

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

DISSERTATION

CT basierte Radiomics-Modelle zur Vorhersage des ^{68}Ga -PSMA PET-
Lymphknotenstatus bei Patienten mit Prostatakarzinom

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

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Falk Konstantin Lübke

aus Bonn

Datum der Promotion: 04.03.2022

Vorwort

Teilergebnisse der vorliegenden Arbeit wurden veröffentlicht in:

A. Hartenstein, F. Lübbe, A. D. J. Baur, M. M. Rudolph, C. Furth, W. Brenner, H. Amthauer, B. Hamm, M. Makowski & T. Penzkofer, *Prostate Cancer Nodal Staging: Using Deep Learning to Predict 68Ga-PSMA-Positivity from CT Imaging Alone*, Sci Rep. 2020 Feb 25;10(1):3398.
doi: 10.1038/s41598-020-60311-z. / 25.02.2020

Inhaltsverzeichnis

ABBILDUNGSVERZEICHNIS	6
TABELLENVERZEICHNIS	7
ABKÜRZUNGSVERZEICHNIS	8
ABSTRACT	11
ENGLISH ABSTRACT	13
1. EINLEITUNG	15
1.1 DAS PROSTATAKARZINOM.....	15
1.1.1 Epidemiologie.....	15
1.1.2 Basisdiagnostik und das PSA – Screening.....	15
1.1.3 MRT in der Primärdiagnostik	17
1.1.4 Staging	18
1.2 ⁶⁸ GA-PSMA PET/CT ZUM LYMPHKNOTENSTAGING	20
1.3 GESCHICHTE DER COMPUTERTOMOGRAPHIE.....	21
1.4 GRUNDSÄTZE DER PET-CT	23
1.4.1 Rohdaten	23
1.4.2 Korrektur der Rohdaten	24
1.5 ZIELMOLEKÜLE.....	25
1.6 PSMA UND DIE GESCHICHTE DER TRACERMOLEKÜLE.....	26
1.7 TUMORHETEROGENITÄT	27
1.8 GRUNDSÄTZE DER TEXTURANALYSE.....	29
1.8.1 Textur.....	29
1.9 RADIOMICS	32
1.10 HYPOTHESE.....	34

2. MATERIAL UND METHODEN	36
2.1 STUDIENPOPULATION	36
2.2 POSITRONEN-EMISSIONS-TOMOGRAPHIE-TRACER.....	37
2.3 BILDGEBUNGSPROTOKOLL.....	37
2.4 BILDANALYSE UND SEGMENTIERUNG DER LYMPHKNOTEN	39
2.5 STATISTISCHE AUSWERTUNG	41
2.5.1 <i>Balancing</i>	42
2.5.2 <i>Bagging</i>	42
2.5.3 <i>Extraktion</i>	43
2.5.4 <i>Featureselektion</i>	43
2.5.5 <i>Stabilität</i>	44
2.5.6 <i>Training</i>	45
2.5.7 <i>Testen</i>	46
2.6 VERGLEICH DER KLASSIFIKATOREN MIT DER EINSCHÄTZUNG VON RADIOLOGEN	46
3. ERGEBNISSE	47
3.1 VERGLEICH DER KLASSIFIKATOREN MIT DER EINSCHÄTZUNG VON RADIOLOGEN	53
4. DISKUSSION	60
4.1 DISKUSSION DER METHODEN	60
4.1.1 <i>Gerät</i>	60
4.1.2. <i>PET-CT zu Beurteilung von Malignität</i>	60
4.1.3 <i>Segmentiermethode</i>	61
4.1.4 <i>Featureselektion und Vorhersagemodelle</i>	62
4.2. DISKUSSION DER ERGEBNISSE	64
4.2.1 <i>Vergleich der Modelle mit Radiologen</i>	65

4.2.1.1 Performance der CT-Lymphknotendiagnostik in anderen Studien	65
4.2.2 Vergleich mit anderen Radiomics – Studien.....	66
4.2.3 Deep learning	68
4.2.4 Ausgewählte Radiomic Features.....	68
4.3 AUSBLICK	70
4.4 LIMITATIONEN	70
5. ZUSAMMENFASSUNG	72
6. LITERATURVERZEICHNIS	75
EIDESSTATTLICHE VERSICHERUNG	85
ANTEILSERKLÄRUNG AN ETWAIGEN ERFOLGTEN PUBLIKATIONEN	87
LEBENS LAUF	88
PUBLIKATIONS LISTE	88
DANKSAGUNG.....	91

Abbildungsverzeichnis

ABBILDUNG 1: VERANSCHAULICHUNG STATISTISCHER METHODEN IN DER QUANTITATIVEN BILDVERARBEITUNG.....	31
ABBILDUNG 2: STUDIENPOPULATION UND ERSTELLEN DER DATENSÄTZE.	37
ABBILDUNG 3: PROZESS DER RADIOMICS-METHODE.....	38
ABBILDUNG 4: MULTIPLANARES PET-CT FUSIONSBILD EINES PATIENTEN MIT RETROPERITONEALEN LYMPHKNOTENMETASTASEN EINES PROSTATAKARZINOMS.	40
ABBILDUNG 5: SEGMENTIERUNG VON LYMPHKNOTEN MITTELS MITK.	41
ABBILDUNG 6: VERGLEICH VERSCHIEDENER FEATURE-SELEKTIONSMETHODEN HINSICHTLICH IHRER STABILITÄT.	45
ABBILDUNG 7: DARSTELLUNG DER AM TRAINIGSDATENSATZ ERZEUGTEN KLASSIFIKATIONSBÄUME.	49
ABBILDUNG 8: GRAFISCHE DARSTELLUNG DER ERGEBNISSE NACH ANWENDUNG DER ALGORITHMEN AUF DEN TESTDATENSATZ.	51
ABBILDUNG 9: KLASSIFIKATIONSLEISTUNG DER DER ZWEI RADIOLOGEN.	53
ABBILDUNG 10: KOMBINATION DER ERGEBNISSE DER RADIOLOGEN UND DER KLASSIFIKATIONALGORITHMEN.	56
ABBILDUNG 11: GRAFISCHE DARSTELLUNG DER ERGEBNISSE NACH BAGGING.	58

Tabellenverzeichnis

TABELLE 1: KLINISCHE TNM-KLASSIFIKATION DES PROSTATAKARZINOMS DER UICC.....	18
TABELLE 2: KLASSIFIZIERUNG DES REZIDIVRISIKOS FÜR LOKAL BEGRENZTE KARZINOME	19
TABELLE 3: SELEKTION DER ZWANZIG RELEVANTESTEN FEATURES.....	47
TABELLE 4: ERGEBNISSE NACH ANWENDUNG DER ALGORITHMEN AUF DEN TESTDATENSATZ....	50
TABELLE 5: STATISTISCHE AUSWERTUNG DER EINSCHÄTZUNG DER ZWEI RADIOLOGEN.....	53
TABELLE 6: KOMBINATION DER ERGEBNISSE DER RADIOLOGEN UND DER KLASSIFIKATIONSALGORITHMEN.	54
TABELLE 7: ERGEBNISSE NACH BAGGING.	57

Abkürzungsverzeichnis

A

AUC *area under the curve*

B

BPH *benigne Prostatahyperplasie*

C

CNN *convolutional neural network*

CT *Computertomographie*

D

DRU *digital-rektale-Untersuchung*

E

EAU *European Association of Urology*

F

FSM *Featureselektionsmethode*

G

glcm *Gray-Level Co-Occurrence Matrix*

gldm *Gray Level Dependence Matrix*

glrlm *Gray Level Run Length Matrix*

glszm *Grey Level Size Zone Matrix*

H

HU *Hounsfield Units*

I

IOV *interobserver variability*

L

lda *lineare Diskriminanzanalyse*

log *logistische Regression*

LOR *Line of Response*

M

mi *mutual information*

mlp *multilayer perceptron*

mrmi *maximum relevance minimum redundancy*

MRT *Magnetresonanztomographie*

N

ngtdm *Neighbourhood Gray-tone Difference Matrix*

NSCLC *non small-cell lung carcinoma*

P

PCa *prostate carcinoma*

PET *Positronenemissionstomographie*

PET/CT *Positronenemissionstomographie/Computertomographie*

PI-RADS *Prostate Imaging Reporting and Data System*

pls *partial least squares*

PSA *prostata-spezifisches Antigen*

PSMA *prostata-spezifisches Membranantigen*

R

ROC *receiver operating characteristic*

rpart *recursive partitioning*

s

svm *support vector machines*

v

VOI *volume of interest*

Abstract

Die Detektion von Lymphknotenmetastasen ist bei Patienten mit Prostatakarzinom im Rahmen der Ausbreitungsdiagnostik von therapieentscheidender Bedeutung. Die standardmäßig eingesetzte Bildgebung mittels CT zeigt nur eine geringe Aussagekraft.

Mit dem ^{68}Ga -PSMA PET/CT existiert nun eine äußerst vielversprechende Methode zum Staging bei Prostatakarzinom. Bislang nur zum Re-Staging leitliniengerecht, stehen die Chancen für eine baldige Empfehlung zum Primärstaging gut. Das ^{68}Ga -PSMA PET/CT ist jedoch nur eingeschränkt verfügbar, so dass viele Staging-Untersuchungen weiterhin mit kontrastmittelgestütztem CT durchgeführt werden. In den letzten Jahren konnten Radiomics-Methoden zur quantitativen Bildanalyse gute Leistungen u.a. in der Unterscheidung von malignen und benignen Veränderungen verschiedener Entitäten zeigen. Wir haben untersucht, ob mittels Radiomics-Methoden der ^{68}Ga -PSMA PET/CT-Status von Lymphknoten nur auf Basis der CT-Bilder vorhergesagt werden kann.

Methodik: Aus einem ^{68}Ga -PSMA PET/CT Bilddatensatz von 447 Patienten mit Prostatakarzinom wurden 369 PET-positive und 1756 PET-negative Lymphknoten identifiziert und segmentiert. Relevante Radiomics-Features wurden mittels vier verschiedener Featureselektionsmethoden (FSM) [Methode nach Wilcoxon, Area under the curve (AUC), mutual information (mi), mutual information minimum redundancy (mrmi)] ermittelt. Anschließend wurden sechs verschiedene Klassifikationsalgorithmen (Klassifikatoren) [lineare und logistische Regression, support vector machines (svm), ein neuronales Netzwerk (multilayer perceptron, mlp) sowie Klassifikationsbäume (recursive partitioning, rpart)] zur Vorhersage des Lymphknotenstatus an einem nach Bagging-Verfahren balancierten Trainingsdatensatz trainiert. 24 FSM-Klassifikator-Kombinationen wurden am separaten Testdatensatz getestet. Zum Vergleich klassifizierten 2 Radiologen denselben Datensatz nach vier Kategorien (0=sicher nicht

maligne, 1= wahrscheinlich nicht maligne, 2= wahrscheinlich maligne, 3= sicher maligne).

Zusätzlich wurde die Vorhersagekraft der Kombination von automatischen Klassifikatoren und Radiologen ausgewertet.

Ergebnisse: Die beste Vorhersageleistung wurde von den Klassifikationsbäumen (rpart) mit vorheriger Featureselektion nach der mrmi-Methode erreicht mit einer Korrektklassifikationsrate von 83% (Sensitivität 88%, Spezifizität 82%, positiver Vorhersagewert 48%, negativer Vorhersagewert 97%). Die Radiologen erreichten in der Beurteilung der Lymphknoten beide eine Korrektklassifikationsrate von 95% (Sensitivität beide 76%, Spezifizität 98% und 97%, positive Vorhersagewerte 88% und 84%, negativer Vorhersagewert beide 96%). Die Kombination von Klassifikatoren und Radiologen verbesserte die Klassifikationsleistung der Radiologen nicht.

Diskussion: Radiomics-Modelle auf Basis von CT-Bildern können den ⁶⁸Ga-PSMA PET-Status von Lymphknoten mit hoher Wahrscheinlichkeit vorhersagen. Im Vergleich zur, in der bisherigen Literatur, ermittelten Leistung der CT-Bildgebung, könnten sie die Lymphknotendiagnostik bei Patienten mit Prostatakarzinom verbessern bzw. die konventionelle Diagnostik oder das ⁶⁸Ga-PSMA PET/CT ergänzen. Die Validierung der Klassifikationsalgorithmen an externen Datensätzen und weitere Studien mit histopathologischer Korrelation sind notwendig, um eine klinische Anwendbarkeit zu ermöglichen.

English Abstract

The detection of lymph node metastases for staging is of critical importance for therapy decisions in patients with prostate cancer (PCa). CT as a standard imaging technique is of limited value.

With ^{68}Ga -PSMA PET/CT, a promising method for staging in PCa now exists. So far only guideline-compliant for re-staging, chances for a timely recommendation for primary staging are good. However, ^{68}Ga -PSMA PET/CT remains limited in availability, so many staging examinations are still performed with contrast-enhanced CT. In recent years, radiomics methods for quantitative image analysis have shown good performance e.g. in differentiating malignant from benign lesions of different entities. We investigated whether radiomics methods could predict the ^{68}Ga -PSMA PET-status of lymph nodes based on CT images only.

METHODS: From a ^{68}Ga -PSMA PET/CT image dataset of 447 patients with prostate cancer, 369 PET-positive and 1756 PET-negative lymph nodes were identified and segmented. Relevant radiomic features were identified using four different feature selection methods (FSM) [method according to Wilcoxon, area under the curve (AUC), mutual information (mi), mutual information minimum redundancy (mrmi)]. Then, six different classification algorithms (classifiers) [linear and logistic regression, support vector machines (svm), a neural network (multilayer perceptron, mlp), and classification trees (recursive partitioning, rpart)] for lymph node status prediction were trained on a training dataset balanced by the bagging method. 24 FSM-classifier combinations were tested on the separate test dataset. For comparison, 2 radiologists classified the same data set according to four categories (0=certainly nonmalignant, 1=probably nonmalignant, 2=probably malignant, 3=certainly malignant). In addition, the predictive power of the combination of automatic classifiers and radiologists was evaluated.

Results: The best predictive performance was achieved by the classification trees (rpart) with prior feature selection using the mrmi-method with an accuracy of 83% (sensitivity 88%, specificity 82%, positive predictive value 48%, negative predictive value 97%). The radiologists both achieved a 95% accuracy (sensitivity both 76%, specificity 98% and 97%, positive predictive values 88% and 84%, negative predictive value both 96%). The combination of classifiers and radiologists did not improve classification performance of radiologists.

Discussion: Radiomics models based on CT images can predict ^{68}Ga -PSMA PET-status of lymph nodes with high probability. Compared to the performance of CT imaging determined in previous literature, they could potentially improve lymph node diagnosis in patients with PCa and/or supplement conventional diagnostics or ^{68}Ga -PSMA PET/CT. Validation of the classifiers on external datasets and further studies with histopathological correlation are necessary to enable clinical applicability.

1. Einleitung

1.1 Das Prostatakarzinom

1.1.1 Epidemiologie

Mit 25,4% aller festgestellten Krebserkrankungen ist das Prostatakarzinom (PCa) die häufigste bei Männern in Deutschland.¹ Die jährliche Inzidenz betrug 2016 etwa 58.780 Neuerkrankungen.² Ca. 3% aller Sterbefälle gehen auf das Prostatakarzinom zurück; es ist damit die sechsthäufigste Todesursache in Deutschland¹. Die altersstandardisierte Erkrankungsrate ist nach einem langen Anstieg seit 2003 konstant und sinkt seit 2011. 2013 betrug sie noch 99 auf 100.000 Personen, 2014 waren es noch 92,7, mit eher stabiler Entwicklungserwartung². Aufgrund der aktuellen demographischen Entwicklung müssen wir aber in Zukunft mit steigenden absoluten Erkrankungsraten rechnen²: Das Erkrankungsrisiko eines 35-jährigen liegt bei unter 0,1%, während das Risiko eines 75-jährigen an Prostatakrebs zu erkranken etwa 5% beträgt. Das mittlere Erkrankungsalter lag 2014 bei 72 Jahren².

Der Anteil der über 60-jährigen wird bis 2050 auf 28 Millionen Männer gestiegen sein, im Vergleich zu heute entspräche dies einer Verdopplung¹. Medizinisch und gesundheitspolitisch könnte das Thema PCa zunehmend an Bedeutung gewinnen³.

1.1.2 Basisdiagnostik und das PSA – Screening

In der Basisdiagnostik zum Ausschluss des Prostatakarzinoms sind die digital-rektale Untersuchung (DRU) sowie die Bestimmung des prostataspezifischen Antigens (PSA) im Labor etabliert. Die Bestimmung des PSA-Wertes ist das primäre Verfahren in der Früherkennung. Die DRU wird in der Früherkennung als alleinige Diagnostik, mit einer Sensitivität von 59% und einer Spezifität von 94%, als nicht ausreichend bewertet, aber zusätzlich empfohlen^{1,4}. Die Kombination aus DRU und PSA-Wert Bestimmung erhöht die Spezifität in der Erkennung von Prostatakarzinomen. Ergänzend kann in der Primärdiagnostik eine transrektale Sonographie durchgeführt werden¹.

Das prostataspezifische Antigen ist zwar organspezifisch, es kann aber auch z.B. im Falle einer Prostatitis oder einer benignen Prostatahyperplasie (BPH) erhöht sein.⁵

Die Bestimmung des PSA-Werts bzw. das PSA-Screening ist eine zuletzt umstrittene Methode zur Früherkennung des Prostatakarzinoms. In einer großen europäischen Studie mit 162.243 Probanden konnte gezeigt werden, dass das PSA-Screening einen signifikanten Rückgang der prostatakrebspezifischen Mortalität zur Folge hat: in der Screeninggruppe starben im Vergleich zur Kontrollgruppe pro 10.000 Probanden 12 weniger am Prostatakarzinom (49 vs. 61)^{1,6}. 781 Patienten müssen gescreent werden, um eine Person zu retten.

Das PSA-Screening verursacht allerdings auch eine hohe Menge an Überdiagnosen. In der Screeninggruppe wurden pro 10.000 Männer 1020 PCA diagnostiziert, in der Kontrollgruppe waren es 680. 340 Patienten erhielten in der Screeninggruppe die Diagnose Prostatakrebs, was ihnen ohne Screening erspart geblieben wäre.¹

Von den Autoren der Studie wird zusätzlich vermutet, dass die Behandlung etwa der Hälfte durch das Screening diagnostizierten Tumore nicht notwendig ist. Um eine Person zu retten, müssen 34 PCA diagnostiziert werden^{1,6}.

Eine Metaanalyse aus der Cochrane Collaboration, die die Ergebnisse aller vorhandenen Daten aus randomisierten kontrollierten Studien zusammenführt, zeigt, dass weder PCa-spezifische Mortalität noch die Gesamtsterblichkeit durch ein PSA-Screening verringert werden. Allerdings treten häufig Überdiagnostik und Übertherapie auf.^{1,7} Von einer systematischen Aufklärung über das PSA-Screening wird somit laut Leitlinienprogramm Onkologie 2019 abgeraten¹.

Bei verdächtigem Befund in der Früherkennung, beispielsweise einem PSA-Wert ≥ 4 ng/ml, einem auffälligen Anstieg des PSA-Werts oder bei suspektem Befund in der DRU, soll laut Leitlinie eine Prostatabiopsie angeraten werden. In der histologischen Untersuchung soll zum

Grading der Gleason Score verwendet werden.^{1,8} Bei einem Gleason-Score von ≥ 8 sollen weitere Staginguntersuchungen folgen¹.

Ein kontroverser diagnostischer Wert ist die PSA-Dichte; der PSA-Wert, geteilt durch das, im transrektalen Ultraschall gemessene, Prostatavolumen. Die PSA-Dichte liefert bei Patienten, die sich in aktiver Überwachung befinden Informationen über das Risiko einer Tumorprogression^{5,9}.

1.1.3 MRT in der Primärdiagnostik

Im Falle einer negativen Biopsie sollte eine MRT (Magnetresonanztomographie) erfolgen. Suspekte Läsionen können dann gezielt biopsiert werden¹. Zur Beurteilung von Prostataläsionen im MRT hat sich das Scoring-system Prostate Imaging Reporting and Data System (PI-RADS, deutsch: Prostatabildgebung Befundungs- und Datensystem) etabliert. Mit PI-RADS kann auf einer Skala von 1-5 die Wahrscheinlichkeit angegeben werden, mit der es sich bei einer detektierten Läsion um ein klinisch signifikantes Prostatakarzinom handelt^{10,11}. Das MRT kann auch schon im Rahmen der Erstbiopsie eingesetzt werden. Zusätzlich zur MRT-gesteuerten Biopsie sollte leitliniengerecht aber auch eine systematische Biopsie der Prostata zusätzlich zur PSA-Kontrolle angeboten werden, da auch ein negatives MRT ein signifikantes Karzinom nicht mit letzter Sicherheit ausschließen kann¹. Durch die Kombination von MRT-gesteuerter Biopsie und systematischer Biopsie kann eine höhere Detektionsrate klinisch signifikanter Karzinome erreicht werden als durch jede Methode allein^{1,12}.

Die Kombination von MRT-gesteuerter Biopsie systematischer Biopsie muss aber nicht in jedem Falle angewendet werden. In einem Review der Cochrane Collaboration wird argumentiert, dass auf die systematische Biopsie in bestimmten Fällen zugunsten eines klinischen und biochemischen Monitorings verzichtet werden könnte.¹⁰ Auch die aktuelle Leitlinie der EAU (European Association of Urology, deutsch: europäische Gesellschaft für Urologie) bietet die Möglichkeit an, im Falle eines negativen MRT und klinisch geringem Verdacht auf ein

Prostatakarzinom, in gemeinsamer Entscheidungsfindung mit dem Patienten auf eine Biopsie zu verzichten⁵.

1.1.4 Staging

Tabelle 1: Klinische TNM-Klassifikation des Prostatakarzinoms der UICC.

Beurteilung der Ausdehnung der Krebserkrankung (T=Primärtumor, N=Lymphknotenmetastasen, M=Fernmetastasen, Quelle: wörtlich nach Wittekind, 2020, S. 245-247¹³).

Kategorie	
TX	Primärtumor kann nicht beurteilt werden
T0	Kein Anhalt für Primärtumor
T1	Klinisch nicht erkennbarer Tumor, der weder tastbar noch in bildgebenden Verfahren sichtbar ist.
T1a	Tumor zufälliger histologischer Befund („incidental carcinoma“) in 5 % oder weniger des resezierten Gewebes
T1b	Tumor zufälliger histologischer Befund („incidental carcinoma“) in mehr als 5 % des resezierten Gewebes
T1c	Tumor durch Nadelbiopsie diagnostiziert (z. B. wegen erhöhtem PSA)
T2	Tumor begrenzt auf Prostata
T2a	Tumor befällt die Hälfte eines Lappens oder weniger
T2b	Tumor befällt mehr als die Hälfte eines Lappens
T2c	Tumor in beiden Lappen
T3	Tumor durchbricht die Prostatakapsel
T3a	Extraprostatische Ausbreitung (einseitig oder beidseitig) eingeschlossen mikroskopisch nachweisbare Infiltration des Blasenhalses
T3b	Tumor infiltriert Samenblase(n)
T4	Tumor ist fixiert oder infiltriert andere benachbarte Strukturen als Samenblasen, z. B. Sphincter externus, Rektum, und/oder Levatormuskel und/oder ist an Beckenwand fixiert
NX	Regionäre Lymphknoten können nicht beurteilt werden
N0	Keine regionären Lymphknotenmetastasen

N1	Regionäre Lymphknotenmetastasen
M0	Keine Fernmetastasen
M1	Fernmetastasen
M1a	Nichtregionäre(r) Lymphknoten
M1b	Knochen
M1c	Andere Lokalisation(en)

Tabelle 2: Klassifizierung des Rezidivrisikos für lokal begrenzte Karzinome

(cT= klinische Primärtumor-Kategorie, s. Tabelle 1, Quelle: in Anlehnung an das Leitlinienprogramm Onkologie, 2019, S. 60¹, basierend auf D'Amico et al., 1998¹⁴ und Thompson et al., 2007¹⁵)

	Niedriges Risiko	Intermediäres Risiko	Hohes Risiko
PSA	≤ 10 ng/mL	> 10 ng/ml - 20 ng/ml oder	> 20 ng/ml oder
Gleason Score	6	7 oder	≥ 8 oder
cT- Kategorie	1c, 2a	2b	2c

Bei Patienten mit einem Gleason Score von ≥ 8 bzw. einer Tumorkategorie cT3/4 (s. Tabelle 1) ist die Detektion von Lymphknotenmetastasen von therapieentscheidender Bedeutung¹. Laut Leitlinie sollen diese Patienten im Rahmen des Primärstaging eine MRT- oder CT (Computertomographie) der Beckenorgane bekommen.¹ Patienten mit positivem Befund für Lymphknotenmetastasen in der CT-Untersuchung kann eine Lymphadenektomie erspart werden.¹ Die Performance der CT ist in der Diagnose von Lymphknotenmetastasen allerdings eingeschränkt^{16,17}. In einem Review von Abuzallouf et al., das Studien mit insgesamt 4264 Patienten nach CT und pelviner Lymphadenektomie einschließt, hatten insgesamt 15,3 % eine histopathologisch bewiesene Lymphadenopathie. Das CT detektierte diese bei nur 2,5 %. Die Sensitivität lag bei einem Median von 7%; die Mediane von Spezifität sowie positivem und

negativem prädikativen Wert lagen bei 100%, 85% und 100%, respektive. Bei Patienten mit Gleason Score von 8 oder darüber lag die Detektionsrate bei 12,5 %, das tatsächliche Risiko für eine pathologische Lymphadenopathie nach pelviner Lymphadenektomie lag aber bei 22,8%. Es wurden also im CT etwas mehr als die Hälfte der Patienten mit Gleason Score von 8 oder darüber und positivem Lymphknotenbefund korrekt klassifiziert.¹⁷

Ist das Prostatakarzinom histologisch gesichert, sollten Patienten mit einem PSA-Wert >10 oder einem Gleason-Score ≥ 8 , oder einer T-Kategorie cT3/4 oder Knochenschmerzen zum Ausschluss von Knochenmetastasen eine Skelettszintigraphie erhalten.¹

Laut aktuell gültiger Leitlinie ist die Rolle die PET/CT (Positronenemissionstomographie/Computertomographie) im Rahmen des primären Stagings noch unklar und sollte im Rahmen klinischer Studien erfolgen. Es gibt allerdings Hinweise, dass das Positronenemissionstomogramm mit Targeting des prostataspezifischen Membranantigens (PSMA-PET/CT) bessere Ergebnisse als konventionelle Diagnostik liefert^{16,18}. Derzeit ist es aufgrund der unzureichenden Studienlage aber nur zum Re-Staging nach Tumorrezidiv empfohlen¹. Die derzeit im Stadium der Konsultationsfassung befindliche neue Version der Leitlinie Prostatakarzinom stellt allerdings eine Änderung in Aussicht. Hiernach könnte das PSMA PET/CT bald zur primären Ausbreitungsdiagnostik bei Hochrisikopatienten (s. Tabelle 2) eingesetzt werden.

1.2 ⁶⁸Ga-PSMA PET/CT zum Lymphknotenstaging

In einer randomisierten Studie mit 302 Patienten mit Prostatakarzinom und hohem Risiko (z.B. PSA >20 ng/mL oder International Society of Urothology (ISUP) grade group 3–5, Stadium $\geq T3$) konnten eine Sensitivität und Spezifität von 85% und 98%, respektive, des mit ⁶⁸Ga markierten PSMA PET-CT für pelvine Lymphknotenmetastasen und Fernmetastasen festgestellt werden. Vor allem zeigte sich ein deutlicher Vorteil im Vergleich zum konventionellen Staging mittels einer Kombination aus CT und Knochenszintigraphie. Die Sensitivität und Spezifität

für diese Untersuchungsmethoden betragen nur 38% und 91%, respektive. Es handelte sich hier um „per-patient“ (deutsch. pro Patient) – Analysen; es wurden also für die statistische Analyse einzelne Fälle als positiv oder negativ klassifiziert, nicht einzelne Läsionen.¹⁸

In einer Studie von Maurer et al. an 130 Patienten mit moderatem bis hochgradigem Prostatakarzinom ergab sich für das Lymphknotenstaging mittels PSMA PET/CT „per-lesion“ (deutsch: pro Läsion), also für einzelne Läsionen, eine Sensitivität, Spezifität und Korrektklassifikationsrate in der Beurteilung von Lymphknotenmetastasen von 73,5%, 99,2% und 95,7% sowie eine „per-patient“-Sensitivität, -Spezifität und -Korrektklassifikationsrate von 65,9%, 98,9% und 88,5%, respektive.¹⁶

Maurer et al. konnten in derselben Studie zeigen, dass Lymphknoten allein mithilfe von CT oder MRT Bildgebung mit einer per-lesion-Sensitivität von nur 28,2%, einer -Spezifität von 97,1% und einer -Korrektklassifikationsrate (accuracy) von 86,1% klassifiziert werden. Die per-patient-Sensitivität, -Spezifität und -Korrektklassifikationsrate lagen bei 43,9%, 85,4% und 72,3%, respektive.¹⁶

Insgesamt zeigen die vorliegenden Studien deutliche Vorteile des PSMA PET/CT gegenüber konventionellen Staginguntersuchungen.

1.3 Geschichte der Computertomographie

Der erste CT-Scanner wurde 1971 von Godfrey N. Hounsfield zur Untersuchung des Schädels entwickelt. Das Prinzip der CT besteht in der Abtastung des Patienten durch Röntgenstrahlen aus verschiedenen Richtungen und der Erfassung der Strahlung durch Detektoren. Durch Kollimation des Röntgenstrahls wird erreicht, dass der Patient Schicht-für-Schicht durchleuchtet werden kann. Aus der aus jeder Richtung gemessenen Schwächung der Röntgenstrahlen wird, basierend auf der Radon-Transformation, die Röntgenschwächung an jedem Punkt der Schicht errechnet. Diese wird in sogenannte CT-Werte umgerechnet, die in hounsfield units (HU,

deutsch: Hounsfield Einheiten) angegeben werden. Wasser wird der Wert 0 HU zugeordnet; Luft der Wert -1000 HU. Diese werden in Graustufen kodiert und so als Bild darstellbar gemacht.^{19,20}

CT-Scanner der ersten und zweiten Generation arbeiteten noch nach dem Translation-Rotation-Prinzip^{19,20}. In Scannern der ersten Generation sandte eine Röntgenröhre einen Nadelstrahl aus und war einem Detektor zugeordnet. Diese musste sich in einer Translationsbewegung einmal über den Querschnitt des Patienten bewegen, worauf eine Rotationsbewegung um 1° folgte. Bei CT-Scannern der zweiten Generation war der Nadelstrahl zu einem Fächer ausgeweitet. Dies ermöglichte eine Rotationsbewegung um 30°, was die Aufnahmezeit schon deutlich verkürzte. Dennoch funktionierte dieser Scanner noch nach dem Translation-Rotation Prinzip²⁰

Erst der Scanner der 3. Generation funktionierte nach dem Rotation-Rotation-Prinzip. Der Strahlenfächer konnte den Körperquerschnitt als Ganzes erfassen. Die Translationsbewegung war somit obsolet. Bei Geräten der vierten Generation, die sich aber nicht durchsetzen konnten, rotierte nur die Röntgenröhre. Ein stationärer Detektorkranz füllte den gesamten Durchmesser der Scanneröffnung aus^{19,20}.

Mitte der 90-er Jahre stieg das Spiral-CT zur mehrheitlich verwendeten Technologie auf. Der Unterschied zu vorherigen Modellen war der kontinuierliche Tischvorschub, was zu einer spiralförmigen, kontinuierlichen Abtastbewegung führte. Dies beschleunigte die Untersuchungszeit so, dass eine Abtastung der untersuchten Region innerhalb eines Atemzuges möglich wurde.¹⁹ Dies macht sich vor allem in der Suche nach Lungen- oder Leberherden bemerkbar. Das Spiral-CT verhindert, dass Schichten in atemverschieblichen Organen verpasst bzw. doppelt gescannt werden²¹.

1.4 Grundsätze der PET-CT

„Die Positronenemissionstomographie (PET) ist ein Verfahren zur nicht-invasiven Bestimmung der räumlichen und zeitlichen Verteilung einer radioaktiven Substanz im menschlichen Körper anhand der emittierten Strahlung“²². Dazu wird ein Radiopharmakon intravenös appliziert. Das Radiopharmakon ist die Kombination eines Radionuklids und eines pharmazeutischen Wirkstoffs, bzw. Pharmakons. Letzteres ist als Adressat zu verstehen, der das Radionuklid gezielt an eine bestimmte Stelle im Körper bringt^{22,23}

Innerhalb des Atomkerns des instabilen Radionuklids findet zunächst eine Kernumwandlung statt, wobei ein Proton in ein Neutron umgewandelt wird. Dies setzt ein Positron frei, das bereits nach kurzer Distanz mit einem Elektron kollidiert. Elektron und Positron vernichten sich in einer sogenannten Annihilation gegenseitig. Energie wird in Form zweier γ -Quanten frei, die sich in einem nahezu exakten 180° -Winkel voneinander entfernen. Die γ -Quanten treffen dann ungefähr zeitgleich auf entgegengesetzte Seiten des Detektorrings²²⁻²⁴. Szintillationskristalle im Detektorring produzieren auf ein Auftreffen eines γ -Quants ein schwaches Lichtsignal, das von einem Photomultiplier in ein elektrisches Signal umgewandelt wird. Durch Bleiblenen wird sichergestellt, dass nur γ -Quanten aus einer Einfallsrichtung auf den Detektor treffen.²²

1.4.1 Rohdaten

Werden zwei γ -Quanten zeitgleich vom Detektorring innerhalb eines gewissen Zeitintervalls detektiert, spricht man von einer Koinzidenz. Auf einer Verbindungslinie zwischen den beiden Detektoren, der Line of Response (LOR, deutsch: Koinzidenzlinie) wird das Annihilationsereignis angenommen. Die Koinzidenzen können entweder wahre Koinzidenzen sein, wenn beide γ -Quanten tatsächlich aus einer Annihilation stammen, oder unerwünschte Koinzidenzen, wenn z.B. zwei γ -Quanten nur zufällig zeitgleich auftreffen und nicht aus einer Annihilation stammen. Weitere unerwünschte Koinzidenzen entstehen durch Streuung, wenn γ -

Quanten auf ihrem direkten Weg zum Detektor abgelenkt werden. Dies kann eine „Verschmierung der Abbildung“²² zur Folge haben. Auch kommt es vor, dass γ -Quanten den Detektor aufgrund von gewebeabhängiger Schwächung nicht erreichen. Je dichter und je dicker das Gewebe, das durchstrahlt werden muss, desto häufiger kommt es zu Streukoinzidenzen und desto größer die Schwächung. Die Rohdaten setzen sich also aus wahren Koinzidenzen, zufälligen Koinzidenzen und Streukoinzidenzen zusammen und werden weiter von der gewebeabhängigen Schwächung beeinflusst^{22,23}.

1.4.2 Korrektur der Rohdaten

Zur Korrektur der Rohdaten, und um möglichst viele unerwünschte Koinzidenzen auszuschließen, werden verschiedene Verfahren angewandt:

Zum Ausschluss zufälliger Koinzidenzen wird die Menge von Koinzidenzen in einem Zeitintervall untersucht, das soweit auseinanderliegt, dass in ihm nur zufällige Koinzidenzen vorkommen können. Die Menge an Koinzidenzen in diesem Intervall, gibt eine Annäherung an den Anteil an zufälligen Koinzidenzen im Intervall, in dem wahre Koinzidenzen erwartet werden.²²

Zum Ausschluss von Streukoinzidenzen wird die gemessene Energie der γ -Quanten genutzt, welche durch die Streuung einen Teil ihrer Energie verlieren. Unterschiede in der Energie zweier γ -Quanten einer Koinzidenz deuten also darauf hin, dass ein γ -Quant gestreut wurde.²² Dieses Verfahren ist allerdings durch die Energieauflösung der Detektoren eingeschränkt, sodass ein erheblicher Anteil der Streukoinzidenzen übersehen wird. Weitere, teils komplexe Verfahren müssen angewandt werden, um die verbleibenden Streukoinzidenzen zu berechnen^{22,25}.

Zur Korrektur der gewebeabhängigen Schwächung der γ -Quanten wird die Computertomographie verwendet. Um die Schwächung entlang einer LOR zu errechnen, werden die Schwächungskoeffizienten aus dem Energiebereich der Röntgenstrahlen in den

Energiebereich der γ -Quanten umgerechnet. LOR mit dichterem Gewebe werden rechnerisch höher bewertet, da hier davon ausgegangen wird, dass mehr γ -Quanten absorbiert wurden und so den Detektor nicht erreicht haben²².

1.5 Zielmoleküle

Um Zellen des Prostatakarzinoms radioaktiv zu markieren wird ein geeignetes Zielmolekül benötigt. Zwei in letzter Zeit häufiger untersuchte Kandidaten sind Cholin und das prostataspezifische Membranantigen (PSMA).

Cholin, Vorläufersubstanz von Phosphatidylcholin und damit wichtiger Bestandteil von Zellmembranen, akkumuliert sich verstärkt in Zellmembranen maligner Tumore. Durch erhöhte Proliferation und stärkeren Metabolismus der Zellmembrankomponenten erhöht sich damit auch die Cholin-Akkumulation^{22,26}. Ein weiterer Kandidat ist das prostataspezifische Membranantigen (PSMA). Es hat als Zielmolekül für Radiotracer mit hohen Detektionsraten von Prostatakrebläsionen vor mehreren Jahren vermehrt Aufmerksamkeit auf sich gezogen^{27,28}. PSMA ist ein membrangebundenes Enzym, auch bekannt als Folathydrolase oder Glutamatcarboxypeptidase.^{28,29}

In einer Meta-Analyse zum Vergleich von ⁶⁸Ga- und ⁶⁴Cu -markierten PSMA PET/CT mit ¹⁸F-Fluoroethylcholin und ¹¹C-cholin PET/CT zeigten die PSMA-PET/CT eine höhere Detektionsrate von Läsionen des Prostatakarzinoms bei Patienten mit biochemischem Rezidiv. Dieser Unterschied trat insbesondere bei Patienten mit geringerem PSA-Wert auf und war bei höheren PSA-Werten statistisch nicht signifikant.³⁰

Pfister et al. verglichen auf der Basis histopathologischer Daten die Performance von ¹⁸F-Fluoroethylcholin PET/CT und ⁶⁸Ga-markiertem PSMA PET/CT und stellten eine höhere Korrekturklassifikationsrate des ⁶⁸Ga-markierten PSMA PET/CTs von 91.9% (vs. ¹⁸F-

Fluoroethylcholin 82.5%) fest. Hier wurden lokoregionäre Rezidive sowie Lymphknotenmetastasen miteinbezogen.³¹

1.6 PSMA und die Geschichte der Tracermoleküle

Das PSMA Protein hat eine zellinterne, eine transmembrane sowie eine externe Partition.²⁹ Sweat et al.³² fanden heraus, dass bei immunohistochemischer Untersuchung mit PSMA Antikörpern sich nur 46,2% der Zellen in benignem Prostatagewebe färben lassen während sich Zellen in malignem Prostatagewebe und Lymphknotenmetastasen zu 79,3% und 76,4%, respektive, färben lassen. Eine mögliche Erklärung ist, dass PSMA in weitaus höherem Maße in malignem Prostatagewebe und Lymphknotenmetastasen als in benignem Gewebe exprimiert wird. Das Molekül wird allerdings auch in Geweben außerhalb der Prostata vermutet: Afshar-Oromieh et al. konnten zeigen, dass es einen hohen Uptake eines PSMA-Radiotracers in der Niere und in den Speicheldrüsen gibt. Weniger hohen Uptake gibt es in den Tränendrüsen, der Leber, der Milz, dem Darm sowie in physiologischem Prostatagewebe.²⁷

Um das PSMA gezielt anzusteuern wird ein geeignetes Tracermolekül benötigt. Zunächst wurden für diesen Zweck monoklonale Antikörper entwickelt. Der monoklonale Antikörper ¹¹¹In-Capromab (ProstaScint®) war der erste Radiotracer, mit ¹¹¹In radioaktiv markiert, und zielte auf das intrazelluläre Epitop des PSMA. Dieser muss entweder von der Zelle internalisiert werden oder erfordert eine beschädigte Membran, also sterbende oder tote Zellen. ¹¹¹In-Capromab konnte durch diese Einschränkung im klinischen Alltag nicht überzeugen.³³ In Folge dessen wurden Antikörper kreiert, die auf das extrazelluläre Epitop des PSMA zielten. Diese Tracer, wie J415, J533, und J591, aber monoklonale Antikörper im Allgemeinen haben folgende Nachteile, die ihre Nutzung limitieren: Sie haben lange Plasmahalbwertszeiten (3-4 Tage), schlechte Tumorpenetration und niedrige Aufnahmeverhältnisse von Tumor zu normalem

Gewebe.^{29,34} Die lange Plasmahalbwertszeit bedeutet auch, dass Radionukleide mit längerer Halbwertszeit genutzt werden müssen, was zu höheren Strahlungsdosen führt.²⁹

Wie Schwarzenboeck et al.³⁵ schreiben, sind die aktuell verbreitetsten Kandidaten ein mit ⁶⁸Ga markierter niedermolekularer PSMA-Inhibitor, ⁶⁸Ga-HBED-CC-PSMA³⁶ sowie die Theranostika ⁶⁸Ga-PSMA-617 und ⁶⁸Ga-PSMA-I&T^{35,37}.

1.7 Tumorheterogenität

Genetische Heterogenität ist ein anerkanntes Merkmal malignen Tumorgewebes³⁸ und mit Resistenzbildung und Therapieversagen assoziiert.³⁹ Heterogenität beschreibt die Variationen genetischer Subtypen, Wachstums- und Angiogenesefaktoren, des Mikroumfelds individueller Tumore sowie zwischen Primär- und Sekundärtumoren. Tumorheterogenität verursacht Variationen in Proliferationsgeschwindigkeit, Vaskularisation, Zelltod und Metabolismus.⁴⁰

Das Prostatakarzinom ist laut Tolkach et al. ein Paradebeispiel für Tumorheterogenität: 95% aller Prostatakarzinome sind azinäre Adenkarzinome, allerdings weisen diese ein sehr heterogenes Spektrum auf. Prostatakarzinome sind morphologisch sehr pluriforme Tumore mit einer Mischung aus benignen Drüsen, trabekelförmigen, soliden und einzelligen Tumormustern. Zur histologischen Kategorisierung nutzen wir den Gleason-Score. Dieser ist der wohl stärkste Vorhersagewert für das Verhalten von Prostatakarzinomen. Weiterhin haben Prostatakarzinome eine nahezu einzigartige Art und Weise der Infiltration mit sehr unscharf abgrenzbarer Tumorberandung. Sie bilden sehr irreguläre Tumorzungen, die sich weit vom Tumorkern entfernen können. Aus genetischer Sicht wird es heute in 7 verschiedene Gruppen basierend auf jeweils einer gemeinsamen genetischen Veränderung eingeteilt. Diese haben allerdings aktuell keine klinische Relevanz, da sie keine relevanten Aussagen über Prognose und

Therapie treffen. Prostatakarzinome weisen auch eine signifikante intratumorale Heterogenität auf; innerhalb eines Tumors gibt es also viele verschiedene genetische Varianten bzw. Klone⁴¹.

Lindberg et. al konnten anhand einer Prostataktomieprobe zeigen, dass in allen 25 von ihnen Untersuchten Proben unterschiedliche Tumorklone zu finden waren^{41,42}. Manche benachbarten Klone auf minimaler Entfernung waren genetisch hochgradig unterschiedlich.

Laut Tolkach et al. gibt es Hinweise, dass alle Metastasen des Prostatakarzinoms aus einem gemeinsamen, monoklonalen Ursprung entstammen. Gleichzeitig gibt es Hinweise, dass Metastasen, wie der Primärtumor, selbst auch polyklonal sind und damit als Klon-Reservoir für neue Metastasen fungieren. Die Polyklonalität von Krebserkrankungen kann therapierelevant sein, denn ein hohes Maß an Tumorheterogenität ist mit Therapieresistenz assoziiert. In Metastasen können unter Therapiedruck neue, therapieresistentere Klone gebildet werden, die andere Klone auskonkurrieren. Maligne Prostatakarzinome können viele Klone mit unterschiedlichen Resistenzprofilen enthalten. Allerdings können wir, wie bereits gesagt, auch davon ausgehen, dass alle Klone auf einen monoklonalen Ursprung zurückzuführen sind, und damit eventuell auf eine gemeinsame genetische Veränderung, die allen Klonen zu eigen ist. Gelingt es, diese gemeinsame Veränderung therapeutisch ins Visier zu nehmen könnte das Prostatakarzinom effektiver therapiert werden.⁴¹

Wie Aerts et. al ausführen, sind durch die örtliche und zeitliche Heterogenität der Tumore histologische Untersuchungen durch Biopsien äußerst limitiert. Die Autoren weisen weiter darauf hin, dass es zur Unterschätzung der genetischen Variation führen kann, wenn man diese durch eine einzige Biopsie darzustellen versucht. Eine ganzheitliche Darstellung des Tumorgewebes sei so nicht möglich.⁴³

Aerts et al. bemerken in ihrer Arbeit auch, dass aufgrund der benannten Beschränkungen eine alternative Methode, die Heterogenität eines Tumors zu beurteilen, für Patienten große Vorteile

bergen könnte. Eine Möglichkeit könnte die Beurteilung der Tumorheterogenität über bildgebende Verfahren sein. Durch Bildgebung ist man in der Lage den gesamten Tumor darzustellen und dessen Progression bzw. Therapieansprechen zu verfolgen. Außerdem sei es ein Vorteil der radiologischen Bildgebung, dass sie im klinischen Alltag während der Behandlung bereits regelmäßig wiederholt wird. Sie muss nicht, wie andere molekulare Charakterisierungsmethoden, aufwendig in den Klinikalltag integriert werden. In Zukunft könnte die Tumorheterogenität durch algorithmische Analyse von Bilddaten, unter anderem durch die Texturanalyse bestimmt werden.⁴³

1.8 Grundsätze der Texturanalyse

Die Nutzung von Texturanalysen zur Interpretation von Bildmaterial ist weder neu, noch ist sie radiologischem Bildmaterial vorbehalten. Schon im Jahr 1973 wurden Texturmerkmale von Haralick et al. zur Interpretation von Mikroaufnahmen sowie Satellitenbildern eingesetzt.⁴⁴

1.8.1 Textur

Der Ursprung des Wortes „Textur“ liegt im lateinischen Wort „textura“, also „Gewebe“. ⁴⁵ Haralick et al. beschreiben Textur als „die angeborene Eigenschaft praktisch jeder Oberfläche – die Maserung von Holz, die Webart von Stoff, die Struktur von Getreide auf einem Feld, etc.“. ⁴⁴ Sie bietet „wichtige Informationen zur strukturellen Anordnung von Oberflächen und ihrer Beziehung zur umgebenden Umwelt“. ⁴⁴

Wichtige Bestandteile der Textur sind laut Rengers und Prinz die sogenannten Objektprimitive und Farbwerte bzw. Grautöne sowie deren besondere räumliche Anordnung und Wiederholung. Primitive können „Punkte, Linien oder einfachste geometrische Formen“ sein, „die zusammengesetzt eine Struktur oder ein Muster bilden“. In der digitalen Bildverarbeitung beschreibt Textur laut Regners und Prinz eine „Bildregion die sich durch die spezielle

Anordnung von Bildpunkten und deren Grauwerte zueinander auszeichnet.“ Die „Anordnung der Grauwerte“ würde dabei „als Variation sich wiederholender Muster von Grauwerten innerhalb lokaler Nachbarschaften angesehen.“⁴⁶

Grob simplifiziert entsteht laut Haralick et al. eine feine Textur, wenn es in der räumlichen Anordnung von Grautönen in einem Bildausschnitt kein Muster und eine große Variation von Grautönen gibt. Ist die räumliche Anordnung der Bildpunkte definierter so resultiert daraus eine gröbere Textur.⁴⁴

1.8.2 Texturanalyse und Texturmerkmale

Zur Ableitung quantifizierbarer Texturmerkmale aus medizinischem Bildmaterial gibt es verschiedene Herangehensweisen. Das können strukturelle, modellbasierte oder auch statistische Texturmodelle sein. In der medizinische Bildanalyse haben sich statistische Texturmodelle durchgesetzt; sie beschreiben die Verteilung von Grautönen im Bild sowie deren Beziehung zueinander.⁴⁷ Statistiken erster Ordnung wie zum Beispiel der Median, der Mittelwert oder die Kurtosis („peakedness“) beschreiben die Verteilung der Grautöne auf dem Histogramm der Pixelintensität. Statistiken erster Ordnung erfassen allerdings nicht die Beziehung einzelner Bildpunkte zueinander.⁴⁸

Haralick et al. nehmen an, dass Informationen über die Textur von Bildmaterial in der durchschnittlichen räumlichen Beziehung von Grautönen im Bild zueinander liegen.⁴⁴ Diese Information kann mittels Statistiken zweiter Ordnung extrahiert werden, basierend beispielsweise auf den Gray-Level Co-Occurrence Matrices (glcm), auch Gray-Tone Spatial Dependence Matrices genannt.⁴⁴ Sie beschreiben die Beziehung zweier Bildpunkte zueinander und bilden die Frequenz ab, mit der die verschiedenen Grauwertintensitäten in verschiedenen Ebenen angrenzend an jede andere Grauwertintensität vorkommen (siehe Abb. 1).⁴⁸ Statistiken erster und zweiter Ordnung können nützlich sein, wenn die die Grautöne der Bildpunkte

normalverteilt sind.⁴⁷ Ein Radiomic feature, das aus GLCM's abgeleitet werden kann, ist zum Beispiel „Joint Entropy“; ein Maß für die Zufälligkeit mit der Grautonintensitäten in der Umgebung auftreten⁴⁹.

Wenn das Histogramm der Bildpunkte aber keiner Gaußkurve entspricht, diese also nicht normalverteilt sind, sind Statistiken höherer Ordnung nützlicher⁴⁷. Mithilfe Neighbourhood Gray-tone Difference Matrices (ngtdm) können die Beziehungen eines Voxels mit mehreren umgebenden Voxeln und so Statistiken höherer Ordnung wie zum Beispiel Kontrast, Grobkörnigkeit (coarseness) und Lebendigkeit (busyness) errechnet werden.^{48,49} Jeder Grauton wird hier als Differenz zwischen sich selbst und dem mittleren Grauwert der benachbarten Voxel beschrieben.^{49,50}

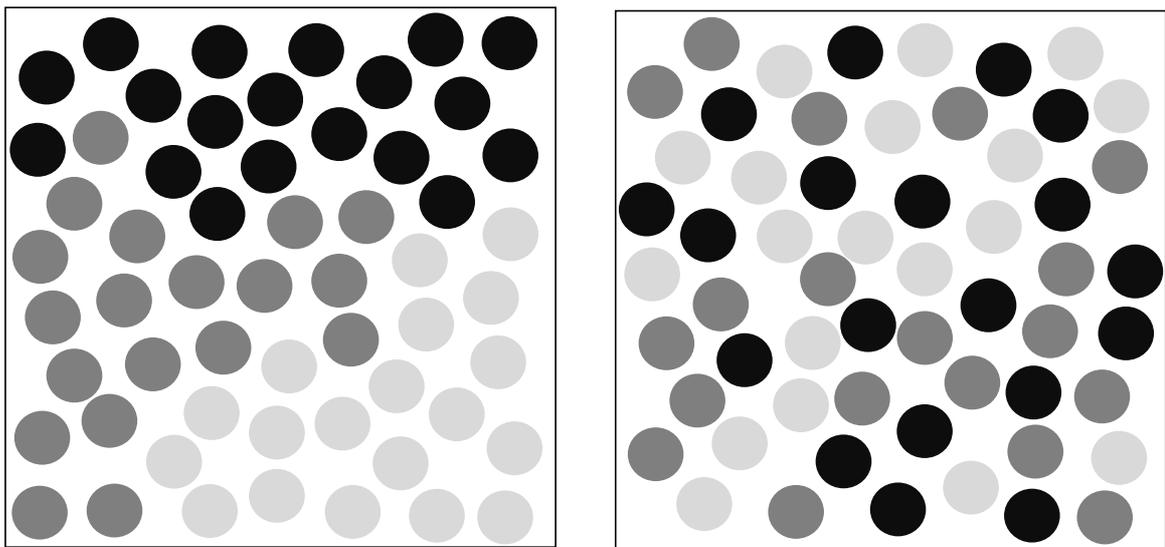


Abbildung 1: Veranschaulichung statistischer Methoden in der quantitativen Bildverarbeitung. Beide Bildausschnitte (Rechtecke) enthalten die gleiche Anzahl von Bildpunkten verschiedener Grauwertintensitäten. Histogrammbasierte Statistiken erster Ordnung wie zum Beispiel der Median oder die Kurtosis erkennen zwischen den beiden Bildausschnitten keinen Unterschied. Statistiken zweiter Ordnung allerdings, mittels Gray-Level-Cooccurrence-Matrices (GLCM) errechnet, bilden die Frequenz ab, mit der die verschiedenen Grauwertintensitäten in verschiedenen Ebenen in der Umgebung aller Grauwertintensität vorkommen. Radiomic-Features die mit GLCM abgeleitet werden, können so die unterschiedliche Verteilung von Grauwertintensitäten in den beiden Bildausschnitten darstellen.⁴⁸ (Quelle: in Anlehnung an Lubner et. al.⁴⁸).

1.9 Radiomics

Radiologisches Bildmaterial wurde in der Vergangenheit auf größtenteils visuelle Art und Weise bewertet.⁵¹ Die visuelle, qualitative Bildbeurteilung ist aber immer auch eine subjektive Beurteilung. Um Befunde quantifizierbarer und somit objektiver zu gestalten, gibt es in der CT-Befundung begrenzte Möglichkeiten. Zum Beispiel kann der HU-Wert oder der Durchmesser von Strukturen genutzt werden.

Im Wissenschaftsgebiet Radiomics können heute große Mengen quantitativer Information aus dem verfügbaren Bildmaterial extrahiert werden. Man erhofft sich aus diesen Daten Rückschlüsse auf die Pathophysiologie des Gewebes ziehen zu können. Radiomics ist auch als eine große Sammlung von Daten zu verstehen, im besten Falle in offener Datenbank geteilt, zur Generierung und zum Testen von Hypothesen. In Einbeziehung anderer relevanter Patientendaten, wie zum Beispiel der Laborwerte, klinischer Befunde oder pathologischer Untersuchungen können mittels Radiomics wichtige zusätzliche Informationen gewonnen werden. Radiomics kann so zu diagnostischen Zwecken verwendet werden, aber auch Voraussagen zu Prognose und Therapieansprechen treffen. Im Bereich der Onkologie fand bisher das Gros der Forschung statt.^{51,52}

Radiomics kann auf verschiedenes radiologisches Bildmaterial, wie aus der CT, MRT oder PET, angewendet werden und basiert auf der Analyse verschiedener Bildmerkmale, der Radiomic-features⁵². Diese setzen sich zusammen aus Statistiken erster (first-order-Statistiken) und höherer Ordnung (z.B. GLCM oder NGTDM), sowie aus den Formmerkmalen (shape-features), wie das Volumen, die maximale Oberfläche oder auch die Sphärizität (sphericity).⁵³ Zur Analyse der Bildmerkmale, so Lambin et al. werden Volumes of Interest (VOI's) manuell oder semi-automatisch/automatisch mithilfe spezieller Programme segmentiert. Aus diesen VOI kann eine nahezu unerschöpfliche Menge Radiomic-features extrahiert werden und gleichzeitig in die Analyse miteinbezogen werden. Diese Modelle versuchen den vorhandenen Datensatz möglichst

genau abzubilden. Wenn zu viele Features mit in die Berechnung eingehen, kann es zum „overfitting“ (deutsch: Überanpassung) des Modells auf den Datensatz kommen. Hierbei bildet das Modell den Datensatz auf zu komplizierte Weise ab und ist damit nicht mehr auf neue Datensätze übertragbar. Um dem zu begegnen, greift man auf eine vorherige oder spätere Featureselektion zurück. Hierbei können mittels unterschiedlicher Methoden relevante Bildmerkmale (features) ausgewählt werden; beispielsweise die, mit der größten Robustheit gegenüber der Veränderung verschiedener Parameter.⁵² Außerdem können an dieser Stelle auch redundante Bildmerkmale, die stark miteinander korrelieren, ausgeschlossen und die Bildmerkmale, mit der höchsten Variabilität zwischen den Patienten ausgewählt werden. Man geht davon aus, dass Bildmerkmale mit großer Variabilität informativer sind.⁵¹ So geht eine vergleichsweise geringe Zahl relevanter, bzw. wenig redundanter Bildmerkmale in die Berechnung ein.

Einen wichtigen Grundstein haben Aerts et al. 2014 gelegt, als sie zeigen konnten, dass mittels Radiomics in Patienten mit Bronchialkarzinom sowie Kopf- und Halskarzinom Bildmerkmale identifiziert werden können, die möglicherweise prognostische Hinweise zum Krankheitsverlauf geben können⁴³.

Auch zur Lymphknotendiagnostik im CT mit Radiomics-Methoden gibt es bereits mehrere Publikationen (s. auch 4.2.1), z.B. kreierten Sha et al. ein mathematisches Modell, dass mit einer Korrektklassifikationsrate von 88,2% mediastinale Lymphknotenmetastasen des NSCLC (non small-cell lung carcinoma) von benignen Lymphknoten unterscheiden kann.⁵⁴

Bei Patienten mit Prostatakarzinom zeigten Untersuchungen der CT-Dichte mittels Radiomics einen signifikanten Dichteunterschied in PET-positiven Lymphknoten im Vergleich zu PET-negativen Lymphknoten⁵⁵

Auch in der MRT-basierten Prostatadiagnostik ist Radiomics in den Fokus gerückt, da die klassische MRT-Diagnostik zwar hilfreich in der Diagnostik klinisch signifikanter PCa ist, aber eine hohe IOV (interobserver variability, deutsch: Variabilität zwischen Beobachtern) aufweist.⁵⁶ Die Maschinenlernen-Vorhersagemodelle von Algohary et al. bieten zum Beispiel eine signifikante Verbesserung der Genauigkeit der Diagnose von PCa anhand von MRT-Bildern in Patienten in Active Surveillance (deutsch: aktive Überwachung) im Vergleich zur Befundung durch Radiologen.^{56,57} Außerdem stellten Antonelli et al. fest, dass im Vergleich mit einem erfahrenen Radiologen, Maschinenlernen-Klassifikatoren Läsionen mit einem Gleason-Score-4 Anteil mit größerer Wahrscheinlichkeit vorhersagen.^{56,58} Bei der Analyse von Texturmerkmalen in MRT-Bildern konnte gezeigt werden, dass mehrere Texturmerkmale signifikante Unterschiede zwischen malignem und karzinomfreiem Prostatagewebe aufweisen.⁵⁹

1.10 Hypothese

Zusammenfassend ergibt sich unsere Fragestellung aus folgender Problematik: Die Leistung der CT ist im Lymphknotenstaging bei Patienten mit Prostatakarzinom sehr eingeschränkt.¹⁷ Das ⁶⁸Ga-markierte PSMA PET/CT, die mögliche Alternative mit, in der bisherigen Literatur, deutlich besserer Leistung¹⁸, ist bislang im Primärstaging nicht leitliniengerecht.¹ Außerdem verursacht das PET-CT eine deutlich höhere Strahlenbelastung als das CT allein.⁶⁰

Da Radiomics-Methoden in der bisherigen Literatur gute Ergebnisse in der Unterscheidung maligner und benigner Lymphknoten, erzielten^{54,61,62}, wollten wir folgende Frage beantworten:

Eignen sich Radiomics-Methoden in der Lymphknotendiagnostik bei Patienten mit Prostatakarzinom als zusätzliche Methoden oder Alternative zu konventionellen Methoden bzw. zum ⁶⁸Ga-PSMA PET/CT?

Zur Beantwortung der Fragestellung stellen wir folgende Hypothesen auf:

Radiomics-basierte Klassifikationsalgorithmen können den ^{68}Ga -PSMA PET/CT-Status von Lymphknoten nur auf Basis von CT-Bildern besser vorhersagen als Radiologen.

Zusätzlich wollten wir untersuchen, ob die Algorithmen einen Mehrwert liefern können, wenn Radiologen keine sichere Entscheidung treffen können.

Wir stellen die zusätzliche Hypothese auf, dass die Kombination von Klassifikationsalgorithmen und Radiologen eine bessere Vorhersageleistung erzielen als letztere allein.

Aufgrund der bislang vielversprechenden Literatur zur Leistung des ^{68}Ga -PSMA PET/CT in der Lymphknotendiagnostik bei Patienten mit PCa und aufgrund des zur Verfügung stehenden großen ^{68}Ga -PSMA PET/CT Datensatzes wählten wir den PET-CT-Status der Lymphknoten als Goldstandard für unsere Untersuchungen. Im Hinblick auf die Fragestellung kann die Verifikation unserer Hypothesen eine erste Tendenz geben. Im Anschluss müssten, um eine klinische Anwendbarkeit zu ermöglichen, deutlich aufwendigere Studien mit histopathologischer Korrelation folgen.

2. Material und Methoden

Diese retrospektive Studie wurde von der Ethikkommission der Charité Universitätsmedizin Berlin genehmigt. Auf die Notwendigkeit einer schriftlichen Einwilligung wurde vom Ethikprüfungsausschuss aufgrund der retrospektiven Gestaltung dieser Studie verzichtet

2.1 Studienpopulation

447 Patienten ($68,7 \pm 7,54$ [45-87] Jahre) die zwischen September 2009 und April 2017 ein ^{68}Ga -PSMA PET/CT mit einer kontrastverstärkten CT-Untersuchung erhielten, wurden in die Studie aufgenommen. Der PSA lag durchschnittlich bei $20,9 \pm 94,6$ [0-1423] ng/ml. Insgesamt wurden 2125 Lymphknoten identifiziert und segmentiert, von denen 369 PET-positiv und 1756 PET-negativ waren. Dies entspricht durchschnittlich etwa 0,83 positiven und 3,93 negativen Lymphknoten pro Patient. Im PET nicht eindeutig beurteilbare Lymphknoten wurden nicht segmentiert.

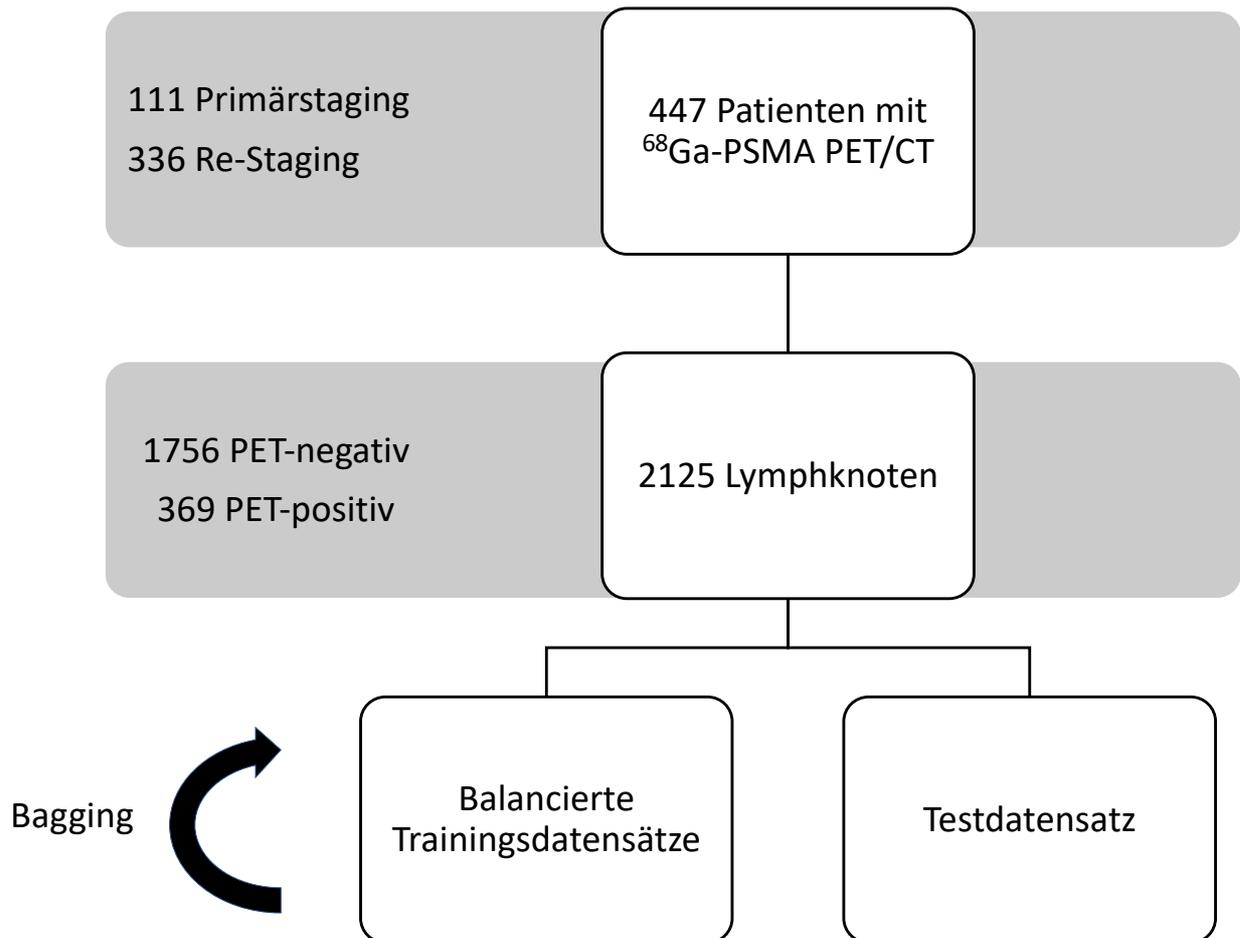


Abbildung 2: Studienpopulation und Erstellen der Datensätze.

Nach Segmentierung der Lymphknoten wurde, um die Überzahl der PET-negativen Lymphknoten auszugleichen ein, bzw. nach Bagging mehrere, balancierte Trainingsdatensätze erstellt.

2.2 Positronen-Emissions-Tomographie-Tracer

^{68}Ga wurde aus einem handelsüblichen $^{68}\text{Ge}/^{68}\text{Ga}$ Generator (Eckert & Ziegler Radiopharma GmbH, Berlin, Deutschland) extrahiert und anschließend mit PSMA- HBED-CC (ABX GmbH, Radeberg, Deutschland) mit ^{68}Ga markiert.²⁷

2.3 Bildgebungsprotokoll

Die PET/CT-Bildgebung wurde ca. 60 min nach intravenöser Injektion von ca. 120 MBq ^{68}Ga -PSMA-HBED-CC durchgeführt. Es wurde ein Gemini Astonish TF 16^ PET/CT Scanner

(Phillips Medical Systems) verwendet.⁶³ Für alle PET-Scans wurde der 3-D-Erfassungsmodus verwendet. Koronale, axiale und sagittale Schichten wurden rekonstruiert (144 Voxel mit 4 mm³, isotrop). Vor dem PET-Scan wurde eine niedrig dosierte CT zur anatomischen Kartierung und Schwächungskorrektur durchgeführt (effektives Röhrenstrom-Zeitprodukt von 30-40 mAs; Röhrenspannung 120 kV). Die CT-Scans wurden anschließend mit der PET fusioniert. Jede Tischposition wurde für etwa 1,5 Minuten mit einer Überlappung von 50% erfasst.

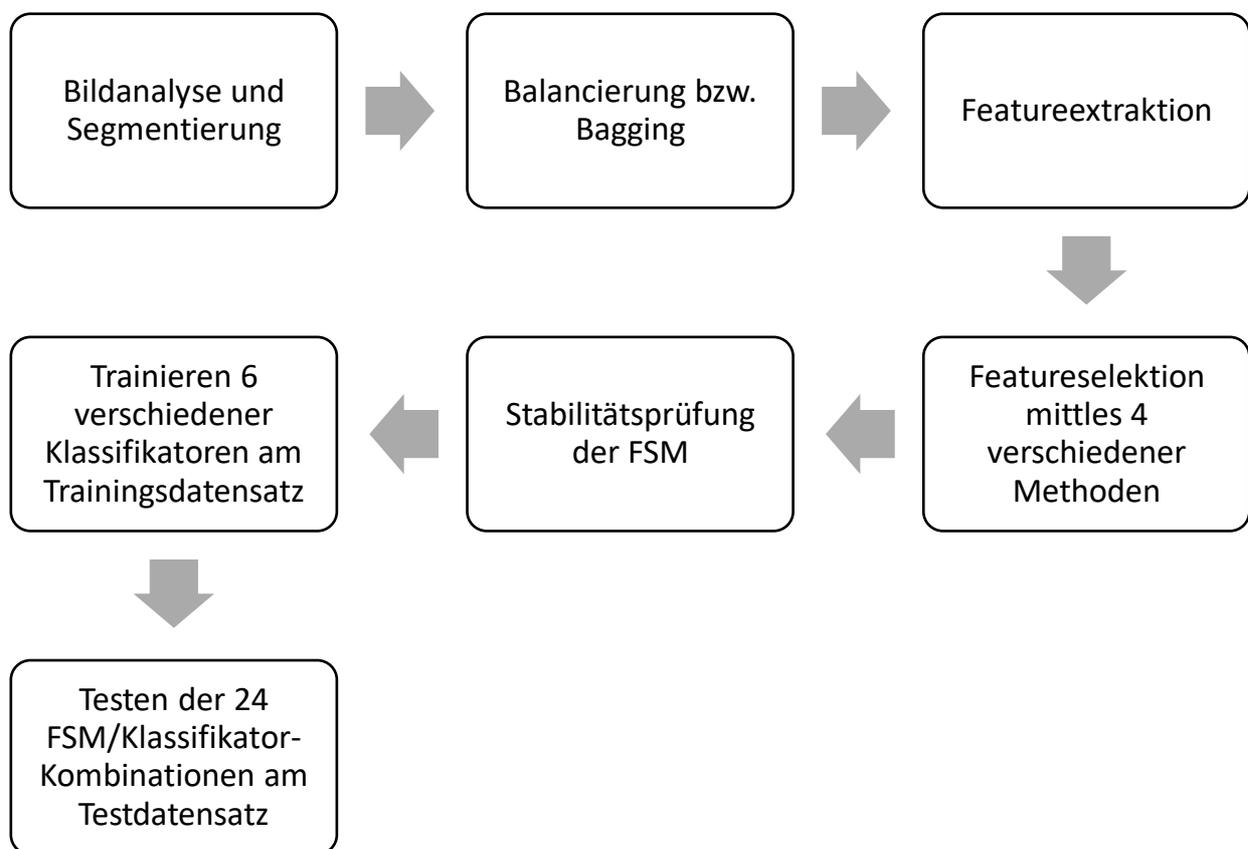


Abbildung 3: Prozess der Radiomics-Methode.

Der Prozess der Radiomics Methode begann mit der Segmentierung ausgewählter Lymphknoten. Der so entstandene Datensatz wurde zu Test- und Trainingszwecken aufgeteilt. Der Trainingsdatensatz wurde, um die Überzahl PET-negativer Lymphknoten auszugleichen, einfach, bzw. später mittels Bagging-Verfahren, mehrfach balanciert. Im nächsten Schritt wurden Radiomics-Features aus den Bildregionen extrahiert um dann die relevantesten Features mittels vier verschiedener Featureselektionsmethoden (FSM) auszuwählen. Nach einer Stabilitätsprüfung der FSM wurden sechs verschiedene Klassifikationsalgorithmen (Klassifikatoren) trainiert. Die 24 so entstandenen FSM/Klassifikator Kombinationen wurden am Testdatensatz getestet. (FSM= Featureselektionsmethode)

2.4 Bildanalyse und Segmentierung der Lymphknoten

Für jeden Patienten wurden maximal fünf Lymphknoten ausgewählt. PET-positivität galt als primäres Auswahlkriterium (siehe Abb. 4 und 5), größere wurden vor kleineren Lymphknoten bevorzugt. Weiterhin wurden systematisch PET-negative Lymphknoten ausgewählt. Jede fokale Erhöhung der Aufnahme von ^{68}Ga -PSMA-HBED-CC höher als die Hintergrundaufnahme, die nicht mit einer physiologischen Aufnahme des Tracers assoziiert werden konnte, wurde als Hinweis auf einen malignen Prozess bewertet. Die Einstufung in maligne und benigne Lymphknoten wurde zusätzlich durch einen Radiologen mit 5-10 Jahren Erfahrung sowie im Konsens mit einem zweiten Radiologen mit mehr als fünf Jahren Erfahrung mit PSMA PET/CT geprüft.

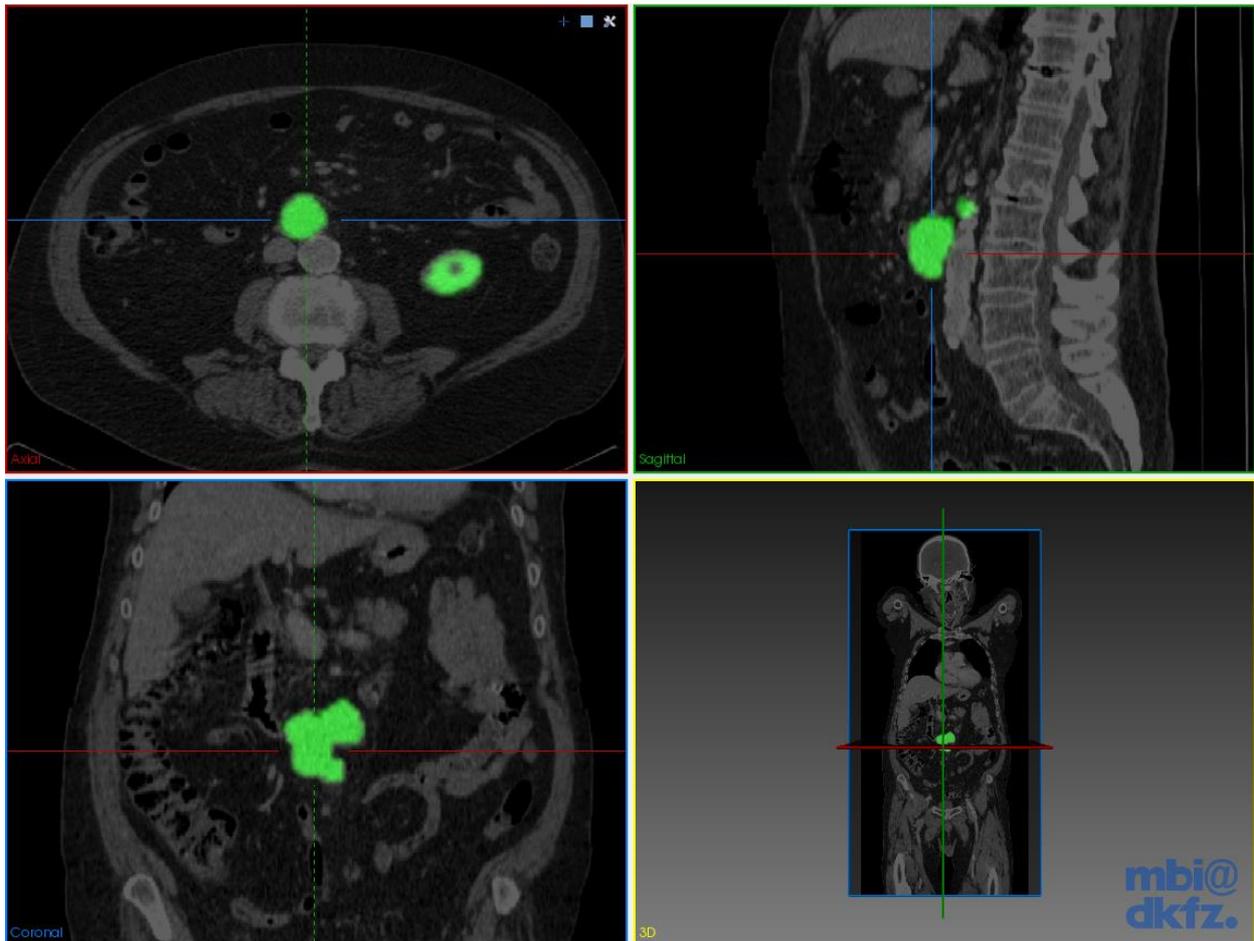


Abbildung 4: Multiplanares PET-CT Fusionsbild eines Patienten mit retroperitonealen Lymphknotenmetastasen eines Prostatakarzinoms.

Für jeden Patienten wurden vom Doktoranden maximal fünf Lymphknoten ausgewählt. PET-positivität (grün) galt als primäres Auswahlkriterium. Zweitrangig wurden größere vor kleineren Lymphknoten bevorzugt. Jede fokale Erhöhung der Aufnahme ^{68}Ga -PSMA-HBED-CC höher als die Hintergrundaufnahme, die nicht mit einer physiologischen Aufnahme des Tracers assoziiert werden konnte, wurde als Hinweis auf einen malignen Prozess bewertet. Beispielfhaft kann man in der axialen Schicht (links oben) eine physiologische Traceraufnahme (grün) der linken Niere erkennen. Zur Darstellung und Segmentierung verwendeten wir das Computerprogramm MITK des deutschen Krebsforschungszentrums.

Das Computerprogramm MITK 2016 3.0 (Deutsches Krebsforschungszentrum, Abteilung Medical Image Computing, Im Neuenheimer Feld 280, 69120 Heidelberg) wurde zur Darstellung und Segmentierung der Lymphknoten verwendet. Das Segmentierungs-Toolkit in MITK bietet dem Nutzer mehrere Optionen zur Segmentierung. Maßgeblich unterscheiden sich dabei die manuelle sowie die semiautomatische Art zu segmentieren. Bei letzterer Methode

werden wenige Schichten einer Struktur manuell segmentiert; die dazwischenliegenden Schichten werden mittels Interpolation ergänzt.

Die Segmentierung wurde vom Autor zu Beginn semiautomatisch durchgeführt. Wir fanden, dass die Grenzen der zu segmentierenden Strukturen in MITK aber mittels Interpolation nur unzureichend beschrieben wurden und entschieden uns dazu, vormalig semiautomatisch generierte Segmentationen Schicht-für Schicht manuell anzupassen. Die große Mehrheit der Lymphknoten wurden anschließend, deutlich zeitintensiver, aber für uns exakter, vollständig manuell segmentiert (siehe Abb. 5).

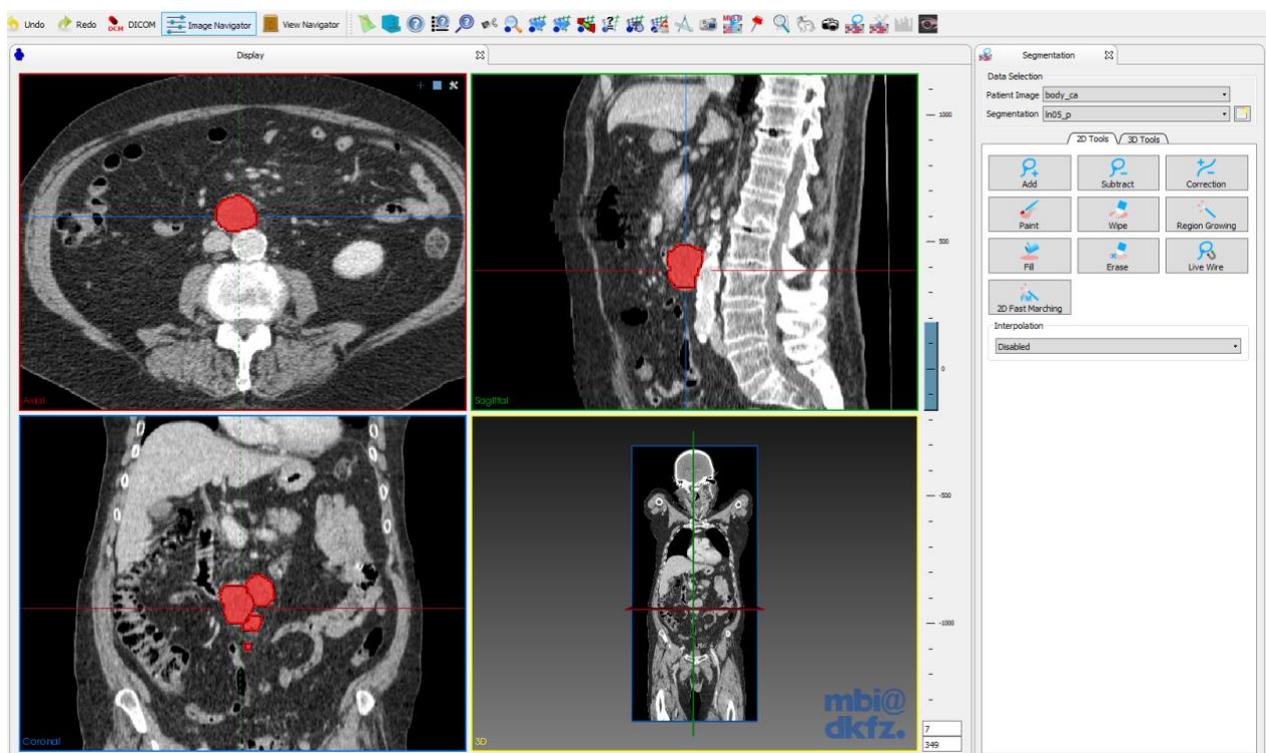


Abbildung 5: Segmentierung von Lymphknoten mittels MITK.

Ausgewählte Lymphknoten wurden mittels des Segmentation-Toolkits des Computerprogramms MITK in der axialen Ebene segmentiert. Die semiautomatische Segmentierung mittels Interpolation beschrieb die Lymphknotengrenzen nur unzureichend, weshalb wir uns für die manuelle Segmentierung entschieden.

2.5 Statistische Auswertung

Die in dieser Arbeit verwendete Methodik der statistischen Analyse von Radiomics-Merkmalen wurden in der Arbeitsgruppe entwickelt und in Zusammenarbeit mit dem Statistiker Dr. Paul

Schmidt mit der statistischen Programmiersprache R⁶⁴ durchgeführt. Der Anteil des Autors liegt in der Mitarbeit an der Studienkonzeption, der Planung und des Aufbaus, sowie der Qualitätskontrolle des Datensatzes inklusive der Ausarbeitung der Segmentierungsmethodik und Durchführung der Segmentierung sowie der Validierung und Interpretation der Ergebnisse.

Zunächst wurde der vorhandene Datensatz nach Zufallsprinzip balanciert. Die Selektion relevanter Features wurde mittels univariater Filtermethoden, bzw. Featureselektionsmethoden (FSM) durchgeführt. Statistische und maschinelle Lernmodelle (Classifier, deutsch: Klassifikatoren) wurden zur Vorhersage des positiv-negativ-Status von Lymphknoten erstellt und ausgewertet. Um die Leistung der Methoden zu beurteilen, haben wir die Methoden an einem Teil der Daten (Trainingsdatensatz) trainiert und an einem anderen Teil (Testdatensatz) getestet.

2.5.1 Balancing

Insgesamt wurden 2125 Lymphknoten identifiziert und segmentiert, 369 davon waren PET-positiv und 1756 PET-negativ. Um das Ungleichgewicht negativer und positiver Lymphknoten auszugleichen, wurden der Trainingsdatensatz nach Zufallsprinzip balanciert.

2.5.2 Bagging

Die zufällige Balancierung kreiert eine gewisse Verzerrung des Trainingsdatensatzes. Um diesen Effekt auszugleichen, kann die Balancierung mehrfach wiederholt werden. Die Algorithmen werden dann am jeweils entstandenen Trainingsdatensatz trainiert und auf den Testdatensatz angewendet. Diese Ergebnisse können dann aggregiert werden.

Es wurde zunächst eine vollständige Auswertung, inklusive Training und Testen der Algorithmen, ohne Bagging durchgeführt. Als zusätzliche Maßnahme wurde der gesamte Prozess im Anschluss unter Anwendung des Bagging-Verfahrens wiederholt. Im Ergebnisteil sind Ergebnisse mit und ohne Bagging aufgeführt.

2.5.3 Extraktion

Radiomics Merkmale wurden mit dem pyradiomics Paket extrahiert (Version 2.0.0, <https://github.com/Radiomics/pyradiomics>)⁶⁵. Die Bilddaten wurden mit dem in Pyradiomics eingebauten sitkBSpline Interpolator auf isotropisch 1.0 x 1.0 x 1.0 mm umgerechnet. Folgende Gruppen von radiomischen Merkmalen wurden extrahiert: shape, first order, glcm, gldm, glrlm and glszm. Die Voxelwerte wurden mit einer Bin-Größe von 25 HU zur Feature-Stabilisierung analysiert.

2.5.4 Featureselektion

Es stehen uns heute unzählige verschiedene radiomic features zur Verfügung. Die Nutzung aller verfügbaren Merkmale für die Modellierung würde unweigerlich zu einer übergenauen Beschreibung des vorhandenen Datensatzes führen; im Umkehrschluss ist ein solches Modell weniger gut auf neue Datensätze anwendbar. Um dieses „overfitting“ (deutsch: Überanpassung) zu vermeiden nutzten wir verschiedene Methoden zur Featureselektion (Featureselektionsmethoden, FSM). Jede Methode sollte aus dem Pool an extrahierten Bildmerkmalen die 20 für unseren Datensatz relevantesten auswählen⁵².

Wir nutzen vier verschiedene Methoden zur Selektion relevanter Features: Die Methode nach Wilcoxon, Area under the Curve (AUC, deutsch: Fläche unter der Kurve), Maximierung der mutual information (mi) und der Ansatz der maximum relevance minimum redundancy (mrmi).

Wilcoxon

Hier werden die metrischen Features auf einen signifikanten Unterschied entlang der Antwort-Kategorien, also dem Lymphknotenstatus untersucht. Bei einem größeren Stichprobenumfang wie dem unseren ist das fast immer der Fall, somit ist eine Auswahl der Features nach Größe des p-Werts ungeeignet. Stattdessen wählen wir hier den Abstand der Lageparameter Der Test nach Wilcoxon prüft den Median der Differenzen zwischen den Antwort-Kategorien.

AUC

Entlang des gesamten Wertebereich eines Features wird, wenn der jeweilige Wert als Grenzwert für die Prognose des Lymphknotenstatus gewählt würde, die Sensitivität und die Falsch-Positiv-Rate (1 - Spezifität) geprüft. Aus diesen werten entsteht die ROC (reciever operating characteristic) -Kurve, mit der Sensitivität auf der y – Achse sowie der Falsch-positiv-Rate auf der x-Achse. Eine Auswahl der Features erfolgt dann anhand der Größe der Fläche unter der so entstandenen ROC-Kurve, der AUC.

Mutual information (mi)

Nach diesem Ansatz werden dann die Features ausgewählt, welche die höchste mutual information mit der Response-Variable (deutsch: Antwort-Variabel), also dem Lymphknotenstatus aufweisen, dieser also am ähnlichsten sind.

Maximum relevance minimum redundancy (mrmi)

Dieser Ansatz basiert auf dem der mutual information. Neben einer Maximierung der mutual information mit der Response-Variable wird aber gleichzeitig versucht, die Redundanz mit anderen, Features zu minimieren.

Die so extrahierten Features wurden anschließend standardisiert.

2.5.5 Stabilität

Um die Selektionsmethoden auf ihre Stabilität hin zu untersuchen, simulierten wir mittels R eine n-malige Teilung des Datensatz und untersuchten inwieweit die Featureselektion beider Partitionen übereinstimmte. Das Ergebnis ist in folgender Grafik (Abbildung 6) festgehalten:

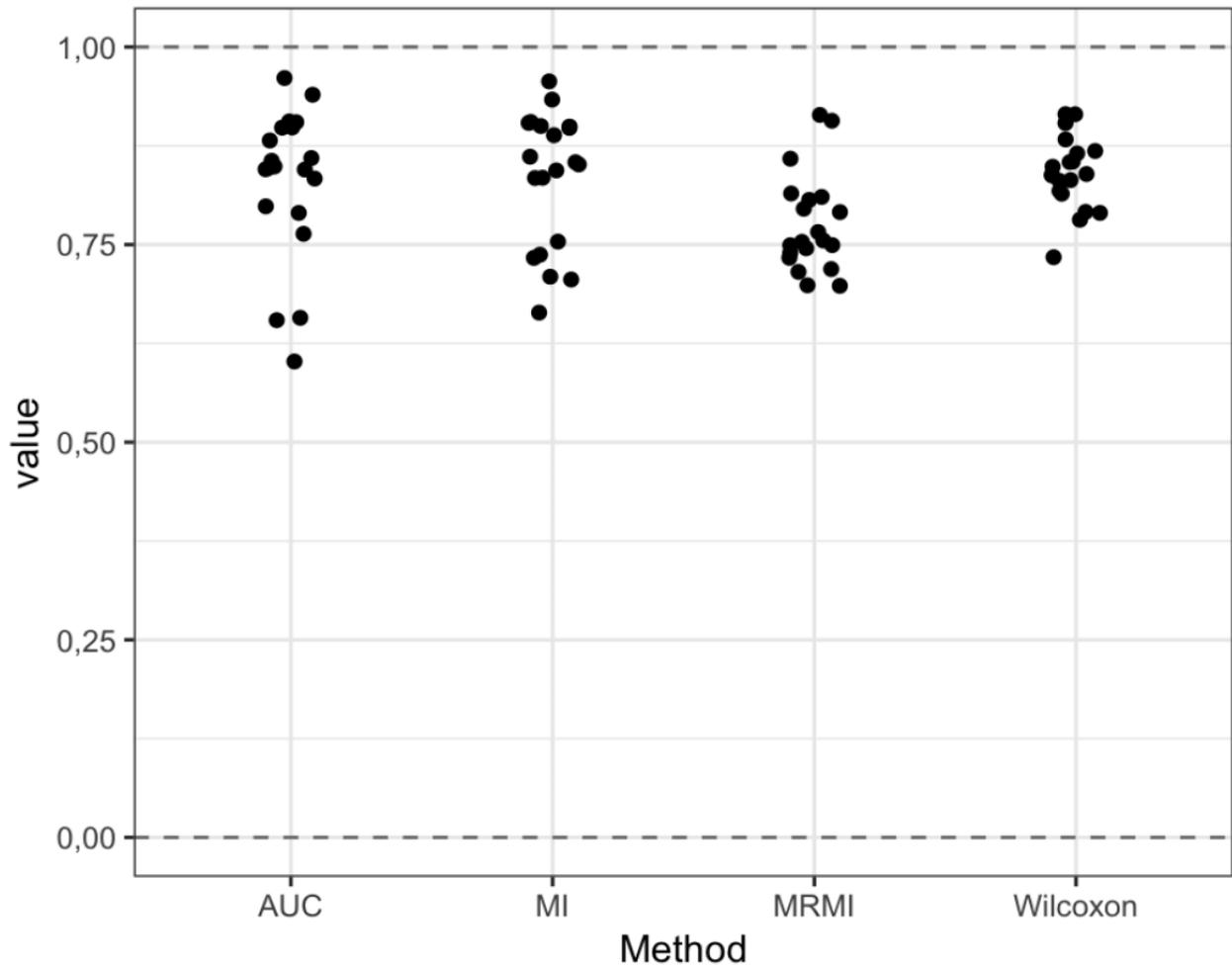


Abbildung 6: Vergleich verschiedener Feature-Selektionsmethoden hinsichtlich ihrer Stabilität.

Der Datensatz wurde n -mal geteilt und die Feature-Selektionsmethoden AUC (area under the curve), MI (mutual information), MRMI (maximum relevance minimum redundancy) sowie Wilcoxon darauf angewendet. Der Grad der Übereinstimmung beider Partitionen nach der Teilung wird hier an der y-Achse als „Value“ dargestellt, mit Wert 1 als maximaler Übereinstimmung. Die Anwendung von Wilcoxon erzielt hier als Feature-Selektionsmethode die stabilsten Ergebnisse. (in Zusammenarbeit mit beauftragtem Statistiker erstellt).

2.5.6 Training

Zur Vorhersage der Malignität von Lymphknoten teilten wir den Datensatz in einen Trainings- und einen Testdatensatz. Am Trainingsdatensatz wurden nach jeder FSM sechs mathematische Vorhersagemodelle (Klassifikatoren) trainiert. So entstanden 24 verschiedene FSM-Klassifikator Kombinationen. Folgende Klassifikatoren wurden verwendet: lineare Diskriminanzanalyse (lda), logistische Regression (log), partial least squares (pls), support

vector machines (svm), ein neuronales Netzwerk (multilayer perceptron, mlp) sowie Klassifikationsbäume (rpart). - mlp) sowie Klassifikationsbäume (rpart, recursive partitioning).

2.5.7 Testen

Die 24 trainierten FSM-Klassifikator Kombinationen können nun auf den separaten Testdatensatz angewendet werden. Dieser Schritt dient der Validierung der Prognosen der erzeugten Algorithmen.

2.6 Vergleich der Klassifikatoren mit der Einschätzung von Radiologen

In einem nächsten Schritt verglichen wir, wie die Klassifikationsalgorithmen im Vergleich mit der Einschätzung von Radiologen abschneiden. Außerdem werteten wir aus ob die Algorithmen einen Mehrwert liefern können, wenn Radiologen keine sichere Entscheidung treffen können. Dafür ließen wir die zwei Radiologen jeden beobachteten Lymphknoten in vier Klassen einordnen (0=sicher nicht maligne, 1= wahrscheinlich nicht maligne, 2= wahrscheinlich maligne, 3= sicher maligne). Für die erste Auswertung, zum Vergleich der Beurteilung der Radiologen mit den Ergebnissen der mathematischen Modelle, fassten wir die Klassen 0 und 1 sowie 2 und 3 in zwei Gruppen zusammen. In einer zweiten Auswertung nutzten wir die Vorhersagen der Algorithmen in den mittleren Kategorien, 1 und 2, und analysierten, ob sie so einen Mehrwert zur Beurteilung der Radiologen beitragen können.

3. Ergebnisse

Wie im Methodikteil erläutert nutzten wir, um ein overfitting zu vermeiden, Methoden zur Selektion der relevantesten Features. Jede Methode sollte die 20 relevantesten Features auswählen, welche in der folgenden Tabelle in absteigender Gewichtung dargestellt sind.

Table 3: Selektion der zwanzig relevantesten Features.

Sortierung je Selektionsmethode in absteigender Relevanz. Um ein overfitting durch die Einbeziehung zu vieler Bildmerkmale zu vermeiden, nutzten wir vier verschiedene Feature-Selektionsmethoden. Diese sollte die 20 relevantesten Features auswählen, welche dann in zur Berechnung genutzt wurden.

Selektionsmethode	Wilcoxon	Area under the curve
Feature	firstorder_Median	firstorder_Median
	firstorder_Mean	firstorder_Mean
	glcm_Idmn	glcm_Idn
	glcm_Idn"	glcm_Idmn
	firstorder_10Percentile	firstorder_10Percentile
	gldm_DependenceNonUniformityNormalized	firstorder_Skewness
	firstorder_90Percentile	firstorder_Kurtosis
	shape_SurfaceVolumeRatio	glrlm_LongRunHighGrayLevelEmphasi
	glrlm_ShortRunLowGrayLevelEmphasis	firstorder_Maximum
	glrlm_RunLengthNonUniformityNormalized	glrlm_ShortRunLowGrayLevelEmphasis
	gldm_LowGrayLevelEmphasis	gldm_LowGrayLevelEmphasis
	glrlm_LowGrayLevelRunEmphasis	gldm_LargeDependenceHighGrayLevelEmphasis
	original_glrlm_RunPercentage	glrlm_LowGrayLevelRunEmphasis
	glszm_ZoneEntropy	firstorder_90Percentile
	gldm_SmallDependenceLowGrayLevelEmphasis	ldm_SmallDependenceLowGrayLevelEmphasis
	glrlm_ShortRunEmphasis	firstorder_Energy
	glcm_Idm	firstorder_TotalEnergy
	glcm_Id	gldm_DependenceNonUniformityNormalized
	glszm_LowGrayLevelZoneEmphasis	gldm_GrayLevelNonUniformity
	gldm_DependenceVariance	glrlm_GrayLevelNonUniformity
Selektionsmethode	Mutual Information	Maximum Relevance Minimum Redundancy
Feature	firstorder_Median	firstorder_Median
	firstorder_Mean	firstorder_Kurtosis

	firstorder_Kurtosis	original_glrlm_LongRunHighGrayLevelEmphasis
	glcm_Idn	glcm_Idmn
	glcm_Idmn	firstorder_Skewness
	firstorder_10Percentile	glszm_LargeAreaHighGrayLevelEmphasis
	firstorder_Skewness	firstorder_10Percentile
	glrlm_LongRunHighGrayLevelEmphasis	firstorder_TotalEnergy
	glrlm_ShortRunLowGrayLevelEmphasis	gldm_LowGrayLevelEmphasis
	gldm_LargeDependenceHighGrayLevelEmphasis	firstorder_Maximum
	gldm_LowGrayLevelEmphasis	firstorder_RootMeanSquared
	firstorder_Maximum	gldm_DependenceNonUniformityNormalized
	firstorder_Energy	glszm_ZoneVariance
	firstorder_TotalEnergy	firstorder_Mean
	LargeAreaHighGrayLevelEmphasis	glszm_SmallAreaHighGrayLevelEmphasis
	glrlm_LowGrayLevelRunEmphasis	gldm_SmallDependenceLowGrayLevelEmphasis
	firstorder_90Percentile	firstorder_InterquartileRange
	gldm_SmallDependenceLowGrayLevelEmphasis	gldm_LargeDependenceHighGrayLevelEmphasis
	gldm_DependenceNonUniformityNormalized	firstorder_90Percentile
	gldm_GrayLevelNonUniformity	glcm_Idn

Anschließend wurden anhand der ausgewählten Bildmerkmale sechs verschiedenartige mathematische Modelle kreiert. Die erstellten Klassifikationsbäume sind graphisch darstellbar (s. Abbildung 7).

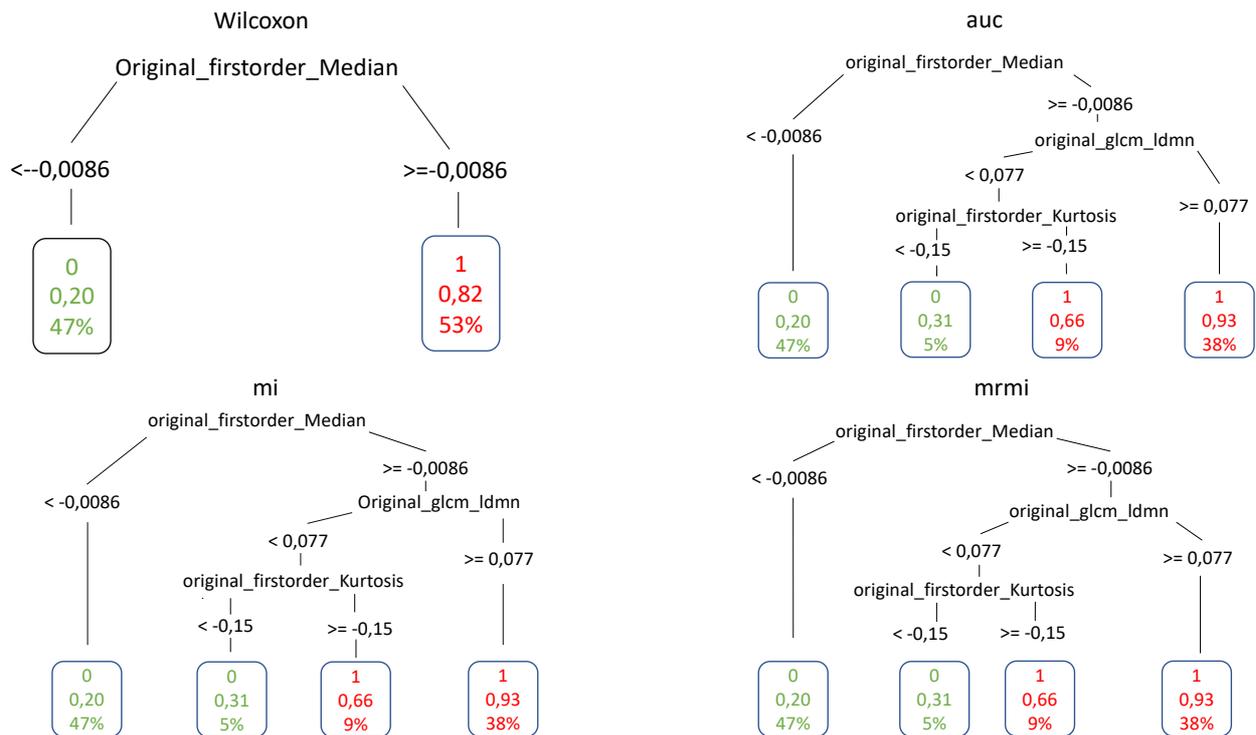


Abbildung 7: Darstellung der am Trainingsdatensatz erzeugten Klassifikationsbäume.

Diese Grafik zeigt die je Featureselektionsmethode ausgewählten Bildmerkmale in unterschiedlicher Gewichtung zur Unterscheidung PET-positiver und PET-negativer Lymphknoten.. Der Median wird methodenübergreifend als erstes Auswahlkriterium gewählt. Im Falle der Selektionsmethoden AUC (area under the curve), MI (mutual information) und MRMI (maximum relevance minimum redundancy) nutzten die Klassifikationsbäume noch zwei weitere Bildmerkmale zur genaueren Unterscheidung: glcm_Idmn und firstorder_Kurtosis. (in Zusammenarbeit mit beauftragtem Statistiker erstellt).

Zur Validierung der Prognosen der sechs Algorithmen wendeten wir sie auf den Testdatensatz an. Zunächst wurden die Features standardisiert. Die Ergebnisse wurden in Tabelle 4 zusammengefasst.

Tabelle 4: Ergebnisse nach Anwendung der Algorithmen auf den Testdatensatz.

Die Algorithmen wurden zunächst am Trainingsdatensatz trainiert und anschließend am Testdatensatz getestet. In dieser Tabelle ist dargestellt, mit welcher Korrektklassifikationsrate (accuracy) eine bestimmte Kombination aus Featureselektionsmethode und Vorhersagemodell (hier: Klassifikator) vorhersagen kann, ob ein Lymphknoten des Testdatensatzes PET-positiv oder PET-negativ ist. Zusätzlich sind jeweils die errechnete Sensitivität, Spezifität, sowie der positive und der negative Vorhersagewert angegeben. Verwendete Featureselektionsmethoden (FSM): Wilcoxon (wlcx), area under the curve (auc), mutual information (mi), maximum relevance minimum redundancy (mrmr).

Verwendete Klassifikatoren: lineare Diskriminanzanalyse (lda), logistische Regression (log), partial least squares (pls), support vector machines (svm), ein neuronales Netzwerk (multilayer perceptron, mlp) sowie Klassifikationsbäume (recursive partitioning, rpart).

Klassifikator	FSM	Korrektklassifikationsrate	Sensitivität	Spezifität	Positiver Vorhersagewert	Negativer Vorhersagewert
lda	wlcx	0,84	0,88	0,84	0,51	0,97
lda	auc	0,83	0,88	0,82	0,49	0,97
lda	mi	0,83	0,9	0,82	0,49	0,98
lda	mrmr	0,83	0,87	0,82	0,48	0,97
log	wlcx	0,82	0,87	0,81	0,47	0,97
log	auc	0,82	0,91	0,8	0,47	0,98
log	mi	0,82	0,91	0,8	0,47	0,98
log	mrmr	0,83	0,88	0,82	0,49	0,97
pls	wlcx	0,85	0,87	0,85	0,53	0,97
pls	auc	0,82	0,87	0,81	0,47	0,97
pls	mi	0,82	0,87	0,81	0,47	0,97
pls	mrmr	0,83	0,87	0,82	0,48	0,97
svm	wlcx	0,85	0,87	0,85	0,52	0,97
svm	auc	0,82	0,87	0,81	0,47	0,97
svm	mi	0,82	0,87	0,81	0,47	0,97
svm	mrmr	0,84	0,87	0,83	0,5	0,97
mlp	wlcx	0,78	0,87	0,76	0,41	0,97
mlp	auc	0,82	0,85	0,81	0,47	0,97
mlp	mi	0,78	0,9	0,75	0,41	0,97
mlp	mrmr	0,8	0,9	0,79	0,45	0,98
rpart	wlcx	0,8	0,9	0,78	0,44	0,98
rpart	auc	0,77	0,91	0,74	0,41	0,98
rpart	mi	0,77	0,91	0,74	0,41	0,98
rpart	mrmr	0,77	0,91	0,74	0,41	0,98

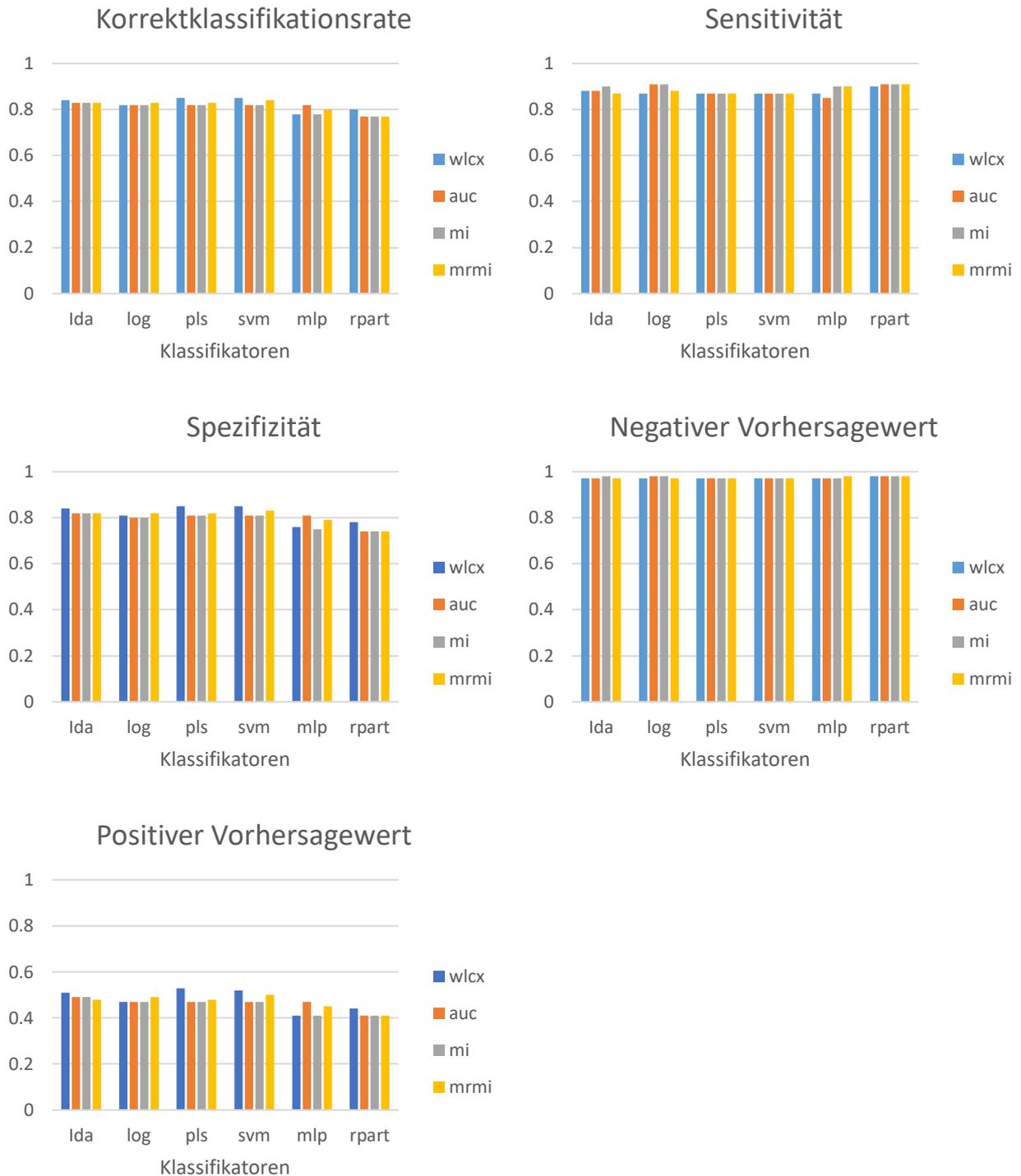


Abbildung 8: Grafische Darstellung der Ergebnisse nach Anwendung der Algorithmen auf den Testdatensatz.

Die Algorithmen wurden zunächst am Trainingsdatensatz trainiert und anschließend am Testdatensatz getestet. In dieser Zusammenfassung ist dargestellt, mit welcher Korrektklassifikationsrate (accuracy) eine bestimmte Kombination aus Feature-Selektionsmethode und Vorhersagemodell (hier: Klassifikator) vorhersagen kann, ob ein Lymphknoten des Testdatensatzes PET-positiv oder PET-negativ ist. Zusätzlich

sind jeweils die errechnete Sensitivität, Spezifität, sowie der positive und der negative Vorhersagewert graphisch dargestellt.

Verwendete Featureselektionsmethoden (FSM): Wilcoxon (wlcx), area under the curve (auc), mutual information (mi), maximum relevance minimum redundancy (mrmi).

Verwendete Klassifikatoren: lineare Diskriminanzanalyse (lda), logistische Regression (log), partial least squares (pls), support vector machines (svm), ein neuronales Netzwerk (multilayer perceptron, mlp) sowie Klassifikationsbäume (recursive partitioning, rpart) (in Zusammenarbeit mit beauftragtem Statistiker erstellt).

Die beste Vorhersageleistung wurde hier von partial least squares (pls) und support vector machines (svm), jeweils mit einer vorherigen Featureselektion nach Wilcoxon erbracht. Beide Klassifikatoren nutzen lineare Trennebenen zur Klassifikation der Lymphknoten. Beide Klassifikator-Selektionsmethode-Kombinationen erreichten eine Korrektklassifikationsrate von 85%, eine Sensitivität von 87%, eine Spezifität von 85%, sowie einen negativen Vorhersagewert von 97%. Sie unterschieden sich nur leicht in ihrem positiven Vorhersagewert; partial least squares + Wilcoxon mit 53% und support vector machines + Wilcoxon mit 52%. Andere Klassifikator-Featureselektion-Kombinationen kamen teilweise auf höhere Sensitivitätswerte, zum Beispiel erreichte das neuronale Netzwerk (mlp) in jeder Kombination eine Sensitivität von $\geq 90\%$. Die Spezifitätswerte des neuronalen Netzwerks lagen aber deutlich niedriger, in einem Fall (Wilcoxon) bei 78% und in drei Fällen bei 74%.

3.1 Vergleich der Klassifikatoren mit der Einschätzung von Radiologen

Tabelle 5: Statistische Auswertung der Einschätzung der zwei Radiologen.

Zunächst vergleichen wir die Gesamtpformance der Einschätzung der zwei Radiologen. Diese teilten die segmentierten Lymphknoten in 4 Kategorien ein (0=sicher nicht maligne, 1= wahrscheinlich nicht maligne, 2= wahrscheinlich maligne, 3= sicher maligne). Um diese Ergebnisse vergleichbar zu machen, fassten wir die Klassen 0 und 1 sowie 2 und 3 zunächst in zwei Klassen zusammen.

Radiologe	Korrektklassifikationsrate	Sensitivität	Spezifizität	Positiver Vorhersagewert	Negativer Vorhersagewert
1	0,95	0,76	0,98	0,88	0,96
2	0,95	0,76	0,97	0,84	0,96

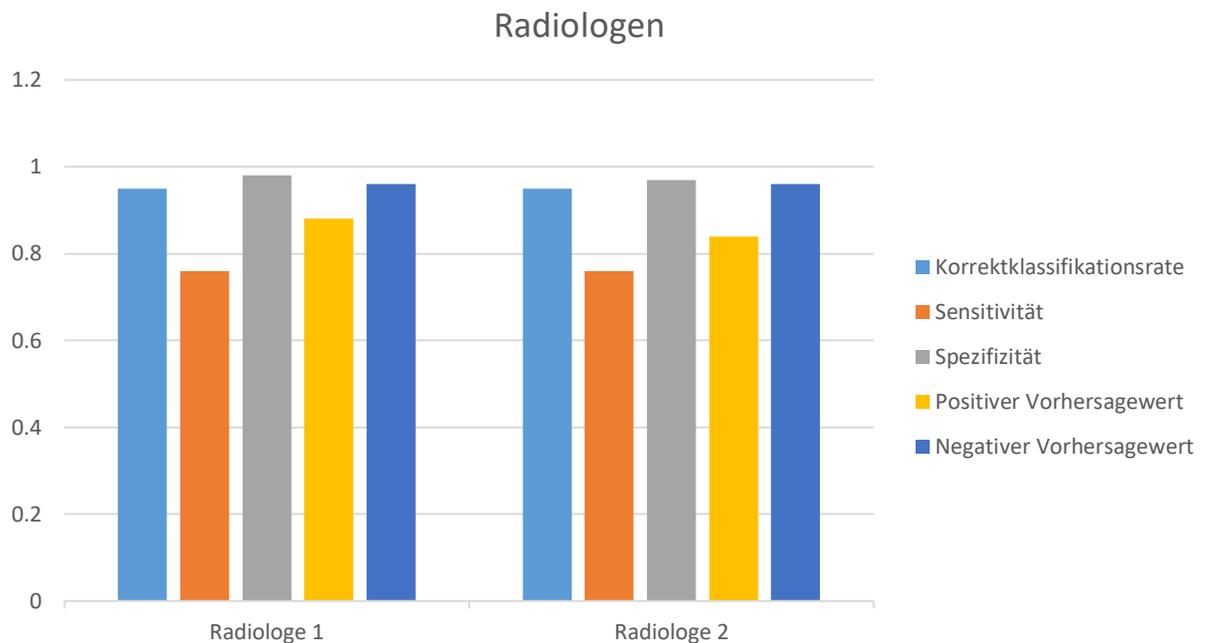


Abbildung 9: Klassifikationsleistung der der zwei Radiologen.

Die Radiologen erreichten in der Beurteilung der Lymphknoten beide eine Korrektklassifikationsrate von 95% und eine Sensitivität von 76%, eine Spezifizität von 98% und 97%, positive Vorhersagewerte von 88% und 84%, respektive, sowie beide einen negativen Vorhersagewert von 96%.

Tabelle 6: Kombination der Ergebnisse der Radiologen und der Klassifikationsalgorithmen.

Zwei Radiologen teilten die segmentierten Lymphknoten in 4 Kategorien ein (0=sicher nicht maligne, 1= wahrscheinlich nicht maligne, 2= wahrscheinlich maligne, 3= sicher maligne). Kombinierte Prognosen der Radiologen (in den Kategorien 0=sicher benigne, und 3=sicher maligne) sowie der Klassifikationsmodelle (in den Zwischenkategorien 1= wahrscheinlich nicht maligne und 2=wahrscheinlich maligne).

Verwendete Featureselektionsmethoden (FSM): Wilcoxon (wlcx), area under the curve (auc), mutual information (mi), maximum relevance minimum redundancy (mrmi).

Verwendete Klassifikatoren: lineare Diskriminanzanalyse (lda), logistische Regression (log), partial least squares (pls), support vector machines (svm), ein neuronales Netzwerk (multilayer perceptron, mlp) sowie Klassifikationsbäume (recursive partitioning, rpart).

Radiologe	Klassifikator	FSM	Korrektklassifikationsrate	Sensitivität	Spezifizität	Positiver Vorhersagewert	Negativer Vorhersagewert
1	Ida	wlcx	0,88	0,87	0,88	0,58	0,97
1	Ida	auc	0,87	0,87	0,87	0,56	0,97
1	Ida	mi	0,87	0,88	0,87	0,56	0,97
1	Ida	mrmi	0,87	0,85	0,87	0,56	0,97
1	log	wlcx	0,86	0,85	0,86	0,55	0,97
1	log	auc	0,86	0,90	0,85	0,54	0,98
1	log	mi	0,86	0,90	0,85	0,54	0,98
1	log	mrmi	0,87	0,87	0,87	0,56	0,97
1	pls	wlcx	0,88	0,85	0,89	0,59	0,97
1	pls	auc	0,86	0,85	0,86	0,55	0,97
1	pls	mi	0,86	0,85	0,86	0,54	0,97
1	pls	mrmi	0,87	0,85	0,88	0,57	0,97
1	svm	wlcx	0,88	0,85	0,89	0,60	0,97
1	svm	auc	0,86	0,85	0,87	0,55	0,97
1	svm	mi	0,86	0,85	0,86	0,54	0,97
1	svm	mrmi	0,87	0,87	0,87	0,57	0,97
1	mlp	wlcx	0,84	0,85	0,84	0,51	0,97
1	mlp	auc	0,86	0,87	0,86	0,55	0,97
1	mlp	mi	0,84	0,88	0,83	0,51	0,97
1	mlp	mrmi	0,87	0,88	0,86	0,56	0,97
1	rpart	wlcx	0,85	0,88	0,84	0,52	0,97
1	rpart	auc	0,84	0,90	0,83	0,50	0,98
1	rpart	mi	0,84	0,90	0,83	0,50	0,98
1	rpart	mrmi	0,84	0,90	0,83	0,50	0,98
2	Ida	wlcx	0,89	0,78	0,91	0,63	0,96

2	Ida	auc	0,89	0,78	0,91	0,62	0,95
2	Ida	mi	0,89	0,78	0,91	0,62	0,95
2	Ida	mrmi	0,89	0,76	0,91	0,63	0,95
2	log	wlcx	0,88	0,78	0,90	0,61	0,95
2	log	auc	0,89	0,78	0,91	0,62	0,96
2	log	mi	0,89	0,78	0,91	0,63	0,96
2	log	mrmi	0,89	0,76	0,92	0,65	0,95
2	pls	wlcx	0,89	0,76	0,91	0,62	0,95
2	pls	auc	0,88	0,76	0,90	0,59	0,95
2	pls	mi	0,88	0,76	0,90	0,60	0,95
2	pls	mrmi	0,89	0,76	0,91	0,63	0,95
2	svm	wlcx	0,89	0,76	0,91	0,63	0,95
2	svm	auc	0,88	0,76	0,91	0,61	0,95
2	svm	mi	0,88	0,76	0,91	0,61	0,95
2	svm	mrmi	0,89	0,76	0,92	0,65	0,95
2	mlp	wlcx	0,88	0,76	0,90	0,59	0,95
2	mlp	auc	0,89	0,76	0,91	0,62	0,95
2	mlp	mi	0,88	0,78	0,90	0,60	0,95
2	mlp	mrmi	0,89	0,78	0,91	0,64	0,96
2	rpart	wlcx	0,87	0,78	0,88	0,56	0,95
2	rpart	auc	0,86	0,78	0,88	0,56	0,95
2	rpart	mi	0,86	0,78	0,88	0,56	0,95
2	rpart	mrmi	0,86	0,78	0,88	0,56	0,95

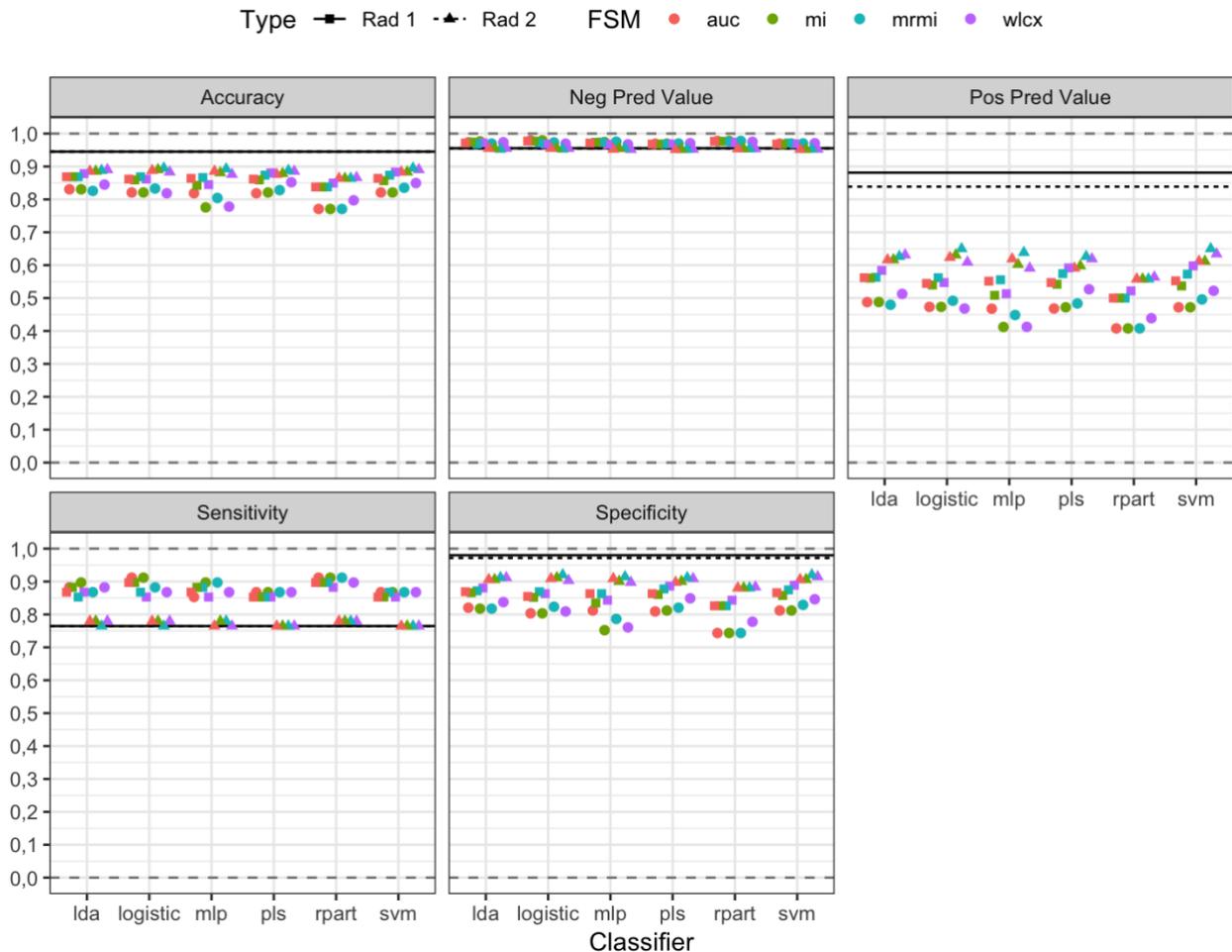


Abbildung 10: Kombination der Ergebnisse der Radiologen und der Klassifikationsalgorithmen.

Zwei Radiologen teilten die segmentierten Lymphknoten in 4 Kategorien ein (0=sicher nicht maligne, 1= wahrscheinlich nicht maligne, 2= wahrscheinlich maligne, 3= sicher maligne). Graphische Darstellung der kombinierten Prognosen der Radiologen (in den Kategorien 0=sicher benigne, und 3=sicher maligne) sowie der mathematischen Modelle (in den Zwischenkategorien 1= wahrscheinlich nicht maligne und 2=wahrscheinlich maligne).

Verwendete Featureselektionsmethoden (FSM): Wilcoxon (wlcx), area under the curve (auc), mutual information (mi), maximum relevance minimum redundancy (mrrmi).

Verwendete Klassifikatoren: lineare Diskriminanzanalyse (lda), logistische Regression (log), partial least squares (pls), support vector machines (svm), ein neuronales Netzwerk (multilayer perceptron, mlp) sowie Klassifikationsbäume (recursive partitioning, rpart).

(Rad= Radiologe, logistic= logistische Regression, Accuracy= deutsch: Korrektklassifikationsrate, Neg Pred Value= deutsch: negativer Vorhersagewert, Pos Pred Value= deutsch: positiver Vorhersagewert, Sensitivity= deutsch: Sensitivität, Specificity= deutsch: Spezifizität) (in Zusammenarbeit mit beauftragtem Statistiker erstellt).

Tabelle 7: Ergebnisse nach Bagging.

Die Balancierung des Trainingsdatensatzes kreiert eine gewisse Verzerrung. Um dies zu umgehen, kann die Balancierung mehrfach wiederholt und die Modelle auf dem jeweiligen Datensatz trainiert werden. Die Ergebnisse können dann aggregiert werden.

Verwendete Featureselektionsmethoden (FSM): Wilcoxon (wlcx), area under the curve (auc), mutual information (mi), maximum relevance minimum redundancy (mrmi).

Verwendete Klassifikatoren: lineare Diskriminanzanalyse (lda), logistische Regression (log), partial least squares (pls), support vector machines (svm), ein neuronales Netzwerk (multilayer perceptron, mlp) sowie Klassifikationsbäume (recursive partitioning, rpart).

Klassifikator	FSM	Korrektklassifikationsrate	Sensitivität	Spezifität	Positiver Vorhersagewert	Negativer Vorhersagewert
Ida	wlcx	0,82	0,88	0,81	0,47	0,97
Ida	auc	0,82	0,90	0,80	0,47	0,98
Ida	mi	0,81	0,88	0,79	0,45	0,97
Ida	mrmi	0,83	0,87	0,82	0,48	0,97
log	wlcx	0,81	0,88	0,79	0,45	0,97
log	auc	0,81	0,90	0,79	0,46	0,98
log	mi	0,82	0,91	0,80	0,47	0,98
log	mrmi	0,81	0,87	0,80	0,46	0,97
pls	wlcx	0,82	0,88	0,81	0,47	0,97
pls	auc	0,81	0,88	0,79	0,45	0,97
pls	mi	0,80	0,88	0,78	0,44	0,97
pls	mrmi	0,82	0,87	0,81	0,48	0,97
svm	wlcx	0,82	0,87	0,81	0,47	0,97
svm	auc	0,81	0,88	0,79	0,45	0,97
svm	mi	0,81	0,88	0,79	0,45	0,97
svm	mrmi	0,81	0,87	0,80	0,46	0,97
mlp	wlcx	0,82	0,87	0,81	0,47	0,97
mlp	auc	0,79	0,90	0,77	0,43	0,97
mlp	mi	0,77	0,87	0,75	0,41	0,97
mlp	mrmi	0,83	0,85	0,82	0,48	0,97
rpart	wlcx	0,76	0,87	0,74	0,39	0,97
rpart	auc	0,82	0,87	0,81	0,48	0,97
rpart	mi	0,82	0,87	0,81	0,48	0,97
rpart	mrmi	0,83	0,88	0,82	0,48	0,97

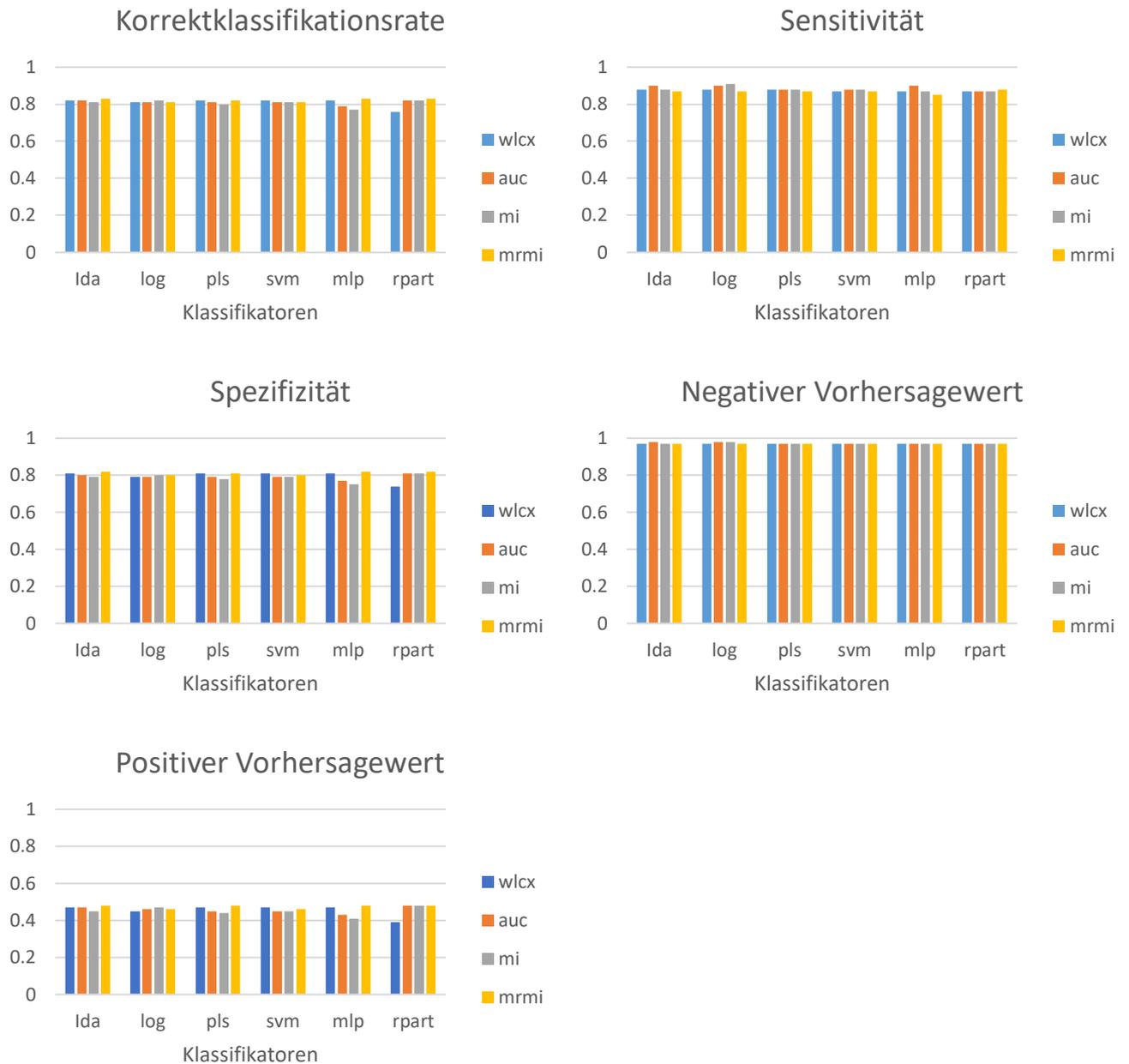


Abbildung 11: Grafische Darstellung der Ergebnisse nach bagging.

Die Algorithmen wurden zunächst an verschiedenen balancierten Trainingsdatensätzen trainiert und anschließend am Testdatensatz getestet. In dieser Zusammenfassung der aggregierten Ergebnisse ist dargestellt, mit welcher Korrektklassifikationsrate (accuracy) eine bestimmte Kombination aus Feature-Selektionsmethode und Vorhersagemodell (hier: Klassifikator) vorhersagen kann, ob ein Lymphknoten des Testdatensatzes PET-positiv oder PET-negativ ist. Zusätzlich sind jeweils die errechnete Sensitivität, Spezifizität, sowie der positive und der negative Vorhersagewert graphisch dargestellt.

Verwendete Featureselektionsmethoden (FSM): Wilcoxon (wlcx), area under the curve (auc), mutual information (mi), maximum relevance minimum redundancy (mirmi).

Verwendete Klassifikatoren: lineare Diskriminanzanalyse (lda), logistische Regression (log), partial least squares (pls), support vector machines (svm), ein neuronales Netzwerk (multilayer perceptron, mlp) sowie Klassifikationsbäume (recursive partitioning, rpart) (in Zusammenarbeit mit beauftragtem Statistiker erstellt).

Nach Bagging erreichen die mathematischen Modelle eine Korrektklassifikationsrate von 76% - 83%, eine Sensitivität von 85%-91%, eine Spezifität von 74% - 82%, einem positiven Vorhersagewert von 39% – 48% sowie einem negativen Vorhersagewert von 97%-98%

Die beste Vorhersageleistung wurde hier von den Klassifikationsbäumen (rpart) mit vorheriger Featureselektion nach der mrmi-Methode erreicht mit einer Korrektklassifikationsrate von 83%, eine Sensitivität von 88%, eine Spezifität von 82%, einem positiven Vorhersagewert von 48% sowie einem negativen Vorhersagewert von 97%.

Andere Klassifikator-Featureselektion-Kombinationen kamen teils auf höhere Einzelwerte, zum Beispiel erreichte das logistische Regressionsmodell mit vorheriger Featureselektion mittels „mutual information“ (mi) eine Sensitivität von 91%.

4. Diskussion

4.1 Diskussion der Methoden

4.1.1 Gerät

Alle PET-CT Bilder unseres Datensatzes wurden am selben Gerät angefertigt. Um eine ähnliche Bildqualität zu gewährleisten variiert am Gerät aufgrund automatischer Anpassungen an das Körpervolumen des Patienten das Röhrenspannungs-Zeit-Produkt (mAs). Außerdem variiert die Kontrastierung des Organparenchyms durch Kontrastmittel in jedem Patienten etwas. Grund hierfür ist das unterschiedliche Herzzeitvolumen jedes Patienten sowie die gewichtsadaptierte injizierte Kontrastmittelmenge. Dadurch entstehen für jeden Patienten etwas unterschiedliche Bildeindrücke. Es ist nicht auszuschließen, dass diese Unterschiede auch die Auswertung der Radiomic Features verfälscht. Für unsere Studie waren diese Parameter aber nicht weiter standardisierbar. Grundsätzlich wäre es für weitere Studien interessant die Robustheit der Ergebnisse bei Verwendung verschiedener Geräte bzw. bei unterschiedlicher Strahlendosis und Kontrastmittelmenge zu überprüfen.

4.1.2. PET-CT zu Beurteilung von Malignität

PET-Positivität wurde in dieser Studie als Hauptkriterium zur Unterscheidung zwischen Lymphknotenmetastasen und Lymphknoten ohne Tumormanifestation genutzt. Histopathologische Daten lagen nicht vor. Das PET-CT ist in seiner Diagnosefähigkeit von Lymphknotenmetastasen allerdings etwas limitiert. Wie bereits in der Einleitung erwähnt, erreicht das PSMA-PET-CT eine Sensitivität von etwa 85% in der Detektion von Lymphknotenmetastasen.¹⁸ Es ist dadurch denkbar, dass bestimmte Lymphknotenmetastasen vom PSMA-Tracer nicht oder in zu geringer Konzentration markiert wurden und dadurch nicht als solche diagnostiziert wurden. Dadurch ist es möglich, dass den Klassifikationsmodellen der gleiche Bias antrainiert wurde. Andere Studien zur Analyse von Lymphknotenmetastasen mittels Radiomics verwenden histopathologische Daten als Grundlage zur Beurteilung von

Malignität^{61,62}. Es muss vermutet werden, dass Vorhersagemodelle auf Basis dieser Daten realitätsnähere Aussagen treffen können. Gleichzeitig muss hervorgehoben werden, dass in diesen Studien ein deutliche kleinerer Datensatz verwendet wurde, z.B. 204 Lymphknoten von Baessler et al.⁶¹ oder 412 von Seidler et. al⁶². Die reine Größe unseres Datensatzes ist hier, mit 2125 segmentierten Lymphknoten, überlegen. In weiteren Studien, bzw. zur Validierung unserer Ergebnisse sollten histopathologische Untersuchungen als Grundlage zur Beurteilung von Malignität herangezogen werden. Die Aussagefähigkeit der Vorhersagemodelle würde so verbessert.

4.1.3 Segmentiermethode

In einem aufwendigen Prozess wurden mittels des Computerprogramms MITK pro Patient etwa fünf Lymphknoten segmentiert. Generell muss zwischen der manuellen und der semiautomatischen Segmentiermethode unterschieden werden. Die Abgrenzung des VOI (volume of interest) kann zwischen verschiedenen Beobachtern sowie zwischen verschiedenen Segmentiermethoden deutlich variieren (interobserver variability, IOV).^{52,66} Beispielsweise konnten Pavic et al. an Patienten mit malignem Pleuramesotheliom bei manueller Segmentierung eine deutliche IOV mit Dice coefficient von 0.29 zeigen.⁶⁷ Der Dice coefficient beschreibt hier die räumliche Überschneidung zweier Segmentationen, mit einem Wert von 1 für identische Konturen.⁶⁷ Der Dice coefficient korrelierte darüber hinaus mit der Stabilität der Radiomics-Merkmale; eine hohes Maß an IOV führte in dieser Studie zu einer deutlich verringerten Stabilität der Radiomics-Merkmale⁶⁷. Bezüglich der Reproduzierbarkeit von Radiomics-Merkmalen scheint die semiautomatische gegenüber der manuellen Methode verlässlichere Ergebnisse zu produzieren⁶⁸. Die Segmentierung wurde zunächst semiautomatisch durchgeführt. Die Grenzen der zu segmentierenden Strukturen wurden unserer Meinung nach in MITK mittels Interpolation aber nur unzureichend beschrieben, was zur Entscheidung führte, vormals semiautomatisch generierte Segmentationen Schicht-für Schicht manuell anzupassen.

Die große Mehrheit der Lymphknoten wurden anschließend vollständig manuell segmentiert (siehe Abb.5).

Zusammenfassend kann die IOV im Segmentierprozess eine Quelle erheblichen Bias sein und zu Stabilitätseinschränkungen der Radiomics-Merkmale führen. Um diese Art des Bias zu reduzieren haben Lambin et al. in ihrem Radiomics Quality Score (RQS) mehrere Methoden vorgeschlagen⁵²: Wiederholung der Segmentation durch mehrere Personen, Algorithmen oder Software, die Störung der Segmentation durch zufälliges Bildrauschen oder die Segmentierung in verschiedenen Atemzyklen. Die Robustheit der extrahierten Merkmale kann anhand des Vergleichs verschiedener Arten der Segmentation überprüft werden. Die vorgeschlagenen Methoden kamen in unserer Studie nicht zur Anwendung, was als Limitation gesehen werden kann. Die Robustheit der von uns generierten Vorhersagemodelle und ihre klinische Anwendbarkeit könnten dadurch eingeschränkt sein. Gleichzeitig kann auch hier wieder die reine Größe unseres Datensatzes, insbesondere im Vergleich zu kleineren Studien (s. 4.1.2) hervorgehoben werden. Die manuelle Segmentierung wurde nachfolgend durch zwei Radiologen überprüft.

Zur besseren Vergleichbarkeit von Studien im Bereich Radiomics wäre es in Zukunft sinnvoll gewisse Qualitätsstandards (wie von Lambin et al. vorgeschlagen⁵²) im Segmentierprozess einzuhalten. Dies würde auch Qualität der Ergebnisse und die Anwendbarkeit der Vorhersagemodelle erhöhen.

4.1.4 Featureselektion und Vorhersagemodelle

Mit pyradiomics verfügen wir über einen umfangreichen Pool an analysierbaren Bildmerkmalen. Um nur die relevanten Bildmerkmale in die Analyse einfließen zu lassen verwendeten wir fünf verschieden Bildmerkmalselektionsmethoden (Feature-selection methods). Eine Featureselektion vor dem eigentlichen Training der Algorithmen vermeidet ein späteres „overfitting“ der Algorithmen auf den Testdatensatz⁵². Anstatt sehr viele Bildmerkmale in die

Berechnung der Algorithmen einfließen zu lassen, die dann den Trainingsdatensatz zu genau beschreiben (overfitting) beschränkten wir uns auf die relevantesten 20. Dies verbessert die Anwendbarkeit der Algorithmen auf neue Datensätze. Die vier von uns verwendeten Methoden zur Featureselektion wählen nach unterschiedlichen Grundsätzen die für den Datensatz relevantesten Features aus. Wir beobachteten, nach mehrfacher Wiederholung jedes Selektionsprozesses, dass die Methode nach Wilcoxon die stabilsten Selektionsergebnisse liefert (siehe Abbildung 6). Im Großen und Ganzen beobachten wir aber auch, dass die gewählte Featureselektionsmethode in unserem Fall nur einen geringen Einfluss auf das Endergebnis hat; alle Selektionsmethoden wählen sehr ähnliche Features zur Unterscheidung PET-positiver und PET-negativer Lymphknoten. Außerdem sehen wir eine ähnliche Gewichtung der ausgewählten Bildmerkmale: Methodenübergreifend wird der Median der Grauwerte als wichtigstes Unterscheidungsmerkmal gewählt. An zweiter Stelle steht in allen Fällen außer im Falle der Methode mirmi (maximum relevance minimum redundancy) der Mittelwert der Grauwerte. Mit der Methode mirmi wird versucht, zusätzlich zur Auswahl von Features mit hoher mutual information redundante Bildmerkmale auszuschließen. Mittelwert und Median werden durch mirmi offenbar als Statistiken erster Ordnung (firstorder) mit ähnlicher Aussage gewertet, wodurch derjenige mit geringerer Aussagekraft (in diesem Fall der Mittelwert) zugunsten einer anderen Statistik erster Ordnung, der Kurtosis ausgeschlossen wird. Interessant ist, dass auch in einer anderen Studie das histogramm-basierte radiomics-Merkmal Kurtosis als wichtiger Parameter genutzt wurde⁶⁹; hier allerdings zur Unterscheidung zwischen Knochenmetastasen des Prostatakarzinoms und bereits therapierten, sklerosierten Läsionen.

Mit den jeweils ausgewählten 20 Features konnten wir nun mathematische Modelle zur Beurteilung neuer Datensätze trainieren. Es stehen heute verschiedenste Modellierungsmethoden zur Verfügung. Wie Lambin et al.⁵² in ihrer Arbeit bemerken kann es von Vorteil sein, mehrere Methoden anzuwenden, da jede Methode ihrer eigenen Limitationen

aufweist. Diesen Ansatz verfolgten auch wir, durch den Einsatz von mehreren statistischen Modellen. Zur internen Validierung der mathematischen Modelle wendeten wir sie auf dem zweiten, vom Modelltraining bisher unabhängigen Teil unseres Datensatzes an.

4.2. Diskussion der Ergebnisse

In dieser Studie haben wir Radiomics-Methoden an einem großen Datensatz von insgesamt 447 Patienten mit 2125 analysierten Lymphknoten auf die Fähigkeit hin getestet, PET-positive von PET-negativen Lymphknoten bei Patienten mit Prostatakarzinom zu unterscheiden.

Um die Robustheit der Ergebnisse auf Veränderungen der statistischen Methoden zu überprüfen haben wir insgesamt 24 verschiedene Kombinationen aus Feature-Selektionsmethode und Klassifikationsalgorithmus an einem Trainingsdatensatz trainiert und einem Testdatensatz getestet.

Insgesamt beobachten wir, dass alle Kombinationen aus Feature-Selektionsmethode und Klassifikationsalgorithmus ähnliche gute Ergebnisse liefern. Die beste Vorhersageleistung wurde von den Klassifikationsbäumen (rpart) mit vorheriger Featureselektion nach der mrmi-Methode erreicht mit einer Korrektklassifikationsrate von 83% (Sensitivität von 88%, Spezifität 82%, positiver Vorhersagewert 48%, negativer Vorhersagewert 97%).

Der positive Vorhersagewert fällt mit 39% – 48% modellübergreifend auffallend deutlich geringer aus als die restlichen statistischen Parameter. Dieser Parameter gibt an, wie viele der als PET-positiv vorhergesagten Lymphknoten auch tatsächlich positiv waren. Insgesamt detektieren unsere Modelle Lymphknotenmetastasen des Prostatakarzinoms mit hoher Sensitivität, produzieren aber auch viele falsch positive.

4.2.1 Vergleich der Modelle mit Radiologen

Wir verglichen die Performance der mathematischen Modelle mit der von zwei Radiologen. Sie erreichten in der Beurteilung der Lymphknoten beide eine Korrektorklassifikationsrate von 95% sowie eine Spezifität von 98% und 97%. Sie produzierten hier bessere Ergebnisse als die von uns kreierte mathematischen Modelle. Ihre Sensitivität war allerdings mit 76% geringer als die der Modelle. Die Radiologen konnten also mehr Lymphknoten korrekt klassifizieren, erkannten aber weniger der PET-positiven Lymphknoten im Vergleich zu den Klassifikationsalgorithmen. Eine Kombination von Radiologen und Klassifikatoren, mit Ersatz der Beurteilung der Lymphknoten, die von den Radiologen in die unsicheren Kategorien 1 und 2 eingeteilt wurden, durch die Beurteilung Klassifikatoren, brachte keine besseren Ergebnisse als die Klassifikation der Radiologen allein.

Wir können daher beide von uns aufgestellten Hypothesen nicht verifizieren. Die Klassifikationsalgorithmen sagen den PET-CT Status der Lymphknoten nicht besser voraus als die Radiologen. Außerdem ist die Vorhersageleistung der Kombination von Klassifikationsalgorithmen und Radiologen nicht besser als die der Letzteren allein.

4.2.1.1 Performance der CT-Lymphknotendiagnostik in anderen Studien

Im Vergleich zu unseren Ergebnissen fanden Abuzallouf et al., in einem Review eine viel geringere Sensitivität des CT für pelvine Lymphknotenmetastasen des Prostatakarzinoms mit einem Median von 7% (Spezifität 100%, Positiver Vorhersagewert 85%).¹⁷ Allerdings wurden die Statistiken in diesem Review „per-patient“ gerechnet, während in unserer Studie die Sensitivität für einzelne Lymphknoten berechnet wurde, also „per-lesion“. Dadurch sind die Ergebnisse nicht direkt vergleichbar. Unsere Ergebnisse unterscheiden sich auch drastisch von den Ergebnissen der Studie von Maurer et al., in der allerdings CT und MRT Bilder beurteilt wurden. Ein Radiologe beurteilte hier Lymphknoten anhand konventioneller CT oder MRT-Bilder mit einer „per-lesion“-Sensitivität von nur 28,2% (Spezifität 97,1%,

Korrektklassifikationsrate 86.1%) und einer „per-patient“-Sensitivität von 43.9% (Spezifität 85,4%, Korrektklassifikationsrate 72,3%).¹⁶ In beiden Studien lagen histopathologische Daten zum Vergleich vor, was in unserer Studie nicht der Fall war.

Nach Aussage der größeren Studien scheint die Leistung des CT in der Ausbreitungsdiagnostik des Prostatakarzinoms sehr eingeschränkt zu sein. Außerdem weisen die unterschiedlichen Ergebnisse auf eine hohe Variabilität zwischen einzelnen Beobachtern (Interobserver Variability) in der Beurteilung von Lymphknoten im CT bei Patienten mit Prostatakarzinom hin.

Ein großer Unterschied von den beschriebenen Untersuchungen zu unserer Studie ist die histopathologische Korrelation, die in unserer Studie nicht durchgeführt wurde. Daher ist ein Vergleich nur eingeschränkt möglich.

Für den großen Unterschied der Vorhersageleistung der Radiologen in der beschriebenen Literatur zu unserer Untersuchung findet sich für uns aber keine eindeutige Erklärung.

Insgesamt können wir also die erste aufgestellte Hypothese (Radiomics-basierte Klassifikationsalgorithmen können den ⁶⁸Ga-PSMA PET/CT-Status von Lymphknoten nur auf Basis von CT-Bildern besser vorhersagen als Radiologen) in Bezug auf die gute Leistung der Radiologen in unserer Studie zwar nicht verifizieren. In Bezug auf die deutlich eingeschränktere Leistung der Radiologen in der vorliegenden Literatur erscheint dies aber weniger eindeutig.

4.2.2 Vergleich mit anderen Radiomics – Studien

Peeken et. al untersuchten an 80 Patienten mit 149 segmentierten Lymphknoten ebenfalls die Leistung von Radiomics-Methoden zur Detektion von Lymphknotenmetastasen des Prostatakarzinoms. 110 der segmentierten Lymphknoten waren positiv, 39 negativ. Peeken et al. bestimmten unter anderem die „balanced accuracy“ (deutsch: balancierte Korrektklassifikationsrate), die bei unbalancierten Datensätzen eine realistischere Aussage über

die Vorhersageleistung trifft als die Korrektklassifikationsrate. Das beste Klassifizierungsmodell von Peeken et al. erreichte eine „balanced accuracy“ von 84%. Wir errechneten die „balanced accuracy“ zur besseren Vergleichbarkeit: Unser bestes Modell erreichte eine balancierte Korrektklassifikationsrate von 85%, welche sogar etwas höher als die reguläre Korrektklassifikationsrate mit 83% lag. Desweiteren erreichten die besten Klassifizierungsmodelle von Peeken et. al. vs. Unserer Studie folgende weitere Werte: eine Sensitivität von 98% vs. 88%, eine Spezifität von 69% vs. 82%, einen positiven Vorhersagewert von 92% vs. 48% und einen negativen Vorhersagewert von 90% vs. 97%, respektive. Die beste Leistung des Parameters „Area under the curve“ (deutsch: Fläche unter der Kurve) erreichte in der Studie von Peeken et. al. ein anderes Modell mit 95%. Diesen Parameter haben wir nicht errechnet und können ihn nicht direkt vergleichen.⁷⁰

Im Vergleich ist die Leistung unseres besten Modells gemessen an der Sensitivität etwas und gemessen am positiven Vorhersagewert (48%) deutlich geringer. Gleichzeitig sind die balancierte Korrektklassifikationsrate, die Spezifität und der negative Vorhersagewert unseres besten Modells höher. Vorteile ergeben sich bei unserer Studie durch den deutlich größeren Datensatz. Vorteile der Studie von Peeken et al. ergeben sich daraus, dass der Lymphknotenstatus histologisch gesichert wurde und dass an einem externen Datensatz getestet wurde.

Desweiteren gibt es einige Studien, in denen Radiomics-Methoden zur Diagnose von Lymphknotenmetastasen anderer Primärtumore herangezogen wurden. Sha et al kreierten ein mathematisches Modell, dass mit 88,2% iger Genauigkeit mediastinale Lymphknotenmetastasen des NSCLC (non small-cell lung carcinoma) von benignen Lymphknoten unterscheiden kann.

54

In weiteren vergleichbaren Studien wurden Methoden des Maschinlernens verwendet: Es konnte an 88 Patienten mit metastasiertem, nicht seminomatösem, testikulären Keimzelltumor

gezeigt werden, dass Maschinenlernen-Klassifikatoren eine Korrektklassifikationsrate von 81% in der Unterscheidung maligner und benigner Lymphknoten erreichen.⁶¹ Eine weitere Studie stellte fest, dass Maschinenlernen-Klassifikatoren dasselbe mit einer Korrektklassifikationsrate von 90% bei Patienten mit Kopf- und Halskarzinom erreichen.⁶²

Unsere Ergebnisse decken sich mit einer Korrektklassifikationsrate von 85% in etwa mit den Ergebnissen der vorgestellten Studien.

4.2.3 Deep learning

Der hier verwendete Datensatz wurde auch in einer weiteren Studie genutzt. Hier wurden Convolutional Neural Networks (CNN) im Rahmen des deep learning trainiert. Ein großer Unterschied des deep learning- Ansatzes ist, dass keine festgelegten Radiomic-Features, wie in unserer Studie, genutzt werden, sondern die Algorithmen Merkmale des Bildmaterials eigenständig erkennen um es zu klassifizieren. Die beste Klassifikationsleistung der CNN wurde mir einer Korrektklassifikationsrate von 89% erreicht. Bei diesem Ansatz haben die Algorithmen aber offenbar zu einem großen Teil die Lymphknoten umgebenden anatomische Strukturen zur Klassifikation genutzt. An einem für die Lymphknotenlokalisierung balancierten sowie einem segmentierten Datensatz, der die umgebenden Strukturen ausschloss, wurden wesentlich schlechtere Klassifikationsleistungen erbracht, mit Korrektklassifikationsraten von 80% und 76%, respektive. Im Vergleich zu diesen letzten Ergebnissen erreichten unsere besten Algorithmen, auf Basis festgelegter Radiomic-Features, etwas bessere Werte. Erstaunlicherweise ist bei den deep learning - Klassifikatoren allerdings, obgleich derselbe Datensatz zu Grunde lag, kein Absinken des positiven Vorhersagewerts (PPV), im Vergleich zu unseren Ergebnissen, zu bemerken.⁷¹

4.2.4 Ausgewählte Radiomic Features

Am Beispiel der Klassifikationsbäume (siehe Abb. 7) können wir die Funktion der mathematischen Modelle analysieren. In den genauen Prozess der übrigen Modelle haben wir

weniger Einsicht. Als wichtigstes Unterscheidungsmerkmal zwischen PET-positiven und PET-negativen Lymphknoten wählten die Klassifikationsbäume den Median der Grauwerte. Für unseren Datensatz ist diese simple Statistik den komplexeren offenbar überlegen. Wenn die Selektionsmethode nach Wilcoxon vorgeschaltet ist wird sogar ausschließlich den Median als Unterscheidungskriterium genutzt. Ist der Median der Grauwerte unterhalb eines bestimmten Werts, klassifiziert der Algorithmus den untersuchten Lymphknoten als PET-negativ und oberhalb als PET-positiv. Möglicherweise lässt sich dies durch den histologischen Unterschied von Lymphknotenmetastasen und Lymphknoten ohne Tumormanifestation erklären. Lymphknotenmetastasen erreichen durch die vermehrte Zellreplikation vermutlich eine höhere Zelldichte als reguläre Lymphknoten. CT-morphologisch wären somit höhere Hounsfield-Einheiten (HU, engl.: hounsfield units) der analysierten Voxel zu erwarten. Dies würde sich in höheren Werten des Medians der Grauwerte niederschlagen. Zusätzlich neigen Lymphknoten ohne Tumormanifestation zu Fetthili, also fettigen Veränderungen des Lymphknotenhilus, welche mit ihren geringen HU den Median der Grauwerte deutlich senken könnten. Die anderen drei Selektionsmethoden (AUC, MI und MRMI) wählen noch zwei weitere Selektionsmethoden zur genaueren Analyse der laut Median-Cut-off-Wert als PET-positiv klassifizierten Lymphknoten: die Kurtosis sowie eine Gray Level Co-occurrence Matrix (glcm), namens inverse difference moment normalized⁴⁹. Die Kurtosis, eine Statistik erster Ordnung, drückt mit einem höheren Wert eine weitere Entfernung der Masse der Werte vom Mittelwert aus⁴⁹. Die zweite Statistik, eine Statistik zweiter Ordnung, ist ein Ausdruck der Homogenität der Grautöne in der ausgewählten Region of Interest (ROI)⁴⁹. Mit den Werten der Kurtosis sowie der glcm_idmn scheint die Heterogenität der Grautöne eine wichtige Rolle in der Unterscheidung der Lymphknoten zu spielen.

4.3 Ausblick

Auch wenn in dieser Untersuchung zwei Radiologen mit einer Sensitivität von 76% PET-positive Lymphknoten anhand von CT-Bildern vorhersagen, beschreiben Abuzallouf et al. in einem Review die Sensitivität des CT für Lymphknotenmetastasen z.B. mit einem Median von nur 7%.¹⁷ Da im beschriebenen Review anders als in dieser Untersuchung histopathologische Korrelationen vorgenommen wurden, sind die Ergebnisse zwar nur eingeschränkt vergleichbar, in unserer Studie erreichte das beste Klassifikationsmodell aber eine deutlich höhere Sensitivität von 88%. Obwohl wir unsere Hypothesen insgesamt nicht bestätigen können, deuten die Ergebnisse darauf hin, dass Radiomics-Methoden zu einer deutlichen Steigerung der Detektionsraten von Lymphknotenmetastasen im CT bei Patienten mit PCa führen könnten. Um dies zu beweisen, sind weitere Untersuchungen mit histopathologischer Korrelation notwendig.

Das PSMA-PET-CT erreicht mit einer Sensitivität und Spezifität von 85% und 98%, zwar bessere Werte als das CT, ist aber derzeit nur zum Re-Staging empfohlen^{1,18}. Auch wenn das PSMA-PET-CT zum Primärstaging empfohlen werden sollte, bleibt es ein Verfahren, das nur in speziellen Zentren zur Verfügung steht. In Situationen, in denen kein PET-CT zur Verfügung steht, könnten Klassifikationsmodelle in Zukunft zur Anwendung kommen bzw. könnten sie die konventionelle Diagnostik oder auch das PET-CT ergänzen.

Radiomics hat das Potential ein wichtiger Teil in der radiologischen Diagnostik zu werden. Unsere Ergebnisse unterstützen das.

4.4 Limitationen

Die Klassifikationsmodelle wurden vollständig an einem Datensatz der Charité trainiert. Sie wurden zwar an einem weiteren Datensatz getestet, für einen allgemeinen Einsatz wäre eine Validierung an einem oder mehreren externen Datensätzen notwendig. Weiterhin können wir

einen systematischen Bias durch die manuelle Segmentierung der Lymphknoten durch nur einen Beobachter nicht ausschließen. Insgesamt könnte sich vom Bildgebungsprotokoll bis hin zur Validierung in zukünftigen Studien noch mehr an eine wie von Lambin et al. vorgeschlagene Vorgehensweise gehalten werden⁵²; beispielsweise könnte die Robustheit der Radiomics-Features unter verschiedenen Arten der Segmentierung überprüft werden. Die Qualität der Vorhersagemodelle könnte so eventuell erhöht werden. Eine weitere wichtige Limitation ist die fehlende histopathologische Korrelation unserer Ergebnisse, jedoch würde dies auch einen Bias hin zu positiven Lymphknoten bedeuten, da negative Lymphknoten in der Regel nicht exstirpiert / biopsiert werden. Unsere Klassifikationsmodelle arbeiten daher auf Grundlage der Aussage des PSMA/PET-CT. Weiterhin sind prospektive Studien zur Validierung durchzuführen.

5. Zusammenfassung

Das Prostatakarzinom ist die häufigste diagnostizierte Krebserkrankung bei Männern in Deutschland. Da sie vorrangig ältere Menschen betrifft, müssen wir im Zuge der aktuellen demographischen Entwicklung mit zukünftig steigenden Erkrankungsraten rechnen.

Bei Patienten mit höherem Gleason-Score (≥ 8) soll laut Leitlinie eine CT- oder MRT-Untersuchung der Beckenorgane zur Detektion von Lymphknotenmetastasen durchgeführt werden. Ein positiver Bildgebungsbefund für Lymphknotenmetastasen kann Patienten eine pelvine Lymphadenektomie inklusive ihrer Nebenwirkungen ersparen. Die Detektionsrate von Lymphknotenmetastasen im CT sowie im MRT ist allerdings unzufriedenstellend. Mit dem ^{68}Ga -PSMA PET/CT ist zuletzt eine deutlich zuverlässigere Methode zur Ausbreitungsdiagnostik aufgekommen, die, bisher nur zum Re-Staging leitliniengerecht, auch bald im Primärstaging empfohlen werden könnte. Das ^{68}Ga -PSMA PET/CT ist allerdings eine nur in spezialisierten Zentren verfügbare Option.

Radiomics-Methoden zeigten in der bisherigen Literatur vielversprechende Ergebnisse in der Detektion von Lymphknotenmetastasen verschiedener Malignome. Wir haben untersucht, ob Radiomics-Methoden auch die Lymphknotendiagnostik im CT bei Patienten mit PCa verbessern können. Wir stellten die Hypothese auf, dass Radiomics-basierte Klassifikationsalgorithmen den ^{68}Ga -PSMA PET/CT-Status von Lymphknoten nur auf Basis von CT-Bildern besser vorhersagen als Radiologen. Zusätzlich stellten wir die Hypothese auf, dass die Kombination von Klassifikationsalgorithmen und Radiologen eine bessere Vorhersageleistung erzielen als letztere allein.

Zu Beginn unserer Studie gab es unseres Wissens nach noch keine vergleichbare Veröffentlichung. Peeken et al.⁷⁰ haben mittlerweile eine ähnliche Arbeit mit guten Ergebnissen publiziert.

Wir untersuchten einen großen Bilddatensatz von 447 Patienten mit Prostatakarzinom. Es wurden Patienten inkludiert, die ein kontrastmittelverstärktes ⁶⁸Ga-PSMA PET/CT entweder im Rahmen eines Primärstaging oder eines Re-Staging erhielten. Insgesamt wurden 2125 Lymphknoten identifiziert. Wir nutzten die Daten der PET-CT-Untersuchungen, um die Lymphknoten in maligne und nicht maligne zu unterteilen. Histopathologische Untersuchungen lagen nicht vor. 369 der segmentierten Lymphknoten waren PET-positiv und 1756 PET-negativ. Die Lymphknoten wurden manuell mittels des Computerprogramms MITK segmentiert. Nach einer Selektion der relevanten Radiomic-Features mittels vier verschiedener Methoden (Featureselektionsmethoden, FSM) trainierten wir sechs unterschiedliche Klassifikationsalgorithmen am Trainingsdatensatz. Die 24 so entstandenen Kombinationen aus Featureselektionsmethode und Klassifikationsalgorithmus wurden am Testdatensatz überprüft.

Die beste Vorhersageleistung wurde von einer Kombination aus Klassifikationsbäumen (recursive partitioning, rpart) und vorheriger Featureselektion nach der mrm-Methode erreicht mit einer Korrektklassifikationsrate von 83%, eine Sensitivität von 88%, eine Spezifität von 82%, einem positiven Vorhersagewert von 48% sowie einem negativen Vorhersagewert von 97%.

Zusätzlich verglichen wir die Performance unserer Klassifikationsalgorithmen mit denen zweier erfahrener Radiologen, welche beide mit einer Korrektklassifikationsrate von 95% (Sensitivität beide 76%, Spezifität von 98% und 97%) PET-positive und PET-negative Lymphknoten identifizierten. Die Vorhersageleistung der Kombination von Klassifikationsalgorithmen und Radiologen war außerdem nicht besser als die der Radiologen allein.

Insgesamt war die Performance der Klassifikationsalgorithmen der der Radiologen unterlegen. Die Sensitivität der Radiologen war allerdings mit 76% geringer als die der Klassifikationsalgorithmen.

In anderen Studien zeigen Radiologen allerdings in der konventionellen Diagnostik eine deutlich geringere Leistung¹⁶⁻¹⁸, auch wenn diese Ergebnisse aufgrund der bei uns fehlenden histopathologischen Korrelation eingeschränkt mit unseren vergleichbar sind.

Wir können zwar die aufgestellten Hypothesen nicht bestätigen, doch weisen die Ergebnisse, im Vergleich zur existierenden Literatur, darauf hin, dass Radiomics-Methoden die konventionellen Methoden zur Detektion von Lymphknotenmetastasen, im Zuge der Ausbreitungsdiagnostik bei Patienten mit PCa, verbessern könnten.

Die Stärken dieser Studie liegen in der mit 2125 Lymphknoten außergewöhnlichen Größe des Datensatzes sowie in der extensiven statistischen Analyse; insgesamt wurden 24 verschiedene Kombinationen aus FSM und Klassifikationsalgorithmen an einem aufwändig nach Bagging-Verfahren balancierten Datensatz trainiert. Um aber eine zukünftige klinische Anwendbarkeit zu ermöglichen sind zusätzliche prospektive Untersuchungen mit histopathologischer Korrelation notwendig.

6. Literaturverzeichnis

1. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft DK, AWMF). Interdisziplinäre Leitlinie der Qualität S3 zur Früherkennung, Diagnose und Therapie der verschiedenen Stadien des Prostatakarzinoms, Langversion 5.1. 2019. <http://www.leitlinienprogramm-onkologie.de/leitlinien/prostatakarzinom/> (Zugriff am 06.11.19).
2. Robert Koch-Institut G. Krebs in Deutschland für 2015/2016. . 12 ed. Berlin: RKI; 2019: 98-101.
3. Brachlow J, Kalin M, Randazzo M, Forster B, John H. Das Prostatakarzinom: aktuelle Diagnostik und Therapiestandart. (1661-8157 (Print)).
4. Hoogendam A, Buntinx F, de Vet HC. The diagnostic value of digital rectal examination in primary care screening for prostate cancer: a meta-analysis. *Fam Pract* 1999; **16**(6): 621-6.
5. EAU Guidelines. Edn. presented at the EAU Annual Congress Amsterdam ed; 2020.
6. Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Zappa M, Nelen V, Kwiatkowski M, Lujan M, Maattanen L, Lilja H, Denis LJ, Recker F, Paez A, Bangma CH, Carlsson S, Puliti D, Villers A, Rebillard X, Hakama M, Stenman UH, Kujala P, Taari K, Aus G, Huber A, van der Kwast TH, van Schaik RH, de Koning HJ, Moss SM, Auvinen A. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. (1474-547X (Electronic)).
7. Ilic D, Neuberger Mm Fau - Djulbegovic M, Djulbegovic M Fau - Dahm P, Dahm P. Screening for prostate cancer. (1469-493X (Electronic)).
8. 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.

9. Loeb S, Bruinsma SM, Nicholson J, Briganti A, Pickles T, Kakehi Y, Carlsson SV, Roobol MJ. Active surveillance for prostate cancer: a systematic review of clinicopathologic variables and biomarkers for risk stratification. (1873-7560 (Electronic)).
10. 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. (1873-7560 (Electronic)).
11. 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. (1873-7560 (Electronic)).
12. Filson CP, Natarajan S, Margolis DJ, Huang J, Lieu P, Dorey FJ, Reiter RE, Marks LS. Prostate cancer detection with magnetic resonance-ultrasound fusion biopsy: The role of systematic and targeted biopsies. (1097-0142 (Electronic)).
13. Wittekind C, Union for International Cancer C, Wiley VCH. TNM - Klassifikation maligner Tumoren : korrigierter Nachdruck 2020 mit allen Ergänzungen der UICC aus den Jahren 2017 bis 2019 / herausgegeben von Christian Wittekind. 8. Auflage ed. Weinheim: Weinheim : Wiley-VCH; 2020.
14. D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K, Broderick GA, Tomaszewski JE, Renshaw AA, Kaplan I, Beard CJ, Wein A. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *Jama* 1998; **280**(11): 969-74.
15. Thompson I, Thrasher JB, Aus G, Burnett AL, Canby-Hagino ED, Cookson MS, D'Amico AV, Dmochowski RR, Eton DT, Forman JD, Goldenberg SL, Hernandez J, Higano CS, Kraus SR, Moul JW, Tangen CM. Guideline for the management of clinically localized prostate cancer: 2007 update. *J Urol* 2007; **177**(6): 2106-31.

16. Maurer T, Gschwend JE, Rauscher I, Souvatzoglou M, Haller B, Weirich G, Wester HJ, Heck M, Kubler H, Beer AJ, Schwaiger M, Eiber M. Diagnostic Efficacy of (68)Gallium-PSMA Positron Emission Tomography Compared to Conventional Imaging for Lymph Node Staging of 130 Consecutive Patients with Intermediate to High Risk Prostate Cancer. (1527-3792 (Electronic)).
17. Abuzallouf S, Dayes I Fau - Lukka H, Lukka H. Baseline staging of newly diagnosed prostate cancer: a summary of the literature. (0022-5347 (Print)).
18. Hofman MS, Lawrentschuk N, Francis RJ, Tang C, Vela I, Thomas P, Rutherford N, Martin JM, Frydenberg M, Shakher R, Wong LM, Taubman K, Ting Lee S, Hsiao E, Roach P, Nottage M, Kirkwood I, Hayne D, Link E, Marusic P, Matera A, Herschtal A, Iravani A, Hicks RJ, Williams S, Murphy DG. Prostate-specific membrane antigen PET-CT in patients with high-risk prostate cancer before curative-intent surgery or radiotherapy (proPSMA): a prospective, randomised, multicentre study. *Lancet* 2020; **395**(10231): 1208-16.
19. Galanski M, Prokop M, Schaefer-Prokop C, Molen AJvd. Ganzkörper-Computertomographie. 2 ed. Stuttgart: Georg Thieme Verlag KG; 2007.
20. Enzweiler CNH. Morphologische und funktionelle Herzbildgebung mit der Computertomographie: Klinischer Stellenwert von Elektronenstrahl- und Mehrzeilen-Technologie. Charite - Universitätsmedizin Berlin; 2007.
21. Kalender WA, Seissler W Fau - Klotz E, Klotz E Fau - Vock P, Vock P. Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation. (0033-8419 (Print)).
22. Schober O, Heindel W. PET-CT. Stuttgart: Georg Thieme Verlag KG; 2008.
23. WorkmanJr RB, Coleman RE. PET/CT: Springer, New York, NY; 2006.

24. Dieplinger AM. Einsatz der Ga-68 DOTATOC PET/CT bei der Detektion unbekannter NET-Primarien: Medizinische Fakultät Charité – Universitätsmedizin Berlin; 2018.
25. Bailey DL. Quantitative Procedures in 3D PET. In: Bendriem B, Townsend DW, eds. *The Theory and Practice of 3D PET*. Dordrecht: Springer Netherlands; 1998: 55-109.
26. Li R, Ravizzini GC, Gorin MA, Maurer T, Eiber M, Cooperberg MR, Alemozzaffar M, Tollefson MK, Delacroix SE, Chapin BF. The use of PET/CT in prostate cancer. (1476-5608 (Electronic)).
27. Afshar-Oromieh A, Malcher A Fau - Eder M, Eder M Fau - Eisenhut M, Eisenhut M Fau - Linhart HG, Linhart Hg Fau - Hadaschik BA, Hadaschik Ba Fau - Holland-Letz T, Holland-Letz T Fau - Giesel FL, Giesel Fl Fau - Kratochwil C, Kratochwil C Fau - Haufe S, Haufe S Fau - Haberkorn U, Haberkorn U Fau - Zechmann CM, Zechmann CM. PET imaging with a [68Ga]gallium-labelled PSMA ligand for the diagnosis of prostate cancer: biodistribution in humans and first evaluation of tumour lesions. (1619-7089 (Electronic)).
28. Eiber M, Maurer T, Souvatzoglou M, Beer AJ, Ruffani A, Haller B, Graner FP, Kubler H, Haberkorn U, Eisenhut M, Wester HJ, Gschwend JE, Schwaiger M. Evaluation of Hybrid (6)(8)Ga-PSMA Ligand PET/CT in 248 Patients with Biochemical Recurrence After Radical Prostatectomy. (1535-5667 (Electronic)).
29. Afshar-Oromieh A, Babich JW, Kratochwil C, Giesel FL, Eisenhut M, Kopka K, Haberkorn U. The Rise of PSMA Ligands for Diagnosis and Therapy of Prostate Cancer. *Journal of Nuclear Medicine* 2016; **57**(Supplement 3): 79S-89S.
30. Treglia G, Pereira Mestre R, Ferrari M, Bosetti DG, Pascale M, Oikonomou E, De Dosso S, Jermini F, Prior JO, Roggero E, Giovanella L. Radiolabelled choline versus PSMA PET/CT in prostate cancer restaging: a meta-analysis. (2160-8407 (Print)).
31. Pfister D, Porres D, Heidenreich A, Heidegger I, Knuechel R, Steib F, Behrendt FF, Verburg FA. Detection of recurrent prostate cancer lesions before salvage lymphadenectomy is

more accurate with (68)Ga-PSMA-HBED-CC than with (18)F-Fluoroethylcholine PET/CT. (1619-7089 (Electronic)).

32. Sweat SD, Pacelli A Fau - Murphy GP, Murphy Gp Fau - Bostwick DG, Bostwick DG. Prostate-specific membrane antigen expression is greatest in prostate adenocarcinoma and lymph node metastases. (0090-4295 (Print)).

33. Psimadas D, Valotassiou V, Alexiou S, Tsougos I, Georgoulas P. Radiolabeled mAbs as Molecular Imaging and/or Therapy Agents Targeting PSMA. *Cancer Investigation* 2018; **36**(2): 118-28.

34. Lutje S, Heskamp S, Cornelissen AS, Poepel TD, van den Broek SA, Rosenbaum-Krumme S, Bockisch A, Gotthardt M, Rijpkema M, Boerman OC. PSMA Ligands for Radionuclide Imaging and Therapy of Prostate Cancer: Clinical Status. (1838-7640 (Electronic)).

35. Schwarzenboeck SM, Rauscher I, Bluemel C, Fendler WP, Rowe SP, Pomper MG, Asfhar-Oromieh A, Herrmann K, Eiber M. PSMA Ligands for PET Imaging of Prostate Cancer. (1535-5667 (Electronic)).

36. von Eyben FE, Picchio M, von Eyben R, Rhee H, Bauman G. (68)Ga-Labeled Prostate-specific Membrane Antigen Ligand Positron Emission Tomography/Computed Tomography for Prostate Cancer: A Systematic Review and Meta-analysis. LID - S2405-4569(16)30160-2 [pii] LID - 10.1016/j.euf.2016.11.002 [doi]. (2405-4569 (Electronic)).

37. Weineisen M, Schottelius M, Simecek J, Baum RP, Yildiz A, Beykan S, Kulkarni HR, Lassmann M, Klette I, Eiber M, Schwaiger M, Wester H-J. 68Ga- and 177Lu-Labeled PSMA I&T: Optimization of a PSMA-Targeted Theranostic Concept and First Proof-of-Concept Human Studies. *Journal of Nuclear Medicine* 2015; **56**(8): 1169-76.

38. Campbell PJ, Yachida S Fau - Mudie LJ, Mudie Lj Fau - Stephens PJ, Stephens Pj Fau - Pleasance ED, Pleasance Ed Fau - Stebbings LA, Stebbings La Fau - Morsberger LA, Morsberger La Fau - Latimer C, Latimer C Fau - McLaren S, McLaren S Fau - Lin M-L, Lin Ml

Fau - McBride DJ, McBride Dj Fau - Varela I, Varela I Fau - Nik-Zainal SA, Nik-Zainal Sa Fau - Leroy C, Leroy C Fau - Jia M, Jia M Fau - Menzies A, Menzies A Fau - Butler AP, Butler Ap Fau - Teague JW, Teague Jw Fau - Griffin CA, Griffin Ca Fau - Burton J, Burton J Fau - Swerdlow H, Swerdlow H Fau - Quail MA, Quail Ma Fau - Stratton MR, Stratton Mr Fau - Iacobuzio-Donahue C, Iacobuzio-Donahue C Fau - Futreal PA, Futreal PA. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. (1476-4687 (Electronic)).

39. 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. *The New England journal of medicine* 2012; **366**(10): 883-92.

40. Nelson DA, Tan Tt Fau - Rabson AB, Rabson Ab Fau - Anderson D, Anderson D Fau - Degenhardt K, Degenhardt K Fau - White E, White E. Hypoxia and defective apoptosis drive genomic instability and tumorigenesis. (0890-9369 (Print)).

41. Tolkach Y Fau - Kristiansen G, Kristiansen G. The Heterogeneity of Prostate Cancer: A Practical Approach. (1423-0291 (Electronic)).

42. Lindberg J, Kristiansen A, Wiklund P, Gronberg H, Egevad L. Tracking the origin of metastatic prostate cancer. (1873-7560 (Electronic)).

43. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebbers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications* 2014; **5**: 4006.

44. Robert M. Haralick KS, Its'hak Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics* 1973; **SMC-3**(6): 610-21.

45. „Textur“ auf Dudenonline. <https://www.duden.de/rechtschreibung/Textur> (Zugriff am 10.09.2018).
46. Norman Rengers TP. JAVA-basierte Texturanalyse mittels Neighborhood Gray-Tone Differenz Matrix (NGTDM) zur Optimierung von Landnutzungsklassifikationen in hoch auflösenden Fernerkundungsdaten. *Geoinformation* 2009: 455-67.
47. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. *Cancer Imaging* 2013; **13**(1): 140-9.
48. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. *RadioGraphics* 2017; **37**(5): 1483-503.
49. Radiomic Features — pyradiomics v3.0. <https://pyradiomics.readthedocs.io/en/latest/features.html> (Zugriff am 18.03.2020).
50. Bashir U, Siddique MM, McLean E, Goh V, Cook GJ. Imaging Heterogeneity in Lung Cancer: Techniques, Applications, and Challenges. (1546-3141 (Electronic)).
51. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. (1527-1315 (Electronic)).
52. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S. Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology* 2017; **14**: 749.
53. Rizzo SA-Ohoo, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, Bellomi M. Radiomics: the facts and the challenges of image analysis. (2509-9280 (Electronic)).

54. Sha X, Gong G, Qiu Q, Duan J, Li D, Yin Y. Discrimination of mediastinal metastatic lymph nodes in NSCLC based on radiomic features in different phases of CT imaging. (1471-2342 (Electronic)).
55. Giesel FL, Schneider F, Kratochwil C, Rath D, Moltz J, Holland-Letz T, Kauczor H-U, Schwartz LH, Haberkorn U, Flechsig P. Correlation Between SUVmax and CT Radiomic Analysis Using Lymph Node Density in PET/CT-Based Lymph Node Staging. *Journal of Nuclear Medicine* 2017; **58**(2): 282-7.
56. Hamm CA, Beetz NL, Savic LJ, Penzkofer T. [Artificial intelligence and radiomics in MRI-based prostate diagnostics]. (1432-2102 (Electronic)).
57. Algohary AA-Ohoo, Viswanath S, Shiradkar R, Ghose S, Pahwa S, Moses DA-Ohoo, Jambor I, Shnier R, Bohm M, Haynes AM, Brenner P, Delprado W, Thompson J, Pulbrock M, Purysko ASA-Ohoo, Verma S, Ponsky L, Stricker P, Madabhushi A. Radiomic features on MRI enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. LID - 10.1002/jmri.25983 [doi]. (1522-2586 (Electronic)).
58. Antonelli M, Johnston EW, Dikaios N, Cheung KK, Sidhu HS, Appayya MB, Giganti F, Simmons LAM, Freeman A, Allen C, Ahmed HU, Atkinson D, Ourselin S, Punwani SA-Ohoo. Machine learning classifiers can predict Gleason pattern 4 prostate cancer with greater accuracy than experienced radiologists. (1432-1084 (Electronic)).
59. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, Zheng J, Goldman D, Moskowitz C, Fine S, 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. *European radiology* 2015; **25**(10): 2840-50.
60. Huang B, Law MW, Khong PL. Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. *Radiology* 2009; **251**(1): 166-74.

61. Baessler BA-Ohoo, Nestler T, Pinto Dos Santos D, Paffenholz P, Zeuch V, Pfister D, Maintz D, Heidenreich A. Radiomics allows for detection of benign and malignant histopathology in patients with metastatic testicular germ cell tumors prior to post-chemotherapy retroperitoneal lymph node dissection. (1432-1084 (Electronic)).
62. Seidler M, Forghani B, Reinhold C, Perez-Lara A, Romero-Sanchez G, Muthukrishnan N, Wichmann JL, Melki G, Yu E, Forghani R. Dual-Energy CT Texture Analysis With Machine Learning for the Evaluation and Characterization of Cervical Lymphadenopathy. (2001-0370 (Print)).
63. Surti S, Kuhn A, Werner ME, Perkins AE, Kolthammer J, Karp JS. Performance of Philips Gemini TF PET/CT scanner with special consideration for its time-of-flight imaging capabilities. *J Nucl Med* 2007; **48**(3): 471-80.
64. Development Core Team R. R: A Language and Environment for Statistical Computing; 2011.
65. 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. (1538-7445 (Electronic)).
66. Larue RTHM, Defraene G, De Ruysscher D, Lambin P, van Elmpt W. Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *The British Journal of Radiology* 2016; **90**(1070): 20160665.
67. 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.
68. Owens CA, Peterson CB, Tang C, Koay EJ, Yu W, Mackin DS, Li J, Salehpour MR, Fuentes DT, Court LE, Yang J. Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLOS ONE* 2018; **13**(10): e0205003.

69. Acar EA-Ohoo, Leblebici A, Ellidokuz BE, Basbinar Y, Kaya GC. Machine learning for differentiating metastatic and completely responded sclerotic bone lesion in prostate cancer: a retrospective radiomics study. (1748-880X (Electronic)).
70. Peeken JC, Shouman MA, Kroenke M, Rauscher I, Maurer T, Gschwend JE, Eiber M, Combs SE. A CT-based radiomics model to detect prostate cancer lymph node metastases in PSMA radioguided surgery patients. *Eur J Nucl Med Mol Imaging* 2020; **47**(13): 2968-77.
71. Hartenstein A, Lübbe F, Baur ADJ, Rudolph MM, Furth C, Brenner W, Amthauer H, Hamm B, Makowski M, Penzkofer T. Prostate Cancer Nodal Staging: Using Deep Learning to Predict (68)Ga-PSMA-Positivity from CT Imaging Alone. *Sci Rep* 2020; **10**(1): 3398.

Eidesstattliche Versicherung

„Ich, Falk Konstantin Lübke, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: CT-basierte Radiomics-Modelle zur Vorhersage des ⁶⁸Ga-PSMA PET-Lymphknotenstatus bei Patienten mit Prostatakarzinom (Englisch: CT-based radiomics models to predict ⁶⁸Ga-PSMA PET-status of lymph nodes in patients with prostate carcinoma) 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) 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.“

24.04.2021

Falk Konstantin Lübke

Anteilerklärung an etwaigen erfolgten Publikationen

Falk Konstantin Lübbe hatte folgenden Anteil an den folgenden Publikationen:

Publikation 1: A. Hartenstein, F. Lübbe, A. D. J. Baur, M. M. Rudolph, C. Furth, W. Brenner, H. Amthauer, B. Hamm, M. Makowski & T. Penzkofer, *Prostate Cancer Nodal Staging: Using Deep Learning to Predict 68Ga-PSMA-Positivity from CT Imaging Alone*, Sci Rep. 2020 Feb 25;10(1):3398. doi: 10.1038/s41598-020-60311-z.

Beitrag: Beteiligung am Studiendesign, Erstellung des zugrundeliegenden Datensatzes, kritische Kommentierung des Artikels.

Die Methodik (Deep Learning auf Basis von Convolutional Neural Networks) und Ergebnisse dieser Publikation sind im Nachgang zu der vorliegenden Arbeit entstanden und nicht Teil dieser Arbeit, lediglich der zugrundeliegende Datensatz ist teilweise identisch. Die Ergebnisse der beiden Arbeiten werden in der Diskussion in Kontext gesetzt.

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

Unterschrift des Doktoranden

Lebenslauf

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

Publikationsliste

A. Hartenstein, F. Lübbe, A. D. J. Baur, M. M. Rudolph, C. Furth, W. Brenner, H. Amthauer, B. Hamm, M. Makowski & T. Penzkofer, Prostate Cancer Nodal Staging: Using Deep Learning to Predict ⁶⁸Ga-PSMA-Positivity from CT Imaging Alone, *Sci Rep.* 2020 Feb 25;10(1):3398. doi: 10.1038/s41598-020-60311-z.

L. Adams, F. Lübbe, K. Bressemer, M. Wagner, B. Hamm, M. Makowski, Non-alcoholic fatty liver disease in underweight patients with inflammatory bowel disease: A case-control study. *PLoS One.* 2018 Nov 14;13(11):e0206450. doi: 10.1371/journal.pone.0206450. eCollection 2018.

Danksagung

Zunächst möchte ich mich ganz herzlich bei meinen Betreuern Herrn PD Dr. Tobias Penzkofer und Herrn Prof. Dr. Markus Makowski bedanken, für die interessanten Einblicke in die wissenschaftliche Arbeit und für die tatkräftige und intensive Betreuung während meiner Promotion.

Für die statistische Begleitung der Arbeit gilt mein Dank Herrn Dr. Paul Schmidt.

Für die Durchsicht meiner Dissertation und die hilfreichen Denkanstöße möchte ich mich herzlich bei Herrn Dr. Boris Gorodetski bedanken.

Mein besonderer Dank gilt meinen Eltern, Frau Dr. Dr. Jutta Kallage-Lübbe und Herrn Dr. Paul Joachim Lübbe, ohne die mein Medizinstudium und meine Promotion nicht möglich gewesen wären.