

Aus dem Fachbereich Medizin
der Johann Wolfgang Goethe-Universität
Frankfurt am Main

betreut am
Zentrum der Radiologie
Institut für Diagnostische und Interventionelle Radiologie
Direktor: Prof. Dr. med. Thomas J. Vogl

**Lipiodol-Anreicherung als bildgebender Biomarker (Radiomics) zur
Vorhersage des Therapieansprechens und Überlebens nach
Transarterieller Chemoembolisation (TACE) bei Patienten mit
hepatozellulärem Karzinom**

Dissertation
zur Erlangung des Doktorgrades der Medizin
des Fachbereiches Medizin
der Johann Wolfgang Goethe-Universität
Frankfurt am Main

vorgelegt von
Oleg Elenberger

aus Berjosowka

Frankfurt am Main, 2023

Dekan:	Prof. Dr. Stefan Zeuzem
Referent:	Prof. Dr. Thomas J. Vogl
Korreferent/in:	Prof. Dr. Dr. Albrecht Piiper
Tag der mündlichen Prüfung:	12.09.2023

Meinen Eltern

Teile dieser Arbeit wurden publiziert in:

Bernatz S, Elenberger O, Ackermann J, et al. CT - radiomics and clinical risk scores for response and overall survival prognostication in TACE HCC patients. *Sci Rep.* 2023;13:1-9. doi:10.1038/s41598-023-27714-0

1. EINLEITUNG	6
1.1 EPIDEMIOLOGIE	6
1.2 RISIKOFAKTOREN	7
1.3 DIAGNOSE	9
1.4 STAGING	13
1.5 THERAPIE	14
1.5.1 Chirurgische Therapie (BCLC 0 oder A)	14
1.5.2 Ablation (BCLC 0 oder A)	15
1.5.3 Transarterielle Therapien (BCLC B)	16
1.5.4 Systemische Therapie (BCLC C)	18
1.6 mRECIST (MODIFIED RESPONSE EVALUATION CRITERIA IN SOLID TUMORS)	18
1.7 PROGNOTISCHE KLINISCHE SCORES	21
1.7.1 Child-Pugh Score	21
1.7.2 MELD-Score (Model of End-stage Liver Disease)	22
1.7.3 ALBI-Score (Albumin-Bilirubin-Score)	23
1.7.4 HAP-Score (Hepatoma arterial-embolisation prognostic score)	24
1.7.5 mHAP-II-Score (modified HAP-II-Score)	25
1.7.6 STATE Score (Selection for TrArterial chemoembolisation TrEatment)	25
1.7.7 6 and 12 Score	26
1.8 KÜNSTLICHE INTELLIGENZ / RADIOMICS	27
1.9 ZIELSETZUNG	32
2. MATERIAL UND METHODEN	34
2.1 PATIENTENKOLLEKTIV	34
2.2 TACE	38
2.3 TRANSFER UND AUFBEREITUNG DER BILDDATEN	40
2.4 SEGMENTIERUNG DER BILDDATENSÄTZE	40
2.5 RADIOMICS ANALYSE (ENGINEERED RADIOMICS FEATURES)	41
2.6 INTERRATER-RELIABILITÄT UND FEATUREREDUNDANZ	43
2.7 BERECHNUNG DER KLINISCHEN SCORES	46
2.8 FEATURE AUSWAHL UND ENTWICKLUNG DES BERECHNUNGSMODELLS (SUPERVISED LEARNING)	47
3. ERGEBNISSE	53
3.1 ANALYSE DES THERAPIEANSPRECHENS	53
3.2 EXPLORATIVE DATENANALYSE MITTELS DIMENSIONALITÄTSREDUKTION	54
3.3 ANSPRECHVERHALTEN AUF EBENE DER TL	55
3.4 PROGNOSE DES GESAMTANSPRECHENS AUF PATIENTENEBENE	58
3.5 PROGNOSE DES GESAMTÜBERLEBENS	59
4. DISKUSSION	62
5. ZUSAMMENFASSUNG	68
6. SUMMARY	70
7. LITERATURVERZEICHNIS	72
8. ANHANG	83
8.1 ABKÜRZUNGSVERZEICHNIS	83
8.2. ABBILDUNGS- UND TABELLENVERZEICHNIS	85
8.3 ANHANG	87
9. DANKSAGUNG	93
10. LEBENS LAUF	94
11. SCHRIFTLICHE ERKLÄRUNG	96

1. Einleitung

1.1 Epidemiologie

Das primäre Leberkarzinom steht mit ca. 906.000 jährlichen Neuerkrankungen an sechster Stelle der häufigsten malignen Erkrankungen weltweit.¹ Als karzinombedingte Todesursache steht das primäre Leberkarzinom mit ca. 830.000 Fällen pro Jahr weltweit an dritter Stelle.¹

Mit einem Anteil von ca. 90 % stellt das hepatozelluläre Karzinom (HCC) die häufigste Entität eines primären Leberkarzinoms dar.^{2,3}

Im Jahr 2020 entfielen in etwa 88.000 Neuerkrankungen auf Europa (ca. 9,2 % der globalen jährlichen Neuerkrankungen) und davon ca. 9.600 Fälle auf Deutschland (1,1 % der globalen jährlichen Neuerkrankungen).¹

Die Lebenszeitprävalenz auf die gesamte Welt betrachtet, liegt bei ca. 2,2%, in Deutschland liegt der Wert bei 1,4%.¹

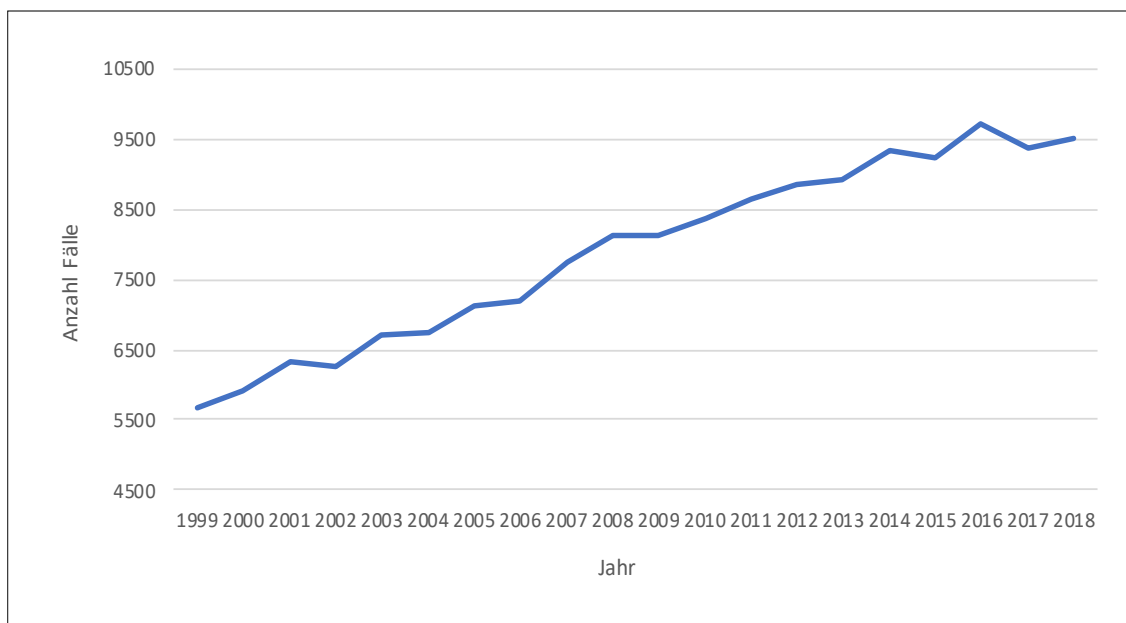


Abbildung 1. Inzidenz in absoluten Zahlen, Entwicklung in Deutschland 1999-2018, nach RKI, <https://www.krebsdaten.de> (abgerufen 25.03.2022)⁴

Im zeitlichen Verlauf von 1999 – 2018 nahm die Inzidenz in Deutschland von 5,2 / 100.000 auf 6,7 / 100.000 zu (siehe Abbildung 1).^{4,5}

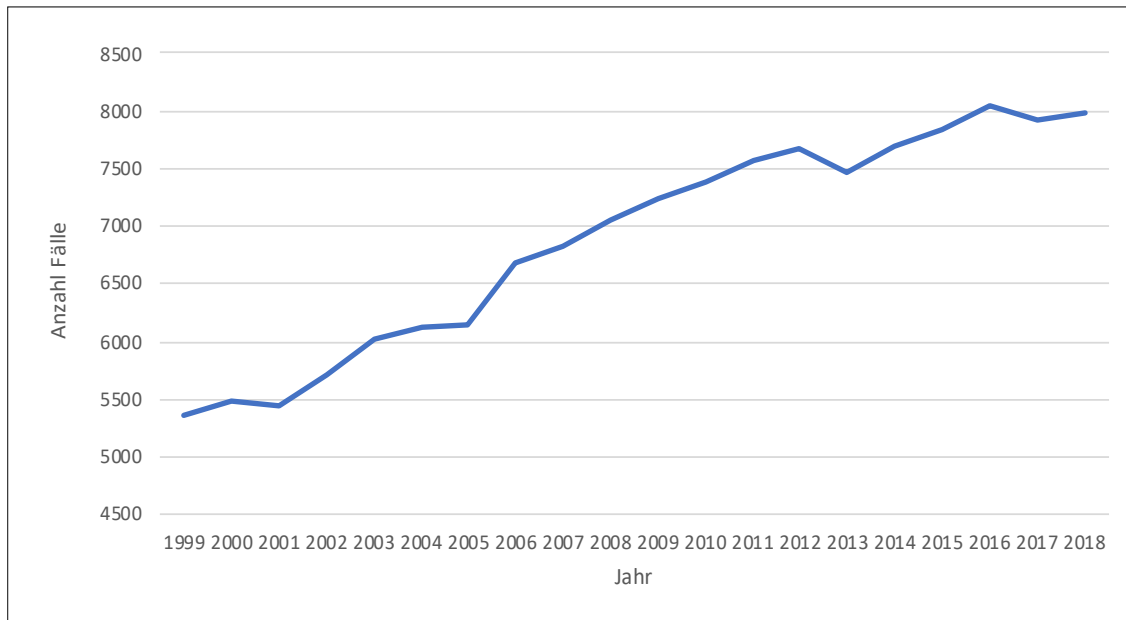


Abbildung 2. Mortalität in absoluten Zahlen, Entwicklung in Deutschland 1999-2018, nach RKI, <https://www.krebsdaten.de> (abgerufen 25.03.2022)⁶

Die Mortalität entwickelte sich in diesem Zeitraum ebenfalls zunehmend, im Jahr 1999 belief sich diese auf 4,8 / 100.000, im Jahr 2018 bereits auf 5,2 / 100.000 (siehe Abbildung 2).^{6,7}

1.2 Risikofaktoren

Weltweit stellt eine Erkrankung mit dem Hepatitis-B-Virus (HBV) die häufigste Ursache (33%) der Entwicklung eines HCC dar, gefolgt von Alkoholabusus (30%). Eine Hepatitis-C-Erkrankung (HCV) steht an dritter Stelle mit 21%. HCC auf Grund von anderen Ursachen, darunter nichtalkoholische Steatohepatitis (NASH) und nichtalkoholische Fettlebererkrankung (NAFLD), stellt die letzte Untergruppe mit 16% dar.⁸

Hierbei zeigt sich jedoch eine erhebliche Varianz abhängig von der geografischen Verteilung der Inzidenzen (siehe Abbildung 3).

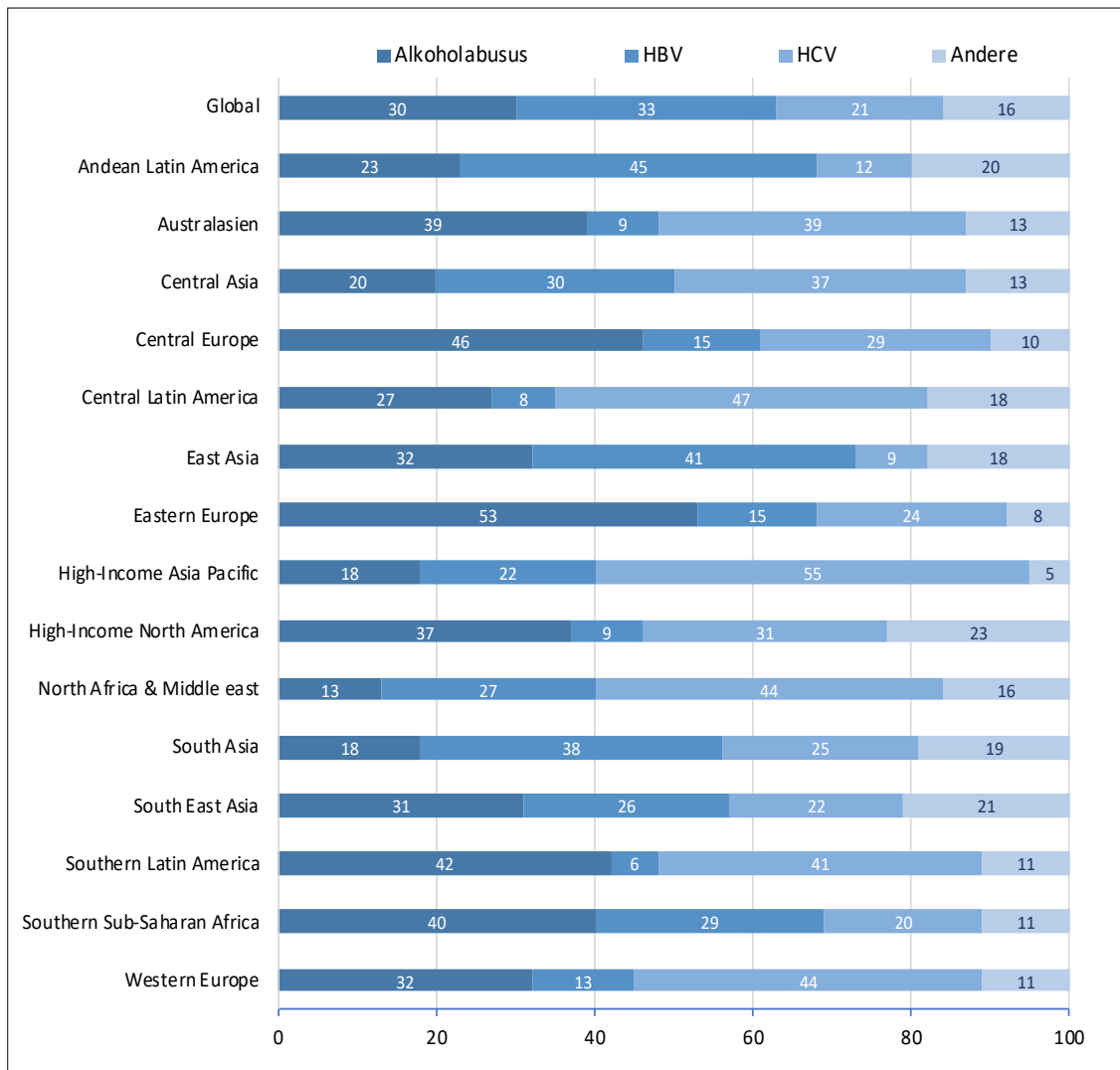


Abbildung 3. Risikofaktoren abhängig von der geografischen Verteilung, nach Fitzmaurice et al.⁸

In Deutschland werden die meisten Fälle dem Alkoholabusus zugeschrieben (44%), eine HCV-Erkrankung hingegen steht für 33% der Fälle als Ursache. NASH, NAFLD und weitere Entitäten (Hämochromatose, α 1-Antitrypsin-Mangel und andere) verursachen 14% der Fälle, gefolgt von einer HBV-Erkrankung mit 8% der HCC-Fälle.⁸

Laut Estes et al.⁹ muss zukünftig von einer weltweiten Zunahme von NAFLD-Fällen ausgegangen werden, wodurch wiederum mit einem Anstieg der HCC-Fälle zu rechnen

ist. Auf der Grundlage von vorliegenden Daten zum metabolischen Syndrom und Typ 2 Diabetes mellitus geht die Studie davon aus, dass alleine in Deutschland im Zeitraum von 2016 – 2030 eine Zunahme von NAFLD Fällen um 13,5% von 18,45 Mio. auf 20,95 Mio. Fälle zu erwarten ist. Ausgehend von diesen Zahlen prognostizieren Estes et al.⁹ in diesem Zeitraum eine Zunahme der HCC-Fälle in Deutschland als Folge einer NAFLD von 1970 auf 4090, dies entspricht einer Steigerung um 108%.⁹

1.3 Diagnose

Die Bildgebung spielt eine führende Rolle in der HCC-Diagnostik. Hierbei macht man sich die Umstellung der Blutversorgung der Hepatozyten, als Folge der Entartung, zu Nutze. Während gesunde Zellen vorwiegend aus dem Pfortaderkreislauf gespeist werden, ändert sich die Hämodynamik im Rahmen der malignen Transformation und maligne Läsionen werden hauptsächlich über die A. hepatica gespeist.¹⁰ Hierdurch erklären sich die typischen bildgebenden Kriterien einer malignen Läsion; arterielles Hyperenhancement mit Auswaschen des Kontrastmittels in der portalvenösen und/oder venösen Phase der Untersuchung.^{11,12} Das Kontrastmittelverhalten kann sowohl in einer multiphasischen Computertomographie (CT), als auch in der Magnetresonanztomographie (MRT) beobachtet werden.

Roberts et al.¹³ geben die Sensitivität der Detektion eines HCC Herdes mit 82% im MRT und 66% im CT an, dies bei nahezu gleicher Spezifität von 92% vs. 91%. Bei Läsionen >2 cm steigt die Sensitivität auf 88% im MRT vs. 79% im CT an, bei einer Spezifität von 87% vs. 90% (siehe Abbildung 4).¹³ Bei Herden unter 1 cm Durchmesser fällt sowohl die Sensitivität als auch Spezifität deutlich. Diese Abnahme demonstriert die Herausforderung bei der Diagnose von kleinen HCC-Herden.

	Gesamt		> 2 cm		1 – 2 cm		< 1 cm	
	CT	MRT	CT	MRT	CT	MRT	CT	MRT
Sensitivität	66 %	82 %	79 %	88 %	64 %	70 %	48 %	69 %
Spezifität	92 %	91 %	90 %	87 %	88 %	87 %	69 %	46 %

Abbildung 4. Sensitivität und Spezifität MRT vs. CT in Abhängigkeit der Herdgröße, nach Roberts et al.¹³

Die deutsche Leitlinie empfiehlt eine multiphasische MRT zur Primärdiagnostik, bei unklarem Befund in der MRT wird eine ergänzende multiphasische CT und/oder CEUS (kontrastmittelverstärkter Ultraschall) empfohlen (siehe Abbildung 5).¹⁴ Eine Bildgebung mittels CT kann dennoch favorisiert werden. So kann z.B. Aszites in der MRT zu erheblichen Artefakten führen. Da eine CT in deutlich kürzerer Zeit erhoben wird, profitieren Patienten mit verminderter Fähigkeit zum Luftanhalten von der kürzeren Untersuchungszeit. Die Verfügbarkeit der Modalität kann ein weiterer Entscheidungsgrund für eine Diagnostik mittels CT sein.¹³

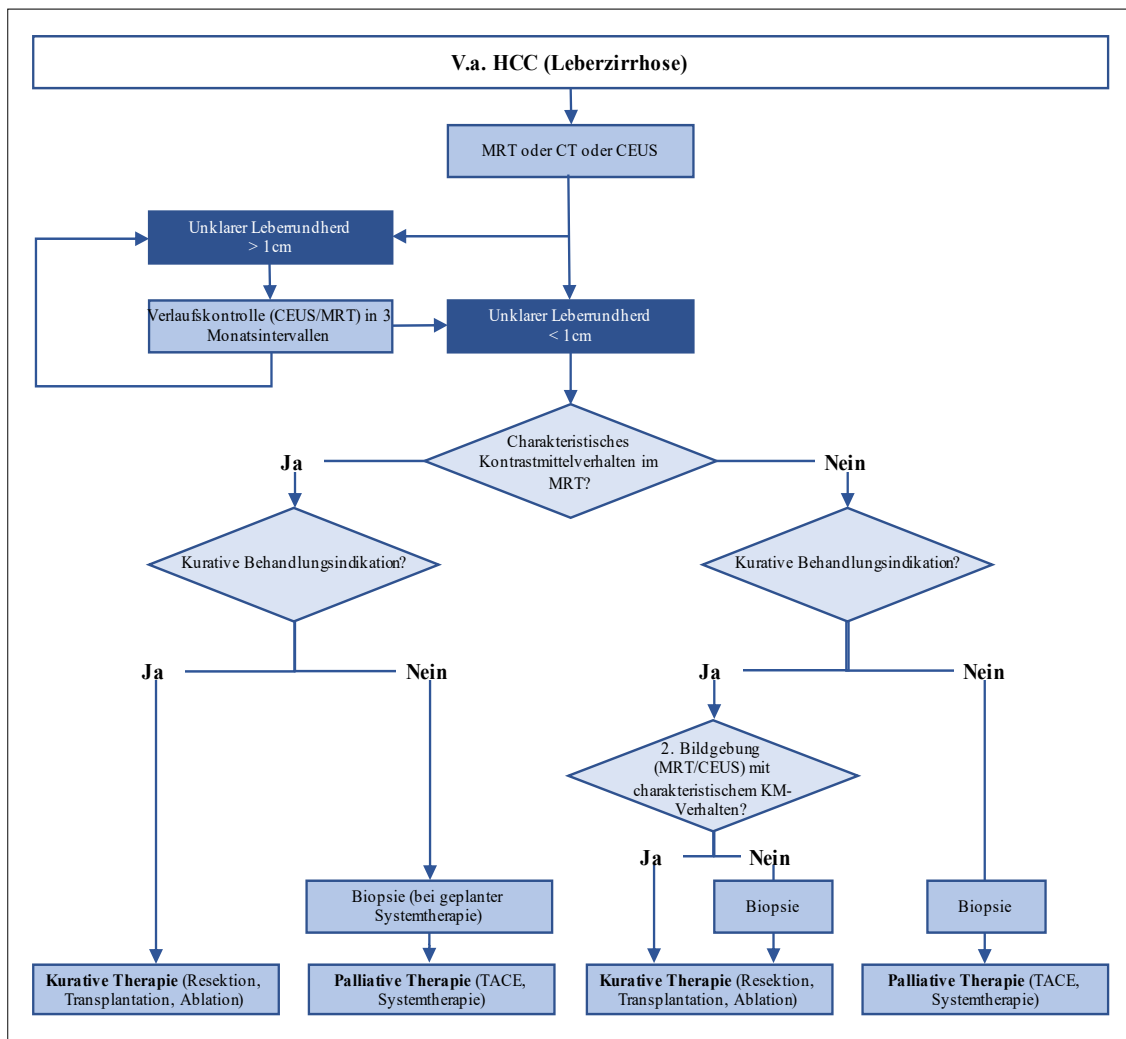


Abbildung 5. Diagnostikalgorithmus HCC, nach S3-Leitlinie Diagnostik und Therapie des Hepatozellulären Karzinoms und biliärer Karzinome.¹⁴

Um eine vergleichbare Befundung und reproduzierbare Ergebnisse zu erhalten wurde vom American College of Radiology (ACR) im Jahr 2011 ein standardisiertes System der Terminologie, Technik, Interpretation und Befundung eingeführt, welches zuletzt im Jahr 2018 aktualisiert wurde: das Liver Imaging Reporting and Data System (LI-RADS).¹⁵ Das System verwendet als Hauptmerkmale (1) das arterielle Hyperenhancement, (2) die Größe der Läsion, (3) eine etwaige kontrastmittelanreichernde Kapsel, (4) das Auswaschen des Kontrastmittels und (5) Schwellenwachstum bei unklaren Läsionen.

Nach LI-RADS werden acht Kategorien definiert mit entsprechenden Wahrscheinlichkeiten für das Vorliegen einer malignen Läsion / eines HCCs (siehe Abbildung 6 und 7).

Kategorie	Details	Wahrscheinlichkeit für das Vorliegen eines HCC bzw. von Malignität
LR-NC	Nicht kategorisierbar aufgrund fehlender oder schlechter Bildgebung	
LR-TIV	Tumorinfiltration in der Vene	
LR-1	definitiv benigne (z.B. Zyste, Hämangiom)	0% HCC / 0% maligne
LR-2	wahrscheinlich benigne	16% HCC / 18% maligne
LR-M	Wahrscheinlich oder definitiv maligne, aber nicht HCC-spezifisch (z.B. Target-Zeichen beim Cholangiokarzinom)	37% HCC / 94% maligne
LR-3	mit mittlerer Wahrscheinlichkeit maligne	37% HCC / 39% maligne
LR-4	wahrscheinlich HCC	74% HCC / 81% maligne
LR-5	definitiv HCC	95% HCC / 98% maligne

Abbildung 6. LI-RADS Kategorien, nach Schreyer et al.¹⁶

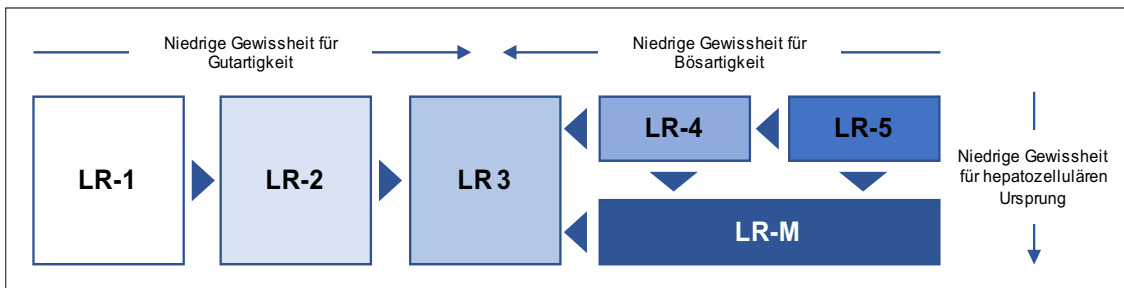


Abbildung 7. Gewissheit für Gutartigkeit abhängig von der LI-RADS Kategorie, nach CT/MRT LI-RADS® v2018.¹⁷

In Einzelstudien werden für LR-5 Läsionen Spezifitäten von 98% und Sensitivitäten von 69% angegeben.¹⁸

Eine Metaanalyse, welche insgesamt 18 Studien im Zeitraum 2013 – 2018 einschließt, ergibt eine gepoolte Spezifität von 93% und Sensitivität von 67% für LR-5 Läsionen. LR-3 Läsionen weisen eine Spezifität von 85% und Sensitivität von 86% auf.¹⁹

1.4 Staging

Für das weitere Vorgehen sind neben der Tumorlast, eine etwaige Fernmetastasierung, Leberfunktion, Nebenerkrankungen, sowie der Allgemeinzustand der Patienten die entscheidenden Kriterien in der Therapieauswahl.

Alle diese Faktoren werden im Barcelona Clinic Liver Score (BCLC) berücksichtigt.²⁰ Im Jahre 1992 vorgestellt, ist das Schema sowohl in den klinischen Guidelines als auch den nationalen Empfehlungen etabliert.³

Nach BCLC werden Patienten in eine von fünf Gruppen eingeordnet. Abhängig von der Gruppe empfiehlt das Schema das geeignete Therapieregime (siehe Abbildung 8).

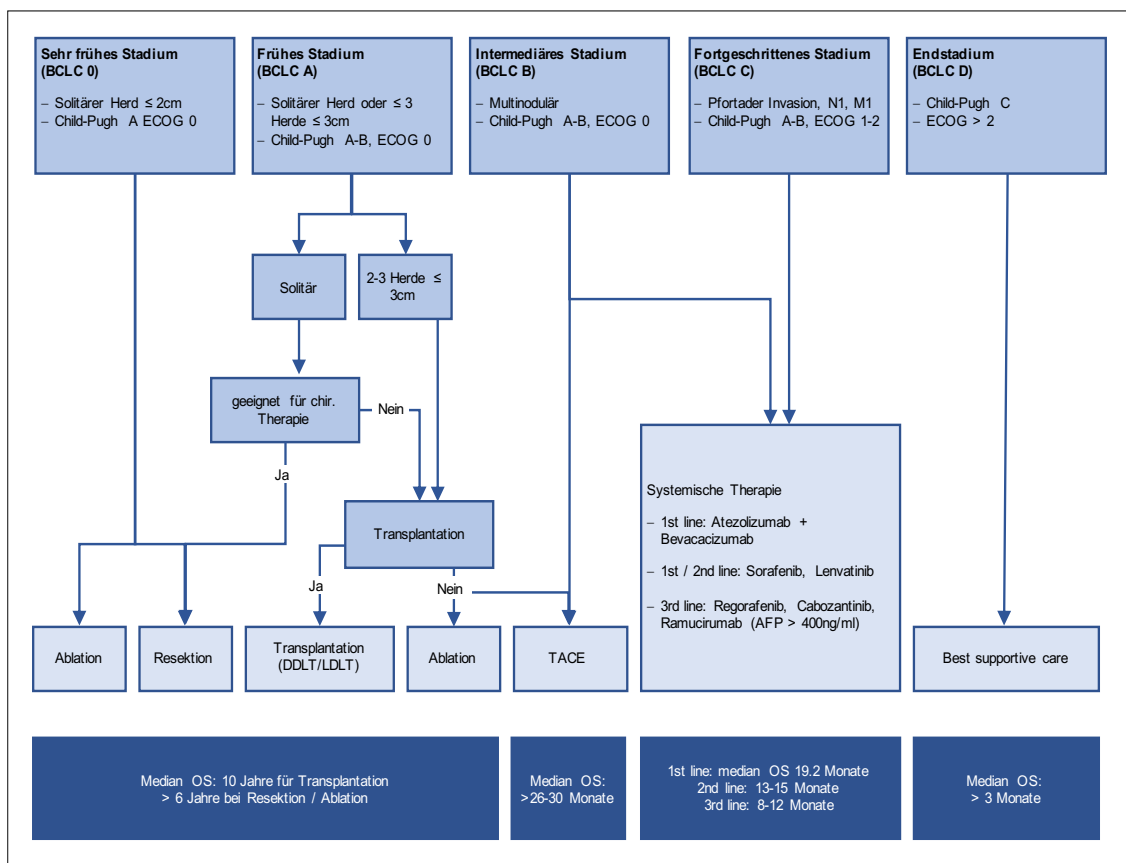


Abbildung 8. BCLC Algorithmus, nach Llovet et al.²

1.5 Therapie

1.5.1 Chirurgische Therapie (BCLC 0 oder A)

Chirurgische Therapien, im Sinne einer Resektion oder Transplantation, kommen für Patienten im frühen Stadium (BCLC 0 oder A) in Frage. Nach einer organerhaltenden Resektion besteht eine 5-Jahres-Überlebensrate von 60%, bei einer Rezidivrate von 70% innerhalb dieses Zeitraumes.²¹

Eine Lebertransplantation ist bei Patienten, die sich nicht für eine organerhaltende Operation eignen, indiziert. Als Inklusionskriterien für eine Listung zur Transplantation gilt ein solitärer HCC-Herd von ≤ 5 cm im Durchmesser oder maximal 3 Herde von jeweils ≤ 3 cm im Durchmesser, gemein als Milan-Kriterien bekannt.²²

Im Jahr 2020 standen auf der Warteliste für eine Lebertransplantation in Deutschland insgesamt 1987 Patienten bei 826 vorgenommenen Transplantationen im selben Jahr (siehe Abbildung 9).²³

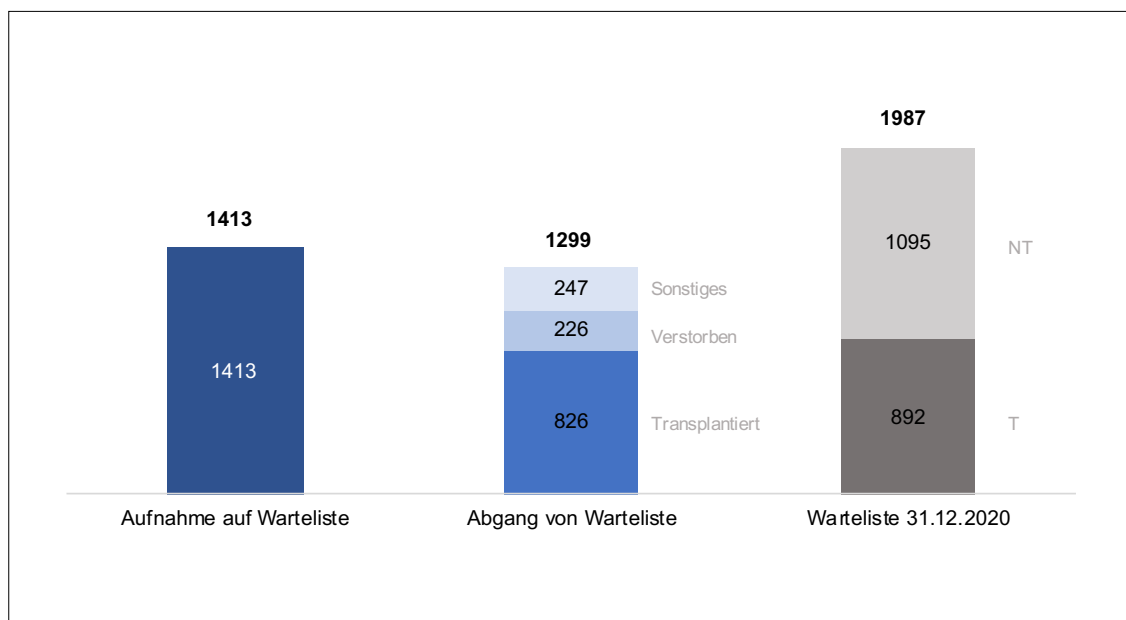


Abbildung 9. Transplantationszahlen Deutschland 2020, nach DSO (Deutsche Stiftung Organtransplantation).²³

(T: transplantabel gemeldete Patienten, NT: nicht transplantabel gemeldete Patienten).

Nach den Milan-Kriterien ausgewählte und transplantierte Patienten weisen eine 5-Jahres-Überlebensrate von 60-80% und eine 10-Jahres-Überlebensrate von 50% auf bei einer Rezidivrate von ca. 15%. Ein positiver Nebeneffekt einer Transplantation ist die Heilung der zugrundeliegenden Lebererkrankung.¹⁰

1.5.2 Ablation (BCLC 0 oder A)

Für Patienten im frühen Stadium der Erkrankung kommt als Alternative für einen operativen Eingriff eine lokale Ablation in Frage.¹⁴

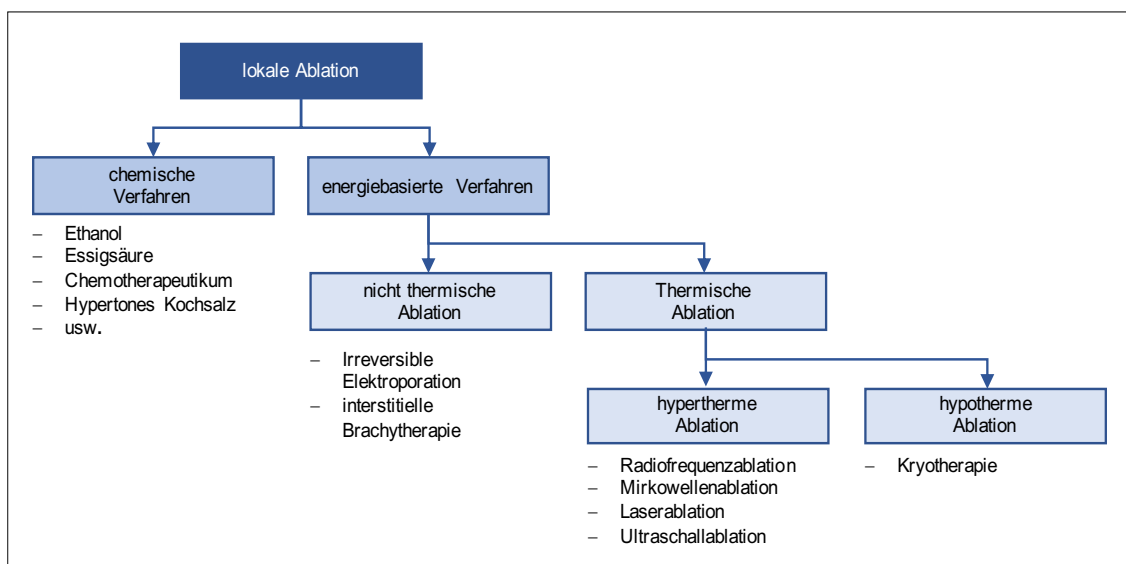


Abbildung 10. Systematik der lokal ablativen Verfahren, nach Mahnken et al.²⁴

Dafür steht ein breites Spektrum lokal ablativer Verfahren zur Wahl (siehe Abbildung 10).²⁵

Am häufigsten wird die thermische Ablation mittels Mikrowellenablation (MWA) und Radiofrequenzablation durchgeführt.²⁶

Bei einer Radiofrequenzablation werden über einen transkutanen Zugang eine bis mehrere Elektroden in das Zielgewebe eingeführt. Über ein lokal erzeugtes elektromagnetisches Feld kommt es durch Oszillationen von Ionen im Gewebe zur

Erzeugung von Reibungswärme, was letztendlich zu einer thermischen Nekrose (Koagulationsnekrose) führt. Die Funktionsweise einer MWA ist der Radiofrequenzablation ähnlich. Auch hier wird die Energie über einen in das Zielorgan eingebrachten Applikator verabreicht. Im Unterschied zur Radiofrequenzablation, bei der hochfrequenter elektrischer Strom erzeugt wird, bedient man sich bei der MWA hochfrequenter Radiowellen. Der Unterschied besteht in den höheren erzielbaren Temperaturen bei einer MWA, sowie der Möglichkeit die Energie in einem deutlich kürzeren Zeitraum zu applizieren. So lassen sich größere Areale in kürzerer Zeit abladieren. Kühleffekte von großen Gefäßen in der Nähe der Zielläsion sind im Rahmen einer MWA weniger relevant.²⁴ So dass einige Autoren der MWA den Vorzug vor der Radiofrequenzablation geben.²⁷

Ein lokal ablatives Verfahren eignet sich bei Irresektabilität des Herdes oder Patienten mit einem erhöhten perioperativen Risiko. Als Vorteile erweisen sich hier neben Kosteneffizienz, die Reduktion der Morbidität, reduzierte postoperative Schmerzsymptomatik und eine verkürzte Krankenhausaufenthaltsdauer.

Zunehmend rücken die ablativen Verfahren als primäre Therapieoption ins Licht.^{2,3,10} Aktuelle Metaanalysen kommen jedoch zu Ergebnissen, die einer chirurgischen Therapie eine Überlegenheit gegenüber ablativen Verfahren sowohl im Gesamtüberleben, als auch in der rezidivfreien Überlebenszeit bescheinigen.^{28,29}

1.5.3 Transarterielle Therapien (BCLC B)

Die Hauptmethode stellt hier die transarterielle Chemoembolisation (TACE) dar. Diese eignet sich für Patienten im intermediären Stadium der Erkrankung, wenn keine kurativen Optionen vorliegen. Die TACE weist hier einen Überlebensvorteil gegenüber best supportive care auf.³⁰

Die mittlere Überlebenszeit nach TACE beträgt 20-37 Monate (Median 30 Monate).³¹

Eine weitere Indikation zur TACE ist Bridging oder Downstaging im Rahmen einer geplanten Transplantation. Postoperative Resttumore und Rezidive stellen eine weitere Indikation dar.

Man unterscheidet die konventionelle TACE (cTACE, im weiteren Verlauf als TACE bezeichnet) von einer Embolisation mit medikamentenfreisetzenden Partikeln (drug eluting beads), kurz DEB-TACE.

Das Verfahren beruht auf einer selektiven Applikation eines Chemotherapeutikums. Hierbei wird ein Mikrokatheter in eine den Tumor versorgende Arterie eingebracht. Nach Applikation des Chemotherapeutikums erfolgt die Partikelembolisation des zuführenden Gefäßes. Das Ziel ist, neben dem Erreichen einer hohen lokalen Konzentration des Chemotherapeutikums, eine Ischämie des Tumorgewebes. Durch die lokale Abgabe reduziert sich die systemische Toxizität, und der Gefäßverschluss verhindert ein Auswaschen des Chemotherapeutikums aus dem Zielort.³²

Der Unterscheid einer DEB-TACE zur TACE ist, dass bei der DEB-TACE das Embolisatpartikel selbst zum Träger des Chemotherapeutikums wird.

In älteren Studien wurde kein relevanter Unterschied im Tumoransprechen und dem Gesamtüberleben zwischen TACE und DEB-TACE nachgewiesen.³³ In einer neueren Metastudie wird dagegen ein Vorteil der DEB-TACE im Gesamtüberleben gegenüber TACE postuliert.³⁴ Aufgrund heterogener Studiendesigns und international heterogener TACE-Durchführung bleibt die Überlegenheit der einen, gegenüber der anderen TACE-Methodik weiterhin kontrovers. Die Entscheidung zur jeweiligen TACE-Methode basiert somit häufig auf der Präferenz des entsprechenden behandelnden Teams.³⁵

Eine weitere Therapiemöglichkeit ist die transarterielle Radioembolisation (TARE), auch als selektive interne Radiotherapie bezeichnet (SIRT). Bei diesem Verfahren werden mit einem β -Strahler, meistens Yttrium-90, beladene Mikrosphären selektiv über einen arteriellen Zugang am Zielort verabreicht. In älteren Studien wird TARE oder SIRT als eine sichere Therapieoption mit vergleichbarem Tumoransprechverhalten zur TACE dargestellt.³⁶ Die bereits erwähnte Metastudie von Yang et al.³⁴ aus dem Jahr 2020 liefert Daten, die einen Vorteil der TARE gegenüber TACE im Gesamtüberleben bezeugen. In der selben Publikation wird TARE eine signifikant niedrigere Komplikationsrate im Vergleich zu sowohl DEB-TACE als auch TACE bescheinigt.³⁴

In den deutschen Leitlinien zur Therapie des HCC wird die TACE (cTACE und DEB-TACE werden dabei als gleichwertige Verfahren angesehen) vor der TARE empfohlen, mit einem Empfehlungsgrad A respektive Empfehlungsgrad 0.¹⁴

1.5.4 Systemische Therapie (BCLC C)

Bis zum Jahr 2008 stand für Patienten im fortgeschrittenen Stadium der Erkrankung keine effektive systemische Therapie zur Verfügung. Mit der SHARP-Studie konnte eine Zunahme der medianen Überlebenszeit durch Sorafenib (Tyrosinkinase-Inhibitor) mit 10,7 Monaten gegenüber einem Placebo mit 7,9 Monaten nachgewiesen werden.³⁷ Als Folge der Veröffentlichung wurde Sorafenib als erstes Mittel zur Therapie des HCC von der Food and Drug Administration (FDA) zugelassen, und spielt bis zum heutigen Tag eine relevante Rolle als Teil der Erstlinientherapie.¹⁴

Es dauerte weitere 10 Jahre bis mit Lenvatinib im Jahr 2018 (Tyrosinkinase-Inhibitor) ein weiteres Mittel zur Erstlinientherapie zugelassen worden ist. Lenvatinib zeigt ein ähnliches Wirkprofil wie Sorafenib mit medianen Überlebenszeiten von 13,6 Monaten gegenüber 12,3 Monaten unter Sorafenib.³⁸

Mittlerweile steht mit einer Kombinationstherapie von Atezolizumab (PD-L1-Antikörper) + Bevacizumab (VEGF-Antikörper) ein Therapie mit einer deutlichen Verbesserung der medianen Überlebenszeit auf 19,2 Monate zur Verfügung.^{39,40}

1.6 mRECIST (modified Response Evaluation Criteria in Solid Tumors)

Die Beurteilung des Therapieansprechens erfolgt anhand der mRECIST-Kriterien (modified Response Evaluation Criteria in Solid Tumors).⁴¹ Dies stellt eine Anpassung der RECIST-Kriterien an die Messung von HCC-Läsionen dar. Im Gegensatz zu den RECIST-Kriterien, wird bei der Messung lediglich der vitale Tumoranteil einbezogen.⁴¹

Anhand der präinterventionellen Bildgebung werden maximal zwei intrahepatische, nach den mRECIST-Kriterien messbare, Referenzläsionen (TL, target lesion) pro Patient bestimmt und die Summe der Längsachsendurchmesser gebildet. Liegen gleichzeitig verschiedene Läsionen mit typischem KM-Verhalten (arterielles Enhancement und venöses wash-out) und atypischem KM-Verhalten vor, so sollen Läsionen mit typischem KM-Verhalten bevorzugt in die Messung eingeschlossen werden, vorausgesetzt die Messung der Zielläsion ist reproduzierbar. Eine Zielläsion

sollte hierbei die Mindestgröße des vitalen Tumoranteiles von 1cm im Längsdurchmesser aufweisen und eine reproduzierbare Messung erlauben. Als vitaler Tumor werden dabei KM-anreichernde Tumoranteile in der arteriellen Phase einer CT oder MRT-Untersuchung angesehen.⁴¹ Bei der Messung sowohl in der präinterventionellen Bildgebung, als auch in der abschließenden Kontrolle werden Nekrosezonen des Tumors ausgespart (siehe Abbildung 11).

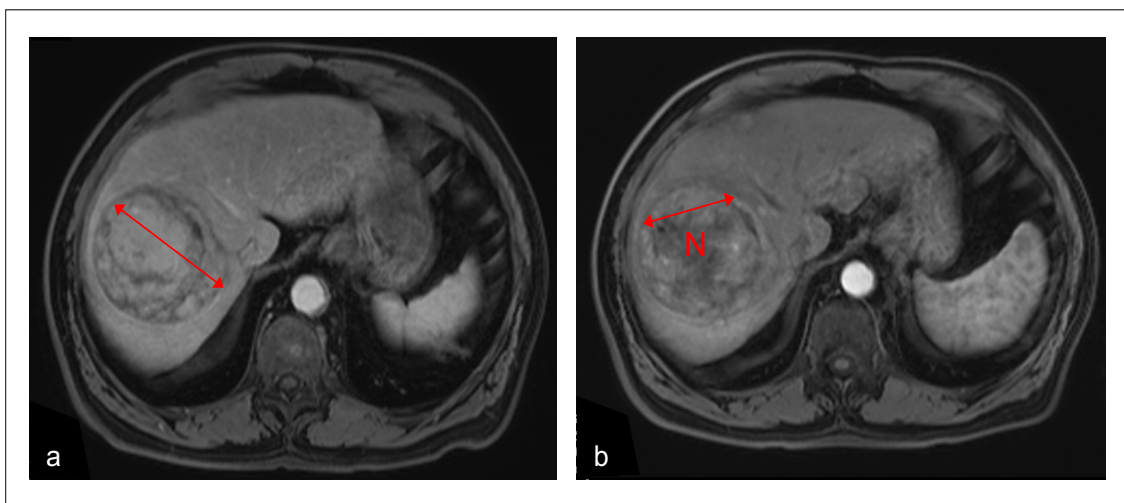


Abbildung 11. Differenzierung vitaler Tumoranteile nach mRECIST;

Messung des vitalen Tumoranteiles in arterieller MRT-Sequenz präinterventionell (a), und nach 3 TACE-Prozeduren (b) unter Aussparung des zentralen Nekroseareales („N“)

Nach Bestimmung der intrahepatischen Läsionen werden weitere extrahepatische TL bestimmt, wobei die Gesamtanzahl der TL nicht mehr als 5 betragen soll. Auch für die extrahepatischen Läsionen gilt die Regel der Reproduzierbarkeit der Zielläsionen bei einer Mindestgröße von ≥ 1 cm. Werden Lymphknoten als extrahepatische TL definiert, so müssen diese eine Mindestgröße von 1,5 cm in der kurzen Achse aufweisen. Als Ausnahme gelten periportal gelegene Lymphknoten, diese müssen eine Mindestgröße von ≥ 2 cm in der kurzen Achse aufweisen.

Neben TL werden auch non-target-Lesions (NTL) erfasst und dokumentiert. Hierunter fallen alle weiteren, eindeutig malignen Leberläsionen, mit typischem und atypischem KM-Verhalten, die nicht als TL definiert worden sind. Eine maligne

Pfortaderthrombose sowie extrahepatische Tumormanifestationen, wie z.B. eindeutige pleurale oder peritoneale Noduli sowie Lymphknoten werden in der initialen Bildgebung als auch in der abschließenden Kontrolle ebenso als NTL definiert. Diese NTL müssen nicht in der Größe dokumentiert werden, hier reicht eine Dokumentation der NTL im Verlauf.

Auf Ebene der Referenzläsionen wird eine komplette Remission (CR, complete response) oder partielle Remission (PR, partial response) als Ansprechen auf die Therapie gewertet. Unter CR wird der vollständige Rückgang einer arteriellen Anreicherung der Zielläsionen in der Kontrolluntersuchung verstanden. Eine PR wird als Rückgang des (summierten) Längsachsendurchmessers der vitalen Tumoranteile der Zielläsionen um mindestens 30% definiert. Ein Progress (PD, progressive disease) wird als Zunahme der vitalen Tumoranteile um mindestens 20% definiert. Der Zustand des stable disease (SD) wird als weder PR noch PD definiert.

Target lesions	Non-target lesions	New lesions		Overall response
CR	CR	No	→	CR
CR	NN	No	→	PR
PR	Non-PD	No	→	PR
SD	Non-PD	No	→	SD
PD	Any	Yes/No	→	PD
Any	PD	Yes/No	→	PD
Any	Any	Yes	→	PD

Abbildung 12. Bewertung des Gesamtansprechens anhand der mRECIST Kriterien, nach Llovet et al.⁴¹

1.7 Prognostische klinische Scores

Die Therapieentscheidung im Rahmen einer HCC-Erkrankung ist oft kompliziert und hängt maßgeblich vom Tumorstatus, der Leberfunktion, sowie dem generellen Gesundheitszustand des Patienten ab.

Um eine therapeutische Handlungsempfehlung zu etablieren, inadäquates Ansprechen zu erkennen, mögliche Übertherapie zu vermeiden und entsprechend unerwünschte Nebenwirkungen potentiell ineffizienter Therapien zu reduzieren, wurden eine Reihe prognostischer Scores über die letzten Jahre entwickelt.

1.7.1 Child-Pugh Score

Zuerst von Child und Turcotte⁴² im Jahr 1964 beschrieben. Im Jahr 1973 von Pugh modifiziert und in die heute bekannte und weit angewandte Form überführt (siehe Abbildung 13).⁴³ Initial wurde der Score zur Risikostratifizierung des perioperativen Risikos nach Blutung von ösophagealen Varizen bei zirrhotischen Patienten angewandt.

Table I.—GRADING OF SEVERITY OF LIVER DISEASE

CLINICAL AND BIOCHEMICAL MEASUREMENTS	POINTS SCORED FOR INCREASING ABNORMALITY		
	1	2	3
Encephalopathy (grade)*	None	1 and 2	3 and 4
Ascites	Absent	Slight	Moderate
Bilirubin (mg. per 100 ml.)	1-2	2-3	> 3
Albumin (g. per 100 ml.)	3.5	2.8-3.5	< 2.8
Prothrombin time (sec. prolonged)	1-4	4-6	> 6
For primary biliary cirrhosis:— Bilirubin (mg. per 100 ml.)	1-4	4-10	> 10

* According to grading of Trey, Burns, and Saunders (1966).

Abbildung 13. Child-Pugh Klassifikation, aus der Originalarbeit von Pugh et al.⁴³

Die hierbei mögliche erreichbare Punktzahl variiert von 5 bis 15. Abhängig der Punktzahl werden die Patienten in eine Klasse kategorisiert:

Child Pugh A (5-6 Punkte), Child Pugh B (7-9 Punkte,) Child Pugh C (10-15 Punkte).

Im zeitlichen Verlauf wurde der Score als prädiktiver Faktor eines generalisierten operativen Eingriffes bei zirrhotischen Patienten etabliert. Schließlich wurde die Anwendung des Scores auf die generelle Prognose von Patienten mit einer Leberzirrhose ausgeweitet.

Der Einschluss von hepatischer Enzephalopathie und Aszites wird beim Child-Pugh Score als ein subjektiv bewertetes Kriterium bemängelt.⁴⁴

1.7.2 MELD-Score (Model of End-stage Liver Disease)

Der MELD-Score wurde zunächst als prädiktiver Score etabliert, um das Outcome nach Anlage eines transjugulären hepatischen portosystemischen Shuntes (TIPS), unabhängig von der zugrunde liegenden Lebererkrankung als Ursache der portalen Hypertension, abzuschätzen.⁴⁵

In der Folge zeigte sich eine bessere Prognoseabschätzung im Vergleich zu Child-Pugh in Hinsicht auf die dekompensierte Leberzirrhose und die sich daraus ergebende bessere Abschätzung der Dringlichkeit zur Lebertransplantation.⁴⁶

Als Vorteil gegenüber Child-Pugh wird die Streichung der bereits erwähnten subjektiven Kriterien für hepatische Enzephalopathie und Aszites angesehen. Zur Berechnung des MELD-Scores werden nur objektiv messbare Parameter herangezogen: Serum-Bilirubin, Serum-Kreatinin und INR (International normalized Ratio). Als Nachteil kann jedoch die komplizierte Formel hinter MELD angesehen werden. Der Score kann im klinischen Alltag letztendlich nur mittels eines Rechners bestimmt werden.

Die Formel zur Berechnung des MELD-Scores lautet:

$$10 \times \{0,957 \times \log_e(\text{Kreatinin mg/dl}) + 0,378 \times \log_e(\text{Bilirubin mg/dl}) + 1,120 \times \log_e(\text{INR}) + 0,643\}$$

Der maximal erreichbare Wert liegt bei 40. In der Berechnung wird auf ganze Zahlen gerundet. Der errechnete Wert korreliert mit der Dreimonatsmortalität (siehe Abbildung 14). Bei einem MELD-Score von 40 liegt diese bei nahezu 100%.⁴⁷

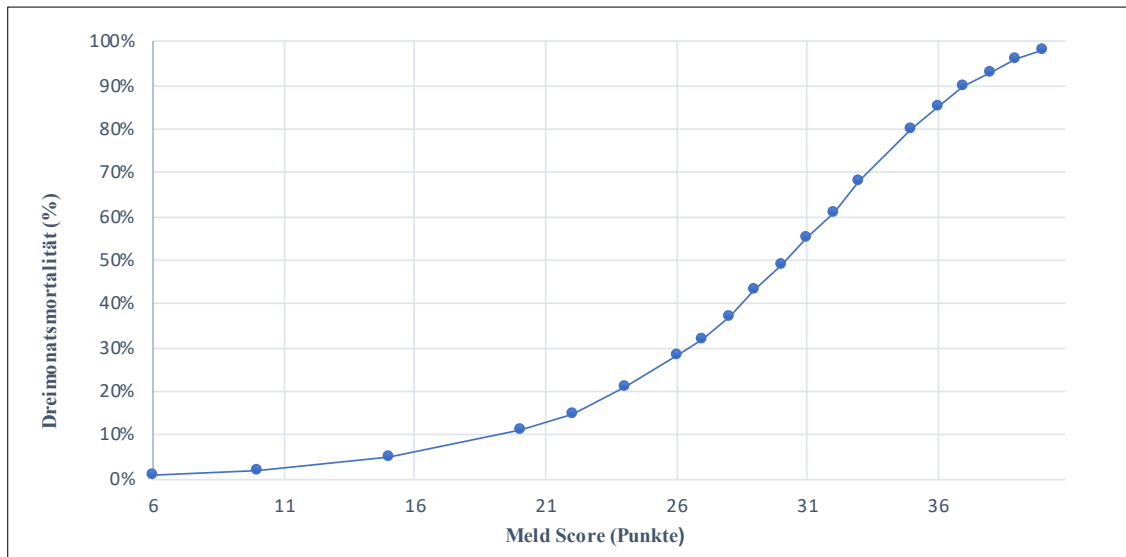


Abbildung 14. Dreimonatsmortalität (y-Achse) in Abhängigkeit vom MELD-Score (X-Achse), nach Kaiser et al.⁴⁷

1.7.3 ALBI-Score (Albumin-Bilirubin-Score)

Ein weiterer Versuch der Vereinfachung und Objektivierung des Child-Pugh-Scores ist der ALBI-Score. Wie der Name impliziert wird der Score nur Anhand des Serum-Albumin und Serum-Bilirubin berechnet.

Das Ziel des Scores ist die Abschätzung der Leberfunktion, unter der Annahme, dass diese signifikant das Überleben beeinflusst. Die Ergebnisse zeigen, in Bezug auf das Überleben, eine dem Child-Pugh-Score ebenbürtige, und teilweise bessere Prognose des Outcomes.⁴⁸⁻⁵⁰

Die Formel zur Berechnung des ALBI-Scores lautet:

$$x = (\log_{10} \text{Bilirubin } (\mu\text{mol/L}) \times 0,66) + (\text{Albumin}(g/L) \times -0,085)$$

Abhängig vom errechneten Wert x wird der Patient in eine von 3 Gruppen eingeteilt:

$$\text{ALBI 1: } x \leq -2,60$$

$$\text{ALBI 2: } x > -2,60 \text{ -- } \leq 1,39$$

$$\text{ALBI 3: } x > -1,39$$

Dabei nimmt die Überlebenszeit von ALBI 1 zu ALBI 3 ab.⁴⁸

1.7.4 HAP-Score (Hepatoma arterial-embolisation prognostic score)

Zuerst 2013 von Kadalayil et al.⁵¹ beschrieben mit dem Ziel einen einfachen Score zu entwickeln, der zuverlässige Aussagen zur Mortalitätsprognose von Patienten mit irresektablem HCC, die sich einer TACE unterziehen, erlaubt.

Dabei stützt sich der Score auf objektive und messbare Größen: Albumin, Alpha-Fetoprotein (AFP), Bilirubin sowie Tumorgroße. Jedem Wert wird ein Grenzwert gegeben. Wird dieser unter- oder überschritten bekommt der Patient einen Punkt. Abhängig der Punktesumme definiert sich die Risikogruppe mit: HAP A als Niedrigrisikogruppe bis HAP D als Höchststrisikogruppe.

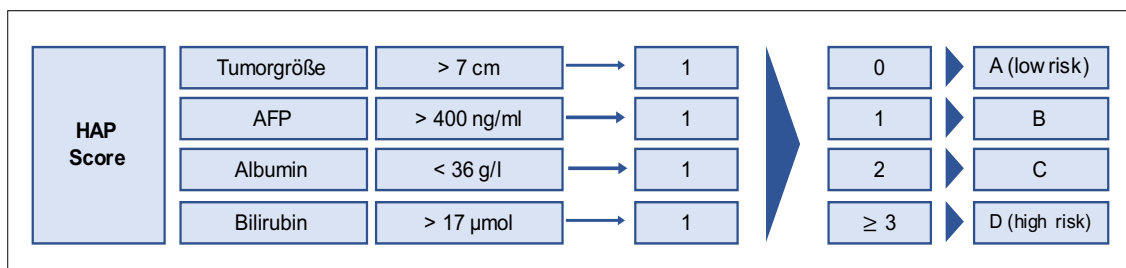


Abbildung 15. HAP Score, nach Müller et al.⁵²

Die Ergebnisse der Studie zeigen, dass der HAP-Score gegenüber den verglichenen Scoringsystemen (unter anderem BCLC, Child-Pugh und MELD) eine deutliche Überlegenheit in der Mortalitätsprognose aufweist.

Das mediane Gesamtüberleben oder Overall Survival (OS) verteilt sich über die entsprechenden HAP-Gruppen folgendermaßen: Gruppe A: 27,6 Monate, B: 18,5 Monate, C: 9,0 Monate, D: 3,6 Monate.⁵¹

1.7.5 mHAP-II-Score (modified HAP-II-Score)

Der bereits beschriebene HAP-Score ist von Park et al.⁵³ um einen zusätzlichen Bestandteil erweitert worden. Bei der Anzahl von Tumoren ≥ 2 wird dem mHAP-II-Score ein Punkt addiert (siehe Abbildung 16). Die weitere Berechnung bleibt zum HAP-Score unverändert. Die Einbeziehung des zusätzlichen Faktors Tumoranzahl führt zu einer signifikanten Verbesserung der Prognoseabschätzung und so einer zuverlässigen Vorauswahl von HCC Patienten, die sich optimal für eine TACE-Behandlung eignen.^{53,54} Die Fläche unter der ROC-Kurve (AUROC) beträgt für das OS nach 3 Jahren 0,717 vs. 0,658 und nach 5 Jahren 0,728 vs. 0,645 (HAP vs. mHAP-II).⁵³

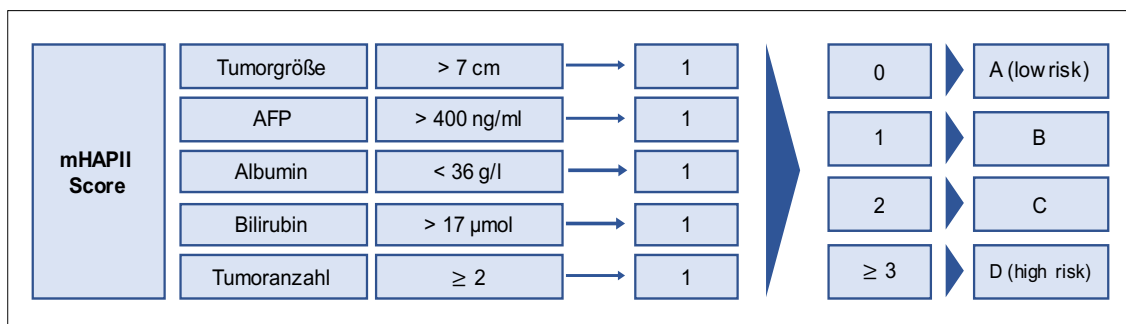


Abbildung 16. mHAP-II Score, nach Müller et al.⁵⁵

1.7.6 STATE Score (Selection for TrAnsarterial chemoembolisation TrEatment)

Das Ziel der Entwicklung des STATE-Scores war es, einen objektiven Score zur Identifizierung von Patienten, die von einer TACE profitieren würden, zu definieren.

Auch wenn eine TACE für Patienten im Stadium BCLC B empfohlen wird, ist es die häufigste Erstlinientherapie die zur Anwendung kommt, obwohl nur 10-15% der Patienten zum Diagnosezeitpunkt dem Stadium BCLC B zuzuordnen sind.⁵⁶ Unter der Annahme, dass innerhalb der BCLC B Kategorie eine erhebliche Heterogenität der Erkrankten herrscht, wurde ein Score entwickelt der 3 unabhängige Variablen vereint. Jede Variable für sich stellt einen prädiktiven Faktor für das Überleben dar.

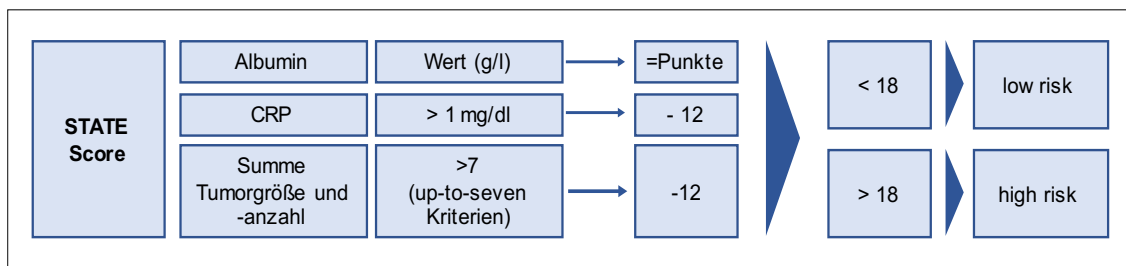


Abbildung 17. STATE Score, nach Müller et al.⁵⁵

Mit dem STATE-Score kann innerhalb der heterogenen BCLC B Gruppe eine Differenzierung zwischen sich für eine TACE-Prozedur geeigneten und ungeeigneten Patienten vorgenommen werden (siehe Abbildung 17). Patienten die als „low risk“ definiert werden haben ein deutliches längeres Gesamtüberleben im Vergleich zur „high risk“ Gruppe (20,5 Monate und 6,1 Monate).⁵⁷

Dabei ist der Score unabhängig von der gewählten TACE Therapie (cTACE oder DEB-TACE), sowie der Ätiologie der zugrundeliegenden Tumorerkrankung (Alkohol, Virushepatitis, usw.).⁵⁷

1.7.7 6 and 12 Score

Der Score soll ähnlich dem STATE-Score eine Vorauswahl der für eine TACE geeigneten Patienten auf eine einfache und schnelle Art ermöglichen.

Mit der Inklusion der Anzahl der Tumoren und der Summe der Tumordurchmesser ist der Score im Vergleich zu anderen vorgestellten Scores einfach zu ermitteln (siehe Abbildung 18).

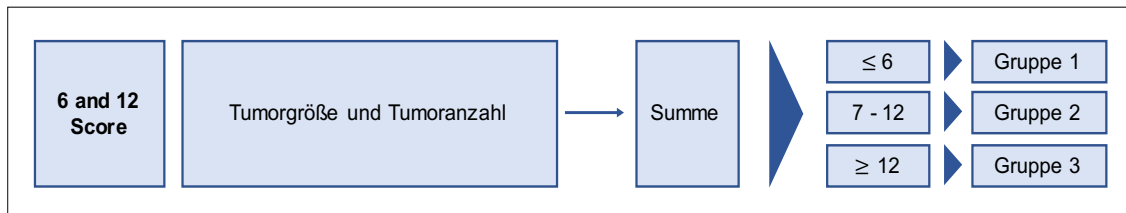


Abbildung 18. 6 and 12 Score, nach Müller et al.⁵⁵

Im Rahmen der Erstbeschreibung des Scores zeigt sich eine signifikante Differenz im Gesamtüberleben abhängig der Eingruppierung der Patienten (Gruppe 1: median OS 49,1 Monate, Gruppe 2: median OS 32,0 Monate, Gruppe 3: median OS 15,8 Monate).⁵⁸ Die Patienten in der Gruppe 3 haben eine kürzere Überlebenszeit mit TACE, als mit einer systemischen Behandlung mit Azetolizumab + Bevacizumab mit einem medianen OS von 19,2 Monaten.^{39,40}

Trotz der Einfachheit des Scores sind die in der Studie verglichenen HAP- und ALBI-Scores dem 6 and 12 Score nicht überlegen. Wang et al.⁵⁸ führen das Ergebnis auf die größere prognostische Wertigkeit der Tumorlast im Vergleich zur Leberfunktion zurück.⁵⁸

In weiteren Studien zeigt sich der 6 and 12 Score in der Abschätzung des Gesamtüberlebens nach TACE den verglichenen HAP-, mHAP-II- und ALBI-Score überlegen.⁵⁹

1.8 Künstliche Intelligenz / Radiomics

Die Weiterentwicklung der Computersysteme, Rechenkapazitäten und Möglichkeiten komplexer bioinformatischer Auswertestrategien ermöglicht es unzählige quantitative

Werte aus vorhandenen Bilddaten, wie beispielsweise MRT oder CT zu extrahieren. Der Ausdruck „Omics“ steht ganz allgemein für die umfassende Charakterisierung von Strukturen durch, unter anderem, Quantifizierung und bioinformatische Integrationen. Bereits länger bekannt ist der Ausdruck Genomics, basierend auf umfassenden Genanalysen.⁶⁰

Unter Radiomics versteht man eine quantitative Herangehensweise an medizinische Bilddaten.⁶¹ Dahinter steht der Gedanke, dass in medizinischen Bilddaten mehr Informationen verborgen sind, als das menschliche Auge wahrnehmen kann. Wie Gillies et al.⁶¹ in einer Arbeit postulieren: „Images are more than pictures, they are data“.⁶¹ Es basiert auf der Annahme, dass durch diese, visuell-semantischen, aktuell nicht in der klinischen Routine genutzten, Informationen Rückschlüsse auf die zugrundeliegenden pathophysiologischen Aspekte wie z.B. Tumorgenetik, Phänotyp, die Pathophysiologie und das Mikroumfeld des Tumors gezogen werden können.⁶¹⁻⁶⁴

Radiomics zielt auf die Extraktion von quantitativen und reproduzierbaren Merkmalen aus medizinischen Bilddatensätzen ab, die für das menschliche Auge nur schwer, oder nicht erfassbar sind.⁶⁵ Die Daten müssen, um sie beurteilbar zu machen, aufbereitet werden. Gillies et al.⁶¹ beschreiben den Begriff als „mineable“, damit wird umschrieben, dass man Daten so aufbereiten muss, dass sie auf unterschiedlichste mathematische Arten immer wieder neu betrachtet werden können, um etwaige, initial nicht erkennbare Zusammenhänge, darstellen zu können.⁶¹ Mit Hilfe von mathematischen Analysen werden räumliche Verteilung von Signalintensitäten und die Beziehung der Pixel/Voxel untereinander darstellbar, und für weitere Analysen verwertbar gemacht.^{64,66} Die extrahierten Merkmale werden durch Methoden des maschinellen Lernens untersucht und lassen Rückschlüsse auf, unter anderem, Diagnose, Therapieansprechen und Outcome zu.^{67,68}

Der Prozess von Datenerhebung bis Analyse erfolgt in mehreren Schritten (siehe Abbildung 19). Nach Erhebung der Bilddaten (CT oder MRT) erfolgt im nächsten Schritt das sogenannte Preprocessing, oder Aufbereitung der Bilddaten. Das Preprocessing soll eine Einheitlichkeit und Konsistenz der Bilddaten gewährleisten. Die Region of Interest (ROI) oder Volume of Interest (VOI) kann wahlweise manuell, semiautomatisch oder vollautomatisch segmentiert werden. Aus der Segmentation erfolgt die Extraktion der Bildmerkmale oder Features. Hierbei werden Features

unterschiedlicher Klassen extrahiert. Dadurch wird eine große Anzahl an Features generiert.

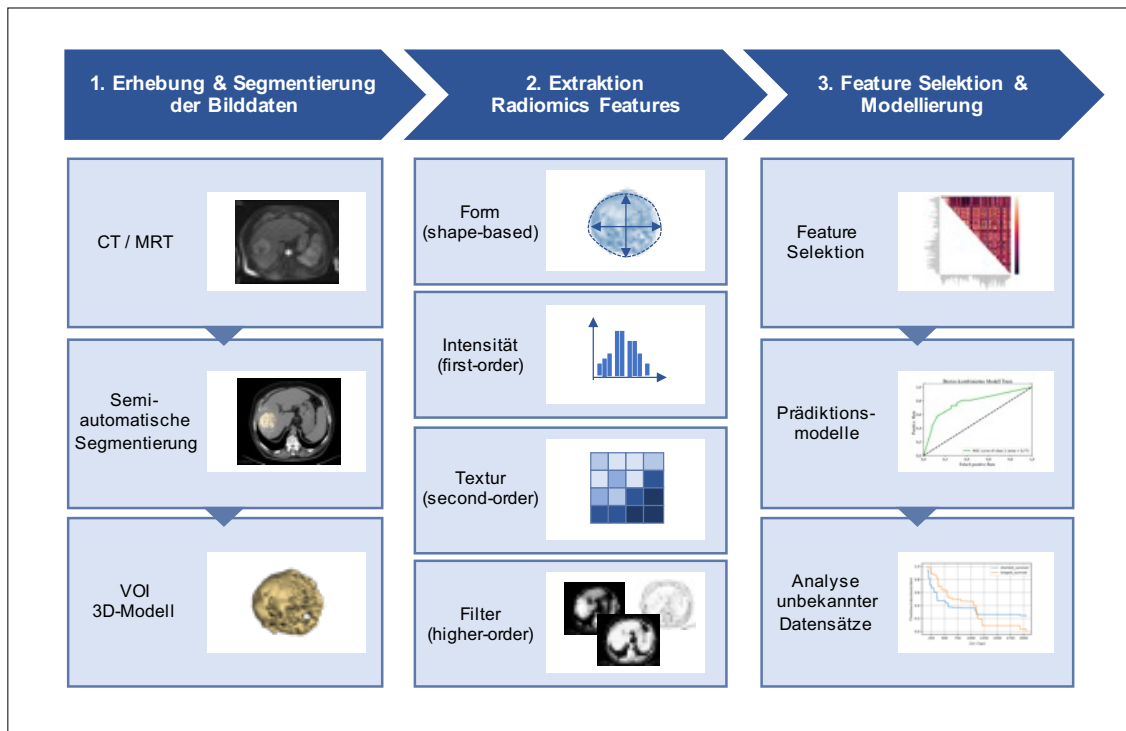


Abbildung 19. Workflow Radiomics

Mit dem Begriff der Feature Extraktion wird die Kernidee von Radiomics beschrieben. Bei diesem Schritt des Radiomics-Workflows werden, die bereits erwähnten, vom Menschen primär nicht wahrnehmbaren, Informationen ausgelesen. Diese werden durch Algorithmen aus dem Bild extrahiert und können in shape-based, first-order, second-order und higher-order Features unterteilt werden.

Unter shape-based versteht man Features welche, unabhängig der Grauwerte, die Größe und Form der segmentierten Läsion beschreiben.

Unter first-order versteht man die Verteilung der Grauwerte insgesamt, unabhängig von der räumlichen Auflösung, auch als Histogramm bekannt. Aus diesem Histogramm können mehrere Features wie z.B.: Streuung (Standard Deviation), Asymmetrie (Skewness), Gipfel der Verteilung (Kurtosis), Streuungsmaß (Entropy) und weitere Parameter berechnet werden.

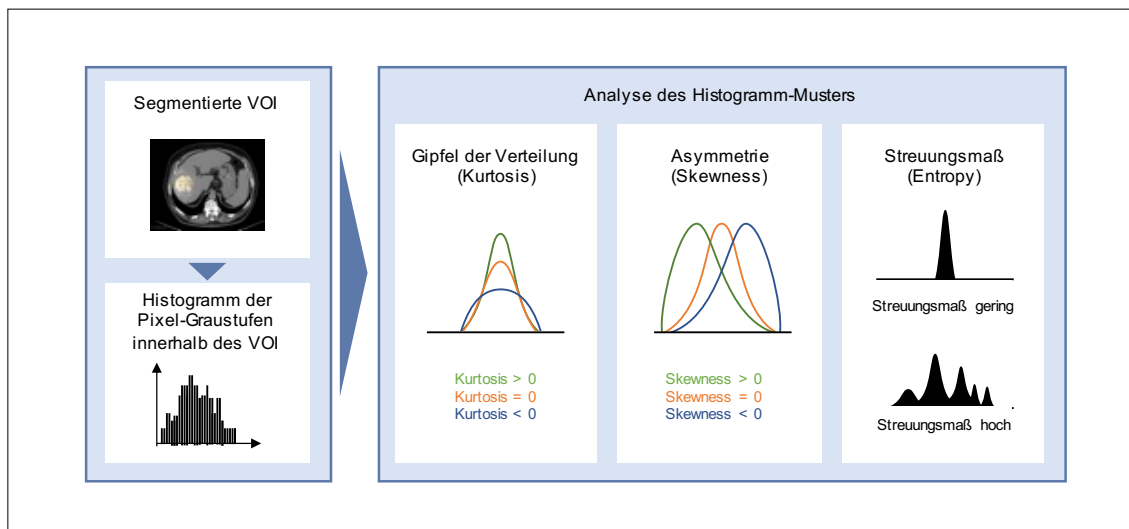


Abbildung 20. First-order Features, nach Park et al.⁶⁹

Second-order Features, oder sogenannte Texturparameter, basieren auf der multivariaten Verteilung von Voxelpaaren. Vereinfacht ausgedrückt handelt es sich dabei um die räumliche Anordnung von Mustern denen relevante Informationen zugrunde liegen können. Für die Auswertung wird aus dem Bildausschnitt eine Matrix der Voxel erstellt. Jedem Voxel dieses Ausschnittes wird, abhängig seines Grauwertes, ein Wert zugeteilt. Auf Basis dieser Werte werden second-order Features extrahiert. Hierzu zählt z.B. die Gray Level Co-Occurance Matrix (GLCM). GLCM beschreibt die Häufigkeit bestimmter benachbarter Voxelpaare. Aus der sich so ergebenden zweiten Matrix werden wiederum Histogramme berechnet mit den sich so erneut ergebenden Werten für Kurtosis, Entropy, Skewness und weitere Parameter berechnet werden können.

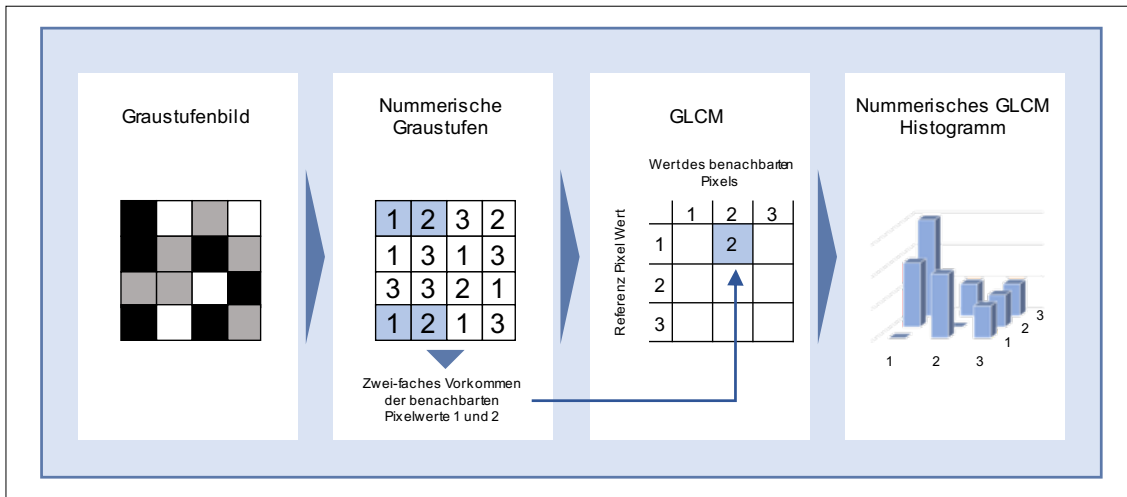


Abbildung 21. Second-order Features, nach Park et al.⁶⁹

Bei higher-order Features werden die Bilddaten mittels Filtern alteriert. Zum Beispiel wird der Laplace-Filter zur Kantenanhebung angewendet. Der Gauß-Filter wird verwendet um eine Glättung durch Reduktion des Rauschens zu erreichen. Beide Filter können kombiniert zur Anwendung kommen. Darüber hinaus gibt es unzählige weitere Filter. Die alterierten Bilddaten werden im Anschluss herangezogen um daraus wiederum sowohl first-order als auch second-order Features zu extrahieren.

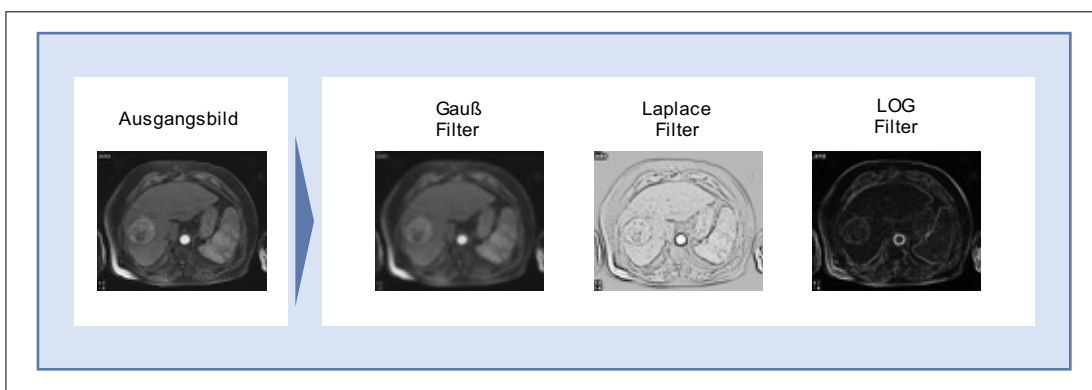


Abbildung 22. Higher-order Features, nach Park et al.⁶⁹

Durch die Anwendung von maschinellem Lernen wird der Pool der zur Verfügung stehenden Features eingegrenzt. Die sogenannte Feature Selektion erfolgt entsprechend

eines durch den Benutzer definierten Problems. Nur die Features mit der höchsten Korrelation verbleiben im Pool zur weiteren Analyse. Anhand der verbleibenden Features wird ein Prädiktionsmodell gebildet. Dieses wird mit den bekannten Features angelernt, mit dem Ziel das zuvor definierte Problem anhand nicht bekannter Datensätze zu evaluieren.

Die Anwendbarkeit, der durch Radiomics ermöglichten Dateninterpretation, auf den klinischen Alltag wurde bereits mehrfach unter Beweis gestellt.⁷⁰⁻⁷⁶ Im Jahr 2021 veröffentlichte die nordamerikanische Gesundheitsbehörde FDA einen Aktionsplan mit der Absicht künstliche Intelligenz und insbesondere Machine Learning in der Medizin zu fördern und fordern, sowie die Akzeptanz in der Bevölkerung zu unterstützen.^{77,78} Schon die Überschrift ist wegweisend: „...Software as a medical device...“.^{77,78} Dies unterstreicht die Notwendigkeit, sich mit der künstlichen Intelligenz in der Medizin nicht nur als Radiologe, sondern generell als Mediziner, unabhängig von der Fachspezialisierung, auseinander zu setzen.

1.9 Zielsetzung

Es existieren multiple Scores zur Einschätzung des Gesamtüberlebens von Patienten die sich einer TACE unterziehen. Diese werden aktuell jedoch nur für klinische Studien empfohlen und die finale Therapieentscheidung wird in interdisziplinären Tumorboards aufgrund von Expertenkonsens getroffen.

Die Zielsetzung dieser Arbeit ist es, mit Hilfe von Radiomics und maschinellem Lernen, ein objektives Modell zur Prognoseabschätzung des Therapieansprechens und Gesamtüberlebens von HCC-Patienten die sich einer TACE unterziehen zu entwickeln.

Im Kontext des Technologiefortschritts durch Radiomics und Machine Learning, stellen wir die Hypothese auf:

- durch die Auswertung und Integration von Lipidolablagerungen in der Zielläsion nach der ersten durchgeführten TACE ist eine zuverlässigere

Prognose zum Therapieansprechen und Gesamtüberleben mit Hilfe von Radiomics möglich, als dies klinische Scores alleine erlauben

Im klinischen Setting erfolgen bisweilen mehrere TACE-Sitzungen vor einer Re-Evaluation der Therapie und einer möglichen Änderung des Therapiekonzeptes.

Bei Bestätigung unserer Hypothese wird die Identifizierung von HCC-Patienten, die von einer TACE profitieren, bzw. die keinen Erfolg nach einer TACE-Behandlung erleben, bereits nach der ersten TACE-Sitzung ermöglicht. Diese Einschätzung kann Nebenwirkungen reduzieren und zu einer Verbesserung des Behandlungsschemas führen.

2. Material und Methoden

2.1 Patientenkollektiv

Die dieser Studie zugrunde liegenden Daten wurden durch das Universitäre Centrum für Tumorerkrankungen Frankfurt (UCT) zur Verfügung gestellt.

Eine schriftliche Einverständniserklärung wurde von allen teilnehmenden Patienten eingeholt. Die Studie wurde durch das Ethik-Komitee des Universitätsklinikum Frankfurt am Main genehmigt.

Im Rahmen der retrospektiven Studie wurden insgesamt 61 HCC-Patienten [12 Frauen (19,7% der Patientenpopulation) und 49 Männer (80,3%), Durchschnittsalter bei Diagnosestellung 66 Jahre, Spannweite 37 – 86 Jahre] eingeschlossen, welche im Zeitraum zwischen Januar 2012 und Januar 2020 eine Behandlung mittels TACE bekommen haben. Das Patientenkollektiv ist in der nachfolgenden Tabelle 1 aufgeschlüsselt.

Merkmal	Gesamt	Train-Set	Test-Set
Anzahl Patienten	61	42	19
Geschlecht (%)			
<i>m</i>	49 (80,3)	35 (83,3)	14 (73,7)
<i>w</i>	12 (19,7)	7 (16,7)	5 (26,3)
Alter bei Diagnose (Jahre)			
<i>Median</i>	66	67	63
<i>Range</i>	37 - 86	37 - 86	50 - 86
Abstand Diagnose zu TACE (Tage)			
<i>Median</i>	43	42	59
<i>Range</i>	3 - 978	3 - 874	17 - 978
Leberläsionen (%)			
<i>1</i>	24 (39,3)	16 (38,1)	8 (42,1)
<i>2</i>	7 (11,5)	4 (6,6)	3 (15,8)
<i>3</i>	8 (13,1)	6 (9,8)	2 (10,5)
<i>>=4</i>	22 (36,1)	16 (26,2)	6 (31,6)
Lokalisation der Leberläsionen (%)			
<i>unilobär</i>	42 (68,9)	29 (69,0)	13 (68,4)
<i>bilobär</i>	19 (31,1)	13 (31,0)	6 (31,6)

Maximale Herdgröße (cm)			
<i>Median</i>	2,9	2,2	3,4
<i>Range</i>	1,0 - 10,1	1,0 - 9,8	1,0 - 10,1
Ätiologie HCC (%)			
<i>Hepatitis B</i>	9 (14,8)	6 (14,3)	3 (15,8)
<i>Hepatitis C</i>	14 (23,0)	10 (23,8)	4 (21,1)
<i>Alkohol</i>	19 (31,1)	14 (33,3)	5 (26,3)
<i>NASH</i>	5 (8,2)	3 (7,1)	2 (10,5)
<i>Alkohol + virale Hepatitis</i>	6 (9,8)	4 (9,5)	2 (10,5)
<i>Alkohol + NASH</i>	1 (1,6)	1 (2,4)	0 (0)
<i>Andere (kryptogene Zirrhose, Autoimmunhepatitis)</i>	7 (11,5)	4 (9,5)	3 (15,8)
T der TNM Klassifikation			
<i>T1</i>	20 (32,8)	15 (35,7)	5 (26,3)
<i>T2</i>	21 (34,4)	14 (33,3)	7 (36,8)
<i>T3</i>	10 (16,4)	7 (16,7)	3 (15,8)
<i>T4</i>	0 (0,0)	0 (0,0)	0 (0,0)
<i>unbekannt</i>	10 (16,4)	6 (14,3)	4 (21,1)
N der TNM Klassifikation			
<i>N0</i>	50 (82,0)	36 (85,7)	14 (73,7)
<i>N1</i>	0 (0,0)	0 (0,0)	0 (0,0)
<i>Nx</i>	2 (3,3)	1 (2,4)	1 (5,3)
<i>unbekannt</i>	9 (14,8)	5 (11,9)	4 (21,1)
M der TNM Klassifikation			
<i>M0</i>	50 (82,0)	35 (83,3)	15 (78,9)
<i>M1</i>	3 (4,9)	2 (4,8)	1 (5,3)
<i>Mx</i>	0 (0,0)	0 (0,0)	0 (0,0)
<i>unbekannt</i>	8 (13,1)	5 (11,9)	3 (15,8)
BCLC vor TACE			
<i>A</i>	18 (29,5)	14 (33,3)	4 (21,1)
<i>B</i>	33 (54,1)	24 (57,1)	9 (47,4)
<i>C</i>	10 (16,4)	4 (9,5)	6 (31,6)
<i>D</i>	0 (0,0)	0 (0,0)	0 (0,0)
Child Pugh Score vor TACE			
<i>A</i>	36 (59,0)	25 (59,5)	11 (57,9)
<i>B</i>	6 (9,8)	5 (11,9)	1 (5,3)
<i>C</i>	0 (0,0)	0 (0,0)	0 (0,0)
<i>unbekannt</i>	19 (31,1)	12 (28,6)	7 (36,8)
MELD-Score vor TACE			
<i><6</i>	1 (1,6)	0 (0,0)	1 (5,3)
<i><10</i>	23 (37,7)	13 (31,0)	10 (52,6)
<i><15</i>	11 (18,0)	10 (23,8)	1 (5,3)

<20	1 (1,6)	1 (2,4)	0 (0,0)
<i>unbekannt</i>	25 (41,0)	18 (42,9)	7 (36,8)
Albumin (g/dl)			
<i>Median</i>	3,8	3,8	3,9
<i>Range</i>	1,8 - 7,2	1,8 - 7,2	3,0 - 4,5
Bilirubin (mg/dl)			
<i>Median</i>	0,8	1,1	0,9
<i>Range</i>	0,3 - 2,2	0,3 - 2,2	0,3 - 1,3
INR			
<i>Median</i>	1,12	1,13	1,12
<i>Range</i>	0,92 - 3,30	0,92 - 3,30	1,00 - 1,79
CRP (mg/dl)			
<i>Median</i>	0,41	0,46	0,32
<i>Range</i>	0,03 - 4,39	0,03 - 4,39	0,07 - 2,89
AFP (ng/ml)			
<i>Median</i>	12,9	9,2	17,3
<i>Range</i>	2,1 - 60500,0	2,1 - 60500,0	2,2 - 9276,0

Tabelle 1. Patientenkollektiv

Die Einschlusskriterien umfassten:

- histologisch gesichertes HCC
- drei aufeinanderfolgende TACE ausschließlich mit dem Wirkstoff Mitomycin C (Medac®, Hamburg, Deutschland) sowie Lipiodol (Guerbet GmbH, Frankreich) ± Degradable Starch Microspheres (EmboCept®S, PharmaCept GmbH, Berlin, Germany)
- alle der drei TACE wurden im gleichen Leberlappen durchgeführt, sowie alle Zielläsionen (Target Lesions, TL) im Rahmen jeder TACE behandelt
- native CT-Verlaufskontrolle 24 Stunden nach TACE
- KM gestützte CT und/oder MRT mit arterieller, venöser und/oder spätvenöser Phase vor erster TACE
- KM gestützte CT und/oder MRT mit arterieller, venöser und/oder spätvenöser Phase nach der dritten TACE im Rahmen der Kontrolle des Therapieansprechens.

Als Ausschlusskriterien gelten hingegen:

- TL die nicht im Rahmen aller drei TACE behandelt worden sind
- Zeitintervall zwischen erster und dritter TACE >6 Monate
- vorausgegangene lokale Therapie der TL
- keine reproduzierbare Messung nach modified RECIST (mRECIST)-Kriterien möglich
- insuffiziente Bildqualität
- alternative Chemotherapeutika zu Mitomycin C in der TACE.

Die angewandten Ein- und Ausschlusskriterien sind in der nachfolgenden Abbildung 19 nach STARD-Kriterien aufgeführt.⁷⁹

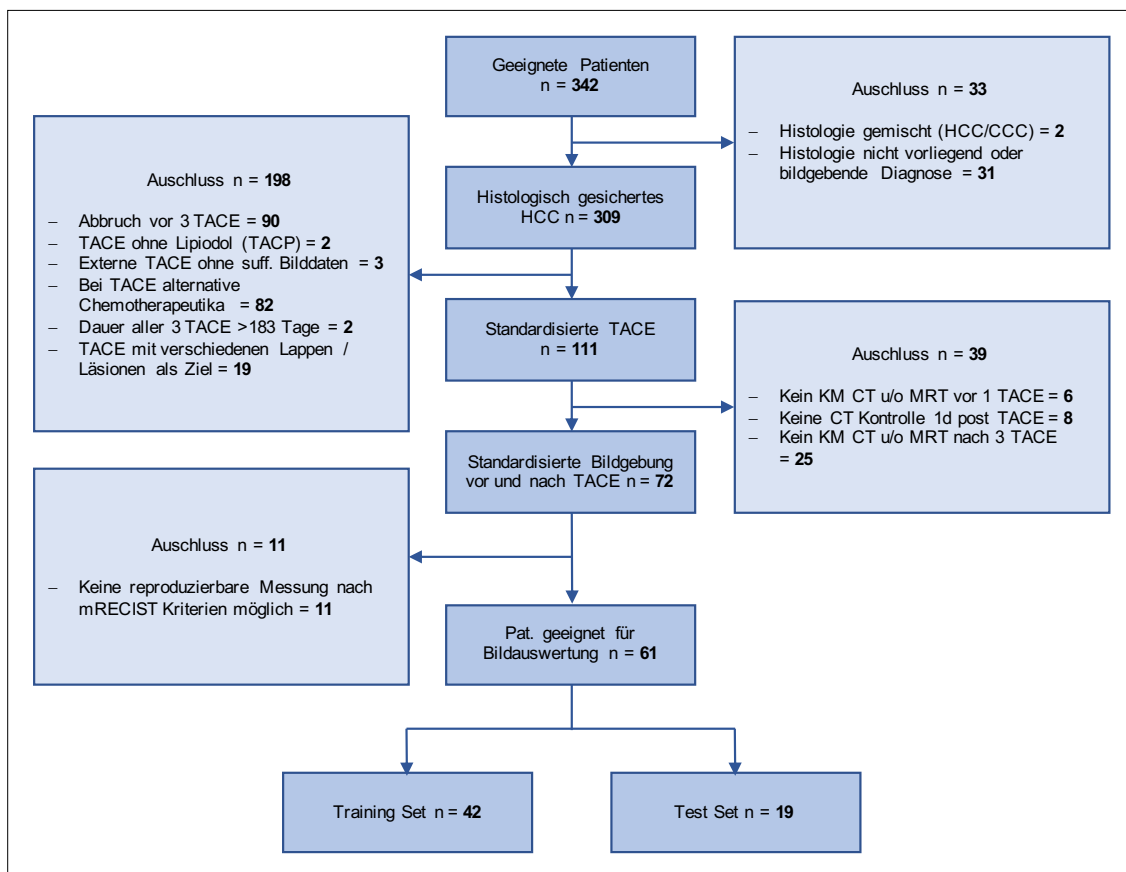


Abbildung 23. STARD-Flowchart der Ein- und Ausschlusskriterien

2.2 TACE

Nach der klinisch indizierten Indikationsstellung erfolgt eine ausführliche Aufklärung des Patienten über Indikation, Vorgehen und mögliche Nebenwirkungen der TACE-Prozedur im Rahmen der klinischen Routine. Die häufigste mögliche Nebenwirkung stellt hierbei das Postembolisationssyndrom dar. Übelkeit, Erbrechen, Fieber und Abdominalschmerzen sind typische Symptome. Die Dauer der Symptomatik kann hierbei von wenigen Stunden bis zu mehreren Tagen anhalten. Je nach Studie wird das Auftreten eines Postembolisationssyndroms mit 30-90% angegeben.^{80,81}

Nach Darstellung der Aorta abdominalis wird mit einem Kobra- oder Sidewinderkatheter die Gefäßanatomie über die Darstellung des Truncus coeliacus und der A. mesenterica superior abgebildet. Hierbei werden die tumorversorgenden Gefäße und mögliche Kollateralen für die Intervention identifiziert. Bei bilobärem Befall wird der Leberlappen mit der größeren Tumorlast behandelt.

Nach superselektiver Sondierung des tumorversorgenden Gefäßes mittels eines Mikrokatheters wird Mitomycin C (Medac®, Hamburg, Deutschland) mit einer maximalen Dosis von 8mg/m² KOF (Median: 9,87 mg ± 0,52) in einer Emulsion mit Lipiodol (Guerbet GmbH, Frankreich) ± Degradable Starch Microspheres (EmboCept®S, PharmaCept GmbH, Berlin, Germany) unter gepulster Durchleuchtung appliziert. Die Applikation erfolgt bis zum Sistieren des Flusses um einen Reflux des Therapeutikums in die Seitäste der A. hepatica propria sowie des Truncus coeliacus zu vermeiden. In 19,7% der Sitzungen (36/183 TACE-Prozeduren, 29,5% der Patienten n = 18/61) wurden zusätzlich abbaubare Stärkemikrosphären verabreicht (EmboCept®S, PharmaCept GmbH, Berlin, Germany) mit einer Dosis von Median 182,63 mg ± 84,97.

Um die Lipiodol-Anreicherung zu dokumentieren wird 24 Stunden nach der TACE-Intervention eine native CT des Oberbauches durchgeführt. Die Untersuchungsparameter sind hierbei 120 ± 13,5 kV und 213 ± 54,7 mAs bei einem Dosis-Längen-Produkt von 238,4 ± 153,1 mGy x cm (jeweils Median ± 1 Standardabweichung). Die Schichtdicke beträgt in der axialen Schichtung jeweils 5 mm.

Bei jedem, in der Studie berücksichtigten Patienten, wurden insgesamt drei TACE-Prozeduren in einem Abstand von maximal 6 Monaten durchgeführt. Im Anschluss

erfolgt eine Kontrolle des Ansprechens auf die Intervention mittels einer KM-gestützten CT und/oder MRT-Untersuchung (siehe Abbildung 24).¹⁴

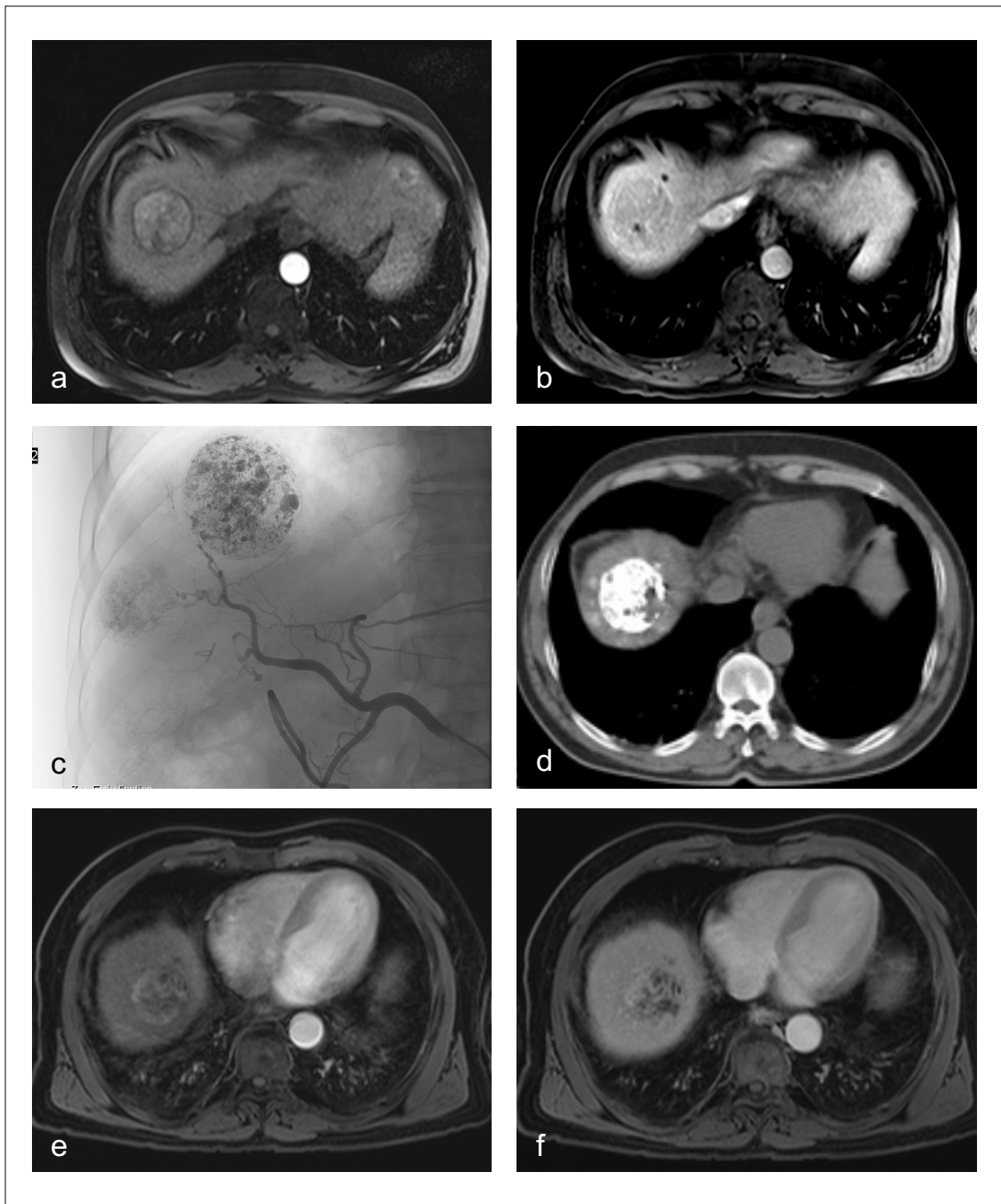


Abbildung 24. Repräsentativer Therapieverlauf; Präinterventionelle MRT in arterieller (a) und portal-venöser (b) Phase, primäre TACE-Prozedur (c), nativ CT 24 nach TACE-Prozedur (d), Verlaufskontrolle nach dritter TACE-Prozedur mittels MRT in arterieller (e) und portal-venöser (f) Phase.⁸²

2.3 Transfer und Aufbereitung der Bilddaten

Der Bilddatensatz, der 24 Stunden nach der ersten TACE-Prozedur erhobenen nativen CT-Untersuchung, wird im DICOM (Digital Imaging and Communications in Medicine) Format exportiert. Der exportierte Datensatz wird in die 3D Slicer Software Plattform (<http://slicer.org>, Version 4.9.0) importiert, mit dem Ziel die Daten mit Hilfe der 3D Slicer Software zu visualisieren und weiterzuverarbeiten.^{83,84}

Vor der Extraktion der Bildmerkmale (Features) wird der importierte Datensatz in eine Schichtdicke von 1mm x 1mm x 1mm umberechnet. Die Imaging Biomarker Standardization Initiative (IBSI) gibt aktuell keine eindeutigen Kriterien der weiterführenden Bildnachbearbeitung vor, entsprechend wurden keine weiteren Bildmanipulationen vorgenommen.^{85,86}

Die IBSI versteht sich als internationale unabhängige Kooperationsinitiative mit dem Ziel, der forschenden Gemeinde einen Leitfaden zur standardisierten Erhebung und Prozessierung von Bilddaten an die Hand zu geben.⁸⁵

2.4 Segmentierung der Bilddatensätze

Der verblindete Doktorand (OE, Facharzt für Radiologie, 10 Jahre Berufserfahrung) definiert pro Studienpatient maximal 2 TLs anhand der präinterventionellen Bildgebung (arterielle Phase CT oder MRT), basierend auf mRECIST Kriterien. Die TLs werden durch den Doktoranden im exportierten CT-Bilddatensatz segmentiert. Dabei wird ein 3-Dimensionales Volume of Interest (VOI) innerhalb der zu interessierenden HCC-Läsion eingezeichnet. Eindeutige Nekroseareale werden hierbei ausgespart. Um eine vollständige Segmentierung des VOI zu erreichen wird der semiautomatische „grow from seeds“ Algorithmus der 3-D Slicer Software verwendet. Der Algorithmus soll hierbei eine Variabilität zwischen verschiedenen Messungen verringern.^{83,87,88}

Die Semiautomatische Segmentierung wird vom Untersucher auf eindeutige Fehler, z.B. Segmentierung von angrenzendem extrahepatischem Gewebe oder eindeutigem Übertritt der Segmentierung über den Rand der TL, untersucht. Diese werden händisch mit dem Brush-erase Werkzeug angepasst (siehe Abbildung 25).

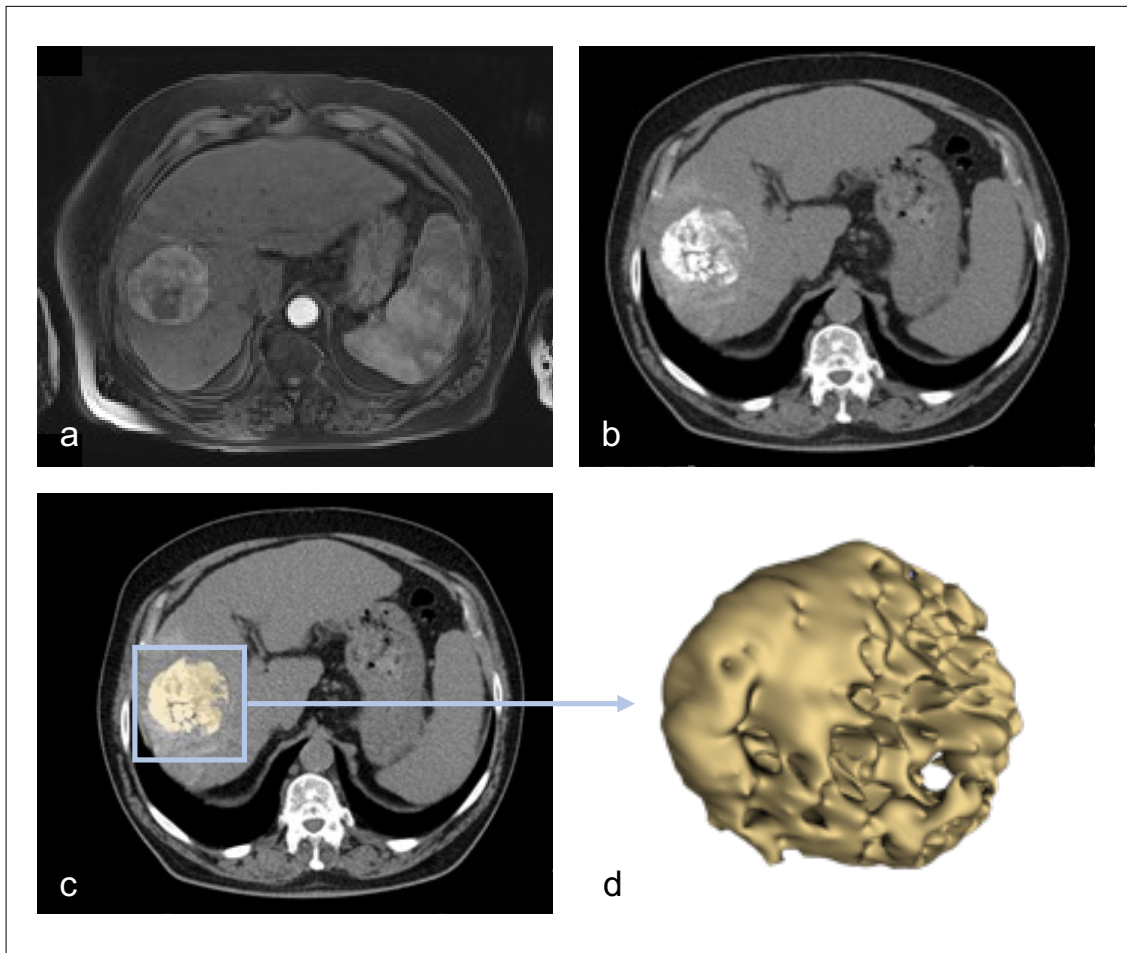


Abbildung 25. Workflow einer repräsentativen Segmentierung; Präinterventionelle MRT in arterieller Phase mit flau anreicherndem HCC-Herd (a), 24 Stunden post-TACE CT nativ mit deutlicher Lipiodol Anreicherung (b), semiautomatische Segmentierung der Lipiodolanreicherung (c), 3D-Modell des segmentierten Herdes (d), nach Bernatz, Elenberger et al.⁸²

Das gleiche Prozedere der Segmentierung wird, unabhängig vom ersten Untersucher, von einem weiteren Arzt durchgeführt (Weiterbildungsassistent der Radiologie, 3,5 Jahre Berufserfahrung).

2.5 Radiomics Analyse (engineered radiomics features)

Um aus den Bilddatensätzen die Radiomics Features zu extrahieren installiert der Doktorand die Open-Source Erweiterung PyRadiomics innerhalb der 3D Slicer Software Plattform.^{63,83}

Mit den voreingestellten Standardeinstellungen extrahiert der Doktorand alle Standard Features (n = 107, Feature Klassen = 7): Shape, First Order Statistics, Second order Statistics (Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Gray Level Dependence Matrix (GLDM), Neighbouring Gray Tone Difference Matrix (NGTDM)).^{89,90}

Die Standardeinstellungen werden gewählt um eine Reproduzierbarkeit durch Dritte zu erhöhen, da eine Vielzahl an Studien die obige open source Software nutzt und mehrheitlich die Standardeinstellungen beibehalten werden.^{89,91} Eine Aufstellung der 7 Featureklassen findet sich als ausführliche Erklärung im Handbuch von PyRadiomics. Eine detaillierte Erläuterung würde außerhalb des Rahmen der aktuellen Dissertation liegen.⁹⁰ In Kürze, unter Shape versteht man die räumliche Darstellung des VOI ohne Bezug zu den Grauwerten. Unter First Order Statistics versteht man ein Histogramm, eine Häufigkeitsverteilung von Grauwerten innerhalb des VOI ohne Bezug der Grauwerte bzw. deren räumliche Verteilung untereinander. Die verbleibenden Featureklassen (GLCM, GLRLM, GLSZM, GLDM und NGTDM) gehören zu den Higher Order Statistics, hier werden die Grauwerte im Raum in Bezug zueinander gebracht und Abhängigkeiten dargestellt.^{69,90}

Siehe hierzu Einleitung Punkt 1.8 Künstliche Intelligenz / Radiomics.

Das vorliegende Studiendesign wird zur Erhöhung der Transparenz und zur nachhaltigen wissenschaftlichen Arbeit anhand des empfohlenen Beurteilungskatalogs für Radiomics-Studien entsprechend des radiomics quality scores (RQS) ausgewertet. Das Studiendesign wird anhand des RQS mit 14 von 36 möglichen Punkten bewertet (<https://radiomics.world/rqs>), Auswertung siehe Anhang A1. Als Vergleich soll hier eine Studie dienen, die den RQS von 189 Radiomics-Studien untersucht, der Median liegt hier bei 11 Punkten, mit einem Maximum von 16 Punkten.⁹² Die vorliegende Studie reiht sich somit in die aktuelle Forschungslandschaft ein.

Wie auch das bereits erwähnte IBSI versteht sich RQS als eine Leitlinie um die Güte von Datenerhebung, Bildgebung, Feature Extraktion, Analyse und Auswertung zu quantifizieren. So soll zum einen für den Studienleiter, als auch für den Außenstehenden, die Wertigkeit der Studie anhand der vorgegebenen Parameter nachvollziehbar sein.⁶²

2.6 Interrater-Reliabilität und Featureredundanz

Um die Reproduzierbarkeit der Segmentierungen zu erfassen, wird für jedes Feature der Intraklassen-Korrelationskoeffizient (ICC) berechnet.⁸⁹ Der ICC wird mit Hilfe des Pingouin-Package in Python berechnet (ICC3).⁹³

Die Grenzwerte werden hierbei entsprechend Baeßler et al.⁹⁴ definiert: ICC 0,75 – 1: exzellent, ICC 0,60 – 0,74: gut, ICC 0,40 – 0,59: moderat, ICC \leq 0,39: schwach.⁹⁴ Diese Werte sind Grenzwerte, die häufig bei Radiomics-Analysen genutzt werden.

Der mediane Intraklassen-Korrelationskoeffizient ICC liegt für das gesamte Set aller Featureklassen bei einem Wert von 0,9, und reicht hierbei von 0,76 (\pm 0,41, Neighbouring Gray Tone Difference Matrix (NGTDM)) bis 0,98 (\pm 0,03, First Order Statistics), siehe Abbildung 26 und 27.

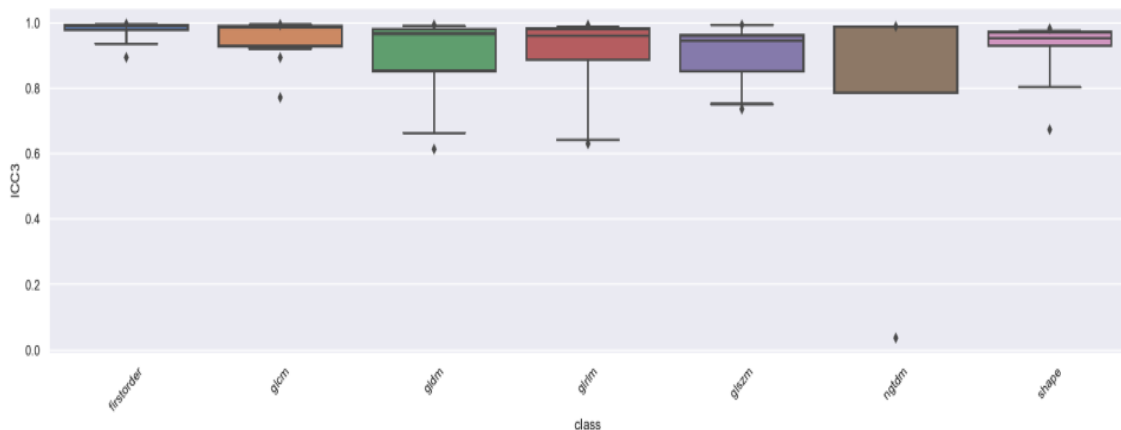


Abbildung 26. Darstellung der Intraklassen-Korrelationskoeffizienten (ICC) der 7 Featureklassen⁸²

Featureklasse	ICC	
	Median	SD
First Order Statistics	0,978	0,027
Gray Level Co-occurrence Matrix (GLCM)	0,958	0,051
Gray Level Dependence Matrix (GLDM)	0,898	0,127
Gray Level Run Length Matrix (GLRLM)	0,897	0,129
Gray Level Size Zone Matrix (GLSZM)	0,905	0,093
Neighbouring Gray Tone Difference Matrix (NGTDM)	0,758	0,413
Shape	0,925	0,086

Abbildung 27. ICC für jede einzelne Featureklasse (SD = Standardabweichung), Aufstellung zur Abbildung 27, nach Bernatz, Elenberger et al.⁸²

Von allen Features zeigen insgesamt 8 einen ICC $<0,75$ (Range 0,04 – 0,74), diese Features werden für die weitere Betrachtung verworfen, sodass nur noch Features mit einer als exzellent definierten Reproduzierbarkeit übrigbleiben ($n = 99$). Ein Swarmplot der ICC sowie die gesamte Tabelle der ICC-Werte für alle einzelnen Features sind im Anhang (Anhang A2 und A3) abgebildet.

Die verbleibenden Features werden nach der Pearson Methode korreliert, mit dem Ziel alle sich stark überschneidenden und somit redundante Features aus dem Set zu eliminieren (Pearson $> 0,95$). Insgesamt werden auf diese Weise weitere 52 Features eliminiert, siehe Abbildung 28.

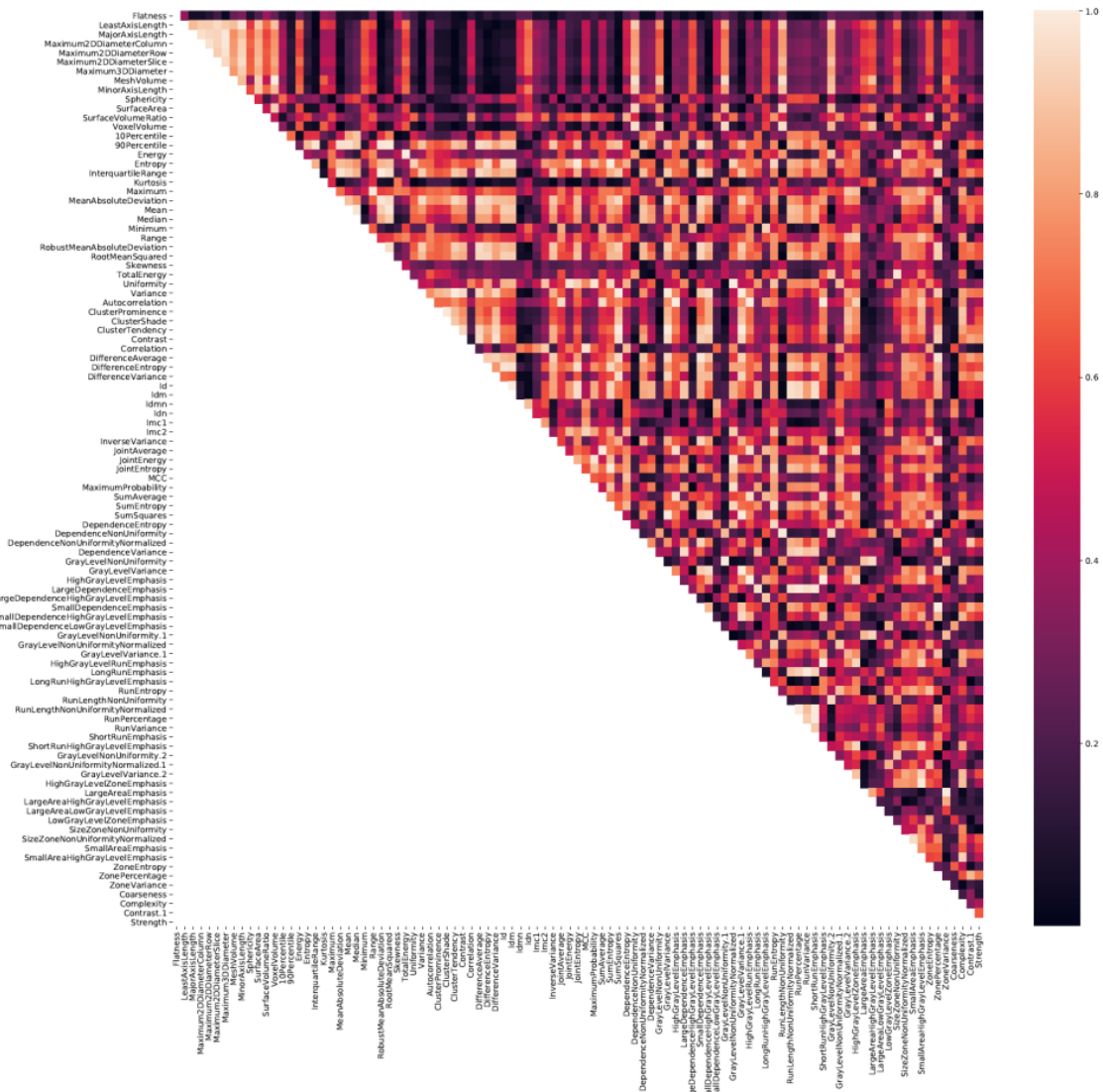


Abbildung 28. Heatmap der Pearsonkorrelationsanalyse mit den reproduzierbaren (ICC ≥ 0,75) Radiomics Features, redundante Features stellen sich hier in der Legende hell dar mit Werten ≥ 0,95

Das verbleibende Featureset aus 47 Features beinhaltet die als nichtredundant und reproduzierbar definierte Features. Dieses Set stellte das finale Set dar mit welchem alle weiteren Modelle dieser Studie entwickelt werden.

Aus Gründen der Übersichtlichkeit befindet sich die Tabelle mit den ausgeschiedenen und verbleibenden Features im Anhang (Anhang A4).

2.7 Berechnung der klinischen Scores

Insgesamt fünf klinische Scores zur Beurteilung des Ansprechens nach TACE oder Risikostratifizierung, werden anhand der vorliegenden Bilddaten und Laborwerte durch den Doktoranden berechnet und evaluiert⁵²:

- (I) Albumin-Bilirubin-Score (ALBI-Score)⁴⁸
- (II) Hepatoma arterial embolization prognostic score (HAP-Score)⁵¹
- (III) modified HAP-II-score (mHAP-II-Score)⁵³
- (IV) Selection for TrAnsarterial chemoembolisation TrEatment (STATE-Score)⁵⁷
- (V) 6 and 12-Score⁵⁸

Klinischer Score	Gesamt	Train-Set	Test-Set
Anzahl Patienten	61	42	19
ALBI-Score (%)			
<i>I</i>	28 (45,9)	21 (50,0)	7 (36,8)
<i>2</i>	30 (49,2)	19 (45,2)	11 (57,9)
<i>3</i>	2 (3,3)	2 (4,8)	0 (0,0)
<i>NA</i>	1 (1,6)	0 (0,0)	1 (5,3)
HAP-Score Gruppe (%)			
<i>A</i>	30 (49,2)	21 (50,0)	9 (47,4)
<i>B</i>	20 (32,8)	14 (33,3)	6 (31,6)
<i>C</i>	9 (14,8)	6 (14,3)	3 (15,8)
<i>D</i>	1 (1,6)	1 (2,4)	0 (0,0)
<i>NA</i>	1 (1,6)	0 (0,0)	1 (5,3)
mHAP-II-Score Gruppe (%)			
<i>A</i>	11 (18,0)	7 (16,7)	4 (21,1)
<i>B</i>	30 (49,2)	21 (50,0)	9 (47,4)
<i>C</i>	11 (18,0)	9 (21,4)	2 (10,5)
<i>D</i>	8 (13,1)	5 (11,9)	3 (15,8)
<i>NA</i>	1 (1,6)	0 (0,0)	1 (5,3)
State-Score-Gruppe (%)			
<i>low risk</i>	7 (11,5)	6 (14,3)	1 (5,3)
<i>high risk</i>	53 (86,9)	36 (85,7)	17 (89,5)
<i>NA</i>	1 (1,6)	0 (0,0)	1 (5,3)
6 and 12 Gruppe (%)			
<i>1</i>	24 (39,3)	18 (42,9)	6 (31,6)
<i>2</i>	25 (41,0)	16 (38,1)	9 (47,4)
<i>3</i>	12 (19,7)	8 (19,0)	4 (21,1)

Tabelle 2: Klinische Scores, Verteilung für das gesamte Kollektiv, sowie für Train- und Test-Set

Als nächstes wird die arterielle Hypervaskularisation der TL von drei unabhängigen Bewertern (Doktorand (OE), SB, SM) beurteilt. Die visuelle Beurteilung der Kontrastmittel-Anreicherung wird auf einer Likert-Skala mit einem Wert von 1 (minimale Anreicherung) bis 5 (starke Anreicherung) erfasst, da jede Läsion eine arterielle Anreicherung zeigt. Die dreifache Auswertung wird bezüglich der Rater-Variabilität untersucht. Der ICC zwischen den 3 Bewertern liegt bei 0,86, und ist somit als exzellent zu bewerten.

Der vom Bewerter 1 (OE, Doktorand) erfasste visuelle Score wird für den Algorithmus für maschinelles Lernen angewendet. Für die Möglichkeit einer umfassenden Analyse, bestehend aus Radiomics und bekannten klinischen Scores, werden die klinischen Scores und die visuellen Scores der arteriellen Mehranreicherung als klinische Benchmark definiert.

2.8 Feature Auswahl und Entwicklung des Berechnungsmodells (supervised learning)

Als erster Schritt werden die Daten zu einer uniformen Varianz skaliert (StandardScaler). Dabei werden die Werte der zu untersuchenden Features auf ein $\text{mean} = x$ mit einer Standardabweichung von 1 gebracht. Das Ziel dabei ist es unterschiedlich skalierte Daten, das heißt Daten mit einer hohen Werteverianz, zu normalisieren und somit vergleichbar zu machen. Dieses wird durch das Etablieren einer gleichen Wertigkeit und gleicher Breite der Werte erreicht. Die Verteilungsform der Werte wird durch die Skalierung nicht beeinflusst.

Hiernach wird das Patientenkollektiv mittels GroupShuffleSplit (<http://scikit-learn.org>) in ein Train-Set (70%) und Test-Set (30%) getrennt. Eine Mischung der Patienten über die Grenzen der Sets ist ab diesem Zeitpunkt nicht mehr möglich. Das Train-Set wird für die Erstellung der späteren Modelle genutzt und das unabhängige Test-Set wird genutzt um die Modelle nach deren Etablierung in zuvor nicht gesehenen Daten bezüglich ihrer Performance zu testen. Dieser Ansatz stellt das typische Vorgehen bei Studien des maschinellen Lernens dar.

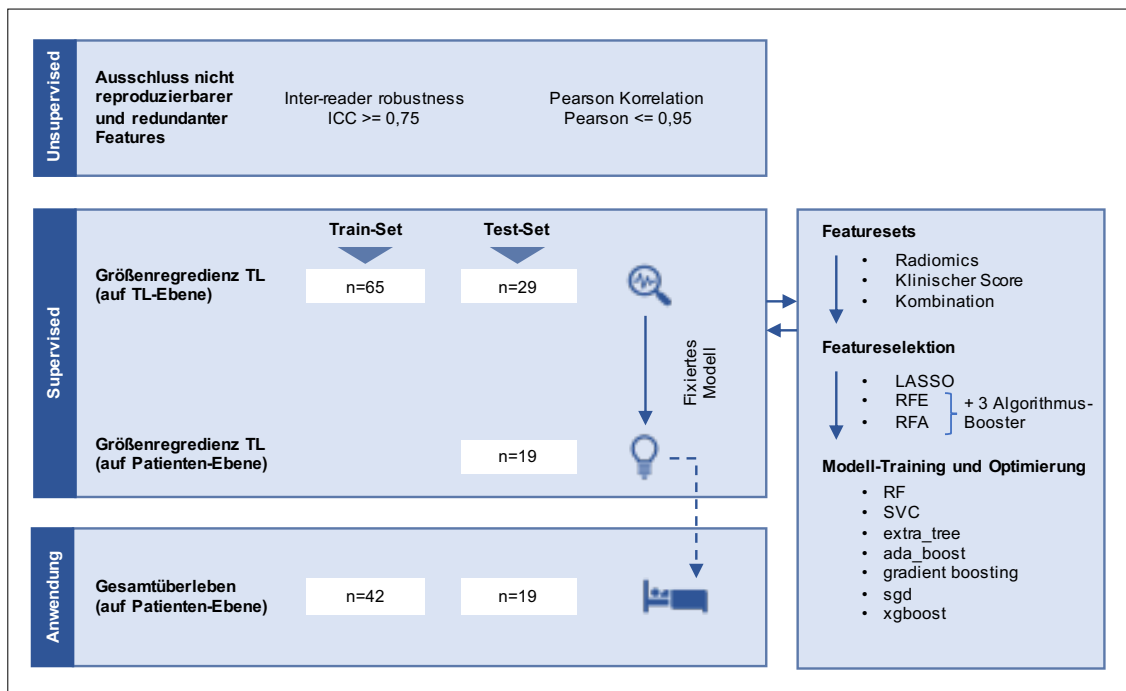


Abbildung 29. Workflow Entwicklung des Modells, nach Bernatz, Elenberger et al.⁸²

Die Zielvariable auf Ebene der TL, Größenregredienz als Folge eines annehmbaren Therapieansprechens, wird zunächst für jede einzelne TL separat untersucht. Hierzu werden die vorhandenen Features des Train-Sets mittels 7 verschiedener Selektionsalgorithmen untersucht und eingegrenzt (siehe Abbildung 29).

Dies erfolgt mittels der Rekursiven Feature Elimination (RFE) und Rekursiven Feature Addition (RFA). RFE und RFA sind Selektionsalgorithmen mit Hilfe derer und der vorgegebenen Zielvariable (supervised learning) aus dem Set von verbleibenden (n = 47) Features die relevantesten selektiert werden.

Im Rahmen der Anwendung des RFE-Algorithmus wird in mehreren Schritten die Gruppe der Features immer wieder verkleinert, indem die am wenigstens mit der Zielvariable korrelierenden Features aus dem Set ausgeschlossen werden. Das auf diese Weise verkleinerte Featureset durchläuft weitere Male den Algorithmus, bis ein vom Algorithmus als optimal definiertes Set an Features verbleibt.

Bei der Addition (RFA) werden alle Features mit der Zielvariable korreliert. Dabei werden die Features nach ihrer Korrelation mit der Zielvariable gerankt. Die Zielvariable wird in Kombination mit dem höchstbewerteten Feature analysiert. Dem Set wird nun das nächst bestbewertete Feature addiert, und die so entstehende Gruppe

aus Features wird erneut mit der Zielvariable korreliert. Steigt die Korrelation des so entstandenen Featuresets, so wird das nächste Feature in die Auswahl aufgenommen. Der Algorithmus wird so lange ausgeführt bis alle Features analysiert wurden und ein entsprechendes Featureset definiert ist.

Die Definition der Wertigkeit eines Features kann durch verschiedene Algorithmen beeinflusst werden. Wir verwenden jeweils 3 verschiedene Algorithmusbooster sowohl im Rahmen der RFE als auch RFA. Im Einzelnen sind dies: Gradient Boosting Decision Tree Model (GBDT), Dropouts meet multiple Additive Regression Trees (DART) und Gradient based One Side Sampling (GOSS). Vereinfacht dargestellt handelt es sich hier um verschiedene Bewertungsmodelle der Features im Rahmen der RFE und RFA, die zu jeweils drei unterschiedlichen Ergebnissen der Featureselektion sowohl bei RFE und RFA führen.

Als siebte Featureselektionsmethode wenden wir LASSO (least absolute shrinkage and selection operator) an. Dabei werden weniger relevante oder irrelevante unabhängige Features mittels einer Schrumpfung verkleinert. Dabei können die unabhängigen Variablen einen Wert von Null einnehmen. So soll der Einfluss von weniger wichtigen unabhängigen Variablen auf die abhängige Variable reduziert werden. In unserem Set verbleiben alle Features mit einem Wert >0 .

Die Featureselektion erfolgt nach dem beschriebenen Vorgehen für drei Gruppen: Radiomics-Gruppe, Klinischer Score und einer Kombination aus klinischem Score und Radiomics.

Nach dem Durchgang aller Featureselektionsmodelle erhalten wir insgesamt 7 Featuresets (RFE-GBDT, RFE-DART, RFE-GOSS, RFA-GBDT, RFA-DART, RFA-GOSS und LASSO) pro Gruppe.

Die so erhaltenen 7 Featuresets werden weiter selektiert, indem aus dem RFE und RFA-Set nur die besten Featuresets beibehalten wurden, gemessen an der Accuracy. Unter Accuracy versteht man den Anteil der korrekten Vorhersagen durch unser Modell.

So bleiben für jede Gruppe insgesamt 3 Featuresets, jeweils ein RFE- und RFA-Modell, sowie das mit LASSO definierte Set. Im Falle von gleichwertiger Accuracy werden Sets mit der kleineren Anzahl an notwendigen Features für diese Performance der Vorzug

gegeben. Beim klinischen Score sind sowohl das Beste RFE- als auch das RFA-Modell identisch.

Das Featureset der kombinierten Gruppe ist nach der automatischen Featureselektion identisch zum Featureset der Radiomics-Gruppe. Aus diesem Grund wird das Featuresets für die kombinierte Gruppe händisch definiert. Hierzu wird eine Kombination aus den besten Features der Radiomics-Gruppe und den besten Features der Gruppe mit dem klinischen Score gebildet. So bleiben letztendlich 6 Featuresets zur weiteren Verarbeitung (siehe Abbildung 30).

Featureselektion	ausgewählte Features
Radiomics	
RFE	<ul style="list-style-type: none"> • Flatness • Sphericity • 10Percentile • Maximum • Skewness • Imc1 • LargeAreaEmphasis • LargeAreaLowGrayLevelEmphasis • LargeDependenceHighGrayLevelEmphasis
RFA	<ul style="list-style-type: none"> • LargeDependenceHighGrayLevelEmphasis
LASSO	<ul style="list-style-type: none"> • Flatness • Minimum • Skewness • LargeAreaLowGrayLevelEmphasis
Klinischer Score	
RFE & RFA	<ul style="list-style-type: none"> • mHAP_II_score_group
LASSO	<ul style="list-style-type: none"> • mHAP_II_score_group • 6_and_12_group
Kombinierte Features (Radiomics und klinisch)	
Beste Kombination	<ul style="list-style-type: none"> • LargeDependenceHighGrayLevelEmphasis • mHAP_II_score_group

Abbildung 30. Definition der besten Features der unterschiedlichen Gruppen abhängig vom Featureselektionsalgorithmus

Jedes der 6 definierten Featuresets wird im nächsten Schritt genutzt um mit Hilfe von 7 verschiedenen Klassifikatoren das beste Vorhersagemodell zu identifizieren. Hierzu wird Hyperopt, eine Python basierte Bibliothek zur automatisierten Hyperparameteroptimierung, verwendet.^{95,96}

Hyperopt umfasst für unsere Analyse die folgenden 7 Klassifikatoren: (1) random forest, (2) support vector classifier, (3) extra-trees classifier, (4) AdaBoost classifier, (5) gradient boosting classifier, (6) stochastic gradient classifier und (7) xgboost classifier. Jeder Klassifikator ist dabei ein differenter Algorithmus mit dem jedes einzelne Featureset mehrmals getestet wird. Der Startpunkt des Algorithmus sowie der Algorithmus selbst (entspricht einem der sieben Klassifikatoren) wird bei jedem Durchlauf zufällig ausgewählt. Hierdurch werden jeweils unterschiedliche Ergebnisse generiert, die im Anschluss nach ihrer Wertigkeit gerankt werden.

Im vorliegenden Setting durchläuft jedes Featureset insgesamt 50-mal den Optimierungszyklus. Das Vorgehen wird sowohl auf das Train-Set als auch auf das Test-Set angewendet.

Das beste Arbeitsmodell ist:

RandomForestClassifier (bootstrap=False, criterion='entropy', max_features=None, min_samples_leaf=2, n_estimators=165, n_jobs=1, random_state=3, verbose=False)

Die nachfolgende Tabelle 3 zeigt einen Überblick über die Ergebnisse.

Klassifikator	Featureselektion	Train-Set		Test-Set	
		Accuracy	AUC	Accuracy	AUC
Radiomics					
GradientBoostingClassifier	RFE	1,000	1,000	0,517	0,596
ExtraTreesClassifier	RFA	0,738	0,881	0,586	0,674
AdaBoostClassifier	LASSO	1,000	1,000	0,586	0,551
Klinischer Score					
SVC	RFE & RFA	0,615	0,606	0,621	0,689
SGDClassifier	LASSO	0,615	0,618	0,621	0,649
Kombinierte Features					
RandomForestClassifier	Beste Kombination	0,862	0,957	0,724	0,705

Tabelle 3. Ergebnisse nach ML mit den besten Modellen auf Ebene der Target Lesions, das kombinierte Modell im blauen Kasten hervorgehoben

Gemessen an der AUC erweist sich das kombinierte Modell (kombiniertes Modell AUC: 0,705, klinischer Score: 0,689, Radiomics: 0,674) als das Beste in Bezug auf die Prädiktion eines Tumoransprechens auf TL-Ebene (siehe Tabelle 3). Das entsprechende Modell mit seinen Parametern wird für die weitere Analyse fixiert. Im nächsten Schritt wird das TACE-Ansprechen auf Patientenebene analysiert.

In der Folge wird die Zielvariable von ‚individuelles Tumoransprechen‘ auf ‚Gesamtüberleben‘ geändert. Hierzu nutzen wir die random survival forest-Analyse, unter Zuhilfenahme des Python Paket Scikit-survival 0.16.1.⁹⁷ So wird die Performance des definierten Featuresets auf die Vorhersage des Gesamtüberlebens geprüft. Die Performance wird mit dem C-Index bewertet, um eine Prüfung der Ausgangshypothese vorzunehmen.

3. Ergebnisse

3.1 Analyse des Therapieansprechens

Die vorliegende Studienpopulation besteht aus 61 Patienten (Median Alter 65,3 Jahre \pm 10 (Standardabweichung), Range 37 – 86 Jahre, 12 Frauen/49 Männer) mit insgesamt 94 TL.

Nach mRECIST Kriterien haben 38,3% (36/94) der TL und 40,1% (25/61) der Patienten ein Ansprechen auf die TACE gezeigt (siehe Tabelle 4).

Merkmal	Gesamt	Train-Set	Test-Set
Anzahl TL	94	65	29
mRECIST Ansprechen der TL			
CR	13 (13,8)	9 (13,8)	4 (13,8)
PR	23 (24,5)	16 (24,6)	7 (24,1)
SD	48 (51,1)	32 (49,2)	16 (55,2)
PD	10 (10,6)	8 (12,3)	2 (6,9)
Anzahl Patienten	61	42	19
mRECIST Gesamtansprechen			
CR	5 (8,2)	4 (9,5)	1 (5,3)
PR	20 (32,8)	13 (30,9)	7 (36,8)
SD	26 (42,6)	18 (42,9)	8 (42,1)
PD	10 (16,4)	7 (16,7)	3 (15,8)

Tabelle 4. Gesamtansprechen auf Ebene der TL und Gesamtansprechen auf Patientenebene nach insgesamt 3 TACE-Behandlungen, nach mRECIST

Per Zufallsverteilung werden 70% der Studienpopulation ($n = 42$, Median Alter 66,1 Jahre \pm 10,3 (Standardabweichung), Range 37 – 86 Jahre, 7 Frauen/35 Männer) mit insgesamt 65 TL dem Train-Set zugewiesen. Die restlichen 30% der Studienpopulation ($n = 19$, Median Alter 63,5 Jahre \pm 9,2 (Standardabweichung), Range 50 – 86, 5 Frauen/14 Männer) werden der unabhängigen Testgruppe mit insgesamt 29 TL zugewiesen. Diese Aufteilung wird für alle Analysen konstant gehalten.

Nach der dritten TACE-Prozedur zeigt sich ein Therapieansprechen auf Ebene der einzelnen TL bei 38,5% der TL im Train-Set (9 CR + 16 PR = 25/65) und bei 37,9% der TL im Test-Set (4 CR + 7 PR = 11/29).

Bezogen auf die Patientenebene zeigte sich bei 40,5% der Patienten im Train-Set (4 CR + 13 PR = 17/42) und 42,1% der Patienten im Test-Set (1 CR + 7 PR = 8/19) ein Therapieansprechen in der Kontrolluntersuchung nach der dritten TACE.

Betrachtet man die gesamte Studienkohorte, so zeigen 38,3% der TL (36/94) und 41% der Patienten (25/61) ein Therapieansprechen.

Die Bewertung des Gesamtansprechens der Therapie erfolgt entsprechend der mRECIST-Kriterien (siehe Abbildung 12) und schließt neben TL auch NTL, sowie auch eindeutige neu aufgetretene tumoröse Läsionen, mit ein. In der vorliegenden Studie werden maximal zwei intrahepatische Referenzläsionen als TL definiert. Weitere, typische und atypische, intrahepatische Läsionen, als auch eindeutig maligne extrahepatische Läsionen und Lymphknoten werden als non-target-lesions (NTL) definiert.

3.2 Explorative Datenanalyse mittels Dimensionalitätsreduktion

Bevor die hypothesengetriebene Datenanalyse durchgeführt wird, wird zunächst eine explorative Datenanalyse betrieben. Die Daten werden ohne Vorgabe der Zielvariable untersucht (unsupervised learning). Dabei sollen im Rahmen der Analyse Zusammenhänge oder Abhängigkeiten innerhalb scheinbar ungleicher Daten aufgedeckt werden. Mit Hilfe eines t-SNE Plots (t-dispergierte stochastische Nachbarschaftsimplantation) werden hierzu die Daten in ihrer Dimensionalität vermindert. Der Pool von den 47 verbliebenen Features wird so auf zwei Featuresets, unter Datenverlust, verringert. Mit der Darstellung sollen etwaige Abhängigkeiten der Daten zueinander visualisiert und somit erkennbar werden. Die Visualisierung der vorliegenden Daten (siehe Abb. 31) zeigt keine Clusterbildung, weshalb von einem Nichtvorhandensein direkter Abhängigkeiten auszugehen ist und im nächsten Schritt mit der Analyse auf Basis komplexerer Algorithmen fortgeföhren werden kann.

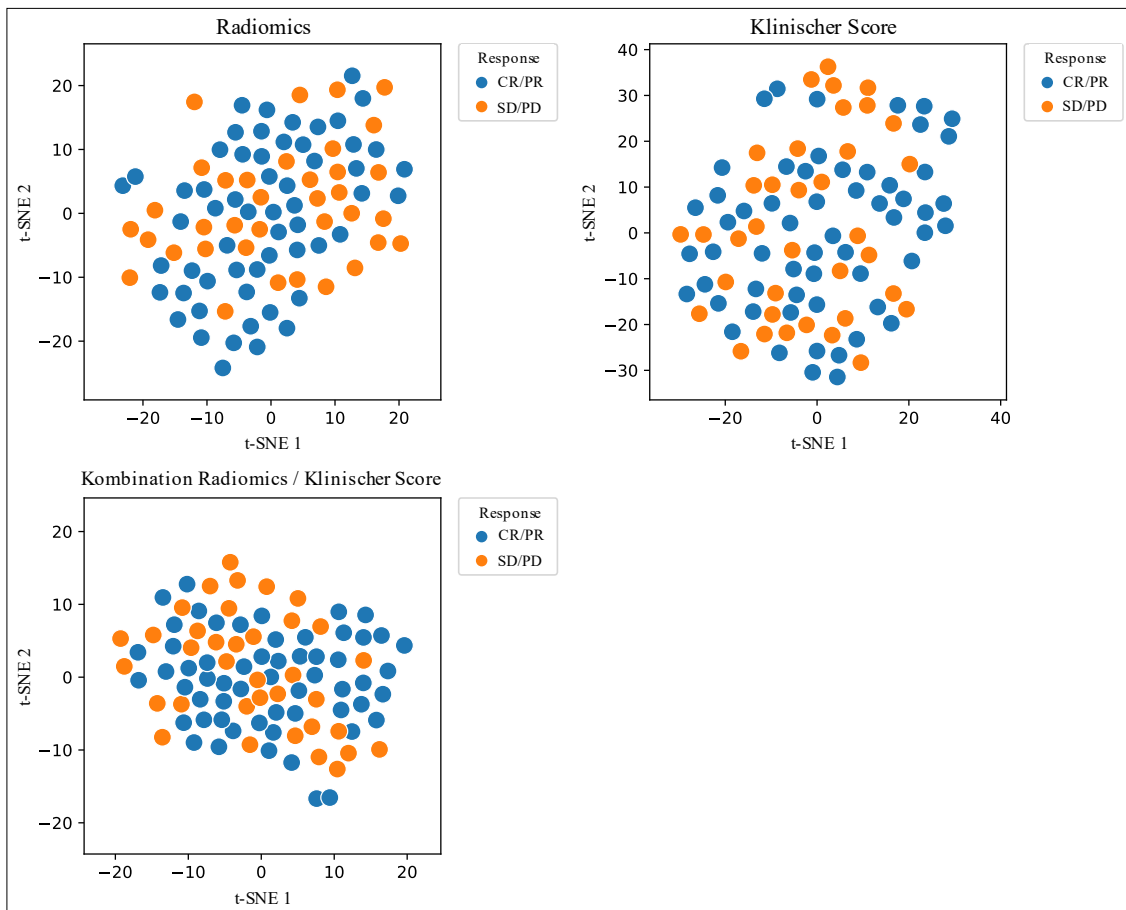


Abbildung 31. t-SNE Plots ohne erkennbarem abhängigem Verteilungsmuster. Zielpunkte definiert als complete response oder partial response (CR/PR = blau) und stable disease oder progressive disease (SD/PD = orange), nach Bernatz, Elenberger et al.⁸²

3.3 Ansprechverhalten auf Ebene der TL

Um die Fähigkeiten eines ML-Modells auf die Vorhersagbarkeit des Tumoransprechens nach einer TACE-Prozedur aufzuzeigen werden insgesamt 7 verschiedene Featureselektionsmethoden und 7 verschiedene ML-Modelle angewendet (siehe 2.8 Material und Methoden).

Sowohl die Featureselektionsmethoden als auch die ML-Modelle werden hierbei auf die generierten Radiomics-Daten, sowie auf die aus den vorliegenden Daten extrahierten klinischen Daten, und eine Kombination aus Radiomics und klinischen Daten angewendet.

Für jedes Featureset wird AUROC als Kriterium für die Performance bestimmt. Zeigen verschiedene Featuresets die gleichen AUC-Werte, wird das Featureset mit der geringeren Anzahl an Features bevorzugt.

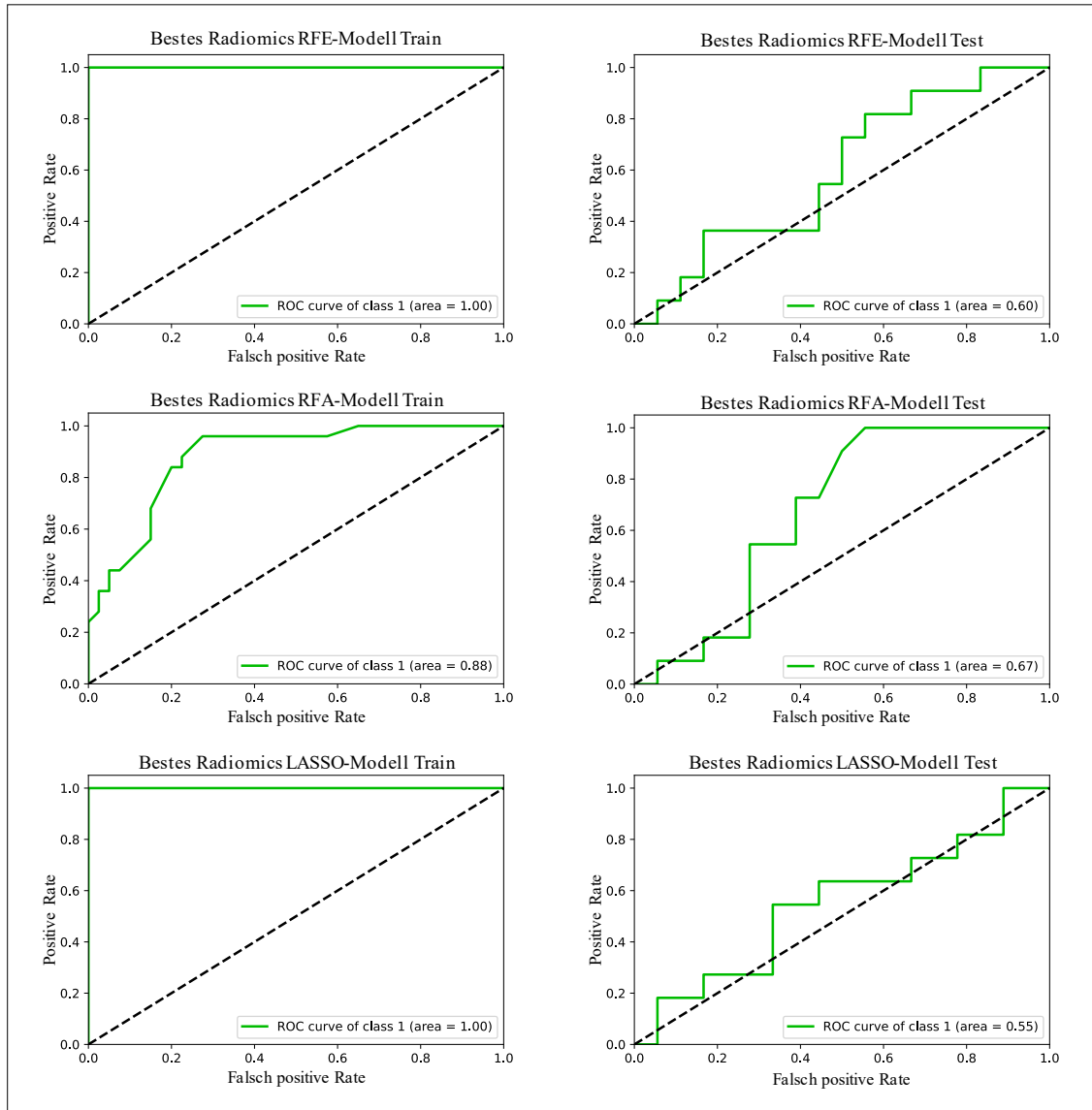


Abbildung 32. ROC-Kurven der Featureselektionsalgorithmen RFE, RFA und LASSO für Radiomics-Features. Das beste Radiomics Modell in der mittleren Zeile.

Das beste Radiomics-Modell (RFA Featureselektion) weist einen Test-AUC von 0,67, bei einem Train-AUC von 0,88, auf (siehe Abbildung 32).

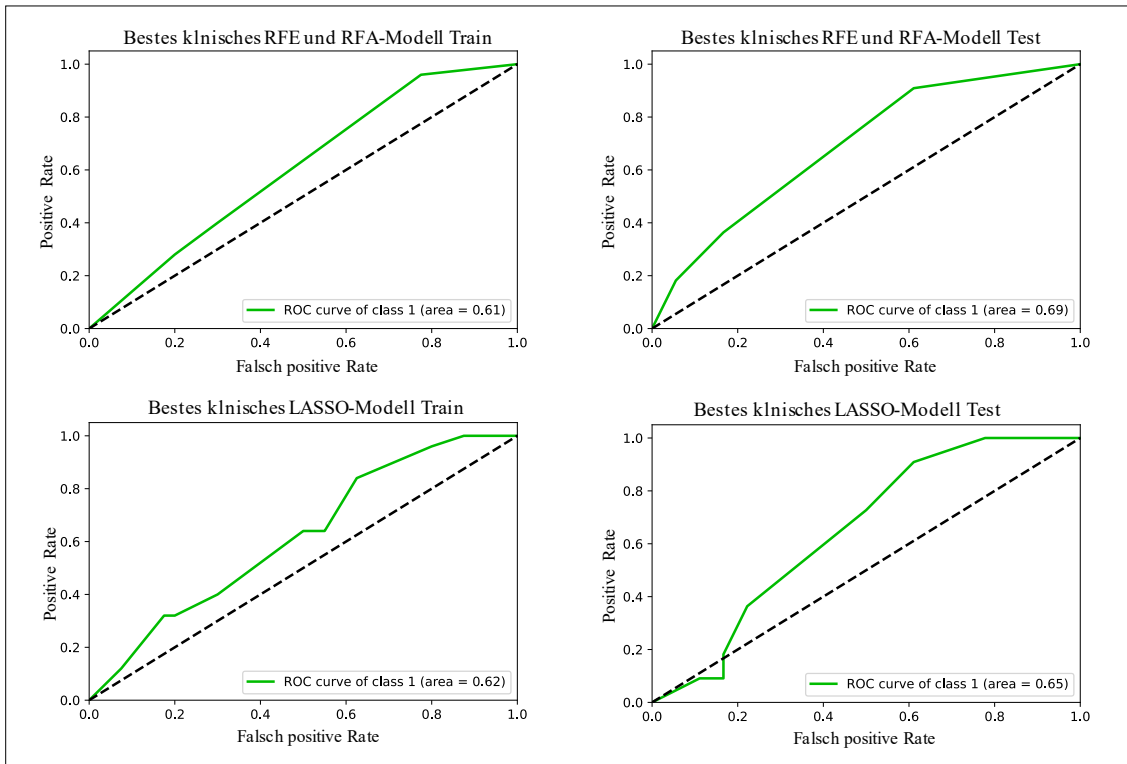


Abbildung 33. ROC-Kurven der Featureselektionsalgorithmen RFE, RFA und LASSO für **klinische Features**. Das beste klinische Modell in der oberen Zeile.

Das beste klinische Modell weist einen Test-AUC von 0,69 auf, dieser fällt besser als der Train-AUC mit 0,61 aus, was einem Bias entspricht.

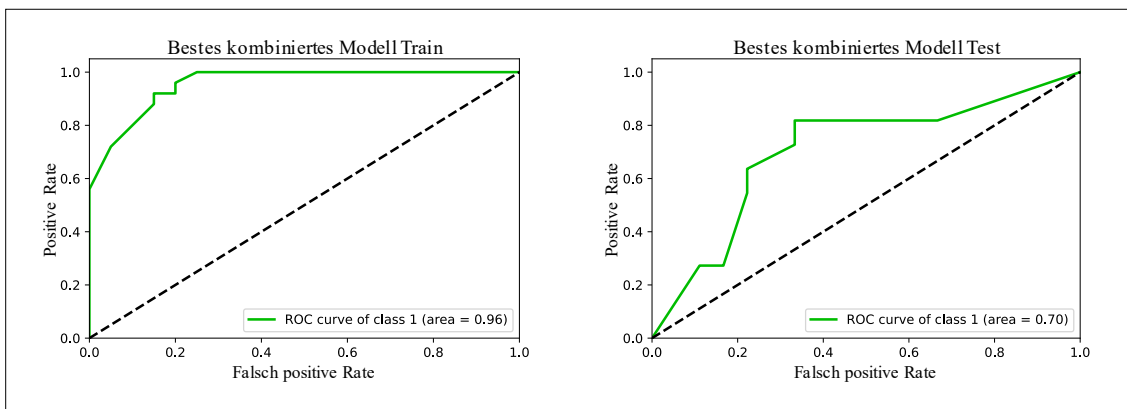


Abbildung 34. ROC-Kurven für das beste **kombinierte Modell** auf Target Lesions Ebene, nach Bernatz, Elenberger et al.⁸²

Das aus Radiomics und dem klinischen Datensatz kombinierte Modell weist, gegenüber dem reinen Radiomics und dem reinen klinischen Score Modell, die beste Performance auf, die Test-AUC liegt hier bei 0,70 und die Train-AUC bei 0,96 (siehe Abbildung 34). Das kombinierte Modell besteht aus dem Feature der Radiomics Gruppe: *LargeDependenceHighGrayLevelEmphasis* und dem Feature aus der klinischen Gruppe: *mHAP_II_score_group*. Der zugrundeliegende ML-Algorithmus ist Random Forest Classifier. Das so definierte Featureset wird für das weitere Vorgehen fixiert und unverändert beibehalten.

3.4 Prognose des Gesamtansprechens auf Patientenebene

Das kombinierte Featureset wird im nächsten Schritt von der individuellen Target Lesion-Ebene auf die nächst höhere Ebene, das Gesamtansprechen des Kollektivs auf die TACE-Therapie angewendet. Das Gesamtansprechen wird nach den mRECIST Kriterien beurteilt, was neben dem Ansprechverhalten der untersuchten TL, non-Target-Lesions sowie auch etwaige neue Läsionen mit einschließt.⁴¹

Bezogen auf das Gesamtansprechen weist das kombinierte Modell eine AUC von 0,77 im Train-Set auf (95% Konfidenzintervall 0,68-0,91). Im Test-Set weist das Modell eine AUC von 0,62 auf (95% Konfidenzintervall 0,41-0,83), siehe Abbildung 35.

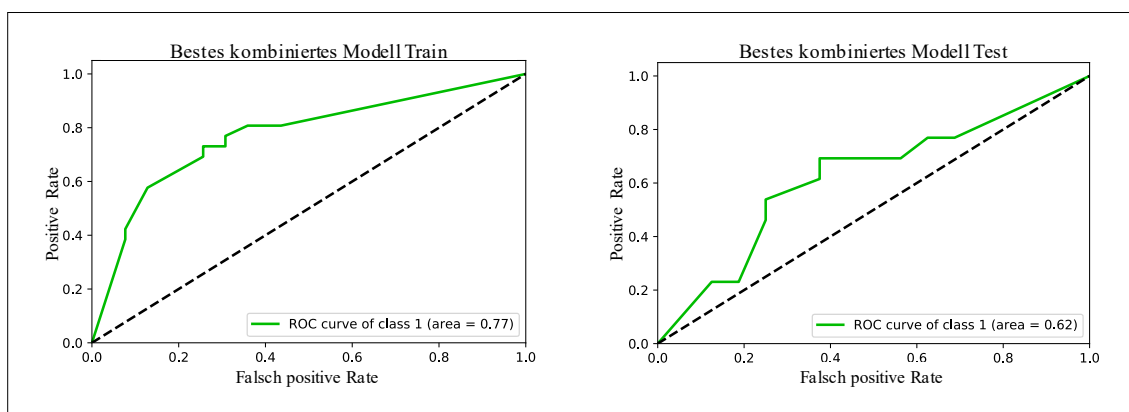


Abbildung 35. ROC-Kurven des kombinierten Featuresets, Performance des Modells bezogen auf das Gesamtansprechen des Kollektivs auf die TACE-Therapie, nach Bernatz, Elenberger et al.⁸²

3.5 Prognose des Gesamtüberlebens

Im nächsten Schritt wird das kombinierte Featureset in Bezug auf die Fähigkeit ein Gesamtüberleben zu prognostizieren beurteilt.

Anders als die Vorhersage des Tumoransprechens auf Patientenebene, das nach mRECIST Kriterien bewertet wird, wird an dieser Stelle kein Surrogat, sondern der C-Index herangezogen und damit eine direkte Aussage zur Prognose des Gesamtüberlebens ermöglicht.

Das kombinierte Set erreicht im Train-Set einen C-Index von 0,71 und im Test-Set einen C-Index von 0,67. Im Vergleich hierzu erreichen die isoliert betrachteten Features einen C-Index von 0,58 für das klinische Feature (*mHAP_II_score_group*) (Train-Set C-Index: 0,70), sowie einen C-Index von 0,60 (Train-Set C-Index: 0,60) für das Radiomics Feature (*LargeDependenceHighGrayLevelEmphasis*). Hierdurch wird die Überlegenheit des kombinierten Modells aus zwei Features gegenüber dem einzelnen besten klinischen, sowie Radiomics Feature gezeigt.

Mit dem Python Paket „lifelines package“ werden Kaplan-Meier-Kurven der gesamten Patientenkohorte, sowie aufgeteilt nach Train- und Test-Set, generiert (siehe Abbildung 36), um das OS der Gruppen zu visualisieren.⁹⁸

Die generierten Kaplan-Meier-Kurven weisen einen vergleichbaren Verlauf auf, was eine Vergleichbarkeit der entsprechenden Kollektive darstellt.

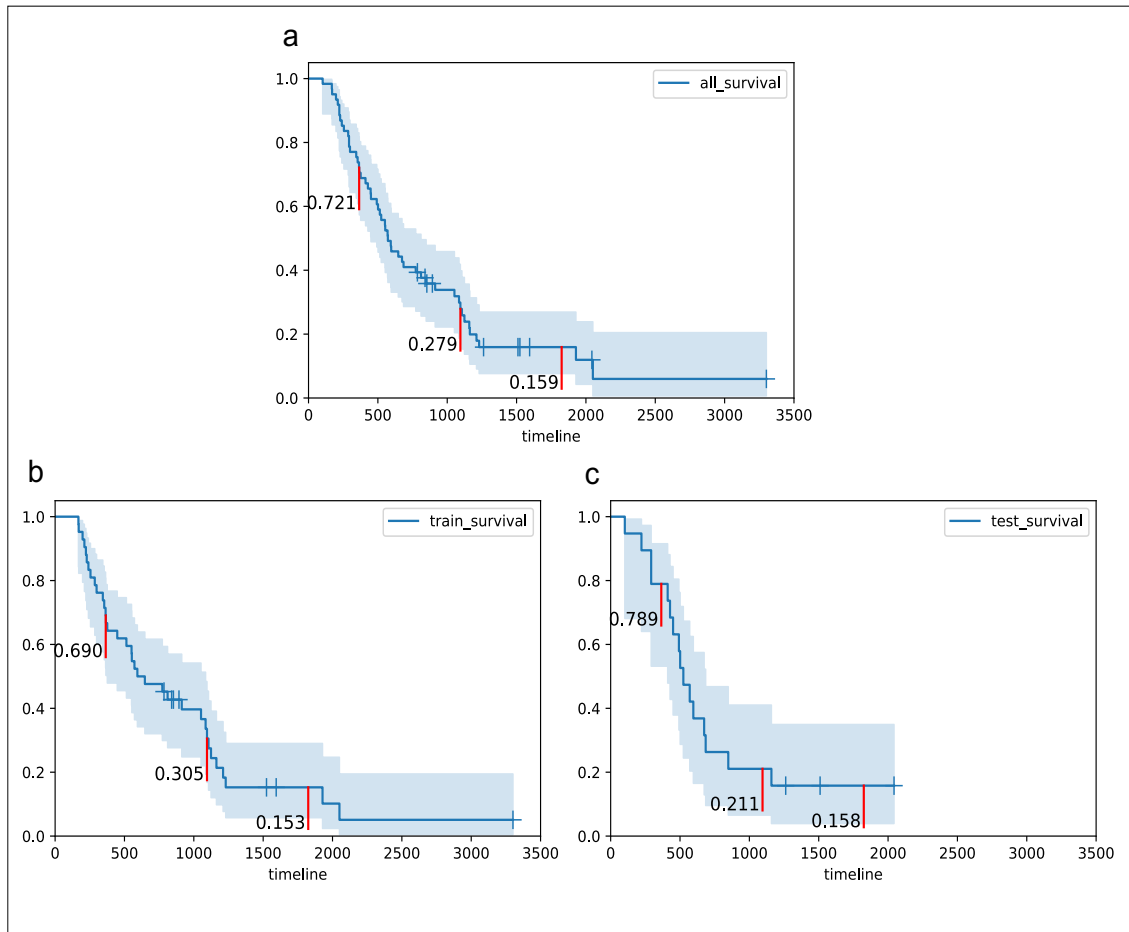


Abbildung 36. Kaplan-Meier OS-Kurven aufgeteilt nach dem gesamten Patientenkollektiv (a), sowie einzeln aufgeteilt in Train-Gruppe (b) und Test-Gruppe (c). Rote Linien zeigen die Anteile der noch lebenden Patienten zu den Zeitpunkten: 1 Jahr, 3 Jahre und 5 Jahre, nach Bernatz, Elenberger et al.⁸²

Als Letztes wird die Fähigkeit des kombinierten Modells zur Prädiktion des Gesamtüberlebens anhand zweier dezidierter Patienten beurteilt. Hierzu wird aus dem Test-Set der Patient mit der kürzesten Überlebenszeit (120 Tage), und der Patient mit der längsten Überlebenszeit (beim Zeitpunkt von 2043 Tagen erfolgte der letzte dokumentierten Kontakt mit dem Patienten) weiter untersucht. Um das individuelle Überleben zu prognostizieren wird der Kaplan-Meier-Schätzer, basierend auf unserem kombinierten Modell, angewendet.

Anhand des berechneten risk scores, welcher die zu erwartenden Anzahl von Ereignissen in einem Random Forest des besagten Patienten darstellt, zeigt der Patient mit einer kürzeren Überlebenszeit einen risk score von 26,89. Der Patient mit einer

längeren Überlebenszeit weist dagegen einen risk score von 23,55 auf. Die Kaplan-Meier-Schätzung weist hier eine signifikante Differenz im Log-Rang-Test (log rank Test) auf ($p = 0,006$), siehe Abbildung 37.

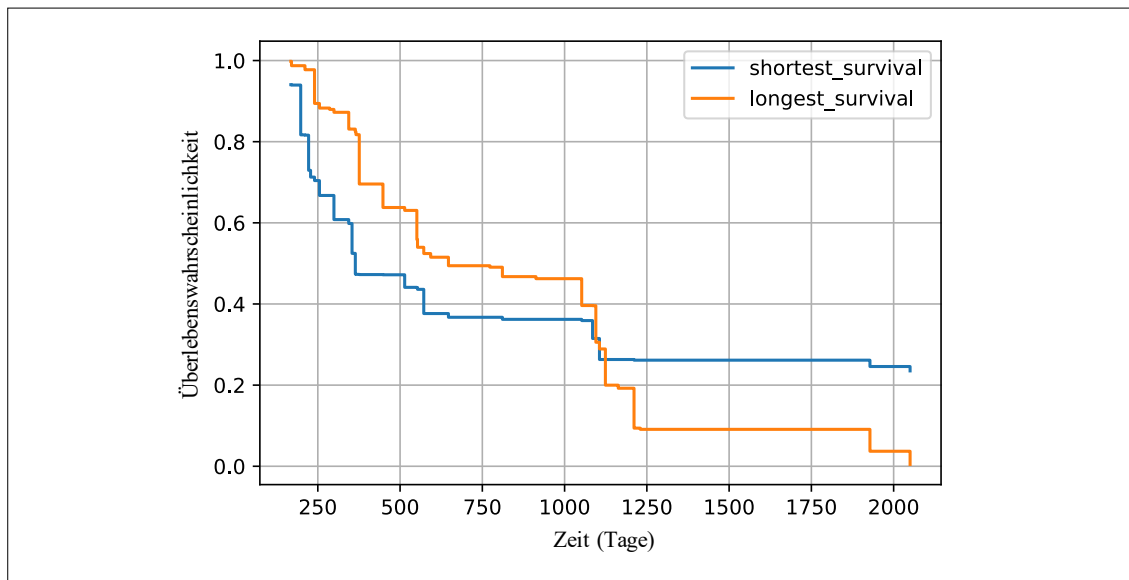


Abbildung 37. Kaplan-Meier-Schätzung, kombiniertes Modell geschätzte OS zweier Patienten, nach Bernatz, Elenberger et al.⁸²

4. Diskussion

Ziel dieser Arbeit ist es ein Modell zu beschreiben, das eine, ergänzend zur gängigen Praxis eines Tumorboard-Konsensus, zusätzliche und objektive Aussage zum Therapieansprechen einer TACE-Prozedur liefern soll. Die Frage ob eine weitere TACE-Prozedur sinnvoll ist, soll durch das neue Modell datenbasiert und damit objektiv untermauert werden.

Dazu wird ein kombiniertes Modell, bestehend aus einem klinischen Score und den Daten in den vorhandenen Schnittbildern, welche mit dem Begriff Radiomics umrissen werden können, definiert.

Die Ergebnisse dieser Studie zeigen, dass die Lipiodolanreicherung in einer Zielläsion nach einer TACE-Prozedur als möglicher Biomarker für eine Vorhersage des Therapieansprechens nach TACE dienen kann.

Eine Vielzahl von Studien beschäftigt sich mit der Frage, wie Patienten, die von einer TACE-Prozedur profitieren, von Patienten, die sich formal für eine TACE eignen, aber letztendlich keinen Vorteil durch die Therapieoption erfahren, unterschieden werden können.^{51–54,57–59,99}

Die Ergebnisse der Studien erscheinen vielversprechend, im klinischen Alltag wird die Indikation zur TACE-Prozedur, sowie zu etwaigen Wiederholungen der TACE, jedoch vorwiegend durch einen Konsens der Expertenmeinungen im Rahmen eines Tumorboards getroffen.

Anhand der oben erwähnten Studien wurden die in dieser Studie evaluierten klinischen Scores ausgewählt.

Mehrere Studien beschäftigen sich mit Lipiodol als möglichem Biomarker zur Evaluation des Therapieansprechens nach einer TACE bei HCC-Patienten. Letzen et al.¹⁰⁰ haben in einer prospektiven Studie bei 32 Patienten (HCC: n=18, andere primäre und sekundäre Lebertumore: n=14) unmittelbar nach erfolgter TACE das Volumen der Lipiodolanreicherung mittels digitaler Volumetomographie (DVT) gemessen. Die Studie kann einen signifikanten Unterschied im Volumen der Lipiodolanreicherung zwischen TACE-respondern und non-respondern aufzeigen.¹⁰⁰

Stark et al.¹⁰¹ untersuchen die Verteilung und Dichte der Lipiodolanreicherung in HCC-Herden nach TACE-Prozeduren bei insgesamt 42 Patienten (HCC, n=26, andere

primäre und sekundäre Lebertumore: n=16). Die Dichte der Lipiodolanreicherung wird in einer 24 Stunden postprozeduralen nativen CT quantifiziert. Stark et al.¹⁰¹ konnten eine Korrelation zur Volumenabnahme von KM-anreichernden Tumorumfängen in Folgeuntersuchungen und damit das therapeutische Ansprechen nach TACE nachweisen.¹⁰¹

Park et al.¹⁰² extrahieren in einer retrospektiven Studie Radiomics Features aus HCC-Herden anhand einer präinterventionellen KM-verstärkten CT. In die Studie werden 88 Patienten eingeschlossen die sich ein bis zwei TACE-Interventionen unterziehen. Park et al.¹⁰² vergleichen drei Modelle: klinischer Score (bestehend aus Child-Pugh-Score, AFP und Tumorgröße), Radiomics-Score und einer Kombination aus klinischem und Radiomics-Score in Bezug auf die Vorhersagbarkeit des OS. In der Studie kann die Überlegenheit des kombinierten Scores gegenüber dem klinischen oder Radiomics Score alleine nachgewiesen werden.¹⁰²

Nach dem besten Wissen des Autors dieser Arbeit gibt es zum aktuellen Zeitpunkt keine Arbeiten welche die Lipiodolanreicherung nach TACE-Prozeduren bei HCC-Patienten, als Interpretation für ein Gesamtüberleben des Patienten, mittels hochdimensionalen Radiomics-Featureanalysen untersucht.

In der vorliegenden Studie werden die Daten von 61 an einem HCC erkrankten Patienten, die sich mehrfachen TACE-Prozeduren unterzogen haben, untersucht. Wir nutzen die Lipiodolanreicherung nach der ersten TACE-Prozedur als Biomarker für ein Therapieansprechen, um anhand dieser Daten eine Vorhersage zum generellen Überleben der Patienten zu treffen.

Um die Daten auszuwerten bedienen wir uns mehreren Schritten der Datenaufbereitung. Um eine Redundanz der Features zu verhindern wenden wir die Pearson Korrelation an. Um eine Stabilität der Features zu gewährleisten, erfolgt die Extraktion der Features durch mehrere unabhängigen Radiologen.

Die so erhaltenen Daten werden wiederum durch diverse ML-Modelle mit der sogenannten Hyperparameteroptimierung trainiert, um eine Aussage zum Tumoransprechen auf die initiale TACE-Prozedur treffen zu können (siehe Kapitel 2.8). Die so erhaltenen Modelle werden miteinander verglichen. Das beste Modell, gemessen an der AUROC erweist sich als ein kombiniertes Modell, welches neben einem klinischen Score (mHAP-II-Score) Radiomics Features

(LargeDependenceHighGrayLevelEmphasis) enthält. Dieses Modell wird verwendet, um das Gesamtansprechen, abhängig von den mRECIST Kriterien, zu beurteilen.

Die Hypothese dieser Arbeit, Lipiodolanreicherung als Biomarker nach einer TACE zur Differenzierung zwischen Patienten, welche von einer TACE profitieren, und Patienten die keinen Nutzen von einer TACE erfahren werden, wird durch die erhaltenen Daten gestützt.

Das Modell kann Risiko Scores berechnen, welche eine Prognose zum Gesamtüberleben erlauben.

Die für ein HCC typischen Bildgebungscharakteristika mit arteriellem Hyperinhancement gefolgt von einem Auswaschen des Kontrastmittels in der portalvenösen und/oder späten Phase der Untersuchung, sowie die Beurteilung der vitalen Tumoranteile mittels der mRECIST Kriterien sind etablierte Parameter im klinischen Alltag.^{11,12,41}

Mehrere Studien haben sich bereits mit der Idee befasst, in prätherapeutisch erlangten kontrastmittelgestützten Schnittbilddatensätzen Biomarker zu definieren, mit welchen sie prädiktive Modelle für ein Ansprechen des HCC nach TACE-Prozeduren entwickeln.¹⁰³⁻¹⁰⁶ Die erwähnten Studien bilden Nomogramme abgeleitet von klinischen und bildgebenden Features. Die C-Indices der erwähnten Studien reichen von 0,70 – 0,77 und zeigen so vielversprechende Ergebnisse in Bezug auf die Vorhersagewertigkeit der Modelle in Bezug auf das Gesamtüberleben, wobei die in dieser Arbeit vorliegenden Daten mit einem C-Index von 0,71 mit diesen Studien vergleichbar sind.^{103,105,106}

Kuang et al.¹⁰⁴ erreicht mit prätherapeutischen MRT Datensätzen und klinischen Parametern eine Prognose des Therapieansprechens auf Läsionsebene (Ansprechverhalten einzelner umschriebener TL) mit einer AUC von 0,81.¹⁰⁴ Die Studie weist jedoch weder eine Aussage zum Gesamtansprechen auf Patientenebene, noch in Bezug auf das Gesamtüberleben, auf. Ebenso bleibt unklar wie viele TACE-Prozeduren vor der Analyse erfolgen.¹⁰⁴ Die hier vorliegende Studie hingegen folgt einem deutlich stringenterem, zweistufigen, Ablauf mit der Definition eines Modells auf Ebene der Läsion. Anschließend wird das Modell auf die Patientenebene angewendet. Hier wird neben dem Ansprechen der TL die Vorhersagekraft des Modells auf das Gesamtansprechen, abhängig nicht nur von der TL, sondern auch von NTL und etwaigen neuen Läsionen, validiert. Zuletzt zeigen wir die Fähigkeiten unseres Modells zur Prädiktion des Gesamtüberlebens auf.

Mit dem Ansatz der Gewinnung von Daten aus einer nativen CT-Untersuchung wird das Problem, der Bewegungsartefaktüberlagerung und zeitlich teilweise problematischen Erfassung der arteriellen Phase der Untersuchung mit entsprechender Minderung der Bildqualität, eliminiert.¹⁰⁷ Bewegungsartefakte und zeitlich differierende Kontrastmittelanreicherungsmuster führen zu einer negativen Beeinflussung der Entwicklung von zuverlässigen AI/ML Modellen. Sie führen zu einem vermehrtem Rauschen in einem System, das auch in einer perfekten experimentellen Umgebung unter Mängeln in Hinblick auf die Robustheit und Belastbarkeit der Modelle leidet.^{89,94,108} Entsprechend vorangegangenen Studienergebnissen unterstützen die Ergebnisse der vorliegenden Studie das Potenzial, und die Vorteile, Lipiodolablagerungen in der Schnittbildgebung als Biomarker zu nutzen.^{109,110} Miszczuk et al.¹¹⁰ haben in der prospektiven Studie insgesamt 39 Patienten mit einem malignen Lebertumor eingeschlossen, 29 der Patienten in der Gruppe mit einem HCC, welche einer TACE zugeführt wurden. Die Studie zeigt, dass die Lipiodolanreicherung in einer 24 Stunden post-interventionell angefertigten nativen CT-Untersuchung einen Zusammenhang zum Therapieansprechen aufweist. Die Lipiodolanreicherung kann als Surrogat für das arterielle Kontrastmittel-Anreicherungsmuster angesehen werden.¹¹⁰

Das Kontrastmittel-Anreicherungsmuster der HCC-Herde hat einen prognostischen Wert für das Gesamtüberleben, was von den vorliegenden Daten unterstützt wird.¹¹¹ In dem in dieser Arbeit definierten Modell hat das Feature LargeDependenceHighGrayLevelEmphasis, ein Feature das zu der Featureklasse Gray Level Dependence Matrix gehört, die höchste prädiktive Kraft. Das erscheint nachvollziehbar nachdem nach Miszczuk et al.¹¹⁰ das Lipiodol-Anreicherungsmuster als höchst prädiktiv für das Therapieansprechen gewertet werden kann, und das Feature LargeDependenceHighGrayLevelEmphasis die Verteilung von Grauwerten widerspiegelt.^{90,110}

Das Feature LargeDependenceHighGrayLevelEmphasis findet sich auch in einer Studie von Brancato et al.¹¹² die anhand von Radiomics Features das Grading der untersuchten HCC vorhersagen.¹¹² Dieses Feature ist Teil des Modells mit der höchsten Vorhersagekraft zur Differenzierung zwischen Grad 1 und Grad 3 Tumoren, was das Potenzial des entsprechenden Features als bildgebender Biomarker unterstreicht.¹¹²

Die aktuellen Guidelines für das hepatozelluläre Karzinom der European Society for Medical Oncology (ESMO) empfehlen prognostische Scores als Entscheidungshilfe für einen Behandlungsalgorithmus nur innerhalb einer klinischen Studie. Für den klinischen

Alltag werden keine Empfehlung für die Nutzung dieser Scores zur Evaluierung des Behandlungsschemas empfohlen. Die Guidelines sehen lediglich den hepatoma arterial-embolisation prognostic Score (HAP-Score) als potenzielles Tool zur Stratifizierung, das heißt Einteilung von Studiengruppen, in zukünftigen Studien zur TACE.¹² Dies bestätigt die in dieser Studie vorliegenden Ergebnisse, da die Auswertung aller klinischer Scores einen Bias aufzeigt. Alle klinischen Scores weisen einen verzerrten Train-/Test-AUC auf, mit jeweils höherem Test-AUC im Vergleich zum Trainingsdatensatz (siehe Abbildung 33).

Die vorliegende Arbeit weist Limitationen auf, die bedacht und diskutiert werden müssen. Zum einen kann auf Grund des retrospektiven Studiendesigns eine Stichprobenverzerrung vorliegen. Dem entgegenwirken könnte man mit einem prospektiven Studiendesign in zukünftigen ähnlich gelagerten Untersuchungen.

Mit 61 Studienpatienten ist zudem die Stichprobe relativ klein, weist aber eine hohe Homogenität auf. Alle eingeschlossenen Patienten haben ein histologisch nachgewiesenes HCC und die gesamte Gruppe erhielt 3 TACE-Behandlungen vor einer bildgebenden Verlaufskontrolle. Desweiteren erhielten alle Patienten die gleichen Chemotherapeutika im Rahmen der TACE. Es muss jedoch erwähnt werden, dass im Rahmen der TACE bei ca. 20% der Patienten zusätzlich zu den standardisierten Chemotherapeutika abbaubare Stärkemikrosphären (EmboCept®S, PharmaCept GmbH, Berlin, Germany) verabreicht worden sind. Diese Tatsache könnte die Lipiodolanreicherung im Patientenkollektiv beeinflusst haben. In einer prospektiven Studie sollte, um den Einfluss der abbaubaren Mikrosphären auszuschließen, entweder darauf verzichtet, oder eine Kontrollgruppe ohne Anwendung der Mikrosphären als Referenzgruppe etabliert werden.

Die Originaldatensätze der 24 Stunden postinterventionellen CT zur Evaluation der Lipiodolanreicherungen, lagen in einer Schichtdicke von 5 mm vor. Obwohl alle Bilder auf eine Schichtdicke von 1 x 1 x 1 mm reformatiert wurden, wäre eine initiale Rekonstruktion der Bilddaten in einer Schichtdicke von 1 mm vorzuziehen.

Die in der Abbildung 37 aufgezeigten vorhergesagten Überlebenskurven weisen eine Überkreuzung auf, welches ein Hinweis für einen möglichen Bias darstellt. Eine Re-Validierung des Vorgehens mit einem größeren Patientenpool, idealerweise in einem multizentrischen Setting wäre sinnvoll um die Generalisierbarkeit der Ergebnisse der Studie zu ratifizieren.

In Summe lässt sich aus der vorliegenden Arbeit schließen, dass Radiomics Features, die aus standardmäßig erhobenen nativen CT-Untersuchungen 24 Stunden nach TACE-Prozedur, generiert werden, ein Potenzial haben als bildgebende Biomarker zu dienen, und einen Rückschluss auf das Ansprechen der TACE bei Patienten mit einem zugrundeliegenden HCC erlauben. Dieses Ergebnis bestätigt die zuvor aufgestellte Hypothese.

Bildgebende Biomarker sind in der vorliegenden Studie klinischen Scores überlegen. Ein kombiniertes Modell, bestehend sowohl aus einem klinischen Score als auch einem bildgebenden Biomarker, zeigt die besten Resultate in Hinblick auf die Vorhersage eines Gesamtüberlebens. Unsere Arbeit liefert Erkenntnisse, dass zukünftig im klinischen Kontext möglicherweise entsprechende Modelle eine Hilfestellung bei der Therapieentscheidung bieten können, jenseits der typischen Entscheidung basierend auf Expertenkonsensus.

Das vorgestellte kombinierte Modell sollte in weiterführenden Studien evaluiert und validiert werden und könnte somit potentiell zukünftige Anwendungsgebiete im Kontext personalisierter Medizin eröffnen, mit dem Ziel der Risikominimierung von Nebenwirkungen bei Patienten, die von einer TACE-Behandlung nicht profitieren würden.

5. Zusammenfassung

Auf Grund einer hohen Inzidenz und Mortalität, welche in den nächsten Jahren voraussichtlich eine deutliche Zunahme erfahren wird, stellt die Behandlung eines HCC an alle beteiligten Fächer der Medizin, sowie an den Patienten und die Patientin, eine enorme Herausforderung dar. In der klinischen Routine hat sich die TACE, nicht nur bei Patienten im intermediären Stadium der Erkrankung, etabliert, sodass im Laufe der Erkrankung nahezu jeder zweite Patient mindestens eine TACE-Behandlung bekommt. Der mit Radiomics betitelte, im medizinischen Bereich relativ junge, Forschungszweig beschäftigt sich mit der Idee, dass in den Schnittbildern eine für das menschliche Auge nicht sichtbare Ebene von Informationen vorliegt, welche mit den richtigen Mitteln extrahiert, relevante Daten und Informationen zur Genetik, Phänotypie und Pathophysiologie des Tumors liefern kann.

Hier greift der Ansatz dieser Arbeit an. In dieser Arbeit wird die Hypothese postuliert, dass durch die Auswertung und Integration von Lipiodolablagerungen in der Zielläsion nach der ersten durchgeführten TACE eine zuverlässigere Prognose zum Therapieansprechen und Gesamtüberleben mit Hilfe von Radiomics möglich ist, als dies klinische Scores alleine erlauben.

Dazu wurde in dieser Arbeit ein Patientenstamm von 61 Patienten untersucht. Alle Patienten litten an einem histologisch gesicherten HCC. Bei allen Patienten wurden innerhalb eines Zeitintervalls von 6 Monaten drei TACE durchgeführt mit einer nachfolgenden Verlaufskontrolle mittels kontrastmittelgestützter MRT oder CT. In einem dezidierten, mehrstufigen Verfahren wurden aus der nativen 24 Stunden postinterventionellen CT-Kontrolle die Lipiodol anreichernden HCC-Herde segmentiert. Aus diesem segmentierten 3-D Bilddatensatz wurde eine Vielzahl von bildgebenden Biomarkern, Features, extrahiert. Die Features wurden im weiteren Prozess selektiert, redundante und nicht reproduzierbare Features wurden für das weitere Vorgehen verworfen.

Aus den vorliegenden Daten der Patienten wurden Informationen selektiert, mit welchen insgesamt 5 klinische Scores berechnet wurden, diese Scores wurden im weiteren Verfahren ebenfalls als Features angesehen.

Mehrere Machine Learning-Algorithmen wurden mit der Zielvariable:

Größenregredienz des Tumors nach TACE als Folge eines annehmbaren Therapieansprechens, angelernt.

Das beste Ergebnis lieferte ein ML-Algorithmus mit einem Random Forrest Klassifikator auf der Grundlage des kombinierten, aus Radiomics-Features und klinischem Score-Features bestehendem Featuresets.

Um die initial aufgestellte Hypothese zu überprüfen wurde die Zielvariable von Größenregredienz der TL auf OS verändert. Die Performance des ML-Algorithmus in Bezug auf die neu definierte Zielvariable OS wurde hierbei mit dem C-index bewertet. Im Test-Set liegt ein C-Index von 0,67 vor. Das kombinierte Modell aus klinischem Score und Radiomics zeigt hierbei eine Überlegenheit gegenüber dem klinischen Score allein (C-Index 0,58) und dem Radiomics score (C-Index 0,60). Dies bestätigt die aufgestellte Hypothese. Das kombinierte Modell hat die Fähigkeit, anhand der Lipiodolanreicherung in der 24 Stunden postinterventionell durchgeführten CT, zur Prädiktion eines Gesamtüberlebens von HCC-Patienten nach einer TACE.

Die Patienten mit der kürzesten und längsten Überlebenszeit innerhalb der Studienpopulation dienten als Grundlage für eine Kaplan-Meier-Schätzung und Berechnung eines Risiko-Scores (siehe Abbildung 37). Dabei zeigt sich eine signifikante Differenz zwischen den Risiko-Scores. Eine Kurve dieser Art könnte zukünftig theoretisch als Schätzung zur Überprüfung der Indikation einer TACE-Wiederholung für einzelne Patienten dienen. Für eine entsprechende Generalisierbarkeit sind weiterführende Studien zur Validierung nötig. Unsere Studie liefert hier erste vielversprechende Hinweise, wobei unsere Limitationen nicht zu vernachlässigen sind, wie im Detail diskutiert.

Zusammenfassend zeigt unsere Arbeit, dass ein von uns definierter kombinierter Score, bestehend aus bildgebenden Biomarkern (Radiomics) und einem klinischen Score (m-HAP-II-Score), eine Prognose zum Gesamtüberleben nach der ersten TACE-Behandlung liefern kann. Mit Hilfe dieses kombinierten Scores war es in unserer Studienkohorte möglich abzuschätzen, ob ein Patient von weiteren TACE-Prozeduren profitieren würde. Der Behandlungsalgorithmus könnte auf dieser Basis individuell angepasst werden.

Der kombinierte Score hätte somit nicht nur das Potenzial Nebenwirkungen zu verhindern und Kosten im System einzusparen, sondern ebenfalls den Patienten potentiell individuell effektiveren Therapiealternativen zuzuführen.

6. Summary

Due to a high incidence and mortality, which is expected to increase significantly in the coming years, the treatment of HCC is an enormous challenge for all medical disciplines involved, as well as for the patient.

In clinical routine, TACE has become established, not only for patients in the intermediate stage of the disease, so that in the course of the disease almost every second patient receives at least one TACE treatment.

Radiomics, a relatively new branch of research in the medical field, deals with the idea that there is a layer of information in the cross-sectional images that is not visible to the human eye, which, extracted with the right tools, can provide relevant data and information on the genetics, phenotype and pathophysiology of the tumor.

This is where the approach of this work comes into play. In this work the hypothesis is postulated that by evaluating and integrating lipiodol deposits in the target lesion after the first performed TACE, a more reliable prognosis on treatment response and overall survival is possible using radiomics than clinical scores alone allow.

For this purpose, a patient population of 61 patients was studied in this work. All patients had histologically confirmed HCC. In all patients, 3 TACE treatments were performed within a time interval of 6 months with subsequent follow-up by contrast-enhanced MRI or CT.

Lipiodol-enhancing HCC foci were segmented from native 24-h post-interventional CT follow-up in a dedicated, multistage procedure. From this segmented 3-D image dataset, a variety of imaging biomarkers, features, were extracted. The features were selected in the further process, redundant and non-reproducible features were discarded for further procedure.

From the available data of the patients, information was selected with which a total of 5 clinical scores were calculated, these scores were also considered as features in the further process.

Several machine learning algorithms were trained with the target variable: tumor size regression after TACE as a consequence of an acceptable treatment response.

The best result was provided by an ML algorithm with a Random Forrest classifier based on the combined feature set consisting of radiomics features and clinical score

features.

In order to test the initial hypothesis, the target variable was changed from size regression of TL to OS. The performance of the ML algorithm with respect to the newly defined target variable OS was evaluated with the C-index.

The combined model of clinical score and radiomics showed superiority over the clinical score alone (C-index 0.58) and the radiomics score (C-index 0.60) with a Test-Set C-index of 0,67. This confirms the hypothesis. The combined model had the ability, based on lipiodol accumulation in 24-h postinterventional CT, to predict overall survival of HCC patients after TACE.

The patients with the shortest and longest survival within our study population served as the basis for a Kaplan-Meier estimate and calculation of a risk score (see Figure 35). This revealed a significant difference between the risk scores. A curve of this type could theoretically serve as an estimate to verify the indication of repeat TACE for individual patients. Further validation studies are needed for appropriate generalizability. Our study provides some promising evidence in this regard, although our limitations are not negligible, as discussed in detail.

In summary, our work shows that a combined score we defined, consisting of imaging biomarkers (radiomics) and a clinical score (m-HAP-II score), can provide a prognosis on overall survival after the first TACE treatment. Using this combined score, it was possible in our study cohort to assess whether a patient would benefit from further TACE procedures. The treatment algorithm could be individually adjusted on this basis. The combined score would thus not only have the potential to prevent side effects and save costs in the system, but also to potentially direct the patient individually to more effective treatment alternatives.

7. Literaturverzeichnis

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209-249. doi:10.3322/caac.21660
2. Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. *Nat Rev Dis Prim.* 2021;7(1). doi:10.1038/s41572-020-00240-3
3. Galle PR, Forner A, Llovet JM, et al. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol.* 2018;69(1):182-236. doi:10.1016/j.jhep.2018.03.019
4. RKI. Inzidenz_Numberof Cases_D_HCC. <https://www.krebsdaten.de>.
5. RKI. Inzidenz_ASRper100tsd_D_HCC. <https://www.krebsdaten.de>.
6. RKI. Mortality_Numberofcases_D_HCC. <https://www.krebsdaten.de>.
7. RKI. Mortality_ASRper100tsd_D_HCC. <https://www.krebsdaten.de>.
8. Fitzmaurice C, Akinyemiju T, Abera S, et al. The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level results from the global burden of disease study 2015. *JAMA Oncol.* 2017;3(12):1683-1691. doi:10.1001/jamaoncol.2017.3055
9. Estes C, Anstee QM, Arias-Loste MT, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. *J Hepatol.* 2018;69(4):896-904. doi:10.1016/j.jhep.2018.05.036
10. Villanueva A. Hepatocellular Carcinoma. Longo DL, ed. *N Engl J Med.* 2019;380(15):1450-1462. doi:10.1056/NEJMra1713263
11. Jiang HY, Chen J, Xia CC, Cao LK, Duan T, Song B. Noninvasive imaging of hepatocellular carcinoma: From diagnosis to prognosis. *World J Gastroenterol.* 2018;24(22):2348-2362. doi:10.3748/wjg.v24.i22.2348
12. Vogel A, Cervantes A, Chau I, et al. Hepatocellular carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2018;29(Supplement 4):iv238-iv255. doi:10.1093/annonc/mdy308
13. Roberts LR, Sirlin CB, Zaiem F, et al. Imaging for the diagnosis of

- hepatocellular carcinoma: A systematic review and meta-analysis. *Hepatology*. 2018;67(1):401-421. doi:10.1002/hep.29487
14. Leitlinienprogramm Onkologie. Konsultationsfassung Diagnostik und Therapie des Hepatozellulären Karzinoms und biliärer Karzinome Wesentliche Neuerungen in der Leitlinie zur Diagnostik und Therapie des hepatozellulären Karzinoms. 2021:1-205.
 15. Chernyak V, Fowler KJ, Kamaya A, et al. Liver Imaging Reporting and Data System (LI-RADS) version 2018: Imaging of hepatocellular carcinoma in at-risk patients. *Radiology*. 2018;289(3):816-830. doi:10.1148/radiol.2018181494
 16. Schreyer AG, Lüth S, Ringe KI. Bildgebung in der zirrhotischen Leber – warum, wie und was ist LI-RADS? *Radiol up2date*. 2022;22(01):69-84. doi:10.1055/a-1561-9413
 17. Algorithm D, Response T, Viewed L. CT/MRT LI-RADS ® v2018.
 18. Cho MJ, An C, Aljoqiman KS, et al. Diagnostic performance of Liver Imaging Reporting and Data System in patients at risk of both hepatocellular carcinoma and metastasis. *Abdom Radiol*. 2020;45(11):3789-3799. doi:10.1007/s00261-020-02581-9
 19. Liang Y, Xu F, Guo Y, et al. Diagnostic performance of LI-RADS for MRI and CT detection of HCC: A systematic review and diagnostic meta-analysis. *Eur J Radiol*. 2021;134(November 2020):109404. doi:10.1016/j.ejrad.2020.109404
 20. Llovet JM, Brú C, Bruix J. Prognosis of hepatocellular carcinoma: The BCLC staging classification. *Semin Liver Dis*. 1999;19(3):329-337. doi:10.1055/s-2007-1007122
 21. Roayaie S, Obeidat K, Sposito C, et al. Resection of hepatocellular cancer ≤ 2 cm: Results from two Western centers. *Hepatology*. 2013;57(4):1426-1435. doi:10.1002/hep.25832
 22. Mazzaferro V, REGALIA E, DOCI R, et al. Carcinomas in Patients With Cirrhosis. *N Engl J Med*. 1996;334(11):693-699.
 23. Deutsche Stiftung Organtransplantation. Grafiken zum Tätigkeitsbericht 2020. <http://www.dso.de>.
 24. Mahnken A, Thomas C WK. *Interventionelle Radiologie*.; 2019. doi:10.1055/b-005-143315

25. Gupta P, Maralakunte M, Kumar-M P, et al. Overall survival and local recurrence following RFA, MWA, and cryoablation of very early and early HCC: a systematic review and Bayesian network meta-analysis. *Eur Radiol.* 2021;31(7):5400-5408. doi:10.1007/s00330-020-07610-1
26. Weinstein JL, Ahmed M. Percutaneous ablation for hepatocellular carcinoma. *Am J Roentgenol.* 2018;210(6):1368-1375. doi:10.2214/AJR.17.18695
27. Vogl TJ, Nour-Eldin NEA, Hammerstingl RM, Panahi B, Naguib NNN. Microwave Ablation (MWA): Basics, Technique and Results in Primary and Metastatic Liver Neoplasms - Review Article. *RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgeb Verfahren.* 2017;189(11):1055-1066. doi:10.1055/s-0043-117410
28. Xuan D, Wen W, Xu D, Jin T. Survival comparison between radiofrequency ablation and surgical resection for patients with small hepatocellular carcinoma: A systematic review and meta-analysis. *Medicine (Baltimore).* 2021;100(7):e24585. doi:10.1097/MD.00000000000024585
29. Shiina S, Sato K, Tateishi R, et al. Percutaneous Ablation for Hepatocellular Carcinoma: Comparison of Various Ablation Techniques and Surgery. *Can J Gastroenterol Hepatol.* 2018;2018. doi:10.1155/2018/4756147
30. Llovet JM, Bruix J. Systematic review of randomized trials for unresectable hepatocellular carcinoma: Chemoembolization improves survival. *Hepatology.* 2003;37(2):429-442. doi:10.1053/jhep.2003.50047
31. Llovet JM, De Baere T, Kulik L, et al. Locoregional therapies in the era of molecular and immune treatments for hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol.* 2021;18(5):293-313. doi:10.1038/s41575-020-00395-0
32. Vogl TJ, Zangos S, Balzer JO, et al. Transarterielle chemoembolisation (TACE) des hepatozellulären karzinoms: Technik, indikationsstellung und ergebnisse. *RoFo Fortschritte auf dem Gebiet der Rontgenstrahlen und der Bildgeb Verfahren.* 2007;179(11):1113-1126. doi:10.1055/s-2007-963285
33. Facciorusso A, Di Maso M, Muscatiello N. Drug-eluting beads versus conventional chemoembolization for the treatment of unresectable hepatocellular carcinoma: A meta-analysis. *Dig Liver Dis.* 2016;48(6):571-577. doi:10.1016/j.dld.2016.02.005

34. Yang B, Liang J, Qu ZY, Yang FY, Liao ZY, Gou HF. Transarterial strategies for the treatment of unresectable hepatocellular carcinoma: A systematic review. *PLoS One*. 2020;15(2):1-20. doi:10.1371/journal.pone.0227475
35. Reig M, Forner A, Rimola J, et al. BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. *J Hepatol*. 2022;76(3):681-693. doi:10.1016/j.jhep.2021.11.018
36. Kolligs FT, Bilbao JI, Jakobs T, et al. Pilot randomized trial of selective internal radiation therapy vs. chemoembolization in unresectable hepatocellular carcinoma. *Liver Int*. 2015;35(6):1715-1721. doi:10.1111/liv.12750
37. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in Advanced Hepatocellular Carcinoma. *N Engl J Med*. 2008;359(4):378-390. doi:10.1056/nejmoa0708857
38. Kudo M, Finn RS, Qin S, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. *Lancet*. 2018;391(10126):1163-1173. doi:10.1016/S0140-6736(18)30207-1
39. Finn RS, Qin S, Ikeda M, et al. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *N Engl J Med*. 2020;382(20):1894-1905. doi:10.1056/nejmoa1915745
40. Cheng AL, Qin S, Ikeda M, et al. Updated efficacy and safety data from IMbrave150: Atezolizumab plus bevacizumab vs. sorafenib for unresectable hepatocellular carcinoma. *J Hepatol*. 2022;76(4):862-873. doi:10.1016/j.jhep.2021.11.030
41. Llovet JM, Lencioni R. mRECIST for HCC: Performance and novel refinements. *J Hepatol*. 2020;72(2):288-306. doi:10.1016/j.jhep.2019.09.026
42. Child CG, Turcotte JG. Surgery and portal hypertension. *Major Probl Clin Surg*. 1964;1:1-85. <http://www.ncbi.nlm.nih.gov/pubmed/4950264>.
43. Pugh RNH, Murray-Lyon IM, Dawson JL, Pietroni MC, Williams R. Transection of the oesophagus for bleeding oesophageal varices. *Br J Surg*. 1973;60(8):646-649. doi:10.1002/bjs.1800600817
44. Cholongitas E, Papatheodoridis G V., Vangeli M, Terreni N, Patch D, Burroughs AK. Systematic review: The model for end-stage liver disease - Should it replace Child-Pugh's classification for assessing prognosis in cirrhosis? *Aliment*

- Pharmacol Ther.* 2005;22(11-12):1079-1089. doi:10.1111/j.1365-2036.2005.02691.x
45. Malinchoc M, Kamath PS, Gordon FD, Peine CJ, Rank J, Ter Borg PCJ. A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts. *Hepatology.* 2000;31(4):864-871. doi:10.1053/he.2000.5852
 46. Forman LM, Lucey MR. Predicting the prognosis of chronic liver disease: An evolution from child to MELD. *Hepatology.* 2001;33(2):473-475. doi:10.1053/jhep.2001.22481
 47. Kaiser T, Zeuzem S, Thiery J. Laborwerte - Das Zünglein an der Waage. *Dtsch Arztebl Int.* 2013;110(3):E62-E70.
 48. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: A new evidence-based approach - The albi grade. *J Clin Oncol.* 2015;33(6):550-558. doi:10.1200/JCO.2014.57.9151
 49. Su TS, Yang HM, Zhou Y, et al. Albumin - Bilirubin (ALBI) versus Child-Turcotte-Pugh (CTP) in prognosis of HCC after stereotactic body radiation therapy. *Radiat Oncol.* 2019;14(1). doi:10.1186/s13014-019-1251-y
 50. Kao W-Y, Su C-W, Chiou Y-Y, et al. Based on the Albumin-Bilirubin Grade to Assess the Outcomes of Radiofrequency Ablation 1 VASCULAR AND INTERVENTIONAL RADIOLOGY: Albumin-Bilirubin Grade for Assessment of the Outcome of Radiofrequency Ablation Kao et al. *Fac Med (W Radiol.* 2017;285(2). <https://doi.org/10.1148/radiol.2017162382>.
 51. Kadalayil L, Benini R, Pallan L, et al. A simple prognostic scoring system for patients receiving transarterial embolisation for hepatocellular cancer. *Ann Oncol.* 2013;24(10):2565-2570. doi:10.1093/annonc/mdt247
 52. Müller L, Stoehr F, Mähringer-Kunz A, Hahn F, Weinmann A, Kloeckner R. Current Strategies to Identify Patients That Will Benefit from TACE Treatment and Future Directions a Practical Step-by-Step Guide. *J Hepatocell Carcinoma.* 2021;Volume 8:403-419. doi:10.2147/jhc.s285735
 53. Park Y, Kim SU, Kim BK, et al. Addition of tumor multiplicity improves the prognostic performance of the hepatoma arterial-embolization prognostic score. *Liver Int.* 2016;36(1):100-107. doi:10.1111/liv.12878

54. Peisen F, Maurer M, Grosse U, et al. Predictive performance of the mHAP-II score in a real-life western cohort with hepatocellular carcinoma following trans-arterial chemoembolisation with drug-eluting beads (DEB-TACE). 2020:3782-3792.
55. Müller L, Stoehr F, Mähringer-kunz A, Hahn F. Current Strategies to Identify Patients That Will Benefit from TACE Treatment and Future Directions a Practical Step-by-Step Guide. 2021:403-419.
56. Park JW, Chen M, Colombo M, et al. Global patterns of hepatocellular carcinoma management from diagnosis to death: The BRIDGE Study. *Liver Int.* 2015;35(9):2155-2166. doi:10.1111/liv.12818
57. Hucke F, Pinter M, Graziadei I, et al. How to STATE suitability and START transarterial chemoembolization in patients with intermediate stage hepatocellular carcinoma. *J Hepatol.* 2014;61(6):1287-1296. doi:10.1016/j.jhep.2014.07.002
58. Wang Q, Xia D, Bai W, et al. Development of a prognostic score for recommended TACE candidates with hepatocellular carcinoma: A multicentre observational study. *J Hepatol.* 2019;70(5):893-903. doi:10.1016/j.jhep.2019.01.013
59. Kaewdech A, Sripongpun P, Cheewasereechon N, Jandee S, Chamroonkul N, Piratvisuth T. Validation of the “Six-and-Twelve” Prognostic Score in Transarterial Chemoembolization-Treated Hepatocellular Carcinoma Patients. *Clin Transl Gastroenterol.* 2021;12(2):e00310. doi:10.14309/ctg.0000000000000310
60. Perakakis N, Yazdani A, Karniadakis GE, Mantzoros C. Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. *Metabolism.* 2018;87:A1-A9. doi:10.1016/j.metabol.2018.08.002
61. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology.* 2016;278(2). doi:10.1148/radiol.2015151169
62. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141

63. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21):e104-e107. doi:10.1158/0008-5472.CAN-17-0339
64. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to radiomics. *J Nucl Med.* 2020;61(4):488-495. doi:10.2967/JNUMED.118.222893
65. Kromrey M-L, Grothe S, Nell C, Rosenberg B. Künstliche Intelligenz in der Radiologie. *Radiol up2date.* 2022;22(02):121-136. doi:10.1055/a-1718-8846
66. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights Imaging.* 2020;11(1). doi:10.1186/s13244-020-00887-2
67. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5. doi:10.1038/ncomms5006
68. Jin Z, Chen L, Zhong B, et al. Machine-learning analysis of contrast-enhanced computed tomography radiomics predicts patients with hepatocellular carcinoma who are unsuitable for initial transarterial chemoembolization monotherapy: A multicenter study. *Transl Oncol.* 2021;14(4):101034. doi:10.1016/j.tranon.2021.101034
69. Park HJ, Park B, Lee SS. Radiomics and Deep Learning : Hepatic Applications. 2020;21(4):387-401.
70. Cuocolo R, Cipullo MB, Stanzione A, et al. Machine learning for the identification of clinically significant prostate cancer on MRI: a meta-analysis. *Eur Radiol.* 2020;30(12):6877-6887. doi:10.1007/s00330-020-07027-w
71. Pérez MJ, Grande RG. Application of artificial intelligence in the diagnosis and treatment of hepatocellular carcinoma: A review. *World J Gastroenterol.* 2020;26(37):5617-5628. doi:10.3748/wjg.v26.i37.5617
72. Gitto S, Cuocolo R, Albano D, et al. MRI radiomics-based machine-learning classification of bone chondrosarcoma. *Eur J Radiol.* 2020;128(December 2019). doi:10.1016/j.ejrad.2020.109043
73. Rudie JD, Rauschecker AM, Bryan RN, Davatzikos C, Mohan S. Emerging Applications of Artificial Intelligence in Neuro-Oncology. *Radiology.* 2019;290(3):607-618. doi:10.1148/radiol.2018181928

74. Cuocolo R, Uggia L, Solari D, et al. Prediction of pituitary adenoma surgical consistency: radiomic data mining and machine learning on T2-weighted MRI. *Neuroradiology*. 2020;62(12):1649-1656. doi:10.1007/s00234-020-02502-z
75. Cuocolo R, Perillo T, De Rosa E, Uggia L, Petretta M. Current applications of big data and machine learning in cardiology. *J Geriatr Cardiol*. 2019;16(8):601-607. doi:10.11909/j.issn.1671-5411.2019.08.002
76. FDA approves stroke-detecting AI software. *Nat Biotechnol*. 2018;36(4):290-290. doi:10.1038/nbt0418-290
77. Artificial-Intelligence-and-Machine-Learning-Software-Medical-Device @ Wwww.Fda.Gov. <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>.
78. Ahmad MA, Overman S, Allen C, Kumar V, Teredesai A, Eckert C. Software as a Medical Device. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. New York, NY, USA: ACM; 2021:4023-4024. doi:10.1145/3447548.3470823
79. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. *BMJ Open*. 2016;6(11):1-17. doi:10.1136/bmjopen-2016-012799
80. Blackburn H, West S. Management of postembolization syndrome following hepatic transarterial chemoembolization for primary or metastatic liver cancer. *Cancer Nurs*. 2016;39(5):E1-E18. doi:10.1097/NCC.0000000000000302
81. Mason MC, Massarweh NN, Salami A, Sultenfuss MA, Anaya DA. Post-embolization syndrome as an early predictor of overall survival after transarterial chemoembolization for hepatocellular carcinoma. *Hpb*. 2015;17(12):1137-1144. doi:10.1111/hpb.12487
82. Bernatz S, Elenberger O, Ackermann J, et al. CT - radiomics and clinical risk scores for response and overall survival prognostication in TACE HCC patients. *Sci Rep*. 2023;13:1-9. doi:10.1038/s41598-023-27714-0
83. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging*. 2012;30(9):1323-1341. doi:10.1016/j.mri.2012.05.001
84. Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges. *Magn*

- Reson Imaging*. 2012;30(9):1234-1248. doi:10.1016/j.mri.2012.06.010
85. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. 2016. doi:10.1148/radiol.2020191145
 86. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328-338. doi:10.1148/radiol.2020191145
 87. Parmar C, Velazquez ER, Leijenaar R, et al. Robust radiomics feature quantification using semiautomatic volumetric segmentation. Woloschak GE, ed. *PLoS One*. 2014;9(7):e102107. doi:10.1371/journal.pone.0102107
 88. Velazquez ER, Parmar C, Jermoumi M, et al. Volumetric CT-based segmentation of NSCLC using 3D-Slicer. *Sci Rep*. 2013;3. doi:10.1038/srep03529
 89. Bernatz S, Zhdanovich Y, Ackermann J, et al. Impact of rescanning and repositioning on radiomic features employing a multi-object phantom in magnetic resonance imaging. *Sci Rep*. 2021;11(1). doi:10.1038/s41598-021-93756-x
 90. Organization P. Pyradiomics Documentation. 2022:130. <http://www.radiomics.io/>.
 91. Bernatz S, Ackermann J, Mandel P, et al. Comparison of machine learning algorithms to predict clinically significant prostate cancer of the peripheral zone with multiparametric MRI using clinical assessment categories and radiomic features. 2020:6757-6769.
 92. Park JE, Kim HS, Kim D, et al. A systematic review reporting quality of radiomics research in neuro-oncology: Toward clinical utility and quality improvement using high-dimensional imaging features. *BMC Cancer*. 2020;20(1):1-11. doi:10.1186/s12885-019-6504-5
 93. Vallat R. Pingouin: statistics in Python. *J Open Source Softw*. 2018;3(31):1026. doi:10.21105/joss.01026
 94. Baeßler B, Weiss K, Santos DP Dos. Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study. *Invest Radiol*. 2019;54(4):221-228. doi:10.1097/RLI.0000000000000530
 95. Komer B, Bergstra J, Eliasmith C. Hyperopt-Sklearn.

96. Bergstra J, Yamins D, Cox DD. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *30th Int Conf Mach Learn ICML 2013*. 2013;(PART 1):115-123.
97. Pölsterl S. Scikit-survival: A library for time-to-event analysis built on top of scikit-learn. *J Mach Learn Res*. 2020;21:1-6.
98. Davidson-Pilon C. lifelines: survival analysis in Python. *J Open Source Softw*. 2019;4(40):1317. doi:10.21105/joss.01317
99. Op Den Winkel M, Nagel D, Op Den Winkel P, et al. The munich-transarterial chemoembolisation score holds superior prognostic capacities compared to TACE-tailored modifications of 9 established staging systems for hepatocellular carcinoma. *Digestion*. 2019;100(1):15-26. doi:10.1159/000493136
100. Letzen BS, Malpani R, Miszczuk M, et al. Lipiodol as an intra-procedural imaging biomarker for liver tumor response to transarterial chemoembolization: Post-hoc analysis of a prospective clinical trial. *Clin Imaging*. 2021;78:194-200. doi:10.1016/j.clinimag.2021.05.007
101. Stark S, Wang C, Savic LJ, et al. Automated feature quantification of Lipiodol as imaging biomarker to predict therapeutic efficacy of conventional transarterial chemoembolization of liver cancer. *Sci Rep*. 2020:1-13. doi:10.1038/s41598-020-75120-7
102. Park H, Kim J, Sj C, Sh L, Hy L, Park H. Chemoembolization : Comparison of Models Using Radiomics. 2018;(November):1026-1034.
103. Dai Y, Jiang H, Feng S-T, et al. Noninvasive Imaging Evaluation Based on Computed Tomography of the Efficacy of Initial Transarterial Chemoembolization to Predict Outcome in Patients with Hepatocellular Carcinoma. *J Hepatocell Carcinoma*. 2022;Volume 9(March):273-288. doi:10.2147/jhc.s351077
104. Kuang Y, Li R, Jia P, et al. MRI-Based Radiomics: Nomograms predicting the short-term response after transcatheter arterial chemoembolization (TACE) in hepatocellular carcinoma patients with diameter less than 5 cm. *Abdom Radiol*. 2021;46(8):3772-3789. doi:10.1007/s00261-021-02992-2
105. Li L, Kan X, Zhao Y, et al. Radiomics signature: A potential biomarker for the prediction of survival in advanced hepatocellular carcinoma. *Int J Med Sci*.

- 2021;18(11):2276-2284. doi:10.7150/ijms.55510
106. Meng XP, Wang YC, Ju S, et al. Radiomics Analysis on Multiphase Contrast-Enhanced CT: A Survival Prediction Tool in Patients With Hepatocellular Carcinoma Undergoing Transarterial Chemoembolization. *Front Oncol.* 2020;10(July):1-12. doi:10.3389/fonc.2020.01196
 107. Ikram NS, Yee J, Weinstein S, et al. Multiple arterial phase MRI of arterial hypervascular hepatic lesions: improved arterial phase capture and lesion enhancement. *Abdom Radiol.* 2017;42(3):870-876. doi:10.1007/s00261-016-0948-8
 108. Berenguer R, Del Rosario Pastor-Juan M, Canales-Vázquez J, et al. Radiomics of CT features may be nonreproducible and redundant: Influence of CT acquisition parameters. *Radiology.* 2018;288(2):407-415. doi:10.1148/radiol.2018172361
 109. Chen CS, Li FK, Guo CY, et al. Tumor vascularity and lipiodol deposition as early radiological markers for predicting risk of disease progression in patients with unresectable hepatocellular carcinoma after transarterial chemoembolization. *Oncotarget.* 2016;7(6):7241-7252. doi:10.18632/oncotarget.6892
 110. Miszczuk MA, Chapiro J, Geschwind JFH, et al. Lipiodol as an Imaging Biomarker of Tumor Response After Conventional Transarterial Chemoembolization: Prospective Clinical Validation in Patients with Primary and Secondary Liver Cancer. *Transl Oncol.* 2020;13(3):100742. doi:10.1016/j.tranon.2020.01.003
 111. Hasdemir DB, Dávila LA, Schweitzer N, et al. Evaluation of CT vascularization patterns for survival prognosis in patients with hepatocellular carcinoma treated by conventional TACE. *Diagnostic Interv Radiol.* 2017;23(3):217-222. doi:10.5152/dir.2016.16006
 112. Brancato V, Garbino N, Salvatore M, Cavaliere C. MRI-Based Radiomic Features Help Identify Lesions and Predict Histopathological Grade of Hepatocellular Carcinoma. *Diagnostics.* 2022;12(5). doi:10.3390/diagnostics12051085

8. Anhang

8.1 Abkürzungsverzeichnis

ACR	American College of Radiology
AFP	Alpha-Fetoprotein
AI	artificial intelligence, dt. künstliche Intelligenz
ALBI	Albumin Bilirubin Score
AUROC	area under ROC, dt. Fläche unter der ROC-Kurve
BCLC	Barcelona Clinic Liver Score
bzw.	beziehungsweise
ca.	circa
CEUS	Contrast Enhanced Ultrasound, dt. kontrastmittelverstärkter Ultraschall
chir.	chirurgisch
cm	Zentimeter
CR	complete response, dt. komplette Remission
CT	Computertomographie
DEB-TACE	drug eluting beads-TACE
DICOM	Digital Imaging and Communications in Medicine
DVT	digitale Volumentomographie
FDA	Food and Drug Administration
HAP-Score	Hepatoma arterial-embolisation prognostic score
HBV	Hepatitis-B-Virus
HCC	hepatozelluläres Karzinom
HCV	Hepatitis-C-Virus
IBSI	Image Biomarker Standardization Initiative
ICC	intraclass correlation coefficient, dt. Intraklassen- Korrelationskoeffizient
INR	International normalized Ratio
KOF	Körperoberfläche
kV	Kilovolt
LI-RADS	Liver Imaging Reporting and Data System

LASSO	least absolute shrinkage and selection operator
mAs	Milliamperesekunden
MELD	Model of End-stage Liver Disease
mGy	Milligray
mHAP-II-Score	modified HAP-II-Score
mRECIST	modified Response Evaluation Criteria in Solid Tumors
MRT	Magnetresonanztomographie
MWA	Mikrowellenablation
NAFLD	non alcoholic fatty liver disease, dt nichtalkoholische Fettlebererkrankung,
NASH	nichtalkoholische Steatohepatitis
NTL	non-target-lesions
OS	Overall Survival, dt.Gesamtüberleben
PD	progressive disease
PD-L1	Programmed cell death 1 ligand 1
PR	partial response, dt. partielle Remission
RFA	Rekursive Feature Addition
RFE	Rekursive Feature Elimination
ROC	receiver operating characteristic, dt.Grenzwertoptimierungskurve
ROI	Region of Interest
RQS	radiomics quality score
SD	stable disease
SHARP	Sorafenib Hepatocellular Carcinoma Assessment Randomized Protocol
SIRT	selektive interne Radiotherapie
STATE-Score	Selection for transarterial chemoembolization treatment
TACE	transarterielle Chemoembolisation
TARE	transarterielle Radioembolisation
TIPS	transjugulärer hepatischer portosystemischer Shunt
TL	Target lesion, dt. Referenzläsion oder Zielläsion
VEGF	Vascular endothelial growth factor
VOI	Volume of Interest
vs.	versus, dt. im Gegensatz zu
z.B.	zum Beispiel

8.2. Abbildungs- und Tabellenverzeichnis

Abbildungen

Abbildung 1. Inzidenz in absoluten Zahlen, Entwicklung in Deutschland 1999-2018	6
Abbildung 2. Mortalität in absoluten Zahlen, Entwicklung in Deutschland 1999-2018	7
Abbildung 3. Risikofaktoren abhängig von der geographischen Verteilung	8
Abbildung 4. Sensitivität und Spezifität MRT vs. CT in Abhängigkeit der Herdgröße	10
Abbildung 5. Diagnostikalgorithmus HCC, nach S3-Leitlinie Diagnostik und Therapie des Hepatozellulären Karzinoms und biliärer Karzinome	11
Abbildung 6. LI-RADS Kategorien	12
Abbildung 7. Gewissheit für Gutartigkeit abhängig von der LI-RADS Kategorie	12
Abbildung 8. BCLC Algorithmus	13
Abbildung 9. Transplantationszahlen Deutschland 2020	14
Abbildung 10. Systematik der lokal ablativen Verfahren	15
Abbildung 11. Differenzierung vitaler Tumoranteile nach mRECIST	19
Abbildung 12. Bewertung des Gesamtansprechens anhand der mRECIST Kriterien	20
Abbildung 13. Child-Pugh Klassifikation	21
Abbildung 14. Dreimonatsmortalität in Abhängigkeit vom MELD-Score	23
Abbildung 15. HAP Score	24
Abbildung 16. mHAP-II Score	25
Abbildung 17. STATE Score	26
Abbildung 18. 6 and 12 Score	27
Abbildung 19. Workflow Radiomics	29
Abbildung 20. First-order Features	30
Abbildung 21. Second-order Features	31
Abbildung 22. Higher-order Features	31
Abbildung 23. STARD-Flowchart der Ein- und Ausschlusskriterien	37
Abbildung 24. Repräsentativer Therapieverlauf	39

Abbildung 25. Workflow einer repräsentativen Segmentierung	41
Abbildung 26. Darstellung der Intraklassen-Korrelationskoeffizienten (ICC)	43
Abbildung 27. ICC für jede einzelne Featureklasse	44
Abbildung 28. Heatmap der Pearsonkorrelationsanalyse	45
Abbildung 29. Workflow Entwicklung des Modells	48
Abbildung 30. Definition der besten Features der unterschiedlichen Gruppen abhängig vom Featureselektionsalgorithmus	50
Abbildung 31. t-SNE Plots	55
Abbildung 32. ROC-Kurven der Featureselektionsalgorithmen RFE, RFA und LASSO für Radiomics-Features	56
Abbildung 33. ROC-Kurven der Featureselektionsalgorithmen RFE, RFA und LASSO für klinische Features	57
Abbildung 34. ROC-Kurven für das beste kombinierte Modell auf Target Lesions Ebene	57
Abbildung 35. ROC-Kurven des kombinierten Featuresets, Performance des Modells bezogen auf das Gesamtansprechen des Kollektivs auf die TACE-Therapie	58
Abbildung 36. Kaplan-Meier OS-Kurven	60
Abbildung 37. Kaplan-Meier-Schätzung	61

Tabellen

Tabelle 1. Patientenkollektiv	34
Tabelle 2. Klinische Scores	46
Tabelle 3. Ergebnisse nach ML mit den besten Modellen auf Ebene der Target Lesions	51
Tabelle 4. Gesamtansprechen auf Ebene der TL und Gesamtansprechen auf Patientenebene	53

8.3 Anhang

A1. Radiomics quality score

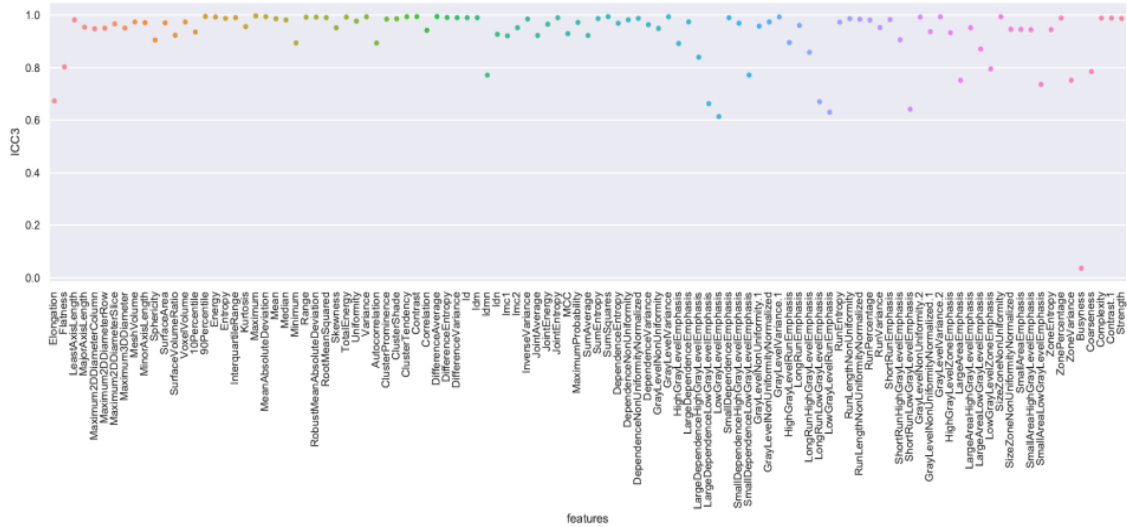
Image protocol quality - well-documented image protocols (for example, contrast, slice thickness, energy, etc.) and/or usage of public image protocols allow reproducibility/replicability	<input checked="" type="checkbox"/> protocols well documented <input type="checkbox"/> public protocol used <input type="checkbox"/> none
Multiple segmentations - possible actions are: segmentation by different physicians/algorithms/software, perturbing segmentations by (random) noise, segmentation at different breathing cycles. Analyse feature robustness to segmentation variabilities	<input checked="" type="radio"/> yes <input type="radio"/> no
Phantom study on all scanners - detect inter-scanner differences and vendor-dependent features. Analyse feature robustness to these sources of variability	<input type="radio"/> yes <input checked="" type="radio"/> no
Imaging at multiple time points - collect images of individuals at additional time points. Analyse feature robustness to temporal variabilities (for example, organ movement, organ expansion/shrinkage)	<input type="radio"/> yes <input checked="" type="radio"/> no
Feature reduction or adjustment for multiple testing - decreases the risk of overfitting. Overfitting is inevitable if the number of features exceeds the number of samples. Consider feature robustness when selecting features	<input checked="" type="radio"/> Either measure is implemented <input type="radio"/> Neither measure is implemented
Multivariable analysis with non radiomics features (for example, EGFR mutation) - is expected to provide a more holistic model. Permits correlating/inferencing between radiomics and non radiomics features	<input checked="" type="radio"/> yes <input type="radio"/> no
Detect and discuss biological correlates - demonstration of phenotypic differences (possibly associated with underlying gene-protein expression patterns) deepens understanding of radiomics and biology	<input type="radio"/> yes <input checked="" type="radio"/> no
Cut-off analyses - determine risk groups by either the median, a previously published cut-off or report a continuous risk variable. Reduces the risk of reporting overly optimistic results	<input type="radio"/> yes <input checked="" type="radio"/> no
Discrimination statistics - report discrimination statistics (for example, C-statistic, ROC curve, AUC) and their statistical significance (for example, p-values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)	<input checked="" type="checkbox"/> a discrimination statistic and its statistical significance are reported <input checked="" type="checkbox"/> a resampling method technique is also applied <input type="checkbox"/> none

Calibration statistics - report calibration statistics (for example, Calibration-in-the-large/slope, calibration plots) and their statistical significance (for example, P-values, confidence intervals). One can also apply resampling method (for example, bootstrapping, cross-validation)	<input type="checkbox"/> a calibration statistic and its statistical significance are reported <input type="checkbox"/> a resampling method technique is applied <input checked="" type="checkbox"/> none
Prospective study registered in a trial database - provides the highest level of evidence supporting the clinical validity and usefulness of the radiomics biomarker	<input type="radio"/> yes <input checked="" type="radio"/> no
Validation - the validation is performed without retraining and without adaptation of the cut-off value, provides crucial information with regard to credible clinical performance	<input type="checkbox"/> No validation <input checked="" type="checkbox"/> validation is based on a dataset from the same institute <input type="checkbox"/> validation is based on a dataset from another institute <input type="checkbox"/> validation is based on two datasets from two distinct institutes <input type="checkbox"/> the study validates a previously published signature <input type="checkbox"/> validation is based on three or more datasets from distinct institutes
Comparison to 'gold standard' - assess the extent to which the model agrees with/is superior to the current 'gold standard' method (for example, TNM-staging for survival prediction). This comparison shows the added value of radiomics	<input checked="" type="radio"/> yes <input type="radio"/> no
Potential clinical utility - report on the current and potential application of the model in a clinical setting (for example, decision curve analysis).	<input checked="" type="radio"/> yes <input type="radio"/> no
Cost-effectiveness analysis - report on the cost-effectiveness of the clinical application (for example, QALYs generated)	<input type="radio"/> yes <input checked="" type="radio"/> no
Open science and data - make code and data publicly available. Open science facilitates knowledge transfer and reproducibility of the study	<input type="checkbox"/> scans are open source <input type="checkbox"/> region of interest segmentations are open source <input type="checkbox"/> the code is open sourced <input type="checkbox"/> radiomics features are calculated on a set of representative ROIs and the calculated features and representative ROIs are open source

Total score **14** (38.89%)

Anhang A1. Radiomics Quality score, nach <https://radiomics.world/rqs>

A2. ICC-Analyse Swarmplot



Anhang A2. ICC-Analyse Swarmplot

A.3 Tabelle der ICC-Werte

class	features	ICC3
firstorder	Maximum	0.997
firstorder	90Percentile	0.995
gldm	DifferenceAverage	0.995
gldm	Contrast	0.995
gldm	SumSquares	0.995
firstorder	MeanAbsoluteDeviation	0.994
gldm	ClusterTendency	0.994
firstorder	Variance	0.994
gldm	GrayLevelVariance	0.994
glszm	SizeZoneNonUniformity	0.994
glszm	GrayLevelVariance.2	0.994
gldm	GrayLevelVariance.1	0.993
firstorder	Energy	0.993
firstorder	TotalEnergy	0.993
glszm	GrayLevelNonUniformity.2	0.993
firstorder	Range	0.993
firstorder	RobustMeanAbsoluteDeviation	0.993
gldm	DifferenceEntropy	0.992
firstorder	RootMeanSquared	0.991
gldm	DifferenceVariance	0.991
firstorder	InterquartileRange	0.991
gldm	Id	0.991
gldm	SmallDependenceEmphasis	0.991
gldm	Idm	0.991

glcm	JointEntropy	0.990
ngtdm	Contrast.1	0.989
glszm	ZonePercentage	0.989
ngtdm	Complexity	0.989
gldm	DependenceNonUniformityNormalized	0.988
firstorder	Entropy	0.988
ngtdm	Strength	0.988
glcm	SumEntropy	0.987
glrlm	RunLengthNonUniformity	0.987
firstorder	Mean	0.987
glcm	ClusterShade	0.987
glcm	InverseVariance	0.986
glcm	ClusterProminence	0.985
glrlm	RunLengthNonUniformityNormalized	0.985
glrlm	ShortRunEmphasis	0.983
gldm	DependenceNonUniformity	0.983
shape	LeastAxisLength	0.983
firstorder	Median	0.982
glrlm	RunPercentage	0.981
firstorder	Uniformity	0.978
gldm	LargeDependenceEmphasis	0.975
glrlm	GrayLevelNonUniformityNormalized	0.975
shape	VoxelVolume	0.974
shape	MeshVolume	0.974
glrlm	RunEntropy	0.974
glcm	MaximumProbability	0.973
shape	MinorAxisLength	0.972
shape	SurfaceArea	0.971
gldm	SmallDependenceHighGrayLevelEmphasis	0.970
gldm	DependenceEntropy	0.969
shape	Maximum2DDiameterSlice	0.967
glcm	JointEnergy	0.966
gldm	DependenceVariance	0.964
glrlm	LongRunEmphasis	0.961
glrlm	GrayLevelNonUniformity.1	0.959
firstorder	Kurtosis	0.957
shape	MajorAxisLength	0.955
glrlm	RunVariance	0.953
glcm	lmc2	0.953
glszm	LargeAreaHighGrayLevelEmphasis	0.952
firstorder	Skewness	0.952
shape	Maximum3DDiameter	0.951
shape	Maximum2DDiameterRow	0.951
gldm	GrayLevelNonUniformity	0.950
shape	Maximum2DDiameterColumn	0.948
glszm	SizeZoneNonUniformityNormalized	0.946
glszm	SmallAreaEmphasis	0.946

glszm	ZoneEntropy	0.945
glszm	SmallAreaHighGrayLevelEmphasis	0.945
glcm	Correlation	0.943
glszm	GrayLevelNonUniformityNormalized.1	0.937
firstorder	10Percentile	0.936
glszm	HighGrayLevelZoneEmphasis	0.934
glcm	MCC	0.930
glcm	Idn	0.927
shape	SurfaceVolumeRatio	0.924
glcm	JointAverage	0.923
glcm	SumAverage	0.923
glcm	Imc1	0.922
glrlm	ShortRunHighGrayLevelEmphasis	0.906
shape	Sphericity	0.906
glrlm	HighGrayLevelRunEmphasis	0.896
firstorder	Minimum	0.895
glcm	Autocorrelation	0.894
gldm	HighGrayLevelEmphasis	0.893
glszm	LargeAreaLowGrayLevelEmphasis	0.872
glrlm	LongRunHighGrayLevelEmphasis	0.859
gldm	LargeDependenceHighGrayLevelEmphasis	0.840
shape	Flatness	0.803
glszm	LowGrayLevelZoneEmphasis	0.795
ngtdm	Coarseness	0.786
gldm	SmallDependenceLowGrayLevelEmphasis	0.772
glcm	Idmn	0.772
glszm	ZoneVariance	0.752
glszm	LargeAreaEmphasis	0.752
glszm	SmallAreaLowGrayLevelEmphasis	0.737
shape	Elongation	0.674
glrlm	LongRunLowGrayLevelEmphasis	0.671
gldm	LargeDependenceLowGrayLevelEmphasis	0.663
glrlm	ShortRunLowGrayLevelEmphasis	0.642
glrlm	LowGrayLevelRunEmphasis	0.631
gldm	LowGrayLevelEmphasis	0.614
ngtdm	Busyness	0.037

Anhang A3. ICC für alle Radiomics Features. Die fett hervorgehobenen Features weisen einen ICC < 0,75 auf und wurden verworfen.

A.4 Auflistung der selektierten und verworfenen Features nach Pearson Analyse

Selektierte Features	Verworfenne Features
Nichtredundante Features	Redundante Features
<ul style="list-style-type: none"> • Flatness • LeastAxisLength • MajorAxisLength • MeshVolume • Sphericity • SurfaceArea • SurfaceVolumeRatio • 10Percentile • 90Percentile • Energy • Entropy • Kurtosis • Maximum • Minimum • Skewness • Uniformity • Variance • Autocorrelation • ClusterProminence • Contrast • Correlation • Idmn • Idn • Imc1 • Imc2 • InverseVariance • DependenceEntropy • DependenceNonUniformityNormalized • DependenceVariance • GrayLevelNonUniformity • LargeDependenceHighGrayLevelEmphasis • SmallDependenceLowGrayLevelEmphasis • LongRunHighGrayLevelEmphasis • RunEntropy • GrayLevelNonUniformity.2 [gldm] • GrayLevelNonUniformityNormalized.1 [glrlm] • GrayLevelVariance.2 [gldm] • HighGrayLevelZoneEmphasis • LargeAreaEmphasis • LargeAreaLowGrayLevelEmphasis • LowGrayLevelZoneEmphasis • SizeZoneNonUniformityNormalized • ZoneEntropy • Coarseness • Complexity • Contrast.1[ngtdm] • Strength. 	<ul style="list-style-type: none"> • Maximum2DDiameterColumn • Maximum2DDiameterRow • Maximum2DDiameterSlice • Maximum3DDiameter • MinorAxisLength • VoxelVolume; • InterquartileRange • MeanAbsoluteDeviation • Mean • Median • Range • RobustMeanAbsoluteDeviation • RootMeanSquared • TotalEnergy • ClusterShade • ClusterTendency • DifferenceAverage • DifferenceEntropy • DifferenceVariance • Id • Idm • JointAverage • JointEnergy • JointEntropy • MCC • MaximumProbability • SumAverage • SumEntropy • SumSquares • DependenceNonUniformity • GrayLevelVariance [gldm] • HighGrayLevelEmphasis • LargeDependenceEmphasis • SmallDependenceEmphasis • SmallDependenceHighGrayLevelEmphasis • GrayLevelNonUniformity.1 [glrlm] • GrayLevelNonUniformityNormalized [glrlm] • GrayLevelVariance.1 [glrlm] • HighGrayLevelRunEmphasis • LongRunEmphasis • RunLengthNonUniformity • RunLengthNonUniformityNormalized • RunPercentage • RunVariance • ShortRunEmphasis • ShortRunHighGrayLevelEmphasis • LargeAreaHighGrayLevelEmphasis • SizeZoneNonUniformity • SmallAreaEmphasis • SmallAreaHighGrayLevelEmphasis • ZonePercentage • ZoneVariance

Anhang A4. Auflistung der selektierten und verworfenen Features nach Pearsonanalyse

11. Schriftliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die dem Fachbereich Medizin der Goethe-Universität in Frankfurt am Main zur Promotionsprüfung eingereichte Dissertation mit dem Titel:

Lipiodol-Anreicherung als bildgebender Biomarker (Radiomics) zur Vorhersage des Therapieansprechens und Überlebens nach Transarterieller Chemoembolisation (TACE) bei Patienten mit hepatozellulärem Karzinom

im Zentrum der Radiologie, Institut für Diagnostische und Interventionelle Radiologie des Universitätsklinikums der Goethe-Universität Frankfurt am Main

unter Betreuung und Anleitung, sowie Unterstützung durch

Herrn Prof. Dr. med. Thomas J. Vogl ohne sonstige Hilfe selbst durchgeführt und bei der Abfassung der Arbeit keine anderen als die in der Dissertation angeführten Hilfsmittel benutzt habe. Darüber hinaus versichere ich, nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Ich habe bisher an keiner in- oder ausländischen Universität ein Gesuch um Zulassung zur Promotion eingereicht. Die vorliegende Arbeit wurde bisher nicht als Dissertation eingereicht.

Vorliegende Ergebnisse der Arbeit wurden in folgendem Publikationsorgan veröffentlicht:

Bernatz S, Elenberger O, Ackermann J, et al. CT - radiomics and clinical risk scores for response and overall survival prognostication in TACE HCC patients. *Sci Rep.* 2023;13:1-9. doi:10.1038/s41598-023-27714-0

(Ort, Datum)

(Unterschrift)