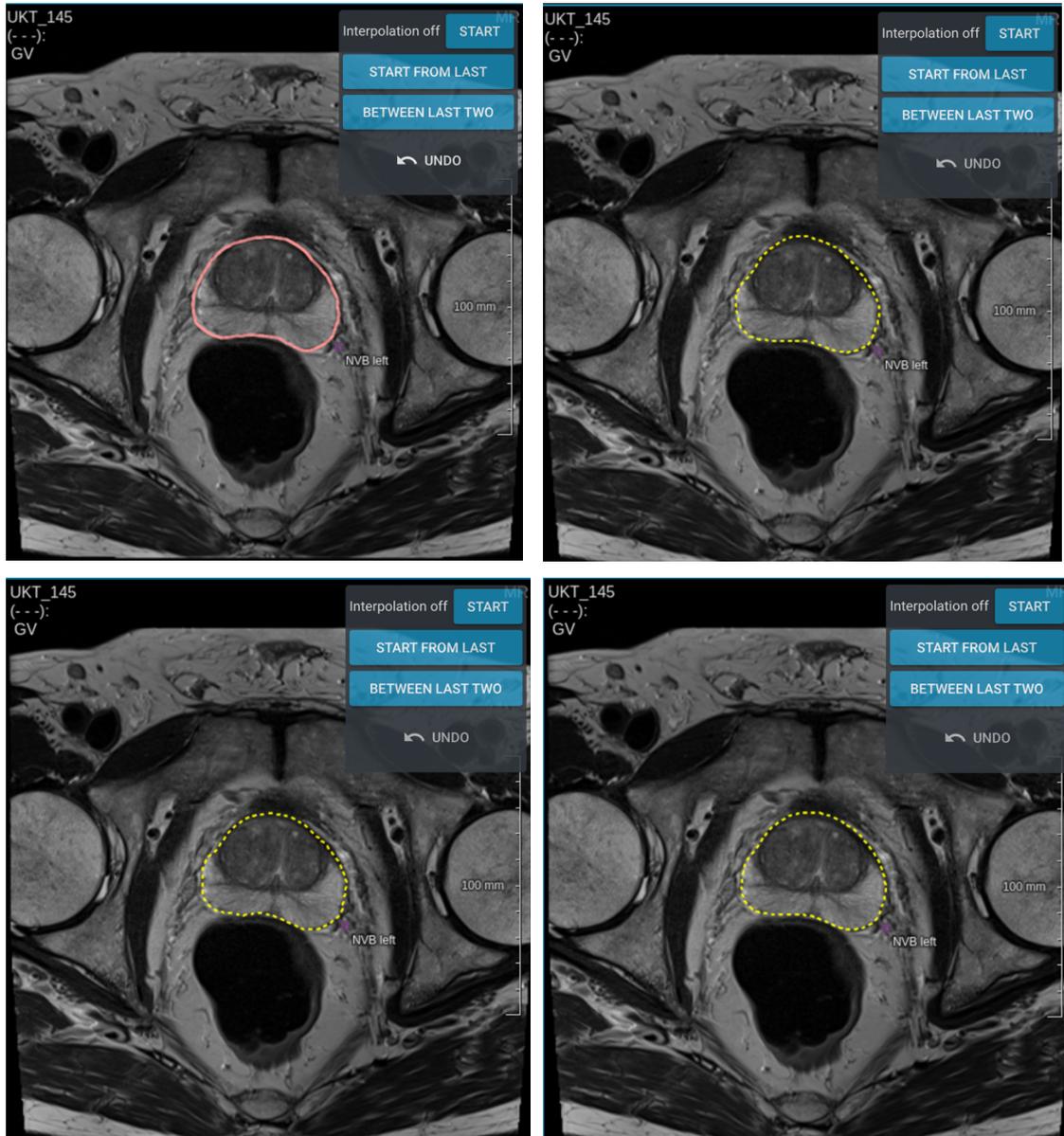


Abbildung 12. Beispiel für die Kategorie 1 auf Höhe der Drüsenmitte: Zu sehen sind auf allen vier Bildern dasselbe Schnittbild eines Patienten. Oben links in Rot ist die manuell erstellte Segmentierung zu sehen. Dem gegenüber sind auf den anderen Bildern gelb gestrichelt die automatisch vorgenommenen Segmentierungen der „KI Tübingen“ (oben rechts), der „KI Mannheim“ (unten links) und der „KI Gemischt“ (unten rechts) dargestellt. Auf allen Bildern ist rechts unterhalb der Segmentierung die manuelle Annotation des neurovaskulären Bündels links („NVB left“, lila Markierung) zu sehen.

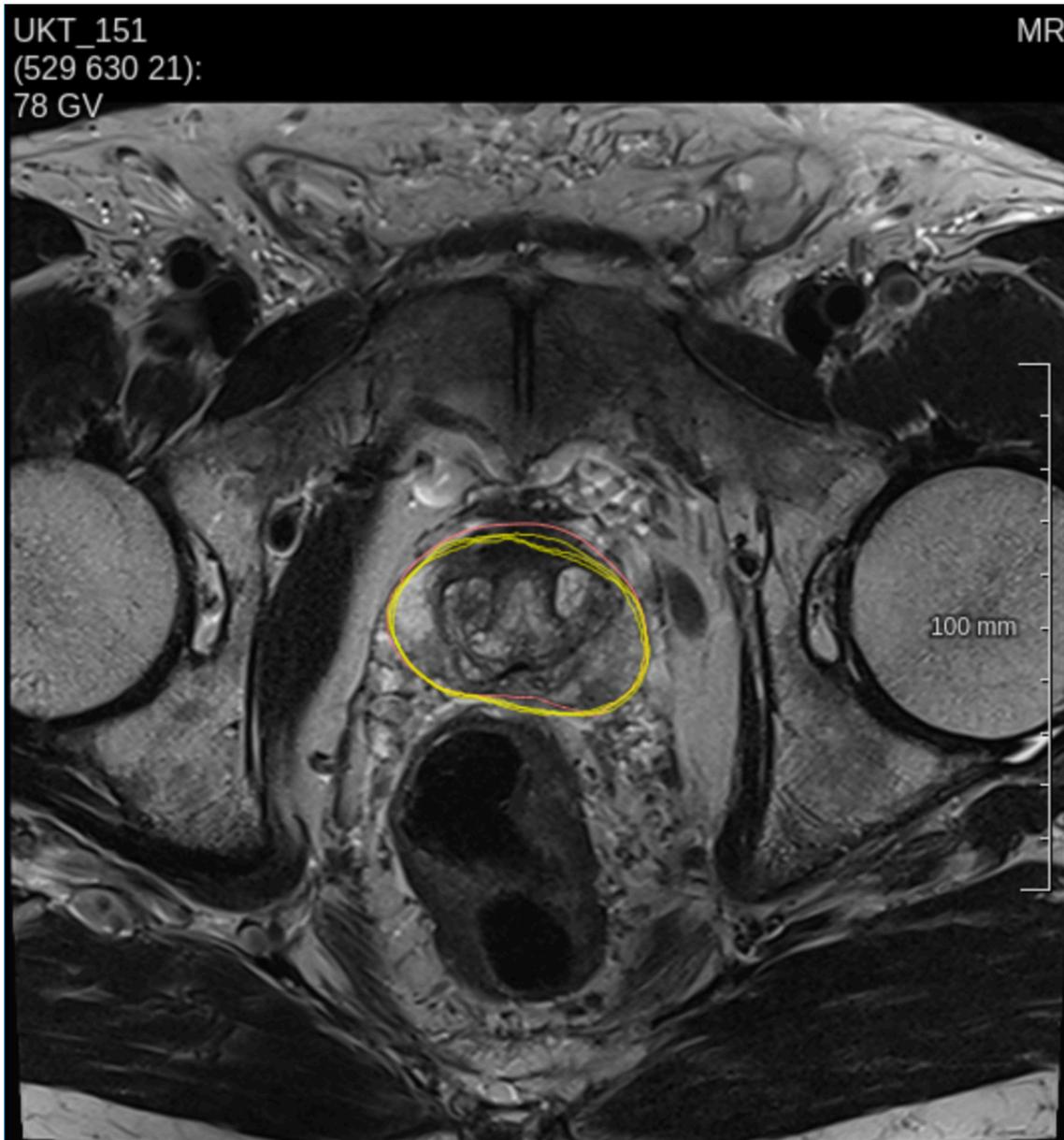


Abbildung 13. Beispiel für die Kategorie 2 am Übergang der Drüsenmitte zur Prostatabasis: Geringe Abweichung der Umrandungslinie (Gelb) aller automatisch vorgenommenen Segmentierungen von der manuellen Umrandungslinie (Rot).

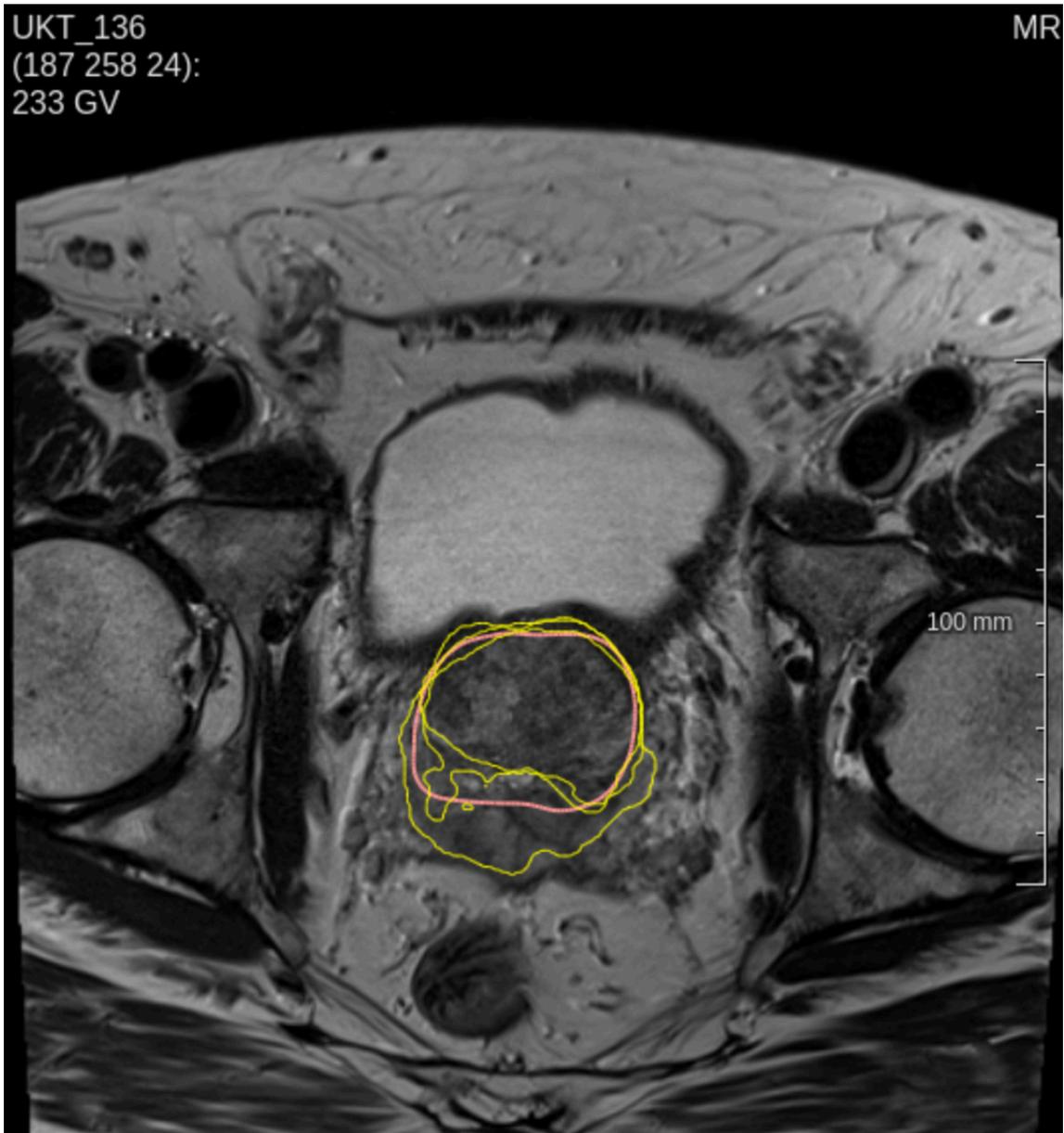


Abbildung 14. Beispiel für die Kategorie 3 im Bereich der Prostatabasis: starke Abweichung der Umrandungslinien aller drei automatisiert vorgenommenen Segmentierungen (Gelb) von der manuellen Segmentierung (Rot).

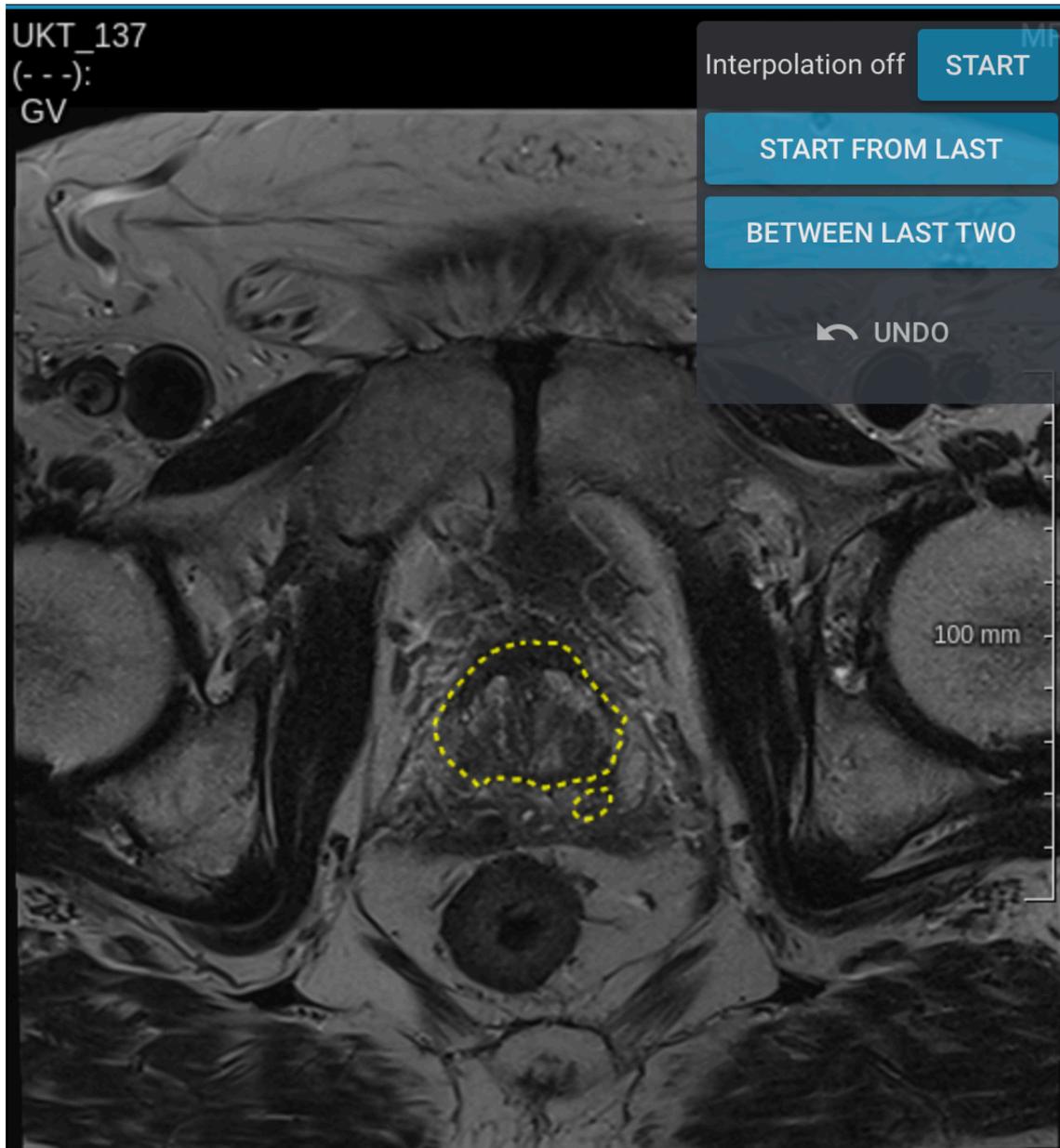


Abbildung 15. Beispiel für die Kategorie 4 im Bereich der Prostatabasis: die beiden gelb gestrichelten Linien entsprechen der automatisiert vorgenommenen Segmentierung der auf Grundlage der Tübinger Stichprobe trainierten KI. Die kleine Umrandung entspricht dabei der Segmentierung einer Struktur, die bildmorphologisch nicht zum Prostatagewebe gehört.

3 Ergebnisse

3.1 Studienpopulation der Stichprobe Tübingen

Im folgenden Abschnitt soll die im Rahmen der RESECT-Studie erhobene Stichprobe bestehend aus den 185 Patienten des Standortes Tübingen zunächst in ihrer Gesamtheit beschrieben werden ($N = 185$). Im Weiteren werden dann die Patientengruppe der Stichprobe Tübingen, die für das Trainieren der KIs herangezogen wurden ($n = 100$) und schließlich diejenige Patientengruppe, die allen drei trainierten KIs als Testdatensatz präsentiert wurden ($n = 52$), deskriptiv analysiert.

Da es sich ausnahmslos um prostatektomierte Patienten handelt, liegen ausschließlich Probanden männlichen Geschlechts vor. Daten zur ethnischen Herkunft konnten retrospektiv nicht erhoben werden.

Das Alter der Patienten bezieht sich auf den Zeitpunkt ihrer mpMRT-Untersuchung. Bei mehreren Untersuchungen wurde die zeitlich am nächsten zur Prostatektomie stehende herangezogen. Dies entspricht auch derjenigen Untersuchung, deren Bildmaterial später zur Segmentierung der Prostaten verwendet wurde.

Der jüngste Patient war zum Zeitpunkt seiner mpMRT-Untersuchung 45 Jahre alt, der älteste 81 Jahre. Das mittlere Alter betrug 66 Jahre ($SD = 7$). Das Alter ist gemäß Shapiro-Wilk-Test normalverteilt ($p = 0.053$). Abbildung 16 zeigt eine graphische Darstellung der Verteilung in Form eines Histogramms mit Normalverteilungskurve. Die Breite eines Balkens entspricht zwei Lebensjahren.

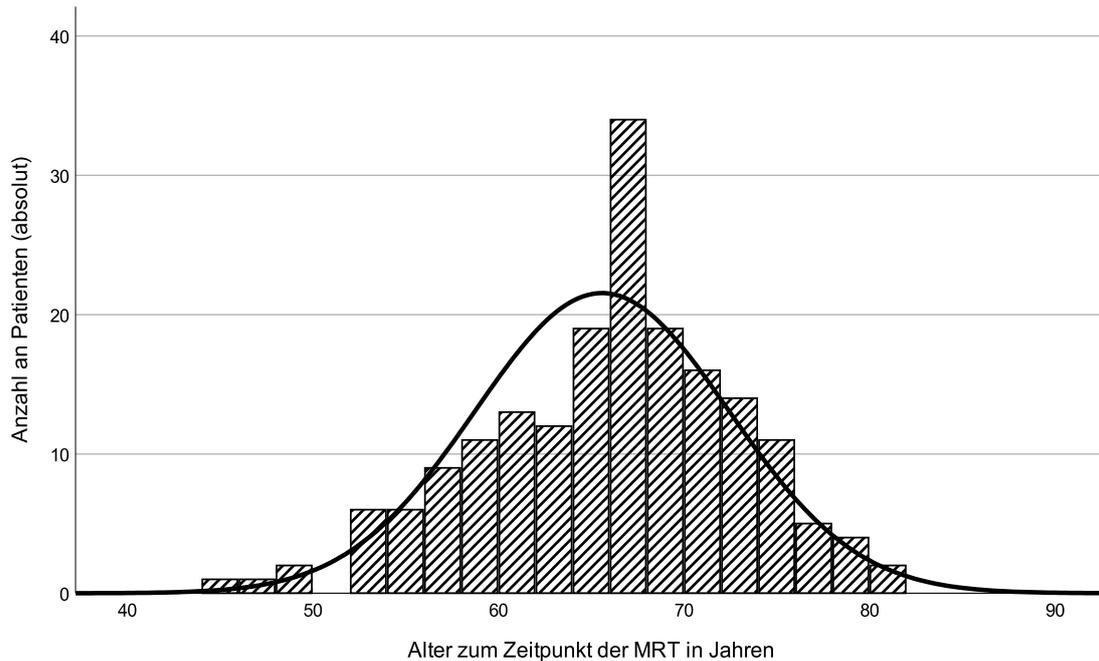


Abbildung 16. Histogramm „Alter zum Zeitpunkt der MRT“ mit Normalverteilungskurve

Tabelle 6 zeigt die Altersverteilung der 185 Patienten analog zu den später präsentierten Zahlen des Robert Koch Instituts eingeteilt in Gruppen zu je 5 Jahren. Die zahlenmäßig stärkste Gruppe innerhalb der Tübinger Stichprobe stellt die der 65 - 69 Jahre alten Patienten ($n = 65, 35.1\%$) dar. Laut den für das Jahr 2018 vom Robert Koch-Institut veröffentlichten Zahlen zur altersspezifischen Neuerkrankungsrate pro 100.000 Patienten in Deutschland hinsichtlich eines Prostatakarzinoms ist die Gruppe der 75 - 79 Jahre alten Männer die größte; das mittlere Erkrankungsalter (Median) lag dem Bericht des Robert Koch-Institutes zufolge im Jahr 2018 bei 71 Jahren. Die Zahl der Erkrankten unter 45 Jahren ist auch im deutschlandweiten Vergleich verschwindend gering: Für einen 35-jährigen Mann liegt das Risiko, in den nächsten 10 Jahren zu erkranken, bei unter 0.1 %. Anders als in der Allgemeinbevölkerung mit - im Jahr 2018 - fast 600 Neuerkrankungen pro 100.000 Männern im Jahr 2018 mit einem Alter von über 85 Jahren gab es in der Tübinger Stichprobe keinen Patienten über 81 Jahren, die Gruppe der 80 bis 84-jährigen zählte zwei Patienten.

Tabelle 6. Alter in Jahren zum Zeitpunkt der MRT, eingeteilt in Gruppen

Alter in Jahren	<i>n</i>	Prozent
45-49	4	2.2 %
50-54	9	4.9 %
55-59	23	12.4 %
60-64	32	17.3 %
65-69	65	35.1 %
70-74	34	18.4 %
75-79	16	8.6 %
80-84	2	1.1 %

Das mittlere Körpergewicht betrug zum Zeitpunkt der Narkoseaufklärung im direkten Vorfeld der Prostatektomie des jeweiligen Patienten 83.43 kg ($SD = 13.17$), die mittlere Körpergröße 1.77 m ($SD = 0.06$) und der daraus errechnete mittlere Body-Mass-Index 26.51 kg/m² ($SD = 3.82$). Der kleinste Patient maß 1.62 m, der größte 1.92 m. Das geringste Körpergewicht wurde mit 50.00 kg angegeben, das höchste mit 160.00 kg. Der BMI reichte von 17.72 kg/m² bis 46.26 kg/m².

Der Shapiro-Wilk-Test ergab für das Körpergewicht und den BMI keine Normalverteilung ($p < 0.001$). Die Körpergröße ist gemäß Shapiro-Wilk-Tests normalverteilt ($p = 0.266$).

Der Gleason-Score der jeweiligen Patienten wurde sowohl anhand der bei Verdacht auf ein Prostatakarzinom vorgesehenen Diagnostik entnommenen Stanzbiopsien als auch laut Pathologiebefund des Prostatektomiepräparates notiert. Letzterer konnte so bei allen 185 Patienten erhoben werden, stanzbiopsische Befunde lagen nur bei 180 der 185 Patienten vor. Bei diesen 5 fehlenden Befunden wurde bei entsprechenden Patienten entweder keine Stanzbiopsie vorgenommen oder bei nicht in domo durchgeführter Stanzbiopsie kein Befund in der jeweiligen Patientenakte hinterlegt.

Der Gleason-Score reichte in den befundeten Stanzbiopsien von 3+3=6 bis 5+5=10. Am häufigsten kam der Befund 3+4=7a vor ($n = 62$, 33.5 %).

In den befundeten Prostatektomiepräparaten reichte der Gleason-Score von 3+3=6 bis 5+4=9b. Der häufigste Score war 4+3=7a ($n = 70$, 38.8 %).

Tabelle 7 zeigt die prozentuale Verteilung der Gleason-Scores aller Stanzbiopsien sowie der Prostatektomiepräparate. Wie oben bereits erwähnt, konnte bei fünf Patienten kein stanzbiopischer Gleason-Score erhoben werden, diese wurden in Tabelle 7 mit „keine Angabe“ (k. A.) aufgeführt.

Tabelle 7. Gleason-Score (Stanzbiopsien und Prostatektomiepräparat)

Gleason-Score	Stanzbiopsien		Prostatektomiepräparat	
	<i>n</i>	Prozent	<i>n</i>	Prozent
3 + 3 = 6	11	5.9 %	2	1.1 %
3 + 4 = 7a	62	33.5 %	64	34.6 %
4 + 3 = 7b	38	20.5 %	70	37.8 %
4 + 4 = 8	47	25.4 %	22	11.9 %
4 + 5 = 9a	18	9.7 %	26	14.1 %
5 + 4 = 9b	2	1.1 %	1	0.5 %
5 + 5 = 10	2	1.1 %	-	-
k. A.	5	2.7 %	-	-

Der PSA-Wert wurde in den meisten Fällen für jeden Patienten zu mehreren Zeitpunkten abgenommen. Der jeweils hier zugrunde liegende PSA-Wert ist der der entsprechenden Prostatektomie zeitlich nächstgelegene. Die PSA-Dichte wurde aus diesem PSA-Wert und dem Prostatavolumen berechnet. Das jeweils dafür benötigte Prostatavolumen entspricht dem des mpMRT-Befundes. In den meisten Fällen konnte das Volumen anhand der drei ausgemessenen und dokumentierten Dimensionen „Breite“ (rechts-links-Abstand in cm), „Höhe“ (apikal-basal-Abstand in cm) und „Tiefe“ (ventral-dorsal-Abstand in cm) errechnet

werden. Smart Radiology hat diese Berechnung sowie die der PSA-Dichte nach Eingabe der entsprechenden Werte automatisch durchgeführt. Bei fehlender Dokumentation der genauen Ausmessung der Prostata im mpMRT-Befund wurde häufig ein approximiertes Volumen angegeben ($n = 46$). Die PSA-Dichte wurde aus dem Quotienten des PSA-Wertes und dieser Volumenangabe errechnet. In zwei der 185 Fällen wurde kein Prostatavolumen im mpMRT dokumentiert, entsprechend konnte auch keine PSA-Dichte errechnet werden.

Der mittlere PSA-Wert lag bei 10.20 ng/ml. Der höchste dokumentierte PSA-Wert lautete 56.00 ng/ml, der niedrigste 0.11 ng/ml ($SD = 8.43$). Die PSA-Dichte reichte von 0.01 ng/ml/cm³ bis 1.36 ng/ml/cm³. Die mittlere PSA-Dichte beträgt 0.25 ng/ml/cm³ ($SD = 0.21$).

PSA-Werte und -Dichte sind gemäß Test auf Normalverteilung nach Shapiro-Wilk nicht normalverteilt ($p < 0.001$). Abbildungen 17 und 18 veranschaulichen die Verteilung der PSA-Werte und -Dichte in Form zweier Histogramme mit eingezeichneter Normalverteilungskurve.

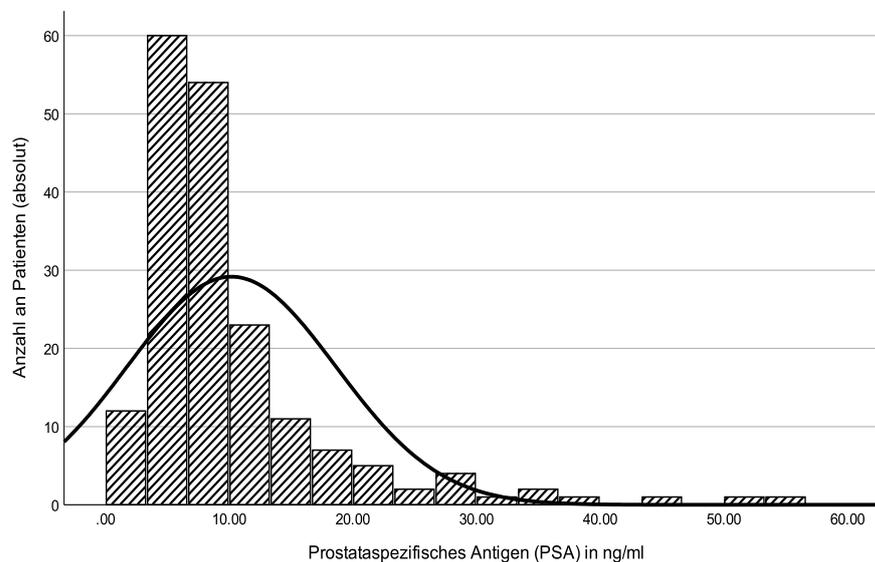


Abbildung 17. Histogramm „Prostataspezifisches Antigen (PSA) in ng/ml“ mit Normalverteilungskurve

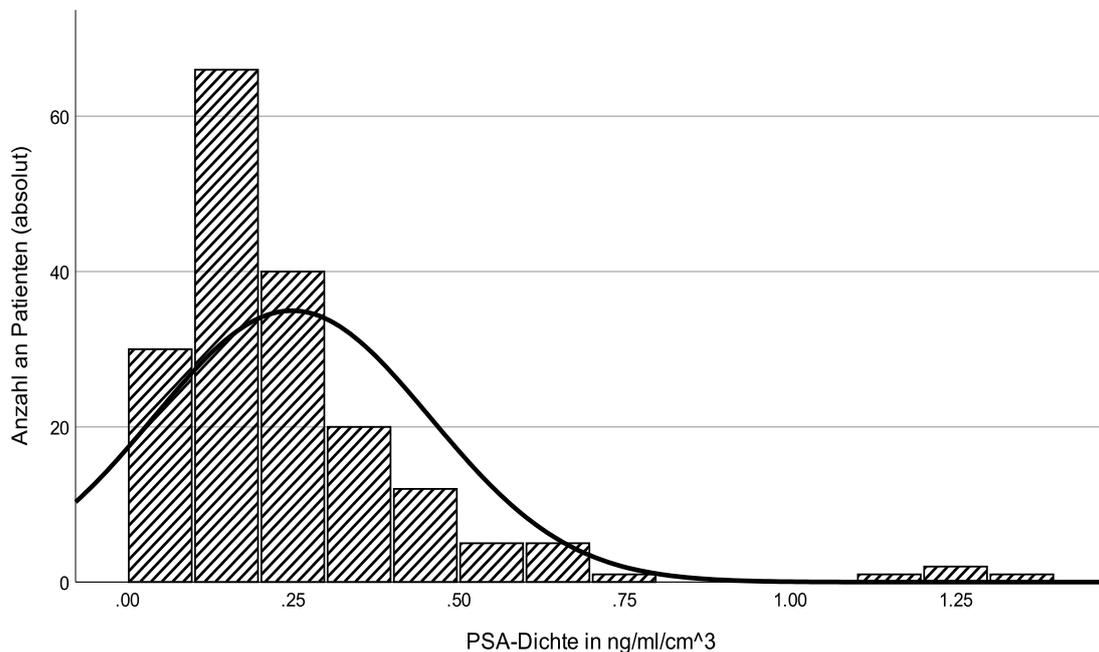


Abbildung 18. Histogramm „PSA-Dichte in ng/ml/cm³“ mit Normalverteilungskurve

Die DRU ist dem körperlichen Untersuchungsbefund des ambulanten oder stationären Arztbriefes des jeweiligen Patienten zum Zeitpunkt seiner Prostatektomie oder zu einer Vorstellung zu einem früheren Zeitpunkt entnommen. In 29 Fällen wurde die DRU nicht durchgeführt oder nicht dokumentiert. Bei 44.9 % ($n = 83$) der 185 Patienten war die DRU auffällig bzw. tastsuspekt, bei 39.5 % ($n = 73$) ergab sie einen unauffälligen Befund.

Der Status des Tumors in Hinblick auf eine extraprostatiche Ausbreitung (pT-Status), der Nodalstatus (pN-Status) sowie der Resektionsstatus (R-Status) wurden dem Pathologiebefund des Prostatektomiepräparates entnommen. Zu „extraprostaticher Ausbreitung“ wird bereits die Infiltration der Prostatakapsel in der Prostata-MRT gezählt.

Der pT-Status reicht von den Stadien pT2a bis pT3b, ein pT2c-Stadium lag in über der Hälfte der Fälle vor ($n = 104$, 65.2%). Tabelle 8 zeigt die detaillierte Verteilung der T-Stadien aller Patienten in Prozent.

Tabelle 8. pT-Status der Prostatektomiepräparate

pT-Status	<i>n</i>	Prozent
pT2a	22	11.9 %
pT2b	2	1.1 %
pT2c	104	56.2 %
pT3a	26	14.1 %
pT3b	31	16.8 %

Knapp über 85 % der Patienten hatten einen negativen Lymphknotenstatus (pN0, $n = 159$, 85.9 %), d.h. die entnommenen und untersuchten Lymphknoten waren tumorfrei. Bei einem Patienten (0.5 %) wurden Mikrometastasen beschrieben (pN1mi), bei 25 Patienten (13.5 %) hatte der Tumor zum Zeitpunkt der Prostatektomie bereits lymphogen gestreut.

Bei 126 der 185 Patienten (68.1 %) konnte die Prostata im Gesunden entfernt werden, d.h. die Schnittränder des Prostatektomiepräparates waren tumorfrei (R0 - Status). Bei einem Patienten (0.5 %) konnte keine klare Aussage über die Zuordnung der Schnittränder des Präparates getroffen werden, weshalb der Resektionsstatus als „unbekannt“ dokumentiert wurde (Rx-Status). Bei 58 Patienten (31.4 %) waren Karzinomreste randbildend (R1-Status).

Bei 19 der 185 Patienten konnte die Prostatektomie nicht nervenerhaltend durchgeführt werden, das entspricht 10,3 % aller Fälle. Knapp 90 % konnten nervenerhaltend operiert werden ($n = 166$, 89.7 %), wobei hier nicht unterschieden wird zwischen einem partiellen (links- oder rechtsseitigen) und vollständigem Nervenerhalt. Die Informationen hierzu wurden ausschließlich den jeweiligen Operationsberichten entnommen.

Die Maße des Prostatektomiepräparates (links-rechts, apikal-basal, ventral-dorsal in cm) sowie das Gewicht in Gramm sind dem Pathologiebefund entnommen. Das Präparatgewicht konnte bei allen 185 Patienten erhoben werden. Die einzelnen Längenangaben wurden im Befund nicht immer betitelt. In Rücksprache mit dem pathologischen Institut des Universitätsklinikums Tübingen

gibt es dort kein belastbares einheitliches Schema. Aus diesem Grund wurden die Dimensionen nur aus den exakt dokumentierten Befunden dokumentiert, bei 13 von 185 Befunden war dies nicht der Fall.

Tabelle 9 gibt eine detaillierte Übersicht über die oben genannten Größen.

Tabelle 9. Gewicht und Maße der Prostatektomiepräparate

	Gewicht in g	rechts-links (cm)	apikal-basal (cm)	ventral-dorsal (cm)
<i>n</i> erhoben bei	185	172	172	172
nicht erhoben bei	0	13	13	13
Mittelwert	52.1	5.0	4.1	3.4
Standardabweichung	22.6	0.7	0.8	0.7
Minimum	15.0	3.0	2.7	2.0
Maximum	190.0	7.3	7.5	6.5

Abbildung 19 zeigt graphisch die Verteilung des Gewichtes des Prostatektomiepräparates in Form eines Histogramms mit eingezeichneter Normalverteilungskurve. Die Werte des Präparatgewichts sind nach Test auf Normalverteilung nach Shapiro - Wilk nicht normalverteilt ($p < 0.001$).

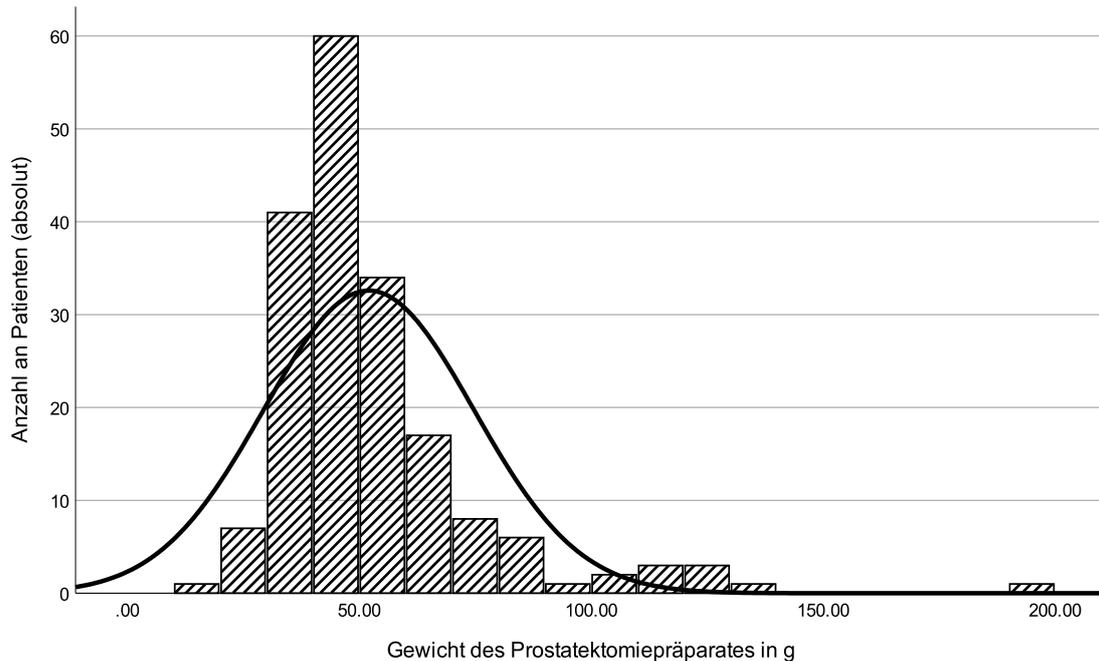


Abbildung 19. Histogramm „Gewicht des Prostatektomiepräparates“ mit Normalverteilungskurve

Im Mittel betrug die Zeit zwischen der zuletzt vor der Prostatektomie durchgeführten MRT-Untersuchung und dem Tag der Prostatektomie 93.6 Tage, die Spannweite betrug 0 bis 834 Tage ($SD = 95.85$).

Tabelle 10 zeigt die Häufigkeit der zur Untersuchung der jeweiligen mpMRT verwendeten Scanner. In über der Hälfte der Fälle ($n = 103, 55.7\%$) wurde mit dem 3 Tesla Magnetresonanztomographen „MAGNETOM Skyra“ gearbeitet. Der Anteil der Untersuchungen mit den beiden 1.5 Tesla Magnetresonanztomographen „Magnetom Avanto fit“ und „MAGNETOM Aera“ (MR62) betrug mit zusammen 5 der insgesamt 185 Fälle weniger als 3% ($n = 5, 2.7\%$).

Tabelle 10. Eingesetzte MRT-Scanner

Scanner	Tesla	<i>n</i>	%
Magnetom Avanto fit	1,5	3	1.6%
Magnetom Skyra	3	103	55.7%
Magnetom Prisma	3	45	24.3%
Magnetom Vida	3	31	16.8%
Magnetom Aera	1,5	2	1.1%
Magnetom Vida fit	3	1	0.5%

Abbildung 20 zeigt ein Histogramm mit der Verteilung der Prostata volumina in der MRT-Bildgebung. Tabelle 11 gibt einen Überblick über die statistischen Kenngrößen dazu. Den angegebenen Volumina und errechneten statistischen Werten liegen die jeweiligen mpMRT-Befunde zugrunde. In zwei der 185 MRT-Befunden wurden weder ein Prostata volumen noch die einzelnen Dimensionen dokumentiert bzw. ausgemessen.

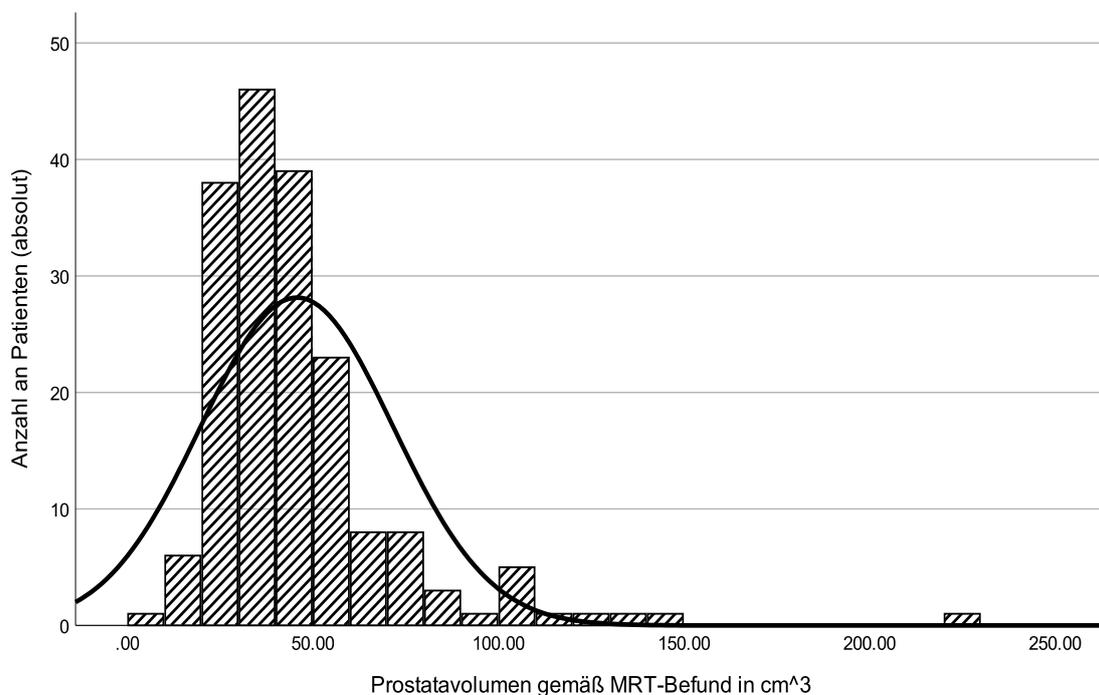


Abbildung 20. Histogramm mit Normalverteilungskurve „Prostata volumen gemäß MRT-Befund“

Tabelle 11. Prostatavolumen (MRT-Befund) in cm³

<i>n</i> erhoben bei	183
nicht erhoben bei	2
<hr/>	
Mittelwert	45.74
Standardabweichung	25.93
Minimum	5.04
Maximum	221.00
<hr/>	

Martins et al. (2020) korrelierten und verglichen in ihrer Studie das Gewicht von Prostatektomiepräparaten mit den gemessenen Volumina im MRT und mittels TRUS. Sie gehen dabei - ähnlich einiger anderer Veröffentlichungen, die sie nennen - von einer Prostatadichte von etwa 1.0 g/ml aus (Martins et al., 2020). Martins et al. (2020) weisen außerdem darauf hin, dass sich das tatsächliche Prostatagewicht durch die Aufbewahrung des Präparates bis zur pathologischen Untersuchung in Formalin verringern kann. Samenbläschen und periprostatiches Fettgewebe könnten ebenfalls Einfluss auf sowohl die Gewebedichte als auch das postoperativ gewogene Gewicht haben (Martins et al., 2020). Auf Grundlage dieser Vereinfachungen werden im Folgenden ebenfalls die in dieser Studie gemessenen MRT-Volumina mit den Gewichten der Prostatektomiepräparaten verglichen.

Die Volumina der Prostata laut MRT-Befund korrelierten stark mit dem Gewicht, das im Pathologiebefund des Prostatektomiepräparates angegeben wurde, $r = 0.903$, $p < 0.001$. Abbildung 21 zeigt ein Streudiagramm der beiden Variablen.

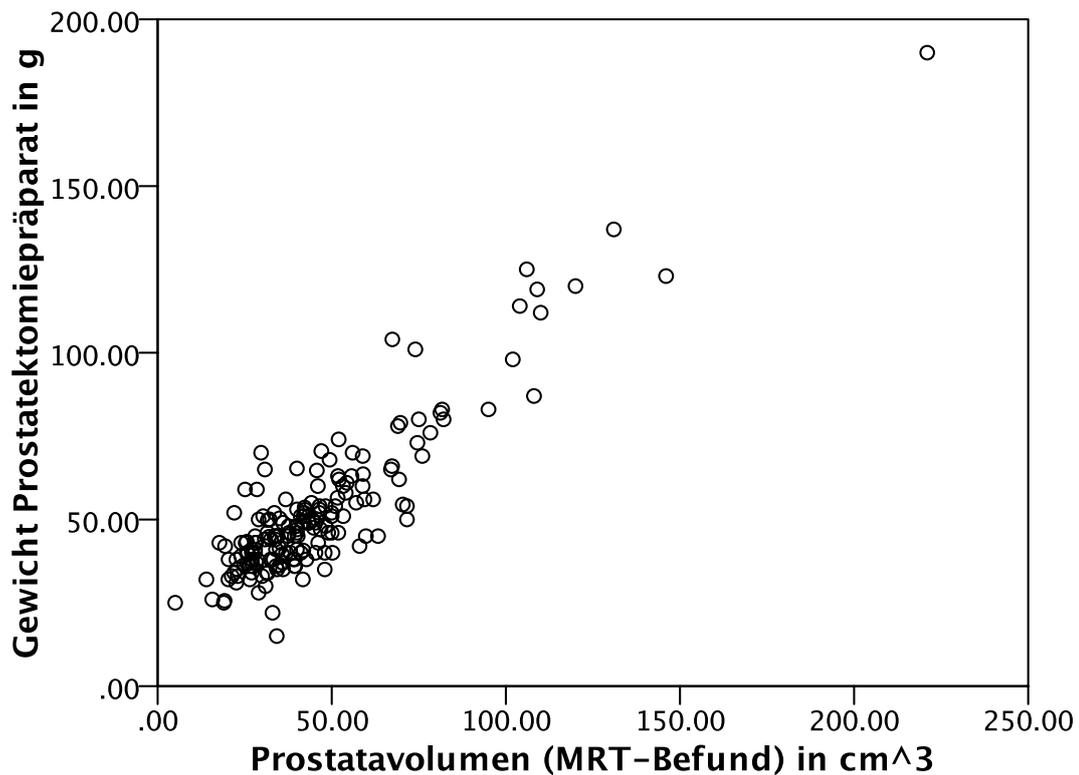


Abbildung 21. Streudiagramm zum Gewicht der Prostatektomiepräparate und den Volumina der Prostatae im MRT

Die in der Prostata-MRT gemessenen Volumina in Kubikzentimetern [cm³] prognostizierten unter Zugrundelegung der oben erläuterten Gewebedichte von 1.0 g/ml signifikant niedrigere Werte als das jeweilige Gewicht des Prostatektomiepräparates in Gramm [g], $N = 183$, $M = 6.495$, $SD = 11.16$, $t(182) = 7.873$, $p < 0.001$. Abbildung 22 zeigt die beiden Variablen dargestellt in einem Boxplot.

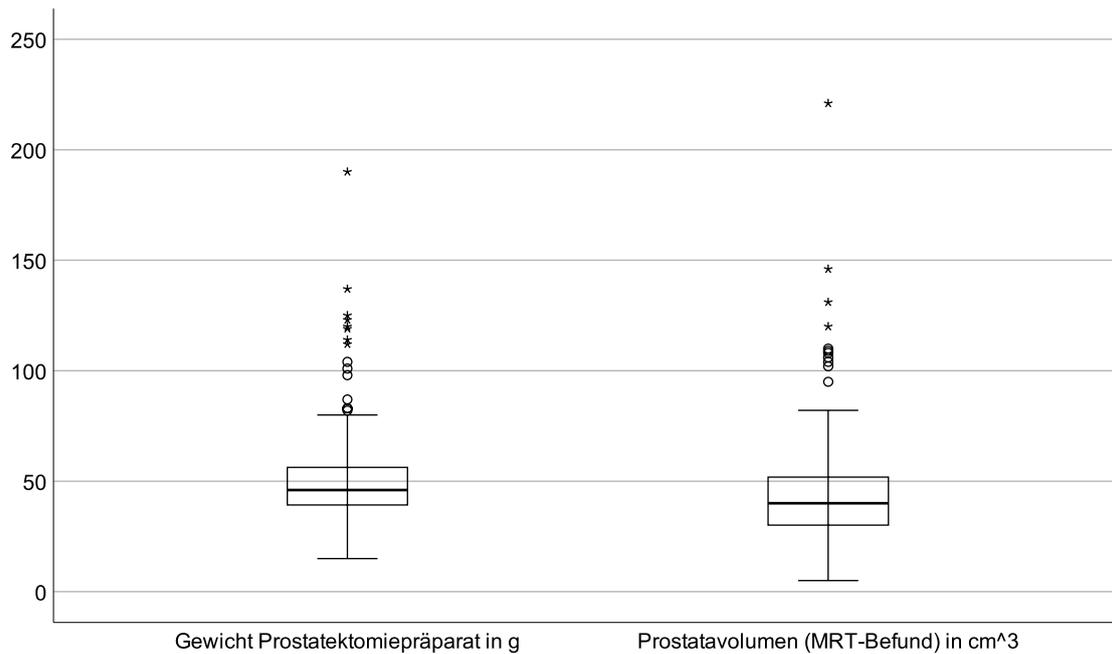


Abbildung 22. Boxplot zur Verteilung nach Gewicht der Prostatektomiepräparate und der Volumina der Prostata in der MRT

Tabelle 12 zeigt die statistischen Merkmale der Parameter „Diameter der Indexläsion in der MRT“ sowie „Anzahl suspekter Läsionen in der MRT“.

Tabelle 12. Anzahl suspekter Läsionen und Diameter der Indexläsion in cm in der MRT

	Anzahl suspekter Läsionen	Ø der Indexläsion (cm)
<i>n</i> erhoben bei	185	167
nicht erhoben bei	0	18
Mittelwert	1.29	1.65
Standardabweichung	0.57	1.87
Minimum	0	0.1
Maximum	3	20.00

Die jeweiligen PI-RADS-Scores wurden dem MRT-Bericht entnommen. In einem Fall (UKT_10) wurde der PI-RADS-Score nachträglich im Rahmen der Erhebung durch den Doktoranden hochkorrigiert. Der PI-RADS-Score wurde im Befund mit 4 beziffert, der Diameter der Indexläsion jedoch mit 1.7 cm angegeben, was formal einem PI-RADS-Score von 5 entspricht. Bei einem der 185 Patienten wurde keine suspekta Läsion im MRT-Befund beschrieben, es konnte entsprechend auch keine Klassifikation nach PI-RADS vorgenommen werden. In 18 Fällen konnte kein Diameter der Indexläsion erhoben werden, da dieser nicht dokumentiert wurde oder keine suspekta Läsion im MRT-Befund beschrieben wurde.

Tabelle 13 zeigt die prozentuale Verteilung der PI-RADS-Scores.

Tabelle 13: Verteilung der PI-RADS-Scores

PI-RADS-Score	<i>n</i>	Prozent
1	-	-
2	2	1.1 %
3	13	7.0 %
4	96	51.9 %
5	73	39.5 %
k. A.	1	0.5 %

In etwas mehr als der Hälfte der 185 Fälle lag ein PI-RADS-Score von 4 vor ($n = 96$, 51.9%), in fast 40 % ein PI-RADS-Score von 5 ($n = 73$, 39.5 %). In Summe lag also bei 90.4 % der Patienten magnetresonanztomographisch ein wahrscheinlich bzw. sehr wahrscheinlich bösartiger Befund vor (PI-RADS 4 + 5, $n = 169$). Der Anteil derer, die in der MRT-Bildgebung wahrscheinlich gutartige Läsionen (PI-RADS 2) zu verzeichnen hatten, lag mit 2 der 185 Fällen bei 1.1 %.

Bei 73.5 % aller Patienten lag ein unifokaler Befund vor, d.h. es gab MRT-morphologisch lediglich eine suspekta Läsion ($n = 136$). Rund ein Viertel wiesen mehr als eine suspekta Läsion auf ($n = 47$, 25.4 %). Bei zwei Patienten ergab die

MRT-Untersuchung keinen Karzinom-verdächtigen Befund (PI-RADS 2), formal lag bei diesen beiden entsprechend auch keine Läsion vor ($n = 2$, 1.1%).

Tabelle 14 gibt die Anzahl der suspekten Läsionen an.

Tabelle 14. Anzahl suspekter Läsionen

Anzahl	<i>n</i>	Prozent
0	2	1.1 %
1	136	73.5 %
2	38	20.5 %
3	9	4.9 %

Bei 26.5 % aller Patienten ($n = 49$) war bildmorphologisch eine extraprostatiche Ausbreitung zu erkennen, das heißt, die Prostatakapsel wurde infiltriert, überschritten oder ein anderes Organ war bereits befallen. In 136 Fällen (73.5 %) war dies nicht der Fall.

3.2 Trainings- und Testdatensatz der Stichprobe Tübingen

Tabelle 15 zeigt die demographischen und klinischen Charakteristika sowohl der 100 Patienten, die zum Trainieren einer der drei KIs herangezogen wurde (Trainingsdatensatz), als auch der 52 Patienten, deren MRT-Bilder der KI nach Beüben durch den jeweiligen Trainingsdatensatz präsentiert wurden (Testdatensatz). Beim Trainingsdatensatz handelt es sich konsequent um die Patienten 1 bis 100 der insgesamt 185 Patienten umfassenden Stichprobe Tübingen, beim Testdatensatz konsequent um die Patienten 134 bis 185.

Ergebnisse

Tabelle 15. Demographische und klinische Charakteristika des Trainings- (UKT_1 bis UKT_100) und Testdatensatzes (UKT_134 bis UKT_185) der Stichprobe Tübingen

	Trainingsdatensatz (<i>n</i> = 100)	Testdatensatz (<i>n</i> = 52)
Alter zum Zeitpunkt der mpMRT (Jahre)		
- Mittelwert (<i>SD</i> , Min. - Max.)	65 (7, 47 - 81)	66 (7, 45 - 79)
Körpergröße (m)		
- Mittelwert (<i>SD</i> , Min. - Max.)	1.77 (0.06, 1.64 - 1.92)	1.79 (0.06, 1.68 - 1.92)
Körpergewicht (kg)		
- Mittelwert (<i>SD</i> , Min. - Max.)	81.90 (11.93, 58 - 121)	86.67 (16.57, 50 - 160)
BMI (kg/m²)		
- Mittelwert (<i>SD</i> , Min. - Max.)	26.20 (3.68, 18.21 - 39.06)	27.09 (4.44, 17.72 - 46.25)
PSA-Wert (ng/dl)		
- Mittelwert (<i>SD</i> , Min. - Max.)	10.76 (8.52, 1.19 - 53.00)	9.99 (9.46, 1.10 - 56.00)
DRU (<i>n</i>, (%))		
- auffällig/tastsuspekt	48 (48 %)	26 (50.0 %)
- unauffällig/nicht tastsuspekt	36 (36 %)	23 (44.2 %)
- nicht dokumentiert	16 (16 %)	3 (5.8 %)
PI-RADS-Score (<i>n</i>, (%))		
- 1	-	-
- 2	2 (2 %)	-
- 3	9 (9 %)	2 (3.8 %)
- 4	48 (48 %)	27 (51.9 %)
- 5	41 (41 %)	23 (44.2 %)
Gleason-Score (Prostatektomiepräparat) (<i>n</i>, (%))		

Ergebnisse

- 3 + 3 = 6	-	2 (3.8 %)
- 3 + 4 = 7a	37 (37 %)	17 (32.7 %)
- 4 + 3 = 7b	35 (35 %)	20 (38.5 %)
- 4 + 4 = 8	12 (12 %)	5 (9.6 %)
- 4 + 5 = 9a	15 (15 %)	8 (15.4 %)
- 5 + 4 = 9b	1 (1 %)	-

Gleason-Score (Stanzbiopsie) (*n*, (%))

- 3 + 3 = 6	6 (6 %)	2 (3.8 %)
- 3 + 4 = 7a	32 (32 %)	18 (34.6 %)
- 4 + 3 = 7b	21 (21 %)	10 (19.2 %)
- 4 + 4 = 8	28 (28 %)	11 (21.2 %)
- 4 + 5 = 9a	8 (8 %)	7 (13.5 %)
- 5 + 4 = 9b	-	2 (3.8 %)
- 5 + 5 = 10	1 (1 %)	1 (1.9 %)
- k. A.	4 (4 %)	1 (1.9 %)

Gewicht Prostatektomiepräparat (g)

- Mittelwert (<i>SD</i> , Min. - Max.)	50.37 (20.35, 22 - 137)	49.20 (17.69, 15.00 - 125.00)
---	-------------------------	-------------------------------

Nerverhaltende Operation (*n*, (%))

- ja	89 (89 %)	47 (90.4 %)
- nein	11 (11 %)	5 (9.6 %)

Postoperativer Resektionsstatus (*n*, (%))

- R0	67 (67 %)	35 (67.3 %)
- R1	32 (32 %)	17 (32.7 %)
- Rx	1 (1 %)	-

Ergebnisse

pT-Status (*n*, (%))

- pT2a	7 (7 %)	10 (19.2 %)
- pT2b	1 (1 %)	1 (1.9 %)
- pT2c	60 (60 %)	21 (40.4 %)
- pT3a	12 (12 %)	13 (25.0 %)
- pT3b	20 (20 %)	7 (13.5 %)

pN-Status (*n*, (%))

- pN0	87 (87 %)	43 (82.7 %)
- pN1	13 (13 %)	8 (15.4 %)
- pN1mi	-	1 (1.9 %)

MRT-Scanner (*n*, (%))

- Magnetom Avanto fit	1 (1 %)	1 (1.9 %)
- Magnetom Skyra	67 (67 %)	19 (36.5 %)
- Magnetom Prisma	18 (18 %)	19 (36.5 %)
- Magnetom Vida	14 (14 %)	10 (19.2 %)
- Magnetom Aera	-	2 (3.8 %)
- Magnetom Vida fit	-	1 (1.9 %)

Feldstärke (*n*, (%))

- 1.5 T	1 (1 %)	3 (5.8 %)
- 3 T	99 (99 %)	49 (94.2 %)

Extraprostatische Ausbreitung (*n*, (%))

- ja	28 (28 %)	15 (28.8 %)
- nein	72 (72%)	37 (71.2 %)

Ergebnisse

Uni- / Multifokale Läsion(en) (n, (%))

- unifokal	70 (70 %)	42 (80.8 %)
- multifokal	28 (19 %)	10 (19.2 %)
- keine Läsion im MRT	2 (2 %)	-

Anzahl suspekter Läsionen (n, (%))

- 0	2 (2 %)	-
- 1	70 (70 %)	42 (80.8 %)
- 2	21 (21 %)	8 (15.4 %)
- 3	7 (7 %)	2 (3.8 %)

Das Prostatavolumen wurde bei 99 der 100 Patienten des Trainingsdatensatzes im MRT-Befund angegeben. Entsprechend konnte bei nur 99 der 100 Patienten die PSA-Dichte errechnet werden. Auch in der Kohorte des Testdatensatzes wurde das Prostatavolumen nur bei 51 der 52 Patienten im MRT-Befund angegeben, die PSA-Dichte konnte bei nur 51 Patienten errechnet werden. Tabelle 16 zeigt die entsprechenden Werte.

Tabelle 16. Prostatavolumen und PSA-Dichte des Trainings- und des Testdatensatzes

	Trainingsdatensatz (n = 99)	Testdatensatz (n = 51)
<hr/>		
Prostatavolumen im MRT-Befund (cm ³)		
- Mittelwert (SD, Min. - Max.)	45.67 (22.94, 17.72 - 131.00)	39.91 (17.14, 5.04 - 106.00)
PSA-Dichte (ng/dl/cm ³)		
- Mittelwert (SD, Min. - Max.)	0.25 (0.20, 0.03 - 1.22)	0.27 (0.26, 0.04 - 1.36)

Der Durchmesser einer Indexläsion konnte bei 92 der 100 Patienten des Trainingsdatensatzes erhoben werden. Bei den fehlenden 8 Personen wurde er entweder nicht angegeben oder es wurde keine karzinomverdächtige Läsion im MRT-Befund beschrieben. Der Mittelwert lag bei 1.80 cm ($sd = 2.45$, *Min.* - *Max.* 0.15 - 20.00). Im Testdatensatz konnte der Durchmesser einer Indexläsion bei 45 der 52 Patienten erhoben werden. Bei den fehlenden 7 Männern wurde er entweder nicht angegeben oder es wurde keine karzinomverdächtige Läsion im MRT-Befund beschrieben. Der Mittelwert lag bei 1.59 cm ($sd = 0.62$, *Min.* - *Max.* 0.80 - 3.80).

Zur Untersuchung auf Unterscheidung zwischen dem Trainingsdatensatz ($n = 100$) und dem Testdatensatz ($n = 52$) wurde bei den metrischen klinischen Parametern ein ungepaarter t-Test durchgeführt. Um die beiden Datensätze auf Varianzgleichheit zu untersuchen, wurde zuvor ein Levene-Test durchgeführt. Demnach ist für die Varianzen der Variablen „Körpergröße“ ($p = 0.548$), „Körpergewicht“ ($p = 0.197$), „BMI“ ($p = 0.319$), „PSA-Wert“ ($p = 0.639$), „PSA-Dichte“ ($p = 0.278$), „Prostatavolumen gemessen im MRT“ ($p = 0.116$), sowie „Durchmesser der Indexläsion“ ($p = 0.083$) Gleichheit anzunehmen. Für die Variablen „Feldstärke des MRT-Scanners“ ($p < 0.001$), konnte keine Gleichheit der Varianzen gezeigt werden.

Wie Tabelle 17 zeigt, unterscheiden sich der Trainingsdatensatz und der Testdatensatz nicht signifikant hinsichtlich der klinischen Parameter „Körpergröße“, „BMI“, „PSA-Wert“, „PSA-Dichte“, „Prostatavolumen gemessen im MRT“, „Durchmesser der Indexläsion“ und „Feldstärke des MRT-Scanners“.

Tabelle 17. Statistische Kennzahlen zur Untersuchung auf Unterscheidung hinsichtlich verschiedener klinischer Parameter zwischen Trainings- und Testdatensatz

	95% - Konfidenzintervall		<i>t</i> (150)	<i>p</i> -Wert
	Untere Grenze	Obere Grenze		
Körpergröße	-0.03797	0.00226	-1.754	0.081
BMI	-2.22495	0.44941	-1.312	0.192
PSA-Wert	-2.21523	3.76191	0.511	0.610
PSA-Dichte	-0.08795	0.06153	-0.349	0.727
Prostatavolumen gemessen im MRT	-1.44407	12.96864	1.580	0.116
Diameter der Indexläsion	-0.52436	0.94538	0.567	0.572
Feldstärke des MRT-Scanners	-0.0309	0.1740	1.397	0.168

Hinsichtlich des klinischen Parameters Körpergewicht gab es einen signifikanten Unterschied zwischen den beiden Datensätzen (95%-CI [-9.38133, -0.13744]), $t(150) = -2.035$, $p = 0.044$), wobei das Körpergewicht der Patienten des Testdatensatzes durchschnittlich 4.76 kg größer war als das der Patienten des Trainingsdatensatzes.

3.3 Vergleich der Dice-Koeffizienten

Der Mittelwert der Dice-Koeffizienten der Stichprobe Tübingen beträgt 0.8905 ($n = 52$, $sd = 0.0421$, $se = 0.0058$, 95% CI [0.8788, 0.9022], $md = 0.9003$). Der Mittelwert der Stichprobe Mannheim 0.8839 ($n = 52$, $sd = 0.0478$, $se = 0.0066$, 95% CI [0.8706, 0.8971], $md = 0.8952$), der der gemischten Stichprobe 0.8920 ($n = 52$, $sd = 0.0392$, $se = 0.0054$, 95% CI [0.8811, 0.9029], $md = 0.9006$). Die Dice-Koeffizienten sind gemäß dem Test nach Shapiro Wilk für keine der drei Stichproben normalverteilt ($p < 0.001$). Der kleinste gemessene Dice-Koeffizient (Minimum) beträgt für die Tübinger Stichprobe 0.7259, der größte (Maximum) 0.9423. Für die Mannheimer Stichprobe betragen diese beiden Werte 0.6806 und 0.9387, für die gemischte Stichprobe 0.7207 und 0.9488. Abbildung 23 veranschaulicht die Verteilung der Dice-Koeffizienten gruppiert nach den drei KIs in einem Box-Plot, Tabelle 18 soll die statistischen Kenngrößen zu den erhobenen Dice-Koeffizienten übersichtlich zusammenfassen.

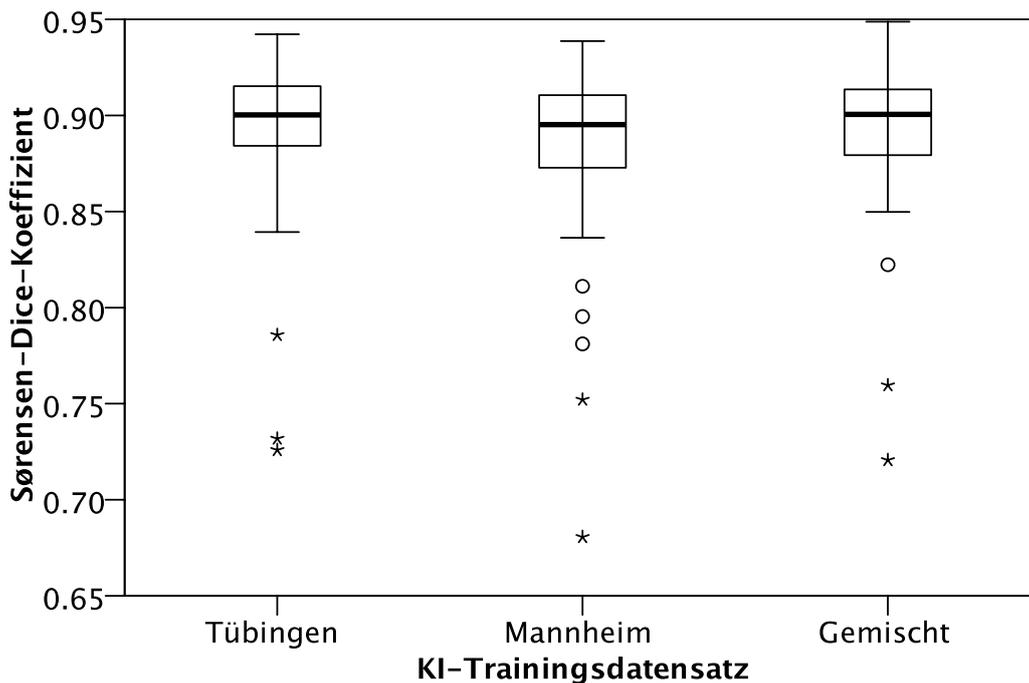


Abbildung 23. Box-Blot zur Verteilung der Dice-Koeffizienten nach Stichprobe

Tabelle 18. Dice-Koeffizienten der drei KIs

Trainingsdatensatz ($n = 51$)	Dice-Koeffizient	<i>SD</i>	<i>MD</i>	95%-Konfidenzintervall
KI 1 „Tübingen“	0.8905	0.0421	0.9003	0.8788 - 0.9023
KI 2 „Mannheim“	0.8839	0.0477	0.8952	0.8706 - 0.8972
KI 3 „gemischt“	0.8920	0.0392	0.9006	0.8811 - 0.9029

Zur Beurteilung der Unterschiedlichkeit der drei KIs untereinander wurde ein Äquivalenztest durchgeführt, zusätzlich wurde eine Korrelation nach Bravais-Pearson berechnet. Für den Äquivalenztest wurden jeweils die 90%-Konfidenzintervalle der Differenzen der Dice-Koeffizienten der gepaarten Stichproben „Tübingen-Mannheim“ ($r = 0.789$), „Tübingen-gemischt“ ($r = 0.913$) und „Mannheim-gemischt“ ($r = 0.752$) berechnet.

Das 90%-Konfidenzintervall der Differenz der Dice-Koeffizienten für das Stichprobenpaar „Tübingen-Mannheim“ beträgt -0.00026 (unterer Wert) bis 0.01353 (oberer Wert). Das 90%-Konfidenzintervall für das Stichprobenpaar

„Tübingen-Mix“ beträgt -0.00545 (unterer Wert) bis 0.00252 (oberer Wert). Das 90%-Konfidenzintervall für das Stichprobenpaar „Mannheim-Mix“ beträgt -0.01545 (unterer Wert) bis -0.00075 (oberer Wert). Alle drei 90%-Konfidenzintervalle liegen innerhalb der Grenzen von ± 0.01 . Damit ist für alle drei Stichprobenpaare untereinander Äquivalenz nachgewiesen.

Um dieses Ergebnis neben dem Sørensen-Dice Koeffizienten vergleichbar mit anderen Methoden zum Überprüfen der Überschneidung zweier Bild-Segmentationen machen zu können, soll die Äquivalenz zusätzlich durch das dimensionslose Maß der Effektstärke d nach Cohens nachgewiesen werden. Dazu wurden nach gängiger Übereinkunft eines „kleinen Effektes“ bei Werten von d kleiner 0.5 (Rice & Harris, 2005) die obere Grenze des 90%-Konfidenzintervalls auf $d = +0.5$, die untere auf $d = -0.5$ festgelegt.

Das 90%-Konfidenzintervall der Cohens-d-Werte für das Stichprobenpaar „Tübingen-Mannheim“ (Punktschätzung für d : 0.224) beträgt -0.008 (unterer Wert) bis 0.454 (oberer Wert). Das 90%-Konfidenzintervall für das Stichprobenpaar „Tübingen-Mix“ (Punktschätzung für d : -0.085) beträgt -0.313 (unterer Wert) bis 0.144 (oberer Wert). Das 90%-Konfidenzintervall für das Stichprobenpaar „Mannheim-Mix“ (Punktschätzung für d : -0.256) beträgt -0.487 (unterer Wert) bis -0.023 (oberer Wert). Alle drei 90%-Konfidenzintervalle liegen auch hier innerhalb der Grenzen von $d = \pm 0.5$. Damit ist auch bezüglich d für alle drei Stichprobenpaare untereinander Äquivalenz nachgewiesen.

3.4 Einflussfaktoren

Die Dice-Koeffizienten (abhängige Antwortvariablen) als Ausdruck der Güte der jeweiligen KI wurden anhand der Prädiktorvariablen BMI, PSA-Dichte, Prostatavolumen, PI-RADS-Score, Anzahl suspekter Läsionen im MRT, Diameter einer Indexläsion sowie einer möglichen extraprostatistischen bildmorphologischen Ausbreitung untersucht. Als methodische Grundlage diente hierfür die verallgemeinernde Schätzgleichung.

Zum Verständnis der Auswertung ist von Bedeutung, dass insgesamt 156 unterschiedliche Dice-Koeffizienten erhoben werden konnten, da jeder der drei auf Grundlage unterschiedlicher Trainingsdatensätze geübten KIs anschließend die 52 gleichen Bilddatensätze zur automatisierten Segmentation präsentiert wurden.

Die 156 Werte für die Dice-Koeffizienten wurden je nach KI, der sie präsentiert wurden, einer von drei Gruppen zugeordnet (Innersubjektvariable). Jeder Patient erhielt zur Auswertung außerdem eine Identifikationsnummer (Subjektvariable, 1 bis 52). Aufgrund eines fehlenden Wertes der Variablen „Prostatavolumen“ musste zur Analyse ein Patiententestdatensatz ausgeschlossen werden ($n = 3$, 1.9 %), die Auswertung bezieht sich folglich auf 51 Dice-Koeffizienten pro KI ($n = 153$, 98.1 %). Pro Subjekt wurden innerhalb der Korrelationsmatrix drei Messungen vorgenommen. Tabelle 19 zeigt die Ergebnisse der Analyse.

Tabelle 19. Ergebnisse der Verallgemeinernden Schätzgleichung

Variable	Regressions- koeffizient B	Standard- fehler	Wald-Chi- Quadrat	df	p
BMI	0.002	0.0015	2.300	1	0.129
Prostatavolumen	0.001	0.0006	4.567	1	0.033
PSA-Dichte	-0.008	0.0167	0.244	1	0.621
PI-RADS-Score	-0.003	0.0121	0.071	1	0.790
Extrapr. Ausbreitung	0.006	0.0149	0.163	1	0.686
Anzahl Läsionen	0.011	0.0099	1.207	1	0.272
Ø der Indexläsion	0.001	0.0091	0.022	1	0.881

Der Regressionskoeffizient gibt an, um wieviel sich die abhängige Variable verändert, wenn sich der entsprechende Parameter um eine Einheit erhöht. Im konkreten Beispiel des Parameters "BMI" würde ein Regressionskoeffizient von 0.002 bei signifikanter Bedeutung ausdrücken, dass sich pro Einheit „BMI“ der Dice-Koeffizient um 0.002 erhöht.

Unter Berücksichtigung des Standardfehlers ($se = 0.0015$), des 95%-Konfidenzintervalls (95% CI [-0.001, 0.005]) und der Signifikanz ($p = 0.129$) hat der BMI als erklärende Variable keinen signifikanten Einfluss auf die Dice-Koeffizienten. Gleiches gilt für die Prädiktoren PSA-Dichte (95% CI [-0.024, 0.244]), PI-RADS-Score (95% CI [-0.027, 0.021]), extraprostatiche Ausbreitung (95% CI -0.023, 0.035), Anzahl suspekter Läsionen (95% CI [-0.008, 0.030]) und Diameter der Indexläsion (95% CI [-0.016, .019]).

Das im MRT ausgemessene und errechnete Prostatavolumen ist ein signifikanter Prädiktor für den Dice-Koeffizienten (95% CI [0.000, 0.003]). Für jede Einheit, die das Prostatavolumen zunimmt, erhöht sich der Dice-Koeffizient um 0.001.

Um im Sinne einer Moderationsanalyse zu zeigen, dass sich die Aussagen zu den vorhersagenden Variablen als mögliche Einflussfaktoren auf die Dice-Koeffizienten unabhängig von der jeweiligen KI treffen lassen, wurde jede dieser Variable zusammen mit der Innersubjektvariable „KI-Gruppe 1, 2 und 3“ zu einem Interaktionsterm zusammengefasst und analysiert. Die Variable „Gruppe“ wurde so als mögliche Moderatorvariable für die jeweiligen Prädiktorvariablen untersucht. Tabelle 20 zeigt die entsprechenden Parameterschätzer.

Tabelle 20. Interaktionsanalyse

Interaktionsterm aus Innersubjektvariable „KI-Gruppe“ mit	Regressions- koeffizient B	Standard- fehler	Wald-Chi- Quadrat	df	p
- BMI	0.000	0.0004	0.599	1	0.439
- Prostatavolumen	-0.000091	0.000074	1.503	1	0.220
- PSA-Dichte	-0.006	0.0025	6.163	1	0.013*
- PI-RADS-Score	0.003	0.0031	0.769	1	0.381
- Extrapr. Ausbreitung	-0.002	0.0029	0.587	1	0.444
- Anzahl Läsionen	0.000037	0.0016	0.001	1	0.981
- Diameter der Indexläsion	0.002	0.0020	0.569	1	0.451

Die Ergebnisse konnten keinen Interaktionseffekt der KI-Zugehörigkeit (KI1, KI2, KI3) auf die Beziehung zwischen den Prädiktorvariablen BMI, Prostatavolumen, PI-RADS-Score, extraprostatiche Ausbreitung, Anzahl der Läsionen und Durchmesser der Indexläsion und den Dice-Koeffizienten finden ($p > 0.05$). Es konnte nachgewiesen werden, dass die Gruppenzugehörigkeit den Effekt zwischen PSA-Dichte und Dice-Koeffizienten signifikant moderiert ($p = 0.013$).

3.5 Visuelle Untersuchung der automatisierten Segmentierungen

Bei der qualitativen Untersuchung der Segmentierungen der drei KIs wurde jedes Schnittbild der T2-gewichteten MRT-Sequenzen der 52 Patienten des Testdatensatzes inspiziert und bezüglich ihrer Genauigkeit im Vergleich zu der manuellen Segmentierung anhand der vier in Kapitel 2.3.4 beschriebenen Kategorien bewertet. Dabei stellte sich heraus, dass neun der 52 MRT-Sequenzen leicht verschwommene Bilder im Sinne von Bewegungsartefakten aufwiesen. Bei einem Patienten konnte ein Katheter abgegrenzt werden; dieser wurde von der „KI Mannheim“ und der „KI Gemischt“ erkannt und umrandet, die „KI Tübingen“ segmentiert die Prostatagrenzen und lässt den Katheter unberücksichtigt. Die Mediane der Kategorien der jeweils einzelnen Prostatabereiche sind bei allen drei KIs identisch. Bei allen drei KIs schneidet die Segmentierung der Drüsenmitte im Median am besten ab, die Segmentierung der Prostatabasis außerdem schlechter als die Segmentierung der Apex.

Die allein von der Tübinger Stichprobe trainierte KI („KI Tübingen“) erzielte bei der qualitativen Untersuchung der automatisierten Segmentierungen des Testdatensatzes ($n = 52$) mittels des oben beschriebenen ordinalen Kategoriensystems von 1 bis 4 im Bereich der Prostatabasis im Median die Kategorie 3 ($SD = 0.693$), im Bereich der Drüsenmitte im Median die Kategorie 1 ($SD = 0.542$) und im Bereich des Apex im Median die Kategorie 2 ($SD = 0.534$). Im Bereich der Prostatabasis wurde in zwei Fällen die Kategorie 1 vergeben (3.8 %), in 21 Fällen die Kategorie 2 (40.4 %), in 25 Fällen die Kategorie 3 (48.1 %) und in vier Fällen die Kategorie 4 (7.7 %). In den Bereichen der Drüsenmitte sowie des Apex wurden lediglich die Kategorien 1 bis 3 vergeben. In der

Drüsenmitte waren das in 28 Fällen die Kategorie 1 (53.8 %), in 23 Fällen die Kategorie 2 (44.2 %) und in einem Fall die Kategorie 3 (1.9 %); in der Apex-Region in einem Fall die Kategorie 1 (1.9 %), in 29 Fällen die Kategorie 2 (55.8 %) und in 22 Fällen die Kategorie 3 (42.3 %).

Die allein von der Mannheimer Stichprobe trainierte KI („KI Mannheim“) erzielte im Bereich der Prostatabasis im Median die Kategorie 3 ($SD = 0.574$), im Bereich der Drüsenmitte im Median die Kategorie 1 ($SD = 0.577$) und im Bereich des Apex im Median die Kategorie 2 ($SD = 0.565$). Im Bereich der Prostatabasis wurde in einem Fall die Kategorie 1 vergeben (1.9 %), in 22 Fällen die Kategorie 2 (42.3 %), in 28 Fällen die Kategorie 3 (53.8 %) und in einem Fall die Kategorie 4 (1.9 %). Im Bereich der Prostatamitte wurde in 28 Fällen die Kategorie 1 vergeben (53.8 %), in 22 Fällen die Kategorie 2 (42.3 %) und in zwei Fällen die Kategorie 3 (3.8 %). Im Bereich des Apex wurde in zwei Fällen die Kategorie 1 vergeben (3.8 %), in 28 Fällen die Kategorie 2 (53.8 %) und in 22 Fällen die Kategorie 3 (42.3 %).

Die von sowohl von der Tübinger als auch der Mannheimer Stichprobe trainierte KI („KI Gemischt“) erzielte im Bereich der Prostatabasis im Median die Kategorie 3 ($SD = 0.753$), im Bereich der Drüsenmitte im Median die Kategorie 1 ($SD = 0.610$) und im Bereich des Apex im Median die Kategorie 2 ($SD = 0.561$). Im Bereich der Prostatabasis wurde in fünf Fällen die Kategorie 1 vergeben (9.6 %), in 17 Fällen die Kategorie 2 (32.7 %), in 27 Fällen die Kategorie 3 (51.9 %) und in drei Fällen die Kategorie 4 (5.8 %). Im Bereich der Drüsenmitte wurde in 28 Fällen die Kategorie 1 (53.8 %), in 21 Fällen die Kategorie 2 (40.4 %) und in drei Fällen die Kategorie 3 (5.8 %) vergeben. Im Bereich des Apex wurde in zwei Fällen die Kategorie 1 vergeben (3.8 %), in 29 Fällen die Kategorie 2 (55.8 %) und in 21 Fällen die Kategorie 3 (40.4 %).

4 Diskussion

4.1 Zusammenfassung der Ergebnisse

In den letzten zehn Jahren sind zahlreiche Studien zur KI-basierten Segmentierung der Prostata und ihrer Zonen auf T2-gewichteten MRT-Bildern erschienen (Bardis et al., 2020). Vorrangiges Ziel dieser Arbeit ist es, mit dem Wissen um mögliche Parameter, die die Ergebnisse einer automatisierten Segmentierung der Prostata beeinflussen, *deep learning*-Algorithmen weiter zu verbessern. Unserem Kenntnisstand nach wurden bislang noch keine Arbeiten veröffentlicht, die Einflussfaktoren auf die Performanz oder Güte einer KI zur automatisierten Erkennung und Segmentierung der Prostata, ihrer Zonen oder suspekter Läsionen untersucht. Im Folgenden sollen die Ergebnisse dieser Studie mit der einschlägigen Literatur verglichen werden sowie Vor- und Nachteile unseres Studiendesigns aufgezeigt werden. Abschließend werden aus dieser Arbeit ableitbare Forschungsgegenstände diskutiert.

4.1.1 Vergleich der Dice-Koeffizienten mit vergleichbaren Studien

Die mittleren Dice-Koeffizienten zur Segmentierung der gesamten Prostata lagen in unserer Studie für die drei auf unterschiedlichen Trainingsdaten beübten KIs bei 0.89, 0.88 und 0.89. Das ist vergleichbar mit den Ergebnissen anderer Studien der vergangenen 10 Jahren auf diesem Gebiet. Karimi et al. (2018) nennen in ihrer Publikation für die Segmentierung der gesamten Prostata einen mittleren Dice-Koeffizienten von 0.88, Zavala-Romero et al. (2020) 0.90 respektive 0.89 und damit ebenfalls Werte um die 0.90. Der Algorithmus in der Studie von Tian et al. (2018) war mit einem mittleren Dice-Koeffizienten von 0.85 schlechter als unserer, ebenso der von Chiali et al. (2016) mit einem Wert von 0.77 respektive 0.82. Die unserer Recherche nach besten mittleren Dice-Koeffizienten für das KI-basierte Segmentieren der Prostata in ihrer Gesamtheit, und auch die Dice-Koeffizienten unserer Studie übertreffenden, liefern Zabihollahy et al. (2019) mit einem mittleren Dice-Koeffizienten von 0.95, Bardis et al. (2021) mit einem mittleren Dice-Koeffizienten von 0.94 sowie Cheng et al. (2016) mit einem mittleren Dice-Koeffizienten von 0.93.

4.1.2 Verallgemeinernde Schätzgleichung zur Analyse der Einflussfaktoren

Wir konnten anhand einer verallgemeinernden Schätzgleichung (*generalized estimating equation*, GEE) zeigen, dass das MRT-diagnostisch gemessene Prostatavolumen einen signifikanten Einfluss auf die Ausprägung der Dice-Koeffizienten hat. Der BMI, die PSA-Dichte, eine extraprostatatische Ausbreitung, die Anzahl suspekter Läsionen, der PI-RADS-Score und der Durchmesser einer Indexläsion hingegen haben keinen signifikanten Einfluss auf die Höhe der jeweiligen Dice-Koeffizienten. Verallgemeinernde Schätzgleichungen liefern eine robuste, komplexe und gleichzeitig elegante Methodik zur simultanen Analyse möglicher Parameter auf eine abhängige Variable. Sie sind in anderen Wissenschaften wie der Epidemiologie, den Wirtschaftswissenschaften und der Biologie bereits deutlich etablierter als in der Medizin (Ghisletta & Spini, 2004). Im Rahmen der Literaturrecherche konnten keine anderen Arbeitsgruppen gefunden werden, die auf diese oder andere Weise mögliche Prädiktorvariablen im Bereich der Segmentierung anatomischer Strukturen analysierten.

4.1.3 Auswahl der Prädiktorvariablen und Interpretation der Regressionsparameter

Bei der Auswahl unserer Prädiktorvariablen sollten verschiedene Aspekte beleuchtet werden, die als mögliche Einflussfaktoren auf die Performanz einer automatisierten Segmentierung der Prostata auf MRT-Bildern denkbar sind. Zunächst ist die Konstitution des jeweiligen Patienten zu nennen. Der BMI gilt als Ausdruck des Ernährungszustandes. Die Vermutung lag nahe, dass Adipositas und damit ein erhöhter BMI zu niedrigeren Überschneidungen der manuellen mit der automatisierten Segmentierung führt, da ein Mehr an Fettgewebe und Körpermasse zu vermehrten Artefakten in der MRT-Bildgebung führen kann. Diesen Zusammenhang konnten wir nicht nachweisen. Da der BMI nichts über die Verteilung des Körperfettes und der Muskelmasse aussagt, wären Parameter wie Bauch- und Taillenumfang ein denkbarer alternativer Ansatz zur Untersuchung der körperlichen Konstitution auf die Performanz der KI.

Die PSA-Dichte ordnet den PSA-Wert der Größe der Prostata entsprechend ein, ist so interindividuell und als möglicher Indikator für Entzündungen, benigne oder maligne Prozesse in der Prostata zu interpretieren. Eine erhöhte PSA-Dichte könnte als möglicher Indikator für akute oder stattgefundene Veränderungen innerhalb des Prostatagewebes auch die Qualität von MRT-Bildern und damit den KI-Algorithmus negativ beeinflussen. Der von uns nicht signifikant nachgewiesene Effekt der PSA-Dichte auf die Dice-Koeffizienten spricht dafür, dass eine erhöhte PSA-Dichte keinen Einfluss auf die Organgrenzen der Prostata in der MRT-Bildgebung hat. Die KI-basierte Abgrenzung und Umrandung des Organs in seiner Gesamtheit scheinen durch eine höhere PSA-Dichte nicht negativ beeinflusst.

Auf ein mögliches Verschwimmen der Prostatakapsel und nicht mehr klar zu erkennende Organgrenzen zielten auch die Prädiktoren „extraprostatistische Ausbreitung“, „Anzahl suspekter Läsionen“ und „Diameter der Indexläsion“ ab. Die Vermutung lag nahe, dass ein Prostatakrebs, der bereits die Kapsel infiltriert oder gar über sie hinauswächst, die Leistung einer KI hinsichtlich des Erkennens und der Segmentierung der Prostata negativ beeinflusst. Auch diesen Zusammenhang konnten wir nicht nachweisen. Möglich wäre, dass die extraprostatistische Ausbreitung eines Prostatakarzinoms in unseren Trainingsdatensätzen die klare Abgrenzung der Prostata als Organ noch zuließ, dies aber bei einer größeren Stichprobe mit einer höheren Anzahl an Patienten oder andersartiger Ausprägung kapselüberschreitenden Wachstums nicht mehr der Fall ist. Die Anzahl suspekter Läsionen konnte ebenfalls nicht als signifikanter Einflussfaktor identifiziert werden. Denkbar wäre hier, dass sich anders als bei einem kapselüberschreitenden Prozess das Geschehen auf die Prostatagrenzen beschränkt, das heißt auf MRT-Bildern zwar keinen Einfluss auf die Umrandung des Organs in seiner Gesamtheit hat, vielleicht aber auf die Segmentierung der einzelnen innerhalb der Kapsel verorteten Prostatazonen oder Läsionen innerhalb des Prostatagewebes. Und auch der Durchmesser einer Indexläsion hat unseren Ergebnissen nach keinen signifikanten Einfluss auf die Performanz der drei KIs hinsichtlich der Segmentierung der gesamten Prostata.

Der PI-RADS-Score trifft als Klassifikationssystem der mpMRT der Prostata Aussagen über die Wahrscheinlichkeit der Malignität einer Läsion. Er berücksichtigt dabei nicht nur die anatomische T2-gewichtete Sequenz, sondern auch die diffusionsgewichtete und kontrastmittelunterstützte. Wir gingen von der Annahme aus, dass ein erhöhter PI-RADS-Score die Höhe der Dice-Koeffizienten negativ beeinflusst, da maligne Prozesse in der Bildgebung oft von unscharfer Berandung gekennzeichnet sind und zu einem unkontrollierten grenzüberschreitenden Wachstum neigen. Auch diesen Zusammenhang konnten wir nicht nachweisen. Ähnlich wie die Variable „extraprostatistische Ausbreitung“ und „Anzahl der Läsionen“ scheint ein erhöhter PI-RADS-Score das Umranden der Prostata in ihrer Gesamtheit nicht negativ zu beeinflussen, da sie in den MRT-Bildern noch klar abgrenzbar scheint. Interessant wäre auch hier eine Untersuchung des PI-RADS-Scores auf die automatisierte zonale Segmentierung der Prostata sowie auf das Detektieren und Umranden suspekter Läsionen.

Ein im Verhältnis kleineres Prostatavolumen, so unsere Vermutung bezüglich dieser Prädiktorvariable, führt zu einer schwierigeren Abgrenzung des Organs gegenüber benachbarten Strukturen. Wir konnten einen signifikanten Einfluss des im MRT gemessenen Prostatavolumens auf die Höhe der Dice-Koeffizienten nachweisen: Für jede Einheit, die das Prostatavolumen zunimmt, erhöht sich der Dice-Koeffizient um 0.001. Das zeigt, dass die KI große Prostaten besser abzugrenzen und präziser zu segmentieren vermag, wenn man für die Genauigkeit der Abgrenzung den Dice-Koeffizienten zu Grunde legt.

4.1.4 Visuelle Analyse der Segmentierungen

Zur qualitativen Untersuchung der automatisiert vorgenommenen Umrandungen durch die drei KIs wurden die Bildsequenzen der 52 Testdatensätze, einzeln und von kranial nach kaudal unterteilt in die Organabschnitte „Prostatabasis“, „Drüsenmitte“ und „Apex“, analysiert. Dabei zeigte sich vergleichbar mit den Studien von Jensen et al. (2019) und Karimi et al. (2018) eine bessere Übereinstimmung der automatisierten mit der manuellen Segmentierung im

Bereich der Drüsenmitte, weniger gute Übereinstimmungen in den Bereichen der Prostatapole, also der Prostatabasis und dem Apex. Bei Karimi et al. (2018) wird der Unterschied in den mittleren Dice-Koeffizienten der jeweiligen Prostataregion deutlich. Diese betragen für den Apex 0.80 und für die Basis 0.79, für die Drüsenmitte hingegen 0.91 (Karimi et al., 2018).

4.1.5 Bizentrisches Studiendesign

Da wir den KI-Algorithmus nicht nur durch einen einzigen Bilddatensatz beübt haben, sondern mit bizentrisch erhobenen Daten auf Grundlage dreier unterschiedlich trainierter KIs mögliche Einflussfaktoren auf deren Performanz analysieren konnten, wurde die Aussagekraft unserer Studie im Vergleich zu anderen Veröffentlichungen erhöht. Bardis et al. (2021) publizierten exemplarisch dafür eine der bis dato aktuellsten Arbeiten im Bereich der automatisierten Segmentierung der Prostata mit einem sehr hohen mittleren Dice-Koeffizient (0.94) für das Segmentieren der gesamten Prostata. Sie nutzten einen Bilddatensatz, dem lediglich Patienten eines Zentrums zugrunde liegen. Eine mögliche Problematik eines solchen unizentrischen Studienaufbaus besteht darin, dass sich vom MRT-Scanner über die Arbeitsabläufe beim Durchführen der mpMRT-Protokolle bis hin zur manuellen Segmentierung bei nur einer Klinik schlechter auf die Allgemeinheit respektive Kliniken im Allgemeinen schließen lässt. Prägnanter ausgedrückt: Der Vorteil eines multi- oder bizentrisch angelegten Studiendesigns, wie es in dieser Arbeit und der RESCT-Studie verwirklicht wird, besteht in der breiteren Erhebung der Patientendaten, deren Analyse einen universelleren Interpretationsspielraum zulassen. Um in einem nächsten Schritt zu zeigen, dass unsere drei unterschiedlich trainierten KIs auch stetig zueinander arbeiten, wurde zwischen den jeweiligen KI-Paaren ein Äquivalenztest auf Grundlage des 90%-Konfidenzintervalls sowohl der Dice-Koeffizienten als auch der entsprechenden Effektstärken nach Cohans durchgeführt. Rundo et al. (2020) nutzen zwar ebenfalls zwei unterschiedliche Datensätze zum Beüben ihrer KI, die dann die Prostatazonen segmentieren soll. Sie testen die Ergebnisse jedoch nicht auf Äquivalenz (Rundo et al., 2020). In

unserer Arbeit konnte für alle drei KI-Paare Äquivalenz für die absoluten Dice-Koeffizienten-Werte und für die Effektstärken nach Cohans d nachgewiesen werden. Daraus lässt sich ableiten, dass es keinen signifikanten Unterschied zwischen unseren drei unterschiedlich trainierten KIs gibt. Weiter kann man schlussfolgern, dass das der KI zugrunde liegende CNN unabhängig vom Trainingsdatensatz stetige Ergebnisse liefert. Um diese daraus abgeleiteten Aussagen noch robuster zu machen, wäre ein multizentrischer Ansatz, wie ihn Sanford et al. (2020) in ihrer Studie verwirklichen, denkbar. Sie postulieren durch die Variabilität der Trainingsdaten, der Augmentation und dem Transferlernen einer KI auf Grundlage der Datensätze mehrerer Einrichtungen das Wettmachen der patienten- und einrichtungsspezifischen Einflussfaktoren auf das Segmentieren der Prostata (Sanford et al., 2020).

4.1.6 Auswahl der Patienten, Stichprobengröße und Tumorstatus der Patienten

Die Auswahl der Patienten unserer Stichprobe verlief nicht in vollem Maße nach dem Zufallsprinzip. Eine chronologische Liste von prostatektomierten Patienten wurde konsekutiv abgearbeitet. Dabei wurde bei jedem einzelnen Patienten nach einer intern stattgefundenen mpMRT geschaut. Im Falle nicht verfügbarer Bilddatensätze wurden Patienten ausgeschlossen und übergangen. Einige der Patienten, deren Bilddatensätze fehlerbehaftet waren, wurden willkürlich ersetzt. Die Liste aller prostatektomierter Patienten hätte zunächst vollständig nach in domo durchgeführten und verfügbaren mpMRT-Bilddatensätzen untersucht werden können, um anschließend nach dem Zufallsprinzip 185 Patienten auszuwählen.

Auf der anderen Seite kann man argumentieren, dass sich der Zeitraum, in dem die Patienten schlussendlich ausgewählt wurden, nicht von anderen Zeiträumen unterscheidet und somit keine Einschränkungen aufgrund der so getroffenen Auswahl der Patienten nach Reihenfolge zu erwarten sind.

Die Anzahl von 185 retrospektiv erhobenen Patientendatensätze wurde mit dem Partner Mannheim im Rahmen der dieser Doktorarbeit übergeordneten

bizentrischen RESECT-Studie vereinbart. Sowohl von der Tübinger als auch der Mannheimer Stichprobe dienten je 100 Patienten-Bilddaten als Trainingsdatensatz; 52 weitere der Stichprobe Tübingen als Testdatensatz. Die im Rahmen dieser Arbeit verwendete Anzahl von in Summe 252 Patienten ist vergleichbar mit der anderer Arbeiten und übersteigt diese sogar gegenüber vielen Veröffentlichungen. So kamen Bardis et al. (2021) in ihrer Studie auf insgesamt 242 Patienten, Zavala-Romero et al. (2020) auf 220 Patienten, Zabihollahy et al. (2019) auf 225. Zhu et al. (2019) arbeiteten mit insgesamt 163 Patienten, also 89 weniger als in der vorliegenden Arbeit. Cheng et al. (2016) kamen auf lediglich 120 Patienten, 132 weniger als in unserer Studie. Die Stichprobe von Karimi et al. (2016) umfasste 75 Patienten, die der Studie von Tian et al. (2018) 140 Patienten, wobei nur 41 davon eigens erhoben, 99 aus einer öffentlich zugänglichen Quelle stammten. Rundo et al. (2020) zählten in ihrer Stichprobe insgesamt 40 Patienten, das sind 112 Patienten weniger als in unserer Arbeit. Der Studie von Chilali et al. (2016) liegen 30 ebenfalls öffentlich zugängliche Bilddatensätze zugrunde, also 122 weniger als in der hier vorliegenden. Nur Sanford et al. (2020) hatten unserer Recherche nach im Bereich automatisierter Prostatasegmentierung auf T2-gewichteten MRT-Bildern in den vergangenen 10 Jahren in ihrer Studie mit 648 Patienten insgesamt 396 mehr als unsere Arbeit. Die RESECT-Studie umfasst neben den insgesamt 185 Patienten der Stichprobe Tübingen insgesamt 255 Patienten der Stichprobe Mannheim. Denkbar wäre in einer Folgearbeit das Analysieren auf Grundlage dieser zusammen 440 Patienten.

Sowohl Zhu et al. (2019) als auch Sanford et al. (2020), Jensen et al. (2019) und Zabihollahy et al. (2019) hatten anders als wir in ihrer Stichprobe sowohl Patienten mit diagnostiziertem Prostatakarzinom als auch prostatagesunde Patienten. Einschlusskriterien der Stichprobe Tübingen für die RESECT-Studie und die vorliegende Arbeit waren prostatektomierte Patienten mit zuvor stattgefundenener mpMRT, bei allen wurde im klinischen Verlauf ein Prostatakarzinom operativ entfernt worden. Um allgemeingültigere Ergebnisse zu erheben, wäre es denkbar, den Trainings-, Validierungs- und Testdatensatz um prostatagesunde Patienten zu erweitern.

4.2 Limitationen

4.2.1 Durchführung der manuellen Segmentierung

Von großer Wichtigkeit bezüglich des Beübens eines automatisierten Segmentationsalgorithmus ist eine präzise vorgenommene manuelle Umschreibung der Prostatagrenzen in den Trainingsbildern - sie dient der KI als Grundwahrheit (Bardis et al., 2021). Anders als in der vorliegenden Studie, in der die manuellen Umrandungen der Prostatae vom Doktoranden selbst ohne Berufserfahrung vorgenommen wurden, weisen die überwiegende Mehrheit der Autorinnen und Autoren vergleichbarer Arbeiten die ausgesprochene und langjährige Expertise ihrer an den Segmentierungen beteiligten Anwenderinnen und Anwender aus. Bardis et al. (2021) ließen die Segmentationen in ihrer Studie von 10 Radiologinnen und Radiologen vornehmen, diese wurden abschließend von einem Radiologen mit 10jähriger Erfahrung in der Befundung von mpMRT-Bildern überprüft. Auch sie hinterfragten das Annotieren der Grundwahrheit kritisch (Bardis et al., 2021). Zavala-Romero et al. (2020) ließen die manuelle Segmentierung ihrer Bilddaten durch Radiologinnen und Radiologen mit zusammen mehr als 25 Jahren Erfahrung erfolgen, deren anschließende Überprüfung geschah durch einen „erfahrenen“ Radioonkologen an. Die Umzeichnung der Prostata nahmen in der Studie von Zabihollay et al. (2019) vier Radiologen und Radiologinnen mit fünf bis 13 Jahren Erfahrung vor. Sie sprachen sich im Vorfeld der eigentlichen Segmentation bezüglich der Vorgehensweise anhand zweier Test-Bilddatensätze ab (Zabihollahy et al., 2019). Tian et al. (2018) machen in ihrer Veröffentlichung keine genaueren Angaben über die Person oder Personen, die die manuellen Segmentationen der Prostata vorgenommen haben; sie sprechen bei der Bezeichnung einer Umrandungslinie in einer Bildunterschrift lediglich von der eines „erfahrenen Radiologen“ (Tian et al., 2018). Bei Chilali et al. (2016) wurde durch einen Radiologen bzw. eine Radiologin mit mehr als 15 Jahren Erfahrung segmentiert, ein weiterer Trainingsdatensatz aus einem öffentlich zugänglichen Prostata-Atlas, dessen manuellen Segmentierungen genau wie bei Karimi et al. (2016) von einem erfahrenen Radiologen bzw. einer erfahrenen Radiologin mit Expertise im Bereich der Diagnosestellung eines Prostatakarzinoms

vorgenommen wurden (Betrouni et al., 2012). Vergleichbar mit dieser Arbeit wurden in der Studie von Cheng et al. (2016) die Prostatae ebenfalls von lediglich geschulten Untersucherinnen und Untersuchern umrandet, anschließend begutachtet von einem erfahrenen Radiologen bzw. einer erfahrenen Radiologin.

Montagne et al. (2021) untersuchten in ihrer Studie diese Grundwahrheit und ihre Qualität in Abhängigkeit von der Geübtheit des oder der die Prostatagrenzen Einzeichnenden. Sie zeigten, dass es insgesamt keinen signifikanten Unterschied zwischen Anwenderinnen und Anwendern unterschiedlicher Erfahrungsstufen gibt, wenn es um die genaue Segmentation der Prostata geht. Sie verglichen die Segmentationen des Organs als Ganzes sowie der Transitionszone auf T2-gewichteten axialen MRT-Bildern ($N = 40$) von sieben Radiologinnen und Radiologen unterteilt in drei Gruppen: mehr als 1000 befundete Prostata-MRT-Untersuchungen, mindestens 500 Begutachtungen und weniger als 100 (Montagne et al., 2021). Die Studie zeigte, dass es am ehesten zu Unterschiedlichkeiten zwischen den Befundungen des jeweiligen Arztes oder der jeweiligen Ärztin an äußeren Abschnitten (Apex und Basis) der Prostata kommt. Interessant wäre ein Vergleich mit jemandem, der keinerlei Erfahrung auf dem Gebiet der radiologischen Befundung von MRT-Bildern speziell im Bereich des kleinen Beckens und der Prostata hat. Die Segmentierungen der Prostatae dieser Arbeit wurden wie bereits erwähnt vom Doktoranden nach abgeschlossenem Medizinstudium und ohne Berufserfahrung vorgenommen. Der Doktorand wurde von seiner Betreuerin, einer Assistenzärztin im dritten Weiterbildungsjahr, in das Programm zur Segmentierung eingeführt und in der Segmentierung geschult. Die Betreuerin supervidierte und überprüfte die Umrandungen. Man hätte an diese Stelle noch weitere Untersucherinnen und Untersucher mit mehr Erfahrung setzen können, was die Präzision der primär vorgenommenen Segmentierungen nach Montagne et al. (2021) jedoch nur fraglich beeinflusst hätte. (Berufs-)Erfahrung scheint vorwiegend für das Segmentieren der Prostatazonen und noch mehr für das Detektieren und Umranden suspekter Läsionen von Bedeutung zu sein; für das Segmentieren der Prostata in ihrer Gesamtheit hingegen weniger (Montagne et al., 2021). Als optimale Lösung denkbar wäre ein konsensueller Entscheid mehrerer geschulter

und erfahrener Radiologinnen und Radiologen bzw. Anwenderinnen und Anwendern im Allgemeinen über die zum Trainieren einer KI verwendeten Segmentierungen als Grundwahrheit.

4.2.2 Dice-Koeffizient als Maß der Performanz einer KI zur Bildsegmentierung

Als Ausdruck der Leistungsfähigkeit der von Fraunhofer MEVIS entwickelten und in dieser Studie genutzten KI wurde der in nahezu allen von uns recherchierten Studien genutzte Sørensen-Dice Koeffizient herangezogen (Bardis et al., 2021; Cheng et al., 2016; Chilali et al., 2016; Karimi et al., 2018; Sanford et al., 2020; Zabihollahy et al., 2019; Zavala-Romero et al., 2020; Zhu et al., 2019). Einzig Allen et al. (2006) geben in den Ergebnissen ihrer sehr frühen Publikation auf dem Gebiet der automatisierten Prostatasegmentierung keinen Sørensen-Dice Koeffizienten an. Sie nehmen stattdessen als Maß der Übereinstimmung zwischen der manuellen und der KI-basierten Umrandung eine absolute Differenz in Millimetern sowie eine Volumendifferenz in Prozent (Allen et al., 2006). Das lässt ihre Ergebnisse schwerer in die gängige Literatur einbetten, außerdem fließen Bildpunkte, die gänzlich außerhalb der Prostata liegen, nicht negativ mit in die Bewertung ein. Die Dice-Koeffizienten als Ausdruck der Güte der KI und als Antwortvariablen zur Analyse in dieser Studie ließe sich Ersetzen durch andere Indizes im Bereich des Vergleichens von Bildern wie dem Jaccard-Index. Er könnte außerdem um beispielsweise die Hausdorff-Metrik, wie sie Karimi et al. (2018) in ihrer Studie anführen, erweitert werden.

4.2.3 Berechnung der Dice-Koeffizienten nach Prostataabschnitt

Zhu et al. berechneten den Dice-Koeffizienten im Gegensatz zu dieser Arbeit nicht nur über die gesamte Prostata hinweg, sondern aufgeteilt in die Abschnitte „Prostatabasis“, „mittlerer Drüsenabschnitt“ und „Apex“. Dadurch lässt sich zeigen, in welcher Organregion ein KI-Algorithmus hinsichtlich der Organabgrenzung besonders gut oder aber weniger gut arbeitet (Zhu et al., 2019). Auch Karimi et al. (2016) gehen bei der Berechnung der Dice-

Koeffizienten nach dieser Gliederung der Prostata vor. Sie zeigen einen Unterschied der Überschneidungen innerhalb der drei Organregionen (Dice-Koeffizient der Prostatamitte: 0.91, der Basis: 0.79, der Apex: 0.80) (Karimi et al., 2018). Um diesem Umstand gerecht zu werden, wurde die Leistung der KI in dieser Arbeit bei jedem Patienten des Testdatensatzes Bild für Bild und gegliedert nach den oben genannten Organregionen qualitativ untersucht und jeder Organbereich für jede der drei KIs in eine von 4 Kategorien eingeteilt.

4.2.4 Redundanz zweier Bilddatensätze

Ein weiterer Kritikpunkt an dieser Arbeit liegt in der Tatsache, dass eine der drei KIs (KI 3, „Gemischt“) Bilddatensätze aus den beiden anderen KI-Trainingsdatensätzen enthält, nämlich die ersten 50 aus je der Stichprobe Mannheim sowie Tübingen. Dadurch wird die Aussagekraft dieser KI im direkten Vergleich mit den jeweils anderen geschwächt (sie überschneiden sich zu je 50 Bilddaten). Rundo et al. (2019) nutzten wie wir als einen von insgesamt drei Trainingsdatensätzen ebenfalls einen gemischten Datensatz. Die Idee unserer Arbeit dabei war es zu überprüfen, ob es bei einem bizenrischen Studiendesign Unterschiede zwischen den Bilddaten der beiden Standorte gibt und ob eine aus beiden Standorten gespeiste KI signifikant bessere oder schlechtere Ergebnisse liefert. Dies konnte mit Hilfe des Äquivalenztests widerlegt werden. Vom Studienaufbau aussagekräftiger wäre jedoch eine „gemischte KI“ mit Bilddaten, die disjunkt zu den Bilddaten der anderen beiden KIs sind, welche jeweils nur Bilddaten eines Standortes verwenden. Dies wird in der RESECT-Studie verwirklicht. Interessant wäre weiterhin, ob es Unterschiede zwischen den Bilddaten der Standorte als Trainingsdatensätze in der Performanz beim Erkennen und Segmentieren der Prostatazonen sowie suspekter Läsionen gibt, später dann auch hinsichtlich der der RESECT-Studie zugrunde liegenden Fragestellungen zur Respektabilität und Möglichkeit einer nervenerhaltenden Operation.

4.3 Künftige und an diese Arbeit anknüpfende Forschung

4.3.1 Untersuchen weiterer Prädiktorvariablen

Es lassen sich noch weitere klinische Parameter hinsichtlich deren Einflusses auf die Dice-Koeffizienten untersuchen. Denkbar wären der Gleason-Score als Ausdruck der zellulären Entdifferenzierung, das Patientenalter und damit ein anderer Stoffwechsel sowie Verteilung des Muskel- und Fettgewebes, der histologische Subtyp eines Prostatakarzinoms als Ausdruck unterschiedlicher Zellverbände, systemische Begleiterkrankungen, urologische oder abdominelle Operationen in der Vergangenheit als mögliche Urheber für Artefakte im MRT sowie der Zeitpunkt der mpMRT im Verlauf einer Tumorerkrankung als Maß unterschiedlicher Tumorstadien.

4.3.2 Evaluierung der manuellen Segmentierung als Grundwahrheit

Auch in der Studie von Montagne et al. (2021) besteht - wenn auch kein substanzieller aber dennoch - ein Unterschied zwischen der Segmentierung unterschiedlicher Anwenderinnen und Anwendern egal welcher Erfahrungsstufe. Die präzise Umrandung der Prostata - der Grundwahrheit - ist wie bereits ausgeführt ein entscheidender Schritt in der KI-basierten Diagnostik, Beurteilung und Therapieentscheidung sowie -ausführung. Ein geeignetes Studiendesign zur Evaluierung dieser Grundwahrheit könnte so aussehen, dass neben der Segmentierung von geübter Hand ein weiterer klinisch erfahrener Radiologe respektive eine klinisch erfahrene Radiologin die Ergebnisse sowohl der KI als auch der manuellen Prostataumrandung beurteilt und diese quasi dreifach justierten Prostataumrandungen als Trainingsdatenset für eine weitere KI dienen. Denkbar wäre auch eine Markierung der Prostatakapsel und -grenzen in Körperspendern eines anatomischen Instituts und das Untersuchen und Auswerten dieser so markierten Bereiche im MRT. Die Grundwahrheit dieser Bildsequenzen entspräche dann der tatsächlichen anatomischen Organbegrenzung. Einen ähnlichen Ansatz gehen Betrouni et al. (2012). Ihrer Studie nach könnte auch das Versehen eines Prostata-Phantoms mit einem

MRT-diagnostisch zu sehenden Marker als Grundwahrheit dienen (Betrouni et al., 2012).

4.3.3 Segmentieren der Prostatazonen und suspekter Läsionen unter Berücksichtigung aller mpMRT-Sequenzen

Das automatisierte Segmentieren der Prostata als Organ in seiner Gesamtheit funktioniert bereits verglichen mit dem manuellen Segmentieren sehr gut, Dice-Koeffizienten erreichen Werte von knapp über 0.95 und damit eine fast vollständige Überlappung (Dice = 1). Die Ergebnisse für das Segmentieren der peripheren und Transitionszone hingegen bleiben noch deutlich darunter (Rundo et al., 2020), noch schlechter sieht es bei der Detektion und Umrandung suspekter Läsionen respektive eines Prostatakarzinoms aus. Ziel künftiger Forschung muss es also sein, die Algorithmen besonders in diesen Bereichen weiter zu verbessern, da in der KI-basierten Analyse neben der Zeitersparnis durch die automatisierte Segmentierung der Prostata und ihrer Zonen ein unter Umständen therapieentscheidendes und lebensverlängerndes Diagnostikum für Patienten liegen könnte. Die in dieser Arbeit untersuchten Einflussfaktoren könnten nach gleicher Methodik auch auf Segmentierungen der Prostatazonen und von der KI detektierter und eingezeichneter Läsionen der Prostata analysiert werden. Grundlage dieser Arbeit war die Segmentierung auf T2-gewichteten sogenannten anatomischen Bildsequenzen. Clark et al. (2017) untersuchten in ihrer Publikation die Segmentierungen einer KI auf DWI-Sequenzen. Einflussfaktoren ließen sich entsprechend auch auf diesen Sequenzen untersuchen.

5 Zusammenfassung

Viele KI-Algorithmen aus dem Bereich des *deep learning* konnten in den letzten Jahren zeigen, dass die automatisierte Segmentierung der Prostata auf MRT-Datensätzen in ihrer Gesamtheit bereits sehr gute und für den klinischen Alltag akzeptable Ergebnisse liefern kann (Bardis et al., 2021). Von großem Interesse ist das Forschungsfeld der strukturierten KI-basierten Befundung der mpMRT der Prostata hinsichtlich dreier Aspekte: der großen Zeitersparnis gegenüber einer manuellen Bearbeitung der Datensätze, einer über Anwenderinnen und Anwendern hinweg vergleichbaren Befundung und der diagnostischen Präzision, die eine KI bei zunehmenden Bilddatenmengen leisten kann.

Die vorliegende Arbeit ist aus einem Projekt heraus entstanden, das eine KI-basierte Entscheidungshilfe zur Behandlung eines Prostatakarzinoms entwickelt. Eine Anwendung soll dabei prätherapeutisch valide Aussagen zur Wahrscheinlichkeit eines Nervenerhalts bei geplanter radikaler Prostatektomie sowie den postoperativen Resektionsstatus treffen. In einer retrospektiven Erhebung wurden dazu bizenitrisch an den beiden Standorten Tübingen und Mannheim 185 und 255 Patientendaten, in Summe also 340, erhoben. Dazu zählen neben den Bildsequenzen der mpMRT der Prostata auch zahlreiche klinische und demographische Parameter.

Vorrangiger Forschungsgegenstand dieser Studie ist es, mögliche Einflussfaktoren auf die Performanz der KI zu untersuchen. Dazu wurde der KI-Algorithmus anhand dreier Trainingsdatensätze zu je 100 Patienten geübt und auf einen Testdatensatz mit 52 Patienten angewandt. Die automatisierten Segmentierungen wurden mit den vom Doktoranden vorgenommenen manuellen Segmentierungen anhand des Sørensen-Dice Koeffizienten verglichen. Die berechneten Sørensen-Dice Koeffizienten dienten als Ausdruck der Güte der KI. Sie wurden als Antwortvariable mittels Verallgemeinernder Schätzgleichung anhand der Prädiktorvariablen Body-Mass-Index, PSA-Dichte, Prostatavolumen im MRT, extraprostatasche Ausbreitung, Anzahl suspekter Läsionen, Diameter einer Indexläsion und dem PI-RADS-Score untersucht. Dabei konnte das Prostatavolumen als signifikanter Einflussfaktor identifiziert werden. Die auf Grundlage unterschiedlicher Trainingsdatensätze geübten KIs wurden mit Hilfe

eines Äquivalenztests auf Unterschiedlichkeiten hinsichtlich ihrer Segmentierungen analysiert. Es stellte sich heraus, dass sich die drei KIs im Rahmen der zuvor definierten Grenzen nicht unterscheiden. Zur qualitativen Analyse wurden die KI-basierten Segmentierungen außerdem anhand von vier Kategorien klassifiziert.

Uns ist keine Veröffentlichung bekannt, die Einflussfaktoren auf die Performanz automatisierter Segmentierungsalgorithmen untersucht. Künftige Studien könnten auf dieser Arbeit aufbauend weitere Einflussfaktoren analysieren. Denkbar und von sicherlich großem Interesse wäre außerdem eine Untersuchung unserer sowie anderer Einflussfaktoren auf die zonale Segmentierung der Prostata sowie der Umrandung suspekter Läsionen oder konkret eines Prostatakarzinoms. In diesem Bereich erzielen die KI-basierten Algorithmen noch keinen so durchschlagenden Erfolg wie es die Segmentierungen der Prostatagrenzen schaffen.

6 Literaturverzeichnis

- Ahmed, H. U., El-Shater Bosaily, A., Brown, L. C., Gabe, R., Kaplan, R., Parmar, M. K., Collaco-Moraes, Y., Ward, K., Hindley, R. G., Freeman, A., Kirkham, A. P., Oldroyd, R., Parker, C., & Emberton, M. (2017). Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): A paired validating confirmatory study. *Lancet*, 389(10071), 815-822. [https://doi.org/10.1016/s0140-6736\(16\)32401-1](https://doi.org/10.1016/s0140-6736(16)32401-1)
- Allen, P. D., Graham, J., Williamson, D. C., & Hutchinson, C. E. (2006). Differential segmentation of the prostate in MR images using combined 3D shape modelling and voxel classification. 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.,
- Aumüller, G.-V., Aust, G.-V., Engele, J.-V., Kirsch, J.-V., Maio, G.-V., Mayerhofer, A.-V., Mense, S.-V., & Reißig, D.-V. (2020). *Duale Reihe Anatomie* (5., korrigierte Auflage ed.). Georg Thieme Verlag KG.
- Ayala, A. G., Ro, J. Y., Babaian, R., Troncoso, P., & Grignon, D. J. (1989). The prostatic capsule: does it exist? Its importance in the staging and treatment of prostatic carcinoma. *Am J Surg Pathol*, 13(1), 21-27.
- Bardis, M., Houshyar, R., Chantaduly, C., Tran-Harding, K., Ushinsky, A., Chahine, C., Rupasinghe, M., Chow, D., & Chang, P. (2021). Segmentation of the Prostate Transition Zone and Peripheral Zone on MR Images with Deep Learning. *Radiol Imaging Cancer*, 3(3), e200024. <https://doi.org/10.1148/rycan.2021200024>
- Bardis, M. D., Houshyar, R., Chang, P. D., Ushinsky, A., Glavis-Bloom, J., Chahine, C., Bui, T.-L., Rupasinghe, M., Filippi, C. G., & Chow, D. S. (2020). Applications of Artificial Intelligence to Prostate Multiparametric MRI (mpMRI): Current and Emerging Trends. *Cancers*, 12(5), 1204. <https://doi.org/10.3390/cancers12051204>
- Barentsz, J. O., Richenberg, J., Clements, R., Choyke, P., Verma, S., Villeirs, G., Rouviere, O., Logager, V., & Fütterer, J. J. (2012). ESUR prostate MR guidelines 2012. *Eur Radiol*, 22(4), 746-757. <https://doi.org/10.1007/s00330-011-2377-y>
- Becerra, M. F., Alameddine, M., Zucker, I., Tamariz, L., Palacio, A., Nemeth, Z., Velasquez, M. C., Savio, L. F., Panizzutti, M., Jue, J. S., Soodana-Prakash, N., Ritch, C. R., Gonzalgo, M. L., Parekh, D. J., & Punnen, S. (2020). Performance of Multiparametric MRI of the Prostate in Biopsy Naïve Men: A Meta-analysis of Prospective Studies. *Urology*, 146, 189-195. <https://doi.org/https://doi.org/10.1016/j.urology.2020.06.102>
- Berges, R., & Oelke, M. (2011). Age-stratified normal values for prostate volume, PSA, maximum urinary flow rate, IPSS, and other LUTS/BPH indicators in the German male community-dwelling population aged 50 years or older. *World J Urol*, 29(2), 171-178. <https://doi.org/10.1007/s00345-010-0638-z>
- Betrouni, N., Iancu, A., Puech, P., Mordon, S., & Makni, N. (2012). ProstAtlas: a digital morphologic atlas of the prostate. *Eur J Radiol*, 81(9), 1969-1975. <https://doi.org/10.1016/j.ejrad.2011.05.001>
- Beyer, T., Schlemmer, H. P., Weber, M. A., & Thierfelder, K. M. (2021). PI-RADS 2.1 - Image Interpretation: The Most Important Updates and Their

- Clinical Implications. *Rofo*, 193(7), 787-796. <https://doi.org/10.1055/a-1324-4010> (PI-RADS 2.1 – Befundinterpretation: Die wichtigsten Neuerungen und ihre klinischen Implikationen.)
- Bhavsar, A., & Verma, S. (2014). Anatomic imaging of the prostate. *Biomed Res Int*, 2014, 728539. <https://doi.org/10.1155/2014/728539>
- Caglic, I., Kovac, V., & Barrett, T. (2019). Multiparametric MRI - local staging of prostate cancer and beyond. *Radiol Oncol*, 53(2), 159-170. <https://doi.org/10.2478/raon-2019-0021>
- Carass, A., Roy, S., Gherman, A., Reinhold, J. C., Jesson, A., Arbel, T., Maier, O., Handels, H., Ghafoorian, M., Platel, B., Birenbaum, A., Greenspan, H., Pham, D. L., Crainiceanu, C. M., Calabresi, P. A., Prince, J. L., Roncal, W. R. G., Shinohara, R. T., & Oguz, I. (2020). Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis. *Scientific Reports*, 10(1), 8242. <https://doi.org/10.1038/s41598-020-64803-w>
- Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N. C., & Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning. *Phys Med*, 83, 9-24. <https://doi.org/10.1016/j.ejomp.2021.02.006>
- Cheng, R., Roth, H. R., Lu, L., Wang, S., Turkbey, B., Gandler, W., McCreedy, E. S., Agarwal, H. K., Choyke, P., & Summers, R. M. (2016). Active appearance model and deep learning for more accurate prostate segmentation on MRI. *Medical imaging 2016: Image processing*, 9781456119517, 1-7. <https://doi.org/10.1117/11.1016/j.crad.2019.03.016>
- Chiam, K., Carle, C., Hughes, S., Kench, J. G., Woo, H. H., Lord, S., & Smith, D. P. (2021). Use of multiparametric magnetic resonance imaging (mpMRI) in active surveillance for low-risk prostate cancer: a scoping review on the benefits and harm of mpMRI in different biopsy scenarios. *Prostate Cancer Prostatic Dis*, 24(3), 662-673. <https://doi.org/10.1038/s41391-021-00320-9>
- Chilali, O., Puech, P., Lakroum, S., Diaf, M., Mordon, S., & Betrouni, N. (2016). Gland and Zonal Segmentation of Prostate on T2W MR Images. *J Digit Imaging*, 29(6), 730-736. <https://doi.org/10.1007/s10278-016-9890-0>
- Clark, T., Zhang, J., Baig, S., Wong, A., Haider, M. A., & Khalvati, F. (2017). Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks. *J Med Imaging (Bellingham)*, 4(4), 041307. <https://doi.org/10.1117/1.Jmi.4.4.041307>
- Cutaia, G., La Tona, G., Comelli, A., Vernuccio, F., Agnello, F., Gagliardo, C., Salvaggio, L., Quartuccio, N., Sturiale, L., Stefano, A., Calamia, M., Arnone, G., Midiri, M., & Salvaggio, G. (2021). Radiomics and Prostate MRI: Current Role and Future Applications. *J Imaging*, 7(2). <https://doi.org/10.3390/jimaging7020034>
- Das, C. J., Razik, A., Sharma, S., & Verma, S. (2019). Prostate biopsy: when and how to perform. *Clin Radiol*, 74(11), 853-864. <https://doi.org/10.1016/j.crad.2019.03.016>
- Deutsche Krebsgesellschaft, D. K., AWMF. (2021). *S3-Leitlinie Prostatakarzinom, Langversion 6.2*. <http://www.leitlinienprogramm-onkologie.de/leitlinien/prostatakarzinom/>

- Franiel, T., Asbach, P., Beyersdorff, D., Blondin, D., Kaufmann, S., Mueller-Lisse, U. G., Quentin, M., Rödel, S., Röthke, M., Schlemmer, H. P., & Schimmöller, L. (2021). mpMRI of the Prostate (MR-Prostatographie): Updated Recommendations of the DRG and BDR on Patient Preparation and Scanning Protocol. *Rofo*, 193(7), 763-777. <https://doi.org/10.1055/a-1406-8477> (mpMRT der Prostata (MR-Prostatografie): Aktualisierte Empfehlungen der DRG und des BDR zur Vorbereitung und Durchführung.)
- Ghisletta, P., & Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics*, 29(4), 421-437.
- Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: Images Are More than Pictures, They Are Data. *Radiology*, 278(2), 563-577. <https://doi.org/10.1148/radiol.2015151169>
- Hotker, A., & Donati, O. F. (2021). [PI-RADS 2.1 and structured reporting of magnetic resonance imaging of the prostate]. *Radiologe*. <https://doi.org/10.1007/s00117-021-00868-6> (PI-RADS 2.1 und strukturierte Befundung der Magnetresonanztomographie der Prostata.)
- Jensen, C., Sørensen, K. S., Jørgensen, C. K., Nielsen, C. W., Høy, P. C., Langkilde, N. C., & Østergaard, L. R. (2019). Prostate zonal segmentation in 1.5 T and 3T T2W MRI using a convolutional neural network. *Journal of Medical Imaging*, 6(1), 014501.
- Karimi, D., Samei, G., Kesch, C., Nir, G., & Salcudean, S. E. (2018). Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models. *Int J Comput Assist Radiol Surg*, 13(8), 1211-1219. <https://doi.org/10.1007/s11548-018-1785-8>
- Kasivisvanathan, V., Rannikko, A. S., Borghi, M., Panebianco, V., Mynderse, L. A., Vaarala, M. H., Briganti, A., Budäus, L., Hellawell, G., Hindley, R. G., Roobol, M. J., Eggener, S., Ghei, M., Villers, A., Bladou, F., Villeirs, G. M., Viridi, J., Boxler, S., Robert, G., . . . Moore, C. M. (2018). MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med*, 378(19), 1767-1777. <https://doi.org/10.1056/NEJMoa1801993>
- Klotz, L., Pond, G., Loblaw, A., Sugar, L., Moussa, M., Berman, D., Van der Kwast, T., Vesprini, D., Milot, L., Kebabdjian, M., Fleshner, N., Ghai, S., Chin, J., & Haider, M. (2020). Randomized Study of Systematic Biopsy Versus Magnetic Resonance Imaging and Targeted and Systematic Biopsy in Men on Active Surveillance (ASIST): 2-year Postbiopsy Follow-up. *Eur Urol*, 77(3), 311-317. <https://doi.org/10.1016/j.eururo.2019.10.007>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259-269.

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4), 541-551. <https://doi.org/10.1162/neco.1989.1.4.541>
- Lepor, H., Gregerman, M., Crosby, R., Mostofi, F. K., & Walsh, P. C. (1985). Precise localization of the autonomic nerves from the pelvic plexus to the corpora cavernosa: a detailed anatomical study of the adult male pelvis. *J Urol*, 133(2), 207-212. [https://doi.org/10.1016/s0022-5347\(17\)48885-9](https://doi.org/10.1016/s0022-5347(17)48885-9)
- Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13-22. <https://doi.org/10.1093/biomet/73.1.13>
- Manfredi, M., Mele, F., Garrou, D., Walz, J., Fütterer, J. J., Russo, F., Vassallo, L., Villers, A., Emberton, M., & Valerio, M. (2018). Multiparametric prostate MRI: technical conduct, standardized report and clinical use. *Minerva Urol Nefrol*, 70(1), 9-21. <https://doi.org/10.23736/s0393-2249.17.02846-6>
- Martins, T., Mussi, T. C., & Baroni, R. H. (2020). Prostate volume measurement by multiparametric magnetic resonance and transrectal ultrasound: comparison with surgical specimen weight. *Einstein (Sao Paulo)*, 18, eAO4662. https://doi.org/10.31744/einstein_journal/2020AO4662
- McBee, M. P., Awan, O. A., Colucci, A. T., Ghobadi, C. W., Kadom, N., Kansagra, A. P., Tridandapani, S., & Auffermann, W. F. (2018). Deep Learning in Radiology. *Acad Radiol*, 25(11), 1472-1480. <https://doi.org/10.1016/j.acra.2018.02.018>
- McLaughlin, P. W., Troyer, S., Berri, S., Narayana, V., Meirowitz, A., Roberson, P. L., & Montie, J. (2005). Functional anatomy of the prostate: implications for treatment planning. *Int J Radiat Oncol Biol Phys*, 63(2), 479-491. <https://doi.org/10.1016/j.ijrobp.2005.02.036>
- McNeal, J. E. (1981). The zonal anatomy of the prostate. *Prostate*, 2(1), 35-49. <https://doi.org/10.1002/pros.2990020105>
- McNeal, J. E. (1988). Normal Histology of the Prostate. *The American Journal of Surgical Pathology*, 12(8). https://journals.lww.com/ajsp/Fulltext/1988/08000/Normal_Histology_of_the_Prostate.3.aspx
- Montagne, S., Hamzaoui, D., Allera, A., Ezziane, M., Luzurier, A., Quint, R., Kalai, M., Ayache, N., Delingette, H., & Renard-Penna, R. (2021). Challenge of prostate MRI segmentation on T2-weighted images: inter-observer variability and impact of prostate morphology. *Insights Imaging*, 12(1), 71. <https://doi.org/10.1186/s13244-021-01010-9>
- Radtke, J. P., Kuru, T. H., Boxler, S., Alt, C. D., Popeneciu, I. V., Huettenbrink, C., Klein, T., Steinemann, S., Bergstraesser, C., Roethke, M., Roth, W., Schlemmer, H. P., Hohenfellner, M., & Hadaschik, B. A. (2015). Comparative analysis of transperineal template saturation prostate biopsy versus magnetic resonance imaging targeted biopsy with magnetic resonance imaging-ultrasound fusion guidance. *J Urol*, 193(1), 87-94. <https://doi.org/10.1016/j.juro.2014.07.098>

- Rice, M. E., & Harris, G. T. (2005). Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen's d, and r. *Law and Human Behavior*, 29(5), 615-620. <https://doi.org/10.1007/s10979-005-6832-7>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* Cham.
- Rundo, L., Han, C., Zhang, J., Hataya, R., Nagano, Y., Militello, C., Ferretti, C., Nobile, M. S., Tangherloni, A., & Gilardi, M. C. (2020). CNN-based prostate zonal segmentation on T2-weighted MR images: A cross-dataset study. In *Neural Approaches to Dynamics of Signal Exchanges* (pp. 269-280). Springer.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
- Sanford, T. H., Zhang, L., Harmon, S. A., Sackett, J., Yang, D., Roth, H., Xu, Z., Kesani, D., Mehralivand, S., Baroni, R. H., Barrett, T., Girometti, R., Oto, A., Purysko, A. S., Xu, S., Pinto, P. A., Xu, D., Wood, B. J., Choyke, P. L., & Turkbey, B. (2020). Data Augmentation and Transfer Learning to Improve Generalizability of an Automated Prostate Segmentation Model. *AJR Am J Roentgenol*, 215(6), 1403-1410. <https://doi.org/10.2214/ajr.19.22347>
- Schulte, E. (2020). Prostata (Vorsteherdrüse). In G. Aumüller, G. Aust, A. Conrad, J. Engele, J. Kirsch, G. Maio, A. Mayerhofer, S. Mense, D. Reißig, J. Salvetter, W. Schmidt, F. Schmitz, E. Schulte, K. Spanel-Borowski, G. Wennemuth, W. Wolff, & L. J. Wurzinger (Eds.), *Duale Reihe Anatomie* (5., korrigierte Auflage ed.). Georg Thieme Verlag. <https://doi.org/10.1055/b-007-170976>
- Schünke, M., Schulte, E., Schumacher, U., Voll, M., & Wesker, K. H. (2018). 5.10 Männliches Genitale: akzessorische Genitaldrüsen. In M. Schünke, E. Schulte, U. Schumacher, M. Voll, & K. H. Wesker (Eds.), *Prometheus LernAtlas - Innere Organe* (5. Auflage ed.). Georg Thieme Verlag KG. <https://doi.org/10.1055/b-006-149645>
- Siddiqui, M. M., Rais-Bahrami, S., Truong, H., Stamatakis, L., Vourganti, S., Nix, J., Hoang, A. N., Walton-Diaz, A., Shuch, B., Weintraub, M., Kruecker, J., Amalou, H., Turkbey, B., Merino, M. J., Choyke, P. L., Wood, B. J., & Pinto, P. A. (2013). Magnetic resonance imaging/ultrasound-fusion biopsy significantly upgrades prostate cancer versus systematic 12-core transrectal ultrasound biopsy. *Eur Urol*, 64(5), 713-719. <https://doi.org/10.1016/j.eururo.2013.05.059>
- Sökeland, J., Schulze, H., Rübber, H., Helpap, B., & Alken, C.-E. (2008). Taschenlehrbuch Urologie 39 Tabellen H1 - Universitätsbibliothek H2 - 48 A 5749. In (14., vollst. überarb. Aufl. ed., pp. 470). Stuttgart [u.a.]: Thieme.

- Tian, Z., Liu, L., Zhang, Z., & Fei, B. (2018). PSNet: prostate segmentation on MRI based on a convolutional neural network. *J Med Imaging (Bellingham)*, 5(2), 021208. <https://doi.org/10.1117/1.Jmi.5.2.021208>
- Turkbey, B., & Haider, M. A. (2022). Deep learning-based artificial intelligence applications in prostate MRI: brief summary. *Br J Radiol*, 95(1131), 20210563. <https://doi.org/10.1259/bjr.20210563>
- Tzelves, L., Protogerou, V., & Varkarakis, I. (2022). Denonvilliers' Fascia: The Prostate Border to the Outside World. *Cancers (Basel)*, 14(3). <https://doi.org/10.3390/cancers14030688>
- Ushinsky, A., Bardis, M., Glavis-Bloom, J., Uchio, E., Chantaduly, C., Nguyentat, M., Chow, D., Chang, P. D., & Houshyar, R. (2021). A 3D-2D Hybrid U-Net Convolutional Neural Network Approach to Prostate Organ Segmentation of Multiparametric MRI. *AJR Am J Roentgenol*, 216(1), 111-116. <https://doi.org/10.2214/ajr.19.22168>
- Villeirs, G. M., & De Meerleer, G. O. (2007). Magnetic resonance imaging (MRI) anatomy of the prostate and application of MRI in radiotherapy planning. *European Journal of Radiology*, 63(3), 361-368. <https://doi.org/https://doi.org/10.1016/j.ejrad.2007.06.030>
- Walsh, P. C., & Donker, P. J. (1982). Impotence following radical prostatectomy: insight into etiology and prevention. *J Urol*, 128(3), 492-497. [https://doi.org/10.1016/s0022-5347\(17\)53012-8](https://doi.org/10.1016/s0022-5347(17)53012-8)
- Walsh, P. C., Lepor, H., & Eggleston, J. C. (1983). Radical prostatectomy with preservation of sexual function: anatomical and pathological considerations. *The Prostate*, 4(5), 473-485.
- Walz, J., Epstein, J. I., Ganzer, R., Graefen, M., Guazzoni, G., Kaouk, J., Menon, M., Mottrie, A., Myers, R. P., Patel, V., Tewari, A., Villers, A., & Artibani, W. (2016). A Critical Analysis of the Current Knowledge of Surgical Anatomy of the Prostate Related to Optimisation of Cancer Control and Preservation of Continence and Erection in Candidates for Radical Prostatectomy: An Update. *European Urology*, 70(2), 301-311. <https://doi.org/https://doi.org/10.1016/j.eururo.2016.01.026>
- Wang, B., Lei, Y., Tian, S., Wang, T., Liu, Y., Patel, P., Jani, A. B., Mao, H., Curran, W. J., Liu, T., & Yang, X. (2019). Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation. *Med Phys*, 46(4), 1707-1718. <https://doi.org/10.1002/mp.13416>
- Weinreb, J. C., Barentsz, J. O., Choyke, P. L., Cornud, F., Haider, M. A., Macura, K. J., Margolis, D., Schnall, M. D., Shtern, F., Tempany, C. M., Thoeny, H. C., & Verma, S. (2016). PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol*, 69(1), 16-40. <https://doi.org/10.1016/j.eururo.2015.08.052>
- Yucel, S., Erdogru, T., & Baykara, M. (2005). Recent neuroanatomical studies on the neurovascular bundle of the prostate and cavernosal nerves: clinical reflections on radical prostatectomy. *Asian J Androl*, 7(4), 339-349. <https://doi.org/10.1111/j.1745-7262.2005.00097.x>
- Zabihollahy, F., Schieda, N., Krishna Jeyaraj, S., & Ukwatta, E. (2019). Automated segmentation of prostate zonal anatomy on T2-weighted

- (T2W) and apparent diffusion coefficient (ADC) map MR images using U-Nets. *Med Phys*, 46(7), 3078-3090. <https://doi.org/10.1002/mp.13550>
- Zanfardino, M., Franzese, M., Pane, K., Cavaliere, C., Monti, S., Esposito, G., Salvatore, M., & Aiello, M. (2019). Bringing radiomics into a multi-omics framework for a comprehensive genotype-phenotype characterization of oncological diseases. *Journal of translational medicine*, 17(1), 337-337. <https://doi.org/10.1186/s12967-019-2073-2>
- Zavala-Romero, O., Breto, A. L., Xu, I. R., Chang, Y. C., Gautney, N., Dal Pra, A., Abramowitz, M. C., Pollack, A., & Stoyanova, R. (2020). Segmentation of prostate and prostate zones using deep learning : A multi-MRI vendor analysis. *Strahlenther Onkol*, 196(10), 932-942. <https://doi.org/10.1007/s00066-020-01607-x>
- Zhu, Y., Wei, R., Gao, G., Ding, L., Zhang, X., Wang, X., & Zhang, J. (2019). Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. *J Magn Reson Imaging*, 49(4), 1149-1156. <https://doi.org/10.1002/jmri.26337>

7 Abbildungsverzeichnis

Abbildung 1. Schematischer Sagittalschnitt der Prostata mit Angabe der Prostataabschnitte und der zonalen Gliederung, abgezeichnet nach Jensen et al. (2019).	2
Abbildung 2. Schematischer Sagittalschnitt der Prostata mit zonaler Gliederung, abgezeichnet nach McLaughlin et al. (2005).....	4
Abbildung 3. Zonale Gliederung der Prostata nach Weinreb et al. (2016)	5
Abbildung 4. Verlauf des posterolateralen Gefäßnervenbündels (rot), abgezeichnet nach McLaughlin et al. (2005)	7
Abbildung 5. MRT-Sequenzen der Prostata, transversal. Von links nach rechts: T2w, ADC-Karte und DWI. Mit roten Pfeilen ist eine ca. 9 x 6 mm messende Läsion in der medialen peripheren Zone rechts markiert (PI-RADS 4).....	11
Abbildung 6. Zusammenhang von Künstlicher Intelligenz, Maschinenlernen und Tiefen Lernen, modifiziert nach Bardis et al (2020)	15
Abbildung 7. Segmentierung der Prostata auf der SATORI-Plattform, Beispiel 1	27
Abbildung 8. Segmentierung der Prostata auf der SATORI-Plattform, Beispiel 2	28
Abbildung 9. Arbeitsfluss bis zur Segmentierung der Prostata.....	29
Abbildung 10. Zusammensetzung der Trainingsdatensätze.....	30
Abbildung 11. Beispiel für die Kategorie 1 im Bereich der Drüsenmitte: Nahezu vollständige Überlappung aller automatisiert vorgenommenen Segmentierungen (Gelb) mit der manuellen Segmentierung (Rot).....	36
Abbildung 12. Beispiel für die Kategorie 1 auf Höhe der Drüsenmitte: Zu sehen sind auf allen vier Bildern dasselbe Schnittbild eines Patienten. Oben links in Rot ist die manuell erstellte Segmentierung zu sehen. Dem gegenüber sind auf den anderen Bildern gelb gestrichelt die automatisiert vorgenommenen Segmentierungen der „KI Tübingen“ (oben rechts), der „KI Mannheim“ (unten links) und der „KI Gemischt“ (unten rechts) dargestellt. Auf allen Bildern ist rechts unterhalb der Segmentierung die manuelle Annotation des neurovaskulären Bündels links („NVB left“, lila Markierung) zu sehen.	37
Abbildung 13. Beispiel für die Kategorie 2 am Übergang der Drüsenmitte zur Prostatabasis: Geringe Abweichung der Umrandungslinie (Gelb) aller automatisiert vorgenommenen Segmentierungen von der manuellen Umrandungslinie (Rot).	38
Abbildung 14. Beispiel für die Kategorie 3 im Bereich der Prostatabasis: starke Abweichung der Umrandungslinien aller drei automatisiert vorgenommenen Segmentierungen (Gelb) von der manuellen Segmentierung (Rot).	39
Abbildung 15. Beispiel für die Kategorie 4 im Bereich der Prostatabasis: die beiden gelb gestrichelten Linien entsprechen der automatisiert vorgenommenen Segmentierung der auf Grundlage der Tübinger Stichprobe trainierten KI. Die kleine Umrandung entspricht dabei der Segmentierung einer Struktur, die bildmorphologisch nicht zum Prostatagewebe gehört.	40
Abbildung 16. Histogramm „Alter zum Zeitpunkt der MRT“ mit Normalverteilungskurve.....	42

Abbildung 17. Histogramm „Prostata-spezifisches Antigen (PSA) in ng/ml“ mit Normalverteilungskurve	45
Abbildung 18. Histogramm „PSA-Dichte in ng/ml/cm ³ “ mit Normalverteilungskurve	46
Abbildung 19. Histogramm „Gewicht des Prostatektomiepräparates“ mit Normalverteilungskurve	49
Abbildung 20. Histogramm mit Normalverteilungskurve „Prostatavolumen gemäß MRT-Befund“	50
Abbildung 21. Streudiagramm zum Gewicht der Prostatektomiepräparate und den Volumina der Prostatae im MRT	52
Abbildung 22. Boxplot zur Verteilung nach Gewicht der Prostatektomiepräparate und der Volumina der Prostatae in der MRT	53
Abbildung 23. Box-Blot zur Verteilung der Dice-Koeffizienten nach Stichprobe	62

8 Tabellenverzeichnis

Tabelle 1. Bildsequenzen der mpMRT nach Empfehlung der DRG und des BDR.....	9
Tabelle 2. PI-RADS-Score Version 2.1.....	12
Tabelle 3. Studien zur CNN-basierten Prostatasegmentation nach Bardis et al. (2020).....	20
Tabelle 4. Verwendete Software.....	24
Tabelle 5. Kategorien (1 bis 4) zur qualitativen Analyse der Dice-Koeffizienten	35
Tabelle 6. Alter in Jahren zum Zeitpunkt der MRT, eingeteilt in Gruppen	43
Tabelle 7. Gleason-Score (Stanzbiopsien und Prostatektomiepräparat).....	44
Tabelle 8. pT-Status der Prostatektomiepräparate	47
Tabelle 9. Gewicht und Maße der Prostatektomiepräparate	48
Tabelle 10. Eingesetzte MRT-Scanner	50
Tabelle 11. Prostatavolumen (MRT-Befund) in cm ³	51
Tabelle 12. Anzahl suspekter Läsionen und Diameter der Indexläsion in cm in der MRT	53
Tabelle 13: Verteilung der PI-RADS-Scores.....	54
Tabelle 14. Anzahl suspekter Läsionen	55
Tabelle 15. Demographische und klinische Charakteristika des Trainings- (UKT_1 bis UKT_100) und Testdatensatzes (UKT_134 bis UKT_185) der Stichprobe Tübingen.....	56
Tabelle 16. Prostatavolumen und PSA-Dichte des Trainings- und des Testdatensatzes	59
Tabelle 17. Statistische Kennzahlen zur Untersuchung auf Unterscheidung hinsichtlich verschiedener klinischer Parameter zwischen Trainings- und Testdatensatz	61
Tabelle 18. Dice-Koeffizienten der drei KIs	62
Tabelle 19. Ergebnisse der Verallgemeinernden Schätzgleichung	64
Tabelle 20. Interaktionsanalyse	65

9 Erklärung zum Eigenanteil

Die Arbeit wurde in der Abteilung Diagnostische und Interventionelle Radiologie der Klinik für Radiologie des Universitätsklinikums Tübingen unter Betreuung von Frau Dr. med. Judith Herrmann, Herrn Prof. Dr. med. Sergios Gatidis sowie meinem Doktorvater Herrn Prof. Dr. med. Konstantin Nikolaou durchgeführt.

Die Konzeption der Studie erfolgte durch mich, Sebastian Reiß, meinem Doktorvater Herrn Prof. Dr. med. Konstantin Nikolaou, meiner Betreuerin Frau Dr. med. Judith Herrmann und meinem Betreuer Herrn Prof. Dr. med. Sergios Gatidis in Rücksprache mit dem Konsortium der Studie „RESECT“. Namentlich zu nennen sind hier der Studienleiter Herr PD Dr. med. Dominik Nörenberg sowie Herr PhD Jan Hendrik Moltz und Herr Dr. Peter Kohlmann vom Fraunhofer Institut für digitale Medizin (MEVIS).

Alle manuellen Segmentierungen der Prostaten der Stichprobe Tübingen wurden von mir, Sebastian Reiß, durchgeführt. Alle Daten, die der statistischen Auswertung und Analyse dieser Arbeit dienten, wurden von mir, Sebastian Reiß, erhoben. Die KI, die SATORI-Plattform sowie die verwendeten Sörensen-Dice Koeffizienten wurden von Fraunhofer MEVIS entwickelt bzw. berechnet und zur Verfügung gestellt.

Die statistische Auswertung erfolgte eigenständig nach Rückversicherung der Methodik und Beratung durch Herrn Dr. Johann Jacoby, Dipl.-Psychologe am Institut für Klinische Epidemiologie und angewandte Biometrie.

Ich versichere, das Manuskript selbständig verfasst zu haben und keine weiteren als die von mir, Sebastian Reiß, angegebenen Quellen verwendet zu haben.

Tübingen, den 31. Dezember 2022