

Aus der
Radiologischen Universitätsklinik Tübingen
Abteilung Diagnostische und Interventionelle Radiologie

**Automatische Charakterisierung metabolischer Effekte
der onkologischen Immuntherapie mittels
18F-FDG-PET/MRT**

**Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Medizin**

**der Medizinischen Fakultät
der Eberhard Karls Universität
zu Tübingen**

vorgelegt von

Vu, Viet Chau

2023

Dekan: Professor Dr. B. Pichler

1. Berichterstatter: Professor Dr. S. Gatidis
2. Berichterstatter: Dr. T. Klag

Tag der Disputation: 13.01.2023

*gewidmet
meiner Familie*

Inhaltsverzeichnis

Inhaltsverzeichnis	4
Abbildungsverzeichnis	6
Tabellenverzeichnis	8
Abkürzungsverzeichnis	9
1 Einleitung	11
1.1 Automatisierung in der Bildgebung	11
1.2 Malignes Melanom	12
1.3 Onkologische Bildgebung	14
1.3.1 Magnetresonanztomographie (MRT)	14
1.3.2 Positronenemissionstomographie (PET)	16
1.3.3 Hybridbildgebung: PET/MRT	17
1.3.4 Multiparametrische Analyse lymphatischer Organe	19
1.4 Künstliche neuronale Netze	19
1.4.1 Convolutional Neural Networks	20
1.4.2 Deep Learning in der Radiologie	21
1.5 Zielsetzung	23
2 Material und Methoden	25
2.1 Patientenkollektiv	25
2.2 Bildgebung	25
2.3 DCNet	26
2.4 Automatische Segmentierungspipeline	28
2.3.1 Manuelle Segmentierung im 2D-Bild	29
2.4.1 Training	33
2.5 Evaluierung der Genauigkeit	33
2.5.1 Qualitative Auswertung	33

2.5.2	Quantitative Auswertung.....	34
2.6	Ableitung von PET/MRT-Parametern	36
2.7	Ableitung einer Therapieresponse.....	36
3	Ergebnisse	38
3.1	Patientenkollektiv	38
3.2	Manuelle Segmentierung	39
3.2.1	Dauer.....	39
3.2.2	3D-Masken	39
3.3	Automatische Segmentierung	40
3.3.1	Dauer.....	40
3.3.2	Visuelle Qualität.....	40
3.3.3	Genauigkeit	51
3.4	Abgeleitete PET/MRT-Parameter.....	52
3.5	Ableitung einer Therapieresponse	60
4	Diskussion.....	63
4.1	Zusammenfassung der Ergebnisse.....	63
4.2	Diskussion der Ergebnisse und Methoden	63
4.4	Limitationen.....	68
4.3	Herausforderungen im klinischen Einsatz von Deep Learning	69
4.5	Ausblick.....	71
5	Zusammenfassung.....	73
Anhang	75
6	Literaturverzeichnis	81
7	Erklärung zum Eigenanteil	92
8	Veröffentlichungen	93
Danksagung	94

Abbildungsverzeichnis

Abbildung 1: Aufbau von DCNet.....	28
Abbildung 2: Importierter DICOM-Datensatz im MITK-Interface nach Anlegen einer Segmentierung.	29
Abbildung 3: Manuelle Segmentierung einer Leber in der Wasser-Phase.	31
Abbildung 4: Manuelle Segmentierung einer Milz in der Wasser-Phase.	31
Abbildung 5: Manuelle Segmentierung der Wirbelkörper BWK 1-LWK 5 und Os sacrum in der Wasser-Phase.	32
Abbildung 6: Beispiele für Artefakte.	33
Abbildung 7: Manuell erstellte 3D-Organmasken von Leber, Milz und Wirbelkörper in MITK.	39
Abbildung 8: Beispiel einer einwandfreien automatischen Segmentierung der Leber im Transversalschnitt.	40
Abbildung 9: Beispiel einer einwandfreien automatischen Segmentierung der Leber im Koronalschnitt.	41
Abbildung 10: Beispiel einer einwandfreien automatischen Segmentierung der Leber im Sagittalschnitt.	41
Abbildung 11: Beispiel einer einwandfreien automatischen Segmentierung der Milz im Transversalschnitt.	42
Abbildung 12: Beispiel einer einwandfreien automatischen Segmentierung der Milz im Koronalschnitt und Sagittalschnitt.	42
Abbildung 13: Beispiel einer einwandfreien automatischen Segmentierung der Wirbelkörper im Sagittalschnitt.	43
Abbildung 14: Beispiel einer einwandfreien automatischen Segmentierung der Wirbelkörper im Koronalschnitt und Transversalschnitt.	43
Abbildung 15: Lokale FK der Leber in der automatischen Segmentierung.	45
Abbildung 16: Lokale FK der Milz in der automatischen Segmentierung.	46

Abbildung 17: Lokale FK der Wirbelkörper einschließlich Os sacrum in der automatischen Segmentierung.	46
Abbildung 18: Globale FK der Leber in der automatischen Segmentierung. ...	47
Abbildung 19: Globale FK der Milz in der automatischen Segmentierung.	47
Abbildung 20: Globale FK der Wirbelkörper einschließlich Os sacrum.	48
Abbildung 21: Sehr schlechte bis fehlende Organerkennung.	48
Abbildung 22: Anzahl der Aufnahmen nach visuellem Score.	49
Abbildung 23: Boxplot über Treffergenauigkeit (accuracy) der automatischen Segmentierungen.	51
Abbildung 24: Boxplot über dice score (DSC) der automatischen Segmentierungen.	51
Abbildung 25: Verteilung des automatisch abgeleiteten ADC gegen die ground truth.	52
Abbildung 26: Verteilung des automatisch abgeleiteten SUV gegen die ground truth.	54
Abbildung 27: Verteilung des automatisch abgeleiteten maximalen SUV gegen die ground truth.	55
Abbildung 28: Verteilung des automatisch abgeleiteten Fettgehalts der Organe gegen die ground truth.	57
Abbildung 29: Verteilung des automatisch abgeleiteten Organvolumens gegen die ground truth.	58
Abbildung 30: Zeitlicher Verlauf der mittleren SUV.	60
Abbildung 31: Differenzen des mittleren SUV nach 3 Monaten.	61
Abbildung 32: Boxplot über Sensitivität (sensitivity) der automatischen Segmentierungen, organweise.	75
Abbildung 33: Boxplot über Spezifität (specificity) der automatischen Segmentierungen, organweise.	75
Abbildung 34: Boxplot über positiven Vorhersagewert (precision) der automatischen Segmentierungen, organweise.	76

Tabellenverzeichnis

Tabelle 1: MRT-Aufnahmeparameter.	26
Tabelle 2: Visueller Score zur qualitativen Beurteilung der Organmasken.	34
Tabelle 3: Patientendaten.	38
Tabelle 4: Anzahl lokaler und globaler FK nach Ort.	50
Tabelle 5: Treffergenauigkeit (accuracy).	51
Tabelle 6: Dice score (DSC).	51
Tabelle 7: ADC, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson.	52
Tabelle 8: Durchschnittlicher ADC.	53
Tabelle 9: SUV, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson.	54
Tabelle 10: Durchschnittlicher SUV.	54
Tabelle 11: Maximaler SUV, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson.	55
Tabelle 12: Maximaler SUV.	56
Tabelle 13: Fettgehalt, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson.	57
Tabelle 14: Fettgehalt der Organe.	57
Tabelle 15: Organvolumen, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson.	58
Tabelle 16: Organvolumen.	59
Tabelle 17: SUV_{diff} , Shapiro-Wilk-Tests auf Normalverteilung.....	61
Tabelle 18: Sensitivität (sensitivity).	75
Tabelle 19: Spezifität (specificity).	75
Tabelle 20: Positive Vorhersagewerte (precision).	76
Tabelle 21: Beurteilung der Fehlklassifikationen, Vergabe des visual scores.	76

Abkürzungsverzeichnis

<i>18F</i>	Fluor-18
<i>2D</i>	zweidimensional
<i>3D</i>	dreidimensional
<i>Abb.</i>	Abbildung
<i>abs.</i>	absolut
<i>ADC</i>	apparent diffusion coefficient
<i>AJCC</i>	American Joint Committee on Cancer
<i>BMI</i>	body mass index
<i>BRAF</i>	B-rapidly accelerated fibrosarcoma
<i>BWK</i>	Brustwirbelkörper
<i>CI</i>	Konfidenzintervall
<i>CNN</i>	Convolutional Neural Network
<i>CR</i>	Complete response
<i>CT</i>	Computertomographie
<i>CTLA-4</i>	Cytotoxic T-Lymphocyte Antigen 4
<i>Diff</i>	Differenz
<i>DL</i>	Deep Learning
<i>DSC</i>	Dice similarity coefficient
<i>DWI</i>	Diffusion weighted imaging
<i>FDG</i>	Fluordesoxyglucose, 2-Fluor-2-Desoxy-D-Glucose
<i>FK</i>	Fehlklassifikation
<i>GK</i>	Ganzkörper
<i>h</i>	Stunde
<i>HWK</i>	Halswirbelkörper
<i>kg</i>	Kilogramm
<i>LWK</i>	Lendenwirbelkörper
<i>MEK</i>	Mitogen-activated protein kinase kinase (= MAP2K)
<i>M</i>	Mittelwert
<i>min</i>	Minute
<i>mm</i>	Millimeter
<i>mmol</i>	Millimol
<i>MRT</i>	Magnetresonanztomographie
<i>mSv</i>	Millisievert
<i>MTV</i>	Metabolisch aktives Tumolvolumen (= MATV)
<i>NIFTI</i>	Neuroimaging Informatics Technology Initiative
<i>NRAS</i>	Neuroblastoma RAS
<i>OSEM</i>	Ordered subset expectation maximization
<i>PD-1</i>	Programmed cell death protein 1
<i>PD</i>	Progressive disease
<i>PET</i>	Positronen-Emissions-Tomographie
<i>PERCIST</i>	Positron Emission Response Criteria in Solid Tumors

<i>PR</i>	Partial response
<i>PWI</i>	Perfusion weighted imaging
<i>RECIST</i>	Response Evaluation Criteria In Solid Tumors
<i>rel.</i>	relativ
<i>ReLU</i>	Rectified linear unit
<i>ROI</i>	Region of interest
<i>s</i>	Sekunde
<i>SD</i>	standard deviation, Standardabweichung
<i>SD</i>	stable disease
<i>SPECT</i>	Single Photon Emission Computed Tomography
<i>SUV</i>	Standardized uptake value
<i>SW</i>	Shapiro-Wilk
<i>T1w</i>	T1-gewichtet
<i>T2w</i>	T2-gewichtet
<i>TLG</i>	Total lesion glycolysis

1 Einleitung

1.1 Automatisierung in der Bildgebung

In den letzten Jahrzehnten ist die medizinische Bildgebung in der Diagnostik unabdingbar geworden und ist seither fester Bestandteil vieler onkologischer Leitlinien. Der stetige Zuwachs an neuen Technologien in Soft- und Hardware ermöglicht eine signifikant höhere Bildqualität und kürzere Untersuchungszeiten (Petralia et al., 2019).

Im Gegensatz zu numerischen Labortests, bei denen durch konsensuale Referenzwerte leicht ein Befund zu erstellen ist, benötigen generierte Bilddaten eine ausführliche Analyse. Bei der multiparametrischen Bildgebung, welche mehrere unterschiedliche Bilddaten innerhalb einer einzelnen Untersuchung erfasst, ist die Auswertung häufig von sehr hohem Zeitaufwand, Personalbedarf, Subjektivität und Missinterpretation geprägt. Auch bei großen epidemiologischen Studien mit zahlreichen Patienten erschweren diese Mängel ein repräsentatives Ergebnis (Küstner et al., 2018). Eine Automatisierung der Auswertung sichert sowohl Objektivität als auch Reproduzierbarkeit.

Vorige Arbeiten zeigten gute Ergebnisse in der automatisierten medizinischen Bildanalyse. Die semantische Segmentierung von Organen, also eine Klassifizierung von Raumpunkten (Voxel) in sich nicht-überlappende Regionen anhand verschiedener Kriterien, bildet hierbei den Grundbaustein zur weiteren Untersuchung organspezifischer Pathologien (Küstner et al., 2018). Um dieses Ziel zu erreichen, hat sich methodisch die Nutzung des *Deep Learnings*, einer Domäne maschinellen Lernens mittels neuronaler Netze, zur Bildanalyse etabliert (Litjens et al., 2017). Die Einteilung der Voxel in mehrere Klassen, wie beispielsweise Organ und Hintergrund, erzeugt eine repräsentative Abbildung und bewirkt durch die Definition eines Suchbereichs eine einfachere und gezieltere weitere Analyse (Anwar et al., 2018, Küstner et al., 2020). Die vorliegende Arbeit behandelt die Anwendung und Validierung eines neuronalen Netzes zur automatischen Beurteilung von metabolischen Veränderungen bei Melanompatienten unter Immuntherapie.

1.2 Malignes Melanom

Beim malignen Melanom handelt es sich um einen hochmalignen, von Melanozyten ausgehenden Tumor, dessen Inzidenz in der westlichen Population in den letzten Jahren und Jahrzehnten stark gestiegen ist (Siegel et al., 2020, Sacchetto et al., 2018). Das mittlere Erkrankungsalter beträgt 57 Jahre, wobei primär Individuen mit hellem Hauttyp betroffen sind. Das maligne Melanom entspricht in den USA der sechsthäufigsten Krebserkrankung bei Frauen, sowie der fünfhäufigsten bei Männern (Carr et al., 2020). Die Neigung zu früher lymphogener und hämatogener Metastasierung erklärt die schlechte Prognose dieser Tumorentität (Pavri et al., 2016), wobei bereits der Nachweis von Mikrometastasen sowie die steigende Tumorlast in Sentinellymphknoten mit einer signifikant schlechteren Prognose einhergehen (Balch et al., 2010, van der Ploeg et al., 2009). Die häufigsten Orte der Fernmetastasierung bilden Haut, Knochen, Leber, ZNS und Lunge (Misir et al., 2016, Kozovska et al., 2016). Seltener können auch Metastasen im Gastrointestinaltrakt, in der Nebenniere oder in der Milz auftreten (Trindade et al., 2009, Ollila et al., 1996, Branum et al., 1991). Die fortgeschrittenen Stadien führen je nach Ort und Tumorlast der Metastasen zu einer äußerst schlechten Prognose: Bei Metastasierung in das zentrale Nervensystem beträgt die mediane Überlebensdauer 2,5 Monate nach Diagnosestellung und ist zwölf Mal schlechter als bei extraregionaler Lymphknoten- oder subkutaner Metastase (Sandru et al., 2014). Unter den verschiedenen Hautkrebstypen ist das maligne Melanom mit 3 % Anteil für 75 % der Todesfälle verantwortlich (Sladden et al., 2009, Pavri et al., 2016).

Die klinische Stadieneinteilung des Tumors erfolgt aktuell nach AJCC 2016 in Stadium I bis IV, wobei Stadium IV durch das Vorhandensein von Fernmetastasen (M1) definiert ist. In Stadium III und IV hat sich in der Ausbreitungsdiagnostik eine Überlegenheit der PET/CT gegenüber anderen Bildgebungsverfahren gezeigt (Xing et al., 2011). Bei kurativer Intention sollte therapeutisch bei jedem Stadium die radikale Exzision des Primarius zur Vermeidung lokaler Rezidive erfolgen (Sladden et al., 2009).

Neben der konventionellen Chemotherapie stellen heute sowohl Immuncheckpoint-Inhibitoren als auch Proteinkinase-Inhibitoren eine wichtige

Domäne in der medikamentösen Therapie des Melanoms dar. Mit dem monoklonalen Antikörper *Ipilimumab* als CTLA-4-Inhibitor wurde in der EU 2011 das erste Immuntherapeutikum zur Behandlung bei Inoperabilität zugelassen. 2015 erfolgte die Zulassung für den PD-1-Inhibitor *Nivolumab*, welches für ein besseres Outcome mit *Ipilimumab* kombiniert werden sollte (Wolchok et al., 2017). Der im gleichen Jahr zugelassene PD-1-Inhibitor *Pembrolizumab* stellt in Monotherapie eine weniger toxische Alternative dar (Robert et al., 2015), und findet genauso wie *Nivolumab* nach vollständiger Resektion auch als Adjuvans Verwendung. Die Ansprechrate dieser Immuntherapeutika beträgt etwa 60 % bei kombinierter anti-CTLA-4- und anti-PD-1-Therapie bei BRAF-Wildtyp, sowie etwa 40 % bei anti-PD1-Monotherapie (Larkin et al., 2015). Als klinisch wichtige Nebenwirkungen der neuen Antikörpertherapien sind schwere Diarrhoen und Hepatitiden als Autoimmunreaktionen zu nennen, die letal verlaufen können (Boutros et al., 2016).

Die BRAF-Inhibitoren erhielten mit *Vemurafenib* 2012 und *Dabrafenib* 2013 die EU-Zulassung. Ihr Einsatz erfolgt bei onkogener BRAF-V600E/K-Mutation, welche bei etwa 40 % der Melanompatienten auftritt, sowie bei NRAS-Mutation, welche bei etwa 15 % auftritt (Jakob et al., 2012, Long et al., 2017b). In Kombination mit MEK-Inhibitoren, wie z.B. *Trametinib*, wird eine höhere Überlebensrate und Verträglichkeit mit einer Ansprechrate von etwa 70 % erreicht (Ascierto et al., 2016, Long et al., 2017a). Zu den häufigsten Nebenwirkungen dieser Substanzklassen zählen leichte bis mittelgradige Dermatotoxizitäten sowie Ödembildung, Diarrhoen, Übelkeit und Erbrechen (Ascierto et al., 2013).

Siegel et al. beschreiben einen signifikanten Rückgang der Mortalität von jährlich 7 % in den letzten vergangenen Jahren, welchen sie auf neue Therapiemöglichkeiten zurückführen (Siegel et al., 2020). Zwar wird angenommen, dass das Hautkrebsscreening als Früherkennungsmaßnahme ebenso zur Mortalitätssenkung beiträgt (Rastrelli et al., 2014), jedoch konnte dieser Effekt bisher retrospektiv in randomisierten kontrollierten Studien nicht nachgewiesen werden (Johansson et al., 2019).

1.3 Onkologische Bildgebung

1.3.1 Magnetresonanztomographie (MRT)

Die MRT ist ein bildgebendes Verfahren, das auf ionisierende Strahlung verzichtet. Dabei macht sie sich das Prinzip der Kernspinresonanz zunutze. Durch den Einsatz eines statischen Magnetfeldes wird der Spin der Atomkerne (v.a. Wasserstoff) im Körper parallel und antiparallel ausgerichtet. Ein Hochfrequenzimpuls lässt die Protonen in der zu untersuchenden Ebene um die Längsachse des Magnetfeldes präzedieren (sog. Larmorpräzession), wodurch eine Quermagnetisierung entsteht. Die Abschaltung des Impulses bewirkt eine longitudinale und transversale Relaxation der Protonen. Die Änderungen der Magnetfelder können über Detektoren als Signal ausgewertet werden, das über eine Fourier-Transformation zu einem Bild verarbeitet wird. Durch Änderung der Systemparameter, wie beispielsweise Repetitionszeit (TR), Echozeit (TE), Inversionszeit (TI) oder Flipwinkel, können verschiedene Wichtungen und Sequenzen erzeugt werden (Lugauer and Wetzl, 2018).

Der hohe Weichteilkontrast, die gute räumliche Auflösung und die fehlende Strahlenbelastung machen die MRT zu einer ausgezeichneten Untersuchungsmethode für Tumorpatienten, die ohnehin einer hohen Untersuchungsfrequenz unterliegen (Morone et al., 2017). Weiterhin können über spezielle Sequenzen funktionelle Parameter wie Diffusion (DWI) und Perfusion (PWI) beurteilt werden. Durch den sog. ADC-Wert (*apparent diffusion coefficient*) können Molekularbewegungen in der diffusionsgewichteten Bildgebung berechnet und quantifiziert werden (Hilario et al., 2012). Somit kann dieser Wert einen prognostischen (Akkus Yildirim et al., 2019) oder diagnostischen (Filograna et al., 2018) Faktor bei onkologischen Therapien darstellen. Die DWI als zusätzliche Sequenz bietet hierbei klinisch wichtige Informationen bei onkologischen Patienten und verbessert maßgebend die Diagnostik, sodass etwa 40 % mehr Tumoren in der DWI-Sequenz – im Vergleich zu einer konventionellen MR-Aufnahme – erkannt werden können (Low and Gurney, 2007, Morone et al., 2017, Petralia et al., 2019). Die Magnetresonanztomographie findet durch Erfassung von

Stoffwechselprodukten und Aufdeckung von Tumorstoffwechseln ebenfalls vielfach diagnostische Anwendung in der Entdeckung und Charakterisierung von Neoplasien (Golder, 2004, Rapalino and Ratai, 2016).

Spezielle Sequenzen, wie beispielsweise die Dixon-Sequenz, können genutzt werden, um Bildartefakte zu mindern. Diese auf *chemical shift* basierende Methode macht sich die unterschiedliche Anregungsfrequenz von Protonen in Wasser und Fett (bzw. Methylen und Methylgruppen der Triglyzeride) zunutze, da deren magnetischer Summenvektor zwischen einer sogenannten in-Phase (bei gleichem Summenvektor) und einer opposed-Phase (bei um 180° gedrehtem Summenvektor, auch: out-of-phase) schwingt (Ma, 2008, Dixon, 1984). Durch Addition bzw. Subtraktion der in diesen Phasen aufgenommenen Bilder ergibt sich die sogenannte Wasser- bzw. Fett-Phase (Leyendecker et al., 2010), sodass durch eine durchgeführte Sequenz vier klinisch verwertbare Bildserien entstehen. Obwohl die Kontrastmittel-gestützte Computertomographie (CT) heute immer noch als Standarduntersuchung zur Detektion von Fernmetastasen beim malignen Melanom eingesetzt wird, bieten Ganzkörper (GK)-MRT und vor allem PET/CT diagnostisch eine signifikant höhere Genauigkeit (Pfannenbergl et al., 2007, Muller-Horvat et al., 2006).

Neben der Detektion von Tumorherden spielt die onkologische Bildgebung jedoch noch weitere Rollen: Durch die potenziell letalen Nebenwirkungen der neuen Immuntherapien, ergibt sich die Notwendigkeit einer Beurteilung der Therapieeffektivität. Bei fehlender Wirksamkeit können durch Therapieabbruch schwere unerwünschte Arzneimittelwirkungen und Folgekosten vermieden werden. Zur Objektivierbarkeit eines Therapieansprechens bei Tumorerkrankungen wurden die RECIST-Kriterien (*response evaluation criteria in solid tumors*) festgelegt, welche eine komplette Remission (*CR, complete response*), eine partielle Remission (*PR, partial response*), eine stabile Erkrankung (*SD, stable disease*), und eine Progression (*PD, progressive disease*) definieren. Zur Einordnung ist unter anderem die Größenverlaufsbestimmung messbarer Läsionen erforderlich. Zusätzlich sind je nach Tumor die Bestimmung von Tumormarkern, sowie die Einbindung von Zyto- und Histologie nötig (Eisenhauer et al., 2009).

Demgegenüber stehen die funktionellen PERCIST-Kriterien (positron emission tomography response criteria in solid tumors), welche metabolische Therapieansprachen über die Positronenemissionstomographie definieren sollen.

1.3.2 Positronenemissionstomographie (PET)

Die aus der Nuklearmedizin stammende PET ist in der Lage, molekulare Prozesse in Schichtbildern darzustellen.

Da Veränderungen auf Molekularebene in der Regel früher als die morphologischen Änderungen auftreten, können auch metabolische Effekte, wie z.B. ein Therapieansprechen, früher erkannt werden (Belmouhand et al., 2019). Für die Untersuchung werden spezielle mit Radionukliden markierte Pharmaka, sogenannte *Tracer*, verwendet. Eine wichtige Rolle spielt die radioaktiv markierte ¹⁸F-Fluordesoxyglucose (¹⁸F-FDG), welche wie die unmarkierte Glucose von Körperzellen verstoffwechselt wird. Das Radioisotop ist ein β⁺-Strahler und emittiert Positronen. Nach Zurücklegen einer kurzen Strecke, annihiliert ein Positron ein Elektron im Gewebe, wodurch es zur sogenannten Vernichtungsstrahlung kommt. Die zwei entstehenden Gammaphotonen, die in einem Winkel von 180° emittiert werden, können von den Detektorzellen im PET-Scanner erfasst werden, um so den Zerfallsort bzw. Vernichtungsort zu bestimmen (Beuthien-Baumann, 2018). Da Malignome in der Regel einen erhöhten Glukoseverbrauch aufweisen, können diese beispielsweise mithilfe der ¹⁸F-FDG-PET lokalisiert werden. Durch das Fehlen präziser anatomischer Orientierungspunkte oder die Glukoseaufnahme nicht-maligner Zellen wird jedoch die Spezifität der PET durch falsch-positive oder mehrdeutige Ergebnisse deutlich beeinträchtigt (Israel and Kuten, 2007).

Diagnostisch wegweisend zeigten sich Parameter wie der SUV (*standardized uptake value*), der Auskunft über die quantitative Traceraufnahme gibt und schon früh aufgrund seiner Abhängigkeit von Tracermenge, Serumglukose (bei FDG-PET) und Körperoberfläche (Kim et al., 1994) in Kritik geraten ist.

Ein weiterer volumetrischer Parameter stellt das *metabolisch aktive Tumervolumen* (MTV oder MATV) dar, das sich auf die FDG-anreichernde

Tumormenge bezieht, die einen SUV-Schwellenwert überschreitet. Die sogenannte *total lesion glycolysis* (TLG) definiert das Produkt aus mittlerer SUV und MTV. Sowohl TLG als auch MTV finden heutzutage Anwendung bei verschiedenen Tumorentitäten und können gute Aussagen zu Therapieansprechen treffen (Bauckneht et al., 2021, Rogasch et al., 2018, Im et al., 2018). Beim malignen Melanom dienen diese Parameter als starke unabhängige Prognosefaktoren für das Gesamtüberleben (Reinert et al., 2020).

Bei kontrollierten Protokollen stellt auch der SUV trotz der noch heute bestehenden Limitationen einen guten Prognosemarker bei vielen Tumorentitäten dar (Anzai, 2020). Die PERCIST-Kriterien beispielsweise, welche ein metabolisches Therapieansprechen definieren sollen, orientieren sich am Verlauf der sogenannten *SUL*, welche einer an der fettfreien Körpermasse korrigierten SUV entspricht (Tirkes et al., 2013). Daneben wurden in der Literatur seit einigen Jahren weitere verbesserte bzw. auf Immuntherapie spezialisierte Schemata, wie PECRIT, PERCIMT, imPERCIST oder iPERCIST erörtert, deren Kriterien sich ebenfalls an der SUL-Dynamik orientieren. Durch geänderte Therapieregimes mit den neuen Wirkungsweisen der Checkpoint-Inhibitoren beschrieben die betreffenden Autoren die Notwendigkeit einer Umstellung der Interpretation durch die neuen Evaluationskriterien. Dies sei vor allem der Fall, wenn die neuen Agenzien statt einer Tumorverkleinerung eine Krankheitsstabilisierung anstrebten (Decazes and Bohn, 2020, Amrane et al., 2019). Die neuen auf PET-basierenden Schemata, vor allem imPERCIST, wiesen in Bezug auf die Voraussage eines progressfreien Gesamtüberlebens eine signifikante Überlegenheit gegenüber konventionellen CT-Kriterien auf (Kitajima et al., 2022).

1.3.3 Hybridbildgebung: PET/MRT

Die PET/MRT-Untersuchung stellt ähnlich wie andere Hybridverfahren, wie PET/CT oder SPECT/CT, die Kombination einer molekularen und morphologischen Analyse in einer Untersuchung dar. So vereinen die Geräte die

Vorteile einer funktionellen und anatomischen Bildgebung mit der Möglichkeit zur Datenbeurteilung während des exakt gleichen pathophysiologischen Zustands. Von den seit 2010 kommerziell erhältlichen PET/MRT-Scannern gibt es weltweit etwa 150 Geräte, die vor allem in der Onkologie ihren Einsatz finden. Dazu gehören beispielsweise die Detektion und das Staging von Tumoren, die Planung einer Strahlentherapie oder Biopsie, oder die Abklärung bei Ungewissheit in anderen Bildgebungsmodalitäten (Umutlu et al., 2019).

Ebenso eine große Rolle spielt die Therapiekontrolle. Von großer Bedeutung ist die frühe Beurteilung des Therapieeffekts, sodass unnötig lange, als auch wirkungslose Therapien schnell erkannt werden können (Kovacs et al., 2018).

Im Vergleich zur PET/CT schneidet die PET/MRT bei der Diagnostik von Tumoren in den meisten Fällen gleich gut oder besser ab (Riola-Parada et al., 2016). Ein großer Vorteil der PET/MRT ist die Bestimmung multiparametrischer Daten bei gleichzeitig kleiner Strahlungsexposition: Im Vergleich zur PET/CT kann eine Dosisreduktion von ca. 80 % bei $3,6 \pm 1,4$ mSv pro GK-PET/MRT-Untersuchung erreicht werden (Martin et al., 2020). Infolgedessen stellte diese Untersuchung in den letzten Jahren auch eine Schlüsselkomponente in der Kinderonkologie dar (Gatidis et al., 2017), da neben strahlenarmer Bildgebung bei gleicher Bildqualität auch weniger Tracer gegeben werden kann (Ehman et al., 2017). Neben den zahlreichen MRT-Parametern sind durch die lange Untersuchungszeit zudem weitere Variablen, wie beispielsweise die Dynamik der Traceraktivität beurteilbar. So konnte bei Brust- und Darmkrebspatienten durch eine dynamische PET-Untersuchung quantitativ das Ansprechen auf eine Chemotherapie vorhergesagt werden (Miles et al., 2018). Weiterhin ermöglicht die simultane MR-Aufnahme eine Bewegungskorrektur (Catana, 2015), sowie eine Partialvolumenkorrektur der PET-Bilder (Belzunce et al., 2019). Da die Mehrkosten einer PET/MRT im Vergleich zu einer PET/CT durch Geräte, Untersuchungszeit und Zusatzkenntnisse derzeit etwa 4.000 USD pro Untersuchung (Miles et al., 2018) betragen und noch nicht von den hiesigen Krankenkassen übernommen werden, ist die Notwendigkeit von weiteren Studien und einer Standardisierung von Arbeitsabläufen und Untersuchungsprotokollen gegeben (Broski et al., 2018, Umutlu et al., 2019, Miles et al., 2018).

1.3.4 Multiparametrische Analyse lymphatischer Organe

Im lymphatischen System werden die primären lymphatischen Organe, wie Knochenmark und Thymus, von den sekundären lymphatischen Organen, wie der Milz, den Lymphknoten und den Lymphfollikeln unterschieden.

Es ist bekannt, dass das Knochenmark des Achsenskeletts unter Immuntherapie aktiviert wird, und in ihm ein erhöhter Glukosemetabolismus nachgewiesen werden kann (Schwenck et al., 2020), welcher in der Hybridbildgebung über den SUV quantifizierbar ist. Auch in der Milz fanden sich Hinweise auf einen erhöhten Glukoseuptake unter Immuntherapie, jedoch fanden Schwenck et al. keine Signifikanz zwischen Respondern und Non-Respondern der Immuntherapie bei Patienten mit malignem Melanom (Schwenck et al., 2020).

Schraml et al. zeigten Zusammenhänge bei der multiparametrischen Analyse der PET/MRT auf, die bei der Beurteilung von lymphatischen Organen bei onkologischen Patienten von Nutzen sein können. Unter anderem ließ sich die metabolische Aktivität des Knochenmarks direkt invers mit dessen Fettgehalt korrelieren. Die über dem Knochenmark abgeleiteten ADC-Werte wiesen ebenfalls eine inverse Korrelation mit dem Fettgehalt auf. Die erhöhte metabolische Aktivität und Diffusion des fettarmen Knochenmarks sei womöglich einer aktiveren Hämatopoese im roten Knochenmark gegenüber dem nicht-blutbildenden fettreicheren gelben Knochenmark zuzuschreiben (Schraml et al., 2015). Dieses Wissen um die Physiologie der Organe bietet eine Grundlage für nachfolgende Analysen.

1.4 Künstliche neuronale Netze

Unter künstlichen neuronalen Netzen versteht man eine Herangehensweise für die Implementierung maschinellen Lernens. Sie sind von der Funktionsweise eines tierischen Hirns inspiriert (Fukushima, 1980) und bestehen aus vielschichtig miteinander verbundenen künstlichen Neuronen. Diese sind in der Lage, erhaltene Informationen zu gewichten, um sie an folgende Neurone

weiterzuleiten und eine Mustererkennung durchzuführen (Michelucci, 2019, Choy et al., 2018).

Bereits in den 1970er Jahren begann mit der Computer-Ära die automatisierte medizinische Bildanalyse durch Randerkennung und Segmentierung mittels *region growing* (Litjens et al., 2017). Ende der 1990er Jahre kam es zu einem Zuwachs an *überwachten* (engl.: supervised) *Lernalgorithmen* (Litjens et al., 2017), welche vor ihrem Einsatz mit Aufgaben trainiert werden müssen, deren Ergebnisse bekannt sind und vorgegeben werden. Im Gegensatz zu regelbasierten Algorithmen können künstliche neuronale Netze noch unbekannte Trainingsdaten nutzen, um Gesetzmäßigkeiten zu erlernen, sich anzupassen, und zu verbessern (Choy et al., 2018, Maleki et al., 2020). Noch heute bildet diese Art der Mustererkennung die Grundlage vieler etablierter *machine learning*-Systeme in der medizinischen Bildgebung (Yasaka et al., 2018, Litjens et al., 2017).

1.4.1 Convolutional Neural Networks

Besonders *convolutional neural networks* (*CNN* oder *ConvNet*, dt.: Faltende neuronale Netzwerke), die der Domäne des sogenannten *Deep Learning* (DL) angehören, finden heute aufgrund guter Ergebnisse bei der Bilderkennung in der Radiologie ihre Anwendung (Yamashita et al., 2018, Anwar et al., 2018). Diese sind in der Lage durch verschiedene Arten von neuronalen Schichten, sog. *convolution layers*, *pooling layers* und *fully connected layers*, nach dem Training die räumliche Hierarchie von Merkmalen erlernen. Ein biologisches Vorbild bilden die optisch rezeptiven Felder, bei denen ein visueller Bereich bzw. eine Gruppe von Sinnesrezeptoren auf ein individuelles Neuron geschaltet werden. Convolution layers dienen primär dem Auslesen von Merkmalen wie Kanten und bestehen mathematisch aus linearen und nicht-linearen Operatoren, meist aus einer diskreten Faltung (*engl. convolution*) und einer Aktivierungsfunktion (*ReLU*). Das Auslesen geschieht mithilfe eines oder mehrerer Kernels, welche strukturiert die diskrete Faltung eines Bildbereiches, genannt *feature map*, mit einer kleinen Matrix einer definierten Größe durchführen. Mit jedem convolutional layer wird in der Regel eine feature map mit anderer Größe erstellt. Mithilfe von

pooling layers wird ein sogenanntes lineares *downsampling* durchgeführt, welches die Größe der feature maps reduziert, die an die folgenden Layers weitergegeben werden (Yamashita et al., 2018). Dies bietet den Vorteil, die Rechenleistung zu reduzieren, jedoch auch dominante Merkmale hervorzuheben. Dadurch wird unter anderem ein *Overfitting* reduziert, ein Szenario, in dem Zusammenhänge in den Trainingsdaten zwar gut erschlossen werden, eine Generalisierbarkeit auf unabhängige Testdaten aufgrund von Überanpassung jedoch nicht vorhanden ist (Choy et al., 2018). Es können mehrere convolutional und pooling layers hintereinandergeschaltet werden, sodass man von einer zunehmenden Tiefe des Netzes spricht. Unter ebenso zunehmender benötigter Rechenleistung wird hierdurch eine detailliertere und tiefgründigere Abbildung der Daten ermöglicht.

Die feature map aus dem letzten convolution bzw. pooling layer wird in eine eindimensionale Matrix transformiert und kann einem oder mehreren fully connected layers zugeführt werden. Über eine letzte nicht-lineare Aktivierungsfunktion werden die Inputs aus den vorigen Layers hier den verschiedenen Klassen (beispielsweise Organ/Nicht-Organ) über eine spezifische Gewichtung zugeordnet (Yamashita et al., 2018).

Das Training eines Netzwerks beschreibt das Anpassen der Kernels in den convolutional layers und die Gewichtung in den eventuell vorhandenen fully connected layers, sodass eine möglichst kleine Diskrepanz zwischen dem Output und den Referenzwerten aus den Trainingsdaten entsteht. Dieser Vorgang wird *backpropagation* genannt (Yamashita et al., 2018).

1.4.2 Deep Learning in der Radiologie

Der große Erfolg der Vorstellung des CNNs namens *AlexNet* von Krizhevsky et al. bei der ImageNet Large Scale Visual Recognition Competition (ILSVRC) (Krizhevsky et al., 2012) war bedeutend für die Entwicklung von DL und das Interesse an CNNs in der Bilderkennung (Yasaka et al., 2018, Yamashita et al., 2018). Vorige Arbeiten zeigten bereits bei ausreichend vorhandenen Trainingsdaten gute und akkuratere Ergebnisse bei zunehmender Netztiefe (He et al., 2016, Sun et al., 2017, Chan et al., 2020). In den letzten Jahren zeigten

sich Erfolge in den Aufgabenbereichen der Erkennung, Segmentierung, Klassifikation und Voraussage (Montagnon et al., 2020). Da die Effizienz und Limitation von computergestützter Diagnosesoftware sowohl von den mathematischen Grundzügen eines Netzes zur numerischen Beschreibung von Bildern, als auch von der Qualität und der Anzahl extrahierter Bildmerkmale zum Training eines Netzes abhängt (Chan et al., 2020), ist eine enge multidisziplinäre Zusammenarbeit zwischen Ingenieuren und Medizinern notwendig, um gute Ergebnisse zu erzielen.

In der Diagnostik wurde DL bereits auf viele Bereiche übertragen. Große Anwendung findet sich in der Onkologie, bei der Erkennung und Klassifizierung von Tumoren, deren Phänotyp und deren Malignitätsstatus. Weiterführende Studien beschreiben die Nutzung zur Voraussage über Therapieansprechen, Prognosebestimmung und Risikoanalyse. Trotz großem Forschungsinteresse und rapid steigender Anzahl an Publikationen, werden diese Systeme in der klinischen Routine noch kaum angewandt. Grund dafür könnte eine fehlende Arbeitseffizienz bei noch fehlendem Workflow und Standard sein. Dennoch bieten die automatisierten DL-Systeme aufgrund der Möglichkeit des schnellen Lernens von Millionen von Datensätzen große Generalisierbarkeit und Stabilität bei Fällen von großer Variabilität (Chan et al., 2020).

Bei besonders zeitaufwändigen Aufgaben, wie der semantischen Segmentierung von Organen, die von hohem Zeitaufwand und Ressourcen, jedoch auch Subjektivität und Missinterpretation geprägt ist, kann ein deutlicher Vorteil in der Automatisierung gesehen werden. Dahingehend wurden in CT und MRT in den letzten Jahren bereits sehr gute Segmentierungsergebnisse mit hoher Genauigkeit in verschiedenen Körperregionen mittels DL erzielt (Isensee et al., 2021, Chen et al., 2020, Kim et al., 2020, van Harten et al., 2019, Küstner et al., 2018). Die Datenlage in allen anderen Bildgebungsmodalitäten ist hingegen noch überschaubar.

Über die Bestimmung und Voraussage einer Therapieresponse könnten schwere Nebenwirkungen und hohe Folgekosten reduziert werden. Eine zentrale Ansatzstelle aktueller Forschung sind die *Radiomics*, ein Gebiet mit der Zielsetzung, mittels maschinellen Lernens mehrdimensionale Features aus

radiologischen Bilddaten zu quantifizieren (Gillies et al., 2016). Einige Algorithmen wurden bereits trainiert, um in CT-Bildern non-invasive radiographische Marker zu charakterisieren, um ein Therapieansprechen unter Immuntherapie voraussagen (Trebeschi et al., 2019, Ligerio et al., 2021). Ein anderer Ansatz von Sun et al. bezog mit dem gleichen Ziel genomische Daten zur Ermittlung einer radiographischen CD8⁺-Zellsignatur mit ein. Brendlin et al. erzielten akkuratere Ergebnisse bei der Verwendung von dual-energy CT-Bildern zur Extraktion diagnostischer Parameter (Brendlin et al., 2021).

Die Datenlage in der Hybridbildgebung ist noch deutlich kleiner, jedoch konnten bereits vielversprechende Ergebnisse zur Responsevoraussage bei Immuntherapie mittels Radiomics in PET/MRT-Bildern gezeigt werden (Liebgott et al., 2021). Sjöholm et al. stellten eine Methodik zur automatischen multiparametrischen Ganzkörperanalyse mit Extraktion metabolischer Parameter vor. Diese sah die Erstellung eines multiparametrischen Atlas vor, welcher mittels sogenannter Imiomics erstellt wurde. Diese dienen dazu, PET/MRT-Daten durch Bildverformung in ein einheitliches Koordinatensystem zu übertragen, um interindividuelle Vergleiche durchführen zu können. Da mit diesem Vorgehen automatische Veränderungen bestimmter tumorassoziierter PET-Parameter einhergehen, wurde für die automatische Tumordiagnostik eine Vorgehensweise mit Analyse innerhalb von Originalbilddaten empfohlen (Sjöholm et al., 2019, Strand et al., 2017).

1.5 Zielsetzung

Durch die Automatisierung der radiologischen Bildgebung und deren Auswertung könnten potenziell Objektivität und Reliabilität gesichert, sowie Zeit und Ressourcen gespart werden. Beispielsweise könnte die automatische Analyse von PET-Daten eine präzisere onkologische Diagnostik durch eine objektivere Beurteilung des Therapieansprechens ermöglichen. Infolgedessen würde dies eine präzisere Therapiesteuerung unterstützen.

Das Ziel dieser prospektiven Arbeit war die Etablierung einer Methodik zur automatischen Extraktion morphologischer und funktioneller Parameter aus

Ganzkörper-PET/MRT-Daten. Die Studienpopulation setzte sich aus Melanompatienten unter Immuntherapie zusammen.

Das Ziel dieser Arbeit bestand aus den folgenden Einzelzielen:

- Die Etablierung einer automatischen Segmentierungspipeline zur Segmentierung von Milz, Knochenmarkräumen und Leber auf PET/MRT-Daten mithilfe maschinellen Lernens
- Die Evaluation der Genauigkeit der automatischen Segmentierungen
- Die darauf basierende automatisierte Extraktion anatomischer und funktioneller Parameter aus PET und MRT und die Bewertung deren Genauigkeit
- Proof of concept: Besteht ein Zusammenhang zwischen dem Verlauf der metabolischen Aktivität von Milz und Knochenmark und dem onkologischen Therapieansprechen unter Immuntherapie bei Patienten mit malignem Melanom?

2 Material und Methoden

2.1 Patientenkollektiv

In die Studie eingeschlossen wurden volljährige Patienten mit nicht-operablem malignem Melanom Stadium IV mit mindestens einer vorhandenen messbaren Läsion. Bei Indikation zur systemischen Chemotherapie wurden diese mit Signaltransduktionsinhibitoren oder Immuncheckpoint-Inhibitoren therapiert. Ausschlusskriterien waren akute Erkrankungen, schwangere oder stillende Frauen, nicht einwilligungsfähige Patienten, sowie Kontraindikationen für MRT-Untersuchungen, wie z.B. MR-untaugliche Implantate, Hörerkrankungen, Klaustrophobie oder Kontrastmittelallergie.

Im Zeitraum vom 29.09.2014 bis zum 23.02.2018 unterliefen 24 Patienten jeweils drei GK-PET/MRT-Untersuchungen. Die Gesamtuntersuchungszahl betrug demzufolge 72. Die erste Untersuchung zum Zeitpunkt t_0 erfolgte vor Therapiestart. Zum Therapiereponse-Monitoring erfolgte zwei Wochen (t_1) und drei Monate (t_2) nach Therapiestart ein weiterer Scan. Diese fanden im MR/PET-Zentrum der Radiologischen Universitätsklinik Tübingen statt. Zusätzlich durchliefen alle Patienten zum Staging und zur Therapiekontrolle zum Zeitpunkt t_0 und t_2 eine GK-PET/CT-Untersuchung.

Zuvor wurde das schriftliche Einverständnis der Patienten bezüglich der Nutzung und Auswertung ihrer Daten für wissenschaftliche Zwecke eingeholt. Die ethische und rechtliche Unbedenklichkeit wurde von der hiesigen Ethikkommission bestätigt (Projektnummer 251/2012BO1).

2.2 Bildgebung

Die Bildaufnahmen wurden mithilfe des *Biograph mMR*-Tomographen der Firma Siemens Healthineers GmbH (Erlangen, Deutschland) mit 3 Tesla-MRT und paralleler molekularer PET-Bildgebung erstellt.

Alle Patienten waren mindestens sechs Stunden vor der Gabe des Tracers ^{18}F -FDG nüchtern. Vor der ersten als auch vor der dritten PET/MRT-Aufnahme erhielten die Patienten eine PET/CT-Untersuchung, sodass sich die Uptake-Phase auf eine Zeit von 120 min belief, welche für die zweite gesonderte

PET/MRT-Untersuchung übernommen wurde. Die GK-PET-Aufnahmen gingen mit einer Aufnahmedauer von vier Minuten pro Bettposition einher. Weiterhin wurde auf eine jederzeit konstante Dosierung des Tracers geachtet. Für die PET-Bildgebung wurde der herstellereigene 3D-OSEM-Algorithmus mit 21 subsets, 2 Iterationen, einer Matrixgröße von 256×256 und einem 4 mm Gauß-Filter verwendet. Zur Korrektur von Weichteilartefakten wurde eine T1w spoiled gradient-echo-Sequenz mit Fett-Wasser-Trennung in Dixon-Technik unter endexpiratorischen Atemstopps durchgeführt. Die genauen MRT-Parameter können Tabelle 1 entnommen werden. Bei dem genutzten Kontrastmittel handelte es sich um Gadobutrol (Gadovist® 1,0 mmol/ml, Bayer Vital GmbH, Leverkusen, Deutschland) (Seith et al., 2018).

Tabelle 1: MRT-Aufnahmeparameter

Feldstärke	3 T
Matrixgröße	312 x 384
In-Plane-Auflösung	1,3 x 1,3 mm
Schichtdicke	3 mm
TE, Echozeit	1,23 / 2,46 ms
TR, Repetitionszeit	3,95 ms
Flipwinkel	10°
Bandwidth	1085 Hz

Anhand der Bildgebung wurde im Anschluss mithilfe der RECIST-Kriterien bei allen Patienten ein Therapieansprechen in fünf Kategorien eingeteilt: complete response (CR), partial response (PR), stable disease (SD), mixed response (mixed) und progressive disease (PD).

2.3 DCNet

DCNet ist ein Convolutional Neural Network (CNN) und gehört somit zu den überwachten Lernalgorithmen. Dem Netz müssen demzufolge Trainingsdaten zugeführt werden, bevor es an Testdaten angewandt werden kann. Der Anwendungszweck besteht hierbei in der voxelweisen Segmentierung von Geweben in Bilddaten. In dieser Studie entspricht dies einer binären Klassifikation der Voxel in die Klassen *Organ* und *Hintergrund*.

Das Netz vereint einige Entwürfe und Methoden aus vorigen Arbeiten zu einem neuen künstlichen neuronalen Netz. Das Grundgerüst bedient sich an Elementen aus *UNet* (Ronneberger et al., 2015) und *VNet* (Milletari et al., 2016), welche bei der zweidimensionalen Lokalisierung sowie volumetrischen Segmentierung halfen (Küstner et al., 2018). Eine Kombination mit Teilen aus *ResNet* (He et al., 2016) und *DenseNet* (Huang et al., 2017) trug zur Optimierung durch unter anderem zunehmende Tiefe des Netzes bei. Das neue Netz zielt durch Veränderung des rezeptiven Feldes auf eine verbesserte Merkmalsextraktion und Positionsbestimmung bei gleichzeitig niedriger Anzahl benötigter Trainingsdaten ab. Insgesamt besteht DCNet aus 152 Layers, in welchen 1,1 Millionen Parameter trainierbar sind (Küstner et al., 2018, Küstner et al., 2020). DCNet ist aus einem absteigenden codierenden und aufsteigenden decodierenden Schenkel aufgebaut (Abb. 1), wobei der aufsteigende Schenkel die Inputform der extrahierten Merkmale wiederherstellen soll. Die auf DenseNet beruhende Struktur setzt sich aus nacheinander folgenden dense blocks zusammen, welche aus einer Anzahl L dense convolutional layers besteht, auf welche im codierenden Schenkel ein *TransitionLayerPool* (pooling layer) und im decodierenden Schenkel ein *TransitionLayerUp* zum Upsampling folgt. Ein fully connected layer gibt es nicht. Die dense blocks sind über Verkettungen (*concatenations*) strukturiert, sodass jeder Layer zusätzlich kollektiven Input von vorgehenden Layern bekommt. Eine lineare Interpolation der letzten Schicht dient der Schätzung einer berücksichtigten Lokalisation der Segmentierung. Als Input werden sowohl die Bilder der Wasser- als auch der Fettphase der T1w Dixon-Sequenz genutzt, wobei jedes einzelne Bild zuvor in einen $32 \times 32 \times 32$ Würfel konvertiert wird. Als Output zeigen sich die Segmentierungen als binäre Masken (Küstner et al., 2018, Küstner et al., 2019, Küstner et al., 2020).

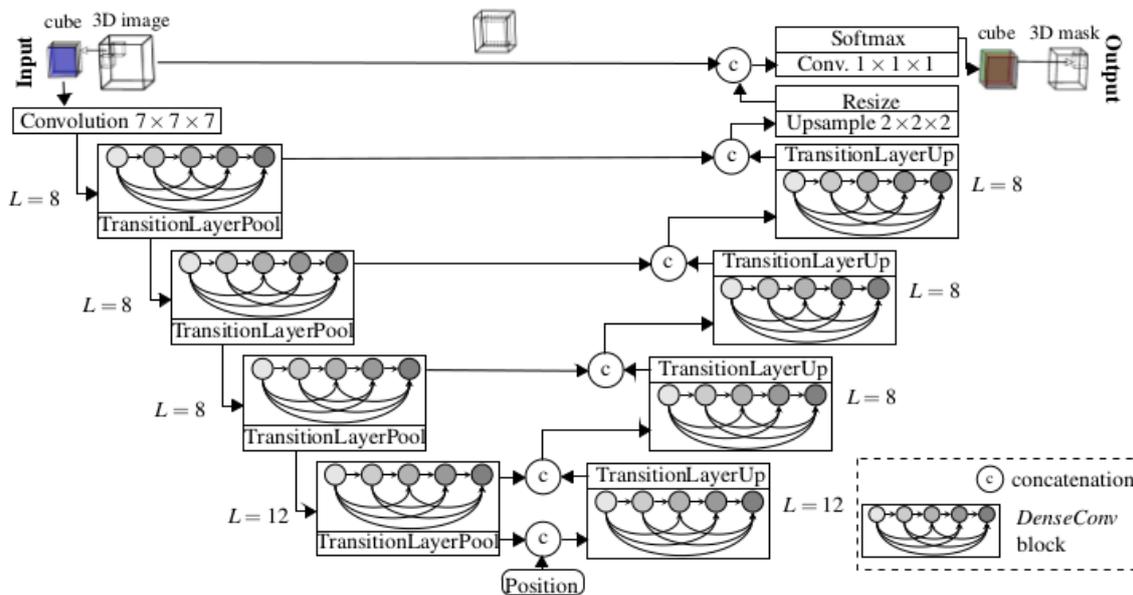


Abbildung 1: **Aufbau von DCNet.** (Küstner et al., 2019) Absteigender codierender (links) und aufsteigender decodierender Schenkel (rechts). Hintereinander geschaltete dense layers bestehen aus einer Anzahl L an dense convolutional layers gefolgt von einem TransitionLayerPool/Up. Verkettungen sind mit (c) beschrieben.

2.4 Automatische Segmentierungspipeline

Der Export der Bilder aus dem Kliniksystem in ein DICOM-Dateiformat, sowie die sonstige Bearbeitung und Verwendung der Daten erfolgte nach Pseudonymisierung der Patienten an einem dedizierten Klinikarbeitsplatz.

Als die zu untersuchenden Körperstrukturen wurden drei Organe festgelegt: Leber, Milz und Wirbelkörper. Einerseits verfügen die Organe über eine geringe interindividuelle Formvariabilität, sodass eine einheitlichere Beurteilung eines Therapieansprechens möglich war. Andererseits wurde von den Organen durch ihre Rolle in der Immunabwehr eine messbare Ansprache durch die Anwendung von Immuntherapeutika erwartet (Liebgott et al., 2021).

Eine automatische Segmentierung durch DCNet setzte ein Training des Netzes voraus. Zur Segmentierung von Leber, Milz und Wirbelkörper wurden die Klassifikatoren daher mit den 3D-Masken der drei Organe trainiert.

2.3.1 Manuelle Segmentierung im 2D-Bild

Die Generierung der Trainingsdatensätze erfolgte durch Erstellung von 3D-Masken mithilfe der Software *MITK Workbench* (Medical Imaging Interaction Toolkit, Version 2016.11) des Deutschen Krebsforschungszentrums (Heidelberg, Deutschland). Hierzu wurden zunächst Leber, Milz und Wirbelkörper ab BWK 1 bis einschließlich Os sacrum (im Folgenden als *Wirbelkörper (WK)* zusammengefasst) manuell im 2D-Bild segmentiert.

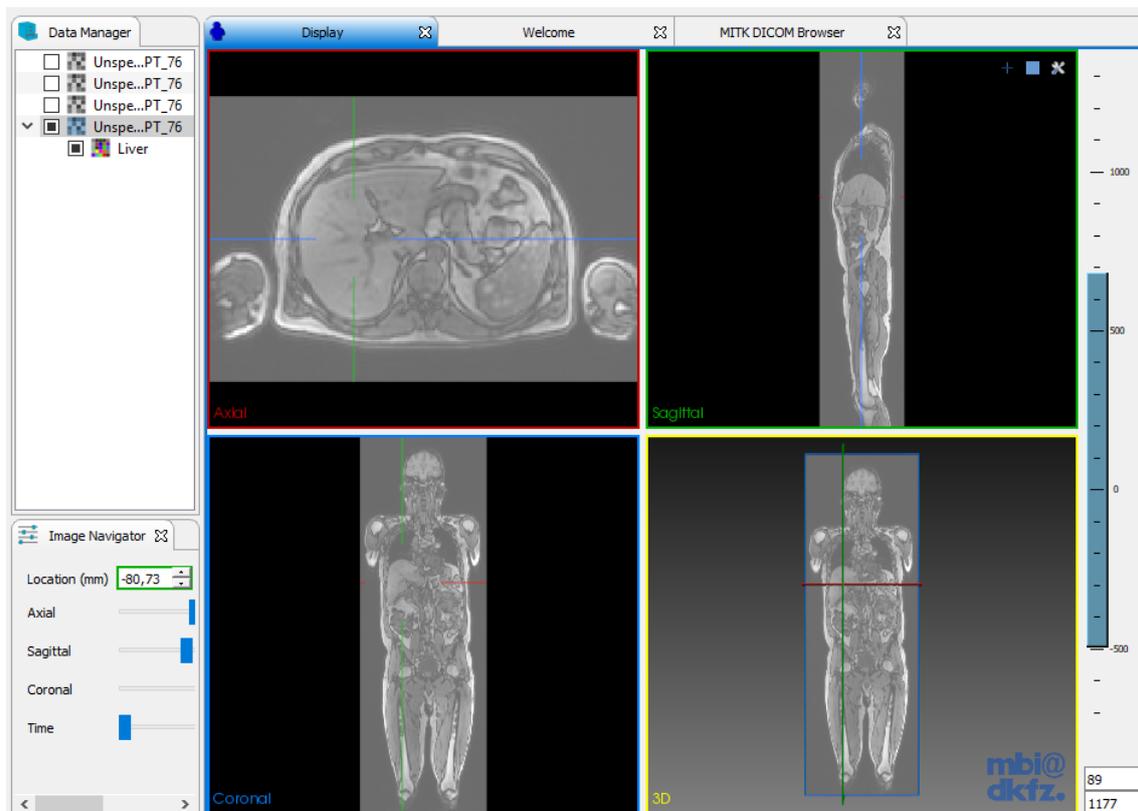


Abbildung 2: Importierter DICOM-Datensatz im MITK-Interface nach Anlegen einer Segmentierung. Es zeigt sich eine nicht ideal gefensterte opposed phase-Aufnahme.

Nach Import und Auswahl der entsprechenden Bildserie wurden die Bilder in das *Display*-Fenster geladen (Abb. 2). Nach dem erstmaligen Laden sowie nach dem Neuöffnen der Bilder nach einer Unterbrechung war eine Reinitialisierung über die *Global Reinit*-Funktion nötig, da ansonsten die Erstellung und Speicherung

neuer Segmentierungen nicht möglich war. Eine passende Fensterung wurde durch die Funktion *Use optimized levelwindow* erreicht.

Über das *Segmentation*-Menü wurden im Folgenden die neuen Segmentierungen für die interessierenden Bildbereiche (*region of interests, ROI*) erstellt. Für die folgenden Auswertungen betraf dies jeweils Leber, Milz und die Wirbelkörper. Da für jede Aufnahme vier Bildserien vorlagen, wurden die Segmentierungen zur Vermeidung von Verschiebung durch chemical shift-Artefakte einheitlich in der Dixon-Wasserphase durchgeführt. Der Abgleich zu den anderen Phasen (Fett-, in-, opposed-) sowie eine manuelle Fensterung zur Beurteilung schlecht erkennbarer anatomischer Verhältnisse bei Bildartefakten erwies sich als hilfreich.

Die Ausgangssegmentierungen für Leber und Milz erfolgten im Axialschnitt (Abb. 3, Abb. 4). Die Wirbelkörper wurden im Sagittal-, das Os sacrum im Koronalschnitt segmentiert (Abb. 5). Eine nachträgliche Revision mit Korrektur zur Qualitätssicherung fand in allen drei Ebenen statt und wurde von erfahrenen Radiologen überprüft.

Für die Segmentierung der Leber wurde die direkt angrenzende Vena cava inferior, sowie alle direkt nach außen angrenzenden großen Lebergefäße ausgeschlossen. Die internen Lebergefäße, die keine Verbindung zur Lebergrenze zeigten, blieben in der Segmentierung erhalten.

Für jede manuelle Segmentierung wurde die Dauer der Arbeit erfasst. Im Anschluss erfolgte die Speicherung der Segmentierungen im nifti-Dateiformat.

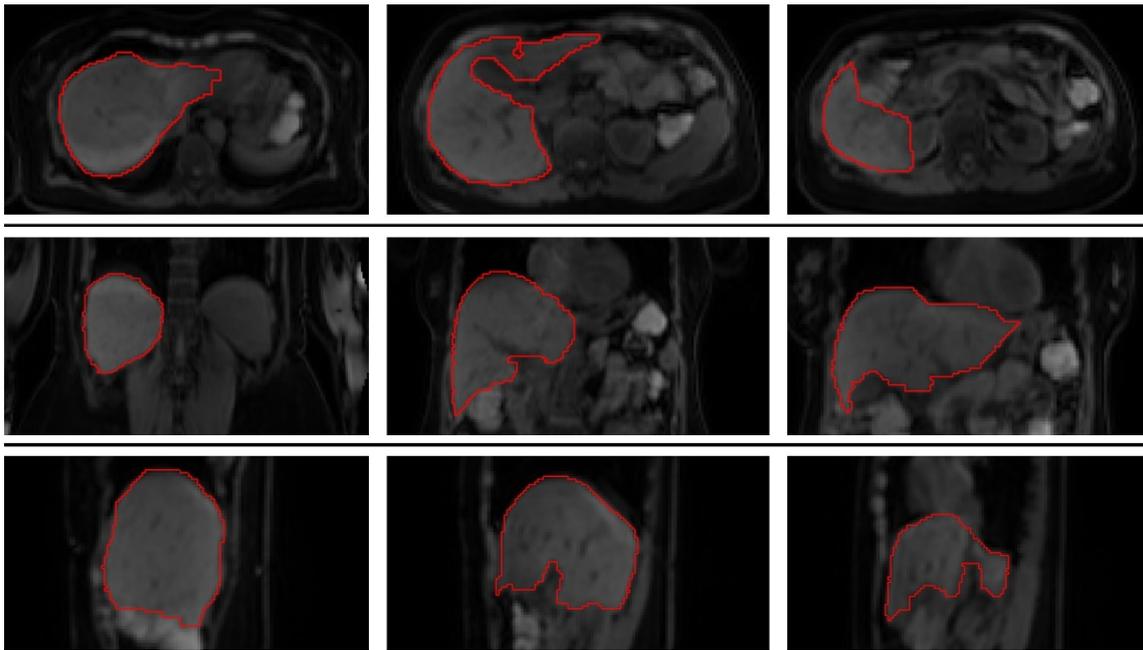


Abbildung 3: Manuelle Segmentierung einer Leber in der Wasser-Phase.
Oben axial, mittig koronal, unten sagittal.

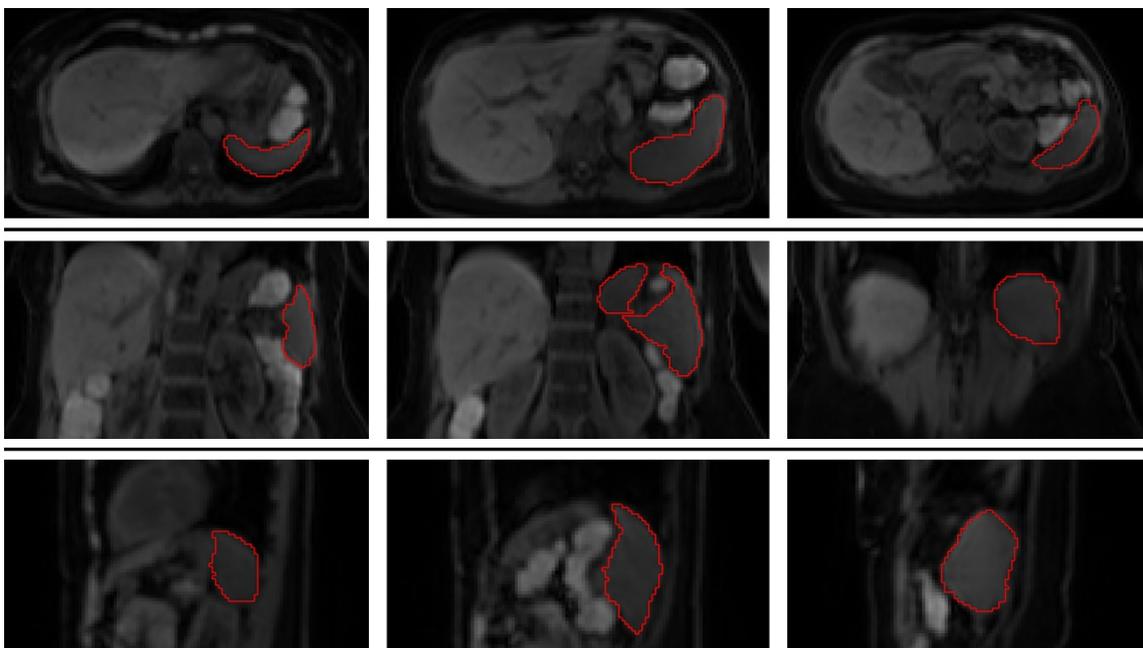


Abbildung 4: Manuelle Segmentierung einer Milz in der Wasser-Phase.
Oben axial, mittig koronal, unten sagittal.

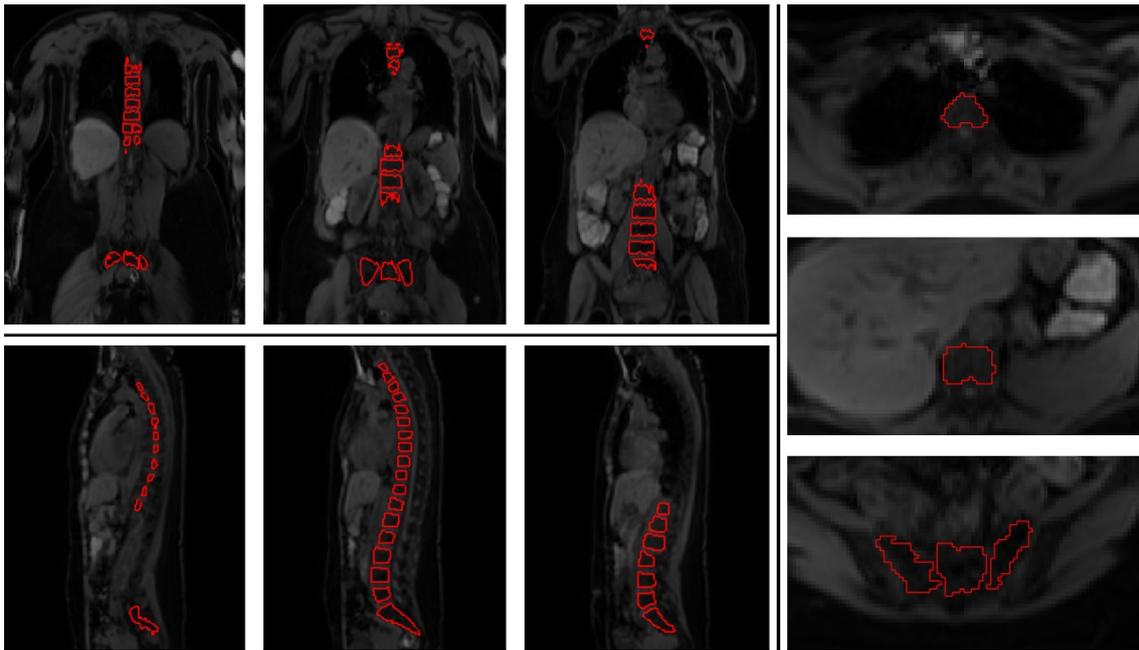
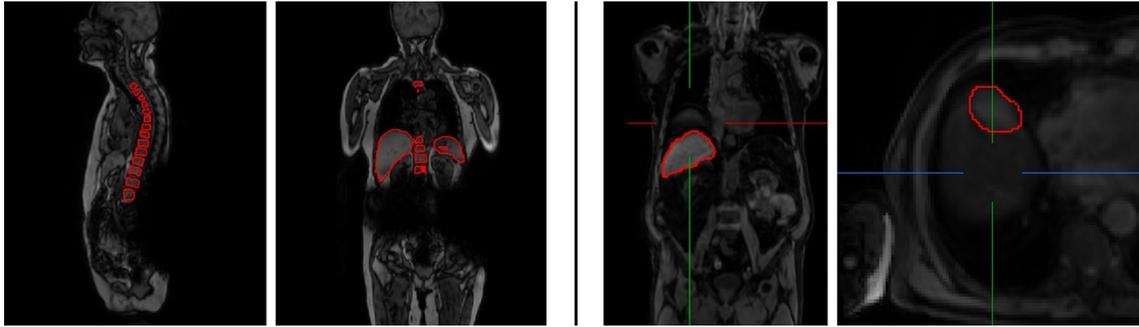


Abbildung 5: Manuelle Segmentierung der Wirbelkörper BWK 1-LWK 5 und Os sacrum in der Wasser-Phase. Oben koronal, unten sagittal, rechts axial.

In den Abb. 3-5 können die finalen manuellen Segmentierungen der drei Organe gesehen werden.

Schwierigkeiten bereitete, neben der anatomischen Variabilität, die Aufnahmequalität der Bilder. Da nicht alle Aufnahmen die gleiche Schärfe vorwiesen, und einige von ihnen erheblich artefaktbehaftet waren (Abb. 6), konnte hier eine akkurate Segmentierung auch bei entsprechender Fensterung nicht vonstattengehen. Entsprechende Stellen wurden von der Segmentierung ausgenommen. Ein Datensatz mit großem Suszeptibilitätsartefakt im Abdomen (Abb. 6 links) wurde im Folgenden von der Studie ausgeschlossen.



*Abbildung 6: **Beispiele für Artefakte.** Links: unmögliche Segmentierung der LWK und Os sacrum durch Suszeptibilitätsartefakt durch Fremdkörper – Rechts: Andeutung eines zweiten Leberberrandes (s. Cursor) durch Bewegungsartefakt.*

2.4.1 Training

Aufgrund der geringen Kohortengröße fand das Training des Klassifikators in einem Kreuzvalidierungsverfahren statt. Ein Großteil der Patientendaten (85%) wurde zufällig ausgewählt und für das Training verwendet, der übrige Teil (15%) für die Evaluation nach dem Training. Die Randomisierung diente dem Ausschluss einer Überrepräsentation verschiedener Phänotypen, beispielsweise betreffend das Alter, Geschlecht oder BMI. Damit für alle Patienten Testmasken erstellt werden konnten, wurde sieben Mal trainiert (*7-fold cross validation*). Für jedes Training wurden andere 15 % der jeweiligen Patientendaten für die Evaluierung zurückbehalten. Da von jedem Patienten jeweils drei Aufnahmen vorlagen, wurde bei diesem Prozess darauf geachtet, dass alle drei Aufnahmen zur Vermeidung eines Overfittings vollständig im Trainings- oder Validierungsdatenset enthalten waren, sodass die Patienten im Evaluierungsset dem Klassifikator gänzlich unbekannt blieben. Nach dem Training konnten erfolgreich die automatischen Segmentierungen erzeugt werden.

2.5 Evaluierung der Genauigkeit

2.5.1 Qualitative Auswertung

Über einen zuvor erstellten numerischen Score wurde die Qualität der automatisch erzeugten Segmentierungen visuell beurteilt.

Als Maß wurde dafür in allen drei Ebenen aller MRT-Bilder samt automatischer Segmentierung der drei Organe visuell nach Fehlklassifikationen (FK) gesucht. Bei den FK handelt es sich um falsch-positiv und falsch-negativ definierte Organabschnitte, die entweder organnah (lokal) oder organfern (global) vorkommen können. Grundsätzlich wurde jede an das Organ grenzende Klassifikation als lokale FK gewertet. Klassifikationen in die unmittelbar benachbarten Organe und Strukturen, die auch nicht an die Organmaske angrenzten, und höchstens 10 cm entfernt waren, wurden ebenfalls als lokale FK betrachtet. Bei den drei Organen (Leber, Milz, Wirbelkörper) unterscheidet sich folglich die Lokalisation der lokalen und globalen FK. Es wurde zwischen kleinen (< 10 Voxel), mittelgroßen (10 - 19 Voxel) und großen FK (≥ 20 Voxel) unterschieden und die Lokalisation notiert.

Die Scores sehen für die drei Organe wie folgt aus:

Tabelle 2: Visueller Score zur qualitativen Beurteilung der Organmasken.

1	sehr grobe (globale) FK, keine Organerfassung
2	grobe globale FK (evtl. mit lokaler FK), schlechte Organerkennung
3	grobe lokale FK ODER kleine globale FK (evtl. mit kleineren lokalen FK)
4	kleine FK außerhalb der Organgrenzen, jedoch lokal angrenzend oder in unmittelbarer Organnähe
5	fehlerfrei

2.5.2 Quantitative Auswertung

Zur quantitativen Beurteilung der Segmentierungen wurde die automatisch segmentierte Organmaske mit der manuell segmentierten Organmaske voxelweise in einer Konfusionsmatrix verglichen. Die manuell erstellten Trainingsdaten wurden hierbei als wahre Referenzwerte (Grundwahrheit, *ground truth*) angesehen. Organweise wurden durch die Masken die beiden Klassen *Organ* (entweder Leber, Milz oder Wirbelkörper) und *Background* (Hintergrund) definiert, aus denen nun die Anzahl der Voxel ermittelt werden konnte. Im Vergleich ergaben sich so für jeden Voxel die Fälle richtig-positiv (*rp*), richtig-

negativ (rn), falsch-positiv (fp) oder falsch-negativ (fn). Die berechneten Kenngrößen sind:

Treffergenauigkeit

$$\text{accuracy} = \frac{rp + rn}{rp + fp + rn + fn}$$

Dice score (auch Sørensen-Dice similarity coefficient (**DSC**), F1-Score)

$$\text{DSC} = \frac{2 \cdot rp}{2 \cdot rp + fp + fn}$$

Sensitivität

$$\text{sensitivity} = \frac{rp}{rp + fn}$$

Spezifität

$$\text{specificity} = \frac{rn}{rn + fp}$$

Positiver Vorhersagewert

$$\text{precision} = \frac{rp}{rp + fp}$$

Diese Metriken wurden organweise für das komplette 3D-Volumen berechnet und mit Boxplots visualisiert.

Eine Gewichtung aller Kenngrößen fand balanciert statt: Die accuracy wurde als sogenannte *balanced accuracy* berechnet, die ihren Vorteil in einer realitätsgetreueren Abbildung bei ungleich großen Klassen findet (García et al., 2009). Da die Klasse *Organ* deutlich weniger Werte als die Klasse *Background* enthielt, wurde nach Anzahl der Bildwerte pro Klasse gewichtet. Das gleiche Prinzip wurde bei sensitivity, specificity und precision angewandt (s. Anhang). Die gesamten gesammelten Daten wurden gründlich auf Übereinstimmung mit den vorliegenden MRT-Bildern geprüft.

Die Metriken wurden in *MATLAB* (Version R2019a; The MathWorks Inc., Natick, Massachusetts, USA) und *Excel 2019* (Version 16.0; Microsoft Corp., Redmond, Washington, USA) tabellarisiert und zur gesamten weiteren statistischen Analyse

und Visualisierung in *SPSS Statistics* (Version 26.0.0.1; IBM, Armonk, New York, USA) importiert.

2.6 Ableitung von PET/MRT-Parametern

Zur zusätzlichen Quantifizierung der Ergebnisse wurden Parameter der PET/MRT-Untersuchungen über den ROIs bestimmt. Organweise und patientenweise wurden ADC, durchschnittlicher und maximaler SUV, Fettgehalt und Organvolumen berechnet. Die Werte wurden jeweils für die manuell, als auch für die automatisch segmentierten Masken samt Mittelwert mit Standardabweichung, Median, Minimum und Maximum bestimmt.

Es wurden Verteilungsdiagramme mit den abgeleiteten Parametern über der automatischen Maske auf der Abszisse, und der ground truth auf der Ordinate erstellt. Bei der ground truth handelt es sich um die als richtig angenommenen Werte, die über den manuell erstellten Organmasken berechnet wurden.

Zum Test auf Normalverteilung wurde ein Shapiro-Wilk-Test angewendet. Bei teils parametrischer und teils nicht-parametrischer Verteilung wurden entsprechend der Korrelationskoeffizient r nach Pearson und der Rangkorrelationskoeffizient ρ nach Spearman berechnet.

Zur Analyse der PET-Metriken wurde die absolute Differenz der jeweiligen Parameter für die jeweiligen Patienten gebildet. Die Werte über den automatischen Masken bildeten den Minuenden, die ground truth den Subtrahenden. Zur Berechnung der relativen Differenz wurde für die einzelnen Patienten die absolute Differenz durch die ground truth dividiert. Für die absolute und die relative Differenz wurden jeweils Mittelwert, Standardabweichung, Median, Minimum und Maximum bestimmt.

2.7 Ableitung einer Therapieresponse

Als proof of concept wurde beispielhaft der SUV von der Leber, der Milz und den Wirbelkörpern analysiert. Mit nach Response und Organ aufgeteilten Liniendiagrammen („before-after graph“) wurde der SUV zu den Zeitpunkten t_0 , t_1 und t_2 dargestellt. Der zeitliche Verlauf des SUV zwischen erstem Zeitpunkt vor

Therapiebeginn (t_0) und drittem Zeitpunkt (t_2) drei Monate nach Therapiebeginn wurde über dessen Differenz $SUV_{diff} = SUV(t_2) - SUV(t_0)$ beschrieben. Diese Differenz wurde, kategorisiert nach Response, mittels Boxplots visualisiert.

Es wurde geprüft, ob ein signifikanter Unterschied von SUV_{diff} bei Respondern (CR + PR) versus Non-Respondern (PD) auftrat. Trotz non-parametrischer Verteilung der Daten (nach Shapiro-Wilk) und kleiner Studienpopulation wurde eine parametrische Testung durchgeführt, da von einer bleibenden Robustheit der Testung nach Simulationsstudien ausgegangen wurde (Rasch and Guiard, 2004). Zur Testung der Hypothese wurde ein zweiseitiger Welch-Test (Welch's t-Test) angewandt.

3 Ergebnisse

3.1 Patientenkollektiv

Insgesamt wurden in 72 MRT-Datensätzen die Leber, die Milz und die Wirbelkörper von BWK 1 bis Os sacrum manuell segmentiert.

Tabelle 3: Patientendaten. Angaben in Mittelwert \pm SD (Median; Spannweite).

Patientenzahl n	24	
♂ Männer	14	
♀ Frauen	10	
Alter [y]	63 \pm 11	(60; 49-84)
BMI bei t ₀ [kg/m ²]	26,72 \pm 4,15	(26,04; 19,93-35,78)
BMI bei t ₁ [kg/m ²]	26,60 \pm 4,02	(25,48; 20,71-35,29)
BMI bei t ₂ [kg/m ²]	26,11 \pm 3,84	(24,75; 20,49-34,07)
Körpergröße [m]	1,73 \pm 0,08	(1,73; 1,60-1,87)
<i>Therapie (Anzahl)</i>		
PD1-Inhibitor	12	
CTLA4-Inhibitor	3	
BRAF- und MEK-Inhibitor	2	
fehlende Angabe	7	
<i>Therapieresponse (Anzahl)</i>		
complete response (CR)	4	
partial response (PR)	4	
stable disease (SD)	2	
mixed response (mixed)	5	
progressive disease (PD)	9	

Insgesamt wurden zwei Aufnahmen aus der Studie ausgeschlossen. Eine Aufnahme zeigte ein großes Suszeptibilitätsartefakt im Abdomen (Abb. 6), sodass eine genaue bzw. komplette Segmentierung der Wirbelkörper nicht durchgeführt werden konnte. Bei einer zweiten Aufnahme kam es durch einen Speicherfehler zu einem fehlerhaften Datenimport.

Infolgedessen erfolgte das Training des Netzes und die automatische Segmentierung mit und an 70 Datensätzen.

3.2 Manuelle Segmentierung

3.2.1 Dauer

Vom Import der DICOM-Dateien in die Software *MITK Workbench* bis zum Laden der Bilder in den Viewer vergingen im Schnitt 130 s.

Die Segmentierungsdauer für die Leber betrug pro Aufnahme durchschnittlich 18,7 ($\pm 8,1$) min, für die Milz 5,0 ($\pm 2,0$) min und für die Wirbelkörper 18,3 ($\pm 6,2$) min. Über 70 Aufnahmen hinweg ergab sich eine Summe von 1.308 min für Leber, 351 min für Milz und 1.278 min für Wirbelkörper.

Die reine Arbeitszeit der manuellen Segmentierung betrug hierdurch 2.937 min (48,95 h) ohne Ladezeit bzw. 3.089 min (51,48 h) mit Ladezeiten. Die Zeit für eine anschließende Qualitätssicherung und eventuelle Verbesserungen wurde nicht mit einberechnet.

3.2.2 3D-Masken

Durch die manuelle zweidimensionale Segmentierung in mehreren hintereinander folgenden Ebenen entstand die dreidimensionale Maske, die in MITK Workbench zur besseren Visualisierung als frei bewegliches 3D-Modell angezeigt werden konnte (Abb. 7).

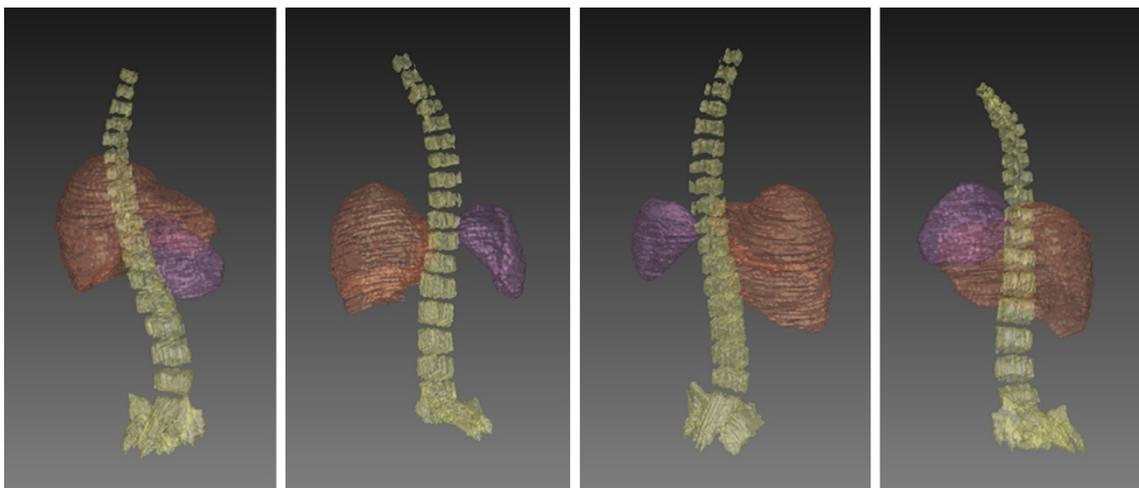


Abbildung 7: **Manuell erstellte 3D-Organmasken von Leber (orange), Milz (lila) und Wirbelkörper (gelb) in MITK.**

3.3 Automatische Segmentierung

3.3.1 Dauer

Die Dauer des Trainings betrug pro Durchgang im Schnitt 85 Stunden. Das Berechnen der Metriken nach Laden der Datensätze dauerte pro Aufnahme durchschnittlich 1,6 s bei der Leber, 0,8 s bei der Milz und 1,8 s bei der Wirbelsäule. So ergab sich eine Gesamtdauer von etwa 4 min für die automatische Segmentierung aller drei Organe an allen 70 Patienten.

3.3.2 Visuelle Qualität

3.3.2.1 Fehlerfreie Segmentierung

Eine visuell fehlerfreie Segmentierung mit dem visual score 5/5 erfolgte bei der Leber 3 Mal, bei der Milz 31 Mal und bei den Wirbelkörpern 21 Mal. Diese sind in den folgenden Abb. 8-14 zu finden.

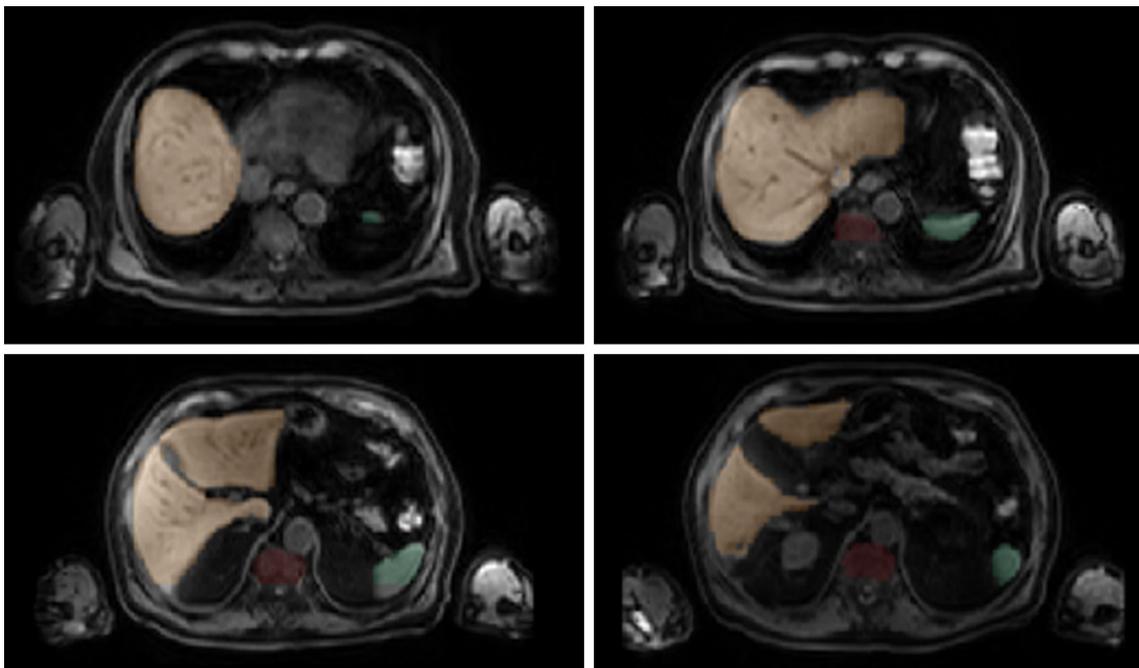


Abbildung 8: Beispiel einer einwandfreien automatischen Segmentierung der Leber (orange) im Transversalschnitt.

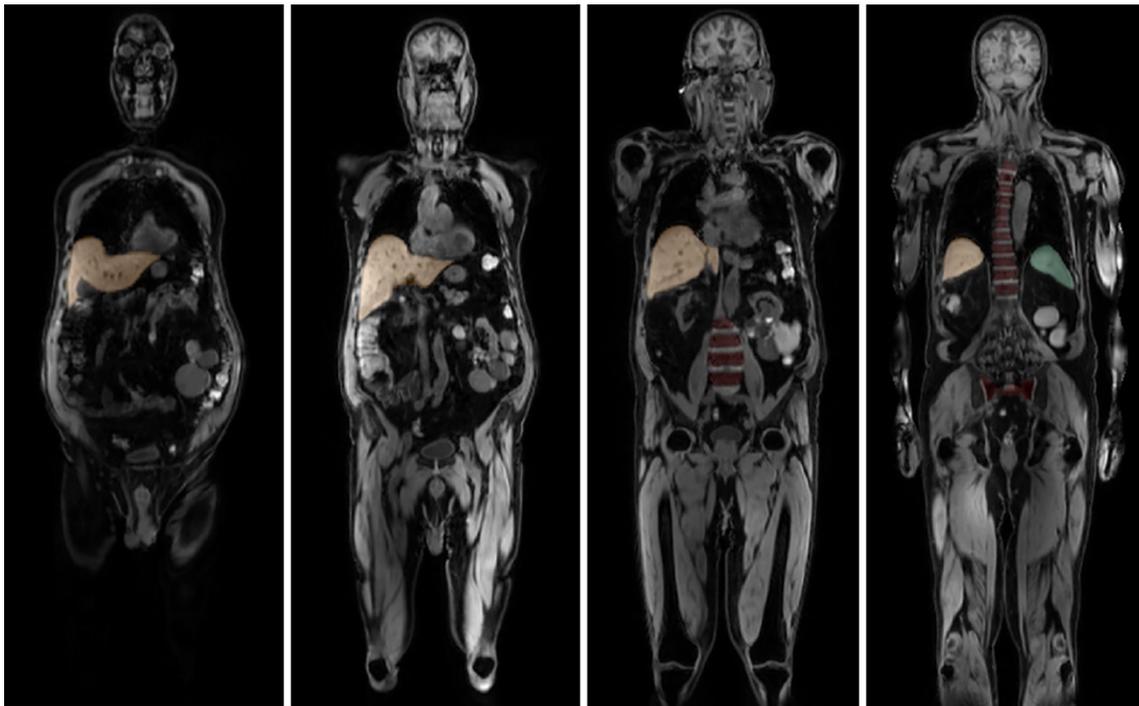


Abbildung 9: Beispiel einer einwandfreien automatischen Segmentierung der Leber (orange) im Koronalschnitt.

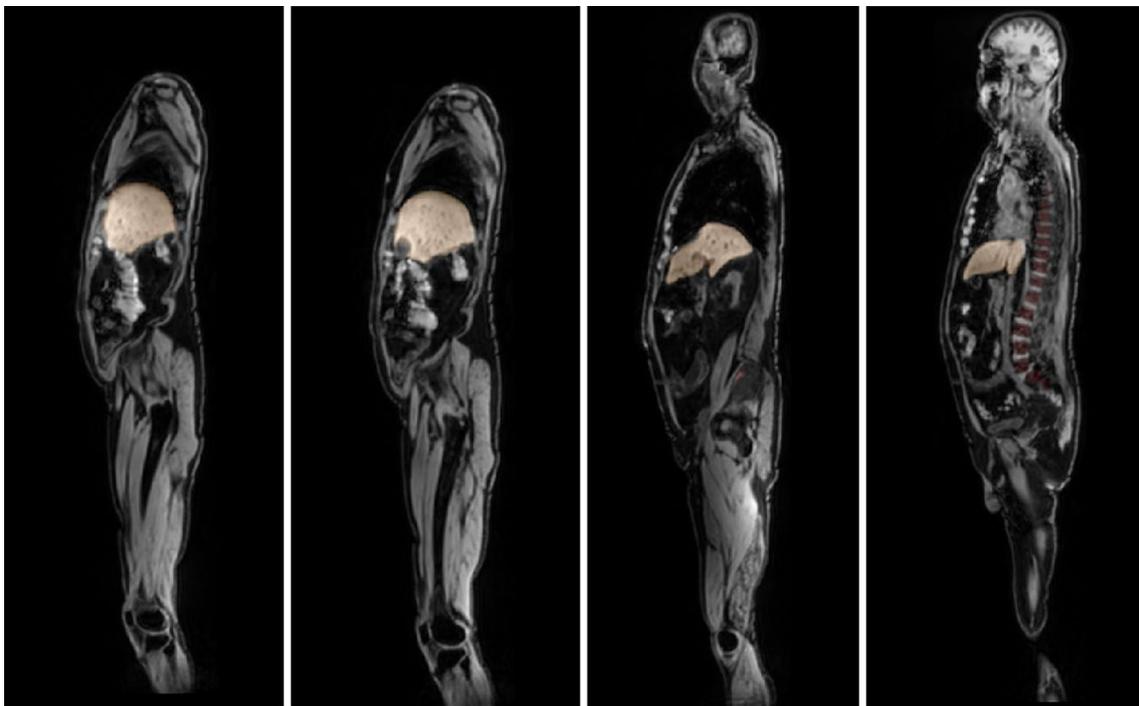


Abbildung 10: Beispiel einer einwandfreien automatischen Segmentierung der Leber (orange) im Sagittalschnitt.

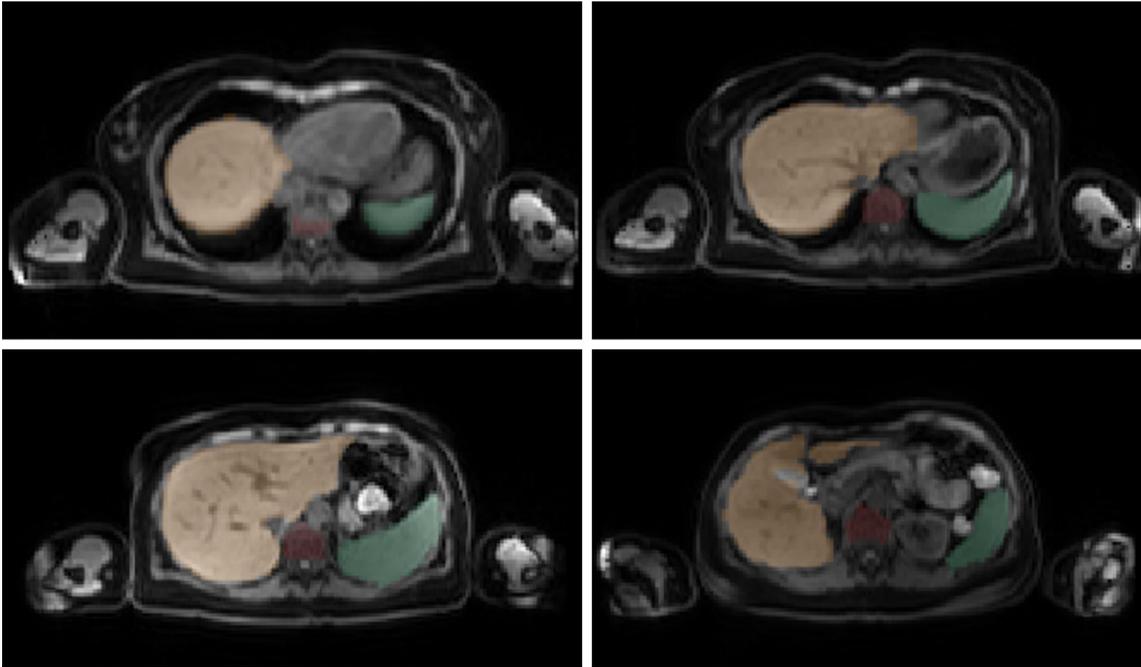


Abbildung 11: Beispiel einer einwandfreien automatischen Segmentierung der Milz (grün) im Transversalschnitt.

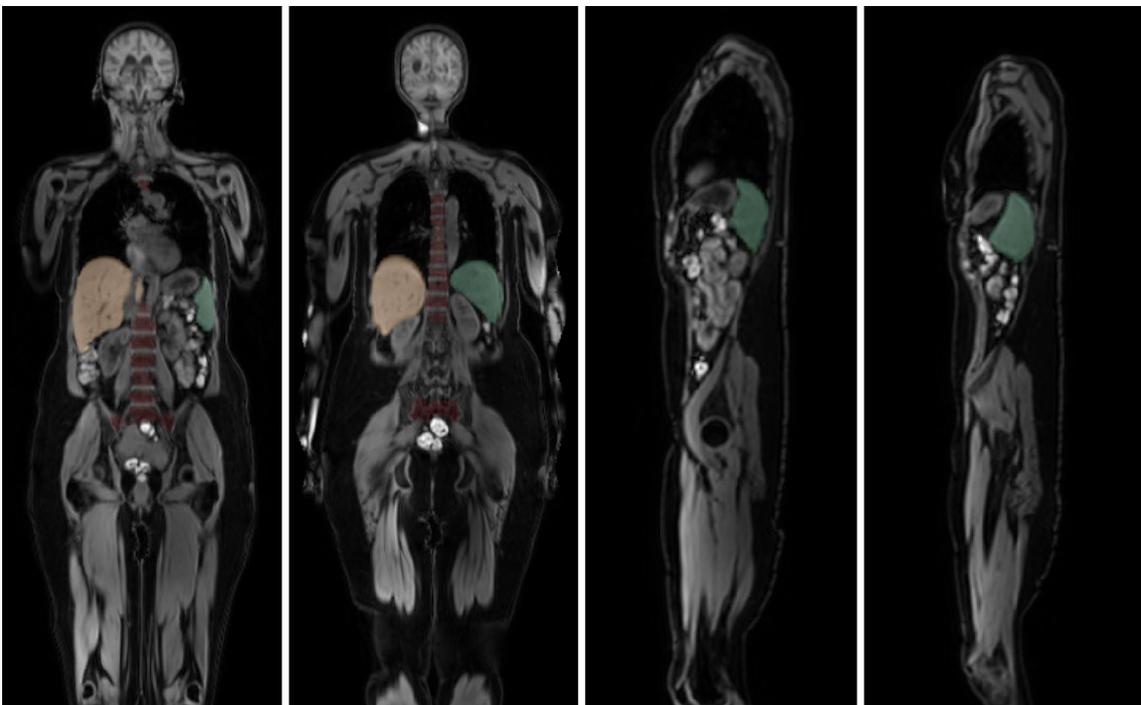


Abbildung 12: Beispiel einer einwandfreien automatischen Segmentierung der Milz (grün) im Koronarschnitt (links) und Sagittalschnitt (rechts).

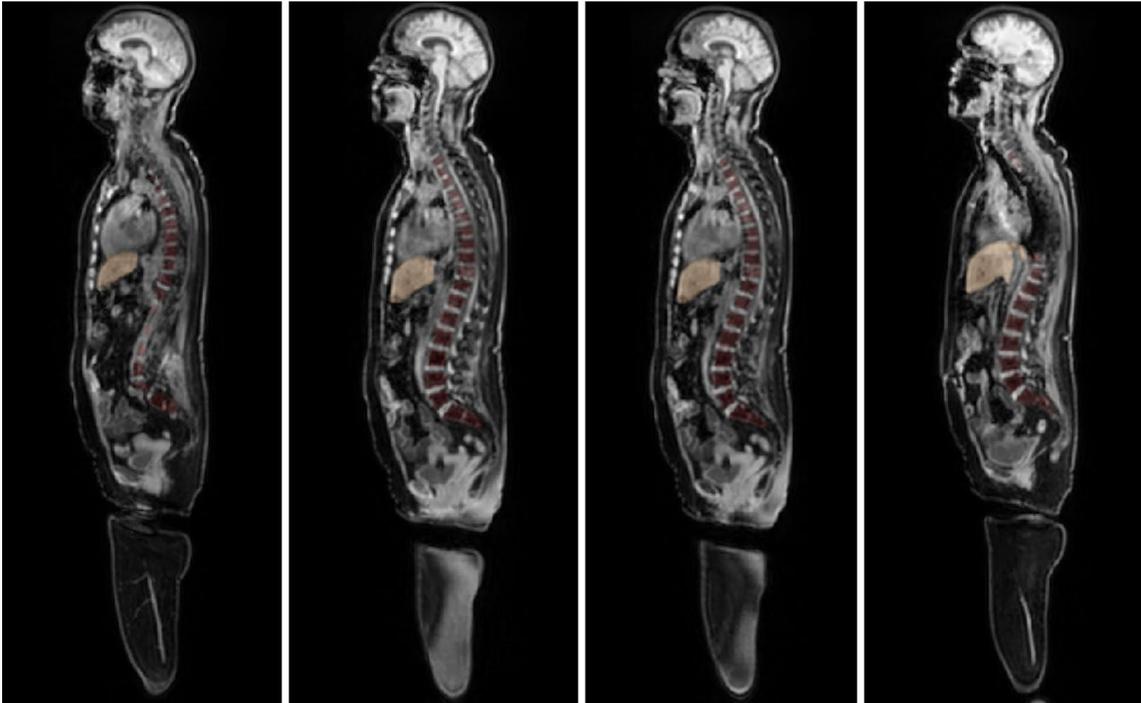


Abbildung 13: Beispiel einer einwandfreien automatischen Segmentierung der Wirbelkörper (rot) im Sagittalschnitt.

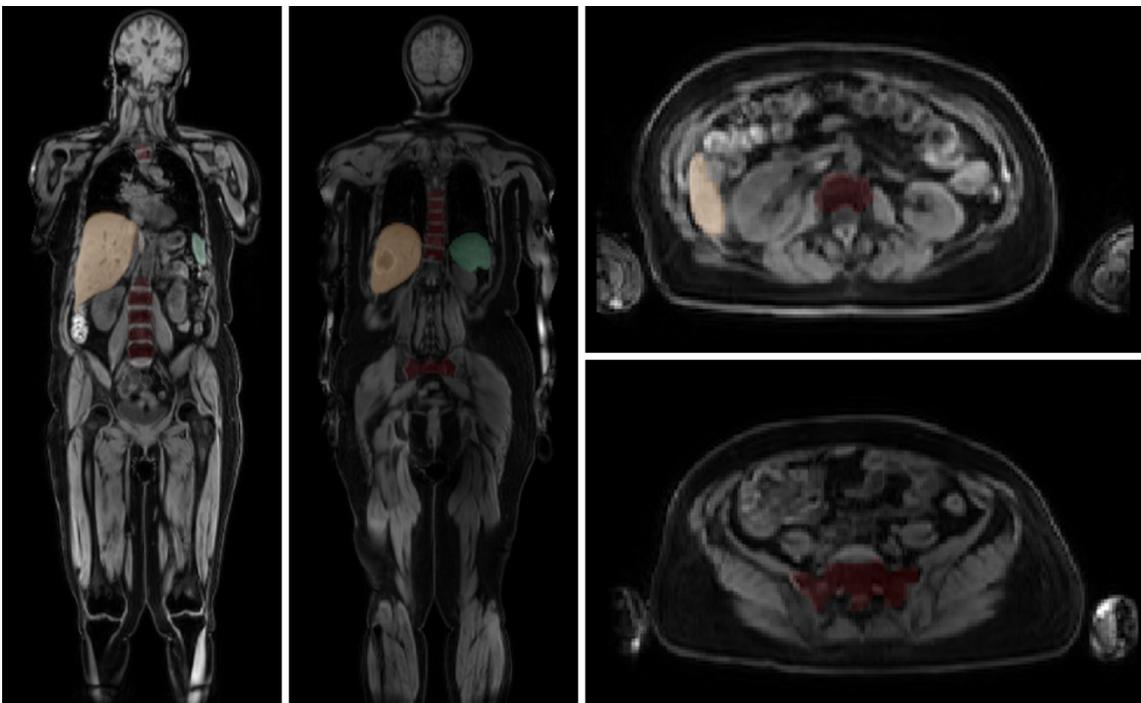


Abbildung 14: Beispiel einer einwandfreien automatischen Segmentierung der Wirbelkörper (rot) im Koronarschnitt (links) und Transversalschnitt (rechts).

3.3.2.2 Lokale Fehlklassifikationen

Grundsätzlich wurden alle Klassifikationen der Organe über die Organgrenze hinweg als lokale FK gewertet. Dabei kam es zu *oversegmentation* (falsch-positive) und *undersegmentation* (falsch-negative). Bei der automatischen Segmentierung der Leber fanden sich weitere lokale (organnahe) FK am knöchernen Thorax, am kaudalen Herz, am Magen, an der Gallenblase, am oberen rechten Nierenpol, sowie in den Darmabschnitten des rechten oberen Quadranten des Abdomens. Bei der Milz fanden sich diese gehäuft am Zwerchfell, am oberen linken Nierenpol, am Herzen, sowie in den Darmschlingen des linken oberen Quadranten des Abdomens. Lokale FK der Wirbelkörper befanden sich im Wirbelkörper HWK 7, in den Disci intervertebrales, im Os ischii, sowie in den Processus paravertebrales (s. Abb. 15-17).

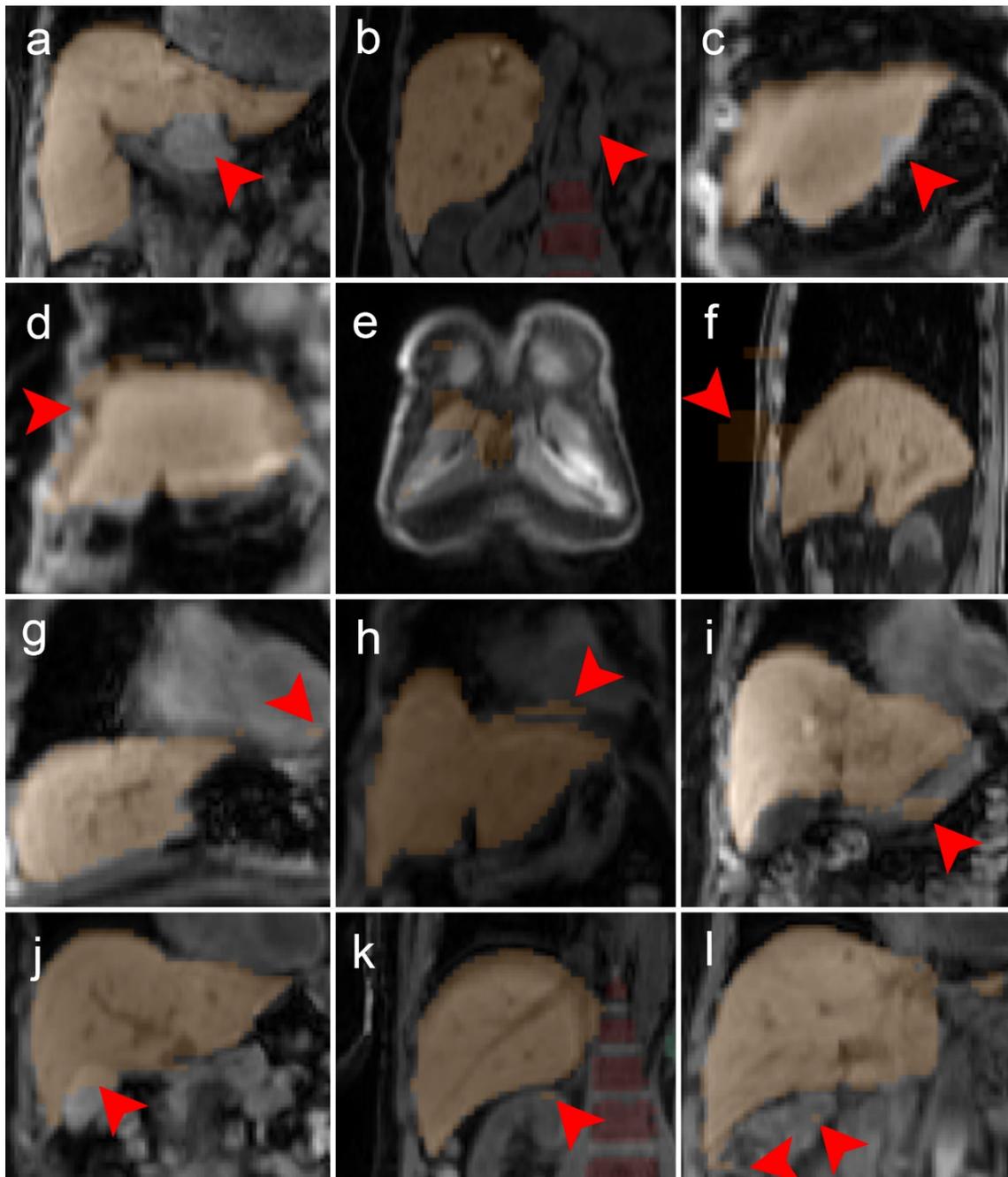


Abbildung 15: Lokale FK der Leber in der automatischen Segmentierung.
 Fehlender Organteil (a, b, c), Thoraxwand (d, e, f), Herz (g, h), Magen (i),
 Gallenblase (j), oberer rechter Nierenpol (k), Kolon (l).

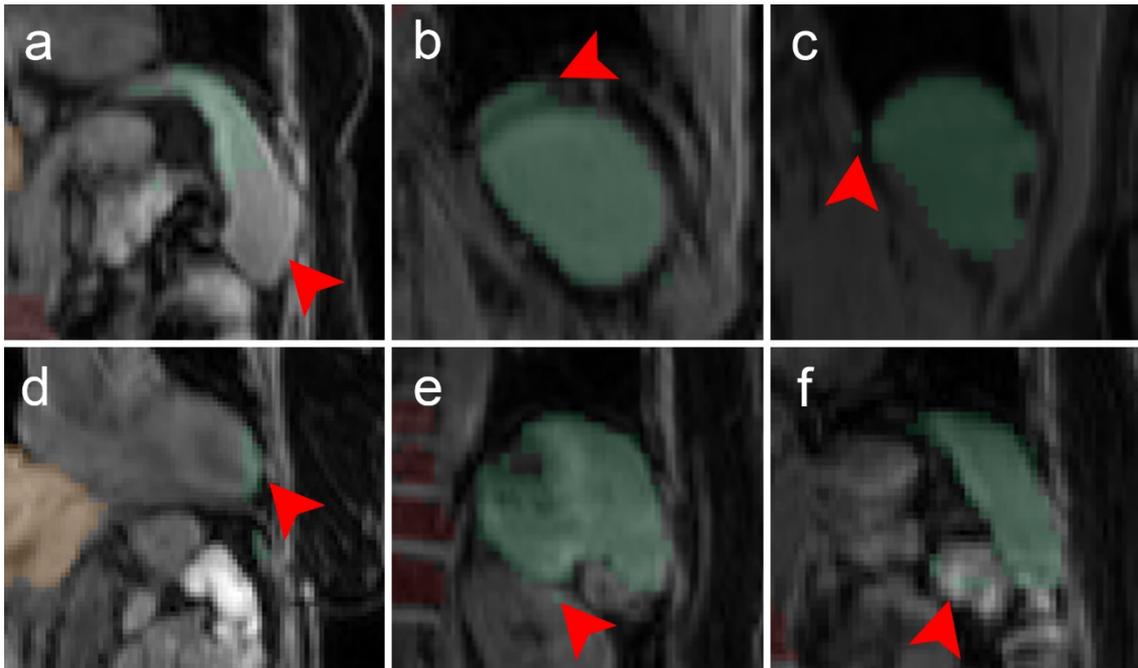


Abbildung 16: Lokale FK der Milz in der automatischen Segmentierung.
 Fehlender kaudaler Organteil (a), Zwerchfell (b), Muskulatur (b, c), Herz (d),
 oberer linker Nierenpol (e), Kolon (f).

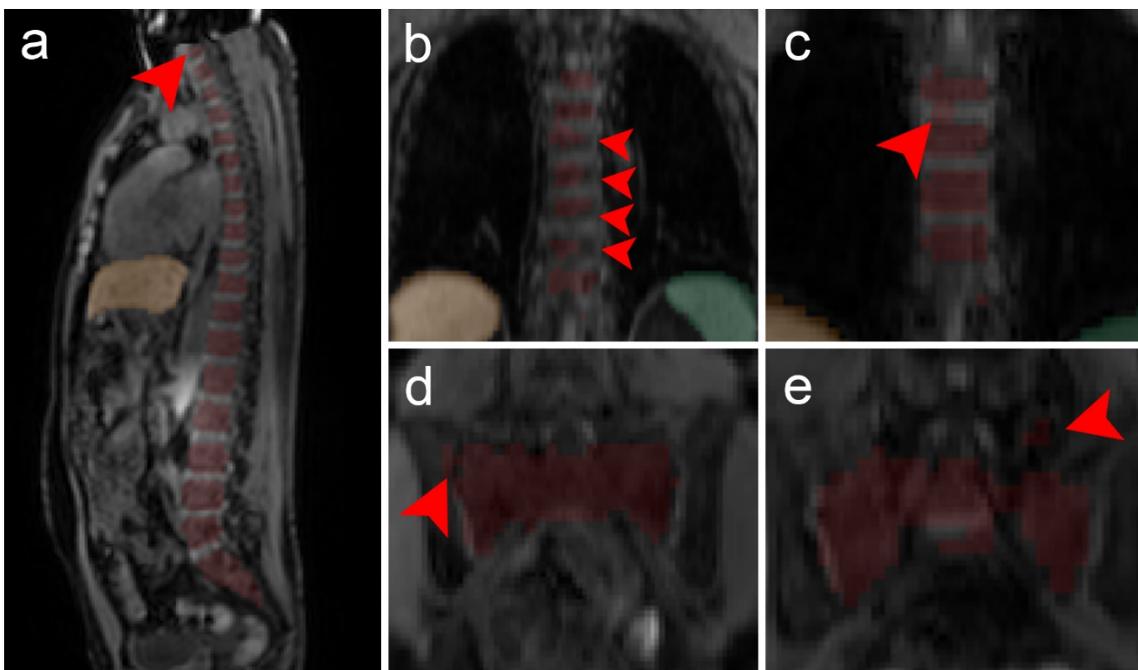


Abbildung 17: Lokale FK der Wirbelkörper einschließlich Os sacrum in der automatischen Segmentierung.
 HWK 7 (a), fehlender Organteil (b),
 Bandscheibe (c), Os ilium (d), Proc. transversus (e).

3.3.2.3 Globale Fehlklassifikationen

Die globalen (organfernen) FK der Leber fanden sich am kaudalen rechten Nierenpol, in der Milz, im Oberschenkel, im Oberarm, sowie im Gehirn. Bei der Milzsegmentierung befand sich eine globale FK an der rechten Niere. Globale FK der Wirbelkörper fanden sich in Herznähe, an der Niere, an der Milz, am Oberschenkel, sowie am Rektum (s. Abb. 18-20).

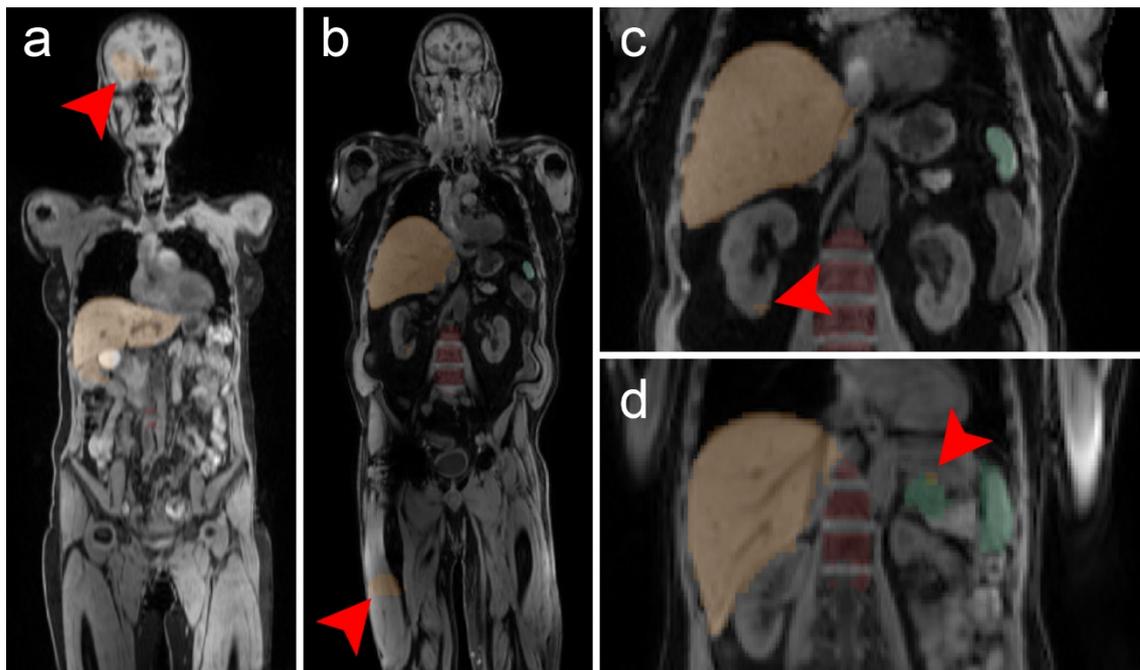


Abbildung 18: Globale FK der Leber in der automatischen Segmentierung.
Gehirn (a), Oberschenkel (b), unterer rechter Nierenpol (b, c), Milz (d).

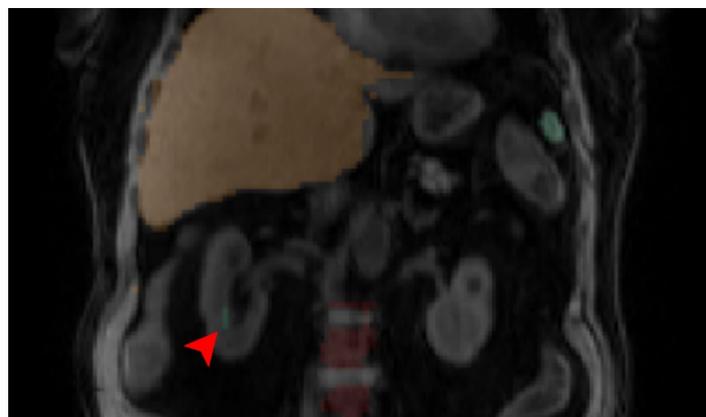


Abbildung 19: Globale FK der Milz in der automatischen Segmentierung an der rechten Niere.

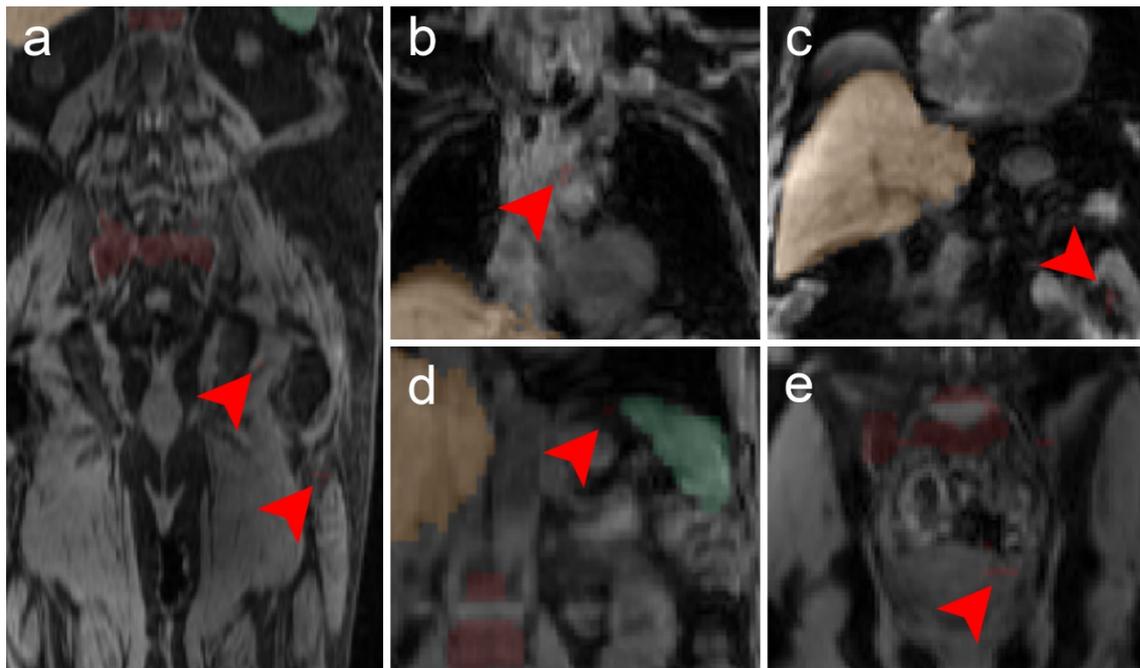


Abbildung 20: **Globale FK der Wirbelkörper einschließlich Os sacrum.** Oberschenkel (a), Aortenbogen (b), Leberartefakt und linke Niere (c), Milz (d), Rektum (e).

3.3.2.4 Weitere grobe Segmentierungsfehler

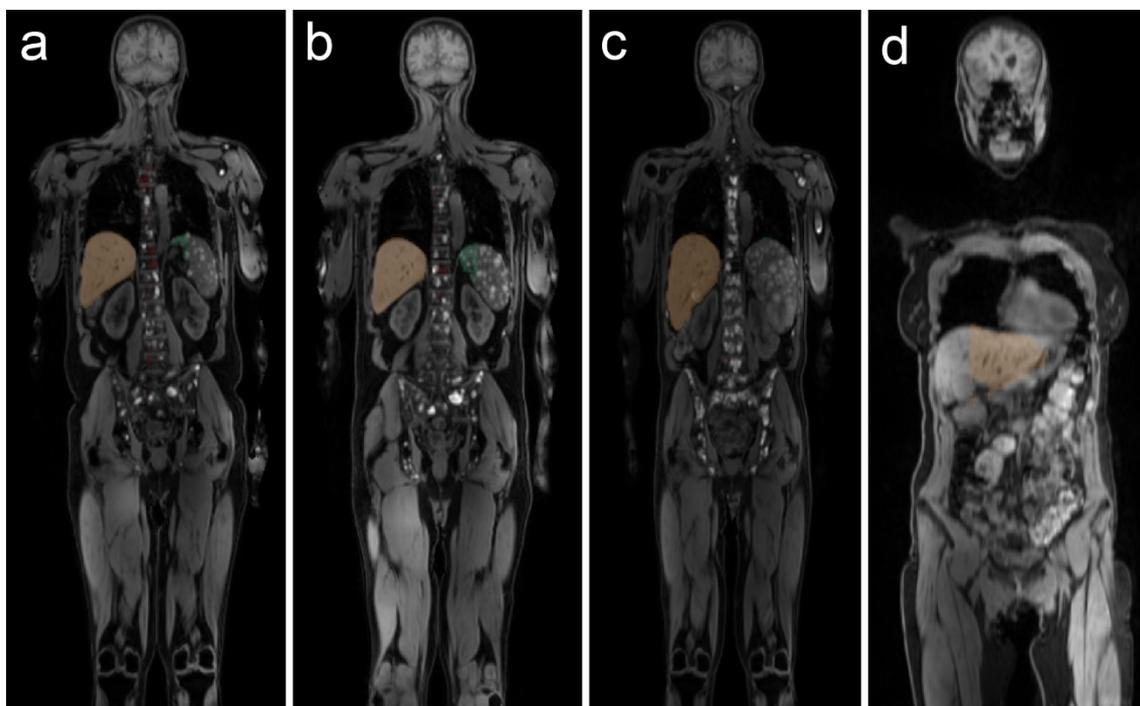


Abbildung 21: **Sehr schlechte bis fehlende Organerkennung von Milz und WK (a-c), großer fehlender Organteil der rechten Leber (d).**

Unter grobe Segmentierungsfehler fiel die sehr schlechte bis fehlende Organerkennung. Dies geschah jeweils dreimal bei Milz und WK. Als Befund fanden sich in Abb. 21a-c in den Wirbelkörpern und der Milz unter anderem zahlreiche Fernmetastasen. Kombiniert mit weiteren globalen groben FK bedingten diese Bilder die Vergabe von visual score 1.

3.3.2.5 Visual score

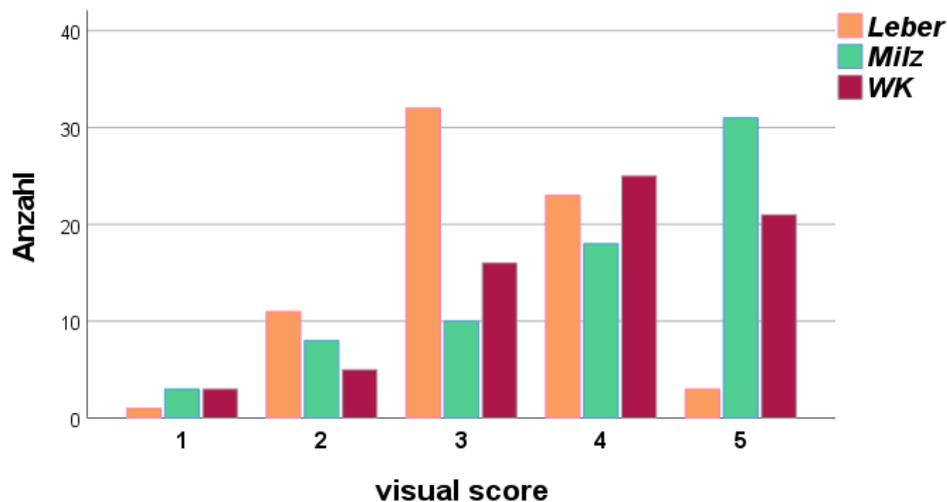


Abbildung 22: Anzahl der Aufnahmen nach visuellem Score, organweise.
 Ein visual score von 5 entspricht einer perfekten Segmentierung, ein score von 1 einer sehr schlechten Segmentierung.

Den besten visuellen Score mit einem Modalwert von 5 erreichte die Milz, den zweitbesten die Wirbelkörper mit 4, und den drittbesten die Leber mit 3.

Einen perfekten visual score 5 erhielten 3 Lebern, 31 Milzen und 21 Wirbelkörper-Datensätze.

Eine Auswertung aller lokalen und globalen FK mit Vergabe des jeweiligen visual scores kann im Anhang gefunden werden.

3.3.2.6 Anatomische Lokalisation der Fehlklassifikationen

Tabelle 4: Anzahl lokaler und globaler FK nach Ort.

Lokalisation	Leber	Milz	WK
Darm	22	11	
Herz	17	1	1
Brustwand/Bauchwand	16		6
Magen	12	4	
Gallenblase	11		
Niere	7	6	2
Milz	3		2
Pankreas	2		
Lunge	2		1
Aorta, A. pulmonalis	2		2
Arm	1		2
Oberschenkelmuskulatur	1		3
Bandscheibe	1		4
Cutis, dorsal	1		
Gehirn	1		
Zwerchfell		4	1
HWK 7			23
Processus transversus			3
Os ilium			2
Rektum			1
ungenauere lokoregionäre Grenzen	30	9	9
fehlender größerer Organteil	8	11	2
keine Organerkennung		3	3
einwandfrei	3	31	21

Die häufigste Lokalisation der Fehlklassifikationen für Leber und Milz war der Darm. Für die Leber entsprach dies meist einem kleinen Bereich der anatomisch nahen rechten Kolonflexur (vgl. Abb. 15l), für die Milz meist der linken Kolonflexur (vgl. Abb. 16f). Die häufigste FK der Wirbelkörper befand sich am 7. Halswirbelkörper (vgl. Abb. 17a).

Oftmals kam es in den automatischen Segmentierungen zu einer ungenauen Bestimmung der Organgrenzen. Bei der Milz hingegen wurden auch größere Organteile mehrfach nicht miterfasst.

3.3.3 Genauigkeit

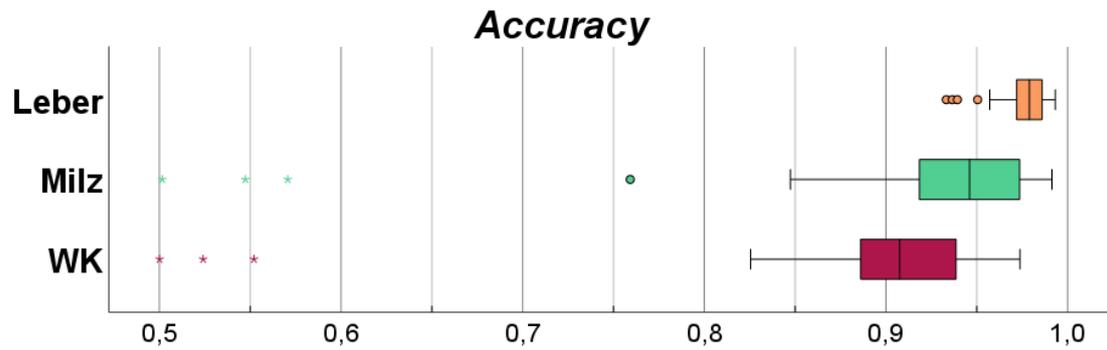


Abbildung 23: **Boxplot über Treffergenauigkeit (accuracy) der automatischen Segmentierungen, organweise.**

Tabelle 5: **Treffergenauigkeit (accuracy), organweise.**

Accuracy	Mittelwert	SD	Median	Minimum	Maximum
Leber	97,95 %	1,30 %	97,90 %	93,33 %	99,33 %
Milz	92,40 %	9,10 %	94,61 %	50,16 %	99,14 %
Wirbelkörper	89,42 %	8,59 %	90,75 %	50,01 %	97,38 %

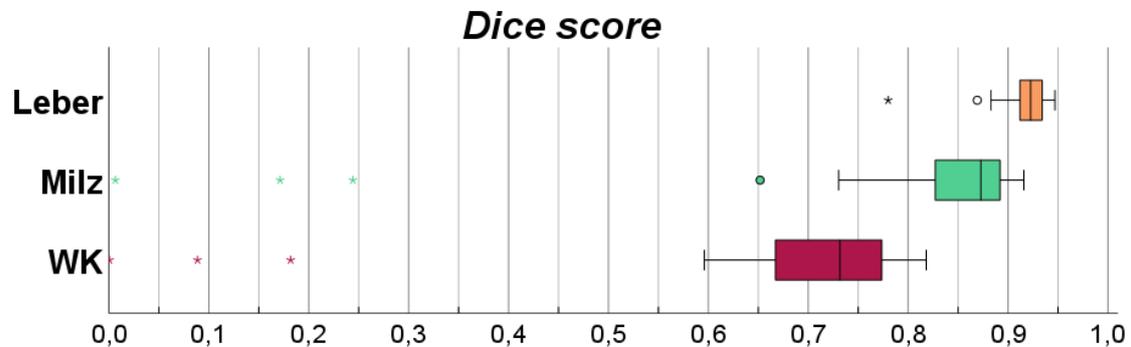


Abbildung 24: **Boxplot über dice score (DSC) der automatischen Segmentierungen, organweise.**

Tabelle 6: **Dice score (DSC), organweise.**

DSC	Mittelwert	SD	Median	Minimum	Maximum
Leber	0,9187	0,235	0,9227	0,7801	0,9470
Milz	0,8282	0,1552	0,8728	0,0065	0,9160
Wirbelkörper	0,6966	0,1415	0,7319	0,0006	0,8181

Die beste accuracy erreichte die Leber mit einem Mittelwert von 97,95 %, danach folgten die Milz mit 92,40 % und die Wirbelkörper mit 89,42 %. Den besten DSC erreichte ebenso die Leber mit 0,9187, gefolgt von der Milz mit 0,8282 und die Wirbelkörper mit 0,6966.

Sowohl bei der accuracy als auch beim DSC waren bei Milz und WK jeweils drei große Ausreißer zu erkennen, welche den sechs Datensätzen mit visual score 1 bei fehlender Organerkennung (vgl. Abb. 21a-c) entsprechen. Ein großer Ausreißer beim DSC der Leber entsprach ebenfalls einem Datensatz mit großem fehlenden Organteil (vgl. Abb. 21d) in Kombination mit großen groben FK, welchem ein visual score von 1 vergeben wurde.

Weitere kleinere Ausreißer entsprachen Datensätzen mit groben FK mit einem visual score von 2.

3.4 Abgeleitete PET/MRT-Parameter

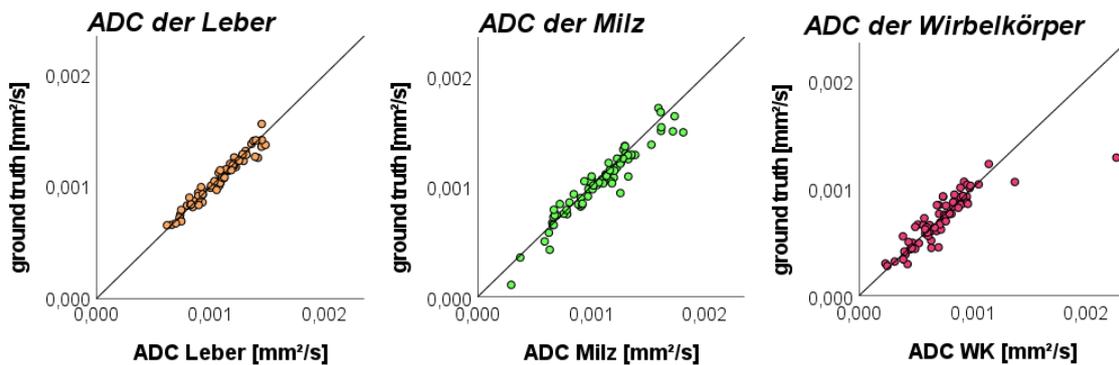


Abbildung 25: Verteilung des automatisch abgeleiteten ADC (X) gegen die ground truth (Y) mit der Winkelhalbierenden $y = x$ als Sollgerade.

Tabelle 7: ADC, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson. Alle Korrelationen signifikant, Niveau $< 0,001$.

		p (SW-Test)	Normalverteilung	r	ρ	
ADC	Leber	manuell	0,3468	Ja	0,977	0,981
		automatisch	0,1208	Ja		
	Milz	manuell	0,7510	Ja	0,958	0,956
		automatisch	0,5137	Ja		
	WK	manuell	0,5488	Ja	0,85	0,917
		automatisch	$< 0,001$	Nein		

Tabelle 8: Durchschnittlicher ADC, absolute Werte in [mm³/s], abgeleitet über der manuellen und automatisch generierten Maske, jeweils mit absoluter und relativer Differenz (Diff), organweise.

ADC	Leber				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	1,07E-03	2,20E-04	1,10E-03	0,64E-03	1,56E-03
automatisch	1,08E-03	2,24E-04	1,11E-03	0,62E-03	1,49E-03
abs. Diff	3,14E-05	2,04E-04	3,30E-06	-0,11E-03	1,66E-03
rel. Diff	-2,20 %	24,18 %	0,18 %	-198,34 %	12,49 %
Milz					
manuell	1,03E-03	3,15E-04	1,05E-03	0,11E-03	1,71E-03
Automatisch	1,08E-03	3,25E-04	1,12E-03	0,29E-03	1,81E-03
abs. Diff	6,74E-05	2,50E-04	3,82E-05	-1,17E-04	1,98E-03
rel. Diff	3,33 %	33,95 %	3,23 %	-204,02 %	175,53 %
Wirbelkörper					
manuell	7,15E-04	2,35E-04	7,01E-04	2,75E-04	12,83E-04
automatisch	7,24E-04	2,85E-04	7,09E-04	2,31E-04	22,70E-04
abs. Diff	2,87E-05	2,04E-04	1,18E-05	-1,83E-04	1,16E-03
rel. Diff	-0,78 %	30,06 %	0,53 %	-201,11 %	76,90 %

Der über der automatisch erstellten Maske abgeleitete ADC wies über allen Organen eine signifikante Korrelation zu der ground truth mit maximal $r = 0,981$ für die normalverteilte Leber bzw. minimal $p = 0,917$ für die nicht-normalverteilten WK auf. Abb. 25 zeigt eine homogene Verteilung entlang der Sollgeraden. Ein Ausreißer in den WK korrelierte mit einem Datensatz mit fehlender Organerkennung und visual score 1/5 (s. Abb. 21c). Der Betrag der relativen Abweichung von der ground truth betrug über alle drei Organe im Mittel $2,10 \% \pm 29,40 \%$.

Die Wirbelkörper zeigten beim ADC im Mittel die kleinste relative Differenz. Im Median bot die Leber die kleinste relative Differenz und die kleinste Standardabweichung.

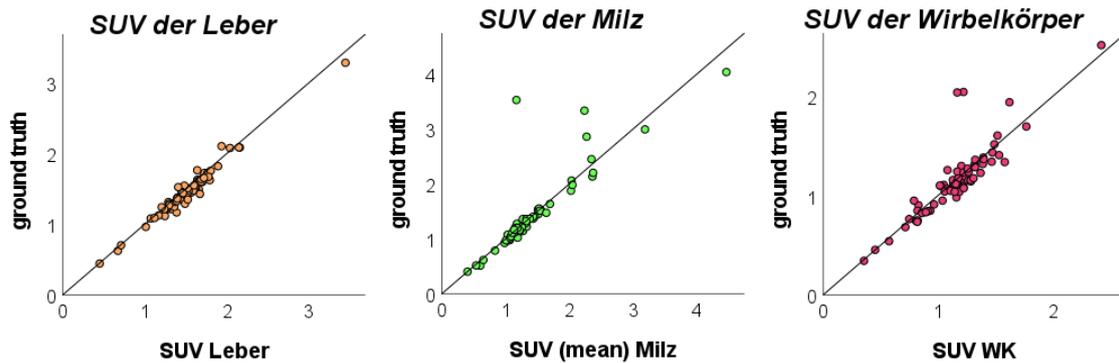


Abbildung 26: Verteilung des automatisch abgeleiteten SUV (X) gegen die ground truth (Y) mit Sollgerade.

Tabelle 9: SUV, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson. Korrelationen signifikant, Niveau $< 0,001$.

		p (SW-Test)	Normalverteilung	r	ρ	
SUV	Leber	manuell	$< 0,001$	Nein	0,981	0,944
		automatisch	$< 0,001$	Nein		
	Milz	manuell	$< 0,001$	Nein	0,862	0,922
		automatisch	$< 0,001$	Nein		
	WK	manuell	$< 0,001$	Nein	0,872	0,890
		automatisch	$< 0,005$	Nein		

Tabelle 10: Durchschnittlicher SUV, dimensionslos, organweise.

SUV	Leber				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	1,4557	0,3738	1,4473	0,4429	3,2811
automatisch	1,5140	0,3773	1,5064	0,4503	3,4425
abs. Diff	0,0564	0,0733	0,0498	-0,1673	0,2442
rel. Diff	4,10 %	5,05 %	3,63 %	-7,96 %	18,97 %
	Milz				
manuell	1,4067	0,6474	1,2095	0,4022	4,0286
Automatisch	1,3950	0,5860	1,2370	0,4032	4,4546
abs. Diff	-0,0226	0,3316	0,0241	-2,3535	0,4259
rel. Diff	0,83 %	10,86 %	2,02 %	-66,83 %	17,88 %
	Wirbelkörper				
manuell	1,1424	0,3510	1,1360	0,3427	2,5106
automatisch	1,1558	0,3019	1,1661	0,3553	2,4190
abs. Diff	0,0048	0,1708	0,0331	-0,8688	0,2437
rel. Diff	1,91 %	10,42 %	2,88 %	-42,70 %	18,26 %

Die abgeleitete SUV wies ebenfalls über allen drei Organen eine signifikante Korrelation zur ground truth mit maximal $\rho = 0,944$ bei der Leber und minimal $\rho = 0,890$ bei den WK auf. Bis auf jeweils drei Ausreißer in Milz und WK zeigt Abb. 26 eine homogene Verteilung entlang der Sollgeraden. Die sechs Ausreißer korrelierten mit sehr schlechter bzw. keiner Organerkennung mit einem visual score von 1/5 (s. Abb. 21). Die durchschnittliche relative Differenz zur ground truth betrug über alle drei Organe $2,28 \% \pm 8,78 \%$. Das positive Vorzeichen der relativen Differenz ließ auf eine Überschätzung der SUV schließen.

Im Mittel war die relative Differenz des SUV bei der Milz am geringsten, die niedrigste Standardabweichung erreichte der SUV der Leber.

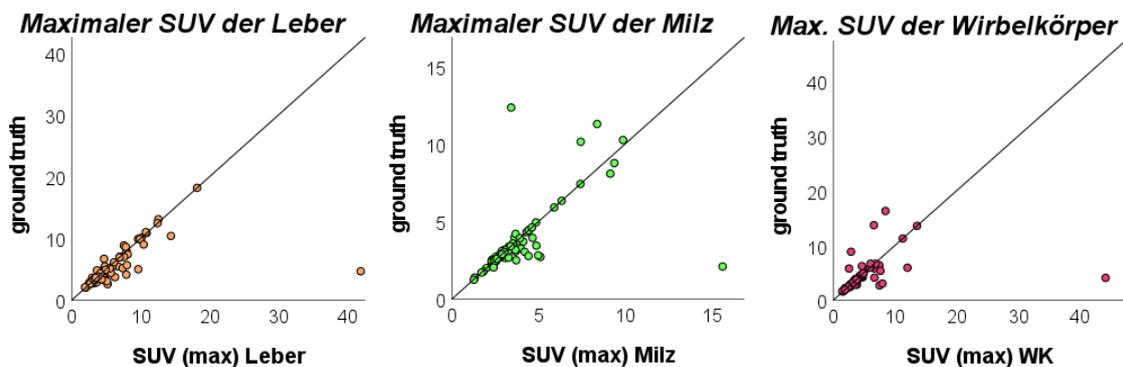


Abbildung 27: Verteilung des automatisch abgeleiteten maximalen SUV (X) gegen die ground truth (Y) mit Sollgerade

Tabelle 11: Maximaler SUV, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson.

Korrelationen signifikant, Niveau < 0,001.

			p (SW-Test)	Normalverteilung	r	ρ
SUVmax	Leber	manuell	< 0,001	Nein	0,522	0,862
		automatisch	< 0,001	Nein		
	Milz	manuell	< 0,001	Nein	0,586	0,803
		automatisch	< 0,001	Nein		
	WK	manuell	< 0,001	Nein	0,287	0,758
		automatisch	< 0,001	Nein		

Tabelle 12: *Maximaler SUV, dimensionslos, organweise*

SUV max	Leber				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	5,6376	3,4012	4,6537	2,0226	18,5985
automatisch	6,4280	5,3233	4,8595	1,9189	41,8962
abs. Diff	0,9265	4,5491	0,0000	-1,9933	37,2256
rel. Diff	21,15 %	97,52 %	0,00 %	-29,83 %	797,03 %
SUV max	Milz				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	3,6951	2,2391	2,9862	1,2391	12,3165
Automatisch	3,9028	2,2980	3,2339	1,2631	15,5878
abs. Diff	0,1883	2,0766	0,0000	-8,9081	13,5001
rel. Diff	14,25 %	79,58 %	0,00 %	-72,33 %	646,68 %
SUV max	Wirbelkörper				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	4,5125	2,6588	3,8102	1,5703	16,2639
automatisch	5,0974	5,2561	3,8516	1,4658	44,1023
abs. Diff	0,5418	5,1692	0,0000	-7,8121	40,0366
rel. Diff	19,39 %	122,50 %	0,00 %	-67,57 %	984,73 %

Der abgeleitete maximale SUV wies über allen Organen eine signifikante Korrelation zur ground truth mit maximal $\rho = 0,862$ bei der Leber und minimal $\rho = 0,758$ bei den WK auf. Abb. 27 zeigt einige Ausreißer unter der Sollgeraden, die vor allem mit kleinen bis großen FK ins Herz, in die Nieren und ins Gehirn korrelierten. Die drei großen Ausreißer über der Sollgeraden entsprachen hingegen den Datensätzen mit sehr schlechter bzw. fehlender Organerkennung mit einem visual score von 1/5 (s. Abb. 21). Insgesamt zeigte sich über alle drei Organe eine mittlere relative Differenz von $18,26 \% \pm 99,87 \%$. Ein positives Vorzeichen bei der mittleren relativen Differenz beschrieb eine Überschätzung der maximalen SUV. Die mediane relative Differenz betrug bei allen Organen 0. Die niedrigste mittlere relative Differenz der maximalen SUV und Standardabweichung fanden sich bei der Milz.

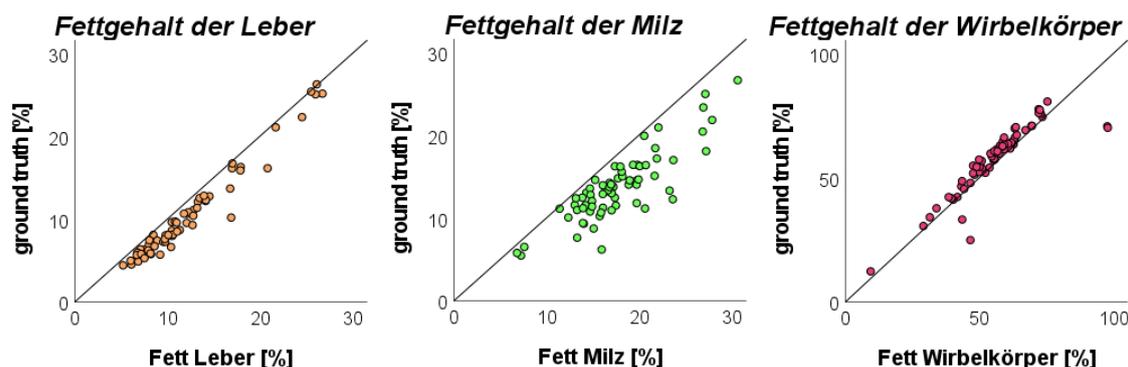


Abbildung 28: Verteilung des automatisch abgeleiteten Fettgehalts der Organe (X) gegen die ground truth (Y) mit Sollgerade.

Tabelle 13: Fettgehalt, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson. Korrelationen signifikant, Niveau $< 0,001$.

		p (SW-Test)	Normalverteilung	r	ρ	
Fat	Leber	manuell	$< 0,001$	Nein	0,981	0,965
		automatisch	$< 0,001$	Nein		
	Milz	manuell	$< 0,01$	Nein	0,854	0,800
		automatisch	$< 0,05$	Nein		
	WK	manuell	$< 0,005$	Nein	0,884	0,977
		automatisch	$< 0,005$	Nein		

Tabelle 14: Fettgehalt der Organe, organweise.

Fat	Leber				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	10,3529 %	5,8919 %	8,1014 %	4,4111 %	30,3489 %
automatisch	11,7342 %	5,3284 %	10,4271 %	5,2102 %	26,7399 %
abs. Diff	1,76 %	1,04 %	1,63 %	-0,12 %	6,71 %
rel. Diff	21,78 %	13,56 %	21,48 %	-0,47 %	65,95 %
Fat	Milz				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	14,1016 %	4,9455 %	13,4402 %	5,4559 %	37,6863 %
Automatisch	17,6909 %	4,6482 %	17,0113 %	6,8070 %	30,6375 %
abs. Diff	3,96 %	2,42 %	3,44 %	0,30 %	11,39 %
rel. Diff	31,92 %	25,11 %	24,79 %	2,73 %	158,30 %
Fat	Wirbelkörper				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	57,6078 %	12,7578 %	60,0143 %	12,2749 %	80,5703 %
automatisch	55,7306 %	13,4951 %	55,8998 %	9,3893 %	97,7890 %
abs. Diff	-1,87 %	6,35 %	-2,94 %	-7,40 %	27,79 %
rel. Diff	-2,79 %	14,44 %	-5,20 %	-23,51 %	87,40 %

Der abgeleitete Fettgehalt wies über allen Organen eine signifikante Korrelation zur ground truth mit maximal $\rho = 0,977$ bei den WK und minimal $\rho = 0,800$ bei der Milz auf. In Abb. 28 lässt sich erkennen, dass die Werte für Leber und Milz eher überschätzt, die Werte für die WK hingegen eher konstant unterschätzt wurden. Die drei Werte der WK unterhalb der Sollgeraden entsprachen den drei Datensätzen mit fehlender Organerkennung und visual score 1/5 (s. Abb. 21). Im Mittel betrug der Betrag der relativen Differenz über alle Organe $18,83\% \pm 17,70\%$. Die niedrigste mittlere relative Differenz fand sich beim Fettgehalt der WK, ein negatives Vorzeichen zeigte eine Unterschätzung der Werte für die WK. Die niedrigste Standardabweichung erreichte der Fettgehalt der Leber.

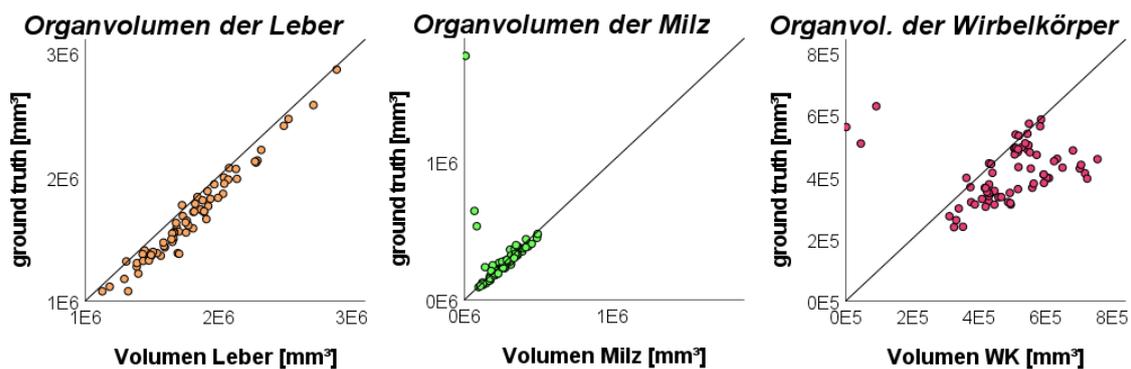


Abbildung 29: Verteilung des automatisch abgeleiteten Organvolumens (X) gegen die ground truth (Y) mit Sollgerade

Tabelle 15: Organvolumen, Shapiro-Wilk-Test und Korrelationskoeffizienten r nach Spearman und ρ nach Pearson. Korrelationen auf dem Niveau $< 0,001$ signifikant, mit Stern (*) markierte Korrelationen sind nicht signifikant.

		p (SW-Test)	Normalverteilung	r	ρ	
Volume	Leber	manuell	$< 0,005$	Nein	0,981	0,972
		automatisch	$< 0,05$	Nein		
	Milz	manuell	$< 0,001$	Nein	0,048*	0,711
		automatisch	0,8941	Ja		
	WK	manuell	0,4839	Ja	0,140*	0,437
		automatisch	$< 0,001$	Nein		

Tabelle 16: **Organvolumen, absolute Werte in [mm³], organweise**

Volume	Leber				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	1,65E6	3,67E5	1,60E6	8,71E5	2,86E6
automatisch	1,79E6	3,45E5	1,74E6	1,13E6	2,89E6
abs. Diff	1,21E5	6,85E+04	1,18E5	-4,29E4	3,27E5
rel. Diff	7,73 %	4,99 %	7,56 %	-2,42 %	23,68 %
	Milz				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	2,89E5	2,08E5	2,65E5	9,24E4	1,77E6
Automatisch	2,65E5	1,05E5	2,74E5	8,46E3	4,96E5
abs. Diff	-2,91E4	2,29E5	1,72E4	-1,76E6	5,93E4
rel. Diff	1,10 %	22,59 %	6,89 %	-99,52 %	23,76 %
	Wirbelkörper				
	Mittelwert	SD	Median	Minimum	Maximum
manuell	4,04E5	9,68E4	4,01E5	1,21E5	6,26E5
automatisch	4,89E5	1,39E5	5,08E5	1,69E3	7,57E5
abs. Diff	7,79E4	1,54E5	8,29E4	-5,58E5	3,32E5
rel. Diff	22,73 %	32,92 %	24,58 %	-99,70 %	84,42 %

Die berechneten Organvolumina wiesen für alle drei Organe eine signifikante Korrelation zur ground truth mit maximal $\rho = 0,972$ bei der Leber und minimal $\rho = 0,437$ bei den WK auf. In Abb. 29 zeigt sich eine homogenere Verteilung der Volumina bei Leber und Milz gegenüber den Wirbelkörpern. Die Werte lagen größtenteils unterhalb der Sollgeraden und wurden demzufolge überschätzt. Jeweils drei große Ausreißer fanden sich bei Milz und WK, welche den sechs Datensätzen mit fehlender Organerkennung und visual score 1/5 entsprachen (s. Abb. 21).

Im Mittel betrug die relative Differenz des Organvolumens für alle drei Organe $10,52 \% \pm 20,17 \%$. Die durchschnittlich kleinste Differenz bot das Volumen der Milz, die kleinste Standardabweichung zeigte das Volumen der Leber.

3.5 Ableitung einer Therapieresponse

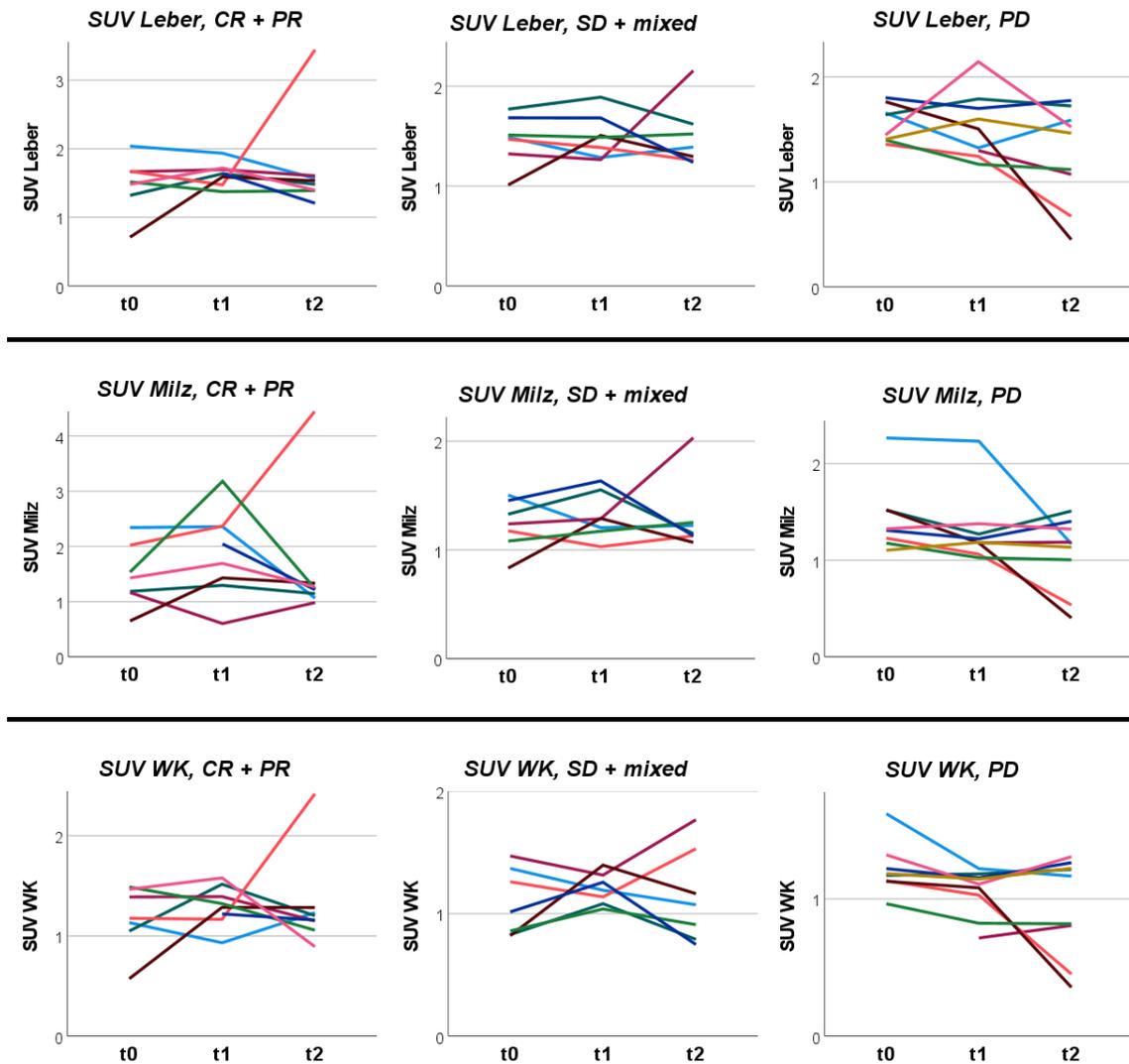


Abbildung 30: **Zeitlicher Verlauf der mittleren SUV** (dimensionslos) nach Organ und Response, abgeleitet über die automatisch erstellten Masken. Jede Linie entspricht einem Patienten, der zu den Zeitpunkten t_0 (vor Therapie) t_1 (2 Wochen nach Therapie) und t_2 (drei Monate nach Therapie) jeweils eine PET/MRT-Bildgebung durchlief. Responder links (complete response CR, partial response PR), stable disease SD und mixed response mittig, Non-Responder rechts (progressive disease PD).

Zur Prüfung auf einen signifikanten Unterschied des zeitlichen Verlaufs des SUV zwischen Respondern und Non-Respondern, wurde zunächst die Differenz SUV_{diff} zwischen dem SUV zum Zeitpunkt t_2 und t_0 berechnet. Es wurden Shapiro-Wilk-Tests zur Verteilungsanalyse der Differenzen SUV_{diff} angewandt (Tab. 17).

Tabelle 17: SUV_{diff} als Differenz der SUV-Werte 3 Monate nach Therapiebeginn, Shapiro-Wilk-Tests auf Normalverteilung

		p (SW-Test)	Normalverteilung
SUV_{diff}	Leber	0,012	Nein
	Milz	0,002	Nein
	WK	0,244	Ja

Da 2 Aufnahmen von der Studie ausgeschlossen wurden, konnten die Differenzen des SUV von 3 Organen von $n = 22$ Patienten berechnet werden, von denen 9 Non-Responder (PD), 8 Responder (CR + PR) und 7 stable disease (SD + mixed) zugeordnet wurden (Abb. 31).

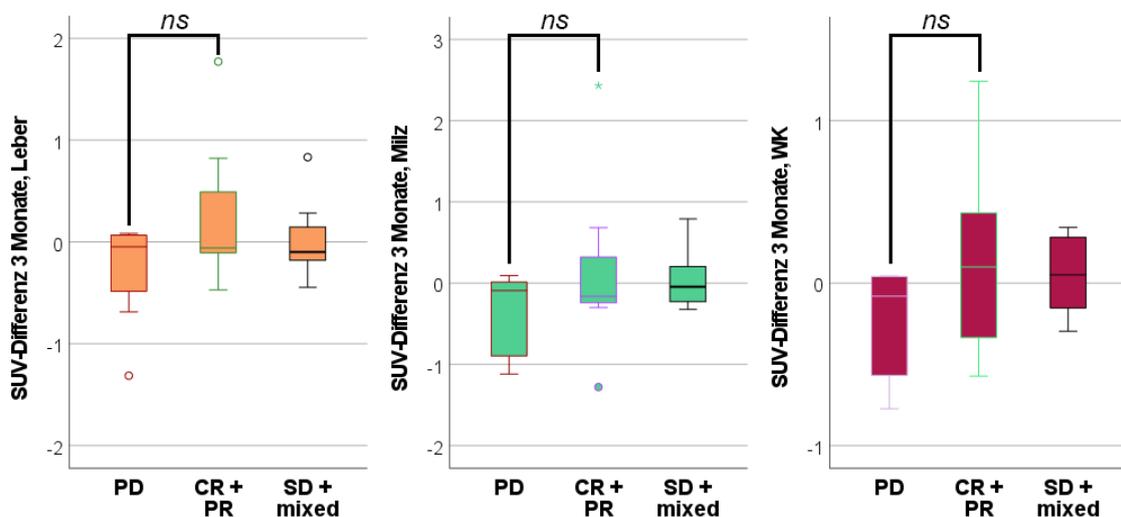


Abbildung 31: Differenzen des mittleren SUV (dimensionslos) nach 3 Monaten, aufgeteilt nach Organ und Response. Die Differenzen zwischen den Respondern und Non-Respondern über allen Organen war non-signifikant (ns).

Bei den Respondern (CR + PR) kam es im Schnitt zu einem Anstieg des SUV 3 Monate nach Therapiebeginn mit Immuntherapeutika in Leber

($M = 0,287$, $SD = 0,765$), Milz ($M = 0,164$, $SD = 1,153$) und Wirbelkörpern ($M = 0,138$, $SD = 0,648$).

Bei den Non-Respondern (PD) hingegen kam es im Schnitt zu einem Abfall des SUV 3 Monate nach Therapiebeginn mit Immuntherapeutika in Leber ($M = -0,270$, $SD = 0,495$), Milz ($M = -0,371$, $SD = 0,518$) und Wirbelkörpern ($M = -0,243$, $SD = 0,342$).

Der SUV war nach 3 Monaten bei den Respondern durchschnittlich um 0,557 (Leber) / 0,535 (Milz) / 0,380 (WK) höher als bei den Non-Respondern (Leber: 95%-CI[-0,195;1,310]), Milz: (95%-CI[-0,553;1,624]), WK: (95%-CI[-0,240;1,000]).

Es gab keinen signifikanten Unterschied zwischen den SUV-Differenzen bei Respondern und Non-Respondern bei der Leber ($t(10,04) = 1,649$, $p = 0,130$), der Milz ($t(8,08) = 1,132$, $p = 0,290$) und den Wirbelkörpern ($t(8,83) = 1,391$, $p = 0,198$).

4 Diskussion

4.1 Zusammenfassung der Ergebnisse

Eine automatische Segmentierungspipeline konnte erfolgreich mittels maschinellen Lernens mit DCNet erstellt werden. Über eine 7-fache Kreuzvalidierung konnte das Netz erfolgreich an 70 Datensätzen trainiert und validiert, und in 70 Datensätzen Leber, Milz und Wirbelkörper automatisch segmentiert werden.

Nach dem Training betrug die mittlere Zeit zur Erstellung der automatischen Segmentierungen 4 min, wohingegen für die manuellen Segmentierungen eine reine Arbeitszeit von etwa 49 h anfiel.

Bei der visuellen Beurteilung der automatischen Segmentierung erreichte die Milz den besten Score, gefolgt von Wirbelköpern und Leber. Den besten dice score (DSC) erreichte die Leber mit 0,9187, gefolgt von der Milz mit 0,8282 und den Wirbelkörper mit 0,6966.

Anatomische und funktionelle Parameter wurden erfolgreich über den segmentierten Organen abgeleitet. Es fand sich in unserem Patientenkollektiv in Leber, Milz und Wirbelkörpern kein signifikanter Unterschied des SUV nach drei Monaten Immuntherapie bei Respondern versus Non-Respondern.

4.2 Diskussion der Ergebnisse und Methoden

DCNet erzielte für die Segmentierung von Leber und Milz vergleichbar gute Ergebnisse wie in der Literatur vorgeschriebene Algorithmen. Für die Segmentierung in T1w-Aufnahmen mittels CNNs fanden sich DSCs zwischen 0,892 und 0,956 für die Leber, sowie zwischen 0,700 und 0,945 für die Milz. (Kart et al., 2021, Conze et al., 2021).

Für die Segmentierung der Wirbelkörper existieren bisher nur wenige Daten. Die meisten Studien beschrieben neuronale Netze, welche mit Wirbeln in CT-Aufnahmen trainiert und validiert wurden, jedoch hohe DSCs von 0,943 erreichen konnten (Vania et al., 2019). Li et al. und Lessmann et al. stellten Algorithmen zur Segmentierung der Lumbalwirbelkörper ohne Quer- und Dornfortsätze in T2w-Aufnahmen vor und erreichten sehr hohe DSCs von 0,980 bzw. 0,949 (Li et

al., 2021, Lessmann et al., 2019). Turella et al. erreichte für die Segmentierung von ganzen Lumbalwirbeln in T1w-Aufnahmen einen DSC von 0,866 (Turella et al., 2021).

Durch eine fehlende Standardisierung in der Methodik zur Erstellung der automatischen Segmentierungen finden sich große Unterschiede und Schwierigkeiten in deren Evaluation.

Bereits bei der Auswahl der Bildgebungsmodalität finden sich diverse Unterschiede. Im Direktvergleich verschiedener CNNs treten bei denselben Patienten bessere Ergebnisse bei der kompletten Organsegmentierung in CT- gegenüber MRT-Bildern auf. Zurückzuführen ist dies womöglich auf meist geringere Auflösungen der MRT-Aufnahmen und niedrigere Kontraste (Conze et al., 2021), insbesondere bei Knochenaufnahmen. Auch eine schwankende Bildqualität bei Magnetfeldinhomogenität könnte eine Rolle spielen. Demgegenüber kann die kürzere Untersuchungsdauer der CT-Untersuchung mit einer Minderung von Bewegungsartefakten einhergehen. Die feste Einteilung der CT-Kontraststufen in die Hounsfield-Skala ermöglichen eine robustere und interindividuell vergleichbare Bildgebung, die eine höhere Segmentierungsqualität bedingen kann.

Zwischen den MRT-Untersuchungen finden sich viele Unterschiede in den Aufnahmeparametern, beispielsweise bezüglich Feldstärken, Auflösungen, und Sequenzen. Zwar wurden viele große Studien zur Segmentierung der Bauchorgane an T1w-Sequenzen durchgeführt (Kart et al., 2021, Conze et al., 2021, Küstner et al., 2018), jedoch finden sich beispielsweise für die Milz und die Nieren durch ihre anatomische Lage stärkere Kontraste, und damit akkuratere Ergebnisse in T2w-Aufnahmen (Conze et al., 2021). Auch bei der Segmentierung der Wirbel bzw. Wirbelkörper scheinen die T2w-Aufnahmen aufgrund höherer Kontraste besser abzuschneiden. Obwohl in dieser Studie die DSC-Werte der Wirbelkörper niedriger als die in der Literatur vorbeschriebenen Werte zu sein scheinen, ist hervorzuheben, dass die oben genannten Studien nur Lumbalwirbel bzw. -Wirbelkörper einschlossen. Dadurch wurden im Vorhinein andere Sequenzen mit höherer Auflösung verwendet, sowie komplexere Strukturen wie

das Os sacrum nicht berücksichtigt. Bei einer komplexen Struktur wie der Wirbelsäule, welche aus vielen untereinander getrennten kleinen Wirbeln besteht, ist es nicht verwunderlich, dass die Segmentierungsgenauigkeit gegenüber großen Organen wie der Leber schlechter ausfiel.

Die Robustheit von DCNet unter verschiedenen Feldstärken, Empfängerspulen, MRT-Sequenzen und Patientenpositionen haben Küstner et al. bereits für die Segmentierung von viszeralem und subkutanem Fettgewebe gezeigt, bei der sehr gute Ergebnisse (DSC 0,94) erzielt wurden (Küstner et al., 2020). Im Gegensatz dazu wurde in unserer Studie die Leistung bei kleineren und anatomisch komplexeren Organen untersucht, wohingegen das Fettgewebe als große und homogenere Struktur leichter zu segmentieren schien (Küstner et al., 2020).

Die besten vorbeschriebenen Segmentierungen wurden mit einem auf nnU-Net-basierenden Algorithmus (Isensee et al., 2021, Kart et al., 2021) erstellt. Die aktuell als „state-of-the-art“ bezeichnete DL-Technik stellt in sich keine neue Netzarchitektur dar, sondern erzielt vielmehr durch vollautomatische Eigenkonfigurierung und Eigenoptimierung eine Systematisierung der sonst aufwändigen manuellen Konfiguration (Isensee et al., 2021). Zuvor wurden zur Effizienzsteigerung des Trainingsprozesses unterschiedlicher CNNs manuell verschiedene Optimierungsalgorithmen wie beispielsweise *Adam* (Zhou et al., 2018) oder *RMSprop* (Küstner et al., 2020, Kim et al., 2020) implementiert.

Doch neben den Trainingsalgorithmen spielen auch die Trainingsdaten eine bedeutende Rolle. Zur Bewertung ihrer Genauigkeit wurden in der Literatur vorwiegend CNNs evaluiert, die an gesunden Patienten und Organen trainiert und validiert wurden, und deren DSCs über Segmentierungen gesunder Organe berechnet wurden (Kart et al., 2021, Conze et al., 2021, Li et al., 2021, Lessmann et al., 2019). Zwar gibt es viele Studien zur Identifizierung und Segmentierung maligner Melanome, jedoch fand sich nach gründlicher Recherche zum Zeitpunkt der Erstellung dieses Manuskripts noch keine Studie zur Organsegmentierung bei Patienten mit metastasiertem malignem Melanom. Letztlich stellen die Metastasierungen eine Schwierigkeit in der Organerkennung dar. Große Fernmetastasen in Milz und Knochen sorgten in einigen Datensätzen dafür, dass

eine adäquate Segmentierung nicht vonstattengehen konnte, und präsentierten sich in den folgenden Ergebnissen als persistierende Ausreißer. An dieser Stelle empfiehlt sich die Validierung der Methodik an einem größeren Patientenkollektiv.

Retrospektiv lässt sich trotz ausgiebiger Kontrollen an drei Trainingsdatensätzen eine Segmentierung des siebten Halswirbelkörpers in den WK-Daten erkennen, der aus der Segmentierung ausgenommen sein sollte. Als zusätzlicher Faktor könnte dies zu einer niedrigeren Segmentierungsqualität beigetragen haben. Der örtlichen Verteilung der FK nach zu urteilen (Tab. 4), fanden sich die meisten FK der WK an HWK 7. Sollten diese FK in den visual scores vernachlässigt werden, würden 17 Datensätze mit einem höheren score eingeordnet werden, unter denen 15 den höchsten Score von 5/5 erreicht hätten. Die starke Abhängigkeit der Ergebnisse und Vulnerabilität durch Beeinflussung von Trainingsdaten wurde bereits vielfach aufgezeigt (Chan et al., 2020). Zunächst wurde die Überschätzung des Organvolumens der WK auch auf die teilweise manuelle Übersegmentierung der Halswirbelkörper zurückgeführt, jedoch fanden die Überschätzungen in fast allen Datensätzen einschließlich der übersegmentierten statt. Aus diesem Grund wurde auf eine Wiederholung des Trainings mit korrigierten Trainingsdaten verzichtet.

Über die automatisch segmentierten Masken wurden erfolgreich anatomische und funktionelle Parameter aus den PET/MRT-Untersuchungen abgeleitet. Bis auf das Organvolumen der Wirbelkörper erreichten die Segmentierungen durchweg sehr hohe Korrelationen. Grundsätzlich wurden, bis auf oben erwähnte Ausreißer, die Organvolumina gegenüber der ground truth überschätzt, was mit dem abgeleiteten erhöhten Fettgehalt in Leber und Milz und erniedrigten Fettgehalt der Wirbelkörper korreliert. Bei der intraperitonealen Lage von Leber und Milz bewirkt die Segmentierung über die Organgrenzen unter anderem die Erfassung des fettreichen Peritoneums, während die Übersegmentierung der Wirbelkörper eher eine Erfassung des fettarmen vertebrealen Bandapparats beinhaltet. Bei akkuraterer Segmentierung mit höherem DSC in Leber und Milz

wurden in diesen Organen auch höhere Korrelationen und niedrigere relative Differenzen gegenüber der ground truth erreicht. Die Leber bot gegenüber den anderen Organen die niedrigste Standardabweichung der Parameter, was auf eine robustere Segmentierung hindeuten könnte. Damit hängt auch zusammen, dass die Leber in Relation zu ihrem Volumen kürzere Organgrenzen aufwies, die überschritten werden konnten.

Die Ableitung der mittleren SUV-Werte war sehr zuverlässig und wies über allen Organen eine geringe relative Differenz zu ground truth auf. Es fielen teilweise deutliche Überschätzungen des maximalen SUV auf. Diese korrelierten jedoch alle mit zum Teil sehr kleinen FK in Strukturen mit sehr hohem Glukosemetabolismus bzw. hoher Traceraufnahme, wie Herz, Gehirn oder Niere.

In unserem Patientenkollektiv ließ sich kein signifikanter Unterschied im dreimonatigen SUV-Verlauf von Respondern versus Non-Respondern bei immuntherapeutisch behandelten Melanompatienten feststellen.

Zwar ist der SUV in allen drei Organen beim 3-Monats-Follow-up im Mittel bei den Respondern gestiegen und bei den Non-Respondern gesunken, was sich mit den Hinweisen in der Literatur decken würde (Schwenck et al., 2020). Das 95%-Konfidenzintervall beinhaltet jedoch ebenso negative Differenzen, welche für den gegenteiligen Verlauf sprechen könnten.

Auch Schwenck et al. beschrieben die Tendenz zum höheren Glukosemetabolismus unter Immuntherapie über der Milz bei jedoch non-signifikanter Differenz zwischen Respondern und Non-Respondern. Neben der Zellanreicherung in der Milz zeigten sie auch die Aktivierung des Knochenmarks auf. Im Achsenskelett wurde über einen signifikanten Anstieg des Glukosemetabolismus im PET/CT nach Immuntherapie berichtet (Schwenck et al., 2020).

Bei unserem kleinen Patientenkollektiv ($n = 20$) gilt es, oben genannte Hinweise in einer größeren Kohorte zu validieren, um zu prüfen, inwiefern der SUV geeignet ist, um ein Therapieansprechen ableiten zu können. Es bietet sich an, im gleichen Zuge auch Patienten heranzuziehen, die aufgrund anderer

Tumorentitäten – abseits des malignen Melanoms – mit Immuntherapeutika behandelt werden.

Andere Ansätze bezüglich einer non-invasiven Responseableitung unter Immuntherapie beschäftigten sich mit der automatischen Ableitung mittels Radiomics. Diese mit Big Data verknüpfte Technik eröffnet ein weites Feld mit großen Möglichkeiten in der radiologischen Grundlagenforschung (Lambin et al., 2017). Trebeschi et al. erstellten und validierten Radiomics-Algorithmen zur Responseableitung in CT-Bildern bei Melanompatienten und erreichten ein signifikantes, wenn auch nur schwaches Ergebnis von 0,55 bis 0,64 AUC (Trebeschi et al., 2019). Es stellt sich zudem die Frage, inwiefern non-invasive Verfahren geeignet sind, um solch eine Fragestellung zu beantworten. Unter PD-1/PD-L1-Inhibition kommt es in verschiedenen Tumorentitäten in rund 10 % der Fälle zur pseudoprogredienten oder hyperprogredienten Entwicklung (Ferrara et al., 2018). Dieses Phänomen beschreibt ein paradoxes Tumorwachstum unter Immuntherapie und sollte bei der Responseableitung stets mitbedacht werden. Bei klinisch sehr heterogenem Erscheinungsbild des Melanoms könnten Algorithmen unter Einbeziehung weiterer klinischer Parameter, unter anderem auch Laborwerte, Histo- und Zytologiebefunde wegweisend werden (Yan et al., 2019).

4.4 Limitationen

Die oben angesprochenen Unterschiede und Probleme decken sich gut mit den Limitationen dieser Studie.

Die hier präsentierten Ergebnisse sind modalitätsgebunden und spezifisch für die erfolgte Bildgebung nach Studienprotokoll, wodurch nur von einer eingeschränkten Generalisierbarkeit ausgegangen werden kann. Für die Segmentierungsgenauigkeit von DCNet gilt, dass die Leber und Milz als große Organe leichter zu segmentieren sind als kleinere und anatomisch komplexere Strukturen, welche in folgenden Studien angegangen werden könnten.

Weiterhin besteht eine große Abhängigkeit gegenüber den Trainingsdaten hinsichtlich ihrer Qualität und Anzahl. Bei den hier eingesetzten 70

Trainingsdatensätzen handelt es sich um ein kleines Kollektiv, welches zudem auf Patienten mit fortgeschrittenen malignen Melanomen beschränkt ist. Gerade im Hinblick auf Artefakte oder interindividuelle anatomische und pathophysiologische Unterschiede würde die Robustheit der Segmentierungen von größeren und diverseren Datenmengen profitieren. Eine Situation mit weniger verfügbaren Trainingsdaten ist in der klinischen Routine vorstellbar und könnte im Weiteren geprüft werden.

Die Ableitung der funktionellen und anatomischen PET/MRT-Parameter ist stark abhängig von der Genauigkeit der Segmentierungen, sodass alle oben genannten Limitationen auch an dieser Stelle zutreffen. Unsere Studie gilt als proof of concept und sollte an einem größeren Datensatz validiert werden.

4.3 Herausforderungen im klinischen Einsatz von Deep Learning

Bei der Nutzung computergestützter Diagnostik ergeben sich verschiedene Hindernisse und Problematiken, welche vor Implementierung in die klinische Routine unbedingt abgeklärt werden sollten. Dabei liegen neben den technischen Herausforderungen ebenso rechtliche und moralische Hürden vor.

Ein wesentlicher, häufig genannter Aspekt ist das Fehlen von Trainingsdaten. Bei kleinen Datenmengen ist die Qualität der Trainingsdaten bei großer Abhängigkeit umso entscheidender. Durch den weiten Einsatz von medizinischen Bildarchiven mittels PACS-Systeme in den Kliniken stünden heutzutage eine enorme Anzahl von Bilddaten zur Verfügung, jedoch ist der Zugriff auf diese für Forschungszwecke beschränkt. Bei der Erhebung der Bilddaten wäre eine Aufklärung zur Verwendung zu Trainingszwecken nötig. Auch bei der pseudonymisierten Weiterverarbeitung dieser Patientendaten wäre zuvor eine Genehmigung durch eine Ethikkommission einzuholen. Zwar steht heutzutage eine Vielzahl an open-source-Bibliotheken zur Weiterverwendung offen, doch wäre eine größere Vielfalt zum akkurateren Training der DL-Systeme durch zusätzliche Daten sehr wünschenswert.

Ein großer Ansatz diesem Problem entgegenzutreten ist das föderale Lernen. Durch multiinstitutionelle Kollaboration können mehrere beteiligte Institute ihren lokalen Algorithmus mit eigenen Daten trainieren, und ihren neuen trainierten Algorithmus auf einen Server hochladen. In diesem werden alle zugeführten Daten akkumuliert und zu einem einheitlichen Algorithmus zusammengeführt, welcher den Instituten zurückgesendet werden kann (Sheller et al., 2020). Erhebliche Vorteile finden sich neben der großen Datenmenge darin, dass keine privaten kritischen Daten ausgetauscht werden, und dadurch rechtliche Hürden verringert werden. Währenddessen können sich serverseitig Schwierigkeiten in der Aggregation der Netze bei etwa um 1 % schlechteren Ergebnissen gegenüber konventionellen Trainingsmethoden ergeben (Chang et al., 2018, Sheller et al., 2020).

Eine weitere Problematik bestünde darin, dass viele der verfügbaren oder potenziell verfügbaren Bilddaten nicht annotiert wären. Auch wenn in den PACS-Systemen häufig eine Befundung beiliegt, müsste ein Großteil der Ressourcen in der zeitintensiven Überführung der Befundtexte in ein verwertbares Format, wie beispielsweise Segmentierungen, eingesetzt werden (Litjens et al., 2017). Wie zeitaufwendig diese sein können, haben wir demonstriert. Durch „AI-Assisted Annotation“ Software wie *NVIDIA Clara*, die mittlerweile auch in MITK Workbench implementiert wurde (German Cancer Research Center, 2022), ist es möglich, semiautomatische Segmentierungen durch vortrainierte Algorithmen zu erstellen. Die Segmentierungen können nach Erstellung bearbeitet und adjustiert werden, sodass die nachträglichen Änderungen an einen Server übermittelt werden, um die zukünftigen Segmentierungen zu verbessern. Ebenso können vorhandene Algorithmen durch sogenanntes Transferlernen mit bereits annotierten Daten augmentiert werden (NVIDIA Corporation, 2022)

Wichtig zu verstehen ist, dass DL-Algorithmen bisher eine Assistenz in der Diagnostik darstellen. Die Methoden wurden in der Literatur vielfach als „black boxes“ beschrieben (Litjens et al., 2017), da es schwierig zu verstehen ist, was zwischen den verschiedenen Layern geschieht. In einem Feld wie der Medizin, in der Verantwortung eine juristische Rolle spielen kann, stellt sich die Frage, inwieweit die Ersteller einer Diagnosesoftware Verantwortung übernehmen

können. Diese Überlegung führte zu Forschungen mit dem Ziel, verschiedene Arten der Unsicherheiten zu quantifizieren, um Fehlerquellen in DL-Netzen zu lokalisieren (Kendall and Gal, 2017). Eine Implementierung in bestehende Algorithmen könnte schnell fehlerhafte Netzarchitekturen herausfiltern und dabei helfen, DL-basierte Anwendungen schneller in die klinische Routine zu integrieren (Litjens et al., 2017, Saba et al., 2019).

4.5 Ausblick

Die Erwartungen an die Zukunft, in der die computergestützte Diagnostik ihren Einsatz in der klinischen Routine findet, sind in den letzten Jahren stark gestiegen. Durch den rapiden technischen Fortschritt in Prozessoren- und Grafikchipleistung wurde erst der Einsatz von tiefen neuronalen Netzwerken ermöglicht. Heutzutage bieten eine Vielzahl von open-source-Bibliotheken Zugriff auf sowohl digitale medizinische Bilddaten, als auch Frameworks, wie Caffe, Torch und TensorFlow, welche zum maschinellen Lernen verwendet werden (Choy et al., 2018).

DL begleitet uns in unserem Alltag in vielen Formen. Von selbstfahrenden Autos zu sozialen Medien bis hin zu Forschungszwecken, beispielsweise in der Bioinformatik, findet das DL heutzutage breite Anwendung. Eine Implementierung in ein technisches Fach wie die Radiologie liegt daher nahe. Nennenswerte Anwendungserfolge sind beispielsweise die Vorhersage des Outcomes mit und ohne rtPA-Therapie eines akuten Schlaganfalls im MRT (Nielsen et al., 2018), die Unterscheidung von Karzinom und non-kanzerösem Prostatagewebe im MRT (Song et al., 2018) oder die histologische Klassifikation von Lungentumoren in CT-Bildern (Chaunzwa et al., 2021).

Einige Autoren spekulierten bereits seit einigen Jahren auf eine Übernahme eines Großteils der Arbeit der Radiologen durch künstliche Intelligenzen in absehbarer Zeit (Obermeyer and Emanuel, 2016). Chan et al. beschreiben die computergestützte Diagnosesoftware mittels DL, trotz der vielversprechenden Ergebnisse mit ihren Millionen an Gewichten eher als ein sehr komplexes mathematisches Modell und weit davon entfernt, intelligent zu sein. Durch die erfolgreiche Anwendung von maschinellem Lernen in anderen Bereichen wie im

Verkehrswesen oder im Marketing, seien die Erwartung in der Anwendung in der Bildgebung stark gestiegen (Chan et al., 2020). Dennoch steigt die Anzahl der Publikationen in diesem Bereich rasant. Der nächste angezeigte Schritt könnte die Implementierung in den Klinikalltag mit der Entwicklung allseits verfügbarer und einfach bedienbarer Algorithmen sein. Nach der Veröffentlichung des vollautomatisierten nnU-Nets, das zudem auch noch hervorragende Ergebnisse präsentiert, scheint dieser Schritt nicht mehr weit entfernt zu sein.

5 Zusammenfassung

Durch die Automatisierung der medizinischen Bildgebung und deren Analyse durch Anwendung maschineller Lernverfahren können potenziell Objektivität und Reliabilität gesichert, sowie Zeit und Ressourcen gespart werden. Ein großer Anwendungsbereich findet sich in der Onkologie, besonders bei der Erkennung und Klassifizierung von Tumoren, deren Phänotyp und deren Malignitätsstatus. Weiterführende Studien beschreiben die Nutzung zur Voraussage über Therapieansprechen, Prognosebestimmung und Risikoanalyse. Den Grundbaustein hierfür bildet die semantische Segmentierung, bei der Organe und Strukturen in Bildgebungen voxelweise klassifiziert werden, um den Suchbereich für weitere Analysen zu definieren.

In der vorliegenden Arbeit wurden PET/MRT-Untersuchungen von Patienten mit malignem Melanom unter Immuntherapie betrachtet. Die frühe und genaue Vorhersage des Therapieansprechens kann gerade in diesem Patientenkollektiv entscheidend sein, da die Ansprechraten mäßig sind und die Therapie mit signifikanten Nebenwirkungen und hohen Kosten verbunden ist. Anatomische und funktionelle Parameter der PET/MRT-Bildgebung können potenziell genutzt werden, um ein Therapieansprechen einzuschätzen und therapeutische Entscheidungen zu unterstützen.

Ziel dieser Arbeit war die automatische Extraktion anatomischer und funktioneller Parameter aus multiparametrischen Ganzkörper-PET/MRT-Daten mittels maschineller Lernverfahren. Dazu wurde ein Algorithmus aus der Klasse der Convolutional Neural Networks (CNN) trainiert und validiert.

Das Patientenkollektiv setzte sich aus 24 Melanompatienten unter Immuntherapie zusammen, die jeweils drei zeitlich versetzte PET/MRT-Untersuchungen durchliefen. In der Summe ergab dies 72 Bildgebungen, von denen zwei aufgrund von Artefakten und Importfehlern von der Studie ausgeschlossen wurden.

Anhand manuell segmentierter Trainingsdaten, welche auf T1-gewichteten MRT-Bildern erzeugt wurden, konnte das CNN für die automatische Segmentierung von Milz, Knochenmark der Wirbelsäule und Leber trainiert werden.

Die quantitative Evaluation dieser automatischen Organsegmentierung im Vergleich zur manuellen Organsegmentierung ergab eine insgesamt hohe Genauigkeit. Die Leber erreichte die höchste Genauigkeit (DSC 0,9187), gefolgt von der Milz (DSC 0,8282) und den Wirbelkörpern (DSC 0,6966). Dabei konnte der Zeitaufwand der Segmentierung von ca. 1 Stunde pro Datensatz (manuell) auf etwa 4 Sekunden (automatisch) beschleunigt werden.

Aus den automatischen Segmentierungen extrahierte Parameter (ADC, durchschnittlicher und maximaler SUV, Fettgehalt und Organvolumen) zeigten eine hohe Übereinstimmung zur manuellen Segmentierung. Beim durchschnittlichen SUV wurde eine mittlere relative Abweichung von $4,10 \% \pm 5,05 \%$ bei der Leber, $0,83 \% \pm 10,86 \%$ bei der Milz und $1,91 \% \pm 10,42 \%$ bei den Wirbelkörpern erreicht.

Als proof of concept wurde geprüft, ob sich der SUV (jeweils über Leber, Milz und Wirbelkörpern) als Parameter eignet, um einen Responsestatus der Immuntherapie abzuleiten. Obwohl sich Hinweise auf den Anstieg des SUV unter Immuntherapie über Leber, Milz und Wirbelkörper fanden, konnte in unserem Patientenkollektiv keine Signifikanz nachgewiesen werden.

Zusammenfassend lässt sich sagen, dass eine automatische Ableitung anatomischer und funktioneller Parameter aus multiparametrischen PET/MRT Daten mittels Deep Learning möglich ist und zufriedenstellende Ergebnisse erreicht wurden. Voraussetzung sind ausreichende und gute Trainingsdaten. Es empfiehlt sich, die Methodik an einem größeren und diverseren Patientenkollektiv zu validieren.

Anhang

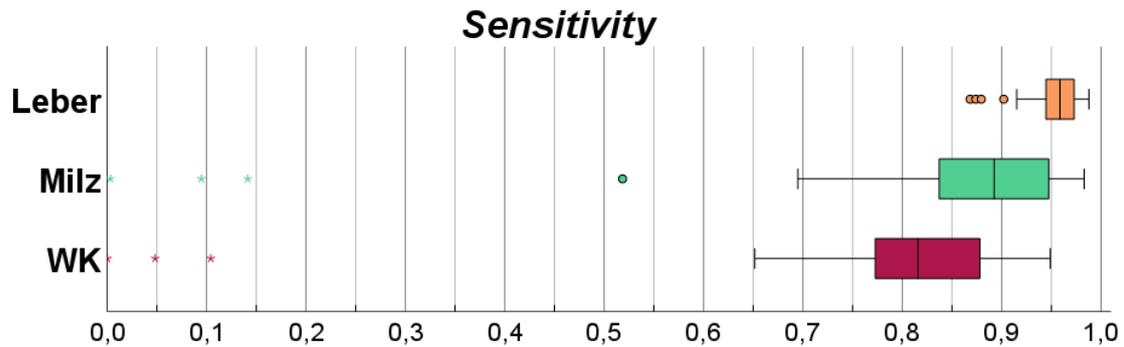


Abbildung 32: **Boxplot über Sensitivität (sensitivity) der automatischen Segmentierungen, organweise**

Tabelle 18: **Sensitivität (sensitivity), organweise**

Sensitivity	Mittelwert	SD	Median	Minimum	Maximum
Leber	95,41 %	2,60 %	95,87 %	86,83 %	98,79 %
Milz	84,82 %	18,21 %	89,25 %	00,33 %	98,31 %
Wirbelkörper	78,92 %	17,20 %	81,59 %	00,03 %	94,90 %

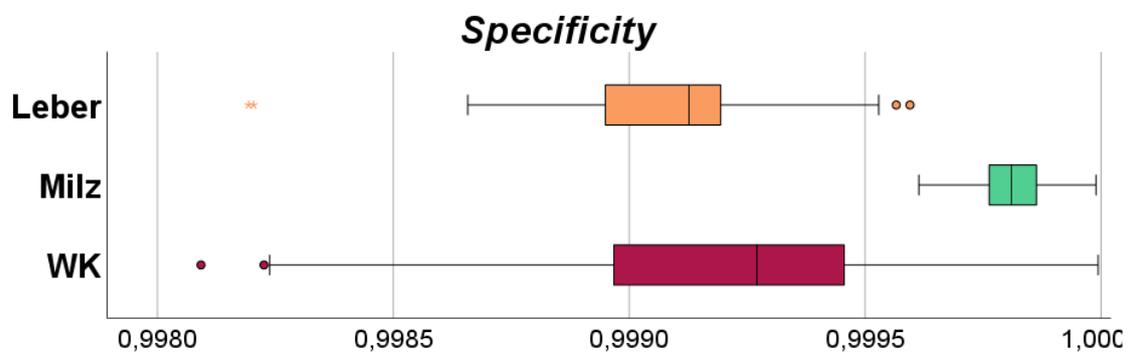


Abbildung 33: **Boxplot über Spezifität (specificity) der automatischen Segmentierungen, organweise**

Tabelle 19: **Spezifität (specificity), organweise**

Specificity	Mittelwert	SD	Median	Minimum	Maximum
Leber	99,91 %	0,25 %	99,91 %	99,82 %	99,96 %
Milz	99,98 %	0,01 %	99,98 %	99,96 %	100 %
Wirbelkörper	99,92 %	0,04 %	99,93 %	99,81 %	100 %

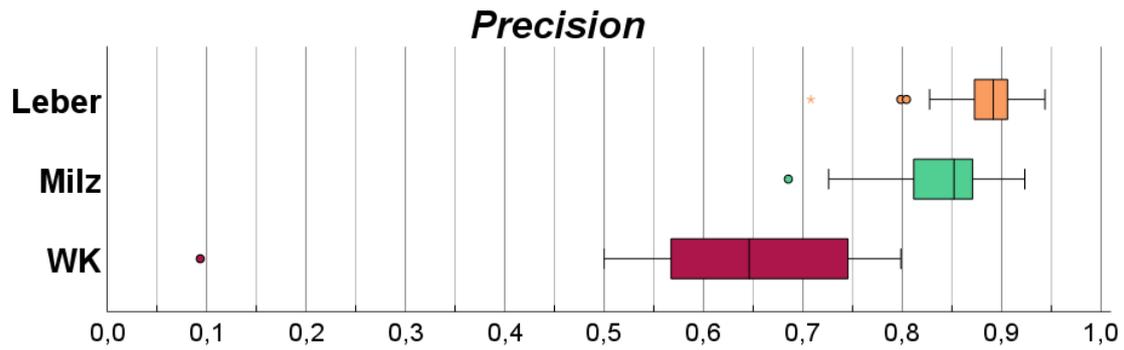


Abbildung 34: **Boxplot über positiven Vorhersagewert (precision) der automatischen Segmentierungen, organweise**

Tabelle 20: **Positive Vorhersagewerte (precision), organweise**

Precision	Mittelwert	SD	Median	Minimum	Maximum
Leber	88,67 %	3,40 %	89,16 %	70,81 %	94,37 %
Milz	84,09 %	5,00 %	85,21 %	68,54 %	92,33 %
Wirbelkörper	64,63 %	11,20 %	64,60 %	9,38 %	79,87 %

Tabelle 21: **Beurteilung der Fehlklassifikationen und Vergabe des visual scores, organweise, nach Datensatz. Ein score von 1 entspricht einer schlechten, ein score von 5 einer sehr guten Segmentierung. Die FK wurden der Größe und Lokalisation nach beurteilt. Es wurde zwischen kleinen (FK1 <10 Voxel), mittelgroßen (FK2, 10-19 Voxel), und großen FK (FK3, ≥20 Voxel) unterschieden und die betreffenden Bilder (c = koronar) vermerkt (s. 2.5.1).**

Index	Leber	Milz	WK	Leber	Milz	Wirbelkörper
1	4	1	1	c62-63 Magen (FK1) c64-70 ungenau im Bereich um V. cava (Segment I, FK1)	nahezu keine Organerkennung	nahezu keine Organerkennung
2	4	1	1	c28-29 Rippe + Bauchdecke re. (FK1), c40-45 Herz (FK1), c64-70 um V. cava leicht ungenau (FK1)	nahezu keine Organerkennung	nahezu keine Organerkennung
3	4	1	1	c64-73 V. cava ungenau, c94-95 untere Rippe (FK1)	nahezu keine Organerkennung	nahezu keine Organerkennung
4	3	3	5	c48-53 re. Arm (FK1), c46-58 Herz (FK3), c60-63 Darm nah (FK1)	c72-83 Magen (FK3)	einwandfrei
5	4	3	5	c60-63 Gallenblase (Segment V, FK1)	c56-63 Organ fehlt (FK3)	einwandfrei

Anhang

6	4	3	5	c60-63 Gallenblase (Segment V, FK1)	c72-83 Magen (FK3)	einwandfrei
7	4	5	3	c26-31 Rippe (FK1), c58-75 Bauchwand re. (FK1/2), c70-71 Herz (FK1)	einwandfrei	c60-63 Aorta thoracica, c78-79 Rippe rechts (FK1), c90-91 Bandscheibe Th5/6 (FK1)
8	4	5	5	c24-31 Bauchwand (FK1/2), c28-29 Rippe (FK1), c64-71 Bauchwand re. (FK1/2),	einwandfrei	einwandfrei
9	2	5	3	c23-31 fehlendes Organ ventral (FK3), c32-39 Magen (FK1), c36-45 Darm Becken fern (FK1), c40-41 Rippe (FK1), c51-63 fehlendes Organ kaudal (FK3), c44-75 Bauchwand re. (FK1/2)	einwandfrei	c66-67 Herz (FK1), c75-79 Rippe rechts (FK1)
10	3	5	4	c40-43 Bauchwand (FK1), c44-45 Magen (FK1), c58-59 Magen (FK1), c64-71 Herz (FK3),	einwandfrei	c64-71 HWK7 (FK2), c90-95 li. L5 Proc. transversus (FK1)
11	4	3	4	c53-59 Gallenblase (FK2), C64-65 Herz (FK1), c64-67 Magen (FK1), c68-69 Herz (FK1)	c56-63 Herz (FK3)	c68-73 HWK7 (FK2)
12	4	4	5	c48-54 Gallenblase/Pankreas (FK2), c64-71 Darm/Bauchwand) (FK1)	c60-63 Darm LOQ (FK1), c80-82 li. oberer Nierenpol angrenzend (FK1)	einwandfrei
13	3	4	5	c36-45 Magen (FK3), c66-67 re. Niere (FK1)	c64-71 Darm angrenzend (FK2)	einwandfrei
14	3	4	5	c30-31 li. Thorax (FK2)	c54-64 Darm LOQ (FK1)	einwandfrei
15	4	5	4	c64-70 Darm (FK2)	einwandfrei	c68-73 HWK7 (FK1)
16	3	4	3	c32-43 Bauchwand rechts unten (FK1), c32-57 grobe FK3 angrenzend kaudal, c34-37 kraniale Lebergrenzen (FK2), c70-77 V. cava inf. (FK1)	c60-63 Zwerchfell angrenzend (FK1), c68-69 Magen (FK1)	c64-75 HWK7 (FK3)
17	4	2	3	c47-49 Magen (FK1), c64-67 Darm (FK1), c74-75 re. oberer Nierenpol (FK1)	c50-63 nicht erkanntes ventrales Organ (FK3)	c64-71 HWK7 (FK3), c68-69 Milz (FK1)
18	2	4	5	c48-51 Pankreas (FK1), c52-79 Darm (FK3, sehr grob und diffus), c68-71 re. oberer Nierenpol (FK1),	c68-69 Magen (FK1)	einwandfrei
19	4	5	5	c49-57 Gallenblase (FK1/2), c64-67 Magen (FK1)	einwandfrei	einwandfrei
20	3	5	4	c46-49 Darm (FK1), c52-55 Gallenblase (FK1), c64-73 Magen (FK1/2), c64-75 Colon ascendens (FK3)	einwandfrei	c76-83 Bandscheibe Th6/7 (FK1)
21	4	4	5	c60-63 Darm (FK1)	c64-75 Zwerchfell (FK1)	einwandfrei
22	3	5	5	c32-36 Organ schlecht erkannt (FK3), c50-63 Gallenblase (FK2)	einwandfrei	einwandfrei

Anhang

23	4	3	5	c46-47 Darm (FK1)	c54-63 ventrales Organ fehlt (FK3), c60-71 Zwerchfell (FK1/2)	einwandfrei
24	3	5	5	c32-63 ungenaue Grenzen Segment II+III und V+VI (FK3), c84-87 oberer re. Nierenpol (FK1)	einwandfrei	einwandfrei
25	3	4	5	c32-75 kaudale Lebergrenzen Segment 4b+V+VI unscharf (FK2/3)	c90-95 oberer li. Nierenpol (FK1)	einwandfrei
26	3	5	5	c52-63 kaudales Organ Segment VI schlecht (FK3) (c78-79 + c88-89 Milz FK1?)	einwandfrei	einwandfrei
27	5	5	3	einwandfrei	einwandfrei	c58-63 li. Niere (FK1), c78-81 HWK7 (FK1), c94-95 Milz (FK1)
28	3	4	3	c44-53 c56-57 Herz (FK1), c64-72 Leber Segment VI teils nicht erkannt (FK3)	c81-89 unsauberer Rand kaudal (FK2)	c50-57 li. Niere und Lunge re. (FK2) c60-63 Zwerchfell li. (FK1)
29	3	4	3	c32-69 Magen (FK3)	c84-95 unsauberer Rand kaudal (FK2)	c62-63 Aorta abdominalis (FK1), c70-71 re. Niere (FK2)
30	5	3	4	einwandfrei	c76-89 Darm LOQ (FK2/3)	c82-85 Bandscheiben Th6-10 ventral (FK1)
31	5	2	5	einwandfrei	c45-63 ventrales Organ nicht erkannt (FK3)	einwandfrei
32	4	5	4	c64-65 Herz (FK1)	einwandfrei	c54-71 HWK7 (FK2)
33	3	5	4	c42-49 Magen (FK3), c57-63 Segment II/III ungenau (FK3), c66-67 Aorta th. (FK1)	einwandfrei	c78-83 L5 paravertebral li. (FK1),
34	3	5	4	c42-57 Magen (FK1/2), c56-63 Segment II/III ungenau (FK3),	einwandfrei	c74-77 li. Os ilium
35	2	5	4	c36-63 Darm diffus (FK1/2), c36-63 fehlendes Organteil mittig (FK3), c50-63 Segment II/III fehlt (FK3)	einwandfrei	c64-67 HWK7 (FK2)
36	4	4	3	c36-39 Herz (FK2)	c76-89 oberer li. Nierenpol (FK2)	c64-65 HWK7 (FK2), c84-88 li. Teil Th6-8 schlecht (FK3), c84-90 Bandscheibe Th7/8 (FK2), c75-83 L4/L5 ungenau (FK2) c68-84 L3 nicht erkannt (FK3), c80-87 Bandscheiben Th6-8 (FK2) + li. Teil Th6-8, Os sacrum sehr unsauber
37	4	4	2	c32-33 Herz (FK1), c58-75 Bauchwand (FK1/2), c72-79 kaudaler Rand unsauber (FK1)	c49-54 Darm LOQ (FK1), c79-88 oberer li. Nierenpol (FK1)	c64-65 HWK7 (FK1), c70-80 L3 schlecht erkannt (FK3), c82-89 Bandscheiben Th6-8 (FK2)
38	3	4	3	c57-88 Organ kaudal abgeschnitten (FK1/2) c64-67 A. pulmonalis dex. (FK1)	c60-63 lat. Grenzen ungenau (FK2), c74-91 Darm LOQ (FK2)	c64-65 HWK7 (FK1), c70-80 L3 schlecht erkannt (FK3), c82-89 Bandscheiben Th6-8 (FK2)
39	3	5	5	c30-49 Darm (FK2), c38-39 Magen (FK1), c42-45 Magen (FK1), c62-63 Darm ROQ (FK1), c68-69 Darm ROQ (FK1)	einwandfrei	einwandfrei
40	3	5	5	c45-64 Segment IVa+b fehlen teilweise (FK3), c65-71 Herz (FK1)	einwandfrei	einwandfrei

Anhang

41	3	4	5	c56-63 Segment IVa+b ungenau (FK3), c64-69 Herz (FK1)	c58-63 Darm LOQ (FK2)	einwandfrei
42	3	4	4	c42-47 Herz (FK1/2), c56-63 Darm LUQ (FK1)	c88-95 kraniale Grenze + Zwerchfell (FK2)	c66-69 HWK7 (FK1)
43	4	5	4	c41-45 Darm (FK2), c59-61 mediokaudale Grenze ungenau (FK2)	einwandfrei	c72-73 HWK7 (FK1)
44	4	5	4	c64-65 Herz + kaudale Grenzen ungenau (FK1)	einwandfrei	c64-71 HWK7 (FK1)
45	2	3	4	c26-31 schlechte Organgrenzen (FK3), c49-69 Segment IVa+b schlecht (FK2/3), c64-69 unterer re. Nierenpol (FK1), c64-85 li. lat. Oberschenkel (FK3)	c56-61 re. Niere (FK1)	c70-73 HWK7 (FK1)
46	3	5	4	c56-63 Herz (FK1/2), c56-63 kaudal leicht abgeschnitten (FK3)	einwandfrei	c72-73 HWK7 (FK1)
47	3	5	4	c28-31 lateral u. kranial leicht abgeschnitten (FK3), c52-62 Segment II+III abgeschnitten (FK3)	einwandfrei	c72-73 HWK7 (FK1)
48	4	5	4	c24-25 Thorax (FK1)	einwandfrei	c52-57 HWK7 (FK1/2), c72-75 Os sacrum re. unsauber (FK1)
49	3	5	4	c64-76 li. Leberlappen ungenau (FK3)	einwandfrei	c64-69 HWK7 (FK1/2)
50	3	5	3	c62-63 Darm ROQ (FK1), c66-67 Herz (FK1), c64-75 li. Leberlappen ungenau (FK2/3)	einwandfrei	c62-63 Aortenbogen (FK1)
51	2	5	4	c50-59 li. Leberlappen teilweise nicht erkannt, c52-68 Darm ROQ (FK3) + LUQ (FK1), c56-63 re. Lunge (FK3), c78-79 Milz (FK2)	einwandfrei	c72-82 Os ilium re.
52	2	5	5	c50-70 li. Leberlappen grob nicht erkannt (FK3)	einwandfrei	einwandfrei
53	3	5	3	c78-79 Milz (FK2), c88-89 re. Niere angrenzend (FK1)	einwandfrei	c80-83 Rektum (FK1)
54	3	4	4	c44-47 Herz (FK1), c65-68 li. Leberlappen (FK3), c78-79 re. Niere (FK1), c82-83 re. Niere (FK1)	c70-71 Darm LOQ (FK1), c80-81 lat. Rand (FK1), c93-95 Darm LOQ (FK1)	c64-69 HWK7 (FK1/2)
55	4	4	4	c44-45 Darm ROQ (FK1), c70-71 Gallenblase (FK1)	c94-95 lat. Rand Rippe (FK1)	c64-65 HWK7 (FK2)
56	3	3	4	c64-75 mediokaudal leicht abgeschnitten (FK3)	c58-63 Darm LOQ (FK2)	c60-65 HWK7 (FK1/2)
57	3	4	4	c44-45 Herz (FK1), c60-63 Darm LOQ (FK2), c96-97 Milz (FK2)	c96-98 lat. abgeschnitten (FK3)	c64-65 unsauberer ventraler Rand L4 (FK1)
58	3	3	2	c59-63 mediokaudal leicht abgeschnitten (FK3)	c96-99 lat. Abgeschnitten (FK3)	c96-105 Th7+8 nicht erkannt (FK3), c96-105 Os sacrum schlecht erkannt (FK3)
59	3	5	5	c57-63 kaudal ungenau (FK2), c58-63 Darm RUQ (FK3)	einwandfrei	einwandfrei

60	4	5	3	c54-64 lat. Rand zu viel (FK2), c62-63 Darm ROQ (FK1), c94-95 lat. Rippe (FK1)	einwandfrei	c76-81 Bandscheibe L4/L5 (FK1), c90-95 paravertebral li. (FK3)
61	4	5	3	c68-75 kaudal abgeschnitten (FK2)	einwandfrei	c64-72 L4 schlecht erkannt (FK3)
62	2	2	2	c4-31 diffus rechts (FK3), c52-63 kaudal abgeschnitten (FK3), c64-69 li. Leberlappen schlecht erkannt (FK2/3), c96-103 randständig abgeschnitten (FK3)	c76-89 med. viel abgeschnitten (FK3), c96-105 med. und kaudal viel abgeschnitten (FK3)	c0-19 diffus (FK3), c96-103 li. Oberschenkel (FK2), c110-115 re. Handgelenk (FK1), c120-121 Rücken rechts + median (FK2)
63	2	3	2	c0-15 li. auf Leberhöhe (FK3), c32-43 re. Rippen (FK2), c42-45 Gallenblase (FK1), c49-63 kaudal abgeschnitten (FK3)	c68-75 mediokranial abgeschnitten (FK3)	c0-19 diffus (FK3), c48-51 über li. Schulter (FK2), c76-77 Oberschenkelkopf li. (FK1), c84-91 Femur + lat. Oberschenkel (FK2), c102-107 Gluteus re. (FK2), c112-115 Rücken (FK2)
64	3	2	2	c0-7 li. auf Leberhöhe (FK3), c64-73 mediokranial abgeschnitten (FK2/3)	c70-93 medial viel abgeschnitten (FK3)	c0-7 diffus (FK3), c90-95 re. Os ilium (FK1), c101-111 diffus Rücken (FK2/3)
65	3	4	4	c34-39 li. Leberhöhe + Thorax (FK2), c44-47 LUQ Abdomen (FK1), c46-51 Darm ROQ (FK1)	c56-63 Darm LOQ (FK2), c90-91 med. Muskel (FK1)	c66-67 C5 (FK1), c92-95 Os sacrum ungenau (FK2)
66	3	5	4	c36-39 + c42-43 median Haut (FK1), c48-51 kaudale Grenzen ungenau (FK3), c60-63 re. Lunge (FK1), c58-63 Gallenblase (FK2), c78-81 Bandscheibe Th11/Th12 (FK1), c94-95 Haut Rücken (FK1)	einwandfrei	c56-59 Aorta abdominalis (FK1), c67-68 HWK7 (FK1), c86-87 Os ilium (FK1),
67	1	2	4	c32-51 diffus median betont (FK3), c52-55 ventral ungenau (FK1), c56-63 re. Drittel nicht erkannt (FK3), c58-63 kaudale Grenzen ungenau (FK3), c64-93 Gallenblase und Darm ROQ im Block erfasst (FK3), c64-91 Gehirn viel (FK3)	c83-95 kaudal viel abgeschnitten (FK3)	c94-95 Os sacrum ungenaue Ränder (FK1)
68	2	2	3	c32-53 Block links auf Leberhöhe (FK3), Organ einwandfrei	c66-75 Darm LOQ (FK1), c72-79 Darm LOQ (FK2), c72-95 Organ zentral nicht erkannt (FK3)	c88-103 L4 + L5 + Os sacrum nur teilweise erkannt (FK3)
69	2	2	3	c40-45 Block links auf Leberhöhe (FK3), c96-105 kraniallateral abgeschnitten (FK3)	c96-103 Organ dorsal randständig viel nicht erkannt (FK3)	c60-63 Aorta thoracica (FK1), c80-91 L4 + L5 nur teilweise erkannt (FK3)
70	2	2	3	c40-45 Rippen re. (FK3), c64-85 lateral abgeschnitten (FK2), c96-104 kraniallateral abgeschnitten (FK3)	c96-102 Organ dorsal randständig viel nicht erkannt (FK3)	c64-65 re. lat. Brustwand (FK1), c80-93 L4 + L5 incl. angrenzende Bandscheiben nur teilweise erkannt (FK3)

6 Literaturverzeichnis

- AKKUS YILDIRIM, B., ONAL, C., ERBAY, G., CEM GULER, O., KARADELI, E., REYHAN, M. & KOC, Z. 2019. Prognostic values of ADCmean and SUVmax of the primary tumour in cervical cancer patients treated with definitive chemoradiotherapy. *Journal of Obstetrics and Gynaecology*, 39, 224-230.
- AMRANE, K., LE GOUPIL, D., QUERE, G., DELCROIX, O., GOUVA, S., SCHICK, U., SALAUN, P. Y., ABGRAL, R., ALAVI, Z., KEROMNES, N. & QUERELLOU, S. 2019. Prediction of response to immune checkpoint inhibitor therapy using 18F-FDG PET/CT in patients with melanoma. *Medicine (Baltimore)*, 98, e16417.
- ANWAR, S. M., MAJID, M., QAYYUM, A., AWAIS, M., ALNOWAMI, M. & KHAN, M. K. 2018. Medical Image Analysis using Convolutional Neural Networks: A Review. *J Med Syst*, 42, 226.
- ANZAI, Y. 2020. ADC on Diffusion Imaging: The MR Analog of SUV on FDG-PET: Perfect Is the Enemy of Good. *J Magn Reson Imaging*, 51, 286-287.
- ASCIERTO, P. A., MCARTHUR, G. A., DRÉNO, B., ATKINSON, V., LISZKAY, G., DI GIACOMO, A. M., MANDALÀ, M., DEMIDOV, L., STROYAKOVSKIY, D., THOMAS, L., DE LA CRUZ-MERINO, L., DUTRIAUX, C., GARBE, C., YAN, Y., WONGCHENKO, M., CHANG, I., HSU, J. J., KORALEK, D. O., ROONEY, I., RIBAS, A. & LARKIN, J. 2016. Cobimetinib combined with vemurafenib in advanced BRAFV600-mutant melanoma (coBRIM): updated efficacy results from a randomised, double-blind, phase 3 trial. *The Lancet Oncology*, 17, 1248-1260.
- ASCIERTO, P. A., SCHADENDORF, D., BERKING, C., AGARWALA, S. S., VAN HERPEN, C. M. L., QUEIROLO, P., BLANK, C. U., HAUSCHILD, A., BECK, J. T., ST-PIERRE, A., NIAZI, F., WANDEL, S., PETERS, M., ZUBEL, A. & DUMMER, R. 2013. MEK162 for patients with advanced melanoma harbouring NRAS or Val600 BRAF mutations: a non-randomised, open-label phase 2 study. *The Lancet Oncology*, 14, 249-256.
- BALCH, C. M., GERSHENWALD, J. E., SOONG, S. J., THOMPSON, J. F., DING, S., BYRD, D. R., CASCINELLI, N., COCHRAN, A. J., COIT, D. G., EGGERMONT, A. M., JOHNSON, T., KIRKWOOD, J. M., LEONG, S. P., MCMASTERS, K. M., MIHM, M. C., JR., MORTON, D. L., ROSS, M. I. & SONDAK, V. K. 2010. Multivariate analysis of prognostic factors among 2,313 patients with stage III melanoma: comparison of nodal micrometastases versus macrometastases. *J Clin Oncol*, 28, 2452-9.
- BAUCKNEHT, M., BERTAGNA, F., DONEGANI, M. I., DURMO, R., MICELI, A., DE BIASI, V., LAUDICELLA, R., FORNARINI, G., BERRUTI, A., BALDARI, S., VERSARI, A., GIUBBINI, R., SAMBUCETI, G., MORBELLI, S. & ALBANO, D. 2021. The prognostic power of 18F-FDG PET/CT extends to estimating systemic treatment response duration in metastatic castration-resistant prostate cancer (mCRPC) patients. *Prostate Cancer Prostatic Dis.*

- BELMOUHAND, M., LOFGREN, J., JOHANNESSEN, H. H., BAEKSGAARD, L., GUTTE, H., TARIQ, K. & ACHIAM, M. P. 2019. Early response evaluation of neoadjuvant therapy with PET/MRI to predict resectability in patients with adenocarcinoma of the esophagogastric junction. *Abdom Radiol (NY)*, 44, 836-844.
- BELZUNCE, M. A., MEHRANIAN, A. & READER, A. J. 2019. Enhancement of Partial Volume Correction in MR-Guided PET Image Reconstruction by Using MRI Voxel Sizes. *IEEE Trans Radiat Plasma Med Sci*, 3, 315-326.
- BEUTHIEN-BAUMANN, B. 2018. [Principles of PET]. *Radiologe*, 58, 487-500.
- BOUTROS, C., TARHINI, A., ROUTIER, E., LAMBOTTE, O., LADURIE, F. L., CARBONNEL, F., IZZEDDINE, H., MARABELLE, A., CHAMPIAT, S., BERDELOU, A., LANOY, E., TEXIER, M., LIBENCIUC, C., EGGERMONT, A. M., SORIA, J. C., MATEUS, C. & ROBERT, C. 2016. Safety profiles of anti-CTLA-4 and anti-PD-1 antibodies alone and in combination. *Nat Rev Clin Oncol*, 13, 473-86.
- BRANUM, G. D., EPSTEIN, R. E., LEIGHT, G. S. & SEIGLER, H. F. 1991. The role of resection in the management of melanoma metastatic to the adrenal gland. *Surgery*, 109, 127-31.
- BRENDLIN, A. S., PEISEN, F., ALMANSOUR, H., AFAT, S., EIGENTLER, T., AMARAL, T., FABY, S., CALVARONS, A. F., NIKOLAOU, K. & OTHMAN, A. E. 2021. A Machine learning model trained on dual-energy CT radiomics significantly improves immunotherapy response prediction for patients with stage IV melanoma. *J Immunother Cancer*, 9.
- BROSKI, S. M., GOENKA, A. H., KEMP, B. J. & JOHNSON, G. B. 2018. Clinical PET/MRI: 2018 Update. *AJR Am J Roentgenol*, 211, 295-313.
- CARR, S., SMITH, C. & WERNBERG, J. 2020. Epidemiology and Risk Factors of Melanoma. *Surg Clin North Am*, 100, 1-12.
- CATANA, C. 2015. Motion correction options in PET/MRI. *Semin Nucl Med*, 45, 212-23.
- CHAN, H.-P., SAMALA, R. K., HADJIISKI, L. M. & ZHOU, C. 2020. Deep Learning in Medical Image Analysis. In: LEE, G. & FUJITA, H. (eds.) *Deep Learning in Medical Image Analysis : Challenges and Applications*. Cham: Springer International Publishing.
- CHANG, K., BALACHANDAR, N., LAM, C., YI, D., BROWN, J., BEERS, A., ROSEN, B., RUBIN, D. L. & KALPATHY-CRAMER, J. 2018. Distributed deep learning networks among institutions for medical imaging. *Journal of the American Medical Informatics Association*, 25, 945-954.
- CHAUNZWA, T. L., HOSNY, A., XU, Y., SHAFER, A., DIAO, N., LANUTI, M., CHRISTIANI, D. C., MAK, R. H. & AERTS, H. J. W. L. 2021. Deep learning classification of lung cancer histology using CT images. *Scientific Reports*, 11, 5471.
- CHEN, Y., RUAN, D., XIAO, J., WANG, L., SUN, B., SAOUAF, R., YANG, W., LI, D. & FAN, Z. 2020. Fully automated multiorgan segmentation in abdominal magnetic resonance imaging with deep neural networks. *Medical physics*, 47, 4971-4982.
- CHOY, G., KHALILZADEH, O., MICHALSKI, M., DO, S., SAMIR, A. E., PIANYKH, O. S., GEIS, J. R., PANDHARIPANDE, P. V., BRINK, J. A. &

- DREYER, K. J. 2018. Current Applications and Future Impact of Machine Learning in Radiology. *Radiology*, 288, 318-328.
- CONZE, P.-H., KAVUR, A. E., CORNEC-LE GALL, E., GEZER, N. S., LE MEUR, Y., SELVER, M. A. & ROUSSEAU, F. 2021. Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks. *Artificial Intelligence in Medicine*, 117, 102109.
- DECAZES, P. & BOHN, P. 2020. Immunotherapy by Immune Checkpoint Inhibitors and Nuclear Medicine Imaging: Current and Future Applications. *Cancers (Basel)*, 12.
- DIXON, W. T. 1984. Simple proton spectroscopic imaging. *Radiology*, 153, 189-94.
- EHMAN, E. C., JOHNSON, G. B., VILLANUEVA-MEYER, J. E., CHA, S., LEYNES, A. P., LARSON, P. E. Z. & HOPE, T. A. 2017. PET/MRI: Where might it replace PET/CT? *J Magn Reson Imaging*, 46, 1247-1262.
- EISENHAUER, E. A., THERASSE, P., BOGAERTS, J., SCHWARTZ, L. H., SARGENT, D., FORD, R., DANCEY, J., ARBUCK, S., GWYTHYER, S., MOONEY, M., RUBINSTEIN, L., SHANKAR, L., DODD, L., KAPLAN, R., LACOMBE, D. & VERWEIJ, J. 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*, 45, 228-47.
- FERRARA, R., MEZQUITA, L., TEXIER, M., LAHMAR, J., AUDIGIER-VALETTE, C., TESSONNIER, L., MAZIERES, J., ZALCMAN, G., BROSSEAU, S., LE MOULEC, S., LEROY, L., DUCHEMANN, B., LEFEBVRE, C., VEILLON, R., WESTEEL, V., KOSCIELNY, S., CHAMPIAT, S., FERTÉ, C., PLANCHARD, D., REMON, J., BOUCHER, M. E., GAZZAH, A., ADAM, J., BRIA, E., TORTORA, G., SORIA, J. C., BESSE, B. & CAMELLA, C. 2018. Hyperprogressive Disease in Patients With Advanced Non-Small Cell Lung Cancer Treated With PD-1/PD-L1 Inhibitors or With Single-Agent Chemotherapy. *JAMA Oncol*, 4, 1543-1552.
- FILOGRANA, L., MAGARELLI, N., CELLINI, F., MANFRIDA, S., LEONE, A., COLOSIMO, C. & VALENTINI, V. 2018. Diffusion weighted imaging (DWI) and apparent diffusion coefficient (ADC) values for detection of malignant vertebral bone marrow lesions. *Eur Rev Med Pharmacol Sci*, 22, 590-597.
- FUKUSHIMA, K. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, 193-202.
- GARCÍA, V., MOLLINEDA, R. A. & SÁNCHEZ, J. S. Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions. 2009 Berlin, Heidelberg. Springer Berlin Heidelberg, 441-448.
- GATIDIS, S., BENDER, B., REIMOLD, M. & SCHAFER, J. F. 2017. PET/MRI in children. *Eur J Radiol*, 94, A64-A70.
- GERMAN CANCER RESEARCH CENTER. 2022. *The Medical Imaging Interaction Toolkit (MITK)* [Online]. Available: [https://www.mitk.org/wiki/The_Medical_Imaging_Interaction_Toolkit_\(MITK\)](https://www.mitk.org/wiki/The_Medical_Imaging_Interaction_Toolkit_(MITK)) [Accessed 28.03.2022].
- GILLIES, R. J., KINAHAN, P. E. & HRICAK, H. 2016. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*, 278, 563-577.

- GOLDER, W. 2004. Magnetic resonance spectroscopy in clinical oncology. *Onkologie*, 27, 304-9.
- HE, K., ZHANG, X., REN, S. & SUN, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 770-778.
- HILARIO, A., RAMOS, A., PEREZ-NUNEZ, A., SALVADOR, E., MILLAN, J. M., LAGARES, A., SEPULVEDA, J. M., GONZALEZ-LEON, P., HERNANDEZ-LAIN, A. & RICOY, J. R. 2012. The added value of apparent diffusion coefficient to cerebral blood volume in the preoperative grading of diffuse gliomas. *AJNR Am J Neuroradiol*, 33, 701-7.
- HUANG, G., LIU, Z., VAN DER MAATEN, L. & WEINBERGER, K. Q. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. 4700-4708.
- IM, H. J., BRADSHAW, T., SOLAIYAPPAN, M. & CHO, S. Y. 2018. Current Methods to Define Metabolic Tumor Volume in Positron Emission Tomography: Which One is Better? *Nucl Med Mol Imaging*, 52, 5-15.
- ISENSEE, F., JAEGER, P. F., KOHL, S. A. A., PETERSEN, J. & MAIER-HEIN, K. H. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*, 18, 203-211.
- ISRAEL, O. & KUTEN, A. 2007. Early Detection of Cancer Recurrence: ¹⁸F-FDG PET/CT Can Make a Difference in Diagnosis and Patient Care. *Journal of Nuclear Medicine*, 48, 28S-35S.
- JAKOB, J. A., BASSETT, R. L., JR., NG, C. S., CURRY, J. L., JOSEPH, R. W., ALVARADO, G. C., ROHLFS, M. L., RICHARD, J., GERSHENWALD, J. E., KIM, K. B., LAZAR, A. J., HWU, P. & DAVIES, M. A. 2012. NRAS mutation status is an independent prognostic factor in metastatic melanoma. *Cancer*, 118, 4014-23.
- JOHANSSON, M., BRODERSEN, J., GOTZSCHE, P. C. & JORGENSEN, K. J. 2019. Screening for reducing morbidity and mortality in malignant melanoma. *Cochrane Database Syst Rev*, 6, CD012352.
- KART, T., FISCHER, M., KÜSTNER, T., HEPP, T., BAMBERG, F., WINZECK, S., GLOCKER, B., RUECKERT, D. & GATIDIS, S. 2021. Deep Learning-Based Automated Abdominal Organ Segmentation in the UK Biobank and German National Cohort Magnetic Resonance Imaging Studies. *Investigative Radiology*, 56, 401-408.
- KENDALL, A. & GAL, Y. 2017. What uncertainties do we need in Bayesian deep learning for computer vision? *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA: Curran Associates Inc.
- KIM, C. K., GUPTA, N. C., CHANDRAMOULI, B. & ALAVI, A. 1994. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *Journal of Nuclear Medicine*, 35, 164-167.
- KIM, H., JUNG, J., KIM, J., CHO, B., KWAK, J., JANG, J. Y., LEE, S. W., LEE, J. G. & YOON, S. M. 2020. Abdominal multi-organ auto-segmentation using 3D-patch-based deep convolutional neural network. *Sci Rep*, 10, 6204.
- KITAJIMA, K., WATABE, T., NAKAJO, M., ISHIBASHI, M., DAISAKI, H., SOEDA, F., TANEMURA, A., KANEKURA, T., YAMAZAKI, N. & ITO, K. 2022. Tumor response evaluation in patients with malignant melanoma

- undergoing immune checkpoint inhibitor therapy and prognosis prediction using (18)F-FDG PET/CT: multicenter study for comparison of EORTC, PERCIST, and imPERCIST. *Jpn J Radiol*, 40, 75-85.
- KOVACS, N., SZIGETI, K., HEGEDUS, N., HORVATH, I., VERES, D. S., BACHMANN, M., BERGMANN, R. & MATHE, D. 2018. Multimodal PET/MRI Imaging Results Enable Monitoring the Side Effects of Radiation Therapy. *Contrast Media Mol Imaging*, 2018, 5906471.
- KOZOVSKA, Z., GABRISOVA, V. & KUCEROVA, L. 2016. Malignant melanoma: diagnosis, treatment and cancer stem cells. *Neoplasma*, 63, 510-7.
- KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. 2012. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. Lake Tahoe, Nevada: Curran Associates Inc.
- KÜSTNER, T., HEPP, T., FISCHER, M., SCHWARTZ, M., FRITSCHKE, A., HÄRING, H.-U., NIKOLAOU, K., BAMBERG, F., YANG, B., SCHICK, F., GATIDIS, S. & MACHANN, J. 2020. Fully Automated and Standardized Segmentation of Adipose Tissue Compartments via Deep Learning in 3D Whole-Body MRI of Epidemiologic Cohort Studies. *Radiology: Artificial Intelligence*, 2, e200010.
- KÜSTNER, T., MÜLLER, S., FISCHER, M., SCHWARTZ, M., MARTIROSIAN, P., YANG, B., SCHICK, F. & GATIDIS, S. 2019. Automated organ segmentation of liver and spleen in whole-body T1-weighted MR images: Transfer learning between epidemiological cohort studies. *Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*. Montreal, Canada.
- KÜSTNER, T., MÜLLER, S., FISCHER, M., WEIB, J., NIKOLAOU, K., BAMBERG, F., YANG, B., SCHICK, F. & GATIDIS, S. Semantic organ segmentation in 3d whole-body mr images. 2018 25th IEEE International Conference on Image Processing (ICIP), 2018. IEEE, 3498-3502.
- LAMBIN, P., LEIJENAAR, R. T. H., DEIST, T. M., PEERLINGS, J., DE JONG, E. E. C., VAN TIMMEREN, J., SANDULEANU, S., LARUE, R. T. H. M., EVEN, A. J. G., JOCHEMS, A., VAN WIJK, Y., WOODRUFF, H., VAN SOEST, J., LUSTBERG, T., ROELOFS, E., VAN ELMPT, W., DEKKER, A., MOTTAGHY, F. M., WILDBERGER, J. E. & WALSH, S. 2017. Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, 14, 749-762.
- LARKIN, J., CHIARION-SILENI, V., GONZALEZ, R., GROB, J. J., COWEY, C. L., LAO, C. D., SCHADENDORF, D., DUMMER, R., SMYLLIE, M., RUTKOWSKI, P., FERRUCCI, P. F., HILL, A., WAGSTAFF, J., CARLINO, M. S., HAANEN, J. B., MAIO, M., MARQUEZ-RODAS, I., MCARTHUR, G. A., ASCIERTO, P. A., LONG, G. V., CALLAHAN, M. K., POSTOW, M. A., GROSSMANN, K., SZNOL, M., DRENO, B., BASTHOLT, L., YANG, A., ROLLIN, L. M., HORAK, C., HODI, F. S. & WOLCHOK, J. D. 2015. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med*, 373, 23-34.
- LESSMANN, N., VAN GINNEKEN, B., DE JONG, P. A. & IŠGUM, I. 2019. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Medical image analysis*, 53, 142-155.

- LEYENDECKER, J. R., BROWN, J. J. & MERKLE, E. M. 2010. *Practical Guide to Abdominal and Pelvic MRI*, Wolters Kluwer/Lippincott Williams & Wilkins Health.
- LI, H., LUO, H., HUAN, W., SHI, Z., YAN, C., WANG, L., MU, Y. & LIU, Y. 2021. Automatic lumbar spinal MRI image segmentation with a multi-scale attention network. *Neural Comput Appl*, 1-14.
- LIEBGOTT, A., GATIDIS, S., VU, V. C., HAUEISE, T., NIKOLAOU, K. & YANG, B. Feature-based Response Prediction to Immunotherapy of late-stage Melanoma Patients Using PET/MR Imaging. 2020 28th European Signal Processing Conference (EUSIPCO), 18-21 Jan. 2021. 1229-1233.
- LIGERO, M., GARCIA-RUIZ, A., VIAPLANA, C., VILLACAMPA, G., RACITI, M. V., LANDA, J., MATOS, I., MARTIN-LIBERAL, J., OCHOA-DE-OLZA, M., HIERRO, C., MATEO, J., GONZALEZ, M., MORALES-BARRERA, R., SUAREZ, C., RODON, J., ELEZ, E., BRAÑA, I., MUÑOZ-COUSELO, E., OAKNIN, A., FASANI, R., NUCIFORO, P., GIL, D., RUBIO-PEREZ, C., SEOANE, J., FELIP, E., ESCOBAR, M., TABERNERO, J., CARLES, J., DIENSTMANN, R., GARRALDA, E. & PEREZ-LOPEZ, R. 2021. A CT-based Radiomics Signature Is Associated with Response to Immune Checkpoint Inhibitors in Advanced Solid Tumors. *Radiology*, 299, 109-119.
- LITJENS, G., KOOI, T., BEJNORDI, B. E., SETIO, A. A. A., CIOMPI, F., GHAFLOORIAN, M., VAN DER LAAK, J., VAN GINNEKEN, B. & SANCHEZ, C. I. 2017. A survey on deep learning in medical image analysis. *Med Image Anal*, 42, 60-88.
- LONG, G. V., FLAHERTY, K. T., STROYAKOVSKIY, D., GOGAS, H., LEVCHENKO, E., DE BRAUD, F., LARKIN, J., GARBE, C., JOUARY, T., HAUSCHILD, A., CHIARION-SILENI, V., LEBBE, C., MANDALA, M., MILLWARD, M., ARANCE, A., BONDARENKO, I., HAANEN, J., HANSSON, J., UTIKAL, J., FERRARESI, V., MOHR, P., PROBACHAI, V., SCHADENDORF, D., NATHAN, P., ROBERT, C., RIBAS, A., DAVIES, M. A., LANE, S. R., LEGOS, J. J., MOOKERJEE, B. & GROB, J. J. 2017a. Dabrafenib plus trametinib versus dabrafenib monotherapy in patients with metastatic BRAF V600E/K-mutant melanoma: long-term survival and safety analysis of a phase 3 study. *Ann Oncol*, 28, 1631-1639.
- LONG, G. V., HAUSCHILD, A., SANTINAMI, M., ATKINSON, V., MANDALA, M., CHIARION-SILENI, V., LARKIN, J., NYAKAS, M., DUTRIAUX, C., HAYDON, A., ROBERT, C., MORTIER, L., SCHACHTER, J., SCHADENDORF, D., LESIMPLE, T., PLUMMER, R., JI, R., ZHANG, P., MOOKERJEE, B., LEGOS, J., KEFFORD, R., DUMMER, R. & KIRKWOOD, J. M. 2017b. Adjuvant Dabrafenib plus Trametinib in Stage III BRAF-Mutated Melanoma. *N Engl J Med*, 377, 1813-1823.
- LOW, R. N. & GURNEY, J. 2007. Diffusion-weighted MRI (DWI) in the oncology patient: value of breathhold DWI compared to unenhanced and gadolinium-enhanced MRI. *J Magn Reson Imaging*, 25, 848-58.
- LUGAUER, F. & WETZL, J. 2018. Magnetic Resonance Imaging. In: MAIER, A., STEIDL, S., CHRISTLEIN, V. & HORNEGGER, J. (eds.) *Medical Imaging Systems: An Introductory Guide*. Cham (CH): Springer
- Copyright 2018, The Author(s).

- MA, J. 2008. Dixon techniques for water and fat imaging. *J Magn Reson Imaging*, 28, 543-58.
- MALEKI, F., OVENS, K., NAJAFIAN, K., FORGHANI, B., REINHOLD, C. & FORGHANI, R. 2020. Overview of Machine Learning Part 1: Fundamentals and Classic Approaches. *Neuroimaging Clin N Am*, 30, e17-e32.
- MARTIN, O., SCHAARSCHMIDT, B. M., KIRCHNER, J., SUNTHARALINGAM, S., GRUENEISEN, J., DEMIRCIOGLU, A., HEUSCH, P., QUICK, H. H., FORSTING, M., ANTOCH, G., HERRMANN, K. & UMUTLU, L. 2020. PET/MRI Versus PET/CT for Whole-Body Staging: Results from a Single-Center Observational Study on 1,003 Sequential Examinations. *J Nucl Med*, 61, 1131-1136.
- MICHELUCCI, U. 2019. Fundamentals of Convolutional Neural Networks. *Advanced Applied Deep Learning : Convolutional Neural Networks and Object Detection*. Berkeley, CA: Apress.
- MILES, K. A., VOO, S. A. & GROVES, A. M. 2018. Additional Clinical Value for PET/MRI in Oncology: Moving Beyond Simple Diagnosis. *J Nucl Med*, 59, 1028-1032.
- MILLETARI, F., NAVAB, N. & AHMADI, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth International Conference on 3D Vision (3DV), 2016. IEEE, 565-571.
- MISIR, A. F., DURMUSLAR, M. C., ZERENER, T. & GUN, B. D. 2016. Primary malignant melanoma. *Saudi Med J*, 37, 446-9.
- MONTAGNON, E., CERNY, M., CADRIN-CHENEVERT, A., HAMILTON, V., DERENNES, T., ILINCA, A., VANDENBROUCKE-MENU, F., TURCOTTE, S., KADOURY, S. & TANG, A. 2020. Deep learning workflow in radiology: a primer. *Insights Imaging*, 11, 22.
- MORONE, M., BALI, M. A., TUNARIU, N., MESSIOU, C., BLACKLEDGE, M., GRAZIOLI, L. & KOH, D. M. 2017. Whole-Body MRI: Current Applications in Oncology. *AJR Am J Roentgenol*, 209, W336-W349.
- MULLER-HORVAT, C., RADNY, P., EIGENTLER, T. K., SCHAFER, J., PFANNENBERG, C., HORGER, M., KHORCHIDI, S., NAGELE, T., GARBE, C., CLAUSSEN, C. D. & SCHLEMMER, H. P. 2006. Prospective comparison of the impact on treatment decisions of whole-body magnetic resonance imaging and computed tomography in patients with metastatic malignant melanoma. *Eur J Cancer*, 42, 342-50.
- NIELSEN, A., HANSEN, M. B., TIETZE, A. & MOURIDSEN, K. 2018. Prediction of Tissue Outcome and Assessment of Treatment Effect in Acute Ischemic Stroke Using Deep Learning. *Stroke*, 49, 1394-1401.
- NVIDIA CORPORATION. 2022. *NVIDIA Clara Imaging* [Online]. Available: <https://developer.nvidia.com/clara-medical-imaging> [Accessed 28.03.2022].
- OBERMEYER, Z. & EMANUEL, E. J. 2016. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *New England Journal of Medicine*, 375, 1216-1219.
- OLLILA, D. W., ESSNER, R., WANEK, L. A. & MORTON, D. L. 1996. Surgical Resection for Melanoma Metastatic to the Gastrointestinal Tract. *Archives of Surgery*, 131, 975-980.

- PAVRI, S. N., CLUNE, J., ARIYAN, S. & NARAYAN, D. 2016. Malignant Melanoma: Beyond the Basics. *Plast Reconstr Surg*, 138, 330e-40e.
- PETRALIA, G., PADHANI, A. R., PRICOLO, P., ZUGNI, F., MARTINETTI, M., SUMMERS, P. E., GRAZIOLI, L., COLAGRANDE, S., GIOVAGNONI, A., BELLOMI, M. & ITALIAN WORKING GROUP ON MAGNETIC, R. 2019. Whole-body magnetic resonance imaging (WB-MRI) in oncology: recommendations and key uses. *Radiol Med*, 124, 218-233.
- PFANNENBERG, C., ASCHOFF, P., SCHANZ, S., ESCHMANN, S. M., PLATHOW, C., EIGENTLER, T. K., GARBE, C., BRECHTEL, K., VONTHEIN, R., BARES, R., CLAUSSEN, C. D. & SCHLEMMER, H. P. 2007. Prospective comparison of 18F-fluorodeoxyglucose positron emission tomography/computed tomography and whole-body magnetic resonance imaging in staging of advanced malignant melanoma. *Eur J Cancer*, 43, 557-64.
- RAPALINO, O. & RATAI, E. M. 2016. Multiparametric Imaging Analysis: Magnetic Resonance Spectroscopy. *Magn Reson Imaging Clin N Am*, 24, 671-686.
- RASCH, D. & GUIARD, V. 2004. The robustness of parametric statistical methods. *Psychology Science*, 46, 175-208.
- RASTRELLI, M., TROPEA, S., ROSSI, C. R. & ALAIBAC, M. 2014. Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In Vivo*, 28, 1005-11.
- REINERT, C. P., GATIDIS, S., SEKLER, J., DITTMANN, H., PFANNENBERG, C., LA FOUGERE, C., NIKOLAOU, K. & FORSCHNER, A. 2020. Clinical and prognostic value of tumor volumetric parameters in melanoma patients undergoing (18)F-FDG-PET/CT: a comparison with serologic markers of tumor burden and inflammation. *Cancer Imaging*, 20, 44.
- RIOLA-PARADA, C., GARCÍA-CAÑAMAQUE, L., PÉREZ-DUEÑAS, V., GARCERANT-TAFUR, M. & CARRERAS-DELGADO, J. L. 2016. Simultaneous PET/MRI vs. PET/CT in oncology. A systematic review. *Revista Española de Medicina Nuclear e Imagen Molecular (English Edition)*, 35, 306-312.
- ROBERT, C., SCHACHTER, J., LONG, G. V., ARANCE, A., GROB, J. J., MORTIER, L., DAUD, A., CARLINO, M. S., MCNEIL, C., LOTEM, M., LARKIN, J., LORIGAN, P., NEYNS, B., BLANK, C. U., HAMID, O., MATEUS, C., SHAPIRA-FROMMER, R., KOSH, M., ZHOU, H., IBRAHIM, N., EBBINGHAUS, S., RIBAS, A. & INVESTIGATORS, K.-. 2015. Pembrolizumab versus Ipilimumab in Advanced Melanoma. *N Engl J Med*, 372, 2521-32.
- ROGASCH, J. M. M., HUNDSDOERFER, P., HOFHEINZ, F., WEDEL, F., SCHATKA, I., AMTHAUER, H. & FURTH, C. 2018. Pretherapeutic FDG-PET total metabolic tumor volume predicts response to induction therapy in pediatric Hodgkin's lymphoma. *BMC Cancer*, 18, 521.
- RONNEBERGER, O., FISCHER, P. & BROX, T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, 2015. Springer, 234-241.
- SABA, L., BISWAS, M., KUPPILI, V., CUADRADO GODIA, E., SURI, H. S., EDLA, D. R., OMERZU, T., LAIRD, J. R., KHANNA, N. N., MAVROGENI,

- S., PROTOGEROU, A., SFIKAKIS, P. P., VISWANATHAN, V., KITAS, G. D., NICOLAIDES, A., GUPTA, A. & SURI, J. S. 2019. The present and future of deep learning in radiology. *Eur J Radiol*, 114, 14-24.
- SACCHETTO, L., ZANETTI, R., COMBER, H., BOUCHARDY, C., BREWSTER, D. H., BROGANELLI, P., CHIRLAQUE, M. D., COZA, D., GALCERAN, J., GAVIN, A., HACKL, M., KATALINIC, A., LARONNINGEN, S., LOUWMAN, M. W. J., MORGAN, E., ROBSAHM, T. E., SANCHEZ, M. J., TRYGGVADOTTIR, L., TUMINO, R., VAN EYCKEN, E., VERNON, S., ZADNIK, V. & ROSSO, S. 2018. Trends in incidence of thick, thin and in situ melanoma in Europe. *Eur J Cancer*, 92, 108-118.
- SANDRU, A., VOINEA, S., PANAITESCU, E. & BLIDARU, A. 2014. Survival rates of patients with metastatic malignant melanoma. *J Med Life*, 7, 572-6.
- SCHRAML, C., SCHMID, M., GATIDIS, S., SCHMIDT, H., LA FOUGERE, C., NIKOLAOU, K. & SCHWENZER, N. F. 2015. Multiparametric analysis of bone marrow in cancer patients using simultaneous PET/MR imaging: Correlation of fat fraction, diffusivity, metabolic activity, and anthropometric data. *J Magn Reson Imaging*, 42, 1048-56.
- SCHWENCK, J., SCHORG, B., FIZ, F., SONANINI, D., FORSCHNER, A., EIGENTLER, T., WEIDE, B., MARTELLA, M., GONZALEZ-MENENDEZ, I., CAMPI, C., SAMBUCETI, G., SEITH, F., QUINTANILLA-MARTINEZ, L., GARBE, C., PFANNENBERG, C., ROCKEN, M., LA FOUGERE, C., PICHLER, B. J. & KNEILLING, M. 2020. Cancer immunotherapy is accompanied by distinct metabolic patterns in primary and secondary lymphoid organs observed by non-invasive in vivo (18)F-FDG-PET. *Theranostics*, 10, 925-937.
- SEITH, F., FORSCHNER, A., SCHMIDT, H., PFANNENBERG, C., GUCKEL, B., NIKOLAOU, K., LA FOUGERE, C., GARBE, C. & SCHWENZER, N. 2018. 18F-FDG-PET detects complete response to PD1-therapy in melanoma patients two weeks after therapy start. *Eur J Nucl Med Mol Imaging*, 45, 95-101.
- SHELLER, M. J., EDWARDS, B., REINA, G. A., MARTIN, J., PATI, S., KOTROTSOU, A., MILCHENKO, M., XU, W., MARCUS, D., COLEN, R. R. & BAKAS, S. 2020. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep*, 10, 12598.
- SIEGEL, R. L., MILLER, K. D. & JEMAL, A. 2020. Cancer statistics, 2020. *CA Cancer J Clin*, 70, 7-30.
- SJÖHOLM, T., EKSTRÖM, S., STRAND, R., AHLSTRÖM, H., LIND, L., MALMBERG, F. & KULLBERG, J. 2019. A whole-body FDG PET/MR atlas for multiparametric voxel-based analysis. *Sci Rep*, 9, 6158.
- SLADDEN, M. J., BALCH, C., BARZILAI, D. A., BERG, D., FREIMAN, A., HANDISIDE, T., HOLLIS, S., LENS, M. B. & THOMPSON, J. F. 2009. Surgical excision margins for primary cutaneous melanoma. *Cochrane Database Syst Rev*, Cd004835.
- SONG, Y., ZHANG, Y.-D., YAN, X., LIU, H., ZHOU, M., HU, B. & YANG, G. 2018. Computer-aided diagnosis of prostate cancer using a deep convolutional neural network from multiparametric MRI. *Journal of Magnetic Resonance Imaging*, 48, 1570-1577.

- STRAND, R., MALMBERG, F., JOHANSSON, L., LIND, L., SUNDBOM, M., AHLSTRÖM, H. & KULLBERG, J. 2017. A concept for holistic whole body MRI data analysis, Imiomics. *PLOS ONE*, 12, e0169966.
- SUN, C., SHRIVASTAVA, A., SINGH, S. & GUPTA, A. Revisiting unreasonable effectiveness of data in deep learning era. Proceedings of the IEEE international conference on computer vision, 2017. 843-852.
- TIRKES, T., HOLLAR, M. A., TANN, M., KOHLI, M. D., AKISIK, F. & SANDRASEGARAN, K. 2013. Response Criteria in Oncologic Imaging: Review of Traditional and New Criteria. *RadioGraphics*, 33, 1323-1341.
- TREBESCHI, S., DRAGO, S. G., BIRKBAK, N. J., KURILOVA, I., CALIN, A. M., DELLI PIZZI, A., LALEZARI, F., LAMBREGTS, D. M. J., ROHAAN, M. W., PARMAR, C., ROZEMAN, E. A., HARTEMINK, K. J., SWANTON, C., HAANEN, J., BLANK, C. U., SMIT, E. F., BEETS-TAN, R. G. H. & AERTS, H. 2019. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. *Ann Oncol*, 30, 998-1004.
- TRINDADE, M. R., BLAYA, R. & TRINDADE, E. N. 2009. Melanoma metastasis to the spleen: Laparoscopic approach. *J Minim Access Surg*, 5, 17-9.
- TURELLA, F., BREDELL, G., OKUPNIK, A., CAPRARA, S., GRAF, D., SUTTER, R. & KONUKOGLU, E. High-Resolution Segmentation of Lumbar Vertebrae from Conventional Thick Slice MRI. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2021. Springer, 689-698.
- UMUTLU, L., BEYER, T., GRUENEISEN, J. S., RISCHPLER, C., QUICK, H. H., VEIT-HAIBACH, P., EIBER, M., PURZ, S., ANTOCH, G., GATIDIS, S., NIKOLAOU, K., SCHAEFER, J. F., RAUSCH, I., HERRMANN, K., VORSTAND DER INTERDISZIPLINAREN, A. G. F. H. B. I. A. R. & VORSTAND DER DEUTSCHEN RONTGENGESELLSCHAFT UND DER DEUTSCHEN GESELLSCHAFT FUR NUKLEARMEDIZIN IN ALPHABETISCHER, R. 2019. Whole-Body [18F]-FDG-PET/MRI for Oncology: A Consensus Recommendation. *Rofo*, 191, 289-297.
- VAN DER PLOEG, I. M., KROON, B. B., ANTONINI, N., VALDES OLMOS, R. A. & NIEWEG, O. E. 2009. Comparison of three micromorphometric pathology classifications of melanoma metastases in the sentinel node. *Ann Surg*, 250, 301-4.
- VAN HARTEN, L. D., NOOTHOUT, J. M., VERHOEFF, J. J., WOLTERINK, J. M. & ISGUM, I. Automatic Segmentation of Organs at Risk in Thoracic CT scans by Combining 2D and 3D Convolutional Neural Networks. SegTHOR@ ISBI, 2019.
- VANIA, M., MUREJA, D. & LEE, D. 2019. Automatic spine segmentation from CT images using Convolutional Neural Network via redundant generation of class labels. *Journal of Computational Design and Engineering*, 6, 224-232.
- WOLCHOK, J. D., CHIARION-SILENI, V., GONZALEZ, R., RUTKOWSKI, P., GROB, J.-J., COWEY, C. L., LAO, C. D., WAGSTAFF, J., SCHADENDORF, D., FERRUCCI, P. F., SMYLIE, M., DUMMER, R., HILL, A., HOGG, D., HAANEN, J., CARLINO, M. S., BECHTER, O., MAIO, M., MARQUEZ-RODAS, I., GUIDOBONI, M., MCARTHUR, G., LEBBÉ, C., ASCIERTO, P. A., LONG, G. V., CEBON, J., SOSMAN, J., POSTOW,

- M. A., CALLAHAN, M. K., WALKER, D., ROLLIN, L., BHOORE, R., HODI, F. S. & LARKIN, J. 2017. Overall Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. *New England Journal of Medicine*, 377, 1345-1356.
- XING, Y., BRONSTEIN, Y., ROSS, M. I., ASKEW, R. L., LEE, J. E., GERSHENWALD, J. E., ROYAL, R. & CORMIER, J. N. 2011. Contemporary diagnostic imaging modalities for the staging and surveillance of melanoma patients: a meta-analysis. *J Natl Cancer Inst*, 103, 129-42.
- YAMASHITA, R., NISHIO, M., DO, R. K. G. & TOGASHI, K. 2018. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, 9, 611-629.
- YAN, Y., ZENG, S., WANG, X., GONG, Z. & XU, Z. 2019. A Machine Learning Algorithm for Predicting Therapeutic Response to Anti-PD1. *Technol Cancer Res Treat*, 18, 1533033819875766.
- YASAKA, K., AKAI, H., KUNIMATSU, A., KIRYU, S. & ABE, O. 2018. Deep learning with convolutional neural network in radiology. *Jpn J Radiol*, 36, 257-272.
- ZHOU, Z., RAHMAN SIDDIQUEE, M. M., TAJBAKHSH, N. & LIANG, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: STOYANOV, D., TAYLOR, Z., CARNEIRO, G., SYEDA-MAHMOOD, T., MARTEL, A., MAIER-HEIN, L., TAVARES, J. M. R. S., BRADLEY, A., PAPA, J. P., BELAGIANNIS, V., NASCIMENTO, J. C., LU, Z., CONJETI, S., MORADI, M., GREENSPAN, H. & MADABHUSHI, A., eds. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018// 2018 Cham*. Springer International Publishing, 3-11.

7 Erklärung zum Eigenanteil

Die Arbeit wurde in der Radiologischen Universitätsklinik Tübingen unter Betreuung von Prof. Dr. Sergios Gatidis durchgeführt.

Die Konzeption der Studie erfolgte in Zusammenarbeit mit Prof. Dr. Sergios Gatidis.

Die Erstellung der Trainingsdaten zur Durchführung einer automatischen Segmentierung, sowie die Qualitätssicherung manueller und automatischer Segmentierungen wurde von mir durchgeführt. Die statistischen Auswertungen samt Erstellung der zugehörigen Tabellen und Grafiken, sowie die visuelle Bildanalyse, der Zusammenschnitt und die Kenntlichmachung der MRT-Bilder und Segmentierungen erfolgten durch mich.

Die technische Durchführung der automatischen Segmentierung erfolgte durch Frau Annika Liebgott, MSc.

Ich versichere, das Manuskript selbstständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen verwendet zu haben.

Tübingen, den 04.07.2022

Viet Chau Vu

8 Veröffentlichungen

Im Kontext dieser Dissertation entstanden folgende Konferenzbeiträge, welche im inhaltlichen Zusammenhang mit der Dissertationsarbeit stehen. Konkrete Ergebnisse aus diesen Beiträgen sind nicht in diese Arbeit eingeflossen:

LIEBGOTT, A., LORENZ, C., GATIDIS, S., VU, V. C., NIKOLAOU, K. & YANG, B. 2021. Automated Multi-Organ Segmentation in Pet Images Using Cascaded Training of a 3d U-Net and Convolutional Autoencoder. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

LIEBGOTT, A., GATIDIS, S., VU, V. C., HAUEISE, T., NIKOLAOU, K. & YANG, B. Feature-based Response Prediction to Immunotherapy of late-stage Melanoma Patients Using PET/MR Imaging. 2020 28th European Signal Processing Conference (EUSIPCO), 18-21 Jan. 2021 2021. 1229-1233.

Danksagung

Ich danke meinem Doktorvater, Herrn Prof. Dr. Sergios Gatidis, für die hervorragende und vertrauensvolle Betreuung, die wegweisenden Anregungen, sowie die Überlassung des Themas.

Meinem Mentor Herrn PD Dr. Christian Reinert danke ich für seine Hilfsbereitschaft, sowie seine wertvollen Anmerkungen und Korrekturen.

Ein besonderer Dank gilt Frau Annika Liebgott für die stets konstruktive Zusammenarbeit.

Weiterhin danke ich meinen Freunden für das Korrekturlesen dieser Arbeit, die hilfreichen Ideen, sowie den emotionalen Rückhalt während des gesamten Studiums.

Abschließend bedanke ich mich bei meiner Familie, ohne deren Unterstützung ich nicht so weit gekommen wäre.