

Aus der Klinik für Radiologie  
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Analyse und Segmentation von MRT-Datensätzen des  
Prostatakarzinoms für die Entwicklung von Computer-basierten  
Algorithmen

Analysis and segmentation of MRI datasets of prostate cancer  
for the development of computer-based algorithms

zur Erlangung des akademischen Grades  
Doctor medicinae (Dr. med.)

vorgelegt der Medizinischen Fakultät  
Charité – Universitätsmedizin Berlin

von

Luise Franz

Datum der Promotion: 30.11.2023

# Vorwort

Teilergebnisse der vorliegenden Arbeit wurden bereits veröffentlicht in:

Makowski MR, Bressemer KK, Franz L, Kader A, Niehues SM, Keller S, Rueckert D, Adams LC. De Novo Radiomics Approach Using Image Augmentation and Features From T1 Mapping to Predict Gleason Scores in Prostate Cancer. Invest Radiol. April 2021. doi: 10.1097/RLI.0000000000000788. Epub ahead of print. PMID: 34047538.

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b> .....	<b>i</b>
<b>Abkürzungsverzeichnis</b> .....	<b>iii</b>
<b>Abbildungs- und Tabellenverzeichnis</b> .....	<b>iv</b>
<b>1. Zusammenfassung</b> .....	<b>1</b>
<b>2. Einleitung</b> .....	<b>5</b>
2.1. Epidemiologie des Prostatakarzinoms.....	5
2.2. Diagnostik des Prostatakarzinoms .....	6
2.2.1. Früherkennung und Basisdiagnostik .....	6
2.2.2. Diagnosesicherung und Staginguntersuchungen .....	9
2.3. Multiparametrische MRT .....	11
2.4. Mapping-MRT als neue quantitative MRT-Technik.....	14
2.5. Radiomics .....	17
2.5.1. Begriffserklärung Radiomics .....	17
2.5.2. Grundgedanke hinter Radiomics.....	18
2.5.3. Prozess der Datengewinnung .....	20
2.5.4. Anwendungsgebiete von Radiomics .....	24
2.5.5. Herausforderungen .....	26
2.6. Fragestellung .....	28
<b>3. Material und Methoden</b> .....	<b>29</b>
3.1. Studienpopulation.....	29
3.2. MRT-Untersuchung.....	29
3.3. Histopathologische Untersuchung .....	31
3.4. Evaluation und Segmentierung der Bilder.....	32
3.5. Augmentation der Bilder und Extraktion der Features .....	34
3.6. Selektion der Features .....	40

3.7. Vorhersagemodelle .....	40
3.8. Statistische Analyse .....	41
<b>4. Ergebnisse.....</b>	<b>43</b>
4.1. Demographische und klinische Daten .....	43
4.2. Evaluation der Vorhersagemodelle.....	48
4.2.1. Random Forest.....	51
4.2.2. Stochastic Gradient Boosting.....	52
4.2.3. L2 Regularized Support Vector Machines .....	53
4.2.4. K-Nearest Neighbors .....	55
<b>5. Diskussion.....</b>	<b>56</b>
5.1. Interpretation der Ergebnisse .....	56
5.2. Vergleich mit vorliegender Literatur .....	57
5.3. Limitationen.....	63
5.4. Schlussfolgerung und Ausblick.....	65
<b>6. Literaturverzeichnis .....</b>	<b>68</b>
<b>Eidesstattliche Versicherung .....</b>	<b>86</b>
<b>Anteilerklärung an Publikationen.....</b>	<b>87</b>
<b>Lebenslauf .....</b>	<b>88</b>
<b>Liste der Publikationen.....</b>	<b>88</b>
<b>Danksagung.....</b>	<b>90</b>
<b>Bescheinigung Statistik.....</b>	<b>91</b>

# Abkürzungsverzeichnis

ADC	.....	apparent diffusion coefficient
AUC	.....	Area under Curve
AWMF	..	Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften
BMI	.....	Body-Mass-Index
BPH	.....	benigne Prostatahyperplasie
bpMRT	.....	biparametrische MRT
CT	.....	Computertomographie
DCE	.....	dynamic contrast enhanced
DRU	.....	digital rektale Untersuchung
DWI	.....	diffusion weighted imaging
ERSPC	.....	European Randomized Study of Screening for Prostate Cancer
ESUR	.....	European Society of Urogenital Radiology
GFR	.....	Glomeruläre Filtrationsrate
IQA	.....	Interquartilsabstand
ISUP	.....	International Society of Urological Pathology
KI	.....	Konfidenzintervall, Konfidenzintervall
KNN	.....	K-Nearest Neighbors
LL	.....	Look-Locker
LoG	.....	Laplacian of Gaussian
MOLLI	.....	Modified Look-Locker Inversion-Recovery
mpMRT	.....	multiparametrische Magnetresonanztomographie
MRT	.....	Magnetresonanztomographie
PCa	.....	Prostatakarzinom
PET	.....	Positronen-Emissions-Tomografie
PI-RADS	.....	Prostate Imaging Reporting and Data System
PIVOT	.....	Prostate Cancer Intervention versus Observation Trial
PSA	.....	prostataspezifisches Antigen
QIN	.....	Quantitative Imaging Network
RF	.....	Random Forest
RFE-SVM	.....	Recursive Feature Selection Support Vector Machines
ROC	.....	Receiver Operating Characteristic
SGB	.....	Stochastic Gradient Boosting
SI	.....	Signalintensität
SMOTE	.....	Synthetic-Minority-Oversampling-Technik
SVM	.....	L2 Regularized Support Vector Machines
t	.....	Zeit
T1WI	.....	T1 weighted imaging
T2WI	.....	T2-weighted imaging
TCGA	.....	The Cancer Genome Atlas
TI	.....	inversion time
TRUS	.....	transrektaler Ultraschall
VOI	.....	Volume of Interest

# Abbildungs- und Tabellenverzeichnis

Abbildung 1: Schema des Radiomics-Prozesses .....	20
Abbildung 2: Schematische Darstellung der MOLLI Sequenz.....	30
Abbildung 3: Manuelle Segmentierung der T2WI, ADC-Maps und T1-Maps. ....	33
Abbildung 4: Entscheidungsbaum innerhalb der Transformationspipeline .....	35
Abbildung 5: Prozess der Datengewinnung .....	39
Abbildung 6: ROC-Kurven und AUC für die Algorithmen RF, SGB und KNN.....	50
Tabelle 1: PSA und DRU in der Früherkennung der AWMF Leitlinie .....	7
Tabelle 2: International Society of Urological Pathology Grading-System.....	10
Tabelle 3: Demographische und klinische Eigenschaften der Population des Trainingsdatensatzes .....	44
Tabelle 4: Demographische und klinische Eigenschaften der Population des Testdatensatzes .....	47
Tabelle 5: Leistung der Vorhersagemodelle des Random Forest Algorithmus.....	51
Tabelle 6: Leistung der Vorhersagemodelle des Stochastic Gradient Boosting Algorithmus .....	52
Tabelle 7: Leistung der Vorhersagemodelle des L2 Regularized Support Vector Machines Algorithmus .....	54
Tabelle 8: Leistung der Vorhersagemodelle des K-Nearest Neighbors Algorithmus.....	55

# 1. Zusammenfassung

**Hintergrund:** Das Prostatakarzinom (PCa) tritt im Laufe des Lebens bei rund einem von neun Männern auf (1). Der histopathologische Gleason-Score (GS) spielt dabei eine entscheidende Rolle in der Wahl der Therapie des PCas (2, 3). Als Alternative zur invasiven Biopsieentnahme wurden in den letzten Jahren Radiomics-Analysen, bei denen große Mengen quantitativer Features aus der Bildgebung extrahiert werden, zum Grading des PCas eingeführt. In dieser Arbeit wurde untersucht, ob das PCa mit Hilfe von Radiomics zuverlässig in der biparametrischen Magnetresonanztomographie (bpMRT) und im T1-Mapping detektiert und zwischen PCas mit einem GS von 6, 7 und  $\geq 8$  unterschieden werden kann.

**Material und Methoden:** In dieser retrospektiven Arbeit wurde eine Radiomics-Analyse von MRT-Bildern (T2-weighted imaging (T2WI), diffusion weighted imaging (DWI) und T1-Mapping) von 66 Patienten mit histopathologisch bestätigtem PCa durchgeführt. Die Tumorkläsionen, die Transitionalzone und die periphere Zone wurden pixelweise manuell segmentiert. Die MRT-Bilder wurden um das zehnfache augmentiert, um die Größe des Datensatzes zu erhöhen, und 1390 Features wurden pro Bild extrahiert. Nach Ausschluss stark korrelierender Features wurden anhand von 876 Features drei Modelle zur Vorhersage des GS (GS 6, GS 7, GS  $\geq 8$ ) durch verschiedene Machine-Learning-Algorithmen trainiert: Ein Modell wurde mit dem Originaldatensatz trainiert, eins mit dem augmentierten Datensatz und eins mit dem augmentierten Datensatz unter Ausschluss von Features aus dem T1-Mapping. Anschließend wurden die Modelle durch einen unabhängigen Testdatensatz evaluiert.

**Ergebnisse:** Die insgesamt beste Leistung mit einer Genauigkeit von 92 % (95% Konfidenzintervall (KI): 0,62 - 1,00) wurde durch den L2 Regularized Support Vector Machines (SVM) gefolgt von dem Random Forest (RF) (83 %; 95 % KI: 0,52 - 0,98), Stochastic Gradient Boosting (SGB) (75 %; 95 % KI: 0,43 - 0,95) und K-Nearest Neighbors (KNN) (50 %; 95 % KI: 0,21 - 0,79) erzielt. Vorhersagemodelle, die mit dem augmentierten Bilddatensatz trainiert wurden, wiesen bei allen vier Algorithmen eine bessere Leistung auf als Modelle, welche mit den Originalbildern trainiert wurden. Der Ausschluss von Features des T1-Mappings ging bei den SVM und dem KNN mit einer

unveränderten und bei dem RF (- 0,16) und dem SGB (- 0,25) mit einer abnehmenden Genauigkeit einher.

**Schlussfolgerung:** In dieser neuartigen Radiomics-Analyse von MRT-Bildern von PCa-Patienten wurde erstmals der Nutzen von Features aus dem T1-Mapping untersucht. Es wurde gezeigt, dass das PCa durch Radiomics zuverlässig in der bpMRT und im T1-Mapping detektiert werden kann. Es konnte mit einer Genauigkeit von bis zu 92 % zwischen PCas mit einem GS von 6, 7 und  $\geq 8$  unterschieden werden. Zudem deuten die Ergebnisse darauf hin, dass das Augmentieren von Bilddaten in Radiomics-Analysen zu einer besseren Leistung von Vorhersagemodellen führen kann.

## Abstract

**Objective:** Prostate carcinoma (PCa) occurs in approximately one in nine men throughout a lifetime (1). The histopathological Gleason score (GS) plays a crucial role in the choice of therapy for PCa (2, 3). As an alternative to invasive biopsy sampling, radiomics analysis, which extracts large amounts of quantitative features from imaging, has been introduced in recent years for grading PCa. In this work, we investigated whether radiomics can reliably detect PCa in biparametric magnetic resonance imaging (bpMRI) and T1 mapping and distinguish between PCas with a GS of 6, 7, and  $\geq 8$ .

**Materials and methods:** In this retrospective study, a radiomics analysis of MRI (T2-weighted imaging (T2WI), diffusion weighted imaging (DWI), and T1 mapping) from 66 patients with histopathologically confirmed PCa was performed. Tumor lesions, the transitional zone, and the peripheral zone were manually segmented pixel by pixel. MR images were augmented tenfold to increase the size of the dataset, and 1390 features were extracted per image. After excluding highly correlating features, 876 features were used to train three models for the prediction of the GS (GS 6, GS 7, GS  $\geq 8$ ) by different machine learning algorithms: one model was trained with the original images, one with the augmented images, and one with the augmented images without features from T1 mapping. Subsequently, the models were evaluated by an independent test dataset.

**Results:** The overall best performance with an accuracy of 92 % (95 % confidence interval (CI): 0.62 - 1.00) was obtained by the L2 Regularized Support Vector Machines (SVM) followed by the Random Forest (RF) (83 %; 95 % CI: 0.52 - 0.98), Stochastic Gradient Boosting (SGB) (75 %; 95 % CI: 0.43 - 0.95) and K-Nearest Neighbors (KNN) (50 %; 95 % CI: 0.21 - 0.79). For all four algorithms, prediction models trained with the augmented image dataset performed better than models trained with the original images. Excluding features from T1 mapping was associated with no change of accuracy for the SVM and KNN and decreased accuracy for the RF (- 0.16) and SGB (- 0,25).

**Conclusion:** In this novel radiomics analysis of MRI from PCa patients, the utility of features from T1 mapping was investigated for the first time. It was shown that PCa could be reliably detected by radiomics in bpMRI and T1 mapping. It was possible to distinguish between PCa with a GS of 6, 7, and  $\geq 8$  with an accuracy up to 92%. In addition, the

results suggest that augmenting image data in radiomics analyses can lead to better performance of predictive models.

## 2. Einleitung

### 2.1. Epidemiologie des Prostatakarzinoms

Mit einer geschätzten Inzidenz von 1,4 Millionen Fällen im Jahr 2020 stellt das Prostatakarzinom (PCa) weltweit die zweithäufigste Krebsdiagnose und zudem die fünfthäufigste zum Tode führende Tumorerkrankung bei Männern dar (4). Ungefähr einer von neun Männern erhält im Laufe seines Lebens die Diagnose eines Prostatakarzinoms (1). In über der Hälfte aller Länder ist es die am häufigsten diagnostizierte Krebserkrankung, so auch in Nord- und Westeuropa, Australien, Neuseeland, Nordamerika sowie vielen Ländern Subsahara-Afrikas (5). In insgesamt 46 Ländern, darunter vermehrt Ländern in Subsahara-Afrika und in der Karibik, ist es die häufigste Ursache für einen durch Krebs verursachten Tod (5). Auf Grund der hohen Inzidenz und der langen Überlebenszeit macht das Prostatakarzinom 20 % aller prävalenten Tumore bei Männern aus und hat die höchste 5-Jahres-Prävalenz aller Tumorerkrankungen (4).

Sowohl die Inzidenz als auch die Mortalitätsrate weisen große geographische Unterschiede auf. So ist in Ländern mit einem hohen soziodemographischen Index die Wahrscheinlichkeit, bis zum Alter von 79 Jahren an einem Prostatakarzinom zu erkranken, im Vergleich mit Ländern mit einem niedrigen bis mittleren Index etwa achtfach erhöht (6). Die höchste Inzidenz ist unter afroamerikanischen Männern in den Vereinigten Staaten zu verzeichnen, während die geringste Inzidenz unter in ihren Heimatländern lebenden asiatischen Männern beobachtet werden kann (7). Der Unterschied in der Inzidenz dieser beiden vorgenannten Gruppen beträgt ein 40-faches (7). Die Variabilität der Inzidenz je nach Region kann unter anderem auch darauf zurückgeführt werden, wie intensiv Früherkennungsmaßnahmen wie der prostataspezifische-Antigen-Test (PSA-Test) durchgeführt werden. Jedoch stellten sich bereits vor der Einführung des PSA-Tests geographische Differenzen dar, sodass zudem Lifestyle-Faktoren eine potenzielle Rolle spielen könnten (8). So nehmen die Inzidenz und Mortalitätsrate des Prostatakarzinoms bei männlichen Migranten, die aus Ländern mit niedrigem Risiko in Länder mit höherem Risiko einwandern, im Vergleich zu Männern ihres Heimatlandes zu, übersteigen jedoch nicht die Raten des Ziellandes (9, 10). Im Zuge der kommerziellen Verfügbarkeit des PSA-Tests in den 1980er Jahren stiegen die Inzidenzraten rasch an und fielen anschließend mit der abnehmenden Anzahl

unentdeckter Fälle wieder ab (5). In Ländern mit späterer Einführung des PSA-Tests ist die Inzidenz weiterhin steigend (5).

Die Entstehung des Prostatakarzinoms ist nicht abschließend geklärt. Sie ist jedoch multifaktoriell und hängt von Risikofaktoren ab, welche die Entstehung eines Prostatakarzinoms begünstigen oder dessen fortschreitenden bzw. tödlichen Verlauf fördern. Bekannte Risikofaktoren, welche die Entstehung eines PCas beeinflussen, sind ein hohes Alter, die afroamerikanische Ethnie und eine positive Familienanamnese bezüglich Prostatakarzinomen (8). Die Inzidenz bei unter 40-jährigen Männern ist sehr gering, während sie ab einem Lebensalter von 55 Jahren deutlich zunimmt (8). Genomweite Assoziationsstudien haben eine genetische Prädisposition von Prostatakarzinomen bestätigt. So wurden in Populationen unterschiedlicher ethnischer Herkunft über 180 genetische Loci identifiziert (11, 12). Wahrscheinlich trägt zudem eine hohe Körpergröße zur Karzinomentstehung bei (13). Da die genannten Faktoren endogen sind, kann die Kenntnis der Risikofaktoren zwar nicht zur Risikominimierung genutzt werden, jedoch zur Identifikation von Risikogruppen herangezogen werden.

Zusätzlich gibt es Risikofaktoren für ein fortgeschrittenes oder tödliches Prostatakarzinom. Hier sind vor allem eine hohe Körpergröße, Übergewicht und Rauchen zu nennen (8). Eine protektive Wirkung haben körperliche Aktivität und wahrscheinlich der Konsum von Tomaten sowie die Einnahme von Statinen (8).

## **2.2. Diagnostik des Prostatakarzinoms**

### *2.2.1. Früherkennung und Basisdiagnostik*

Der Wert der Früherkennungsmaßnahmen des Prostatakarzinoms ist umstritten, was sich in der Variabilität der Screening-Strategien von klinischen Leitlinien widerspiegelt. Die Leitlinie der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) von 2019 empfiehlt, Männer ab 45 Jahren über die Möglichkeit von Früherkennungsmaßnahmen zu informieren, sofern ihre mutmaßliche Lebenserwartung mehr als 10 Jahre beträgt. Diese Altersgrenze kann bei Männern mit erhöhtem Risiko für ein Prostatakarzinom um fünf Jahre vorverlegt werden. Entscheidet sich der Patient für die Früherkennungsmaßnahmen, so sollen der PSA-Wert bestimmt und eine digital rektale Untersuchung (DRU) durchgeführt werden (s. Tabelle 1) (14).

**Tabelle 1:** Prostataspezifisches Antigen (PSA) und digital rektale Untersuchung (DRU) in der Früherkennung der AWMF Leitlinie (14)

<b>Information über die Früherkennung:</b>	
<ul style="list-style-type: none"> <li>• Männer ab 45 Jahren mit einer Lebenserwartung &gt; 10 Jahre</li> <li>• Ausnahme: Männer mit erhöhtem Risiko für ein Prostatakarzinom ab 40 Jahren</li> </ul>	
<b>Weiteres Vorgehen je nach PSA-Wert (wenn keine Indikation zur Biopsie vorliegt):</b>	
PSA < 1 ng/ml	Nächste PSA-Bestimmung in vier Jahren*
PSA 1–2 ng/ml	Nächste PSA-Bestimmung in zwei Jahren
PSA > 2 ng/ml	Jährliche PSA-Bestimmung
*Ab 70 Jahren ist keine weitere PSA-gestützte Früherkennung empfohlen	
<b>Empfehlen einer Prostatabiopsie bei mindestens einem der folgenden Kriterien:</b>	
<ul style="list-style-type: none"> <li>• PSA <math>\geq</math> 4 ng/ml</li> <li>• Ergebnis bei der DRU, das auf ein Karzinom hindeutet</li> <li>• auffälliger PSA-Anstieg</li> </ul>	

Die Erhebung des PSA-Werts wird auf Grund seiner Vor- und Nachteile kontrovers diskutiert. Dabei hat die Höhe des PSA-Werts weitreichende Folgen für die weitere Diagnostik. So ist der weit verbreitete Grenzwert für die Durchführung einer Biopsie ein PSA-Wert von 4,0 ng/ml. Dieser Grenzwert ist mit einem positiven Vorhersagewert von 30 % für Männer ab 50 Jahren (15) und einem negativen Vorhersagewert von 85 % für Männer mit einem medianen Alter von 69 Jahren zum Zeitpunkt der Biopsie assoziiert (16).

In einem Cochrane Review, welcher fünf randomisierte kontrollierte Studien in einer Metaanalyse evaluierte, konnte kein statistisch signifikanter Unterschied der Prostatakarzinom-spezifischen Mortalität und der allgemeinen Mortalität zwischen der Screening-Gruppe durch das PSA und der Kontrollgruppe festgestellt werden (17). Gleichzeitig wies die Screening-Gruppe eine höhere Inzidenz auf. Dabei wurden lokale Prostatakarzinome vermehrt diagnostiziert, während der Anteil an fortgeschrittenen Prostatakarzinomen abnahm (17). Im Widerspruch dazu zeigte die European Randomized Study of Screening for Prostate Cancer (ERSPC) eine signifikante Reduktion der Mortalität des Prostatakarzinoms durch das PSA-Screening. So lag das Verhältnis der Inzidenzraten der Screening-Gruppe im Vergleich zu der Kontrollgruppe nach einem Follow-Up von 16 Jahren bei 0,8 (95 % KI, 0,72 - 0,89) mit einer absoluten Mortalität von 0,18 % (18). Die Number needed to Screen zur Vermeidung eines Todesfalls lag nach 16 Jahren bei 570 und die Number needed to Treat bei 18 (18).

Studien, welche die Langzeitfolgen von Vorsorgemaßnahmen auf Grundlage der ERSPC berechnet haben, ergaben mit einer längeren Follow-Up Periode nach dem Screening eine zunehmend sinkende Mortalität und geringere Number needed to Screen und Treat (19, 20). Eine Reduktion der Mortalität setzt möglicherweise allerdings erst nach 10 Jahren Follow-Up ein, sodass Männer mit einer Lebenserwartung unter 10 - 15 Jahren darüber aufgeklärt werden sollten, dass die Vorsorgeuntersuchung für sie wahrscheinlich nicht von Vorteil ist (17). Keine der genannten Studien untersuchte die Auswirkungen von Früherkennungsmaßnahmen auf die Lebensqualität.

Die Prostate Cancer Intervention versus Observation Trial (PIVOT)-Studie verglich die radikale Prostatektomie mit der Strategie des Watchful Waitings und implizierte dabei wichtige Rückschlüsse für die Vorsorgeuntersuchung (21). Im Fall einer Therapieentscheidung zur radikalen Prostatektomie war eine sinkende Mortalität erst ab einem PSA-Wert über 10 ng/ml zu verzeichnen (21). Somit könnte der bislang auf 4,0 ng/ml festgelegte Grenzwert zur Durchführung einer Biopsie möglicherweise ohne wesentliche Auswirkungen auf die Mortalitätsrate erhöht werden (22).

Nachteile der Früherkennungsmaßnahmen sind deren Risiken sowie die Überdiagnose und Übertherapie von Patienten, deren Karzinome klinisch unauffällig geblieben wären. So zeigt die ERSPC eine um 40 % höhere Inzidenz von gescreenten Patienten im Vergleich zur Kontrollgruppe nach 16 Jahren Verlaufskontrollen (18). Die Risiken des Screenings umfassen falsch positive Ergebnisse sowie Komplikationen der Biopsie und der Therapie. In einem Review von *Hayes et al.* liegt das kumulative Risiko, ein falsch positives PSA-Testergebnis nach drei bis vier Screening-Runden zu erhalten, bei 12 - 13 % (22).

Werden die sinkende Mortalität und die steigende Rate an überdiagnostizierten und übertherapierten Fällen gegeneinander abgewogen, so wird von einem systematischen Screening abgeraten und weniger intensives Screening einer definierten Zielgruppe von Männern über 50 Jahren oder Männern mit einer positiven Familienanamnese empfohlen. Die Entscheidung zum Screening sollte mit dem Patienten zusammen getroffen werden. Vorher sollte über Vor- und Nachteile aufgeklärt und die individuelle Lebenserwartung in die Entscheidungsfindung einbezogen werden (17, 22). Die optimalen Zeitabstände der Früherkennungsmaßnahmen sind ungewiss, sodass die Empfehlungen je nach Leitlinie variieren (s. Tabelle 1).

### *2.2.2. Diagnosesicherung und Staginguntersuchungen*

Bei einem erhöhten PSA-Wert, einem karzinomverdächtigen Ergebnis bei der DRU oder einem auffälligen PSA-Anstieg wird zur Abklärung eine Stanzbiopsie empfohlen (s. Tabelle 1). Ein erhöhter PSA-Wert allein sollte nach einigen Wochen unter standardisierten Bedingungen im gleichen Labor bestätigt werden, bevor eine Biopsie in Betracht gezogen wird (23). Standardmäßig wird eine systematische Prostatabiopsie mit transrektalem Ultraschall (TRUS) durchgeführt, bei der zehn bis zwölf Gewebeproben in einem gitterartigen Muster gewonnen werden (14, 23). Zur präventiven Antibiose werden trotz steigender Resistenz und damit vermehrten schwerwiegenden Infektionen orale oder intravenöse Chinolone genutzt (23). Häufige leichte bis mittelschwere Nebenwirkungen der Biopsieentnahme sind Blutungen, Hämatome und kurzzeitige Angstzustände. Häufige schwere Nebenwirkungen umfassen Infektionen, Blutungen mit Transfusionsbedarf, Pneumonien, erektile Dysfunktion und Inkontinenz (22).

Vor der Entscheidung zur Biopsie können molekulare Biomarker und die multiparametrische Magnetresonanztomographie (mpMRT) genutzt werden, um die Spezifität des PSA-Tests zu erhöhen und so die Rate an unnötigen Biopsien zu senken (24). Beide werden in den NCCN-Leitlinien von 2018 vor der Durchführung einer Biopsie empfohlen (24). Inzwischen gibt es prognostische Biomarker (z.B. Decipher, Prolaris, Oncotype DX), mit Hilfe derer die Tumoraggressivität, die klinische Progression, die prostatakarzinomspezifische Mortalität und das Risiko zur Metastasierung eingeschätzt werden können (25). Zudem können potenzielle falsch-negative Ergebnisse der Biopsie mit molekularen Biomarkern (z.B. 4Kscore, Prostate Health Index, Prostate Cancer Antigen 3 Test, ConfirmMDx) identifiziert werden (25).

Mit der oben beschriebenen systematischen Prostatabiopsie werden 21 - 28 % der Prostatakarzinome übersehen und 14 – 17 % im Grading zu niedrig eingestuft (25). An diese Stelle spielt die mpMRT, welche morphologische und funktionelle Techniken kombiniert, eine bedeutende Rolle im Diagnoseverfahren. Mit Hilfe der mpMRT können gezielte Biopsien auffälliger Läsionen durchgeführt werden. Hierfür gibt es drei verschiedene Methoden: 1) Fusion von MRT-Bildern mit transrektalem Ultraschall unter Verwendung einer Computer-Software, 2) perkutane Biopsie während der eigentlichen MRT-Untersuchung und 3) kognitive Prostatabiopsie per TRUS nach Begutachtung der MRT-Bilder (25). Metaanalysen von Studienergebnissen zu der MRT-gestützten gezielten versus der systematischen Biopsie bei zuvor biopsienaiven Männern deuten

darauf hin, dass die gezielte im Vergleich zur systematischen Biopsie häufiger klinisch signifikante Karzinome entdeckt (26, 27). Jedoch zeigt die Mehrheit der Studienergebnisse, dass sowohl die systematische als auch die gezielte Biopsie klinisch signifikante Karzinome verfehlt (28), sodass die aktuelle AWMF Leitlinie eine MRT-gestützte, gezielte Biopsie in Kombination mit einer systematischen Biopsie der suspekten Herde empfiehlt (14).

In der anschließenden histopathologischen Untersuchung der Gewebeproben werden die Tumorlokalisation, der International Society of Urological Pathology (ISUP)-Gleason-Score und der ISUP-Grad bestimmt (23). Der Gleason-Score ist ein prognostischer Parameter und spielt eine entscheidende Rolle in der Wahl der Therapie des PCas (2, 3). So profitieren Patienten mit einem niedrigen Gleason-Score von zunächst abwartendem Vorgehen mit regelmäßigen Kontrolluntersuchungen, während bei Patienten mit einem höheren Gleason-Score ein radikale Prostatektomie oder eine Androgendeprivation erwogen werden sollte (24, 29). Zur Bestimmung des Gleason-Scores wird die Drüsenmorphologie begutachtet. Dabei wird die Entdifferenzierung der Drüsen mit Hilfe des Gleason-Grads in fünf Grade eingeteilt. Anschließend errechnet sich der Gleason-Score bei der Prostatastanzbiopsie aus der Addition des häufigsten und des höchsten Gleason-Grads und bei einem Prostatektomiepräparat aus der Addition des am häufigsten und des am zweithäufigsten vorkommenden Gleason-Grads (25). Das Grading des Prostatakarzinoms erfolgt als internationaler Standard nach dem Grading-System der ISUP und unterteilt fünf verschiedene Grade in Abhängigkeit vom Gleason-Score, die Tabelle 2 zu entnehmen sind (3).

**Tabelle 2:** International Society of Urological Pathology (ISUP) Grading-System des Prostatakarzinoms (3)

<b>Grad 1</b>	Gleason-Score von 3 + 3 - ausschließlich einzelne, gut umschriebene Drüsen
<b>Grad 2</b>	Gleason-Score von 3 + 4 - überwiegend einzelne, gut umschriebene Drüsen mit geringem Anteil an malformierten, fusionierten oder kribriformen Drüsen
<b>Grad 3</b>	Gleason-Score von 4 + 3 - überwiegend malformierte, fusionierte oder kribriforme Drüsen mit geringem Anteil an gut umschriebenen Drüsen

<b>Grad 4</b>	Gleason-Score von 4 + 4, 3 + 5 und 5 + 3 <ul style="list-style-type: none"> <li>- ausschließlich malformierte, fusionierte oder kribriforme Drüsen oder gut umschriebene Drüsen plus Areale ohne Drüsen</li> </ul>
<b>Grad 5</b>	Gleason-Score von 4 + 5, 5 + 4 und 5 + 5 <ul style="list-style-type: none"> <li>- keine Drüsen erkennbar, ggf. Nekrosen, mit oder ohne malformierte, fusionierte oder kribriforme Drüsen</li> </ul>

Auf Grund der hohen Inzidenz, der niedrigen Mortalitätsrate und der zahlreichen Therapieoptionen ist im Rahmen der Diagnose die Risikostratifizierung des Prostatakarzinoms von großer Bedeutung. Hierzu dient die weitverbreitete Risikogruppeneinteilung nach d'Amico. Sie beschreibt das Risiko für einen biochemischen Rückfall nach radikaler Prostatektomie oder Radiotherapie und hat sich auf Grund ihrer Generalisierbarkeit gegen andere Modelle durchgesetzt (30). Die Einteilung basiert auf dem klinischen TNM-Stadium, dem prätherapeutischen PSA-Wert und dem Gleason-Score. Bei lokal begrenzten Prostatakarzinomen indiziert ein Gleason-Score von 6 ein geringes Risiko, ein Gleason-Score von 7 ein mittelhohes Risiko und ein Score ab 8 ein hohes Risiko (23, 30).

Das anschließende Staging des PCas hängt von den vorhandenen Therapieoptionen, den Präferenzen des Patienten und seinen Komorbiditäten ab. Die European Association of Urology empfiehlt keine weiteren Staging-Untersuchungen bei einem niedrigem Risiko, während ab einem mittelhohem Risiko mit ISUP-Grad 3 eine mpMRT der Prostata für das lokale Staging und eine Metastasensuche mit einer Bildgebung (Computertomographie (CT) oder Magnetresonanztomographie (MRT)) des Abdomens und Beckens und einer Knochenszintigraphie durchgeführt werden sollte (23).

### **2.3. Multiparametrische MRT**

Im Allgemeinen dient die MRT der Prostata der Beurteilung der Anatomie, der Gewebeschaffenheit sowie funktioneller Gewebeeigenschaften. Zur Darstellung von Raumforderungen der Prostata werden Feldstärken von 1,5 oder 3 Tesla mit Surface-Phased-Array-Spulen genutzt, jedoch wird auf Grund des höheren Signal-zu-Rauschen-Verhältnisses und der kürzeren Scanzeit eine Feldstärke von 3 Tesla bevorzugt (31).

Im Fall des Prostatakarzinoms wird ein multiparametrisches Protokoll für die MRT-Bildgebung empfohlen. Die mpMRI umfasst eine hochauflösende T2-gewichtete (englisch: T2-weighted imaging, T2WI) Sequenz sowie mindestens zwei funktionelle MRT-Techniken, die eine Diffusionsbildgebung (englisch: diffusion weighted imaging, DWI) und die dynamische Kontrastmitteluntersuchung (englisch: dynamic contrast enhanced, DCE) umfassen (32, 33). Mittels der T2WI-Sequenzen werden morphologische Informationen über die Prostata gewonnen, sodass die Kapsel, die zonale Anatomie, das Samenbläschen und eine mögliche extraprostatiche Tumorerkrankung evaluiert werden können (34). Das zumeist von der peripheren Zone ausgehende Prostatakarzinom stellt sich dort in der T2WI-Sequenz hypointens im Vergleich zum hyperintensiven glandulären Gewebe der peripheren Zone dar. Allerdings erschwert die unspezifische Darstellung von morphologisch ähnlichen Entitäten wie Prostatitis, intraepithelialen Neoplasien, posttherapeutischen Veränderungen und postbiopsischen Narben die Diagnose, sodass die T2WI-Sequenz nicht ohne die funktionellen MRT-Techniken genutzt werden sollte (34). Zudem sind Tumore in der Transitionalzone der Prostata schwer zu detektieren, da sich die Signalintensitäten (SI) des karzinogenen Gewebes und der Transitionalzone überschneiden (35). So sind Prostatakarzinome in der Transitionalzone lediglich anhand unspezifischer Charakteristika wie einer linsenförmigen oder unscharfen Begrenzung sowie einer homogenen bis moderat hypointensen Signalintensität identifizierbar (31). Vor allem bei Patienten ab dem mittleren Alter können knotige Veränderungen durch eine benigne Prostatahyperplasie (BPH) in der Transitionalzone die Detektion eines Prostatakarzinoms erschweren (35). Hochgradige Karzinome weisen im Vergleich zu niedriggradigen Karzinomen eher eine geringere SI auf (36). Da postbiopsische Hämorrhagien Artefakte hervorrufen können, sollte zwischen der Biopsieentnahme und der MRT ein Abstand von 4 - 6 Wochen eingehalten werden (36).

Die DWI ermöglicht die Darstellung freien Wassers in den Zellzwischenräumen, sodass für jedes Volumenelement ein Diffusionskoeffizient (englisch: apparent diffusion coefficient, ADC) bestimmt werden kann (37). Da im Tumorgewebe eine höhere Zelldichte und damit einhergehend geringere freie Molekularbewegungen bestehen, sind die ADC-Werte dort oft reduziert (38, 39), während sich das Karzinom in der DWI bei hohem b-Wert in vielen Fällen mit hoher SI darstellt. Dies gilt sowohl für die periphere als auch für die Transitionalzone (31). Um das Hintergrundsignal des normalen

Prostatagewebes zu unterdrücken, sind große b-Werte erforderlich (34). Vorangegangene Studien zeigten bei Prostatakarzinomen eine negative Korrelation der ADC-Werte mit dem Gleason-Score und somit der Tumoraggressivität (40-42). Eine Metaanalyse über den klinischen Wert von T2-gewichteten Sequenzen in Kombination mit DWI-Sequenzen zur Diagnose des Prostatakarzinoms hat gezeigt, dass die Nutzung beider Techniken zusammen im Vergleich zur Nutzung von T2-gewichteten Sequenzen eine Erhöhung der Sensitivität von 62 % auf 72 % und der Spezifität von 77 % auf 81 % herbeiführen kann (43). Jedoch variieren die publizierten ADC-Werte je nach gewählten Feldstärken und b-Werten, sodass der Vergleich absoluter ADC-Werte bei einer quantitativen Auswertung nur eingeschränkt möglich erscheint (34).

Die kontrastmittelabhängige DCE-Sequenz dient zur Beurteilung der Vaskularisierung der Prostata und der Neoangiogenese von Prostatakarzinomen (44). Die DCE-MRT ermittelt T1W-Signalintensitäts(SI)-Zeit(t)-Verläufe im Prostatagewebe in axialer Schichtführung nach intravenöser Injektion von gadoliniumhaltigem Kontrastmittel (45). Die Auswertung der SI kann qualitativ und semiquantitativ erfolgen, der Verlauf der SI-t-Kurve auch quantitativ (37). Durch die Verbindung T2-gewichteter Sequenzen mit DCE-Sequenzen wird die Sensitivität der Karzinomdetektion gesteigert (44, 46), jedoch ermöglicht die DCE-Sequenz ebenfalls keine zuverlässige Differenzierung zwischen Prostatakarzinom, Prostatitis und BPH (34). Aktuell wird der Nutzen der DCE unter Abwägung zusätzlicher Scanzeit und Kosten (47) sowie des Nebenwirkungsprofils der gadoliniumhaltigen Kontrastmittel (48) kontrovers diskutiert. Vor allem bei biopsienaiven Männern wird die biparametrische MRT (bpMRT) als Alternative zur mpMRT in Erwägung gezogen. Die bpMRT beinhaltet keine DCE und nutzt nur multiplanare T2-gewichtete Sequenzen sowie eine axiale DWI (49). Die Studienlage zum Vergleich der mpMRT und bpMRT ist sehr heterogen, sodass aktuell noch keine einheitliche Empfehlung ausgesprochen werden kann (49).

Als international anerkannter Standard zur Durchführung und objektiver Interpretation der mpMRT wird das Prostate Imaging Reporting and Data System (PI-RADS) angewandt, dessen aktualisierte Version V2 2015 von der European Society of Urogenital Radiology (ESUR) entwickelt wurde (50). Das PI-RADS V2 ermöglicht es den Untersuchenden, alle bei der mpMRT identifizierten fokalen Läsionen der Prostata zu beurteilen und anhand einer 5-stufigen Likert-Skala zu bewerten. Diese gibt die Wahrscheinlichkeit eines klinisch signifikanten Prostatakarzinoms an, wobei ein Gleason-Score ab 7, ein Volumen über

0,5 ml oder extraprostatitisches Wachstum als Kriterien für ein klinisch signifikantes Karzinom gelten (50). Abhängig von der Zone, in der die Läsion lokalisiert ist, ist eine bestimmte MR-Sequenz entscheidend für den Wert auf der Likert-Skala. So sind in der peripheren Zone vorerst die DWI und die korrespondierende ADC-Map ausschlaggebend, während in der Transitionalzone vorrangig T2-gewichtete Sequenzen über den Wert entscheiden (50).

Insgesamt gewinnt die mpMRT, vor deren Einführung die MRT vor allem zum Staging, zur Therapieplanung und zum Monitoring des Prostatakarzinoms verwendet wurde, zunehmend an Bedeutung. Da das multiparametrische Protokoll unter Anwendung des PI-RADS V2 eine hohe Sensitivität und Spezifität zur Detektion klinisch signifikanter Prostatakarzinome aufweist, wie in verschiedenen systematischen Reviews und Metaanalysen bestätigt wurde (51-53), empfehlen aktuelle Leitlinien nun eine mpMRT-Untersuchung im Rahmen der Initialdiagnostik vor der Biopsieentnahme (14, 24). So dient die MR-Bildgebung heutzutage auch der Tumordetektion, Lokalisation und Charakterisierung. Darüber hinaus wird aktuell die zukünftige Nutzung der mpMRT zur Ermittlung des prognose- und therapieentscheidenden Gleason-Scores als akkurate und nichtinvasive Alternative zur transrektalen Biopsie diskutiert, um die Komplikationen und den Kostenaspekt der MRT-gestützten oder MRT-Ultraschall-Fusions-Biopsie zu vermeiden (54).

#### **2.4. Mapping-MRT als neue quantitative MRT-Technik**

Die technologischen Fortschritte der letzten Jahre haben eine Quantifizierung von konventioneller T1- (englisch: T1 weighted imaging, T1WI) und T2-gewichteter MR-Bildgebung ermöglicht. Hierfür werden Kartierungen generiert, welche die Signalintensitäten eines jeden Voxels aus der T1WI bzw. T2WI in Form von farbkodierten Pixeln darstellen (55). So entstehen farbkodierte Kartierungen, die eine quantitative Annäherung an die T1- bzw. T2-Relaxationszeiten ermöglichen und den objektiven Vergleich von Gewebeeigenschaften mit Referenzwerten sowie mit longitudinalen Verlaufswerten gestatten (56).

Die T1-Kartierung stellt die longitudinale T1-Relaxationszeit dar, also die benötigte Zeit (in ms) der Kernspin-Magnetisierung für eine Rückkehr in ihren Gleichgewichtszustand nach Applikation eines Radiofrequenzpulses und dadurch bedingter Inversion um 180

Grad (55). Sie kann durch die exponentielle Funktion der Inversion-Recovery-Curve beschrieben werden (55).

Es werden verschiedene Techniken zur Kartierung angewandt, wobei die nach ihren Entwicklern benannte Look-Locker-Technik als eine der effizientesten gilt, da die zur Kartierung benötigten Daten während eines einzigen Atemzuges generiert werden können (57). Jedoch findet die Datengenerierung in unterschiedlichen Phasen des Herzzyklus statt, sodass die fortlaufend generierten Bewegungsartefakte die kardiale T1-Kartierung verhindern (57). Einen großen Fortschritt in der kardialen MRT mit signifikanten Vorteilen lieferte die Weiterentwicklung der inzwischen weit verbreiteten Modified Look-Locker Inversion-Recovery-Technik (MOLLI) durch *Messroghli et al.* (58). Die MOLLI-Sequenz ist EKG-getriggert und generiert in der Diastole mehrerer Herzschläge drei bis fünf Inversion-Recovery-Bildserien zu unterschiedlichen Zeiten nach dem Inversionspuls (55). Nach einer Pause von drei Herzschlägen wird eine erneute Inversion ausgelöst und mit einer leichten zeitlichen Verzögerung werden weitere drei Bilder generiert, sodass zusätzliche Punkte auf der Inversion-Recovery-Curve abgebildet werden können (55). Anschließend wird aus den Signalintensitäten der Bilder mit Hilfe der exponentiellen Funktion die T1-Relaxationszeit berechnet und farbkodiert in die T1-Kartierung übertragen (58). Durch die Anwendung der MOLLI-Sequenz können im Gegenteil zu bisherigen Quantifizierungsmethoden T1-Kartierungen mit hoher räumlicher Auflösung während eines einzigen Atemzugs generiert werden (57). Weitere Stärken sind die schnelle Generierung reproduzierbarer T1-Kartierungen, eine hohe Objektivierbarkeit und ein überzeugendes Signal-zu-Rauschen-Verhältnis (58, 59).

Der wesentliche Vorteil von T1-Kartierungen ist die Möglichkeit der objektiven quantitativen Auswertung. Konventionelle MR-Techniken wie die T1WI bestehen aus qualitativen nichtparametrischen Sequenzen und erlauben somit lediglich eine semiquantitative Auswertung, die zwischen verschiedenen Betrachtenden und Untersuchungszeitpunkten variiert (55). In der Prostata-MRT werden konventionelle T1-gewichtete Sequenzen zur Detektion von entzündungs- oder biopsiebedingten intraprostatischen Hämorrhagien und ggf. metastasensuspекt vergrößerten parailiakalen und lokoregionären Lymphknoten genutzt, spielen aber keine Rolle in der Tumorcharakterisierung (60, 61).

T1-Kartierungen haben bislang vor allem in der kardialen MR Anwendung gefunden, da das T1-Mapping die auf fokale Pathologien limitierte T1WI-Sequenz um die Darstellbarkeit und Quantifizierung diffuser Veränderungen wie reaktive Fibrose, Amyloidose und entzündliche Myopathie erweitert (55, 62, 63). Das native T1-Mapping dient außerdem der Darstellung von Myokardinfarkten, myokardialen Ödemen, Eisenüberladung und longitudinalen myokardialen Veränderungen (55).

Bei der Diagnose des Prostatakarzinoms findet die T1-Kartierung aktuell keine standardmäßige Anwendung, wurde aber bereits von einigen Studiengruppen mit unterschiedlichen technischen Herangehensweisen untersucht. *Kjaer et al.* konnten 1987 keinen signifikanten Unterschied der T1-Relaxationszeit zwischen Prostatakarzinomen und benigner Prostatahyperplasie bei 18 Patienten feststellen (64). 2011 stellten *Foltz et al.* ein signifikant kürzeres T1 in PCa-Läsionen im Vergleich zu gesundem Prostatagewebe bei 13 Patienten fest (65). *Yu et al.* kamen bei der Untersuchung von 109 Patienten mit PCa unter Anwendung von MR-Fingerprinting zu dem gleichen Ergebnis, konnten aber keine signifikanten Unterschiede zwischen niedrig- und hochgradigen PCa's zeigen (66). Die Studie beschränkte sich jedoch auf Gewebe der peripheren Zone. Neueste Erkenntnisse von *Baur et al.* zeigten bei 23 PCa-Patienten unter Nutzung der MOLLI-Sequenz ebenfalls eine signifikant niedrigere T1-Relaxationszeit im PCa im Vergleich zu Referenzregionen der peripheren und Transitionalzone (33). Dabei wiesen sie eine große Area under Curve (AUC) sowie eine hohe Sensitivität und Spezifität zur Unterscheidung zwischen PCa-Läsionen und Referenzgebieten der peripheren Zone bzw. Transitionalzone auf (33). Die AUC der Kombination von T1- und ADC-Werten war signifikant höher als die der alleinigen ADC-Werte. Es konnte allerdings konform zu *Yu et al.* kein signifikanter Unterschied zwischen der T1-Relaxationszeit von niedrig- und hochgradigem PCa ermittelt werden (33). Damit übereinstimmend sind die Ergebnisse von *Panda et al.* in Bezug auf die periphere Zone (67). Jedoch konnte für klinisch signifikante Prostatakarzinome in der Transitionalzone eine signifikant niedrigere T1-Relaxationszeit im Vergleich zu niedriggradigen Läsionen nachgewiesen werden (32). Hierbei erzielte die kombinierte Anwendung mit ADC-Werten wiederum eine höhere AUC als T1 allein.

Die heterogenen Studienergebnisse sprechen für eine weitere Untersuchung des potenziellen Nutzens der T1-Relaxationszeit zur Bestimmung des Tumorgads von

Prostatakarzinomen und zur Unterscheidung des Prostatakarzinoms von gesundem Gewebe.

## 2.5. Radiomics

Der zunehmende technische Fortschritt in der radiologischen Bildgebung in den letzten Jahrzehnten ermöglichte die Entwicklung neuer, automatisierter Analysemethoden, um größere Mengen an Informationen aus den Bildmerkmalen zu extrahieren und die wachsenden Datensätze auszuwerten. Zu diesen Innovationen in der Bildanalyse zählen auch Radiomics (68). Mit Hilfe von Radiomics können hochdimensionale Datensätze in Form von quantitativen Bildmerkmalen aus tomographischen Bildern der Computertomographie, MRT und Positronen-Emissions-Tomographie extrahiert werden (69).

### 2.5.1. Begriffserklärung Radiomics

Das Suffix *-omics* hat seinen Ursprung in verschiedenen Disziplinen der Molekularbiologie. Während das darin enthaltene Suffix *-ome* zur Beschreibung einer Gesamtheit ähnlicher Einzelelemente genutzt wird, bedeutet *-omics* die Charakterisierung von Gesamtheiten ähnlicher biologischer Moleküle. Darauf basierend entstanden die Begriffe Genomics zur Beschreibung des Genoms, also der Gesamtheit aller Gene eines Organismus, Transcriptomics in Bezug auf das Transkriptom, was die Gesamtheit aller RNA bedeutet, sowie Proteomics zur Charakterisierung des Proteoms und Metabolomics zur Beschreibung des Metaboloms (69).

Heutzutage wird das Suffix auch bei der Generierung komplexer mehrdimensionaler Daten aus Einzelelementen in anderen medizinischen Forschungsbereichen angewandt (70). So auch in der Radiologie, wo Radiomics einen Prozess der Auswertung digitaler radiologischer Bilder beschreibt, bei dem eine große Anzahl quantitativer Eigenschaften aus den Bildern extrahiert wird (68, 69). Das *omics*-Konzept trifft in mehreren Aspekten auf die quantitative Bildgebung zu, da ein dreidimensionales Bild einerseits mehrere Millionen Voxel beinhalten und zudem pathologische Entitäten hunderte von Merkmalen wie Größe, Form und Struktur aufweisen können (69). Der extrahierte Datensatz an Radiomics-Merkmalen kann anschließend mit anderen Patientendaten kombiniert und durch Data-Mining, also der computergestützten systematischen Anwendung statistischer Methoden auf große Datenbestände, auf bestimmte Zusammenhänge,

Regelmäßigkeiten und Gesetzmäßigkeiten untersucht werden (68, 69). Der entscheidende Vorteil von *-omics*-Datensätzen ist die Möglichkeit der Analyse durch Data-Mining zur Hypothesengenerierung und -überprüfung (69). Die Bildmerkmale bzw. bildbasierten Biomarker können im Zusammenspiel mit Genomics-Datensätzen oder mit anderen *-omics*-Datensätzen mit klinischen Endpunktergebnissen korreliert werden und nach abgeschlossener Analyse zur evidenzbasierten klinischen Entscheidungsunterstützung in Bezug auf Diagnose, Einschätzung von Prognose und Therapieansprechen sowie Monitoring genutzt werden (69).

### *2.5.2. Grundgedanke hinter Radiomics*

Die zugrundeliegende Hypothese von Radiomics besagt, dass quantitative Bildmerkmale Informationen über Prozesse und Veränderungen auf zellulärer Ebene beinhalten und dass folglich pathophysiologische Vorgänge durch die quantitative Bildanalyse erfasst werden können (69, 71). Dies kann auch auf zelluläre pathologische Vorgänge in Tumorzellen angewandt werden, was einen entscheidenden Vorteil in der Beurteilung der Tumorbilogie darstellt. Die quantitativen Bildmerkmale beziehen sich auf Charakteristika wie Form, Größe, Volumen, Gewebestruktur und Signalintensität und liefern ergänzend zu den Informationen aus Laboruntersuchungen, klinischen Berichten, Genomics und Proteomics zusätzliche Informationen zum Tumorphänotyp und zur Mikroumgebung (69).

Mittlerweile ist bekannt, dass die Mehrzahl klinisch signifikanter solider Tumore stark heterogen in Bezug auf Phänotyp, Proteom und Genom ist (72-75). Außerdem unterliegen Tumore Veränderungen im zeitlichen Verlauf, weshalb das Ansprechen auf Targeted Therapies in vielen Fällen in monatlichen und nicht in jährlichen Zeitintervallen untersucht werden sollte (76-78). Die Heterogenität des Genoms von Tumoren und Metastasen innerhalb eines Individuums kann maßgeblich zum Ausbleiben von Therapieerfolgen und zu einer Therapieresistenz beitragen (79). Mit Radiomics kann die Heterogenität eines Tumors sowie die Heterogenität zwischen verschiedenen Tumoren quantitativ erfasst werden (69). Somit stellen die durch Radiomics extrahierten Bildmerkmale vorteilhafte potenzielle in-vivo Biomarker dar (71). Sie beziehen sich im Gegenteil zu aus Biopsien generierten molekularen in-vitro Biomarkern auf den gesamten Tumor, sind hierdurch robuster gegenüber heterogenen Merkmalsausprägungen innerhalb des Tumors und ermöglichen zudem Therapiemonitoring und aktive Überwachung einer Tumorerkrankung (69, 71). Nicht zuletzt hat die Anwendung von Radiomics bei Krebserkrankungen den Vorteil, dass die standardmäßige radiologische

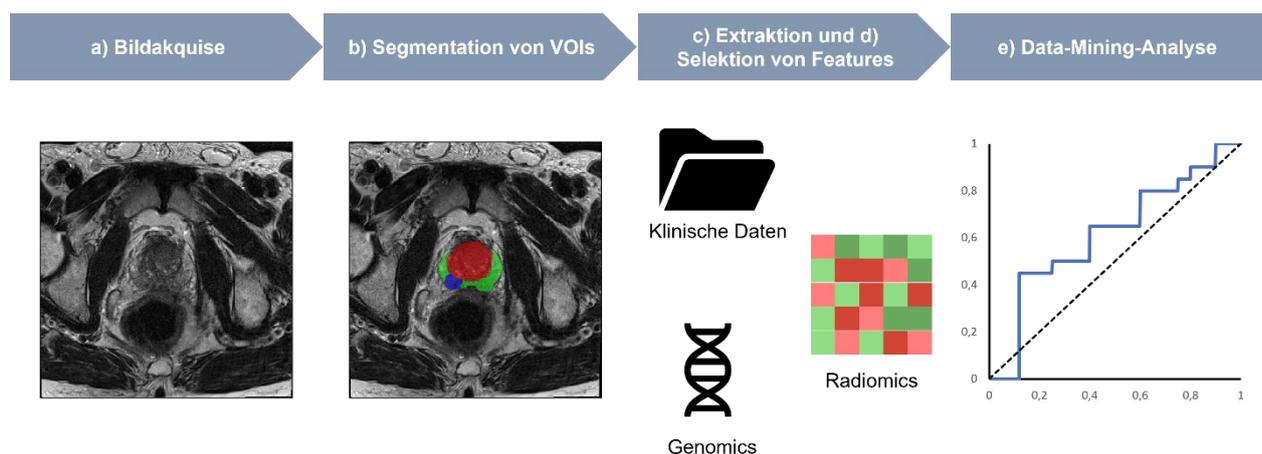
Bildgebung genutzt und diese auch bei fast allen an Krebs Erkrankten durchgeführt wird. Zukünftig könnte die routinemäßige Extraktion von Radiomics-Merkmalen Datenbanken erheblichen Ausmaßes generieren und die Durchführung von Big Data Analysen ermöglichen (69).

Entsprechend der Hypothese könnte die Heterogenität des Tumorgenoms und dadurch bedingte phänotypische Heterogenität durch quantitative Bildgebung darstellbar sein. Die Verbindung von Radiomics und Genomics wird durch den Begriff Radiogenomics beschrieben und geht von einem Zusammenhang von Bildfeatures mit Genexpressionsmustern aus (68, 69, 80). Durch Radiogenomics besteht die Möglichkeit, Genexpressionsmuster und Mutationsstatus durch Radiomics-Daten zu erfassen, um die klinische Entscheidungsfindung zu unterstützen (69). Dies wurde in Studien der Arbeitsgruppe *Kuo et al.* bestätigt. Sie konnte zeigen, dass eine Kombination von 28 Bildfeatures ausreicht, um in hepatozellulären Karzinomen 78 % der Genexpressionsmuster zu rekonstruieren und so die Zellproliferation, synthetische Leberfunktion und Patientenprognose zu erfassen (81). Darüber hinaus stellte sie fest, dass Bildfeatures aus kontrastverstärkter MRT von Glioblastomen die Aktivierung spezifischer Hypoxie- bzw. Proliferations-Genexpressionsprogramme sowie die Überexpression des Epidermal Growth Factor Receptor prognostizieren können, welcher einen potenziellen therapeutischen Ansatz bietet (82).

Grundsätzlich ist zu betonen, dass durch Radiomics-Analysen Korrelationen und keine Kausalzusammenhänge ermittelt werden können. Deshalb sollten die gewonnenen Erkenntnisse zur Unterstützung bei Entscheidungen dienen, anstatt sie isoliert zu betrachten. Beispielsweise kann eine Korrelation mit Genomics- oder anderen *-omics*-Datensätzen nicht nur die Auswahl des Ortes der Biopsieentnahme und die Entscheidung für oder gegen einen Test auf bestimmte Genveränderungen mitbestimmen, sondern auch histopathologische Ergebnisse verifizieren (54, 69). Letzteres ist wichtig, da die Fehlerrate der Tumorhistopathologie durch Fehler bei der Biopsieentnahme und durch die Objektivität bei der Beurteilung hoch ist (69). Beim Prostatakarzinom beträgt sie bis zu 62,8 %, wobei in den meisten Fällen ein zu niedriger Gleason-Score bestimmt wird (83, 84).

### 2.5.3. Prozess der Datengewinnung

Der Radiomics-Workflow umfasst verschiedene Schritte mit unterschiedlichen Herausforderungen, die es zu bewältigen gilt (80) und die im Anschluss genauer beschrieben werden: a) Bildakquise, b) Identifikation und Segmentation des zu untersuchenden Volumens (englisch: Volume of Interest, VOI), c) Extraktion deskriptiver Features aus den VOIs, d) Selektion von Features und e) Data-Mining-Analyse der Datenbanken zur diagnostischen, prognostischen oder prädiktiven Modellentwicklung (69). Eine schematische Darstellung des Workflows ist Abbildung 1 zu entnehmen.



**Abbildung 1:** Schema des Radiomics-Prozesses. Radiomics beginnt mit der Akquise radiologischer Bilder (Computertomographie, Magnetresonanztomographie, Positronen-Emissions-Tomographie). In diesen werden anschließend Volumes of Interest (VOIs), die z.B. Tumore enthalten, identifiziert und segmentiert. Quantitative Features werden von den gerenderten Volumen extrahiert und mit anderen Daten wie z.B. klinischen Daten, Genomics und Proteomics ausgewertet. Anhand dieser Daten werden dann durch Data-Mining diagnostische, prognostische oder prädiktive Modelle entwickelt.

#### a) Bildakquise

Die in medizinischen Bildgebungszentren angewandten Protokolle zur Bildakquise und Bildrekonstruktion moderner CT-, MRT- und Positronen-Emissions-Tomografie(PET)-Bildgebung sind sehr vielfältig und unterliegen keiner Standardisierung (69, 80). So variieren die Protokolle je nach Hersteller in Bezug auf Schichtdicke, Pixelgröße, Positionierung der zu Untersuchenden und Rekonstruktionsalgorithmus (80). Dies stellt ein Problem für die quantitative Bildanalyse dar, weil Abweichungen durch Akquisitions- und Rekonstruktionsparameter auftreten können, sodass Vergleiche von Untersuchungsergebnissen verschiedener Bildgebungszentren mit unterschiedlichen

Scannern erschwert werden (69). In MR-Bildern hängen die Signalintensitäten von einem komplexen Zusammenspiel gewebeeigener Eigenschaften wie der Relaxationszeit sowie von Akquisitionsparametern ab, weshalb eine alleinige Auswertung der Signalintensitäten keinen Informationsgewinn über physikalische Gewebeeigenschaften ermöglicht (80). Während quantitative Sequenzen wie DWI, DCE und T1- bzw. T2-Mapping Rückschlüsse auf physiologische Gewebeeigenschaften erlauben (s. Kapitel 2.3 und 2.4), bleibt ihre Reproduzierbarkeit abhängig von Akquisitionsparametern (80). Als Reaktion auf die fehlende Vergleichbarkeit gab es in den letzten Jahrzehnten mehrere Ansätze zur Festlegung von Akquise- und Rekonstruktionsstandards im Gebiet der quantitativen Bildgebung (69, 80, 85), wobei vor allem das Quantitative Imaging Network (QIN) des National Cancer Instituts als Vorreiter für neue bildgebende Verfahren gilt.

#### *b) Identifikation und Segmentation der Volumes of Interest*

Im Anschluss an die Bildakquise werden VOIs wie Tumore, gesundes Gewebe und andere anatomische Strukturen identifiziert (80). In der Onkoradiologie werden heterogene Teilvolumina innerhalb eines Tumors, die einen prognostischen Wert haben könnten, in radiologischen Berichten oft nicht erfasst. So berücksichtigen komplexe Befundungs- und Datensysteme zwar Bilddeskriptoren bzgl. der Textur, aber keine Heterogenität (69). Beispielhaft sind hier das PI-RADS für die Prostata (s. Kapitel 2.3) sowie ähnliche Systeme zur Befundung von Brust und Lunge zu nennen (60, 86, 87). In Radiomics werden heterogene Teilvolumina miterfasst. Anstatt eingangs Annahmen über den prognostischen Wert von Merkmalen zu treffen und Daten diesbezüglich zu filtern, werden zunächst so viele Daten wie möglich gesammelt und anschließend durch Data-Mining die Merkmale mit dem höchsten prognostischen Wert identifiziert (69).

Das Segmentieren der Bilder in VOIs kann entweder manuell durch erfahrene Radiolog\*innen oder (halb)automatisiert erfolgen und ist der kritischste Schritt des Radiomics-Workflows, da die Daten der Features aus den segmentierten VOIs extrahiert werden (69). Während physiologische anatomische Strukturen wie Knochenelemente und Organe automatisch segmentiert werden können, müssen pathologischen Entitäten wie Tumore auf Grund der intra- und intertumoralen Heterogenität manuell beurteilt werden (69). Die manuelle Segmentation wird oft als ideale Methode gesehen, bringt jedoch eine nicht zu vernachlässigende Variabilität zwischen Untersuchenden (88, 89) und einen hohen Arbeitsaufwand mit sich (80). Die Anwendung auf Radiomics-Analysen

ist daher oft auf kleine Datensätze limitiert (80). Zur Verbesserung der Reproduzierbarkeit und Verringerung der Arbeitslast wurden in den letzten 30 Jahren viele automatische und halbautomatische Segmentierungsmethoden für individuellen Bildgebungsverfahren und anatomische Regionen entwickelt (69, 80). Allerdings herrscht keine übereinstimmende Meinung darüber, welche Metriken zur Evaluierung der Reproduzierbarkeit und Konsistenz der Algorithmen genutzt werden sollten (80). So werden die Segmentierungsmethoden weiterhin kontrovers diskutiert, wobei zunehmend eine computergestützte Kantenerkennung mit anschließender manueller Segmentierung eingesetzt wird (69).

### c) *Extraktion von Features*

Die Extraktion von hochdimensionalen quantitativen Datensätzen über Bildmerkmale der segmentierten VOIs ist der den Radiomics-Workflow definierende Schritt. Sie können entweder direkt aus den Bildern extrahiert werden oder aber aus modifizierten Bildern nach Anwendung bestimmter Filter und Transformationen (90). Es gibt verschiedene Kategorien von Features, von denen die Folgenden für die hier zugrundeliegende Fragestellung von besonderer Bedeutung sind:

1) First-order-Features beschreiben die Verteilung der Werte individueller Voxel ohne Berücksichtigung räumlicher Zusammenhänge (69). Sie reduzieren die dreidimensionalen Daten des Tumolvolumens auf einzelne Werte deskriptiver statistischer Größen wie Median, Mittelwert, Maximum und Minimum sowie auf die Uniformität oder Entropie der Bildintensitäten (69, 80). Zudem können Schiefe und Kurtosis aus einem Signalintensitäts-Histogramm ermittelt werden, welches den prozentualen Anteil an Voxeln (y-Achse) für eine bestimmten Signalintensität (x-Achse) abbildet (69, 80).

2) Shape-based-Features werden aus der dreidimensionalen Oberfläche des gerenderten VOIs extrahiert und beschreiben dessen Form. So können wichtige Features wie das Volumen, der Oberflächeninhalt sowie das Oberflächen-zu-Volumen-Verhältnis berechnet werden (80). Beispielsweise ist letzteres bei einem runden Tumor kleiner als bei einem spikulierten Tumor (80).

3) Matrix-based-Features (auch Second-order-Features) beinhalten Features zur Charakterisierung der Textur und werden durch die Berechnung statistischer Zusammenhänge benachbarter Voxel ermittelt (90). Sie beschreiben die räumliche

Anordnung der Voxel-Signalintensitäten und somit die intratumorale Heterogenität (90). Textur-basierte-Features können auf der Grundlage der Grauwertematrix errechnet werden, welche die Häufigkeit der einzelnen Intensitätswerte in einem festgelegten Abstand in eine bestimmte Richtung angibt (80). In der Praxis finden multiple Methoden zur Extraktion von Matrix-basierten Features Anwendung (91, 92).

4) Higher-order-Features werden durch statistische Methoden entsprechend 1) - 3) ermittelt, nachdem die Bilder durch Filter oder mathematische Transformationen modifiziert wurden, welche dem Erkennen repetitiver oder nicht-repetitiver Muster, dem Hervorheben von Details oder der Rauschunterdrückung dienen (90). Zu den Methoden gehören die Fraktalanalyse, das Minkowski-Funktional, die Wavelet-Transformation und die Laplace-Transformation von Bildern, auf die ein Gauß-Filter angewandt wurde (69).

#### *d) Selektion von Features*

Die Menge der Features zur Charakterisierung der VOIs nach Abschluss der Extraktion ist vergleichbar mit dem Datenvolumen aus Genexpressionsanalysen und kann durch die hohe Komplexität zu Overfitting des zu generierenden Modells an die vorhandenen Daten führen (69, 80). Overfitting tritt auf, wenn der Fehler bezüglich der Trainingsdatensätze sinkt, während gleichzeitig der Fehler bezüglich der Testdatensätze steigt. Zur Vermeidung des Overfittings muss die Dimensionalität des Datensatzes verringert werden, wobei die Selektion von reproduzierbaren und nicht redundanten Features sowie von Features mit einem möglichst großen Kontrastumfang, also einer hohen Variabilität zwischen den Patienten, angestrebt werden sollte (80, 93). Die Herangehensweisen an die Selektion der Features sind vielfältig, da jedes Tool eine unterschiedliche Anzahl an Features unterschiedlicher Kategorien extrahiert und es keine standardisierte Selektionsmethode gibt (90). Eine mögliche Vorgehensweise ist es, zuerst stark miteinander korrelierende und damit redundante Features auszuschließen (69). Hierfür wird eine Korrelationsmatrix berechnet und Features mit einem Korrelationskoeffizienten über einem festgelegten Schwellenwert (z.B.  $r > 0,95$ ) werden ausgeschlossen (69, 80, 93). Anschließend können hoch reproduzierbare Features durch Nutzung eines Test-Retest-Datensatzes und durch Bestimmung des Konkordanzkorrelationskoeffizienten selektiert werden (94, 95). Zukünftig könnten Machine-Learning-Algorithmen auf komplexe Radiomics-Datensätze angewandt werden, um schwer zu erkennende Muster zu identifizieren und so geeignete Features zu selektieren (90).

#### e) *Data-Mining-Analyse der Datenbanken zur Modellentwicklung*

Das Ziel von Radiomics ist je nach Anwendungsfall die Verknüpfung der selektierten Bildfeatures mit klinischen Daten, Genomics-Profilen, histopathologischen Informationen, Serummarkern, Biomarkern und/oder der Krankengeschichte. Nach der Einpflege der Daten in eine Datenbank kann diese durch Data-Mining-Analysen auf Grundlage von Machine-Learning-Methoden oder statistischen Methoden auf Zusammenhänge untersucht werden, um Vorhersagen über eine Ausgabevariable zu treffen (69). Im Machine-Learning wird unüberwachtes Lernen (englisch: unsupervised learning) und überwachtes Lernen (englisch: supervised learning) unterschieden. Algorithmen des unüberwachten Lernens werden mit nicht gelabelten Daten trainiert, das heißt dem Modell werden ausschließlich Eingabedaten und keine Ausgabedaten zur Verfügung gestellt (80). Der Algorithmus fasst die Eingabedaten zusammen, gruppiert sie oder stellt sie graphisch dar und kann so beispielsweise der Selektion von Features dienen (90). Im Gegenteil dazu werden überwachte Lernalgorithmen mit gelabelten Daten trainiert. Sie verfügen somit über sowohl Eingabe- als auch Ausgabedaten, sodass Vorhersagen bezüglich eines bestimmten Ereignisses oder Phänotyps getroffen werden können (80). Hierzu gehören auch die in dieser Arbeit angewandten Algorithmen wie Random Forest (RF), Stochastic Gradient Boosting (SGB), L2 Regularized Support Vector Machines (SVM) sowie K-Nearest Neighbors (KNN).

#### *2.5.4. Anwendungsgebiete von Radiomics*

Durch Radiomics gewonnene in-vivo Biomarker stellen eine nichtinvasive, schnelle und kostengünstige Methode zur Untersuchung phänotypischer Informationen dar, welche die personalisierte Medizin maßgeblich vorantreiben könnte (71). Sie können zur klinischen Entscheidungsfindung in der Diagnose, Prognose und Therapieauswahl bei Krebserkrankungen beitragen:

*Anwendung in der Diagnostik:* Verschiedene Studien haben gezeigt, dass Radiomics-Analysen zur Unterscheidung von Prostatakarzinomen zu benignem Prostatagewebe genutzt werden und zusätzliche Informationen zur Tumoraggressivität liefern können (96-98). *Wibmer et al.* konnten in einer retrospektiven Studie von 147 Männern mit Prostatakarzinomen durch Textur-basierte Features Prostatakarzinome von benignem Prostatagewebe in T2-gewichteten Sequenzen und in ADC-Maps abgrenzen (96). In einer anschließenden Studie wurden dieselben Features zur Ermittlung des Gleason-

Grads genutzt und konnten mit einer Genauigkeit von 93 % zwischen einem Gleason-Score von 6 (3 + 3) und einem Gleason-Score von 7 oder höher unterscheiden sowie mit einer Genauigkeit von 92 % zwischen den beiden Gleason-Scores von 7 (4 + 3 und 3 + 4) (98). *Qi et al.* entwickelten in einer retrospektiven Studie ein Radiomics-Modell, um das Vorliegen eines Prostatakarzinoms bei Männern mit einem erhöhten PSA-Wert zu beurteilen und so unnötige Biopsien zu vermeiden (99). Darüber hinaus könnten Radiomics-Analysen zur Entscheidungsfindung über den Ort der Biopsie oder Resektion beitragen, damit das entnommene Gewebe möglichst viele diagnostische, prognostische und prädiktive Informationen enthält (69).

*Anwendung zur Prognosestellung:* Durch die Möglichkeit der Darstellung der Tumorheterogenität bietet sich eine Anwendung von Radiomics zur Einschätzung der Tumoraggressivität und Prognose an (69). Dies wurde bereits für die Evaluation verschiedener Tumore gezeigt (81, 82, 100). *Aerts et al.* entwickelten durch Radiomics-Analysen einen prognostischen Phänotyp des Lungenzellkarzinoms und konnten diesen anschließend auch auf Kopf- und Nackentumore anwenden (71). Dies warf die Möglichkeit eines generellen prognostischen Phänotyps auf, der in verschiedenen Krebsarten exprimiert wird (71). In einer prospektiven Studie von *Vignati et al.* wurden bei 49 Patienten mit Prostatakarzinom durch Radiomics-Analyse Features aus ADC-Maps und T2-gewichteten Sequenzen gewonnen, um zwischen einem pathologischen Gleason-Score von 6 und 7 oder höher zu unterscheiden (101). Diese Grenze ist entscheidend, da Patienten mit einem Gleason-Score von 6 einer Active Surveillance unterzogen werden können (101). In Vergleich zu standardmäßigen ADC-Grenzwerten mit einer AUC von 0,82 - 0,85 erzielten die Features in ADC- und T2-Maps Werte von 0,92 - 0,96. Diese und weitere Forschungsergebnisse deuten auf eine mögliche Nutzung von Radiomics-Analysen zur Ermittlung der Tumoraggressivität und zur klinischen Entscheidungsfindung hin (96-98, 101, 102).

*Anwendung zur Therapieentscheidung:* Bildfeatures könnten im Sinne der Präzisionsmedizin zur individuellen Vorhersage der zu einem spezifischen Zeitpunkt am besten geeigneten Therapie dienen (69). Für das Hepatozelluläre Karzinom konnten phänotypische Informationen in CT-Bildern identifiziert werden, die mit einem Genexpressionsprogramm korrelieren, welches nach Ansprechen auf eine Doxorubicin-Gabe durchlaufen wird (103). Bei lokal fortgeschrittenen Mammakarzinomen konnten Textur-basierte Features in DCE-Sequenzen in einer Studie von 58 Frauen unterstützend

das voraussichtliche Ansprechen auf eine neoadjuvante Chemotherapie vor deren Beginn abschätzen (104). Beim Prostatakarzinom könnten Therapieentscheidungen entsprechend des durch Radiomics-Analysen ermittelten Gleason-Scores getroffen werden. *Zhang et al.* konnten in einer Studie das Potenzial von auf multiparametrischer MRT basierenden Radiomics-Modellen zur Identifizierung von Prostatakarzinomen, die im Grading nach einer radikalen Prostatektomie im Vergleich zur Biopsie hochgestuft werden, zeigen (54).

#### 2.5.5. Herausforderungen

Obwohl die meisten Schritte des Radiomics-Workflows durch spezielle Software-Tools unterstützt werden können, sind die Qualitätsprüfung der Eingabedaten, die tumorgenaue Segmentierung sowie die Selektion der optimalen Features eine technische Herausforderung und gleichzeitig von großer Bedeutung, da sie eine Voraussetzung für die Generierung von zuverlässigen und reproduzierbaren Ergebnissen sind (90). Neben der technischen Komplexität des Radiomics-Workflows ist das mangelhafte Studiendesign von Data-Mining-Analysen und hierbei insbesondere die gleichzeitige Hypothesengenerierung und -testung problematisch und schränkt die Reproduzierbarkeit von Studienergebnissen ein (69). Die Reproduzierbarkeit von Radiomics-Analysen sollte durch einen unabhängigen Datensatz validiert werden, der vorzugsweise aus einer anderen medizinischen Einrichtung stammt (69). Im Idealfall sollte eine prospektive multizentrische Studie mit dem Radiomics-Biomarker als einen der primären Endpunkte zur Bestätigung der Ergebnisse durchgeführt werden (105, 106). Darüber hinaus sollten detaillierte Daten zur Klinik und zu Risikofaktoren der Patienten in die Radiomics-Analyse einbezogen werden, da diese vor allem in retrospektiven Studien die Bildfeatures durch Confounding beeinflussen können (80, 90). Zudem sollten Standards für die Durchführung von Radiomics-Studien und für die Veröffentlichung von Ergebnissen in der Literatur eingeführt werden, um die Generalisierbarkeit von Studienergebnissen zu erhöhen und Metaanalysen zu ermöglichen (69).

Bei der Bildakquise der konventionellen MRT ist die große Varianz der Akquisitions- und Rekonstruktionsparameter und damit der Signalintensitäten zwischen unterschiedlichen Geräteherstellern sowie verschiedenen medizinischen Institutionen zu beachten (s. Kapitel 2.5.3). So können die Signalintensitäten in zwei nacheinander durchgeführten Aufnahmen beim selben Patienten in unveränderter Position mit dem selben Scanner und Aufnahmeprotokoll unterschiedliche Werte annehmen (107). Die Abhängigkeit der

Radiomics-Features von den numerischen Werten der Signalintensitäten erschwert den direkten Vergleich zwischen verschiedenen Patienten und erfordert eine Normalisierung der Daten (107). Bestehende Standardisierungsmethoden, die auf Signalintensitäts-Histogrammen basieren, konnten die Variation der durchschnittlichen Bildintensitäten innerhalb einer gewebspezifischen Region in verschiedenen Studien unter Nutzung unterschiedlicher Scanner signifikant reduzieren (108, 109), finden aber gegenwärtig auf Grund der Komplexität der Normalisierungsmethoden und mangelnder Erfahrungswerte noch keine Anwendung in Radiomics. Anders verhält es sich bei quantitativer Bildgebung wie den Mapping-Sequenzen, welche hoch reproduzierbar sind (55) und somit die Generalisierbarkeit der Radiomics-Analyse verbessern könnten.

Die statistische Power von prädiktiven, auf Radiomics basierenden Modellen ist maßgeblich von der Größe der verwendeten Datenmenge abhängig (69). Die meisten publizierten Studien weisen eine kleine Fallzahl auf, da die Erfassung von großen, qualitativ hochwertigen Datensätzen aus retrospektiven Daten zeitaufwändig ist. So betrug die retrospektive Datenerhebung in einer Studie von *Gillies et al.* drei Stunden pro Patient (69). Die Durchführung von multizentrischen Studien zur Erhöhung der statistischen Power, das Zusammenfassen der veröffentlichten Studien in Form von Metaanalysen sowie die Validierung von Forschungsergebnissen durch institutionenexterne Daten erfordert die gemeinsame Nutzung von Bilddaten und Metadaten über mehrere Standorte hinweg und ist die größte Herausforderung bei der Etablierung von Radiomics-Modellen als Biomarker (69, 80, 90). Um dies zu gewährleisten, müssen sich multiinstitutionale, nationale oder internationale Zusammenschlüsse auf zentralisierte oder dezentralisierte Netzwerke zum Teilen von Daten einigen (69). Hierfür gibt es bereits einige Lösungsansätze wie The Cancer Genome Atlas (TCGA) (110), welcher umfassende multidimensionale Genomics-Daten und klinische Annotationen von mehr als 30 Krebsarten enthält, sowie das Cancer Imaging Archive (111), welches die Bilddaten von Patienten der TCGA zur Verfügung stellt. Zukünftig sollten Bilddaten und Metadaten für Radiomics-Analysen prospektiv und unter Einhaltung von Standards als Teil eines strukturierten Berichts von elektronischen Krankenaktensystemen erhoben werden, um Datenverluste zu reduzieren und den Prozess effizienter zu gestalten (69).

## 2.6. Fragestellung

In dieser Arbeit wird der Nutzen von Radiomics-Analysen von MRT-Datensätzen zur Diagnostik des Prostatakarzinoms untersucht. Dabei wird eine retrospektive Studie vorgestellt, in der eine neuartige Radiomics-Analyse zur Bestimmung des Gleason-Scores beim PCa entwickelt und validiert wird, welche Image Augmentation und Features aus dem quantitativen T1-Mapping verwendet (112). Zur Einteilung des Gleason-Scores wurden entsprechend der Risikogruppeneinteilung nach d'Amico die Gruppen  $GS = 6$ ,  $GS = 7$  und  $GS \geq 8$  festgelegt.

Im Rahmen der vorliegenden Untersuchung wurden MRT-Datensätze der Prostata für die Entwicklung von Computer-basierten Algorithmen analysiert und segmentiert, um folgende Fragestellungen zu beantworten:

1. Können mit Hilfe von Radiomics spezifische Merkmale von Tumorgraden in den quantitativen Bildfeatures der biparametrischen MRT (ADC und T2WI) identifiziert werden, um zwischen Prostatakarzinomen mit einem Gleason-Score von 6, 7 und  $\geq 8$  zu unterscheiden?
2. Inwieweit können Features aus dem T1-Mapping zur Unterscheidung zwischen Prostatakarzinomen mit einem Gleason-Score von 6, 7 und  $\geq 8$  beitragen?

## **3. Material und Methoden**

### **3.1. Studienpopulation**

Die Studienkohorte der Querschnittsstudie umfasste zu Anfang der Untersuchung 69 Patienten mit histopathologisch bestätigtem PCa. Die beschriebene Patientengruppe erhielt zwischen Februar 2016 und Februar 2020 eine bpMRT-Untersuchung und ein zusätzliches T1-Mapping in 3-Tesla Kernspintomographen an der Charité – Universitätsmedizin Berlin. Es wurden nur Patienten eingeschlossen, die eine histopathologische Diagnose eines PCas unter Angabe des Gleason-Scores sowie eine MRT-Untersuchung mit Verwendung von T2-gewichteter Sequenz, DWI und T1-Mapping erhalten hatten. Außerdem durften zwischen dem Zeitpunkt der MRT-Untersuchung und dem der gezielten Biopsie durch MRT-Ultraschall-Fusion nicht mehr als drei Monaten liegen, um die Aktualität des histopathologischen Befunds zu gewährleisten. Patienten, die allgemeine Kontraindikationen für die MRT-Untersuchung wie Herzschrittmacher, Cochlea-Implantate und Klaustrophobie aufwiesen, wurden von der Studie ausgeschlossen. Zudem galt eine eingeschränkte Bildqualität oder keine in den MRT-Bildern auffindbare Läsion als Ausschlusskriterium. Insgesamt wurden 3 Patienten auf Grund einer mangelhaften Bildqualität von der Studie ausgeschlossen, sodass die anschließende Radiomics-Analyse anhand von Bildern von 66 Männern durchgeführt wurde. Die MRT-Aufnahmen wurden retrospektiv identifiziert und die klinischen Daten der jeweiligen Patienten in eine Datenbank eingepflegt. Ein Ethikantrag mit der Referenznummer EA1/076/16 wurde im Voraus durch die Ethikkommission bewilligt.

### **3.2. MRT-Untersuchung**

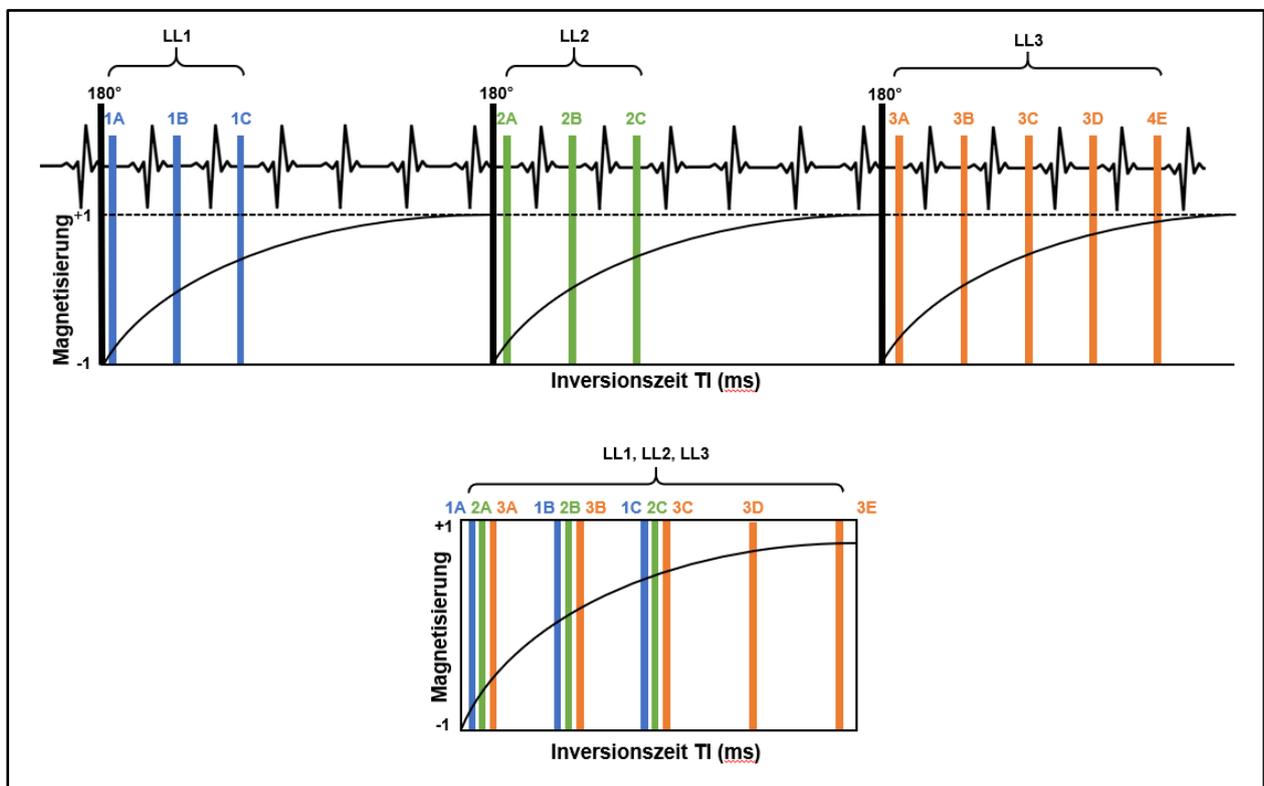
Die MRT-Untersuchungen wurden in der oben genannten Zeitspanne anhand von zwei verschiedenen hausinternen 3-Tesla Kernspintomographen (Siemens Magnetom Skyra und Vida, Erlangen) unter Verwendung von pelvinen Surface-Phased-Array-Spulen durchgeführt. Um Artefakte durch Luft oder Stuhl im Rektum vorzubeugen, wurden die Patienten gebeten, vor der Untersuchung ihr Rektum zu leeren (113).

Die MRT-Protokolle der T2WI und DWI entsprachen den Empfehlungen aktueller Leitlinien. Die Sequenzparameter der axialen, sagittalen und koronaren T2WI waren:

Anzahl an Schichten 24, Voxelgröße 0,52 x 0,47 x 3,00 mm<sup>3</sup>, Sichtfeld 180 x 180 mm, Echozeit / Repetitionszeit 116 ms / 4040 ms, Turbofaktor 25, Kippwinkel 160° und Gesamtakquisitionszeit 3 Minuten und 56 Sekunden.

Für die axiale DWI waren die Sequenzparameter: Anzahl der Schichten 23, Voxelgröße 1,4 x 1,4 x 3,0 mm<sup>3</sup>, Sichtfeld 220 x 220 mm, Echozeit / Repetitionszeit 61 ms / 4400 ms, b-Werte 5, 100, 500, 1000 s/mm<sup>3</sup> und Gesamtakquisitionszeit 5 Minuten und 5 Sekunden. Mit Hilfe einer auf dem MRT-Scanner vorinstallierten Software (Version VE11A) wurden ADC-Maps berechnet. Hierfür wurden b-Werte von 50 und 1000 s/mm<sup>3</sup> sowie Bilder mit b-Werten von 1400 s/mm<sup>3</sup> genutzt.

Für das T1-Mapping wurde die von *Messroghli et al.* weiterentwickelte MOLLI-Sequenz verwendet (s. Kapitel 2.4) (58).



**Abbildung 2:** Schematische Darstellung der MOLLI Sequenz (58). Drei Messungen (LL1, LL2, LL3) wurden nacheinander mit unterschiedlichen Inversionszeiten (TI) nach einem 180° Inversionsimpuls durchgeführt (obere Abbildung). Zur Berechnung der T1-Relaxationszeit wurden die Bilder entsprechend ihrer Inversionszeit geordnet (untere Abbildung). MOLLI = Modified Look-Locker Inversion-Recovery.

Hierbei wurden nach einem 180° Inversionsimpuls nacheinander drei Sätze von Look-Locker-Messungen (LL1, LL2 und LL3) mit zunehmender Inversionszeit (englisch: inversion time, TI) durchgeführt (s. Abbildung 2). Die Messungen erfolgten EKG-getriggert in der Enddiastole und innerhalb eines Atemzuges. Entsprechend dem Schema 3(3)3(3)5 wurden nach einem 180° Inversionsimpuls jeweils 3, 3 und 5 Einzelmessungen durchgeführt, welche zur Wiederherstellung der Magnetisierung von einer Ruhephase mit variabler Dauer je nach Herzfrequenz des Patienten von mindestens 4 s gefolgt waren. Während die Inversionszeit von LL1 100 ms betrug, waren die Inversionszeiten von LL2 mit 200 ms und LL3 mit 350 ms leicht versetzt, um zusätzliche Punkte entlang der Inversion-Recovery-Curve zu erfassen. Anschließend wurden die Bilder entsprechend der Inversionszeit sortiert und die Signalintensitäten der einzelnen Pixel der Inversion-Recovery-Curve hinzugefügt. Anhand der exponentiellen Funktion der Inversion-Recovery-Curve konnte anschließend die T1-Relaxationszeit berechnet werden, bei der definitionsgemäß 63 % der ursprünglichen Längsmagnetisierung wieder erreicht sind.

Die einzelnen Bilder der Sequenzen wurden nacheinander mit folgenden Sequenzparametern generiert: Anzahl der Schichten 20, Voxelgröße 1,3 x 1,3 x 3,0 mm<sup>3</sup>, Sichtfeld 281 x 281 mm, Echozeit / Repetitionszeit 2,45 ms / 808,6 ms, Kippwinkel 35° und Gesamtakquisitionszeit 4 Minuten und 28 Sekunden. Die farbkodierten T1-Maps wurden automatisch pixelweise berechnet und nach der Akquise in multiplen Schichten unter Verwendung einer benutzerdefinierten 12-Bit Look-Up-Tabelle angezeigt.

### **3.3. Histopathologische Untersuchung**

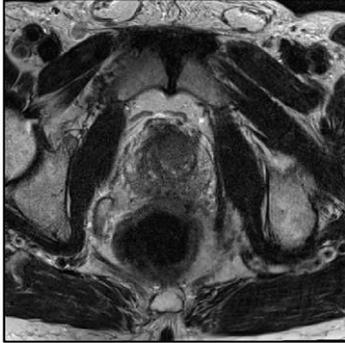
Das Vorliegen eines PCAs wurde durch die histopathologische Untersuchung einer systematischen sonographiegesteuerten Prostatabiopsie bestätigt, wobei bei Angabe mehrerer Grade der höhere Grad gewählt wurde. Patienten mit Läsionen, die in der radiologischen Befundung einen PI-RADS-Score  $\geq 3$  aufwiesen und karzinomverdächtig oder nicht eindeutig waren, erhielten zusätzlich eine gezielte MRT/TRUS-fusionierte Biopsie der Läsionen (114). Anschließend wurden die Stanzbiopsien durch erfahrene hausinterne Uropatholog\*innen nach aktuellen Empfehlungen feingeweblich aufgearbeitet und befundet. Hierbei wurde der Tumorgrad entsprechend den aktuellen Empfehlungen der ISUP zum Grading des Prostatakarzinoms bestimmt (3).

### **3.4. Evaluation und Segmentation der Bilder**

Die biparametrischen MRT-Bilder wurden entsprechend der zweiten PI-RADS-Version durch zwei Radiolog\*innen mit 4 bzw. 5 Jahren Erfahrung in urogenitaler Radiologie klassifiziert (113). Fälle mit unklaren Befunden oder nicht eindeutigen PI-RADS-Scores wurden erneut evaluiert und mit dem schriftlichen radiologischen Befund abgeglichen. Anschließend wurden die T2-gewichteten Sequenzen, ADC-Maps und T1-Maps manuell segmentiert. Hierbei wurden als VOIs Läsionen, deren Wahrscheinlichkeit für das Vorliegen eines klinisch signifikanten Prostatakarzinoms unklar oder wahrscheinlich bis sehr wahrscheinlich war (PI-RADS  $\geq$  3), sowie die Transitionalzone und die periphere Zone identifiziert und pixelweise manuell segmentiert (s. Abbildung 3). Für die Segmentation wurde die Open Source Software „ITK-SNAP“ ([www.itk-snap.org](http://www.itk-snap.org)) genutzt. Anschließend wurden nach Rücksprache mit den Radiolog\*innen bei Bedarf Korrekturen an den Annotationen vorgenommen. Die Annotationen und MRT-Bilder wurden als DICOM-Serien auf einem Netzwerk der hausinternen IT-Abteilung gespeichert.

**Segmentierung der Bilder:**

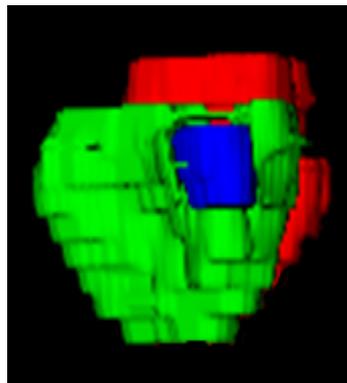
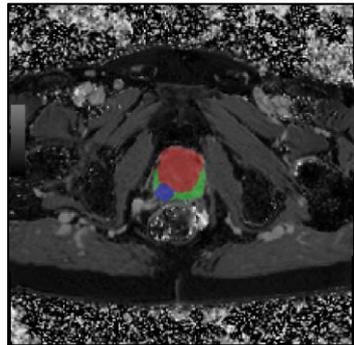
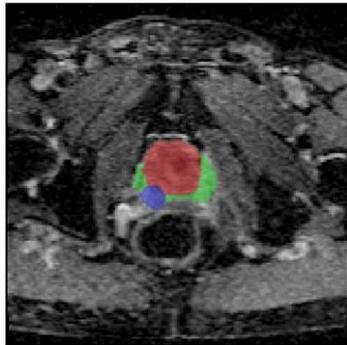
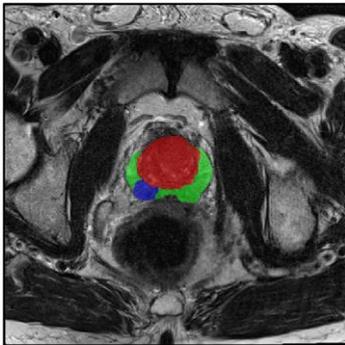
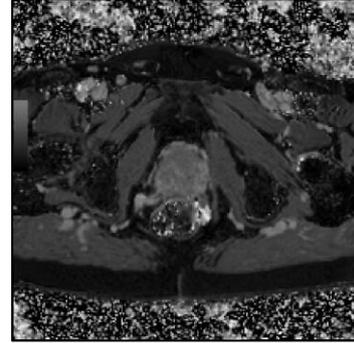
**T2WI**



**ADC-Maps**



**T1-Maps**



**3D-Ansicht:**

**Abbildung 3:** Manuelle Segmentierung der axialen T2-gewichteten Sequenzen (T2WI), Apparent-Diffusion-Coefficient-Maps (ADC-Maps) und T1-Maps. Die periphere Zone, die Transitionalzone und die Tumorkläsion wurden pixelweise segmentiert.

### 3.5. Augmentation der Bilder und Extraktion der Features

Nach der Akquise und Segmentierung der MRT-Bilder erfolgte zunächst eine Korrektur der Daten, um niederfrequente Unregelmäßigkeiten in der Signalintensität auszugleichen. Bei den Unregelmäßigkeiten handelt es sich um anatomisch irrelevante Intensitätsschwankungen, welche durch die Wahl der Radiofrequenzspule, die Pulssequenz sowie durch die Geometrie des abgebildeten Gewebes selbst entstehen können. Sie stellen einen potenziellen Störfaktor bei der Radiomics-Analyse dar (115). Die Korrektur dieses Verstärkungsfeld-Bias erfolgte durch den „N4ITK“-Algorithmus, eine Variante des standardmäßig genutzten N3-Algorithmus (116).

In dem folgenden Radiomics-Workflow wurden zusätzlich zu den Originalbildern augmentierte Bilder analysiert (s. Abbildung 5). Das Augmentieren von Bilddaten ist gleichzusetzen mit der Generierung neuer Bilddaten durch das Variieren existierender Bilddaten. Diese Methode wird bei einigen Machine-Learning-Modellen angewandt, um die Genauigkeit des Vorhersagemodells zu verbessern sowie um Overfitting zu vermeiden (117). Für das Augmentieren der Bilder wurden die Open-Source-Programmbibliotheken „Pytorch“ und „fastai“ verwendet (118, 119). Dabei wurden die MRT-Bilder und Annotationen zunächst mit Hilfe der Software „SimpleITK“ im DICOM-Format geladen (120) und in 3D-Volumina konvertiert, welche anschließend durch eine Transformationspipeline in „Pytorch“ und „fastai“ in Torch-Tensoren konvertiert wurden. Die Transformationspipeline bestand aus der Kombination vier unterschiedlicher Transformationen a), b), c) und d) (s. Abbildung 4). Dabei wurden nicht alle vier Transformationen auf jedes Bild angewandt, sondern jede Transformation per Zufallsprinzip mit einer jeweiligen Wahrscheinlichkeit von 0,25. Die einzelnen Sequenzen einer MRT-Untersuchung (T1-Mapping, T2WI und DWI) durchliefen die Pipeline unabhängig voneinander, sodass mit einer Wahrscheinlichkeit von 97 % ein neues MRT-Bild am Ende eines Durchlaufs entstand. Insgesamt erfolgten zehn Durchläufe des Augmentationsprozesses pro ursprünglichem MRT-Bild, was in etwa einer Verzehnfachung der Anzahl der MRT-Bilder entsprach.



### c) *Hinzufügen von Gauß'schem Rauschen*

Zum Hinzufügen von zufälligem Gauß'schem Rauschen wurde ein Volumenobjekt  $A_{Gauß}$  gebildet, dessen Größe der des 3D-Volumens der segmentierten Bilder  $B_{alt}$  entsprach und welches aus zufällig generierten Gleitkommazahlen mit einem Mittelwert von 0 und einer Standardabweichung von 1 bestand. Anschließend wurde das neu generierte Volumen mit dem Faktor  $s = \frac{1}{100}$  skaliert, voxelweise mit dem 3D-Volumen der MRT-Bilder  $B_{alt}$  multipliziert und schließlich zu diesem addiert, um das transformierte Bild  $B_{neu}$  zu erzeugen:

$$B_{neu} = B_{alt} (1 + sA_{Gauß}) \quad [2]$$

### d) *Perspektivische Verzerrung*

Bei der manuellen Segmentierung entstehen oft kleine Abweichungen, wodurch die Grenzen der Annotation oft nicht den genauen Grenzen des Tumors entsprechen. Um diese Ungenauigkeiten bei der Segmentierung zu simulieren, wurden die Bilder perspektivisch verzerrt. Hierfür wurde ein quellenfreies Vektorfeld generiert. Anschließend wurde das Vektorfeld genutzt, um das 3D-Volumen zu verzerren, indem die ursprünglichen Positionen der Voxel leicht verändert wurden.

Die Extraktion von Features erfolgte mit Hilfe von Python und der Open-Source-Programmbibliothek „PyRadiomics“. „PyRadiomics“ wurde entwickelt, um ein Standardwerkzeug für die Radiomics-Community bereitzustellen und so die Reproduzierbarkeit und Vergleichbarkeit von Radiomics-Ergebnissen zu verbessern (121). Zum Laden der MRT-Bilder in „PyRadiomics“ wurde „SimpleITK“ genutzt. Die Extraktion der Features wurde jeweils vor und nach der Bildaugmentation durchgeführt. Insgesamt wurden 1390 Features aus jedem 3D-Volumen extrahiert.

Unter den 1390 Features waren 18 First-order-Features, 14 Shape-based-Features, 68 Matrix-based-Features und 1290 Higher-order-Features. Higher-order-Features sind per Definition Features, welche nach der Anwendung eines Bildfilters aus den Bildern extrahiert werden. Zur Generierung der Higher-order-Features wurden in der durchgeführten Radiomics-Analyse 15 verschiedene Filter auf die Bilder angewandt.

Anschließend wurden die gleichen 18 First-order-Features und 68 Matrix-based-Features extrahiert, welche bereits vor der Anwendung der Filter extrahiert wurden. Somit ergaben sich 1290 Higher-order-Features. Shape-based-Features beschreiben die Morphologie des Tumors und wurden deshalb nur aus den ungefilterten Bildern extrahiert.

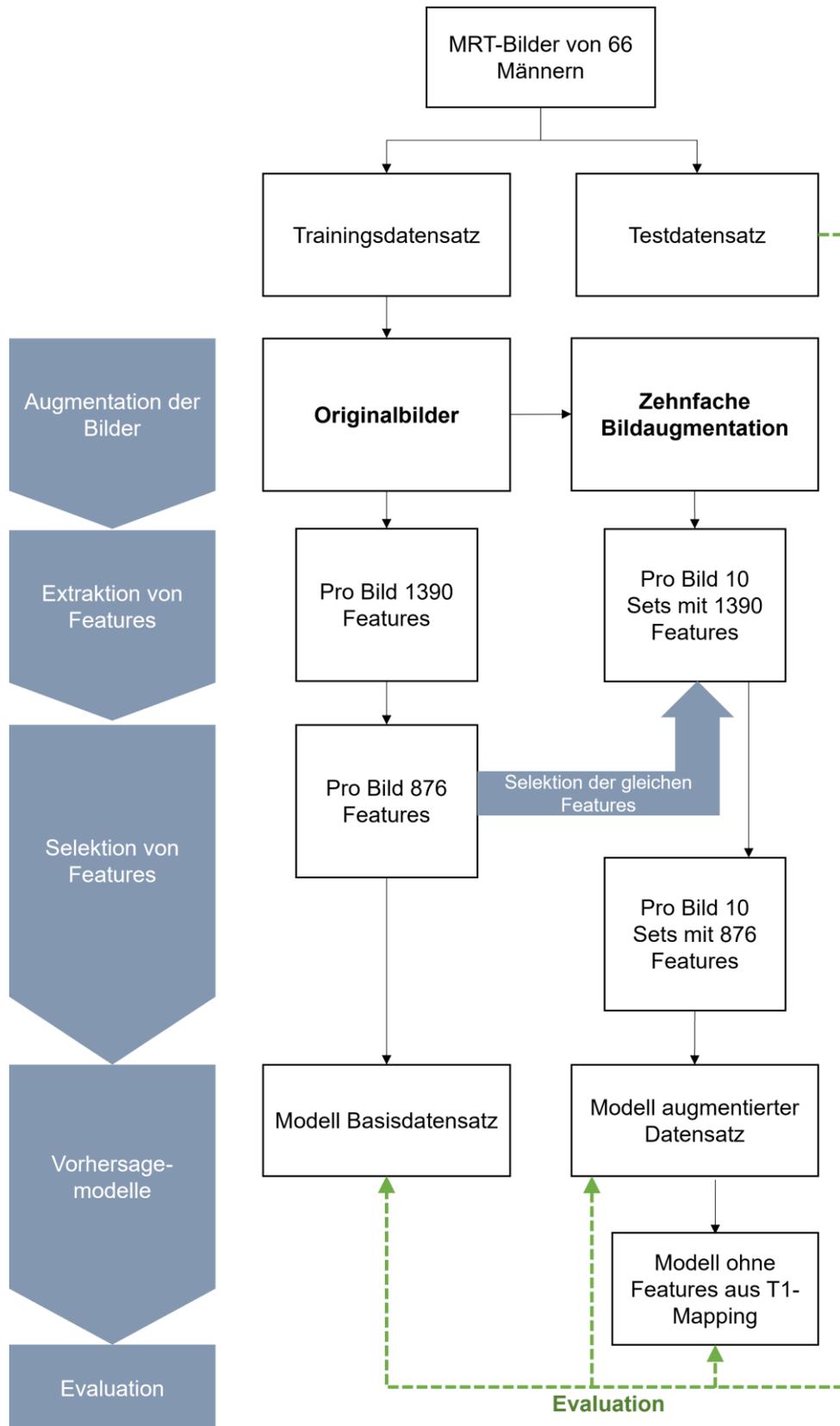
In der Radiomics-Analyse dient die Nutzung von Filtern zur Extraktion von Higher-order-Features dem Hervorheben bestimmter Bildeigenschaften (69). Die in dieser Radiomics-Analyse angewandten Filter waren Wavelets, Laplacian-of-Gaussian(LoG)-Filter und ein Logarithmus-Filter. Durch die Veränderung der Parameter dieser Filtertypen wurden insgesamt 15 verschiedene Filter angewandt.

Der Wavelet-Filter filtert bestimmte Texturinformationen, indem er das Originalbild in niedrige und hohe Frequenzen mit unterschiedlichen Auflösungen zerlegt (122). Hierfür werden Tiefpassfilter und Hochpassfilter auf das Bild angewandt. Tiefpassfilter lassen nur niedrigfrequente Komponenten eines Signals passieren. Hochpassfilter lassen nur hochfrequente Komponenten durch und können genutzt werden, um horizontale, vertikale und diagonale Kanten zu betonen (122). Durch den Wavelet-Filter können Features bestimmter Frequenzbereiche extrahiert werden, sodass diese Informationen über die Textur eines Tumors liefern können (71). In „PyRadiomics“ werden in der axialen, transversalen und frontalen Ebene jeweils acht Zerlegungen generiert, wobei in jeder der drei Ebenen alle möglichen Kombinationen eines Hoch- und Tiefpassfilters möglich sind. Die Größe der acht Zerlegungen entspricht dem Originalbild, sodass die ursprünglichen Tumorgrenzen direkt auf die Zerlegungen nach der Wavelet-Transformation angewandt werden können (71). Auf Wavelets basierende Features wurden bereits in vielen Studien für die Klassifizierung von gesundem Gewebe sowie von Tumoren verwendet (122-127).

Der LoG-Filter dient der Detektion von Kanten und kann zur Hervorhebung der Tumorgrenzen genutzt werden (128). Er setzt sich aus zwei verschiedenen Schritten zusammen: Zuerst wird das Bild durch die Nutzung eines Gauß'schen Kerns geglättet. Anschließend werden Bereiche mit schnellen Veränderungen der Grauwerte durch die Anwendung einer Laplace-Transformation hervorgehoben, was vor allem zu einer Hervorhebung von Kanten führt (121, 129).

Der Logarithmus-Filter in „PyRadiomics“ berechnet aus dem Originalbild den Logarithmus des Absolutwerts + 1 und skaliert die resultierenden Werte anschließend auf

den Bereich des Originalbildes. Dabei werden negative Originalwerte nach Anwendung des Filters wieder negativ gemacht (121). Durch den Logarithmus-Filter wird eine kleine Spannbreite niedriger Werte im Originalbild auf eine größere Spannbreite im gefilterten Bild abgebildet, während die höheren Werte des Originalbilds im gefilterten Bild komprimiert werden.



**Abbildung 5:** Prozess der Datengewinnung. Aus 66 MRT-Bildern wurden ein Trainingsdatensatz (80 %) und ein Testdatensatz (20 %) gebildet. Mit den Bildern des Trainingsdatensatzes wurde die zehnfache Augmentierung der Bilder durchgeführt. Anschließend wurden 1390 Features pro Bild extrahiert. Die Selektion der Features erfolgte anhand der Originalbilder und wurde auf die augmentierten Bilder übertragen. Mit Hilfe des Trainingsdatensatzes wurden drei verschiedene Modelle trainiert, die anschließend durch den Testdatensatz evaluiert wurden. MRT = Magnetresonanztomographie.

### 3.6. Selektion der Features

Zur Selektion der Features wurden die Programmiersprache „R“ (Version 3.7) und das Paket „tidyverse“ genutzt (130, 131). Hierfür wurden die extrahierten Features vorerst mit den klinischen Daten der Patienten kombiniert und entsprechend de Gleason-Scores gruppiert. Anschließend wurde aus 20 % der Originalbilder ein balancierter Testdatensatz gebildet, indem per Zufallsprinzip aus jeder Gruppe Bilder ausgewählt wurden. Die restlichen 80 % bildeten den Trainingsdatensatz.

Nach der Einteilung in den Test- und Trainingsdatensatz wurden geeignete Features aus der Gesamtmenge an 1390 Features ausgewählt. Dies erfolgte ausschließlich anhand der Originalbilder. Hierfür wurden stark korrelierende Features im Trainingsdatensatz identifiziert, indem für die Features der Originalbilder eine Korrelationsmatrix erstellt wurde. Features mit einem Korrelationskoeffizienten  $r > 0,75$  wurden aus der Radiomics-Analyse ausgeschlossen. Anschließend wurden die in den Originalbildern identifizierten stark korrelierenden Features auch aus den augmentierten Bildern ausgeschlossen. Insgesamt wurde so die Anzahl der Features von 1390 auf 314 für die T2-gewichtete Sequenz, 280 für T1-Maps und 282 für ADC-Maps reduziert.

### 3.7. Vorhersagemodelle

Mit Hilfe des Trainingsdatensatzes wurden Vorhersagemodelle zur Prognose des Gleason-Scores trainiert und anschließend durch den Testdatensatz evaluiert. Die Modelle sollten zwischen den drei binären Klassifikatoren eines Gleason-Scores = 6 vs. > 6, eines Gleason-Scores = 7 vs.  $\neq 7$  und eines Gleason-Score  $\geq 8$  vs. < 8 unterscheiden. Das Erstellen der Modelle erfolgte in „R“ unter Verwendung des Pakets „caret“ (132). Die verwendeten Algorithmen waren Random Forest (RF), Stochastic Gradient Boosting (SGB), L2 Regularized Support Vector Machines (SVM) sowie K-Nearest Neighbors (KNN).

Mit jedem der vier Algorithmen wurden mit Hilfe des Trainingsdatensatzes drei verschiedene Modelle trainiert:

- 1) Zunächst wurde mit den Features der Originalbilder ein Basismodell generiert, mit welchem die anderen Modelle im Anschluss verglichen werden konnten. Zur

wiederholten Kreuzvalidierung wurden willkürlich 15 % der Daten zur Bildung eines Validierungsdatensatzes selektiert.

- 2) Anschließend wurde mit jedem Algorithmus ein Modell mit den Features der augmentierten Bilder trainiert. Dabei wurden die Features der Originalbilder zur Validierung genutzt.
- 3) Zur Beurteilung der prognostischen Aussagekraft von T1-Maps wurden im Anschluss daran alle Features aus den T1-Maps exkludiert und ein weiteres Modell anhand der verbleibenden Features aus den augmentierten Bildern trainiert. Die Kreuzvalidierung wurde wiederum mit den Features der Originalbilder durchgeführt.

### **3.8. Statistische Analyse**

Die deskriptive Analyse der Studienpopulation wurde mit Hilfe von „IBM SPSS Statistics 27.0“ durchgeführt. Sämtliche ermittelten Gruppenergebnisse der Trainings- und Testdaten wurden als Mediane mit Angabe des Interquartilsabstands (IQA) angegeben. Die Populationen der drei Gruppen unterschiedlicher Gleason-Scores (GS = 6, GS = 7 und  $GS \geq 8$ ) wurden auf Unterschiede bzgl. klinischer und demographischer Eigenschaften untersucht. Dies wurde gesondert für den Trainings- und Testdatensatz durchgeführt. Hierfür wurde auf Grund der kleinen Fallzahlen der Kruskal-Wallis-Test auf metrische Messgrößen wie Alter, Größe, Gewicht, Nierenfunktionsparameter, PSA-Wert und PSA-Dichte sowie auf ordinal skalierte Messgrößen wie den Charlson-Komorbiditätsindex und den PI-RADS-Score angewandt. Ein möglicher Unterschied zwischen den beiden Gruppen bezüglich der Lokalisation des Prostatakarzinoms wurde mit Hilfe des exakten Fisher-Tests untersucht. Das Signifikanzniveau lag für alle Tests bei 0,05.

Um die Generalisierbarkeit und Genauigkeit der vier trainierten Modelle zu bestimmen, wurden die Modelle durch den Testdatensatz evaluiert. Da zur Prognose des Gleason-Scores ein One-Against-All-Ansatz genutzt wurde, ergab sich eine Prädiktionsmatrix von  $m \times 3$ , in der  $m$  die Anzahl der Beobachtungen im Testdatensatz (Anzahl der Patienten) und die drei Spalten die binären Klassifikatoren eines Gleason-Scores von = 6 vs. > 6,

= 7 vs.  $\neq 7$  und  $\geq 8$  vs.  $< 8$  beschreiben. Um einen ternären Klassifikator des Gleason-Scores mit den Ausprägungen 6, 7 und  $\geq 8$  zu erhalten, wurde die Prädiktionsmatrix zu einem einzigen Vektor reduziert. Hierzu wurde eine Argmax-Funktion auf jede Zeile der Matrix angewandt und die Spalte, in der das Maximum einer Zeile und damit die höchste Wahrscheinlichkeit der Ausprägung eines Klassifikators lag, als finale Prognose bestimmt. Das SVM-Modell war eine Ausnahme, da dieses die Objekte in eine der zwei Klassen einteilt und keine Wahrscheinlichkeiten für die Ausprägung eines Klassifikators generiert werden können. Aus diesem Grund wurde der höchste prognostizierte Gleason-Score der drei Klassifikatoren als finale Prognose ausgewählt.

Zur Beurteilung des ternären Klassifikators wurde eine 3 x 3 Wahrheitsmatrix erstellt, durch welche die Gesamtgenauigkeit sowie die klassenspezifische Sensitivität, Spezifität und Genauigkeit berechnet wurden. Mit Hilfe des R-Pakets „pROC“ (133) wurden die Receiver-Operating-Characteristic-(ROC)-Kurven und die entsprechende AUC-Kurve ermittelt. Als Signifikanzniveau wurde 0,05 festgelegt.

## 4. Ergebnisse

### 4.1. Demographische und klinische Daten

Die durchgeführte Radiomics-Analyse umfasste MRT-Bilder von 66 Männern mit einem histopathologisch bestätigten Prostatakarzinom. Die Studienpopulation wurde in einen Trainingsdatensatz mit 54 Patienten (80 %) und einen Testdatensatz mit 12 Patienten (20 %) aufgeteilt. Eine Gegenüberstellung der demographischen und klinischen Eigenschaften der drei verschiedenen Gruppen mit einem Gleason-Score von 6, 7 und  $\geq 8$  ist den Tabellen 3 und 4 zu entnehmen.

Im Trainingsdatensatz wiesen 16 PCa-Patienten (29,6 %) einen Gleason-Score von 6, 24 (44,5 %) einen Gleason-Score von 7 und 14 (25,9 %) einen Gleason-Score von  $\geq 8$  auf. Im Testdatensatz zeigte die histopathologische Untersuchung bei 4 Patienten (33,3 %) einen Gleason-Score von 6, bei 5 Patienten (41,7 %) einen Gleason-Score von 7 und bei 3 Patienten (25,0 %) einen Gleason-Score von  $\geq 8$ .

Zur Zeit der erhobenen Befunde betrug das mediane Alter der Studienpopulation des Trainingsdatensatzes 71 Jahre (IQA 8 Jahre). Dieses lag in der Studienpopulation des Testdatensatzes bei 74 Jahren (IQA 12 Jahre). Bei einem Signifikanzniveau von 0,05 ergab sich in beiden Datensätzen kein statistisch signifikanter Unterschied zwischen dem Durchschnittsalter der drei Gruppen unterschiedlicher Gleason-Scores ( $p = 0,477$  bzw.  $p = 0,430$ ). Die PCa-Läsion im MRT lag in 89 % der Population des Trainingsdatensatzes und in 67 % der Population des Testdatensatzes in der peripheren Zone. In den restlichen Fällen befand sich das PCa in der Transitionalzone. Auch hier zeigte sich in beiden Datensätzen kein statistisch signifikanter Unterschied zwischen den drei Gruppen unterschiedlicher Gleason-Scores ( $p = 0,126$  bzw.  $p = 0,333$ ). Keiner der Patienten wies multiple maligne Läsionen im MRT auf.

Die Evaluation der MRT-Bilder nach PI-RADS 2 ergab in der Studienpopulation des Trainingsdatensatzes einen Score von 3 bei einem Patienten (1,9 %), einen Score von 4 bei 21 Patienten (38,9 %) und einen Score von 5 bei 32 Patienten (59,2 %). Im Testdatensatz wiesen keine Patienten einen PI-RADS von 3, 4 Patienten (33,3 %) einen PI-RADS von 4 und 8 Patienten (66,7 %) einen PI-RADS von 5 auf. Zwischen den drei Gruppen verschiedener Gleason-Scores bestand sowohl im Trainings- als auch im

Testdatensatz ein statistisch signifikanter Unterschied ( $p = 0,003$  bzw.  $p = 0,01$ ). Dies ist darauf zurückzuführen, dass der PIRADS-Score mit dem Gleason-Score korreliert (134).

**Tabelle 3:** Demographische und klinische Eigenschaften der Population des Trainingsdatensatzes

<b>Variable</b>	<b>Gleason-Score = 6 (n = 16)</b>	<b>Gleason-Score = 7 (n = 24)</b>	<b>Gleason-Score <math>\geq</math> 8 (n = 14)</b>	<b>Gesamter Trainingsdatensatz (n = 54)</b>
Alter (Jahre)	70 (IQA 7)	71 (IQA 13)	72 (IQA 10)	71 (IQA 8)
	$p = 0,477$			
Ort der Läsion im MRT (%)				
<i>Periphere Zone</i>	12 (75,0)	23 (95,8)	13 (92,9)	48 (88,9)
<i>Transitionalzone</i>	4 (25,0)	1 (4,2)	1 (7,1)	6 (11,1)
	$p = 0,126$			
Evaluation im MRT (%)				
<i>PI-RADS 3</i>	1 (6,3)	0 (0,0)	0 (0,0)	1 (1,9)
<i>PI-RADS 4</i>	10 (62,5)	10 (41,7)	1 (7,1)	21 (38,9)
<i>PI-RADS 5</i>	5 (31,2)	14 (58,3)	13 (92,9)	32 (59,2)
	$p = 0,003$			
Prostatavolumen (ml)	42,4 (IQA 21,7)	41,3 (IQA 40,6)	37,6 (IQA 36,6)	42,4 (IQA 34,1)
	$p = 0,813$			
PSA-Wert (ng/ml)	7,0 (IQA 2,5)	9,1 (IQA 7,6)	12,0 (IQA 31,9)	8,4 (IQA 7,5)
	$p = 0,359$			
PSA-Dichte (ng/ml <sup>2</sup> )	0,13 (IQA 0,14)	0,19 (IQA 0,22)	0,28 (IQA 1,0)	0,19 (IQA 0,24)
	$p = 0,173$			
Nierenfunktionsparameter				
<i>Kreatinin (mg/dl)</i>	1,02 (IQA 0,18)	0,98 (IQA 0,29)	0,96 (IQA 0,16)	0,99 (IQA 0,28)
	$p = 0,059$			
<i>GFR (ml/min)</i>	80 (IQA 11)	84 (IQA 25)	90 (IQA 9)	80 (IQA 23)
	$p = 0,891$			
<i>Harnstoff (mg/dl)</i>	28 (IQA 9)	31 (IQA 9)	34 (IQA 9)	31 (IQA 10)

	p = 0,739			
Gewicht (kg)	85,5 (IQA 17,3)	83,5 (IQA 13,5)	79,0 (IQA 10,0)	80,0 (IQA 15,0)
	p = 0,102			
Größe (cm)	180 (IQA 16)	179 (IQA 8)	177 (IQA 5)	178 (IQA 7)
	p = 0,467			
BMI (kg/m <sup>2</sup> )	24,7 (IQA 5,5)	26,3 (IQA 3,1)	24,5 (IQA 3,3)	25,6 (IQA 4,2)
	p = 0,243			
Zusatzerkrankungen (%)				
<i>Herzinfarkt</i>	1 (6,3)	0 (0,0)	0 (0,0)	1 (1,9)
<i>Systolische Herzinsuffizienz</i>	0 (0,0)	0 (0,0)	1 (7,1)	1 (1,9)
<i>Zerebrovaskuläre Erkrankung</i>	2 (12,5)	0 (0,0)	2 (14,3)	4 (7,4)
<i>Chron. Lungenerkrankung</i>	2 (12,5)	2 (8,3)	0 (0,0)	4 (7,4)
<i>Kollagenose</i>	0 (0,0)	0 (0,0)	1 (7,1)	1 (1,9)
<i>Nierenerkrankung**</i>	0 (0,0)	0 (0,0)	1 (7,1)	1 (1,9)
<i>Diabetes Mellitus***</i>	0 (0,0)	4 (16,7)	2 (14,3)	6 (11,1)
<i>Weitere Tumorerkrankung</i>	0 (0,0)	2 (8,3)	1 (7,1)	3 (5,6)
<i>Leukämie</i>	1 (6,3)	2 (8,3)	0 (0,0)	3 (5,6)
<i>Metastasierter solider Tumor</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
Charlson-Komorbiditätsindex (%)				
<i>0 Punkte</i>	11 (68,8)	16 (66,7)	10 (71,4)	37 (68,5)
<i>1 – 2 Punkte</i>	5 (31,2)	6 (25,0)	3 (21,4)	14 (25,9)
<i>3 – 4 Punkte</i>	0 (0,0)	2 (8,3)	1 (7,1)	3 (5,6)
<i>&gt; 5 Punkte</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
	p = 0,918			

Als Durchschnittswert ist der Median angegeben. Abkürzungen: PCa = Prostatakarzinom, MRT = Magnetresonanztomographie, PI-RADS = Prostate Imaging Reporting and Data System, PSA = prostataspezifisches Antigen, BMI = Body-Mass-Index, GFR = Glomeruläre Filtrationsrate, IQA = Interquartilsabstand, p = Signifikanzwert (Signifikanzniveau 0,05). \*nach PI-RADS Version 2 \*\*mäßig schwere oder schwere Nierenerkrankung \*\*\*ohne Endorganschäden.

Der mediane PSA-Wert in der Population des Trainingsdatensatzes lag bei 8,4 ng/ml (IQA 7,5 ng/ml) und das mediane Prostatavolumen bei 42,4 ml (IQA 34,1 ml). Der Median der PSA-Dichte, welche durch die Division des PSA-Werts durch das Prostatavolumen berechnet wird und somit eine Einheit von ng/ml<sup>2</sup> hat, betrug im Trainingsdatensatz

0,19 ng/ml<sup>2</sup> (IQA 0,24 ng/ml<sup>2</sup>). Bei der Population des Testdatensatzes betrug der mediane PSA-Wert 10,7 ng/ml (IQA 6,0), die mediane PSA-Dichte 0,22 ng/ml<sup>2</sup> (IQA 0,10 ng/ml<sup>2</sup>) und das mediane Prostatavolumen 35,9 ml (IQA 24,3 ml). Bei allen drei Messgrößen war der Unterschied zwischen den drei Gruppen verschiedener Gleason-Scores nicht signifikant.

Alle drei Gleason-Score-Gruppen wiesen als Mittelwerte Normwerte für die Nierenfunktionsparameter Kreatinin, Harnstoff und Glomeruläre Filtrationsrate (GFR) auf. In Bezug auf die mutmaßlichen PCa-Risikofaktoren Übergewicht und hohe Körpergröße lag die Studienpopulation mit einem mittleren Body-Mass-Index (BMI) von 25,5 kg/m<sup>2</sup> (IQA 4,0 kg/m<sup>2</sup>) im unteren Bereich einer Präadipositas und mit einer mittleren Größe von 178 cm (IQA 7 cm) über der durchschnittlichen Körpergröße deutscher Männer von 176 cm (135). Der Unterschied zwischen den drei Gleason-Score-Gruppen im Trainings- bzw. Testdatensatz bezüglich Größe ( $p = 0,467$  bzw.  $p = 0,864$ ), Gewicht ( $p = 0,102$  bzw.  $p = 0,182$ ) und BMI ( $p = 0,243$  bzw.  $p = 0,738$ ) war statistisch nicht signifikant.

Einige der PCa-Patienten wiesen Komorbiditäten auf, welche durch den Charlson-Komorbiditätsindex erhoben wurden (136). Von den insgesamt 66 Patienten wies ein Patient einen stattgehabten Herzinfarkt auf, zwei wiesen eine systolische Herzinsuffizienz auf, vier eine zerebrovaskuläre Erkrankung, fünf eine chronische Lungenerkrankung, jeweils einer eine Kollagenose und eine mäßig schwere bis schwere Nierenerkrankung und sieben einen medikamentös behandelten Diabetes Mellitus ohne Endorganschäden. Zudem erlitten drei Patienten eine zusätzliche solide Tumorerkrankung ohne Metastasierung in den letzten fünf Jahren, drei eine Leukämie und ein Patient eine metastasierte solide Tumorerkrankung. Insgesamt erhielten 68 % der Patienten im Charlson-Komorbiditätsindex 0 Punkte, 26 % 1 - 2 Punkte, 5 % 3 - 4 Punkte und 1 % über 5 Punkte. Es war kein statistisch signifikanter Unterschied zwischen den drei Gruppen unterschiedlicher Gleason-Scores im Trainings- bzw. Testdatensatz zu verzeichnen ( $p = 0,918$  bzw.  $p = 0,591$ ).

**Tabelle 4:** Demographische und klinische Eigenschaften der Population des Testdatensatzes

<b>Variable</b>	<b>Gleason-Score = 6 (n = 4)</b>	<b>Gleason-Score = 7 (n = 5)</b>	<b>Gleason-Score ≥ 8 (n = 3)</b>	<b>Gesamter Testdatensatz (n = 12)</b>
Alter (Jahre)	78 (IQA 8)	71 (IQA 11)	74 (IQA 6)	74 (IQA 12)
	p = 0,430			
Ort der Läsion im MRT (%)				
<i>Periphere Zone</i>	3 (75,0)	2 (40,0)	3 (100,0)	8 (66,7)
<i>Transitionalzone</i>	1 (25,0)	3 (60,0)	0 (0,0)	4 (33,3)
	p = 0,333			
Evaluation im MRT (%)				
<i>PI-RADS 3</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
<i>PI-RADS 4</i>	1 (25,0)	0 (0,0)	3 (100,0)	4 (33,3)
<i>PI-RADS 5</i>	3 (75,0)	5 (100,0)	0 (0,0)	8 (66,7)
	p = 0,010			
Prostatavolumen (ml)	33,2 (IQA 15,8)	45,3 (IQA 30,1)	37,6 (IQA 15,5)	35,9 (IQA 24,3)
	p = 0,822			
PSA-Wert (ng/ml)	7,6 (IQA 1,6)	11,2 (IQA 1,3)	8,0 (IQA 97,6)	10,7 (IQA 6,0)
	p = 0,703			
PSA-Dichte (ng/ml <sup>2</sup> )	0,17 (IQA 0,10)	0,17 (IQA 0,10)	0,22 (IQA 1,78)	0,22 (IQA 0,10)
	p = 0,638			
Nierenfunktionsparameter				
<i>Kreatinin (mg/dl)</i>	0,94 (IQA 0,00)	1,09 (IQA 0,19)	0,97 (IQA 0,22)	0,97 (IQA 0,14)
	p = 0,268			
<i>GFR (ml/min)</i>	83 (IQA 9)	72 (IQA 19)	86 (IQA 10)	79 (IQA 18)
	p = 0,800			
<i>Harnstoff (mg/dl)</i>	46 (IQA 0)	37 (IQA 7)	35 (IQA 1)	35 (IQA 9)
	p = 0,600			
Gewicht (kg)	88,0 (IQA 8,5)	80,0 (IQA 10,0)	80,0 (IQA 3,5)	80,5 (IQA 8,8)
	p = 0,182			

Größe (cm)	180 (IQA 9)	176 (IQA 4)	177 (IQA 7)	177 (IQA 7)
	p = 0,864			
BMI (kg/m <sup>2</sup> )	26,1 (IQA 3,5)	24,7 (IQA 2,5)	25,5 (IQA 3,0)	25,3 (IQA 2,9)
	p = 0,738			
Zusatzerkrankungen (%)				
<i>Herzinfarkt</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
<i>Systolische Herzinsuffizienz</i>	0 (0,0)	1 (20,0)	0 (0,0)	1 (8,3)
<i>Zerebrovaskuläre Erkrankung</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
<i>Chron. Lungenerkrankung</i>	0 (0,0)	1 (20,0)	0 (0,0)	1 (8,3)
<i>Kollagenose</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
<i>Nierenerkrankung**</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
<i>Diabetes Mellitus***</i>	0 (0,0)	1 (20,0)	0 (0,0)	1 (8,3)
<i>Weitere Tumorerkrankung</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
<i>Leukämie</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
<i>Metastasierter solider Tumor</i>	0 (0,0)	0 (0,0)	1 (33,3)	1 (8,3)
Charlson-Komorbiditätsindex (%)				
<i>0 Punkte</i>	4 (100,0)	2 (40,0)	2 (66,7)	8 (66,7)
<i>1 – 2 Punkte</i>	0 (0,0)	3 (60,0)	0 (0,0)	3 (25,0)
<i>3 – 4 Punkte</i>	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
<i>&gt; 5 Punkte</i>	0 (0,0)	0 (0,0)	1 (33,3)	1 (8,3)
	p = 0,591			

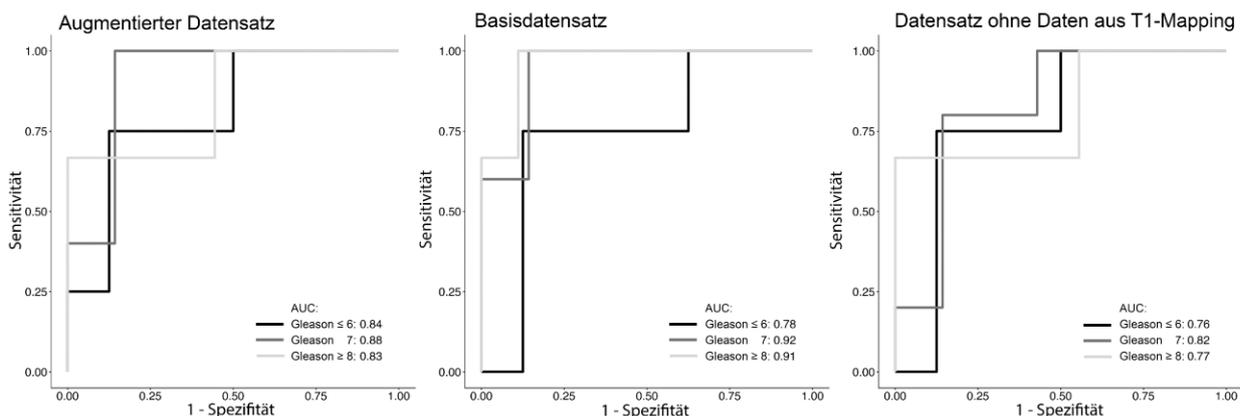
Als Durchschnittswert ist der Median angegeben. Abkürzungen: PCa = Prostatakarzinom, MRT = Magnetresonanztomographie, PI-RADS = Prostate Imaging Reporting and Data System, PSA = prostataspezifisches Antigen, BMI = Body-Mass-Index, GFR = Glomeruläre Filtrationsrate, IQA = Interquartilsabstand, p = Signifikanzwert (Signifikanzniveau 0,05). \*nach PI-RADS Version 2 \*\*mäßig schwere oder schwere Nierenerkrankung \*\*\*ohne Endorganschäden.

## 4.2. Evaluation der Vorhersagemodelle

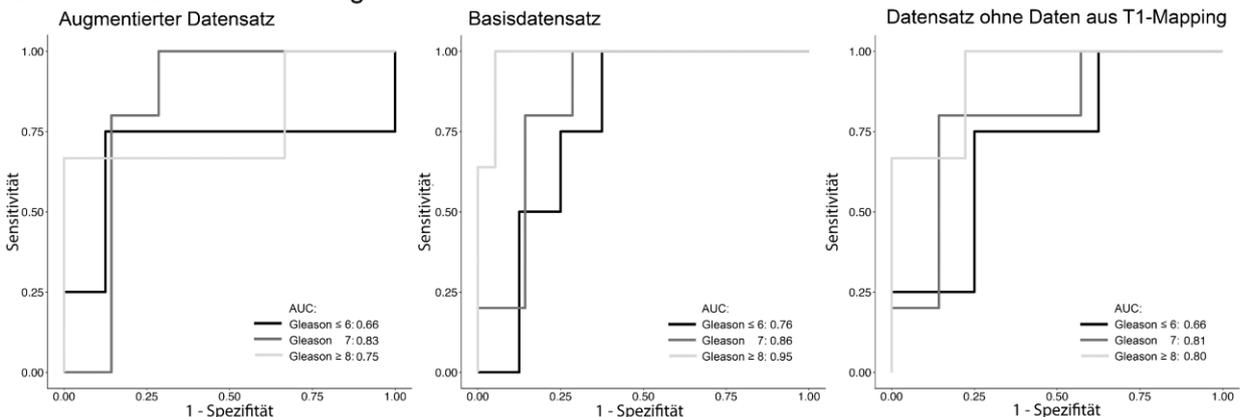
Die vier verschiedenen Machine-Learning-Algorithmen wurden mit Hilfe des Trainingsdatensatzes trainiert, um Vorhersagemodelle zur Prognose des Gleason-Scores zu entwickeln. Die Vorhersagemodelle wurden anschließend durch den Testdatensatz evaluiert. Eine Übersicht der Receiver-Operating-Characteristics-Kurven (ROC-Kurven) und der jeweiligen Flächen unter der Kurve (Englisch: Area Under the Curve, AUC) ist Abbildung 6 zu entnehmen. Der Grenzwert wurde durch die Maximierung der Summe aus Sensitivität und Spezifität festgelegt.

Die Vorhersagemodelle, welche mit den augmentierten Bildern trainiert wurden, zeigten bei allen vier Algorithmen eine bessere Leistung als die Basismodelle, die mit den Originalbildern trainiert wurden. Ein Ausschluss der Features des T1-Mappings aus dem Datensatz ging bei dem Random Forest und dem Stochastic Gradient Boosting mit einer abnehmenden Genauigkeit einher, während die Genauigkeit bei den L2 Regularized Support Vector Machines und dem K-Nearest Neighbors unverändert blieb. Die insgesamt beste Leistung wurde durch den L2 Regularized Support Vector Machines gefolgt von dem Random Forest, Stochastic Gradient Boosting und K-Nearest Neighbors erzielt.

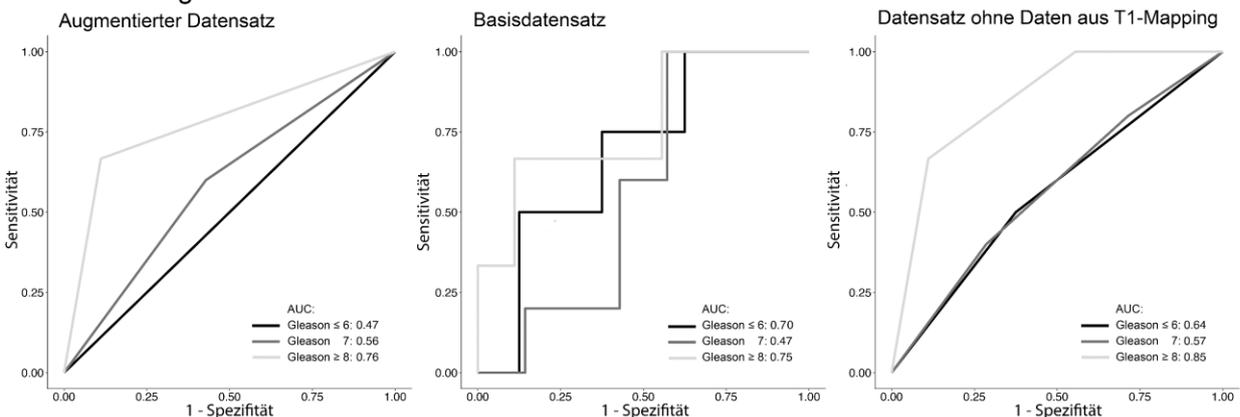
## Random Forest



## Stochastic Gradient Boosting



## k-Nearest Neighbors



**Abbildung 6:** Receiver-Operating-Characteristics-Kurven (ROC-Kurven) und die jeweiligen Flächen unter der Kurve (Englisch: Area Under the Curve, AUC) für die Algorithmen Random Forest (RF), Stochastic Gradient Boosting (SGB) und k-Nearest Neighbors (KNN). Die beste Leistung zeigten die L2 Regularized Support Vector Machines (SVM), jedoch konnten für die SVM keine ROC-Kurven erstellt werden, da sie keine Wahrscheinlichkeiten für das Vorliegen eines Klassifikators berechnen. Ein Vergleich der AUC der anderen drei Algorithmen zeigte, dass nach dem SVM der RF gefolgt vom SGB und KNN die beste Leistung erbrachte.

Adapted with permission from Wolters Kluwer Health, Inc.: [Makowski MR, Bresslem KK, Franz L, Kader A, Niehues SM, Keller S, Rueckert D, Adams LC, De Novo Radiomics Approach Using Image Augmentation and Features From T1 Mapping to Predict Gleason Scores in Prostate Cancer, *Investigative Radiology*, [https://journals.lww.com/investigativeradiology/Abstract/9000/De\\_Novo\\_Radiomics\\_Approach\\_Using\\_Image.98703.aspx](https://journals.lww.com/investigativeradiology/Abstract/9000/De_Novo_Radiomics_Approach_Using_Image.98703.aspx)]. Wolters Kluwer Health, Inc. and its Societies take no responsibility for the accuracy of the translation from the published English original and are not liable for any errors which may occur.

#### 4.2.1. Random Forest

Bei dem Random Forest Algorithmus wurden Modelle mit unterschiedlich vielen Entscheidungsbäumen angewandt. Die Anzahl der Entscheidungsbäume wurde zwischen fünf und 30 variiert. Mit zunehmender Anzahl an Entscheidungsbäumen bis zu 10 - 15 Bäumen stieg die Genauigkeit der Vorhersagemodelle. Mit weiterhin zunehmender Anzahl an Bäumen fiel die Genauigkeit jedoch wieder ab. Ein Vergleich der Leistung der drei verschiedenen Modelle (basierend auf dem Basisdatensatz, dem augmentierten Datensatz und dem Datensatz ohne Features aus dem T1-Mapping) ist Tabelle 5 zu entnehmen. Das Random Forest Modell mit der höchsten Genauigkeit wurde durch den Datensatz der augmentierten Bilder generiert. Für dieses betrug die Genauigkeit 0,83 (95 % KI 0,52 - 0,98). Die Sensitivität und Spezifität für die Klassifizierung eines Gleason-Scores von 6 betrugen 0,75 bzw. 0,88, für einen Gleason-Score von 7 lagen sie bei 1,00 bzw. 0,86 und für einen Gleason-Score von  $\geq 8$  bei 0,67 bzw. 1,00. Zudem wurde die Balanced Accuracy bestimmt, welche die Genauigkeit zwischen positiven und negativen Ergebnissen getrennt betrachtet und die Ergebnisse mittelt. Diese lag bei 0,81 für einen Gleason-Score von 6, 0,93 für einen Gleason-Score von 7 und 0,83 für einen Gleason-Score von  $\geq 8$ . Das auf dem Basisdatensatz basierende Modell sowie das Modell, bei dem die Features aus dem T1-Mapping ausgeschlossen wurden, erzielten beide eine Genauigkeit von 0,67 (95 % KI: 0,35 - 0,90).

**Tabelle 5:** Leistung der Vorhersagemodelle des Random Forest Algorithmus

	<b>Gleason-Score = 6</b>	<b>Gleason-Score = 7</b>	<b>Gleason-Score <math>\geq 8</math></b>
<b>Basisdatensatz</b>			
Sensitivität	0,50	1,00	0,33
Spezifität	0,88	0,57	1,00
Balanced Accuracy	0,69	0,79	0,67
Genauigkeit	0,67 (95 % KI: 0,35 - 0,90)		
<b>Augmentierter Datensatz</b>			
Sensitivität	0,75	1,00	0,67
Spezifität	0,88	0,86	1,00
Balanced Accuracy	0,81	0,93	0,83

Genauigkeit	0,83 (95 % KI: 0,52 - 0,98)		
<b>Ausschluss von T1-Mapping</b>			
Sensitivität	0,75	0,60	0,67
Spezifität	0,62	0,86	1,00
Balanced Accuracy	0,69	0,73	0,83
Genauigkeit	0,67 (95 % KI: 0,35 - 0,90)		

Das auf dem augmentierten Datensatz basierende Modell erzielte eine höhere Genauigkeit als das Basismodell. Ein Ausschluss der Features des T1-Mappings aus dem Datensatz bewirkte eine Abnahme der Genauigkeit. KI = Konfidenzintervall.

#### 4.2.2. Stochastic Gradient Boosting

Bei der Anwendung des Stochastic Gradient Boosting Algorithmus wurde zur Hyperparameteroptimierung eine Rastersuche durchgeführt. Das Raster wurde durch eine Tiefe der Entscheidungsbäume zwischen 1 und 13, eine Anzahl der Boosting-Wiederholungen zwischen 100 und 1250, eine Lernrate von 0,1, 0,2 oder 0,3 und eine Knotengröße von 10 definiert. Die Vorhersagemodelle mit der besten Leistung wurden durch eine mittlere Tiefe der Entscheidungsbäume, eine Anzahl an Boosting-Wiederholungen von 300 und kleinere Lernraten erzielt.

**Tabelle 6:** Leistung der Vorhersagemodelle des Stochastic Gradient Boosting Algorithmus

	<b>Gleason-Score = 6</b>	<b>Gleason-Score = 7</b>	<b>Gleason-Score ≥ 8</b>
<b>Basisdatensatz</b>			
Sensitivität	0,50	0,80	0,67
Spezifität	0,75	0,71	1,00
Balanced Accuracy	0,62	0,76	0,83
Genauigkeit	0,67 (95 % KI: 0,21 - 0,79)		
<b>Augmentierter Datensatz</b>			
Sensitivität	0,75	0,80	0,67
Spezifität	0,75	0,86	1,00
Balanced Accuracy	0,75	0,83	0,83
Genauigkeit	0,75 (95 % KI: 0,43 - 0,95)		
<b>Ausschluss von T1-Mapping</b>			

Sensitivität	0,25	0,80	0,33
Spezifität	0,75	0,43	1,00
Balanced Accuracy	0,50	0,61	0,67
Genauigkeit	0,50 (95 % KI: 0,21 - 0,79)		

*Das auf dem augmentierten Datensatz basierende Modell erzielte eine höhere Genauigkeit als das Basismodell. Ein Ausschluss der Features des T1-Mappings aus dem Datensatz bewirkte eine Abnahme der Genauigkeit. KI = Konfidenzintervall.*

Das Stochastic Gradient Boosting Modell mit der höchsten Genauigkeit wurde ebenfalls durch den Datensatz der augmentierten Bilder generiert, bei dem die Genauigkeit 0,75 (95 % KI 0,43 - 0,95) betrug (s. Tabelle 6). Hierbei lagen die Sensitivität und Spezifität für die Klassifizierung eines Gleason-Scores von 6 beide bei 0,75, für einen Gleason-Score von 7 bei 0,80 bzw. 0,86 und für einen Gleason-Score von  $\geq 8$  bei 0,67 bzw. 1,00. Die Balanced Accuracy betrug 0,75 für einen Gleason-Score von 6, 0,83 für einen Gleason-Score von 7 und 0,83 für einen Gleason-Score von  $\geq 8$ . Das auf dem Basisdatensatz basierende Modell wies mit einer Genauigkeit von 0,67 (95 % KI: 0,21 - 0,79) eine schlechtere Leistung auf als das auf dem augmentierten Datensatz basierende Modell. Bei einem Ausschluss der Features aus dem T1-Mapping aus dem Datensatz verschlechterte sich die Genauigkeit zu 0,50 (95 % KI: 0,21 - 0,79).

#### 4.2.3. L2 Regularized Support Vector Machines

Es wurden sowohl L1 Support Vector Machines als auch L2 Support Vector Machines genutzt, welche sich lediglich durch einen kleinen Unterschied in der Zielfunktion zur Optimierung der SVM unterscheiden. Die L2-Regulierung wirkte sich stärker auf eine Verbesserung der Genauigkeit der SVM aus. Außerdem wurden Gewichtungen hinzugefügt, welche Ausreißer weniger stark gewichteten als essenzielle Datenpunkte. Jedoch hatte die Gewichtung keinen erheblichen Einfluss auf die Klassifizierungsleistung.

Die Vorhersagemodelle der SVM erzielten insgesamt die beste Leistung (s. Tabelle 7). Das auf dem augmentierten Datensatz basierende Modell wies eine hohe Genauigkeit von 0,92 (95 % KI: 0,62 - 1,00) auf, wobei ein Ausschluss der Features des T1-Mappings keinen Einfluss auf die Leistung des Modells hatte. Die Genauigkeit des Basismodells lag bei 0,58 (95 % KI: 0,28 - 0,86).

Sowohl bei dem Modell des augmentierten Datensatzes als auch bei dem Modell, das auf dem Datensatz ohne Features aus dem T1-Mapping basierte, wurde lediglich eins der zwölf Prostatakarzinome aus dem Testdatensatz falsch klassifiziert. Auch die Sensitivität und Spezifität war bei den zuletzt genannten Modellen hoch und lag bei 1,00 bzw. 0,88 für die Klassifizierung eines Gleason-Scores von 6, bei 1,00 für beide Werte für die Klassifizierung eines Gleason-Scores von 7 und bei 0,67 bzw. 1,00 für die Klassifizierung eines Gleason-Scores von  $\geq 8$ . Die Balanced Accuracy betrug 0,94 für einen Gleason-Score von 6, 1,00 für einen Gleason-Score von 7 und 0,83 für einen Gleason-Score von  $\geq 8$ .

**Table 7:** Leistung der Vorhersagemodelle des L2 Regularized Support Vector Machines Algorithmus

	<b>Gleason-Score = 6</b>	<b>Gleason-Score = 7</b>	<b>Gleason-Score <math>\geq 8</math></b>
<b>Basisdatensatz</b>			
Sensitivität	1,00	0,40	0,33
Spezifität	0,62	0,71	1,00
Balanced Accuracy	0,81	0,56	0,67
Genauigkeit	0,58 (95 % KI: 0,28 - 0,86)		
<b>Augmentierter Datensatz</b>			
Sensitivität	1,00	1,00	0,67
Spezifität	0,88	1,00	1,00
Balanced Accuracy	0,94	1,00	0,83
Genauigkeit	0,92 (95 % KI: 0,62 - 1,00)		
<b>Ausschluss von T1-Mapping</b>			
Sensitivität	1,00	1,00	0,67
Spezifität	0,88	1,00	1,00
Balanced Accuracy	0,94	1,00	0,83
Genauigkeit	0,92 (95 % KI: 0,62 - 1,00)		

Das auf dem augmentierten Datensatz basierende Modell erzielte eine höhere Genauigkeit als das Basismodell. Ein Ausschluss der Features des T1-Mappings aus dem Datensatz bewirkte keinen Unterschied in der Leistung. KI = Konfidenzintervall.

#### 4.2.4. K-Nearest Neighbors

Beim Trainieren des K-Nearest Neighbors Algorithmus wurde die Anzahl an Nachbarn zwischen 4 und 20 und die Distanz zwischen 1 und 5 variiert. Dabei führte eine kürzere Distanz zu Modellen mit höherer Genauigkeit. Die Modifizierung der Anzahl an Nachbarn resultierte in keiner maßgeblichen Erhöhung der Genauigkeit.

Insgesamt war die Leistung der Vorhersagemodelle des K-Nearest Neighbors die schlechteste der vier Algorithmen (s. Tabelle 8). Wie auch bei den L2 Regularized Support Vector Machines wies das durch den augmentierten Datensatz trainierte Vorhersagemodell die höchste Genauigkeit auf, wobei der Ausschluss von Features des T1-Mappings keinen Einfluss hatte. Hierbei betrug die Genauigkeit 0,50 (95 % KI: 0,21 - 0,79), während sie bei dem Basismodell nur 0,33 (95 % KI: 0,10 - 0,65) betrug.

**Tabelle 8:** Leistung der Vorhersagemodelle des K-Nearest Neighbors Algorithmus

	<b>Gleason-Score = 6</b>	<b>Gleason-Score = 7</b>	<b>Gleason-Score ≥ 8</b>
<b>Basisdatensatz</b>			
Sensitivität	0,25	0,20	0,67
Spezifität	0,50	0,57	0,89
Balanced Accuracy	0,38	0,39	0,78
Genauigkeit	0,33 (95 % KI: 0,10 - 0,65)		
<b>Augmentierter Datensatz</b>			
Sensitivität	0,25	0,60	0,67
Spezifität	0,75	0,57	0,89
Balanced Accuracy	0,50	0,59	0,78
Genauigkeit	0,50 (95 % KI: 0,21 - 0,79)		
<b>Ausschluss von T1-Mapping</b>			
Sensitivität	0,50	0,40	0,67
Spezifität	0,62	0,71	0,89
Balanced Accuracy	0,56	0,56	0,78
Genauigkeit	0,50 (95 % KI: 0,21 - 0,79)		

Die höchste Genauigkeit wurde durch das auf dem augmentierten Datensatz basierende Modell erzielt. Ein Ausschluss der Features des T1-Mappings aus dem Datensatz bewirkte keinen Unterschied in der Genauigkeit. KI = Konfidenzintervall.

## 5. Diskussion

### 5.1. Interpretation der Ergebnisse

In dieser retrospektiven Studie wurde gezeigt, dass das Prostatakarzinom durch Radiomics in der biparametrischen MRT (ADC und T2WI) und im T1-Mapping zuverlässig detektiert werden kann. Insbesondere können mit Hilfe von Radiomics spezifische Merkmale vom Tumorgraden in den quantitativen Bildfeatures identifiziert werden, um mit einer Genauigkeit von bis zu 92 % zwischen Prostatakarzinomen mit einem Gleason-Score von 6, 7 und  $\geq 8$  zu unterscheiden.

Zudem wurde in der Studie der mögliche Nutzen von Features aus dem T1-Mapping für Radiomics-Analysen von Prostatakarzinomen untersucht. Features des T1-Mappings sind anders als Features der konventionellen MRT quantifizierbar (55), sodass sie möglicherweise die Genauigkeit von Radiomics-Modellen steigern könnten. Die Ergebnisse der Studie zeigen, dass durch Hinzufügen von Features aus dem T1-Mapping eine Zunahme der Genauigkeit bei zwei von vier Vorhersagemodellen zu verzeichnen war. Dazu gehörten der RF (+ 0,16) und der SGB (+ 0,25), wohingegen die Genauigkeit der anderen zwei Vorhersagemodelle (KNN, SVM) unverändert blieb. Eine mögliche Erklärung für den ausbleibenden Effekt der Features aus dem T1-Mapping auf die Leistung des SVM- und KNN-Modells könnte die niedrige Fallzahl des Trainingsdatensatzes sein. Der potenzielle Nutzen von Features aus dem T1-Mapping sollte deshalb in Studien mit größeren Fallzahlen genauer untersucht werden.

Die Radiomics-Analyse in der vorliegenden Arbeit stellt einen neuen methodischen Ansatz dar, da im Zuge der Extraktion der Features die Bilddaten durch zufällige Modifikation von Helligkeit und/oder Kontrast, Hinzufügen von Gauß'schem Rauschen und perspektivische Verzerrung augmentiert wurden. Unseres Wissens wurde bislang keine vergleichbare Studie durchgeführt und publiziert. Das Augmentieren von Bilddatensätzen ist eine gängige Methode des Machine-Learning, um die Genauigkeit des Vorhersagemodells zu verbessern und Overfitting zu vermeiden (117). Bei allen vier Algorithmen führte eine zehnfache Augmentation zu einer Erhöhung der Genauigkeit um durchschnittlich 19 % sowie der Balanced Accuracy des Vorhersagemodells. Die Ergebnisse unserer Studie zeigen somit, dass das Augmentieren von Bilddaten die

Leistung von Vorhersagemodellen des Gleason-Scores beim Prostatakarzinom steigern könnte.

## 5.2. Vergleich mit vorliegender Literatur

Analysen von hochdimensionalen *omics*-Datensätzen zur Identifikation von molekularen Biomarkern werden zunehmend durchgeführt und wenden dabei unterschiedliche Studiendesigns an. In unserer Studie wurde die insgesamt beste Leistung durch den Algorithmus der SVM gefolgt von RF, SGB und KNN erzielt. Dies stimmt mit den Ergebnissen einer Studie überein, in der die Leistung von vier verschiedenen Klassifizierungsalgorithmen zur Analyse von *omics*-Datensätzen miteinander verglichen wurde (137). Dort wurde beobachtet, dass bei einer großen Anzahl an Features, die die Anzahl an Patienten übersteigt, durch SVM und RF eine höhere Genauigkeit als durch KNN erreicht wird (137). Zudem wurde festgestellt, dass SVM gefolgt von RF weniger anfällig für eine unausgewogene Klassenverteilung als KNN ist (137). Zwar wurde in unserer Studie eine ausgeglichene Klassenverteilung angestrebt, jedoch war die Klassenverteilung mit 29,6 % PCa-Patienten mit einem Gleason-Score von 6, 44,5 % mit einem Gleason-Score von 7 und 25,9 % mit einem Gleason-Score von  $\geq 8$  nicht gänzlich ausbalanciert, wobei der SVM davon möglicherweise am wenigsten beeinflusst wurde.

Während die meisten Studien zu Radiomics-Analysen des Prostatakarzinoms der Unterscheidung maligner Neubildungen von gesundem Prostatagewebe dienen (96, 99, 138-140), wurden auch einige Studien durchgeführt, um den Nutzen von Radiomics-Analysen für die Klassifizierung des Gleason-Scores von Prostatakarzinomen zu untersuchen:

*Fehr et al.* entwickelten und analysierten ein Radiomics-Modell anhand von 147 PCa-Patienten, in welchem Textur-Features aus bpMRT unter Anwendung zwei verschiedener Augmentierungsmethoden extrahiert wurden (98). Ziel war die Unterscheidung zwischen Prostatakarzinomen und gesundem Prostatagewebe, zwischen einem GS von 6 und einem GS von  $\geq 7$  sowie zwischen einem GS von 7 (3 + 4) und einem GS von 7 (4 + 3). In der Studienpopulation waren vor allem Patienten mit hochgradigen Prostatakarzinomen vertreten. Jedoch können bei stark unausgewogenen Klassenverteilungen viele Algorithmen eine geringere Vorhersagegenauigkeit für die seltene Klasse haben (141), da alle Fälle in die am häufigsten vertretene Gruppe

klassifiziert werden und so zwar eine gute Genauigkeit erreicht wird, dies aber mit Einbußen in der Spezifität und/oder Sensitivität einher geht (98). Um die Minderheitsklasse zu vergrößern, nutzten *Fehr et al.* zwei verschiedene Augmentierungsmethoden: Für eines der Vorhersagemodelle wurden die Daten durch die Synthetic-Minority-Oversampling-Technik (SMOTE) augmentiert. Die SMOTE erhöht in einem unausgeglichenen Datensatz die Anzahl der unterrepräsentierten Fälle, indem „synthetische“ Datenpunkte innerhalb des Bereichs eines Feature-Vektors aus der Minderheitsklasse und seiner  $k$  nächsten Nachbarn geschaffen werden (142). Allerdings besteht bei der SMOTE die Möglichkeit, dass die Anzahl der unterrepräsentierten Fälle nicht ausreicht, um die genaue Wahrscheinlichkeitsverteilung für die tatsächlichen Daten zu schätzen (143).

Für ein anderes Vorhersagemodell wandten *Fehr et al.* Gibbs-Sampling an (144). Bei diesem werden die Werte von Features verschiedener Patienten miteinander kombiniert, indem aus der bedingten Wahrscheinlichkeitsverteilung eines Features bei einem bestimmten Wert eines anderen Features neue Datenpunkte gebildet werden (145). Jedoch werden durch Gibbs-Sampling die Werte der Features nicht verändert, sondern neu kombiniert, sodass potenzielle Bias der Bilder fortbestehen. Im Gegenteil dazu wurden in unserer Augmentierungsmethode sowohl Bias durch Unterschiede in der Helligkeit und im Kontrast als auch Bias durch Gauß'sches Rauschen und Bewegungsartefakte simuliert. Zudem wurde bei der Erstellung der Studienpopulation darauf geachtet, eine annähernd ausgeglichene Verteilung auf die einzelnen Gruppen (GS 6, GS 7,  $GS \geq 8$ ) zu erhalten und so den Störfaktor einer unausgewogenen Klassenverteilung auf die Leistung der Vorhersagemodelle zu minimieren.

In der Studie von *Fehr et al.* betrug die Genauigkeit zur Unterscheidung zwischen einem GS von 6 und einem  $GS \geq 7$  je nach Lokalisation des Prostatakarzinoms 92 - 93 %. Die gleiche Genauigkeit wurde bei der Unterscheidung zwischen einem GS von 7 (3 + 4) und 7 (4 + 3) erreicht. Somit war die Genauigkeit ähnlich hoch wie in unserer Radiomics-Analyse. Jedoch wurde die Leistung der Vorhersagemodelle durch ein Kreuzvalidierungsverfahren mit den augmentierten Bildern anstatt durch einen unabhängigen Testdatensatz ermittelt. Im Gegenteil dazu wurde in unserer Studie zuerst eine Kreuzvalidierung mit 15 % des Originaldatensatzes durchgeführt, um bei der Validierung potenzielle Bias durch augmentierte Daten zu vermeiden. Im Vergleich zur Kreuzvalidierung ist die Validierung durch einen unabhängigen Testdatensatz die

idealere Methode zur Ermittlung der Genauigkeit eines Vorhersagemodells (98), weshalb in unserer Studie die Daten zu Beginn in einen Trainings- (80 %) und einen Testdatensatz (20 %) geteilt wurden, um im Anschluss an das Training der Algorithmen die Vorhersagemodelle durch den unabhängigen Testdatensatz zu evaluieren. Für eine bessere Einschätzung der Leistung der Vorhersagemodelle von *Fehr et al.* wäre eine Evaluierung durch einen unabhängigen Testdatensatz von Interesse gewesen.

Zur Selektion von Features gibt es verschiedene Methoden, von denen sich bislang keine als Standardmethode durchgesetzt hat. Grundsätzlich kann zwischen Filtermethoden vor Anwendung eines Machine-Learning-Algorithmus und in Methoden, die in den Machine-Learning-Algorithmus integrierte sind, unterschieden werden (146). *Fehr et al.* nutzten zur Selektion der Features und zur Klassifizierung Recursive Feature Selection Support Vector Machines (RFE-SVM), bei welchen in jedem Durchlauf Features mit dem geringsten Einfluss auf eine Leistungsverbesserung des Vorhersagemodells rückwärts eliminiert werden (147). Bei der Rückwärtselimination wird anfangs ein Modell mit allen Features gebildet, von welchem nach und nach Features ausgeschlossen werden, bis die verbliebenen Features alle einen signifikanten Einfluss auf das Outcome haben (148). Ein Nachteil bei der Rückwärtselimination ist jedoch, dass ein eliminiertes Feature im weiteren Verlauf nicht wieder in Betracht gezogen wird, obwohl es im finalen Modell eine größere Signifikanz haben könnte (148).

Im Gegenteil dazu wurden in unserer Studie Features vor dem Training der Algorithmen eliminiert, indem für die Features der Originalbilder des Trainingsdatensatzes eine Korrelationsmatrix erstellt wurde und Features mit einem Korrelationskoeffizienten  $r > 0,75$  aus der Radiomics-Analyse ausgeschlossen wurden. Die optimale Methode zur Selektion von Features wird kontrovers diskutiert, wobei die a priori Exklusion von stark korrelierenden Features, die durch die Korrelation keine oder wenige zusätzliche Informationen beitragen, eine Möglichkeit ist (149). Als Alternative zur Rückwärtselimination gibt es die Herangehensweise des Full-Model-Approach, bei welcher alle vorselektierten Features in das Modell einbezogen werden. Diese haben wir in unserer Studie angewandt, wobei mit den vorselektierten Features vier verschiedenen Algorithmen trainiert wurden. Die Methode des Full-Model-Approach kann die Vorteile haben, Overfitting und Bias durch Selektion von Features zu vermeiden, jedoch wird sie in vielen Fällen nicht angewandt, da eine große Anzahl an Features die Arbeit mit dem Modell erschwert und das Modell oft schwer definierbar ist (149). In den von uns

genutzten Algorithmen SVM, RF und SGB ist die Selektion von Features integriert, indem Variablen bestraft werden (SVM) oder in Bäumen entsprechend ihrer Wichtigkeit angeordnet werden (RF, SGB) (146). Bei KNN basiert die Vorhersage ausschließlich auf dem Abstand zwischen den Datenpunkten, jedoch kann die Gewichtung von Features durch eine zusätzliche Selektionsmethode von extern vorgegeben werden (150).

Die generierten Vorhersagemodelle von *Fehr et al.* sind binär und ermöglichen lediglich eine Unterscheidung in einen GS von 6 und einen  $GS \geq 7$  oder aber eine Unterscheidung zwischen einem GS von 7 (3 + 4) und 7 (4 + 3). Jedoch wurde keine Genauigkeit für die Vorhersage der einzelnen Gruppen bestimmt. In unserer Studie wurden ebenfalls binäre Klassifikatoren angewandt. Allerdings wurde anschließend eine Unterscheidung aller Gruppen voneinander ermöglicht, indem eine Vorhersagematrix erstellt und die maximale Wahrscheinlichkeit bestimmt wurde. Dadurch konnten die jeweiligen Genauigkeiten für einen GS von 6, einen GS von 7 und einen  $GS \geq 8$  berechnet werden.

*Chaddad et al.* untersuchten in einer Studie von 99 PCa-Patienten ein neuartiges Modell zur Vorhersage des Gleason-Scores, welches auf der Joint Intensity Matrix (JIM) basierte (151). Konventionelle, durch die Grauwertematrix definierte Textur-basierte-Features können nur die Textur innerhalb eines Bildes, nicht aber die Textur mehrerer verschiedener Sequenzen (z.B. T2WI und DWI) charakterisieren. Um die Textur der bpMRT zu untersuchen, kamen *Chaddad et al.* mit einer JIM auf. Während die Grauwertematrix die gemeinsame Wahrscheinlichkeit angibt, gleiche Intensitätswerte in zwei verschiedenen Voxels in einem festgelegten Abstand in eine bestimmte Richtung zu beobachten, gibt JIM die gemeinsame Intensitätsverteilung von 3D-Bildern verschiedener Modalitäten (z.B. T2WI und DWI) an.

Ein binäres Vorhersagemodell, welches zwischen einem  $GS = 6$  vs.  $GS > 6$ , einem  $GS = 3 + 4$  vs.  $GS \neq 3 + 4$  sowie einem  $GS \geq 4 + 3$  vs.  $GS < 4 + 3$  unterscheiden sollte, wurde mit Hilfe eines Random Forest Algorithmus erstellt. Der Random Forest kann bei einem begrenzten Trainingsdatensatz gute Ergebnisse erzielen und bietet zudem die Möglichkeit der Berechnung von Wichtigkeitsmaßen, welche die Wichtigkeit eines einzelnen Features für die Vorhersage des Gleason-Scores messen (152). So nutzten *Chaddad et al.* den Random Forest nicht nur zur Generierung eines Vorhersagemodells, sondern auch zur Analyse prädiktiver Features. In unserer Studie wurden keine

prädiktiven Features analysiert, wobei dies vor allem für das T1-Mapping von Interesse ist und in zukünftigen Studien untersucht werden sollte.

*Chaddad et al.* nutzten einen ausgeglichenen Trainings- und Testdatensatz, um Störfaktoren einer unausgewogenen Klassenverteilung auf die Leistung der Vorhersagemodelle zu verhindern. Jedoch war die ursprüngliche Studienpopulation in Bezug auf die Gleason-Score-Gruppen unausgeglichen, sodass die Generierung eines ausgeglichenen Trainingsdatensatzes zu einer Verringerung des Datenumfangs auf 40 Patienten führte. Zudem kann willkürliches Undersampling, welches den Vorgang der zufälligen Wahl von Daten aus der Mehrheitsklasse bis zum Erreichen eines Gleichgewichts mit der Minderheitsklasse beschreibt, zum Verlust von Informationen aus der Mehrheitsklasse führen (153). In unserer Studie war die Klassenverteilung annähernd ausgewogen und die Stichprobengröße umfangreicher, da alle vorhandenen Daten für den Trainings- bzw. Testdatensatz genutzt wurden.

In der Studie von *Chaddad et al.* wurde die beste Leistung durch eine Kombination von Features basierend auf der JIM und der Grauwertematrix erhalten. Dabei betragen die AUCs für die einzelnen Gruppen mit 65 - 78 % etwas weniger als die AUCs unserer Studie. Zudem wurde auch in dieser Studie keine Genauigkeit für die Vorhersage der einzelnen Gruppen, sondern die Genauigkeiten der binären Vorhersagemodelle bestimmt.

Sowohl *Fehr et al.* als auch *Chaddad et al.* schlossen Features, die anfällig für Abweichungen bei der manuellen Segmentation sein könnten, aus ihrer Studie aus (98, 151). Dazu gehörten Features wie Volumen, Form, Festigkeit, Konvexität und Exzentrizität. In unserer Studie wurden Shape-based-Features in die Radiomics-Analyse eingeschlossen, da sie wichtige Informationen über die Aggressivität des Prostatakarzinoms liefern könnten (154). Dabei wurden die manuell erstellten Annotationen in mehreren Durchläufen von erfahrenen Radiolog\*innen revidiert.

In anderen Radiomics-Analysen (54, 155) wurde nach abgeschlossener manueller Segmentation der PCa's die Genauigkeit der Annotationen evaluiert. Dafür wurden die MRT-Bilder mehrmals durch die gleiche untersuchende Person sowie durch weitere Untersuchende annotiert (155). Anschließend wurde die Robustheit von Features aus den Annotationen durch Bestimmung von Korrelationskoeffizienten untersucht und Features mit einem Korrelationskoeffizienten unter einem bestimmten Grenzwert wurden

aus der Radiomics-Analyse ausgeschlossen (155). Hierfür ist die Bestimmung des Intraklassen-Korrelationskoeffizienten eine gängige Methode, dessen Wert die Stabilität des Features auf einer Skala von 0 bis 1 angibt (156). Eine weitere Möglichkeit zur Bewertung von Unterschieden in der Annotation verschiedener Untersuchenden ist der Dice-Koeffizient, welcher ebenfalls Werte von 0 bis 1 annimmt und die räumliche Überlappung angibt (156). Jedoch gibt es trotz der bekannten Variabilität bei der manuellen Segmentation bislang keine etablierten Standards zur Untersuchung der Robustheit von Features.

In einigen Radiomics-Modellen wurde logistische Regression zum Erstellen von Vorhersagemodellen genutzt (54, 155). *Gong et al.* generierten ein Radiomics-Modell zur Unterscheidung hochgradiger Prostatakarzinome (definiert als  $GS > 7$ ) von niedriggradigen Prostatakarzinomen (definiert als  $GS \leq 7$ ) (155). Dafür wurden First-order-Features, Shape-based-Features und Matrix-based-Features aus bpMRT von 489 PCa-Patienten extrahiert. Anschließend wurden zur Selektion prognostischer Features drei Schritte zur Eliminierung von redundanten Features durchlaufen. Zuerst wurde der Kruskal-Wallis-Test zur Identifikation von Features, die mit den festgelegten Grenzen des Gleason-Scores signifikant korrelieren, durchgeführt. Anschließend wurde die Minimum-Redundancy-Maximum-Relevance-Methode angewandt. Dabei handelt es sich um einen Algorithmus zur systematischen Auswahl von Features mit der höchsten Relevanz und der niedrigsten Redundanz für ein Klassifizierungsproblem (157). Zuletzt erfolgte die Selektion von Features und die Klassifizierung durch logistische Regression und sequenzielle Rückwärtselimination. Wie oben beschrieben weist die Rückwärtseliminierung einige Nachteile auf. So erzielte das Modell der bpMRT eine Genauigkeit und Balanced Accuracy von 70 %. Die vergleichsweise niedrige Genauigkeit könnte dadurch erklärt werden, dass bei der logistischen Regression im Gegenteil zu komplexeren Machine-Learning-Algorithmen ein linearer Zusammenhang der unabhängigen Variablen mit dem natürlichen Logarithmus des Odd Ratio angenommen wird (158) und dadurch nichtlineare Zusammenhänge von Daten nicht erfasst werden.

In der Literatur werden zur Gruppierung von Gleason-Scores mit der Annahme einer ähnlichen Prognose verschiedene Grenzen gewählt. In Radiomics-Analysen erschwert dies einen aussagekräftigen Vergleich zwischen den Studien. *Chaddad et al.* wählten eine Unterteilung in  $GS = 6$ ,  $GS = 3 + 4$  und  $GS \geq 4 + 3$  (151). Sie unterschieden somit innerhalb eines GS von 7 zwischen den beiden Möglichkeiten  $3 + 4$  und  $4 + 3$ , was dem

auf der ISUP Konsensuskonferenz von 2014 festgelegten Grading entspricht (3). Dabei stellt ein GS von 3 + 4 einen ISUP-Grad 2 und ein GS von 4 + 3 einen ISUP-Grad 3 dar. Eine Unterscheidung zwischen ISUP-Grad 3 und ISUP-Grad 4 ist sinnvoll, weil Studien eine signifikant schlechtere Prognose für letzteren nachgewiesen haben (159-164). Jedoch umfasst die Gruppe eines  $GS \geq 4 + 3$  sowohl Patienten mit einem mittelhohen Risiko (4 + 3) (30) als auch Patienten mit einer großen Wahrscheinlichkeit für einen Tod durch das Prostatakarzinom (5 + 5) (3). Demgegenüber nutzten *Gong et al.* eine Gruppierung in  $GS \leq 7$  und  $GS > 7$  (155), was der Unterteilung entspricht, die auch in der Prostate Cancer Outcomes Study (PCOS) angewandt wurde (165). Durch eine Unterscheidung in lediglich zwei Gruppen ist die der Datensatzes pro Gruppe umfangreicher, allerdings werden auch in diesem Fall Patienten mit sehr unterschiedlichen Prognosen miteinander gruppiert, sodass die Aussagekraft des Vorhersagemodells sinkt.

*Fehr et al.* nutzten eine Klassifizierung in einen GS von 6 vs. einen  $GS \geq 7$  und eine weitere Klassifizierung in einen GS von 7 (3 + 4) vs. einen GS von 7 (4 + 3) (98). Es wurde also zwischen ISUP-Grad 3 und ISUP-Grad 4 unterschieden. Jedoch stellen Prostatakarzinome mit einem  $GS \geq 8$  eine Limitation dieser Gruppierung dar, da sie im ersten Schritt als  $GS \geq 7$  und im zweiten Schritt fälschlicherweise als  $GS = 7$  (3 + 4) oder  $GS = 7$  (4 + 3) klassifiziert werden. Somit muss vor Durchlaufen des zweiten Schritts bekannt sein, dass es sich um ein Prostatakarzinom mit einem GS von 7 handelt.

In unserer Studie wurde entsprechend vorliegender Literatur der Scandinavian Prostate Cancer Group (166) zwischen einem GS von 6, einem GS von 7 und einem  $GS \geq 8$  unterschieden. Diese Einteilung richtet sich nach der Risikostratifizierung nach d'Amico, welche eine niedrige (GS 2 - 6), mittlere (GS = 7) und hohe Risikogruppe (GS 8 – 10) unterscheidet (30), und den risikoabhängigen Therapieempfehlungen. Jedoch wurde in unserer Radiomics-Analyse nicht in einen GS von 7 (3 + 4) und 7 (4 + 3) unterschieden, da dies zu einer geringen Datenmenge und zu einer unausgeglichene Datensatz geführt hätte.

### **5.3. Limitationen**

Eine wichtige Limitation der Arbeit ist die kleine Studienpopulation ( $n = 66$ ), welche bei der Aufteilung in den Trainings- und Testdatensatz zu kleinen Fallzahlen in den einzelnen

Fallgruppen eines Gleason-Scores von 6, 7 und  $\geq 8$  führte. Auf Grund der kleinen Fallzahlen müssen die Ergebnisse der Machine-Learning-Methoden mit Vorsicht interpretiert werden und sind schwer generalisierbar. Außerdem ist dadurch keine Effektquantifizierung möglich. In der vorliegenden Arbeit wurde allerdings bereits ein unabhängiger Testdatensatz verwendet, um eine mögliche Generalisierbarkeit auf einen neuen Datensatz zu zeigen. Eine weitere Limitation ist der retrospektive Charakter, auf Grund dessen in einigen Fällen nicht alle klinischen Daten für die deskriptive Analyse vorlagen. In Zukunft sollten prospektive Studien mit höheren Fallzahlen durchgeführt werden, um generalisierbare Ergebnisse zu gewinnen. Hinzu kommt, dass die Bilder in einer einzigen Institution und von nur zwei verschiedenen Scannern akquiriert wurden. Für eine bessere Generalisierbarkeit sollten multizentrische Studien unter Einschluss mehrerer Scanner durchgeführt werden.

Die histopathologische Beurteilung der Gewebeproben zur Bestimmung des Gleason-Scores hängt von der Qualität der Gewebeprobe ab. Bei der histopathologischen Beurteilung steht bei einer Biopsieentnahme deutlich weniger Gewebe zur Verfügung als bei einer Prostatektomie, sodass Gleason-Scores nach einer Prostatektomie häufig nach oben korrigiert werden (84). Somit könnten die in dieser Studie verwendeten Gleason-Scores zu niedrig eingestuft worden sein. Allerdings muss in vielen Fällen auf den in der Biopsie bestimmten Gleason-Score zurückgegriffen werden, da vor allem Patienten mit niedriggradigen Prostatakarzinomen keine Prostatektomie erhalten.

Die manuelle Segmentation der Prostatakarzinome und der zonalen Anatomie der Prostata kann der Subjektivität der Untersuchenden unterliegen. Die manuelle Segmentation könnte vor allem einen Einfluss auf Shape-based-Features haben, welche die Morphologie des Tumors wie das Volumen und die Form beschreiben. In anderen Studien wurden diese Features ausgeschlossen, um Bias durch die manuelle Segmentation zu vermeiden (98, 151). In unserer Studie wurden die Annotationen mehrfach durch erfahrene Radiolog\*innen revidiert, um Ungenauigkeiten bei der Annotation zu minimieren. In zukünftigen Studien könnte eine (teil)automatisierte Segmentation, z.B. eine computergestützte Kantenerkennung mit anschließender manueller Segmentation, angewandt werden.

#### 5.4. Schlussfolgerung und Ausblick

Bei der Diagnose des Prostatakarzinoms können Modelle, die durch die Vorhersage des Gleason-Scores den Grad des Prostatakarzinoms bestimmen können, ein wertvolles diagnostisches und prognostisches Tool darstellen. Sie haben das Potenzial, unnötige Biopsien und deren Nebenwirkungen zu verhindern und bei der Entscheidungsfindung der geeigneten Therapieoption zu helfen, die von Watchful Waiting bei niedriggradigen Prostatakarzinomen bis zur radikalen Prostatektomie bei höhergradigen Prostatakarzinomen reichen kann. In der bpMRT und im T1-Mapping sind quantitative Features enthalten, die Informationen über die Aggressivität eines Tumors beinhalten können. Während diese in der Befundung durch Radiolog\*innen nicht erfasst werden können, ermöglichen Radiomics-Analysen die Extraktion, Selektion und Analyse quantitativer Features.

In dieser Arbeit wurde der Nutzen einer neuartigen Radiomics-Analyse von MRT-Datensätzen zur Diagnostik des Prostatakarzinoms untersucht, welche Datenaugmentation und quantitative Features aus dem T1-Mapping zusätzlich zu Features aus der bpMRT verwendet. Abschließend konnte gezeigt werden, dass Vorhersagemodelle, die auf den augmentierten Bilddaten der bpMRT basieren, eine Klassifizierung in drei verschiedene Gruppen von Gleason-Scores (6, 7 und  $\geq 8$ ) ermöglichen. Die Nutzung von Features aus dem T1-Mapping könnte die Genauigkeit von Vorhersagemodellen durch Radiomics-Analysen erhöhen und sollte in prospektiven, multizentrischen Studien genauer evaluiert werden.

Die Annotation von radiologischen Bildern ist in vielen Fällen mit einem hohen Kosten- und Zeitaufwand verbunden und benötigt die Beurteilung durch Expert\*innen. Gleichzeitig ist die statistische Power des Vorhersagemodells maßgeblich von der Größe des verwendeten Datensatzes abhängig (69). Vor diesem Hintergrund kann das Augmentieren von Bilddaten bei der Durchführung von Radiomics-Analysen von großem Nutzen sein, da die Menge der Daten vervielfacht werden kann und dies zur Erhöhung der Genauigkeit von Vorhersagemodellen beitragen könnte. Eine weitere Möglichkeit für die Generierung größerer Datensätze ist das Teilen von Daten innerhalb multiinstitutionaler, nationaler oder internationaler Zusammenschlüsse wie das Cancer Imaging Archive, einer Open-Access-Datenbank medizinischer Bilder. Die dort zur Verfügung gestellten Datensätze könnten zudem zur Nutzung als unabhängiger

Testdatensatz dienen, um die Scanner-übergreifende und Zentren-übergreifende Generalisierbarkeit von Radiomics-Analysen zu evaluieren.

In dieser Studie wurden Vorhersagemodelle generiert, die den Gleason-Score des Tumors als Ganzes betreffen. Jedoch ist es zukünftig von Interesse, Modelle zu erstellen und zu untersuchen, die eine voxelweise Vorhersage des Gleason-Scores und somit eine Generierung von Gleason-Score-Maps ermöglichen. Mit Hilfe dieser Maps könnten Aussagen über Teilgebiete des Tumors getroffen werden und so unter anderem gezielte Biopsien ermöglicht werden (98). In verschiedenen Studien wurden bereits voxelweise Vorhersagemodelle erstellt (139, 167, 168). So generierten *Viswanath et al.* aus T2WI ein Modell zur Detektion und anschließenden voxelweisen Klassifikation von Prostatakarzinomen (167). Dabei wurde die Ausdehnung des Prostatakarzinoms automatisch von Prostatektomiepräparaten auf die entsprechenden MRT-Schichten übertragen. Eine voxelweise Bestimmung des Gleason-Scores auf Grundlage von bpMRT und T1-Mapping wurde jedoch bislang nicht untersucht.

Studienergebnisse über den potenziellen Nutzen der T1-Relaxationszeit zur Bestimmung des Tumorgrads von Prostatakarzinomen und zur Unterscheidung des Prostatakarzinoms von gesundem Gewebe sind heterogen und unterscheiden sich vor allem in Abhängigkeit von der Lokalisation des Prostatakarzinoms in der peripheren oder Transitionalzone (32, 33, 66, 67). Auch bei Features der bmMRT könnte die Aussagekraft verschiedener Features je nach Lokalisation des Prostatakarzinoms variieren (102, 169). In unserer Radiomics-Analyse wurde die zonale Verteilung aufgrund der kleinen Fallzahl von Prostatakarzinomen in der Transitionalzone ( $n = 10$ ) nicht berücksichtigt, sodass dies für die Durchführung zukünftiger Studien von Interesse bleibt.

Abschließend ist zu betonen, dass Ergebnisse aus Radiomics-Analysen nicht isoliert, sondern im Zusammenspiel mit anderen *-omics*-Daten und klinischen Daten evaluiert werden sollten. *Qi et al.* führten eine Radiomics-Studie durch, in der ein klinisch-radiologisches Vorhersagemodell, ein Radiomics-Vorhersagemodell und eine Kombination der beiden genannten Modelle zur Vorhersage eines Prostatakarzinoms miteinander verglichen wurden (99). Die Radiomics-Analyse basierte auf T2WI, DCE und DWI von 199 PCa-Patienten. Das kombinierte Modell umfasste die Radiomics-Daten, klinische Daten wie das Alter und die PSA-Dichte sowie radiologische Daten wie den PI-RADS-Score. Es erzielte im Vergleich zu den anderen beiden Modellen die höchste

Genauigkeit (93 %) zur Unterscheidung eines Prostatakarzinoms von gesundem Prostatagewebe. Die Studie könnte darauf hinweisen, dass die Kombination unserer Radiomics-Analyse mit klinischen und radiologischen Daten die Leistung des Vorhersagemodells weiter steigern könnte.

## 6. Literaturverzeichnis

1. Lee HH, Shaw NM, Mohammed S, Kowalczyk KJ, Stamatakis L, Krasnow RE. Prostate Cancer in Men With Treated Advanced Heart Failure: Should we Keep Screening? *Urology*. 2020;136:46-50.
2. Etzioni R, Gulati R, Falcon S, Penson DF. Impact of PSA screening on the incidence of advanced stage prostate cancer in the United States: a surveillance modeling approach. *Med Decis Making*. 2008;28(3):323-31.
3. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol*. 2016;40(2):244-52.
4. World Health Organization. Global Cancer Observatory. Lyon: World Health Organization; 2020 [08.01.2021]. Available from: <https://gco.iarc.fr/today/home>.
5. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424.
6. Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, Dicker DJ, Chimed-Orchir O, Dandona R, Dandona L, Fleming T, Forouzanfar MH, Hancock J, Hay RJ, Hunter-Merrill R, Huynh C, Hosgood HD, Johnson CO, Jonas JB, Khubchandani J, Kumar GA, Kutz M, Lan Q, Larson HJ, Liang X, Lim SS, Lopez AD, MacIntyre MF, Marczak L, Marquez N, Mokdad AH, Pinho C, Pourmalek F, Salomon JA, Sanabria JR, Sandar L, Sartorius B, Schwartz SM, Shackelford KA, Shibuya K, Stanaway J, Steiner C, Sun J, Takahashi K, Vollset SE, Vos T, Wagner JA, Wang H, Westerman R, Zeeb H, Zoeckler L, Abd-Allah F, Ahmed MB, Alabed S, Alam NK, Aldhahri SF, Alem G, Alemayohu MA, Ali R, Al-Raddadi R, Amare A, Amoako Y, Artaman A, Asayesh H, Atnafu N, Awasthi A, Saleem HB, Barac A, Bedi N, Bensenor I, Berhane A, Bernabé E, Betsu B, Binagwaho A, Boneya D, Campos-Nonato I, Castañeda-Orjuela C, Catalá-López F, Chiang P, Chibueze C, Chittheer A, Choi JY, Cowie B, Damtew S, das Neves J, Dey S, Dharmaratne S, Dhillon P, Ding E, Driscoll T, Ekwueme D, Endries AY, Farvid M, Farzadfar F, Fernandes J, Fischer F, TT GH, Gebru A, Gopalani S, Hailu A, Horino M, Horita N, Hussein A, Huybrechts I, Inoue M, Islami F, Jakovljevic M, James S, Javanbakht M, Jee SH, Kasaeian A, Kedir MS, Khader YS, Khang YH, Kim D, Leigh J, Linn S, Lunevicius R, El Razek HMA, Malekzadeh R, Malta DC, Marcenes W, Markos D, Melaku YA, Meles KG, Mendoza W, Mengiste DT, Meretoja TJ, Miller TR, Mohammad KA, Mohammadi A, Mohammed S, Moradi-Lakeh M, Nagel G, Nand D, Le Nguyen Q, Nolte S, Ogbo FA, Oladimeji KE, Oren E, Pa M, Park EK, Pereira DM, Plass D, Qorbani M, Radfar A, Rafay A, Rahman M, Rana SM, Søreide K, Satpathy M, Sawhney M, Sepanlou SG, Shaikh MA, She J, Shiue I, Shore HR, Shrimme MG, So S, Soneji S, Stathopoulou V, Stroumpoulis K, Sufiyan MB, Sykes BL, Tabarés-Seisdedos R, Tadese F, Tedla BA, Tessema GA, Thakur JS, Tran BX, Ukwaja KN, Uzochukwu BSC, Vlassov VV, Weiderpass E, Wubshet Terefe M, Yeboyo HG, Yimam HH, Yonemoto N, Younis MZ,

Yu C, Zaidi Z, Zaki MES, Zenebe ZM, Murray CJL, Naghavi M. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol.* 2017;3(4):524-48.

7. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA. SEER Cancer Statistics Review, 1975-2017. Bethesda, MD: National Cancer Institute; [08.01.2021]. Available from: [https://seer.cancer.gov/csr/1975\\_2017/](https://seer.cancer.gov/csr/1975_2017/).

8. Pernar CH, Ebot EM, Wilson KM, Mucci LA. The Epidemiology of Prostate Cancer. *Cold Spring Harb Perspect Med.* 2018;8(12).

9. Shimizu H, Ross RK, Bernstein L, Yatani R, Henderson BE, Mack TM. Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. *Br J Cancer.* 1991;63(6):963-6.

10. Yu H, Harris RE, Gao Y-T, Gao R, Wynder EL. Comparative Epidemiology of Cancers of the Colon, Rectum, Prostate and Breast in Shanghai, China versus the United States. *International Journal of Epidemiology.* 1991;20(1):76-81.

11. Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, Ghousaini M, Luccarini C, Dennis J, Jugurnauth-Little S, Dadaev T, Neal DE, Hamdy FC, Donovan JL, Muir K, Giles GG, Severi G, Wiklund F, Gronberg H, Haiman CA, Schumacher F, Henderson BE, Le Marchand L, Lindstrom S, Kraft P, Hunter DJ, Gapstur S, Chanock SJ, Berndt SI, Albanes D, Andriole G, Schleutker J, Weischer M, Canzian F, Riboli E, Key TJ, Travis RC, Campa D, Ingles SA, John EM, Hayes RB, Pharoah PD, Pashayan N, Khaw KT, Stanford JL, Ostrander EA, Signorello LB, Thibodeau SN, Schaid D, Maier C, Vogel W, Kibel AS, Cybulski C, Lubinski J, Cannon-Albright L, Brenner H, Park JY, Kaneva R, Batra J, Spurdle AB, Clements JA, Teixeira MR, Dicks E, Lee A, Dunning AM, Baynes C, Conroy D, Maranian MJ, Ahmed S, Govindasami K, Guy M, Wilkinson RA, Sawyer EJ, Morgan A, Dearnaley DP, Horwich A, Huddart RA, Khoo VS, Parker CC, Van As NJ, Woodhouse CJ, Thompson A, Dudderidge T, Ogden C, Cooper CS, Lophatananon A, Cox A, Southey MC, Hopper JL, English DR, Aly M, Adolfsson J, Xu J, Zheng SL, Yeager M, Kaaks R, Diver WR, Gaudet MM, Stern MC, Corral R, Joshi AD, Shahabi A, Wahlfors T, Tammela TL, Auvinen A, Virtamo J, Klarskov P, Nordestgaard BG, Røder MA, Nielsen SF, Bojesen SE, Siddiq A, Fitzgerald LM, Kolb S, Kwon EM, Karyadi DM, Blot WJ, Zheng W, Cai Q, McDonnell SK, Rinckleb AE, Drake B, Colditz G, Wokolorczyk D, Stephenson RA, Teerlink C, Muller H, Rothenbacher D, Sellers TA, Lin HY, Slavov C, Mitev V, Lose F, Srinivasan S, Maia S, Paulo P, Lange E, Cooney KA, Antoniou AC, Vincent D, Bacot F, Tessier DC, Kote-Jarai Z, Easton DF. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet.* 2013;45(4):385-91, 91e1-2.

12. Al Olama AA, Kote-Jarai Z, Berndt SI, Conti DV, Schumacher F, Han Y, Benlloch S, Hazelett DJ, Wang Z, Saunders E, Leongamornlert D, Lindstrom S, Jugurnauth-Little S, Dadaev T, Tymrakiewicz M, Stram DO, Rand K, Wan P, Stram A, Sheng X, Pooler LC,

Park K, Xia L, Tyrer J, Kolonel LN, Le Marchand L, Hoover RN, Machiela MJ, Yeager M, Burdette L, Chung CC, Hutchinson A, Yu K, Goh C, Ahmed M, Govindasami K, Guy M, Tammela TL, Auvinen A, Wahlfors T, Schleutker J, Visakorpi T, Leinonen KA, Xu J, Aly M, Donovan J, Travis RC, Key TJ, Siddiq A, Canzian F, Khaw KT, Takahashi A, Kubo M, Pharoah P, Pashayan N, Weischer M, Nordestgaard BG, Nielsen SF, Klarskov P, Røder MA, Iversen P, Thibodeau SN, McDonnell SK, Schaid DJ, Stanford JL, Kolb S, Holt S, Knudsen B, Coll AH, Gapstur SM, Diver WR, Stevens VL, Maier C, Luedeke M, Herkommer K, Rinckleb AE, Strom SS, Pettaway C, Yeboah ED, Tettey Y, Biritwum RB, Adjei AA, Tay E, Truelove A, Niwa S, Chokkalingam AP, Cannon-Albright L, Cybulski C, Wokołorczyk D, Kluźniak W, Park J, Sellers T, Lin HY, Isaacs WB, Partin AW, Brenner H, Dieffenbach AK, Stegmaier C, Chen C, Giovannucci EL, Ma J, Stampfer M, Penney KL, Mucci L, John EM, Ingles SA, Kittles RA, Murphy AB, Pandha H, Michael A, Kierzek AM, Blot W, Signorello LB, Zheng W, Albanes D, Virtamo J, Weinstein S, Nemesure B, Carpten J, Leske C, Wu SY, Hennis A, Kibel AS, Rybicki BA, Neslund-Dudas C, Hsing AW, Chu L, Goodman PJ, Klein EA, Zheng SL, Batra J, Clements J, Spurdle A, Teixeira MR, Paulo P, Maia S, Slavov C, Kaneva R, Mitev V, Witte JS, Casey G, Gillanders EM, Seminara D, Riboli E, Hamdy FC, Coetzee GA, Li Q, Freedman ML, Hunter DJ, Muir K, Gronberg H, Neal DE, Southey M, Giles GG, Severi G, Cook MB, Nakagawa H, Wiklund F, Kraft P, Chanock SJ, Henderson BE, Easton DF, Eeles RA, Haiman CA. A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet.* 2014;46(10):1103-9.

13. MacInnis RJ, English DR. Body size and composition and prostate cancer risk: systematic review and meta-regression analysis. *Cancer Causes Control.* 2006;17(8):989-1003.

14. Leitlinienprogramm Onkologie. Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms, Kurzversion 5.1: Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF); 2019 [17.01.2021]. Available from: <https://www.leitlinienprogramm-onkologie.de/leitlinien/prostatakarzinom/>.

15. Catalona WJ, Richie JP, deKernion JB, Ahmann FR, Ratliff TL, Dalkin BL, Kavoussi LR, MacFarlane MT, Southwick PC. Comparison of prostate specific antigen concentration versus prostate specific antigen density in the early detection of prostate cancer: receiver operating characteristic curves. *J Urol.* 1994;152(6 Pt 1):2031-6.

16. Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS, Parnes HL, Minasian LM, Ford LG, Lippman SM, Crawford ED, Crowley JJ, Coltman CA, Jr. Prevalence of prostate cancer among men with a prostate-specific antigen level  $\leq$  4.0 ng per milliliter. *N Engl J Med.* 2004;350(22):2239-46.

17. Ilic D, Neuberger MM, Djulbegovic M, Dahm P. Screening for prostate cancer. *Cochrane Database Syst Rev.* 2013(1):Cd004720.

18. Hugosson J, Roobol MJ, Månsson M, Tammela TLJ, Zappa M, Nelen V, Kwiatkowski M, Lujan M, Carlsson SV, Talala KM, Lilja H, Denis LJ, Recker F, Paez A,

Puliti D, Villers A, Rebillard X, Kilpeläinen TP, Stenman UH, Godtman RA, Stinesen Kollberg K, Moss SM, Kujala P, Taari K, Huber A, van der Kwast T, Heijnsdijk EA, Bangma C, De Koning HJ, Schröder FH, Auvinen A. A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. *Eur Urol.* 2019;76(1):43-51.

19. Gulati R, Mariotto AB, Chen S, Gore JL, Etzioni R. Long-term projections of the harm-benefit trade-off in prostate cancer screening are more favorable than previous short-term estimates. *J Clin Epidemiol.* 2011;64(12):1412-7.

20. Loeb S, Vonesh EF, Metter EJ, Carter HB, Gann PH, Catalona WJ. What is the true number needed to screen and treat to save a life with prostate-specific antigen testing? *J Clin Oncol.* 2011;29(4):464-7.

21. Wilt TJ, Brawer MK, Jones KM, Barry MJ, Aronson WJ, Fox S, Gingrich JR, Wei JT, Gilhooly P, Grob BM, Nsouli I, Iyer P, Cartagena R, Snider G, Roehrborn C, Sharifi R, Blank W, Pandya P, Andriole GL, Culkin D, Wheeler T. Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med.* 2012;367(3):203-13.

22. Hayes JH, Barry MJ. Screening for prostate cancer with the prostate-specific antigen test: a review of current evidence. *Jama.* 2014;311(11):1143-9.

23. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, Fossati N, Gross T, Henry AM, Joniau S, Lam TB, Mason MD, Matveev VB, Moldovan PC, van den Bergh RCN, Van den Broeck T, van der Poel HG, van der Kwast TH, Rouviere O, Schoots IG, Wiegel T, Cornford P. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment with Curative Intent. *Eur Urol.* 2017;71(4):618-29.

24. Carroll PH, Mohler JL. NCCN Guidelines Updates: Prostate Cancer and Prostate Cancer Early Detection. *J Natl Compr Canc Netw.* 2018;16(5s):620-3.

25. Litwin MS, Tan HJ. The Diagnosis and Treatment of Prostate Cancer: A Review. *Jama.* 2017;317(24):2532-42.

26. Schoots IG, Roobol MJ, Nieboer D, Bangma CH, Steyerberg EW, Hunink MG. Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. *Eur Urol.* 2015;68(3):438-50.

27. Monni F, Fontanella P, Grasso A, Wiklund P, Ou YC, Randazzo M, Rocco B, Montanari E, Bianchi G. Magnetic resonance imaging in prostate cancer detection and management: a systematic review. *Minerva Urol Nefrol.* 2017;69(6):567-78.

28. Haider MA, Yao X, Loblaw A, Finelli A. Multiparametric Magnetic Resonance Imaging in the Diagnosis of Prostate Cancer: A Systematic Review. *Clin Oncol (R Coll Radiol)*. 2016;28(9):550-67.
29. Hassan O, Han M, Zhou A, Paulk A, Sun Y, Al-Harbi A, Alrajjal A, Baptista Dos Santos F, Epstein JI. Incidence of Extraprostatic Extension at Radical Prostatectomy with Pure Gleason Score 3 + 3 = 6 (Grade Group 1) Cancer: Implications for Whether Gleason Score 6 Prostate Cancer Should be Renamed "Not Cancer" and for Selection Criteria for Active Surveillance. *J Urol*. 2018;199(6):1482-7.
30. Hernandez DJ, Nielsen ME, Han M, Partin AW. Contemporary evaluation of the D'amico risk classification of prostate cancer. *Urology*. 2007;70(5):931-5.
31. Hamm B, Asbach P. Magnetic Resonance Imaging of the Prostate in the PI-RADS Era. In: Hodler J, Kubik-Huch RA, von Schulthess GK, editors. *Diseases of the Abdomen and Pelvis 2018-2021: Diagnostic Imaging - IDKD Book*. Cham (CH): Springer Copyright 2018, The Author(s). 2018. p. 99-115.
32. Panda A, Obmann VC, Lo WC, Margevicius S, Jiang Y, Schluchter M, Patel IJ, Nakamoto D, Badve C, Griswold MA, Jaeger I, Ponsky LE, Gulani V. MR Fingerprinting and ADC Mapping for Characterization of Lesions in the Transition Zone of the Prostate Gland. *Radiology*. 2019;292(3):685-94.
33. Baur ADJ, Hansen CM, Rogasch J, Posch H, Elezkurtaj S, Maxeiner A, Erb-Eigner K, Makowski MR. Evaluation of T1 relaxation time in prostate cancer and benign prostate tissue using a Modified Look-Locker inversion recovery sequence. *Sci Rep*. 2020;10(1):3121.
34. Barentsz JO, Richenberg J, Clements R, Choyke P, Verma S, Villeirs G, Rouviere O, Logager V, Fütterer JJ. ESUR prostate MR guidelines 2012. *Eur Radiol*. 2012;22(4):746-57.
35. Akin O, Sala E, Moskowitz CS, Kuroiwa K, Ishill NM, Pucar D, Scardino PT, Hricak H. Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging. *Radiology*. 2006;239(3):784-92.
36. Wang L, Mazaheri Y, Zhang J, Ishill NM, Kuroiwa K, Hricak H. Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with Gleason grade after radical prostatectomy. *Radiology*. 2008;246(1):168-76.
37. Röhke M, Blondin D, Schlemmer H. PI-RADS-Klassifikation: Strukturiertes Befundungsschema für die MRT der Prostata. *RöFo*. 2013;185(3):253-61.

38. Zelhof B, Pickles M, Liney G, Gibbs P, Rodrigues G, Kraus S, Turnbull L. Correlation of diffusion-weighted magnetic resonance data with cellularity in prostate cancer. *BJU Int.* 2009;103(7):883-8.
39. Sato C, Naganawa S, Nakamura T, Kumada H, Miura S, Takizawa O, Ishigaki T. Differentiation of noncancerous tissue and cancer lesions by apparent diffusion coefficient values in transition and peripheral zones of the prostate. *J Magn Reson Imaging.* 2005;21(3):258-62.
40. Surov A, Meyer HJ, Wienke A. Correlations between Apparent Diffusion Coefficient and Gleason Score in Prostate Cancer: A Systematic Review. *Eur Urol Oncol.* 2020;3(4):489-97.
41. Meyer HJ, Wienke A, Surov A. Discrimination between clinical significant and insignificant prostate cancer with apparent diffusion coefficient - a systematic review and meta analysis. *BMC Cancer.* 2020;20(1):482.
42. Shaish H, Kang SK, Rosenkrantz AB. The utility of quantitative ADC values for differentiating high-risk from low-risk prostate cancer: a systematic review and meta-analysis. *Abdom Radiol (NY).* 2017;42(1):260-70.
43. Wu LM, Xu JR, Ye YQ, Lu Q, Hu JN. The clinical value of diffusion-weighted imaging in combination with T2-weighted imaging in diagnosing prostate carcinoma: a systematic review and meta-analysis. *AJR Am J Roentgenol.* 2012;199(1):103-10.
44. Franiel T, Hamm B, Hricak H. Dynamic contrast-enhanced magnetic resonance imaging and pharmacokinetic models in prostate cancer. *Eur Radiol.* 2011;21(3):616-26.
45. Engelbrecht MR, Huisman HJ, Laheij RJ, Jager GJ, van Leenders GJ, Hulsbergen-Van De Kaa CA, de la Rosette JJ, Blickman JG, Barentsz JO. Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using dynamic contrast-enhanced MR imaging. *Radiology.* 2003;229(1):248-54.
46. Sciarra A, Panebianco V, Ciccariello M, Salciccia S, Cattarino S, Lisi D, Gentilucci A, Alfarone A, Bernardo S, Passariello R, Gentile V. Value of magnetic resonance spectroscopy imaging and dynamic contrast-enhanced imaging for detecting prostate cancer foci in men with prior negative biopsy. *Clin Cancer Res.* 2010;16(6):1875-83.
47. Boesen L, Nørgaard N, Løgager V, Balslev I, Bisbjerg R, Thestrup KC, Winther MD, Jakobsen H, Thomsen HS. Assessment of the Diagnostic Accuracy of Biparametric Magnetic Resonance Imaging for Prostate Cancer in Biopsy-Naive Men: The Biparametric MRI for Detection of Prostate Cancer (BIDOC) Study. *JAMA Netw Open.* 2018;1(2):e180219.

48. American College of Radiology. ARC Manual on Contrast Media 2021 [01.04.2021]. Available from: [https://www.acr.org/-/media/ACR/Files/Clinical-Resources/Contrast\\_Media.pdf](https://www.acr.org/-/media/ACR/Files/Clinical-Resources/Contrast_Media.pdf).
49. Bass EJ, Pantovic A, Connor M, Gabe R, Padhani AR, Rockall A, Sokhi H, Tam H, Winkler M, Ahmed HU. A systematic review and meta-analysis of the diagnostic accuracy of biparametric prostate MRI for prostate cancer in men at risk. *Prostate Cancer Prostatic Dis.* 2020.
50. Hassanzadeh E, Glazer DI, Dunne RM, Fennessy FM, Harisinghani MG, Tempany CM. Prostate imaging reporting and data system version 2 (PI-RADS v2): a pictorial review. *Abdom Radiol (NY).* 2017;42(1):278-89.
51. Drost FH, Osses D, Nieboer D, Bangma CH, Steyerberg EW, Roobol MJ, Schoots IG. Prostate Magnetic Resonance Imaging, with or Without Magnetic Resonance Imaging-targeted Biopsy, and Systematic Biopsy for Detecting Prostate Cancer: A Cochrane Systematic Review and Meta-analysis. *Eur Urol.* 2020;77(1):78-94.
52. Woo S, Suh CH, Kim SY, Cho JY, Kim SH, Moon MH. Head-to-Head Comparison Between Biparametric and Multiparametric MRI for the Diagnosis of Prostate Cancer: A Systematic Review and Meta-Analysis. *AJR Am J Roentgenol.* 2018;211(5):W226-w41.
53. Fütterer JJ, Briganti A, De Visschere P, Emberton M, Giannarini G, Kirkham A, Taneja SS, Thoeny H, Villeirs G, Villers A. Can Clinically Significant Prostate Cancer Be Detected with Multiparametric Magnetic Resonance Imaging? A Systematic Review of the Literature. *Eur Urol.* 2015;68(6):1045-53.
54. Zhang GM, Han YQ, Wei JW, Qi YF, Gu DS, Lei J, Yan WG, Xiao Y, Xue HD, Feng F, Sun H, Jin ZY, Tian J. Radiomics Based on MRI as a Biomarker to Guide Therapy by Predicting Upgrading of Prostate Cancer From Biopsy to Radical Prostatectomy. *J Magn Reson Imaging.* 2020;52(4):1239-48.
55. Taylor AJ, Salerno M, Dharmakumar R, Jerosch-Herold M. T1 Mapping: Basic Techniques and Clinical Applications. *JACC Cardiovasc Imaging.* 2016;9(1):67-81.
56. Moon JC, Messroghli DR, Kellman P, Piechnik SK, Robson MD, Ugander M, Gatehouse PD, Arai AE, Friedrich MG, Neubauer S, Schulz-Menger J, Schelbert EB. Myocardial T1 mapping and extracellular volume quantification: a Society for Cardiovascular Magnetic Resonance (SCMR) and CMR Working Group of the European Society of Cardiology consensus statement. *J Cardiovasc Magn Reson.* 2013;15(1):92.
57. Nacif MS, Turkbey EB, Gai N, Nazarian S, van der Geest RJ, Noureldin RA, Sibley CT, Ugander M, Liu S, Arai AE, Lima JA, Bluemke DA. Myocardial T1 mapping with MRI: comparison of look-locker and MOLLI sequences. *J Magn Reson Imaging.* 2011;34(6):1367-73.

58. Messroghli DR, Radjenovic A, Kozerke S, Higgins DM, Sivananthan MU, Ridgway JP. Modified Look-Locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. *Magn Reson Med*. 2004;52(1):141-6.
59. Messroghli DR, Plein S, Higgins DM, Walters K, Jones TR, Ridgway JP, Sivananthan MU. Human myocardium: single-breath-hold MR T1 mapping with high spatial resolution--reproducibility study. *Radiology*. 2006;238(3):1004-12.
60. American College of Radiology. Prostate Imaging Reporting & Data System (PI-RADS) version 2.1 [07.04.2021]. Available from: <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/PI-RADS>.
61. Barrett T, Vargas HA, Akin O, Goldman DA, Hricak H. Value of the hemorrhage exclusion sign on T1-weighted prostate MR images for the detection of prostate cancer. *Radiology*. 2012;263(3):751-7.
62. Yu L, Sun J, Sun J, Li J, Dong Y, Zhou X, Greiser A, Han Y, Zhang Q, Xie Q, Chen Y. Early detection of myocardial involvement by T(1) mapping of cardiac MRI in idiopathic inflammatory myopathy. *J Magn Reson Imaging*. 2018;48(2):415-22.
63. Fontana M, Banypersad SM, Treibel TA, Maestrini V, Sado DM, White SK, Pica S, Castelletti S, Piechnik SK, Robson MD, Gilbertson JA, Rowczenio D, Hutt DF, Lachmann HJ, Wechalekar AD, Whelan CJ, Gillmore JD, Hawkins PN, Moon JC. Native T1 mapping in transthyretin amyloidosis. *JACC Cardiovasc Imaging*. 2014;7(2):157-65.
64. Kjaer L, Thomsen C, Iversen P, Henriksen O. In vivo estimation of relaxation processes in benign hyperplasia and carcinoma of the prostate gland by magnetic resonance imaging. *Magn Reson Imaging*. 1987;5(1):23-30.
65. Foltz WD, Haider MA, Chung P, Bayley A, Catton C, Ramanan V, Jaffray D, Wright GA, Ménard C. Prostate T(1) quantification using a magnetization-prepared spiral technique. *J Magn Reson Imaging*. 2011;33(2):474-81.
66. Yu AC, Badve C, Ponsky LE, Pahwa S, Dastmalchian S, Rogers M, Jiang Y, Margevicius S, Schluchter M, Tabayoyong W, Abouassaly R, McGivney D, Griswold MA, Gulani V. Development of a Combined MR Fingerprinting and Diffusion Examination for Prostate Cancer. *Radiology*. 2017;283(3):729-38.
67. Panda A, O'Connor G, Lo WC, Jiang Y, Margevicius S, Schluchter M, Ponsky LE, Gulani V. Targeted Biopsy Validation of Peripheral Zone Prostate Cancer Characterization With Magnetic Resonance Fingerprinting and Diffusion Mapping. *Invest Radiol*. 2019;54(8):485-93.

68. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-6.

69. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-77.

70. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical T, Board on Health Care S, Board on Health Sciences P, Institute of M. Evolution of Translational Omics: Lessons Learned and the Path Forward. In: Micheel CM, Nass SJ, Omenn GS, editors. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington (DC): National Academies Press (US)

Copyright 2012 by the National Academy of Sciences. All rights reserved.; 2012.

71. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.

72. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA, Velculescu VE, Kinzler KW, Vogelstein B, Iacobuzio-Donahue CA. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*. 2010;467(7319):1114-7.

73. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012;366(10):883-92.

74. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, Curtis C, Watts C, Tavaré S. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A*. 2013;110(10):4009-14.

75. Van Meter T, Dumur C, Hafez N, Garrett C, Fillmore H, Broaddus WC. Microarray analysis of MRI-defined tissue samples in glioblastoma reveals differences in regional expression of therapeutic targets. *Diagn Mol Pathol*. 2006;15(4):195-205.

76. Burstein HJ, Harris LN, Marcom PK, Lambert-Falls R, Havlin K, Overmoyer B, Friedlander RJ, Jr., Gargiulo J, Strenger R, Vogel CL, Ryan PD, Ellis MJ, Nunes RA, Bunnell CA, Campos SM, Hallor M, Gelman R, Winer EP. Trastuzumab and vinorelbine as first-line therapy for HER2-overexpressing metastatic breast cancer: multicenter phase

II trial with clinical outcomes, analysis of serum tumor markers as predictive factors, and cardiac surveillance algorithm. *J Clin Oncol*. 2003;21(15):2889-95.

77. Wagle N, Emery C, Berger MF, Davis MJ, Sawyer A, Pochanard P, Kehoe SM, Johannessen CM, Macconail LE, Hahn WC, Meyerson M, Garraway LA. Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol*. 2011;29(22):3085-96.

78. Sequist LV, Martins RG, Spigel D, Grunberg SM, Spira A, Jänne PA, Joshi VA, McCollum D, Evans TL, Muzikansky A, Kuhlmann GL, Han M, Goldberg JS, Settleman J, Iafrate AJ, Engelman JA, Haber DA, Johnson BE, Lynch TJ. First-line gefitinib in patients with advanced non-small-cell lung cancer harboring somatic EGFR mutations. *J Clin Oncol*. 2008;26(15):2442-9.

79. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidias P, Bergethon K, Shaw AT, Gettinger S, Cospers AK, Akhavanfard S, Heist RS, Temel J, Christensen JG, Wain JC, Lynch TJ, Vernovsky K, Mark EJ, Lanuti M, Iafrate AJ, Mino-Kenudson M, Engelman JA. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med*. 2011;3(75):75ra26.

80. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, Balagurunathan Y, Gatenby RA, Gillies RJ. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012;30(9):1234-48.

81. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, Chan BK, Matcuk GR, Barry CT, Chang HY, Kuo MD. Decoding global gene expression programs in liver cancer by noninvasive imaging. *Nat Biotechnol*. 2007;25(6):675-80.

82. Diehn M, Nardini C, Wang DS, McGovern S, Jayaraman M, Liang Y, Aldape K, Cha S, Kuo MD. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc Natl Acad Sci U S A*. 2008;105(13):5213-8.

83. Nguyen PL, Schultz D, Renshaw AA, Vollmer RT, Welch WR, Cote K, D'Amico AV. The impact of pathology review on treatment recommendations for patients with adenocarcinoma of the prostate. *Urol Oncol*. 2004;22(4):295-9.

84. Djavan B, Kadesky K, Klopukh B, Marberger M, Roehrborn CG. Gleason scores from prostate biopsies obtained with 18-gauge biopsy needles poorly predict Gleason scores of radical prostatectomy specimens. *Eur Urol*. 1998;33(3):261-70.

85. Clarke LP, Sriram RD, Schilling LB. Imaging as a Biomarker: Standards for Change Measurements in Therapy workshop summary. *Acad Radiol*. 2008;15(4):501-30.

86. Kazerooni EA, Armstrong MR, Amorosa JK, Hernandez D, Liebscher LA, Nath H, McNitt-Gray MF, Stern EJ, Wilcox PA. ACR CT Accreditation Program and the Lung Cancer Screening Program Designation. *J Am Coll Radiol*. 2016;13(2 Suppl):R30-4.
87. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, Mendelson EB, Wilcox PA, Butler PF, D'Orsi CJ. The ACR BI-RADS experience: learning from history. *J Am Coll Radiol*. 2009;6(12):851-60.
88. Rios Velazquez E, Aerts HJ, Gu Y, Goldgof DB, De Ruysscher D, Dekker A, Korn R, Gillies RJ, Lambin P. A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol*. 2012;105(2):167-73.
89. van Dam IE, van Sörnsen de Koste JR, Hanna GG, Muirhead R, Slotman BJ, Senan S. Improving target delineation on 4-dimensional CT scans in stage I NSCLC using a deformable registration tool. *Radiother Oncol*. 2010;96(1):67-72.
90. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, Bellomi M. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp*. 2018;2(1):36.
91. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. *Clin Cancer Res*. 2015;21(2):249-57.
92. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, Ganeshan B, Miles KA, Cook GJ, Goh V. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging*. 2012;3(6):573-89.
93. Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, Goldgof DB, Hall LO, Korn R, Zhao B, Schwartz LH, Basu S, Eschrich S, Gatenby RA, Gillies RJ. Test-retest reproducibility analysis of lung CT image features. *J Digit Imaging*. 2014;27(6):805-23.
94. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255-68.
95. Carrasco JL, King TS, Chinchilli VM. The concordance correlation coefficient for repeated measures estimated by variance components. *J Biopharm Stat*. 2009;19(1):90-105.
96. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, Zheng J, Goldman D, Moskowitz C, Fine SW, Reuter VE, Eastham J, Sala E, Vargas HA. Haralick texture analysis of prostate MRI: utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason scores. *Eur Radiol*. 2015;25(10):2840-50.

97. Nketiah G, Elschot M, Kim E, Teruel JR, Scheenen TW, Bathen TF, Selnaes KM. T2-weighted MRI-derived textural features reflect prostate cancer aggressiveness: preliminary results. *Eur Radiol.* 2017;27(7):3050-9.
98. Fehr D, Veeraraghavan H, Wibmer A, Gondo T, Matsumoto K, Vargas HA, Sala E, Hricak H, Deasy JO. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proc Natl Acad Sci U S A.* 2015;112(46):E6265-73.
99. Qi Y, Zhang S, Wei J, Zhang G, Lei J, Yan W, Xiao Y, Yan S, Xue H, Feng F, Sun H, Tian J, Jin Z. Multiparametric MRI-Based Radiomics for Prostate Cancer Screening With PSA in 4-10 ng/mL to Reduce Unnecessary Biopsies. *J Magn Reson Imaging.* 2020;51(6):1890-9.
100. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, Velazquez ER, Lambin P, Gu Y, Balagurunathan Y, Eikman E, Gatenby RA, Eschrich S, Gillies RJ. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS One.* 2015;10(3):e0118261.
101. Vignati A, Mazzetti S, Giannini V, Russo F, Bollito E, Porpiglia F, Stasi M, Regge D. Texture features on T2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness. *Phys Med Biol.* 2015;60(7):2685-701.
102. Niu XK, Chen ZF, Chen L, Li J, Peng T, Li X. Clinical Application of Biparametric MRI Texture Analysis for Detection and Evaluation of High-Grade Prostate Cancer in Zone-Specific Regions. *AJR Am J Roentgenol.* 2018;210(3):549-56.
103. Kuo MD, Gollub J, Sirlin CB, Ooi C, Chen X. Radiogenomic analysis to identify imaging phenotypes associated with drug response gene expression programs in hepatocellular carcinoma. *J Vasc Interv Radiol.* 2007;18(7):821-31.
104. Teruel JR, Heldahl MG, Goa PE, Pickles M, Lundgren S, Bathen TF, Gibbs P. Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *NMR Biomed.* 2014;27(8):887-96.
105. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101(21):1446-52.
106. Hayes DF, Bast RC, Desch CE, Fritsche H, Jr., Kemeny NE, Jessup JM, Locker GY, Macdonald JS, Mennel RG, Norton L, Ravdin P, Taube S, Winn RJ. Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst.* 1996;88(20):1456-66.

107. Madabhushi A, Udupa JK. New methods of MR image intensity standardization via generalized scale. *Med Phys*. 2006;33(9):3426-34.
108. Ge Y, Udupa JK, Nyúl LG, Wei L, Grossman RI. Numerical tissue characterization in MS via standardization of the MR image intensity scale. *J Magn Reson Imaging*. 2000;12(5):715-21.
109. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging*. 2000;19(2):143-50.
110. Lee H, Palm J, Grimes SM, Ji HP. The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations. *Genome Med*. 2015;7:112.
111. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26(6):1045-57.
112. Makowski MR, Bressemer KK, Franz L, Kader A, Niehues SM, Keller S, Rueckert D, Adams LC. De Novo Radiomics Approach Using Image Augmentation and Features From T1 Mapping to Predict Gleason Scores in Prostate Cancer. *Invest Radiol*. 2021.
113. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, Margolis D, Schnall MD, Shtern F, Tempany CM, Thoeny HC, Verma S. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol*. 2016;69(1):16-40.
114. Maxeiner A, Fischer T, Schwabe J, Baur ADJ, Stephan C, Peters R, Slowinski T, von Laffert M, Marticorena Garcia SR, Hamm B, Jung EM. Contrast-Enhanced Ultrasound (CEUS) and Quantitative Perfusion Analysis in Patients with Suspicion for Prostate Cancer. *Ultraschall Med*. 2019;40(3):340-8.
115. Belaroussi B, Milles J, Carme S, Zhu YM, Benoit-Cattin H. Intensity non-uniformity correction in MRI: existing methods and their validation. *Med Image Anal*. 2006;10(2):234-46.
116. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310-20.
117. Bloice M, Stocker C, Holzinger A. Augmentor: An Image Augmentation Library for Machine Learning. *Journal of Open Source Software*. 2017;2(19):432.
118. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Köpf A, Yang E, DeVito Z, Raison M, Tejani A,

Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. 2019.

119. Howard J, Gugger S. Fastai: A Layered API for Deep Learning. *Information*. 2020;11(2):108.

120. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The Design of SimpleITK. *Front Neuroinform*. 2013;7:45.

121. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts H. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;77(21):e104-e7.

122. Moradmand H, Setayeshi S, Karimian AR, Sirous M, Akbari ME. Comparing the performance of image enhancement methods to detect microcalcification clusters in digital mammography. *Iran J Cancer Prev*. 2012;5(2):61-8.

123. Dettori L, Semler L. A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography. *Comput Biol Med*. 2007;37(4):486-98.

124. Wiltgen M, Bloice M, Koller S, Hoffmann-Wellenhof R, Smolle J, Gerger A. Computer-aided diagnosis of melanocytic skin tumors by use of confocal laser scanning microscopy images. *Anal Quant Cytol Histol*. 2011;33(2):85-100.

125. Weyn B, van de Wouwer G, van Daele A, Scheunders P, van Dyck D, van Marck E, Jacob W. Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. *Cytometry*. 1998;33(1):32-40.

126. Hilado S, Gan Lim L, Naguib R, Dadios E, Avila J. Implementation of Wavelets and Artificial Neural Networks in Colonic Histopathological Classification. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. 2014;18:792-7.

127. Chaddad A, Daniel P, Niazi T. Radiomics Evaluation of Histological Heterogeneity Using Multiscale Textures Derived From 3D Wavelet Transformation of Multispectral Images. *Front Oncol*. 2018;8:96.

128. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, Leijenaar RT, Haibe-Kains B, Lambin P, Gillies RJ, Aerts HJ. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife*. 2017;6.

129. Moradmand H, Aghamiri SMR, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. *J Appl Clin Med Phys*. 2020;21(1):179-90.

130. R Core Team. R: A language and environment for statistical computing. Wien: R Foundation for Statistical Computing; 2018.
131. Wickham H. Welcome to the tidyverse. 2019;4(43):1686.
132. Kuhn M. caret: Classification and Regression Training. Astrophysics Source Code Library; 2015. p. ascl:1505.003.
133. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12(1):77.
134. Quentin M, Schimmöller L, Hiestler A, Arsov C, Rabenalt R, Albers P, Antoch G, Blondin D. Korrelation des PI-RADS Summenscore mit dem Gleason score von gezielt MR bioptisch gesicherten Prostatakarzinomen. Rofo. 2014;186(S 01):VO402\_5.
135. Statistisches Bundesamt. Körpermaße nach Altersgruppen: Männer 2017 [26.06.2021]. Available from: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Gesundheitszustand-Relevantes-Verhalten/Tabellen/koerpermasse-maenner.html>.
136. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83.
137. Guo Y, Graber A, McBurney RN, Balasubramanian R. Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. BMC Bioinformatics. 2010;11:447.
138. Moradi M, Salcudean SE, Chang SD, Jones EC, Buchan N, Casey RG, Goldenberg SL, Kozlowski P. Multiparametric MRI maps for detection and grading of dominant prostate tumors. J Magn Reson Imaging. 2012;35(6):1403-13.
139. Madabhushi A, Feldman MD, Metaxas DN, Tomaszewski J, Chute D. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI. IEEE Trans Med Imaging. 2005;24(12):1611-25.
140. Niaf E, Rouvière O, Mège-Lechevallier F, Bratan F, Lartizien C. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Phys Med Biol. 2012;57(12):3833-51.
141. Ling CX, Sheng VS. Class Imbalance Problem. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning. Boston, MA: Springer US; 2010. p. 171-.

142. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Int Res.* 2002;16(1):321–57.
143. Nguyen GH, Bouzerdoum A, Phung SL, editors. *Learning Pattern Classification Tasks with Imbalanced Data Sets* 2009.
144. Resnik P, Hardisty E. *Gibbs sampling for the uninitiated.* 2009.
145. van Ravenzwaaij D, Cassey P, Brown SD. A simple introduction to Markov Chain Monte-Carlo sampling. *Psychon Bull Rev.* 2018;25(1):143-54.
146. Sun P, Wang D, Mok VC, Shi L. Comparison of Feature Selection Methods and Machine Learning Classifiers for Radiomics Analysis in Glioma Grading. *IEEE Access.* 2019;7:102010-20.
147. Rakotomamonjy A. Variable selection using svm based criteria. *J Mach Learn Res.* 2003;3(null):1357–70.
148. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. *Fam Med Community Health.* 2020;8(1):e000262.
149. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *Bmj.* 2009;338:b604.
150. Li S, Harner EJ, Adjero DA. Random KNN feature selection - a fast and stable alternative to Random Forests. *BMC Bioinformatics.* 2011;12(1):450.
151. Chaddad A, Kucharczyk MJ, Niazi T. Multimodal Radiomic Features for the Predicting Gleason Score of Prostate Cancer. *Cancers (Basel).* 2018;10(8).
152. Archer KJ, Kimes RV. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis.* 2008;52(4):2249-60.
153. Lin W-C, Tsai C-F, Hu Y-H, Jhang J-S. Clustering-based undersampling in class-imbalanced data. *Information Sciences.* 2017;409-410:17-26.
154. Cuocolo R, Stanzione A, Ponsiglione A, Romeo V, Verde F, Creta M, La Rocca R, Longo N, Pace L, Imbriaco M. Clinically significant prostate cancer detection on MRI: A radiomic shape features study. *European Journal of Radiology.* 2019;116:144-9.

155. Gong L, Xu M, Fang M, Zou J, Yang S, Yu X, Xu D, Zhou L, Li H, He B, Wang Y, Fang X, Dong D, Tian J. Noninvasive Prediction of High-Grade Prostate Cancer via Biparametric MRI Radiomics. *J Magn Reson Imaging*. 2020;52(4):1102-9.
156. Pavic M, Bogowicz M, Würms X, Glatz S, Finazzi T, Riesterer O, Roesch J, Rudofsky L, Friess M, Veit-Haibach P, Huellner M, Opitz I, Weder W, Frauenfelder T, Guckenberger M, Tanadini-Lang S. Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncologica*. 2018;57(8):1070-4.
157. Hanchuan P, Fuhui L, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005;27(8):1226-38.
158. Peng CYJ, Lee KL, Ingersoll GM. An introduction to logistic regression analysis and reporting. *Journal of Educational Research*. 2002;96(1):3-14.
159. Chan TY, Partin AW, Walsh PC, Epstein JI. Prognostic significance of Gleason score 3+4 versus Gleason score 4+3 tumor at radical prostatectomy. *Urology*. 2000;56(5):823-7.
160. Stark JR, Perner S, Stampfer MJ, Sinnott JA, Finn S, Eisenstein AS, Ma J, Fiorentino M, Kurth T, Loda M, Giovannucci EL, Rubin MA, Mucci LA. Gleason Score and Lethal Prostate Cancer: Does 3+4=4+3? *Journal of Clinical Oncology*. 2009;27(21):3459-64.
161. Gonzalgo ML, Bastian PJ, Mangold LA, Trock BJ, Epstein JI, Walsh PC, Partin AW. Relationship between primary Gleason pattern on needle biopsy and clinicopathologic outcomes among men with Gleason score 7 adenocarcinoma of the prostate. *Urology*. 2006;67(1):115-9.
162. Kang DE, Fitzsimons NJ, Presti JC, Kane CJ, Terris MK, Aronson WJ, Amling CL, Freedland SJ, SEARCH Database Study G. Risk stratification of men with gleason score 7 to 10 tumors by primary and secondary gleason score: Results from the SEARCH database. *Urology*. 2007;70(2):277-82.
163. Lau WK, Blute ML, Bostwick DG, Weaver AL, Sebo TJ, Zincke H. Prognostic factors for survival of patients with pathological Gleason score 7 prostate cancer: Differences in outcome between primary Gleason grades 3 and 4. *Journal of Urology*. 2001;166(5):1692-7.
164. Wright JL, Salinas CA, Lin DW, Kolb S, Koopmeiners J, Feng Z, Stanford JL. Differences in Prostate Cancer Outcomes Between Cases With Gleason 4 + 3 and Gleason 3 + 4 Tumors in a Population Based Cohort. *Journal of Urology*. 2009;182(6):2702-7.

165. Resnick MJ, Koyama T, Fan K-H, Albertsen PC, Goodman M, Hamilton AS, Hoffman RM, Potosky AL, Stanford JL, Stroup AM, Van Horn RL, Penson DF. Long-Term Functional Outcomes after Treatment for Localized Prostate Cancer. *New England Journal of Medicine*. 2013;368(5):436-45.
166. Vickers A, Bennette C, Steineck G, Adami HO, Johansson JE, Bill-Axelson A, Palmgren J, Garmo H, Holmberg L. Individualized estimation of the benefit of radical prostatectomy from the Scandinavian Prostate Cancer Group randomized trial. *Eur Urol*. 2012;62(2):204-9.
167. Viswanath SE, Bloch NB, Chappelw JC, Toth R, Rofsky NM, Genega EM, Lenkinski RE, Madabhushi A. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery. *J Magn Reson Imaging*. 2012;36(1):213-24.
168. Tiwari P, Kurhanewicz J, Madabhushi A. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. *Med Image Anal*. 2013;17(2):219-35.
169. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol*. 2017;27(10):4082-90.

## Eidesstattliche Versicherung

„Ich, Luise Franz, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: Analyse und Segmentation von MRT-Datensätzen des Prostatakarzinoms für die Entwicklung von Computer-basierten Algorithmen (Analysis and segmentation of MRI datasets of prostate cancer for the development of computer-based algorithms) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; [www.icmje.org](http://www.icmje.org)) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

Datum:

Unterschrift:

# Anteilerklärung an Publikationen

Luise Franz hatte folgenden Anteil an der folgenden Publikation:

Publikation 1: Makowski M, Bressem KK, Franz L, Kader A, Niehues SM, Keller S, Rueckert D, Adams LC, De novo radiomics approach using image augmentation and features from T1 mapping to predict Gleason scores in prostate cancer, Investigative Radiology, 2021.

Beitrag im Einzelnen:

- Retrospektive Identifikation von MRT-Aufnahmen und Einpflegen klinischer und demographischer Daten in eine Datenbank
- Ein- und Ausschluss von Patienten-MRTs in die Studie entsprechend den Ein- und Ausschlusskriterien
- Deskriptive Datenanalyse, aus der die Tabelle 1 in der Publikation entstanden ist
- Manuelle softwaregestützte Segmentation von Tumoren und Organregionen der Prostata (periphere Zone, Transitionalzone) sowie mehrfache manuelle Korrektur nach Revision der Annotationen zusammen mit Radiolog\*innen
- Extraktion von Radiomics-Features aus den Bilddaten und Diskussion der geeigneten Augmentationsstrategie während des Extraktionsvorgangs
- Zuarbeiten bei Aufstellung der Vorhersagemodelle, insbesondere Aufarbeiten der Input-Daten für den Algorithmus (Gleason-Score, Radiomics-Features)
- Mitwirken beim Verfassen der Publikation sowie Zuarbeit bei der Revision

---

Unterschrift, Datum und Stempel des/der erstbetreuenden Hochschullehrers/in

---

Unterschrift des Doktoranden/der Doktorandin

## **Lebenslauf**

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

## Liste der Publikationen

Makowski MR, Bressemer KK, Franz L, Kader A, Niehues SM, Keller S, Rueckert D, Adams LC. De Novo Radiomics Approach Using Image Augmentation and Features From T1 Mapping to Predict Gleason Scores in Prostate Cancer. *Invest Radiol*. April 2021. doi: 10.1097/RLI.0000000000000788. Epub ahead of print. PMID: 34047538.

## Danksagung

Ich danke meiner Erstbetreuerin PD Dr. med. Lisa Adams für die sehr gute und unkomplizierte Betreuung, die schnelle Beantwortung meiner Fragen und die ständige Erreichbarkeit bei auftretenden Problemen. Es hat Spaß gemacht, so engmaschig betreut zu werden und gleichzeitig den Freiraum zu haben, selbstständig zu arbeiten und eigene Schwerpunkte zu setzen. Danke auch an meinen Zweitbetreuer PD Dr. med. Keno Bressemer für die detaillierten Erklärungen zu den Prozessen der Radiomics-Analyse und an Rouven Glauert, dessen Wissen über Machine-Learning-Algorithmen mich sehr bereichert hat. Nicht zuletzt möchte ich meinen Eltern danken, die mich während meines Zweitstudiums fortlaufend unterstützt und motiviert haben und mir so die Anfertigung dieser Arbeit ermöglicht haben.

# Bescheinigung Statistik



CharitéCentrum für Human- und Gesundheitswissenschaften

Charité | Campus Charité Mitte | 10117 Berlin

Institut für Biometrie und Klinische Epidemiologie (IBiKE)

Direktorin: Prof. Dr. Geraldine Rauch

Name, Vorname: Franz, Luise  
Emailadresse: luise.franz@charite.de  
Matrikelnummer: 223109  
PromotionsbetreuerIn: Adams, Lisa  
Promotionsinstitution/Klinik: Radiologie CCM

Postanschrift:  
Charitéplatz 1 | 10117 Berlin  
Besucheranschrift:  
Reinhardtstr. 58 | 10117 Berlin

Tel. +49 (0)30 450 562171  
geraldine.rauch@charite.de  
<https://biometrie.charite.de/>



## Bescheinigung

Hiermit bescheinige ich, dass Frau *Luise Franz* innerhalb der Service Unit Biometrie des Instituts für Biometrie und Klinische Epidemiologie (IBiKE) bei mir eine statistische Beratung zu einem Promotionsvorhaben wahrgenommen hat. Folgende Beratungstermine wurden wahrgenommen:

- *Termin 1: 11.01.2021*

Folgende wesentliche Ratschläge hinsichtlich einer sinnvollen Auswertung und Interpretation der Daten wurden während der Beratung erteilt:

- *Schwerpunkt auf Deskription der Daten legen*
- *Aufgrund kleiner Fallzahl die Ergebnisse der Machine-Learning-Methoden vorsichtig interpretieren*

Diese Bescheinigung garantiert weder die richtige Umsetzung der in der Beratung gemachten Vorschläge, die korrekte Durchführung der empfohlenen statistischen Verfahren noch die richtige Darstellung und Interpretation der Ergebnisse. Die Verantwortung hierfür obliegt allein der Promovierenden. Das Institut für Biometrie und Klinische Epidemiologie übernimmt hierfür keine Haftung.

Datum: 11.01.2021

Name des Beraters: Claus Nowak

Unterschrift Berater, Institutsstempel



Claus Peter Nowak  
Digital unterschrieben von Claus Peter Nowak  
Datum: 2021.01.11 08:48:56 +01'00'