

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

DISSERTATION

**Radiomics-basierte CT-Bilddaten-Analyse zur nicht-invasiven Vorhersage des
histopathologischen Tumorgrads klarzelliger Nierenzellkarzinome**

Radiomics-based CT image analysis to predict the pathological grade in clear cell
renal cell carcinoma

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

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Maximilian Winter

aus Mannheim

Datum der Promotion: 30.11.2023

Inhaltsverzeichnis

Tabellenverzeichnis	4
Abbildungsverzeichnis	5
Zusammenfassung	6
Abstract	8
1 Einleitung	10
1.1 Epidemiologie des Nierenzellkarzinoms	10
1.2 Histologische Subtypen des Nierenzellkarzinoms und benigne Alternativbefunde	10
1.3 Graduierung des Nierenzellkarzinoms	11
1.4 Stadieneinteilung, Prognose und Therapie des Nierenzellkarzinoms.....	13
1.5 Überblick zur Diagnostik des Nierenzellkarzinoms	15
1.6 Die Bedeutung der Computertomographie für die Diagnostik des Nierenzellkarzinoms.....	16
1.6.1 Bildeigenschaften verschiedener renaler Raumforderungen in der computertomographischen Bildgebung	17
1.6.2 Bisherige technische Entwicklungen der computertomographischen Bildgebung	19
1.7 Künstliche Intelligenz in der Radiologie	20
1.7.1 Maschinelles Lernen	21
1.8 Radiomics	22
1.8.1 Akquisition der Bilddatensätze	23
1.8.2 Segmentierung	24
1.8.3 Extraktion der quantitativen Bildmerkmale	25
1.8.4 Selektion und Vorverarbeitung der quantitativen Bildmerkmale.....	28
1.8.5 Trainieren eines Modells des maschinellen Lernens	28
1.8.6 Erhebung statistischer Maße zur Evaluation eines Modells des maschinellen Lernens	30

1.9 Fragestellung	32
2 Methodik.....	34
2.1 Auswahl der Studienpopulation	34
2.2 Erhebung der Patientendaten.....	35
2.3 Segmentierung.....	35
2.4 Extraktion der quantitativen Bildmerkmale.....	39
2.5 Selektion von Bildmerkmalen mit hoher Reproduzierbarkeit	39
2.5.1 Voxel Dropout.....	39
2.5.2 Simulation von zwei zusätzlichen Annotatoren	40
2.5.3 Rauschreduktion/Rauscherhöhung	40
2.5.4 Bildschärfung/Weichzeichnung	40
2.6 Datenunterteilung und Oversampling	41
2.7 Ausschluss von Bildmerkmalen mit starker paarweiser Korrelation	42
2.8. Selektion von Bildmerkmalen mittels ANOVA.....	44
2.9 Trainieren von Modellen des maschinellen Lernens.....	45
2.9.1 Überblick über die vier verwendeten Modelle	45
2.9.2 Training des Random Forest Modells	49
2.9.3 Training des Gradient Boosting Modells	50
2.9.4 Training des L1/L2 regularisierten Regressionsmodells	50
2.9.5 Training des k-Nearest Neighbor Modells.....	50
3 Ergebnisse	51
3.1 Studienpopulation	51
3.2 Ergebnisse des Trainings der Modelle des maschinellen Lernens.....	55
3.2.1 Ergebnisse des Random Forest Modells	58
3.2.2 Ergebnisse des Gradient Boosting Modells	59
3.2.3 Ergebnisse des L1/L2 regularisierten Regressionsmodells	60
3.2.4 Ergebnisse des k-Nearest Neighbor Modells	61
4 Diskussion	62

4.1 Interpretation der Ergebnisse.....	62
4.1.1 Alleinstellungsmerkmale der vorliegenden Arbeit	62
4.1.2 Einordnung der Studie in die Literatur.....	62
4.2 Limitationen der vorliegenden Arbeit	68
4.2.1 Gruppierung der Tumorgade und Klassenimbalance	68
4.2.2 Monozentrität der Studie.....	70
4.2.3 Manuelle Segmentierung	71
4.2.4 Verwendung von ein- bis dreiphasigen CT-Aufnahmen.....	72
4.2.5 Einschluss biopsierter Tumoren	72
4.3 Möglichkeiten und Grenzen von Radiomics - ein Ausblick	73
4.3.1 Weitere Anwendungen von Radiomics in der computertomographischen Diagnostik des klarzelligen Nierenzellkarzinoms	73
4.3.2 Radiogenomics.....	73
4.3.3 Zukünftige Herausforderungen von Radiomics	74
4.3.4 Deep Learning	75
4.3.5 Deep Learning zur nicht-invasiven Bestimmung des histopathologischen Tumorgads klarzelliger Nierenzellkarzinome	78
Literaturverzeichnis	79
Eidesstattliche Versicherung	91
Lebenslauf.....	92
Danksagung	93
Bescheinigung der statistischen Beratung	94

Tabellenverzeichnis

Tabelle 1: Verteilung der Tumorgrade in den Trainings- und Testdaten.	41
Tabelle 2: Patienten- und Tumorcharakteristika.	52
Tabelle 3: Nach Tumorgraden und CT-Phasen aufgeteilte Ergebnisse des Random Forest Modells bezüglich der Tumorgradklassifizierung.	58
Tabelle 4: Nach CT-Phasen aufgeteilte Ergebnisse des Random Forest Modells bezüglich der Tumorgradklassifizierung.	58
Tabelle 5: Nach Tumorgraden und CT-Phasen aufgeteilte Ergebnisse des Gradient Boosting Modells bezüglich der Tumorgradklassifizierung.	59
Tabelle 6: Nach CT-Phasen aufgeteilte Ergebnisse des Gradient Boosting Modells bezüglich der Tumorgradklassifizierung.	59
Tabelle 7: Nach Tumorgraden und CT-Phasen aufgeteilte Ergebnisse des L1/L2 regularisierten Regressionsmodells bezüglich der Tumorgradklassifizierung.	60
Tabelle 8: Nach CT-Phasen aufgeteilte Ergebnisse des L1/L2 regularisierten Regressionsmodells bezüglich der Tumorgradklassifizierung.	60
Tabelle 9: Nach Tumorgraden und CT-Phasen aufgeteilte Ergebnisse des k-Nearest Neighbor Modells bezüglich der Tumorgradklassifizierung.	61
Tabelle 10: Nach CT-Phasen aufgeteilte Ergebnisse des k-Nearest Neighbor Modells bezüglich der Tumorgradklassifizierung.	61

Abbildungsverzeichnis

Abbildung 1: Exemplarische Darstellung einer Grey-Level Co-Occurrence-Matrix.	26
Abbildung 2: Exemplarische Darstellung einer Grauwerte-Lauflängenmatrix.	27
Abbildung 3: Exemplarische Darstellung zweier ROC-Kurven.	31
Abbildung 4: Darstellung der Tumorsegmentierung in ITK-Snap.	36
Abbildung 5: Bildschirmaufnahmen des in der Abbildung 4 segmentierten klarzelligen Nierenzellkarzinoms (WHO/ISUP G1) der rechten Niere in derselben Schnittebene in drei verschiedenen Phasen.	38
Abbildung 6: Farbkodierte Korrelationsmatrix der nativen CT-Aufnahmen.	43
Abbildung 7: Farbkodierte Korrelationsmatrix der arteriellen CT-Aufnahmen.	43
Abbildung 8: Farbkodierte Korrelationsmatrix der venösen CT-Aufnahmen.	44
Abbildung 9: Grundlegender Aufbau eines Entscheidungsbaums.	46

Zusammenfassung

Hintergrund

Der histopathologische Tumorgrad klarzelliger Nierenzellkarzinome wird aktuell durch die Untersuchung eines Tumorresektats oder einer Stanzzyylinderbiopsie ermittelt. Radiomics-basierte CT-Bilddaten-Analysen könnten eine nicht-invasive Bestimmung des Tumorgrads erlauben. Die Prinzip von Radiomics beruht darauf, dass sich Veränderungen auf mikroskopischer Ebene auch in strukturellen, quantitativ messbaren Bilddatenunterschieden manifestieren, diese Veränderungen in radiologischen Datensätzen erkannt und folglich zu klinischen Parametern in Beziehung gesetzt werden können.

Methoden

Untersucht wurden die computertomographischen Aufnahmen von 382 Patienten mit 388 klarzelligem Nierenzellkarzinomen. Auf die manuelle Segmentierung der Tumore folgte die Extraktion von 1221 quantitativen Bildmerkmalen. Nach der Merkmalsselektion mittels einer Reproduzierbarkeitsanalyse, einer Korrelationsanalyse und ANOVA verblieben 4 Merkmale für die native CT, 52 Merkmale für die arterielle CT und 17 Merkmale für die venöse CT. Es wurden vier Modelle des maschinellen Lernens bezüglich der Unterscheidung zwischen niedriggradigen (WHO/ISUP / Fuhrman G1), mittelgradigen (WHO/ISUP / Fuhrman G2-G3) und hochgradigen (WHO/ISUP / Fuhrman G4) Tumoren trainiert: Ein Random Forest Modell, ein Gradient Boosting Modell, ein L1/2 regularisiertes Regressionsmodell und ein k-Nearest-Neighbor Modell. Das Training der Modelle erfolgte jeweils getrennt für die native, die arterielle und die venöse CT.

Ergebnisse

Das Gradient Boosting Modell (Accuracy: arteriell 77%, venös 81%, nativ 80%) und das Random Forest Modell (Accuracy: arteriell 86%, venös 83%, nativ 50%) erzielten die besten Ergebnisse bei der Klassifizierung des Tumorgrads. Das L1/2 regularisierte Regressionsmodell (Accuracy: arteriell 77%, venös 56%, nativ 80%) und das k-Nearest-Neighbor Modell (Accuracy: arteriell 50%, venös 62%, nativ 70%) ergaben jeweils schlechtere Ergebnisse.

Schlussfolgerung

Mit der vorliegenden Studie konnte gezeigt werden, dass mittels Radiomics-basierter CT-Analysen der histopathologische Tumorgrad klarzelliger Nierenzellkarzinome mit teilweise über 80% Accuracy vorhergesagt werden kann. Die besten Ergebnisse gelangen hierbei anhand kontrastmittelgestützter Aufnahmen. Es sind in Zukunft umfangreichere, multizentrische Studien notwendig, um noch bessere Radiomics-Modelle zu entwickeln und somit gegebenenfalls eine klinische Anwendung von Radiomics zu ermöglichen. Der Einsatz von Deep Learning könnte hierzu entscheidend beitragen.

Abstract

Background

The histopathologic tumor grade of clear cell renal cell carcinomas is currently determined after partial/radical nephrectomy or biopsy. Radiomics-based CT image data analysis might enable non-invasive assessment of tumor grade. Radiomics is based on the fact that changes at the microscopic level also manifest in structural, quantitatively measurable image data differences, which can be detected in radiological data sets and consequently related to clinical parameters.

Methods

The CT images of 382 patients with 388 clear cell renal cell carcinomas were studied. Manual segmentation of the tumors was followed by extraction of 1221 quantitative image features. After feature selection based on reproducibility analysis, correlation analysis, and ANOVA, 4 features remained for unenhanced CT, 52 features for arterial CT, and 17 features for venous CT. Four machine learning models were trained for discrimination between low-grade (WHO/ISUP/Fuhrman G1), intermediate-grade (WHO/ISUP/Fuhrman G2-G3), and high-grade (WHO/ISUP/Fuhrman G4) tumors. This included a random forest model, a gradient boosting model, an L1/2 regularized regression model, and a k-nearest-neighbour model. Training of each model was performed separately for native, arterial, and venous CT.

Results

The gradient boosting model (accuracy: arterial 77%, venous 81%, native 80%) and the random forest model (accuracy: arterial 86%, venous 83%, native 50%) achieved the best results in classifying tumor grade. The L1/2 regularized regression model (Accuracy: arterial 77%, venous 56%, native 80%) and the k-nearest-neighbor model (Accuracy: arterial 50%, venous 62%, native 70%) yielded comparatively worse results, respectively.

Conclusion

The present study demonstrates that radiomics-based CT analysis can predict the histopathological grade of clear cell renal cell carcinoma with more than 80% accuracy. The best results were obtained using contrast-enhanced images. In the future, however, more extensive, multicenter studies are needed to further improve radiomics models and

pave the way for translation to clinical practice. In this process, deep learning could play a crucial role.

1 Einleitung

1.1 Epidemiologie des Nierenzellkarzinoms

Nierenzellkarzinome entstehen aus dem Tubulusepithel der Nephrone und sind verantwortlich für über 90% aller malignen Nierentumore (1). Global betrachtet haben Neoplasien der Niere einen Anteil von circa 2,2% an der Gesamtheit aller malignen Tumorerkrankungen und sind ursächlich für ungefähr 1,8% aller krebsbedingten Todesfälle. Das entspricht jährlich in etwa 430.000 Neuerkrankungen und 180.000 Todesfällen (2). In Deutschland liegt die jährliche Inzidenz bei circa 14.900 Fällen, die Anzahl der Todesfälle beträgt ungefähr 5300 pro Jahr. (3) Die höchste Inzidenz lässt sich in Nordamerika und Europa finden, am niedrigsten ist sie in Asien und Afrika (4). Während die Inzidenz weltweit zunimmt, bleibt die Mortalität überwiegend stabil und verringert sich sogar in vielen westlichen Ländern (4). Männer erkranken ungefähr 1,5 mal häufiger als Frauen (2). Das durchschnittliche Alter zum Zeitpunkt der Diagnosestellung beträgt circa 63 Jahre (5).

Es sind verschiedene Risikofaktoren für die Entstehung von Nierenzellkarzinomen bekannt. Nikotinkonsum gehört dazu, wobei das Risiko mit Zunahme der Konsummenge ansteigt (6). Übergewicht und Hypertension stellen weitere Risikofaktoren dar (7). Genetische Syndrome wie das Von-Hippel-Lindau-Syndrom, das hereditäre papilläre Nierenzellkarzinom, das Birt-Hogg-Dubé Syndrom, die tuberöse Sklerose und die hereditäre Leiomyomatose gehen ebenfalls mit einem erhöhten Auftreten von Nierenzellkarzinomen einher (8). Ein weiterer Zusammenhang konnte zu Trichlorethenexposition hergestellt werden (8).

1.2 Histologische Subtypen des Nierenzellkarzinoms und benigne Alternativbefunde

Die 2016 erschienene WHO-Klassifikation für Nierenzellkarzinome teilt diese in mehrere Subtypen ein. Sie werden anhand verschiedener Aspekte wie der Histologie, Chromosomenaberrationen und molekularen Mechanismen unterschieden (9). Klarzellige Nierenzellkarzinome stellen mit circa 80% den häufigsten Tumorsubtyp dar (10). Mit einem Anteil von etwa 15% bilden papilläre Nierenzellkarzinome den zweithäufigsten Subtyp, chromophobe Nierenzellkarzinome folgen darauf mit ungefähr 5% (10, 11). Daneben gibt es noch zahlreiche seltenere histologische Subtypen (9). Die

Klassifikation von Nierenzellkarzinomen unterliegt seit Jahrzehnten Veränderungen, wodurch es regelmäßig zur Definition neuer Tumorsubtypen und zur Änderung der Einteilung der bereits definierten Subtypen kommt (12).

Weitere maligne Raumforderungen der Niere sind unter anderem Lymphome und Metastasen anderer maligner Tumore, zum Beispiel Lungenkarzinommetastasen (13).

Nierenzysten stellen die häufigste Raumforderung der Niere und somit auch den häufigsten benignen Alternativbefund zu Nierenzellkarzinomen dar (14). Es besteht jedoch die Möglichkeit, dass die Zysten maligne entarten (14). Während die Prävalenz von Nierenzysten in einer Studie, die Computertomographien auf deren Vorliegen untersucht hat, bei circa 41% lag (15), konnte eine Autopsiestudie bei Patienten über 50 Jahren sogar eine Prävalenz von über 50% zeigen (16). Grundsätzlich steigt die Häufigkeit von Nierenzysten mit dem Alter (14).

Neben Zysten gibt es verschiedene solide benigne Raumforderungen der Niere, deren Ursprung entweder im Nierenparenchym oder im Bindegewebe liegt. Dazu gehören Onkozytome, die mit einem Anteil von in etwa 3-7% aller Neoplasien der Niere den häufigsten soliden benignen Nierentumor darstellen (17). Sie finden ihren Ursprung in den interkalierenden Zellen des Nierenparenchyms (17). Ungefähr 3% aller Nierentumore werden durch Angiomyolipome gebildet (18). Auch sie sind benigne und enthalten Muskelgewebe, Fettgewebe und Gefäße (18).

1.3 Graduierung des Nierenzellkarzinoms

1982 wurde durch *Fuhrman et al.* ein Graduierungssystem für Nierenzellkarzinome vorgestellt, das anhand der Zellkerngröße, der Form der Zellkerne und der gegebenenfalls vorhandenen Prominenz der Nukleolen vier verschiedene Grade unterscheidet. Ein Tumor mit runden Zellkernen, circa 10 Mikrometern Kerngröße und nicht sichtbaren Nukleolen entspricht Grad 1. Grad 2 liegt vor, wenn die Zellkerne eine unregelmäßige Form annehmen, ungefähr 15 Mikrometer groß sind und kleine Nukleolen bei 400-facher Vergrößerung sichtbar sind. Bei Grad 3 besitzen die Zellkerne weiterhin eine unregelmäßige Form, sind in etwa 20 Mikrometer groß und weisen bei 100-facher Vergrößerung prominente Nukleolen, auf. Das Vorhandensein von pleomorphen Zellkernen, Spindelzellen und Chromatinklumpen lässt auf Grad 4 schließen (19). Die Nutzung des Fuhrman-Graduierungssystems ist in Europa und Nordamerika weit verbreitet (20, 21). Hinsichtlich der krankheitsspezifischen Mortalität stellt es für

klarzellige Nierenzellkarzinome einen unabhängigen Prädiktor dar (22). Es wurde festgestellt, dass die Prognose sich mit steigendem Tumorgrad verschlechtert, wobei die signifikanten Unterschiede zwischen den einzelnen Graden in den meisten Studien nur durch Gruppierung, beispielsweise in hochgradige (Grad 3 + 4) und niedriggradige Fälle (Grad 1 + 2), demonstriert werden konnten (23). Für chromophobe Nierenzellkarzinome konnte kein signifikanter Zusammenhang zwischen Prognose und Tumorgrad gezeigt werden (24). Ein umstrittener Aspekt der Einteilung nach Fuhrman ist, dass für eine eindeutige Zuteilung zu einem Grad jeweils alle drei Parameter zutreffen müssen. Es wird nicht eindeutig beschrieben, welcher Parameter mehr gewichtet werden sollte, wenn die jeweils erfassten Parameter verschiedenen Graden zuzuordnen sind (25). Für ungefähr 20% aller klarzelligen Nierenzellkarzinome lässt sich daher kein eindeutiger Grad bestimmen (23). Dadurch lässt sich möglicherweise die ebenfalls regelmäßig kritisierte schlechte Interrater-Reliabilität zumindest teilweise erklären (26).

Die International Society of Urological Pathology (ISUP) hat 2012 ein neues System zur Graduierung von Nierenzellkarzinomen vorgestellt, das seit 2015 leicht angepasst auch von der WHO empfohlen wird und folglich als WHO/ISUP-Graduierungssystem etabliert wurde (9, 23, 27). Es beruht ebenfalls auf einer Einteilung in vier verschiedene Tumorgrade, wobei für Grad 1-3 die Sichtbarkeit der Nukleolen entscheidend ist. Bei einer Vergrößerung um den Faktor 400 kaum beziehungsweise nicht sichtbare, basophile Nukleolen entsprechen Grad 1. Grad 2 trifft zu, wenn die Nukleolen bei 400-facher Vergrößerung zu sehen und eosinophil sind. Ausschlaggebend für eine Zuteilung zu Grad 3 ist, dass bereits bei einer Vergrößerung um den Faktor 100 Nukleolen eindeutig sichtbar sind. Ein Tumor wird dem Grad 4 zugeordnet, wenn die Zellen starke Dysmorphien zeigen und sich mit einer rhabdoiden und/oder sarkomatoiden Differenzierung präsentieren (27). Für klarzellige und papilläre Nierenzellkarzinome fand bereits eine Validierung des WHO/ISUP-Graduierungssystems statt, bei chromophoben Nierenzellkarzinomen sollte es jedoch nicht angewendet werden (9, 23, 27). Bei klarzelligen Nierenzellkarzinomen geht der Grad 1 nach bisheriger Datenlage mit einer sehr guten Prognose einher, die krankheitsspezifische 5-Jahres-Überlebensrate liegt bei circa 95%. Bei den Graden 2, 3 und 4 beträgt sie in derselben Reihenfolge ungefähr 80%, 45% und 20%, wobei die Unterschiede bezüglich der Prognose zwischen den einzelnen Graden signifikant sind (28). Im Gegensatz zum Fuhrman-Graduierungssystem benötigt man nach bisherigen Erkenntnissen bei dem WHO-Graduierungssystem also keine Gruppierung der Tumorgrade, um eine prognostische Aussage treffen zu können.

Hierdurch und durch die vereinfachte Zuteilung eines Tumors zu einem der vier Grade scheint das WHO-Graduierungssystem dem Fuhrman-Graduierungssystem bisher überlegen (23).

1.4 Stadieneinteilung, Prognose und Therapie des Nierenzellkarzinoms

Nierenzellkarzinome werden anhand der durch die *Union internationale contre le cancer (UICC)* entwickelten TNM-Klassifikation in vier verschiedene Stadien eingeteilt (10). Die Stadien können entweder nur anhand der Tumorausdehnung (Stadium T1-T4) oder anhand einer Kombination aus den Parametern Tumorausdehnung (T), Lymphknotenbefall (N) und Fernmetastasen (M) (Stadium 1-4) definiert werden (10).

Wenn der Tumor unter 7 cm misst und auf eine Niere begrenzt ist, liegt das Stadium T1 vor. Dieses wird weiter aufgeteilt in das Stadium T1a, für das der Tumor kleiner als 4cm sein muss, und das Stadium T1b, dem ein Tumor von 4cm bis einschließlich 7cm Größe zugeteilt wird. Ein auf eine Niere begrenzter Tumor von mindestens 7cm Größe entspricht dem Stadium T2. Wenn der Tumor über 7cm, aber unter 10cm misst, entspricht dies dem Stadium T2a. Das Stadium T2b trifft zu, wenn der Tumor über 10cm misst. Im Falle einer Infiltration in die Vena renalis, die Vena cava oder in das die Niere umgebende Fettgewebe spricht man von dem Stadium T3. Auch dieses wird weiter unterteilt. T3a liegt vor, wenn der Tumor das perirenale Fettgewebe oder die Vena renalis befällt. Bei einem subphrenischen Befall der Vena cava wird der Tumor dem Stadium T3b zugeteilt. T3c liegt vor, wenn die Vena cava subphrenisch infiltriert wird. Wenn der Tumor in die ipsilaterale Nebenniere hineinwächst oder ein Wachstum außerhalb der Gerota-Faszie erfolgt, wird er dem Stadium T4 zugeordnet (10).

Von N1 spricht man bei einem befallenen regionären Lymphknoten, für N2 müssen mehrere regionäre Lymphknoten befallen sein. Sobald Fernmetastasen vorliegen, handelt es sich um ein Metastasenstadium M1 (10).

Die Kombination von T1, N0 und M0 entspricht einem Stadium 1. T2, N0, M0 entspricht einem Stadium 2. Bei T3, N0, M0 oder T1-3, N1, M0 liegt ein Stadium 3 vor. Ein Stadium 4 ist definiert durch T4, N2 oder durch das Vorhandensein von Metastasen, also M1 (10). Mit fortschreitendem TNM-Stadium nimmt die krankheitsspezifische 5-Jahres-Überlebensrate signifikant ab (29). Im klinischen Alltag werden verschiedene Prognosescores verwendet, die neben der Stadieneinteilung noch weitere Parameter mit prognostischem Aussagewert berücksichtigen und dadurch eine differenziertere

Risikoabschätzung bei Patienten mit Nierenzellkarzinom ermöglichen (10). Bei bereits metastasierten Nierenzellkarzinomen kann zum Beispiel das Memorial Sloan Kettering Cancer Center (MSKCC) System, auch Motzer-Score genannt, zur Abschätzung des krankheitsspezifischen Überlebens verwendet werden (30). Auf diesem Score basiert der International Metastatic RCC Database Consortium (IMDC) Score, der sich aus dem Karnofsky-Index, dem Hämoglobin, dem korrigierten Calcium, der Laktatdehydrogenase, der Thrombozytenzahl und dem Zeitraum zwischen Diagnosestellung und Therapieeinleitung zusammensetzt (30).

Der sogenannte Stage, Size, Grade and Necrosis (SSIGN) Score und der University of California Los Angeles Score (UISS) können bei nicht-metastasierten Nierenzellkarzinomen verwendet werden (10, 31).

Für das Nierenzellkarzinom stehen verschiedene Therapieansätze zur Verfügung. Eine chirurgische Entfernung wird in der großen Mehrzahl der Fälle angestrebt und ist obligat bei kurativer Therapieabsicht, wobei entweder nierenerhaltend eine Teilnephrektomie oder alternativ eine radikale Nephrektomie durchgeführt werden kann (10). Beide Verfahren können sowohl laparoskopisch als auch offen-chirurgisch umgesetzt werden (10, 32). Die Wahl des Verfahrens hat jedoch keinen Einfluss auf das krankheitsspezifische Überleben der Patienten (10, 32). Während im Stadium T1 tendenziell die Teilnephrektomie bevorzugt wird, wird ab dem Stadium T2 aufgrund der Größe des Tumors die radikale Nephrektomie häufiger durchgeführt (10). Bei einem fehlenden pathologischen Befund in der Bildgebung und fehlenden intraoperativen Auffälligkeiten wird im Rahmen der Nephrektomie keine gleichzeitige Entfernung der Lymphknoten empfohlen (10, 32). Die Nebennieren sollten nicht routinemäßig, sondern nur bei einem wahrscheinlichen oder nachgewiesenen Befall mitentfernt werden (32). Bei vorliegenden Metastasen sollte immer eine chirurgische Entfernung oder eine Radiatio der Metastasen in Betracht gezogen werden (32). Im Falle einer geringen Lebenserwartung und einer Vielzahl von Begleiterkrankungen können als palliative Therapieoption eine aktive Überwachung des Tumors oder, wenn gleichzeitig ein kleiner Tumor vorliegt, auch eine Entfernung mittels Kryoablation oder Thermoablation erwogen werden (32).

Hinsichtlich weiterer Therapieverfahren des Nierenzellkarzinoms lässt sich feststellen, dass sowohl klassische Chemotherapeutika als auch Bestrahlungstherapien nur wenig bis gar nicht wirksam sind (33). Aus diesem Grund werden sowohl in fortgeschrittenen Tumorstadien als auch bei vorhandenen Metastasen neuere Substanzen eingesetzt,

deren Angriffspunkte beispielsweise einzelne Rezeptoren oder Signalwege der Zellen darstellen, was auch als zielgerichtete Therapie bezeichnet wird (10, 32, 33). Hierbei werden unter anderem VEGF-Antikörper wie Bevacizumab, mTOR-Inhibitoren wie Everolimus, Tyrosinkinase-Inhibitoren wie Sunitinib, Multikinase-Inhibitoren wie Sorafenib, Interferon-alpha und PD1-Rezeptor-Inhibitoren wie Nivolumab eingesetzt (10, 32). Eine Kombination mehrerer Wirkstoffe ist zur Zeit nicht empfohlen und wird aktuell nur in klinischen Studien verabreicht (32). Die Wahl des Medikaments hängt maßgeblich vom Stadium und vom Risikoprofil des Patienten ab (32). Im Falle eines höheren Stadiums und eines gleichzeitig maximal intermediärem Risikoprofils im IMDC-Score beispielsweise ist das Mittel der ersten Wahl entweder Pazopanib, Bevacizumab in Kombination mit Interferon-alpha oder Sunitinib (32). Es sei an dieser Stelle erwähnt, dass die Therapieempfehlungen regelmäßigen Veränderungen unterliegen und sich hierzu zahlreiche Studien in der Durchführung befinden (32).

Nach kurativer Therapie entwickeln circa 30% der Patienten ein Tumorrezidiv (33).

1.5 Überblick zur Diagnostik des Nierenzellkarzinoms

Über 50% aller Nierenzellkarzinomdiagnosen sind Zufallsbefunde (10). Bei 25-30% aller Patienten hat bereits zum Zeitpunkt der Diagnose eine Metastasierung stattgefunden (34). Die Prognose bezüglich der mittleren 5-Jahres-Überlebensrate ist bei Personen mit Zufallsbefund signifikant besser als bei Personen, die zum Zeitpunkt der Diagnose bereits Symptome entwickelt haben (35). Zu möglichen klinischen Zeichen bei symptomatischen Patienten gehören Makrohämaturie, Flankenschmerzen und ein palpabler Tumor im Oberbauch (36). Diese drei Symptome werden auch als klassische Trias bezeichnet, wobei diese nur bei circa 10% aller Betroffenen zu finden ist (36). Paraneoplastische Syndrome wie beispielsweise Hyperkalzämie oder Polyglobulie lassen sich bei ungefähr 20% aller Patienten finden (37).

Neben der körperlichen Untersuchung und der Bestimmung von Laborparametern wie beispielsweise Kreatinin, Hämoglobin und Laktatdehydrogenase nimmt die Ultraschalluntersuchung als bildgebendes Verfahren eine zentrale Rolle in der Stellung der Verdachtsdiagnose von Nierenzellkarzinomen ein (10). Zur weiteren Abklärung dient im Verdachtsfall die Computertomographie (CT) (32). Falls eine Venenbeteiligung vermutet wird, sollten die Patienten eine Magnetresonanztomographie (MRT) erhalten (32). Darüber hinaus ist eine Kopfbildgebung, üblicherweise mittels MRT, bei klinischem

Verdacht auf Metastasierung des Tumors in das Gehirn indiziert (32). Die Sinnhaftigkeit von Biopsien in der Diagnostik von Nierenzellkarzinomen wird kontrovers diskutiert. Sowohl der histologische Subtyp als auch der histopathologische Tumorgrad lassen sich mithilfe einer Biopsie bestimmen (38). Es besteht bei diesem Eingriff jedoch ein geringes Risiko für Blutungen und für die Verschleppung von Tumorzellen (10, 38). Ein weiterer Nachteil von Biopsien von Nierenzellkarzinomen besteht darin, dass die Tumoren teilweise an verschiedenen Lokalisationen eine unterschiedliche Differenzierung des Gewebes und damit einhergehend einen abweichenden histopathologischen Tumorgrad aufweisen (39). Laut der aktuellen *S-3 Leitlinie Diagnostik, Therapie und Nachsorge des Nierenzellkarzinoms* ist eine Biopsie nur angezeigt, wenn sie relevant für die weitere Therapieform ist oder wenn ein perkutanes Ablationsverfahren als Therapie geplant ist (32). Das etablierte Biopsieverfahren stellt beim Verdacht auf Nierenzellkarzinom die Stanzzyylinderbiopsie dar (10).

1.6 Die Bedeutung der Computertomographie für die Diagnostik des Nierenzellkarzinoms

Mithilfe der CT können das lokale Tumorwachstum, ein Befall von Lymphknoten und eine Fernmetastasierung beurteilt werden (10).

Zur gezielten Diagnostik von Nierenraumforderungen werden CT-Aufnahmen regelhaft in drei verschiedenen Phasen aufgenommen (40). Zunächst wird vor Gabe des Kontrastmittels eine native Aufnahme von der Leberkuppe bis zur Symphyse angefertigt (32, 40). Ungefähr 25-80 Sekunden nach einer Kontrastmitteinjektion von 150 ml beginnt die arterielle (kortikomedulläre) Phase, während der die Nierenrinde und das Nierenmark, deren Dichtewerte sich in dieser Phase um bis zu 100 Hounsfield-Einheiten (HU) unterscheiden, deutlich voneinander abgrenzbar sind und die Nierenarterien eine starke Kontrastierung zeigen (40). Die Aufnahme während der arteriellen Phase sollte den Oberbauch abbilden (32). Nach insgesamt circa 85-120 Sekunden startet die frühvenöse (nephrographische) Phase, während der erneut eine Aufnahme von der Leberkuppe bis zur Symphyse angefertigt wird (32, 40). In dieser Phase stellen sich Rinde und Mark eher homogen dar (40). In einigen Fällen wird auch eine Aufnahme in der spätvenösen (exkretorischen) Phase erstellt. Während dieser Phase, die in der Regel nach 3-5 Minuten abläuft, findet sich das Kontrastmittel bereits im Nierenkelchsystem und die Nieren stellen sich hypodenser dar als in den vorigen Phasen (40). Zusätzlich zu den CT-

Aufnahmen des Abdomens und des Beckens sollte für das die Stadieneinteilung des Nierenzellkarzinoms eine CT-Aufnahme des Thorax durchgeführt werden (10).

1.6.1 Bildeigenschaften verschiedener renaler Raumforderungen in der computertomographischen Bildgebung

In der CT besitzen die verschiedenen Subtypen des Nierenzellkarzinoms unterschiedliche Bildeigenschaften (41). Es gilt hierbei jedoch zu beachten, dass sich diese Eigenschaften zwischen den einzelnen Subtypen teilweise überlappen, wodurch eine eindeutige Zuteilung zu einem Subtyp allein anhand der CT aktuell nicht möglich ist (41, 42).

Die einzelnen Subtypen lassen sich am präzisesten anhand der Heterogenität der Kontrastmittelaufnahme unterscheiden (41, 42). Klarzellige Nierenzellkarzinome weisen oft eine heterogene Kontrastmittelaufnahme auf, wobei die hyperdensen Areale durch hypervaskularisiertes Parenchym und die hypodensen Areale durch Nekrosen oder zystische Veränderungen verursacht werden (41). Papilläre und chromophobe Nierenzellkarzinome nehmen das Kontrastmittel hingegen meist homogen auf (41). Diese Unterschiede werden in der frühvenösen Phase der CT-Aufnahme besonders deutlich und verringern sich in der spätvenösen Phase (41). Da Onkozytome jedoch ähnlich heterogen Kontrastmittel aufnehmen wie klarzellige Nierenzellkarzinome, gestaltet sich die Differenzierung zwischen diesen beiden Raumforderungen schwierig (41).

Des Weiteren finden sich nach Kontrastmittelgabe auch Unterschiede bezüglich der Strahlenabschwächung und somit der Kontrastmittelanreicherung der verschiedenen Subtypen des Nierenzellkarzinoms (42, 43). Zur Beurteilung der Kontrastmittelaффinität werden die HU-Werte einzelner Tumorareale vor und nach Gabe des Kontrastmittels gemessen, wobei eine Erhöhung von 10-20 HU als relevante Kontrastmittelanreicherung gilt (13). Je heterogener der Tumor ist, desto kleiner muss das Areal sein, in dem die Kontrastmittelanreicherung gemessen wird (13). Klarzellige Nierenzellkarzinome unterscheiden sich aufgrund ihrer ausgeprägten Kontrastmittelanreicherung, die sich aus der Hypervaskularität dieser Tumoren ergibt, von den anderen Subtypen (41, 42). Während sich chromophobe Nierenzellkarzinome etwas hypodenser darstellen und somit eine moderate Kontrastmittelanreicherung aufweisen, demonstrieren papilläre Nierenzellkarzinome im Durchschnitt aufgrund ihrer Hypovaskularität die geringste Kontrastmittelanreicherung (41). Letztere zeigen in der arteriellen Phase teilweise noch

gar keine Kontrastmittelanreicherung, sodass eine Unterscheidung zwischen Tumor und Nierenmark gegebenenfalls nicht möglich ist (13). Sie können in der frühvenösen Phase genauer erkannt werden (13).

Außerdem kann die Messung der Strahlenabschwächung in nativen CT-Aufnahmen dazu verwendet werden, um zwischen benignen und malignen Nierenraumforderungen zu unterscheiden (44). Tumoren, deren Dichtewerte ausschließlich unterhalb von 20 HU oder oberhalb von 70 HU liegen, sind in der Regel benigne (43, 44). Bei Dichtewerten zwischen 20 und 70 HU lässt sich in circa 10% der Fälle ein maligner Nierentumor nachweisen (44). Angiomyolipome zeigen bei vorhandenem Fettgehalt oftmals Dichtewerte von unter -10 HU (45). Bei atypischen Angiomyolipomen, die keinen oder einen nur sehr geringen Fettgehalt aufweisen, ist eine Abgrenzung von malignen Nierenzelltumoren anhand der Dichtewerte nicht möglich, was die Diagnose dieser Tumoren erschwert (45).

Papilläre und chromophobe Nierenzellkarzinome weisen mit 32% und 38% häufiger Kalzifizierungen auf als klarzellige Nierenzellkarzinome, bei denen sich nur in 11% der Fälle Kalzifizierungen abgrenzen lassen (42). Von dem Vorhandensein von Kalzifizierungen kann nicht abgeleitet werden, um welchen Subtyp des Nierenzellkarzinoms es sich handelt (41). Die Kalzifizierung renaler Raumforderungen ist jedoch ein starker Hinweis darauf, dass es sich tatsächlich um einen malignen Tumor handelt (41).

Unkomplizierte Nierenzysten definieren sich durch Dichtewerte in nativen CT-Aufnahmen von unter 20 HU, einer fehlenden Kontrastmittelanreicherung nach Kontrastmittelgabe, fehlenden Septierungen und einer nicht sichtbaren Zystenwand (46). Nierenzysten können anhand der Bosniak-Klassifikation in vier Kategorien eingeteilt werden, wobei mit jeder höheren Kategorie die Wahrscheinlichkeit der malignen Entartung zunimmt, ab Kategorie 4 ist sogar mit Sicherheit von dieser auszugehen (13). Die Zuteilung zu einer Kategorie erfolgt anhand der CT-Aufnahme, wobei unter anderem die Wanddicke der Zyste, die Form der Wand, die Dichte der Zyste, eine eventuell vorhandene Septierung, Kalzifizierungen und die Kontrastmittelanreicherung der Wand und Septierung als Kriterien dienen (13). Laut *Israel et Bosniak* kann die Bosniak-Klassifikation auch in MRT-Aufnahmen angewendet werden, was allerdings, unter anderem bedingt durch die bessere Darstellung von Septierungen und die stärkere Aufnahme des Gadoliniumhaltigen Kontrastmittels, tendenziell zu einer höheren Einstufung der Zysten führt (13). Während Zysten der Kategorien 1-2 als benigne gelten, sollten Zysten ab Kategorie 3

entfernt werden (13). Zysten der Kategorie 2F sollten regelmäßig kontrolliert werden, um eine maligne Entartung frühzeitig zu erkennen (13). Unkomplizierte Nierenzysten können jedoch eine durch die Bildrekonstruktion bedingte vorgetäuschte Kontrastmittelanreicherung von mindestens 10 HU aufweisen (13). Dieser Effekt, auch Pseudoenhancement genannt, wird durch eine geringe Größe der Zysten und eine starke Kontrastmittelanreicherung des umgebenden Nierenparenchyms begünstigt und muss bei der Beurteilung der Raumforderung in der CT-Aufnahme beachtet werden (47).

1.6.2 Bisherige technische Entwicklungen der computertomographischen Bildgebung

Die Computertomographie wurde 1972 von *G. N. Hounsfield*, einem britischen Elektromechaniker, zum ersten Mal durchgeführt (48). Im selben Jahr fand sie auch ihren weltweit ersten klinischen Einsatz im Atkinson Morley Hospital in London (49). Für die Erfindung der CT erhielt *Hounsfield* gemeinsam mit *A. M. Cormack*, einem südafrikanischen Physiker, der ebenfalls bedeutend zur Entwicklung der CT beigetragen hatte, 1979 den Nobelpreis für Medizin (49). Seit 1972 wurde die CT kontinuierlich weiterentwickelt, wobei zu den Hauptzielen stets eine schnellere Durchführung, eine höhere Auflösung und eine Reduktion der Strahlendosis gehörten (49).

Einen Meilenstein stellte die Entwicklung der erstmals 1989 vorgestellten Spiral-CT dar (50). Bei der Spiral-CT erfolgte nun, im Gegensatz zum vorher üblichen schrittweisen Vorschub nach jeder Aufnahme, ein kontinuierlicher Vorschub des Tisches, was zu kürzeren Untersuchungszeiten, einer besseren Bildqualität, einer zuverlässigeren Erfassung von Läsionen und einer verbesserten 3D-Bildrekonstruktion führte (49, 50). Während die Detektoren zunächst nur aus einer Zeile bestanden, führte Ende der 1990er Jahre die Entwicklung von mehrzeiligen Detektoren dazu, dass mehrere Schichten gleichzeitig aufgenommen werden konnten. Dies führte zu einer Verringerung der Schichtdicke und zu einer erheblichen Verkürzung der Untersuchungszeit (49). Mittlerweile werden im klinischen Alltag Detektoren mit bis zu 320 Detektorzeilen eingesetzt (51). Ein weiterer wichtiger Fortschritt war die Entwicklung der Dual-Source-CT, die 2006 zum ersten Mal klinisch evaluiert wurde und bei der sowohl zwei Röntgeneinheiten als auch zwei Detektoren zum Einsatz kommen (49, 52). Sie ermöglicht den gleichzeitigen Einsatz von Röntgenstrahlung zweier unterschiedlicher Energieniveaus, was als Dual-Energy-CT bezeichnet wird (53). Die Dual-Energy-CT erlaubt bei bestimmten Fragestellungen eine präzisere Diagnostik, da die HU-Werte

eines Gewebes je nach Energieniveau der Strahlendosis variieren (54). Während die Aufnahme einer einzelnen Schicht 1972 noch 300 Sekunden andauerte, ist dies heute innerhalb von 100 Millisekunden möglich (49). Die Aufnahmezeit betrug bei den ersten CT-Untersuchungen insgesamt 35 Minuten, wonach weitere 9 Tage folgten, bis der Computer die Verarbeitung der CT-Aufnahmen abgeschlossen hatte (49). Heute sind beide Prozesse innerhalb weniger Sekunden möglich (49).

Im Jahr 2009 wurden in Deutschland insgesamt 9,4 Millionen CT-Untersuchungen durchgeführt, was je 100 Einwohnern 11,4 CT-Untersuchungen entspricht (55). CT-Untersuchungen sind in Deutschland für mehr als 60% der medizinisch verursachten Strahlendosis verantwortlich (56). Folglich wird vor allem seit Beginn des Jahrtausends ein Zusammenhang zwischen der Strahlenexposition bei computertomographischen Untersuchungen und eines dadurch möglicherweise erhöhten Krebsrisikos kontrovers diskutiert (57). Eine australische Studie aus dem Jahr 2012 kommt beispielsweise zum Ergebnis, dass CT-Untersuchungen von Kindern mit einer kumulierten Dosis von 50 mGy das Leukämierisiko verdreifachen (58). Zur Reduktion der Strahlendosis wurde unter anderem Ende der 1990er Jahre die von der Gewebisdichte abhängige automatische Modulation des Röhrenstroms in CT-Geräten eingeführt, die bei konstanter Bildqualität eine um 68% verringerte Strahlendosis erlaubt (56, 57). Des Weiteren kommen seit 2009 in CT-Geräten sogenannte iterative Rekonstruktionsalgorithmen zum Einsatz (59). Hierdurch lassen sich mit reduzierter Strahlendosis aufgenommene CT-Aufnahmen nachträglich so bearbeiten, dass sowohl das Bildrauschen als auch die Artefakte verringert werden, was eine Verringerung der Strahlendosis um bis zu 75% ermöglicht (57, 59).

1.7 Künstliche Intelligenz in der Radiologie

Künstliche Intelligenz beschreibt die Wissenschaft der Entwicklung von Maschinen, die in der Lage sind, Probleme eigenständig durch den Einsatz von Intelligenz zu lösen (60). Hierbei versuchen sie teilweise, Lösungsansätze zu verwenden, die den Lösungsansätzen menschlicher Intelligenz ähneln (60). Es existiert in der Literatur jedoch keine einheitliche Definition für den Begriff Künstliche Intelligenz (61). Deren Erforschung begann kurz nach dem Ende des zweiten Weltkriegs (60). Alan Turing, ein britischer Mathematiker, gilt als einer der Pioniere auf dem Gebiet und hielt bereits 1947 die wahrscheinlich erste offizielle Vorlesung zum Thema Künstliche Intelligenz (60). Seine

Definition von Künstlicher Intelligenz, der sogenannte Turing-Test, beschreibt eine Maschine als künstlich intelligent, wenn eine Person sie in einer Konversation, in der die Person nicht sehen kann, ob sie mit einem Mensch oder einer Maschine spricht, nicht von einem echten Menschen unterscheiden kann (62). In den letzten Jahren gelangen zahlreiche Durchbrüche auf dem Gebiet der Künstlichen Intelligenz, beispielsweise bei der Spracherkennung, der Klassifizierung von Gegenständen, dem autonomen Fahren oder gar dem Erlernen von komplexen Brettspielen (63).

1.7.1 Maschinelles Lernen

Maschinelles Lernen beschreibt das Lernen eines Modells anhand der Analyse eines Datensatzes durch einen Computer, der dann aus dem Erlernten Vorhersagen für ähnliche, aber bisher nicht analysierte Daten treffen kann (64). Es stellt dementsprechend eine Form der Künstlichen Intelligenz dar (64). Durch maschinelles Lernen können Datensätze, deren Analyse aufgrund ihrer Komplexität und Größe nicht oder nur eingeschränkt durch Menschen erfolgen kann, auf Muster untersucht werden (64). Hierdurch erhofft sich die Wissenschaft bei den verschiedensten Problemen neue Erkenntnisse, zu denen man ohne Zuhilfenahme des maschinellen Lernens nicht gelangen könnte, und damit einhergehend auch die Entwicklung neuer Lösungsansätze für eben solche Probleme (64). Allgemein wird zwischen überwachtem Lernen und nicht überwachtem Lernen differenziert (64). Beim überwachten Lernen lernt der Computer mit Datensätzen, die bereits durch einen Menschen klassifiziert wurden, also beispielsweise mit Bildern, die bereits korrekt einer Klasse zugeteilt wurden, und wendet das Erlernte dann bei noch nicht klassifizierten Datensätzen an (64). Beim nicht überwachten Lernen lernt der Computer ohne die Verwendung einer bereits erfolgten Klassifizierung des Datensatzes und findet somit selbst einen Weg, die jeweiligen Daten anhand bestimmter Merkmale einzuteilen (64).

Der Einsatz von Künstlicher Intelligenz und maschinellem Lernen wird auch in der Medizin aktuell breit erforscht (65). Beispielsweise existieren bereits zahlreiche Studien, die die Vorteile des Einsatzes von maschinellem Lernen bei der Diagnostik und Prognoseabschätzung von Tumorpatienten oder bei der Bestimmung der bestmöglichen Therapie bei verschiedenen Erkrankungen demonstrieren konnten (65). In der Radiologie kann maschinelles Lernen zum Beispiel bei der Erkennung und Klassifizierung von Tumoren oder der Extraktion von quantitativen Bildmerkmalen aus Bilddatensätzen verwendet werden (66). Ein Teilgebiet dieser Forschung stellt die Diagnostik von

Nierenraumforderungen dar (67). Hierbei wird unter anderem untersucht, inwiefern maschinelles Lernen bei der Differenzierung zwischen verschiedenen Raumforderungen der Niere, der Prognose von Nierenzellkarzinomen und deren nicht-invasiver histopathologischer Graduierung hilfreich sein kann (67).

1.8 Radiomics

Unter Radiomics versteht man eine Subdisziplin der Radiologie, die sich mit der Extraktion und Analyse quantitativer Daten aus CT-, MRT-, Positron-Emissions-Tomographie (PET)-, Röntgen- und Ultraschallaufnahmen beschäftigt (68). Diese extrahierten Merkmale beschreiben unter anderem die Größe, die Form und die Graustufen einer Struktur, wie beispielsweise eines Tumors oder eines Organs (68). Im Gegensatz zur ansonsten vor allem visuellen, subjektiven Interpretation radiologischer Bilddatensätze stellen die quantitativen Bildmerkmale eine objektive Beurteilungsgrundlage dar. Die Bildmerkmale beispielsweise eines Tumors können zu weiteren Daten aus klinischen Berichten, genetischen Untersuchungen, Laboruntersuchungen und histologischen Untersuchungen in Beziehung gesetzt werden (69). Die Idee hinter Radiomics ist also, dass sich beispielsweise Veränderungen auf mikroskopischer Ebene oder Veränderungen in der Genexpression teilweise auch in strukturellen, quantitativ messbaren Bilddatenunterschieden manifestieren und diese Veränderungen somit in radiologischen Bilddatensätzen erkannt werden können, sodass gegebenenfalls auf weitere, invasive diagnostische Maßnahmen verzichtet werden kann (39, 69).

Bereits ein radiologischer Bilddatensatz eines einzelnen Patienten enthält mehrere Millionen zu analysierende Voxel und erlaubt die Extrahierung hunderter verschiedener Bildmerkmale (39, 69). Da sich nahezu jede krebserkrankte Person mehreren diagnostische Bildgebungen unterzieht, wird hierdurch eine große Anzahl an Datensätzen erzeugt (69). Die zahlreichen Bilddatensätze der Patienten können daher, insofern sie in vergleichbarer Art und Weise aufgenommen wurden, in Datenbanken gesammelt und analysiert werden. In Zukunft könnten hierdurch voraussichtlich zahlreiche neue Zusammenhänge und somit neue Biomarker entdeckt werden (69). Die Extraktion der quantitativen Bildmerkmale eines Tumors kann unter anderem zur Abschätzung der Prognose, zur Verlaufskontrolle oder auch zur Beurteilung eines eventuellen Therapieerfolgs eingesetzt werden (69).

Radiologische Aufnahmen werden im klinischen Alltag bislang vor allem visuell beurteilt. Hierbei werden auch quantitativ messbare Bildmerkmale beschreiben, wie beispielsweise die Heterogenität eines Tumors. Diese Art der Befundung unterliegt jedoch einer hohen inter- und intraindividuellen Variabilität (68). Radiomics-basierte Analysen eröffnen neue Möglichkeiten, um aus den Bilddatensätzen präzisere und weitere diagnostisch relevante Informationen zu gewinnen.

Zahlreiche aktuelle Studien beschäftigten sich mit Radiomics-basierten Analysen radiologischer Datensätze. So konnten *Wibmer et al.* beispielsweise zeigen, dass durch die Analyse quantitativer Merkmale in MRT-Aufnahmen nicht nur zwischen normalem Prostatagewebe und Prostatakarzinomgewebe unterschieden werden kann, sondern auch eine Aussage über den Gleason-Grad, also den Grad der Entdifferenzierung des Karzinomgewebes, getroffen werden kann (70). Eine weitere Studie von *Coroller et al.* konnte demonstrieren, dass man in CT-Aufnahmen durch die Analyse der quantitativen Bildmerkmale eines Lungenkarzinoms abschätzen kann, ob Fernmetastasen vorliegen (71). Ferner konnten *Liang et al.* in ihrer Studie demonstrieren, dass durch die Verwendung von Radiomics zwischen hochgradigen (Grad 3 + 4) und niedriggradigen (Grad 1 + 2) kolorektalen Karzinomen differenziert werden kann (72). Dies könnte die präoperative, nicht-invasive Prognose ermöglichen und somit auch eine differenziertere Auswahl der geeigneten Therapieform erlauben (72).

Es sind mehrere Schritte notwendig, um eine Radiomics-basierte Analyse einer Läsion durchzuführen (39). Diese Schritte umfassen die Akquisition der Bilddatensätze, die Segmentierung der Läsion, die Extraktion der quantitativen Bildmerkmale, die Selektion der relevantesten Bildmerkmale und abschließend das Trainieren eines auf maschinellem Lernen basierenden Modells (39). Im Folgenden sollen die einzelnen Schritte genauer erläutert werden.

1.8.1 Akquisition der Bilddatensätze

Zunächst gilt es, die geeignete Bildgebungsmodalität für die zu untersuchende Struktur auszuwählen (39). Die quantitative Analyse der Bilddaten gestaltet sich einfacher, wenn keine oder nur sehr wenige Artefakte vorhanden sind und die Aufnahmen verschiedener Patienten mit ähnlichen beziehungsweise den gleichen Scangeräten und Protokollen durchgeführt wurden (39). Auch eine zeitlich vergleichbare Applikation einer ähnlichen Menge an Kontrastmittel erleichtert die Analyse der Bildmerkmale (39). Des Weiteren kann die Analyse einer Läsion zu unterschiedlichen Zeitpunkten oder in unterschiedlichen

Stadien eine Fehlerquelle darstellen, da hierdurch eine zu starke Heterogenität der zu vergleichenden Bilddaten bedingt wird (39). Insgesamt sind für eine Radiomics-basierte Bilddatenanalyse also standardisierte Aufnahmen von hoher Qualität von Vorteil, was technisch nur durch moderne Scan-Geräte möglich wird - dies erklärt unter anderem, warum der Begriff Radiomics erst 2012 das erste Mal in einer wissenschaftlichen Veröffentlichung Verwendung fand (73).

Der Akquisition der Bilddatensätze schließt sich deren Normalisierung an. Computertomographische Aufnahmen müssen beispielsweise entsprechend der im DICOM-Header enthaltenen Angaben bezüglich des Rescale Intercepts und des Rescale Slopes linear transformiert werden, um die jeweiligen Hounsfield-Einheiten der einzelnen Pixel zu erhalten und somit den Bilddatensatz für die Radiomics-basierte Analyse vorzubereiten.

1.8.2 Segmentierung

In dem ausgewählten Bilddatensatz wird dann die Region identifiziert, die quantitativ analysiert werden soll (39). Diese Region, beispielsweise ein Tumor oder eine andere Läsion, wird auch als Region of Interest (ROI) oder Volume of Interest (VOI) bezeichnet (39). Daraufhin erfolgt die Markierung, also Segmentierung, der ROI (39). Hierbei werden mithilfe einer speziellen Software die Grenzen der ROI exakt umrandet, um eine klare Abgrenzung zu umgebenden Strukturen zu gewährleisten und den zu analysierenden Bereich präzise zu definieren (39). Dies wird in jeder einzelnen Schicht der Aufnahme wiederholt, was schließlich eine dreidimensionale Darstellung der segmentierten Struktur erlaubt (39).

Die Segmentierung kann entweder manuell oder auch automatisiert mithilfe von Algorithmen erfolgen, wobei zu beachten ist, dass nicht für alle Tumoren bereits Algorithmen zur automatisierten Segmentierung Verfügung stehen und die manuelle Ausführung der Segmentierung einer ROI je nach Untersucher variiert (39, 74, 75). Diese Variabilität bei der manuellen Durchführung zeigt sich sogar bei der erneuten Segmentierung einer Struktur durch dieselbe Person (39). Außerdem ist die manuelle Segmentierung von Läsionen mit einem hohen zeitlichen Aufwand verbunden, was deren Anwendbarkeit im klinischen Alltag einschränkt (39). Darüber hinaus wird die automatisierte Segmentierung, die sich vor allem bei großen Bilddatensätzen anbietet, durch bestimmte Eigenschaften wie einer nicht ausreichenden Schärfe der Grenzen der ROI oder vorliegenden Artefakten erschwert (39). Als Kompromiss aus der manuellen

und automatisierten Segmentierung einer ROI bietet sich auch eine semi-automatisierte Segmentierung an, bei der die ROI zunächst automatisiert segmentiert wird und diese Segmentierung anschließend manuell korrigiert wird (39).

1.8.3 Extraktion der quantitativen Bildmerkmale

Der nächste Schritt umfasst die Extraktion quantitativer Merkmale (39). Diese erfolgt durch verschiedene Algorithmen und umfasst mehrere hundert bis tausend numerische Variablen (39, 76). Die extrahierten Merkmale können in die Kategorien Formmerkmale, Merkmale erster Ordnung, Merkmale zweiter Ordnung und Merkmale höherer Ordnung eingeteilt werden (76). Formmerkmale werden auch als sogenannte semantische Merkmale bezeichnet, da sie in der Radiologie gebräuchliche Parameter darstellen. Sie umfassen beispielsweise den längsten gemessenen Durchmesser der ROI, deren Volumen, deren Form oder deren Lokalisation (69, 76). Die Merkmale höherer Ordnungen werden auch agnostische Merkmale genannt (69). Sie umfassen Parameter, die in der Radiologie im klinischen Alltag bisher nicht gebräuchlich sind (69).

Quantitative Bildmerkmale erster Ordnung beschreiben die Häufigkeit der einzelnen Intensitäten der Bildpunkte in einer ROI (69, 76). Diese Werte, auch Grauwerte genannt, werden hierbei nicht in räumliche Beziehung zueinander gesetzt (69, 76). Aus der Häufigkeitsverteilung der Grauwerte der einzelnen Pixel lassen sich statistische Parameter wie beispielsweise der maximale Intensitätswert, der minimale Intensitätswert, der Mittelwert, der Modus, die Entropie, die Kurtosis und die Schiefe berechnen (69, 76).

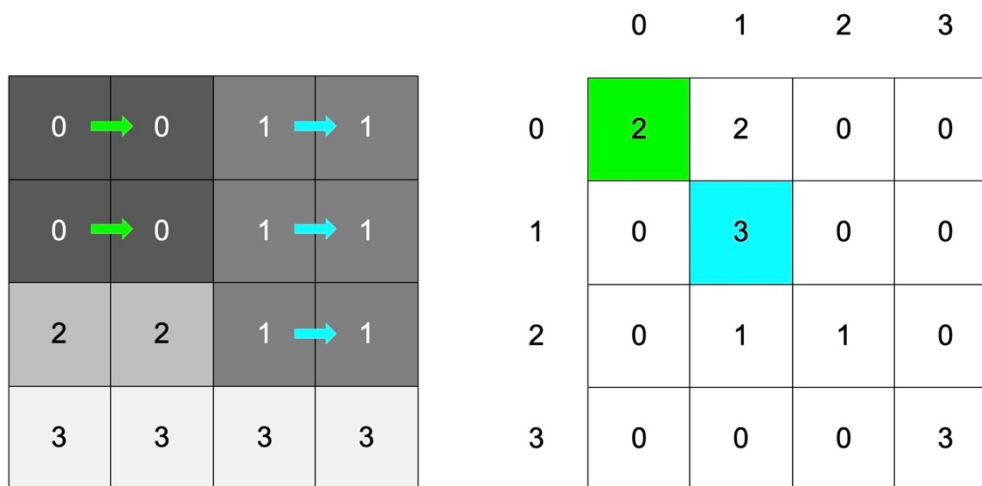


Abbildung 1: Exemplarische Darstellung einer Grey-Level Co-Occurrence-Matrix.

Links eine Beispieldarstellung der Grauwerte von 16 zufällig ausgewählten Bildpunkten, rechts die zugehörige Grey-Level Co-Occurrence Matrix. Die grünen Pfeile kennzeichnen, dass in einer Bildrichtung, die hierbei durch die Pfeile angezeigt wird, dreimal auf einen Grauwert 0 ebenfalls ein Grauwert 0 folgt, was eine Eintragung der grün markierten Zahl 2 in der Matrix zur Folge hat. Analog dazu erfolgt die Eintragung der blau markierten Zahl 3 der Matrix, um zu kennzeichnen, dass in der mit blauen Pfeilen markierten Bildrichtung dreimal auf einen Grauwert 1 ein weiterer Grauwert 1 folgt.

Merkmale zweiter Ordnung werden auch als Texturmerkmale bezeichnet (69, 76). Bei Texturmerkmalen werden nun die Grauwerte angrenzender Bildpunkte zueinander in räumliche Beziehung gesetzt, wodurch die Heterogenität einer ROI erfasst werden kann (69, 76). Für die Berechnung der Texturmerkmale werden auf Basis der Grauwerte Matrizen erstellt, die die Relationen der Grauwerte auf verschiedene Weise abbilden (69, 76). Zum besseren Verständnis ist in Abbildung 1 exemplarisch die sogenannte Grey-Level Co-Occurrence Matrix von 16 zufällig definierten Bildpunkten mit vier unterschiedlichen Grauwerten dargestellt. In der Grey-Level Co-Occurrence Matrix wird dargestellt, wie häufig zwei bestimmte Grauwerte in einer Bildrichtung in derselben Konstellation auftreten, woraus schließlich verschiedene quantitative Merkmale zweiter Ordnung wie unter anderem die Entropie der Matrix (Entropy) oder die Homogenität der Matrix (Homogeneity) berechnet werden können (77).

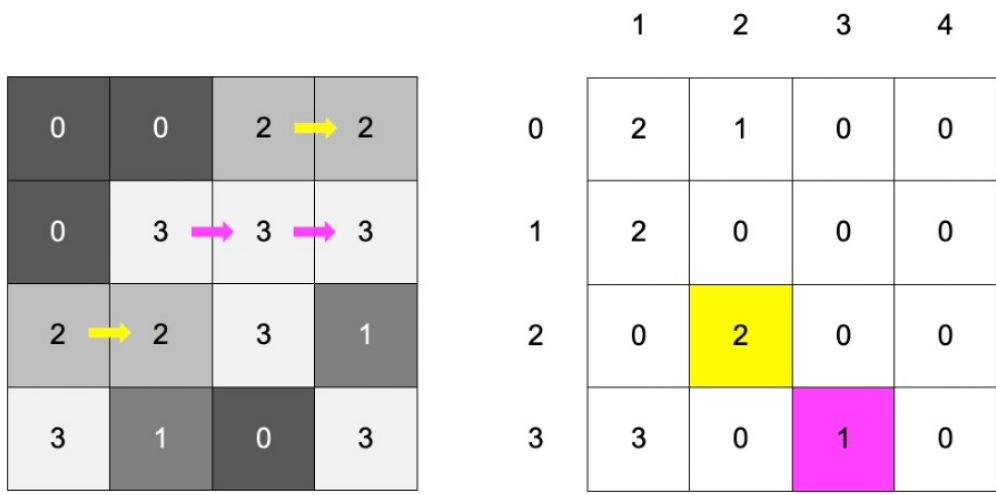


Abbildung 2: Exemplarische Darstellung einer Grauwerte-Lauflängenmatrix.

Links eine Beispieldarstellung der Grauwerte von 16 zufällig ausgewählten Bildpunkten, rechts die zugehörige Grauwerte-Lauflängenmatrix. Die gelben Pfeile zeigen, dass in der durch die Pfeile angezeigten Bildrichtung zweimal ein Lauf der Länge 2 des Grauwerts 2 vorliegt, was die Eintragung der gelb markierten Zahl 2 im passenden Feld der Matrix zur Folge hat. Die lila markierten Pfeile demonstrieren einen Lauf der Länge 3 des Grauwerts 3, weshalb im dazugehörigen Feld der Matrix die lila markierte Zahl 1 eingetragen wird.

Eine weitere häufig verwendete Matrix ist die Grauwerte-Lauflängenmatrix (Grey-Level Run-Length Matrix), in der abgebildet wird, wie häufig derselbe Intensitätswert in einer Bildrichtung auftritt, bis diese Abfolge durch einen abweichenden Grauwert unterbrochen wird (77). Hieraus können Rückschlüsse auf die Grob- oder Feinkörnigkeit der ROI gezogen werden, wobei grobkörnige Strukturen durch längere Abfolgen derselben Grauwerte, feinkörnige Strukturen durch kürzere Abfolgen derselben Grauwerte gekennzeichnet sind (77). In Abbildung 2 ist ein Beispiel einer Grauwerte-Lauflängenmatrix von 16 per Zufall ausgewählten Bildpunkten dargestellt. Die verschiedenen Matrizes können in mehreren Bildrichtungen erstellt werden, wodurch die Heterogenität der gesamten ROI analysiert werden kann (76).

Die Berechnung von quantitativen Merkmalen höherer Ordnung erfolgt durch die Anwendung von Filtern, mit Hilfe derer aus den Originalbildern transformierte Bildversionen erstellt werden (76). Hierfür häufig verwendete Filter stellen beispielsweise sogenannte Wavelet-Filter oder Gabor-Filter dar (78). Aus den transformierten Bildern können dann die quantitativen Bildmerkmale abgeleitet werden können (76).

1.8.4 Selektion und Vorverarbeitung der quantitativen Bildmerkmale

Der Merkmalsextraktion schließt sich die Merkmalsselektion an. Hier gilt es unter anderem, die Bildmerkmale so zu filtern, dass schließlich nur noch die für die Beantwortung der Fragestellung relevanten zu analysierenden Merkmale verbleiben (39, 79). Dies kann folglich die Performance des Modells des maschinellen Lernens erhöhen (79). Um die quantitative Bildmerkmale, die eine besonders starke Assoziation zu den Zielparametern aufweisen, auszuwählen, stehen verschiedene statistische Methoden wie beispielsweise die ANOVA zur Verfügung (80).

Zahlreiche quantitative Bildmerkmale korrelieren in der Regel untereinander, was ohne eine Merkmalsreduktion zu einer übermäßigen Gewichtung einzelner Merkmale führen würde (80). Dies trifft beispielsweise oftmals auf Merkmale zweiter Ordnung, die aus derselben Matrix abgeleitet wurden, zu (80). Eine Möglichkeit der Merkmalsreduktion ist es folglich, durch eine Korrelationsanalyse stark korrelierende Bildmerkmale zu identifizieren und auszuschließen.

Darüber hinaus können Bildmerkmale anhand ihrer Reproduzierbarkeit selektiert werden, wobei verschiedene Methoden zur Verfügung stehen (80). Die Wiederholung der computertomographischen Untersuchung mit erneuter anschließender Merkmalsextraktion stellt eine Möglichkeit dar (80). Hierbei können Bildmerkmale identifiziert und selektiert werden, deren Werte bei wiederholter Aufnahme möglichst übereinstimmen (80). Des Weiteren können Läsionen von verschiedenen Untersuchern segmentiert werden, wobei auch hier die quantitativen Merkmale selektiert werden, deren Werte unabhängig von der Segmentierung stabil bleiben (80). Weitere Möglichkeiten, um die Reproduzierbarkeit von quantitativen Bildmerkmalen zu überprüfen, stellen beispielsweise die Verringerung der Schärfe der einzelnen Bilder oder das geringfügige Verschieben der Segmentierungsmaske dar.

Im Rahmen der Vorverarbeitung der Daten sollten die selektierten quantitativen Merkmale vor dem Training der statistischen Modelle normalisiert und skaliert werden (79).

1.8.5 Trainieren eines Modells des maschinellen Lernens

Im nächsten Schritt werden die selektierten quantitativen Bildmerkmale durch Modelle des maschinellen Lernens auf eine zuvor definierte Zielvariable bezogen (39). Hierfür stehen verschiedene statistische Methoden des maschinellen Lernens zur Verfügung,

darunter beispielsweise Random Forest Modelle oder die k-Nearest Neighbor Modelle (76). Die Auswahl des Verfahrens hängt unter anderem von der gesuchten Zielvariable, der Anzahl der quantitativen Bildmerkmale und dem Stichprobenumfang ab (76).

Zunächst wird der segmentierte Datensatz in zwei separate Untergruppen eingeteilt, die als Trainingsdatensatz und Testdatensatz bezeichnet werden (79). Um das Modell des maschinellen Lernens anhand des Trainingsdatensatzes zu trainieren, werden für jede ROI des Trainingsdatensatzes die selektierten quantitativen Bildmerkmale und die bereits definierten Zielparameter benötigt (76, 79). Anhand der Analyse der Trainingsdaten soll das Modell des maschinellen Lernens lernen, welchem Zielparameter ein Objekt bei einer bestimmten Konstellation von quantitativen Bildmerkmalen zugeordnet werden kann (76, 79). Hierbei kann das Modell durch Anpassung der Hyperparameter, beispielsweise im Rahmen einer Kreuzvalidierung, optimiert werden, um die Genauigkeit der Klassifizierung zu verbessern (76, 79). Es wird also ein Teil der Daten, die das Modell nicht zum Training verwendet, sondern auf die es zum Zwecke der Modelloptimierung überprüft wird, vom Trainingsdatensatz abgespalten; auch: Validierungsdatensatz. Im Rahmen einer mehrfachen Kreuzvalidierung kann der Trainingsdatensatz in eine beliebige Anzahl an Untergruppen aufgeteilt werden und das Trainieren des Modells des maschinellen Lernens wiederholt werden, sodass jede Untergruppe einmal den Validierungsdatensatz darstellt, wodurch einem Overfitting entgegengewirkt werden kann (79, 80). Overfitting meint, dass ein Modell des maschinellen Lernens zwar sehr gute Voraussagen über den Datensatz treffen kann, auf dessen Basis es trainiert wurde, dies sich aber nicht auf neue, unbekannte Bilddatensätze übertragen lässt (79). Die Kreuzvalidierung kann mehrfach wiederholt werden, wobei jeweils unterschiedlich zusammengesetzte Gruppen erstellt werden. Durch die abschließende Analyse des bisher ungesehenen Testdatensatzes kann beurteilt werden, inwiefern das Modell des maschinellen Lernens auch bei unbekanntem Bilddaten eine Vorhersage über die Zielvariable treffen kann (76). Am Ende der Radiomics-basierten Bildanalyse soll folglich ein Modell stehen, das durch die Analyse der quantitativen Bildmerkmale einer ROI beziehungsweise eines VOI einen Zielparameter, wie beispielsweise einen Tumorsubtyp oder einen histopathologischen Tumorgrad, richtig klassifizieren und somit vorhersagen kann (76). Die Durchführung aller Einzelschritte einer Radiomics-basierten Bildanalyse, einschließlich der Vorhersage eines Zielparameters, wird auch als Radiomics-Pipeline bezeichnet (76).

1.8.6 Erhebung statistischer Maße zur Evaluation eines Modells des maschinellen Lernens

Um die Performance eines Modells des maschinellen Lernens bewerten zu können und somit auch unterschiedliche Modelle miteinander vergleichen zu können, können verschiedene statistische Maße wie beispielsweise die Accuracy, der Kappa-Wert oder die Fläche unter der ROC-Kurve (Receiver Operator Characteristics Curve), auch genannt AUC (Area under the ROC-Curve), berechnet werden (76, 81, 82). Da diese statistischen Maße vor allem im Bereich des maschinellen Lernens genutzt werden und in der medizinischen Literatur nicht zwangsläufig häufig Verwendung finden, soll im Folgenden ein kurzer Überblick über deren Berechnung und Verwendung gegeben werden.

Die Accuracy beschreibt, in wie viel Prozent der Fälle ein statistisches Modell des maschinellen Lernens ein Objekt richtig zuordnet (Berechnung:

$$\frac{\text{Richtig positive Fälle} + \text{Richtig negative Fälle}}{\text{Richtig positive Fälle} + \text{Falsch positive Fälle} + \text{Richtig negative Fälle} + \text{Falsch negative Fälle}} \quad (83).$$

Eine hohe Accuracy geht mit der hohen Performance eines Modells einher. Bei aus unterschiedlich großen Gruppen bestehenden Datensätzen kann anstelle der Accuracy auch die Balanced Accuracy berechnet werden (84). Die Balanced Accuracy bezieht sich im Gegensatz zur Accuracy jeweils auf die Relation der richtig positiven Fälle zu allen positiven Fällen und der richtig negativen Fälle zu allen negativen Fällen (Berechnung: $\frac{\text{Sensitivität} + \text{Spezifität}}{2}$), wodurch bei ungleichen Datensätzen eine präzisere Angabe bezüglich der Performance eines Modells des maschinellen Lernens getroffen werden kann (84).

Des Weiteren kann der sogenannte Kappa-Wert berechnet werden. Er ist ein Maß dafür, wie wahrscheinlich es ist, dass die korrekte Klassifizierung eines Objekts durch ein Modell des maschinellen Lernens nicht zufällig zustande gekommen ist (Berechnung:

$$\kappa = \frac{p_0 - p_e}{1 - p_e},$$

wobei p_0 der Accuracy und p_e der Wahrscheinlichkeit, dass die Übereinstimmung der Klassifizierung zufällig entstanden ist, entspricht) (82). Die Berechnung des Kappa-Werts ist ebenso wie die Berechnung der Balanced Accuracy vor allem bei Datensätzen mit deutlich abweichender Gruppengröße sinnvoll, da die Accuracy eines Modells des maschinellen Lernens hierbei zufallsbedingt sehr hoch sein und somit gegebenenfalls fälschlicherweise den Eindruck einer guten Performance des Modells entstehen lassen kann.

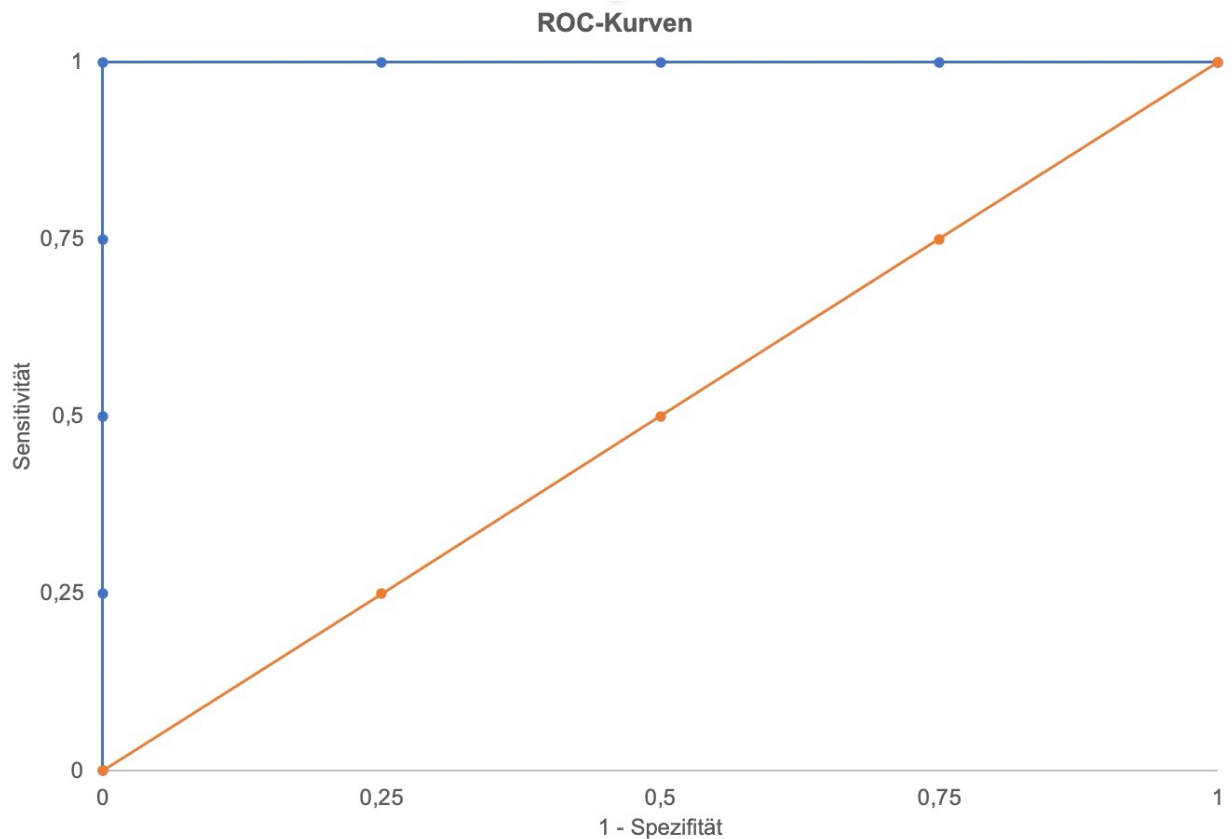


Abbildung 3: Exemplarische Darstellung zweier ROC-Kurven.

Die Abbildung zeigt exemplarisch zwei ROC-Kurven. Die orangefarbene ROC-Kurve, eine Diagonale vom Punkt (0,0) zum Punkt (1,1) des Koordinatensystems, zeigt beispielsweise, dass der Anteil an richtig positiv identifizierten Fällen bei jedem Datenpunkt, also jedem Schwellenwert, dem Anteil an falsch positiv identifizierten Fällen entspricht, was für ein zufälliges Zustandekommen der Entscheidung des statistischen Modells spricht. Im Gegensatz dazu spricht die blaue ROC-Kurve für eine optimale Trennschärfe eines statistischen Modells des maschinellen Lernens, wobei dieses am Punkt (1,0) mit 100% Sensitivität und 100% Spezifität klassifiziert.

Eine ROC-Kurve setzt die Rate an richtig positiven Fällen (= *Sensitivität*) und die Rate an falsch positiven Fällen (= $1 - \text{Spezifität}$) zueinander in Relation (81). Jeder einzelne Datenpunkt auf einer ROC-Kurve stellt ein bestimmtes Paar aus *Sensitivität* und $1 - \text{Spezifität}$ dar und repräsentiert somit jeweils einen Schwellenwert, ab dem ein Modell des maschinellen Lernens eine Entscheidung bezüglich der Klassifizierung eines Objekts trifft (81). In der Abbildung 3 sind exemplarisch zwei ROC-Kurven dargestellt. Die AUC kann Werte zwischen 0 und 1 annehmen, wobei ein hoher Wert für die hohe Performance eines Modells des maschinellen Lernens spricht (81).

1.9 Fragestellung

Aktuell wird der histopathologische Tumorgrad von klarzelligen Nierenzellkarzinomen im klinischen Alltag postoperativ anhand der histologischen Untersuchung des entnommenen Nierengewebes bestimmt. Seltener erfolgt die Graduierung mittels Stanzzyylinderbiopsie, wobei zu beachten ist, dass aufgrund der Untersuchung von nur wenigen Lokalisationen im Tumor gegebenenfalls ein abweichender Tumorgrad festgestellt wird (39, 67). Es lässt sich also feststellen, dass bislang ausschließlich invasive Methoden zur Graduierung klarzelliger Nierenzellkarzinomen zur Verfügung stehen, die dementsprechend das Risiko von Komplikationen bergen (38, 85). An dieser Stelle seien sowohl typische postoperative Komplikationen wie Blutungen oder Wundinfektionen nach (Teil)nephrektomie sowie auch eine geringe Gefahr der Verschleppung von Tumorzellen bei der Stanzzyylinderbiopsie erwähnt (38, 85).

Sowohl der Tumorgrad nach Fuhrman als auch der WHO/ISUP-Tumorgrad besitzen eine signifikante prognostische Aussagefähigkeit bezüglich der krankheitsspezifischen Mortalität von Patienten mit klarzelligen Nierenzellkarzinomen (22, 23, 28). Da bei niedriggradigen Nierenzellkarzinomen ein nur geringeres Metastasierungsrisiko vorliegt, bietet sich in einigen Fällen statt der operativen Tumorentfernung eine aktive Überwachung des Tumors an (86). Es ist folglich erstrebenswert, nicht-invasive diagnostische Methoden zur Bestimmung des Tumorgrads klarzelliger Nierenzellkarzinome zu finden. Die Radiomics-basierte Bestimmung des histopathologischen Tumorgrads könnte eine solche Methode darstellen. Dies wurde bereits in verschiedenen Studien untersucht. Allerdings zeigten diese einen sehr heterogenen Aufbau und eine nahezu ausnahmslos sehr geringe Stichprobengröße (87-96).

Da die Durchführung einer Computertomographie bei Patienten mit klarzelligen Nierenzellkarzinomen ohnehin ein zentraler Bestandteil der Diagnostik ist, wären für die klinische Anwendung einer Radiomics-basierten Tumorgradbestimmung keine zusätzlichen Untersuchungen notwendig (10). Somit könnten Patienten in Zukunft bereits präoperativ eine präzisere Prognose hinsichtlich des krankheitsspezifischen mittleren Überlebens erhalten. Bei niedriggradigen Nierenzellkarzinomen könnte gegebenenfalls sogar anstelle der Operation eine aktive Überwachung erfolgen.

Die vorliegende Arbeit untersucht dementsprechend die folgende zentrale Fragestellung:

Können durch eine Radiomics-basierte Analyse mehrphasiger CT-Datensätze Modelle des maschinellen Lernens trainiert werden, mithilfe derer der histopathologische Tumorgrad nach WHO/ISUP beziehungsweise nach Fuhrman vorhergesagt werden kann und somit eine nicht-invasive Graduierung der Tumoren ermöglicht würde?

2 Methodik

2.1 Auswahl der Studienpopulation

In der vorliegenden retrospektiven Studie wurden die CT-Daten von Patienten mit klarzelligem Nierenzellkarzinomen analysiert, die zwischen dem 01.01.2009 und dem 30.05.2020 in der Charité Universitätsmedizin Berlin dokumentiert wurden. Hierfür wurde eine Stichwortsuche in den Befundtexten der Charité Universitätsmedizin Berlin mit den Suchbegriffen „Nierenzellkarzinom“, „Nierentumor“, „Nierenkarzinom“, „Karzinom der Niere“, „Nierenmalignom“, „Malignom der Niere“, „Nierenraumforderung“, „Raumforderung der Niere“, „Tumor der Niere“, „RF der Niere“, „Nieren RF“, „NZK“ und „NCC“ durchgeführt, bei der insgesamt 6676 Patienten identifiziert wurden. Es erfolgte anschließend die Analyse von 3338 zufällig ausgewählten Patienten. Anhand dieser Analyse konnten 382 Patienten mit 388 klarzelligem Nierenzellkarzinomen ausgewählt werden, die die vorab festgelegten Einschlusskriterien erfüllten:

1. Es lag mindestens ein histopathologisch nachgewiesenes klarzelliges Nierenzellkarzinom vor.
2. Es lag die initiale computertomographische Untersuchung des klarzelligem Nierenzellkarzinoms vor. Hierbei musste mindestens eine CT-Aufnahme in der nativen, der arteriellen oder der venösen Phase vorliegen.
3. Es lag ein histopathologischer Befund des klarzelligem Nierenzellkarzinoms vor, in dem eindeutig, entweder nach (Teil)nephrektomie oder nach Biopsie, der Tumorgrad nach Fuhrman oder nach WHO/ISUP angegeben war. Die Operation oder Biopsie, anhand der der Tumor histopathologisch untersucht wurde, musste innerhalb von maximal 12 Monaten nach der initialen CT-Aufnahme des Primärtumors stattgefunden haben.

Alle Patienten ohne ein histopathologisch nachgewiesenes klarzelliges Nierenzellkarzinom wurden ausgeschlossen. Patienten wurden nicht inkludiert, wenn keine initiale computertomographische Untersuchung des klarzelligem Nierenzellkarzinoms vorlag. Eine fehlende abdominelle CT-Aufnahme führte ebenfalls zum Ausschluss aus der Studie. Des Weiteren wurden alle Patienten exkludiert, die keine histopathologische Graduierung des Primärtumors besaßen. Ein ausschließlich

vorliegender histopathologischer Befund einer Metastase des Primärtumors führte ebenfalls zum Ausschluss der Patienten. Darüber hinaus wurden keine Patienten inkludiert, deren Operation oder Biopsie später als ein Jahr nach der computertomographischen Aufnahme des Primärtumors erfolgte.

2.2 Erhebung der Patientendaten

Die Erhebung der Patientendaten erfolgte entsprechend den Vorgaben der Ethikkommission pseudonymisiert, indem allen Patienten randomisiert eine Nummer, ein sogenannter Patienten-Key, zugewiesen wurde. Zunächst wurden die DICOM-Dateien der CT-Aufnahmen der Patienten exportiert und auf einem Netzlaufwerk gespeichert, wobei je nach Verfügbarkeit eine native, eine arterielle und entweder eine früh- oder eine spätvenöse Phase ausgewählt wurden. Im Falle des Vorliegens mehrerer venöser Phasen wurde die frühvenöse Phase verwendet. Für die Erhebung weiterer Patientendaten wurde eine Datenbank angelegt, in der weitere, für die Analyse relevante Patientendaten eingetragen wurden. Es wurden folgende Patientendaten dokumentiert:

1. Das Geschlecht der Patienten
2. Der histopathologische Tumorgrad des klarzelligen Nierenzellkarzinoms
3. Die Anzahl der vorliegenden klarzelligen Nierenzellkarzinome
4. Das Vorliegen weiterer Tumore
5. Das Datum der initialen CT-Untersuchung des Nierenzellkarzinoms
6. Ob eine (partielle) Nephrektomie oder eine Biopsie durchgeführt wurde
7. Das Datum der (partiellen) Nephrektomie oder Biopsie
8. Das Alter der Patienten zum Zeitpunkt der (partiellen) Nephrektomie oder Biopsie
9. Der längste Durchmesser des klarzelligen Nierenzellkarzinoms
10. Die Lage des klarzelligen Nierenzellkarzinoms
11. Das TNM-Stadium des klarzelligen Nierenzellkarzinoms
12. Das eventuelle Vorliegen einer Veneninvasion
13. Das eventuelle Vorliegen einer Lymphgefäßinvasion

2.3 Segmentierung

Die klarzelligen Nierenzellkarzinome wurden in den exportierten CT-Aufnahmen zunächst identifiziert. Sowohl die Tumorgröße als auch die Tumorage waren in der

Mehrzahl der Fälle in den histopathologischen Befunden angegeben, was die Identifikation der Tumoren beispielsweise bei unscharfen Tumorgrenzen, multiplen Tumoren oder gleichzeitig vorliegenden Zysten erleichterte.

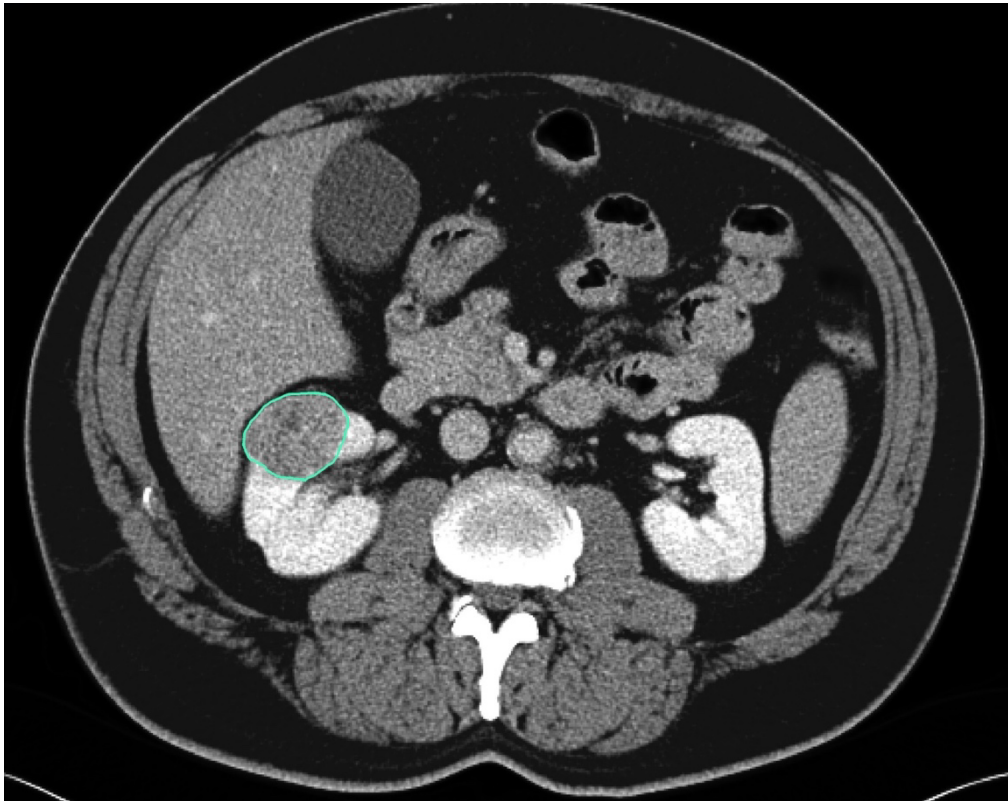


Abbildung 4: Darstellung der Tumorsegmentierung in ITK-Snap.

Bildschirmaufnahme der Segmentierung eines klarzelligen Nierenzellkarzinoms der rechten Niere (WHO/ISUP G1). Es handelt sich hierbei um eine frühvenöse CT-Aufnahme.

Anschließend erfolgte die Segmentierung der klarzelligen Nierenzellkarzinome mithilfe der Software ITK-Snap (Version 3.8.0) (97). Diese erfolgte durch einen Untersucher (Maximilian Winter), wobei alle Segmentierungen von einer radiologischen Fachärztin kontrolliert und gegebenenfalls angepasst wurden. Für die Segmentierung der klarzelligen Nierenzellkarzinome wurden in jeder einzelnen Schicht, in der der Tumor vorhanden war, die Tumorgrenzen manuell mit einem digitalen Annotationswerkzeug umrandet. Dieses Verfahren wurde bei mehreren vorliegenden Phasen für jede Phase einzeln wiederholt. Eventuell vorliegende Infiltrationen der V. renalis oder vorhandene

Tumorthromben in der Vena cava wurden ebenfalls annotiert. Es wurde darauf geachtet, sowohl Gefäße als auch das Nierenkelchsystem bei der Segmentierung auszusparen. In der CT-Aufnahme sichtbare Metastasen des Primärtumors, wie beispielsweise Metastasen in der Nebenniere, wurden nicht annotiert.

Bei mehreren klarzelligen Nierenzellkarzinomen mit demselben histopathologischen Grad wurde dieselbe farbkodierte Maske zur Annotation verwendet. Wenn zwei klarzellige Nierenzellkarzinome vorlagen, die einen unterschiedlichen histopathologischen Tumorgrad aufwiesen, wurde für eines der Karzinome eine abweichende farbkodierte Maske verwendet. Für die Segmentierung der betroffenen Nieren wurde ebenfalls eine Maske mit eigener Farbkodierung verwendet.

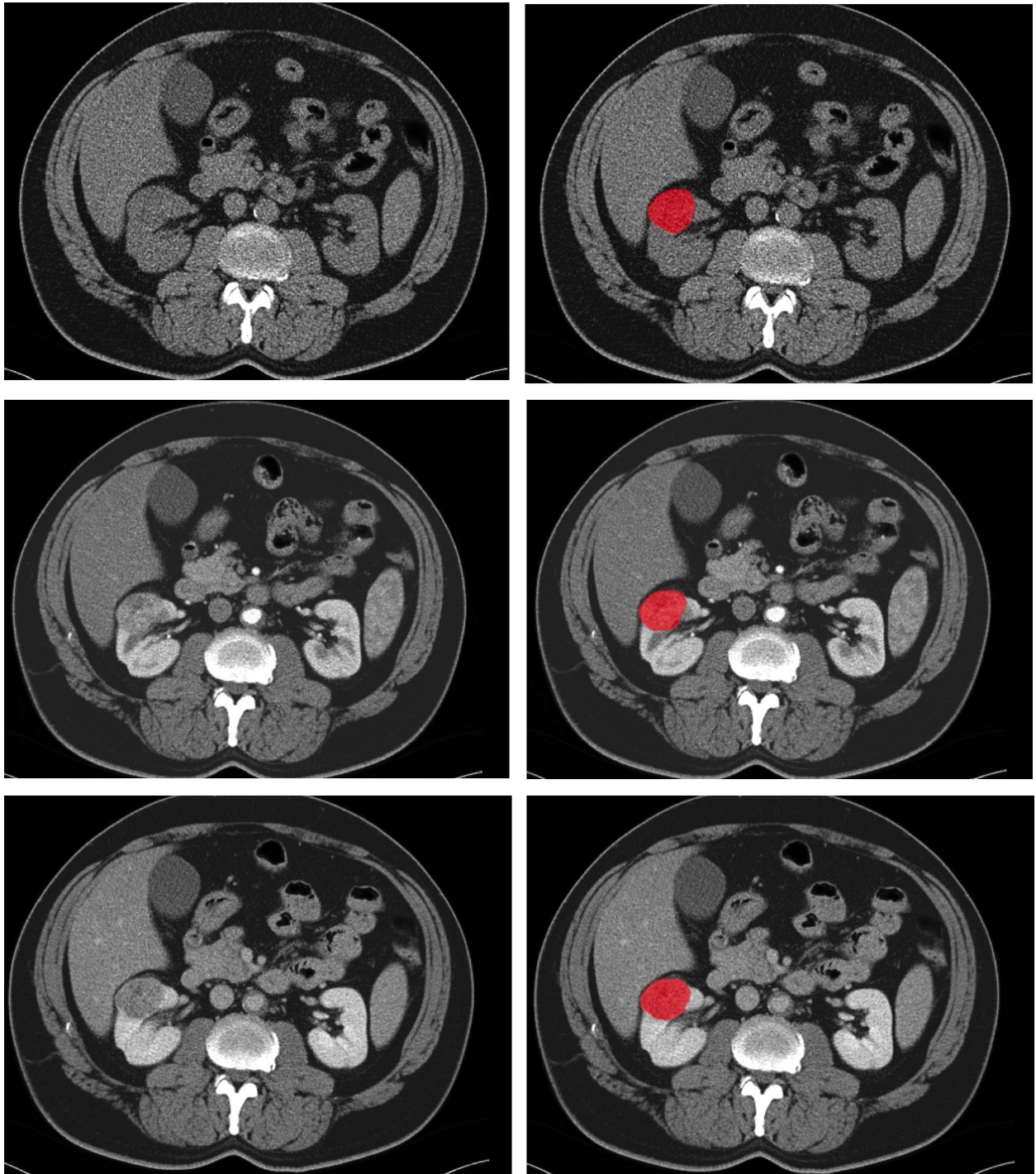


Abbildung 5: Bildschirmaufnahmen des in der Abbildung 4 segmentierten klarzelligen Nierenzellkarzinoms (WHO/ISUP G1) der rechten Niere in derselben Schnittebene in drei verschiedenen Phasen.

In der oberen Reihe ist die native Phase, in der mittleren Reihe die arterielle Phase und in der unteren Reihe die frühvenöse Phase abgebildet. Links ist jeweils das Nierenzellkarzinom vor der Annotation, rechts jeweils nach erfolgter Annotation zu sehen.

2.4 Extraktion der quantitativen Bildmerkmale

Die Extraktion der quantitativen Bildmerkmale erfolgte mit der Programmiersprache Python (Version 3.8.) und der Bibliothek „Pyradiomics“ (98). Insgesamt wurden 1221 Bildmerkmale extrahiert. Die Bildmerkmale wurden hierbei sowohl aus den Originaldaten als auch nach der Anwendung multipler Bildfilter (Wavelet- sowie Log-transformation-Filter) extrahiert. Es wurden 14 formbasierte Bildmerkmale aus den Originaldaten sowie je 18 Bildmerkmale erster Ordnung und 68 matrixbasierte Bildmerkmale aus den Originaldaten und den durch Bildfilter veränderten Bilddaten extrahiert.

Die Extraktion der Bildmerkmale erfolgte für jedes VOI in der Segmentationsmaske separat, sodass jeweils 1221 Bildmerkmale der Nierenraumforderungen und 1221 Bildmerkmale des gesunden Nierengewebes extrahiert wurden. Anschließend wurden die Merkmale tabellarisch im sogenannten Comma-separated values (CSV) für nachfolgende Analysen gespeichert.

2.5 Selektion von Bildmerkmalen mit hoher Reproduzierbarkeit

Ein Bildmerkmal hat eine hohe Reproduzierbarkeit, wenn seine Ausprägung sich trotz unterschiedlicher Bildgebungsprotokolle oder Bildrekonstruktionsparameter nur geringfügig ändert. Des Weiteren sollte, wenn ein Tumor von mehreren Annotatoren segmentiert wird und sich die Segmentationsmaske hierdurch geringfügig ändert, die Ausprägung des Merkmals ebenfalls konstant bleiben.

Da in dieser retrospektiven Auswertung eine erneute Bildakquise mit gering veränderten Bildgebungsprotokollen oder eine Rekonstruktion der Bilder mit anderen Parametern nicht möglich war, wurden diese Umstände simuliert. Ebenfalls erfolgte eine Simulation von zwei zusätzlichen Annotatoren.

Für die Simulation wurden die Originalbilddaten oder die Segmentationsmasken zuerst geringfügig abgeändert und anschließend die oben genannten Bildmerkmale erneut extrahiert. Folgende Änderungen der Segmentationsmasken und Bilddaten fanden Anwendung.

2.5.1 Voxel Dropout

Es wurden zufällig 2% der Voxel aus der Maske entfernt, wodurch diese Maskenareale bei der erneuten Extraktion der Bildmerkmale nicht miteinbezogen wurden.

Bildmerkmale, deren Ausprägung sehr stark von einzelnen Voxelwerten abhängt, würden hierdurch eine geringere Reproduzierbarkeit zeigen.

2.5.2 Simulation von zwei zusätzlichen Annotatoren

Durch geringe Verkleinerung des Radius der Segmentationsmasken um 1 beziehungsweise 2 Voxel, einem folgenden geringen Voxel Dropout und der anschließend erneuten Vergrößerung des Radius wurde eine geringfügige Formveränderung der Masken erzeugt. Hierdurch konnten ein zweiter und ein dritter Annotator der Daten simuliert werden.

2.5.3 Rauschreduktion/Rauscherhöhung

Unterschiedliche Bildrekonstruktionsalgorithmen, verschiedene CT-Scannergeräte und auch die angewandte Strahlendosis während der Untersuchung führen zu einer unterschiedlichen Ausprägung von Bildrauschen. Durch die Reduktion oder das Erhöhen des Bildrauschens können diese unterschiedlichen Gegebenheiten simuliert werden.

Angewandt für die Rauschreduzierung wurde der „curvature driven image denoising“ Algorithmus der SimpleITK Programmbibliothek (99). Die Addition von Bildrauschen wurde durch die Änderung der Pixelintensitäten um +/- 1% erwirkt.

2.5.4 Bildschärfung/Weichzeichnung

Rekonstruktionsparameter wie beispielsweise der Faltungskernel in der computertomographischen Bildgebung können zu einem schärferen oder weicheren Bildeindruck führen. Um dies zu simulieren, wurde eine nachträgliche Bildschärfung mittels Unschärfmaskierung durchgeführt. Eine Weichzeichnung des Bildes wurde bereits bei der Anwendung des Algorithmus für die Rauschreduzierung erreicht.

Die Reproduzierbarkeit der Bildmerkmale wurde über den Intraklassen-Korrelationskoeffizienten (ICC) bestimmt. Angelehnt an die Interpretation des ICC für die Interrater-Übereinstimmung von *Koo und Li* wurde eine gute Reproduzierbarkeit des Bildmerkmals bei einem ICC von ≥ 0.75 festgelegt (100).

Um eine höhere Generalisierbarkeit der Ergebnisse zu gewährleisten, wurden zehn zufällig generierte Teilmengen gebildet, die jeweils 80% der Beobachtungen der

Originaldaten enthielten. Die ICC wurden für jede dieser Teilmengen berechnet. Anschließend wurde das arithmetische Mittel der ICC aller Teilmengen gebildet. 420 Bildmerkmale zeigten eine schlechte Reproduzierbarkeit mit einem mittleren ICC von < 0.75 und wurden von der weiteren Analyse ausgeschlossen.

2.6 Datenunterteilung und Oversampling

Vor der Durchführung der weiteren Merkmalsselektion erfolgte die Unterteilung der Daten in einen Testdatensatz und einen Trainingsdatensatz. 20% der Patienten wurden zufällig dem Testdatensatz zugeordnet, während die restlichen 80% der Patienten dem Trainingsdatensatz zugeordnet wurden. Hierbei wurde darauf geachtet, dass die Distribution der Tumorgade im Test- und Trainingsdatensatz identisch war.

Tabelle 1: Verteilung der Tumorgade in den Trainings- und Testdaten.

Die Tabelle zeigt die Verteilung der einzelnen Tumorgade. Patienten konnten multiple Tumoren aufweisen und eine multiphasische CT erhalten haben, weshalb die Summe an Tumorgaden nicht der Anzahl der Patienten entspricht.

Tumorgrad	n (Trainingsdaten)	n (Testdaten)	n (Oversampling)
Fuhrman G1	29	8	186
Fuhrman G2	161	25	169
Fuhrman G3	76	11	183
Fuhrman G4	19	7	133
WHO/ISUP G1	42	21	187
WHO/ISUP G2	173	5	173
WHO/ISUP G3	70	3	189
WHO/ISUP G4	33	8	191

Die Verteilung der verschiedenen Tumorgrade im Trainingsdatensatz zeigte ein ausgeprägtes Ungleichgewicht der Klassen (siehe Tabelle 1), weshalb ein Oversampling der unterrepräsentierten Klassen angewendet wurde.

Für das Simulieren der unterrepräsentierten Klassen wurden die augmentierten Bildmerkmale aus der Reproduzierbarkeitsanalyse (Abschnitt 2.5) gewählt. Hierdurch konnte erreicht werden, dass die Bildmerkmale einer Klasse eine ähnliche Distribution aufwiesen, ohne eine identische Replikation zu sein. Die Datenpunkte wurden zufällig gewählt, bis die Größe einer Klasse der Mehrheitsklasse ähnelte oder bis alle augmentierten Daten der Klasse in das Training eingeschlossen wurden. Ein Oversampling des Testdatensatzes fand nicht statt.

2.7 Ausschluss von Bildmerkmalen mit starker paarweiser Korrelation

Die Korrelationsanalyse wurde nur auf dem Trainingsdatensatz durchgeführt. Anschließend wurden die Ergebnisse auf den Testdatensatz übertragen. Aufgrund des Vorliegens von teilweise multiplen Kontrastmittelphasen in einer CT-Untersuchung wurde die Korrelationsanalyse getrennt für Aufnahmen ohne Kontrastmittel oder Bildakquise in der venösen und arteriellen Kontrastmittelphase durchgeführt. Die Bildmerkmale wurden untereinander korreliert, wobei bei dem Vorliegen eines Korrelationskoeffizienten (Pearson) von $r > 0.9$ der Ausschluss eines Bildmerkmals erfolgte. Zusätzlich wurden die paarweisen Korrelationen der Bildmerkmale untereinander in einer farbcodierten Korrelationsmatrix dargestellt (siehe Abbildungen 6-8).

Durch die Korrelationsanalyse konnten 562 Bildmerkmale für die venöse Kontrastmittelphase, sowie je 567 Merkmale für die arterielle und native Kontrastmittelphase ausgeschlossen werden.

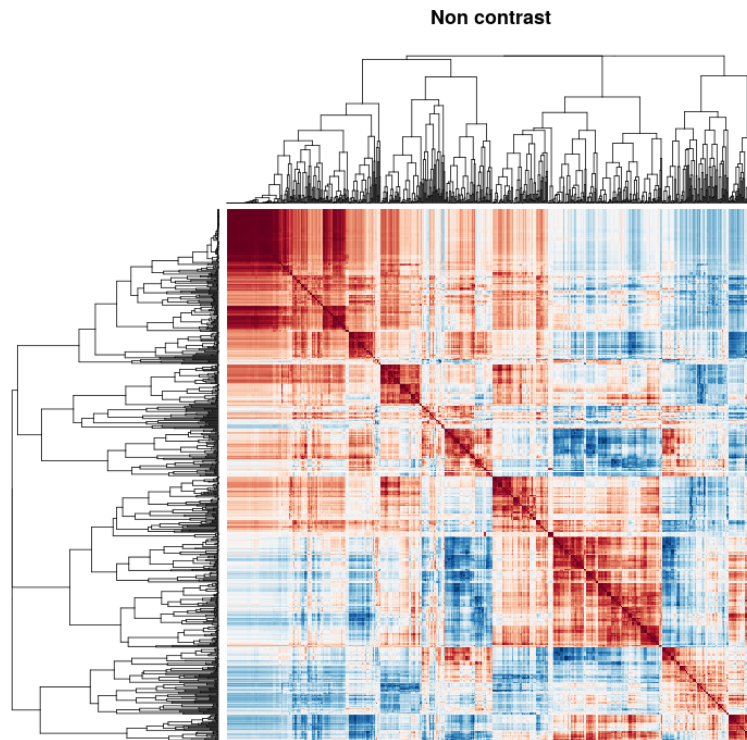


Abbildung 6: Farbkodierte Korrelationsmatrix der nativen CT-Aufnahmen.

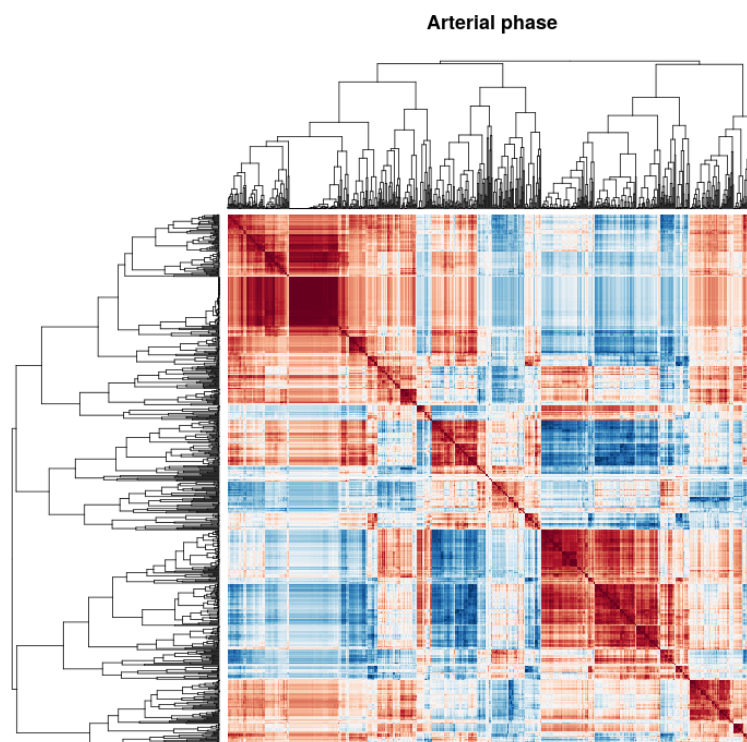


Abbildung 7: Farbkodierte Korrelationsmatrix der arteriellen CT-Aufnahmen.

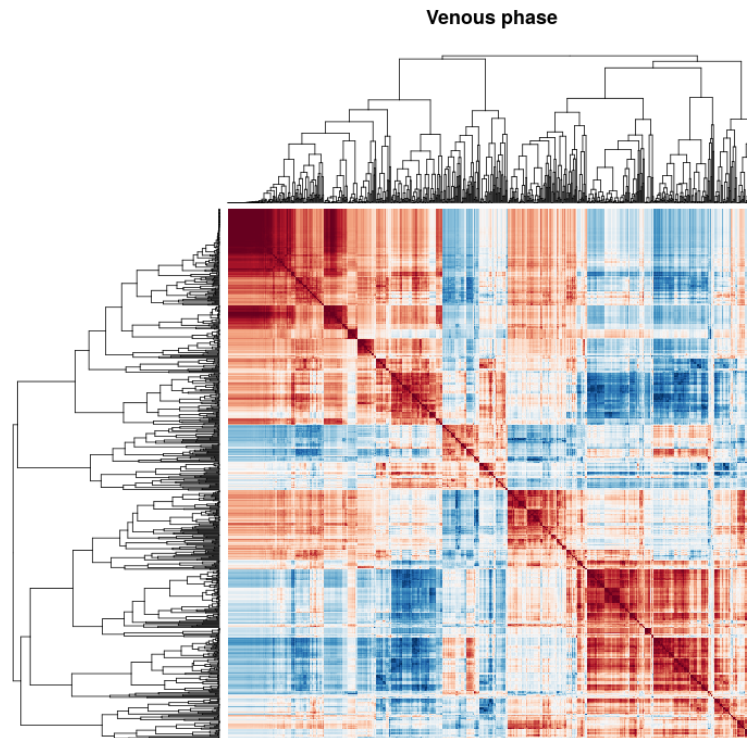


Abbildung 8: Farbkodierte Korrelationsmatrix der venösen CT-Aufnahmen.

2.8. Selektion von Bildmerkmalen mittels ANOVA

Nach Reproduzierbarkeitsanalyse und Korrelationsanalyse verblieben 234 Bildmerkmale für die venöse CT, 229 Merkmale für die arterielle CT und 229 Merkmale für die native CT. Ein weiterer Schritt der Merkmalsselektion untersuchte deshalb, ob es signifikante Unterschiede in der Verteilung der Merkmale zwischen den Gruppen gab. Hierfür wurde für jedes der Merkmale eine Varianzanalyse zwischen den Klassen durchgeführt. Die Varianztestung erfolgte erneut nicht auf dem gesamten Trainingsdatensatz, sondern auf 10 zufällig zusammengesetzten, jeweils 80% der Größe des Trainingsdatensatzes entsprechenden, Teilmengen. Zeigte ein Merkmal in mindestens fünf der zehn Teilmengen keinen signifikanten Unterschied, wurde es von der weiteren Analyse ausgeschlossen. Die Testung der Unterschiede erfolgte hierbei nur auf den Originaldaten, nicht auf den durch Oversampling hinzugefügten Daten. Die Anzahl der Bildmerkmale konnte somit auf 17 für die venöse CT, 52 für die arterielle CT und 4 für die native CT reduziert werden.

Die Durchführung der Reproduzierbarkeitssanalyse, der Korrelationsanalyse und der ANOVA erfolgte mit der Programmiersprache „R“ (Version 4.0.1) (101) und den Programmbibliotheken „tidyverse“ (102) und „irr“ (103).

2.9 Trainieren von Modellen des maschinellen Lernens

Es wurden vier verschiedene Verfahren des maschinellen Lernens bezüglich der Performance für die Unterscheidung von niedriggradigen klarzelligen Nierenzellkarzinomen (definiert als WHO/ISUP G1 oder Fuhrman G1), mittelgradigen Karzinomen (WHO/ISUP G2/G3 oder Fuhrman G2/3) sowie hochgradigen Karzinomen (WHO/ISUP G4 oder Fuhrman G4) evaluiert.

Vor dem Training der Modelle erfolgte eine Normalisierung und Zentrierung der Daten auf einen Mittelwert von 0 und eine Standardabweichung von 1. Das Training jedes Modells erfolgte mittels 10-fach wiederholter 10-facher Kreuzvalidierung. Die Modelle waren ein Random Forest Modell, ein Gradient Boosting Modell, ein L1/L2 regularisiertes logistisches Regressionsmodell und ein k-Nearest Neighbor Modell.

Alle Modelle wurden mit der Programmiersprache „R“ (Version 4.0.1) (101) und der Programmbibliothek „caret“ (104) trainiert. Das Training erfolgte getrennt auf den Merkmalen der venösen CT, der arteriellen CT und der nativen CT.

2.9.1 Überblick über die vier verwendeten Modelle

2.9.1.1 Random Forest Modell und Gradient Boosting Modell

Die Grundlage sowohl des Random Forest Modells als auch des Gradient Boosting Modells bilden sogenannte Entscheidungsbäume (Decision Trees) (105, 106). Grundsätzlich ist in einem Entscheidungsbaum eine Reihenfolge von Entscheidungen festgelegt, die jeweils auf Basis der Ausprägung eines Merkmals, hier auch Prädiktor genannt, getroffen werden. Durch eine Entscheidung wird ein Objekt einer von zwei Gruppen zugeordnet, was im Entscheidungsbaum auch als Aufzweigung zu erkennen ist (siehe Abbildung 9). Durch jede weitere Entscheidung soll enger eingegrenzt werden, wie das Objekt schließlich zu klassifizieren ist, es soll also eine möglichst eindeutige Zuteilung zu einer der beiden Gruppen erfolgen. Sobald kein weiterer Prädiktor eine genauere Eingrenzung ermöglicht, erfolgt die Klassifizierung des Objekts, also dessen

Zuordnung zu einer der vorab definierten Zielvariablen. Die Anzahl der Aufzweigungen entlang eines Baums wird als Tiefe des Entscheidungsbaums bezeichnet.

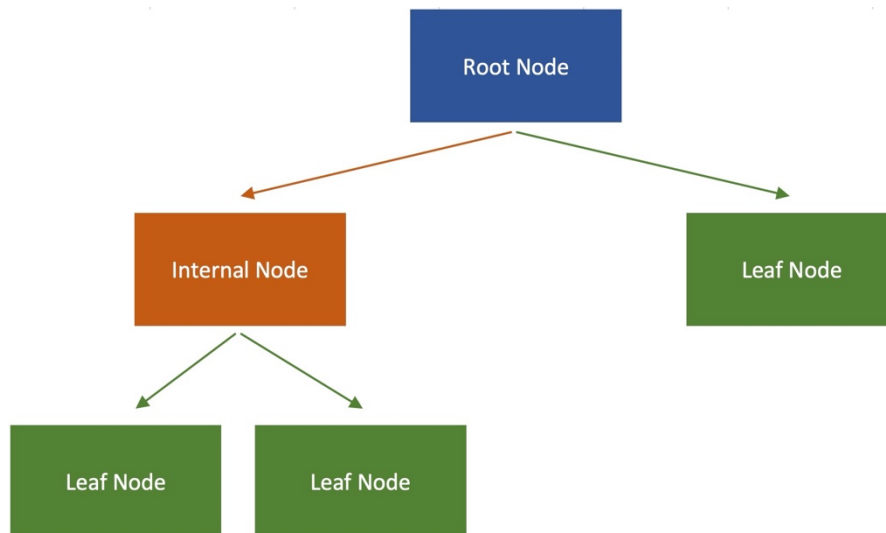


Abbildung 9: Grundlegender Aufbau eines Entscheidungsbaums.

Die Leserichtung eines Entscheidungsbaums ist von oben nach unten festgelegt, wobei der ganz oben stehende Prädiktor auch als Root Node bezeichnet wird. Der weitere Verlauf des Entscheidungsbaums ist unter anderem von der Merkmalskonstellation des jeweiligen Datensatzes abhängig. Entweder ermöglicht ein weiterer Prädiktor, dann als Internal Node definiert, eine erneute Aufzweigung (siehe linke Seite des Entscheidungsbaums), oder es folgt stattdessen direkt anhand der Root Node die Klassifizierung des Objekts (siehe rechte Seite des Entscheidungsbaums) und somit das Ende des Entscheidungsbaums, auch Leaf Node genannt. Je nach Datensatz und verwendetem Modell des maschinellen Lernens können zahlreiche weitere Internal Nodes zur Klassifizierung eines Objekts verwendet werden, was mehr Aufzweigungen und somit eine größere Baumtiefe zur Folge hat.

Random Forest und Gradient Boosting zählen zu den sogenannten Ensemble Methoden (79). Ensemble Methoden nutzen mehrere simple statistische Modelle und fassen die Aussagen dieser Modelle zusammen, um eine Entscheidung bezüglich der Klassifizierung eines Objekts zu treffen (79). Random Forest und Gradient Boosting bestehen jeweils aus mehreren Entscheidungsbäumen, wobei sie sich grundsätzlich in der Art und Weise, wie diese Bäume erstellt und wie die Entscheidungen der einzelnen Bäume zusammengefasst werden, unterscheiden.

Bei einem Random Forest Modell werden mehrere, nicht aufeinander basierende Entscheidungsbäume parallel erstellt (105). Ein Entscheidungsbaum eines Random Forest Modells wird nicht anhand des Originaldatensatzes erstellt, sondern auf Basis eines durch Bootstrapping erstellten gleichgroßen Datensatzes (105). Außerdem wird bei der Baumerstellung bei jeder Entscheidung jeweils nur eine definierte Anzahl an zufällig ausgewählten Prädiktoren verwendet (105). Die Klassifizierung eines Objekts durch ein Random Forest Modell wird durch Mittelung der Entscheidungen aller einzelnen Bäume gefällt (105). Jeder einzelne Entscheidungsbaum hat bei der Klassifizierung denselben Anteil (79).

Beim Gradient Boosting Modell werden die Entscheidungsbäume schrittweise nacheinander auf Basis des Fehlers des jeweils vorangegangenen Entscheidungsbaums erstellt, wobei der Fehler beispielsweise die Abweichung des vorhersagten Werts eines Objekts von dessen tatsächlich gemessenem Wert darstellt (106). Des Weiteren wird eine Lernrate zwischen 0 und 1 definiert, damit der Fehler des vorangegangenen Entscheidungsbaumes nur teilweise in die Erstellung des nächsten Baums eingeht und jeder Entscheidungsbaum nur teilweise zur Klassifizierung des Objekts beiträgt (79). Auf diese Weise kann ein Overfitting des statistischen Modells vermieden werden (106). Die Klassifizierung eines Objekts erfolgt schließlich durch Addition der einzelnen Entscheidungsbäume (106). Dadurch, dass der Fehler jedes Entscheidungsbaums unterschiedlich groß ist, folgt, dass sich schließlich der Beitrag zu der Klassifizierung zwischen den einzelnen Entscheidungsbäumen unterscheidet (79). Die Anzahl der Bäume in einem Gradient Boosting Modell entspricht der Anzahl der Boosting Durchläufe (79). Diese Anzahl wird durch den Verwender des jeweiligen Modells definiert (106).

Sowohl Gradient Boosting Modelle als auch Random Forest Modelle zeigen oftmals eine ähnliche Performance bei der Klassifizierung von Objekten (79). Random Forest Modelle benötigen jedoch in der Regel weniger Computerleistung als Gradient Boosting Modelle, da sie pro Baum nur einen Teil der Merkmale verwenden und alle Entscheidungsbäume parallel prozessiert werden können (79). Darüber hinaus haben beide Modelle den Vorteil, dass ihre Performance nicht sehr abhängig von einer Vorverarbeitung der quantitativen Merkmale ist (79).

2.9.1.2 L1/L2 regularisiertes Regressionsmodell

L1 Regularisierung (Lasso Regression) und L2 Regularisierung (Ridge Regression) gehören beide zu den sogenannten penalisierten Modellen (79, 107, 108). Um auf Basis

von Trainingsdaten einen Testdatensatz vorhersagen zu können, ist es möglich, beispielsweise mithilfe der Methode der kleinsten Quadrate eine Funktion zu bestimmen, die möglichst gut an die Trainingsdaten angepasst ist, und diese Funktion dann auf unbekannte Daten anzuwenden (79). Bei der Methode der kleinsten Quadrate wird die Funktion bestimmt, deren Datenpunkte in Summe den geringsten quadrierten Abstand zu den Trainingsdaten haben (79). Hierbei kann jedoch das Problem bestehen, dass eine Funktion zwar sehr gut an die Trainingsdaten angepasst ist, also einen sehr geringen oder keinen Bias hat, diese Funktion jedoch schlechte Vorhersagen bezüglich unbekannter Daten bietet, also eine große Varianz hat, was insgesamt also einem Overfitting entspricht (79).

Um bessere Vorhersagen für unbekannte Daten treffen zu können, bietet es sich an, eine Funktion zu erstellen, die zwar einen etwas größeren Bias hat, dafür aber einen Testdatensatz deutlich besser klassifizieren kann (79). Auf dieser Grundlage basieren penalisierte Modelle (79). Sowohl bei der L1 Regularisierung als auch bei der L2 Regularisierung wird ein Bias eingeführt, indem die Methode der kleinsten Quadrate durch einen Penalisierungsterm erweitert wird (79). Dieser Penalisierungsterm bewirkt ein Schrumpfen der Koeffizienten der einzelnen quantitativen Merkmale (79). Die Gewichtung der Penalisierung wird durch den Wert λ bestimmt (79). Während bei der Ridge Regression die Koeffizienten nur gegen 0 schrumpfen, können die Koeffizienten bei der Lasso Regression auch den Wert 0 annehmen (79). Durch Lasso Regression kann folglich zusätzlich zur Regularisierung eine Selektion der quantitativen Merkmale erfolgen (79).

Das Elastic Net vereint die L1 Regularisierung und L2 Regularisierung miteinander (109). Bei dem in der vorliegenden Arbeit verwendeten Elastic Net Modell von "glmnet" wird durch den Wert α definiert, wie die jeweilige Regularisierungsmethode gewichtet wird (79). Bei einem α -Wert von 1 wird ausschließlich die L1 Regularisierung verwendet, bei einem Wert von 0 hingegen ausschließlich die L2 Regularisierung (79). Wenn α zwischen 0 und 1 festgelegt ist, erfolgt eine Kombination beider Methoden (79). Durch den Parameter λ wird darüber hinaus die Gewichtung der Penalisierung insgesamt festgelegt (79).

Die optimalen Parameter-Werte der penalisierten Modelle können durch Kreuzvalidierung gefunden werden (79).

2.9.1.3 k-Nearest Neighbor Modell

Das k-Nearest Neighbor Modell gehört zu den nichtlinearen Klassifikationsmodellen (79, 110). Beim k-Nearest Neighbor Modell werden unbekannte Daten klassifiziert, indem eine vorab bestimmte Zahl k an benachbarten Trainingsdatenpunkten analysiert wird (79). Wenn beispielsweise die drei am nächsten benachbarten Trainingsdatenpunkte eines bisher unbekanntes Datenpunkts alle derselben Klasse angehören und für k die Zahl 3 definiert wurde, wird der unbekannte Datenpunkt ebenfalls dieser Klasse zugeordnet. Falls die k -nächsten Nachbarn eines unbekanntes Datenpunkts unterschiedlichen Klassen angehören, so wird dieser Datenpunkt der am häufigsten vertretenen Klasse zugeordnet; es findet also eine Mehrheitsentscheidung statt (79). Die Abstände zu den Datenpunkten können hierbei mithilfe unterschiedlicher Maße definiert werden, am häufigsten wird hierfür die euklidische Distanz verwendet (79).

Je niedriger k gewählt wird, desto eher besteht die Gefahr eines Overfittings (79). Wenn k beispielsweise mit 1 definiert wird, kann ein einzelner Ausreißer im Trainingsdatensatz zur Klassifizierung des unbekanntes Datenpunkts führen (79). Je höher k gewählt wird, desto mehr Nachbarn werden in die Klassifizierung miteinbezogen. Bei hohen k -Werten kann daher die Trennschärfe des Modells abnehmen. Der optimale k -Wert kann durch Kreuzvalidierung gefunden werden (79).

Eine Skalierung der Daten ist beim k-Nearest Neighbor Modell besonders wichtig, da ansonsten Prädiktoren mit großem Maßstab einen höheren Einfluss auf die Berechnung des Abstands zwischen zwei Datenpunkten und somit auch auf die die Klassifizierung durch das Modell hätten (79). Des Weiteren ist auch eine bereits erfolgte Selektion der quantitativen Merkmale von großer Bedeutung, da irrelevante Merkmale gegebenenfalls zu großen Abständen zwischen sonst ähnlichen Datenpunkten führen können und somit die Performance des Modells verschlechtern würden (79).

Je mehr Datenpunkte zur Verfügung stehen, desto größer wird auch die erforderliche Computerleistung zur Berechnung der Abstände eines zu klassifizierenden Datenpunktes zu den Trainingsdatenpunkten (79).

2.9.2 Training des Random Forest Modells

Für das Training des Random Forest Modells wurde die Implementierung von *Liaw und Wiener* (111) genutzt. Verschiedene Modelle wurden über eine Rastersuche evaluiert, wobei die Anzahl an Prädiktoren pro Modell zwischen 2, 9 und 17 variiert wurde. Das

finale Modell wurde anhand der besten Performance in der Kreuzvalidierung gewählt und nutzte 2 Prädiktoren pro Entscheidungsbaum.

2.9.3 Training des Gradient Boosting Modells

Das Gradient Boosting Modell wurde unter Nutzung des „R“ pakets „gbm“ (112) durchgeführt. Eine Rastersuche verschiedener Modelle wurde durchgeführt mit Variation folgender Hyperparameter: Anzahl der Boosting-Durchläufe zwischen 50, 100 oder 150, Tiefe der Entscheidungsbäume zwischen 1, 2 oder 3, einer Lernrate von 0.1 und einer Mindestanzahl von 10 Beobachtungen pro Endpunkt des Entscheidungsbaums. Das finale Modell wurde durch die Performance in der Kreuzvalidierung gewählt und wies 150 Boosting-Durchläufe und eine Baumtiefe von 3 auf.

2.9.4 Training des L1/L2 regularisierten Regressionsmodells

Für das Training wurde das „R“ paket „glmnet“ (113) verwendet. Eine Rastersuche wurde durchgeführt, wobei der Regularisierungsparameter α zwischen 0 und 1 mit einem Inkrement von 0.25 und der Regularisierungsparameter λ zwischen $1/e^1$ und $1/e^{10}$ mit einem Inkrement des Exponenten von 2,25 festgelegt wurde. Die beste Performance wurde durch das Modell mit den Parametern $\alpha = 0.25$ und $\lambda = 0.004$ erzielt.

2.9.5 Training des k-Nearest Neighbor Modells

Für das Training des k-Nearest Neighbor Algorithmus wurde die Implementierung des „caret“-Pakets (104) genutzt. Die Anzahl an Nachbarn wurde zwischen 5, 7 und 9 variiert. Die beste Performance ließ sich mit 5 Nachbarn erzielen.

3 Ergebnisse

3.1 Studienpopulation

Insgesamt konnten 382 Patienten mit 388 klarzelligem Nierenzellkarzinomen eingeschlossen werden. Während bei 373 Patienten ausschließlich ein klarzelliges Nierenzellkarzinom vorlag, wurden bei fünf Patienten zwei klarzellige Nierenzellkarzinome und bei einem Patienten drei klarzellige Nierenzellkarzinome diagnostiziert. Darüber hinaus konnten bei vier Patienten weitere maligne Raumforderungen der Niere erfasst werden. Bei drei Patienten handelte es sich um ein papilläres Nierenzellkarzinom, während bei einem weiteren Patienten zwei Urothelkarzinome vorlagen.

260 Patienten (68,1%) waren männlich und 122 Patienten (31,9%) weiblich. Das Alter der Patienten zum Zeitpunkt der Operation beziehungsweise der Biopsie betrug im Mittel 63,2 Jahre, wobei der jüngste Patient 27 Jahre und der älteste Patient 90 Jahre alt war. Die Analyse der Tumorgrade nach WHO/ISUP ergab in 29 Fällen den Tumorgrad WHO/ISUP G1, in 100 Fällen den Tumorgrad WHO/ISUP G2, in 44 Fällen den Tumorgrad WHO/ISUP G3 und in 20 Fällen den Tumorgrad WHO/ISUP G4. Die Analyse der Tumorgrade nach Fuhrman ergab in 22 Fällen den Tumorgrad Fuhrman G1, in 108 Fällen den Tumorgrad Fuhrman G2, in 54 Fällen den Tumorgrad Fuhrman G2 und in 11 Fällen den Tumorgrad Fuhrman G4.

Der längste Durchmesser der klarzelligem Nierenzellkarzinome betrug im Mittel 59,07 mm (Standardabweichung: 33,80 mm, Spannweite: 10mm-170 mm, Median: 50 mm), wobei hochgradige Nierenzellkarzinome im Durchschnitt einen längeren Durchmesser als niedriggradige Nierenzellkarzinome aufwiesen (WHO/ISUP G1 / Fuhrman G1: 44,41 mm / 40,50 mm, WHO/ISUP G2 / Fuhrman G2: 52,41 mm / 51,56 mm, WHO/ISUP G3 / Fuhrman G3: 70,25 mm / 78,04 mm, WHO/ISUP G4 / Fuhrman G4: 79,50 mm / 94,36 mm).

Von den 388 klarzelligem Nierenzellkarzinomen konnten 176 Karzinome der linken Niere und 212 Karzinome der rechten Niere zugeordnet werden. 102 Karzinome befanden sich am Oberpol, 77 am Unterpol der jeweiligen Niere. Weitere 101 Karzinome befanden sich im Mittelgeschoss der Niere. Die Anzahl der Tumoren, die sowohl den Oberpol als auch das Mittelgeschoss befielen, belief sich auf 33. Analog hierzu konnten 44 klarzellige Nierenzellkarzinome identifiziert werden, die sowohl den Unterpol als auch das

Mittelgeschoss der betroffenen Niere befielen. 31 Karzinome erstreckten sich über die ganze Niere inklusive Nierenoberpol, Nierenmittelgeschoss und Nierenunterpol. Insgesamt gab es 120 native CT-Aufnahmen, 207 arterielle CT-Aufnahmen und 327 venöse CT-Aufnahmen der klarzelligen Nierenzellkarzinome, wobei mittelgradige Nierenzellkarzinome (WHO/ISUP / Fuhrman G2/G3) mit 91 nativen, 171 arteriellen und 257 venösen CT-Aufnahmen jeweils die größte Untergruppe darstellten. Niedriggradige Karzinome (WHO/ISUP / Fuhrman G1) waren jeweils die zweithäufigste Untergruppe mit 16 nativen, 22 arteriellen und 42 venösen CT-Aufnahmen. Am seltensten waren jeweils hochgradige Nierenzellkarzinome (WHO/ISUP G4 / Fuhrman G4) vertreten; hier lagen 13 native, 14 arterielle und 28 venöse CT-Aufnahmen vor. Die Gesamtheit aller erhobenen Patienten- und Tumorcharakteristika ist in Tabelle 2 dargestellt.

Tabelle 2: Patienten- und Tumorcharakteristika.

Patienten- / Tumorcharakteristika	n (%) oder Mittelwert (\pmSD, Spannweite)
	Patienten (n = 382)
	Klarzellige Nierenzellkarzinome (n = 388)
Geschlecht der Patienten	
Männlich	260 (31,9)
Weiblich	122 (68,1)
Patientenalter bei OP / Biopsie in Jahren	
Alle Patienten	63,2 (\pm 11,2, 27-90)
27-45 Jahre	23 (6%)
46-60 Jahre	129 (33,8%)
61-75 Jahre	172 (45,0%)
76-90 Jahre	58 (15,2%)
Nephrektomie / Biopsie	
Nephrektomie	223 (58,4)
Teilnephrektomie	147 (38,5)

Biopsie	12 (3,1)
T-Stadium	
T1	3 (0,8)
T1a	131 (34,3)
T1b	61 (16,0)
T2	2 (0,5)
T2a	26 (6,8)
T2b	7 (1,8)
T3	5 (1,3)
T3a	99 (25,9)
T3b	18 (4,7)
T3c	7 (1,8)
T4	9 (2,4)
NA	14 (3,7)
N-Stadium	
N0	50 (13,1)
N1	10 (2,6)
N2	4 (1,0)
NA	318 (83,2)
M-Stadium	
M0	16 (4,2)
M1	40 (10,5)
NA	326 (85,3)
L-Stadium	
L0	303 (79,3)
L1	7 (1,8)
NA	72 (18,8)

V-Stadium

V0	249 (65,2)
V1	70 (18,3)
V2	26 (6,8)
NA	37 (9,7)

Histopathologischer Tumorgrad

WHO/ISUP G1	29 (15,0)
WHO/ISUP G2	100 (51,8)
WHO/ISUP G3	44 (22,8)
WHO/ISUP G4	20 (10,4)
Fuhrman G1	22 (11,3)
Fuhrman G2	108 (55,4)
Fuhrman G3	54 (27,7)
Fuhrman G4	11 (5,6)

Tumurlage (Höhe)

Oben	102 (26,3)
Mittig + Oben	33 (8,5)
Mittig	101 (26,0)
Mittig + Unten	44 (11,3)
Unten	77 (19,8)
Gesamte Niere	31 (8,0)

Tumurlage (Seite)

Links	176 (45,4)
Rechts	212 (54,6)

Tumorgröße In mm

Alle klarzelligen Nierenzellkarzinome	59,07 (±33,80, 10-170)
--	------------------------

Native Phase

Niedriggradig (G1)	16 (13,0)
---------------------------	-----------

Mittelgradig (G2/G3)	91 (76,0)
Hochgradig (G4)	13 (11,0)
Arterielle Phase	
Niedriggradig (G1)	22 (10,6)
Mittelgradig (G2/G3)	171 (82,6)
Hochgradig (G4)	14 (6,8)
Venöse Phase	
Niedriggradig (G1)	42 (12,8)
Mittelgradig (G2/G3)	257 (78,6)
Hochgradig (G4)	28 (8,6)

3.2 Ergebnisse des Trainings der Modelle des maschinellen Lernens

Das Random Forest Modell schnitt mit 86% Accuracy (Kappa 61%) bei arteriellen beziehungsweise 83% Accuracy (Kappa 60%) bei venösen CT-Aufnahmen am besten ab. Das Gradient Boosting Modell wies bei arteriellen und venösen CT-Aufnahmen mit 79% Accuracy (Kappa 35%) und 81% Accuracy (Kappa 62%) ähnliche Werte auf. Das L1/L2 regularisierte Regressionsmodell konnte zwar in arteriellen CT-Aufnahmen mit einer Accuracy von 77% (Kappa 35%) ähnlich oft den histopathologischen Tumorgrad korrekt vorhersagen, zeigte jedoch mit 56% Accuracy (Kappa -4%) bei venösen CT-Aufnahmen eine schlechtere Leistung. Das k-Nearest-Neighbor Modell konnte in arteriellen CT-Aufnahmen in 50% der Fälle (Kappa -1%) und in venösen CT-Aufnahmen in 62% der Fälle (Kappa 37%) den richtigen Tumorgrad bestimmen. Durch das Training des Random Forest Modells auf Merkmalen der nativen CT-Aufnahmen konnte dieses nur in 50% der Fälle den richtigen Tumorgrad determinieren, der Kappa-Wert lag bei 14%. Hier konnten die anderen Modelle mit jeweils 80% Accuracy (Kappa 56%) des Gradient Boosting Modells und des L1/L2 regularisierten Regressionsmodells (Kappa 0%) beziehungsweise 70% Accuracy (Kappa 42%) des k-Nearest Neighbor Modells deutlich mehr klarzellige Nierenzellkarzinome dem korrekten histopathologischen Tumorgrad zuordnen.

In nativen Aufnahmen konnten niedriggradige und mittelgradige Karzinome am besten durch das Gradient Boosting Modell mit einer Balanced Accuracy von 94% (niedriggradig) beziehungsweise 88% (mittelgradig) klassifiziert werden. Die Sensitivität und Spezifität betragen 100% und 88% für die niedriggradigen Karzinome und 75% und 100% für die mittelgradigen Tumoren. Das k-Nearest Neighbor Modell konnte in derselben CT-Phase aufgenommene niedriggradige klarzellige Nierenzellkarzinome mit einer Balanced Accuracy von 88% (Sensitivität 100%, Spezifität 75%) und mittelgradige Karzinome mit einer Balanced Accuracy von 81% (Sensitivität 62%, Spezifität 100%) klassifizieren. Das Random Forest Modell zeigte Balanced Accuracy Werte von 75% (niedriggradig) und 50% (mittelgradig). Die Sensitivität und Spezifität lag hier bei 50% und 100% (niedriggradig) beziehungsweise 50% und 50% (mittelgradig). Das L1/L2 regularisierte Regressionsmodell demonstrierte ebenfalls niedrige Balanced Accuracy Werte von jeweils 50% (niedriggradig + mittelgradig). Die Sensitivität und Spezifität betrug 0% und 100% (niedriggradig) und 100% und 0% (mittelgradig).

Des Weiteren schnitten in arteriellen Aufnahmen sowohl das L1/L2 regularisierte Regressionsmodell mit einer Balanced Accuracy von 95% (Sensitivität 100%, Spezifität 90%) als auch das Random Forest Modell mit einer Balanced Accuracy von 92% (Sensitivität 100%, Spezifität 85%) am besten bei der Klassifizierung von niedriggradigen Tumoren ab. Sowohl das Gradient Boosting Modell als auch das k-Nearest Neighbor Modell zeigten hierbei geringere Balanced Accuracy Werte von 65% (Sensitivität 50%, Spezifität 80%) und 60% (Sensitivität 50%, Spezifität 70%). Der Tumorgrad mittelgradiger Karzinome konnte in arteriellen Aufnahmen durch das Random Forest Modell am genauesten bestimmt werden (Balanced Accuracy 92%, Sensitivität 84%, Spezifität 100%). Das Gradient Boosting Modell und das L1/L2 regularisierte Regressionsmodell wiesen hinsichtlich desselben Tumorgrads in arteriellen Aufnahmen Balanced Accuracy Werte von jeweils 73% auf (Sensitivität 79%, Spezifität 67%). Das k-Nearest Neighbor Modell schnitt mit 43% Balanced Accuracy (Sensitivität 53%, Spezifität 33%) deutlich schlechter ab. Hochgradige klarzellige Nierenzellkarzinome konnten sowohl durch das Random Forest Modell als auch durch das Gradient Boosting Modell durch die Analyse arterieller CT-Merkmale mit einer Balanced Accuracy von 100% (Sensitivität jeweils 100%, Spezifität jeweils 100%) klassifiziert werden. Deutlich geringere Balanced Accuracy Werte wurden mit 45% (Sensitivität 0%, Spezifität 90%) und 43% (Sensitivität 0%, Spezifität 86%) durch das L1/L2 regularisierte Regressionsmodell und durch das k-Nearest Neighbor Modell erreicht.

Weiterhin zeigte das Gradient Boosting Modell bei niedriggradigen klarzelligen Nierenzellkarzinomen in venösen Aufnahmen die beste Performance mit einer Balanced Accuracy von 89% (Sensitivität 91%, Spezifität 86%). Das Random Forest Modell konnte einen Wert von 81% erreichen (Sensitivität 73%, Spezifität 89%). Während das k-Nearest Neighbor Modell eine Balanced Accuracy von 77% demonstrierte (Sensitivität 54%, Spezifität 92%), schnitt das L1/L2 regularisierte Regressionsmodell mit 50% (Sensitivität 27%, Spezifität 67%) schlechter ab. Die höchste Balanced Accuracy bei der Klassifizierung mittelgradiger und hochgradiger Nierenzellkarzinome konnte in venösen CT-Aufnahmen durch das Gradient Boosting Modell mit 85% (mittelgradig) und 97% (hochgradig) erreicht werden. Die Sensitivität und Spezifität lagen bei diesem Modell für mittelgradige Tumoren bei 77% und 92%, für hochgradige Tumoren bei 100% und 93%. Mittelgradige Karzinome konnten anhand venöser CT-Merkmale durch das Random Forest Modell mit 79% Accuracy (Sensitivität 89%, Spezifität 69%) klassifiziert werden. Das k-Nearest Neighbor Modell wies hierbei eine Balanced Accuracy von 73% (Sensitivität 54%, Spezifität 92%) auf, während das L1/L2 regularisierte Regressionsmodell eine Balanced Accuracy von 46% (Sensitivität 69%, Spezifität 23%) aufwies. Das k-Nearest Neighbor Modell konnte bei der Klassifizierung hochgradiger klarzelliger Nierenzellkarzinome eine hohe Balanced Accuracy von 92% (Sensitivität 100%, Spezifität 85%) zeigen. Das L1/L2 regularisierte Regressionsmodell schnitt mit einer Accuracy von 49% (Sensitivität 0%, Spezifität 97%) am schlechtesten ab. Das Random Forest Modell zeigte bei denselben Tumoren in der venösen Phase hingegen eine Balanced Accuracy von 75% (Sensitivität 50%, Spezifität 100%). Die Ergebnisse der statistischen Auswertung des Trainings der vier Modelle des maschinellen Lernens sind in den Tabellen 3 bis 10 dargestellt.

3.2.1 Ergebnisse des Random Forest Modells

Tabelle 3: Nach Tumorgraden und CT-Phasen aufgeteilte Ergebnisse des Random Forest Modells bezüglich der Tumorgradklassifizierung.

Histopathologischer Tumorgrad	CT-Phase	Balanced Accuracy	Sensitivität	Spezifität	Positiver prädiktiver Wert	Negativer prädiktiver Wert
Niedriggradig (WHO/ISUP und Fuhrman G1)	Nativ	0,75	0,50	1,00	1,00	0,89
	Arteriell	0,92	1,00	0,85	0,40	1,00
	Venös	0,81	0,73	0,89	0,67	0,92
Mittelgradig (WHO/ISUP und Fuhrman G2/G3)	Nativ	0,50	0,50	0,50	0,80	0,20
	Arteriell	0,92	0,84	1,00	1,00	0,50
	Venös	0,79	0,89	0,69	0,89	0,69
Hochgradig (WHO/ISUP und Fuhrman G4)	Nativ	NA	NA	NA	NA	NA
	Arteriell	1,00	1,00	1,00	1,00	1,00
	Venös	0,75	0,50	1,00	1,00	0,98

Tabelle 4: Nach CT-Phasen aufgeteilte Ergebnisse des Random Forest Modells bezüglich der Tumorgradklassifizierung.

CT-Phase	Accuracy	Kappa
Nativ	0,5 (0,19 - 0,81)	0,14
Arteriell	0,86 (0,65 - 0,97)	0,61
Venös	0,83 (0,70 - 0,93)	0,60

3.2.2 Ergebnisse des Gradient Boosting Modells

Tabelle 5: Nach Tumorgraden und CT-Phasen aufgeteilte Ergebnisse des Gradient Boosting Modells bezüglich der Tumorgradklassifizierung.

Histopathologischer Tumorgrad	CT-Phase	Balanced Accuracy	Sensitivität	Spezifität	Positiver prädiktiver Wert	Negativer prädiktiver Wert
Niedriggradig (WHO/ISUP und Fuhrman G1)	Nativ	0,94	1,00	0,88	0,67	1,00
	Arteriell	0,65	0,50	0,80	0,20	0,94
	Venös	0,89	0,91	0,86	0,67	0,97
Mittelgradig (WHO/ISUP und Fuhrman G2/G3)	Nativ	0,88	0,75	1,00	1,00	0,50
	Arteriell	0,73	0,79	0,67	0,94	0,33
	Venös	0,85	0,77	0,92	0,96	0,60
Hochgradig (WHO/ISUP und Fuhrman G4)	Nativ	NA	NA	NA	NA	NA
	Arteriell	1,00	1,00	1,00	1,00	1,00
	Venös	0,97	1,00	0,93	0,40	1,00

Tabelle 6: Nach CT-Phasen aufgeteilte Ergebnisse des Gradient Boosting Modells bezüglich der Tumorgradklassifizierung.

CT-Phase	Accuracy	Kappa
Nativ	0,8 (0,44 - 0,97)	0,56
Arteriell	0,77 (0,55 - 0,92)	0,35
Venös	0,81 (0,67 - 0,91)	0,62

3.2.3 Ergebnisse des L1/L2 regularisierten Regressionsmodells

Tabelle 7: Nach Tumorgraden und CT-Phasen aufgeteilte Ergebnisse des L1/L2 regularisierten Regressionsmodells bezüglich der Tumorgradklassifizierung.

Histopathologischer Tumorgrad	CT-Phase	Balanced Accuracy	Sensitivität	Spezifität	Positiver prädiktiver Wert	Negativer prädiktiver Wert
Niedriggradig (WHO/ISUP und Fuhrman G1)	Nativ	0,50	0,00	1,00	NA	0,80
	Arteriell	0,95	1,00	0,90	0,50	1,00
	Venös	0,5	0,27	0,73	0,23	0,77
Mittelgradig (WHO/ISUP und Fuhrman G2/G3)	Nativ	0,50	1,00	0,00	0,80	NA
	Arteriell	0,73	0,79	0,67	0,94	0,33
	Venös	0,46	0,69	0,23	0,71	0,21
Hochgradig (WHO/ISUP und Fuhrman G4)	Nativ	NA	NA	NA	NA	NA
	Arteriell	0,45	0,00	0,90	0,00	0,95
	Venös	0,49	0,00	0,97	0,00	0,98

Tabelle 8: Nach CT-Phasen aufgeteilte Ergebnisse des L1/L2 regularisierten Regressionsmodells bezüglich der Tumorgradklassifizierung.

CT-Phase	Accuracy	Kappa
Nativ	0,8 (0,44 - 0,97)	0,00
Arteriell	0,77 (0,55 - 0,92)	0,35
Venös	0,56 (0,41 - 0,70)	-0,04

3.2.4 Ergebnisse des k-Nearest Neighbor Modells

Tabelle 9: Nach Tumorgraden und CT-Phasen aufgeteilte Ergebnisse des k-Nearest-Neighbor Modells bezüglich der Tumorgradklassifizierung.

Histopathologischer Tumorgrad	CT-Phase	Balanced Accuracy	Sensitivität	Spezifität	Positiver prädiktiver Wert	Negativer prädiktiver Wert
Niedriggradig (WHO/ISUP und Fuhrman G1)	Nativ	0,88	1,00	0,75	0,5	1,0
	Arteriell	0,60	0,50	0,70	0,14	0,93
	Venös	0,77	0,82	0,73	0,47	0,93
Mittelgradig (WHO/ISUP und Fuhrman G2/G3)	Nativ	0,81	0,62	1,00	1,00	0,40
	Arteriell	0,43	0,53	0,33	0,83	0,10
	Venös	0,73	0,54	0,92	0,95	0,43
Hochgradig (WHO/ISUP und Fuhrman G4)	Nativ	NA	NA	NA	NA	NA
	Arteriell	0,43	0,00	0,86	0,00	0,95
	Venös	0,92	1,00	0,85	0,22	1,00

Tabelle 10: Nach CT-Phasen aufgeteilte Ergebnisse des k-Nearest Neighbor Modells bezüglich der Tumorgradklassifizierung.

CT-Phase	Accuracy	Kappa
Nativ	0,7 (0,35 - 0,93)	0,42
Arteriell	0,50 (0,29 - 0,72)	-0,01
Venös	0,62 (0,47 - 0,76)	0,37

4 Diskussion

4.1 Interpretation der Ergebnisse

4.1.1 Alleinstellungsmerkmale der vorliegenden Arbeit

In dieser monozentrischen, retrospektiven Studie wurden vier Modelle des maschinellen Lernens trainiert, um den histopathologischen Tumorgrad klarzelliger Nierenzellkarzinome nach WHO/ISUP beziehungsweise nach Fuhrman vorherzusagen. Hierzu zählten ein Random Forest Modell, ein Gradient Boosting Modell, ein L1/L2 regularisiertes Regressionsmodell und ein k-Nearest-Neighbor Modell. Es fand sich in der Literatur keine vergleichbare Studie, die diese vier Modelle mithilfe eines Datensatzes trainierte.

Für die Radiomics-Analyse wurden 388 klarzellige Nierenzellkarzinome von insgesamt 382 Patienten verwendet. Keine der verfügbaren Studien, die ebenfalls mittels Radiomics die Vorhersage des Tumorgrads klarzelliger Nierenzellkarzinome untersuchten, konnte eine größere Fallzahl aufweisen (87-96). Die Fallzahl der Studien bewegte sich zwischen 54 klarzelligen Nierenzellkarzinomen in der Studie von *Bektas et al.* (89) bis hin zu 357 von *Cui et al.* untersuchten klarzelligen Nierenzellkarzinomen (94). Somit stellt die vorliegende Studie bezüglich des Stichprobenumfangs die bisher umfangreichste Radiomics-Analyse zur Vorhersage des Tumorgrads klarzelliger Nierenzellkarzinome dar.

Der große Stichprobenumfang, die Verwendung der vier Modelle des maschinellen Lernens und die Gruppierung der Tumorgade in drei Subgruppen, die in Abschnitt 4.2.1 diskutiert wird, stellen im Vergleich zu den anderen dieses Thema untersuchenden Studien allesamt Alleinstellungsmerkmale der in dieser Arbeit durchgeführten Radiomics-Analyse dar.

4.1.2 Einordnung der Studie in die Literatur

Im Rahmen der Literaturrecherche konnten zehn Studien gefunden werden, die ebenfalls mittels Radiomics-basierter Analysen von CT-Datensätzen die nicht-invasive Klassifizierung des histopathologischen Tumorgrads klarzelliger Nierenzellkarzinome untersuchten (87-96). Diese Studien sind sowohl untereinander als auch mit der vorliegenden Studie aufgrund methodischer Unterschiede nur eingeschränkt

vergleichbar. Im folgenden Abschnitt erfolgt neben der zusammenfassenden Erläuterung der verfügbaren Literatur eine Einordnung der vorliegenden Arbeit in diesen Kontext.

In ihrer Studie verwendeten *Ding et al.* eine logistische Regression, um den histopathologischen Tumorgrad klarzelliger Nierenzellkarzinome nach Fuhrman zu klassifizieren (90). Ihr Datensatz bestand aus 206 Patienten mit insgesamt 206 klarzelligem Nierenzellkarzinomen. Sie teilten die Tumore in niedriggradig (Fuhrman Grad 1 + 2) und hochgradig (Fuhrman Grad 3 + 4) ein. Insgesamt lagen 145 niedriggradige und 61 hochgradige Tumore vor. Alle Patienten erhielten dreiphasige CT-Aufnahmen (nativ, arteriell, portalvenös) mit vier verschiedenen CT-Scannern. Die Analyse der CT-Datensätze erfolgte ausschließlich anhand von Aufnahmen der arteriellen und der frühvenösen Phasen. Die Tumoren wurden in dieser Studie manuell segmentiert. *Ding et al.* trainierten das Modell mit drei verschiedenen Merkmalskonstellationen. Die Merkmalsselektion erfolgte mittels L1 regularisierter Regression. Für Variante 1 verwendeten sie ausschließlich Merkmale, die nicht zur Gruppe der Texturmerkmale gehören, bei Variante 2 wurden sowohl Texturmerkmalen als auch quantitative Merkmale anderer Ordnungen eingeschlossen und in einer dritten Variante wurden ausschließlich Texturmerkmale ausgewählt. In dieser Studie wurden keine Accuracy-Werte, sondern ausschließlich die AUC-Werte erfasst. Diese ergaben für Modell 1 eine AUC von 67,1%, für Modell 2 eine AUC von 77,1% und für Modell 3 eine AUC von 78%.

Bektas et al. untersuchten in ihrer Studie, inwiefern man mithilfe fünf verschiedener Modelle des maschinellen Lernens hochgradige (Fuhrman Grad 3 + 4) von niedriggradigen (Fuhrman Grad 1 + 2) klarzelligem Nierenzellkarzinomen unterscheiden kann (89). Sie verwendeten für ihre Radiomics-Analyse ebenfalls CT-Aufnahmen, nutzten jedoch nur Texturmerkmale, also quantitative Bildmerkmale zweiter Ordnung. Aus diesem Grund sind die Ergebnisse der Studie von *Bektas et al.* nur eingeschränkt mit den Ergebnissen der vorliegenden Studie vergleichbar. Es wurden 53 Patienten untersucht, die insgesamt 54 nachgewiesene klarzellige Nierenzellkarzinome hatten. Hiervon waren 31 Karzinome niedriggradig und die restlichen 23 Karzinome hochgradig. In der Studie wurden ausschließlich während der portalvenösen Phase aufgenommene CT-Aufnahmen verwendet. Nach der Segmentierung, die in dieser Studie manuell erfolgte, wurden initial 279 Texturmerkmale extrahiert, wovon nach Merkmalsselektion je nach Modell des maschinellen Lernens noch 7-18 Merkmale verblieben. Die Merkmalsselektion erfolgte anhand von deren Reproduzierbarkeit durch Analyse der Segmentierungen zweier verschiedener Annotatoren, anschließend erfolgte eine weitere

Merkmalsselektion durch die Verwendung des "Waikato Environment for Knowledge Analysis (WEKA) Toolkits". Im Rahmen der Analyse wurden fünf verschiedene Modelle des maschinellen Lernens verglichen, nämlich ein Support Vector Machine Modell, ein k-Nearest Neighbor Modell, ein Naïve Bayes Modell, ein Random Forest Modell und ein Multilayer Perceptron Modell. Am erfolgreichsten schnitt die Analyse mittels Support Vector Machine Modell ab. Die Accuracy betrug hierbei 85,1% bei einer Sensitivität von 91,3% und einer Spezifität von 80,6%. Das k-Nearest Neighbor Modell und das Random Forest Modell zeigten Accuracy-Werte von 79,6% (Sensitivität 82,6%, Spezifität 77,4%) beziehungsweise 75,9% (Sensitivität 78,3%, Spezifität 74,2%). An den beiden Studien von *Bektas et al.* und *Ding et al.* lässt sich erkennen, dass mittels der Analyse von ausschließlich Texturmerkmalen gute Ergebnisse bei der Radiomics-basierten Klassifizierung des Tumorgrads klarzelliger Nierenzellkarzinome erzielen lassen (89, 90). Da jedoch nur *Ding et al.* ihr auf Texturmerkmalen basierendes Modell auch mit auf Nicht-Texturmerkmalen basierenden und gemischten Modellen verglichen, und hierbei das gemischte Modell nur geringfügig schlechter als das ausschließlich texturmerkmalbasierte Modell abschnitt, lässt sich nicht abschließend feststellen, mit welcher Art von Kombination von Merkmalsklassen künftig die besten Ergebnisse bei der Klassifizierung des histopathologischen Tumorgrads klarzelliger Nierenzellkarzinome erzielt werden können (90).

Kocak et al. führten eine Studie durch, in der sie versuchten, mittels logistischer Regression und mittels eines Multilayer Perceptron Modells anhand einer CT-Bilddaten-Analyse zwischen niedriggradigen und hochgradigen klarzelligen Nierenzellkarzinomen zu unterscheiden (88). Hierbei wurde ebenfalls das Graduierungssystem nach Fuhrman verwendet und es erfolgte eine Einteilung in hochgradige (Fuhrman Grad 3 + 4) und niedriggradige (Fuhrman Grad 1 + 2) Karzinome. Im Unterschied zur vorliegenden Studie erfolgte die Analyse ausschließlich anhand nativer CT-Aufnahmen. Insgesamt wurden 25 niedriggradige und 56 hochgradige Karzinome von 81 Patienten analysiert. Die CT-Untersuchungen der Tumoren wurden mit unterschiedlichen Scanprotokollen durchgeführt. Auch in dieser Studie erfolgte die Segmentierung der Tumoren manuell. Von anfänglich 744 extrahierten Texturmerkmalen wurden schließlich fünf Merkmale für das Multilayer Perceptron Modell beziehungsweise sechs Merkmale für die logistische Regression selektiert. Die Selektion der Merkmale erfolgte durch eine Reproduzierbarkeitsanalyse anhand der Segmentierungen zweier Annotatoren, durch eine Korrelationsanalyse und schließlich durch Anwendung des "WEKA Toolkits". Die

logistische Regression wies eine Accuracy von 75,3% (Sensitivität 75,3%, Spezifität 53,5%) beziehungsweise 62,5% nach durchgeführtem Oversampling (Sensitivität 62,5%, Spezifität 62,5%) auf. Das Multilayer Perceptron klassifizierte hingegen 81,5% (Sensitivität 81,5%, Spezifität 65,2%) beziehungsweise 70,5% der klarzelligen Nierenzellkarzinome nach Oversampling (Sensitivität 70,5%, Spezifität 70,5%) korrekt. *Lin et al.* trainierten in ihrer Studie ein Gradient Boosting Modell der Open Source Bibliothek "CatBoost", um zwischen niedriggradigen (Fuhrman Grad 1 + 2) und hochgradigen (Fuhrman Grad 3 + 4) klarzelligen Nierenzellkarzinomen zu differenzieren (87). Insgesamt untersuchten sie 231 Patienten mit 232 histologisch gesicherten klarzelligen Nierenzellkarzinomen. Es lagen 189 niedriggradige (Fuhrman G1: 103, Fuhrman G2: 86) und 43 hochgradige (Fuhrman G3: 38, Fuhrman G4: 5) Tumoren vor. Alle Patienten erhielten dreiphasige CT-Aufnahmen mit einem einheitlichen Scanprotokoll. Die Segmentierung der Tumoren erfolgte manuell. Es erfolgte sowohl aus der nativen, der arteriellen und der portalvenösen Phase eine Merkmalsextraktion, wobei für die native Phase 35, für die arterielle Phase 36, für die portalvenöse Phase 41 und phasenübergreifend 22 quantitative Merkmale extrahiert wurden. Die Accuracy des Gradient Boosting Modells lag mit 74% bei der phasenübergreifenden Klassifizierung am höchsten (Sensitivität 67%, Spezifität 88%). Die Accuracy-Werte für die native, arterielle und frühvenöse Phase lagen bei 72% (Sensitivität 64%, Spezifität 86%), 71% (Sensitivität 64%, Spezifität 83%) und 72% (Sensitivität 67%, Spezifität 88%). Insgesamt konnte das Gradient Boosting Modell von *Lin et al.* schlechtere Ergebnisse bei der Klassifizierung der Tumorgrade als das in der vorliegenden Arbeit trainierte Gradient Boosting Modell erzielen. Die Gründe hierfür könnten einerseits die geringere Fallzahl, andererseits die unterschiedliche Wahl des Algorithmus sein.

He et al. führten eine Studie durch, in der sie ein L1 regularisiertes Regressionsmodell zur Vorhersage des histopathologischen Tumorgrads von klarzelligen Nierenzellkarzinomen trainierten (95). In dieser Studie wurde das Graduierungssystem nach WHO/ISUP verwendet. Es erfolgte ebenfalls eine Gruppierung der Grade in niedriggradige (WHO/ISUP G1 + G2) und hochgradige (WHO/ISUP G3 + G4) Karzinome. Die CT-Aufnahmen erfolgten mit zwei verschiedenen Scangeräten in der nativen, der arteriellen und der frühvenösen Phase. Die Studienkohorte bestand aus 227 Patienten mit 227 klarzelligen Nierenzellkarzinomen. Von diesen 227 Tumoren hatten 69 eine G1, 67 eine G2, 39 eine G3 und 52 eine G4 Graduierung nach WHO/ISUP erhalten. Initial wurden 570 quantitative Bildmerkmale aus den CT-Daten extrahiert. Das beste Ergebnis

wurde durch das Training des L1 regularisierten Regressionsmodells mit 19 selektierten Merkmalen in der arteriellen Phase erreicht. Die AUC lag hier bei 98,6% bei einer Sensitivität von 94,5% und einer Spezifität von 92,6%. Die Accuracy des Modells wurde nicht berechnet.

Cui et al. führten eine Radiomics-basierte CT-Bilddaten-Analyse mit insgesamt 367 Patienten durch, um den histopathologischen Tumorgrad vorherzusagen (94). Es wurde das Graduierungssystem nach WHO/ISUP angewandt, wobei zwischen niedriggradigen (WHO/ISUP G1 + G2) und hochgradigen (WHO/ISUP G3 + G4) Fällen unterschieden wurde. Insgesamt lagen 226 niedriggradige und 131 hochgradige Tumoren vor. Die Bildaquisition erfolgte mit drei unterschiedlichen Scangeräten in jeweils der nativen, der arteriellen und der frühvenösen Kontrastmittelphase. Die Segmentierung wurde manuell durch zwei Radiologen durchgeführt, wodurch auf Basis der zwei unterschiedlichen Annotationsmasken eine Reproduzierbarkeitsanalyse der quantitativen Bildmerkmale durchgeführt werden konnte. Die Merkmalsextraktion fand für alle drei Kontrastmittelphasen statt, wobei insgesamt 276 Merkmale extrahiert wurden. Zur weiteren Merkmalsselektion und zum Training des in dieser Studie verwendeten Gradient Boosting Modells wurde "CatBoost" verwendet. Von den initial 276 extrahierten Merkmalen wurden je nach Kontrastmittelphase 81-86 Merkmale selektiert. Die Accuracy des Gradient Boosting Modells lag für die nativen CT-Aufnahmen bei 77% (Sensitivität 77%, Spezifität 77%), die arteriellen und frühvenösen Aufnahmen bei 78% (Sensitivität jeweils 78%, Spezifität 77% beziehungsweise 82%) und unter Einbeziehung aller Aufnahmen bei 79% (Sensitivität 78%, Spezifität 83%). Die Studie von *Cui et al.* hat eine ähnlich hohe Fallzahl wie die vorliegende Studie, verwendete ebenfalls einen Gradient Boosting Algorithmus und konnte hierbei sehr ähnliche Accuracy-Werte erzielen. Ein Vorteil der Studie von *Cui et al.* besteht in der größeren Anzahl der vorliegenden Aufnahmen in den verschiedenen Kontrastmittelphasen, da jeder Patient eine dreiphasige CT-Aufnahme erhalten hatte. Da hierdurch jedoch keine besseren Ergebnisse erzielt werden konnten, stellt sich die Frage, ob Gradient Boosting Modelle bei der Klassifizierung des Tumorgrads klarzelliger Nierenzellkarzinome von höheren Fallzahlen profitieren.

In ihrer Studie verwendeten *Sun et al.* ein Support Vector Machine Modell, um zwischen niedriggradigen (WHO/ISUP G1 + G2) und hochgradigen (WHO/ISUP G3 + 4) klarzelliger Nierenzellkarzinomen zu unterscheiden (93). Die Studienkohorte bestand aus 227 Patienten mit insgesamt 227 histologisch gesicherten klarzelligen

Nierenzellkarzinomen. Es lagen 155 niedriggradige und 72 hochgradige Tumoren vor. Die Bildakquisition erfolgte mit einheitlichem dreiphasigen (nativ, arteriell, frühvenös) Kontrastmittelprotokoll mit stets demselben CT-Gerät. Die Radiomics-Analyse fand jedoch ausschließlich anhand der in der arteriellen und frühvenösen Phase aufgenommenen Scans statt. Die manuelle Segmentierung der Tumoren fand durch zwei Radiologen statt. Nach der Extraktion von 1029 quantitativen Bildmerkmalen wurde mittels ANOVA, SelectKBest und L1 regularisierter Regression eine Merkmalsselektion durchgeführt, wodurch schließlich 5-7 Merkmale je nach Kontrastmittelphase verblieben. Die AUC betrug für das Support Machine Vector Modell in der arteriellen Phase 85% (Sensitivität 85%, Spezifität 86%), in der frühvenösen Phase 78% (Sensitivität 75%, Spezifität 60%) und phasenübergreifend 88% (Sensitivität 85%, Spezifität 89%).

Shu et al. (91) setzten in ihrer Studie ebenfalls das Graduierungssystem nach Fuhrman ein. Auch sie unterteilten die Tumoren in eine niedriggradige (Fuhrman G1 + G2) und eine hochgradige (Fuhrman G3 + G4) Subgruppe, wobei ersterer Gruppe 161 Patienten mit 161 klarzelligen Nierenzellkarzinomen und letzterer Gruppe 99 Patienten mit 99 klarzelligen Nierenzellkarzinomen zugeteilt werden konnten. Die Segmentierung erfolgte manuell und durch zwei Annotatoren. Es erfolgte die anfängliche Extraktion von 1029 Radiomics-Bildmerkmalen. Die Selektion der Merkmale erbrachte 11 Merkmale für die arterielle und 24 Merkmale für die frühvenöse Kontrastmittelphase. Sie erfolgte mittels der Bestimmung des ICC und mittels LASSO. Auf Basis der selektierten Merkmale wurde für die jeweilige Kontrastmittelphase und auch für die Kombination beider Phasen ein logistisches Regressionsmodell mittels 5-facher Kreuzvalidierung trainiert. Die Accuracy betrug für die arterielle Phase 71,9% (Sensitivität 60,2%, Spezifität 83,8%), für die frühvenöse Phase 73,8% (Sensitivität 69,3%, Spezifität 83,8%) und phasenübergreifend 77,7% (Sensitivität 67,7%, Spezifität 83,9%).

In einer weiteren Studie verwendeten *Shu et al.* das WHO/ISUP Graduierungssystem und teilten insgesamt 271 klarzellige Nierenzellkarzinome in eine niedriggradige Subgruppe (WHO/ISUP G1 + G2) und eine hochgradige Subgruppe (WHO/ISUP G3 + G4) (92). Insgesamt lagen 164 niedriggradige und 107 hochgradige klarzellige Nierenzellkarzinome vor. Die Bildakquise erfolgte bei allen Tumoren in der arteriellen und der frühvenösen Phase. Die Annotation der Tumoren wurde manuell durchgeführt. Die Merkmalsextraktion erbrachte initial 1029 Bildmerkmale, erfolgte mittels LASSO und resultierte in 7 ausgewählten Merkmalen. Es wurden fünf verschiedene Modelle des maschinellen Lernens trainiert, und zwar ein k-Nearest-Neighbor Modell, ein logistisches

Regressionsmodell, ein Multilayer Perceptron Modell, ein Random Forest Modell und ein Support Vector Machine Modell. Die Accuracy-Werte betragen für die arterielle Phase, die frühvenöse Phase und die kombinierten Phasen für das k-Nearest Neighbor Modell 83,4%, 84,0% und 86,5%, für das logistische Regressionsmodell 84,0%, 83,4% und 90,2%, für das Multilayer Perceptron Modell 82,8%, 82,8% und 90,8%, für das Random Forest Modell 85,9%, 94,5% und 89,6%, und für das Support Vector Machine Modell 82,2%, 85,9% und 90,8%.

In der Studie von *Nazari et al.* fand das Graduierungssystem nach Fuhrman Anwendung (96). Es wurden 31 Karzinome einer niedriggradigen (Fuhrman Grad 1 + 2) und 31 Karzinome der hochgradigen Gruppe (Fuhrman Grad 3 + 4) zugewiesen. Alle Patienten hatten sich einem dreiphasigen CT-Scan mit einer nativen, einer arteriellen und einer frühvenösen Kontrastmittelphase unterzogen. Die Tumoren wurden manuell segmentiert. Nach anfänglicher Extraktion von 2530 Radiomics-Merkmalen erfolgte die weitere Selektion dieser. Diese erfolgte mittels einer Korrelationsanalyse, mittels LASSO, mittels t-Test und mittels des sogenannten "Minimum Redundancy Maximum Relevance Algorithmus". Es wurden drei verschiedene Modelle des maschinellen Lernens trainiert: Ein logistisches Regressionsmodell, ein Random Forest Modell und ein Support Vector Machine Modell. Hierbei konnten durch das Random Forest Modell ein AUC-Wert von 62%, durch das logistische Regressionsmodell ein AUC-Wert von 78% und durch das Support Vector Machine Modell ein AUC-Wert von 83% erreicht werden. Da die Accuracy der Modelle nicht berechnet wurde, lässt sich diese Studie nur eingeschränkt mit der vorliegenden Arbeit vergleichen. Die geringe AUC des Random Forest Modells von 62% könnte jedoch durch die geringe Fallzahl der Studie von *Nazari et al.* begründet sein.

4.2 Limitationen der vorliegenden Arbeit

4.2.1 Gruppierung der Tumorgade und Klassenimbalance

Da im vorliegenden Datensatz sowohl klarzellige Nierenzellkarzinome, die eine Graduierung nach WHO/ISUP erhalten hatten, als auch solche, die nach Fuhrman graduiert wurden, vorlagen, musste eine Gruppierung der Tumorgade erfolgen. Dies ist darin begründet, dass beide Graduierungssysteme die Karzinome zwar in vier verschiedene Grade einteilen, diese Grade aber nach anderen Kriterien verteilt werden.

Vor allem zwischen den Graden 2 und 3 bestehen bei den beiden Graduierungssystemen größere Diskrepanzen (114-116). Folglich wurde entschieden, alle Tumoren mit den Tumorgraden G2 und G3 gemeinsam zu gruppieren; sie bildeten somit die Gruppe der mittelgradigen klarzelligen Nierenzellkarzinome. Die Tumoren mit dem histopathologischen Tumorgrad G1 nach Fuhrman beziehungsweise G1 nach WHO/ISUP wurden ebenfalls gemeinsam gruppiert, obwohl auch hier teilweise in Studien Diskrepanzen vorlagen (114-116). Alle Karzinome mit dem Tumorgrad G4 nach Fuhrman und G4 nach WHO/ISUP bildeten gemeinsam die Gruppe der hochgradigen klarzelligen Nierenzellkarzinome, wobei hier in den dazu publizierten Studien überwiegend Überschneidungen gefunden worden waren (114-116). Alle anderen verfügbaren vergleichbaren Studien führten ebenfalls eine Gruppierung der Tumorgrade durch, wobei hier stets eine Einteilung in niedriggradige (Grad 1 + 2) beziehungsweise hochgradige (Grad 3 + 4) Tumoren erfolgte (87-96). Keine der Studien verwendete einen Datensatz mit zwei verschiedenen Graduierungssystemen. Sechs der Studien verwendeten das System nach Fuhrman (87-91, 96). Bei vier der Studien fand das WHO/ISUP System Anwendung (92-95). Eine Trennung der Tumoren in zwei Gruppen war in der vorliegenden Studie aufgrund der angesprochenen Diskrepanzen zwischen den Graduierungssystemen nicht sinnvoll durchführbar. Durch die gemeinsame Gruppierung aller Tumoren war folglich zu erwarten, dass ein kleiner Teil der Nierenzellkarzinome der falschen Subgruppe zugeteilt wird, was letztlich die Performance der Modelle des maschinellen Lernens beeinflusst haben könnte. Des Weiteren könnte die bereits angesprochene niedrige Interrater-Reliabilität des Graduierungssystems nach Fuhrman die korrekte Zuteilung zu einer der drei Gruppen erschwert haben (26).

In der vorliegenden Arbeit lagen deutlich mehr Tumoren mit den Tumorgraden G2 und G3 vor. Um diese Unausgewogenheit des Datensatzes auszugleichen und letztlich eine bessere Aussagekraft der Radiomics-Modelle zu erreichen, wurde ein Oversampling durchgeführt. Durch das Verwenden der augmentierten Bildmerkmale aus der Reproduzierbarkeitsanalyse (Abschnitt 2.5) wurde gewährleistet, dass die hinzugefügten Fälle den restlichen Fällen der Klasse ähnelten, jedoch nicht identisch waren. Da die augmentierten Bildmerkmale jedoch trotzdem auf den ursprünglichen Daten basierten, konnte ein Bias dieser Daten in der Radiomics-Analyse durch das Oversampling gegebenenfalls verstärkt worden sein. Beispielsweise könnte ein außergewöhnlicher Befund, der statistisch betrachtet selten auftritt, durch das Oversampling ungewöhnlich häufig im Trainingsdatensatz vorhanden gewesen sein und somit die Performance des

Modells des maschinellen Lernens beeinflusst haben. Aufgrund dieses Umstands stellt das Oversampling eine Limitation der vorliegenden Arbeit dar.

4.2.2 Monozentrität der Studie

Aufgrund der monozentrischen Durchführung der Studie entstammen sowohl die Trainings- als auch die Testdatensätze demselben Datensatz. Hierin könnte insofern eine Limitation bestehen, als dass die Modelle des maschinellen Lernens teilweise zwar gute Vorhersagen in ähnlichen Datensätzen treffen können, diese aber gegebenenfalls nicht auf die CT-Bilddaten anderer Institutionen anwendbar sind. Andererseits war durch den Einschluss aller klarzelliger Nierenzellkarzinome in der Datenbank der Charité, die die Einschlusskriterien der Studie erfüllten, gewährleistet, dass Aufnahmen durch unterschiedliche CT-Scangeräte und mit abweichenden Kontrastmittellprotokollen im Datensatz der vorliegenden Arbeit vorhanden waren. Die hierdurch bedingte Heterogenität des Datensatzes könnte zwar einerseits die Radiomics-Analyse erschwert haben und somit zu einem höheren Bias geführt haben, andererseits könnte hierdurch die Varianz verringert worden sein und somit die Übertragbarkeit auf unbekanntere Datensätze erhöht worden sein. *Larue et al.* stellten fest, dass sowohl das verwendete CT-Gerät als auch die Schichtdicke Auswirkungen auf die Ausprägung von Radiomics-Bildmerkmalen haben (117). In der Literatur fanden sich bezüglich der Tumorgradbestimmung klarzelliger Nierenzellkarzinome vier Studien, die CT-Aufnahmen von einem CT-Scangerät mit stets denselben Scanprotokollen verwendeten (87, 89, 91, 93). Sechs weitere Studien verwendeten unterschiedliche CT-Scangeräte und verfügten somit über heterogenere Datensätze (88, 90, 92, 94-96).

Um Radiomics-basierte Analysen langfristig in den Klinikalltag einführen zu können, empfiehlt es sich, die jeweils trainierten Modelle des maschinellen Lernens auch auf unbekannteren Datensätzen aus anderen Institutionen zu testen. Multizentrische Studien sind daher in Zukunft sinnvoll, um die Performance der Radiomics-basierten Analysen zu optimieren und diese somit gegebenenfalls für die nicht-invasive Tumordiagnostik im klinischen Alltag verwenden zu können (68). Dies ist jedoch mit einem hohen organisatorischen und finanziellen Aufwand verbunden.

4.2.3 Manuelle Segmentierung

Eine weitere Limitation der vorliegenden Arbeit besteht in der manuellen Segmentierung der Nierenzellkarzinome durch nur eine Person. Die Segmentierung im Rahmen einer Radiomics-Analyse gestaltet sich bei Raumforderungen der Niere vor allem in CT-Aufnahmen als aufwändig (67). Dies ist unter anderem darin begründet, dass diese oft eine sehr heterogene Form besitzen, sich teilweise nur schwach vom umliegenden Nierenparenchym abheben und in einigen Fällen auch Gefäße infiltrieren, in denen oft keine klare Abgrenzung zwischen Tumor und Gefäßlumen möglich ist (67). Durch die erfolgte Simulation zweier weiterer Annotatoren wurde jedoch versucht, nur die quantitativen Merkmale in die Analyse einzuschließen, die trotz veränderter Segmentationsmasken reproduzierbar waren. In drei Studien, die ebenfalls eine Radiomics-Analyse zur Prädiktion des histopathologischen Tumorgrads klarzelliger Nierenzellkarzinome durchführten, fand die Segmentierung durch zwei Radiologen statt (91, 93, 94). Dies könnte insofern von Vorteil sein, als dass durch teilweise noch stärkere und zufällige Abweichungen der verschiedenen Segmentationsmasken die anhand der Reproduzierbarkeit durchgeführte Merkmalsselektion zielführendere Ergebnisse bringen könnte. Eine Simulation hat gegenüber der tatsächlichen erneuten Segmentierung den Vorteil, dass hierdurch Personal und somit wertvolle Arbeitszeit eingespart werden kann, was den Prozess der Radiomics-Analyse insgesamt beschleunigt.

Die manuelle Segmentierung der Tumoren nahm in der vorliegenden Studie circa 70 Minuten für jeden Patienten in Anspruch. Ein großer Zeitaufwand bei diesem Teilschritt der Radiomics-Analyse stellt für künftige Studien mit größeren Stichprobenumfängen eine Einschränkung dar. Darüber hinaus scheint ein solch hoher zeitlicher Aufwand im klinischen Alltag nicht praktikabel. Bei der Ermöglichung einer routinemäßigen Anwendung von Radiomics oder auch bei der Durchführung größerer Studien könnten folglich automatisierte Segmentierungen eine entscheidende Rolle spielen, da somit der durch die manuelle Segmentierung verursachte Zeitaufwand nahezu wegfiel. Mithilfe von Deep Learning-Algorithmen konnten *Isensee et al.* beispielsweise eine automatisierte Segmentierung von Nierenzellkarzinomen und Nierengewebe mit einem zusammengesetzten Dice-Score von 91,2 entwickeln, der eine hohe Übereinstimmung zwischen manuellen Segmentationen von Tumoren und Nieren und den automatisierten Segmentationen anzeigt (118). Hierbei verwendeten sie die sogenannte U-Net-Architektur, also ein für die Segmentierung von medizinischer Bildgebung entwickeltes

neuronaales Netzwerk, das momentan weltweit die besten Ergebnisse bei der Entwicklung von automatisierten Segmentierungsalgorithmen erreicht (67, 118).

4.2.4 Verwendung von ein- bis dreiphasigen CT-Aufnahmen

In der vorliegenden Arbeit wurden Patienten eingeschlossen, wenn sie mindestens eine CT-Aufnahme in entweder einer nativen, einer arteriellen oder einer venösen Kontrastmittelphase erhalten hatten. Hierdurch wurde bedingt, dass einphasige, zweiphasige und dreiphasige CT-Aufnahmen vorlagen, was wiederum zur Folge hatte, dass sich die Fallzahl je nach Kontrastmittelphase unterschied. Somit lagen insgesamt 120 klarzellige Nierenzellkarzinome in der nativen CT-Phase, 207 in der arteriellen Phase und 327 in der venösen Phase vor. Folglich wurde für jede Kontrastmittelphase separat eine Analyse durchgeführt.

Des Weiteren wurden sowohl die während der frühvenösen Kontrastmittelphase als auch die während der spätvenösen Phase aufgenommenen CT-Bilder als venöse CT-Aufnahmen zusammengefasst. Dies könnte gegebenenfalls aufgrund unterschiedlicher Merkmalsausprägungen das Training der Modelle beeinflusst haben und somit eine weitere Limitation der hier durchgeführten Radiomics-Analyse darstellen. Darüber hinaus wird hierdurch auch die Vergleichbarkeit zu anderen Studien, die oftmals ausschließlich frühvenöse CT-Aufnahmen, nicht jedoch spätvenöse CT-Aufnahmen verwendeten, eingeschränkt (87, 89-96).

4.2.5 Einschluss biopsierter Tumoren

In der vorliegenden Arbeit wurden 12 klarzellige Nierenzellkarzinome eingeschlossen, deren histopathologischer Tumorgrad nicht mittels der histopathologischen Untersuchung eines Tumor(teil)resektats, sondern mittels der Untersuchung einer Stanzzyylinderbiopsie bestimmt wurde. Da die Bestimmung des Tumorgrads anhand einer Biopsie ungenauer ist als die Untersuchung anhand eines Resektats, könnten hierdurch gegebenenfalls Tumoren mit einem fälschlicherweise zu hohen oder einem fälschlicherweise zu niedrigen Tumorgrad verwendet worden sein (39, 67). Da diese Tumoren sowohl Teil des Trainings- als auch des Validierungsdatensatzes waren, könnte dies das Training der Modelle des maschinellen Lernens und somit auch deren Ergebnisse negativ beeinflusst haben.

4.3 Möglichkeiten und Grenzen von Radiomics - ein Ausblick

4.3.1 Weitere Anwendungen von Radiomics in der computertomographischen Diagnostik des klarzelligen Nierenzellkarzinoms

Neben der nicht-invasiven Bestimmung des histopathologischen Tumorgrads eröffnet die Verwendung von Radiomics noch weitere Möglichkeiten in der computertomographischen Diagnostik klarzelliger Nierenzellkarzinome.

Radiomics-Analysen eignen sich unter anderem zur Unterscheidung zwischen benignen und malignen renalen Raumforderungen (119). In ihrer Studie konnten *Sun et al.* zum Beispiel demonstrieren, dass mithilfe des von ihnen entwickelten, auf maschinellem Lernen basierenden Radiomics-Modells in 86% der Fälle in CT-Aufnahmen zwischen klarzelligen Nierenzellkarzinomen und atypischen Angiomyolipomen und Onkozytomen differenziert werden konnte (119). Mit einer Accuracy von circa 80% konnten auf diese Weise außerdem chromophobe und papilläre Nierenzellkarzinome von atypischen Angiomyolipomen und Onkozytomen unterschieden werden (119). Hierbei wurde festgestellt, dass vier erfahrene Radiologen, die ebenfalls zwischen benignen und malignen Raumforderungen unterscheiden sollten, hier teilweise eine geringere Accuracy erzielten und insgesamt eine hohe Interrater-Reliabilität aufwiesen (119). Durch Zuhilfenahme von Radiomics-basierten Bilddatenanalysen lässt sich diese Interrater-Reliabilität in Zukunft gegebenenfalls reduzieren (119). Durch *Zhou et al.* konnte ein Radiomics-Modell entwickelt werden, das in 97% der Fälle anhand von kontrastmittelgestützten CT-Aufnahmen klarzellige Nierenzellkarzinome von Onkozytomen unterscheiden konnte (120).

Darüber hinaus lassen sich Radiomics-Analysen auch zur Differenzierung zwischen verschiedenen Subtypen des Nierenzellkarzinoms verwenden (67). So konnten *Sun et al.* diesbezüglich zeigen, dass durch das von ihnen entwickelte Modell in CT-Aufnahmen in circa 90% der Fälle erfolgreich zwischen klarzelligen Nierenzellkarzinomen und nicht-klarzelligen Nierenzellkarzinomen differenziert werden konnte (119).

4.3.2 Radiogenomics

Die Korrelation von Radiomics-Merkmalen mit genetischen Mustern wird auch als Radiogenomics bezeichnet (69). Radiogenomics ermöglicht es, anhand der Untersuchung der quantitativen Bildmerkmale eines Tumors Aussagen über die

Wahrscheinlichkeit einer vorliegenden Genexpression beziehungsweise vorliegenden Mutationen in diesem Tumor zu treffen (69). Radiogenomics könnte somit in Zukunft die Kombination radiologischer und genetischer Daten erlauben, was neue Möglichkeiten bei der Behandlung von Patienten eröffnen könnte (121).

Da sich die Genexpression von Tumoren ändern kann und hierdurch Resistenzen gegenüber zielgerichteten Therapieformen entstehen können, könnte Radiogenomics durch eine nicht-invasive Untersuchung der Genexpression zu einer adäquaten Tumorthherapie beitragen (69, 122). Der Informationsgehalt der radiologischen Bilddatensätze ist bei Radiogenomics-Analysen jedoch auch begrenzt, sodass nicht jede genetische Fragestellung allein anhand der quantitativen Analyse der Bilddatensätze beantwortet werden kann (39).

In ihrer Studie konnten *Jamshidi et al.* zum Beispiel einen Radiogenomics-basierten, aus vier quantitativen Bildmerkmalen bestehenden Risikoscore entwickeln, der das krankheitsspezifische Überleben von Patienten mit klarzelligen Nierenzellkarzinomen abschätzt (123).

4.3.3 Zukünftige Herausforderungen von Radiomics

Eine große Herausforderung für künftige Radiomics-Analysen stellt beispielsweise die Standardisierung der Bildakquise dar. Bereits die Positionierung der Patienten und die Verwendung unterschiedlicher Rekonstruktionsalgorithmen kann beispielsweise Einfluss auf die extrahierten quantitativen Bildmerkmale und somit die Qualität der Radiomics-Studie haben (121). Darüber hinaus werden aktuell in verschiedenen Kliniken unterschiedliche CT-Scangeräte, Rekonstruktionsalgorithmen und Kontrastmittelprotokolle verwendet, was ebenfalls zur Variabilität der Merkmalsextraktion beiträgt. Eine Standardisierung der Bildakquise ist jedoch mit hohen Kosten und einem hohen organisatorischen Aufwand verbunden.

Auch die Standardisierung der verwendeten Bildmerkmale in Radiomics-Analysen kann zur Vergleichbarkeit und somit zur Qualität von Radiomics-Studien beitragen. Dies könnte die klinische Einführung der Algorithmen erleichtern. So erarbeiteten *Zwanenburg et al.* im Rahmen ihrer *Image Biomarker Standardization Initiative* die Standardisierung von insgesamt 169 Radiomics-Merkmalen (124).

Des Weiteren stellt auch die Verknüpfung von Radiomics-Algorithmen mit den in Krankenhäusern üblicherweise verwendeten Picture archiving and communication (PACS) - Plattformen eine Herausforderung dar (125). Diese Verknüpfung würde

einerseits die Entwicklung von Algorithmen vereinfachen, da der Datenzugriff erleichtert würde, andererseits könnte hiermit auch die Grundlage für die Integration von Radiomics-Algorithmen in die Prozesse der klinischen Entscheidungsfindung gelegt werden (125).

4.3.4 Deep Learning

Es wurde bereits festgestellt, dass Deep Learning für die automatisierte Tumorsegmentierung künftig eine entscheidende Rolle spielen könnte. Darüber hinaus könnten in Zukunft durch die Verwendung von Deep Learning beispielsweise bessere Ergebnisse bei der Klassifizierung von histopathologischen Tumorgraden als durch klassische Algorithmen des maschinellen Lernens erzielt werden. Letztere zeigen zwar allgemein bei vielen medizinischen Problemen gute Leistungen, diese reichen jedoch oftmals nicht für eine klinische Anwendung aus (126). Es wird jedoch angenommen, dass Deep Learning Algorithmen die Leistungen der klassischen Modelle des maschinellen Lernens künftig übertreffen und somit auch die Einführung dieser auf künstlicher Intelligenz basierenden Methoden in den Klinikalltag ermöglichen können (126).

Deep Learning ist ein Teilbereich des Maschinellen Lernens, der künstliche neuronale Netzwerke mit multiplen Schichten nutzt (127). Ein klassisches neuronales Netzwerk, das sogenannte feed-forward-Netzwerk, besteht aus einer Eingangsschicht, multiplen Zwischenschichten und einer Ausgangsschicht (127). Die Informationsverarbeitung erfolgt hierbei von der Eingangsschicht zur Ausgangsschicht (127).

Deep Learning-Algorithmen können umfangreiche und komplexe Bilddatensätze ohne ausgiebige Vorverarbeitung der Daten effizient klassifizieren (128). Beim herkömmlichen Maschinellen Lernen ist hierfür hingegen oft eine manuelle, umfangreiche Vorverarbeitung der zu analysierenden Merkmale notwendig (128). Für das Training von Deep Learning Modellen ist das Vorliegen von großen Datensätzen obligat (126). Das Training der Modelle dauert in der Regel wesentlich länger als das Training von Modellen des maschinellen Lernens und erfordert dabei deutlich mehr Rechenleistung. Eine tiefgehende Erläuterung der Funktionsweise von Deep Learning Modellen kann der entsprechenden Literatur entnommen werden.

Ein sehr simples neuronales Netzwerk, das sogenannte Multilayer Perceptron, wurde in den bereits diskutierten Studien von *Bektas et al.*, *Kocak et al* und *Lin et al.* verwendet, konnte jedoch im Vergleich zu den in denselben Datensätzen ebenfalls getesteten Modellen des maschinellen Lernens keine deutlich höhere Performance erzielen (88, 89, 92).

Allgemein lassen sich Deep Learning Algorithmen in der Medizin gut und vielseitig einsetzen, da oftmals sehr große Datensätze vorliegen (127). So konnte von *Choi et al.* demonstriert werden, dass durch die Analyse von digitalisierten Patientendaten, die unter anderem Informationen wie Vorerkrankungen und Medikation enthielten, eine Herzinsuffizienz früher diagnostiziert werden kann (129). Durch eine frühere Detektion der Herzinsuffizienz könnte diese rascher therapiert werden und eine Progression somit verlangsamt werden (129). Hierdurch könnten auch Behandlungskosten gespart werden (129). Des Weiteren konnte gezeigt werden, dass Deep Learning Algorithmen zum Beispiel bei der Früherkennung des Morbus Alzheimer helfen können (130). Dies würde das frühzeitige Ergreifen von Präventionsmaßnahmen erlauben, wodurch das Fortschreiten der Erkrankung verlangsamt werden könnte (130). In einer weiteren Studie konnte gezeigt werden, dass durch den Einsatz von Deep Learning Algorithmen das Vorliegen Mamma-Karzinomen anhand von digitalisierten Gewebepräparaten mit einer hohen Genauigkeit bestimmt werden kann (131).

Auch in den Fachgebieten Radiologie und Onkologie gibt es zahlreiche Einsatzmöglichkeiten für Deep Learning Algorithmen (63). Beispielsweise konnten *Mohsen et al.* demonstrieren, dass Gehirntumoren in MRT-Aufnahmen mittels Deep Learning in die verschiedenen Tumorentitäten eingeteilt werden können (132). Hierbei wurde eine AUC von 98,4% erreicht (132). In einer weiteren Studie von *Nie et al.*, in der ebenfalls MRT-Aufnahmen analysiert wurden, konnte gezeigt werden, dass die mittlere Überlebenszeit von Patienten mit hochgradigen Gliomen durch die Zuhilfenahme von Deep Learning Algorithmen mit Accuracy-Werten von bis zu 89,85% vorhergesagt werden kann (133).

Ein Nachteil von Deep Learning Modellen ist deren mangelnde Transparenz, weshalb in diesem Zusammenhang auch von sogenannter Black-Box-Medizin gesprochen wird (126). Es ist bei Deep Learning Modellen kaum möglich, einzelne Bildmerkmale zu identifizieren, die einen besonders hohen Einfluss auf die Klassifizierung der jeweiligen Objekte haben (126). Die Entscheidungswege der Modelle sind nur sehr begrenzt bis gar nicht nachvollziehbar (126). Bei fehlerhaften Entscheidungen gestaltet es sich daher auch schwierig, bestimmte Fehlerquellen ausfindig zu machen (126). Diese Intransparenz nimmt mit ansteigender Netzwerktiefe zu (126).

Dies bringt auch Komplikationen bezüglich der gesetzlichen Regulierung von Deep Learning Systemen mit sich. Ein Problem könnte beispielsweise die Tatsache darstellen, dass selbst nach der Markteinführung eines Deep Learning Systems dieses noch

weiterhin durch Analyse neuerer Daten lernt und dessen Entscheidungsgenauigkeit sich somit verbessern, aber temporär natürlich auch verschlechtern kann (126). Eine solche Verschlechterung des Modells könnte beispielsweise negative Konsequenzen bezüglich einer Therapieentscheidung nach sich ziehen, was ein ethisches Problem darstellt. Ein weiteres Problem könnte darin bestehen, den Verursacher einer falschen Diagnose oder falschen Therapieentscheidung zu identifizieren. Es liegt nicht auf der Hand, inwiefern das medizinische Personal, das den Algorithmus verwendet, oder die Entwickler des jeweiligen Algorithmus hierfür rechtlich belangt werden könnten.

Auch hinsichtlich des Datenschutzes ist die Einführung von Deep Learning gestützten Radiomics-Analysen problematisch (126). Für die Entwicklung der Algorithmen werden zahlreiche Daten benötigt, die oftmals auch in verschiedenen klinischen Zentren gespeichert sind und deren Weitergabe für Forschungszwecke nicht nur mit einem erheblichen bürokratischen und organisatorischem Aufwand verbunden ist, sondern auch aus datenschutzrechtlichen Gründen teilweise unmöglich ist (126). Ein möglicher Lösungsansatz hierfür könnte sein, dass verschiedene Zentren am Training desselben Algorithmus teilnehmen können, ohne dafür zwingend ihre Trainingsdaten, die in der Regel aus sensiblen Patientendaten bestehen, bereitstellen zu müssen (126). Stattdessen muss nur das jeweils trainierte Modell bereitgestellt werden (126). Ein weiterer Ansatz sind sogenannte CryptoNets (134). Deep Learning Netzwerke können hierbei in verschlüsselte Netzwerke, CryptoNets, umgewandelt werden (134). Diese sind in der Lage, verschlüsselte Daten zu verarbeiten (134). Hierdurch können verschlüsselte Patientendaten in Cloudservices geteilt werden, um dort mittels der jeweiligen Algorithmen analysiert zu werden (134). Die Analyseergebnisse können dann erneut verschlüsselt weitergeleitet werden und nur mit bestimmten Passwörtern entschlüsselt werden, was in Zukunft eine datenschutzkonforme Entwicklung von Deep Learning Algorithmen erlauben könnte (134).

Aktuell herrscht kein Konsens darüber, wie lange die Implementierung von vollautomatisierten Deep Learning Modellen in den radiologischen Klinikalltag dauern wird. Hierbei bewegen sich die spekulierten Zeiträume zwischen wenigen Jahren bis hin zu mehreren Jahrzehnten (126).

4.3.5 Deep Learning zur nicht-invasiven Bestimmung des histopathologischen Tumorgrads klarzelliger Nierenzellkarzinome

Im Rahmen der Literaturrecherche fand sich eine Studie von *Lin et al.*, die untersuchte, ob ein Deep Learning Modell den histopathologischen Tumorgrad klarzelliger Nierenzellkarzinome mittels der Analyse von CT-Aufnahmen vorhersagen kann (135). Sie verwendeten hierbei das WHO/ISUP-Graduierungssystem und teilten den Datensatz in niedriggradige (WHO/ISUP G1 + G2) und hochgradige (WHO/ISUP G3 + G4) Tumoren ein. Insgesamt umfasste die Studie 410 Karzinome, von denen 261 niedriggradig und 149 hochgradig waren. Neben einem lokalen Testdatensatz, der aus 20 Tumoren bestand, lag auch ein externer, ebenfalls 20 Tumoren umfassender, Testdatensatz vor. Die Analyse erfolgte anhand dreiphasiger CT-Aufnahmen, wobei jeweils eine native, eine arterielle und eine frühvenöse Aufnahme vorliegen mussten. Das Deep Learning Netzwerk ResNet wurde in verschiedenen Variationen für die Klassifizierung der Tumoren verwendet. Hierbei konnten Accuracy-Werte von bis zu 72,5% erreicht werden, was im Vergleich zu den auf maschinellem Lernen basierenden Studien ein vergleichbares, jedoch kein besseres Ergebnis darstellt. Da die Fallzahl für das Training eines Deep Learning Modells als gering einzustufen ist, kann davon ausgegangen werden, dass Studien mit einem größeren Stichprobenumfang bessere Ergebnisse bei der Klassifizierung des Tumorgrads erreichen können.

Es sind folglich weitere umfangreiche, multizentrische und gegebenenfalls prospektive Studien erforderlich, um die Ergebnisse von Radiomics-Modellen bei der Klassifizierung des histopathologischen Tumorgrads klarzelliger Nierenzellkarzinome weiter zu verbessern. Der Einsatz von Deep Learning könnte hierzu gegebenenfalls entscheidend beitragen und den Weg für die Einführung von Radiomics-Modellen in den klinischen Alltag ebnen.

Literaturverzeichnis

1. Hsieh JJ, Purdue MP, Signoretti S, Swanton C, Albiges L, Schmidinger M, Heng DY, Larkin J, Ficarra V. Renal cell carcinoma. Nature reviews Disease primers. 2017;3(1):1-19.
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2021.
3. Zentrum für Krebsregisterdaten im Robert Koch-Institut: Datenbankabfrage mit Schätzung der Inzidenz, Prävalenz und des Überlebens von Krebs in Deutschland auf Basis der epidemiologischen Landeskrebsregisterdaten. Mortalitätsdaten bereitgestellt vom Statistischen Bundesamt. [Internet]. [Date accessed 14.03.2021]. Available from: www.krebsdaten.de/abfrage.
4. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International variations and trends in renal cell carcinoma incidence and mortality. European urology. 2015;67(3):519-30.
5. Nepple KG, Yang L, Grubb RL, Strobe SA. Population based analysis of the increasing incidence of kidney cancer in the United States: evaluation of age specific trends from 1975 to 2006. The Journal of urology. 2012;187(1):32-8.
6. Hunt JD, Van Der Hel OL, McMillan GP, Boffetta P, Brennan P. Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. International journal of cancer. 2005;114(1):101-8.
7. Yuan J, Castela J, Gago-Dominguez M, Ross R, Yu M. Hypertension, obesity and their medications in relation to renal cell carcinoma. British journal of cancer. 1998;77(9):1508-13.
8. Ljungberg B, Campbell SC, Cho HY, Jacqmin D, Lee JE, Weikert S, Kiemeny LA. The epidemiology of renal cell carcinoma. European urology. 2011;60(4):615-21.
9. Moch H. The WHO/ISUP grading system for renal carcinoma. Der Pathologe. 2016;37(4):355-60.
10. Escudier B, Porta C, Schmidinger M, Rioux-Leclercq N, Bex A, Khoo V, Grünwald V, Gillessen S, Horwich A. Renal cell carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann Oncol. 2019;30(5):706-20.
11. Network CGAR. Comprehensive molecular characterization of papillary renal-cell carcinoma. New England Journal of Medicine. 2016;374(2):135-45.

12. Srigley JR, Delahunt B, Eble JN, Egevad L, Epstein JI, Grignon D, Hes O, Moch H, Montironi R, Tickoo SK, Zhou M, Argani P, Panel TIRT. The International Society of Urological Pathology (ISUP) Vancouver Classification of Renal Neoplasia. *The American Journal of Surgical Pathology*. 2013;37(10):1469-89.
13. Israel GM, Bosniak MA. How I do it: evaluating renal masses. *Radiology*. 2005;236(2):441-50.
14. Eknayan G. A clinical view of simple and complex renal cysts. *Journal of the American Society of Nephrology*. 2009;20(9):1874-6.
15. Carrim ZI, Murchison JT. The prevalence of simple renal and hepatic cysts detected by spiral computed tomography. *Clinical radiology*. 2003;58(8):626-9.
16. Kissane JM. The morphology of renal cystic disease. *Perspectives in nephrology and hypertension*. 1976;4:31-63.
17. Lopez-Beltran A, Scarpelli M, Montironi R, Kirkali Z. 2004 WHO classification of the renal tumors of the adults. *European urology*. 2006;49(5):798-805.
18. Nelson CP, Sanda MG. Contemporary diagnosis and management of renal angiomyolipoma. *The Journal of urology*. 2002;168(4):1315-25.
19. Fuhrman SA, Lasky LC, Limas C. Prognostic significance of morphologic parameters in renal cell carcinoma. *The American journal of surgical pathology*. 1982;6(7):655-63.
20. Becker A, Hickmann D, Hansen J, Meyer C, Rink M, Schmid M, Eichelberg C, Strini K, Chromecki T, Jesche J. Critical analysis of a simplified Fuhrman grading scheme for prediction of cancer specific mortality in patients with clear cell renal cell carcinoma—Impact on prognosis. *European Journal of Surgical Oncology (EJSO)*. 2016;42(3):419-25.
21. Delahunt B, Egevad L, Samaratunga H, Martignoni G, Nacey JN, Srigley JR. Gleason and Fuhrman no longer make the grade. *Histopathology*. 2016;68(4):475-81.
22. Rioux-Leclercq N, Karakiewicz PI, Trinh QD, Ficarra V, Cindolo L, de la Taille A, Tostain J, Zigeuner R, Mejean A, Patard JJ. Prognostic ability of simplified nuclear grading of renal cell carcinoma. *Cancer: Interdisciplinary International Journal of the American Cancer Society*. 2007;109(5):868-74.
23. Dagher J, Delahunt B, Rioux-Leclercq N, Egevad L, Srigley JR, Coughlin G, Duglinson N, Gianduzzo T, Kua B, Malone G. Clear cell renal cell carcinoma: validation of World Health Organization/International Society of Urological Pathology grading. *Histopathology*. 2017;71(6):918-25.

24. Delahunt B, Sika-Paotonu D, Bethwaite PB, McCredie MR, Martignoni G, Eble JN, Jordan TW. Fuhrman grading is not appropriate for chromophobe renal cell carcinoma. *The American journal of surgical pathology*. 2007;31(6):957-60.
25. Delahunt B. Advances and controversies in grading and staging of renal cell carcinoma. *Modern Pathology*. 2009;22(2):S24-S36.
26. Al-Aynati M, Chen V, Salama S, Shuhaibar H, Treleaven D, Vincic L. Interobserver and intraobserver variability using the Fuhrman grading system for renal cell carcinoma. *Archives of pathology & laboratory medicine*. 2003;127(5):593-6.
27. Delahunt B, Cheville JC, Martignoni G, Humphrey PA, Magi-Galluzzi C, McKenney J, Egevad L, Algaba F, Moch H, Grignon DJ. The International Society of Urological Pathology (ISUP) grading system for renal cell carcinoma and other prognostic parameters. *The American journal of surgical pathology*. 2013;37(10):1490-504.
28. Delahunt B, Eble JN, Egevad L, Samaratunga H. Grading of renal cell carcinoma. *Histopathology*. 2019;74(1):4-17.
29. Tsui K-H, SHVARTS O, SMITH RB, FIGLIN RA, deKERNION JB, BELLDEGRUN A. Prognostic indicators for renal cell carcinoma: a multivariate analysis of 643 patients using the revised 1997 TNM staging criteria. *The Journal of urology*. 2000;163(4):1090-5.
30. Motzer RJ, Mazumdar M, Bacik J, Berg W, Amsterdam A, Ferrara J. Survival and prognostic stratification of 670 patients with advanced renal cell carcinoma. *Journal of clinical oncology*. 1999;17(8):2530-.
31. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, Weaver AL, Parker AS, Zincke H. Prediction of progression after radical nephrectomy for patients with clear cell renal cell carcinoma: a stratification tool for prospective clinical trials. *Cancer: Interdisciplinary International Journal of the American Cancer Society*. 2003;97(7):1663-71.
32. AWMF. S3-Leitlinie Diagnostik, Therapie und Nachsorge des Nierenzellkarzinoms, AWMF-Registernummer: 043/017OL, Kurzversion 2.02020.
33. Makhov P, Joshi S, Ghatalia P, Kutikov A, Uzzo RG, Kolenko VM. Resistance to systemic therapies in clear cell renal cell carcinoma: mechanisms and management strategies. *Molecular cancer therapeutics*. 2018;17(7):1355-64.
34. Gupta K, Miller JD, Li JZ, Russell MW, Charbonneau C. Epidemiologic and socioeconomic burden of metastatic renal cell carcinoma (mRCC): a literature review. *Cancer treatment reviews*. 2008;34(3):193-205.

35. Tsui K-H, Shvarts O, Smith RB, FIGLIN R, de KERNION JB, BELLDEGRUN A. Renal cell carcinoma: prognostic significance of incidentally detected tumors. *The Journal of urology*. 2000;163(2):426-30.
36. Jayson M, Sanders H. Increased incidence of serendipitously discovered renal cell carcinoma. *Urology*. 1998;51(2):203-5.
37. Gold P, Fefer A, Thompson J, editors. *Paraneoplastic manifestations of renal cell carcinoma*. Seminars in urologic oncology; 1996.
38. Marconi L, Dabestani S, Lam TB, Hofmann F, Stewart F, Norrie J, Bex A, Bensalah K, Canfield SE, Hora M. Systematic review and meta-analysis of diagnostic accuracy of percutaneous renal tumour biopsy. *European urology*. 2016;69(4):660-73.
39. Scheckenbach K. Radiomics: Big Data statt Biopsie in der Zukunft? *Laryngo-Rhino-Otologie*. 2018;97(S 01):S114-S41.
40. Yuh BI, Cohan RH. Different phases of renal enhancement: role in detecting and characterizing renal masses during helical CT. *AJR American journal of roentgenology*. 1999;173(3):747-55.
41. Zhang J, Lefkowitz RA, Ishill NM, Wang L, Moskowitz CS, Russo P, Eisenberg H, Hricak H. Solid renal cortical tumors: differentiation with CT. *Radiology*. 2007;244(2):494-504.
42. Kim JK, Kim TK, Ahn HJ, Kim CS, Kim K-R, Cho K-S. Differentiation of subtypes of renal cell carcinoma on helical CT scans. *American Journal of Roentgenology*. 2002;178(6):1499-506.
43. Pooler BD, Pickhardt PJ, O'Connor SD, Bruce RJ, Patel SR, Nakada SY. Renal cell carcinoma: attenuation values on unenhanced CT. *American Journal of Roentgenology*. 2012;198(5):1115-20.
44. O'Connor SD, Pickhardt PJ, Kim DH, Oliva MR, Silverman SG. Incidental finding of renal masses at unenhanced CT: prevalence and analysis of features for guiding management. *American Journal of Roentgenology*. 2011;197(1):139-45.
45. Simpson E, Patel U. Diagnosis of angiomyolipoma using computed tomography—region of interest ≤ -10 HU or 4 adjacent pixels ≤ -10 HU are recommended as the diagnostic thresholds. *Clinical radiology*. 2006;61(5):410-6.
46. Coulam CH, Sheafor DH, Leder RA, Paulson EK, DeLong DM, Nelson RC. Evaluation of pseudoenhancement of renal cysts during contrast-enhanced CT. *American Journal of Roentgenology*. 2000;174(2):493-8.

47. Maki DD, Birnbaum BA, Chakraborty DP, Jacobs JE, Carvalho BM, Herman GT. Renal cyst pseudoenhancement: beam-hardening effects on CT numbers. *Radiology*. 1999;213(2):468-72.
48. Hounsfield GN. Computerized transverse axial scanning (tomography): Part 1. Description of system. *The British journal of radiology*. 1973;46(552):1016-22.
49. Kalender WA. X-ray computed tomography. *Physics in medicine & Biology*. 2006;51(13):R29.
50. Kalender WA, Seissler W, Klotz E, Vock P. Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation. *Radiology*. 1990;176(1):181-3.
51. Silverman J, Paul N, Siewerdsen J. Investigation of lung nodule detectability in low-dose 320-slice computed tomography. *Medical physics*. 2009;36(5):1700-10.
52. Flohr TG, McCollough CH, Bruder H, Petersilka M, Gruber K, Süß C, Grasruck M, Stierstorfer K, Krauss B, Raupach R. First performance evaluation of a dual-source CT (DSCT) system. *European radiology*. 2006;16(2):256-68.
53. Johnson TR. Dual-energy CT: general principles. *American Journal of Roentgenology*. 2012;199(5_supplement):S3-S8.
54. Ginat DT, Gupta R. Advances in computed tomography imaging technology. *Annual review of biomedical engineering*. 2014;16:431-53.
55. Grobe T, Dörning H, Schwartz F. Barmer GEK Arztreport 2011: Auswertungen zu Daten bis 2009; Schwerpunkt: bildgebende Diagnostik, Computer-und Magnetresonanztomographie. St Augustin: Asgard-Verlag. 2011.
56. Lell M, Wucherer M, Kachelrieß M. Dosis und Dosisreduktion in der Computertomografie. *Radiologie up2date*. 2017;17(02):163-78.
57. Euler A, Szücs-Farkas Z, Schindera S. Möglichkeiten der Strahlenreduktion bei der CT des Körperstamms. *Radiologie up2date*. 2014;14(02):e1-e.
58. Pearce MS, Salotti JA, Little MP, McHugh K, Lee C, Kim KP, Howe NL, Ronckers CM, Rajaraman P, Craft AW. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *The Lancet*. 2012;380(9840):499-505.
59. Padole A, Ali Khawaja RD, Kalra MK, Singh S. CT radiation dose and iterative reconstruction techniques. *American journal of roentgenology*. 2015;204(4):W384-W92.
60. McCarthy J. What is artificial intelligence? 1998.

61. Wang P. On defining artificial intelligence. *Journal of Artificial General Intelligence*. 2019;10(2):1-37.
62. Dobrev D. A definition of artificial intelligence. arXiv preprint arXiv:12101568. 2012.
63. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML. Deep learning in medical imaging and radiation therapy. *Medical physics*. 2019;46(1):e1-e36.
64. Shalev-Shwartz S, Ben-David S. *Understanding machine learning: From theory to algorithms*: Cambridge university press; 2014.
65. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *New England Journal of Medicine*. 2019;380(14):1347-58.
66. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *Journal of medical imaging and radiation sciences*. 2019;50(4):477-87.
67. Kocak B, Kus EA, Yardimci AH, Bektas CT, Kilickesmez O. Machine Learning in Radiomic Renal Mass Characterization: Fundamentals, Applications, Challenges, and Future Directions. *American Journal of Roentgenology*. 2020;215(4):920-8.
68. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Physics in Medicine & Biology*. 2016;61(13):R150.
69. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563-77.
70. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, Zheng J, Goldman D, Moskowitz C, Fine SW. 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.
71. Coroller TP, Grossmann P, Hou Y, Velazquez ER, Leijenaar RT, Hermann G, Lambin P, Haibe-Kains B, Mak RH, Aerts HJ. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiotherapy and Oncology*. 2015;114(3):345-50.
72. Liang C, Huang Y, He L, Chen X, Ma Z, Dong D, Tian J, Liang C, Liu Z. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer. *Oncotarget*. 2016;7(21):31401.
73. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A. Radiomics: extracting more information

from medical images using advanced feature analysis. *European journal of cancer*. 2012;48(4):441-6.

74. Yang G, Li G, Pan T, Kong Y, Wu J, Shu H, Luo L, Dillenseger J-L, Coatrieux J-L, Tang L, editors. Automatic segmentation of kidney and renal tumor in ct images based on 3d fully convolutional neural network with pyramid pooling module. 2018 24th International Conference on Pattern Recognition (ICPR); 2018: IEEE.

75. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin P-M, Larochelle H. Brain tumor segmentation with deep neural networks. *Medical image analysis*. 2017;35:18-31.

76. Murray JM, Kaissis G, Braren R, Kleesiek J. Wie funktioniert Radiomics? *Der Radiologe*. 2020;60(1):32-41.

77. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, Ganeshan B, Miles KA, Cook GJ, Goh V. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights into imaging*. 2012;3(6):573-89.

78. Starmans MP, van der Voort SR, Tovar JMC, Veenland JF, Klein S, Niessen WJ. Radiomics: data mining using quantitative medical image features. *Handbook of Medical Image Computing and Computer Assisted Intervention*: Elsevier; 2020. p. 429-56.

79. Kuhn M, Johnson K. *Applied predictive modeling*: Springer; 2013.

80. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and Generalizability in Radiomics Modeling: Possible Strategies in Radiologic and Statistical Perspectives. *Korean J Radiol*. 2019;20(7):1124-37.

81. Hoo ZH, Candlish J, Teare D. *What is an ROC curve? : BMJ Publishing Group Ltd and the British Association for Accident ...*; 2017.

82. Ben-David A. About the relationship between ROC curves and Cohen's kappa. *Engineering Applications of Artificial Intelligence*. 2008;21(6):874-82.

83. Zhu W, Zeng N, Wang N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG proceedings: health care and life sciences*, Baltimore, Maryland. 2010;19:67.

84. Brodersen KH, Ong CS, Stephan KE, Buhmann JM, editors. The balanced accuracy and its posterior distribution. 2010 20th international conference on pattern recognition; 2010: IEEE.

85. Burgess NA, Koo BC, Calvert RC, Hindmarsh A, Donaldson PJ, Rhodes M. Randomized trial of laparoscopic v open nephrectomy. *Journal of endourology*. 2007;21(6):610-3.

86. Patel HD, Semerjian A, Gupta M, Pavlovich CP, Johnson MH, Gorin MA, Allaf ME, Pierorazio PM, editors. Surgical removal of renal tumors with low metastatic potential based on clinical radiographic size: a systematic review of the literature. *Urologic Oncology: Seminars and Original Investigations*; 2019: Elsevier.
87. Lin F, Cui E-M, Lei Y, Luo L-p. CT-based machine learning model to predict the Fuhrman nuclear grade of clear cell renal cell carcinoma. *Abdominal Radiology*. 2019;44(7):2528-34.
88. Kocak B, Durmaz ES, Ates E, Kaya OK, Kilickesmez O. Unenhanced CT texture analysis of clear cell renal cell carcinomas: a machine learning–based study for predicting histopathologic nuclear grade. *American Journal of Roentgenology*. 2019;212(6):W132-W9.
89. Bektas CT, Kocak B, Yardimci AH, Turkcanoglu MH, Yucetas U, Koca SB, Erdim C, Kilickesmez O. Clear cell renal cell carcinoma: machine learning-based quantitative computed tomography texture analysis for prediction of Fuhrman nuclear grade. *European radiology*. 2019;29(3):1153-63.
90. Ding J, Xing Z, Jiang Z, Chen J, Pan L, Qiu J, Xing W. CT-based radiomic model predicts high grade of clear cell renal cell carcinoma. *European journal of radiology*. 2018;103:51-6.
91. Shu J, Tang Y, Cui J, Yang R, Meng X, Cai Z, Zhang J, Xu W, Wen D, Yin H. Clear cell renal cell carcinoma: CT-based radiomics features for the prediction of Fuhrman grade. *European journal of radiology*. 2018;109:8-12.
92. Shu J, Wen D, Xi Y, Xia Y, Cai Z, Xu W, Meng X, Liu B, Yin H. Clear cell renal cell carcinoma: machine learning-based computed tomography radiomics analysis for the prediction of WHO/ISUP grade. *European journal of radiology*. 2019;121:108738.
93. Sun X, Liu L, Xu K, Li W, Huo Z, Liu H, Shen T, Pan F, Jiang Y, Zhang M. Prediction of ISUP grading of clear cell renal cell carcinoma using support vector machine model based on CT images. *Medicine*. 2019;98(14).
94. Cui E, Li Z, Ma C, Li Q, Lei Y, Lan Y, Yu J, Zhou Z, Li R, Long W. Predicting the ISUP grade of clear cell renal cell carcinoma with multiparametric MR and multiphase CT radiomics. *European radiology*. 2020;30(5):2912-21.
95. He X, Zhang H, Zhang T, Han F, Song B. Predictive models composed by radiomic features extracted from multi-detector computed tomography images for predicting low- and high-grade clear cell renal cell carcinoma: A STARD-compliant article. *Medicine*. 2019;98(2).
96. Nazari M, Shiri I, Hajianfar G, Oveisi N, Abdollahi H, Deevband MR, Oveisi M, Zaidi H. Noninvasive Fuhrman grading of clear cell renal cell carcinoma using computed

tomography radiomic features and machine learning. *La radiologia medica*. 2020;125(8):754-62.

97. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006;31(3):1116-28.

98. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RG, Fillion-Robin J-C, Pieper S, Aerts HJ. Computational radiomics system to decode the radiographic phenotype. *Cancer research*. 2017;77(21):e104-e7.

99. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The design of SimpleITK. *Frontiers in neuroinformatics*. 2013;7:45.

100. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*. 2016;15(2):155-63.

101. Team RC. *R: A language and environment for statistical computing*. 2013.

102. Wickham H. *The tidyverse. R package ver.* 2017;1(1):836.

103. Gamer M, Lemon J, Gamer MM, Robinson A, Kendall's W. Package 'irr'. Various coefficients of interrater reliability and agreement. 2012;22.

104. Kuhn M. Building predictive models in R using the caret package. *Journal of statistical software*. 2008;28(1):1-26.

105. Breiman L. Random forests. *Machine learning*. 2001;45(1):5-32.

106. Friedman JH. Greedy function approximation: a gradient boosting machine. *Annals of statistics*. 2001:1189-232.

107. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 1970;12(1):55-67.

108. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996;58(1):267-88.

109. Zou H, Hastie T. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*. 2005;67(2):301-20.

110. Fix E, Hodges JL. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *International Statistical Review/Revue Internationale de Statistique*. 1989;57(3):238-47.

111. Liaw A, Wiener M. Classification and regression by randomForest. *R news*. 2002;2(3):18-22.
112. Ridgeway G. Generalized Boosted Models: A guide to the gbm package. Update. 2007;1(1):2007.
113. Hastie T, Qian J. Glmnet vignette. Retrieved June. 2014;9(2016):1-30.
114. Liu N, Gan W, Qu F, Wang Z, Zhuang W, Agizamhan S, Xu L, Yin J, Guo H, Li D. Does the Fuhrman or World Health Organization/International Society of Urological Pathology Grading System apply to the Xp11. 2 translocation renal cell carcinoma?: a 10-year single-center study. *The American journal of pathology*. 2018;188(4):929-36.
115. Rabjerg M, Gerke O, Engvad B, Marcussen N. Comparing World Health Organization/International Society of Urological Pathology Grading and Fuhrman Grading with the Prognostic Value of Nuclear Area in Patients with Renal Cell Carcinoma. *Uro*. 2021;1(1):2-13.
116. Kim H, Inomoto C, Uchida T, Furuya H, Komiyama T, Kajiwara H, Kobayashi H, Nakamura N, Miyajima A. Verification of the International Society of Urological Pathology recommendations in Japanese patients with clear cell renal cell carcinoma. *International journal of oncology*. 2018;52(4):1139-48.
117. Larue RT, van Timmeren JE, de Jong EE, Feliciani G, Leijenaar RT, Schreurs WM, Sosef MN, Raat FH, van der Zande FH, Das M. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta oncologica*. 2017;56(11):1544-53.
118. Isensee F, Maier-Hein KH. An attempt at beating the 3D U-Net. *arXiv preprint arXiv:190802182*. 2019.
119. Sun X-Y, Feng Q-X, Xu X, Zhang J, Zhu F-P, Yang Y-H, Zhang Y-D. Radiologic-radiomic machine learning models for differentiation of benign and malignant solid renal masses: comparison with expert-level radiologists. *American Journal of Roentgenology*. 2020;214(1):W44-W54.
120. Zhou L, Zhang Z, Chen Y-C, Zhao Z-Y, Yin X-D, Jiang H-B. A deep learning-based radiomics model for differentiating benign and malignant renal tumors. *Translational oncology*. 2019;12(2):292-300.
121. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton J, Snyder A, Weigelt B, Vargas H. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clinical radiology*. 2017;72(1):3-10.
122. Sequist LV, Waltman BA, Dias-Santagata D, Digumarthy S, Turke AB, Fidiyas P, Bergethon K, Shaw AT, Gettinger S, Cospers AK. Genotypic and histological evolution of

lung cancers acquiring resistance to EGFR inhibitors. *Science translational medicine*. 2011;3(75):75ra26-75ra26.

123. Jamshidi N, Jonasch E, Zapala M, Korn RL, Brooks JD, Ljungberg B, Kuo MD. The radiogenomic risk score stratifies outcomes in a renal cell cancer phase 2 clinical trial. *European radiology*. 2016;26(8):2798-807.

124. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJ, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-38.

125. Shaikh FA, Kolowitz BJ, Awan O, Aerts HJ, von Reden A, Halabi S, Mohiuddin SA, Malik S, Shrestha RB, Deible C. Technical challenges in the clinical application of radiomics. *JCO clinical cancer informatics*. 2017;1:1-8.

126. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. *Nature Reviews Cancer*. 2018;18(8):500-10.

127. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow P-M, Zietz M, Hoffman MM. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*. 2018;15(141):20170387.

128. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electronic Markets*. 2021:1-11.

129. Choi E, Schuetz A, Stewart WF, Sun J. Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*. 2017;24(2):361-70.

130. Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D, editors. Early diagnosis of Alzheimer's disease with deep learning. 2014 IEEE 11th international symposium on biomedical imaging (ISBI); 2014: IEEE.

131. Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NN, Tomaszewski J, González FA, Madabhushi A. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Scientific reports*. 2017;7(1):1-14.

132. Mohsen H, El-Dahshan E-SA, El-Horbaty E-SM, Salem A-BM. Classification using deep learning neural networks for brain tumors. *Future Computing and Informatics Journal*. 2018;3(1):68-71.

133. Nie D, Zhang H, Adeli E, Liu L, Shen D, editors. 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. *International conference on medical image computing and computer-assisted intervention*; 2016: Springer.

134. Gilad-Bachrach R, Dowlin N, Laine K, Lauter K, Naehrig M, Wernsing J, editors. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. International conference on machine learning; 2016: PMLR.
135. Lin F, Ma C, Xu J, Lei Y, Li Q, Lan Y, Sun M, Long W, Cui E. A CT-based deep learning model for predicting the nuclear grade of clear cell renal cell carcinoma. *European Journal of Radiology*. 2020;129:109079.

Eidesstattliche Versicherung

„Ich, Maximilian Winter, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema: *Radiomics-basierte CT-Bilddaten-Analyse zur nicht-invasiven Vorhersage des histopathologischen Tumorgrads klarzelliger Nierenzellkarzinome / Radiomics-based CT image analysis to predict the pathological grade in clear cell renal cell 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.“

30.10.2021

Maximilian Winter

Lebenslauf

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

Danksagung

Mein besonderer Dank gilt meinen Betreuern PD Dr. med. Lisa Adams und PD Dr. med. Keno Bressemer für die Vergabe des Promotionsthemas und die kompetente, zuverlässige, und motivierende Unterstützung bei der Durchführung meiner Promotion. Ich hätte mir keine bessere Betreuung wünschen können.

Außerdem möchte ich mich an dieser Stelle bei meiner Familie bedanken, die mich auf meinem bisherigen Lebensweg und bei der Erstellung dieser Arbeit immer bedingungslos unterstützt hat.

Bescheinigung der statistischen Beratung



CharitéCentrum für Human- und Gesundheitswissenschaften

Charité | Campus Charité Mitte | 10117 Berlin

Institut für Biometrie und klinische Epidemiologie (iBikE)

Direktor: Prof. Dr. Geraldine Rauch

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

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



Name, Vorname: Winter, Maximilian
Emailadresse: maximilian.winter@charite.de
Matrikelnummer: 222520
PromotionsbetreuerIn: Dr. Lisa Adams
Promotionsinstitution / Klinik: Klinik für Radiologie CCM

Bescheinigung

Hiermit bescheinige ich, dass Herr *Maximilian Winter* innerhalb der Service Unit Biometrie des Instituts für Biometrie und klinische Epidemiologie (iBikE) bei mir eine statistische Beratung zu einem Promotionsvorhaben wahrgenommen hat. Folgende Beratungstermine wurden wahrgenommen:

- Termin 1: 13.11.2020
- Termin 2: 21.06.2021

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

- Überprüfung von Normalverteilungsannahmen
- Verwendung geeigneter statistischer Tests
- Die Arbeit entspricht den gängigen Anforderungen der Statistik an eine medizinische Dissertation

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

Datum: 29.09.21

Name des Beraters/ der Beraterin: Anna Pöhlmann



Unterschrift BeraterIn, Institutsstempel

