

Aus dem Institut für Diagnostische und Interventionelle Radiologie
(Prof. Dr. med. J. Lotz)
der Medizinischen Fakultät der Universität Göttingen

CT-basierte Radiomic Feature Analyse zur Unterscheidung von malignen und benignen Nierentumoren

INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades
der Medizinischen Fakultät der
Georg-August-Universität zu Göttingen

vorgelegt von

Laura Maria Delonge

aus
Aachen

Göttingen 2023

Dekan: Prof. Dr. med. W. Brück

Betreuungsausschuss

Betreuer/in: PD Dr. med. J. Uhlig

Ko-Betreuer/in: PD Dr. med. M. Leitsmann

Prüfungskommission

Referent/in: PD Dr. med. J. Uhlig

Ko-Referent/in: PD Dr. med. M. Leitsmann

Drittreferent/in: Prof. Dr. R. Dressel

Datum der mündlichen Prüfung: 14.11.2023

Hiermit erkläre ich, die Dissertation mit dem Titel " CT-basierte Radiomic Feature Analyse zur Unterscheidung von malignen und benignen Nierentumoren" eigenständig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Göttingen, den
.....
(Unterschrift)

Die Daten, auf denen die vorliegende Arbeit basiert, wurden teilweise publiziert:

Uhlig J, Leha A, **Delonge LM**, Haack AM, Shuch B, Kim HS, Bremmer F, Trojan L, Lotz J, Uhlig A (2020b): Radiomic Features and Machine Learning for the Discrimination of Renal Tumor Histological Subtypes: A Pragmatic Study Using Clinical-Routine Computed Tomography. *Cancers* 12, 3010

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis.....	IV
Abkürzungsverzeichnis	V
1 Einleitung	1
1.1 Epidemiologie und Ätiologie von Nierentumoren	1
1.2 Klassifikation von Nierentumoren	2
1.2.1 Histologie.....	2
1.2.2 Tumorstadien.....	3
1.3 Klinische Präsentation und Prognose von Nierentumoren	5
1.4 Diagnostik von Nierentumoren	6
1.4.1 Computertomographie	6
1.4.2 Magnetresonanztomographie	7
1.4.3 Biopsie.....	8
1.5 Therapie von Nierentumoren.....	9
1.5.1 Operation.....	9
1.5.2 Ablation.....	10
1.5.3 Aktive Überwachung	10
1.6 Radiomics und maschinelles Lernen	11
1.7 Zielsetzung der Arbeit	13
2 Material und Methoden	14
2.1 Zusammensetzung des Patientenkollektivs.....	14
2.2 Einschlusskriterien	14
2.2.1 Präoperative Bildgebung	14
2.2.2 Histologische Diagnose.....	15
2.3 Segmentation.....	15
2.4 Radiomic Features	16
2.5 Vorverarbeitung.....	21
2.6 Kreuzvalidierung	21
2.7 Diagnostische Genauigkeit	22
2.8 Skalierung.....	23
2.9 Feature selection.....	23
2.10 Ausgleich des Klassenungleichgewichts	23
2.11 Machine learning.....	24
2.11.1 Random forest	24
2.11.2 Neural network.....	26

2.11.3	Support vector machine	27
2.11.4	K-nearest-neighbors.....	28
2.12	Statistische Analysen	29
3	Ergebnisse	30
3.1	Patientenkollektiv	30
3.2	Vorbereitung der Algorithmen.....	32
3.2.1	Vorverarbeitung.....	32
3.2.2	SMOTE.....	35
3.3	Leistung der Algorithmen	36
4	Diskussion	40
4.1	Zusammensetzung des Patientenkollektivs.....	41
4.2	Histopathologische Diagnosen	41
4.3	Bildgebung.....	41
4.4	Methodik.....	43
4.4.1	Segmentationen und Radiomic Features	43
4.4.2	Vorverarbeitung der machine learning-Algorithmen.....	43
4.5	Maschinelle Algorithmen	45
4.6	Limitationen der Studie	47
4.7	Ausblick.....	48
5	Zusammenfassung.....	50
6	Literaturverzeichnis	52

Abbildungsverzeichnis

Abbildung 1: Flussdiagramm Radiomics	12
Abbildung 2: Segmentationsbeispiele	16
Abbildung 3: Beispiel Kreuzvalidierung	22
Abbildung 4: Beispiel Entscheidungsbaum	24
Abbildung 5: Beispiel <i>random forest</i> -Algorithmus	25
Abbildung 6: Beispiel <i>neural network</i> -Algorithmus	26
Abbildung 7: Beispiel <i>support vector machine</i> -Algorithmus	27
Abbildung 8: Beispiel <i>k-nearest-neighbors</i> -Algorithmus	28
Abbildung 9: Flowchart Patienteneinschluss	30
Abbildung 10: Heatmap zur Darstellung der korrelierenden Variablen	35
Abbildung 11: Mediane AUC der Algorithmen	38
Abbildung 12: ROC-Kurve des nnet-Algorithmus	39

Tabellenverzeichnis

Tabelle 1: TNM-Klassifikation von Nierenzellkarzinomen.....	3
Tabelle 2: Tumorstadien bei Nierenzellkarzinomen	4
Tabelle 3: Auflistung der General Info Features	17
Tabelle 4: Auflistung der Radiomic Features nach Subgruppe	18
Tabelle 5: Charakteristika der Studienkohorte	31
Tabelle 6: Zusammenfassung über Bildartefakte und den Ort der Bildgebung	32
Tabelle 7: Auflistung der final verwendeten Radiomic Features	32
Tabelle 8: Charakteristika der Studienkohorte nach der Anwendung von SMOTE	36
Tabelle 9: Übersicht über die Leistung der Algorithmen	37

Abkürzungsverzeichnis

AUC	<i>area under the receiver operating characteristics curve</i> , Fläche unter der Kurve
CT	Computertomographie
HU	<i>hounsfield units</i> , Hounsfield Einheiten
IBSI	Image Biomarker Standardization Initiative
knn	<i>k-nearest-neighbors</i> , Nächste Nachbarn Klassifikation
max	maximal
min	minimal
ML	<i>machine learning</i> , Maschinelles Lernen
MRT	Magnetresonanztomographie
nnet	<i>neural network</i> , Neuronales Netzwerk
rf	<i>random forest</i> , Zufallswald
RFE	<i>recursive feature elimination</i>
ROC	<i>receiver operating characteristics curve</i> , Operationscharakteristik eines Beobachters
ROI	<i>region of interest</i> , Bereich von Interesse
SD	<i>standard deviation</i> , Standardabweichung
SMOTE	<i>synthetic minority over-sampling technique</i>
SRM	<i>small renal masses</i> , Kleine Nierentumoren
svm	<i>support vector machine</i>
UMG	Universitätsmedizin Göttingen

1 Einleitung

1.1 Epidemiologie und Ätiologie von Nierentumoren

Mit etwa 96 % aller Nierentumoren ist das Nierenzellkarzinom der häufigste Nierentumor im Erwachsenenalter (Robert Koch-Institut 2019). Mit etwa 400000 Neudiagnosen im Jahr 2018 ist es weltweit die sechsthäufigste (für Männer) beziehungsweise die zehnthäufigste (für Frauen) diagnostizierte Krebsart (Capitani et al. 2019). Das mittlere Erkrankungsalter beträgt 68 Jahre für Männer und 72 Jahre für Frauen (Robert Koch-Institut 2019).

In den vergangenen 20 Jahren ist die Inzidenz sowohl in Europa als auch weltweit pro Jahr um etwa 2 % angestiegen (Ljungberg et al. 2019). Dies ist vor allem auf die steigende Inzidenz kleiner Nierentumoren (*small renal masses, SRM*), also Tumoren mit einer Größe von unter 4 cm, zurückzuführen (Kane et al. 2008; AWMF 2020). In Deutschland erkranken jährlich etwa 15000 Patienten an einem Nierenzellkarzinom, zwei Drittel der Patienten sind Männer. Die absolute Zahl der Sterbefälle ist für beide Geschlechter in den vergangenen 15 Jahren etwa gleichbleibend (etwa 3300 Männer und 2000 Frauen pro Jahr) (Robert Koch-Institut 2019).

Für das Nierenzellkarzinom sind mehrere Risikofaktoren bekannt. Hierunter fallen unter anderem Tabakkonsum, ein erhöhter Blutdruck (sowohl systolisch als auch diastolisch), ein erhöhter *Body-Mass-Index* und eine chronische Niereninsuffizienz (Hunt et al. 2005; Renehan et al. 2008; Hidayat et al. 2017; Ljungberg et al. 2019).

Schätzungsweise 5 bis 8 % aller Nierenzellkarzinome zeigen eine genetische Komponente (Ljungberg et al. 2019). Häufig besteht eine Assoziation mit einem speziellen Subtyp. Beispiele hierfür sind das von-Hippel-Lindau-Syndrom mit einer Prädisposition für das klarzellige Nierenzellkarzinom, das hereditäre papilläre Nierenzellkarzinomsyndrom mit einer Prädisposition für das papilläre Nierenzellkarzinom (Typ 1) und das Birt-Hogg-Dubé-Syndrom für das chromophobe Nierenzellkarzinom (Przybycin et al. 2013). Ein Beispiel für eine Assoziation mit einem benignen Nierentumor ist die tuberöse Sklerose. Hier liegt bei etwa 50 bis 70 % der betroffenen Patienten gleichzeitig auch ein Angiomyolipom vor (Stillwell et al. 1987; Bissler und Kingswood 2004).

1.2 Klassifikation von Nierentumoren

1.2.1 Histologie

Die World Health Organisation klassifiziert Nierentumoren in mehrere Subtypen. Die für unsere Arbeit relevanten Tumoren werden als Nierenzelltumoren (die malignen klarzelligen-, papillären-, chromophoben Nierenzellkarzinome und das benigne Onkozytom) sowie als mesenchymale Tumoren (das benigne Angiomyolipom) klassifiziert (Moch et al. 2016).

Das klarzellige Nierenzellkarzinom ist mit 80 bis 90 % aller Nierenzellkarzinome der häufigste maligne Subtyp (Ljungberg et al. 2019). Die Ursprungszellen des Tumors sind das Epithel des proximalen Nierentubulus (Cairns 2011). Etwa 80 % der Tumoren zeigen eine funktionelle Inaktivierung des von-Hippel-Lindau-Gens (Brugarolas 2014). Die Immunhistochemie zeigt eine Anfärbung von Vimentin (Epithelmarker) und Carboanhydrase IX (Hypoxiemarker) (Reuter et al. 2014).

Das papilläre Nierenzellkarzinom ist mit einer Häufigkeit von etwa 6 bis 15 % aller Nierenzellkarzinome der zweithäufigste Subtyp. Wie das klarzellige Nierenzellkarzinom entstehen auch diese Tumoren aus den Epithelzellen des proximalen Nierentubulus (Cairns 2011). Eine immunhistochemische Färbung des papillären Nierenzellkarzinoms zeigt eine starke Expression von Cytokeratin 7 (Epithelmarker), Vimentin und Alpha-Methylacyl-CoA Racemase (Reuter et al. 2014).

Die Häufigkeit des chromophoben Nierenzellkarzinom beträgt etwa 5 % aller Nierenzellkarzinome. Der Ursprung des Tumors sind Epithelzellen des distalen Tubulus, genauer des Sammelrohrs (Cairns 2011). Relevante immunhistochemische Marker sind CD117 (Tyrosinkinase), Ksp-cadherin (Tubulusepithelmarker), Cytokeratin 7 und Vimentin (Reuter et al. 2014).

Die häufigsten (und für diese Studie relevanten) benignen Nierentumoren sind das Onkozytom und das Angiomyolipom.

Ein Onkozytom liegt bei etwa 3 bis 9 % aller renaler Tumoren vor (Ljungberg et al. 2019). Der Ursprung sind, wie auch beim chromophoben Nierenzellkarzinom, die Epithelzellen des Sammelrohrs (Cairns 2011). Charakteristische Marker sind CD117 und Ksp-cadherin (Reuter et al. 2014).

Ungefähr 1 bis 3 % aller renaler Tumoren sind Angiomyolipome. Die Prävalenz in der Allgemeinbevölkerung liegt bei etwa 0,44 % (Ljungberg et al. 2019). Es wird eine sporadische Form (etwa 90 % aller Angiomyolipome) von einer mit der tubulösen Sklerose assoziierten

Form unterschieden (Bissler und Kingswood 2004). Histologisch zählen Angiomyolipome zur Gruppe der *perivascular epithelioid cells tumors* und bestehen zu unterschiedlichen Anteilen aus Blutgefäßen (-angio), glatten Muskelzellen (-myo) und Fettzellen (-lipom). Histologische Marker sind HMB45 (Melanozytärer Marker) und Cadhepsin K (Reuter et al. 2014).

1.2.2 Tumorstadien

Die TNM Klassifikation (Tumor, Lymphknoten, Metastasen) für maligne Tumoren wird durch das American Joint Committee of Cancer festgelegt. Analog zu anderen Tumorentitäten werden die lokale Ausdehnung der Tumoren, sowie deren lymphogener Befall und Metastasierung nach den in Tabelle 1 und Tabelle 2 dargelegten Kriterien eingeteilt.

Tabelle 1: TNM-Klassifikation von Nierenzellkarzinomen

Klassifikation	Stadium	Beschreibung
T (Tumor)	T1	Tumor 7 cm oder weniger in größter Ausdehnung, begrenzt auf die Niere
	T1a	Tumor 4 cm oder weniger in größter Ausdehnung
	T1b	Tumor > 4 cm bis 7 cm
	T2	Tumor mehr als 7 cm in größter Ausdehnung, begrenzt auf die Niere
	T2a	Tumor > 7 cm bis 10 cm
	T2b	Tumor > 10 cm
	T3	Tumor breitet sich in die größeren Venen aus oder infiltriert direkt perirenales Fettgewebe, jedoch nicht in ipsilaterale Nebenniere und nicht über die Gerota-Faszie hinaus
	T3a	Tumor mit makroskopischer Ausbreitung in die Nierenvene oder ihre segmentalen Äste (mit muskulärer Wand) oder mit Infiltration des perirenales und/oder peripelvinen Fettgewebes, aber nicht über die Gerota-Faszie hinaus
	T3b	Tumor mit makroskopischer Ausbreitung in die Vena Cava unterhalb des Zwerchfells

Klassifikation	Stadium	Beschreibung
	T3c	Tumor mit makroskopischer Ausbreitung in die Vena cava oberhalb des Zwerchfells oder mit Infiltration der Wand der Vena cava
	T4	Tumor infiltriert über die Gerota-Faszie hinaus (eingeschlossen die kontinuierliche Ausbreitung in die ipsilateralen Nebennieren)
N (regionale Lymphknoten)	N0	Keine regionalen Lymphknotenmetastasen
	N1	Metastase(n) in regionärem Lymphknoten
M (Metastasen)	M0	Keine Fernmetastasen
	M1	Fernmetastasen

Tabelle übernommen aus (Edge und American Joint Committee on Cancer 2010)

Tabelle 2: Tumorstadien bei Nierenzellkarzinomen

Stadium	Primärtumor	Lymphknoten	Fernmetastasen
1	T1	N0	M0
	T1a		
	T1b		
2	T2a	N0	M0
	T2b		
3	T3a	N0	M0
	T3b		
	T3c		
	T1-3	N1	M0
4	T4	N0, N1	M0
	alle T	alle N	M1

Tabelle übernommen aus: (Edge und American Joint Committee on Cancer 2010)

1.3 Klinische Präsentation und Prognose von Nierentumoren

50 bis 60 % der heute diagnostizierten Tumoren sind zufällige, asymptomatische Befunde, die meist im Rahmen einer abdominellen Ultraschall- oder Computertomographieuntersuchung festgestellt werden (Jayson und Sanders 1998; Luciani et al. 2000; Lee et al. 2002; Thorstenson et al. 2014).

In den 1970er Jahren wiesen etwa 60 % der Patienten mit einem Nierenzellkarzinom bei Erstdiagnose eine Makrohämaturie auf und etwa 10 % zeigten die klassische Symptomtrias aus Flankenschmerzen, Makrohämaturie und palpablem Tumor (Skinner et al. 1971). Heutzutage ist das Auftreten von Symptomen (am häufigsten sind eine Makrohämaturie, abdominelle Schmerzen, ein palpabler Tumor, Varikozelen und paraneoplastische Symptome) zu meist ein Hinweis auf das Vorliegen einer fortgeschrittenen Tumorerkrankung. Symptomatische Tumoren sind meist größer, haben einen aggressiveren histologischen Subtyp und ein höheres TNM-Stadium als asymptomatische Tumoren (Lee et al. 2002).

Bei Angiomyolipomen schwankt die Häufigkeit von Schmerzen, Hämaturie und palpablem Tumor bei Diagnose in der Literatur stark (zwischen 85 und 14 % (Seyam et al. 2008; Sooriakumaran et al. 2010)). Die schwerwiegendste Komplikation, die spontane und atraumatische Ruptur des Angiomyolipoms mit massiver Blutung bis hin zum hypovolämischen Schock (Wunderlich Syndrom), zeigt sich nur bei etwa 6 % der Patienten (Seyam et al. 2008; Sooriakumaran et al. 2010).

Auch Onkozytome werden in der Regel inzidentell diagnostiziert, etwa ein Drittel der Patienten klagt über lokale Symptome (Romis et al. 2004).

Während über 50 % der malignen Tumoren im Stadium 1 diagnostiziert werden, liegt bei etwa 25 % bereits ein Stadium 3 (eine Lymphknotenmetastasierung oder eine Ausbreitung in die regionalen Gefäße) und bei etwa 15 % ein Stadium 4 mit Metastasen vor (Robert Koch-Institut 2019).

Die Prognose des Nierenzellkarzinom ist sowohl abhängig vom Subtyp, als auch vom Stadium. Die Fünf-Jahres-Überlebensraten der Subtypen gibt die aktuelle Leitlinie der European Association of Urology mit 91 % für das papilläre-, 88 % für das chromophobe- und 71 % für das klarzellige Nierenzellkarzinom an (Ljungberg et al. 2019). Das relative (im Vergleich zur Gesamtbevölkerung im gleichen Zeitraum) Fünf-Jahres-Überleben in Deutschland sinkt unabhängig vom Subtyp von etwa 95 % für das Stadium 1 auf etwa 15 % bei einem Stadium 4 (Robert Koch-Institut 2019).

1.4 Diagnostik von Nierentumoren

1.4.1 Computertomographie

Die Computertomographie (CT) ist der Goldstandard zur Diagnostik von Nierentumoren (Catalano et al. 2003).

Für die optimale Darstellung der Niere sind verschiedene Kontrastmittelphasen notwendig: Die native Phase (ohne Kontrastmittel) ist hilfreich, um einen Grundwert für eine spätere Kontrastmittelaufnahme zu definieren.

In der kortikomedullären Phase (etwa 25 bis 50 Sekunden nach Kontrastmittelgabe, auch arterielle Phase genannt) werden vor allem die arteriellen Gefäße und der Nierenkortex dargestellt. Dies erlaubt Aussagen über eine mögliche Gefäßinvasion von Nierentumoren. Die nephrographische Phase (etwa 80 bis 180 Sekunden nach Kontrastmittelgabe, auch als venöse Phase bezeichnet) ist am besten für die Darstellung der Nierenvenen und für die Unterscheidung zwischen Nierenparenchym und Nierentumor geeignet (Sheth et al. 2001).

Auch wenn bei der diagnostischen Genauigkeit von Nierenzellkarzinomen eine durchschnittliche Sensitivität von 88 % und Spezifität von 75 % erreicht werden (Vogel et al. 2019), gibt es bis dato keine bildmorphologischen Kriterien die eine sichere Differenzierung von malignen und benignen Nierentumoren anhand der CT ermöglichen (Millet et al. 2011). Jedoch gibt es einige bildmorphologische Charakteristika, die auf die unterschiedlichen histologischen Nierentumorsubtypen hinweisen können.

Der wichtigste Parameter zur Differenzierung der verschiedenen Nierentumoren ist die Kontrastmittelaufnahme, das heißt die Signalveränderung von Geweben nach der Gabe von intravenösem iodhaltigen Kontrastmittel (Israel und Bosniak 2005).

Der einzige Tumorsubtyp, der sicher mittels CT diagnostiziert werden kann, ist das klassische, fettreiche Angiomyolipom. Pathognomonisch ist hier makroskopisches Fett, in einer nativen CT-Untersuchung entspricht dies einer Abschwächung unter -10 Hounsfield-Einheiten (HU) (Song et al. 2016). Das fettarme Angiomyolipom hingegen zeigt ähnliche Eigenschaften wie das klarzellige Nierenzellkarzinom, so dass eine sichere Unterscheidung mittels CT nicht möglich ist (Jinzaki et al. 2014).

98 % der klarzelligen Nierenzellkarzinome zeigen eine Kontrastmittelaufnahme von über 15 HU im Vergleich zur nativen Aufnahme. Die maximale Aufnahme ist bei 81 % der Tumoren in der kortikomedullären Phase (Fu et al. 2016) und die Kontrastmittelaufnahme ist deutlich höher als bei nicht-klarzelligen Karzinomen ($149 \text{ HU} \pm \sim 50 \text{ HU}$) (Kim et al. 2002).

Bei einem Grenzwert von 84 HU können klarzellige Karzinome mit einer Sensitivität von 74 % und einer Spezifität von 100 % von nicht-klarzelligen Karzinomen unterschieden werden (Kim et al. 2002). Die Kontrastmittelaufnahme ist überwiegend heterogen. Kalzifikationen finden sich in etwa 10 bis 20 % der Fälle (Kim et al. 2002; Fu et al. 2016).

Das papilläre Nierenzellkarzinom ist meist klar begrenzt, überwiegend homogen und solide. Im Gegensatz zu klarzelligen Nierenzellkarzinomen zeigen etwa ein Drittel der papillären Nierenzellkarzinome keine nennenswerte Kontrastmittelaufnahme (definiert als Aufnahme über 15 HU), und die Phase der maximalen Aufnahme ist in über 90 % der Fälle die nephrographische Phase (Fu et al. 2016). Die Aufnahmewerte sind niedriger als für das klarzellige Nierenzellkarzinom, zeigen jedoch keinen signifikanten Unterschied zu chromophoben Nierenzellkarzinomen. Papilläre Nierenzellkarzinome zeigen in der Hälfte der Fälle eine heterogene Kontrastmittelaufnahme, bei einer Tumorgöße zwischen drei und sieben Zentimetern sogar in 88 % der Fälle (Kim et al. 2002).

Das chromophobe Nierenzellkarzinom ist, wie das papilläre Karzinom, meist klar begrenzt. Alle Tumoren haben eine nennenswerte Kontrastmittelaufnahme, in 62 % der Fälle ist die maximale Aufnahme in der nephrographischen Phase zu verzeichnen (Fu et al. 2016). Die Kontrastmittelaufnahme ist, im Gegensatz zu den anderen Subtypen, in 69 % der Fälle homogen (Kim et al. 2002). In Aufnahmen des chromophoben Nierenzellkarzinoms finden sich in über ein Drittel der Fälle Kalzifikationen (Kim et al. 2002; Fu et al. 2016). Als möglicher, jedoch unspezifischer Hinweis finden sich in etwa 20 % der Fälle eine zentrale, sternförmige Narbe (Rosenkrantz et al. 2010; Fu et al. 2016).

Der am schwierigsten zu diagnostizierende Nierentumor ist das Onkozytom. Hinweisgebende Eigenschaften sind unter anderem eine zentrale, sternförmige Narbe (in 10 bis 30 % der Fälle (Quinn et al. 1984; Choudhary et al. 2009)), eine segmentale Kontrastumkehr während der kortikomedullären und nephrographischen Phase (Kim et al. 2009) oder eine hohe Kontrastmittelaufnahme (Bird et al. 2011; Young et al. 2013; Choi et al. 2015).

1.4.2 Magnetresonanztomographie

Auf Grund von schlechterer Verfügbarkeit, höherer Kosten und einer längeren Untersuchungszeit gegenüber der CT ist eine Untersuchung mittels Magnetresonanztomographie (MRT) nur indiziert, wenn eine genauere Untersuchung bezüglich einer möglichen Gefäßinvasion eines Tumors erforderlich ist oder eine CT aus anderen Gründen (beispielsweise einer Schwangerschaft) nicht durchgeführt werden kann (Ljungberg et al. 2019).

Die diagnostische Genauigkeit von Nierenzellkarzinomen betreffend zeigt sich eine zur CT

vergleichbare Sensitivität (87,5 %) und Spezifität (89 %) (Vogel et al. 2019). In der Unterscheidung von benignen und malignen Tumoren zeigt sich eine hohe Sensitivität (93,8 %), jedoch eine geringere Spezifität von 71,4 % in der MRT beziehungsweise von 68,4 % in der CT (Beer et al. 2006).

Einige MRT-morphologische Kriterien können für das Vorliegen eines klarzelligen (hohe Signalintensität in T2-gewichteten Aufnahmen) oder papillären (niedrige Signalintensität in T2-gewichteten Aufnahmen) Nierenzellkarzinom sprechen. Eine sichere Unterscheidung ist aber auch hier nicht in allen Fällen möglich. Für das chromophobe Nierenzellkarzinom und das Onkozytom ist die Definition bildmorphologischer Kriterien schwieriger, so dass eine sichere Unterscheidung in der MRT nicht möglich ist (Rosenkrantz et al. 2010; Lopes Vendrami et al. 2017).

Einzig für die Diagnose des fettarmen Angiomyolipoms bietet die MRT gegenüber der CT definitive Vorteile. Hier sind eine niedrige Signalintensität in einer T2-gewichteten Aufnahme und eine starke, frühe Kontrastmittelaufnahme mit anschließendem Kontrastmittel-Auswaschen (*wash-out kinetics*) suggestive Merkmale für ein fettarmes Angiomyolipom (Lopes Vendrami et al. 2017).

Eine große Rolle spielt die MRT außerdem in der Einschätzung ob eine Gefäßinvasion des Tumors, vor allem in der Vena cava inferior, vorhanden ist (Reznek 2004).

1.4.3 Biopsie

Die Biopsie von Nierentumoren konnte sich nicht als diagnostischer Goldstandard etablieren. Gründe hierfür sind zum einen Zweifel an der Sicherheit und der diagnostischen Genauigkeit von Biopsien, zum anderen die Annahme, dass alle soliden Nierentumoren potenziell maligne sind und somit eine OP-Indikation vorliegt (Volpe et al. 2012).

In einer Metaanalyse konnte eine insgesamt hohe diagnostische Rate von Malignomen mittels Biopsien (92 %) gezeigt werden. Histologische Subtypen wurden in 90,3 % der Fälle korrekt diagnostiziert und auch die Komplikationsrate bei Nierenbiopsien war gering (8 %). Die häufigsten Komplikationen waren Hämatome (5 %) und Lumbalschmerzen (3 %). Als Nachteil zeigte sich der Anteil an nicht-diagnostischen Biopsien, also Prozeduren, bei denen aus dem gewonnenen Gewebe keine eindeutige Diagnose gestellt werden kann. Dieser lag, abhängig von der verwendeten Methode, bei bis zu 35 %. Als Ursache wurden technische Limitationen oder Probleme, die Tumormasse korrekt zu treffen, genannt (Marconi et al. 2016). In einer anderen Studie wurden nahezu alle (90 %) nicht-diagnostischen Biopsien postoperativ als maligne klassifiziert (Patel et al. 2016).

Vor allem in der Diagnostik von Onkozytomen zeigen sich Schwierigkeiten. So berichteten

Patel et al. (2017), dass bei ungefähr einem Drittel der durch Biopsie diagnostizierten Onkozytome die Diagnose inkorrekt ist und dass bei 25 % postoperativ ein maligner Nierentumor diagnostiziert wird.

Zusammenfassend spielt die invasive Biopsie nur eine untergeordnete Rolle für die primäre Diagnostik von Nierentumoren und wird aktuell nur für ausgewählte Patientengruppen empfohlen. Hierzu gehören Patienten, für die das Therapiekonzept der aktiven Überwachung (siehe Kapitel 1.5.3) in Frage kommt und Patienten, bei denen die Histologie vor einer Ablationstherapie bestimmt werden soll. Auch für Patienten mit einem bereits metastasierten Nierenzellkarzinom aber noch unbekannter Histopathologie kann eine Biopsie vor Einleitung einer systemischen Therapie hilfreich sein, um die beste medikamentöse und chirurgische Therapieentscheidung zu treffen. Für Patienten mit Komorbiditäten, für die nur die Strategie des *watchful waiting* (palliative Strategie, wenn eine aktive Therapie auf Grund von Komorbiditäten kontraindiziert ist) in Frage kommt, und bei denen das Biopsieergebnis die Therapieentscheidung nicht beeinflusst, ist eine Biopsie ebenso nicht angezeigt wie für Patienten, bei denen in einer bildgebenden Untersuchung der Verdacht auf einen potenziell malignen Tumor geäußert wurde und damit eine Operation indiziert ist (Ljungberg et al. 2019).

1.5 Therapie von Nierentumoren

1.5.1 Operation

Seit ihrer Erstbeschreibung 1963 wird die Nephrektomie als Standardeingriff zur kurativen Therapie des lokal begrenzten Nierenzellkarzinoms angesehen (Robson 1963).

Heute unterscheidet man zwischen der organerhaltenden Nierenteilresektion (partielle Nephrektomie) und der radikalen Nephrektomie. Diese Operationsverfahren können sowohl offen-chirurgisch als auch laparoskopisch, beziehungsweise laparoskopisch-roboterassistiert durchgeführt werden.

Da für die partielle Nephrektomie ein längeres Überleben (für Patienten mit einem T1a Tumorstadium), eine bessere postoperative Nierenfunktion und eine bessere Lebensqualität im Vergleich zur radikalen Nephrektomie gezeigt werden konnte und Rückfall- beziehungsweise Metastasen-Raten vergleichbar waren, wird diese Methode als Therapie der Wahl für Patienten mit T1 Tumoren empfohlen (MacLennan et al. 2012a; MacLennan et al. 2012b; Ljungberg et al. 2019). Insbesondere die bessere postoperative Nierenfunktion ist ein wichtiges Kriterium, da eine verminderte Nierenfunktion als Risikofaktor für kardiovaskuläre Ereignisse gilt (Go et al. 2004).

Für Patienten mit Tumoren über 7 cm Größe (Tumorstadium T2), welche mit einer partiellen Nephrektomie nicht sicher entfernt werden können, wird eine radikale Nephrektomie empfohlen (Ljungberg et al. 2019).

Auch für die aktive Therapie von Angiomyolipomen und Onkozytomen wird die partielle Nephrektomie, sofern möglich, bevorzugt (Ljungberg et al. 2019).

1.5.2 Ablation

Eine weitere Therapieoption ist die Ablation. Das Ziel ist eine lokale Gewebsdestruktion. Bei der Radiofrequenzablation wird CT-gesteuert, mit Hilfe eines perkutan eingebrachten Applikators, eine starke, lokale Hitzeentwicklung erreicht und so eine Nekrose ausgelöst. Bei der Kryoablation wird das Gewebe durch eine Kältesonde auf bis zu minus 70° C heruntergekühlt und eine Koagulationsnekrose und damit Apoptose des Gewebes und des Tumors ausgelöst.

Durch die stereotaktisch-ablative Radiotherapie wird, unter Nutzung CT-gesteuerter Bildgebung und hochenergetischer Strahlen, das Tumorgewebe von außen hochpräzise zerstört (AWMF 2020).

Im Vergleich zur partiellen Nephrektomie zeigt sich für die Ablationsverfahren eine höhere allgemeine Mortalität während bei der krebsspezifischen Mortalität bezogen auf die Behandlungsverfahren kein Unterschied zu erkennen ist. Dies ist wahrscheinlich auf das höhere Alter und die höhere Anzahl an vorhandenen Komorbiditäten der Patienten, die mit einer Ablation behandelt wurden, zurückzuführen. Die Komplikationsraten sind bei Ablationsverfahren niedriger als bei der Nephrektomie, die lokale Rezidivrate jedoch höher (Uhlig et al. 2019).

Diese Therapieoptionen sind deswegen für Patienten mit kleinen Tumoren und Komorbiditäten beziehungsweise einer geringen Lebenserwartung indiziert (Ljungberg et al. 2019).

1.5.3 Aktive Überwachung

Das Konzept der aktiven Überwachung umfasst die engmaschige Kontrolle des kleinen Nierentumors (durch Bildgebung) und ein therapeutisches Eingreifen erst bei Progress der Erkrankung (AWMF 2020).

Hintergrund sind Studien, die zeigen, dass bis zu 30 % der operierten Nierentumoren postoperativ eine benigne Histologie aufweisen und nur etwa 25 % der SRMs als aggressive maligne Tumoren klassifiziert werden (Lane et al. 2012; Kim et al. 2019).

SRMs zeigen außerdem eine geringere Wachstumsrate und metastasieren sehr selten (Chawla

et al. 2006; Smaldone et al. 2012). Auch das relative Fünf-Jahres-Überleben von Patienten mit T1 Tumoren beträgt 95 %.

Leitlinien empfehlen deswegen, diese Therapiestrategie auf Patienten mit *SRMs*, die nicht für die Standardtherapie einer Operation in Frage kommen (zum Beispiel auf Grund von Komorbiditäten) oder diese ablehnen, zu begrenzen und den möglichen Tumorprogress alle drei Monate zu kontrollieren (Jewett et al. 2011; Lane et al. 2012; Ljungberg et al. 2019; AWMF 2020).

Für histologisch gesicherte Onkozytome ist die aktive Überwachung unabhängig von der Tumorgröße die Therapie der Wahl (Richard et al. 2016). Jedoch muss hierbei bedacht werden, dass etwa ein Viertel der bioptisch gesicherten Onkozytome postoperativ als Nierenzellkarzinom umklassifiziert werden (Patel et al. 2017).

Die meisten Angiomyolipome werden mittels aktiver Überwachung kontrolliert (Ljungberg et al. 2019). Während die Indikation für eine aktive Therapie, aus Sorge vor Rupturen und lebensbedrohlichen Blutungen, früher bei einer Größe von über 4 cm lag, ist dieser klare Grenzwert heutzutage umstritten und eine Intervention nur noch bei starken Schmerzen, akuten oder rezidivierenden Blutungen und sehr großen Tumoren indiziert (Bhatt et al. 2015; Ljungberg et al. 2019; Fernández-Pello et al. 2020).

1.6 Radiomics und maschinelles Lernen

Radiomics beschreibt die Umwandlung von medizinischen Bildern in multidimensionale Daten. Ziel ist es, die erworbenen Datenmengen zur Erkennung von Mustern zu nutzen und so Entscheidungshilfen zur Diagnose, Therapie und Prognose von Krankheiten zu entwickeln. Die Grundlage dieses Prozesses ist die Annahme, dass Bilder nicht-sichtbare Details enthalten, die wichtige klinische Informationen über beispielsweise den Subtyp, die Aggressivität oder das Therapieansprechen eines Tumors liefern können (Gillies et al. 2015).

Abbildung 1 zeigt die verschiedenen Schritte, die für die Anwendung von Radiomics notwendig sind. Auf Grundlage einer Bildgebung (meist Schnittbildgebung mit CT oder MRT-Bildern) wird die *region of interest* (ROI) identifiziert. Dies können entweder die komplette Raumforderung oder nur bestimmte Bereiche innerhalb eines Tumors sein. Die ROI wird als dreidimensionales Objekt dargestellt, auf Grundlage dessen werden Eigenschaften (Radiomic Features) extrahiert. Hiermit gewonnene Daten können zur statistischen Auswertung, zur Erstellung von Datenbanken oder zur weiteren Analyse durch maschinelles Lernen verwendet werden (Gillies et al. 2015).

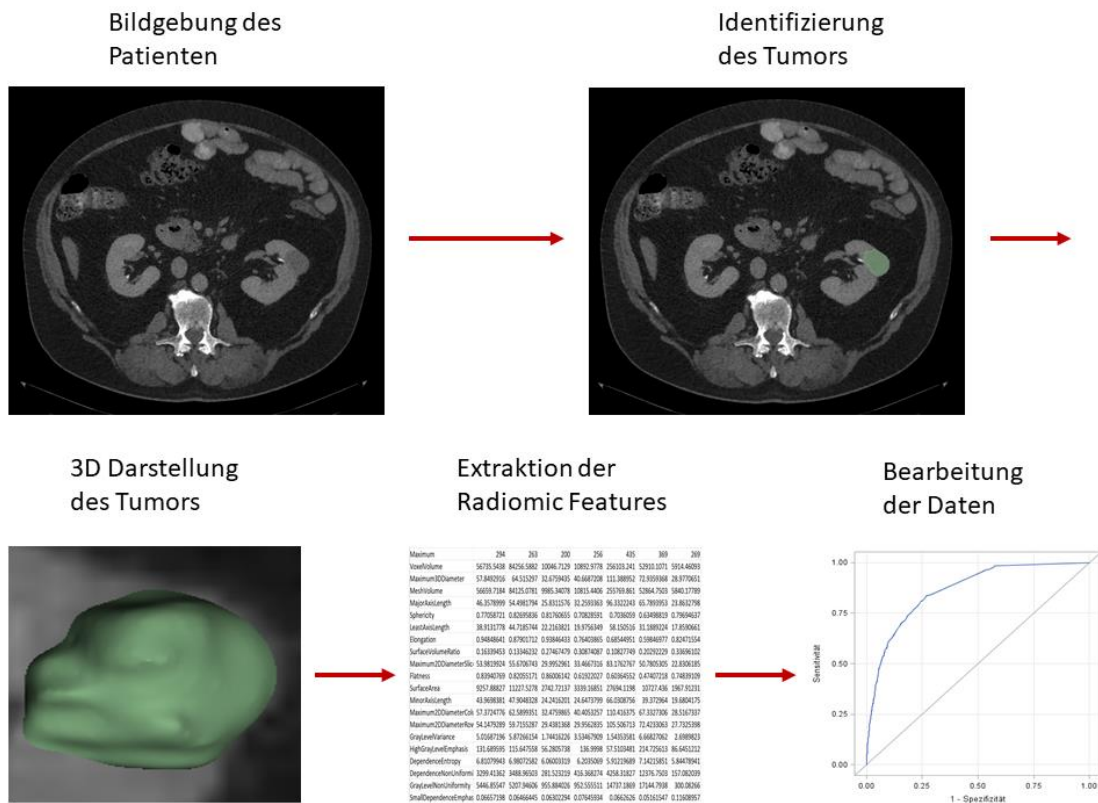


Abbildung 1: Flussdiagramm Radiomics

Maschinelles Lernen (*machine learning*, ML), allein und in Kombination mit Radiomics, ist ein vielversprechender Ansatz in der Optimierung von radiologischen Bewertungen.

In den meisten Fällen, wie auch in dieser Studie, wird das supervisierte maschinelle Lernen verwendet. Hierbei ist das Ergebnis (zum Beispiel die Dignität eines Nierentumors) bereits bekannt, und die eingesetzten Algorithmen (ML-Algorithmen) lernen und optimieren sich selbst anhand dieses Ergebnisses. Ziel ist es, Vorhersagen an unbekanntem Daten zu treffen. So könnte zum Beispiel bei Patienten mit identifizierten Tumoren durch einen Algorithmus eine zusätzliche Einschätzung abgegeben werden, ob die Tumoren maligne oder benigne sind. So könnte in Zusammenschau der Situation eine geeignete, individuelle Therapieentscheidung für jeden Patienten getroffen werden.

Während die Architektur der ML-Algorithmen bekannt ist (für genauere Beschreibungen der verwendeten Algorithmen, siehe Methodik-Teil), ist die genaue Berechnung und Ergebnisfindung nicht vollständig nachzuvollziehen.

Bisher wurden nur wenige Studien zum Themenkomplex „Radiomics und maschinellem Lernen bei Nierentumoren“ publiziert, die zumeist auf die Unterscheidung der unterschiedlichen Subtypen fokussiert waren.

Beispiele sind hier unter anderem die Studie von Sun et al. (2020), die die Leistung eines

maschinellen Algorithmus (*support vector machine*, svm) mit der von Radiologen in Bezug auf die Unterscheidung verschiedener Subtypen verglichen. Es konnte gezeigt werden, dass der Algorithmus im Vergleich zu den erfahrenen Radiologen zumindest gleichwertige Ergebnisse liefert.

Auch Raman et al. (2014) zeigten, dass mit Hilfe eines *random forest* (rf)-Algorithmus und CT-Texturanalysen hohe diagnostische Genauigkeiten (über 90 %) in der Unterscheidung von Subtypen erreicht werden können. Kunapuli et al. (2018) erreichten in ihrer Studie zur reinen Dignitätsunterscheidung eine AUC (*area-under-the receiver operating curve*) von 83 %.

Ein großer Unterschied zu unserer Studie liegt darin, dass die meisten Studien mit Hilfe von standardisierten CT-Protokollen durchgeführt wurden, das heißt, die Untersuchungen wurden an dem gleichen Computertomographen mit einem spezifischen Protokoll durchgeführt. Außerdem wurden in den meisten Fällen mehrphasige CT-Untersuchungen verwendet, das heißt, es wurden sowohl eine native als auch eine nephrographische und kortikomedulläre Aufnahme durchgeführt.

1.7 Zielsetzung der Arbeit

Derzeit ist mittels der radiologischen Schnittbildgebung keine sichere Unterscheidung zwischen benignen und malignen Nierentumoren möglich.

Bis zu 30 % der operativ resezierten Nierentumoren werden postoperativ als benigne klassifiziert, was die derzeitige Übertherapie von kleinen Nierentumoren und den klinischen Bedarf nach einer optimierten Bildgebung verdeutlicht. Auch im Hinblick auf eine alternde Gesellschaft, in der Patienten häufig viele therapielimitierende Komorbiditäten haben, ist eine Operation nicht immer der richtige Behandlungsansatz. Es fehlt jedoch eine zuverlässige nicht-invasive Diagnostik, um eine definitive Diagnose zu stellen und somit eine differenzierte Therapie anbieten zu können.

Auf der Grundlage der bisherigen Literatur zur Diagnostik von Nierentumoren stellen sich für die vorliegende Arbeit folgende Fragen:

- 1) Können Radiomics und maschinelle Lernverfahren auf CT-Daten angewendet werden, die im Rahmen der radiologischen Routinebildgebung generiert wurden?
- 2) Welche diagnostischen Genauigkeiten zur Dignitätsunterscheidung können hierbei erreicht werden?

Das Ziel dieser Arbeit ist die Entwicklung eines prädiktiven statistischen Modells, welches anhand von CT-Bildern eine Vorhersage über die Dignität eines Nierentumors treffen kann.

2 Material und Methoden

Der Ethikantrag zur folgenden Studie wurde am 24.05.2017 von der Ethikkommission der Universitätsmedizin Göttingen (UMG) bewilligt (Antragsnummer 2/4/17). Am 18.01.2019 erfolgte die Bewilligung einer zusätzlich eingereichten Ergänzung. Auf eine Einwilligung der Patienten wurde auf Grund des retrospektiven Studiendesigns verzichtet. Die Studie ist konform mit der Deklaration von Helsinki.

2.1 Zusammensetzung des Patientenkollektivs

In der vorliegenden Studie wurden retrospektiv alle erwachsenen Patienten eingeschlossen, die im Zeitraum zwischen 2012 und 2018 mit den Operationen- und Prozedurenschlüsselcodes „5-552, 5-553, 5-554“ (partielle Resektion der Niere, radikale Nephrektomie) in der Klinik für Urologie der UMG operiert wurden und für die eine präoperative Bildgebung, welche den Nierentumor erfasst hat, vorlag.

Eine Liste der Klinik für Urologie mit den Daten der operierten Patienten wurde mit den histopathologischen Diagnosen aus dem Klinikinformationssystem der UMG (ixserv, ix.mid Software Technologie GmbH, Köln) und der präoperativen Bildgebung aus dem radiologischen Bildverarbeitungsprogramm (Picture Archiving and Communications System, PACS; R1000, General Electric, Boston, MA, USA) abgeglichen.

Ausgeschlossen wurden Patienten, bei denen postoperativ histopathologisch kindliche Nierentumoren (beispielsweise Nephroblastome), nicht-tumoröse Veränderung der Niere (beispielsweise inflammatorisch verändertes Nierenparenchym, Nephropathie, Fibrose, Zysten) oder diffus infiltrative Nierentumore (beispielsweise Lymphome) diagnostiziert wurden.

Es wurden alle Tumorgrößen und -stadien eingeschlossen.

2.2 Einschlusskriterien

2.2.1 Präoperative Bildgebung

Voraussetzung für den Studieneinschluss war das Vorhandensein einer präoperativen CT-Untersuchung, welche den Nierentumor erfasst hat. Eingeschlossen wurden sowohl Aufnahmen, die im Institut für diagnostische und interventionelle Radiologie an der UMG entstanden, als auch auswärtige Untersuchungen. Bei mehreren präoperativen CT-Untersuchungen wurde mit der am aktuellsten verfügbaren präoperativen Untersuchung gearbeitet. Die CT-Untersuchung musste mit intravenösem jodhaltigem Kontrastmittel durchgeführt

und eine portal-venöse Phase akquiriert worden sein.

Die Qualität der Untersuchung beziehungsweise der entstandenen Bilder (beispielsweise Bildartefakte wie Aufhärtungen oder Bewegungen) war kein Ausschlusskriterium. Unterschiedliche Bildqualitäten waren ausdrücklich erwünscht und wurden zur Validierung der Methodik verwendet. Einziges Einschlusskriterium war die vollständige Darstellung des später operierten Tumors. Der Zeitraum zwischen Bildgebung und Operation spielte keine Rolle.

2.2.2 Histologische Diagnose

Sämtliche operierte Nierentumoren unterliefen einer standardisierten histopathologischen Analyse im Institut für Pathologie der UMG.

Mit Hilfe von immunhistochemischen Färbungen (Vimentin, Carboanhydrase IX, Cytokeratin 7, Alpha-Methylacyl-CoA Racemase, CD117, Ksp-cadherin, HMB 45, Cadhepsin K) wurden maligne (klarzellig, papillär und chromophob) und benigne (Angiomyolipom und Onkozytom) Nierentumore klassifiziert.

Diese Diagnose wurde als Goldstandard für spätere Analysen definiert.

2.3 Segmentation

Für die weitere Verarbeitung wurden die Bilddaten pseudonymisiert. Die Daten wurden zentral auf einem Server der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen gespeichert.

Die Nierentumoren der ausgewählten Patienten wurden mit Hilfe der *open source software platform* 3D Slicer segmentiert (<http://www.slicer.org>) (Fedorov et al. 2012).

Im ersten Schritt wurden die CT-Bilder der Patienten anonymisiert eingelesen und die Bilder in axialer Rekonstruktion mit der geringsten vorhandenen Schichtdicke (zwischen 0,75 und 5,0 mm) ausgewählt. Der Tumor wurde auf allen Schichten, auf denen er sichtbar ist, mit Hilfe des Werkzeugs Segmentation editor manuell als ROI markiert. Die Anzahl der markierten Schichten variierte mit der Schichtdicke und der Größe des Tumors.

Zusammenhängende Komponenten mit einer Größe von weniger als 1000 Voxeln (dreidimensionale Pixel) wurden mit dem Werkzeug remove small islands entfernt. Auf eine Glättung (*smoothing*) der Segmentationsgrenzen wurde verzichtet.

Bei Tumoren bei denen Schwierigkeiten zum Beispiel bezüglich eines möglichen Gefäßeinbruchs oder der Ausdehnung bestanden wurde eine Konsensentscheidung mit einem Radiologen mit vier Jahren Erfahrung in abdomineller Bildgebung getroffen. Die segmentierten

Tumoren wurden anschließend als „seg.nrrd“ für die Segmentation und „nrrd“ für das dazugehörige Volumen gespeichert.

Abbildung 2 zeigt Beispiele von CT-Bildern von sowohl einem malignen klarzelligem Nierenzellkarzinom als auch von einem benignen Onkozytom.

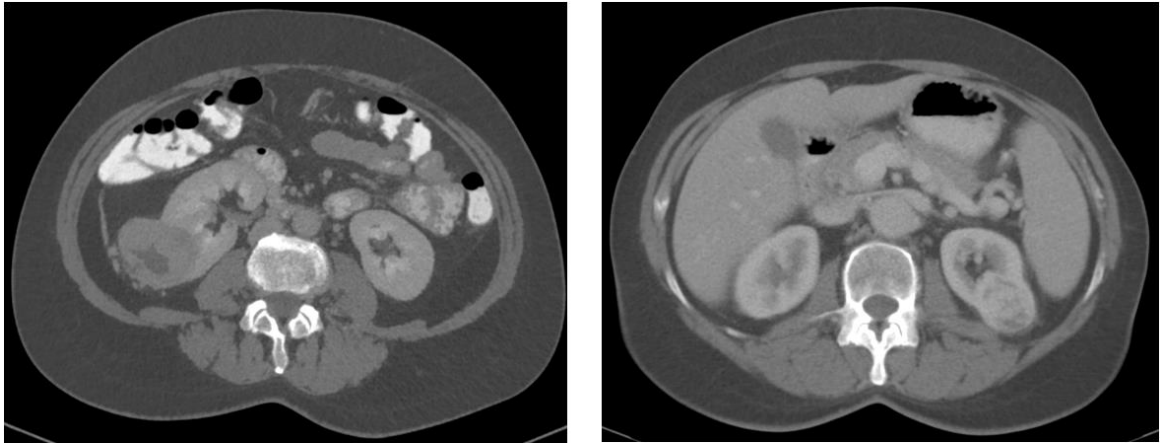


Abbildung 2: Segmentationsbeispiele. Links ein klarzelliges Nierenzellkarzinom, rechts ein Onkozytom.

2.4 Radiomic Features

Nach abgeschlossener Segmentation wurden durch das open source Paket „PyRadiomics“ innerhalb des 3D Slicers (van Griethuysen et al. 2017) insgesamt 132 Features extrahiert. Hiervon sind 107 Radiomic Features und 25 General Info Features.

Beide basieren auf standardisierten Algorithmen in Übereinstimmung mit Definitionen der Image Biomarker Standardization Initiative (IBSI) (van Griethuysen et al. 2017; Zwanenburg et al. 2020).

Für die Analyse wurde eine Bin-Weite von 25 gewählt.

Die in Tabelle 3 aufgelisteten General Info Features beinhalten unter anderem Informationen über die Größe, den Pixelabstand und den mittleren, minimalen und maximalen Voxelintensitätswert des Bildes sowieso Informationen über die Segmentation.

Tabelle 3: Auflistung der General Info Features

Subgruppe	Enthaltene Features
Versions	PyRadiomics
	Numpy
	Simple ITK
	Py Wavelet
	Python
Configuration	Settings
	Enabled Image Types
Image-original	Hash
	Dimensionality
	Spacing
	Size
	Mean
	Minimum
	Maximum
Mask-original	Hash
	Spacing
	Size
	Bounding Box
	Voxel Num
	Volumen Num
	Center of Mass Index
	Center of Mass
Mask-corrected	Spacing
	Size
	Bounding Box
	Voxel Num
	Volume Num
	Center of Mass Index
	Center of Mass
	Mean
	Minimum
	Maximum

Die insgesamt 107 unterschiedlichen Radiomic Features werden in sieben Subgruppen aufgeteilt. Tabelle 4 listet die einzelnen Features, die in den unterschiedlichen Subgruppen extrahiert wurden auf.

Tabelle 4: Auflistung der Radiomic Features nach Subgruppe

Subgruppe	Enthaltene Features
<p>First Order Statistics:</p> <ul style="list-style-type: none"> - Beschreiben geometrische Eigenschaften, beispielsweise Fläche oder Volumen der ROI (Zwanenburg et al. 2020). 	Interquartile Range
	Skewnes
	Uniformity
	Mean
	Energy
	Robust Mean Absolute
	Deviation
	Minimum
	Mean Absolute Deviation
	Total Energy
	Maximum
	Root Mean Squared
	90 Percentile
	Entropy
	Range
	Variance
<p>Shape Features (3D):</p> <ul style="list-style-type: none"> - Beschreiben dreidimensionale Größe und Form der ROI (van Griethuysen et al. 2017). 	Voxel Volume
	Maximum 3D Diameter
	Mesh Volume
	Major Axis Length
	Sphericity
	Least Axis Length
	Elongation
	Surface Volume Ratio
	Maximum 2D Diameter Slice
	Flatness

Subgruppe	Enthaltene Features
	Surface Area
	Minor Axis Length
	Maximum 2D Diameter Column
	Maximum 2D Diameter Row
<p>GLCM Features:</p> <ul style="list-style-type: none"> - beschreiben wie Kombinationen von Graustufen in benachbarten Voxeln entlang einer Bildrichtung verteilt sind (Zwanenburg et al. 2020). 	Joint Average
	Sum Average
	Joint Entropy
	Cluster Shade
	Maximum Prability
	Idmn
	Joint Energy
	Contrast
	Difference Entropy
	Inverse Variance
	Difference Variance
	Idn
	Idm
	Correlation
	Autocorrelation
	Sum Entropy
	MCC
	Sum Squares
	Cluster Prominence
	Imc2
	Imc1
	Difference Average
	Id
	Cluster Tendency
<p>GLSZM Features:</p> <ul style="list-style-type: none"> - Zählen die verknüpften (d.h. gleiche Graustufe wie benachbarte) Voxelgruppen (Zwanenburg et al. 2020). 	Gray Level Variance
	Zone Variance
	Grey Level Non Uniformity Normalized
	Size Zone Non Uniformity Normalized
	Size Zone Non Uniformity
	Grey Level Non Uniformity

Subgruppe	Enthaltene Features
	Large Area Emphasis
	Small Area High Gray Level Emphasis
	Zone Percentage
	Large Area Low Gray Level Emphasis
	Large Area High Gray Level Emphasis
	High Gray Level Zone Emphasis
	Small Area Emphasis
	Low Gray Level Zone Emphasis
	Zone Entropy
	Small Area Low Gray Level Emphasis
<p>GLRLM Features:</p> <ul style="list-style-type: none"> - Beschreiben die Verteilung von Graustufen entlang einer Lauflänge (definiert als die Länge einer zusammenhängenden Sequenz von Voxeln mit der gleichen Graustufe entlang einer Vektorrichtung) (Zwanenburg et al. 2020). 	Short Run Low Gray Level Emphasis Gray Level Variance Low Gray Level Run Emphasis Gray Level Non Uniformity Normalized Run Variance Gray Level Non Uniformity Long Run Emphasis Short Run High Gray Level Emphasis Run Length Non Uniformity Short Run Emphasis Long Run High Gray Level Emphasis Run Percentage Long Run Low Gray Level Emphasis Run Entropy High Gray Level Run Emphasis Run Length Non Uniformity Normalized
<p>NGTDM Features:</p> <ul style="list-style-type: none"> - Quantifizieren den Unterschied zwischen einer Graustufe und der durchschnittlichen Graustufe der Umgebung (van Griethuysen et al. 2017) 	Coarseness Complexity Strength Contrast Busyness
<p>GLDM Features:</p>	Gray Level Variance High Gray Level Emphasis

Subgruppe	Enthaltene Features
- Quantifizieren die Graustufen-Abhängigkeit (Anzahl an verknüpften Voxeln, die innerhalb einer definierten Distanz δ abhängig von einem zentralen Voxel sind) mittels einer Matrix. (van Griethuysen et al. 2017).	Dependence Entropy
	Dependence Non Uniformity
	Small Dependence Emphasis
	Small Dependence High Gray Level Emphasis
	Dependence Non Uniformity Normalized
	Large Dependence Emphasis
	Large Dependence Low Gray Level Emphasis
	Depence Variance
	Large Dependence High Gray Level Emphasis
	Small Dependence Low Gray Level Emphasis
	Low Gray Level Emphasis

GLCM: Gray Level Co-occurrence Matrix, GLSZM: Gray Level Size Zone Matrix, GLRLM: Gray Level Run Length Matrix, NGTDM: Neighbouring Gray Tone Difference Matrix Features, GLDM: Gray Level Dependence Matrix

2.5 Vorverarbeitung

Für einen optimalen Einsatz der maschinellen Lernverfahren (siehe Kapitel 2.13) wurde der vorhandene Datensatz zunächst vorbereitet. Es wurden Variablen, die für alle Patienten identisch und die für alle Patienten nahezu gleich waren (*near-zero-variance*) gelöscht. Außerdem wurden Duplikate und lineare Abhängigkeiten (Variablen die andere exakt erklären) gelöscht.

Multidimensionale Radiomic Features (beispielsweise eine Variable „3x5x7“) wurden gesplittet, so dass mehrere eindimensionale entstanden (beispielsweise je eine Variable mit Ausprägung „3“, „5“ und „7“).

Zur Darstellung der Korrelation einzelner Radiomic Features wurde für jede mögliche binäre Kombination von Variablen ein Korrelationskoeffizient (Spearman/Pearson) berechnet und dieser mittels *heatmap* visualisiert. Hier wird die Korrelation mit Hilfe von Farben von stark korrelierend ($\rho = 1$, rot) zu negativ korrelierend ($\rho = -1$, blau) dargestellt.

2.6 Kreuzvalidierung

Um zu verhindern, dass der Algorithmus mit bekannten Daten deutlich bessere Leistungen als mit unbekanntem erzielt (*overfitting*), wird beim Training der Algorithmen ein Teil der Daten zur späteren Validierung und Leistungsüberprüfung zurückgehalten.

Für die vorliegende Arbeit wurde eine zehnfach-Kreuzvalidierung angewendet. Diese wird in Abbildung 3 verbildlicht. Die Daten werden in zehn etwa gleich große Gruppen aufgeteilt und der ML-Algorithmus mit neun dieser Gruppen trainiert (*Trainset*). Die zehnte Gruppe (*Testset*) dient der Validierung. Dieser Vorgang wird zehnmal wiederholt und jedes Mal wird eine andere Gruppe zur Validierung zurückgehalten. Die sich so ergebenden Fehlerraten werden nun gemittelt. Bei der hier angewandten externen Kreuzvalidierung kann nun die diagnostische Genauigkeit des Algorithmus anhand der zurückgehaltenen Daten (*out-of-bag-samples*) ermittelt werden.

Die in den Punkten 2.8 bis 2.10 dargestellten Schritte werden als interne Kreuzvalidierung innerhalb der bei der externen Kreuzvalidierung erstellten neun Testgruppen durchgeführt.

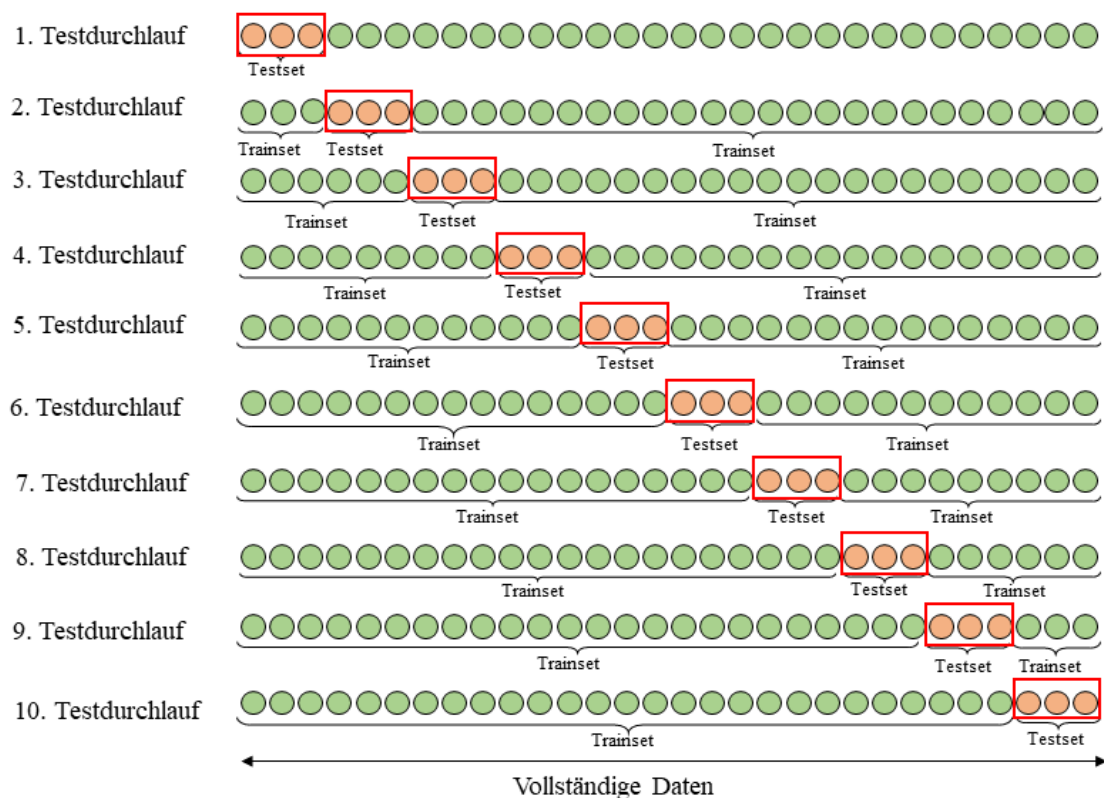


Abbildung 3: Beispiel Kreuzvalidierung. In jedem Testdurchlauf wird das Trainingsset (grün) und Testset (orange) aus jeweils anderen Daten aufgebaut, um ein *overfitting* der Algorithmen zu verhindern.

2.7 Diagnostische Genauigkeit

Die diagnostische Genauigkeit wurde bei der binären Unterscheidung zwischen malignen und benignen Nierentumoren mit der *receiver operating characteristics curve* (ROC-Kurve) dargestellt. Die Ergebnisse der Algorithmen wurden mittels der *out-of-bag-samples* berechnet und mit Hilfe einer 4-Felder-Tafel aufgetragen. Die berechneten Werte Sensitivität und Spezifität

wurden als x-Achse (1-Spezifität) und y-Achse (Sensitivität) bestimmt.

Die AUC wird als Fläche unter der ROC-Kurve berechnet. Es können hier Werte zwischen 0 und 1 erreicht werden, wobei ein Ergebnis von 0,5 das schlechteste Ergebnis ist, da hier von einer rein zufälligen Verteilung ausgegangen werden muss. Ein perfekter diagnostischer Test weist eine AUC von 1 (beziehungsweise 100%) auf.

2.8 Skalierung

Da die Algorithmen nur Zahlen erkennen und deuten können und keine Unterscheidung zwischen Einheiten möglich ist, wird eine Skalierung durchgeführt, um zu verhindern, dass Variablen mit großen Zahlenwerten eine größere Bedeutung beigemessen werden als Variablen mit niedrigeren Zahlenwerten.

2.9 Feature selection

Feature selection ist eine Technik, um die relevantesten Radiomic Features in einem Modell herauszuarbeiten. Es werden sich wiederholende oder nicht informative Radiomic Features identifiziert und ausgeschlossen. Diese können negative Auswirkungen auf die Leistung und Effektivität der Algorithmen haben (Kuhn und Johnson 2013).

Es wurde ein *backward selection* Algorithmus (*recursive feature elimination*, RFE) angewendet, bei dem das initiale Modell alle Radiomic Features enthält. Es wird anschließend berechnet, welche Radiomic Features signifikant zur Leistung des Modells beitragen und eine Rangliste erstellt. Dies wird auch mit Hilfe einer separaten zehnfach-Kreuzvalidierung (siehe Kapitel 2.8) durchgeführt. Die Radiomic Features werden aufgeteilt und dann jeweils die diagnostische Genauigkeit ermittelt. Die am wenigsten relevanten Features werden gelöscht (Guyon et al. 2002).

2.10 Ausgleich des Klassenungleichgewichts

Die Studie umfasst deutlich mehr Patienten mit malignen als mit benignen Nierentumoren. Dieses Klassenungleichgewicht (*class imbalance*) sorgt potenziell für eine schlechtere Leistung der ML-Algorithmen. Folgen sind häufig eine gute Spezifität, aber eine schlechte Sensitivität (Kuhn und Johnson 2013).

Die synthetic minority over-sampling technique (SMOTE) (Chawla et al. 2002) ist eine Möglichkeit dieses Ungleichgewicht auszugleichen. Um mehr Fälle für die Minderheitsgruppe (*minority class*) zu generieren, synthetisiert SMOTE neue Daten.

Es wird zufällig ein Datenpunkt in der Minderheitsgruppe ausgewählt und die am nächsten liegenden Nachbardaten ermittelt. Die neu synthetisierten Daten sind zufällige Kombinationen aus dem Datenpunkt und den Nachbardaten (Kuhn und Johnson 2013).

2.11 Machine learning

Wolperts „no free lunch theorem“ (Wolpert 1996) besagt, dass es kein universell gutes maschinelles Lernverfahren zum Lösen eines Problems gibt und ein Algorithmus, der in einem Teilbereich bessere Leistungen liefert als ein anderer dafür in einem anderen Bereich häufig schlechter ist. Basierend auf dieser Annahme wurden in der vorliegenden Arbeit mehrere unterschiedliche ML-Algorithmen eingesetzt.

2.11.1 Random forest

Der *random forest*-Algorithmus basiert auf Entscheidungsbäumen, die Daten mit Hilfe von „wenn-dann“ Entscheidungen in Klassen aufteilen (beispielsweise: wenn $x > 3$ dann Klasse A, wenn $x < 3$ dann Klasse B). Abbildung 4 zeigt eine Beispielarchitektur eines Entscheidungsbaums.

Ein Baum besteht aus einer Wurzel (*root*), die die Anfangsdaten enthält und Ästen (*branches*) die zu Knoten (*nodes*) führen. Ein Knoten, bei dem keine weitere Unterscheidung möglich ist wird Endknoten (*leaf node*) genannt. Das Ziel ist es, Daten in kleine und homogene Gruppen aufzuteilen. Das heißt: In jeder Entscheidung wird der Test ausgewählt, der die Gruppe am homogensten unterteilt.

Dies wird so oft durchgeführt, bis keine weitere Unterteilung möglich ist.

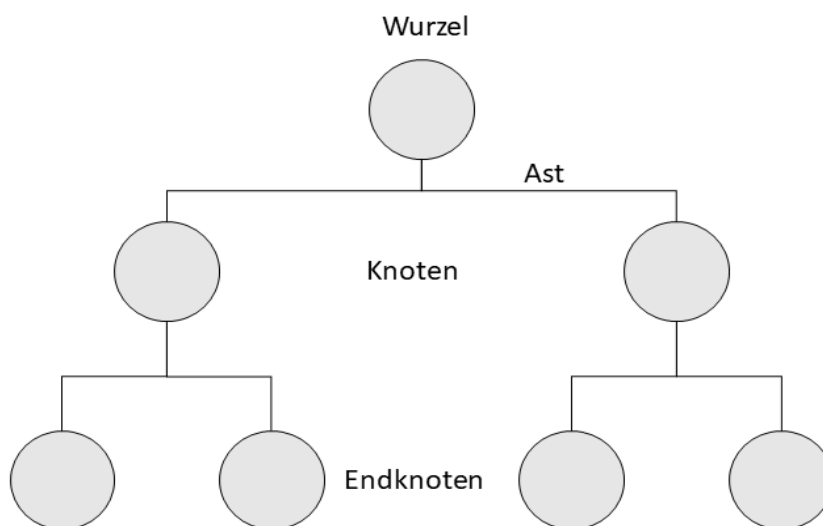


Abbildung 4: Beispiel Entscheidungsbaum. Die Wurzel enthält alle vorhandenen Daten, die durch „wenn-dann“ Entscheidungen in Klassen (Knoten) unterteilt wird. Am Endknoten ist keine weitere Unterscheidung möglich.

Der in Abbildung 5 gezeigte rf-Algorithmus erstellt mehrere Entscheidungsbäume für die Daten und mittelt die entstanden Vorhersagen. Die Bäume werden als *bootstrap samples* erstellt. Das bedeutet, dass alle Bäume mit der gleichen Menge Daten wie der Ursprungsdatensatz arbeiten, die Daten jedoch zufällig ausgewählt werden und danach auch für weitere Bäume zur Verfügung stehen. So werden einige Daten mehrfach, andere dagegen gar nicht berücksichtigt.

Es wird für jeden Baum nur ein Teil der Radiomic Features zufällig bereitgestellt (meist \sqrt{n} mit $n = \text{Anzahl aller Features}$) um eine zu hohe Gewichtung einflussreicher Features zu verhindern. An jedem Knotenpunkt wird entschieden welches Radiomic Feature am besten zur Unterscheidung von malignen und benignen Tumoren geeignet ist.

Der zu beeinflussende Parameter zur Leistungsverbesserung des Algorithmus (*tuning parameter*) ist *mtry*. Dieser beschreibt die Anzahl der Radiomic Features, die an jedem Knoten zufällig ausgewählt werden.

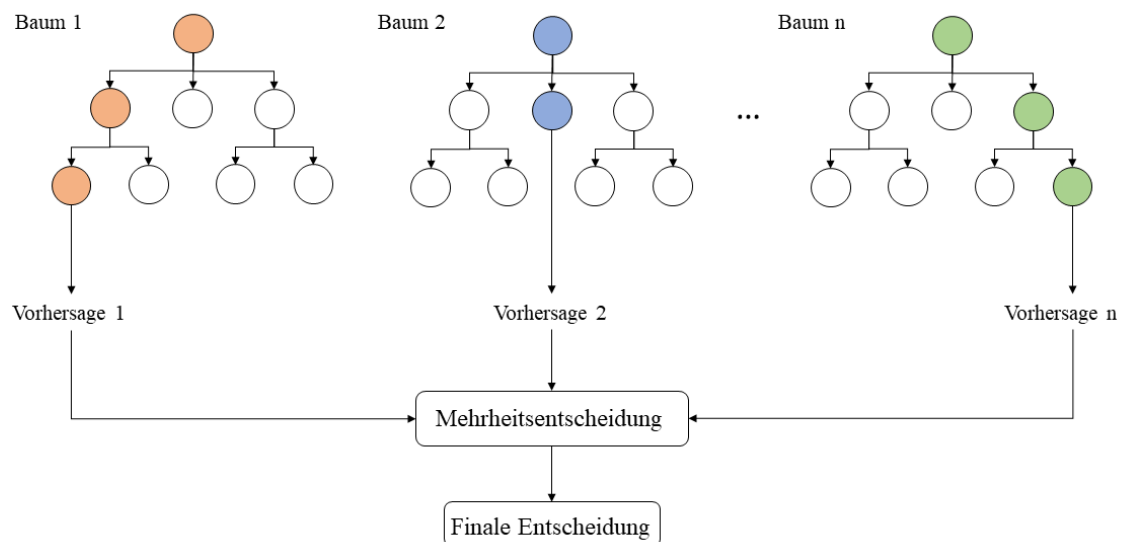


Abbildung 5: Beispiel *random forest*-Algorithmus. Jeder Entscheidungsbaum trifft eine Vorhersage aus denen dann mittels Mehrheitsentscheidung eine finale Vorhersage getroffen wird.

2.11.2 Neural network

Neuronale Netzwerke (nnet) imitieren im weitesten Sinne biologische Systeme aus Nervenzellen. Abbildung 6 zeigt die Struktur des Algorithmus, in dem es einen Input gibt, der in der verborgenen Ebene (*hidden layer*) verarbeitet wird und dann zu einem Output führt. Je größer die Anzahl der Ebenen ist, desto differenzierter ist der Algorithmus.

Jede Ebene (*layer*) besteht aus mehreren Knoten (*nodes*). Die Verbindungen zwischen den verschiedenen Knoten werden unterschiedlich gewichtet. Je höher die Gewichtung ist, desto wichtiger ist diese Verbindung für das Ergebnis.

Die initiale Gewichtung erfolgt willkürlich, da das optimale Gewicht im Vorhinein nicht bekannt ist. Der Algorithmus passt diese auf Grundlage von falschen Vorhersagen mit früheren Gewichtungen an.

Die beeinflussbaren Parameter sind *size*, womit die Anzahl an Knoten in den *hidden layers* bestimmt wird und *decay*, womit die gelernte Gewichtung stabilisiert wird, so dass Ausreißerdaten weniger stark gewichtet werden und der Einfluss auf die Leistung vermindert wird.

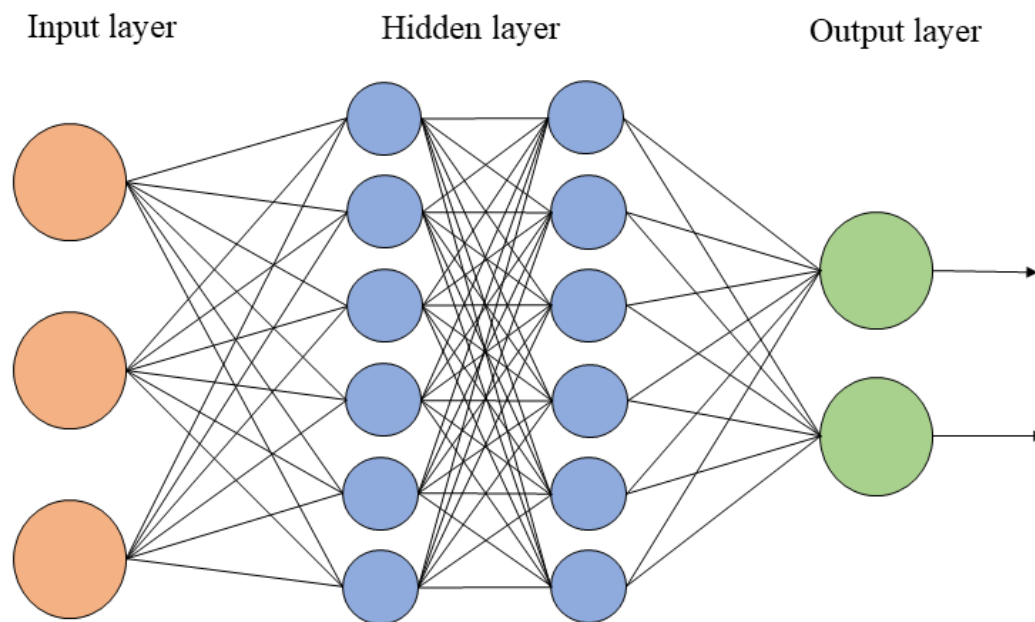


Abbildung 6: Beispiel *neural network*-Algorithmus. Die verfügbaren Daten werden als *input layer* eingefügt, in mehreren *hidden layers* verarbeitet und das Ergebnis als *output layer* präsentiert.

2.11.3 Support vector machine

Support vector machines kreieren mit Hilfe eines Radialbasisfunktionskern einen multidimensionalen Raum, in dem jedes Radiomic Feature eine Dimension darstellt. Eine Hyperebene (mehrdimensionale Ebene) teilt, wie in Abbildung 7 dargestellt, zwei Gruppen ein und versucht den größtmöglichen Abstand (*margin*) zwischen beiden zu erzielen. Die Daten, die auf beiden Seiten am nächsten am Rand liegen, werden Hilfsvektoren (*support vectors*) genannt.

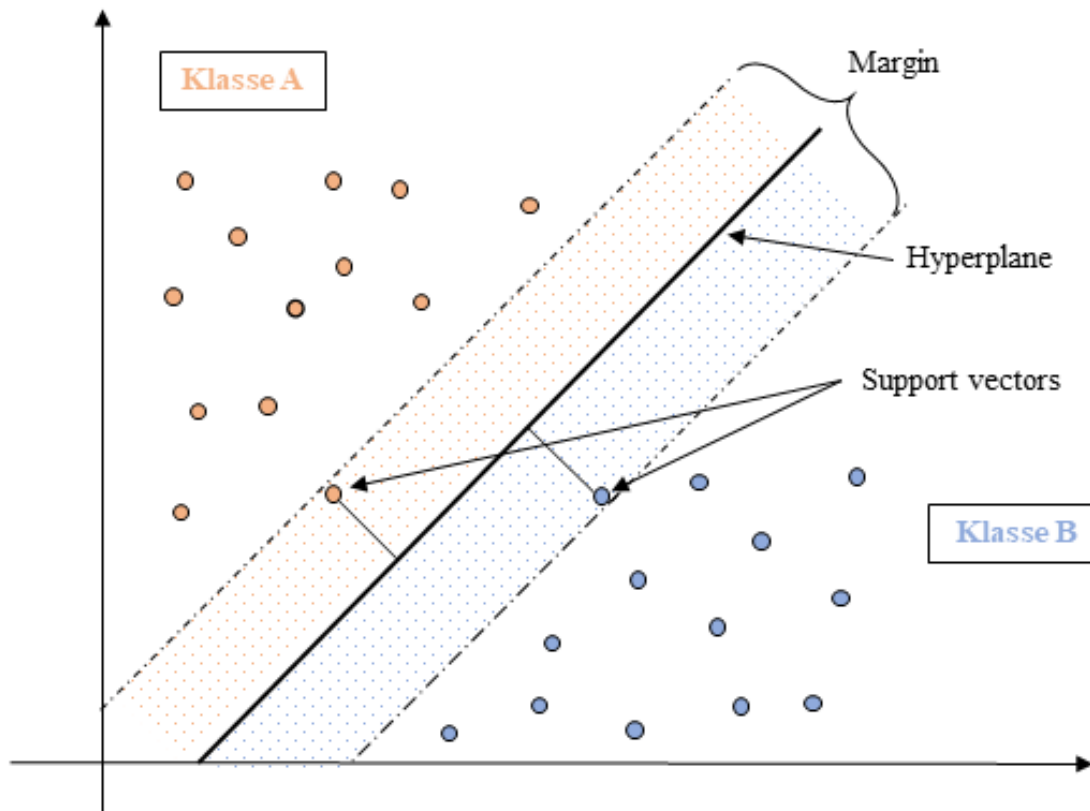


Abbildung 7: Beispiel *support vector machine*-Algorithmus. Die Daten werden durch die Hyperplane in zwei Gruppen mit dem größtmöglichen Abstand eingeteilt.

2.11.4 K-nearest-neighbors

Die Grundidee des *k-nearest-neighbors*-Algorithmus (knn) ist, dass ähnliche Datenpunkte sich räumlich nah sind. Die bekannten Daten werden in einem multidimensionalen Raum (eine Dimension für jede Variable) aufgeteilt und unbekannte Daten mit den am nächsten liegenden Datenpunkten verglichen. Die Anzahl der Datenpunkte wird durch den *tuning parameter* k festgelegt (k ist hierbei eine willkürlich festgelegte Zahl, je höher k ist, desto genauer ist die Vorhersage). Die Entscheidung zu welcher Gruppe die unbekanntenen Daten gehören ist nun eine Mehrheitsentscheidung (beispielsweise: $k =$ zehn, von den zehn nächstgelegenen Daten sind acht maligne und zwei benigne, das bedeutet die unbekanntenen Daten sind maligne). Abbildung 8 zeigt eine Beispieldarstellung mit einem *tuning parameter* von drei.

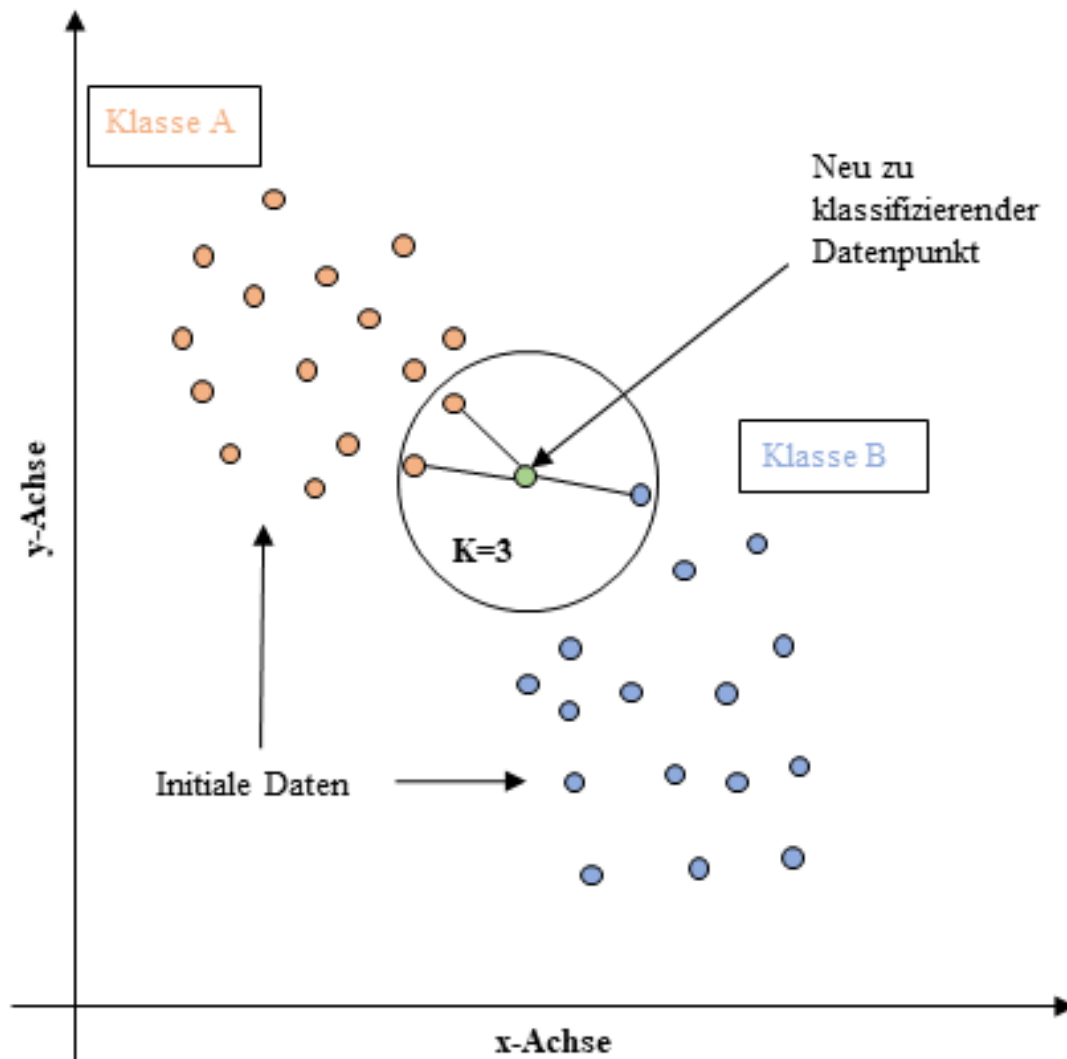


Abbildung 8: Beispiel *k-nearest-neighbors*-Algorithmus. Bekannte Daten werden in einem multidimensionalen Raum aufgeteilt und die neu zu klassifizierenden Daten werden mit den am nächsten liegenden bekannten verglichen.

2.12 Statistische Analysen

Alle statistischen Analysen dieser Arbeit wurden mit der Software R (Version 3.6.0, R core development team, Wien, Österreich) und RStudio (Version 1.3.959, RStudio Inc., Boston, MA, USA) und dem R Paket caret durchgeführt (Kuhn 2008).

Es wurden unterschiedliche Signifikanztests durchgeführt. Für die metrische Unterscheidung der Variablen (Alter in Jahren und maximaler 3D Durchmesser in mm) wurde der *Welch Two Sample t-test* angewendet.

Für die nominale Unterscheidung zweier Variablen (Geschlecht) wurde *Fisher's Exact Test* verwendet.

Die nominale Unterscheidung mehrerer Variablen (Subtypen) wurde mit Hilfe von *Pearson's Chi-squared test* durchgeführt.

Alle p-Werte sind zweiseitig. Es wurde ein Alpha Level von 0,05 als statistisch signifikant festgesetzt.

3 Ergebnisse

3.1 Patientenkollektiv

Abbildung 9 zeigt den Prozess des Patienteneinschlusses. Von ursprünglich 543 Patienten wurden 342 auf Grund von fehlenden präoperativen CT-Bildern und einer, nicht die Einschlusskriterien treffenden, Histologie ausgeschlossen.

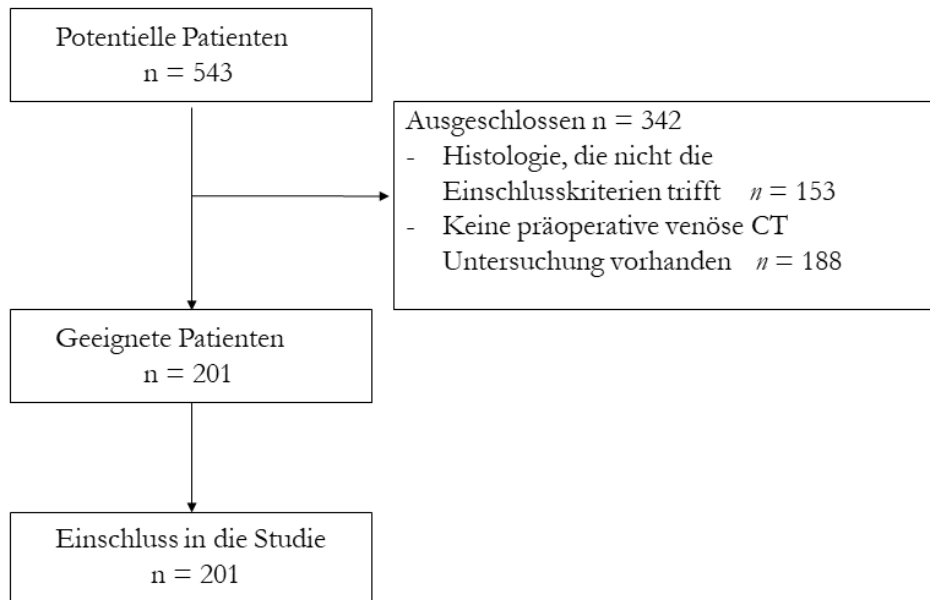


Abbildung 9: Flowchart Patienteneinschluss mit Überblick über die Ein- und Ausschlusskriterien und insgesamt eingeschlossenen Patienten.

Wie in Tabelle 5 gezeigt, waren unter den 201 eingeschlossenen Patienten 78 Frauen (38,8 %) und 123 Männer (61,2 %). Die histologische Analyse ergab insgesamt 171 (85 %) maligne Tumoren. Hiervon waren 131 (65,2 %) klarzellige-, 29 (14,4 %) papilläre- und 11 (5,5 %) chromophobe Nierenzellkarzinome. 30 Tumoren (15 %) zeigten in der histologischen Analyse eine benigne Entität, darunter waren 14 (7 %) Angiomyolipome und 16 (8 %) Onkozytome. 76 % der Tumoren (insgesamt 154 Tumoren) entsprachen einem Stadium T1 (≤ 70 mm). 28 % der eingeschlossenen Tumoren (58 Tumoren) entsprachen, mit einer Größe von ≤ 40 mm, einem Tumorstadium T1a. Das mittlere Alter der Studienkohorte betrug 65 Jahre (SD ± 11 Jahre), wobei Patienten mit malignen Tumoren im Schnitt älter waren als Patienten mit benignen Tumoren (Mittelwert 66 versus 60 Jahre, $p < 0,01$).

Der Unterschied im maximalen 3D Durchmesser der Tumoren war statistisch nicht signifikant ($p = 0,56$) und betrug durchschnittlich 58 mm (± 28) für maligne Tumoren und 55 mm (± 32) für benigne Tumore.

Tabelle 5: Charakteristika der Studienkohorte

Parameter		Total	Maligne Tumoren	Benigne Tumoren	p-Wert
n		201	171	30	
Alter (in Jahren)					0,02
	Mittelwert \pm SD	65 \pm 11	66 \pm 11	60 \pm 12	
	Median (min; max)	66 (31;86)	68 (31; 86)	63 (31; 77)	
Geschlecht					< 0,01
	Weiblich	78 (38,8 %)	58 (33,9 %)	20 (66,7 %)	
	Männlich	123 (61,2 %)	113 (66,1 %)	10 (33,3 %)	
Subtyp					< 0,01
	Klarzelliges Nierenzellkarzinom	131 (65,2 %)	131 (76,6 %)	0 (0,0 %)	
	Papilläres Nierenzellkarzinom	29 (14,4 %)	29 (17 %)	0 (0,0 %)	
	Onkozytom	16 (8 %)	0 (0,0 %)	16 (53,3 %)	
	Angiomyolipom	14 (7 %)	0 (0,0 %)	14 (46,7 %)	
	Chromophobes Nierenzellkarzinom	11 (5,5 %)	11 (6,4 %)	0 (0,0 %)	
Maximaler 3D Durchmesser (in mm)					0,56
	Mittelwert \pm SD	58 \pm 28	58 \pm 28	55 \pm 32	
	Median (min; max)	52 (13; 192)	54 (16; 192)	48 (13; 141)	

n = Anzahl SD = Standardabweichung min = minimal max = maximal

Tabelle 6 zeigt die Unterschiede in der eingeschlossenen Bildgebung. Der Großteil (83,1 %) der verwendeten CT-Untersuchungen wurde in externen Instituten durchgeführt. Hier traten häufiger Bildartefakte auf als bei Aufnahmen, die an der UMG durchgeführt wurden (34 % versus 0,06 %, $p < 0,01$).

Auch die Schichtdicke war bei einer externen Bildgebung höher als bei Untersuchungen an der UMG. Der Median betrug hier 3 mm, während er für Untersuchungen an der UMG bei 1 mm lag.

Die Schichtdicke war bei Untersuchungen mit Bildartefakten außerdem signifikant höher als bei Untersuchungen ohne Bildartefakte (5 mm versus 1,2 mm, $p < 0,01$).

Während die verwendeten Scanner an den externen Instituten nicht bekannt sind, wurde die Bildgebung in der UMG an einem der folgenden vier Siemens Scanner durchgeführt: Somatom Force, Somatom Flash, Somatom Definition AS+ und Sensation 16.

Tabelle 6: Zusammenfassung über Bildartefakte und den Ort der Bildgebung

	Level	Total	Bildartefakte vorhanden	Keine Bildartefakte vorhanden	p-Wert
n		201	60	141	
Ort der Bildgebung	Extern	167 (83,1 %)	58 (96,7 %)	109 (77,3 %)	<0,01
	Universitätsmedizin Göttingen	34 (16,9 %)	2 (3,3 %)	32 (22,7 %)	

n = Anzahl

3.2 Vorbereitung der Algorithmen

3.2.1 Vorverarbeitung

Es wurden zunächst 129 Radiomic Features extrahiert. Nach dem Aufteilen der multidimensionalen Features (Spacing, Size, Spacing.1, Size.1, BoundingBox, CenterOfMass, Spacing.2, Size.2, BoundingBox.1, CenterOfMassIndex.1, CenterOfMass.1, CenterOfMassIndex) ergaben sich insgesamt 159 eindimensionale Radiomic Features.

Tabelle 7 zeigt die 125 Radiomic Features die nach dem Löschen von identischen und *near-zero variance* Variablen (Size.dim.1, Size.dim.2, Minimum) für die weiteren Auswertungen verwendet wurden. In Abbildung 10 werden die korrelierenden Radiomic Features mittels einer *Heatmap* dargestellt.

Tabelle 7: Auflistung der final verwendeten Radiomic Features

Bezeichnung der final verwendeten Radomic features	
Spacing.dim.1	DependenceNonUniformity
Spacing.dim.2	GrayLevelNonUniformity
Spacing.dim.3	SmallDependenceEmphasis
Size.dim.3	SmallDependenceHighGrayLevelEmphasis
Mean	DependenceNonUniformityNormalized
Maximum	LargeDependenceEmphasis

Bezeichnung der final verwendeten Radomic features	
Size.1.dim.1	LargeDependenceLowGrayLevelEmphasis
Size.1.dim.2	DependenceVariance
Size.1.dim.3	LargeDependenceHighGrayLevelEmphasis
VoxelNum	SmallDependenceLowGrayLevelEmphasis
CenterOfMassIndex.dim.1	LowGrayLevelEmphasis
CenterOfMassIndex.dim.2	JointAverage
CenterOfMassIndex.dim.3	SumAverage
CenterOfMass.dim.1	JointEntropy
CenterOfMass.dim.2	ClusterShade
CenterOfMass.dim.3	MaximumProbability
BoundingBox.1.dim.1	Idmn
BoundingBox.1.dim.2	JointEnergy
BoundingBox.1.dim.3	Contrast
Mean.1	DifferenceEntropy
Minimum.1	InverseVariance
Maximum.1	DifferenceVariance
VoxelVolume	Idn
Maximum3DDiameter	Idm
MeshVolume	Correlation
MajorAxisLength	Autocorrelation
Sphericity	SumEntropy
LeastAxisLength	MCC
Elongation	SumSquares
SurfaceVolumeRatio	ClusterProminence
Maximum2DDiameterSlice	Imc2
Flatness	Imc1
SurfaceArea	DifferenceAverage

Bezeichnung der final verwendeten Radomic features	
MinorAxisLength	Id
Maximum2DDiameterColumn	ClusterTendency
Maximum2DDiameterRow	InterquartileRange
GrayLevelVariance	Skewness
HighGrayLevelEmphasis	Uniformity
DependenceEntropy	Median
Energy	RunEntropy
RobustMeanAbsoluteDeviation	HighGrayLevelRunEmphasis
MeanAbsoluteDeviation	ZoneVariance
TotalEnergy	GrayLevelNonUniformityNormalized.1
RootMeanSquared	SizeZoneNonUniformityNormalized
X90Percentile	SizeZoneNonUniformity
Entropy	GrayLevelNonUniformity.2
Variance	LargeAreaEmphasis
X10Percentile	SmallAreaHighGrayLevelEmphasis
Kurtosis	ZonePercentage
ShortRunLowGrayLevelEmphasis	LargeAreaLowGrayLevelEmphasis
GrayLevelVariance.1	LargeAreaHighGrayLevelEmphasis
LowGrayLevelRunEmphasis	HighGrayLevelZoneEmphasis
GrayLevelNonUniformityNormalized	SmallAreaEmphasis
RunVariance	LowGrayLevelZoneEmphasis
GrayLevelNonUniformity.1	ZoneEntropy
LongRunEmphasis	SmallAreaLowGrayLevelEmphasis
ShortRunHighGrayLevelEmphasis	Coarseness
RunLengthNonUniformity	Complexity
ShortRunEmphasis	Strength
LongRunHighGrayLevelEmphasis	Contrast.1

Tabelle 8: Charakteristika der Studienkohorte nach der Anwendung von SMOTE

Parameter		Total	Maligne Tumoren	Benigne Tumoren	p-Wert
n		261	171	90	
Alter (in Jahren)					0,02
	Mittelwert \pm SD	64 \pm 11	66 \pm 11	60 \pm 9,8	
	Median (min; max)	64 (31; 86)	68 (31; 86)	62 (31; 77)	
Geschlecht					< 0,01
	Weiblich	112 (42,9 %)	58 (33,9 %)	54 (60 %)	
	Männlich	149 (57,1 %)	113 (66,1 %)	36 (40 %)	
Subtyp					< 0,01
	Klarzelliges Nierenzellkarzinom	131 (50,2 %)	131 (76,6 %)	0 (0,0 %)	
	Papilläres Nierenzellkarzinom	29 (11,1 %)	29 (17 %)	0 (0,0 %)	
	Onkozytom	52 (19,6 %)	0 (0,0 %)	52 (57,8 %)	
	Angiomyliplom	38 (14,6 %)	0 (0,0 %)	38 (42,2 %)	
	Chromophobes Nierenzellkarzinom	11 (4,2 %)	11 (6,4 %)	0 (0,0 %)	
Maximaler 3D Durchmesser (in mm)					0,33
	Mittelwert \pm SD	57 \pm 28	58 \pm 28	55 \pm 29	
	Median (min; max)	50 (13; 192)	54 (16; 192)	48 (13; 141)	

n = Anzahl SD = Standardabweichung min = minimal max = maximal

3.3 Leistung der Algorithmen

Nach Training der unterschiedlichen ML-Algorithmen in Kombination mit SMOTE und RFE ergaben sich in den *out-of-bag-samples* die in Tabelle 9 detailliert aufgeführten diagnostischen Genauigkeiten.

Generell ist zu beobachten, dass die diagnostische Genauigkeit der meisten Algorithmen ohne SMOTE besser ist als mit. Nur bei dem svmRadial-Algorithmus kann ein besseres Ergebnis erzielt werden (68,6 % vor SMOTE versus 69,6 % danach). Die RFE führt im Fall

des knn-Algorithmus und des smvRadial-Algorithmus zu einer Verbesserung von jeweils etwa 10 %. Abbildung 11 zeigt die diagnostische Genauigkeit der einzelnen Algorithmen in Abhängigkeit von der verwendeten Optimierungsmethode.

Tabelle 9: Übersicht über die Leistung der Algorithmen

<i>Imbalance handling</i>	<i>Feature selection</i>	<i>ML-Algorithmus</i>	<i>Mediane AUC</i>
<i>None</i>	<i>None</i>	knn	50,5 %
<i>None</i>	<i>None</i>	nnet	78,9 %
<i>None</i>	<i>None</i>	rf	73 %
<i>None</i>	<i>None</i>	svmRadial	68,6 %
<i>None</i>	RFE	knn	60,8 %
<i>None</i>	RFE	nnet	72,4 %
<i>None</i>	RFE	rf	68,6 %
<i>None</i>	RFE	svmRadial	78,4 %
SMOTE	<i>None</i>	knn	47,5 %
SMOTE	<i>None</i>	nnet	70,6 %
SMOTE	<i>None</i>	rf	67,5 %
SMOTE	<i>None</i>	svmRadial	69,6 %
SMOTE	RFE	knn	46,6 %
SMOTE	RFE	nnet	66,7 %
SMOTE	RFE	rf	61,8 %
SMOTE	RFE	svmRadial	73,5 %

SMOTE: *synthetic minority oversampling technique* RFE: *recursive feature elimination*,

knn: *k-nearest-neighbors-Algorithmus*, nnet: *neural network-Algorithmus*, rf: *random forest-Algorithmus*, svmRadial: *support vector machine-Algorithmus*

Der Algorithmus mit der besten Leistung ist der *neural network-Algorithmus*. Hier zeigt sich eine AUC von 78,9 % ohne SMOTE und RFE, welche als ROC-Kurve in Abbildung 12 dargestellt ist.

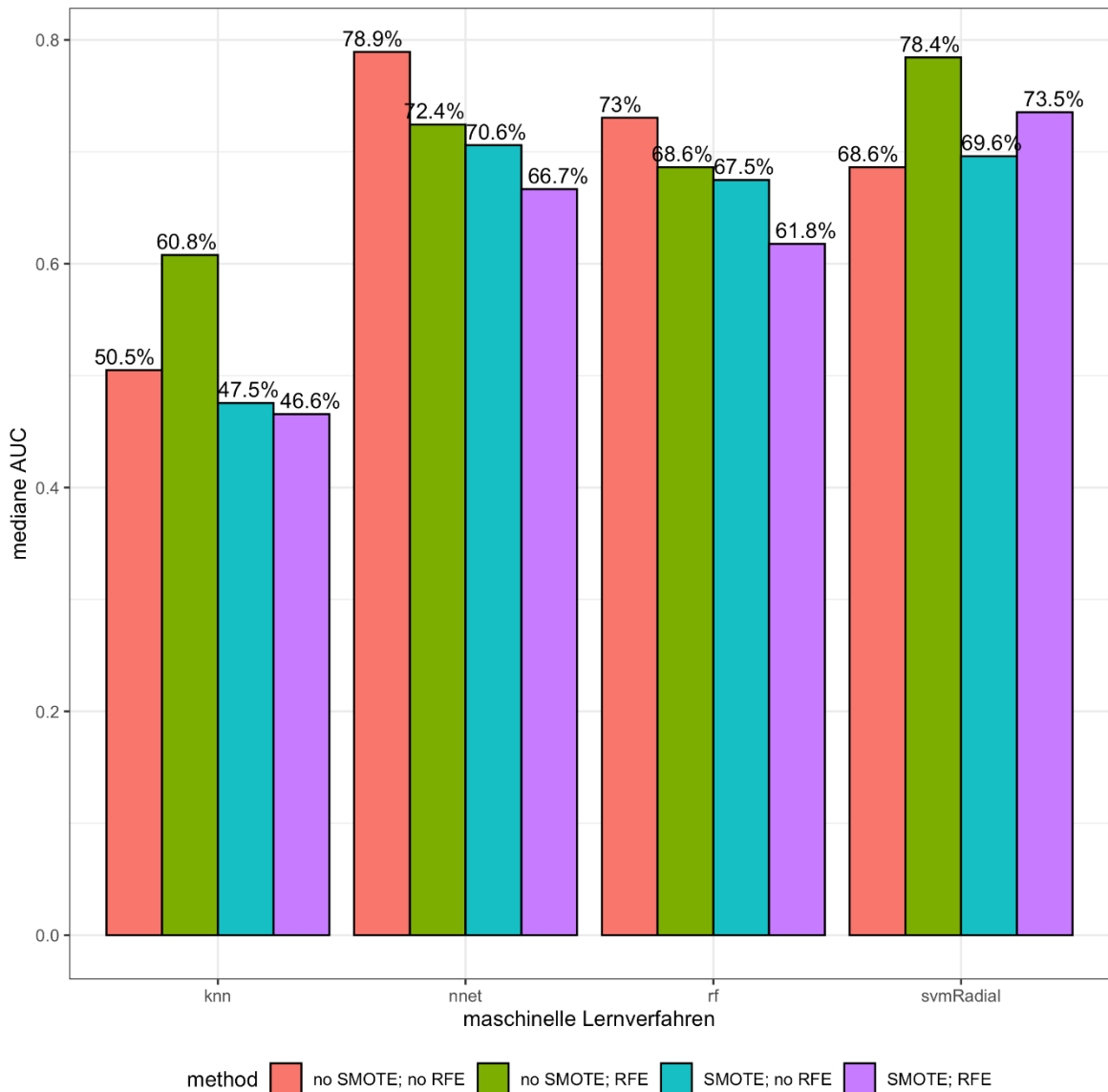


Abbildung 11: Mediane AUC der Algorithmen. Die Abbildung zeigt die medianen AUC Werte aufgeschlüsselt nach den verwendeten maschinellen Lernverfahren und den angewendeten Methoden SMOTE: *synthetic minority oversampling technique*, RFE: *recursive feature elimination*, knn: *k-nearest-neighbors*-Algorithmus, nnet: *neural network*-Algorithmus, rf: *random forest*-Algorithmus, svmRadial: *support vector machine*-Algorithmus

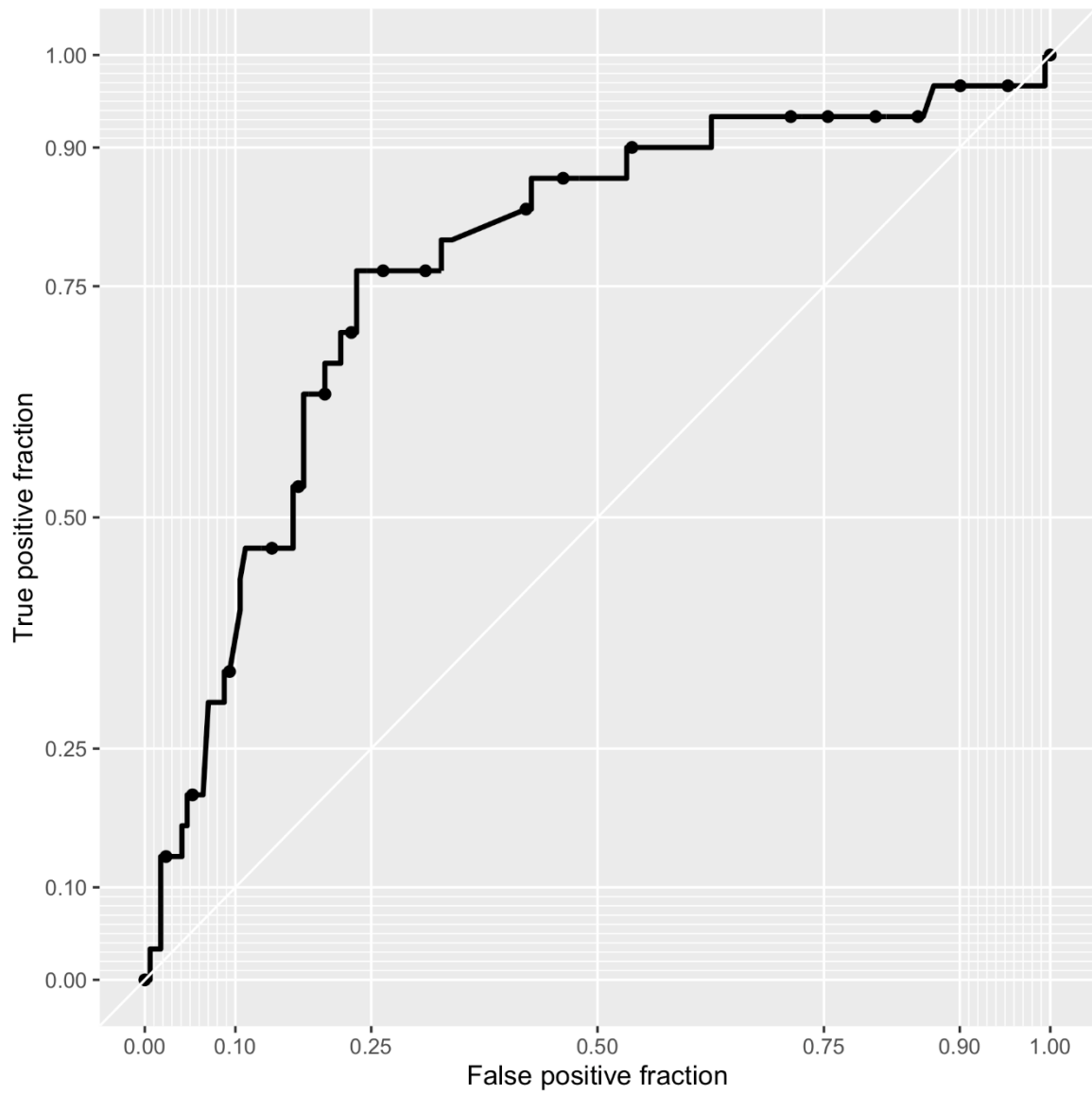


Abbildung 12: ROC-Kurve des nnet-Algorithmus ohne SMOTE und RFE

4 Diskussion

Die Inzidenz von Nierentumoren steigt durch den Einsatz moderner Bildgebungsverfahren kontinuierlich an (Ljungberg et al. 2019). Dennoch ist eine sichere Unterscheidung von malignen und benignen Tumoren anhand der radiologischen Bildgebung weiterhin nicht möglich (Millet et al. 2011).

Eine unmittelbare Konsequenz ist eine Übertherapie von Patienten mit gutartigen Nierentumoren (Kim et al. 2019). Ein potenzieller Lösungsansatz, um Patienten eine invasive Diagnostik oder möglicherweise unnötige Operationen zu ersparen, ist der Einsatz von fortgeschrittenen radiologischen Analyseverfahren wie Radiomic Features und maschinellen Lernverfahren (Kunapuli et al. 2018; Sun et al. 2020). Die Annahme hierbei ist, dass radiologische Bilder Informationen enthalten, die für das menschliche Auge nicht sichtbar sind, aber wichtige Hinweise für die Unterscheidung zwischen benignen und malignen Nierentumoren liefern können.

In der vorliegenden retrospektiven Studie wurde die Möglichkeit einer nicht-invasiven Diagnosehilfe zur Unterscheidung von malignen und benignen Nierentumoren untersucht. Es wurden 201 Patienten mit malignen und benignen Nierentumoren eingeschlossen und anhand von realistischen, nicht standardisierten CT-Untersuchungen mehrere ML-Algorithmen auf die Dignitätsunterscheidung trainiert. Die beste diagnostische Genauigkeit konnte durch einen *neural network*-Algorithmus mit einer AUC von 78,9 % erreicht werden.

Der Fokus dieser Studie lag zum einen auf dem expliziten Einschluss von CT-Daten jeglicher Qualität, um ein möglichst realitätsnahes Szenario zu kreieren und so die klinische Generalisierbarkeit der Ergebnisse zu gewährleisten.

Zum anderen wurde bewusst die Entscheidung getroffen die Unterscheidung zwischen malignen und benignen Tumoren und nicht zwischen histologischen Tumorsubtypen zu untersuchen, da dies die größte klinische Relevanz hat. Maligne Tumoren jeglicher Subtypen werden in den meisten Fällen operiert, so dass eine genaue Klassifikation des Tumors zwar Gewissheit über Prognose und Aggressivität liefert, an der ersten Therapiestrategie jedoch nichts ändert.

Ähnlich verhält es sich mit benignen Tumoren. Während auch hier einige Tumoren auf Grund von Beschwerden oder der Tumorgroße operativ entfernt werden, so ist doch bei histologisch gesicherten benignen Tumoren die aktive Überwachung eine geeignete Strategie. Um dementsprechend sowohl eine Überversorgung benigner Tumoren als auch eine Unter-versorgung maligner Tumoren zu vermeiden, ist die Dignitätsunterscheidung das in unseren Augen wichtigste Ziel in der Diagnostik von Nierentumoren.

4.1 Zusammensetzung des Patientenkollektivs

Die Zusammensetzung des Patientenkollektivs dieser Arbeit ist vergleichbar mit den in der Literatur beschriebenen Daten. Sowohl die Geschlechterverteilung von circa 60 % Männern und 40 % Frauen als auch das mediane Alter von 65 Jahren wurden sowohl für Deutschland als auch für Europa in früheren Studien beschrieben (Ferlay et al. 2013; Ljungberg et al. 2019; Robert Koch-Institut 2019).

76 % der eingeschlossenen Tumoren entsprachen einem Tumorstadium T1. Dies ist etwas mehr als in der Literatur beschrieben (meist um die 50 %), spiegelt aber möglicherweise die steigende Inzidenz von in niedrigen Stadien diagnostizierten Tumoren wider (Kane et al. 2008). Da in unserer Studie nur operierte Patienten eingeschlossen wurden, kann auch eine mögliche Verzerrung durch Patienten mit weit fortgeschrittenen Tumoren, die nicht mehr operiert wurden, nicht ausgeschlossen werden.

4.2 Histopathologische Diagnosen

In unserer Studiengruppe betrug das Verhältnis von malignen zu benignen Tumoren 85 % zu 15 %. Dies ist eine geringere Zahl von benignen Nierentumoren im Vergleich zu der in der Literatur berichteten Inzidenz von bis zu 30 % (Kim et al. 2019).

Dies kann einerseits auf die möglicherweise unterschiedlichen Studienpopulationen zurückzuführen sein, andererseits auf die Tatsache, dass in unsere Studie nur Patienten eingeschlossen wurden, die einer Operation unterzogen wurden. Das bedeutet, dass diese Patienten bereits eine Beurteilung ihrer radiologischen Bildgebung erhielten, die einen möglicherweise malignen Tumor nicht ausschließen konnte. Patienten mit eindeutigen Hinweisen auf einen benignen Tumor, wie zum Beispiel beim klassischen, fettreichen Angiomyolipom ohne Hinweis auf eine erhöhte Blutungsgefahr, erhalten meist keine Operation und wurden somit nicht in unsere Studie eingeschlossen.

4.3 Bildgebung

Der Einschluss in unsere Studie erfolgte ausdrücklich mit CT-Daten jeglicher Qualität, um einerseits die klinische Realität widerzuspiegeln und andererseits eine hohe Generalisierbarkeit unserer Ergebnisse zu gewährleisten. Aus den gleichen Gründen wurde für die vorliegende Arbeit auch nur die portalvenöse Kontrastmittelpphase der verfügbaren CT-Untersuchungen ausgewertet. Dies ist das, unserer Erfahrung nach, häufigste CT-Protokoll in der klinischen Praxis bei dem akzidentelle Nierentumoren detektiert werden.

Für eine vollständige Beurteilung von Nierentumoren wird idealerweise eine mehrphasige CT-Untersuchung durchgeführt (Sheth et al. 2001). Da die meisten Nierentumoren allerdings akzidentelle Befunde sind, ist die Bildgebung häufig nicht für Nierentumoren optimiert. Das heißt; Wichtige Kontrastmittelphasen zur Beurteilung fehlen und auch Bildartefakte sind häufig. Da eine Wiederholung der Untersuchung einerseits Ressourcen bindet und Patienten andererseits einer weiteren Strahlenbelastung aussetzt, sollte diese, wenn möglich, verhindert werden.

Die CT-Untersuchungen in unserer Studie wurden sowohl an externen Zentren als auch im Institut für diagnostische und interventionelle Radiologie der UMG angefertigt. Da keine vorgegebenen und einheitlichen Untersuchungsprotokolle verwendet wurden, zeigen sich große Unterschiede in der Schichtdicke (zwischen 1 und 5 mm). Etwa 30 % der Untersuchungen enthielten zudem Bildartefakte, die möglicherweise großen Einfluss auf die Auswertbarkeit durch maschinelle Algorithmen haben können. Der Großteil der Artefakte trat bei Untersuchungen an externen Instituten auf. Die genauen Gründe für die Unterschiede in Schichtdicke und der Häufigkeit von Bildartefakte sind nicht bekannt, möglich wäre aber einerseits die Verwendung unterschiedlicher CT-Scanner und andererseits eine größere Anzahl an Untersuchungen mit ursprünglich anderer Fragestellung und dadurch nicht optimierter Bildgebung an externen Zentren.

Die Verwendung nicht-standardisierter Protokolle weicht von den meisten bisher zu diesem Thema durchgeführten Studien ab (Kunapuli et al. 2018; Sun et al. 2020). Hier werden die Untersuchungen zumeist an einem Zentrum durchgeführt und ein einheitliches Untersuchungsprotokoll verwendet. Untersuchungen mit Bildartefakten wurden mutmaßlich ausgeschlossen, jedoch wird dies von den meisten Studien nicht explizit erwähnt. Eine standardisierte Bildgebung ist auch Teil des von Lambin et al. (2017) vorgeschlagenen Radiomics Quality Score, der eine standardisierte Bildgebung als eine der Voraussetzung für die Reproduktion von Studien sieht. Eine nicht-standardisierte Bildgebung spiegelt jedoch deutlich realitätsnaher die tägliche Arbeit von klinischen Radiologen, vor allem in großen Zentren, wider, wo oftmals eine externe Bildgebung bereits erfolgte und anhand dieser eine Überweisung der Patienten erfolgte.

Damit die Anwendung von Radiomics zur Klassifikation von Tumoren den Sprung aus der Forschung mit oftmals optimierten Bedingungen in den klinischen Alltag schafft, ist die Anwendbarkeit auf alltägliche Daten eine wichtige Voraussetzung.

4.4 Methodik

4.4.1 Segmentationen und Radiomic Features

Die Segmentation der Nierentumoren erfolgte im Rahmen dieser Studie manuell. Diese Methode wird, in Bezug auf die Genauigkeit der Segmentationen, zumeist als Goldstandard gesehen.

Nachteile sind jedoch der hohe Zeitaufwand, der einen Einsatz für große Datenmengen nahezu unmöglich macht, und eine niedrige Interreader-Übereinstimmung (Kumar et al. 2012). Eine Alternative stellt die automatische beziehungsweise semiautomatische Segmentation dar. So konnten Lin et al. (2012) zeigen, dass eine automatische 3D-Segmentierung von Nierentumoren eine hohe Übereinstimmung mit der manuellen aufweist.

Die Extraktion der Radiomic Features erfolgte mit pyradiomics. Durch die Nutzung einer mit den IBSI-Kriterien übereinstimmenden Plattform kann einerseits die Zuverlässigkeit der extrahierten Radiomic Features erhöht werden (Fornacon-Wood et al. 2020). Andererseits wird so eine standardisierte und reproduzierbare Feature Analyse gewährleistet.

Während es in der Literatur keine Angaben zur Nutzung von pyradiomics zur Analyse von Nierentumoren gibt, wurde einige Studien zur Analyse von Nasopharynxkarzinomen, Pankreastumoren und Lungentumoren veröffentlicht (Khalvati et al. 2019; Liang et al. 2019; Yang et al. 2019).

4.4.2 Vorverarbeitung der machine learning-Algorithmen

Zahlreiche statistische Techniken wurden in dieser Arbeit verwendet, um eine möglichst exakte Aussage hinsichtlich der diagnostischen Genauigkeit der ML-Algorithmen bei der Unterscheidung zwischen malignen und benignen Nierentumoren treffen zu können.

Zunächst ist die Zehnfach-Kreuzvalidierung zu nennen, die angewendet wurde, um ein *overfitting* der ML-Algorithmen zu verhindern. Der Datensatz wurden in zehn Gruppen geteilt und die Algorithmen mit jeweils neun Gruppen trainiert. Die zehnte Gruppe (*out-of-bag-samples*) diente der Validierung der Ergebnisse.

Sämtliche Schritte zur Vorverarbeitung der Daten wurden ausschließlich auf dem jeweiligen Trainingsdatensatz angewendet (interne Kreuzvalidierung). Somit kann der Test-Datensatz als unabhängig betrachtet werden.

Dennoch gibt es Autoren, die ein komplett unabhängiges und möglichst externes Datenset zur Validierung präferieren (Gillies et al. 2015; Rizzo et al. 2018).

In unserer Studie wurde auf diese Art der Validierung bewusst verzichtet, da aufgrund der

relativ gesehenen Größe der Studienpopulation eine weitere Reduktion der Daten mutmaßlich zu einer reduzierten diagnostischen Genauigkeit geführt hätte. Jedoch kann ein statistisches *overfitting* der ML-Algorithmen und eine möglicherweise fehlende Anwendbarkeit auf andere Studienkohorten nicht ausgeschlossen werden.

Derzeit laufende Projekte der Arbeitsgruppe „urologische Radiologie“ der UMG beschäftigen sich genau mit diesem Aspekt und schließen in einer multizentrischen Studie Daten des Universitätsklinikums des Saarlands mit ein, um eine große Studienpopulation und damit ein vollständig unabhängiges und externes Datenset zu erhalten.

Die verwendeten Schritte zur Vorbereitung der Algorithmen (SMOTE und RFE) führten zu unterschiedlichen Ergebnissen.

Die Anwendung von SMOTE, wodurch eine höhere Zahl benigner Tumoren synthetisiert wurde, führte bei drei von vier Algorithmen zu einer Verschlechterung der diagnostischen Genauigkeit. Kocak et al. (2019) nutzen SMOTE in ihrer Studie zur CT-Diagnostik von Tumorgraden in klarzelligen Nierenzellkarzinomen und konnten mit einem *neural network*-Algorithmus keinen statistisch signifikanten Unterschied der AUC-Werte mit oder ohne SMOTE feststellen.

Auch Hu et al. (2020) erreichten bei der Diagnostik von Tumorgraden in Meningeomen im MRT mithilfe eines rf-Algorithmus keine statistisch signifikanten Unterschiede.

Als Grund hierfür kann diskutiert werden, dass die genaue Zusammensetzung der synthetisch entstandenen Daten nicht bekannt ist und der Fokus hier möglicherweise zu stark auf für den Algorithmus später nicht relevante Daten gelegt wurde. Auch der geringe Effekt von SMOTE auf hoch-dimensionale Daten, wie von Blagus und Lusa (2013) beschrieben, könnte eine Erklärung liefern.

Die RFE, bei der die am wenigsten relevanten Radiomic Features identifiziert und gelöscht wurden, führte bei zwei Algorithmen zu einer deutlichen Verbesserung der Leistung, bei zwei jedoch auch zu einer Verschlechterung.

Auch hierfür könnte der Grund darin liegen, dass möglicherweise nicht alle Radiomic Features für alle Subtypenanalysen gleich relevant sind. Es ist möglich, dass Radiomic Features gelöscht wurden, die für bestimmte Unterscheidungen wichtig gewesen wären.

Außerdem kann diese Art der Feature selection zu einem *overfitting* der Daten und damit zu einer schlechteren Leistung der Algorithmen führen (Kuhn 2019).

Zum Beispiel könnte die bessere diagnostische Genauigkeit des rf-Algorithmus ohne RFE darauf hinweisen, dass, durch die zufällige Komponente des Algorithmus bedingt, dieser besser mit unverarbeiteten Daten arbeitet.

Dies wird unterstützt durch die Studie von Wang et al. (2020), die zur Prädiktion des Differenzierungsgrades von Weichteilsarkomen mit Radiomics und einem rf-Algorithmus ebenfalls eine bessere AUC ohne die Verwendung von RFE erreichten.

4.5 Maschinelle Algorithmen

Die von uns angewendeten ML-Algorithmen unterschieden sich deutlich in ihrer Leistung. Die beste Genauigkeit zeigte der nnet-Algorithmus mit einer AUC von 78,9 % (ohne SMOTE und RFE), die schlechteste Leistung erzielte der knn-Algorithmus mit einer AUC von 46,6 % (mit SMOTE und RFE).

Die diagnostische Güte neuronaler Netzwerke wurde bereits in zahlreichen Publikationen beschrieben, unter anderem zur Diagnostik von Brust- und Lungenkrebs sowie von Gehirntumoren (Hua et al. 2015; Pereira et al. 2016; Becker et al. 2017).

Einer der Gründe für die gute diagnostische Genauigkeit des nnet-Algorithmus könnte in der Struktur des Algorithmus mit den Verarbeitungsschritten in *hidden layers* liegen. Hierdurch wird besonderes Gewicht auf für die Unterscheidung wichtige Informationen gelegt, irrelevante Informationen werden im gleichen Schritt unterdrückt (LeCun et al. 2015).

Bisher wurden nur wenige Studien zur Unterscheidung von malignen und benignen Nierentumoren mit Hilfe von Radiomics und maschinellem Lernen publiziert. Es fällt jedoch auf, dass die in dieser Studie erreichten AUC-Werte im Vergleich zu ähnlichen Studien niedriger sind. Beispielsweise untersuchten Kunapuli et al. (2018) 150 Patienten mit klarzelligen-, papillären- und chromophoben Nierenzellkarzinomen sowie Angiomyolipomen und Onkozytomen und konnten mittels Radiomics und *relational function gradient boosting* eine AUC von 83 % erreichen. Auch wenn diese AUC numerisch gering oberhalb der in unserer Studie beschriebenen AUC von 79 % liegt, ist die Größenordnung vergleichbar und der Unterschied vermutlich nicht statistisch signifikant. Einschränkend kommt hinzu, dass Kunapuli et al. für ihre Studie ein standardisiertes Protokoll für die CT-Untersuchung verwendeten und außerdem vier Kontrastmittelphasen bewerteten. Erdim et al. (2020) bewerteten CT-Untersuchungen sowohl ohne Kontrastmittel als auch in der kortikomedullären Phase. Es wurden alle auch in unserer Studie verwendeten Subtypen eingeschlossen und relevante Radiomic Features mittels einer CT-Texturanalyse ermittelt. Hier konnte durch einen rf-Algorithmus eine AUC von 91,5 % erreicht werden. Für diese Studie wurde ein standardisiertes Untersuchungsprotokoll mit zwei Phasen verwendet und die Studie an nur einem Zentrum

durchgeführt. Zudem limitiert die geringe Kohorte von $n = 79$ Patienten die Aussagekraft und Generalisierbarkeit der Studie.

Eine weitere Variante des sogenannten *deep machine learning*, welche Radiomics Analysen umgeht und ML-Algorithmen direkt auf Bilddaten anwendet, wurde von Zhou et al. (2019) bei $n = 192$ Patienten zur Unterscheidung von malignen und benignen Nierentumoren anhand von CT-Untersuchungen in der kortikomedullären Phase verwendet. Es konnte eine *accuracy* (bezeichnet die Wahrscheinlichkeit einer korrekten Vorhersage) von 97 % erreicht werden. Es wurden jedoch auch Zysten in die Gruppe der benignen Tumoren eingeschlossen.

Jenseits der klinisch relevanten Unterscheidung zwischen malignen und benignen Nierentumoren wurden zudem zahlreiche Studien publiziert, welche auf die Unterscheidung renaler Tumorsubtypen eingehen. Diese Studien können größtenteils in eine von drei Gruppen eingeordnet werden.

Erstens die Unterscheidung zwischen fettarmen Angiomyolipomen und malignen Nierenzellkarzinomen. Beispielsweise konnten Feng et al. (2018), mit Hilfe von CT-Texturanalysen und ML-Algorithmen, bei SRMs eine diagnostische Genauigkeit von 0,955 erreichen.

Zweitens die Unterscheidung zwischen Onkozytomen und Nierenzellkarzinomen. Li et al. (2020) zeigten bei der Unterscheidung von chromophoben Nierenzellkarzinomen und Onkozytomen bei fünf ML-Algorithmen eine AUC von jeweils über 85 %. Coy et al. (2019) erreichten mit Radiomics und *deep machine learning* eine Genauigkeit von 75,4 % in der Unterscheidung zwischen Onkozytomen und klarzelligen Nierenzellkarzinomen.

Drittens die Unterscheidung verschiedener maligner Subtypen. Hier konnte etwa von Kocak et al. (2018) mit einer CT-Texturanalyse und ML-Algorithmen eine Genauigkeit von 84,6 % bei der Unterscheidung von klarzelligen und nicht-klarzelligen Nierenzellkarzinomen erreicht werden.

Der Vorteil der Subtypisierung von Nierentumoren liegt in den Erkenntnissen, welche Tumoren eine besondere diagnostische Herausforderung darstellen.

In Subgruppenanalysen der in dieser Arbeit präsentierten Daten konnten wir zeigen, dass anhand der portalvenösen Phase Angiomyolipome und chromophobe Nierenzellkarzinome mit einer AUC von 85 % unterschieden werden können.

Es zeigte sich auch, dass insbesondere die Differenzierung von Onkozytomen zu chromophoben Nierenzellkarzinomen (AUC = 45 %) beziehungsweise zu Angiomyolipomen (AUC = 55 %) eine niedrige diagnostische Genauigkeit aufwies (Uhlig et al. 2020b).

Dies deutet darauf hin, dass insbesondere die Beurteilung von Onkozytomen anhand der vorhandenen Bildgebung schwer ist.

Dies zeigt auch eine Studie von Coy et al. (2017), die durch die Analyse der stärksten Abschwächung innerhalb einer Läsion zur Unterscheidung von Nierentumorsubtypen AUCs zwischen 79 % und 96 % erreichten. Die niedrigste AUC von 79 % zeigte sich für die Unterscheidung von klarzelligen Nierenzellkarzinomen und Onkozytomen.

Die in dieser Studie präsentierten Ergebnisse sind vergleichbar mit initialen Resultaten der Arbeitsgruppe urologische Radiologie der UMG, welche anhand von 94 Patienten mit kleinen Nierentumoren (T1, ≤ 70 mm) eine Unterscheidung zwischen malignen und benignen Tumoren mit einer AUC von 83 % treffen konnte. Die vergleichbare diagnostische Genauigkeit legt dabei nahe, dass auch größere Nierentumoren durch ML-Algorithmen robust in benigne und maligne unterschieden werden können. Hervorzuheben ist auch, dass zwei Radiologen in der Unterscheidung derselben Tumoren nur eine AUC von 68 % erreichten. Somit liegt die diagnostische Genauigkeit der in dieser Studie verwendeten ML-Algorithmen numerisch über der von erfahrenen Radiologen (Uhlig et al. 2020a).

Auch Sun et al. (2020) verglichen die diagnostische Genauigkeit von Radiomics und ML-Algorithmen mit der von vier erfahrenen Radiologen in Bezug auf die Unterscheidung von Nierentumoren. In der Subgruppe klarzelliges Nierenzellkarzinom versus Onkozytom und fettarmes Angiomyolipom wurde eine Genauigkeit von 85,5 % durch den Algorithmus und eine zwischen 73 und 95 % liegende Genauigkeit durch die Radiologen erreicht. Die Genauigkeit in der Unterscheidung zwischen papillärem und chromophobe Nierenzellkarzinom versus Onkozytom und fettarmen Angiomyolipom lag bei 80 % für den Algorithmus und zwischen 50 bis 71 % für die Radiologen.

Der Anspruch unserer Studie, durch die Verwendung realitätsnaher CT-Bilder eine verbesserte klinische Inklusion der Daten und damit eine bessere Anwendbarkeit zu schaffen, erwies sich einerseits als Limitation, da eine gewisse Verzerrung der idealen Ergebnisse hierdurch anzunehmen ist. Jedoch konnte gezeigt werden, dass im Unterschied zu standardisierter Bildgebung schlechtere Ergebnisse erzielt wurden und es weitere Konzepte bedarf bis die Anwendung von Radiomics zur Unterscheidung von malignen und benignen Tumoren in der klinischen Praxis zum Standard gehört.

4.6 Limitationen der Studie

Mit Hilfe von Radiomics und ML-Algorithmen konnte in dieser Studie bei der Dignitätsunterscheidung von Nierentumoren eine diagnostische Genauigkeit von 78,9 % erreicht werden. Als Limitation der Studie ist zunächst der retrospektive Ansatz zu nennen. Es kann

nicht ausgeschlossen werden, dass dadurch Daten verloren gingen oder die Ergebnisse verfälscht wurden. Auch wenn über 90 % aller Nierentumoren einen der in dieser Studie untersuchten Subtypen aufweisen, spiegelt dies die klinische Wirklichkeit nur begrenzt wider und limitiert somit die generalisierte Anwendbarkeit der Ergebnisse, da seltene Nierentumoren möglicherweise falsch klassifiziert würden.

Des Weiteren wurden nur Patienten die eine Operation erhielten in die Studie eingeschlossen. Dadurch ist es möglich, dass Patienten mit eindeutig benignen Tumoren, zum Beispiel fettreichen Angiomyolipomen und Zysten, nicht berücksichtigt wurden und dadurch die Anwendbarkeit des Algorithmus eingeschränkt ist. Möglicherweise würde sich durch einen Einschluss auch dieser Tumoren die diagnostische Genauigkeit verbessern.

Durch die geringe Anzahl an benignen Tumoren und des dadurch erfolgten Ausgleiches des Klassenungleichgewichts ist eine Verfälschung der Ergebnisse nicht ausgeschlossen.

Auf Grund der kleinen Patientenzahl wurde keine unabhängige Validationsgruppe eingesetzt und die Ergebnisse der Algorithmen basieren auf der durchgeführten Kreuzvalidierung. Dies könnte zu überangepassten und zu optimistischen Ergebnissen geführt haben und die Generalisierbarkeit einschränken.

4.7 Ausblick

Basierend auf den von uns präsentierten Ergebnissen und der derzeitigen Literatur weist die Anwendung von Radiomics und ML-Algorithmen das Potenzial auf, die Diagnostik bei Patienten mit Nierentumoren zu verbessern.

Jedoch ist die verwendete Methode für viele Radiologen unbekannt und wird dadurch häufig mit Skepsis gesehen. Das Verständnis von maschinellem Lernen als „Blackbox“, bei der die Entscheidungsfindung unbekannt und nicht nachvollziehbar ist, und die Angst vor einem Bedeutungsverlust von Radiologen können hier als Gründe genannt werden.

Perspektivisch müssen die hier präsentierten Methoden für eine Implementierung in den klinischen Alltag optimiert werden, um einen flächendeckenden Einsatz zu ermöglichen. Während eine hundertprozentige Sicherheit bei der Diagnostik wünschenswert wäre, ist diese nur schwer zu erreichen. Als Vergleich zur diagnostischen Genauigkeit von ML-Algorithmen sollten deshalb die erreichten Genauigkeiten von Radiologen verwendet werden, um eine bessere Vergleichbarkeit und klinische Einordnung zu erreichen.

Denkbar wäre ML-Algorithmen als Zusatz zur Befundung durch Radiologen einzusetzen, beispielsweise als unabhängige Meinung bei nicht sicher zu diagnostizierbaren Tumoren.

Auch der Einsatz in Tumorboards zur Unterstützung der Therapieplanung oder als Argument gegen eine operative Therapie wäre denkbar.

Des Weiteren sollte eine externe Validierung unserer Ergebnisse erfolgen. Nur so ist eine Generalisierbarkeit zu gewährleisten und eine Überoptimierung auf den Trainingsdatensatz zu vermeiden.

Außerdem sind prospektive Studien notwendig, um zu erörtern, in wie fern sich das operative Management von Patienten mit Nierentumoren verändert und ob der Einsatz von ML-Algorithmen tatsächlich zu weniger Operationen führt.

Weitere Möglichkeiten der nicht-invasiven Tumordiagnostik sind der Einsatz sogenannter *liquid biopsies*, die beispielsweise anhand von Tumor-DNA in Blutproben eine Diagnose ermöglichen. In diesem Zusammenhang ist hervorzuheben, dass sich die von uns in dieser Arbeit herausgestellten Methoden dazu eignen, weitere Informationen wie die Ergebnisse einer *liquid biopsy* einzuschließen, um die Tumordiagnostik so weiter zu optimieren.

5 Zusammenfassung

Das Nierenzellkarzinom ist der häufigste maligne Tumor der Niere, mit etwa 15000 Neuerkrankungen pro Jahr in Deutschland. Die Inzidenz kleiner Nierentumoren mit häufig benigner Dignität ist in den letzten Jahren steigend.

Nierentumoren sind häufig Zufallsbefunde und werden im Rahmen einer, aus anderen Gründen durchgeführten, Bildgebung des Abdomens entdeckt. Diese Bildgebung ist für eine korrekte Diagnose häufig nicht optimal. Die radiologische Diagnostik liefert wichtige Hinweise für eine Dignitätsunterscheidung, kann jedoch häufig keine eindeutige Diagnose liefern, so dass die meisten Patienten schlussendlich operiert werden. Jedoch werden etwa 30 % der operierten Tumoren postoperativ als benigne klassifiziert und die betroffenen Patienten somit übertherapiert.

Der Einsatz von Radiomics und *machine learning*-Algorithmen wird als mögliche Alternative zur invasiven und oft unzureichenden Diagnostik mittels Tumorbiopsien angesehen.

Unsere Studie schloss operierte Patienten mit den häufigsten Nierentumorsubtypen (klarzelliges-, papilläres-, chromophobes Nierenzellkarzinom, Onkozytom, Angiomyolipom) ein. Anhand einer präoperativen Computertomographie-Bildgebung wurden die Nierentumoren manuell segmentiert und Radiomics berechnet. Explizit wurden nicht-optimierte Computertomographie-Aufnahmen aus mehreren Zentren eingeschlossen, um die klinische Realität besser abzubilden. Mittels *machine learning*-Algorithmen wurde anhand der errechneten Radiomic Features eine Vorhersage über die Dignität der Nierentumoren erstellt, wobei interne Kreuzvalidierung, *recursive feature selection* und SMOTE zum Einsatz kamen.

Wir konnten zeigen, dass verschiedenen *machine learning*-Algorithmen unterschiedliche diagnostische Genauigkeiten erzielten. Die beste Leistung wurde durch einen *neural network*-Algorithmus erzielt, der eine diagnostische Genauigkeit (AUC, *area-under-the-curve*) von 78,9 % erreichte.

Die Stärken dieser Arbeit liegen zum einen in der Größe des Patientenkollektivs, da deutlich mehr Patienten eingeschlossen wurden als in vergleichbaren Studien. Des Weiteren ist die Studie multizentrisch. So wurde durch den Einschluss von Computertomographie-Daten unterschiedlicher Herkunft mit einer realitätsnahen Bildgebung gearbeitet. Dies ist unserer Meinung nach einer der wichtigsten Punkte für eine Implementierung einer solchen Anwendung in den Klinikalltag.

Als Limitation ist der retrospektive Ansatz der Studie zu nennen. Auch das Einschlusskriterium einer Operation, wodurch Patienten mit eindeutig benignen Tumoren ausgeschlossen wurden, limitiert die generalisierte Anwendbarkeit des *machine learning*-Algorithmus.

Perspektivisch werden Studien mit größeren Patientenkollektiven notwendig sein, um die Methodik zu verbessern. Auch sollte ein prospektiver Ansatz verfolgt werden, um eine Reduktion der Zahl an Operationen und damit eine Reduktion der Überversorgung beurteilen zu können.

6 Literaturverzeichnis

- AWMF (2020): Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF): Diagnostik, Therapie und Nachsorge des Nierenzellkarzinoms, Langversion 2.0, AWMF Registernummer: 043/017O.
- Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A (2017): Deep Learning in Mammography: Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer. *Invest Radiol* 52, 434–440
- Beer AJ, Dobritz M, Zantl N, Weirich G, Stollfuss J, Rummeny EJ (2006): Comparison of 16-MDCT and MRI for characterization of kidney lesions. *AJR Am J Roentgenol* 186, 1639–1650
- Bhatt NR, Davis NF, Flynn R, McDermott T, Thornhill JA, Manecksha RP (2015): Dilemmas in diagnosis and natural history of renal oncocytoma and implications for management. *Can Urol Assoc J* 9, E709–E712
- Bird VG, Kanagarajah P, Morillo G, Caruso DJ, Ayyathurai R, Leveillee R, Jorda M (2011): Differentiation of oncocytoma and renal cell carcinoma in small renal masses (<4 cm): the role of 4-phase computerized tomography. *World J Urol* 29, 787–792
- Bissler JJ, Kingswood JC (2004): Renal angiomyolipomata. *Kidney Int* 66, 924–934
- Brugarolas J (2014): Molecular genetics of clear-cell renal cell carcinoma. *J Clin Oncol* 32, 1968–1976
- Cairns P (2011): Renal cell carcinoma. *Cancer Biomark* 9, 461–473
- Capitanio U, Bensalah K, Bex A, Boorjian SA, Bray F, Coleman J, Gore JL, Sun M, Wood C, Russo P (2019): Epidemiology of Renal Cell Carcinoma. *Eur Urol* 75, 74–84
- Catalano C, Fraioli F, Laghi A, Napoli A, Pediconi F, Danti M, Nardis P, Passariello R (2003): High-resolution multidetector CT in the preoperative evaluation of patients with renal cell carcinoma. *AJR Am J Roentgenol* 180, 1271–1277
- Chawla N, Bowyer K, Hall L, Kegelmeyer W (2002): SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res* 16, 321–357
- Chawla SN, Crispen PL, Hanlon AL, Greenberg RE, Chen DYT, Uzzo RG (2006): The natural history of observed enhancing renal masses: meta-analysis and review of the world literature. *J Urol* 175, 425–431
- Choi JH, Kim JW, Lee JY, Han WK, Rha KH, Choi YD, Hong SJ, Yoon YE (2015): Comparison of computed tomography findings between renal oncocytomas and chromophobe renal cell carcinomas. *Korean J Urol* 56, 695–702
- Choudhary S, Rajesh A, Mayer NJ, Mulcahy KA, Haroon A (2009): Renal oncocytoma: CT features cannot reliably distinguish oncocytoma from other renal neoplasms. *Clin Radiol* 64, 517–522
- Edge SB, American Joint Committee on Cancer (Hrsg.): AJCC cancer staging manual. 7. Auflage; Springer, New York 2010

- Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, et al. (2012): 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* **30**, 1323–1341
- Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JWW, Comber H, Forman D, Bray F (2013): Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* **49**, 1374–1403
- Fernández-Pello S, Hora M, Kuusk T, Tahbaz R, Dabestani S, Abu-Ghanem Y, Albiges L, Giles RH, Hofmann F, Kuczyk MA, et al. (2020): Management of Sporadic Renal Angiomyolipomas: A Systematic Review of Available Evidence to Guide Recommendations from the European Association of Urology Renal Cell Carcinoma Guidelines Panel. *Eur Urol Oncol* **3**, 57–72
- Fornacon-Wood I, Mistry H, Ackermann CJ, Blackhall F, McPartlin A, Faivre-Finn C, Price GJ, O'Connor JPB (2020): Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *Eur Radiol* **30**, 6241–6250
- Fu W, Huang G, Moloo Z, Girgis S, Patel VH, Low G (2016): Multimodality Imaging Characteristics of the Common Renal Cell Carcinoma Subtypes: An Analysis of 544 Pathologically Proven Tumors. *J Clin Imaging Sci* **6**, 50
- Gillies RJ, Kinahan PE, Hricak H (2015): Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **278**, 563–577
- Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C (2004): Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* **351**, 1296–1305
- Guyon I, Weston J, Barnhill S, Vapnik V (2002): Gene Selection for Cancer Classification using Support Vector Machines. *Mach Learn* **46**, 389–422
- Hidayat K, Du X, Zou S-Y, Shi B-M (2017): Blood pressure and kidney cancer risk: meta-analysis of prospective studies. *J Hypertens* **35**, 1333–1344
- Hua K-L, Hsu C-H, Hidayati SC, Cheng W-H, Chen Y-J (2015): Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *Onco Targets Ther* **8**, 2015–2022
- Hunt JD, Hel OL van der, McMillan GP, Boffetta P, Brennan P (2005): Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. *Int J Cancer* **114**, 101–108
- Israel GM, Bosniak MA (2005): How I Do It: evaluating renal masses. *Radiology* **236**, 441–450
- Jayson M, Sanders H (1998): Increased incidence of serendipitously discovered renal cell carcinoma. *Urology* **51**, 203–205
- Jewett MAS, Mattar K, Basiuk J, Morash CG, Pautler SE, Siemens DR, Tanguay S, Rendon RA, Gleave ME, Drachenberg DE, et al. (2011): Active surveillance of small renal masses: progression patterns of early stage kidney cancer. *Eur Urol* **60**, 39–44
- Jinzaki M, Silverman SG, Akita H, Nagashima Y, Mikami S, Oya M (2014): Renal angiomyolipoma: a radiological classification and update on recent developments in diagnosis and management. *Abdom Imaging* **39**, 588–604

- Kane CJ, Mallin K, Ritchey J, Cooperberg MR, Carroll PR (2008): Renal cell cancer stage migration: analysis of the National Cancer Data Base. *Cancer* 113, 78–83
- Khalvati F, Zhang Y, Baig S, Lobo-Mueller EM, Karanicolas P, Gallinger S, Haider MA (2019): Prognostic Value of CT Radiomic Features in Resectable Pancreatic Ductal Adenocarcinoma. *Sci Rep* 9, 5449
- Kim JH, Li S, Khandwala Y, Chung KJ, Park HK, Chung BI (2019): Association of Prevalence of Benign Pathologic Findings After Partial Nephrectomy With Preoperative Imaging Patterns in the United States From 2007 to 2014. *JAMA Surg* 154, 225–231
- Kim JI, Cho JY, Moon KC, Lee HJ, Kim SH (2009): Segmental enhancement inversion at biphasic multidetector CT: characteristic finding of small renal oncocytoma. *Radiology* 252, 441–448
- Kim JK, Kim TK, Ahn HJ, Kim CS, Kim K-R, Cho K-S (2002): Differentiation of subtypes of renal cell carcinoma on helical CT scans. *AJR Am J Roentgenol* 178, 1499–1506
- Kuhn M (2008): Building Predictive Models in R Using the caret Package. *J Stat Softw* 28, 1–26
- Kuhn M, Johnson K: Applied Predictive Modeling. Springer New York, New York 2013
- Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJWL, Dekker A, Fenstermacher D, et al. (2012): Radiomics: the process and the challenges. *Magn Reson Imaging* 30, 1234–1248
- Kunapuli G, Varghese BA, Ganapathy P, Desai B, Cen S, Aron M, Gill I, Duddalwar V (2018): A Decision-Support Tool for Renal Mass Classification. *J Digit Imaging* 31, 929–939
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, et al. (2017): Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14, 749–762
- Lane BR, Tobert CM, Riedinger CB (2012): Growth kinetics and active surveillance for small renal masses. *Curr Opin Urol* 22, 353–359
- LeCun Y, Bengio Y, Hinton G (2015): Deep learning. *Nature* 521, 436–444
- Lee CT, Katz J, Fearn PA, Russo P (2002): Mode of presentation of renal cell carcinoma provides prognostic information. *Urol Oncol* 7, 135–140
- Liang Z-G, Tan HQ, Zhang F, Rui Tan LK, Lin L, Lenkowitz J, Wang H, Wen Ong EH, Kusumawidjaja G, Phua JH, et al. (2019): Comparison of radiomics tools for image analyses and clinical prediction in nasopharyngeal carcinoma. *Br J Radiol* 92, 20190271
- Lin Z, Cui Y, Liu J, Sun Z, Ma S, Zhang X, Wang X (2021): Automated segmentation of kidney and renal mass and automated detection of renal mass in CT urography using 3D U-Net-based deep convolutional neural network. *Eur Radiol (im Druck)*
- Ljungberg B, Albiges L, Abu-Ghanem Y, Bensalah K, Dabestani S, Fernández-Pello S, Giles RH, Hofmann F, Hora M, Kuczyk MA, et al. (2019): European Association of Urology Guidelines on Renal Cell Carcinoma: The 2019 Update. *Eur Urol* 75, 799–810

- Lopes Vendrami C, Parada Villavicencio C, DeJulio TJ, Chatterjee A, Casalino DD, Horowitz JM, Oberlin DT, Yang G-Y, Nikolaidis P, Miller FH (2017): Differentiation of Solid Renal Tumors with Multiparametric MR Imaging. *Radiographics* 37, 2026–2042
- Luciani LG, Cestari R, Tallarigo C (2000): Incidental renal cell carcinoma-age and stage characterization and clinical implications: study of 1092 patients (1982-1997). *Urology* 56, 58–62
- MacLennan S, Imamura M, Lapitan MC, Omar MI, Lam TBL, Hilvano-Cabungcal AM, Royle P, Stewart F, MacLennan G, MacLennan SJ, et al. (2012a): Systematic review of oncological outcomes following surgical management of localised renal cancer. *Eur Urol* 61, 972–993
- MacLennan S, Imamura M, Lapitan MC, Omar MI, Lam TBL, Hilvano-Cabungcal AM, Royle P, Stewart F, MacLennan G, MacLennan SJ, et al. (2012b): Systematic review of perioperative and quality-of-life outcomes following surgical management of localised renal cancer. *Eur Urol* 62, 1097–1117
- Marconi L, Dabestani S, Lam TB, Hofmann F, Stewart F, Norrie J, Bex A, Bensalah K, Canfield SE, Hora M, et al. (2016): Systematic Review and Meta-analysis of Diagnostic Accuracy of Percutaneous Renal Tumour Biopsy. *Eur Urol* 69, 660–673
- Millet I, Doyon FC, Hoa D, Thuret R, Merigeaud S, Serre I, Taourel P (2011): Characterization of small solid renal lesions: can benign and malignant tumors be differentiated with CT? *AJR Am J Roentgenol* 197, 887–896
- Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM (2016): The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. *Eur Urol* 70, 93–105
- Patel HD, Johnson MH, Pierorazio PM, Sozio SM, Sharma R, Iyoha E, Bass EB, Allaf ME (2016): Diagnostic Accuracy and Risks of Biopsy in the Diagnosis of a Renal Mass Suspicious for Localized Renal Cell Carcinoma: Systematic Review of the Literature. *J Urol* 195, 1340–1347
- Patel HD, Druskin SC, Rowe SP, Pierorazio PM, Gorin MA, Allaf ME (2017): Surgical histopathology for suspected oncocytoma on renal mass biopsy: a systematic review and meta-analysis. *BJU Int* 119, 661–666
- Pereira S, Pinto A, Alves V, Silva CA (2016): Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Trans Med Imaging* 35, 1240–1251
- Przybycin CG, Magi-Galluzzi C, McKenney JK (2013): Hereditary syndromes with associated renal neoplasia: a practical guide to histologic recognition in renal tumor resection specimens. *Adv Anat Pathol* 20, 245–263
- Quinn MJ, Hartman DS, Friedman AC, Sherman JL, Lautin EM, Pyatt RS, Ho CK, Csere R, Fromowitz FB (1984): Renal oncocytoma: new observations. *Radiology* 153, 49–53
- Renehan AG, Tyson M, Egger M, Heller RF, Zwahlen M (2008): Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet* 371, 569–578

- Reuter VE, Argani P, Zhou M, Delahunt B (2014): Best practices recommendations in the application of immunohistochemistry in the kidney tumors: report from the International Society of Urologic Pathology consensus conference. *Am J Surg Pathol* **38**, e35–e49
- Reznek RH (2004): CT/MRI in staging renal cell carcinoma. *Cancer Imaging* **4**, S25–S32
- Richard PO, Jewett MAS, Bhatt JR, Evans AJ, Timilsina N, Finelli A (2016): Active Surveillance for Renal Neoplasms with Oncocytic Features is Safe. *J Urol* **195**, 581–586
- Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, Bellomi M (2018): Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp* **2**, 36
- Robson CJ (1963): Radical nephrectomy for renal cell carcinoma. *J Urol* **89**, 37–42
- Romis L, Cindolo L, Patard JJ, Messina G, Altieri V, Salomon L, Abbou CC, Chopin D, Lobel B, de La Taille A (2004): Frequency, clinical presentation and evolution of renal oncocytomas: multicentric experience from a European database. *Eur Urol* **45**, 53–57; discussion 57
- Rosenkrantz AB, Hindman N, Fitzgerald EF, Niver BE, Melamed J, Babb JS (2010): MRI features of renal oncocytoma and chromophobe renal cell carcinoma. *AJR Am J Roentgenol* **195**, W421–427
- Seyam RM, Bissada NK, Kattan SA, Mokhtar AA, Aslam M, Fahmy WE, Mourad WA, Binmahfouz AA, Alzahrani HM, Hanash KA (2008): Changing Trends in Presentation, Diagnosis and Management of Renal Angiomyolipoma: Comparison of Sporadic and Tuberous Sclerosis Complex-associated Forms. *Urology* **72**, 1077–1082
- Sheth S, Scatarige JC, Horton KM, Corl FM, Fishman EK (2001): Current concepts in the diagnosis and management of renal cell carcinoma: role of multidetector CT and three-dimensional CT. *Radiographics* **21**, S237–S254
- Skinner DG, Colvin RB, Vermillion CD, Pfister RC, Leadbetter WF (1971): Diagnosis and management of renal cell carcinoma. A clinical and pathologic study of 309 cases. *Cancer* **28**, 1165–1177
- Smaldone MC, Kutikov A, Egleston BL, Canter DJ, Viterbo R, Chen DYT, Jewett MA, Greenberg RE, Uzzo RG (2012): Small renal masses progressing to metastases under active surveillance: a systematic review and pooled analysis. *Cancer* **118**, 997–1006
- Song S, Park BK, Park JJ (2016): New radiologic classification of renal angiomyolipomas. *Eur J Radiol* **85**, 1835–1842
- Sooriakumaran P, Gibbs P, Coughlin G, Attard V, Elmslie F, Kingswood C, Taylor J, Corbishley C, Patel U, Anderson C (2010): Angiomyolipomata: challenges, solutions, and future prospects based on over 100 cases treated. *BJU Int* **105**, 101–106
- Stillwell TJ, Gomez MR, Kelalis PP (1987): Renal lesions in tuberous sclerosis. *J Urol* **138**, 477–481
- Sun X-Y, Feng Q-X, Xu X, Zhang J, Zhu F-P, Yang Y-H, Zhang Y-D (2020): Radiologic-Radiomic Machine Learning Models for Differentiation of Benign and Malignant Solid Renal Masses: Comparison With Expert-Level Radiologists. *AJR Am J Roentgenol* **214**, W44–W54

- Thorstenson A, Bergman M, Scherman-Plogell A-H, Hosseinnia S, Ljungberg B, Adolfsson J, Lundstam S (2014): Tumour characteristics and surgical treatment of renal cell carcinoma in Sweden 2005-2010: a population-based study from the national Swedish kidney cancer register. *Scand J Urol* 48, 231–238
- Uhlig J, Strauss A, Rücker G, Seif Amir Hosseini A, Lotz J, Trojan L, Kim HS, Uhlig A (2019): Partial nephrectomy versus ablative techniques for small renal masses: a systematic review and network meta-analysis. *Eur Radiol* 29, 1293–1307
- Uhlig J, Biggemann L, Nietert MM, Beißbarth T, Lotz J, Kim HS, Trojan L, Uhlig A (2020a): Discriminating malignant and benign clinical T1 renal masses on computed tomography: A pragmatic radiomics and machine learning approach. *Medicine (Baltimore)* 99, e19725
- Uhlig J, Leha A, DeLonge LM, Haack A-M, Shuch B, Kim HS, Bremmer F, Trojan L, Lotz J, Uhlig A (2020b): Radiomic Features and Machine Learning for the Discrimination of Renal Tumor Histological Subtypes: A Pragmatic Study Using Clinical-Routine Computed Tomography. *youg* 12, 3010
- van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RG, Fillion-Robin J-C, Pieper S, Aerts HJ (2017): Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 77, e104–e107
- Vogel C, Ziegelmüller B, Ljungberg B, Bensalah K, Bex A, Canfield S, Giles RH, Hora M, Kuczyk MA, Merseburger AS, et al. (2019): Imaging in Suspected Renal-Cell Carcinoma: Systematic Review. *Clin Genitourin Cancer* 17, e345–e355
- Wolpert DH (1996): The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Comput* 8, 1341–1390
- Yang M, She Y, Deng J, Wang T, Ren Y, Su H, Wu J, Sun X, Jiang G, Fei K, et al. (2019): CT-based radiomics signature for the stratification of N2 disease risk in clinical stage I lung adenocarcinoma. *Transl Lung Cancer Res* 8, 876–885
- Young JR, Margolis D, Sauk S, Pantuck AJ, Sayre J, Raman SS (2013): Clear cell renal cell carcinoma: discrimination from other renal cell carcinoma subtypes and oncocytoma at multiphasic multidetector CT. *Radiology* 267, 444–453
- Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, Ashrafinia S, Bakas S, Beukinga RJ, Boellaard R, et al. (2020): The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* 295, 328–338
- Kuhn M. <https://topepo.github.io/caret/recursive-feature-elimination.html>; abgerufen am 07.12.2020
- Robert Koch-Institut. https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Nierenkrebs/nierenkrebs_node.html; abgerufen am 02.10.2019

Danksagung

Ein besonderer Dank gilt meinem Doktorvater Herrn PD Dr. med Johannes Uhlig für die Überlassung des Themas und die außergewöhnlich gute Betreuung. Danke für die fachliche und organisatorische Unterstützung, das immer hilfreiche Feedback und das unermüdete Beantworten zahlreicher Fragen.

Außerdem bedanken möchte ich mich bei Frau PD Dr. med Leitsmann für die konstruktiven Anmerkungen welche erheblich zum Gelingen dieser Arbeit beigetragen haben.

