

Aus der
Radiologischen Universitätsklinik Tübingen
Abteilung Diagnostische und Interventionelle Radiologie

**Automatische Gewebesegmentierung
der Nieren und des Pankreas im Ganzkörper-MRT
mittels Deep Learning**

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

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

vorgelegt von

Westenberger, Jasmin Barbara

2021

Dekan: Professor Dr. B. Pichler

1. Berichterstatter: Professor Dr. S. Gatidis

2. Berichterstatter: Professorin Dr. S. Joos

Tag der Disputation: 04.08.2021

Inhaltsverzeichnis

Inhaltsverzeichnis.....	3
Abbildungsverzeichnis	5
Tabellenverzeichnis	7
Abkürzungsverzeichnis	8
1 Einleitung	9
1.1 NAKO Gesundheitsstudie.....	9
1.1.1 MRT-Bildgebung der NAKO	11
1.2 Künstliche neuronale Netzwerke zur automatischen Bildanalyse	13
1.3 Ziel und Fragestellung	16
2 Material und Methoden	17
2.1 MRT-Datensätze.....	17
2.2 Manuelle Segmentierung.....	18
2.3 Automatische Organsegmentierung durch ein Convolutional Neural Network	20
2.4 Datenauswertung.....	24
2.4.1 Vergleich Trainings- und Testdaten.....	25
2.4.2 Vergleich automatische und manuelle Segmentierung.....	25
2.4.2.1 Visuell qualitativer Vergleich	26
2.4.2.2 Quantitativer Vergleich.....	29
2.4.2.3 Vergleich sekundärer Parameter.....	30
3 Ergebnisse.....	31
3.1 Daten über die Population der MRT-Datensätze	31
3.2 Manuelle Segmentierung.....	31
3.3 Algorithmus zur automatischen Organsegmentierung	32

3.4	Vergleich Trainings- und Testdaten	34
3.5	Vergleich automatische und manuelle Segmentierung	37
3.5.1	Visuell qualitative Auswertung	37
3.5.2	Quantitative Auswertung.....	43
3.5.3	Sekundäre Parameter zur Auswertung.....	45
4	Diskussion.....	53
4.1	Diskussion und Ausblick	53
4.2	Schlussfolgerung	59
5	Zusammenfassung	60
6	Literaturverzeichnis.....	62
7	Anhang.....	69
8	Erklärungen zum Eigenanteil	72
9	Danksagung.....	73

Abbildungsverzeichnis

Abbildung 1: Aufteilung der Probanden der NAKO-Studie in die drei verschiedenen Untersuchungsgruppen (Level 1 bis Level 3).....	11
Abbildung 2: Beispielhafte Darstellung der manuellen Segmentierung der drei Organe	19
Abbildung 3: Schematische Darstellung der Trainingsphase eines CNNs	21
Abbildung 4: Darstellung des verwendeten Convolutional Neural Networks....	22
Abbildung 5: Aufteilung der 100 Bilddatensätze (MRT-Aufnahmen) in 80 Trainingsdaten und 20 Testdaten.....	23
Abbildung 6: Schematische Darstellung der Testphase eines CNNs.....	24
Abbildung 7: Beispiel eines Datensatzes zur Darstellung der automatischen Organsegmentierung.....	27
Abbildung 8: Score „Organsegmentierung“ der visuell qualitativen Auswertung	28
Abbildung 9: Score „Organsegmentierung“ der visuell qualitativen Auswertung	28
Abbildung 10: Graphische Darstellung der zeitlichen Dauer der manuellen Segmentierung der 100 MRT-Datensätze.....	32
Abbildung 11: Beispiel einer „Loss Curve“ zur Beschreibung des Lernvorganges eines automatischen Lernalgorithmus.....	34
Abbildung 12: Linke Niere: Vergleich zwischen den Trainings- und Testdatensätzen in Bezug auf die Organvolumina	36
Abbildung 13: Rechte Niere: Vergleich zwischen den Trainings- und Testdatensätzen in Bezug auf die Organvolumina	36
Abbildung 14: Pankreas: Vergleich zwischen den Trainings- und Testdatensätzen in Bezug auf die Organvolumina	37
Abbildung 15: Beispielhafte Darstellung der automatischen Organsegmentierung der drei Organe.....	38
Abbildung 16: Beispielhafte Darstellung der Organsegmentierung der linken und der rechten Niere.....	38
Abbildung 17: Beispiele der automatischen Organsegmentierung der Nieren.	39

Abbildung 18: Beispielhafte Darstellungen der automatischen Segmentierung des Pankreas	40
Abbildung 19: Beispielhafte Darstellung der automatischen Segmentierung der drei Organe	40
Abbildung 20: Score „Organsegmentierung“. Balkendiagramm zur Veranschaulichung der absoluten Häufigkeiten der visuell qualitativen Auswertung der Testdatensätze.....	41
Abbildung 21: Score „sonstige Fehlsegmentierungen“. Balkendiagramm zur Veranschaulichung der absoluten Häufigkeiten der visuell qualitativen Auswertung der Testdatensätze.....	42
Abbildung 22: Dice-Index der automatischen Segmentierung der 20 Testdatensätze.....	45
Abbildung 23: Linke Niere: Vergleich der Organvolumina zwischen der manuellen und der automatischen Segmentierung	47
Abbildung 24: Rechte Niere: Vergleich der Organvolumina zwischen der manuellen und der automatischen Segmentierung	47
Abbildung 25: Pankreas: Vergleich der Organvolumina zwischen der manuellen und der automatischen Segmentierung	48
Abbildung 26: Linke Niere: Vergleich des Fettgehalts zwischen der manuellen und der automatischen Segmentierung	48
Abbildung 27: Rechte Niere: Vergleich des Fettgehalts zwischen der manuellen und der automatischen Segmentierung	49
Abbildung 28: Pankreas: Vergleich des Fettgehalts zwischen der manuellen und der automatischen Segmentierung.....	49
Abbildung 29: Linke Niere: Korrelation der Organvolumina zwischen den manuell und automatisch erstellten Masken.....	51
Abbildung 30: Rechte Niere: Korrelation der Organvolumina zwischen den manuell und automatisch erstellten Masken.....	51
Abbildung 31: Pankreas: Korrelation der Organvolumina zwischen den manuell und automatisch erstellten Masken.....	52

Tabellenverzeichnis

Tabelle 1: Thorakoabdominelle Sequenzparameter der MRT- Ganzkörperaufnahmen der T1-Wichtung der NAKO-Studie.....	18
Tabelle 2: Score der visuell qualitativen Auswertung bezogen auf das jeweilige Organ. Score: „Organsegmentierung“	27
Tabelle 3: Ergänzender Score zur Erfassung des Auftretens von Fehlsegmentierungen außerhalb des Organs. Score: „sonstige Fehlsegmentierungen“	28
Tabelle 4: Verwendete Messgrößen der quantitativen Auswertung	29
Tabelle 5: Demographische Daten des Probandenkollektivs der verwendeten MRT-Aufnahmen der NAKO-Studie	31
Tabelle 6: Zeitliche Dauer der manuellen Segmentierung der drei Organe der 100 MRT-Datensätze	32
Tabelle 7: Demographische und MRT-bezogene Daten des Probandenkollektivs	35
Tabelle 8: Hauptsächlich aufgetretene Fehlsegmentierungen außerhalb des segmentierten Organs.....	43
Tabelle 9: Messgrößen der quantitativen Auswertung	44
Tabelle 10: Vergleich der Organvolumina und des Fettgehalts.....	46
Tabelle 11: Zuordnung der automatisch segmentierten Organe zu den Scores „Organsegmentierung“ und „sonstige Fehlsegmentierung“	69

Abkürzungsverzeichnis

BMI	Body-Mass-Index
CNN	Convolutional Neural Network
CT	Computertomographie
DICOM	Digital Imaging and Communications in Medicine
FG	Fettgehalt
GPU	Graphics Processing Unit (Grafikprozessor)
GTX 1080 TI	Produktbezeichnung für eine Nvidia Grafikkarte
IBM	International Business Machines
m	männlich
Max.	Maximum
Min.	Minimum
MITK	Medical Imaging Interaction Toolkit
MRT	Magnetresonanztomographie
NAKO	NAKO Gesundheitsstudie
NIFTI	Neuroimaging Informatics Technology Initiative
NRRD	Nearly Raw Raster Data
RNN	Rekurrente neuronale Netzwerke
SPSS	Statistical Package for the Social Sciences
TE	Time To Echo (Echozeit)
TR	Time To Repeat = Repetition Time (Repetitionszeit)
T1w	T1-Wichtung
VIBE	Volume Interpolated Breath-hold Examination
w	weiblich

1 Einleitung

1.1 NAKO Gesundheitsstudie

In der medizinischen Wissenschaft gibt es unterschiedliche Studientypen. Zunächst wird dabei zwischen der Primärforschung - also der Gewinnung primärer Daten infolge einer Studie - und der Sekundärforschung, bei der bereits vorhandene Daten neu analysiert oder der momentane Wissensstand zu einem bestimmten Thema wiedergegeben werden kann, unterschieden. Die Grundlagenforschung, die klinische und die epidemiologische Forschung sind weitere Untergliederungsmöglichkeiten der Primärforschung. [(Röhring et al., 2009a); (Röhring et al., 2009b)]

Epidemiologische Studien lassen sich in experimentelle Arbeiten und Beobachtungsstudien unterteilen. Bei Beobachtungsstudien werden Probanden¹ ausschließlich in ihrem Lebensumfeld beobachtet und nach ausgewählten Aspekten untersucht. Sie erhalten keine Behandlung. Eine Untergliederung anhand des zeitlichen Ablaufs von Beobachtungsstudien ist unter anderem in Fall-Kontroll-Studien und Kohortenstudien möglich. Kohortenstudien erfolgen meist prospektiv. Das heißt, dass eine bestimmte Anzahl an Probanden, die dann als Kohorte bezeichnet wird, ausgewählt und über einen bestimmten Zeitraum beobachtet, befragt und untersucht wird. Ziel solcher Studien kann es sein, bestimmte Zusammenhänge zwischen möglichen Auslösern und dem Entstehen von Krankheiten und Pathologien zu erforschen. Insbesondere durch die Auswertung von Kohortenstudien gelingt eine genaue Schätzung des Risikos infolge einer gewissen Exposition. [(Belbasis and Bellou, 2018); (Held, 2010); (Röhring et al., 2009b)] Ein bekanntes Beispiel für eine epidemiologische Kohortenstudie ist die Framingham-Studie. Durch diese Studie konnten erstmals entscheidende Risikofaktoren für Herz-Kreislauf-Erkrankungen festgestellt und entsprechend zugeordnet werden (Tsao and Vasan, 2015).

¹ Wenn im Folgenden zur Bezeichnung von Personengruppen nur die maskuline Form verwendet wird, geschieht dies ausschließlich aus Gründen der Lesbarkeit. Die Bezeichnung schließt - sofern nicht anders spezifiziert - alle Geschlechter mit ein.

Momentan wird in Deutschland bundesweit eine epidemiologische Kohortenstudie durchgeführt. Es handelt sich hierbei um die NAKO Gesundheitsstudie (NAKO). Dies ist eine von der Helmholtz-Gemeinschaft, der Leibniz-Gesellschaft, Universitäten und anderen Forschungseinrichtungen organisierte Studie und Deutschlands größte derzeit laufende Gesundheitsstudie. Insgesamt nehmen an dieser Studie 200.000 Probanden im Alter zwischen 20 und etwa 70 Jahren teil, die zufällig ausgewählt werden. Hiervon sind jeweils 100.000 Probandinnen Frauen und 100.000 Probanden Männer. Die Auswahl geschieht repräsentativ und soll somit als Ausschnitt die Gesamtbevölkerung Deutschlands darstellen. Es handelt sich um eine Beobachtungsstudie, die prospektiv mit dem Beginn im Jahr 2014 und einer geplanten Dauer von 25-30 Jahren erfolgt. In der NAKO-Studie werden Personen multimodal untersucht, mit dem Ziel, verschiedene Daten über die allgemeine Gesundheit zu erhalten und Faktoren von Krankheiten, insbesondere den Volkskrankheiten, zu erforschen. Dadurch soll deren Prävention, Erkennung und Behandlung verbessert werden. [(Schlett et al., 2016); (German National Cohort, 2014); (Wichmann et al., 2012)]

Die Probanden der NAKO-Studie werden in Deutschland an 18 verschiedenen Studienzentren untersucht. Dabei erfolgt eine Unterteilung in drei verschiedene Gruppen, die als Level 1, Level 2 und Level 3 bezeichnet werden (Abb.1). Diese Gruppen sind durch den jeweiligen Umfang der Untersuchungen gekennzeichnet. Die sogenannte Basisuntersuchung (Level 1) bezeichnet hierbei die grundlegende Datenerhebung. Diese Untersuchung durchläuft jeder Teilnehmer. Als Ergänzung sind für 40.000 Probanden die Untersuchungen des Level 2 vorgesehen, bei der diese noch umfassender befragt und untersucht werden. Von den Probanden des Level 2 erhalten 30.000 Personen zusätzlich die Untersuchung des Level 3. Level 3 umfasst Magnetresonanztomographie-Aufnahmen (MRT-Aufnahmen). Fünf der 18 Studienzentren sind MRT-Standorte. Dazu zählen Augsburg, Berlin, Essen, Mannheim und Neubrandenburg. An jedem dieser fünf Standorte werden etwa 6000 Teilnehmer untersucht. Dabei soll die Durchführung der MRT-Aufnahme einen Abstand von maximal zwölf Wochen

zur Basisuntersuchung nicht überschreiten. [(Schlett et al., 2016); (Bamberg et al., 2015); (German National Cohort, 2014)]

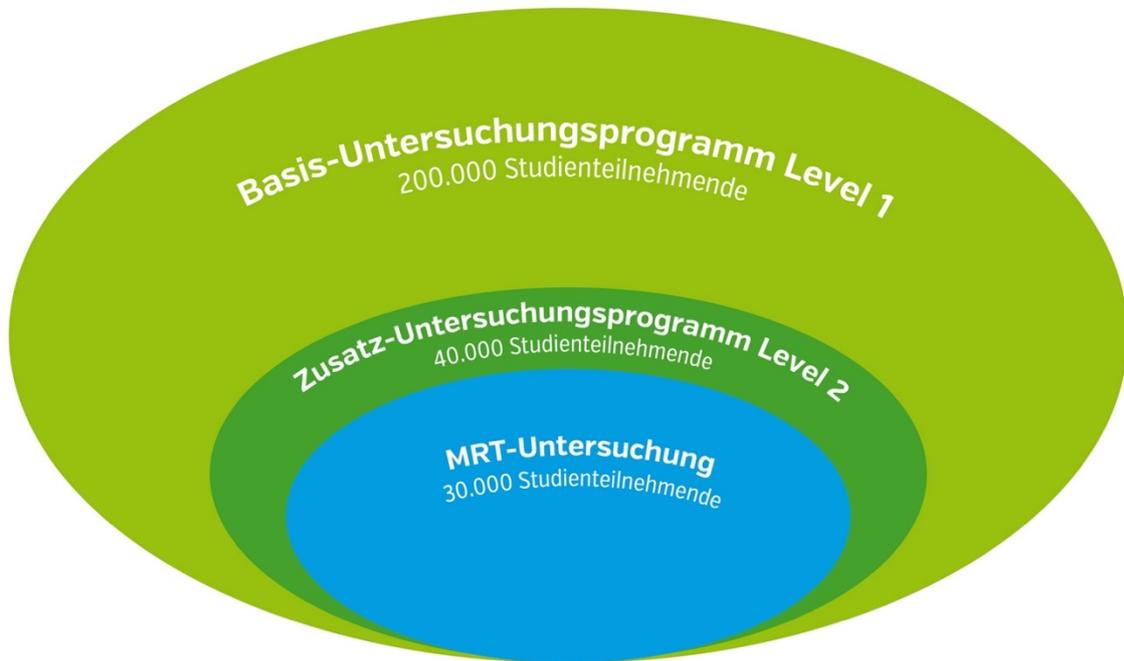


Abbildung 1: Aufteilung der Probanden der NAKO-Studie in die drei verschiedenen Untersuchungsgruppen (Level 1 bis Level 3); Copyright NAKO Gesundheitsstudie.

1.1.1 MRT-Bildgebung der NAKO

Es gibt verschiedene bildgebende Verfahren wie beispielsweise das MRT, die Computertomographie (CT) oder auch die Sonographie, die im klinischen Alltag vordergründig zur Diagnostik genutzt werden. Das MRT zählt wie das CT zu den Schnittbildverfahren. Mithilfe des MRT lassen sich vor allem Weichteile und Nervengewebe sehr gut darstellen. Ein Vorteil von MRT-Untersuchungen ist im Gegensatz zu anderen bildgebenden Verfahren, dass diese bei detailreicher Bildinformation keine wesentlichen Nebenwirkungen aufweisen und sich daher auch für die Anwendung in groß angelegten Studien wie der NAKO anbieten. (Schlett et al., 2016) CT-Untersuchungen hingegen haben aufgrund der ionisierenden Strahlung ein größeres Nebenwirkungsprofil und können so beispielsweise neue Krankheiten induzieren, die dann zu einer Verzerrung des Studienendpunktes führen könnten (Bamberg et al., 2015). Zusätzlich sind MRT-

Aufnahmen verhältnismäßig gut - wenn auch schlechter als CT-Aufnahmen - reproduzierbar, insbesondere im Gegensatz zu beispielsweise der ebenfalls nebenwirkungsarmen Sonographie, die stark vom jeweiligen Untersucher abhängt (Schlett et al., 2016). Implantate aus ferromagnetischen Materialien von Probanden beziehungsweise Patienten können jedoch zu Wechselwirkungen mit dem starken Magnetfeld des MRT-Gerätes führen, weshalb das MRT in solchen Fällen nicht als Untersuchungsmethode geeignet und demnach kontraindiziert ist (Nitz et al., 2011).

In der NAKO-Studie werden die Aufnahmen an 3.0-Tesla Ganzkörper-MRT-Geräten durchgeführt (Bamberg et al., 2015). Die MRT-Aufnahmen, die im Rahmen des Level 3 der NAKO-Studie erfolgen, bilden die Grundlage für die vorliegende Untersuchung. Insgesamt werden zwölf MRT-Sequenzen aufgenommen. Grob lassen sich diese in die neurologische, die kardiovaskuläre, die thorakoabdominelle und die muskuloskelettale Bildgebung unterscheiden. (Schlett et al., 2016) Die Untersuchungsdauer beträgt insgesamt etwa eine Stunde je Proband. Um einen Umgang mit Nebenbefunden innerhalb der Studie zu etablieren, wurde ein Gremium gebildet, welches zufällige Befunde in ihrer klinischen Relevanz mit Auswirkung auf das Wohlbefinden der Teilnehmer und in der Spezifität für die jeweilige Krankheit bewertet. Auf dieser Grundlage wird zwischen mitteilungspflichtigen und nicht-mitteilungspflichtigen Befunden unterschieden. Dies stellt eine Herausforderung dar, da die Probanden in der Regel asymptomatisch sind und keine Anamnese oder sonstige Untersuchungen vorliegen. Hinzu kommt, dass das MRT-Studienprotokoll der NAKO nicht auf die klinisch relevante Diagnostik, sondern vor allem auf die Erhebung von epidemiologischen Parametern ausgelegt ist. (Bamberg et al., 2015)

MRT-Aufnahmen einer großen Anzahl an Probanden, wie dies im Rahmen der NAKO-Studie erfolgt, führen zu erweiterten Möglichkeiten hinsichtlich der Sammlung und Analyse von Patientendaten. Hierbei geht es nicht primär um individuelle Diagnostik, sondern um die Zusammenstellung von phänotypischen Profilen ergänzend zur Erfassung von Blutwerten oder Sequenzen des menschlichen Genoms. Es soll somit anhand der MRT-Aufnahmen ermöglicht werden, neue radiologische Biomarker zu identifizieren. Dies kann zu einem

besseren und genaueren Verständnis von Gesundheit und Krankheit in der Bevölkerung beitragen und erklären, wie verschiedene Risikofaktoren, Verhaltensweisen, molekulare Profile oder auch Expositionen sich gegenseitig beeinflussen. Die in den letzten Jahren zunehmend an Bedeutung gewonnene personalisierte Medizin lässt sich damit fortführen und ergänzen, sowohl in präventiver Hinsicht als auch bei behandelnden Maßnahmen. [(Bamberg et al., 2015); (Meisinger et al., 2016); (Schlett et al., 2016); (Ahrens and Jockel, 2015)] Aufgrund des Designs der NAKO-Studie als Längsschnittstudie werden sich auch später auftretende Pathologien der Probanden auf gegebenenfalls früher erhobene Parameter zurückführen lassen. Gleichzeitig können aufgrund der breitgefächerten Untersuchungsmethoden (MRT, Anamnese, Bioproben, etc.) auch Zusammenhänge von zeitgleich auftretenden Auffälligkeiten identifiziert und erkannt werden. Viele Krankheiten, insbesondere Volkskrankheiten, haben eine multifaktorielle Genese. Ziel der NAKO-Studie ist es somit unter anderem, neue Verknüpfungspunkte der einzelnen Ursachen aufzudecken und daraus die entsprechenden Schlüsse in Bezug auf Prävention, Screening und Behandlungsmethoden zu ziehen. [(Schlett et al., 2016); (Meisinger et al., 2016); (Ahrens and Jockel, 2015)]

Die in diesem Zusammenhang entstehende große Menge an Bilddaten infolge der MRT-Aufnahmen erfordert spezialisierte Methoden zur Auswertung. Die manuelle Bildauswertung ist sehr zeitaufwendig und vor allem bei großen Datenmengen schwer realisierbar. Deshalb kann es insbesondere bei Bildaufnahmen hilfreich sein, andere Mittel zur Datenanalyse als unterstützende Maßnahmen zu etablieren. Eine Möglichkeit sind die Methoden des maschinellen Lernens.

1.2 Künstliche neuronale Netzwerke zur automatischen Bildanalyse

Maschinelles Lernen beschreibt einen Vorgang zur Entwicklung eines autonomen Algorithmus. Der Lernvorgang basiert auf ausgewähltem Datenmaterial, sogenannten Beispieldaten, die auch als Trainingsdaten bezeichnet werden. Für diese Art des Lernens ist keine genaue Programmierung

eines Algorithmus durch den Anwender mit einzelnen Programmierschritten notwendig. Stattdessen entwickelt ein Algorithmus selbstständig ein Verfahren, um ein Problem oder eine bestimmte Aufgabe zu lösen. Bei Lernalgorithmen geschieht dies anhand von Trainingsdaten, indem die jeweiligen Beispiele verallgemeinert werden und dadurch ein Berechnungsvorgang entsteht, der wiederum auf ähnliche Daten angewendet werden kann. [(Döbel et al., 2018); (The, 2018)]

„Deep Learning“, auch tiefes Lernen genannt, ist eine Unterform des maschinellen Lernens. Hierbei erfolgt das Lernen in tiefen künstlichen neuronalen Netzwerken. Die einzelnen Lernschritte und die Ergebnisfindung geschehen in inneren Schichten der künstlichen neuronalen Netzwerke. Es handelt sich um sogenannte Black-Box-Modelle. Das heißt, dass für den Anwender der Lösungsweg verschlossen bleibt und von außen nur schwer nachvollziehbar ist. Künstliche neuronale Netzwerke orientieren sich am Aufbau des menschlichen Gehirns. Ein Vorteil der „Deep Learning“-Methoden ist, dass diese die erforderlichen Merkmale für die Erstellung eines Algorithmus direkt auf Grundlage der eingelesenen Trainingsdaten lernen können. Bevor der maschinelle Lernvorgang beginnen kann, müssen die ausgewählten Trainingsdaten zunächst manuell erstellt werden. Das heißt, dass die Beispieldaten entsprechend ihrer Merkmale mit einer passenden Bezeichnung gekennzeichnet werden und damit eine Lösung vorgegeben wird. Dieser Schritt nennt sich „Labeling“ und ein Merkmal, welches mit der richtigen Lösung beschriftet ist, wird als „Label“ bezeichnet. Bei Bilddaten spricht man auch von Segmentierung. Segmentierung bedeutet demnach, dass ein Bild in unterschiedliche Teile und Objekte zerlegt wird, die inhaltlich zusammengehören. Der maschinelle Lernalgorithmus entwickelt dann einen Vorgang, bei dem aus den markierten Trainingsdaten durch Verallgemeinerung der jeweils eingelesenen Daten eine Klassenbeschriftung entsteht. Diese Klassenbeschriftung kann dann nach Abschluss des Trainingsvorganges auf weitere beliebige Daten angewendet und diese somit automatisiert segmentiert werden. [(Döbel et al., 2018); (Guo et al., 2018); (Küstner et al., 2018c); (Chartrand et al., 2017); (Lee et al., 2017)] Der Begriff Organsegmentierung

bezeichnet demnach einen Vorgang, bei dem dem ausgewählten Organ eine Klassenbeschriftung zugeordnet wird, zum Beispiel der linken Niere, die Klassenbeschriftung „linke Niere“. Unterscheiden lässt sich hierbei zwischen der manuellen Organsegmentierung durch eine individuelle Person und der automatischen Organsegmentierung, die durch einen maschinellen Algorithmus durchgeführt wird.

Ein wichtiger Schritt in der Auswertung von MRT-Aufnahmen ist die inhaltliche Bildanalyse. Die manuelle oder visuelle Bildanalyse wird individuell durch einen menschlichen Betrachter, zum Beispiel einen Radiologen, durchgeführt, erfordert viel Erfahrung - insbesondere bei komplexen Fragestellungen - und kann bei hohem zeitlichen Aufwand auch fehleranfällig sein. Bei großen Datensätzen wie etwa in der NAKO-Studie ist der Aufwand extrem hoch und die visuelle Auswertung nahezu unmöglich. Eine andere Möglichkeit ist, die Bildanalyse automatisiert von Programmen durchführen zu lassen, wobei sich Methoden des maschinellen Lernens anbieten.

Zur automatischen Bildanalyse gibt es verschiedene Methoden. Im Bereich der Medizin wurden und werden diese auf ihre jeweilige Anwendung hin in den unterschiedlichen Bereichen, wie zum Beispiel zur Auswertung von CT- und MRT-Aufnahmen, erprobt. Als dafür geeignet haben sich die Methoden des „Deep Learning“ erwiesen (Chartrand et al., 2017). Bei der automatischen Organsegmentierung konnten vor allem die Convolutional Neural Networks (CNN), welche den künstlichen neuronalen Netzwerken zugerechnet werden, gute Ergebnisse erzielen [(Lavdas et al., 2017); (Hu et al., 2017); (Dou et al., 2017); (Shelhamer et al., 2017), (Milletari et al., 2016)].

Neben CNNs werden in der Literatur unter anderem die Multi-Atlas-Segmentierung, Random Forests und rekurrente neuronale Netzwerke (RNN) zur automatischen Gewebesegmentierung erwähnt [(Lavdas et al., 2017); (Litjens et al., 2017); (Iglesias and Sabuncu, 2015)]. Letztere werden auch den künstlichen neuronalen Netzwerken zugeordnet. Zwar funktionieren diese Verfahren bereits im begrenzten Rahmen und für ausgewählte Fragestellungen, dennoch stellt die automatische Segmentierung weiterhin eine Herausforderung dar und kann fehleranfällig sein. Insofern gilt es diese weiter zu optimieren.

1.3 Ziel und Fragestellung

Die verschiedenen Methoden des maschinellen Lernens lassen sich auch für die Datenauswertung in Kohortenstudien wie der NAKO-Studie nutzen. Bei solchen Studien gibt es eine Vielzahl von erhobenen Daten, die auf bestimmte Parameter analysiert werden können, um beispielsweise Einflussfaktoren über die Gesundheit und die Entstehung von Krankheiten zu erhalten. Ein Beispiel ist die Auswertung von MRT-Aufnahmen. Im Gegensatz zur Analyse von Blutwerten stellt die Analyse von Bilddaten eine größere Herausforderung dar. Bei einer Reihe von Blutwerten, die beispielsweise auf gewisse Parameter hin verglichen werden sollen, ist dies mithilfe von Tabellen oder Anwendungen einer Statistiksoftware möglich. Zur Auswertung und Analyse von Bilddatensätzen infolge von MRT-Untersuchungen sind andere Vorgehensweisen notwendig, da die Bildinformation erst extrahiert werden muss. Neben der manuellen Auswertung haben sich hierbei die Methoden des „Deep Learning“ in künstlichen neuronalen Netzwerken bewährt (Chartrand et al., 2017). Die manuelle Auswertung von Aufnahmen aus bildgebenden Verfahren ist extrem aufwendig. Bei individuellen Fragestellungen wie beispielsweise in der Diagnostik ist dies sinnvoll und üblich. Handelt es sich aber wie bei der NAKO-Studie mit einer Anzahl von 30.000 MRT-Aufnahmen um eine Vielzahl von Bilddaten, ist dies manuell nicht umsetzbar, weshalb automatische Auswertungsschritte notwendig sind. Ein erster Schritt kann die automatische Organsegmentierung sein, die darauf abzielt, mithilfe eines Algorithmus Organe selbstständig zu erkennen. Im Anschluss daran ist auch eine tiefergehende Analyse der Aufnahmen beispielsweise auf mögliche Pathologien der jeweiligen Organe hin möglich. [(Küstner et al., 2018c); (Schlett et al., 2016); (Bamberg et al., 2015)]

Vor diesem Hintergrund hat die vorliegende Untersuchung das Ziel, die folgende Fragestellung zu beantworten: Inwieweit kann ein maschineller Lernalgorithmus, welcher zunächst implementiert und dann validiert wird, in MRT-Ganzkörperaufnahmen die Nieren und das Pankreas automatisiert erkennen, lokalisieren und markieren? Diese Fragestellung wird für 3.0-Tesla MRT-Ganzkörperaufnahmen der T1-Wichtung, die im Rahmen der NAKO-Studie entstanden sind, untersucht.

2 Material und Methoden

2.1 MRT-Datensätze

Das Datenmaterial umfasste 100 im Rahmen der NAKO-Studie erzeugte Ganzkörper-MRT-Aufnahmen mit den dazugehörigen demographischen Daten der jeweiligen Probanden.

Die Datenauswertung erfolgte retrospektiv. Das positive Votum zur Durchführung dieser Studie durch die Ethik-Kommission an der Medizinischen Fakultät der Eberhard-Karls-Universität und am Universitätsklinikum Tübingen unter der Prüfnummer 120/2018BO2 liegt vor. Es wurden pseudonymisierte Daten verwendet, welche von Seiten der NAKO übergeben wurden. Alle Probanden wurden zum Zeitpunkt der jeweiligen Untersuchung im Rahmen der NAKO-Studie in den verschiedenen Studienzentren aufgeklärt und deren Einverständnis zur Teilnahme und Datenverarbeitung eingeholt.

Die Studienpopulation bestand aus den Datensätzen der jeweiligen Probanden, die im Rahmen der NAKO-Studie repräsentativ ausgewählt und für die MRT-Untersuchungen (Level 3) bestimmt wurden. Der Erfassungszeitraum der für die vorliegende Untersuchung verwendeten Daten erstreckte sich von 2015 bis 2016. Aus der Gesamtzahl dieser aufgenommenen Datensätze wurde eine Stichprobe zufällig ausgewählt. Diese schließt 100 Probanden ein. Es wurden sowohl MRT-Aufnahmen von weiblichen als auch von männlichen Probanden verwendet. Die MRT-Untersuchungen erfolgten an einem 3.0-Tesla Ganzkörper-MRT (Magnetom Skyra der Siemens Healthineers). Die Untersuchungsdauer insgesamt betrug im Durchschnitt eine Stunde (Bamberg et al., 2015), für die hier verwendete Aufnahme etwa 5 Minuten.

Die in der vorliegenden Arbeit ausgewerteten Bild-Datensätze liegen als MRT-Ganzkörperaufnahmen in der T1-Wichtung (T1w) vor: T1w 3D VIBE 2-Punkt Dixon mit $1,4 \times 1,4 \times 3,0 \text{ mm}^3$ Voxelgröße in axialer Akquisition und CAIPIRINHA Beschleunigung (Schlett et al., 2016). Weitere Parameter sind in Tabelle 1 dargestellt.

Tabelle 1: Thorakoabdominelle Sequenzparameter der MRT-Ganzkörperaufnahmen der T1-Wichtung der NAKO-Studie [(Küstner et al., 2019); (Schlett et al., 2016)].

Sequenzparameter	Einheit	Wert
Anzahl der Schichten		286
Bandbreite	Hz/px	975
Echozeit 1 und 2 (TE)	ms	1.23 und 2.46
Field of view	%	81,25
Flipwinkel	Grad	9
Matrixgröße	Pixel	320x208
Repetitionszeit (TR)	ms	4.36
Schichtdicke	mm	3

Den Probanden wurde kein Kontrastmittel verabreicht und die Aufnahmen erfolgten unter Atemstopp [(Küstner et al., 2019); (Schlett et al., 2016)].

Zusätzlich zu den MRT-Aufnahmen wurden die demographischen Parameter Geschlecht, Alter und BMI der Probanden erfasst und von Seiten der NAKO übermittelt. Diese Angaben beziehen sich auf die Werte zum Zeitpunkt der jeweiligen Untersuchung der Probanden im Rahmen der NAKO-Studie in den Jahren 2015 und 2016. Bestimmt wurde der Mittelwert mit Standardabweichung sowie das Minimum und das Maximum. Beim Geschlecht wurde die prozentuale Verteilung in Bezug auf weibliche und männliche Probanden untersucht.

2.2 Manuelle Segmentierung

Zunächst wurde die Bildqualität der 100 MRT-Ganzkörperaufnahmen in der T1-Wichtung beurteilt. Anschließend wurde die manuelle Segmentierung dieser Bilddatensätze durchgeführt. Es wurden die Nieren und das Pankreas segmentiert.

Zur manuellen Segmentierung (Labeling) wurden die Datensätze mit der Software Medical Imaging Interaction Toolkit (MITK) der Version 2015.05.2 bearbeitet. Dieses Programm bietet eine Vielzahl von Möglichkeiten der medizinischen Bildbearbeitung. Hierzu zählen u.a. die interaktiven

Segmentierungswerkzeuge, Mess- und Bildstatistik sowie die Volumenvisualisierung [(Nolden et al., 2013); (Wolf et al., 2005)].

In dieser Studie wurde die grundlegende Bildanalyse zur manuellen Segmentierung verwendet. Die jeweiligen MRT-Ganzkörperaufnahmen der einzelnen Datensätze wurden nacheinander in 2D-Ansicht in dem beschriebenen Programm geöffnet und schichtweise bearbeitet. Die Bearbeitung der Datensätze erfolgte in der axialen Ansicht. Die Grenzen der Organe wurden in jeder Schicht der jeweiligen Aufnahme manuell umrandet (2D Tools) und unter den entsprechenden Masken (linke Niere, rechte Niere, Pankreas) gespeichert (Abb.2). Die Speicherung erfolgte als NRRD-Dateiformat. Jedem Pixel der 2D-MRT-Aufnahmen wurde somit beim manuellen Labeling eine der drei Klassen (linke Niere, rechte Niere, Pankreas) zugeordnet und als Grundwahrheit angenommen (Küstner et al., 2018c).

Bei der manuellen Segmentierung der Nieren wurde das Hilum renale miteingeschlossen. Die Gefäße (A. und V. renalis) sowie der Ureter wurden kurz vor beziehungsweise nach dem Ein- beziehungsweise Austritt aus den Nieren nicht mehr miterfasst. Zur Segmentierung des Pankreas war eine genaue Abgrenzung zum umgebenden Gewebe wichtig: Das Caput pancreatis weist eine anatomische Nähe zum Duodenum auf, indem es sich sozusagen in dessen C-Kurve schmiegt. Je nach Ausmaß des Pankreas kann die Cauda pancreatis bis in die Nähe der Milz reichen. Die A. und V. mesenterica superior sowie die A. und V. splenica wurden bei der Segmentierung nicht miteingeschlossen. (Abb.2)

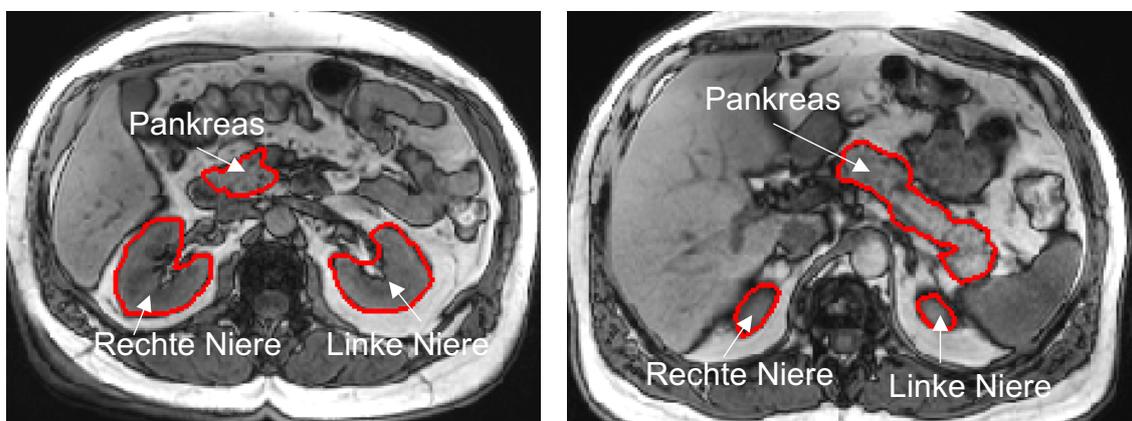


Abbildung 2: Beispielhafte Darstellung der manuellen Segmentierung der drei Organe. Die manuelle Segmentierung wurde mithilfe der Software MITK durchgeführt.

Während der manuellen Segmentierung der 100 Bilddatensätze durch die Software MITK wurde die Zeit in Minuten für jedes Organ der einzelnen MRT-Aufnahmen erfasst. Daraus wurde die Gesamtzeit der manuellen Segmentierung für alle 100 Bilddatensätze für die drei Organe zusammen berechnet. Getrennt nach den drei Organen wurden der Mittelwert mit Standardabweichung sowie das Minimum und das Maximum für die manuelle Segmentierung eines Bilddatensatzes bestimmt. Eine Veranschaulichung der zeitlichen Dauer der manuellen Segmentierung erfolgte in Boxplots.

2.3 Automatische Organsegmentierung durch ein Convolutional Neural Network

Um die automatische Organsegmentierung durchführen zu können, wurde ein „Deep Learning“-Modell, ein Convolutional Neural Network, trainiert.

CNNs bestehen aus mehreren Schichten. Es kann zwischen drei Arten von Schichten unterschieden werden. Diese sind die Faltungsschicht („convolutional layer“), die Pooling-Schicht („pooling layer“) und die vollständig verknüpfte Schicht („fully connected layer“). Zusätzlich gibt es eine Eingangs- sowie eine Ausgangsschicht. Die Faltungsschicht beschreibt die eigentliche Faltungsebene, in der die eingelesenen und markierten Merkmale aus den Eingabedaten extrahiert werden. In der Pooling-Schicht wird die Auflösung der erkannten Merkmale reduziert und verdichtet. Abschließend können in vollständig verknüpften Schichten alle Ausgangsmerkmale der vorherigen Schichten, also der Faltungs- und der Pooling-Schichten, gewichtet und summiert sowie mit dieser Schicht verknüpft werden. Eine typische CNN-Architektur umfasst einen Stapel von mehreren Faltungsschichten, gefolgt von einer Pooling-Schicht. Diese Abfolge wiederholt sich meist mehrfach. Zum Schluss können unter Umständen eine oder mehrere vollständig verknüpfte Schichten folgen. [(Yamashita et al., 2018); (Yasaka et al., 2018); (Litjens et al., 2017)]

Um ein CNN auf die Aufgabe des jeweiligen Zielkollektivs anzupassen, ist zunächst eine Trainingsphase notwendig (Abb.3). Der Lernvorgang erfolgt über ein sogenanntes Ende-zu-Ende-Prinzip (Ronneberger et al., 2015). Die Trainingsdaten werden als Eingangsdaten eingelesen und nach der internen

Verarbeitung, das heißt automatisiert durch das CNN, als Ausgangsdaten präsentiert. Die Ausgangsdaten werden dann mit den Trainingsdaten einer Fehlerfunktion zugeführt. Das CNN passt daraufhin interne Parameter an bis sich der Fehler weiter minimiert. Dieser Vorgang wiederholt sich für alle vorgesehenen Trainingsdaten bis sich der Fehler an der Ausgabe minimiert hat. [(Yasaka et al., 2018); (Jung, 2016); (Ronneberger et al., 2015)]

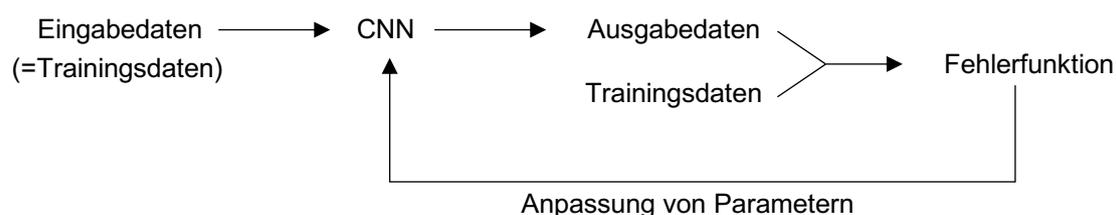


Abbildung 3: Schematische Darstellung der Trainingsphase eines CNNs. (Yasaka et al., 2018)

Das in dieser Arbeit verwendete CNN [(Küstner et al., 2019); (Küstner et al., 2018a); (Küstner et al., 2018c)] basierte auf einer Kombination an Faltungsnetzwerken von UNet, VNet, ResNet und DenseNet (Küstner et al., 2018c). UNets (Ronneberger et al., 2015) dienen hierbei der pixelweisen Lokalisierung und VNets (Milletari et al., 2016) der volumetrischen Segmentierung (Küstner et al., 2018c). Verschwindende Gradienten oder Abbauprobleme infolge von möglichen weitergegebenen Fehlern innerhalb des CNNs sollen durch ResNets (He et al., 2015) verringert werden (Küstner et al., 2018c). DenseNets führen schließlich dazu, dass der Lernvorgang überwacht geschieht (Küstner et al., 2018c). Das verwendete CNN (Abb.4) bestand aus einem absteigendem Input-Pfad zur Codierung und einem aufsteigendem Output-Pfad zur Decodierung. In Abbildung 4 ist linksseitig der Input-Pfad und rechtsseitig der Output-Pfad dargestellt. Der aufsteigende Decodierungszweig hatte die Aufgabe, die Eingabeform aus den extrahierten Merkmalen des absteigenden Codierungszweiges wiederherzustellen. Jede Stufe der beiden Zweige bestand aus mehreren Blöcken. Jeder Block umfasste dabei jeweils einen DenseConvolutional-Block mit einer Anzahl L an Schichten, gefolgt von einem TransitionalLayerUp-Block, welcher neben DenseConvolutional-

Schichten auch Pooling-Schichten enthielt. Dies führte zu insgesamt 152 Schichten und etwa 1,1 Millionen selbstständig trainierbaren Parametern innerhalb des Netzwerkes. Vollständig verknüpfte Schichten (“fully connected layers”) wurden nicht verwendet. Zur Verknüpfung zwischen den einzelnen Stufen des In- und Output-Pfades gab es Verkettungen. Gekennzeichnet ist dies in der Abbildung 4 mit „c“, was für „concatenations“ steht. Bei der Organsegmentierung ist die Lokalisierung von zentraler Bedeutung, weshalb das verwendete CNN zusätzliche Informationen über die relative Positionierung des jeweiligen Organs am Ende des Input-Pfades erhielt. [(Küstner et al., 2019); (Küstner et al., 2018c)]

Die Trainingsdaten wurden als Eingabedaten auf 3D-Würfeln mit einer Größe von 32x32x32 zugeschnitten. Verwendet wurde als zweikanalige Eingabe sowohl das Fett- als auch das Wasserbild aus der T1-gewichteten Doppelsecho-Sequenz. Dies hatte den Vorteil, dass sich die Organe besser voneinander abgrenzen ließen. (Küstner et al., 2019) Die Ausgabe präsentierte binäre Masken der segmentierten Organe (Küstner et al., 2018c).

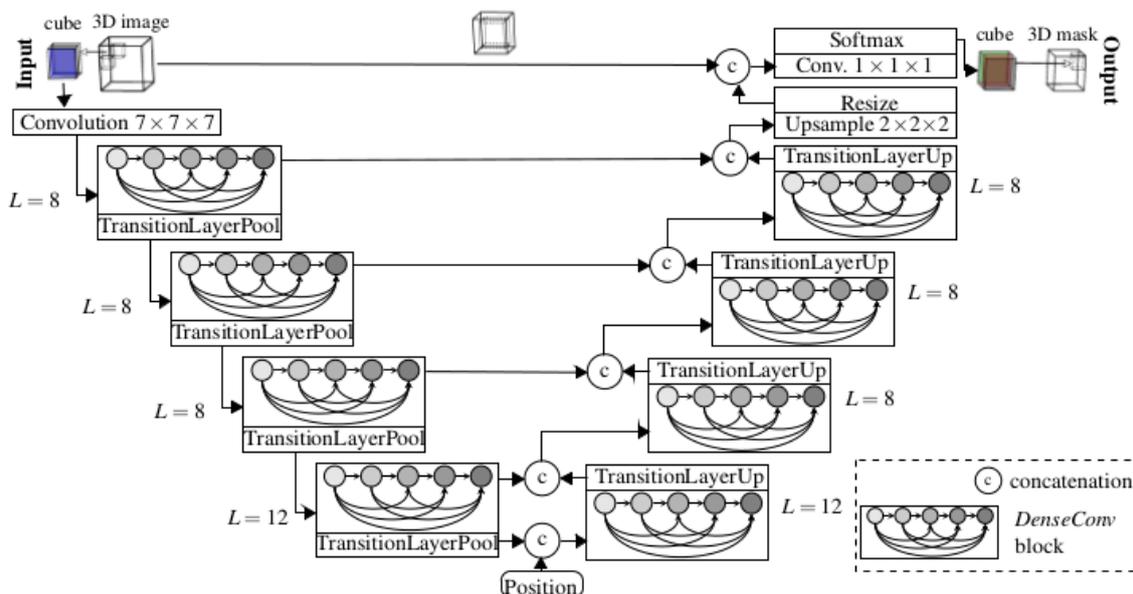


Abbildung 4: Darstellung des verwendeten Convolutional Neural Networks (CNN); (Küstner et al., 2019). Linksseitig ist der Input-Pfad (Codierung) und rechtsseitig der Output-Pfad (Decodierung) dargestellt. Zur Verknüpfung zwischen den einzelnen Stufen der beiden Pfade dienen die mit „c“ gekennzeichneten Verkettungen (concatenation). Jede Stufe der beiden Pfade bestand aus mehreren Blöcken. Es gab jeweils einen DenseConvolutional-Block mit einer Anzahl L an Schichten, an den sich ein

TransitionalLayerUp-Block (DenseConvolutional-Schichten und Pooling-Schichten) anschloss. Das CNN bestand aus 152 Schichten und ~1,1 Millionen trainierbaren Parametern. Am Ende des Input-Pfads (Codierungsweig) wurden Positionsinformationen in das Netzwerk übertragen. [(Küstner et al., 2019); (Küstner et al., 2018c)]

Zur Optimierung des maschinellen Lernalgorithmus waren verschiedene Zwischenschritte notwendig. Zunächst erfolgte die zufällige Aufteilung der 100 Datensätze in 80 Trainingsdaten und 20 Testdaten (Abb.5). Die Trainingsdaten dienten zum Training des CNNs. Die 20 Testdaten wurden ausschließlich für die spätere Evaluation des maschinellen Lernalgorithmus verwendet.

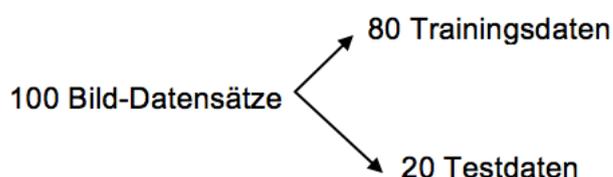


Abbildung 5: Aufteilung der 100 Bilddatensätze (MRT-Aufnahmen) in 80 Trainingsdaten und 20 Testdaten.

Es fanden zwei verschiedene Trainingsvorgänge mit den 80 Trainingsdaten statt:

- i. Training bei gleichzeitiger Validierung innerhalb der 80 Trainingsdaten (80% (n=64) Trainingsdaten, 20% (v=16) Validierungsdaten) zur Bestimmung der optimalen Anzahl an Epochen bei 5-facher Kreuzvalidierung.
- ii. Training des endgültigen Modells mit der optimalen Anzahl an Epochen aus „i“ mit allen 80 Trainingsdaten.

Die optimale Anzahl an Epochen wurde anhand der Validierungsfehlerkurven bestimmt. Eine Epoche umfasste dabei den einmaligen Durchlauf der Trainingsdaten.

Für die Trainingsvorgänge wurde als Grafikprozessor (GPU) ein „one device, GTX 1080 Ti“ verwendet. Für das Training bei gleichzeitiger Validierung innerhalb der Trainingsdaten zur Bestimmung der optimalen Epochendauer wurden folgende Hyperparameter verwendet: „patient_buffer_capacity“: 15, „batch_size“: 48, „batches_per_shift“: 35. Für das Training des endgültigen

Modells wurden folgende Hyperparameter verwendet: „patient_buffer_capacity“: 20, „batch_size“: 48 und „batches_per_shift“: 45.

Nach Abschluss der Trainingsvorgänge wurde das trainierte Modell („ii“) mit den verbliebenen 20 Testdaten evaluiert (Abb.6). Das trainierte Modell wurde hierzu auf die noch unbekannt, nicht für das Training verwendeten MRT-Aufnahmen angewendet.

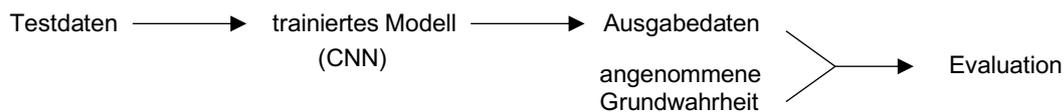


Abbildung 6: Schematische Darstellung der Testphase eines CNNs. Zur Evaluation des trainierten Modells dienen die Testdaten. Diese wurden mit der angenommenen Grundwahrheit (siehe 2.2.) verglichen. (Yasaka et al., 2018)

Die zeitliche Dauer sowohl des Trainings als auch der Evaluation wurde erfasst. Die Zeiten sind abhängig von der verwendeten Hardware, den Hyperparametern und dem geschriebenen Code. Im Gegensatz zu den sonstigen Einstellungen, die im Prozess des „Deep Learning“ automatisiert durch den Algorithmus entwickelt und gefunden werden, können diese von außen gewählt und verändert werden. Hyperparameter bezeichnen all diejenigen Parameter, die vor Beginn des Lernvorganges festgelegt werden (Chartrand et al., 2017).

Zur graphischen Veranschaulichung eines maschinellen Lernvorganges wurde eine „Loss Curve“ erstellt. Diese beschreibt die Fehlerkurven der Trainings- und Validierungsdaten eines typischen Trainingsvorganges zur Ermittlung der optimalen Anzahl an Epochen bei gleichzeitiger Validierung innerhalb des Lernprozesses (80% Trainingsdaten, 20% Validierungsdaten).

2.4 Datenauswertung

Zur Analyse und Auswertung der Ergebnisse wurde die Statistiksoftware IBM SPSS Statistics sowie Microsoft Excel verwendet. Mithilfe von SPSS Statistics wurden auch die Diagramme zur Visualisierung der Ergebnisse erstellt. Die Anfertigung der Tabellen erfolgte mit Microsoft Word.

2.4.1 Vergleich Trainings- und Testdaten

Unabhängig von der Verwendung des Algorithmus ließen sich die Datensätze der Trainingsdaten, mit denen der Testdaten vergleichen. Die 80 Trainingsdaten wurden mit den 20 Testdaten hinsichtlich demographischer und MRT-bezogener Daten verglichen. Untersucht wurde das Alter, das Geschlecht und der BMI als demographische Daten. Als MRT-bezogene Daten wurde jeweils das Organvolumen der drei Organe herangezogen, welches hierbei allein auf die manuelle Segmentierung zurückzuführen war.

In Bezug auf das Alter und den BMI wurde das Minimum, das Maximum und der Mittelwert mit Standardabweichung berechnet, in Bezug auf das Geschlecht die prozentuale Verteilung hinsichtlich männlicher und weiblicher Probanden berechnet. Das Organvolumen wurde getrennt nach den drei Organen untersucht und hierbei der Mittelwert mit Standardabweichung bestimmt. Die P-Werte für das Alter, den BMI und die Organvolumina wurden mithilfe von T-Tests für die jeweilige Mittelwertgleichheit für unabhängige Stichproben analysiert. Es lagen jeweils Varianzgleichheit laut des Levene-Tests und normalverteilte Daten nach Shapiro-Wilk vor. Somit waren die Voraussetzungen für die Anwendung des T-Tests erfüllt. Mit dem Chi-Quadrat-Test nach Pearson wurde auf Gleichheit der Geschlechterverteilung in den beiden Gruppen untersucht. Das Organvolumen der drei Organe wurde graphisch verglichen. Visualisiert wurde dies mithilfe von Boxplots getrennt nach den drei Organen. Zusätzlich wurde jeweils das durchschnittliche Organvolumen der drei Organe der gesamten 100 Datensätze mit der jeweiligen Standardabweichung angegeben.

2.4.2 Vergleich automatische und manuelle Segmentierung

Um zu untersuchen, inwiefern der trainierte Lernalgorithmus in der Lage ist, in noch unbekanntem MRT-Bilddaten Organe korrekt zu erkennen und zu markieren, wurde er nach dem Training auf die 20 Testdatensätze angewendet. Der Erfolg des Algorithmus ließ sich evaluieren, indem die Ergebnisse der automatischen Organsegmentierung der 20 Testdaten durch den maschinellen Lernalgorithmus mit den Ergebnissen der manuellen Organsegmentierung – also der angenommenen Grundwahrheit (siehe 2.2.) – verglichen wurden.

Die Ergebnisse wurden anhand von drei Ansätzen miteinander verglichen: Einerseits visuell qualitativ, andererseits quantitativ und zusätzlich unter weiteren sekundären Parametern.

2.4.2.1 Visuell qualitativer Vergleich

Der visuell qualitative Vergleich der Ergebnisse der automatischen Organsegmentierung, die im Zuge der Analyse der 20 Testdaten durch den trainierten Algorithmus erzielt wurden, mit den Ergebnissen der manuellen Segmentierung erfolgte anhand von zwei dafür entwickelten Scores (Tab.2, Tab.3). Die Auswertung erfolgte unter qualitativen Aspekten. Der Score „Organsegmentierung“ (Tab.2) beschreibt die Präzision der Segmentierung bezogen auf das jeweilige Organ. Durch einen weiteren Score (Tab.3) wurden Fehlsegmentierungen außerhalb der eigentlich zu segmentierenden Organe erfasst, inklusive der Information, in wie vielen Schichten diese aufgetreten waren. Wenn keine Fehlsegmentierungen außerhalb des zu segmentierenden Organs vorlagen, wurde dies durch diesen Score ebenfalls erfasst. Es wurden somit sonstige Fehlsegmentierungen unabhängig von der eigentlichen Organsegmentierung untersucht. Im Folgenden wird dieser zweite Score als „sonstige Fehlsegmentierungen“ bezeichnet.

Die Auswertung geschah getrennt nach den drei Organen (linke Niere, rechte Niere, Pankreas). Dazu wurden die Bilddaten als Formate in NIFTI (MRT-Aufnahme) und NRRD (Segmentierung der Organe durch den Algorithmus) erzeugt und anschließend nacheinander in dem Programm Aliza Medical Imaging & DICOM Viewer geöffnet und miteinander fusioniert, um die Organsegmentierungen zu evaluieren. Dabei wurde zur besseren Unterscheidung der einzelnen Organe für die linke Niere die Farbe „Syngo 1792“ (weißlich), für die rechte Niere „Rainbow B 1536“ (rötlich) und für das Pankreas „Labels 4096“ (grünlich) gewählt (Abb.7).

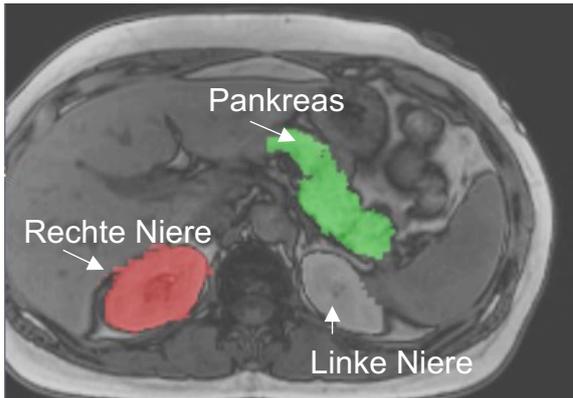


Abbildung 7: Beispiel eines Datensatzes zur Darstellung der automatischen Organsegmentierung. Die Abbildung zeigt eine MRT-Aufnahme in 2D-Ansicht. Die drei Organe wurden durch das Programm Aliza Medical Imaging & DICOM Viewer farblich markiert.

Tabelle 2: Score der visuell qualitativen Auswertung bezogen auf das jeweilige Organ. Score: „Organsegmentierung“.

Nummer	Erläuterung
1	Keine wesentliche, visuell erfassbare Abweichung.
2	Minimal visuell erfassbare Abweichungen: Einzelne Pixel fehlen im Organ beziehungsweise sind diesem außen angelagert. Vereinzelt auftreten (Abb.8).
3	Geringe visuell erfassbare Abweichungen: Kleine Überschreitungen der Organgrenze beziehungsweise kleine Aussparungen im Organ selbst. Mehrere, zusammenhängende Pixel sind betroffen, aber überwiegend seltenes Auftreten (Abb.8).
4	Deutlich visuell erkennbare Abweichungen: Deutliche Aussparungen im Organ (mind. 1/5 betroffen) beziehungsweise Organ wurde fehlerhaft in mehrere Anteile aufgegliedert. Diese Abweichungen treten wiederholt auf (Abb.9).
5	Größere visuell erkennbare Abweichungen: Organ in Grundzügen erkennbar, aber Einbeziehung benachbarter Organe beziehungsweise großer Anteil des Organs (mind. 1/3) wird nicht erkannt (Abb.9).

6	Keine beziehungsweise minimale Organsegmentierung, falsche Lokalisierung, andere Organerkennung beziehungsweise Einbeziehung anderer Organe in ähnlichem Ausmaß wie das ursprüngliche Organ (Abb.9).
----------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

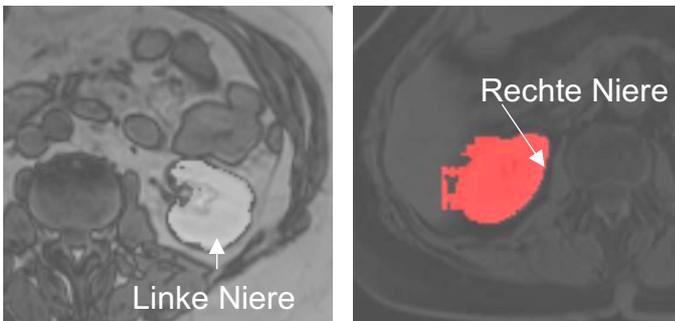


Abbildung 8: Score „Organsegmentierung“ der visuell qualitativen Auswertung. Beispielhafte Abbildung anhand der linken Niere für die Nummer 2 und der rechten Niere für die Nummer 3 des Scores „Organsegmentierung“. Erkennbar ist die Organsegmentierung an der weißlichen (linke Niere) beziehungsweise der rötlichen (rechte Niere) Markierung.



Abbildung 9: Score „Organsegmentierung“ der visuell qualitativen Auswertung. Beispielhafte Abbildung anhand des Pankreas (linke Abbildung) für die Nummer 4, der linken Niere (in der Mitte) für die Nummer 5 und des Pankreas (rechte Abbildung) für die Nummer 6 des Scores „Organsegmentierung“. Erkennbar ist die Organsegmentierung an der weißlichen (linke Niere) beziehungsweise der grünlichen (Pankreas) Markierung.

Tabelle 3: Ergänzender Score zur Erfassung des Auftretens von Fehlsegmentierungen außerhalb des Organs. Score: „sonstige Fehlsegmentierungen“.

Nummer	Anzahl der betroffenen Schichten
0	= 0
I	≤ 10
II	> 10

Zur graphischen Veranschaulichung wurden Balkendiagramme erstellt. Diese bilden die absoluten Häufigkeiten der einzelnen Nummern des Scores „Organsegmentierung“ sowie des Scores „sonstige Fehlsegmentierungen“ ab. Des Weiteren wurden die hauptsächlich aufgetretenen Fehlsegmentierungen unabhängig vom eigentlichen Organ mit Angabe des jeweiligen Ortes tabellarisch aufgelistet.

2.4.2.2 Quantitativer Vergleich

Der quantitative Vergleich der Ergebnisse der automatischen Organsegmentierung mit denen der manuellen Organsegmentierung geschah durch die Bestimmung von Metriken (Tab.4).

Analysiert wurden jeweils der Hintergrund der MRT-Aufnahme und die Organsegmentierung der drei Organe. Die tabellarisch angegebenen Messgrößen sind gemittelte Werte der 20 Testdatensätze. Um die Genauigkeit der Segmentierung hinsichtlich der einzelnen Voxel beurteilen zu können, dienten zwei Ähnlichkeitskoeffizienten, der Dice-Index und der Jaccard-Index. Da es sich hierbei um Maße der Korrelation handelt, entspricht der größte Wert dem ähnlichsten Paar (Vorontsov et al., 2013).

Es erfolgte zusätzlich eine graphische Darstellung anhand von drei Boxplots für die Werte des Dice-Index der 20 Testdaten getrennt nach den drei Organen.

Tabelle 4: Verwendete Messgrößen der quantitativen Auswertung. Verglichen wurden die Ergebnisse der 20 Testdaten der automatischen Segmentierung mit denen der manuellen Segmentierung.

Untersuchte Messgrößen
Sensitivität
Spezifität
Präzision
Falsch-Negativ-Rate
Falsch-Positiv-Rate
Dice-Index
Jaccard-Index

2.4.2.3 Vergleich sekundärer Parameter

Als weitere Parameter, um die Ergebnisse der automatischen Segmentierung mit denen der manuellen Segmentierung zu vergleichen, dienten das Organvolumen und der Fettgehalt der drei Organe nach der jeweiligen Segmentierung. Das Organvolumen wurde aus den manuell und automatisch erstellten Organmasken bestimmt und der Fettgehalt für jedes Organ als Mittelwert berechnet.

Die Fettquantifizierung erfolgte durch die Methode der Dixon-Sequenz. Durch Dixon-Sequenzen entstehen in MRT-Aufnahmen verschiedene Bilddatensätze wie beispielsweise reine Fett- oder Wasserbilder, die den jeweils anderen Kontrast unterdrücken. Indem das Fett-Signal zum Gesamtsignal ins Verhältnis gesetzt wird, kann der Fettanteil im Gewebe ermittelt werden. Um den Fettgehalt der Nieren und des Pankreas zu erhalten, wurde für jedes Voxel innerhalb der jeweiligen Organmaske der zugehörige Fettgehalt aus den Intensitätswerten des korrespondierenden Voxels bestimmt und dann der Mittelwert daraus gebildet. [(Perkins et al., 2015); (Ma, 2008); (Dixon, 1984)]

Es wurde jeweils für die linke Niere, die rechte Niere und das Pankreas sowohl für das Organvolumen als auch für den Fettgehalt die durchschnittliche prozentuale Differenz zwischen manuell und automatisch erstellten Masken als Mittelwert mit Standardabweichung berechnet. Tabellarisch dienten als Vergleichswerte für das Organvolumen und den Fettgehalt jeweils die Angabe des Mittelwertes mit der Standardabweichung. Die P-Werte wurden mithilfe von T-Tests für die jeweilige Mittelwertgleichheit für verbundene Stichproben analysiert. Es erfolgte eine graphische Darstellung in Form von Boxplots getrennt nach den drei Organen.

Zusätzlich wurde die Korrelation der Organvolumina getrennt nach den drei Organen für manuell und automatisch erstellte Masken untersucht. Berechnet wurde jeweils die Korrelation nach Pearson sowie die Signifikanz (2-seitig) durch Angabe des P-Wertes. Graphisch dargestellt wurde dies in Form von Streudiagrammen.

3 Ergebnisse

Im Folgenden sind die Ergebnisse dieser Untersuchung unter den oben beschriebenen Gesichtspunkten aufgegliedert.

3.1 Daten über die Population der MRT-Datensätze

Unter den 100 Probanden, die zwischen 22 und 72 Jahre alt waren, waren sowohl Männer als auch Frauen. Das durchschnittliche Alter betrug 53 Jahre. Die Geschlechtsverteilung war mit 51% weiblichen und 49% männlichen teilnehmenden Personen ausgeglichen. Die untersuchte Population wies einen mittleren BMI von 27,3 kg/m² mit einer Standardabweichung von 5,0 kg/m² auf (Tab.5).

Table 5: Demographische Daten des Probandenkollektivs der verwendeten MRT-Aufnahmen der NAKO-Studie.

Untersuchte Parameter	Einheit	Werte der 100 Probanden
Alter	Jahre	53 ± 9,6 Min.: 22; Max.: 72
Geschlecht	%	w: 51; m: 49
BMI	kg/m ²	27,3 ± 5,0 Min.: 17,8; Max.: 43,8

Legende:

w = weiblich

m = männlich

Min. = Minimum

Max. = Maximum

3.2 Manuelle Segmentierung

Alle MRT-Aufnahmen der 100 Datensätze waren von guter diagnostischer Bildqualität.

Die Gesamtzeit für die manuelle Segmentierung der drei Organe (linke Niere, rechte Niere, Pankreas) betrug etwa 86 Stunden. Dabei nahm die Zeit für die manuelle Segmentierung des Pankreas in etwa doppelt bis vierfach so viel Zeit in Anspruch wie die Segmentierung einer Niere (Tab.6, Abb.10). Im Durchschnitt wurden für die Segmentierung jeweils einer Niere ca. 11 - 13 Minuten benötigt, während dies für das Pankreas knapp 30 Minuten dauerte (Tab.6).

Tabelle 6: Zeitliche Dauer der manuellen Segmentierung der drei Organe der 100 MRT-Datensätze.

Organ	Mittelwert in min	Standardabweichung in min	Minimum in min	Maximum in min
Linke Niere	11,62	4,93	4	31
Rechte Niere	12,24	4,91	5	32
Pankreas	27,87	9,81	10	60

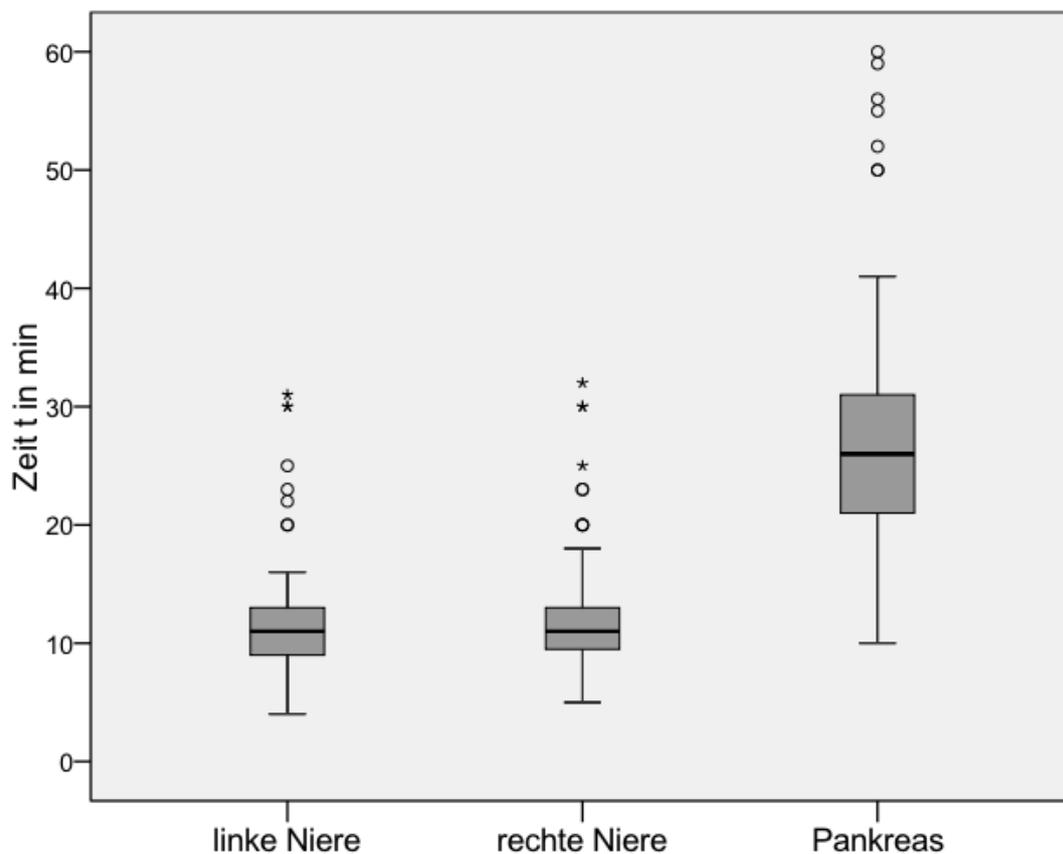


Abbildung 10: Graphische Darstellung der zeitlichen Dauer der manuellen Segmentierung der 100 MRT-Datensätze.

3.3 Algorithmus zur automatischen Organsegmentierung

Zur Optimierung des maschinellen Lernalgorithmus erfolgten die zwei im Methodikteil beschriebenen Trainingsvorgänge:

- i. Training zur Bestimmung der optimalen Anzahl an Epochen bei 5-facher Kreuzvalidierung:
Eine Epoche: ~ 35 Minuten, 5 x 50 Epochen: ~ 5 Tage.
- ii. Training des endgültigen Modells:
Eine Epoche: ~ 30 Minuten, Gesamtzeit: ~ 20 Stunden.

Im Trainingsvorgang „i“ wurden 50 Epochen durchlaufen. Aus den Validierungsfehlerkurven (Abb.11) wurde die Anzahl 40 als optimale Epochendauer bestimmt. Somit fand das Training „ii“, das Training des endgültigen Modells, mit 40 Epochen statt. Das letzte Modell nach den 40 Epochen wurde gespeichert. Mit diesem Modell wurde mithilfe der 20 Testdaten die Evaluation durchgeführt. Die Dauer für die Inferenz einer MRT-Aufnahme eines Probanden betrug etwa 5 bis 10 Minuten.

Die „Loss Curve“ (Abb.11) zeigt die Fehlerkurven der Trainings- und Validierungsdaten eines typischen Trainingsvorganges zur Bestimmung der optimalen Anzahl an Epochen. Diese beziehen sich auf einen Lernvorgang mit 50 Epochen bei gleichzeitiger Validierung innerhalb des Lernprozesses (80% Trainingsdaten, 20% Validierungsdaten). Im Verlauf des in der „Loss Curve“ gezeigten Trainings kam es zu sichtbaren Veränderungen. Zunächst wurde eine steigende Verbesserung sichtbar. Gekennzeichnet ist diese durch den Abfall der „loss“, was eine Abnahme des Fehlers beschreibt. Ab etwa 20 Trainingsdurchgängen wurde langsam eine Sättigung erreicht. Es war kein „overfitting“ sichtbar. Dies lässt sich daran erkennen, dass sowohl der Trainingsfehler („training loss“) als auch der Validierungsfehler („validation loss“) abnahmen.

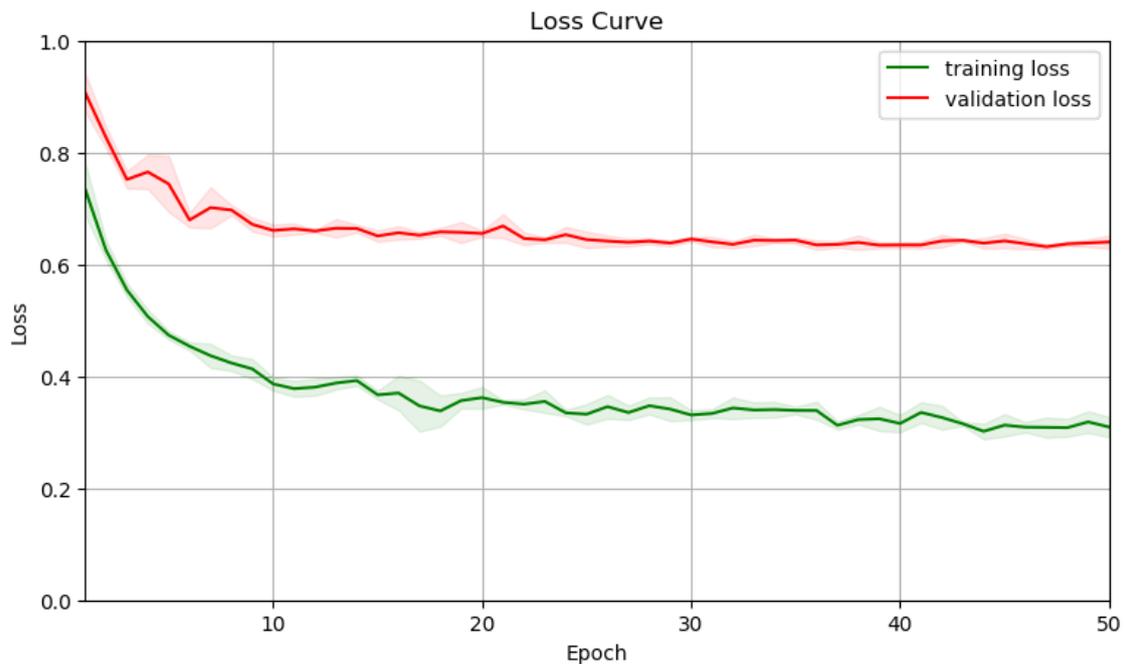


Abbildung 11: Beispiel einer „Loss Curve“ zur Beschreibung des Lernvorganges eines automatischen Lernalgorithmus. Dargestellt ist ein Trainingsvorgang bei gleichzeitiger Validierung innerhalb des Lernprozesses (80% Trainingsdaten, 20% Validierungsdaten). Anhand dieser Validierungsfehlerkurven („training loss“, „validation loss“) ließ sich die optimale Anzahl an Epochen des Lernvorganges bestimmen.

Legende:

Epoch = Epochen

„training loss“ = Trainingsfehler

„Loss“ = Verlust, Fehler

„validation loss“ = Validierungsfehler

3.4 Vergleich Trainings- und Testdaten

Die 80 Trainingsdaten wurden mit den 20 Testdaten hinsichtlich demographischer und MRT-bezogener Daten verglichen (Tab.7).

Die Alters- und Geschlechtsverteilung sowie der BMI waren in den beiden untersuchten Gruppen (Trainings-, Testdaten) ähnlich. Erkennen lässt sich das an dem jeweiligen Mittelwert mit Standardabweichung beziehungsweise beim Geschlecht an der prozentualen Verteilung (Tab.7). In der Gruppe der Trainingsdaten lag in Bezug auf das Alter ein Mittelwert von 52 Jahren mit einer Standardabweichung von 9,5 Jahren vor, in der Gruppe der Testdaten ein Mittelwert von 55 Jahren mit einer Standardabweichung von knapp 10 Jahren. Etwa jeweils die Hälfte der Probanden war in beiden Gruppen weiblich und die andere Hälfte männlich. Der durchschnittliche BMI betrug 27,0 kg/m² in der Gruppe der Trainingsdaten beziehungsweise 28,5 kg/m² in der Gruppe der

Testdaten. Das Organvolumen wich zwischen den beiden Gruppen leicht ab. In Bezug auf die drei Organe war dies aber sehr ähnlich (Tab.7). In den Abbildungen 12 bis 14 wurde das Organvolumen der drei Organe jeweils zwischen Trainings- und Testdaten graphisch verglichen.

Das Organvolumen - über die 100 Datensätze zusammen betrachtet - betrug im Durchschnitt Folgendes:

Linke Niere: 176,9 ml \pm 52,9 ml

Rechte Niere: 165,7 ml \pm 43,7 ml

Pankreas: 80,0 ml \pm 24,0 ml.

Tabelle 7: Demographische und MRT-bezogene Daten des Probandenkollektivs. Unterschieden wurde zwischen den Trainings- und Testdatensätzen.

Erhobene Parameter	Einheit	Werte der 80 Trainingsdaten	Werte der 20 Testdaten	P-Wert
Alter	Jahre	52 \pm 9,5 Min.: 22; Max.: 72	55 \pm 9,8 Min.: 40; Max.: 70	p = .262
Geschlecht	%	w: 52,5; m: 47,4	w: 45,0; m: 55,0	p = .548
BMI	kg/m ²	27,0 \pm 5,1 Min.: 17,8; Max.: 43,8	28,5 \pm 4,9 Min.: 23,1; Max.: 42,6	p = .219
Organvolumen linke Niere	ml	173,3 \pm 54,0	191,3 \pm 46,7	p = .173
Organvolumen rechte Niere	ml	162,3 \pm 41,5	179,2 \pm 50,6	p = .120
Organvolumen Pankreas	ml	77,9 \pm 21,2	88,2 \pm 32,5	p = .189

Legende:

w = weiblich

m = männlich

Min. = Minimum

Max. = Maximum

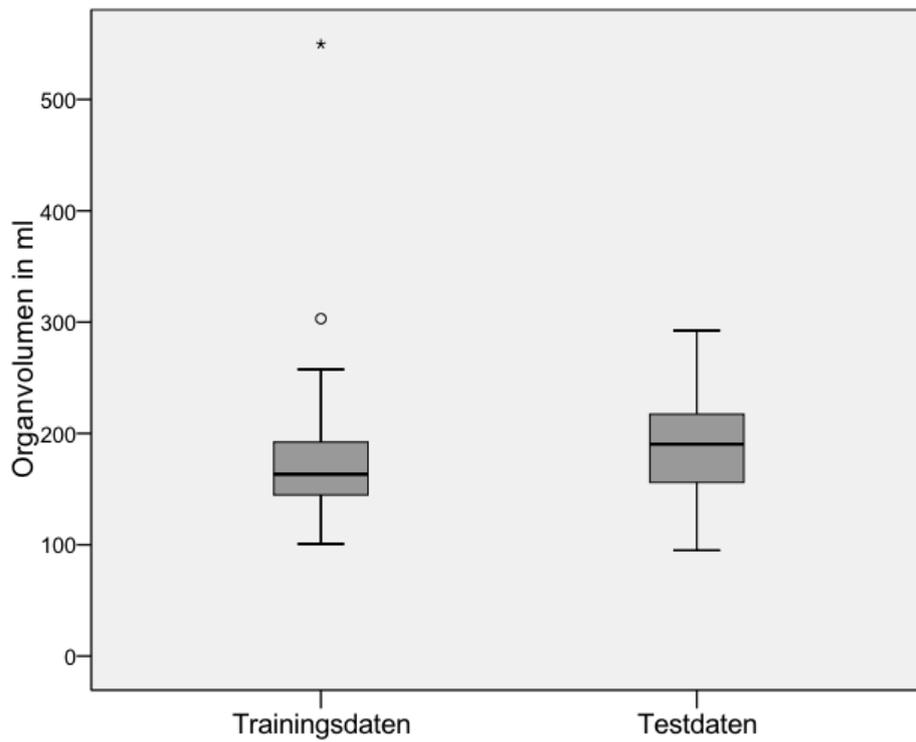


Abbildung 12: Linke Niere: Vergleich zwischen den Trainings- und Testdatensätzen in Bezug auf die Organvolumina. Die Organvolumina waren auf die manuelle Segmentierung zurückzuführen.

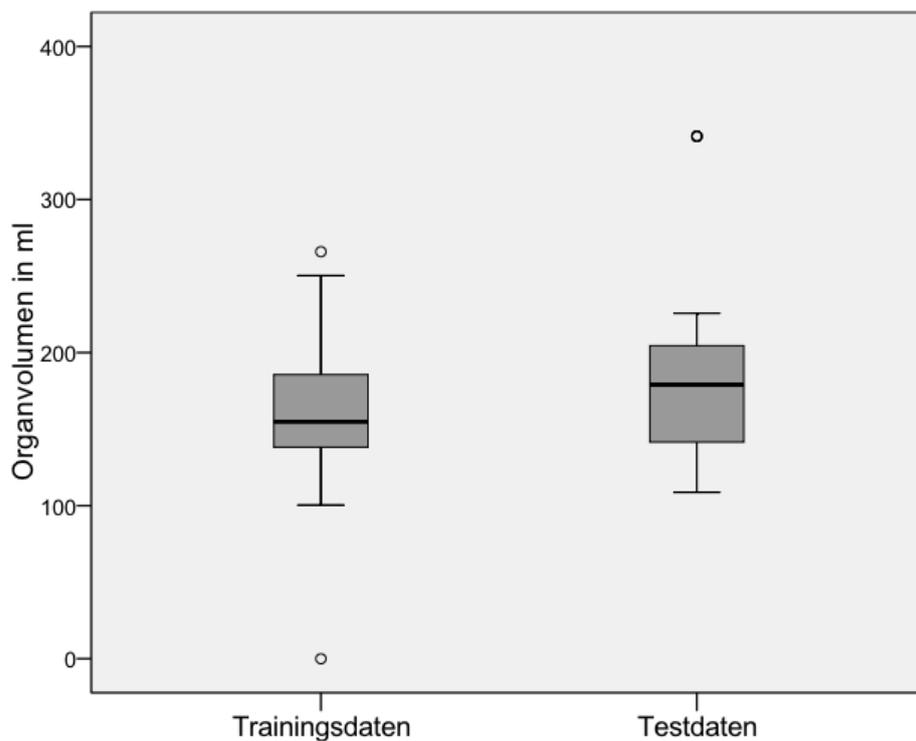


Abbildung 13: Rechte Niere: Vergleich zwischen den Trainings- und Testdatensätzen in Bezug auf die Organvolumina. Die Organvolumina waren auf die manuelle Segmentierung zurückzuführen.

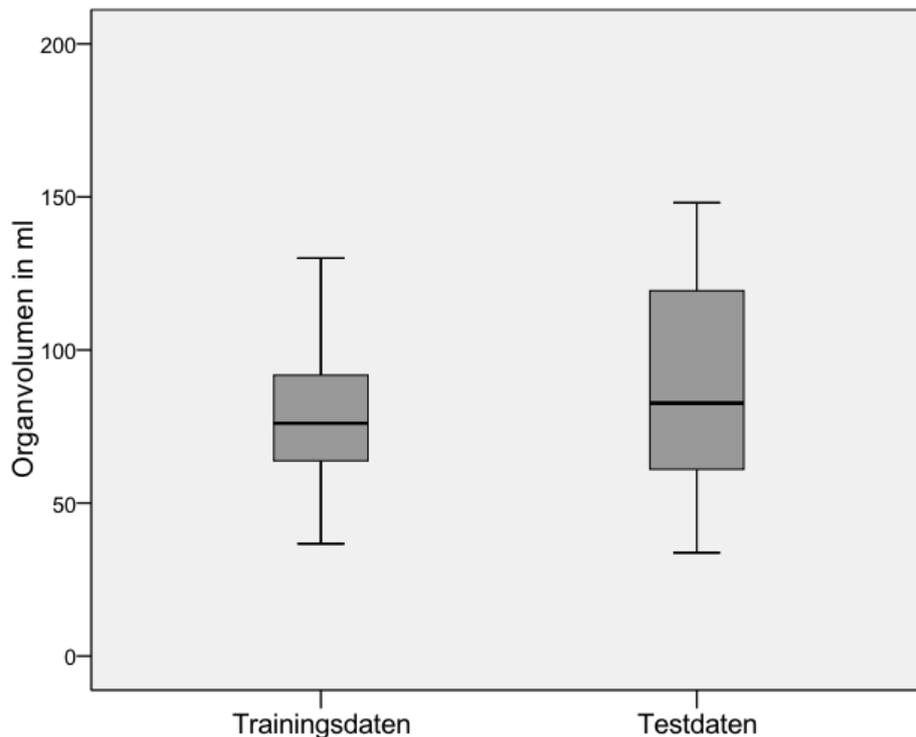


Abbildung 14: Pankreas: Vergleich zwischen den Trainings- und Testdatensätzen in Bezug auf die Organvolumina. Die Organvolumina waren auf die manuelle Segmentierung zurückzuführen.

3.5 Vergleich automatische und manuelle Segmentierung

Die Auswertung zum Vergleich der Ergebnisse der automatischen Segmentierung mit denen der manuellen Segmentierung bezog sich auf die 20 Testdatensätze.

Insgesamt ließ sich erkennen, dass die beiden Nieren genauer als das Pankreas durch den maschinellen Lernalgorithmus segmentiert werden konnten. Dies wird im Folgenden genauer erörtert.

3.5.1 Visuell qualitative Auswertung

Die Ergebnisse der automatischen Segmentierung durch den maschinellen Lernalgorithmus der 20 Testdaten wurden einzeln anhand von visuell qualitativen Kriterien ausgewertet und den zwei dafür entwickelten Scores (Tab.2, Tab.3) zugeordnet. In der folgenden Abbildung (Abb.15) ist anhand von drei verschiedenen Ansichten die automatische Organsegmentierung durch den Algorithmus nach farblicher Markierung beispielhaft dargestellt.

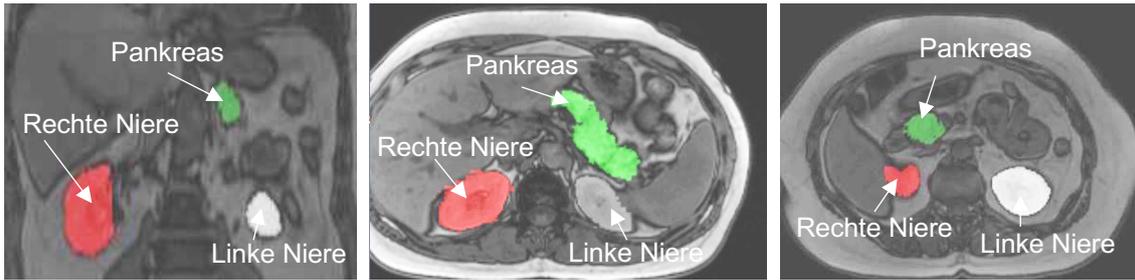


Abbildung 15: Beispielhafte Darstellung der automatischen Organsegmentierung der drei Organe. Die linke Niere (weißliche Markierung), die rechte Niere (rötliche Markierung) und das Pankreas (grünliche Markierung) sind in verschiedenen Ansichten dargestellt.

Die Ergebnisse des visuell qualitativen Vergleichs wiesen eindeutige Unterschiede zwischen der Segmentierung der beiden Nieren und der Segmentierung des Pankreas auf. Die erzielte Genauigkeit der automatischen Segmentierung der Nieren durch den Algorithmus war signifikant höher (Abb.20, Abb.21). Dies zeigte sich in den beiden zur Auswertung dienenden Scores („Organsegmentierung“ und „sonstige Fehlsegmentierungen“). Der Vergleich der automatischen Segmentierung der beiden Nieren zeigte, dass die linke Niere im Durchschnitt am genauesten automatisch segmentiert werden konnte (Abb.20, Abb.21). In der folgenden Abbildung (Abb. 16) ist beispielhaft die automatische Organsegmentierung der linken und der rechten Niere dargestellt.

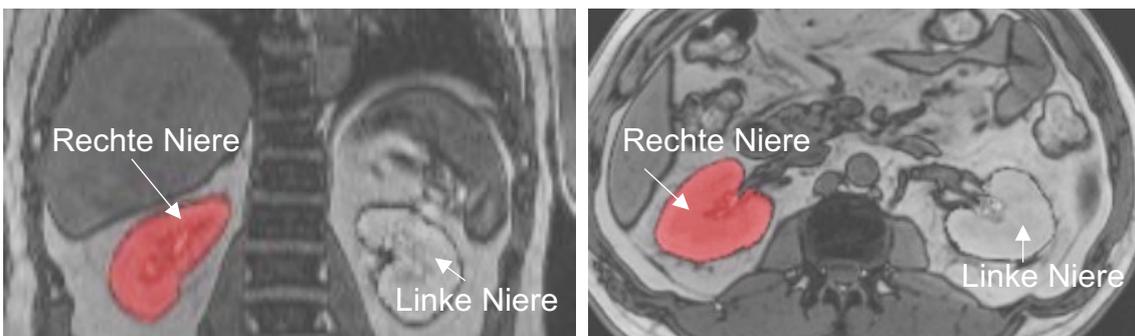


Abbildung 16: Beispielhafte Darstellung der Organsegmentierung der linken und der rechten Niere. Die beiden Nieren sind in koronarer beziehungsweise in axialer Ansicht der MRT-Aufnahme dargestellt.

Die Auswertung des Scores „Organsegmentierung“ ließ sich wie folgt beschreiben (Abb.20): Die automatische Segmentierung der linken Niere war hauptsächlich gut bis sehr gut (Nummer 1 und 2 des Scores

„Organsegmentierung“; Abb.17, Abb.20). Bei einem Datensatz lag einmalig eine deutlich schlechtere Segmentierung der linken Niere vor (Nummer 5 des Scores „Organsegmentierung“). Die Ergebnisse der automatischen Segmentierung der rechten Niere ließen sich überwiegend als gut (Nummer 3 des Scores „Organsegmentierung“), teilweise auch als sehr gut (Nummer 1 und 2 des Scores „Organsegmentierung“) beschreiben (Abb.17, Abb.20).

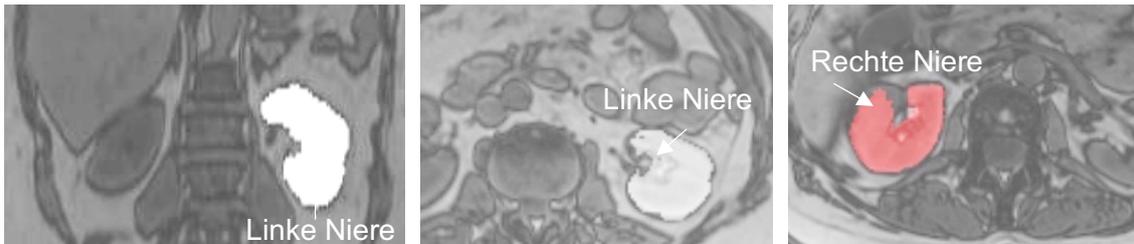


Abbildung 17: Beispiele der automatischen Organsegmentierung der Nieren. Dargestellt ist jeweils ein Beispiel für die sehr gute Segmentierung der linken Niere (linke Abbildung) und für typische Fehler bei der Segmentierung der Nieren (hier sowohl am Beispiel der linken Niere (mittlere Abbildung) als auch der rechten Niere (rechte Abbildung).

Die automatische Segmentierung des Pankreas erreichte unter den 20 Testdaten in keinem Fall die Anforderungen für eine sehr gute Organsegmentierung (Nummer 1 oder 2 des Scores „Organsegmentierung“). Wenn das Pankreas gut bis sehr gut segmentiert wurde, bezog sich dies allein auf einzelne Schichten und nicht auf alle Schichten der jeweiligen MRT-Aufnahme. Deshalb konnte bei der automatischen Segmentierung des Pankreas keine Zuordnung zur Nummer 1 oder 2 des Scores „Organsegmentierung“ erfolgen. Die automatische Segmentierung des Pankreas ließ sich überwiegend als durchschnittlich beschreiben (Nummer 4 des Scores „Organsegmentierung“; Abb.18, Abb.20). Es kam jedoch auch zu Ergebnissen einer schlechten Segmentierung (Nummer 5 und 6 des Scores „Organsegmentierung“; Abb.18, Abb.20).

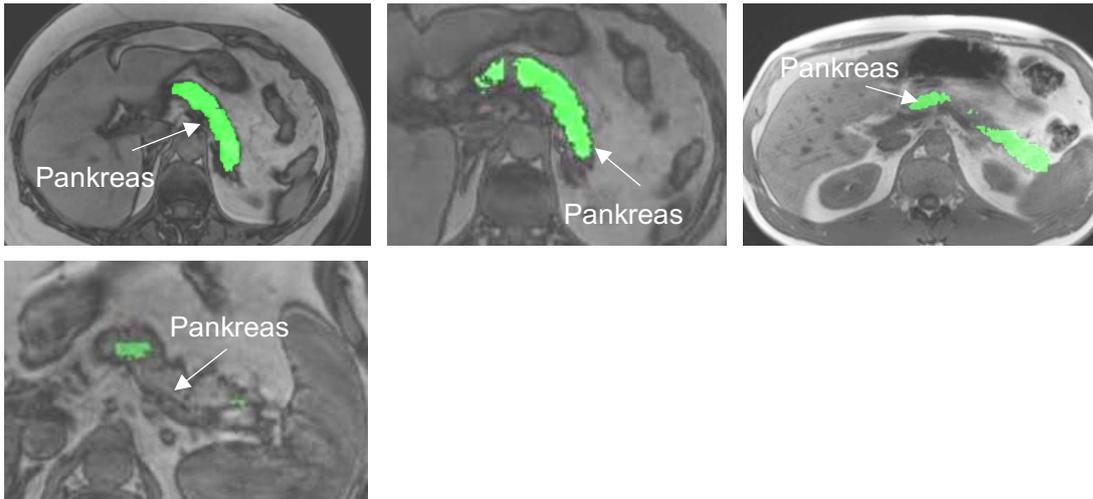


Abbildung 18: Beispielhafte Darstellungen der automatischen Segmentierung des Pankreas. Dargestellt sind jeweils Beispiele für die automatische Segmentierung des Pankreas in unterschiedlichen Genauigkeiten (grünliche Markierung).

Die Ergebnisse der visuell qualitativen Auswertung bezogen auf den Score „Organsegmentierung“ werden in Abbildung 19 beispielhaft verdeutlicht. Zusammenfassend lässt sich sagen, dass die beiden Nieren durch den maschinellen Lernalgorithmus gut bis sehr gut segmentiert werden konnten, bei der automatischen Segmentierung des Pankreas hingegen kam es zu unterschiedlich starken Abweichungen.

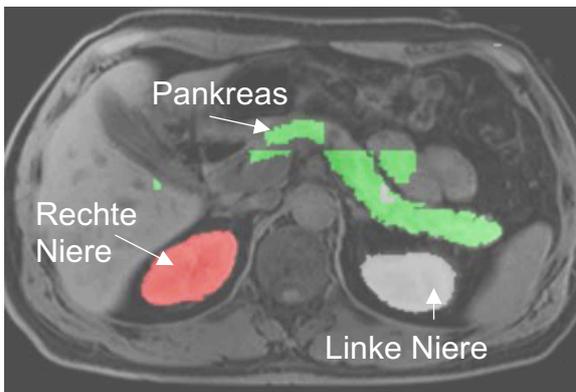


Abbildung 19: Beispielhafte Darstellung der automatischen Segmentierung der drei Organe. Die Auswahl geschah visuell qualitativ nach der Genauigkeit, betrachtet über die gesamten Testdaten: Linke und rechte Niere: gute bis sehr gute Segmentierung; Pankreas: überwiegend durchschnittliche Segmentierung.

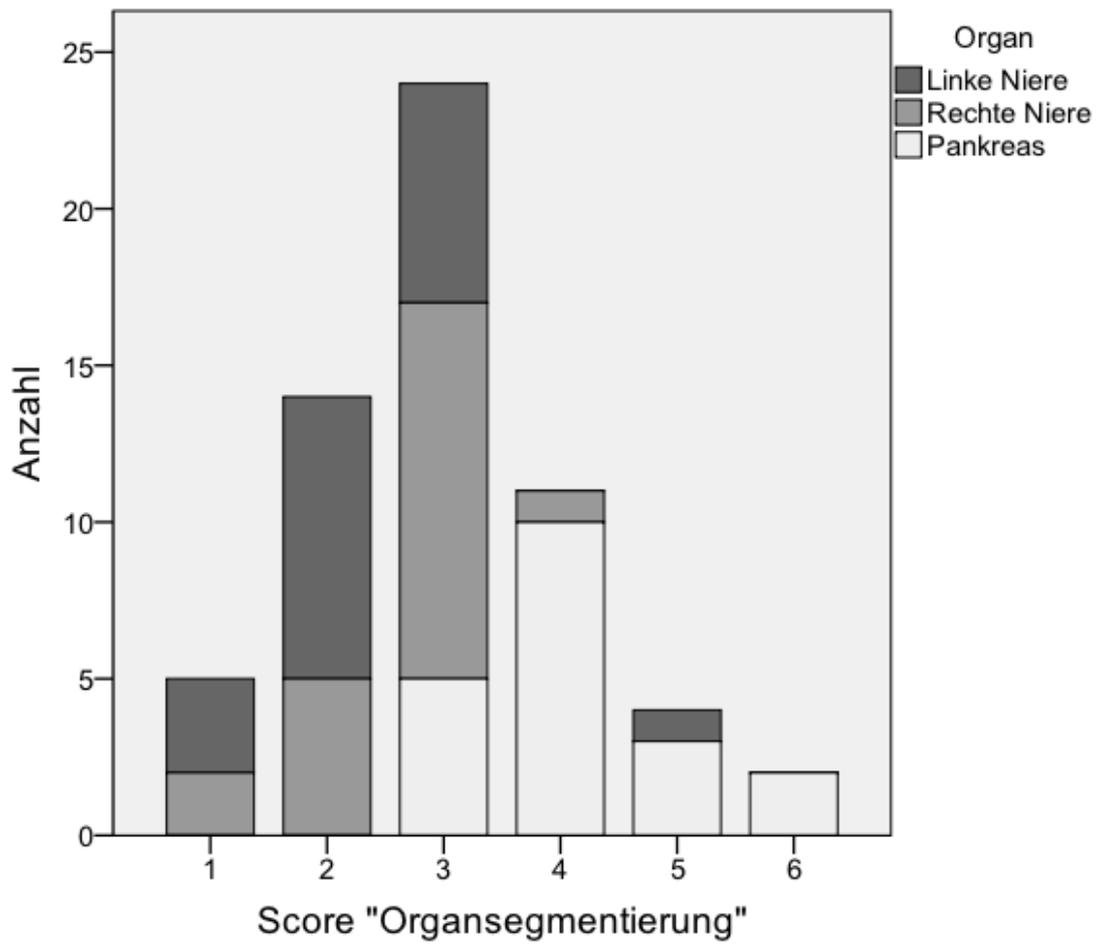


Abbildung 20: Score „Organsegmentierung“. Balkendiagramm zur Veranschaulichung der absoluten Häufigkeiten der Nummern 1 bis 6 des Scores „Organsegmentierung“ (Tab.2) der visuell qualitativen Auswertung der Testdatensätze.

Anhand des Scores „sonstige Fehlsegmentierungen“ ließ sich erkennen, dass bei der automatischen Segmentierung der Nieren seltener Fehlsegmentierungen außerhalb des eigentlichen Organs auftraten als bei der automatischen Segmentierung des Pankreas (Abb.21). Am seltensten traten bei der automatischen Segmentierung der linken Niere Fehlsegmentierungen außerhalb des eigentlichen Organs auf (Abb.21).

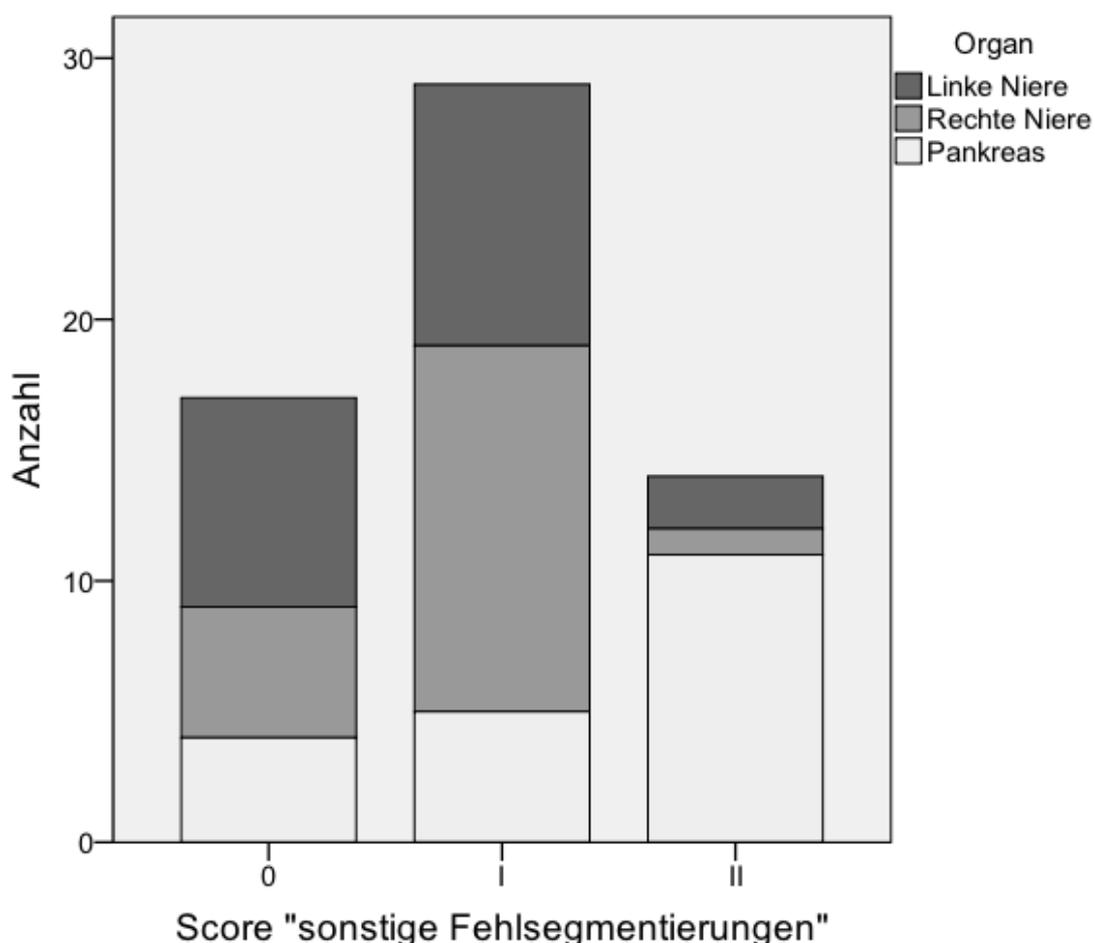


Abbildung 21: Score „sonstige Fehlsegmentierungen“. Balkendiagramm zur Veranschaulichung der absoluten Häufigkeiten der einzelnen Nummern (0,I,II) des Scores „sonstige Fehlsegmentierungen“ (Tab.3) der visuell qualitativen Auswertung der Testdatensätze.

Bei Fehlsegmentierungen außerhalb des ursprünglich beabsichtigten Organs wurden insbesondere drei Organe miterfasst (Tab.8): Dazu zählten bei der automatischen Segmentierung aller drei Organen der Darm, bei der

Segmentierung der linken Niere und des Pankreas zusätzlich die Milz sowie bei der Segmentierung der rechten Niere und des Pankreas zusätzlich die Leber. Die Fehlsegmentierungen reichten von einzelnen Pixeln, die punktförmig sichtbar waren, bis zu kleinen flächigen Bereichen, die maximal etwa die Fläche einer Hälfte einer Niere ausmachten. Diese Fehlsegmentierungen traten in zwei bis maximal 25 Schichten je MRT-Aufnahme auf. Einmalig wurde fehlerhaft anstelle der rechten Niere zusätzlich die linke Niere in den entsprechenden Schichten anteilig segmentiert. Eine seltene Abweichung war bei der Segmentierung des Pankreas in einem Datensatz zu verzeichnen: Hier wurde die Wirbelsäule in einigen Schichten anteilig mitsegmentiert. Ähnliches trat bei anderen Datensätzen auf: Bei der Segmentierung einer linken Niere wurde sowohl Rückenmuskulatur als auch das Pankreas anteilig mitsegmentiert beziehungsweise bei der Segmentierung einer anderen linken Niere die Bauchmuskulatur.

Tabelle 8: Hauptsächlich aufgetretene Fehlsegmentierungen außerhalb des segmentierten Organs. Angegeben wurden diese in absteigender Häufigkeit bezogen auf den jeweiligen Ort.

Segmentiertes Organ	Ort der Fehlsegmentierungen
Linke Niere	Darm, Milz
Rechte Niere	Leber, Darm
Pankreas	Leber, Darm, Milz

3.5.2 Quantitative Auswertung

Auch die quantitative Auswertung bestätigte die zu Beginn dieses Kapitels beschriebenen unterschiedlichen Ergebnisse der automatischen Segmentierung der drei Organe durch den maschinellen Lernalgorithmus in Bezug auf ihre erzielte Genauigkeit (Tab.9, Abb.22). Die automatische Segmentierung der beiden Nieren wies mit jeweils 87% eine deutlich höhere Sensitivität auf als die Segmentierung des Pankreas mit 58%. Dementsprechend fiel auch die Falsch-Negativ Rate aus. Die Ähnlichkeitskoeffizienten Dice und Jaccard waren bei der Segmentierung der beiden Nieren höher als bei der Segmentierung des Pankreas (Tab.9). Der Dice-Index betrug für die Segmentierung der linken Niere 0,8943, für die Segmentierung der rechten Niere 0,9059 und für die

Segmentierung des Pankreas 0,6722. Die Werte des Jaccard-Index betragen für die Segmentierung der linken Niere 0,8088, für die Segmentierung der rechten Niere 0,8280 und für die Segmentierung des Pankreas 0,5063.

Tabelle 9: Messgrößen der quantitativen Auswertung. Verglichen wurden die Ergebnisse der automatischen Segmentierung durch den Algorithmus mit denen der manuellen Segmentierung bezogen auf die 20 Testdaten.

Metriken	Einheit	Hintergrund	Linke Niere	Rechte Niere	Pankreas
Sensitivität	%	99,79	87,46	87,37	58,39
Spezifität	%	90,83	99,99	99,99	99,99
Präzision	%	99,80	91,49	94,05	79,20
Falsch-Negativ-Rate	%	0,205	12,54	12,63	41,61
Falsch-Positiv-Rate	%	9,172	0,014	0,014	0,012
Dice-Index		0,9980	0,8943	0,9059	0,6722
Jaccard-Index		0,9960	0,8088	0,8280	0,5063

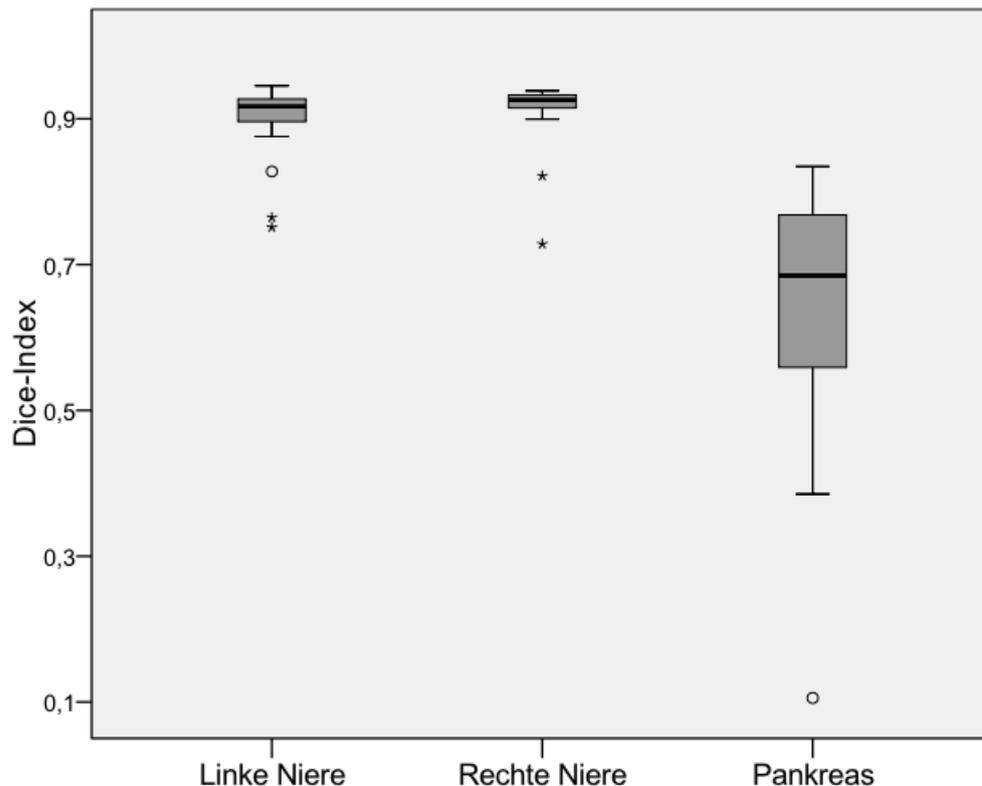


Abbildung 22: Dice-Index der automatischen Segmentierung der 20 Testdatensätze. Graphische Darstellung der Werte des Dice-Index der 20 Testdaten der drei Organe bezogen auf die Ergebnisse der automatischen Segmentierung im Vergleich zu denen der manuellen Segmentierung.

3.5.3 Sekundäre Parameter zur Auswertung

Zusätzlich zur visuell qualitativen und quantitativen Auswertung wurde das Organvolumen und der Fettgehalt der drei Organe zum Vergleich der Ergebnisse der automatischen Segmentierung mit denen der manuellen Segmentierung herangezogen.

Die Auswertung des Volumens und des relativen Fettgehalts ließ unter den drei Organen leichte Unterschiede für manuell und automatisch erstellte Masken erkennen (Tab.10). Insgesamt lässt sich sagen, dass die Analyse inkonsistente Ergebnisse lieferte. Das Organvolumen der rechten Niere zeigte keine signifikanten Unterschiede zwischen manuell und automatisch erstellten Masken. Für die ermittelten Organvolumina der rechten Niere und des Pankreas sowie für den relativen Fettgehalt der drei Organe lagen jeweils signifikante

Abweichungen zwischen manuellen und automatischen Segmentierungsergebnissen vor.

Es zeigte sich, dass bei der Segmentierung des Pankreas sowohl das Organvolumen als auch der Fettgehalt größere Abweichungen zwischen den manuell und automatisch erstellten Masken aufwiesen als bei der Segmentierung der Nieren (Tab.10). Die prozentuale Differenz zwischen manuell und automatisch erstellten Masken war im Durchschnitt wie folgt:

1. Organvolumen: Linke Niere: $7,73\% \pm 9,81\%$; Rechte Niere: $4,68\% \pm 9,26\%$;
Pankreas: $27,10\% \pm 24,70\%$

2. Fettgehalt: Linke Niere: $14,91\% \pm 5,39\%$; Rechte Niere: $10,99\% \pm 5,74\%$;
Pankreas: $24,19\% \pm 10,69\%$

Tabelle 10: Vergleich der Organvolumina und des Fettgehalts. Unterschieden wurde zwischen den Organmasken der manuellen und den Organmasken der automatischen Segmentierung der 20 Testdaten.

	Einheit	manuell	automatisch	P-Wert
Organvolumen der linken Niere	ml	$191,3 \pm 46,7$	$174,1 \pm 35,8$	$p = .016$
Organvolumen der rechten Niere	ml	$179,2 \pm 50,6$	$170,5 \pm 36,8$	$p = .233$
Organvolumen des Pankreas	ml	$88,2 \pm 32,5$	$66,5 \pm 28,7$	$p = .002$
Fettgehalt der linken Niere	%	$18,8 \pm 4,4$	$16,0 \pm 3,4$	$p \leq .001$
Fettgehalt der rechten Niere	%	$18,6 \pm 4,7$	$16,4 \pm 3,9$	$p \leq .001$
Fettgehalt des Pankreas	%	$19,8 \pm 8,2$	$14,8 \pm 6,1$	$p \leq .001$

In den folgenden Abbildungen wird sowohl das Organvolumen (Abb.23 bis Abb.25) als auch der Fettgehalt (Abb.26 bis Abb.28), welcher als Mittelwert der jeweiligen Organmaske berechnet wurde, zum graphischen Vergleich getrennt nach den drei Organen jeweils als Boxplot dargestellt.

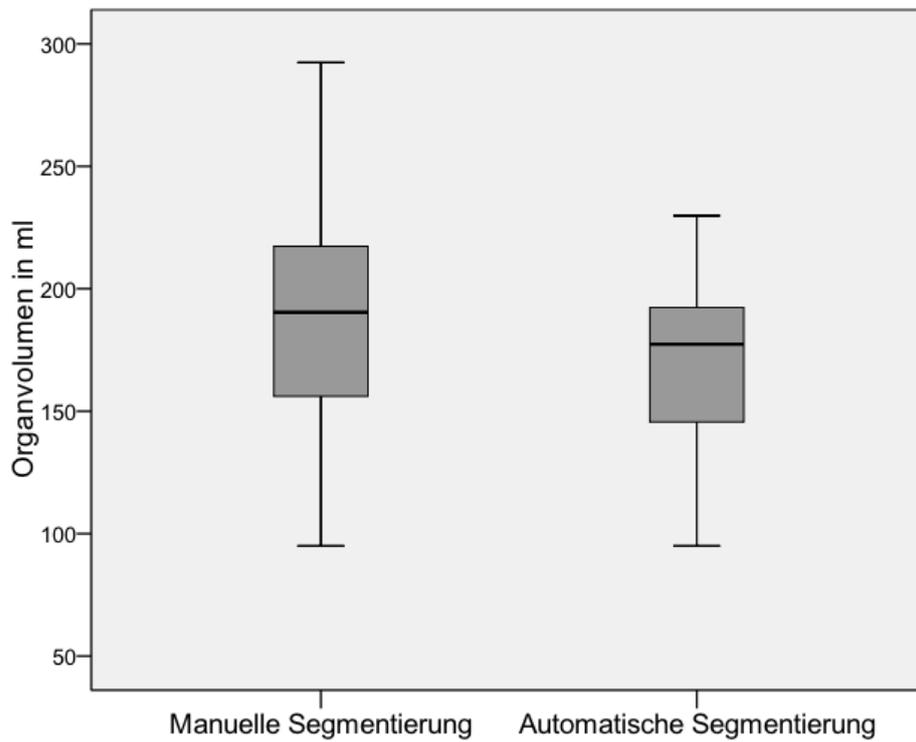


Abbildung 23: Linke Niere: Vergleich der Organvolumina zwischen der manuellen und der automatischen Segmentierung. Die Auswertung bezog sich auf die 20 Testdaten.

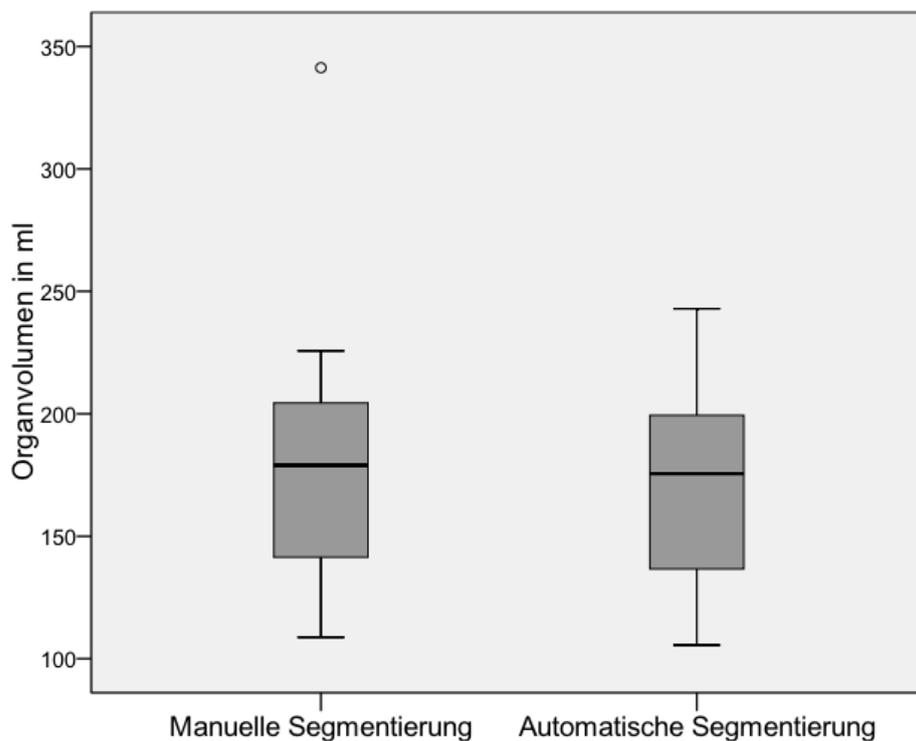


Abbildung 24: Rechte Niere: Vergleich der Organvolumina zwischen der manuellen und der automatischen Segmentierung. Die Auswertung bezog sich auf die 20 Testdaten.

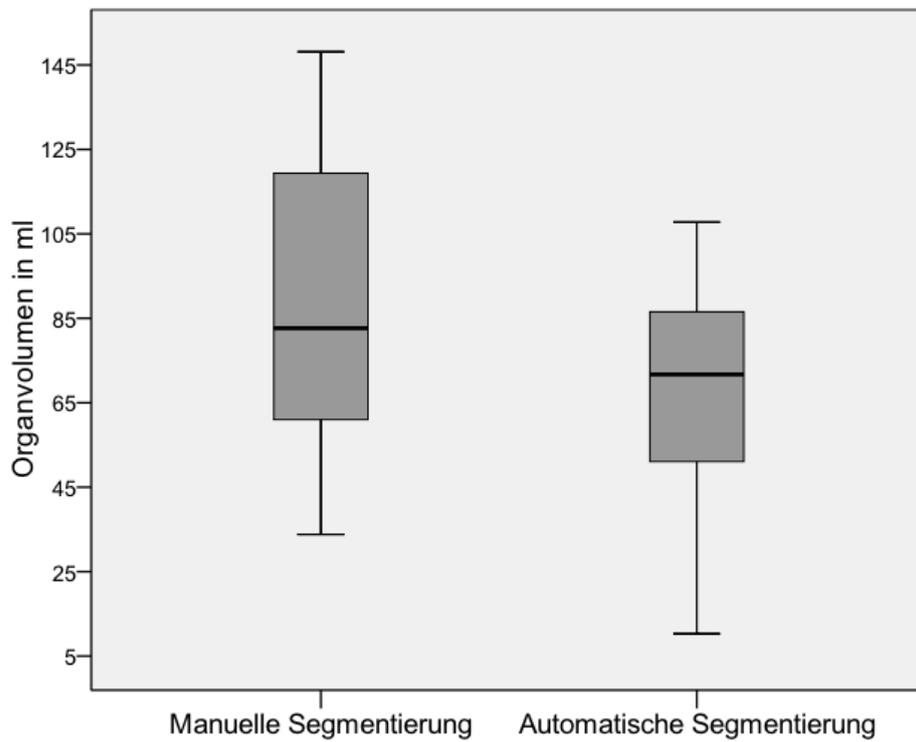


Abbildung 25: Pankreas: Vergleich der Organvolumina zwischen der manuellen und der automatischen Segmentierung. Die Auswertung bezog sich auf die 20 Testdaten.

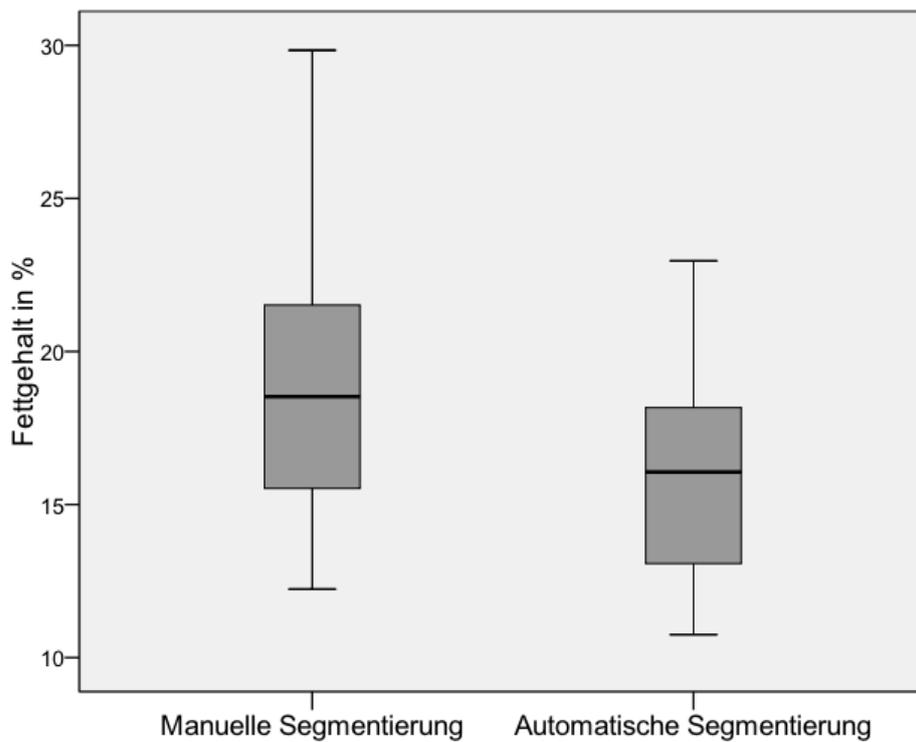


Abbildung 26: Linke Niere: Vergleich des Fettgehalts zwischen der manuellen und der automatischen Segmentierung. Die Auswertung bezog sich auf die 20 Testdaten.

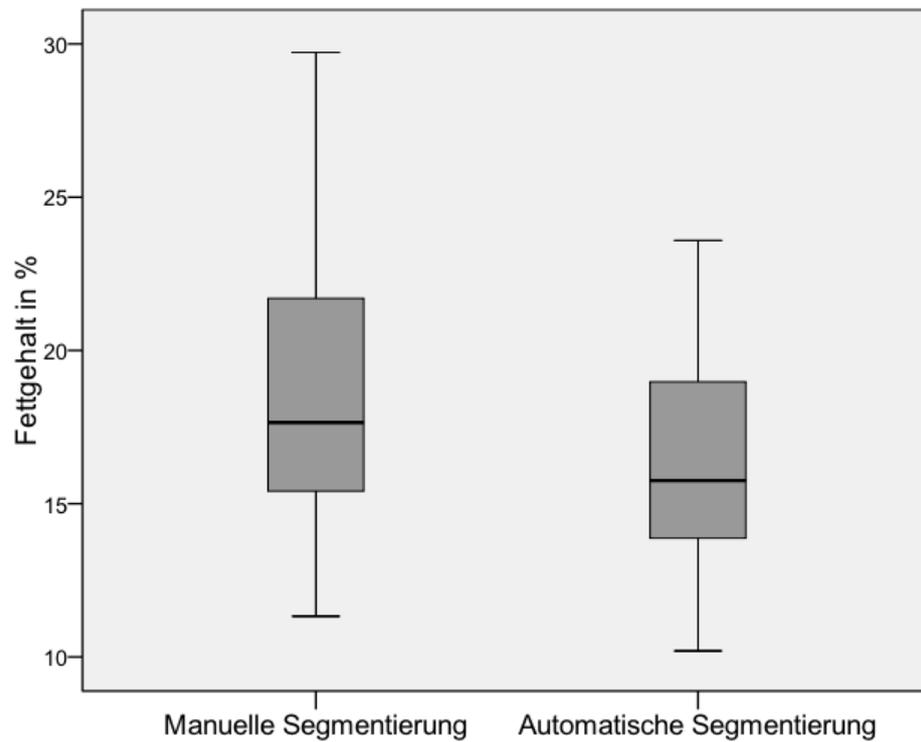


Abbildung 27: Rechte Niere: Vergleich des Fettgehalts zwischen der manuellen und der automatischen Segmentierung. Diese Auswertung bezog sich auf die 20 Testdaten.

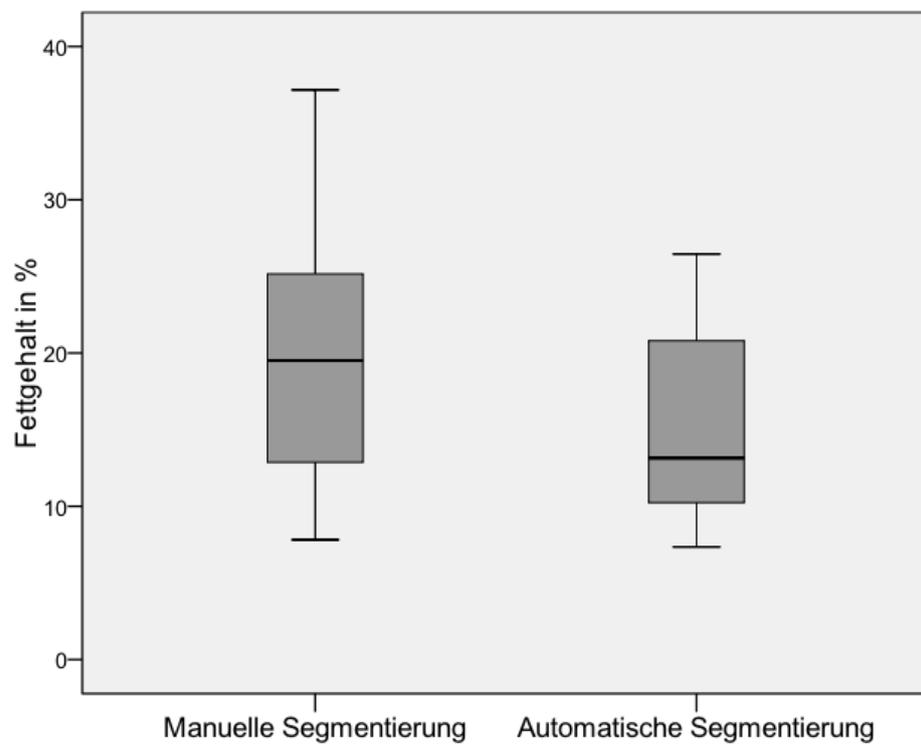


Abbildung 28: Pankreas: Vergleich des Fettgehalts zwischen der manuellen und der automatischen Segmentierung. Diese Auswertung bezog sich auf die 20 Testdaten.

Für die Organvolumina der drei Organe ließ sich ein eindeutiger Zusammenhang zwischen manuell und automatisch erstellten Masken erkennen (Tab.11, Abb.29 bis Abb.31). Die untersuchte Korrelation nach Pearson zeigte einen signifikanten Zusammenhang (Tab.11) und war für das Organvolumen der Nieren zwischen manuell und automatisch erstellten Masken höher (Tab.11, Abb.29 bis Abb.31).

Tabelle 11: Korrelation des Organvolumens zwischen manuell und automatisch erstellten Masken. Untersucht wurde die Korrelation nach Pearson. Die Auswertung bezog sich auf die 20 Testdaten.

		Linke Niere: Manuell Erstellte Masken	Rechte Niere: Manuell Erstellte Masken	Pankreas: Manuell Erstellte Masken
Linke Niere: Automatisch erstellte Masken	Korrelation nach Pearson	r = 0,780		
	P-Wert	p ≤.001		
Rechte Niere: Automatisch erstellte Masken	Korrelation nach Pearson		r = 0,779	
	P-Wert		p ≤.001	
Pankreas: Automatisch erstellte Masken	Korrelation nach Pearson			r = 0,626
	P-Wert			p =.003

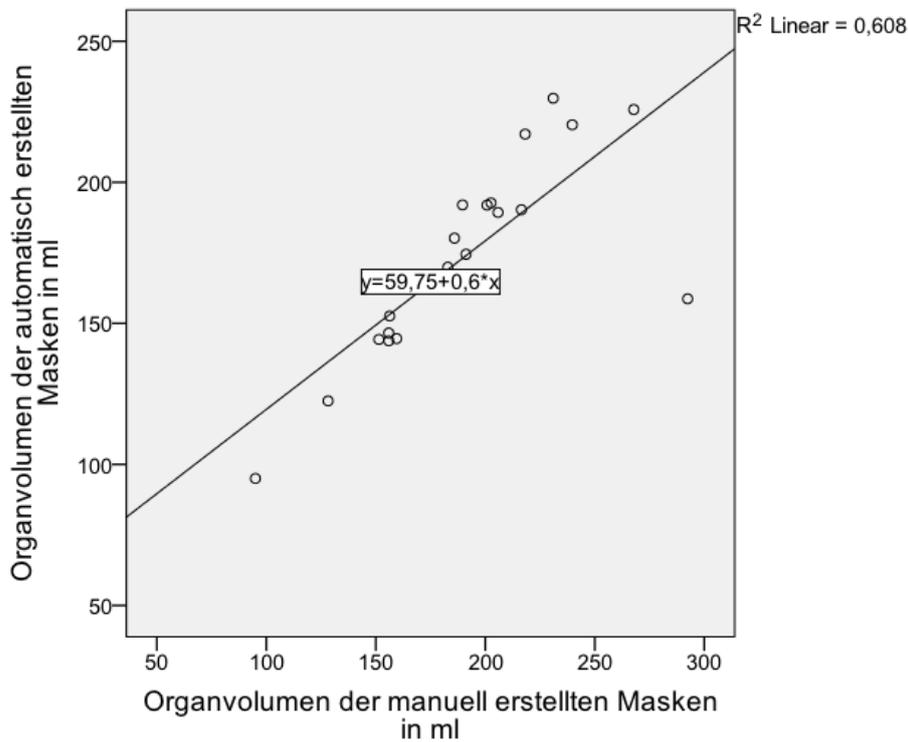


Abbildung 29: Linke Niere: Korrelation der Organvolumina zwischen den manuell und automatisch erstellten Masken. Die Auswertung bezog sich auf die 20 Testdaten.

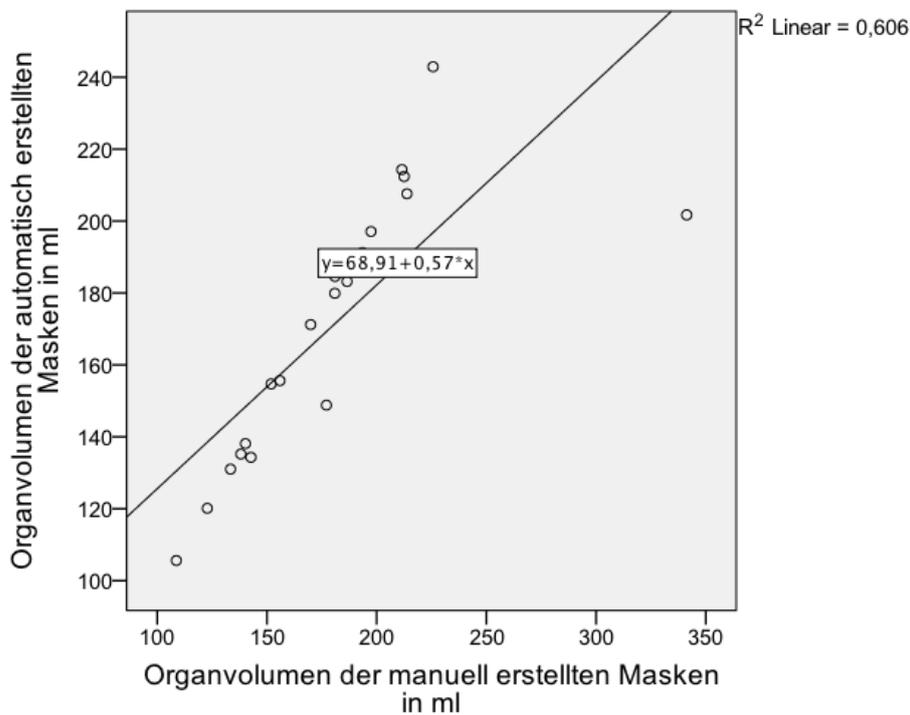


Abbildung 30: Rechte Niere: Korrelation der Organvolumina zwischen den manuell und automatisch erstellten Masken. Die Auswertung bezog sich auf die 20 Testdaten.

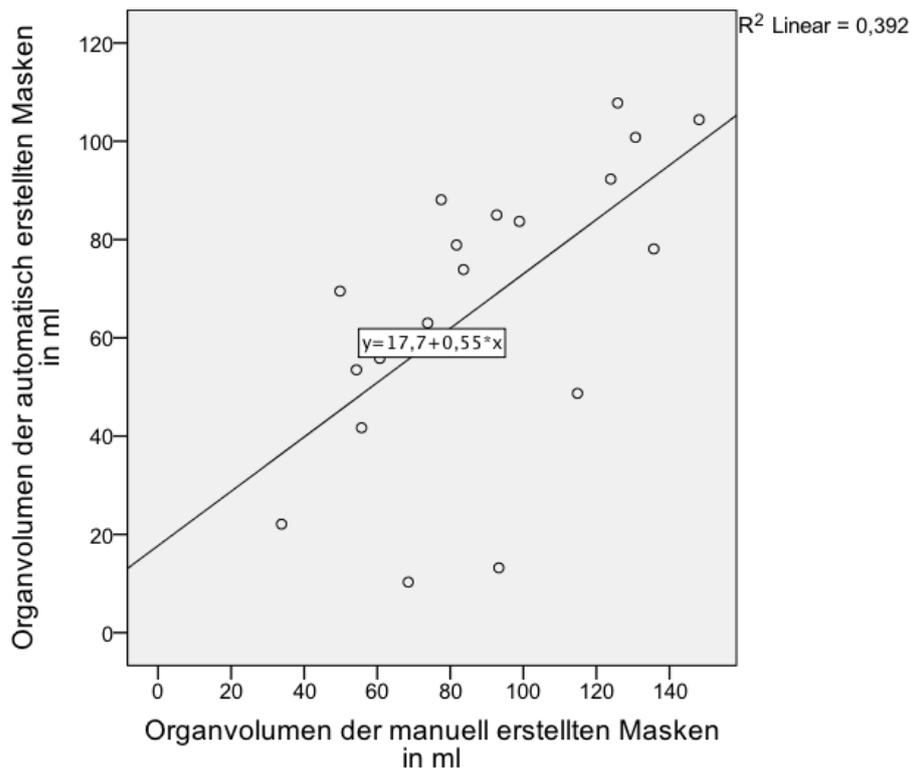


Abbildung 31: Pankreas: Korrelation der Organvolumina zwischen den manuell und automatisch erstellten Masken. Die Auswertung bezog sich auf die 20 Testdaten.

4 Diskussion

In dieser Arbeit wurde untersucht, inwieweit ein maschineller Lernalgorithmus basierend auf CNNs, welcher zunächst implementiert und dann validiert wurde, in 3.0-Tesla MRT-Ganzkörperaufnahmen der NAKO-Studie in der T1-Wichtung die Nieren und das Pankreas automatisiert segmentieren kann.

Insgesamt konnte die technische Umsetzbarkeit der automatischen Organsegmentierung für das untersuchte Kollektiv erfolgreich gezeigt werden. Dabei wurden die Nieren und das Pankreas durch den maschinellen Lernalgorithmus automatisiert erkannt, lokalisiert und markiert. Die erzielte Genauigkeit war hierbei für die automatische Segmentierung der Nieren deutlich höher als für die automatische Segmentierung des Pankreas.

4.1 Diskussion und Ausblick

Die Ergebnisse dieser Arbeit stimmen im Wesentlichen mit denen in der Literatur angegebenen Ergebnissen anderer Arbeiten überein, die ebenfalls eine automatische Segmentierung der Nieren beziehungsweise des Pankreas in MRT- oder CT-Aufnahmen durch die Anwendung von „Deep Learning“ durchführten [(Zhao et al., 2019); (Fu et al., 2018a); (Fu et al., 2018b); (Haghighi et al., 2018); (Jackson et al., 2018); (Hu et al., 2017); (Cai et al., 2016); (Roth et al., 2015)]. Relativierend muss aber die eingeschränkte Vergleichbarkeit von Methoden auf unterschiedlichen Datensätzen beachtet werden.

Der Vergleich der Ergebnisse der automatischen Segmentierung mit denen der manuellen Segmentierung zeigte, dass gerade die automatische Segmentierung der Nieren mit einem mittleren Dice-Koeffizienten von etwa 90% erfolgreich durchgeführt werden konnte. Es hat sich ebenfalls gezeigt, dass dieses Ergebnis innerhalb des untersuchten Kollektivs zuverlässig war und nur eine geringe Variabilität unter den untersuchten Testdatensätzen vorlag. Die in der vorliegenden Arbeit erzielten Ergebnisse sind somit mit in der Literatur angegebenen Dice-Koeffizienten zwischen 86% und 96% für die automatische Segmentierung der Nieren in MRT- beziehungsweise CT-Aufnahmen mittels „Deep Learning“ vergleichbar [(Fu et al., 2018b); (Haghighi et al., 2018); (Jackson et al., 2018); (Hu et al., 2017); (Roth et al., 2015)]. Fu et al. (2018) beispielsweise

untersuchten ebenfalls die automatische Segmentierung verschiedener Organe des Abdomens durch die Anwendung von CNNs an MRT-Aufnahmen und erreichten für die Nieren einen Dice-Koeffizienten von 92%, mit Einbeziehung eines Korrekturnetzwerkes sogar von 93,1%. Hu et al., (2018) hingegen wendeten die automatische Segmentierung mithilfe von CNNs in CT-Aufnahmen unter anderem für die Nieren an. Der erzielte Dice-Koeffizient lag bei 95,4%.

Aufgrund der sehr starken anatomischen Variabilität des Pankreas und der im Folgenden beschriebenen Herausforderungen bei der manuellen Segmentierung dieses Organs, war es nicht möglich, ähnlich gute Ergebnisse wie bei der automatischen Segmentierung der Nieren zu erreichen. Die starke interindividuelle Variabilität des Pankreas kann auch erklären, weshalb es - neben der im Durchschnitt schlechteren automatischen Segmentierung - auch zu einer stärkeren Streuung der Ergebnisse kam und somit die erzielte Genauigkeit der automatischen Segmentierung des Pankreas individuell abhängig vom Datensatz war. Durch die starke interindividuelle Variabilität von Größe, Form und Lage des Pankreas sowie durch die anatomische Nähe zu angrenzenden Organen [(Cai et al., 2016); (Roth et al., 2015)] wird die Abgrenzung des Pankreas zur Umgebung erschwert und es kann zu Verwechslungen mit Darmgewebe, Gefäßen, Bauchfett oder anderen Gewebetypen kommen (Cai et al., 2016). Dies erklärt auch den deutlich höheren Zeitaufwand der manuellen Segmentierung dieses Organs gegenüber den Nieren. In der Literatur finden sich für die automatische Segmentierung des Pankreas bereits optimierte organspezifische Ansätze, bei denen das Pankreas in höherer Präzision segmentiert werden konnte [(Zhao et al., 2019); (Fu et al., 2018a); (Cai et al., 2016); (Roth et al., 2015)]. Der Dice-Koeffizient lag bei den erwähnten Arbeiten zwischen 71,8% und 85,99%. Hierbei handelte es sich neben MRT-Aufnahmen vor allem um CT-Bilder. Ein Vorteil von CT- gegenüber MRT-Aufnahmen ist, dass diese unter anderem weniger Artefakte beispielsweise durch Bewegung und zusätzlich absolute Grauwerte anhand definierter Hounsfield-Einheiten aufweisen, was zu einer insgesamt homogeneren Bildaufnahme führt. Dies kann sich insbesondere bei der Organsegmentierung von komplexen und individuell

sehr variablen Organen wie des Pankreas als vorteilhaft erweisen und somit die Genauigkeit der Segmentierung erhöhen.

Die Ergebnisse der vorliegenden Arbeit sind für die Auswertung der MRT-Aufnahmen in der NAKO Gesundheitsstudie von Bedeutung. Es wird die Möglichkeit aufgezeigt, dass eine automatisierte Auswertung der Aufnahmen im Hinblick auf bestimmten Fragestellungen, welche sich noch spezifizieren und erweitern ließen, möglich ist. Dies kann, neben dem daraus resultierenden großen Zeitgewinn gegenüber der manuellen Auswertung, neue Erkenntnisse über Krankheiten liefern, was eines der Ziele der NAKO-Studie ist.

Für die vorliegende Arbeit war die Verwendung von MRT-Aufnahmen der NAKO Gesundheitsstudie vorteilhaft, da diese standardisiert laut des Studienprotokolls aufgenommen wurden und eine ausreichende Menge an Bilddaten zur Verfügung standen. Zusätzlich ist davon auszugehen, dass es sich überwiegend um eine Auswahl eines Normalkollektivs handelte.

Die vorliegende Arbeit hat Limitationen. Zum einen war die manuelle Segmentierung der Trainingsdaten zeitaufwendig. Um in Zukunft den Zeitaufwand der manuellen Annotierung verringern zu können, finden sich bereits verschiedene Ansätze in der Literatur. Dazu zählen unter anderem das Transferlernen und „domain adaptation“, bei denen angenommene Grundwahrheiten auf ähnlichen Trainingsdaten verwendet beziehungsweise vortrainierte Netzwerke übertragen werden [(Falk et al., 2019); (Knoll et al., 2019); (Küstner et al., 2019); (Chartrand et al., 2017); (Shin et al., 2016); (Xie et al., 2016)]. Ein weiteres Beispiel ist die semiautomatische Segmentierung, bei der eine Segmentierung anhand vortrainierter Daten automatisch durchgeführt wird und diese vorgeschlagene Segmentierung anschließend manuell verbessert werden kann. Dies ist beispielsweise mithilfe der Software NVIDIA Clara möglich (NVIDIA Corporation, 2019).

Die Inferenz der automatischen Segmentierung für die Testdaten der Organe im MRT-Bild war insgesamt mit einer Dauer von 5-10 Minuten pro Datensatz sehr

zeitaufwendig. Dies kann bei der Anwendung auf sehr große Datensätze mit mehreren tausend Probanden oder auch beim Einsatz in der klinischen Routine eine Limitation sein. Hierbei kann der Zeitaufwand in Zukunft durch weitere Optimierung von Hyperparametern und des Modells sowie durch Erweiterung der Hardware verringert werden.

Zum anderen handelte es sich mit einem Umfang von 80 Trainingsdatensätzen um eine verhältnismäßig geringe Anzahl an Trainingsdaten für Methoden des „Deep-Learning“. Da es sich bei der vorliegenden Arbeit zunächst um eine Machbarkeitsstudie handelte, wurde diese Anzahl an Trainingsdaten sowie auch die Anzahl der 20 Testdatensätze bewusst gewählt und könnte in weiteren Untersuchungen erweitert werden. Gleichzeitig erfolgte die Anwendung auf einem sehr konkreten Datensatz und nur für ausgewählte Organe. Hier könnte weiterführend untersucht werden, inwieweit sich die Ergebnisse sowohl für andere Datensätze als auch für andere Organe verallgemeinern lassen.

Des Weiteren zeigte sich, dass bei der automatischen Segmentierung der Organe kleinere fehlsegmentierte Gebiete, die sich räumlich nicht dem jeweiligen Organ zuordnen ließen, segmentiert wurden. CNNs segmentieren die jeweiligen Organe nicht in ihrer Gesamtheit, sondern beurteilen jedes Pixel (2D) beziehungsweise Voxel (3D) des gesamten Bildes für sich eigenständig und führen es einer der Klassenbeschriftungen zu (Fu et al., 2018b). Daraus entsteht dann die präsentierte Segmentierung. Die pixelweise Segmentierung erklärt auch, warum es teilweise zu Fehlsegmentierungen außerhalb des eigentlich segmentierten Organs gekommen ist. Solche Fehler lassen sich aber verhältnismäßig leicht durch anschließende Korrekturmechanismen infolge der Kodierung von Positioneingaben oder durch die Auswahl der größten Zusammenhangskomponente der vorhergesagten Organmaske beheben [(Küstner et al., 2019); (Fu et al., 2018b)].

Die Vergleichbarkeit der sekundären Messgrößen wie Organvolumen und Fettgehalt zwischen den automatisch und manuell erstellten Organmasken zeigte, dass diese trotz kleinerer Segmentierungsfehler des eigentlichen Organs noch relativ gut extrahiert werden konnten. Bei der automatischen Segmentierung von beispielsweise Organen ist in der Regel nicht die eigentliche

Organsegmentierung das Ziel, sondern die darauffolgende Extraktion von Biomarkern wie Volumen, Form oder Fettgehalt.

Künstliche Neuronale Netzwerke sind für vielfältige Anwendungen auch außerhalb der Medizin sehr gut geeignet. So sind viele der in den letzten Jahren gemachten Fortschritte in autonomen Vorgängen wie beispielsweise bei der Sprach- und Bilderkennung, der Übersetzung von Texten oder auch bei selbstfahrenden Fahrzeugen auf die Leistungsfähigkeit von Neuronalen Netzwerken zurückzuführen (The, 2018).

Bei der automatischen Segmentierung von Organen in medizinischen Bildaufnahmen haben sich Convolutional Neural Networks als besonders geeignet erwiesen und die Leistungsfähigkeit konventioneller Segmentierungsansätze wie beispielsweise der Multi-Atlas-Segmentierung oder von Random Forests in vielen Bereichen übertroffen [(Furat et al., 2019); (Bobo et al., 2018); (Hu et al., 2017); (Lavdas et al., 2017); (Litjens et al., 2017); (Çiçek et al., 2016); (Milletari et al., 2016)]. Durch das Training lernt das CNN an den Trainingsdaten selbstständig geeignete Muster zu filtern und entsprechend zu verarbeiten, um aus dem Eingangsbild eine passende Organmaske zu generieren (Soomro et al., 2019). Auf diese Weise kann die Übertragbarkeit und Generalisierbarkeit der erlernten Bildcharakteristika ermöglicht werden [(Soomro et al., 2019); (Falk et al., 2019)]. Wird jedoch durch die Trainingsdaten das Zielkollektiv nur unzureichend abgebildet beziehungsweise sind die manuell erstellen Trainingsdaten fehlerbehaftet oder in sich widersprüchlich, gelingt die Verallgemeinerung des Modells meist nur unzureichend [(Falk et al., 2019); (Selvikvåg Lundervold and Lundervold, 2018)]. Sowohl die gegenüber anderen „Deep Learning“-Methoden zum Training des Netzwerkes verhältnismäßig geringe Anzahl an notwendigen Trainingsdaten als auch die guten Ergebnisse bei der automatischen Segmentierung von Bilddaten - sowohl von medizinischen als auch von nichtmedizinischen - sind weitere Vorteile von CNNs [(Falk et al., 2019); (Furat et al., 2019); (Guo et al., 2018); (Chartrand et al., 2017); (Bojarski et al., 2016); (Li et al., 2014); (Krizhevsky et al., 2012)]. UNet basierte CNN-Architekturen [(Falk et al., 2019); (Ronneberger et al., 2015)] konnten in

vorangehenden Studien sehr gute Ergebnisse bei der automatischen Segmentierung von Geweben in medizinischen Bildaufnahmen erreichen [(Furat et al., 2019); (Küstner et al., 2018c); (Norman et al., 2018); (Çiçek et al., 2016)]. Ebenfalls ist bei verschiedenen Aufgabenstellungen die dreidimensionale automatische Segmentierung für die meisten Datensätze vorteilhaft und der zweidimensionalen Segmentierung tendenziell überlegen [(Furat et al., 2019); (Küstner et al., 2018b)].

In verschiedenen Fachbereichen der Medizin wird bereits an klinischen Anwendungsmöglichkeiten von Convolutional Neural Networks geforscht. So kann beispielweise mit CNNs die bildgestützte Diagnostik von Alzheimer-Erkrankungen [(Ding et al., 2019); (Duc et al., 2019)] und Retinopathien [(De Fauw et al., 2018); (Zhou et al., 2018); (Shuang et al., 2017)] unterstützt werden, anhand von Röntgen-Thorax-Aufnahmen ein Risiko-Score zur Langzeitmortalität von Patienten bestimmt werden (Lu et al., 2019) oder auch die Malignität von Hautläsionen beurteilt werden [(Haenssle et al., 2018); (Esteva et al., 2017)].

Die Möglichkeiten, die durch die Methoden des „Deep Learning“ entstehen, können Ärzte und Ärztinnen bei ihrer Arbeit unterstützen und durch Reduktion repetitiver Aufgaben dazu beitragen, dass mehr Zeit für die Patientenbetreuung zur Verfügung steht. Für ausgewählte medizinische Anwendungen, wie beispielsweise die visuelle Klassifikation von malignitätssuspekten Hautbefunden, lassen sich bereits jetzt Genauigkeiten feststellen, die an die menschliche Diagnostik durch einen Arzt oder eine Ärztin heranreichen können [(Haenssle et al., 2018); (Esteva et al., 2017)]. „Deep Learning“ soll hierbei in Zukunft als Ergänzung sowie Unterstützung in der medizinischen Diagnostik eingesetzt werden und kann vor allem auch zur Analyse von Bildaufnahmen und anderen erhobenen Parametern in Studien mit einer großen Anzahl an Daten vorteilhaft sein. Die abschließende Beurteilung sowie Interpretation der Ergebnisse im klinischen Kontext bleibt dabei aber Aufgabe des Arztes oder der Ärztin.

Da es immer wieder zu Fehlern bei der Anwendung von „Deep Learning“ kommen kann, ist es wichtig zu verstehen, welche Auswirkungen diese haben

können (Guo et al., 2018). So sollte insbesondere betrachtet werden, welche Konsequenzen möglicherweise fehlerhafte Ergebnisse durch die Anwendung von maschinellen Lernalgorithmen im klinischen Alltag haben können, wer dafür die Verantwortung tragen wird und welche ethischen Fragen dies aufwerfen kann (The, 2018).

4.2 Schlussfolgerung

Es lässt sich zusammenfassen, dass eine automatische Gewebesegmentierung abhängig vom jeweiligen Organ in Bildaufnahmen in hoher Präzision möglich ist und somit insbesondere als unterstützende Maßnahme in groß angelegten epidemiologischen Studien wie der NAKO Gesundheitsstudie genutzt werden kann (Küstner et al., 2018c). Dies ist wichtig, da die manuelle Auswertung extrem aufwendig ist und daher bei großen Datenmengen nur schwer bis gar nicht realisierbar wäre. Die Gewebesegmentierung ist ein grundlegender Schritt in der Bildanalyse (Küstner et al., 2018a). Auf Grundlage der Gewebesegmentierung ist es möglich, auch weiterführende Analysen der jeweiligen Bilddaten automatisiert durchführen zu lassen [(Küstner et al., 2018c); (Farag et al., 2014)]. Ein Beispiel ist die volumetrische Segmentierung von Tumoren anhand einzelner Tumormerkmale [(Soomro et al., 2019); (Parmar et al., 2014)] oder auch die Erkennung von bildgebenden Biomarkern im Allgemeinen - unter anderem in Studien wie der NAKO - zur Identifikation von möglichen Risikofaktoren beziehungsweise Pathologien der jeweiligen Organe (Selvikvåg Lundervold et al., 2017). Beginnend von der Gewebesegmentierung können so weiterführende automatische Auswertungspipelines entwickelt werden, die neben dem Ziel einer Diagnose von pathologischen Prozessen auch die Auswertung von einer Vielzahl von Bildaufnahmen in Studien unterstützen können.

5 Zusammenfassung

Die NAKO Gesundheitsstudie ist eine zurzeit in Deutschland durchgeführte epidemiologische Kohortenstudie. Ziel dieser Studie ist es, verschiedene Daten über die allgemeine Gesundheit zu erhalten und Ursachen von Krankheiten zu verstehen. Die Daten, die im Zusammenhang mit solch groß angelegten Studien entstehen, gilt es auszuwerten und zu analysieren. Die Auswertung von Bilddaten, wie beispielsweise den bei der NAKO-Studie erhobenen MRT-Aufnahmen, stellt eine neue Herausforderung gegenüber der Auswertung von sonstigen Parametern wie beispielsweise von Blutwerten dar. Neben der manuellen Auswertung von Bildaufnahmen kann es insbesondere bei einer großen Menge an Daten hilfreich sein, Methoden aus dem Bereich des maschinellen Lernens zur Datenanalyse anzuwenden. Bewährt haben sich hierbei die Methoden des „Deep Learning“ in künstlichen neuronalen Netzwerken, einer Unterform des maschinellen Lernens. Ein grundlegender Schritt in der automatischen Bildanalyse ist die Gewebesegmentation. Verschiedene Ansätze finden sich hierfür in der Literatur. Convolutional Neural Networks (CNNs) haben sich bei der automatischen Segmentierung von Organen in medizinischen Bildaufnahmen als besonders geeignet erwiesen.

In der vorliegenden Arbeit wurde untersucht, inwieweit ein maschineller Lernalgorithmus in MRT-Ganzkörperaufnahmen die Nieren und das Pankreas automatisiert erkennen und markieren kann. Es wurden 3.0-Tesla MRT-Ganzkörperaufnahmen der T1-Wichtung verwendet, die im Rahmen der NAKO-Studie entstanden sind. Der automatische Lernvorgang basierte auf CNNs.

Zunächst wurden 100 MRT-Datensätze der NAKO-Studie aus den Jahren 2015 und 2016 zufällig ausgewählt. Zur Implementierung und Validierung des maschinellen Lernalgorithmus waren anschließend verschiedene Zwischenschritte notwendig. Daraus ergab sich folgendes Vorgehen:

- i. Manuelle Segmentierung der 100 MRT-Aufnahmen mit der Software Medical Imaging Interaction Toolkit (MITK).
- ii. Zufällige Aufteilung der Datensätze in 80 Trainings- und 20 Testdaten.

- iii. Durchführung von zwei Trainingsvorgängen mit den 80 Trainingsdaten:
 - Training bei gleichzeitiger Validierung innerhalb der 80 Trainingsdaten (80% Trainings-, 20% Validierungsdaten) zur Bestimmung der optimalen Epochenanzahl bei 5-facher Kreuzvalidierung.
 - Training des endgültigen Modells mit der optimalen Anzahl an Epochen mit allen 80 Trainingsdaten.
- iv. Evaluation des trainierten Modells anhand der 20 Testdaten durch den Vergleich der Ergebnisse der automatischen Segmentierung mit denen der manuellen Segmentierung.

Die technische Umsetzbarkeit der automatischen Organsegmentierung konnte für das untersuchte Kollektiv erfolgreich gezeigt werden. Die Nieren und das Pankreas wurden dabei durch den maschinellen Lernalgorithmus selbstständig erkannt und markiert. Die erzielte Genauigkeit war für die Nieren (Dice: ~ 90%) deutlich höher als für das Pankreas (Dice: ~ 67%). Dieses Ergebnis zeigte sich in allen drei untersuchten Aspekten der Auswertung (visuell qualitativ, quantitativ, Sekundärparameter).

Die starke interindividuelle Variabilität des Pankreas kann erklären, weshalb die erzielte Genauigkeit der automatischen Segmentierung dieses Organs geringer war. Hier gilt es, optimierte organspezifische Ansätze weiterzuentwickeln, die die Besonderheiten wie eines Organs des Pankreas bei der automatischen Segmentierung berücksichtigen.

Zusammenfassend lässt sich sagen, dass eine automatische Gewebesegmentierung abhängig vom jeweiligen Organ in Bildaufnahmen möglich ist. In groß angelegten epidemiologischen Studien wie der NAKO-Studie kann diese somit unterstützend in der Auswertung genutzt werden. Auf Grundlage der Organsegmentierung lassen sich weiterführende Algorithmen entwickeln, die sowohl als ergänzende Maßnahmen in der Diagnostik von pathologischen Prozessen als auch bei Studien beispielsweise zur Identifikation von bildgebenden Biomarkern und möglichen Risikofaktoren eingesetzt werden können.

6 Literaturverzeichnis

- AHRENS, W. & JOCKEL, K. H. 2015. [The benefit of large-scale cohort studies for health research: the example of the German National Cohort]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 58, 813-821.
- BAMBERG, F., KAUCZOR, H. U., WECKBACH, S., SCHLETT, C. L., FORSTING, M., LADD, S. C., GREISER, K. H., WEBER, M. A., SCHULZ-MENGER, J., NIENDORF, T., PISCHON, T., CASPERS, S., AMUNTS, K., BERGER, K., BULOW, R., HOSTEN, N., HEGENSCHIED, K., KRONCKE, T., LINSEISEN, J., GUNTHER, M., HIRSCH, J. G., KOHN, A., HENDEL, T., WICHMANN, H. E., SCHMIDT, B., JOCKEL, K. H., HOFFMANN, W., KAKS, R., REISER, M. F., VOLZKE, H. & GERMAN NATIONAL COHORT, M. R. I. S. I. 2015. Whole-Body MR Imaging in the German National Cohort: Rationale, Design, and Technical Background. *Radiology*, 277, 206-20.
- BELBASIS, L. & BELLOU, V. 2018. Introduction to Epidemiological Studies. *Methods Mol Biol*, 1793, 1-6.
- BOBO, M. F., BAO, S., HUO, Y., YAO, Y., VIROSTKO, J., PLASSARD, A. J., LYU, I., ASSAD, A., ABRAMSON, R. G., HILMES, M. A. & LANDMAN, B. A. 2018. Fully Convolutional Neural Networks Improve Abdominal Organ Segmentation. *Proc SPIE Int Soc Opt Eng*, 10574.
- BOJARSKI, M., DEL TESTA, D., DWORAKOWSKI, D., FIRNER, B., FLEPP, B., GOYAL, P., JACKEL, L. D., MONFORT, M., MULLER, U., ZHANG, J., ZHANG, X., ZHAO, J. & ZIEBA, K. 2016. End to End Learning for Self-Driving Cars. *arXiv e-prints* [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv160407316B> [Accessed April 01, 2016].
- CAI, J., LU, L., ZHANG, Z., XING, F., YANG, L. & YIN, Q. 2016. Pancreas Segmentation in MRI using Graph-Based Decision Fusion on Convolutional Neural Networks. *Med Image Comput Comput Assist Interv*, 9901, 442-450.
- CHARTRAND, G., CHENG, P. M., VORONTSOV, E., DROZDZAL, M., TURCOTTE, S., PAL, C. J., KADOURY, S. & TANG, A. 2017. Deep Learning: A Primer for Radiologists. *Radiographics*, 37, 2113-2131.
- ÇIÇEK, Ö., ABDULKADIR, A., LIENKAMP, S. S., BROX, T. & RONNEBERGER, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. 2016 Cham. Springer International Publishing, 424-432.
- DE FAUW, J., LEDSAM, J. R., ROMERA-PAREDES, B., NIKOLOV, S., TOMASEV, N., BLACKWELL, S., ASKHAM, H., GLOROT, X., O'DONOGHUE, B., VISENTIN, D., VAN DEN DRIESSCHE, G., LAKSHMINARAYANAN, B., MEYER, C., MACKINDER, F., BOUTON, S., AYOUB, K., CHOPRA, R., KING, D., KARTHIKESALINGAM, A., HUGHES, C. O., RAINE, R., HUGHES, J., SIM, D. A., EGAN, C., TUFAIL, A., MONTGOMERY, H., HASSABIS, D., REES, G., BACK, T., KHAW, P. T., SULEYMAN, M., CORNEBISE, J., KEANE, P. A. &

- RONNEBERGER, O. 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*, 24, 1342-1350.
- DING, Y., SOHN, J. H., KAWCZYNSKI, M. G., TRIVEDI, H., HARNISH, R., JENKINS, N. W., LITUIEV, D., COPELAND, T. P., ABOIAN, M. S., MARI APARICI, C., BEHR, S. C., FLAVELL, R. R., HUANG, S. Y., ZALOCUSKY, K. A., NARDO, L., SEO, Y., HAWKINS, R. A., HERNANDEZ PAMPALONI, M., HADLEY, D. & FRANC, B. L. 2019. A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using (18)F-FDG PET of the Brain. *Radiology*, 290, 456-464.
- DIXON, W. T. 1984. Simple proton spectroscopic imaging. *Radiology*, 153, 189-94.
- DÖBEL, I., LEIS, M., MOLINA VOGELSANG, M., WELZ, J., NEUSTROEV, D., PETZKA, H., RIEMER, A., PÜPING, S., VOSS, A. & WEGELE, M. 2018. *Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung*. [Online]. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. Available: https://www.bigdata.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer_Studie_ML_201809.pdf [Accessed 08.04.2019].
- DOU, Q., YU, L., CHEN, H., JIN, Y., YANG, X., QIN, J. & HENG, P. A. 2017. 3D deeply supervised network for automated segmentation of volumetric medical images. *Med Image Anal*, 41, 40-54.
- DUC, N. T., RYU, S., QURESHI, M. N. I., CHOI, M., LEE, K. H. & LEE, B. 2019. 3D-Deep Learning Based Automatic Diagnosis of Alzheimer's Disease with Joint MMSE Prediction Using Resting-State fMRI. *Neuroinformatics*.
- ESTEVA, A., KUPREL, B., NOVOA, R. A., KO, J., SWETTER, S. M., BLAU, H. M. & THRUN, S. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115-118.
- FALK, T., MAI, D., BENSCH, R., CICEK, O., ABDULKADIR, A., MARRAKCHI, Y., BOHM, A., DEUBNER, J., JACKEL, Z., SEIWALD, K., DOVZHENKO, A., TIETZ, O., DAL BOSCO, C., WALSH, S., SALTUKOGLU, D., TAY, T. L., PRINZ, M., PALME, K., SIMONS, M., DIESTER, I., BROX, T. & RONNEBERGER, O. 2019. U-Net: deep learning for cell counting, detection, and morphometry. *Nat Methods*, 16, 67-70.
- FARAG, A., LU, L., TURKBEY, E., LIU, J. & SUMMERS, R. M. 2014. A Bottom-Up Approach for Automatic Pancreas Segmentation in Abdominal CT Scans. *arXiv e-prints* [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2014arXiv1407.8497F> [Accessed July 01, 2014].
- FU, M., WU, W., HONG, X., LIU, Q., JIANG, J., OU, Y., ZHAO, Y. & GONG, X. 2018a. Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images. *BMC Syst Biol*, 12, 56.
- FU, Y., MAZUR, T. R., WU, X., LIU, S., CHANG, X., LU, Y., LI, H. H., KIM, H., ROACH, M. C., HENKE, L. & YANG, D. 2018b. A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. *Med Phys*, 45, 5129-5137.
- FURAT, O., WANG, M., NEUMANN, M., PETRICH, L., WEBER, M., KRILL, C. E. & SCHMIDT, V. 2019. Machine Learning Techniques for the

- Segmentation of Tomographic Image Data of Functional Materials. *Frontiers in Materials*, 6.
- GERMAN NATIONAL COHORT, C. 2014. The German National Cohort: aims, study design and organization. *Eur J Epidemiol*, 29, 371-82.
- GUO, Y., LIU, Y., GEORGIU, T. & LEW, M. S. 2018. A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval*, 7, 87-93.
- HAENSSLE, H. A., FINK, C., SCHNEIDERBAUER, R., TOBERER, F., BUHL, T., BLUM, A., KALLOO, A., HASSEN, A. B. H., THOMAS, L., ENK, A., UHLMANN, L., READER STUDY LEVEL, I. & LEVEL, I. I. G. 2018. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol*, 29, 1836-1842.
- HAGHIGHI, M., WARFIELD, S. K. & KURUGOL, S. 2018. Automatic Renal Segmentation in Dce-Mri Using Convolutional Neural Networks. *Proc IEEE Int Symp Biomed Imaging*, 2018, 1534-1537.
- HE, K., ZHANG, X., REN, S. & SUN, J. 2015. Deep Residual Learning for Image Recognition. *arXiv e-prints* [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2015arXiv151203385H> [Accessed December 01, 2015].
- HELD, U. 2010. Welche Arten von Studiendesigns gibt es und wie werden sie korrekt eingesetzt? *Swiss Medical Forum*, 10(41), 712-714.
- HU, P., WU, F., PENG, J., BAO, Y., CHEN, F. & KONG, D. 2017. Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. *Int J Comput Assist Radiol Surg*, 12, 399-411.
- IGLESIAS, J. E. & SABUNCU, M. R. 2015. Multi-atlas segmentation of biomedical images: A survey. *Med Image Anal*, 24, 205-219.
- JACKSON, P., HARDCASTLE, N., DAWE, N., KRON, T., HOFMAN, M. S. & HICKS, R. J. 2018. Deep Learning Renal Segmentation for Fully Automated Radiation Dose Estimation in Unsealed Source Therapy. *Front Oncol*, 8, 215.
- JUNG, K. 2016. Semantische Segmentierung mit Hilfe neuronaler Netzwerke zur effizienten Verarbeitung digitalisierter Dokumente. *Tagungsband EVA Berlin 2016. Elektronische Medien & Kunst, Kultur und Historie*, 54-61.
- KNOLL, F., HAMMERNIK, K., KOBLER, E., POCK, T., RECHT, M. P. & SODICKSON, D. K. 2019. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magn Reson Med*, 81, 116-128.
- 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., FISCHER, M., MÜLLER, S., GUTTMANN, D., NIKOLAOU, K., BAMBERG, F., YANG, B., SCHICK, F. & GATIDIS, S. 2018a. Automated segmentation of abdominal organs in T1-weighted MR images using a deep learning approach: application on a large epidemiological MR study

- Proceedings of the International Society for Magnetic Resonance in Medicine (ISMRM)*. Paris, France.
- KÜSTNER, T., JANDT, M., LIEBGOTT, A., MAUCH, L., MARTIROSIAN, P., BAMBERG, F., NIKOLAOU, K., GATIDIS, S., SCHICK, F. & YANG, B. 2018b. Automatic Motion Artifact Detection for Whole-Body Magnetic Resonance Imaging. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 995-999.
- KÜSTNER, T., MÜLLER, S., FISCHER, M., SCHWARTZ, M., MARTIROSIAN, P., YANG, B., SCHICK, F. & GATIDIS, S. 2019. Automated organ segmentation 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., WEIS, J., NIKOLAOU, K., BAMBERG, F., YANG, B., SCHICK, F. & GATIDIS, S. 2018c. Semantic Organ Segmentation in 3D Whole-Body MR Images. *25th IEEE International Conference on Image Processing (ICIP)*. Athens, Greece: IEEE.
- LAVDAS, I., GLOCKER, B., KAMNITSAS, K., RUECKERT, D., MAIR, H., SANDHU, A., TAYLOR, S. A., ABOAGYE, E. O. & ROCKALL, A. G. 2017. Fully automatic, multiorgan segmentation in normal whole body magnetic resonance imaging (MRI), using classification forests (CFs), convolutional neural networks (CNNs), and a multi-atlas (MA) approach. *Med Phys*, 44, 5210-5220.
- LI, Q., WEIDONG, C., WANG, X., ZHOU, Y., FENG, D. D. & CHEN, M. 2014. Medical image classification with convolutional neural network. *13th International Conference on Control Automation Robotics & Vision (ICARCV), Singapore, 2014*, 844-848.
- LITJENS, G., KOOI, T., BEJNORDI, B. E., SETIO, A. A. A., CIOMPI, F., GHAFORIAN, 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.
- LU, M. T., IVANOV, A., MAYRHOFER, T., HOSNY, A., AERTS, H. & HOFFMANN, U. 2019. Deep Learning to Assess Long-term Mortality From Chest Radiographs. *JAMA Netw Open*, 2, e197416.
- MA, J. 2008. Dixon techniques for water and fat imaging. *J Magn Reson Imaging*, 28, 543-58.
- MEISINGER, C., PETERS, A. & LINSEISEN, J. 2016. [From the MONICA-project via KORA to the NAKO-study: Practical Utility of Epidemiological Studies in Augsburg Region]. *Gesundheitswesen*, 78, 84-90.
- MILLETARI, F., NAVAB, N. & AHMADI, S.-A. 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *arXiv e-prints* [Online]. Available: <https://ui.adsabs.harvard.edu/#abs/2016arXiv160604797M> [Accessed June 01, 2016].
- NITZ, W., RUNGE, V. & SCHMEETS, S. 2011. MR-Sicherheit: Statisches Magnetfeld. In: SCHMEETS, S., RUNGE, V. & NITZ, W. (eds.) *Praxiskurs MRT. 2., überarbeitete und erweiterte Auflage*. ed.

- NOLDEN, M., ZELZER, S., SEITEL, A., WALD, D., MULLER, M., FRANZ, A. M., MALEIKE, D., FANGERAU, M., BAUMHAUER, M., MAIER-HEIN, L., MAIER-HEIN, K. H., MEINZER, H. P. & WOLF, I. 2013. The Medical Imaging Interaction Toolkit: challenges and advances : 10 years of open-source development. *Int J Comput Assist Radiol Surg*, 8, 607-20.
- NORMAN, B., PEDOIA, V. & MAJUMDAR, S. 2018. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. *Radiology*, 288, 177-185.
- NVIDIA CORPORATION. 2019. *Clara Medical Imaging* [Online]. Available: <https://developer.nvidia.com/clara-medical-imaging> [Accessed 18.11.2019].
- PARMAR, C., RIOS VELAZQUEZ, E., LEIJENAAR, R., JERMOUMI, M., CARVALHO, S., MAK, R. H., MITRA, S., SHANKAR, B. U., KIKINIS, R., HAIBE-KAINS, B., LAMBIN, P. & AERTS, H. J. 2014. Robust Radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS One*, 9, e102107.
- PERKINS, T., DUIJNDAM, A., EGGERS, H., WEERDT, E. & RIJCKAERT, Y. 2015. *The next generation fat-free imaging*.
- RÖHRING, B., PREL, J.-B. D. & BLETTNER, M. 2009a. Studiendesign in der medizinischen Forschung. *Dtsch Arztebl International*, 106(11), 184-9.
- RÖHRING, B., PREL, J.-B. D., WACHTLIN, D. & BLETTNER, M. 2009b. Studientypen in der medizinischen Forschung. *Dtsch Arztebl International*, 106(15), 262-8.
- RONNEBERGER, O., FISCHER, P. & BROX, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints* [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2015arXiv150504597R> [Accessed May 01, 2015].
- ROTH, H. R., LU, L., FARAG, A., SHIN, H.-C., LIU, J., TURKBEY, E. B. & SUMMERS, R. M. DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation. 2015 Cham. Springer International Publishing, 556-564.
- SCHLETT, C. L., HENDEL, T., WECKBACH, S., REISER, M., KAUCZOR, H. U., NIKOLAOU, K., GUNTHER, M., FORSTING, M., HOSTEN, N., VOLZKE, H. & BAMBERG, F. 2016. Population-Based Imaging and Radiomics: Rationale and Perspective of the German National Cohort MRI Study. *Rofo*, 188, 652-61.
- SELVIKVÅG LUNDERVOLD, A. & LUNDERVOLD, A. 2018. An overview of deep learning in medical imaging focusing on MRI. *arXiv e-prints* [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018arXiv181110052S> [Accessed November 01, 2018].
- SELVIKVÅG LUNDERVOLD, A., RØRVIK, J. & LUNDERVOLD, A. 2017. Fast semi-supervised segmentation of the kidneys in DCE-MRI using convolutional neural networks and transfer learning. *2nd International Scientific Symposium, Functional Renal Imaging: Where Physiology, Nephrology, Radiology and Physics Meet*. Max Delbrück Communications Center, Berlin.

- SHELHAMER, E., LONG, J. & DARRELL, T. 2017. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans Pattern Anal Mach Intell*, 39, 640-651.
- SHIN, H. C., ROTH, H. R., GAO, M., LU, L., XU, Z., NOGUES, I., YAO, J., MOLLURA, D. & SUMMERS, R. M. 2016. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans Med Imaging*, 35, 1285-98.
- SHUANG, Y., DI, X. & KANAGASINGAM, Y. 2017. Exudate detection for diabetic retinopathy with convolutional neural networks. *Conf Proc IEEE Eng Med Biol Soc*, 2017, 1744-1747.
- SOOMRO, M. H., COPPOTELLI, M., CONFORTO, S., SCHMID, M., GIUNTA, G., DEL SECCO, L., NERI, E., CARUSO, D., RENGO, M. & LAGHI, A. 2019. Automated Segmentation of Colorectal Tumor in 3D MRI Using 3D Multiscale Densely Connected Convolutional Neural Network. *J Healthc Eng*, 2019, 1075434.
- THE, L. 2018. Artificial intelligence in health care: within touching distance. *Lancet*, 390, 2739.
- TSAO, C. W. & VASAN, R. S. 2015. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *Int J Epidemiol*, 44, 1800-13.
- VORONTSOV, I. E., KULAKOVSKIY, I. V. & MAKEEV, V. J. 2013. Jaccard index based similarity measure to compare transcription factor binding site models. *Algorithms Mol Biol*, 8, 23.
- WICHMANN, H. E., KAAKS, R., HOFFMANN, W., JOCKEL, K. H., GREISER, K. H. & LINSEISEN, J. 2012. [The German National Cohort]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 55, 781-7.
- WOLF, I., VETTER, M., WEGNER, I., BOTTGER, T., NOLDEN, M., SCHOBINGER, M., HASTENTEUFEL, M., KUNERT, T. & MEINZER, H. P. 2005. The medical imaging interaction toolkit. *Med Image Anal*, 9, 594-604.
- XIE, G.-S., ZHANG, X.-Y., YAN, S. & LIU, C.-L. 2016. Hybrid CNN and Dictionary-Based Models for Scene Recognition and Domain Adaptation. *arXiv e-prints* [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2016arXiv160107977X> [Accessed January 01, 2016].
- 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.
- 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.
- ZHAO, N., TONG, N., RUAN, D. & SHENG, K. 2019. Fully Automated Pancreas Segmentation with Two-stage 3D Convolutional Neural Networks. *arXiv e-prints* [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2019arXiv190601795Z> [Accessed June 01, 2019].

ZHOU, K., GU, Z., LIU, W., LUO, W., CHENG, J., GAO, S. & LIU, J. 2018.
Multi-Cell Multi-Task Convolutional Neural Networks for Diabetic
Retinopathy Grading. *Conf Proc IEEE Eng Med Biol Soc*, 2018, 2724-
2727.

7 Anhang

Tabelle 12: Zuordnung der automatisch segmentierten Organe zu den Scores „Organsegmentierung“ und „sonstige Fehlsegmentierung“. Aufgelistet wurden diese sortiert nach den jeweiligen Datensätzen.

Fortlaufende Nummerierung Datensätze	Organ	Score „Organsegmentierung“	Score „sonstige Fehlsegmentierungen“
1	Linke Niere	2	I
	Rechte Niere	3	I
	Pankreas	4	II
2	Linke Niere	2	0
	Rechte Niere	3	I
	Pankreas	5	0
3	Linke Niere	1	II
	Rechte Niere	3	I
	Pankreas	4	II
4	Linke Niere	2	0
	Rechte Niere	3	0
	Pankreas	5	II
5	Linke Niere	3	0
	Rechte Niere	3	I
	Pankreas	4	II
6	Linke Niere	5	I
	Rechte Niere	3	0
	Pankreas	6	II
7	Linke Niere	2	II
	Rechte Niere	2	I
	Pankreas	4	II
8	Linke Niere	1	0
	Rechte Niere	1	I
	Pankreas	4	II
9	Linke Niere	3	I

	Rechte Niere	2	I
	Pankreas	3	I
10	Linke Niere	2	0
	Rechte Niere	2	I
	Pankreas	4	II
11	Linke Niere	3	I
	Rechte Niere	2	I
	Pankreas	4	II
12	Linke Niere	3	I
	Rechte Niere	4	I
	Pankreas	6	0
13	Linke Niere	2	0
	Rechte Niere	3	0
	Pankreas	4	0
14	Linke Niere	2	0
	Rechte Niere	3	I
	Pankreas	3	I
15	Linke Niere	3	I
	Rechte Niere	3	I
	Pankreas	4	I
16	Linke Niere	3	I
	Rechte Niere	2	II
	Pankreas	3	0
17	Linke Niere	1	I
	Rechte Niere	1	I
	Pankreas	5	I
18	Linke Niere	2	I
	Rechte Niere	3	I
	Pankreas	4	I
19	Linke Niere	3	0
	Rechte Niere	3	0

	Pankreas	3	II
20	Linke Niere	2	I
	Rechte Niere	3	0
	Pankreas	3	II

8 Erklärungen zum Eigenanteil

Die Arbeit wurde innerhalb der Radiologischen Universitätsklinik Tübingen in der Abteilung für Diagnostische und Interventionelle Radiologie unter der Betreuung von PD Dr. Sergios Gatidis durchgeführt.

Die Konzeption der Studie erfolgte durch PD Dr. Sergios Gatidis. Er hat die Arbeit betreut und die Erstellung des Manuskripts begleitet.

Die MRT-Aufnahmen sowie die dazugehörigen demographischen Daten wurden von der NAKO Gesundheitsstudie zur Verfügung gestellt.

Die Entwicklung, Implementierung und der Trainings- sowie Inferenzprozess des maschinellen Lernalgorithmus mithilfe der Daten der NAKO-Studie sowie die automatische Evaluation erfolgten durch Sarah Müller, ebenfalls die Berechnung der Metriken unter Punkt 3.5.2 sowie die Erstellung der Abbildung 11 im Ergebnisteil. Die Datengewinnung und Auswertung unter Punkt 3.3 erfolgte in Zusammenarbeit mit Sarah Müller.

Die Segmentierungsanalysen wurden von Mitarbeitern der Abteilung für Diagnostische und Interventionelle Radiologie am Universitätsklinikum Tübingen durchgeführt. Die Bestimmung des Fettgehalts und der Organvolumina aus den manuell und automatisch erstellten Masken erfolgte durch Dr. Tobias Hepp.

Die statistische Auswertung erfolgte mit Unterstützung durch PD Dr. Sergios Gatidis und Dr. Tobias Hepp.

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

Tübingen,

9 Danksagung

Zuallererst möchte ich meinem Doktorvater, PD Dr. Sergios Gatidis, für die fachliche sowie stets motivierende Unterstützung und Betreuung herzlich danken. Für die Überlassung des Themas bin ich ihm ebenfalls sehr dankbar.

Mein Dank gilt Dr. Tobias Hepp für die Unterstützung während des Schreibprozesses bei inhaltlichen Fragen sowie insbesondere bei der Beantwortung all meiner Fragen zur statistischen Auswertung.

Sarah Müller danke ich für die technische Umsetzbarkeit und Durchführung der automatischen Segmentierung sowie für die Entwicklung des maschinellen Lernalgorithmus. Für ihre ausführliche Hilfe bei Fragen zur Durchführung und Beschreibung möchte ich ihr ebenfalls danken.

Allen Mitarbeiterinnen und Mitarbeitern der Abteilung für Diagnostische und Interventionelle Radiologie der Radiologischen Universitätsklinik Tübingen, insbesondere denen im PET-MRT, möchte ich hiermit für die angenehme und stets freundliche sowie hilfsbereite Atmosphäre während der Durchführung meiner Arbeit danken.

Den Mitarbeitern der NAKO Gesundheitsstudie bin ich dankbar für die Überlassung der MRT-Aufnahmen, der demographischen Daten und der in der Einleitung verwendeten Abbildung 1.

Ich bin Philina Schmidt für das Korrekturlesen sowie all meinen Freunden, die mein Leben und Studium so sehr bereichern, dankbar.

Meinen Eltern, durch die mein Studium überhaupt erst möglich ist, möchte ich danken. Für ihre stete Unterstützung und für ihr Vertrauen in meine Pläne und Vorhaben bin ich ihnen ebenfalls sehr dankbar.