

AUS DEM INSTITUT FÜR DIAGNOSTISCHE UND INTERVENTIONELLE RADIOLOGIE,
KINDER- UND NEURORADIOLOGIE

DIREKTOR: PROF. DR. MED. MARC-ANDRÉ WEBER, M.SC.

INNOVATIVE QUANTITATIVE ANALYSEVERFAHREN IN DER HERZ-MRT:

ENTWICKLUNG, VALIDIERUNG UND DIAGNOSTISCHE RELEVANZ

KUMULATIVE INAUGURALDISSERTATION

ZUR ERLANGUNG DES AKADEMISCHEN GRADES

DOKTOR DER MEDIZINWISSENSCHAFTEN (DR. RER. HUM.)

IM FACHGEBIET MEDIZINISCHE INFORMATIK

AN DER UNIVERSITÄTSMEDIZIN ROSTOCK

VORGELEGT VON

MATHIAS MANZKE | GEB. AM 13.03.1980 IN BERGEN

ROSTOCK | 03.04.2025

https://doi.org/10.18453/rosdok_id00005666

DISSERTATION

UNIVERSITÄT ROSTOCK



Gutachter:

Prof. Dr. med. Felix Meinel, Universitätsmedizin Rostock,
Institut für Diagnostische und Interventionelle Radiologie, Kinder- und
Neuroradiologie

Prof. Dr. rer. nat. Georg Füllen, Universitätsmedizin Rostock,
Institut für Biostatistik und Informatik in Medizin und Altersforschung

Prof. Dr. med. Matthias May, Universitätsklinikum Erlangen, Radiologisches Institut

Jahr der Einreichung: 2025

Jahr der Verteidigung: 2025

INHALTSVERZEICHNIS

I.	EINLEITUNG	1
1.1	HERAUSFORDERUNGEN UND PROBLEME IN DER KARDIALEN BILDGEBUNG	1
1.2	LÖSUNGSANSÄTZE DURCH KI-GESTÜTZTE TECHNOLOGIEN IN DER MRT	3
1.3	ZIELSETZUNG UND AUFBAU DER ARBEIT	4
II.	MATERIAL UND METHODEN	6
2.1	STUDIENDESIGN UND PATIENTENAUSWAHL	6
2.2	KI-BASIERTE BESCHLEUNIGUNG VON MRT-SEQUENZEN	7
2.3	ENTWICKLUNG EINES DL-BASIERTEN KARDIALEN SEGMENTIERUNGSVERFAHRENS.....	11
2.3.1	DATENPARTITIONEN UND MODELLARCHITEKTUR.....	11
2.3.2	VALIDIERUNG UND VERGLEICH ZUR MANUELLEN SEGMENTIERUNG	15
2.4	UNTERSUCHUNG DER REPRODUZIERBARKEIT VON RADIOMICS-FEATURES	16
2.4.1	BILDVERARBEITUNG UND RADIOMICS-EXTRAKTION	16
2.4.2	STATISTISCHE METHODEN ZUR ANALYSE DER REPRODUZIERBARKEIT	17
III.	ERGEBNISSE	19
3.1	OPTIMIERUNG DER MRT-AUFNAHMEN DURCH KI.....	19
3.1.1	AUSWIRKUNGEN AUF BILDQUALITÄT UND AUFNAHMEZEIT	19
3.2	AUTOMATISIERTE SEGMENTIERUNG DES MYOKARDS BEIM T1-MAPPING	20
3.2.1	AUSWAHL DES OPTIMALEN MODELLS BASIEREND AUF DER MODELLELEISTUNG.....	20
3.2.2	GENAUIGKEIT DES DL-MODELLS	21
3.2.3	VERGLEICH ZUR MANUELLEN SEGMENTIERUNG	21
3.3	REPRODUZIERBARKEIT VON RADIOMICS-FEATURES.....	23
3.3.1	ERGEBNISSE FÜR T1- UND T2-MAPPING	23
3.3.2	EINFLUSS VON AUFLÖSUNG UND BILDORIENTIERUNG.....	24
IV.	DISKUSSION	26
4.1	BEWERTUNG DER ERGEBNISSE	26
4.1.1	NUTZEN DER KI-GESTÜTZTEN MRT-BESCHLEUNIGUNG UND KLINISCHE RELEVANZ.....	26
4.1.2	STÄRKEN UND SCHWÄCHEN DER AUTOMATISIERTEN SEGMENTIERUNG	26
4.1.3	BEDEUTUNG DER REPRODUZIERBARKEIT VON RADIOMICS-FEATURES.....	28

4.2	LIMITATIONEN DER STUDIEN	30
V.	ZUSAMMENFASSUNG	31
5.1	KERNERGEBNISSE DER STUDIEN	31
5.2	IMPLIKATIONEN UND PERSPEKTIVEN FÜR DIE KARDIALE BILDGEBUNG	32
VI.	LITERATURVERZEICHNIS	33
VII.	THESEN	39
VIII.	DANKSAGUNG	41
IX.	LISTE ALLER VERÖFFENTLICHUNGEN UND FACHVORTRÄGE	42
X.	ANHANG	45

ABKÜRZUNGSVERZEICHNIS

2D-Shape	2D-Formmerkmale
4Ch	Vierkammerblick
AHA	American Heart Association
API	Apikal
BAS	Basal
CCC	Konkordanz-Korrelationskoeffizient
CMR	Kardiale Magnetresonanztomographie
CNN	Convolutional Neural Network
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
DSC	Sørensen-Dice-Coefficient / Dice score
EDV	Enddiastolisches Volumen
EF	Ejektionsfraktion
ESV	Endsystolisches Volumen
GLCM	Gray Level Co-occurrence Matrix features
GLDM	Gray Level Dependence Matrix features
GLRLM	Gray Level Run Length Matrix features
GLSZM	Gray Level Size Zone Matrix features
ICC	Intraklassenkorrelationskoeffizient
IOU	Intersection over Union (Jaccard Koeffizient)
IQR	Interquartilsabstand
IR	Inversion-Recovery
LAX	Langachsenschnitt
LeakyReLU	Leaky Rectified Linear Unit
LGE	Late-Gadolinium-Enhancement
LV	Linksventrikulär
MID	Mittventrikulär
MOLLI	Modified Look-Locker Inversion recovery
MRT	Magnetresonanztomographie
NGTDM	Neighboring Gray Tone Difference Matrix features
ReLU	Rectified Linear Unit
ROI	Region of Interest
SAX	Kurzachsenschnitt
SV	Schlagvolumen
UMR	Unimedizin Rostock

ABBILDUNGSVERZEICHNIS

ABBILDUNG 1 – ÜBERSICHT DES WORKFLOWS ZUR VERARBEITUNG KARDIALER MRT-DATEN....	4
ABBILDUNG 2 – SCHEMATISCHE DARSTELLUNG DER DREI STUDIENKOHORTEN.	6
ABBILDUNG 3 – VISUALISIERUNG VON AKQUISITIONSSCHEMATA.	8
ABBILDUNG 4 – ABTASTMUSTER.	9
ABBILDUNG 5 – QUANTIFIZIERUNG DER KANTENSCHÄRFE.....	10
ABBILDUNG 6 – U-NET ARCHITEKTURMODELL FÜR BILDSEGMENTIERUNG.	14
ABBILDUNG 7 – MODELLVORHERSAGEN VS. REFERENZ.	22
ABBILDUNG 8 – EMPFOHLENE FEATURES FÜR ALLE T1- UND T2-MAPS.	24

TABELLENVERZEICHNIS

TABELLE 1 – AUFNAHMEPARAMETER:	8
TABELLE 2 – ÜBERSICHT ÜBER DATENPARTITIONEN.....	12
TABELLE 3 – AUFNAHMEPARAMETER UND BILDQUALITÄT.....	19
TABELLE 4 – MODELLVERGLEICH FÜR ALLE ANSICHTEN.	21
TABELLE 5 – METRIKEN DES BESTEN MODELLS UND INTERRATER-VERGLEICH.	22

I. EINLEITUNG

Herz-Kreislauf-Erkrankungen sind die häufigste Todesursache weltweit und verantwortlich für etwa 32 % aller Todesfälle [1]. Die kardiale Magnetresonanztomographie (CMR) ist ein unverzichtbares diagnostisches Verfahren zur Beurteilung von Herzkrankheiten und bietet eine nicht-invasive Möglichkeit, die Struktur, Funktion und Gewebeeigenschaften des Herzens detailliert darzustellen.

Ein wesentlicher Vorteil der CMR liegt in ihrer Vielseitigkeit. Sie erlaubt die präzise Visualisierung kardialer Pathologien wie Myokardinfarkte, Myokarditis, Kardiomyopathien und angeborene Herzfehler. Darüber hinaus bietet sie wertvolle Informationen zur Perfusion, Narbenbildung und zu Entzündungsprozessen im Myokard, was für die Diagnosestellung und Therapieplanung von entscheidender Bedeutung ist.

Die Gewebecharakterisierung mittels T1- und T2-Relaxationszeiten sowie der späten Gadolinium-Anreicherung (LGE) erlaubt eine differenzierte Unterscheidung zwischen ischämischen und nicht-ischämischen Herzerkrankungen und trägt entscheidend zur Identifikation von Patienten mit erhöhtem Risiko für Arrhythmien oder Herzinsuffizienz bei.

Zudem ermöglicht die Herz-MRT eine exakte Volumen- und Funktionsanalyse der Ventrikel, was für die Verlaufskontrolle bei Patienten mit Herzinsuffizienz oder valvulären Erkrankungen essenziell ist.

Die Integration von Künstlicher Intelligenz (KI) in die MRT bringt zusätzliche Vorteile: Sie vereinfacht die Bilderzeugung und Diagnostik, reduziert die Variabilität zwischen Untersuchern und erhöht die Effizienz im klinischen Alltag.

Insgesamt stellt die CMR ein unverzichtbares Tool in der kardiovaskulären Diagnostik dar.

1.1 HERAUSFORDERUNGEN UND PROBLEME IN DER KARDIALEN BILDGEBUNG

Trotz signifikanter Fortschritte in der CMR bestehen weiterhin Herausforderungen, die die Effektivität und Effizienz dieser diagnostischen Methode beeinträchtigen. Eine der zentralen Hürden der CMR ist die vergleichsweise lange Untersuchungsdauer, die für Patienten mit Arrhythmien oder Atemproblemen problematisch sein kann.

Insbesondere Mapping-Sequenzen ermöglichen eine präzise quantitative Gewebecharakterisierung des Myokards [2, 3]. Die myokardiale T1/T2-Relaxation gibt Aufschluss über die biophysikalischen Eigenschaften der Myokardmuskelzellen und des umgebenden extrazellulären Raums. Veränderungen der T1- und T2-Relaxationszeiten dienen als diagnostische Marker für myokardiale Pathologien wie Fibrose, Entzündungen, Ödeme oder Speicherkrankheiten [4]. In der klinischen Praxis erfolgt die Bestimmung dieser

Relaxationszeiten häufig durch manuelle Segmentierung relevanter Bildbereiche (ROIs), die als Binärmasken, also Bilder mit nur zwei Werten (z. B. 0 = außerhalb, 1 = innerhalb der ROI), dargestellt werden. Diese manuelle Vorgehensweise ist jedoch zeitaufwändig und unterliegt einer gewissen Beurteiler-Variabilität.

Obwohl es umfangreiche Literatur zur automatisierten Segmentierung von Herz-Cine-Sequenzen für die volumetrische Analyse gibt [5–13], befassten sich bisher nur wenige Studien mit der Herausforderung der automatisierten Segmentierung von Mapping-Daten [14–20].

Neben der globalen oder regionalen Auswertung der durchschnittlichen Relaxationszeiten bieten quantitative Merkmale (Radiomics-Features), die mit bloßem Auge nicht erkennbar sind, eine Möglichkeit, detaillierte quantitative Informationen aus myokardialen T1- und T2-Mappingdaten zu extrahieren.

Radiomics-Features sind numerische Parameter aus segmentierten ROIs. Sie analysieren Textur, Form, Intensität und Muster und dienen als potenzielle Bildbiomarker für Myokarderkrankungen [21, 22].

Radiomics-Analysen auf myokardialen nativen T1-Mappingdaten wurden bereits als diagnostische Biomarker bei zahlreichen kardiologischen Erkrankungen beforscht. Beispielsweise wurden T1-Mapping-Radiomics-Features verwendet, um Myokarditis zu diagnostizieren [23], die extrazelluläre Matrix bei dilatativer Kardiomyopathie (DCM) zu charakterisieren [24], hypertrophe Kardiomyopathie-Genotypen [25] und Phänotypen hypertropher Kardiomyopathie zu klassifizieren [26, 27], hypertrophe Kardiomyopathie von hypertensiver Herzkrankheit zu unterscheiden [28, 29] sowie Herztumoren von Thromben abzugrenzen [30].

Auch als prognostische Biomarker zeigen T1-Radiomics-Features Potenzial. Sie wurden eingesetzt, um den klinischen Verlauf bei dilatativer Kardiomyopathie [31, 32], Herzinsuffizienz bei systemischem Lupus erythematodes [33], die funktionelle Erholung nach einem Myokardinfarkt [34] und den Therapieerfolg bei pulmonaler arterieller Hypertonie [35] vorherzusagen.

Die Anwendung der Radiomics-Analyse von T2-Maps konzentrierte sich bisher hingegen auf die Diagnose von akuter und chronischer Myokarditis [23].

Eine wesentliche Voraussetzung für diagnostische oder prognostische Signifikanz ist die Reproduzierbarkeit der Messergebnisse. Obwohl die Reproduzierbarkeit von Radiomics-Features erhebliche Beachtung erhalten hat [36–38] und für kardiale Cine-Sequenzen untersucht wurde [39], ist die Datenlage zur Wiederholbarkeit von Radiomics-Features beim myokardialen T1- und T2-Mapping bislang unzureichend.

Bisher hat lediglich eine Studie die Reproduzierbarkeit myokardialer T1- und T2-Radiomics-Features spezifisch untersucht. Diese fand heraus, dass lediglich eine kleine Teilmenge der Merkmale reproduzierbar war [40]. Allerdings konzentrierte sich diese Analyse auf die Gesamtleistung von Merkmalsklassen, nicht auf einzelne Merkmale. Auch die möglichen Einflüsse der Schichtorientierung in Kurzachsenansicht (SAX) vs. Langachsenansicht (LAX) und der räumlichen Auflösung wurden bisher nicht adressiert.

1.2 LÖSUNGSANSÄTZE DURCH KI-GESTÜTZTE TECHNOLOGIEN IN DER MRT

Aktuelle Forschungsarbeiten zielen darauf ab, diese Herausforderungen über den gesamten Workflow, von der Aufnahme über die Nachverarbeitung bis hin zur diagnostischen Auswertung und diagnostischen Relevanz mithilfe von KI zu adressieren. Methodisch unterscheidet man zwischen überwachtem Lernen (auf Basis gelabelter Daten), unüberwachtem Lernen (zur Mustererkennung in unstrukturierten Daten) und bestärkendem Lernen (mittels Belohnungs- und Bestrafungssignalen). Ein wichtiger Bereich des maschinellen Lernens ist das Deep Learning (DL), das auf tiefen neuronalen Netzwerken basiert. Eine spezielle Architektur sind Convolutional Neural Networks (CNNs), die insbesondere für die Bild- und Videoverarbeitung optimiert sind. CNNs verwenden Faltungsoperationen (Convolutions), um charakteristische Merkmale aus Eingabedaten, insbesondere Bildern, zu extrahieren und diese schrittweise auf verschiedenen Ebenen zu analysieren. Sie sind essenziell für Anwendungen wie Objekterkennung, medizinische Bildanalyse und autonome Systeme.

DL-basierte Bildrekonstruktionstechniken können eine Beschleunigung der kardialen MRT ermöglichen, um die Untersuchungszeit zu verkürzen und gleichzeitig die Bildqualität zu erhalten. CNNs verbessern zudem die Bildqualität durch Rauschreduktion – besonders bei niedrigen Signal-zu-Rausch-Verhältnissen und geringeren Magnetfeldstärken [3, 41, 42].

Mithilfe vollautomatisierter Algorithmen zur Myokard-Segmentierung auf T1-Mapping-MRT-Daten, kann der Bedarf an manuellen Eingriffen reduziert und die Konsistenz der Ergebnisse erhöht werden. Während die manuelle Segmentierung zeitaufwändig, mühsam und anfällig für subjektive Variationen zwischen verschiedenen Radiologen ist, ermöglichen DL-gestützte Algorithmen eine vollautomatische und reproduzierbare Konturierung. Solche Segmentierungen bilden die Grundlage für Radiomics-Analysen und die Erforschung neuer diagnostischer Bildbiomarker. Allerdings ist für die diagnostische Relevanz dieser Marker die numerische Stabilität unter verschiedenen Randbedingungen essenziell.

Um die Standardisierung und Vergleichbarkeit dieser Merkmale zu verbessern, ist eine Reproduzierbarkeitsanalyse unerlässlich.

Die Integration dieser neuen Ansätze in die klinische Praxis könnte dazu beitragen, bestehende Herausforderungen in der kardialen MRT zu bewältigen und sowohl die diagnostische Genauigkeit als auch die Effizienz nachhaltig zu steigern.

1.3 ZIELSETZUNG UND AUFBAU DER ARBEIT

Diese kumulative Dissertation basiert auf folgenden drei Originalarbeiten:

1. Klemenz AC, Reichardt L, Gorodezky M, **Manzke M**, Zhu X, Dalmer A, Lorbeer R, Lang CI, Weber MA, Meinel FG. Accelerated Cardiac MRI with Deep Learning-based Image Reconstruction for Cine Imaging. Radiol Cardiothorac Imaging. 2024 Dec;6(6):e230419. [43]
2. **Manzke M**, Iseke S, Böttcher B, Klemenz AC, Weber MA, Meinel FG. Development and performance evaluation of fully automated deep learning-based models for myocardial segmentation on T1 mapping MRI data. Sci Rep. 2024 Aug 14;14(1):18895. [45]
3. **Manzke M**, Laqua FC, Böttcher B, Klemenz AC, Weber MA, Baeßler B, Meinel FG. Repeatability of radiomic features in myocardial T1 and T2 mapping. Eur Radiol. 2025 Jan 15. [44]

Diese drei aufeinander aufbauenden Studien beleuchten den Einsatz Künstlicher Intelligenz zur Optimierung der CMR in drei wesentlichen Bereichen der Bildverarbeitungskette: der Beschleunigung der Bildaufnahme, der Entwicklung eines automatisierten Segmentierungsverfahrens und der Analyse der Reproduzierbarkeit von Radiomics-Features, wie schematisch in **Abbildung 1** dargestellt ist. Diese Studien bilden das Fundament für eine präzisere und effizientere Diagnostik sowie eine optimierte Untersuchungs- und Therapieplanung.

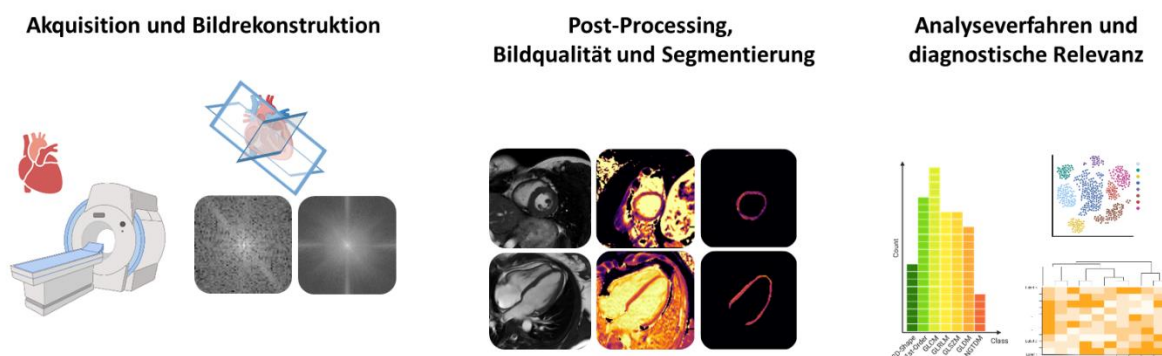


ABBILDUNG 1 – ÜBERSICHT DES WORKFLOWS ZUR VERARBEITUNG KARDIALER MRT-DATEN.

VON LINKS NACH RECHTS: (1) AKQUISITION UND BILDREKONSTRUKTION, EINSCHLIEßLICH DER AUFNAHME VON MRT-SCANS UND DER BILDWIEDERHERSTELLUNG. (2) POST-PROCESSING, BILDQUALITÄT UND SEGMENTIERUNG, MIT FOKUS AUF DIE VERBESSERUNG DER BILDQUALITÄT UND DIE AUTOMATISCHE MYOKARD-SEGMENTIERUNG. (3) ANALYSEVERFAHREN UND DIAGNOSTISCHE RELEVANZ, WELCHE DIE EXTRAKTION VON RADIOMICS-FEATURES, MASCHINELLES LERNEN UND DIE KLINISCHE INTERPRETATION DER DATEN UMFASST.

Studie 1: Akquisition und Bildrekonstruktion

Die erste Originalarbeit [43] untersuchte eine DL-gestützte Optimierung der CMR-Bildaufnahme. Die Cine-Sequenz basierte MRT gilt als Referenzstandard zur Beurteilung der Herzfunktion [46]. Sie erfasst den Herzzyklus zeitlich aufgelöst und wird aus k-Raum-Daten rekonstruiert.

Trotz technischer Fortschritte bleiben lange Aufnahmezeiten und Artefakte durch Arrhythmien zentrale Limitationen [47]. Beschleunigte Sequenzen existieren, sind jedoch aufgrund von geringem Signal-Rausch-Verhältnis und reduzierter Bildqualität klinisch eingeschränkt. DL-basierte Rekonstruktionsverfahren könnten diese Schwächen überwinden [48, 49].

Da herkömmliche Cine-Sequenzen über 10–12 Herzzyklen aufgenommen werden, prüften wir die Hypothese, dass DL-gestützte Rekonstruktionen eine schnellere, potenziell echtzeitfähige Bildgebung bei adäquater Qualität ermöglichen. Ziel war die Bewertung ihres Einflusses auf Erfassungszeit, Volumenmessung und Bildqualität.

Studie 2: Post-Processing

In der zweiten Originalarbeit [44] wurde ein Deep-Learning-Modell entwickelt und validiert, das die Segmentierung des linksventrikulären Myokards auf nativen T1-Maps der Herz-MRT automatisiert und mit hoher Genauigkeit durchführt. Dabei wurden sowohl LAX- als auch SAX-Ansichten berücksichtigt.

Diese Arbeit adressiert die Herausforderungen der zeitaufwändigen und fehleranfälligen manuellen Segmentierung und verfolgt das Ziel, die Konsistenz und Effizienz bei der Analyse von MRT-Daten zu verbessern sowie die Übereinstimmung mit menschlichen Beurteilern zu evaluieren.

Studie 3: Analyseverfahren und diagnostische Relevanz

Die dritte Originalarbeit [45] befasste sich mit der systematischen Analyse der Reproduzierbarkeit von Radiomics-Features in manuell segmentierten ROIs bei T1- und T2-Mapping-Daten. Ziel der Untersuchung war es, den Einfluss von Bildorientierung, räumlicher Auflösung und Sequenzparametern auf die Robustheit von Radiomics-Features in wiederholten Messungen zu evaluieren. Die Ergebnisse sollen zur Standardisierung radiomics-basierter Verfahren beitragen und eine verlässliche klinische Anwendung unterstützen.

Die drei Studien zeigen, wie KI-basierte Verfahren die CMR durch optimierte Bildaufnahme, automatisierte Segmentierung und standardisierte Radiomics-Analyse verbessern können.

II. MATERIAL UND METHODEN

2.1 STUDIENDESIGN UND PATIENTENAUSWAHL

Die drei monozentrischen Studien wurden jeweils an der Universitätsmedizin Rostock durchgeführt und inkludierten drei disjunkte Teilnehmergruppen (Probandengruppe I und II, Patientengruppe). Aus diesen Teilnehmergruppen wurden die jeweiligen Studienkohorten zusammengestellt, wie in folgender **Abbildung 2** dargestellt ist.

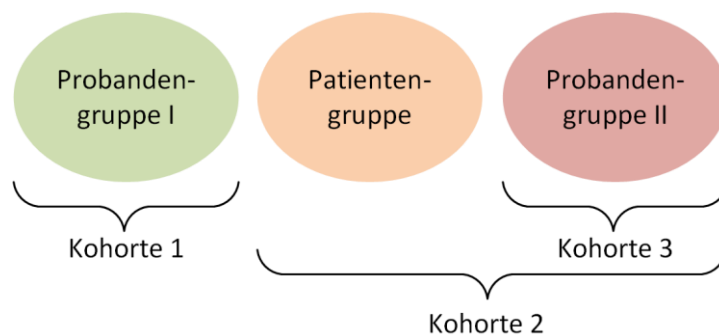


ABBILDUNG 2 – SCHEMATISCHE DARSTELLUNG DER DREI STUDIENKOHORTEN.

STUDIE 1 UMFASST AUSSCHLIEßLICH PROBANDENGRUPPE I, STUDIE 2 UNTERSUCHTE DIE VEREINIGUNGSMENGE DER PATIENTENGRUPPE UND DER PROBANDENGRUPPE II, WÄHREND SICH STUDIE 3 AUF PROBANDENGRUPPE II KONZENTRIERTE.

Die **erste Studie** war eine prospektive Kohortenstudie (**Kohorte I**) zur Evaluation der beschleunigten CMR-Aufnahmesequenz [43], die von Januar bis März 2023 durchgeführt wurde. Die Teilnehmer wurden aus einem Pool freiwilliger gesunder Erwachsener ausgewählt. Ausschlusskriterien umfassten Schwangerschaft, allgemeine Kontraindikationen für die MRT (z. B. Klaustrophobie, nicht MRT-kompatible Implantate) sowie eine bekannte Vorgeschichte von Herzerkrankungen. Insgesamt wurden 55 gesunde Erwachsene (> 18 Jahre) in die Studie eingeschlossen (mittleres Alter: 31 Jahre; Spannweite: 21–66 Jahre; 32 Frauen, 23 Männer).

Die **zweite Studie** war retrospektiv angelegt und umfasste drei Phasen:

- **Phase 1:** Definition der Einschlusskriterien für die Datenerfassung und Export der Bilddaten aus den klinischen Systemen.
- **Phase 2:** Implementierung eines Trainingsloops für mehrere DL-Modelle
- **Phase 3:** Evaluation der Modellleistungsfähigkeit anhand eines Testdatensatzes und Vergleich der Ergebnisse mit einem Referenzdatensatz, der von zwei Beurteilern erstellt wurde.

Die **Kohorte 2** umfasste 50 gesunde Teilnehmer (**Probandengruppe II**: 29 Frauen, 21 Männer, Durchschnittsalter: $39,4 \pm 13,7$ Jahre), deren Bilddaten aus einer früheren prospektiven Freiwilligenstudie [50] stammten, sowie 75 Patienten (**Patientengruppe**), die zwischen 2020 und 2022 an der Universitätsmedizin Rostock (UMR) einer klinisch indizierten CMR unterzogen wurden. Die häufigsten CMR-Indikationen der Patienten waren:

- Myokarditis (26 Patienten, 34,7 %),
- Kardiomyopathie (15 Patienten, 20,0 %) und
- ischämische Herzkrankheit (12 Patienten, 16,0 %).

Die Patienten wurden retrospektiv durch eine Suche im Radiologie-Informationssystem (RIS) identifiziert. Ausschlusskriterien umfassten schwere Bildartefakte ($n = 12$), Kinder ($n = 3$), weniger als drei Kurzachsen-T1-Maps ($n = 4$) sowie Datensatzfehler ($n = 4$).

Die **dritte Studie**, eine retrospektive Post-hoc-Analyse, nutzte ebenfalls die Daten aus der prospektiven Studie [50] (**Kohorte 3 = Probandengruppe II**). Der durchschnittliche BMI betrug $23,9 \pm 4,0$ kg/m², und die durchschnittliche Herzfrequenz lag bei 72 ± 10 Schlägen pro Minute. Jede Teilnehmerin und jeder Teilnehmer unterzogen sich zwei identischen CMR-Untersuchungen.

Die demografischen Verteilungen von Alter, Geschlecht, BMI und Herzfrequenz der drei Teilnehmergruppen werden in **Zusatzabbildung 1** im Anhang dargestellt.

2.2 KI-BASIERTE BESCHLEUNIGUNG VON MRT-SEQUENZEN

MRT Protokoll

Die Methodik basiert auf [43]. Alle Probanden wurden an einem 1,5T-MRT-System (Signa Artist, GE Healthcare, Milwaukee, WI, USA) untersucht. Das Protokoll umfasste vier SAX-Stapel nach klinischem Standard der UMR zur vollständigen Erfassung des Ventrikels.

Zusätzlich wurden drei DL-basierte Cine-Sequenzen mit unterschiedlichen Erfassungsdauern aufgenommen: 1RR, 3RR und 6RR, mit Beschleunigungsfaktoren von 15 bzw. 8. Die Beschleunigung wurde wie folgt definiert:

$$acc = \frac{K_y}{k_y} = \frac{K_y}{vps * \#RR}$$

Dabei bezeichnet acc den Beschleunigungsfaktor, K_y die vollständig abgetastete und k_y die erfasste Anzahl an Phasenlinien, vps die Ansichten pro Segment und $\#RR$ die Anzahl der für Bildgebung verwendeten Herzzyklen(siehe **Tabelle 1**).

TABELLE 1 – AUFNAHMEPARAMETER:

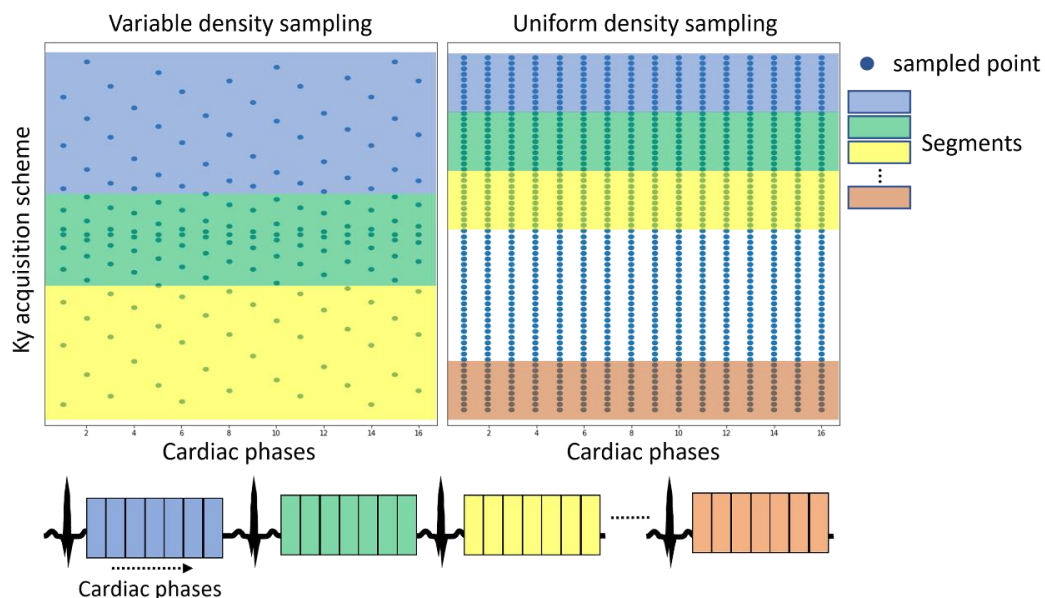
TR = WIEDERHOLUNGSZEIT, TE = ECHOZEIT, FOV=FIELD OF VIEW

MRT Parameter	Referenz	1RR Cine	3RR Cine	6RR Cine
FOV [mm]	340 x 340	340 x 340	340 x 340	340 x 340
In-plane resolution [mm ²]	1.7 x 1.5	1.9 x 1.9	1.7 x 1.7	1.5 x 1.5
Image pixel matrix	200 x 224	180 x 180	200 x 200	224 x 224
Schichtdicke [mm]	10	10	10	10
Frames / Herzzyklus	30	30	30	60
TR [ms]	3.7	3.1	3.2	3.4
TE [ms]	1.7	1.2	1.2	1.2
Ansichten pro Segment (vps)	16	12	8	5

Beschleunigte Cine-Sequenzen

Zwei Hauptkomponenten sind in der speziell untersuchten Aufnahmesequenz von GE enthalten: Erfassung von Sampling mit variabler Dichte und ein auf Deep Learning basierender Rekonstruktionsalgorithmus.

Während bei der herkömmlichen ASSET-Cine-Erfassung die Signalabtastung mit gleichmäßiger Dichte verwendet wird, kommt für die beschleunigte Cine-Sequenz eine Abtaststrategie mit variabler Dichte k - t zum Einsatz (**Abbildung 3**).

**ABBILDUNG 3 – VISUALISIERUNG VON AKQUISITIONSSCHEMATA.**

FÜR EINE CINE-AKQUISITION WIRD DER K-RAUM IN SEGMENTE (FARBIGE KÄSTCHEN) UNTERTEILT, DIE IN ALLEN HERZPHASEN ABGETASTET WERDEN MÜSSEN, WIE IM UNTEREN TEIL DER ABBILDUNG ZU SEHEN IST. BEI EINER STANDARDSEQUENZ (RECHTS) WIRD DER K-RAUM GLEICHMÄßIG ABGETASTET, SODASS DIE SEGMENTE GLEICH GROß SIND. BEI DER ABTASTUNG MIT VARIABLER DICHTHE HABEN DIE ÄUßEREN K-RAUM-SEGMENTE WENIGER PUNKTE, SODASS IN DER GLEICHEN ZEIT EIN GRÖßERES SEGMENT ERFASST WERDEN KANN. DIE SEGMENTGRÖßE WIRD AN DIE ABTASTDICHTHE ANGEPAßT, WODURCH INSGESAMT EINE GERINGERE SEGMENTANZAHL UND DAMIT EINE BESCHLEUNIGTE AKQUISITION MÖGLICH SIND. ABBILDUNG ÜBERNOMMEN AUS [43].

Für die beschleunigte Sequenz verwendet jede Herzphase dieselbe Abtastdichte und denselben Beschleunigungsfaktor um ein konsistentes Abtastmuster zu erzeugen. Zusätzlich wird für jede Herzphase eine zufällige Verschiebung in der Phasenkodierungsdimension angewendet, um eine Sampling-Inkohärenz entlang der Herzphasendimension zu erzeugen. Nach der Generierung des Abtastmusters werden die Abtastpunkte gemäß ihrer *ky*-Position in verschiedenen Herzzyklen sortiert. Zur Rekonstruktion der Cine-Aufnahmen aus den unterabgetasteten k-Raum-Daten kommt im Sequenzprototyp ein entrolltes neuronales Netzwerk zum Einsatz, das in [43] ausführlicher beschrieben ist.

Die genauen Abtastmuster für die 1RR-, 3RR- und 6RR-Cine-Sequenzen werden in folgender **Abbildung 4** dargestellt.

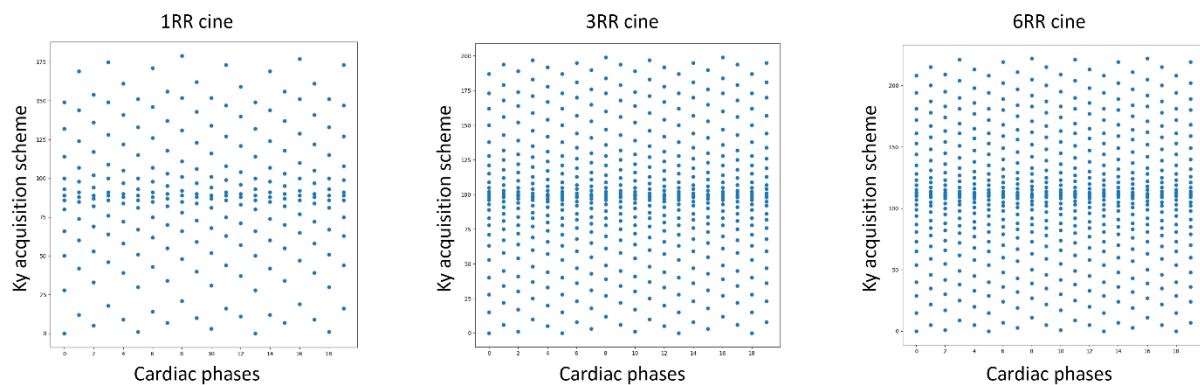


ABBILDUNG 4 – ABTASTMUSTER.

FÜR 1RR (MATRIXGRÖÖE KY=180, BESCHLEUNIGUNG 15); 3RR (MATRIXGRÖÖE KY=200, BESCHLEUNIGUNG 8); 6RR (MATRIXGRÖÖE KY=224, BESCHLEUNIGUNG 8) ÜBER 20 HERZPHASEN. JEDER PUNKT STELLT EINE VOLLSTÄNDIG ABGETASTETE LINIE IN KX-RICHTUNG DAR. ABBILDUNG ÜBERNOMMEN AUS [43].

Bewertung der Aufnahmezeit

Bei zehn Teilnehmern wurden Sequenzlaufzeit, Atemanhalte-Dauer und -Anzahl sowie die Pausen dazwischen erfasst. Die Gesamtzeit setzte sich aus diesen drei Komponenten zusammen.

Bewertung der Bildqualität

Zwei Radiologen (5 bzw. 10 Jahre CMR-Erfahrung) bewerteten die *subjektive Bildqualität* anhand einer fünfstufigen Skala (1 = nicht diagnostisch, 2 = schlecht, 3 = ausreichend, 4 = gut, 5 = exzellent) hinsichtlich Kontrast, Unschärfe und Artefakten. Eine vollständige Verblindung war aufgrund der unterschiedlichen Bildcharakteristika nicht durchführbar.

Zur *objektiven Quantifizierung der Bildqualität* wurde die Kantenschärfe in den Bildern bestimmt. Diese wurde als Steigung der Signalintensitätsänderung beim Übergang vom Blutpool zum Myokard gemessen, wie in folgender **Abbildung 5** dargestellt ist.

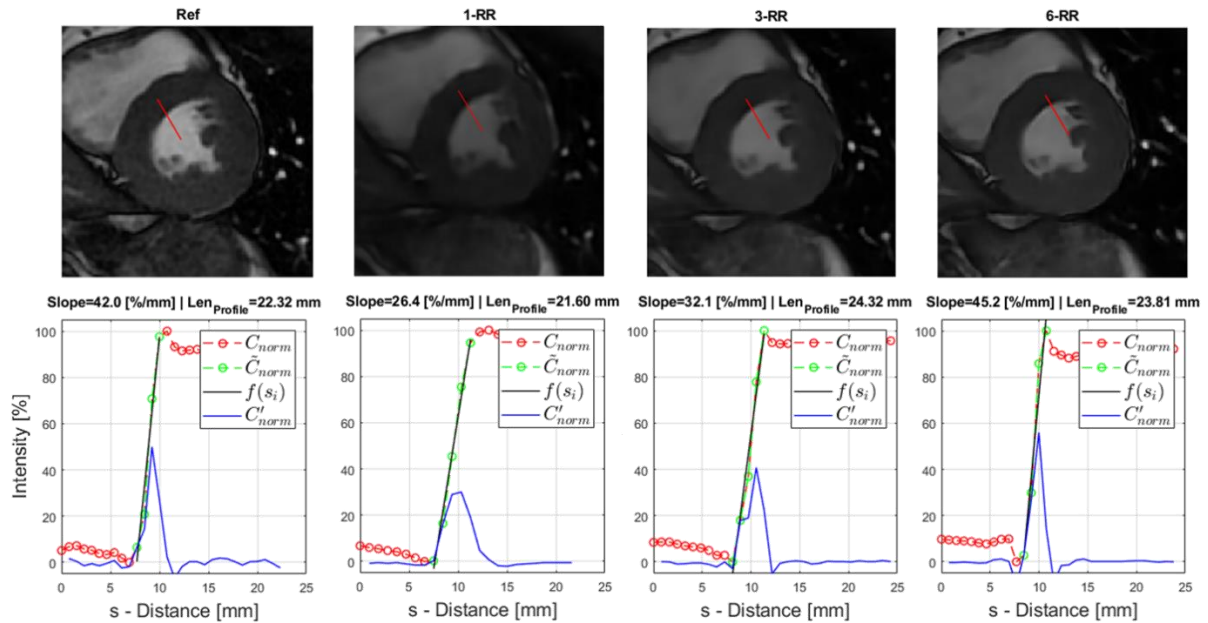


ABBILDUNG 5 – QUANTIFIZIERUNG DER KANTENSCHÄRFE.

(OBEN) AXIALE CINE-MR-BILDER DES HERZENS FÜR JEDEN SEQUENZTYP MIT Roter PROFILLINIE. (UNTEN) DIE DIAGRAMME ZEIGEN DAS NORMALISIERTE INTENSITÄTSPROFIL UND DIE REGRESSIONSLINIE ÜBER DEM ÜBERGANG VOM BLUT ZUM MYOKARD ZUR BESTIMMUNG DER KANTENSCHÄRFE. C_{norm} = NORMALISIERTES INTENSITÄTSPROFIL, $\tilde{C}_{norm}$ = INTENSITÄTSPROFIL, DAS ZUR BERECHNUNG DER KANTENSCHÄRFE VERWENDET WIRD, C'_{norm} = ERSTE ABLEITUNG VON C_{norm} , $f(s)$ = LINEARE REGRESSION DES INTENSITÄTSPROFILS, DAS ZUR BERECHNUNG DER KANTENSCHÄRFE VERWENDET WIRD, $LEN_{profile}$ = LÄNGE DER PROFILLINIE VOM BLUTPOOL ZUM MYOKARD, 1RR = CINE-SEQUENZ MIT EINEM HERZZYKLUS, REF = REFERENZSEQUENZ, s = PROFILLÄNGE, 6RR = CINE-SEQUENZ MIT SECHS HERZZYKLEN, 3RR = CINE-SEQUENZ MIT DREI HERZZYKLEN.

Die Kantenschärfe kann hier als betragsmäßiger Signal-Intensitäts-Anstieg beim Übergang vom Myokardium zum Blutpool aufgefasst werden. Für die Messungen wurden semi-automatisch in allen Bildern der drei Sequenzen 1-RR, 3-RR und 6-RR sowie der Referenzsequenz Profillinien auf den Übergang vom Blutpool zum Myokardium gelegt und das Intensitätsprofil $C(s)$ entlang der Strecke s zwischen zwei zuvor definierten Endpunkten aufgenommen. Für die anschließende Berechnung der Kantenschärfe wurden mithilfe der ersten Ableitung $C'(s) = dC/ds$ alle Stellen $\tilde{C}(s)$ detektiert.

Um eine stabile und von den tatsächlichen Intensitäten unabhängige Erkennung der Abtastpunkte $\tilde{C}$ zu gewährleisten, wurde das Intensitätsprofil $C(s)$ in allen Bildern auf den Wertebereich $[0,100]$ normiert $C_{norm}(s)$.

Die normalisierte Profilkurve ist somit definiert als:

$$C_{norm}(s) := \frac{C(s) - \min(C(s))}{\max(C(s)) - \min(C(s))} \cdot 100, \forall s.$$

Mit dem empirischen Kriterium $C'_{norm}(s) > 8\%$ ließen sich so alle Positionen s bestimmen, bei denen es sich um einen Übergang von Blutpool zu Myokardium handelt.

Über die so gefundenen Stellen $\tilde{C}(s)$ wurde mithilfe einer eindimensionalen linearen Regression die Regressionsgerade $f(s) = as + b + \varepsilon_i$ bestimmt. Der Koeffizient a (Geradenanstieg) wird hier als Kantenschärfe definiert. Das Absolutglied b der Regressionsgeraden, sowie die Regressionsfehler ε_i sind für die weitere Betrachtung nicht relevant. Die Modellentwicklung und Bildanalysen wurden mit kommerzieller Software (MATLAB®, Version R2022a Update 4, Mathworks Inc., Massachusetts, USA) implementiert.

Nachbearbeitung

Mit kommerzieller Software (cvi42, Version 5.16, Circle Cardiovascular Imaging Inc., Kanada) wurden EDV, ESV, SV, EF sowie die LV-Masse für LV und RV bestimmt und auf die Körperoberfläche normiert. Die automatische Segmentierung der endo- und epikardialen Konturen (inkl. Papillarmuskeln) wurde durch einen erfahrenen Radiologen gemäß Böttcher et al. [7] überprüft und ggf. korrigiert. Dabei wurden Korrekturzeit, Anzahl korrigierter Schnitte und Phasenadjustierungen dokumentiert, um den potenziellen Mehraufwand durch manuelle Nachbearbeitung zu bewerten.

Statistische Analyse

Die Teilnehmermerkmale und Bildgebungsparameter wurden als Median innerhalb der Wertebereichsgrenzen dargestellt. Unterschiede zwischen den DL-basierten Cine-Sequenzen und der Referenzsequenz wurden mit dem Wilcoxon-Vorzeichen-Rang-Test unter Anwendung eines Bonferroni-korrigierten Signifikanzniveaus von $p < 0,0167$ ($0,05/3$) analysiert, um Mehrfachtests zu berücksichtigen. Die Inter- und Intrareader-Reproduzierbarkeit wurde für eine Teilstichprobe von 20 Personen bestimmt und mithilfe des Intraklassen-Korrelationskoeffizienten (ICC) mit 95 % Konfidenzintervall (CI) basierend auf einem zweiseitigen Zufallseffekte-Modell berechnet (siehe **Definition 1 in Zusatzmaterial**).

2.3 ENTWICKLUNG EINES DL-BASIERTEN KARDIALEN SEGMENTIERUNGSVERFAHRENS

Ausgehend von der zuvor definierten Kohorte 2 (**Abschnitt 2.1**) erfolgte die systematische Aufbereitung der MRT-Bilddaten zur Entwicklung und Evaluierung des DL-basierten Segmentierungsmodells. Dies umfasste die strukturierte Erstellung des Referenzdatensatzes durch erfahrene Radiologen sowie die standardisierte Vorverarbeitung der T1-Maps.

2.3.1 DATENPARTITIONEN UND MODELLARCHITEKTUR

Im ersten Schritt wurden MRT-Cine-Aufnahmen von 50 gesunden Freiwilligen (Probandengruppe II) in zwei Durchgängen erfasst und die vom Scanner generierten nativen T1-Maps exportiert. Der Datensatz umfasste 300 SAX- und 100 4Ch-Orientierungen.

Im zweiten Schritt wurden zusätzlich Daten von 75 Patienten (Patientengruppe) extrahiert, darunter 225 SAX-Bilder, die aus klinischen Systemen gewonnen wurden. Diese wurden unabhängig voneinander von zwei Radiologen (R1 und R2) segmentiert:

- R1 segmentierte alle Bilder der Probandengruppe 2.
- R2 segmentierte alle Bilder der Patientengruppe.

Eine Teilmenge von 30 Patienten (90 Bilder) wurde von beiden Radiologen segmentiert, um die Variabilität zwischen den Beurteilern zu bewerten. Die manuelle Segmentierung erfolgte mit ITK-Snap (v4.0.0, Open Source).

Der von den beiden Radiologen annotierte und segmentierte Datensatz wurde in Trainings-, Validierungs- und Testdatensätze aufgeteilt wie in **Tabelle 2** zu sehen ist. Der Testdatensatz blieb für die Modellvalidierung unberührt. Jede dieser Teilmengen enthielt vier verschiedene Schnitte Vierkammerblick (4Ch), apikal (API), mittventrikulär (MID) und basal (BAS).

TABELLE 2 – ÜBERSICHT ÜBER DATENPARTITIONEN.

DER VON ZWEI BEURTEILERN GESAMMELTE UND ANNOTIERTE KOMPLEMENTÄRE DATENSATZ IST IN EINEN TRAININGS-, EINEN VALIDIERUNGS- UND EINEN TESTDATENSATZ UNTERTEILT. JEDE DIESER TEILMENGEN ENTHÄLT VIER VERSCHIEDENE ANSICHTEN (VIERKAMMERANSICHT, APIKALE, MITTELVENTRIKULÄRE UND BASALE ANSICHT).

Cohort \ View	All	4Ch	API	MID	BAS
<i>Training</i>	500 (80%)	79 (12.64%)	138 (22.08%)	146 (23.36%)	137 (21.92%)
<i>Validation</i>	62 (10%)	8 (1.28%)	17 (2.72%)	14 (2.24%)	23 (3.68%)
<i>Test</i>	63 (10%)	13 (2.08%)	20 (3.20%)	15 (2.40%)	15 (2.40%)

Im ersten Vorverarbeitungsschritt wurde die Bildauflösung der MRT-Daten von 256×218 Pixeln (Pixelgröße: $1,41 \times 1,41 \text{ mm}^2$) auf ein einheitliches Format von 512×512 Pixeln (Pixelgröße: $0,70 \times 0,60 \text{ mm}^2$) skaliert. Diese Skalierung dient der praktischen Umsetzung, da sie die Daten in ein standardisiertes Eingabeformat überführt und eine Unabhängigkeit von Bildgrößen und Pixelabständen ermöglicht.

Es wurde zunächst erwogen, ein Bild-Padding anzuwenden, um Verzerrungen zu vermeiden und quadratische Pixel beizubehalten. Diese Methode führte jedoch zu einer reduzierten Segmentierungsgenauigkeit und wurde daher nicht weiterverfolgt.

Um sicherzustellen, dass der verbleibende Interpolationsfehler für jede einzelne Segmentierungsmaske möglichst gering bleibt, wurden verschiedene Skalierungsverfahren – bilineare Interpolation, Lanczos, bikubische Interpolation, Pixel-Area-Relation und Nearest-Neighbor-Interpolation – evaluiert und ihr Einfluss auf den T1-Fehler ($\Delta T1$) untersucht [51–53]. Zu diesem Zweck wurden die mittleren T1-Zeiten zwischen den Originalbildern (Auflösung: 256×218 Pixel) und den skalierten Bildern (Auflösung: 512×512 Pixel) verglichen.

Die Nearest-Neighbor-Interpolation führte aufgrund ihrer werteerhaltenden Eigenschaft zu den geringsten Interpolations-bedingten Fehlern.

Zusätzlich wurden keine weiteren Datenaugmentationstechniken zur Erhöhung der Datenmenge eingesetzt. Es kamen Graustufenbilder ohne Intensitätsnormalisierung zum Einsatz.

Netzwerkarchitektur

Das verwendete Deep-Learning Modell basiert auf einer U-Net-Architektur, einer CNN-Architektur, die speziell für Bildsegmentierungsaufgaben entwickelt wurde. Die Grundidee hinter U-Net ist es, Features aus einem Bild im Encoder-Pfad zu extrahieren und diese dann zu verwenden, um eine pixelweise Vorhersage im Decoder-Pfad zu treffen.

In dem in dieser Arbeit entwickelten Algorithmus werden als Eingangsdaten monochrome DICOM-Bilder mit einer Auflösung von 512×512 Pixel und einem Grauwertbereich von 0 bis 2^{16} erwartet (**Abbildung 6**).

Der **Encoder** besteht aus einer Abfolge von Faltungsschichten (Convolutional Layers) gefolgt von Max-Pooling-Schichten. Jeder Schritt im Encoder reduziert die räumliche Dimension des Bildes (durch Max-Pooling) und erhöht die Anzahl der Kanäle durch Faltungsschichten.

Der **Bottleneck**-Block folgt dem Encoder und verbindet ihn mit dem Decoder-Block. Hier werden tiefere Features mit der höchsten Kanalanzahl (512) verarbeitet, während die räumliche Dimension am kleinsten ist (32×32 , wenn Input 512×512).

Der **Decoder** rekonstruiert die Segmentierung, indem die räumliche Dimension durch transponierte Faltung wiederhergestellt wird. In jedem Schritt werden die Ausgaben der transponierten Faltung mit den entsprechenden Encoder-Ausgaben (Skip-Connections) kombiniert. Ziel ist es, die detaillierten Informationen aus den Encoder-Schichten mit den rekonstruierten Features zu verbinden.

Faltungsschichten wurden sowohl im Encoder- als auch im Decoder-Teil verwendet, gefolgt von Rectified Linear Unit (ReLU)- oder LeakyReLU-Aktivierungsfunktionen. Die letzte Schicht hatte nur einen Ausgabekanal, auf den die Sigmoid-Aktivierungsfunktion angewendet wurde, um die Pixelklassenwahrscheinlichkeiten zu generieren, wodurch binäre Segmentierungsmasken erstellt werden konnten.

Die Auswirkungen zweier Aktivierungsfunktionen wurden während der Bewertung der Modellleistung verglichen, wie in einer früheren Arbeit [54] vorgeschlagen. Bei der Standardaktivierungsfunktion ReLU entspricht die Ausgabe des Neurons dem Aktivierungsgrad, wenn dieser im positiven Wertebereich liegt. Die Ausgabe eines Neurons kann also mit zunehmender Aktivierung beliebig ansteigen. Bei Aktivierungen $x < 0$ bleibt das

Neuron deaktiviert. Außerdem wurde die LeakyReLU-Aktivierungsfunktion, eine Variante der ReLU verwendet. Anders als bei der klassischen ReLU gewichtet sie Informationen mit einem zusätzlichen Skalierungsfaktor $\alpha=0,1$ für $x<0$.

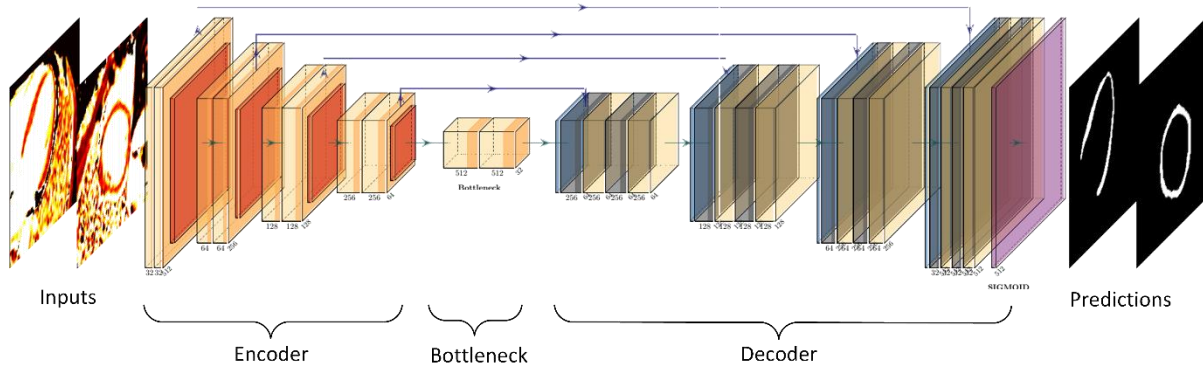


ABBILDUNG 6 – U-NET ARCHITEKTURMODELL FÜR BILDSEGMENTIERUNG.

DIE GRAFIK ZEIGT DIE ART DER EINGABEBILDER (INPUTS) UND AUSGABEN DER SEGMENTIERUNGSMASKEN (PREDICTIONS) SOWIE DIE DREI HAUPTKOMPONENTEN DER U-NET NETZWERKARCHITEKTUR – ENCODER, BOTTLENECK UND DECODER – SOWIE DIE ZWISCHENSCHRITTE MIT VERBINDUNGSPFEILEN, WELCHE DIE SKIP-CONNECTIONS REPRÄSENTIEREN.

Training

Zum Trainieren des Modells wurden kardiale T1-Maps des linken Ventrikels in Kurzachsenansicht (SAX) und Vierkammeransicht (4Ch) mit den entsprechenden Segmentierungsmasken als Eingabe verwendet. Das Modelltraining wurde auf einer Workstation (CPU i7-12700 2,10 GHz / 32 GB RAM; NVIDIA RTX A2000 / 6 GB GRAM) mit CUDA 11.6, Python Tensorflow 2.12.0, Numpy 1.23.5, PyDICOM 2.4.2 und Medpy 0.3.0 durchgeführt. Während der Trainingsschleife wurde die Leistung jedes Modells anhand der binären Genauigkeit (pixelweise Genauigkeit) und einer der beiden Metriken Sørensen-Dice-Koeffizient (DSC) oder Intersection over Union (IOU) bewertet.

Insbesondere wurden in der einschlägigen Literatur [54–56] beide Metriken $DSC(S_1, S_2) = 2|S_1 \cap S_2| / |S_1| + |S_2|$ und $IOU(S_1, S_2) = |S_1 \cap S_2| / |S_1 \cup S_2|$ vorgeschlagen, um die Übereinstimmung zweier Pixelsätze S_1 und S_2 [53] zu messen.

Der Trainingsdatensatz wurde in Batches zu je 16 aufgeteilt. Eine Trainingsschleife von 5 bis 70 Trainingsepochen wurde in Schritten von 5 Epochen unter Verwendung des ADAM-Optimierers [57] in Kombination mit variablen Aktivierungsebenen und einer abnehmenden Lernrate implementiert.

2.3.2 VALIDIERUNG UND VERGLEICH ZUR MANUELLEN SEGMENTIERUNG

Tests und Leistungsbewertung

Die mittleren T1-Zeiten der ROIs für alle trainierten Modelle wurden anhand des zuvor ungesehenen Testdatensatzes, bestehend aus 63 Bildern in verschiedenen Ansichten (4Ch und SAX), berechnet. Darüber hinaus wurden die absoluten mittleren Abweichungen $\Delta T1$ zwischen den vorhergesagten Segmentierungen und die Referenz-Segmentierungen sowie die Metriken DSC und IOU bestimmt. Das beste Modell wurde als das Modell betrachtet, das den Mittelwert aller Fehler $\overline{\Delta T1}$ über alle Bilder des Testdatensatzes minimiert und beide Metriken der Segmentierungsgenauigkeit ($\overline{DSC}$ und $\overline{IOU}$) über alle Bilder des Testdatensatzes maximiert. Zu diesem Zweck wird in der nachfolgenden **Gleichung (1)** eine Funktion $D(\Delta T1, DSC, IOU)$ definiert, die für alle 56 Modelle und alle 63 Bilder aus dem Testdatensatz berechnet wurde. Die arithmetischen Mittelwerte von $\overline{\Delta T1}$, $\overline{DSC}$ und $\overline{IOU}$ wurden über alle 63 absoluten Werte für jedes Modell berechnet. Da $\overline{\Delta T1}$ möglichst klein und $\overline{DSC}$ und $\overline{IOU}$ möglichst groß sein sollten, wurde anschließend der kleinste Mittelwert $\overline{\Delta T1}_j$ sowie die größten Mittelwerte $\overline{DSC}_j$ und $\overline{IOU}_j$ gesucht und das Minimum der Funktion D mit der Methode der kleinsten Quadrate bestimmt. Wir haben das leistungsfähigste Modell als dasjenige bestimmt, das die Funktion D_j mit der Definition minimiert

$$D_j^2 := \left(\overline{\Delta T1}_j - \min(\overline{\Delta T1}_j) \right)^2 + \left(\overline{DSC}_j - \max(\overline{DSC}_j) \right)^2 + \left(\overline{IOU}_j - \max(\overline{IOU}_j) \right)^2, \quad (1)$$

wobei $j = 1, \dots, 56$ die Modellnummer ist.

Im zweiten Schritt wurde das identifizierte beste Modell mit den Referenz-Segmentierungen des ersten Beurteilers (R1) und des zweiten Beurteilers (R2) sowie der Interrater-Variabilität verglichen.

Statistische Analyse

Die Genauigkeit der Myokard-Segmentierung wurde ebenfalls anhand der Vergleichsparameter $\Delta T1$, DSC und der IOU bewertet. Zur Analyse der Übereinstimmung zwischen Modellergebnissen wurde eine Bland-Altman-Analyse durchgeführt.

Zur Bewertung der Modellleistung sowie der Interrater-Variabilität wurden die Unterschiede in $\Delta T1$, DSC und IOU zwischen ‚Modell vs. Mensch‘ und ‚Mensch vs. Mensch‘ (Interrater-Variabilität) mittels des nichtparametrischen Mann-Whitney-U-Tests analysiert.

Ein p-Wert $< 0,05$ wurde als statistisch signifikant angesehen.

Die statistischen Analysen wurden mit Python (Version 3.11.2) unter Verwendung der Bibliotheken Scipy (1.10.1) und Seaborn (0.12.2) durchgeführt.

2.4 UNTERSUCHUNG DER REPRODUZIERBARKEIT VON RADIOMICS-FEATURES

Alle Studienteilnehmer absolvierten am selben Tag zwei MRT-Untersuchungen mit einer Pause von mindestens 20 Minuten zwischen den Scans. Die Untersuchungen wurden mit einem 1,5-T-MRT-Scanner (Avanto^{fit}, Siemens Healthineers, Erlangen, Deutschland) durchgeführt.

2.4.1 BILDVERARBEITUNG UND RADIOMICS-EXTRAKTION

Das MRT-Protokoll beinhaltete natives T1- und T2-Mapping in SAX- und 4Ch-Ansichten.

Für das T1-Mapping kam standardmäßig eine Modified Look-Locker Inversion Recovery (MOLLI) 5(3)3-Sequenz zum Einsatz, die in zwei verschiedenen räumlichen Auflösungen aufgenommen wurde:

- *Standardauflösung (SR)*: $1,9 \times 1,9 \text{ mm}^2$ Pixelgröße
- *Hohe Auflösung (HR)*: $1,4 \times 1,4 \text{ mm}^2$ Pixelgröße

Weitere Einzelheiten zum Bildgebungsprotokoll wurden bereits veröffentlicht [50, 58].

Die manuelle Myokard-Segmentierung des linken Ventrikels erfolgte mit ITK-Snap (v4.0.0, Open Source) durch einen Radiologen auf allen Mapping-Daten.

Anschließend wurden 100 Radiomics-Features aus dem linken Ventrikel extrahiert.

Die Radiomics-Extraktion erfolgte mit den Open-Source-Python-Paketen Pyradiomics (Version 3.0.1), SimpleITK (Version 2.0.2) und NumPy (Version 1.21.5).

Alle extrahierten Merkmale wurden analysiert und sieben Merkmalsklassen zugeordnet, die jeweils unterschiedliche Eigenschaften der Segmentierungen adressieren:

1. *First-Order Features (1st-Order)*:

Beschreiben die statistische Verteilung der Grauwerte innerhalb der ROI, z. B. Mittelwert, Standardabweichung oder Schiefe – unabhängig von räumlicher Ausrichtung.

2. *2D Shape Features*:

Erfassen die geometrische Form und Größe der ROI, z. B. Fläche, Umfang oder Kompaktheit – unabhängig von Intensitätswerten.

3. *Gray Level Co-Occurrence Matrix Features (GLCM)*:

Analysieren die räumliche Beziehung zwischen Pixelpaaren mit bestimmten Grauwertkombinationen und liefern Informationen zu Texturmustern wie Homogenität, Kontrast oder Korrelation.

4. *Gray Level Size Zone Matrix Features (GLSZM):*

Quantifizieren zusammenhängende Zonen gleicher Grauwerte innerhalb der ROI, unabhängig von ihrer Orientierung, und beschreiben z. B. die Variabilität oder Größe homogener Regionen.

5. *Gray Level Run Length Matrix Features (GLRLM):*

Erfassen aufeinanderfolgende Pixel mit identischen Grauwerten in einer bestimmten Richtung und charakterisieren damit die Konsistenz und Richtungstextur der ROI.

6. *Neighboring Gray Tone Difference Matrix Features (NGTDM):*

Berücksichtigen die Differenz eines Pixels zum lokalen Mittelwert seiner Nachbarschaft und liefern Informationen zu lokaler Textur, wie z. B. Coarseness oder Busyness.

7. *Gray Level Dependence Matrix Features (GLDM):*

Beschreiben die Anzahl benachbarter Pixel mit ähnlichem Grauwert, wobei die Abhängigkeit eines Pixels von seiner Umgebung analysiert wird – dies gibt Aufschluss über die lokale Grauerthomogenität.

Eine detaillierte Auflistung aller untersuchten Radiomics-Features befindet sich in Anhang **Zusatztabelle 1**.

2.4.2 STATISTISCHE METHODEN ZUR ANALYSE DER REPRODUZIERBARKEIT

Bewertung der Reproduzierbarkeit von Test und Retest

Die Reliabilität zwischen Test- und Retest-Messungen wurde primär anhand des Intraklassen-Korrelationskoeffizienten (ICC) gemäß der Definition von Shrout und Fleiss [59] berechnet.

Im Gegensatz dazu misst der von Lin [61] entwickelte Konkordanz-Korrelationskoeffizient (CCC) die Übereinstimmung ohne Annahmen über ein ANOVA-Modell (siehe auch Definition 2 im Anhang). Er kombiniert die Pearson-Korrelation (Konsistenz) mit einem Maß für systematischen Bias und bewertet sowohl die Nähe der Daten zur Regressionslinie als auch deren Übereinstimmung mit der Idealgeraden [62].

Zur Validierung der Ergebnisse wurden der CCC sowie der Pearson-Korrelationskoeffizient r ergänzend herangezogen.

Um die Beziehungen zwischen diesen drei Reproduzierbarkeitsmaßen zu überprüfen, wurden sie grafisch gegeneinander aufgetragen (siehe auch Zusatzmaterial in [45]).

Für alle Radiomics-Features wurden die ICCs mit 95% CI unter Verwendung eines zweiseitigen Zufallseffekte-Modells ICC(2,1) für beide Messungen berechnet.

Die numerischen ICC-Werte wurden gemäß folgender Klassifikation interpretiert [63]:

- $< 0,50$: schlecht,
- $0,50 - 0,75$: mittelmäßig,
- $0,75 - 0,90$: gut,
- $> 0,90$: exzellent.

Visualisierung der Reproduzierbarkeit

Detaillierte Ergebnisse für alle Merkmale wurden tabellarisch dargestellt – separat für die Standardauflösung (SR) und die SAX-Ansicht, da diese als klinischer Standard gelten. Zusätzlich wurden die Ergebnisse für das T1- und T2-Mapping in der 4Ch-Ansicht sowie für das hochauflösende T1-Mapping (HR) analysiert.

Für jede Kombination aus Orientierung und Auflösung wurde die Anzahl der Merkmale mit exzellenter, guter, mäßiger und schlechter Reproduzierbarkeit in Balkendiagrammen für alle Merkmalsklassen visualisiert [45].

Zusätzlich wurde eine Teilmenge von Merkmalen mit guter bis exzellenter Reproduzierbarkeit unabhängig von Bildorientierung und räumlicher Auflösung identifiziert.

Ermittlung robuster Features

- Für das T1-Mapping wurde die Schnittmenge von Features mit guter bis exzellenter Reproduzierbarkeit ($ICC > 0,75$) in allen vier Sequenzvarianten bestimmt:
 - Kurzachsen- und Langachsenansicht (SAX und 4Ch)
 - Standard- und hohe räumliche Auflösung (4Ch/SR, 4Ch/HR, SAX/SR, SAX/HR)
- Für das T2-Mapping wurde die Schnittmenge der Merkmale mit $ICC > 0,75$ in den Kurzachsen- und Langachsenansichten bestimmt (4Ch/SR, SAX/SR).

Die Auswahl robuster Merkmale wurde in Heatmaps und Sunburst Diagrammen visualisiert und entsprechend ihrer ICC-Werte über alle Merkmalsklassen hinweg dargestellt.

Analyse geschlechts- und altersspezifischer Unterschiede

Für die als robust identifizierten Merkmale wurden geschlechts- und altersabhängige Unterschiede untersucht. Hierzu wurden die Mittelwerte der jeweiligen Merkmalsausprägungen über beide Messzeitpunkte gebildet und anschließend nach Geschlecht sowie Altersgruppen stratifiziert. Die statistische Analyse erfolgte mithilfe des nichtparametrischen Mann-Whitney-U-Tests (siehe [45]).

III. ERGEBNISSE

3.1 OPTIMIERUNG DER MRT-AUFNAHMEN DURCH KI

Aufnahmezeit

Die schnellste Erfassung wurde mit der 1RR-Cine-Sequenz erreicht (reine Sequenzlaufzeit: 19 Sekunden). Die klinische Standardsequenz wies die längste Laufzeit auf (Median: 84 s), gefolgt von der 6RR-Cine-Sequenz (Median: 68 s) und der 3RR-Cine-Sequenz (Median: 36 s) (**Tabelle 3**).

Da die 1RR-Cine-Sequenz in lediglich 1,9 Atempausen (Median) mit einer durchschnittlichen Atemanhaltdauer von 12,5 Sekunden abgeschlossen werden konnte, betrug die gesamte Erfassungszeit nur 45 Sekunden (Spannweite: 24–67 s). Dies entspricht einer Zeitersparnis von 80 % im Vergleich zur Referenzsequenz ($p = 0,005$).

Die 3RR-Cine-Sequenz hatte eine mittlere Gesamterfassungszeit (inkl. Pausen) von 109 Sekunden (Bereich: 65–132 s) und war damit über 50 % schneller als die klinische Referenzsequenz (228 s, Bereich: 168–275 s, $p = 0,005$).

Die 6RR-Cine-Sequenz zeigte eine Zeitersparnis von 20 % im Vergleich zur klinischen Cine, mit einer Gesamtaufnahmezeit von 183 Sekunden (Bereich: 158–220 s).

Eine Übersicht aller Ergebnisse findet sich in **Tabelle 3**.

TABELLE 3 – AUFNAHMEPARAMETER UND BILDQUALITÄT.

ALLE DEEP-LEARNING-BASIERTEN CINE-SEQUENZEN WURDEN MIT DER REFERENZSEQUENZ VERGLEICHEN. DIE P-WERTE WURDEN MIT DEM WILCOXON-VORZEICHEN-RANGSUMMENTEST ERMITTELT. BH = ATEM ANHALTEN; MD = MEDIAN

	Referenz		1RR Cine			3RR Cine			6RR Cine		
Aufnahme- parameter	Md	Range	Md	Range	p	Md	Range	p	Md	Range	p
<i>t</i> -Sequenz (ohne BH) (min)	01:24	01:07 - 01:40	00:19	00:13- 00:26	0.005	00:36	00:31 - 00:51	0.005	01:08	00:54- 01:25	0.005
<i>t</i> -Aufnahme (with BH) (min)	03:47	02:48 - 04:35	00:47	00:34- 01:07	0.005	01:48	01:05 - 02:12	0.005	03:04	02:38- 03:40	0.013
Anzahl BH	5.5		1.9		0.005	3.4		0.005	5.3		0.047
Dauer BH (s)	15.5		12.5		0.073	15		0.794	13.5		0.234
Bildqualität	Md	Range	Md	Range	p	Md	Range	p	Md	Range	p
subjektiv	4	3.5-5	3	1.5-4	<0.001	4	3.5-5	0.374	5	4-5	<0.001
Kanten-schärfe	30.6	15.8- 42.0	26.4	15.5- 34.8	<0.001	32.1	17.0- 45.8	0.062	37.4	16.4- 51.5	<0.001

3.1.1 AUSWIRKUNGEN AUF BILDQUALITÄT UND AUFNAHMEZEIT

Bildqualität

Je ein Beispielbild aller vier Aufnahmesequenzen (Referenz, 1RR-Cine, 3RR-Cine und 6RR-Cine) ist bereits in (**Abbildung 5**) dargestellt. Die subjektive Bildqualität wurde am höchsten

bei der 6RR-Cine-Sequenz bewertet (Median: 5, $p < 0,001$), während die 1RR-Cine-Sequenz die niedrigste Bewertung erhielt (Median: 3, $p < 0,001$).

Auch die normalisierte Kantenschärfe als objektiver Bildqualitätsparameter war bei der 6RR-Cine-Sequenz am höchsten (Median: 37,4) und bei der 1RR-Cine-Sequenz am niedrigsten (Median: 26,4).

Der Unterschied in der Kantenschärfe zwischen der Referenzsequenz (Median: 30,6) und der 3RR-Cine-Sequenz (Median: 32,1) war nicht statistisch signifikant (**Tabelle 3**).

Ergebnisse der volumetrischen Analyse

Die 1RR-Cine-Sequenz unterschätzte LV-EDV, LV-SV und LV-EF, während die LV-Masse signifikant überschätzt wurde (jeweils $p < 0,001$). Bei 3RR und 6RR zeigten sich keine signifikanten Unterschiede zur Referenz. Die Segmentierungsreproduzierbarkeit war hoch (Intrareader-ICC: 0,95–1,00; Interreader-ICC: 0,93–1,00).

Für den RV ergab sich ein vergleichbares Muster: nur 1RR wies moderate Abweichungen auf, während 3RR und 6RR der Referenz entsprachen.

Weitere Details zu Volumina, Reproduzierbarkeit und Zusatzmaterial finden sich in [43].

Nachbearbeitungszeit

Im Vergleich zur Referenzsequenz war die zur Korrektur der automatischen Segmentierung erforderliche Zeit bei der 1RR-Cine-Sequenz signifikant länger (45 s gegenüber 25,5 s, $p < 0,001$). Für die 3RR- und 6RR-Cine-Sequenzen hingegen zeigte sich kein signifikanter Unterschied.

Die Unterschiede im Nachbearbeitungsaufwand – einschließlich Korrekturzeit, Anzahl korrigierter Schnitte und Phasenkorrekturen – sind in Arbeit [43] zusammengefasst.

3.2 AUTOMATISIERTE SEGMENTIERUNG DES MYOKARDS BEIM T1-MAPPING

Als Ergebnis der Bildvorverarbeitung betrug der maximale absolute Interpolationsfehler $\Delta T1$ nach der Skalierung der Bildauflösung von 256×218 auf 512×512 Pixel ± 4 ms. Dieser Wert wurde als klinisch irrelevant eingestuft, da die kardialen T1-Relaxationszeiten typischerweise zwischen 900 und 1200 ms liegen und die Übereinstimmungsgrenzen für die mittlere Abweichung etwa -39 bis $+34$ ms betragen [50].

3.2.1 AUSWAHL DES OPTIMALEN MODELLS BASIEREND AUF DER MODELLELEISTUNG

Wie in Abschnitt 2.2 beschrieben, erfolgte die Auswahl des optimalen Modells auf Basis einer Strategie zur Minimierung des quantitativen Fehlers in den T1-Werten ($\Delta T1$) sowie zur

Maximierung der Segmentierungsgenauigkeit, gemessen anhand von DSC und IoU. Die Leistungsentwicklung aller vier neuronalen Netzwerke über die Trainingsepochen hinweg wurde in [44] detailliert analysiert. Dabei zeigten die Metriken eine Konvergenz mit zunehmender Epochenanzahl – insbesondere hinsichtlich der Segmentierungsgenauigkeit, wobei die Leistungsunterschiede zwischen 25 und 55 Epochen nur noch gering ausfielen [44].

3.2.2 GENAUIGKEIT DES DL-MODELLS

Basierend auf Gleichung (1) wurde [EP35DSCORE] als das Modell mit der besten Gesamtleistung in Bezug auf $\Delta T1$, DSC und IoU identifiziert (siehe **Tabelle 4**).

TABELLE 4 – MODELLVERGLEICH FÜR ALLE ANSICHTEN.

DIE TABELLE ZEIGT DIE AGGREGIERTEN METRIKEN UND DIE RELAXATIONSZEITEN (TR) FÜR DIE ZEHN BESTEN MODELLE FÜR ALLE 63 TESTBILDER IN ALLEN ANSICHTEN (4Ch UND SAX). IN DIESEM ZUSAMMENHANG BEDEUTET „AM BESTEN“ EINE MINIMUM DER GEMITTELTEN ABSOLUTEN DIFFERENZEN ($\Delta T1$) UND DEN GEMITTELTEN MAXIMALEN METRIKEN.

<i>Model</i>	$\overline{\Delta T1}$	$\overline{DSC}$	$\overline{IOU}$	<i>D</i>
<i>EP35DSCORE</i>	10.571 (17.891)	0.879 (0.069)	0.790 (0.088)	0.013
<i>EP50IOULR</i>	10.714 (15.839)	0.885 (0.049)	0.797 (0.071)	0.143
<i>EP55IOURE</i>	10.746 (14.612)	0.877 (0.080)	0.787 (0.096)	0.175
<i>EP45IOURE</i>	11.063 (18.483)	0.887 (0.061)	0.801 (0.082)	0.492
<i>EP25IOURE</i>	11.095 (18.278)	0.881 (0.064)	0.791 (0.083)	0.524
<i>EP50IOURE</i>	11.222 (17.111)	0.881 (0.073)	0.794 (0.091)	0.651
<i>EP45IOULR</i>	11.302 (17.875)	0.884 (0.060)	0.796 (0.080)	0.730
<i>EP70DSCORE</i>	11.317 (15.861)	0.881 (0.073)	0.793 (0.091)	0.746
<i>EP35IOULR</i>	11.333 (19.683)	0.885 (0.058)	0.798 (0.080)	0.762
<i>EP55DSCORE</i>	11.413 (16.531)	0.879 (0.073)	0.789 (0.092)	0.841

3.2.3 VERGLEICH ZUR MANUELLEN SEGMENTIERUNG

Die Übereinstimmungsgrenzen (IQR) zwischen den Modellergebnissen und den Referenzdaten lagen zwischen –35,5 und +36,1 ms und zeigten damit eine bessere Übereinstimmung als zwischen den beiden menschlichen Beurteilern (–34,7 bis +59,1 ms). Zudem war der Vorhersagefehler Bias des Modells (0,29 ms) deutlich geringer als der Interrater-Bias (–12,2 ms) [44].

Abbildung 7 zeigt die Vorhersageleistung des besten Modells (EP35DSCORE) im Vergleich zum Referenzdatensatz anhand einer Teilmenge des ungesehenen Testdatensatzes – dargestellt für verschiedene Bildorientierungen: 4Ch- und SAX sowie Kurzachsenschnitte (API, MID, BAS).

Die Segmentierungsgenauigkeit korrelierte mit der Interreader-Übereinstimmung: Beide waren bei API geringer und bei BAS sowie MID höher (siehe auch [44]).

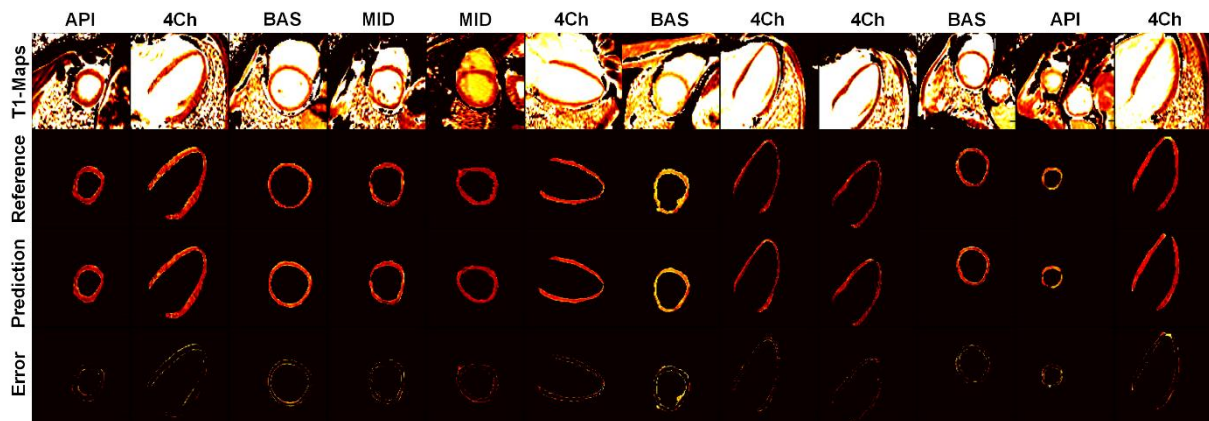


ABBILDUNG 7 – MODELLVORHERSAGEN VS. REFERENZ.

VERGLEICH ZWISCHEN REFERENZDATEN (GROUND-TRUTH-BINÄRMASKEN) UND VORHERSAGEN DES BESTEN MODELLS [EP35DSCORE] AUF BASIS DER EINGABEDATEN (T1-MAPS) FÜR EINE KLEINE TEILMENGE DES TESTDATENSATZES UND SEGMENTIERUNGSFEHLER (ERROR), DARGESTELLT MIT EINEM ZOOMFAKTOR VON 2,3.

Bezüglich $\Delta T1$ zeigte sich, dass der Fehler zwischen dem DL-Modell und der menschlichen Referenz signifikant geringer war als der Unterschied zwischen zwei menschlichen Beurteilern ($p = 0,013$), insbesondere bei API-Schnitten. Für MID- und BAS- Schnitte war dieser Unterschied nicht signifikant (**Tabelle 5**). Die geometrische Segmentierungsgenauigkeit zwischen Modell und Mensch war für alle SAX-Schnitte signifikant höher als die Übereinstimmung zwischen zwei menschlichen Segmentierungen.

TABELLE 5 – METRIKEN DES BESTEN MODELLS UND INTERRATER-VERGLEICH.

ALLE METRIKEN DES BESTEN MODELLS (PRED) WURDEN MIT INTERRATER-METRIKEN (R1vsR2) FÜR SAX-SCHNITTE VERGLEICHEN. P-WERTE STAMMEN AUS DEM MANN-WHITNEY-U-TEST.

Metriken		$\Delta T1$				DSC		IOU		
Schnitt	Maske	Md	Range	p	Md	Range	p	Md	Range	p
SAX	PRED	-4.5	-33 - 107	0.006	0.892	0.43 - 0.94	<0.001	0.806	0.27 - 0.89	<0.001
	R1vsR2	-11.0	-109 - 31		0.842	0.65 - 0.93		0.729	0.48 - 0.87	
API	PRED	-5.0	-33 - 107	0.013	0.863	0.43 - 0.92	0.003	0.759	0.27 - 0.86	0.003
	R1vsR2	-15.0	-89 - 27		0.799	0.67 - 0.88		0.666	0.51 - 0.79	
MID	PRED	-6.0	-25 - 28	0.11	0.893	0.85 - 0.93	<0.001	0.807	0.74 - 0.87	<0.001
	R1vsR2	-14.5	-109 - 20		0.842	0.65 - 0.92		0.728	0.48 - 0.85	
BAS	PRED	-3.0	-19 - 18	0.56	0.916	0.86 - 0.94	<0.001	0.845	0.75 - 0.89	<0.001
	R1vsR2	-5.5	-48 - 31		0.864	0.75 - 0.93		0.760	0.60 - 0.87	

3.3 REPRODUZIERBARKEIT VON RADIOMICS-FEATURES

In der Arbeit [45] wurden T1 und T2-Maps mit Standardauflösung (SR - $1,9 \times 1,9 \text{ mm}^2$) in Kurzachsenansicht (SAX) als klinischer Standard definiert.

3.3.1 ERGEBNISSE FÜR T1- UND T2-MAPPING

T1-Mapping: Es wurden deutliche Unterschiede in der Reproduzierbarkeit der Radiomics-Features festgestellt. Über alle Feature-Klassen hinweg zeigten 6 Merkmale eine exzellente, 29 eine gute, 19 eine mittelmäßige und 46 eine schlechte Reproduzierbarkeit.

Innerhalb jeder Feature-Klasse war die Reproduzierbarkeit einzelner Features sehr heterogen und reichte – je nach Feature – von schlecht bis gut oder exzellent.

Die gepaarten Abhängigkeiten zwischen den Maßen (ICC, CCC und r) weisen ebenfalls auf eine lineare Beziehung mit einem geringen Streuungsgrad hin (siehe [45]).

Auswahl der robustesten Radiomics-Features beim T1-Mapping

In dieser Analyse wurden 15 Merkmale aus sechs Feature-Klassen (**Abbildung 8**) identifiziert [45], die sowohl in unterschiedlichen räumlichen Auflösungen (SR und HR) als auch in SAX- und 4Ch-Ansichten eine gute bis exzellente Reproduzierbarkeit zeigten.

Eine detaillierte Auflistung dieser besonders robusten Features ist in [45] zu finden.

T2-Mapping: Auch beim T2-Mapping zeigten sich erhebliche Unterschiede in der Reproduzierbarkeit der analysierten Merkmale. Über alle Merkmalsklassen hinweg wurde die Reproduzierbarkeit wie folgt klassifiziert: 6 Merkmale mit exzellenter, 25 Merkmale mit guter, 23 Merkmale mit mittelmäßiger und 46 Merkmale mit schlechter Reproduzierbarkeit.

Innerhalb der meisten Merkmalsklassen reichte die Reproduzierbarkeit der einzelnen Features von schlecht bis gut oder exzellent.

Auswahl der robustesten Radiomics-Features beim T2-Mapping

Insgesamt wurden 12 Merkmale aus fünf Merkmalsklassen (**Abbildung 8**) identifiziert [45], die sowohl in Kurzachsen- als auch Langachsenansichten beim T2-Mapping eine gute bis exzellente Reproduzierbarkeit ($\text{ICC} > 0,75$) aufwiesen.

Schnittmenge zwischen den robustesten Radiomics-Features bei T1- und T2-Mapping

Sieben der 12 Features, die als am reproduzierbarsten im T2-Mapping identifiziert wurden, gehören auch zu den 15 robustesten Features im T1-Mapping.

Diese Überschneidung umfasst:

- drei Formmerkmale (2D-Shape): *MinorAxisLength*, *Maximum2DDiameterSlice*, *Maximum2DDiameterColumn*
- drei GLRLM-Merkmale: *RunPercentage*, *RunLengthNonUniformityNormalized*, *GrayLevelNonUniformity*
- ein GLDM-Merkmal: *GrayLevelNonUniformity*

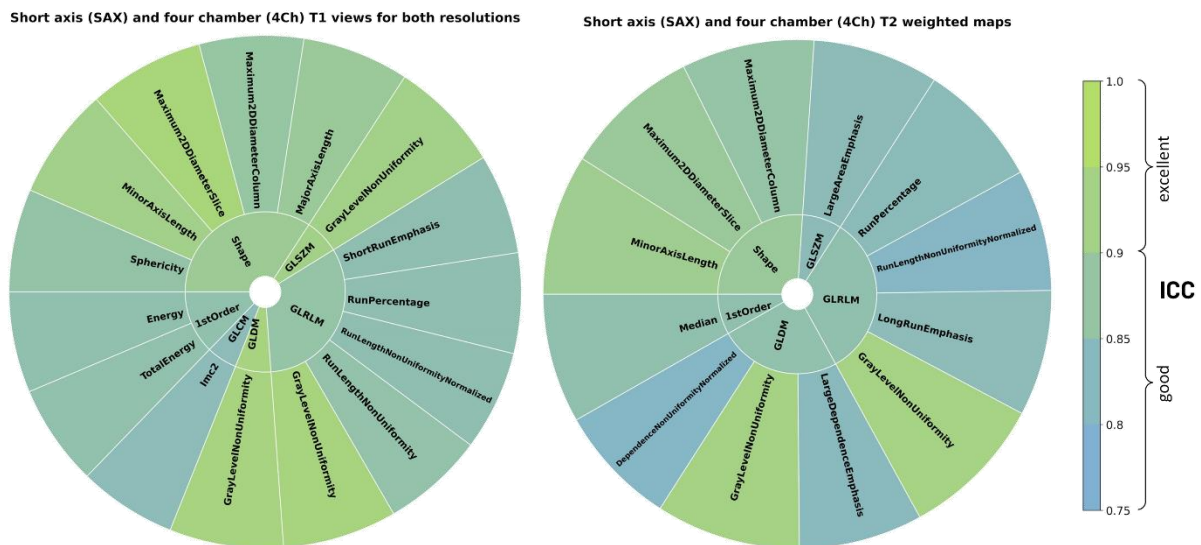


ABBILDUNG 8 – EMPFOHLENE FEATURES FÜR ALLE T1- UND T2-MAPS.

DARGESTELLT SIND DIE SCHNITTMENGEN DER GEMITTELTEN ICC-WERTE FÜR DIEJENIGEN FEATURES, DIE EINE GUTE BIS EXZELLENT REPRODUZIERBARKEIT IN ALLEN UNTERSUCHTEN ANSICHTEN (4Ch UND SAX) UND RÄUMLICHEN AUFLÖSUNGEN (SR UND HR) ZEIGTEN – FÜR T1 (LINKS) UND T2 (RECHTS). DIE FARBSKALA AUF DER RECHTEN SEITE GIBT DIE ROBUSTHEIT DER MERKMALE AN: GRÜN STEHT FÜR EXZELLENT, BLAU FÜR GUTE REPRODUZIERBARKEIT (SIEHE COLORMAP).

Alters- und geschlechtsspezifische Unterschiede

Alle Features mit Ausnahme der *Sphericity* wiesen signifikante Geschlechtsunterschiede beim T1-Mapping auf. Altersabhängige Unterschiede wurden hingegen nicht beobachtet, mit Ausnahme von zwei Formmerkmalen, die möglicherweise durch den höheren Anteil weiblicher Teilnehmer in der älteren Altersgruppe beeinflusst wurden. Eine vollständige Analyse der gruppenspezifischen Unterschiede ist in [45] dokumentiert.

3.3.2 EINFLUSS VON AUFLÖSUNG UND BILDORIENTIERUNG

Unabhängig von Bildorientierung und Auflösung waren die 2D-Shape-Features diejenigen Merkmale mit der höchsten Reproduzierbarkeit: Mindestens 55,5 % der Features dieser Klasse erreichten einen $ICC > 0,75$. NGTDM- und GLCM-Features waren am wenigsten robust, mit nur 0–20 % bzw. 9,1–18,2 % reproduzierbaren Features.

Für das **T1-Mapping** ergab sich eine insgesamt höhere Reproduzierbarkeit in SAX-Ansichten im Vergleich zu 4Ch. In SAX wiesen 32 Features aus sieben Klassen eine gute bis exzellente Reproduzierbarkeit auf, in 4Ch lediglich 16 Features aus sechs Klassen. Unter Berücksichtigung der Auflösung zeigten sich bei Standardauflösung 35 reproduzierbare Features in SAX und 26 in 4Ch, bei hoher Auflösung 40 in SAX und 22 in 4Ch.

Auch im **T2-Mapping** waren die 2D-Shape-Features am robustesten: In SAX-Ansichten erfüllten 77,7 %, in 4Ch-Ansichten 33,3 % der Shape-Features das Kriterium guter bis exzellenter Reproduzierbarkeit. Die übrigen Feature-Klassen zeigten deutlich geringere Werte: Maximal 50 % der Features in SAX und 25 % in 4Ch waren reproduzierbar. Kein GLCM-Feature im T2-Mapping erreichte auch nur eine mittelmäßige Reproduzierbarkeit.

Zusammenfassend war die Reproduzierbarkeit der untersuchten Radiomics-Features für SAX sowohl im T1- als auch im T2-Mapping durchgängig besser als für 4Ch. Beim T2-Mapping wiesen 31 Features aus sechs Klassen in SAX eine gute bis exzellente Reproduzierbarkeit auf, während es in 4Ch lediglich 14 Features aus fünf Klassen waren.

Insgesamt zeigten 2D-Shape- und GLRLM-Features die höchste Reproduzierbarkeit, insbesondere in SAX-Ansichten. Die Orientierung hatte größeren Einfluss auf die Robustheit als die räumliche Auflösung.

IV. DISKUSSION

4.1 BEWERTUNG DER ERGEBNISSE

Die drei durchgeführten Studien decken zentrale Schritte im Workflow der kardialen MRT ab – von der Akquisition über die Nachverarbeitung bis zur Analyse. DL-basierte Bildrekonstruktion reduzierte die Erfassungszeit erheblich, vollautomatische Segmentierung erwies sich als präzise und konsistent, und es konnten robuste Radiomics-Features mit hoher Reproduzierbarkeit identifiziert werden. Diese Ergebnisse unterstreichen das Potenzial KI-gestützter Verfahren zur Effizienzsteigerung und Standardisierung in der kardiovaskulären MRT-Bildgebung.

4.1.1 NUTZEN DER KI-GESTÜTZTEN MRT-BESCHLEUNIGUNG UND KLINISCHE RELEVANZ

Die klinisch relevante Zeitersparnis der 3RR-Sequenz könnte insbesondere für fragile Patientengruppen einen echten Fortschritt darstellen.

Die ultraschnelle 1RR-Cine-Sequenz ermöglichte eine Reduktion der Aufnahmezeit um bis zu 80 %, ging jedoch mit einer geringeren Bildqualität und leichten Abweichungen der volumetrischen und diagnostischen Parameter einher. Diese Sequenz könnte insbesondere bei Patienten mit Herzrhythmusstörungen oder solchen, die keine Atemanhaltephasen tolerieren, eine sinnvolle Alternative darstellen [64].

Die 6RR-Cine-Sequenz lieferte die beste Bildqualität aller getesteten Sequenzen und übertraf den klinischen Standard – bei gleichzeitiger Verkürzung der Erfassungszeit. Die beobachtete Verbesserung der Bildqualität dürfte auf die erhöhte Kantenschärfe und bessere räumliche Auflösung zurückzuführen sein, was bei komplexen diagnostischen Fragestellungen wie der Klappenbeurteilung oder der Analyse der myokardialen Dehnung von Vorteil sein kann.

Diese Ergebnisse stehen im Einklang mit den Beobachtungen von Chen et al. [65], die keine Unterschiede in den volumetrischen Parametern zwischen DL-rekonstruierten und konventionellen Cine-Sequenzen fanden, bei gleichzeitig besserer Bildqualität der DL-Varianten.

4.1.2 STÄRKEN UND SCHWÄCHEN DER AUTOMATISIERTEN SEGMENTIERUNG

Die vollautomatische Segmentierung in der kardialen MRT wurde bereits in zahlreichen Studien untersucht – allerdings primär im Kontext von Cine-Sequenzen zur Analyse ventrikulärer und atrialer Funktionen. Im Gegensatz dazu liegt bisher nur eine begrenzte Zahl an Arbeiten zur Segmentierung auf nativen T1-Maps vor.

Im Gegensatz zu unserer Analyse beschränkten sich Howard et al. auf Kurzachsenansichten und verwendeten eine Segmentanalyse auf Basis des 16-Segment-Modells der AHA [19]. Wie in unserer Studie war die Übereinstimmung bei der Segmentierung zwischen dem neuronalen Netzwerk und menschlichen Experten (DSC: 0,82 - 0,86) vergleichbar mit der Übereinstimmung zwischen zwei menschlichen Experten (DSC: 0,84).

Bhatt et al. entwickelten einen Deep-Learning-Algorithmus unter Verwendung synthetischer Kontrastverstärkung und erreichten eine hohe Genauigkeit bei der Myokardsegmentierung (DSC: 0,81) und T1-Berechnung ($R = 0,87$) [15]. Trotz leicht verbesserter Genauigkeit führte dieser Ansatz zu erhöhter Komplexität der Netzwerkarchitektur und einem gesteigerten Rechenaufwand. Unser Ziel war es daher, ein Modell zu entwickeln, das mit rohen T1-Maps arbeitet – ohne künstliche Kontrastverstärkung [44].

Ein wesentlicher Unterschied zu bisherigen Studien besteht darin, dass unsere Arbeit sowohl Kurzachsen- (SAX) als auch Langachsenansichten (4Ch) berücksichtigt.

Unser bestes Modell erzielte in 4Ch-Ansichten eine vergleichbare Genauigkeit wie in SAX-Ansichten (DSC: $0,89 \pm 0,03$ vs. $0,88 \pm 0,08$). Zudem war die T1-Quantifizierung in 4Ch insgesamt präziser ($\Delta T1$: $6,77 \pm 8,3$ ms vs. $11,6 \pm 19,7$ ms). Der Unterschied resultierte vor allem aus den apikal gelegenen SAX-Schnitten. Dort führt das sehr dünne Myokard dazu, dass bereits geringe Segmentierungsfehler große Auswirkungen auf die berechneten T1-Werte haben können.

Klinische Relevanz

In der heutigen klinischen Praxis werden T1-Maps oft analysiert, indem eine ROI im interventrikulären Septum manuell gezeichnet wird. Dies führt zu einer erheblichen Interrater-Variabilität [50]. Die klinische Verwendung des T1-Mapping in der Herz-MRT ist sehr variabel. Am häufigsten werden ein oder drei SAX-Schnitte aufgenommen, obwohl einige Zentren LAX-Schnitte oder beides aufnehmen. Hier schlagen wir einen auf Deep Learning basierenden Algorithmus vor, der das Myokard auf T1-Karten segmentieren und mittlere quantitative Ergebnisse für die myokardiale T1-Relaxationszeit ohne menschliche Interaktion liefern kann. Der Algorithmus ist für die Integration in den klinischen Arbeitsablauf von Spezialisten für kardiovaskuläre Bildgebung vorgesehen. Die Integration des Algorithmus entweder in die Auswertungssoftware oder in eine Nachbearbeitungssoftware mit bidirektionalem Zugriff auf das Picture Archiving and Communication System (PACS) könnte sicherstellen, dass die Ergebnisse der T1-Analyse direkt an das PACS gesendet werden können und bereit sind, bevor die Ärzte überhaupt mit dem Lesen der Studie beginnen.

4.1.3 BEDEUTUNG DER REPRODUZIERBARKEIT VON RADIOMICS-FEATURES

Radiomics-Analysen in der kardialen MRT sind Gegenstand zahlreicher Studien, wobei der Schwerpunkt bislang meist auf diagnostischen und prognostischen Fragestellungen lag. Die Reproduzierbarkeit der extrahierten Merkmale wurde hingegen bislang nur in wenigen Arbeiten thematisiert [37, 66].

In einer früheren Studie [66] wurde die Reproduzierbarkeit von Radiomics-Merkmalen im quantitativen T1- und T2-Mapping bei 6 gesunden Teilnehmern und 18 Patienten für die Kurzachsenansicht und eine Auflösung untersucht.

Mit Ausnahme des 90. Perzentils zeigten 8/9 Radiomics-Features, die in ihrer Studie für das T1-Mapping als am reproduzierbarsten ermittelt wurden, auch in unserer Kohorte [45] eine gute bis exzellente Reproduzierbarkeit in SAX und Standardauflösung. Für das T2-Mapping in SAX können wir auch bestätigen, dass die 4/4 Features, die in dieser vorherigen Studie als reproduzierbar ermittelt wurden, in unserer Kohorte eine gute bis exzellente Reproduzierbarkeit im T2-Mapping in der Kurzachsenansicht zeigten. Im Gegensatz zu dieser vorherigen Studie konnten wir zeigen, dass die Robustheit der Features je nach Auflösung (SR oder HR) und Ansicht variiert, was zu einer geringeren Reproduzierbarkeit für die 4Ch-Ansicht führte. Wir stellten fest, dass die Anzahl reproduzierbarer Features in Kurzachsenansicht höher war als in der Langachsenansicht (4-Kammer-Ansicht). Dies kann mit der Orientierung der Myokardmuskelfasern relativ zur Bildebene oder mit der Geometrie des kreisförmigen ROI bei der SAX-Ansicht im Vergleich zur quasi hufeisenförmigen ROI in der 4Ch-Ansicht zusammenhängen. Bei der räumlichen Auflösung gab es keinen klaren Trend zugunsten von Maps mit Standardauflösung oder höherer Auflösung. Wir kommen zu dem Schluss, dass die in der klinischen CMR verwendete Standardauflösung wahrscheinlich für die Radiomics-Analyse ausreicht.

Während wir uns auf die Reproduzierbarkeit von Test und Retest konzentrierten, haben andere die Auswirkungen der Bildverarbeitung auf Radiomics-Features bei T1- und T2-Mapping untersucht. Bei Patienten mit hypertropher Kardiomyopathie stellten Marfisi und Kollegen fest, dass Bild-Neuabtastung, Diskretisierung und Filterung einen signifikanten Einfluss auf die Radiomics-Features hatten [67]. In ihrer Studie reagierten Merkmale beim T2-Mapping weniger empfindlich auf Änderungen in der Bildverarbeitung als beim T1-Mapping.

In der klinischen Praxis und bei kardiovaskulären Bildgebungsstudien werden CMR-Untersuchungen häufig an verschiedenen Scannern mit unterschiedlichen Bildgebungsparametern durchgeführt. Es ist plausibel, dass die Reproduzierbarkeit in einer derart weniger kontrollierten Umgebung noch eingeschränkter ist. Eine frühere Studie ergab, dass die meisten Merkmale tatsächlich empfindlich auf Änderungen der Sequenzparameter

reagieren [40]. Für das myokardiale T1- und T2-Mapping fanden sie heraus, dass Radiomics-Features empfindlicher auf Änderungen des Flip-Winkels und der Schichtdicke und weniger empfindlich auf Änderungen der räumlichen Auflösung in der Ebene reagierten [40]. Um solche Unterschiede in den Bildaufnahmeparametern zu berücksichtigen, wurden Normalisierungsmethoden entwickelt, die die Robustheit der Radiomics-Analyse verbessern konnten [68].

Aufgrund der hohen Segmentierungsübereinstimmung durch denselben Radiologen war die Übereinstimmung zwischen den geometrischen Eigenschaften der ROIs sehr hoch, weshalb die 2D-Shape-Features überwiegend gute Ergebnisse zeigten. GLRLM quantifiziert die Länge aufeinanderfolgender Pixel mit gleichem Grauwert – sogenannte Graustufenläufe. Die relativ gute Reproduzierbarkeit in dieser Feature-Klasse kann mit der Studienkohorte (Kohorte III) zusammenhängen, die nur gesunde Probanden mit relativ homogenen T1- und T2-Karten umfasste.

Empfehlung für die klinische Forschung

In der Arbeit [45] haben wir 32 Features für die T1-Maps in SAX-Ansicht und 15 Features für die T1-Maps unabhängig von Ansicht und Auflösung sowie 12 Radiomics-Features für die T2-Maps, ebenfalls unabhängig von der Ansicht, als besonders robuste Features mit guter bis exzellenter Reproduzierbarkeit identifiziert. Wir empfehlen diese Parameter für die weitere Forschung an Patienten mit Herzerkrankungen.

Klinische Relevanz

Radiomics-Features könnten helfen, zwischen verschiedenen Herzerkrankungen zu unterscheiden, indem sie subtile Unterschiede in der Gewebetextur, Morphologie und Funktion aufdecken, die für das menschliche Auge nicht sichtbar sind. Ein weiterer Anwendungsfall könnte sich für die Überwachung des Therapieerfolgs ergeben, indem die Funktionen verwendet werden, um Veränderungen des Herzgewebes und der Herzfunktion während und nach der Behandlung zu quantifizieren. Dies kann helfen, die Therapie anzupassen und die Wirksamkeit zu maximieren. Dennoch sind weitere Untersuchungen von Radiomics-Daten erforderlich, um neue Biomarker zu entdecken und innovative diagnostische und therapeutische Ansätze zu entwickeln. Die breite Integration von Radiomics-Analysen in klinische Studien könnte dazu beitragen, neue Therapien zu validieren und die Patientenversorgung zu verbessern. Sie müssen daher bei wiederholten Messungen unter gleichen Bedingungen konsistente Ergebnisse liefern. Diese Studie leistet einen wesentlichen Beitrag zur Sicherstellung der Zuverlässigkeit kardialer Diagnoseaussagen.

4.2 LIMITATIONEN DER STUDIEN

Alle drei Studien basierten überwiegend auf Daten gesunder Freiwilliger, was die Übertragbarkeit der Ergebnisse auf Patientenkollektive mit fokalen oder diffusen Myokarderkrankungen einschränkt. Einzig die Arbeit [44] schloss zusätzlich auch Routinepatienten ein. Zukünftige Studien sollten daher auch klinische Kohorten einbeziehen, um die Generalisierbarkeit zu verbessern.

Die Bildgebung erfolgte ausschließlich an einem 1,5T-MRT-Scanner eines Herstellers. Somit ist eine hersteller- und feldstärkenübergreifende Validierung notwendig, um die Robustheit und Übertragbarkeit der Ergebnisse auf verschiedene klinische Systeme zu gewährleisten.

Die 3RR-Cine-Sequenz scheint ein idealer Kompromiss zu sein, mit deutlich reduzierter Aufnahmezeit und gleichbleibender Bildqualität und volumetrischen Ergebnissen im Vergleich zu klinischen Standardsequenzen, erfordert aber weitere Validierung im klinischen Setting.

Weitere Einschränkungen betreffen die verwendeten Sequenztypen und Bildorientierungen: In der Cine-Rekonstruktionsstudie [43] wurden ausschließlich SAX-Sequenzen verwendet; die Segmentierungsstudie [44] beschränkte sich auf native T1-Maps, während T1-Maps nach Kontrastmittelgabe, T2-Maps oder ECV-Karten nicht untersucht wurden. Zudem bezog sich die Segmentierung auf globale Mittelwerte, wodurch fokale pathologische Veränderungen unter Umständen nicht erkannt werden. Eine visuelle Überprüfung der T1-Karten bleibt daher erforderlich.

Technische Limitationen ergaben sich unter anderem aus Unterschieden in der räumlichen Auflösung zwischen Sequenzen, die durch unterschiedliche Beschleunigungsfaktoren bedingt waren. In der Entwicklung des Segmentierungsalgorithmus wurden gezielt einzelne Architekturparameter (z. B. Aktivierungsfunktion) variiert, während andere Faktoren wie Lernrate, Anzahl der Layer oder Filteranzahl nicht systematisch untersucht wurden. Eine umfassendere Hyperparameteroptimierung könnte in zukünftigen Arbeiten die Modellleistung weiter verbessern.

Zudem erfolgte die Anwendung der Segmentierungsalgorithmen retrospektiv. Ob eine vollautomatische Inline-Analyse im klinischen Alltag umsetzbar ist, wurde in dieser Arbeit nicht untersucht. Hierzu wären technische Integrationen in PACS-Systeme oder direkt in die MRT-Konsole erforderlich und sollten im Rahmen zukünftiger Studien evaluiert werden.

Für eine breite klinische Anwendung ist insbesondere die externe Validierung in multizentrischen Studien der nächste logische Schritt.

V. ZUSAMMENFASSUNG

Die vorliegende Arbeit analysiert den Einsatz von KI zur Optimierung der kardialen Magnetresonanztomographie (CMR) in drei zentralen Bereichen: der Beschleunigung der Bildaufnahme, der Entwicklung eines automatisierten Segmentierungsverfahrens und der Analyse der Reproduzierbarkeit von Radiomics-Features.

5.1 KERNERGEBNISSE DER STUDIEN

1. **Beschleunigung der Bildaufnahme durch KI-gestützte Rekonstruktion:** Eine Deep-Learning-basierte Methode wurde zur Optimierung der Bildaufnahme von Cine-Sequenzen in der MRT evaluiert. Die Ergebnisse deuten darauf hin, dass KI die Aufnahmezeit um bis zu 80 % reduzieren kann, ohne die diagnostische Bildqualität signifikant zu beeinträchtigen. Besonders die 6RR-Cine-Sequenz bot eine exzellente Bildqualität bei gleichzeitig verkürzter Untersuchungsdauer. Die 3RR-Cine-Sequenz war mehr als 50 % schneller als die klinische Referenz bei hinreichend diagnostischer Bildqualität .
2. **Automatisierte Segmentierung des Myokards:** Zur Reduktion der zeitintensiven manuellen Segmentierung wurde ein Modell zur vollautomatischen Segmentierung von nativen T1-Maps des Herzens entwickelt. Das Modell zeigte sowohl bei den Überlappungsmaßen als auch bei den Intensitätswertigkeiten eine hohe Genauigkeit gegenüber den menschlichen Beurteilern. Es wurde festgestellt, dass die Modellgenauigkeit in den basalen und mittventrikulären Schnitten höher ist als in apikalen Schnitten.
3. **Reproduzierbarkeit von Radiomics-Features:** Die Studie analysierte die Stabilität und Wiederholbarkeit von Radiomics-Features in T1- und T2-Mapping-Daten. Die Ergebnisse zeigten, dass nur eine begrenzte Anzahl von Merkmalen eine hohe Reproduzierbarkeit aufweist, insbesondere bei standardisierter Bildauflösung und -orientierung. Die Bildorientierung (SAX vs. 4Ch) hatte einen erheblichen Einfluss auf die Reproduzierbarkeit der extrahierten Features.

5.2 IMPLIKATIONEN UND PERSPEKTIVEN FÜR DIE KARDIALE BILDGEBUNG

Die Ergebnisse dieser Arbeit belegen das Potenzial von KI, die CMR auf verschiedenen Ebenen zu optimieren. Daraus ergeben sich mehrere Implikationen für klinische Anwendung und zukünftige Forschung.

1. KI-gestützte Bildrekonstruktion ermöglicht eine signifikante Verkürzung der Aufnahmezeit bei Erhalt der diagnostischen Bildqualität. Dies erhöht nicht nur den Patientenkomfort, sondern verbessert auch die Effizienz klinischer Abläufe. Besonders bei Patienten mit Arrhythmien oder eingeschränkter Atemfähigkeit kann diese Technologie einen wichtigen Beitrag leisten.

2. Deep-Learning-gestützte Segmentierungsalgorithmen bieten eine objektive und konsistente Alternative zur manuellen Segmentierung, insbesondere für T1-Mapping-Daten. Während sich viele frühere Arbeiten auf Cine-Sequenzen konzentrierten, zeigt diese Studie, dass eine zuverlässige vollautomatische Segmentierung auch bei Mapping-Sequenzen möglich ist. Dies könnte künftig eine nahtlose Integration in klinische Workflows erlauben – vorausgesetzt, Interoperabilität und Inline-Auswertbarkeit werden sichergestellt.

3. Radiomics-Analysen eröffnen neue Möglichkeiten zur Entwicklung präziserer diagnostischer und prognostischer Marker in der kardialen Bildgebung. Die Ergebnisse der dritten Studie tragen zur Validierung dieser Methodik bei. Sie zeigen aber auch, dass viele Merkmale eine unzureichende Reproduzierbarkeit aufweisen und standardisierte Erhebungs- und Analyseverfahren unerlässlich sind. Die klinische Umsetzung erfordert daher eine engere Verzahnung zwischen algorithmischer Entwicklung, scannerübergreifender Standardisierung und regulatorischer Bewertung.

4. Ausblick:

Zukünftige Forschung sollte sich auf die Entwicklung robuster und generalisierbarer KI-Modelle konzentrieren, die stabil auf verschiedenen MRT-Geräten, Feldstärken und Parametern arbeiten. Ebenso wichtig ist die externe Validierung in multizentrischen Studien, um die klinische Nutzbarkeit nachzuweisen. Eine enge Zusammenarbeit zwischen Informatik, Radiologie und Kardiologie ist hierfür unerlässlich.

VI. LITERATURVERZEICHNIS

1. World Health Organisation Cardiovascular diseases (CVDs) fact sheet; Available via [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).
 2. Messroghli DR, Moon JC, Ferreira VM et al. (2017) Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). *J Cardiovasc Magn Reson* 19(1):75.
 3. Klemenz A-C, Manzke M, Meinel FG (2024) Künstliche Intelligenz in der kardiovaskulären Radiologie Bildakquisition, Bildrekonstruktion und Workflowoptimierung. *Radiologie (Heidelb)* 64(10):766–72.
 4. Puntmann VO, Peker E, Chandrashekhar Y, Nagel E (2016) T1 Mapping in Characterizing Myocardial Disease: A Comprehensive Review. *Circ Res* 119(2):277–99.
 5. Bai W, Sinclair M, Tarroni G et al. (2018) Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson* 20(1):65.
 6. Bernard O, Lalande A, Zotti C et al. (2018) Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Trans Med Imaging* 37(11):2514–25.
 7. Böttcher B, Beller E, Busse A et al. (2020) Fully automated quantification of left ventricular volumes and function in cardiac MRI: clinical evaluation of a deep learning-based algorithm. *Int J Cardiovasc Imaging* 36(11):2239–47.
 8. Dangi S, Linte CA, Yaniv Z (2019) A distance map regularized CNN for cardiac cine MR image segmentation. *Med Phys* 46(12):5637–51.
 9. Ma Z, Wu X, Wang X et al. (2019) An iterative multi-path fully convolutional neural network for automatic cardiac segmentation in cine MR images. *Med Phys* 46(12):5652–65.
 10. Queirós S, Barbosa D, Engvall J et al. (2016) Multi-centre validation of an automatic algorithm for fast 4D myocardial segmentation in cine CMR datasets. *Eur Heart J Cardiovasc Imaging* 17(10):1118–27.
 11. Ruijsink B, Puyol-Antón E, Oksuz I et al. (2020) Fully Automated, Quality-Controlled Cardiac Analysis From CMR: Validation and Large-Scale Application to Characterize Cardiac Function. *JACC Cardiovasc Imaging* 13(3):684–95.
 12. Suinesiaputra A, Sanghvi MM, Aung N et al. (2018) Fully-automated left ventricular mass and volume MRI analysis in the UK Biobank population cohort: evaluation of initial results. *Int J Cardiovasc Imaging* 34(2):281–91.
-

13. Vigneault DM, Xie W, Ho CY, Bluemke DA, Noble JA (2018) Ω -Net (Omega-Net): Fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. *Med Image Anal* 48:95–106.
14. Arega TW, Bricq S, Legrand F, Jacquier A, Lalande A, Meriaudeau F (2023) Automatic uncertainty-based quality controlled T1 mapping and ECV analysis from native and post-contrast cardiac T1 mapping images using Bayesian vision transformer. *Med Image Anal* 86:102773.
15. Bhatt N, Ramanan V, Orbach A et al. (2022) A Deep Learning Segmentation Pipeline for Cardiac T1 Mapping Using MRI Relaxation-based Synthetic Contrast Augmentation. *Radiol Artif Intell* 4(6):e210294.
16. Fahmy AS, El-Rewaify H, Nezafat M, Nakamori S, Nezafat R (2019) Automated analysis of cardiovascular magnetic resonance myocardial native T1 mapping images using fully convolutional neural networks. *J Cardiovasc Magn Reson* 21(1):7.
17. Farrag NA, Lochbihler A, White JA, Ukwatta E (2021) Evaluation of fully automated myocardial segmentation techniques in native and contrast-enhanced T1-mapping cardiovascular magnetic resonance images using fully convolutional neural networks. *Med Phys* 48(1):215–26.
18. Hann E, Popescu IA, Zhang Q et al. (2021) Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. *Med Image Anal* 71:102029.
19. Howard JP, Chow K, Chacko L et al. (2023) Automated Inline Myocardial Segmentation of Joint T1 and T2 Mapping Using Deep Learning. *Radiol Artif Intell* 5(1):e220050.
20. Puyol-Antón E, Ruijsink B, Baumgartner CF et al. (2020) Automated quantification of myocardial tissue characteristics from native T1 mapping using neural networks with uncertainty-based quality-control. *J Cardiovasc Magn Reson* 22(1):60.
21. Hassani C, Saremi F, Varghese BA, Duddalwar V (2020) Myocardial Radiomics in Cardiac MRI. *AJR Am J Roentgenol* 214(3):536–45.
22. Ponsiglione A, Stanzione A, Cuocolo R et al. (2022) Cardiac CT and MRI radiomics: systematic review of the literature and radiomics quality score assessment. *Eur Radiol* 32(4):2629–38.
23. Baessler B, Luecke C, Lurz J et al. (2019) Cardiac MRI and Texture Analysis of Myocardial T1 and T2 Maps in Myocarditis with Acute versus Chronic Symptoms of Heart Failure. *Radiology* 292(3):608–17.

24. Nakamori S, Amyar A, Fahmy AS et al. (2024) Cardiovascular Magnetic Resonance Radiomics to Identify Components of the Extracellular Matrix in Dilated Cardiomyopathy. *Circulation* 150(1):7–18.
 25. Wang J, Yang F, Liu W et al. (2020) Radiomic Analysis of Native T1 Mapping Images Discriminates Between MYH7 and MYBPC3-Related Hypertrophic Cardiomyopathy. *J Magn Reson Imaging* 52(6):1714–21.
 26. Antonopoulos AS, Boutsikou M, Simantiris S et al. (2021) Machine learning of native T1 mapping radiomics for classification of hypertrophic cardiomyopathy phenotypes. *Sci Rep* 11(1):23596.
 27. Neisius U, El-Rewaidy H, Kucukseymen S et al. (2020) Texture signatures of native myocardial T1 as novel imaging markers for identification of hypertrophic cardiomyopathy patients without scar. *J Magn Reson Imaging* 52(3):906–19.
 28. Wang Z-C, Fan Z-Z, Liu X-Y et al. (2024) Deep Learning for Discrimination of Hypertrophic Cardiomyopathy and Hypertensive Heart Disease on MRI Native T1 Maps. *J Magn Reson Imaging* 59(3):837–48.
 29. Neisius U, El-Rewaidy H, Nakamori S, Rodriguez J, Manning WJ, Nezafat R (2019) Radiomic Analysis of Myocardial Native T1 Imaging Discriminates Between Hypertensive Heart Disease and Hypertrophic Cardiomyopathy. *JACC Cardiovasc Imaging* 12(10):1946–54.
 30. Son J, Hong YJ, Kim S et al. (2022) Radiomics Feature Analysis Using Native T1 Mapping for Discriminating Between Cardiac Tumors and Thrombi. *Acad Radiol* 29 Suppl 4:S1-S8.
 31. Chang S, Han K, Kwon Y et al. (2023) T1 Map-Based Radiomics for Prediction of Left Ventricular Reverse Remodeling in Patients With Nonischemic Dilated Cardiomyopathy. *Korean J Radiol* 24(5):395–405.
 32. Zhang J, Xu Y, Li W et al. (2023) The Predictive Value of Myocardial Native T1 Mapping Radiomics in Dilated Cardiomyopathy: A Study in a Chinese Population. *J Magn Reson Imaging* 58(3):772–9.
 33. Wu L-M, Shi R-Y, Wu C-W et al. (2022) A Radiomic MRI based Nomogram for Prediction of Heart Failure with Preserved Ejection Fraction in Systemic Lupus Erythematosus Patients: Insights From a Three-Center Prospective Study. *J Magn Reson Imaging* 56(3):779–89.
 34. Ma Q, Ma Y, Yu T, Sun Z, Hou Y (2021) Radiomics of Non-Contrast-Enhanced T1 Mapping: Diagnostic and Predictive Performance for Myocardial Injury in Acute ST-Segment-Elevation Myocardial Infarction. *Korean J Radiol* 22(4):535–46.
-

35. Wang Y, Lin L, Li X et al. (2022) Native T1 Mapping-Based Radiomics for Noninvasive Prediction of the Therapeutic Effect of Pulmonary Arterial Hypertension. *Diagnostics (Basel)* 12(10).
 36. Akinci D'Antonoli T, Cuocolo R, Baessler B, Pinto Dos Santos D (2024) Towards reproducible radiomics research: introduction of a database for radiomics studies. *Eur Radiol* 34(1):436–43.
 37. Baeßler B, Weiss K, Pinto Dos Santos D (2019) Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study. *Invest Radiol* 54(4):221–8.
 38. Woznicki P, Laqua F, Bley T, Baeßler B (2022) AutoRadiomics: A Framework for Reproducible Radiomics Research. *Front Radiol* 2:919133.
 39. Raisi-Estabragh Z, Gkontra P, Jaggi A et al. (2020) Repeatability of Cardiac Magnetic Resonance Radiomics: A Multi-Centre Multi-Vendor Test-Retest Study. *Front Cardiovasc Med* 7:586236.
 40. Jang J, El-Rewaidy H, Ngo LH et al. (2021) Sensitivity of Myocardial Radiomic Features to Imaging Parameters in Cardiac MR Imaging. *J Magn Reson Imaging* 54(3):787–94.
 41. Jiang D, Dou W, Vosters L, Xu X, Sun Y, Tan T (2018) Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network. *Jpn J Radiol* 36(9):566–74.
 42. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans Image Process* 26(7):3142–55.
 43. Klemenz A-C, Reichardt L, Gorodezky M et al. (2024) Accelerated Cardiac MRI with Deep Learning-based Image Reconstruction for Cine Imaging. *Radiol Cardiothorac Imaging* 6(6):e230419.
 44. Manzke M, Iseke S, Böttcher B, Klemenz A-C, Weber M-A, Meinel FG (2024) Development and performance evaluation of fully automated deep learning-based models for myocardial segmentation on T1 mapping MRI data. *Sci Rep* 14(1):18895.
 45. Manzke M, Laqua FC, Böttcher B et al. (2025) Repeatability of radiomic features in myocardial T1 and T2 mapping. *Eur Radiol*. DOI:10.1007/s00330-024-11337-8.
 46. Kramer CM, Barkhausen J, Bucciarelli-Ducci C, Flamm SD, Kim RJ, Nagel E (2020) Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update. *J Cardiovasc Magn Reson* 22(1):17.
 47. Emrich T, Halfmann M, Schoepf UJ, Kreitner K-F (2021) CMR for myocardial characterization in ischemic heart disease: state-of-the-art and future developments. *Eur Radiol Exp* 5(1):14.
-

48. Mazurowski MA, Buda M, Saha A, Bashir MR (2019) Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *J Magn Reson Imaging* 49(4):939–54.
49. Currie G, Hawk KE, Rohren E, Vial A, Klein R (2019) Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *J Med Imaging Radiat Sci* 50(4):477–87.
50. Böttcher B, Lorbeer R, Stöcklein S et al. (2021) Global and Regional Test-Retest Reproducibility of Native T1 and T2 Mapping in Cardiac Magnetic Resonance Imaging. *J Magn Reson Imaging* 54(6):1763–72.
51. Khaledyan D, Amirany A, Jafari K, Moaiyeri MH, Khuzani AZ, Mashhadi N (2020) Low-Cost Implementation of Bilinear and Bicubic Image Interpolation for Real-Time Image Super-Resolution. *arXiv*.
52. Anand S, Priya L (2019) *A Guide for Machine Vision in Quality Control*. Chapman and Hall/CRC.
53. Handels H (2009) *Medizinische Bildverarbeitung: Bildanalyse, Mustererkennung und Visualisierung für die computergestützte ärztliche Diagnostik und Therapie*, 2nd edn. Vieweg+Teubner.
54. Shen T, Huang F, Zhang X (2023) CT medical image segmentation algorithm based on deep learning technology. *Math Biosci Eng* 20(6):10954–76.
55. Furtado P (2021) Testing Segmentation Popular Loss and Variations in Three Multiclass Medical Imaging Problems. *J Imaging* 7(2).
56. Müller D, Soto-Rey I, Kramer F (2022) Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res Notes* 15(1):210.
57. Kingma DP, Ba J (2014) Adam: A Method for Stochastic Optimization.
58. Dalmer A, Meinel FG, Böttcher B et al. (2024) Native myocardial T1 mapping: influence of spatial resolution on quantitative results and reproducibility. *Quant Imaging Med Surg* 14(1):20–30.
59. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86(2):420–8.
60. Fisher RA (1992) Statistical Methods for Research Workers. In: Kotz S, Johnson NL (eds) *Breakthroughs in Statistics*. Springer New York. New York, NY, pp 66–70.
61. Lin LI-K (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989:45:255.

62. Morgan CJ, Aban I (2016) Methods for evaluating the agreement between diagnostic tests. *J Nucl Cardiol* 23(3):511–3.
63. Koo TK, Li MY (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 15(2):155–63.
64. Monteuuis D, Bouzerar R, Dantoing C, Poujol J, Bohbot Y, Renard C (2024) Prospective Comparison of Free-Breathing Accelerated Cine Deep Learning Reconstruction Versus Standard Breath-Hold Cardiac MRI Sequences in Patients With Ischemic Heart Disease. *AJR Am J Roentgenol* 222(5):e2330272.
65. Chen EZ, Chen X, Lyu J et al. (2021) Cardiac Functional Analysis with Cine MRI via Deep Learning Reconstruction.
66. Marfisi D, Giannelli M, Marzi C et al. (2024) Test-retest repeatability of myocardial radiomic features from quantitative cardiac magnetic resonance T1 and T2 mapping. *Magn Reson Imaging* 113:110217.
67. Marfisi D, Tessa C, Marzi C et al. (2022) Image resampling and discretization effect on the estimate of myocardial radiomic features from T1 and T2 mapping in hypertrophic cardiomyopathy. *Sci Rep* 12(1):10186.
68. Hu Z, Zhuang Q, Xiao Y et al. (2021) MIL normalization -- prerequisites for accurate MRI radiomics analysis. *Comput Biol Med* 133:104403.

VII. THESEN

1. KI-gestützte Methoden zur Bildrekonstruktion können die Erfassungszeit in der kardialen MRT erheblich reduzieren, ohne dass die diagnostische Bildqualität signifikant beeinträchtigt wird.
2. Die Reduktion der Scanzeit durch KI-basierte Rekonstruktion erhöht den Patientenkomfort und kann insbesondere für Patientengruppen mit Atemproblemen oder Arrhythmien von Vorteil sein.
3. Trotz der Vorteile von KI-gestützter Rekonstruktion besteht weiterhin die Notwendigkeit einer umfassenden klinischen Validierung, um eine flächendeckende Implementierung sicherzustellen.
4. DL-Modelle zur automatischen Segmentierung zeigen eine hohe Genauigkeit und können die Interrater-Variabilität zwischen menschlichen Beurteilern verringern.
5. Die automatische Segmentierung ermöglicht eine schnellere und konsistentere Analyse von kardialen MRT-Daten und reduziert den manuellen Bearbeitungsaufwand erheblich.
6. KI-basierte Segmentierungsmodelle könnten langfristig als klinischer Standard etabliert werden, müssen jedoch in Hinblick auf Robustheit und Generalisierbarkeit weiterentwickelt werden.
7. Radiomics-Analysen liefern detaillierte Bildinformationen und haben das Potenzial, neue Biomarker für kardiologische Erkrankungen zu identifizieren.
8. Die Reproduzierbarkeit von Radiomics-Features ist stark von der Bildauflösung und der Bildorientierung abhängig, was eine Standardisierung der Analysemethoden erfordert.
9. Für eine Translation von Radiomics aus der experimentellen Radiologie in die Klinikumgebung bedarf es weiterer multizentrischer und Multi-Vendor Untersuchungen, die zur Erforschung von Referenzwerten beitragen.
10. Nur eine begrenzte Anzahl von Merkmalen ist in verschiedenen Scanner-Einstellungen stabil genug für eine klinische Anwendung, was die derzeitige Nutzung in der Routinepraxis einschränkt.
11. Die Weiterentwicklung von Radiomics-Methoden erfordert eine enge Zusammenarbeit zwischen Informatik, Medizin und Bildverarbeitung, um robuste, diagnostisch verwertbare Features zu identifizieren.
12. Bildartefakte können die Aussagekraft von Radiomics-Features wesentlich beeinträchtigen.
13. Die Kombination aus Deep Learning, Radiomics und KI-gestützter Rekonstruktion hat das Potenzial, neue personalisierte Therapieansätze zu ermöglichen.

14. Trotz der vielversprechenden Ergebnisse erfordert die breite Implementierung von KI in der kardiologischen Bildgebung eine umfassende regulatorische Prüfung und Standardisierung der Verfahren.

VIII. DANKSAGUNG

Mein aufrichtiger und tief empfundener Dank gilt zuallererst meiner geliebten Frau Olga sowie meinen wunderbaren Kindern Liana und Matteo. Für all die entbehrungsreichen Abende und Wochenenden, für Eure Geduld, Euer Verständnis und Eure unermüdliche Liebe – Ihr habt mich stets motiviert, weiterzumachen und nach vorne zu blicken.

Mein besonderer Dank gilt meinem Doktorvater, Prof. Dr. med. Felix Meinel, für die hervorragende Betreuung, seine stets offene Tür, seine Geduld sowie die zahlreichen konstruktiven Gespräche, die maßgeblich zum Gelingen dieser Arbeit beigetragen haben.

Ebenso danke ich meinen geschätzten Kolleg:innen Dr. Ann-Christin Klemenz, Dr. Benjamin Böttcher, Linda Reichardt und Simon Iseke – ohne Euren unermüdlichen Einsatz und Euer Engagement wären die Materialien für diese Arbeit nicht zustande gekommen.

Mein Dank gilt auch Dr. Margarita Gorodetzky von GE Healthcare für die großartige technische Unterstützung bei der Inbetriebnahme neuester Technologien sowie bei der Durchführung der Studie – Ihre Expertise war eine große Bereicherung.

Ein weiterer herzlicher Dank geht an das gesamte Team des Instituts für Diagnostische und Interventionelle Radiologie, Kinder- und Neuroradiologie der Universitätsmedizin Rostock – insbesondere an Herrn Prof. Dr. Marc-André Weber – für die fachliche Begleitung sowie die stets angenehme und unterstützende Zusammenarbeit. Das kollegiale Miteinander und der lebendige fachliche Austausch tragen täglich dazu bei, neugierig zu bleiben und neue Ideen zu entwickeln – ein unschätzbarer Wert.

IX. LISTE ALLER VERÖFFENTLICHUNGEN UND FACHVORTRÄGE

Originalarbeiten

1. **Manzke M**, Laqua FC, Böttcher B, Klemenz AC, Weber MA, Baeßler B, Meinel FG. Repeatability of radiomic features in myocardial T1 and T2 mapping. Eur Radiol. 2025 Jan 15. doi: 10.1007/s00330-024-11337-8.
2. Klemenz AC, Reichardt L, Gorodezky M, **Manzke M**, Zhu X, Dalmer A, Lorbeer R, Lang CI, Weber MA, Meinel FG. Accelerated Cardiac MRI with Deep Learning-based Image Reconstruction for Cine Imaging. Radiol Cardiothorac Imaging. 2024 Dec;6(6):e230419. doi: 10.1148/ryct.230419.
3. **Manzke M**, Iseke S, Böttcher B, Klemenz AC, Weber MA, Meinel FG. Development and performance evaluation of fully automated deep learning-based models for myocardial segmentation on T1 mapping MRI data. Sci Rep. 2024 Aug 14;14(1):18895. doi: 10.1038/s41598-024-69529-7.
4. Klemenz AC, Albrecht L, **Manzke M**, Dalmer A, Böttcher B, Surov A, Weber MA, Meinel FG. Improved image quality in CT pulmonary angiography using deep learning-based image reconstruction. Sci Rep. 2024 Jan 30;14(1):2494. doi: 10.1038/s41598-024-52517-2.
5. Klemenz AC, Beckert L, **Manzke M**, Lang CI, Weber MA, Meinel FG. Influence of Deep Learning Based Image Reconstruction on Quantitative Results of Coronary Artery Calcium Scoring. Acad Radiol. 2024 Jun;31(6):2259-2267. doi: 10.1016/j.acra.2024.03.020.
6. Heinrich A, Yücel S, Böttcher B, Öner A, **Manzke M**, Klemenz AC, Weber MA, Meinel FG. Improved image quality in transcatheter aortic valve implantation planning CT using deep learning-based image reconstruction. Quant Imaging Med Surg. 2023 Feb 1;13(2):970-981. doi: 10.21037/qims-22-639. Epub 2022 Dec 20.
7. Dalmer A, Meinel FG, Böttcher B, **Manzke M**, Lorbeer R, Weber MA, Baeßler B, Klemenz AC. Native myocardial T1 mapping: influence of spatial resolution on quantitative results and reproducibility. Quant Imaging Med Surg. 2024 Jan 3;14(1):20-30. doi: 10.21037/qims-23-943. Epub 2023 Nov 14.
8. Steuwe A, Kamp B, Afat S, Akinina A, Aludin S, Bas EG, Berger J, Bohrer E, Brose A, Büttner SM, Ehrengut C, Gerwing M, Grosu S, Gussew A, Güttler F, Heinrich A, Jiraskova P, Kloth C, Kottlors J, Kuennemann MD, Liska C, Lubina N, **Manzke M**, Meinel FG, Meyer HJ, Mittermeier A, Persigehl T, Schmill LP, Steinhardt M, The Racoon Study Group, Antoch G, Valentin B. Standardization of a CT Protocol for

Imaging Patients with Suspected COVID-19-A RACOON Project. Bioengineering (Basel). 2024 Feb 22;11(3):207. doi: 10.3390/bioengineering11030207.

Übersichtsarbeiten

9. Klemenz AC, Manzke M, Meinel FG. Künstliche Intelligenz in der kardiovaskulären Radiologie : Bildakquisition, Bildrekonstruktion und Workflowoptimierung [Artificial intelligence in cardiovascular radiology : Image acquisition, image reconstruction and workflow optimization]. Radiologie (Heidelb). 2024 Oct;64(10):766-772. German. doi: 10.1007/s00117-024-01335-8.

Sonstige Veröffentlichungen

10. Schröder M, Bürger J, Krüger F, Waltemath D, Dörr M, Winterhalter C, Spors S, Matela J, Al-Suadi S, Baldauf M, Becker M, Birke V, Biskup T, Bläsing D, Budde-Sagert K, Cap CH, Eggert A, Escobar HM, Feistel S, Fennel T, Fischer H, Fischer S, Fromm A, Fuellen G, Garbe LA, Glende U, Gloede C, Gröber P, Hassenrück C, Hillmann F, Hoff K, Huhs N, Jürgensmann S, Jürß H, Kaunaite D, Kerntopf M, Kiesendahl J, Kirste T, Köhling R, Kragl U, Krause T, Kroll S, Labahn K, Lange K, Linke C, Freiherr von Lukas U, Mahler E, **Manzke M**, Michael J, Otto J, Placke M, Raab-Düsterhöft A, Scheel S, Schick S, Schmidt S, Schmitt F, Schnell S, Schöner M, Schwenk C, Schwerdt D, Söll M, Stanke M, Steinbrecher C, Theise A, Weber MA, Wego A, Wiegand-Hoffmeister B, Winkelmann F, Yordanova K, Zeller-Plumhoff B, Zielinski O. Gemeinschaftliches Konzept für eine Landesinitiative zum Forschungsdatenmanagement in Mecklenburg-Vorpommern. doi: 10.5281/ZENODO.13889465

Wissenschaftliche Fachvorträge:

1. Präsenzvortrag auf dem *DIZ Vernetzungstreffen Unimedizin Greifswald/Unimedizin Rostock 2023*: „RACOON-Radiological Cooperative Network“
2. Forschungs-Pitch auf dem *Forschungs-Workshop im Innovationport Wismar 2023*: „Medizin zwischen Physik und Data Science“
3. Präsenzvortrag auf dem *Tag der KI der Universitätsmedizin und Universität Rostock 2024*: „Chancen und Potential der KI in der Radiologie“
4. Präsenzvortrag auf dem jährlichen Treffen der *European Society of Cardiovascular Radiology 2024*: „Development and performance evaluation of fully automated deep

learning-based models for short and long axis segmentation of the myocardium using T1 mapping MRI data”

5. Präsenzvortrag auf dem *European Congress of Radiology 2025*: „Test-retest reliability of radiomic features in myocardial T1 and T2 mapping”
6. Präsenzvortrag auf dem *European Congress of Radiology 2025*: Projektvorstellung „RACOON-CORE_PE”
7. Online-Vortrag auf dem *106. Deutschen Röntgenkongress 2025*: „Test-Retest-Zuverlässigkeit von Radiomics-Features beim T1- und T2-Mapping des Myokards“

X. ANHANG

Definition 1 (Intraklassenkorrelationskoeffizient (ICC)):

Der ICC ist eine statistische Maßzahl zur Bewertung der Übereinstimmung oder Zuverlässigkeit von Messungen innerhalb von Gruppen oder Klassen. Er quantifiziert, wie stark Messwerte innerhalb derselben Gruppe oder Kategorie ähnlicher sind als Messwerte zwischen verschiedenen Gruppen. Der ICC basiert auf der Zerlegung der Gesamtvarianz einer Stichprobe in zwei Komponenten und ist definiert als:

$$ICC(1,1) = \frac{MS_B - MS_W}{MS_B + (k-1)MS_W} \quad (1)$$

$$ICC(1,k) = \frac{MS_B - MS_W}{MS_B} \quad (2)$$

$$ICC(2,1) = \frac{MS_B - MS_{err}}{MS_B + (k-1)MS_{err} + \frac{k(MS_{BM} - MS_{err})}{n}} \quad (3)$$

mit

- k – Anzahl der Messungen pro Gruppe bei wiederholten Messungen
- n – Anzahl der Messobjekte (Fälle),
- MS_B – Varianz zwischen Gruppen,
- MS_W – Varianz innerhalb der Gruppen,
- MS_B – Varianz zwischen Gruppen,
- MS_{BM} – Varianz zwischen den Messungen pro Gruppe,
- MS_{err} – Restvarianz.

Definition 2 (Konkordanz-Korrelationskoeffizient (CCC)):

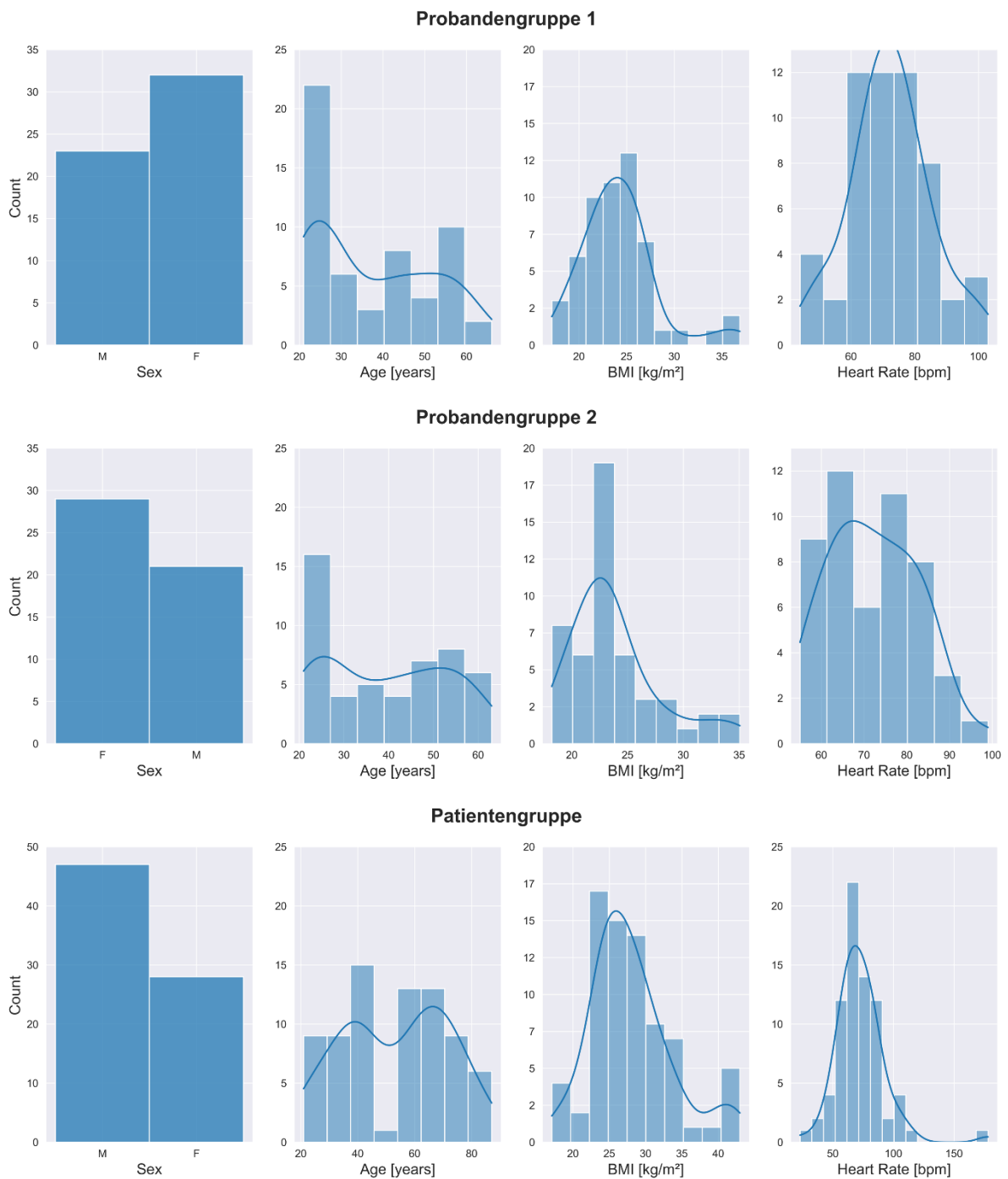
Der CCC misst die Übereinstimmung zwischen zwei Variablen und berücksichtigt sowohl Korrelation als auch Übereinstimmung in den Mittelwerten.

$$CCC = \frac{2\sigma_{12}}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}$$

mit

- μ_i – Mittelwert der jeweiligen Stichprobe ($i \in \{1,2\}$),
 - σ_i^2 – Varianz der jeweiligen Stichprobe ,
 - σ_{12} – Kovarianz zwischen beiden Stichproben
-

Zusatzmaterialien

**ZUSATZABBILDUNG 1 – VERTEILUNGEN DER DEMOGRAFISCHEN INFORMATIONEN.**

DIE GRAFIK ZEIGT DIE STATISTISCHEN VERTEILUNGEN VON ALTER, GESCHLECHT, BODY-MASS-INDEX (BMI) UND HERZFREQUENZ ALLER DREI TEILNEHMERGRUPPEN. PROBANDGRUPPE 1 (OBEN), PROBANDENGRUPPE 2 (MITTE) UND DIE PATIENTENGRUPPE (UNTEN).

ZUSATZTABELLE 1 – VOLLSTÄNDIGE LISTE ALLER UNTERSUCHTEN RADIOMICS-FEATURES

Index	Feature Klasse	Feature Name
1	2D-Shape	original_shape_Elongation
2	2D-Shape	original_shape_Flatness
3	2D-Shape	original_shape_LeastAxisLength
4	2D-Shape	original_shape_MajorAxisLength
5	2D-Shape	original_shape_Maximum2DDiameterColumn
6	2D-Shape	original_shape_Maximum2DDiameterRow
7	2D-Shape	original_shape_Maximum2DDiameterSlice
8	2D-Shape	original_shape_MinorAxisLength
9	2D-Shape	original_shape_Sphericity
10	First_Order	original_firstorder_10Percentile
11	First_Order	original_firstorder_90Percentile
12	First_Order	original_firstorder_Energy
13	First_Order	original_firstorder_Entropy
14	First_Order	original_firstorder_InterquartileRange
15	First_Order	original_firstorder_Kurtosis
16	First_Order	original_firstorder_Maximum
17	First_Order	original_firstorder_MeanAbsoluteDeviation
18	First_Order	original_firstorder_Mean
19	First_Order	original_firstorder_Median
20	First_Order	original_firstorder_Minimum
21	First_Order	original_firstorder_Range
22	First_Order	original_firstorder_RobustMeanAbsoluteDeviation
23	First_Order	original_firstorder_RootMeanSquared
24	First_Order	original_firstorder_Skewness
25	First_Order	original_firstorder_TotalEnergy
26	First_Order	original_firstorder_Uniformity
27	First_Order	original_firstorder_Variance
28	GLCM	original_glcm_Autocorrelation
29	GLCM	original_glcm_JointAverage
30	GLCM	original_glcm_ClusterProminence
31	GLCM	original_glcm_ClusterShade
32	GLCM	original_glcm_ClusterTendency
33	GLCM	original_glcm_Contrast
34	GLCM	original_glcm_Correlation
35	GLCM	original_glcm_DifferenceAverage
36	GLCM	original_glcm_DifferenceEntropy
37	GLCM	original_glcm_DifferenceVariance
38	GLCM	original_glcm_JointEnergy
39	GLCM	original_glcm_JointEntropy
40	GLCM	original_glcm_Imc1
41	GLCM	original_glcm_Imc2

42	GLCM	original_glcm_Idm
43	GLCM	original_glcm_Idmn
44	GLCM	original_glcm_Id
45	GLCM	original_glcm_Idn
46	GLCM	original_glcm_InverseVariance
47	GLCM	original_glcm_MaximumProbability
48	GLCM	original_glcm_SumEntropy
49	GLCM	original_glcm_SumSquares
50	GLRLM	original_glrlm_GrayLevelNonUniformity
51	GLRLM	original_glrlm_GrayLevelNonUniformityNormalized
52	GLRLM	original_glrlm_GrayLevelVariance
53	GLRLM	original_glrlm_HighGrayLevelRunEmphasis
54	GLRLM	original_glrlm_LongRunEmphasis
55	GLRLM	original_glrlm_LongRunHighGrayLevelEmphasis
56	GLRLM	original_glrlm_LongRunLowGrayLevelEmphasis
57	GLRLM	original_glrlm_LowGrayLevelRunEmphasis
58	GLRLM	original_glrlm_RunEntropy
59	GLRLM	original_glrlm_RunLengthNonUniformity
60	GLRLM	original_glrlm_RunLengthNonUniformityNormalized
61	GLRLM	original_glrlm_RunPercentage
62	GLRLM	original_glrlm_RunVariance
63	GLRLM	original_glrlm_ShortRunEmphasis
64	GLRLM	original_glrlm_ShortRunHighGrayLevelEmphasis
65	GLRLM	original_glrlm_ShortRunLowGrayLevelEmphasis
66	GLSZM	original_glszm_GrayLevelNonUniformity
67	GLSZM	original_glszm_GrayLevelNonUniformityNormalized
68	GLSZM	original_glszm_GrayLevelVariance
69	GLSZM	original_glszm_HighGrayLevelZoneEmphasis
70	GLSZM	original_glszm_LargeAreaEmphasis
71	GLSZM	original_glszm_LargeAreaHighGrayLevelEmphasis
72	GLSZM	original_glszm_LargeAreaLowGrayLevelEmphasis
73	GLSZM	original_glszm_LowGrayLevelZoneEmphasis
74	GLSZM	original_glszm_SizeZoneNonUniformity
75	GLSZM	original_glszm_SizeZoneNonUniformityNormalized
76	GLSZM	original_glszm_SmallAreaEmphasis
77	GLSZM	original_glszm_SmallAreaHighGrayLevelEmphasis
78	GLSZM	original_glszm_SmallAreaLowGrayLevelEmphasis
79	GLSZM	original_glszm_ZoneEntropy
80	GLSZM	original_glszm_ZonePercentage
81	GLSZM	original_glszm_ZoneVariance
82	GLDM	original_gldm_DependenceEntropy
83	GLDM	original_gldm_DependenceNonUniformity
84	GLDM	original_gldm_DependenceNonUniformityNormalized

85	GLDM	original_gldm_DependenceVariance
86	GLDM	original_gldm_GrayLevelNonUniformity
87	GLDM	original_gldm_GrayLevelVariance
88	GLDM	original_gldm_HighGrayLevelEmphasis
89	GLDM	original_gldm_LargeDependenceEmphasis
90	GLDM	original_gldm_LargeDependenceHighGrayLevelEmphasis
91	GLDM	original_gldm_LargeDependenceLowGrayLevelEmphasis
92	GLDM	original_gldm_LowGrayLevelEmphasis
93	GLDM	original_gldm_SmallDependenceEmphasis
94	GLDM	original_gldm_SmallDependenceHighGrayLevelEmphasis
95	GLDM	original_gldm_SmallDependenceLowGrayLevelEmphasis
96	NGTDM	original_ngtdm_Busyness
97	NGTDM	original_ngtdm_Coarseness
98	NGTDM	original_ngtdm_Complexity
99	NGTDM	original_ngtdm_Contrast
100	NGTDM	original_ngtdm_Strength

Accelerated Cardiac MRI with Deep Learning–based Image Reconstruction for Cine Imaging

Ann-Christin Klemen, PhD • Linda Reichardt • Margarita Gorodezky, PhD • Mathias Manzke, Dipl-Math Techn • Xucheng Zhu, PhD • Antonia Dalmer, MD • Roberto Lorbeer, PhD • Cajetan I. Lang, MD • Marc-André Weber, MD, MSc • Felix G. Meinel, MD

From the Institute of Diagnostic and Interventional Radiology, Pediatric Radiology and Neuroradiology (A.C.K., L.R., M.M., A.D., M.A.W., F.G.M.), and Department of Cardiology (C.I.L.), Rostock University Medical Center, Schillingallee 36, 18057 Rostock, Germany; GE HealthCare, Munich, Germany (M.G.); GE HealthCare, Menlo Park, Calif (X.Z.); and Department of Radiology, Ludwig-Maximilian University, Munich, Germany (R.L.). Received August 2, 2023; revision requested October 20; revision received September 26, 2024; accepted October 7. Address correspondence to F.G.M. (email: felix.meinel@med.uni-rostock.de).

Authors declared no funding for this work.

Conflicts of interest are listed at the end of this article.

Radiology: Cardiothoracic Imaging 2024; 6(6):e230419 • <https://doi.org/10.1148/ryct.230419> • Content codes:   

Purpose: To assess the influence of deep learning (DL)–based image reconstruction on acquisition time, volumetric results, and image quality of cine sequences in cardiac MRI.

Materials and Methods: This prospective study (performed from January 2023 to March 2023) included 55 healthy volunteers who underwent a noncontrast cardiac MRI examination at 1.5 T. Short-axis stack DL cine sequences of the left ventricle (LV) were performed over one (1RR), three (3RR), and six cardiac (6RR) cycles and compared with a standard cine sequence (without DL, performed over 10–12 cardiac cycles) in regard to acquisition time, subjective image quality, edge sharpness, and volumetric results.

Results: Total acquisition time (median) for a short-axis stack was 47 seconds for the 1RR cine, 108 seconds for 3RR cine, 184 seconds for 6RR cine, and 227 seconds for the standard sequence. Volumetric results showed no difference for the conventional cine (median LV ejection fraction [EF] 63%), 6RR cine (median LVEF, 62%), and 3RR cine (median LVEF, 61%). The 1RR cine sequence significantly underestimated EF (57%) because of a different segmentation of the papillary muscles. Subjective image quality ($P = .37$) and edge sharpness ($P = .06$) of the three-heartbeat DL cine did not differ from the reference standard, while both metrics were lower for single-heartbeat DL cine and higher for six-heartbeat DL cine.

Conclusion: For DL-based cine sequences, acquisition over three cardiac cycles appears to be the optimal compromise, with no evidence of differences in image quality, edge sharpness, and volumetric results, but with a greater than 50% reduced acquisition time compared with the reference sequence.

Supplemental material is available for this article.

© RSNA, 2024

Cardiac MRI with cine sequences is the reference standard for assessment of cardiac function (1). Cine sequences generate images of the cardiac cycle at different times, which can be assembled into a moving image sequence using various reconstruction algorithms. Despite extensive technical advancement in the field of cardiovascular MRI, long acquisition times and arrhythmia-related motion artifacts remain major challenges of this imaging modality (2).

Accelerated cine sequences have been developed but are not commonly used due to low signal-to-noise ratio and reduced image quality. These sequences could be fundamentally improved by deep learning (DL)–based image reconstruction algorithms (3,4). While a conventional cine acquisition is typically acquired over 10–12 heartbeats, we hypothesized that the improved image reconstruction techniques based on DL enable cine imaging in MRI over fewer heartbeats, possibly even in real time, with sufficient image quality. Therefore, the purpose of our study was to assess the influence of DL-based image reconstruction on acquisition time, volumetric results, and image quality of cardiac MRI cine sequences.

Materials and Methods

GE HealthCare provided the prototype MRI sequence and technical support. Two authors (M.G. and X.Z.) are employees

of GE HealthCare. Those authors who are not employees of GE HealthCare had control of the inclusion of any data and information that might present a conflict of interest for those authors who are employees of GE HealthCare.

Study Design and Participants

This prospective, single-center cohort study (performed from January 2023 to March 2023) was approved by the responsible institutional review board (University Medical Center Rostock). Written informed consent was obtained from all participants. Participants were chosen from a pool of volunteers. Exclusion criteria for the study were: pregnancy, general contraindications for MRI (claustrophobia, non–MRI-compatible implants), and a history of heart disease. We included 55 healthy adult volunteers (>18 years).

MRI Protocol

Each volunteer underwent noncontrast cardiac MRI with a 1.5-T MRI system (Signa Artist; GE HealthCare). The MRI protocols included four short-axis stacks covering the whole left ventricle (LV) acquired with our clinical standard (a conventional balanced steady-state free precession, array coil spatial sensitivity encoding [ASSET]–accelerated cine sequence) and three versions of the DL-based cine sequence performed over

Abbreviations

ASSET = array coil spatial sensitivity encoding, DL = deep learning, EDV = end-diastolic volume, EF = ejection fraction, ESV = end-systolic volume, LV = left ventricle, SV = stroke volume

Summary

Deep learning–based cardiac cine sequences performed over three cardiac cycles led to substantially reduced acquisition time, while no differences in image quality and volumetric results compared with a standard clinical sequence could be detected.

Key Points

- Deep learning (DL)–based real-time acquisition had the shortest acquisition time, with slightly inferior image quality and edge sharpness.
- DL-based cine acquisition over three cardiac cycles had substantially reduced acquisition time (>50%), with no evidence of differences in image quality, edge sharpness, and volumetric results of the left ventricle compared with the reference standard sequence.
- DL-based cine acquisition over six cardiac cycles showed significantly enhanced image quality without increased acquisition time.

Keywords

MR Imaging, Cardiac, Heart, Technical Aspects, Cardiac MRI, Deep Learning, Clinical Imaging, Accelerated Imaging

one cardiac cycle (1RR), three cardiac cycles (3RR), and six cardiac cycles (6RR) with acceleration factors of 15, eight, and eight, respectively. Acceleration was calculated using the following equation:

$$\text{acc} = \frac{K_y}{k_y} = \frac{K_y}{\text{vps} \cdot \#RR},$$

where acc is the acceleration factor; K_y and k_y are the fully sampled and acquired number of phase lines, respectively; vps is view per segment, which determines the temporal resolution; and #RR is the number of cardiac cycles for imaging. A more detailed overview of MRI parameters can be found in Table S1.

Accelerated Cine Sequences

Two major components are included in DLCine (SONIC DL prototype; GE HealthCare): variable density sampling acquisition and a DL-based reconstruction algorithm similar to the one used by Orii et al (5). The DLCine prototype had been trained on multiple short- and long-axis sections acquired using both 1.5-T and 3-T systems; images were from 156 volunteers drawn from a volunteer pool with an age range of 26–64 years.

While uniform density sampling is used in conventional ASSET cine acquisition, a variable density k-t sampling strategy is used for DLCine (Fig 1). The precise sampling pattern for the 1RR, 3RR, and 6RR cine sequences can be found in Figure S1. For DLCine, each cardiac phase uses the same sampling density and acceleration factor to generate patterns. Random shift for each cardiac phase in the phase-encoding dimension is applied to create sampling incoherence in the cardiac phase dimension. After sampling patterns are generated, sampling points will be sorted into different cardiac cycles by k_y location.

An unrolled neural network is used in DLCine to reconstruct cine images from undersampled k-space data (Fig 2). The network includes multiple iterations (unrolled blocks), and each

block contains a data consistency module and a neural network–based regularizer. The network is trained with fully sampled cine datasets. The multichannel cine dataset is retrospectively undersampled in k-space and fed into the network as input, and the fully sampled cine images are the label. An L1 loss between reconstructed cine images and fully sampled cine images is used as the loss function in model training. The neural network was trained with a wide range of acceleration factors (from one to 16).

The variable density k-t acquisition is optimized based on previous work (6). Edge k-space is more aggressively undersampled compared with the center k-space.

The DL reconstruction method, using an unrolled neural network reconstruction architecture, is similar to previous work from Sandino et al (7) and Zucker et al (8). Our model uses the same model architecture as Sandino et al (7), but with 12 unroll blocks instead of five. The same network is used for the different RR protocols. Temporal interpolation was applied in both reference and investigational sequences.

Assessment of Acquisition Time

The number and duration of breath holds and the total time of the sequence (runtime of sequence + duration of all breath holds + pauses between breath holds) was recorded for a subset of 10 participants.

Image Quality Assessment

Subjective image quality was assessed by two radiologists (A.D. and F.G.M.) with 5 and 10 years of experience in cardiac MRI, respectively. Blinding was not possible due to the differences in image impression between the sequences. Image quality was rated on a five-point scale regarding image contrast, blurring, and artifacts: 5 = excellent, 4 = good, 3 = sufficient, 2 = poor, 1 = non-diagnostic. To objectively quantify image quality, we determined edge sharpness in the images as the steepness of signal intensity change at the transition from the blood pool to the myocardium (Fig 3). Methodologic details are described in Appendix S1.

Postprocessing

Commercial software (cvi42, version 5.16; Circle Cardiovascular Imaging) was used to obtain LV and right ventricular end-diastolic volume (EDV) and end-systolic volume (ESV), stroke volume (SV), ejection fraction (EF), and mass. EDV, ESV, SV, and mass were also indexed to the body surface area. Endo- and epicardial contours were automatically segmented by the software. Papillary muscles were segmented and counted toward LV mass. All contours were subsequently checked and corrected by an experienced cardiovascular radiologist, as described by Böttcher et al (9). We recorded the time needed for checking and, if necessary, correcting the fully automated contours. We also documented the number of individual sections needing correction and how often the choice of end-systolic or end-diastolic phase needed correction. This analysis was performed to investigate whether increases in acquisition time would be offset by the longer time required for the correction of fully automated volumetric analysis.

Statistical Analysis

Participant characteristics and imaging parameters are presented as medians and ranges. We used the Shapiro-Francia

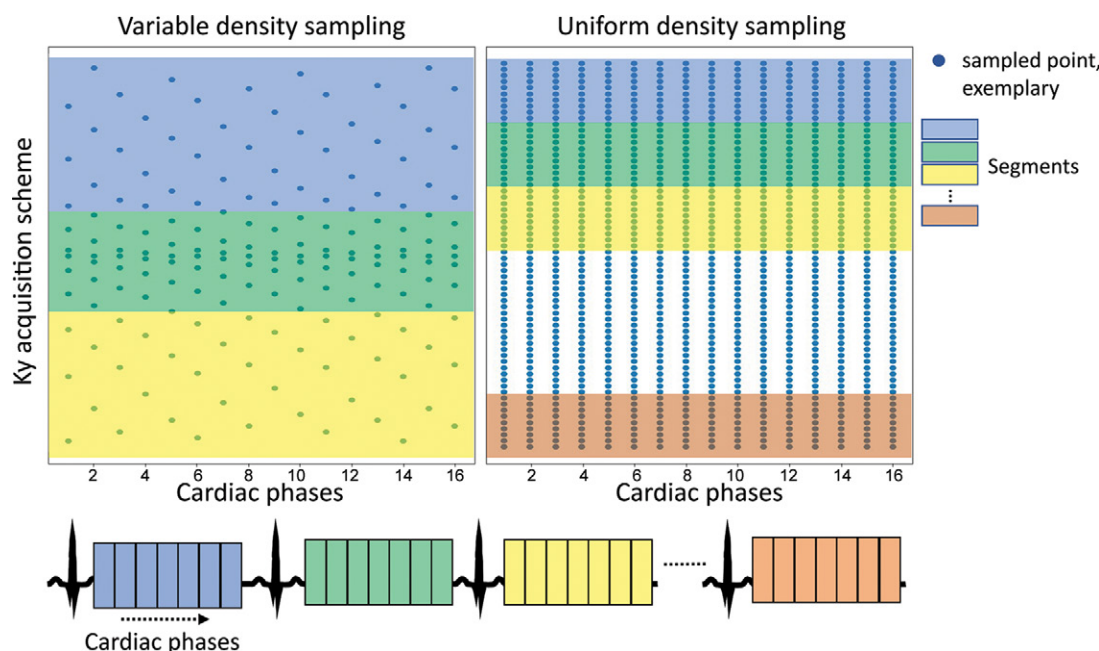


Figure 1: Visualization of acquisition schemes. For a cardiac MRI cine acquisition, the k-space is divided into segments (colored boxes), and each segment needs to be sampled in all cardiac phases, as seen in the lower part of the figure. For a standard sequence (right), the k-space is sampled uniformly, resulting in the segments having equal size. For variable density sampling (left), the outer k-space segments have fewer points, and hence, a larger segment can be acquired within the same time. The segment size is adjusted to the sampling density, resulting in fewer segments overall and hence, an accelerated acquisition.

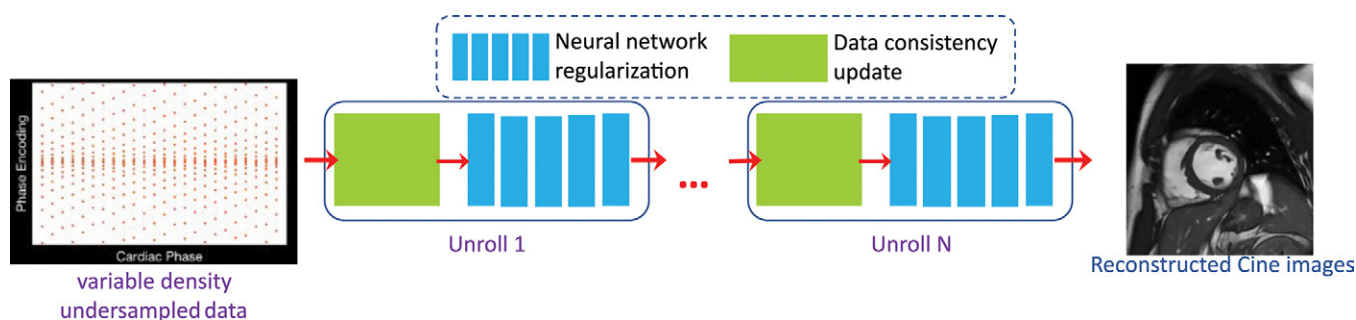


Figure 2: Schematic representation of deep learning-based image reconstruction. An unrolled neural network is used in DLCine to reconstruct cine images from the undersampled k-space data. The network includes multiple iterations (unrolled blocks), with each block containing a data consistency module and a neural network-based regularizer.

test to check normal distributions of the variables that revealed several rejections of the normal distribution. Differences between the DL cine sequences and the reference sequence were assessed by Wilcoxon matched pairs signed rank sum test, with the Bonferroni-adjusted significance level of P less than .0167 (.05/3) to account for multiple testing. Inter- and intrareader reproducibility was performed for a subset of 20 individuals and calculated using the intraclass correlation coefficient with 95% CI from two-way random-effects model. Statistical analyses were performed using Stata 16.1 (Stata Corporation).

Results

Participant Characteristics

Fifty-five healthy volunteers (median age, 31 years [range: 21–66 years]; 32 female, 23 male) were enrolled in the study. Characteristics of study participants are provided in Table S2.

Acquisition Time

Acquisition was fastest for the 1RR cine sequence (pure sequence runtime: 19 seconds). The clinical standard sequence had the longest runtime (median: 84 seconds), followed by the 6RR cine and the 3RR cine sequences, with median runtimes of 68 seconds and 36 seconds, respectively (Table 1). Because the 1RR cine was completed in 1.9 breath holds (median) with a median duration of 12.5 seconds, this sequence had the fastest overall acquisition time of 45 seconds (range: 24–67 seconds), with a time reduction of 80% compared with the reference sequence ($P = .005$). The 3RR cine sequence (median acquisition time with pauses: 109 seconds [range: 65–132 seconds]) was more than 50% faster than the clinical reference (median acquisition time with pauses: 228 seconds [range: 168–275 seconds], $P = .005$). The 6RR cine sequence showed a 20% time reduction compared with the clinical cine sequence, with a total acquisition time of 183 seconds (range: 158–220 seconds). The detailed results are shown in Table 1.

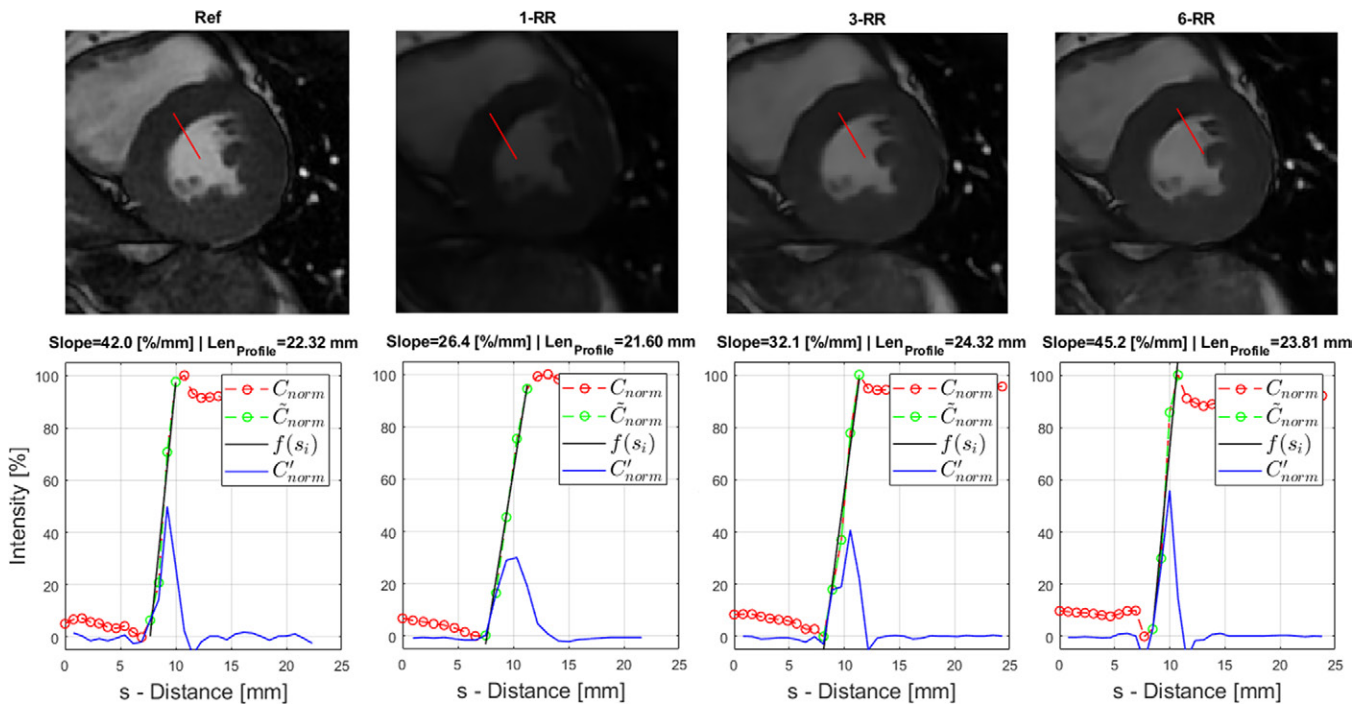


Figure 3: Quantification of edge sharpness. (Top) Axial noncontrast cardiac cine MR images for each sequence type, with red profile line. (Bottom) Graphs show the normalized intensity profile and regression line over the blood to myocardium transition for edge sharpness determination. C_{norm} = normalized intensity profile, $\tilde{C}_{norm}$ = intensity profile used for edge sharpness calculation, C'_{norm} = first derivative of C_{norm} , $f(s_i)$ = linear regression of intensity profile used for edge sharpness calculation, $Len_{Profile}$ = length of profile line from blood pool to myocardium, 1RR = one cardiac cycle cine sequence, Ref = reference sequence, s = distance, 6RR = six cardiac cycles cine sequence, 3RR = three cardiac cycles cine sequence.

Table 1: Acquisition Parameters and Image Quality for Each Cardiac MRI Sequence

Parameter	Reference		1RR Cine			3RR Cine			6RR Cine		
	Median	Range	Median	Range	<i>P</i> Value	Median	Range	<i>P</i> Value	Median	Range	<i>P</i> Value
Acquisition parameter											
Sequence time, without BH (sec)	84	67–100	19	13–26	.005	36	31–51	.005	68	54–85	.005
Acquisition time, with BH (sec)	227	168–275	47	34–67	.005	108	65–132	.005	184	158–220	.01
No. of BH	5.5		1.9		.005	3.4		.005	5.3		.047
Duration of BH (sec)	15.5		12.5		.07	15		.794	13.5		.23
Image quality											
Subjective image quality	4	3.5–5	3	1.5–4	<.001	4	3.5–5	.37	5	4–5	<.001
Edge sharpness	30.6	15.8–42.0	26.4	15.5–34.8	<.001	32.1	17.0–45.8	.06	37.4	16.4–51.5	<.001

Note.—All deep learning–based cine sequences were compared with the reference sequence. *P* values are from Wilcoxon matched pairs signed rank test. BH = breath hold, 1RR = one cardiac cycle, 6RR = six cardiac cycles, 3RR = three cardiac cycles.

Image Quality

Figure 4 shows examples of all four acquisition sequences (reference, 1RR cine, 3RR cine, and 6RR cine). Video files from the same healthy volunteers are provided in Movies 1–4. Subjective image quality was best for the 6RR cine (median score: 5, $P < .001$) and lowest for the 1RR cine (median score: 3, $P < .001$). Normalized edge sharpness as an objective parameter of image quality was also highest for the 6RR cine (median: 37.4) and lowest for the 1RR cine (median: 26.4). There was no evidence of a difference in edge sharpness between the reference cine (median: 30.6) and the 3RR cine (median: 32.1) (Table 1). Body mass index–stratified analysis of the image quality also showed no difference between reference cine and 3RR cine (see Table S3).

Postprocessing Time

Compared with the reference standard, the time required to correct the automated segmentation was significantly longer for the 1RR cine sequence (45 seconds vs 25.5 seconds, $P < .001$) but not for the 3RR and 6RR cines. The differences in the post-processing effort (correction time, number of section and phase corrections) are summarized in Table S4.

Volumetric Analysis

The median LV EDV, LV ESV, and LV SV of all participants determined with the reference sequence was 147 mL (range: 101–228 mL), 56 mL (range: 34–92 mL), and 92 mL (range: 58–148 mL), respectively. The median LV EF was 63% (range:

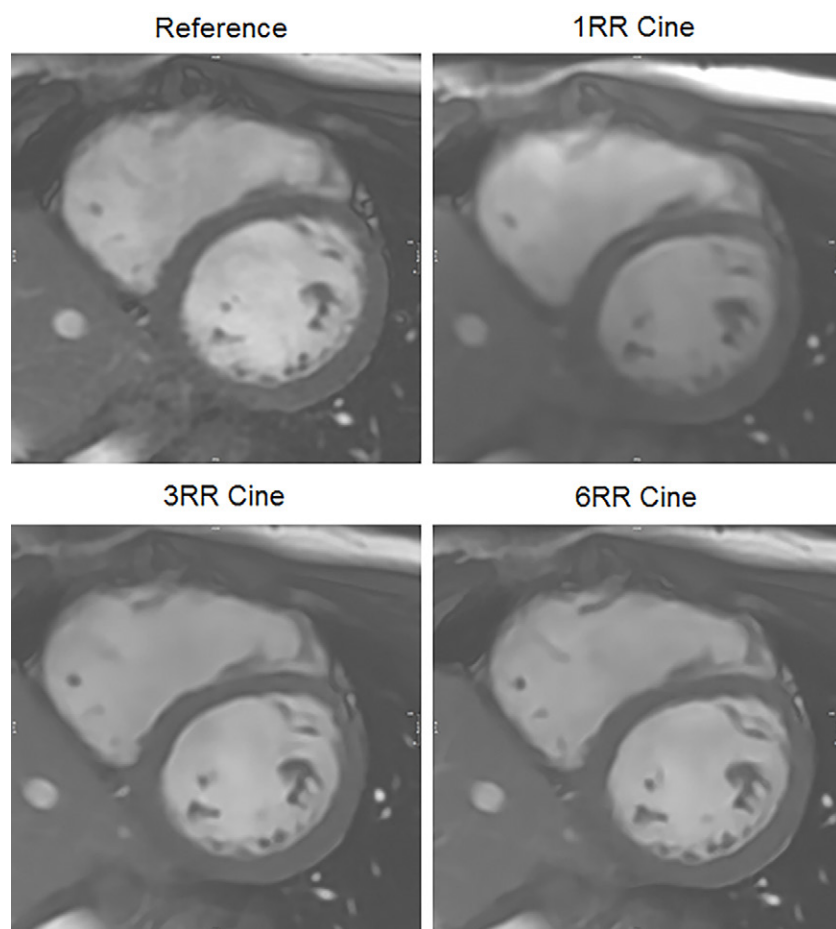


Figure 4: Image examples in a healthy volunteer (31-year-old male). Axial noncontrast cardiac MR images from reference cine sequence, one cardiac cycle (1RR) cine, three cardiac cycles (3RR) cine, and six cardiac cycles (6RR) cine are shown. Cine video files from this case are available as Movies 1–4.

50%–74%), and median LV mass was 91 g (range: 61–141 g) based on the reference sequence. The 1RR cine results significantly underestimated LV EDV (median: 138 mL), LV SV (median: 79 mL), and LV EF (median: 59%) compared with the reference sequence ($P < .001$ for all) and significantly overestimated LV mass (median: 95 g, $P < .001$). Volumetric results of the 3RR cine and 6RR cine showed no evidence of differences compared with the reference. LV EDV, LV ESV, LV mass, and LV SV (all indexed to body surface area) are visualized in Figure 5 and displayed in Table 2. Inter- and intra-reader reproducibility were overall very good, ranging from an intraclass correlation coefficient of 0.95 to 1.0 for intrareader reproducibility and from 0.93 to 1.00 for the interreader reproducibility (Table S5).

Right ventricular volumetric analysis showed similar trends as the LV analysis with moderate deviations for the 1RR cine sequence and no evidence of differences for the 3RR and 6RR cine sequences compared with the reference sequence (see Table S6).

Discussion

Most DL applications in cardiac MRI focus on segmentation tasks (10–15). There are scarce data on DL-based image reconstruction in cardiac MRI. To our knowledge, this is the first prospective cardiac MRI study comparing three different versions of an emerging undersampling prototype DL-based cine sequence with different acceleration rates (acquired over

one, three, or six cardiac cycles) with the clinical reference standard sequence. DL-based cardiac cine sequences performed over three cardiac cycles led to substantially reduced acquisition time, while no differences in image quality and volumetric results compared with a standard clinical sequence could be detected.

For the single-heartbeat cine, acquisition time was drastically reduced (up to 80%) and image quality was lower, while diagnostic and volumetric parameters were slightly different compared with the reference sequence. This ultrafast 1RR cine sequence could be an attractive option especially in patients with arrhythmia or who are unable to hold their breath (16). Zucker et al (8) showed similar results using free-breathing DL-based cine acquisitions. Sandino et al (7) also found that DL-based single-heartbeat cardiac cine imaging is feasible with small differences in volumetric parameters. We observed that DL-based single-heartbeat cine underestimated cardiac volumes and overestimated myocardial mass. This was caused by our approach of counting the papillary muscles toward myocardial mass and not toward the LV cavity. The slightly blurrier image quality in single-heartbeat cine leads to an overestimation of papillary muscle volume.

The 3RR cine seems to be the best compromise, with a reduction in acquisition time by more than 50% and no evidence of differences in volumetric parameters compared with the clinical reference. The 6RR cine delivered the best image quality, exceeding the clinical standard in image quality while still decreasing

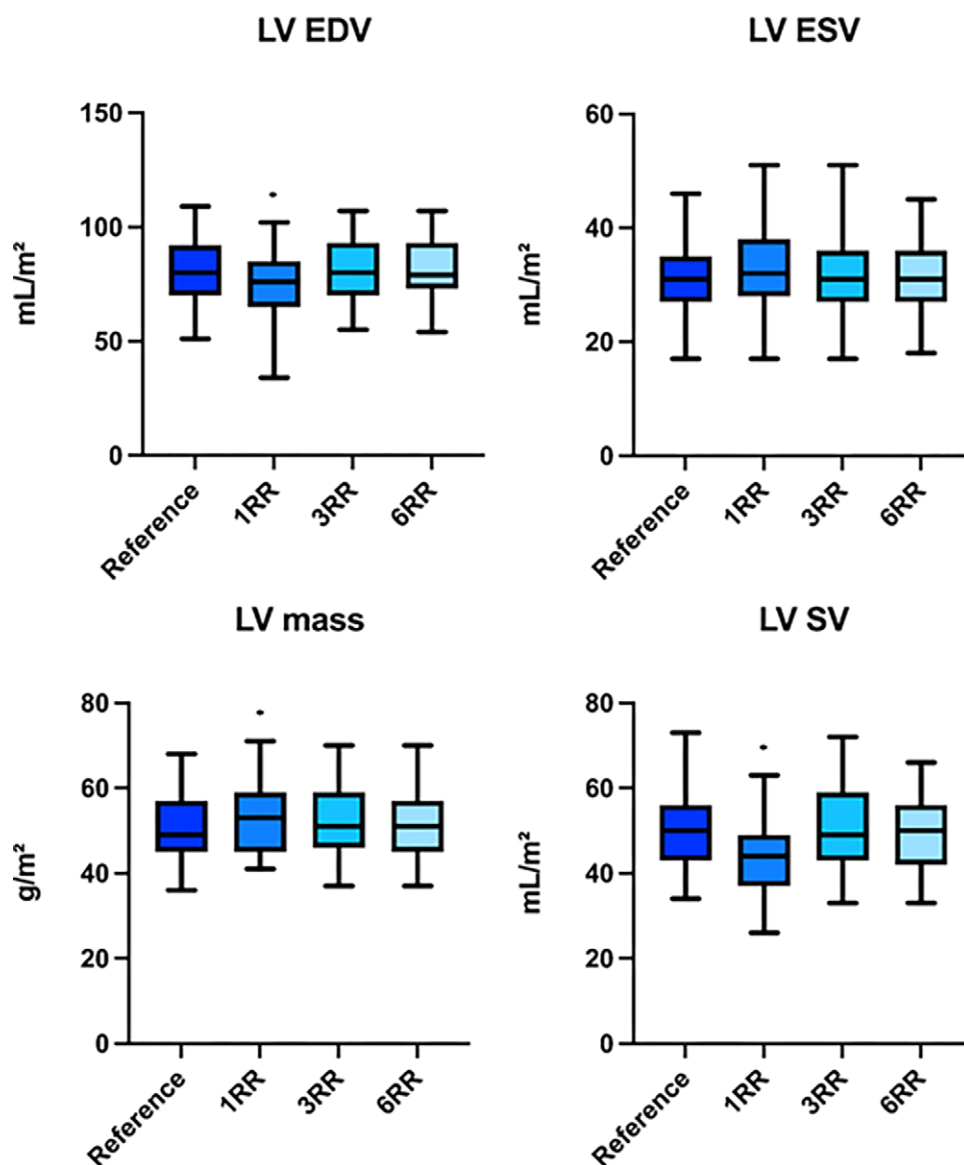


Figure 5: Volumetric results. Box plots show the results for LV EDV, LV ESV, LV SV, and LV mass indexed to body surface area for reference, one cardiac cycle (1RR), three cardiac cycles (3RR), and six cardiac cycles (6RR) cine sequences. Boxes represent the 5%–95% interval. The line inside each box indicates the median. Whiskers indicate the minimum and maximum values. LV EDV = left ventricular end-diastolic volume, LV ESV = left ventricular end-systolic volume, LV SV = left ventricular stroke volume.

the acquisition time. The subjectively improved image quality observed in 6RR cine is likely due to increased edge sharpness and spatial resolution. This may be beneficial for advanced diagnostic tasks, such as assessment of valvular morphology or strain analysis. Chen et al (11) previously found no evidence of differences in volumetric parameters between DL-reconstructed cine sequences and the clinical standard, with superior image quality for DL-based cine sequences. This is in line with our results for the 3RR and 6RR cine sequences.

Compared with conventional end-to-end networks, unrolled neural networks are more robust to varied acceleration and have lower risk of hallucination. Furthermore, a customized spatial-temporal convolution design is used as previously mentioned (7,8). This network uses less trainable parameters for more efficient training and inference performance.

This study had limitations. Only healthy volunteers and relatively young individuals who showed only a range of normal EF

values participated in our cardiac MRI study. Furthermore, the number of experienced readers for subjective image quality was limited, and we only tested one MRI device from one vendor. Additionally, our study was limited to cine images acquired in short-axis orientation. A technical limitation was the difference of in-plane resolution between the different sequences. These differences were due to the limit in acceleration. Further studies are required to shed light on the exact influence of temporal interpolation on temporal blurring, as this may vary across different acceleration rates.

In conclusion, our study demonstrates that DL-based image reconstruction enables highly accelerated cine imaging with preserved image quality and comparable volumetric results. Real-time cine imaging with DL reconstruction is feasible with very short acquisition time and good diagnostic image quality. The three-heartbeat DL cine sequence seems to be an ideal compromise, with substantially reduced acquisition time and preserved

Table 2: Volumetric Results for Left Ventricle

Volumetric Results	Reference		1RR Cine			3RR Cine			6RR Cine		
	Median	Range	Median	Range	<i>P</i> Value	Median	Range	<i>P</i> Value	Median	Range	<i>P</i> Value
Absolute value											
LV EDV (mL)	147	101–228	138	87–206	<.001	148	96–235	.41	145	99–230	.67
LV ESV (mL)	56	34–92	57	34–96	.05	58	34–100	.02	56	33–96	.11
LV SV (mL)	92	58–148	79	51–124	<.001	89	55–167	.26	87	58–152	.09
LV EF (%)	63	50–74	59	45–69	<.001	61	48–72	.04	62	50–71	.10
LV mass (g)	91	61–141	95	64–149	<.001	90	62–148	.22	92	61–147	.23
Indexed to body surface area											
LV EDV (mL/m ²)	79.8	51.4–109.1	75.6	34.2–101.5	<.001	79.9	55.5–107.3	.44	78.9	53.8–106.8	.73
LV ESV (mL/m ²)	30.7	16.6–46	32	16.6–51.5	.02	31	16.6–50.6	.02	31.3	18.3–45.1	.12
LV SV (mL/m ²)	50.1	34–73.3	43.5	25.5–63.2	<.001	48.5	33.2–71.9	.26	50.2	32.8–65.7	.09
LV mass (g/m ²)	48.9	36.1–68.3	53.1	40.6–70.7	<.001	50.7	37.2–70.2	.21	50.7	36.6–69.8	.02

Note.—All deep learning–based cine sequences were compared with the reference sequence. *P* values are from Wilcoxon matched pairs signed rank sum test. EDV = end-diastolic volume, EF = ejection fraction, ESV = end-systolic volume, LV = left ventricular, 1RR = one cardiac cycle, 6RR = six cardiac cycles, SV = stroke volume, 3RR = three cardiac cycles.

image quality and volumetric results compared with standard clinical sequences. Further investigations are necessary to evaluate the robustness and accuracy of DL-based image reconstruction in the clinical setting.

Author contributions: Guarantors of integrity of entire study, **X.Z., A.D., F.G.M.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **A.C.K., L.R., A.D., F.G.M.**; clinical studies, **A.C.K., M.G., A.D., F.G.M.**; statistical analysis, **M.M., A.D., R.L., F.G.M.**; and manuscript editing, all authors

Disclosures of conflicts of interest: **A.C.K.** No relevant relationships. **L.R.** No relevant relationships. **M.G.** Employee of GE HealthCare. **M.M.** No relevant relationships. **X.Z.** Employee of GE HealthCare. **A.D.** No relevant relationships. **R.L.** No relevant relationships. **C.I.L.** No relevant relationships. **M.A.W.** No relevant relationships. **F.G.M.** No relevant relationships.

References

- Kramer CM, Barkhausen J, Bucciarelli-Ducci C, Flamm SD, Kim RJ, Nagel E. Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update. *J Cardiovasc Magn Reson* 2020;22(1):17.
- Emrich T, Halfmann M, Schoepf UJ, Kreitner K-F. CMR for myocardial characterization in ischemic heart disease: state-of-the-art and future developments. *Eur Radiol Exp* 2021;5(1):14.
- Mazurowski MA, Buda M, Saha A, Bashir MR. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on MRI. *J Magn Reson Imaging* 2019;49(4):939–954.
- Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci* 2019;50(4):477–487.
- Orii M, Sone M, Osaki T, et al. Reliability of respiratory-gated real-time two-dimensional cine incorporating deep learning reconstruction for the assessment of ventricular function in an adult population. *Int J Cardiovasc Imaging* 2023;39(5):1001–1011.
- Lai P, Brau A. Improving cardiac cine MRI on 3T using 2D k-t accelerated auto-calibrating parallel imaging. *J Cardiovasc Magn Reson* 2014;16(Supplement 1):W3.
- Sandino CM, Lai P, Vasawala SS, Cheng JY. Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction. *Magn Reson Med* 2021;85(1):152–167.
- Zucker EJ, Sandino CM, Kino A, Lai P, Vasawala SS. Free-breathing accelerated cardiac MRI using deep learning: validation in children and young adults. *Radiology* 2021;300(3):539–548.
- Böttcher B, Beller E, Busse A, et al. Fully automated quantification of left ventricular volumes and function in cardiac MRI: clinical evaluation of a deep learning-based algorithm. *Int J Cardiovasc Imaging* 2020;36(11):2239–2247.
- Ammann C, Hadler T, Gröschel J, Kolbitsch C, Schulz-Menger J. Multilevel comparison of deep learning models for function quantification in cardiovascular magnetic resonance: On the redundancy of architectural variations. *Front Cardiovasc Med* 2023;10:1118499.
- Chen EZ, Chen X, Lyu J, et al. Cardiac functional analysis with cine MRI via deep learning reconstruction. *arXiv 2105.08157* [preprint] <https://arxiv.org/abs/2105.08157>. Posted May 17, 2021. Accessed December 1, 2023.
- Das N, Das S. Epoch and accuracy based empirical study for cardiac MRI segmentation using deep learning technique. *PeerJ* 2023;11:e14939.
- Govil S, Crabb BT, Deng Y, et al. A deep learning approach for fully automated cardiac shape modeling in tetralogy of Fallot. *J Cardiovasc Magn Reson* 2023;25(1):15.
- Åkesson J, Ostenfeld E, Carlsson M, Arheden H, Heiberg E. Deep learning can yield clinically useful right ventricular segmentations faster than fully manual analysis. *Sci Rep* 2023;13(1):1216.
- Wang L, Su H, Liu P. Automatic right ventricular segmentation for cine cardiac magnetic resonance images based on a new deep atlas network. *Med Phys* 2023;50(11):7060–7070.
- Monteuuis D, Bouzerar R, Dantoing C, Poujol J, Bohbot Y, Renard C. Prospective comparison of free-breathing accelerated cine deep learning reconstruction versus standard breath-hold cardiac MRI sequences in patients with ischemic heart disease. *AJR Am J Roentgenol* 2024;222(5):e2330272.



OPEN Development and performance evaluation of fully automated deep learning-based models for myocardial segmentation on T1 mapping MRI data

Mathias Manzke, Simon Iseke, Benjamin Böttcher, Ann-Christin Klemenz, Marc-André Weber & Felix G. Meinel

To develop a deep learning-based model capable of segmenting the left ventricular (LV) myocardium on native T1 maps from cardiac MRI in both long-axis and short-axis orientations. Models were trained on native myocardial T1 maps from 50 healthy volunteers and 75 patients using manual segmentation as the reference standard. Based on a U-Net architecture, we systematically optimized the model design using two different training metrics (Sørensen-Dice coefficient = DSC and Intersection-over-Union = IOU), two different activation functions (ReLU and LeakyReLU) and various numbers of training epochs. Training with DSC metric and a ReLU activation function over 35 epochs achieved the highest overall performance (mean error in T1 10.6 ± 17.9 ms, mean DSC 0.88 ± 0.07). Limits of agreement between model results and ground truth were from -35.5 to $+36.1$ ms. This was superior to the agreement between two human raters (-34.7 to $+59.1$ ms). Segmentation was as accurate for long-axis views (mean error T1: 6.77 ± 8.3 ms, mean DSC: 0.89 ± 0.03) as for short-axis images (mean error $\Delta T1$: 11.6 ± 19.7 ms, mean DSC: 0.88 ± 0.08). Fully automated segmentation and quantitative analysis of native myocardial T1 maps is possible in both long-axis and short-axis orientations with very high accuracy.

Keywords Deep learning, U-Net, Mapping, Cardiac magnetic resonance imaging, Long axis, Short axis

Abbreviations

4Ch	Four-chamber view
AHA	American heart association
API	Apical
BAS	Basal
CNN	Convolutional neural network
DSC	Sørensen-dice-coefficient/dice score
DICOM	Digital imaging and communications in medicine
ECG	Electrocardiogram
IOU	Intersection over union (Jaccard Coefficient)
IQR	Interquartile range
IR	Inversion-recovery
LAX	Long axis
LeakyReLU	Leaky rectified linear unit
MID	Mid-ventricular
MOLLI	MODified Look-Locker Inversion recovery
MRI	Magnetic resonance imaging
mSASHA	Modified saturation recovery single-shot acquisition

Institute of Diagnostic and Interventional Radiology, Pediatric Radiology and Neuroradiology, University Medical Centre Rostock, Ernst-Heydemann-Str. 6, 18057 Rostock, Germany. email: felix.meinel@med.uni-rostock.de

SAX Short axis
ReLU Rectified linear unit

Cardiovascular diseases account for approximately 32% of all deaths worldwide, making them the leading cause of death worldwide¹. Cardiac magnetic resonance imaging (MRI) has provided various tools for pathology detection for many years and is routinely used to assess the structure and function of the cardiovascular system. In cardiac MRI, mapping sequences are used for quantitative tissue characterization of the myocardium². The myocardial T1 relaxation reflects biophysical properties of myocardial muscle cells and the surrounding extracellular space. Altered myocardial T1 relaxation times typically result from fibrosis, inflammation, edema or storage disease³. Quantification of T1 relaxation times is therefore helpful to diagnose pathologies affecting the myocardium.

In clinical practice, cardiovascular imaging specialists often draw regions of interest (ROI) within the T1 maps manually to obtain T1 relaxation times. This process is time-consuming, tedious and prone to variation between readers. Therefore, fully automated contouring of mapping data would be beneficial to efficiently and objectively quantify T1 relaxation times.

There is extensive literature on automated segmentation of cardiac cine sequences for volumetric analysis^{4–12}. Only few publications have addressed the challenge of automated segmentation of mapping data^{13–19}. All of these studies have been limited to T1 maps acquired in the short-axis (SAX) plane perpendicular to the long axis of the left ventricle and have not investigated whether this is also feasible for long-axis (LAX) images. Thus, there is a paucity of evidence for fully automated segmentation of T1 mapping evidence and a lack of data on the successful development of neural networks for segmenting both SAX and LAX T1 maps with high accuracy.

Therefore, the purpose of this study was to develop and validate a deep learning-based model capable of segmenting the left ventricular myocardium on native T1 maps from cardiac MRI both in LAX and SAX orientation.

Material and methods

The study was conducted in three phases. In the first phase, the inclusion criteria for data collection were defined and the data was exported from the clinical systems. In the second phase, four different models based on a U-Net architecture were trained over several training epochs. In the third phase, the results of all models were evaluated using a test data set and compared with the ground truth data set of both evaluators. An overview of the overall study design can be seen in Fig. 1.

Patient selection and study design

We included 50 healthy participants whose imaging data from a previous prospective volunteer study²⁰ were used and 75 patients who underwent a clinically indicated CMR at 1.5 T between 2020 and 2022 at our institution. All MRI examinations were performed on a 1.5 T MRI scanner (Avanto^{fit} Siemens Healthineers, Erlangen, Germany). Details of imaging protocols and the process of creating the T1 maps have been published previously²⁰. Patients were retrospectively selected by searching our institution's radiology information system. Patients were excluded for the following reasons: severe image artifacts ($n = 12$), children ($n = 3$), number of short-axis T1 maps $\neq$ three ($n = 4$) and data set errors ($n = 4$). Figure 2 shows the data collection process.

In the first step, MRI cine images were collected from the 50 healthy volunteers in two rounds and the native T1 maps generated by the scanner were exported. The volunteer dataset consisted of 300 SAX and 100 4Ch image data. In the second step, data from another 75 patients (225 SAX images) were extracted from the clinical systems and segmented by two assessors R1 and R2. In addition, a subset of the patient data-set, with 30 patients ($n = 90$ images), was segmented by both assessors to evaluate interrater variability.

The Table 1 below shows how each view is distributed across the training, validation, and test data partitions.

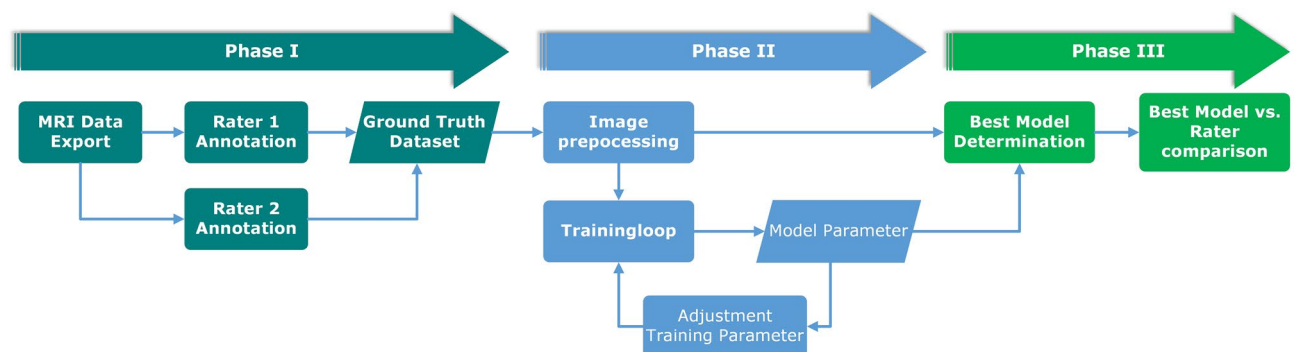


Figure 1. Entire study design flowchart divided in three phases: Phase I (left) contains the data inclusion criteria definition process as well as the data collection and annotation procedure performed by two independent raters who created the ground truth data set; Phase II (center) contains image preprocessing and file operations as well as model specifications and the training loops; During Phase III (right) all post-processing procedures and final analysis were evaluated.

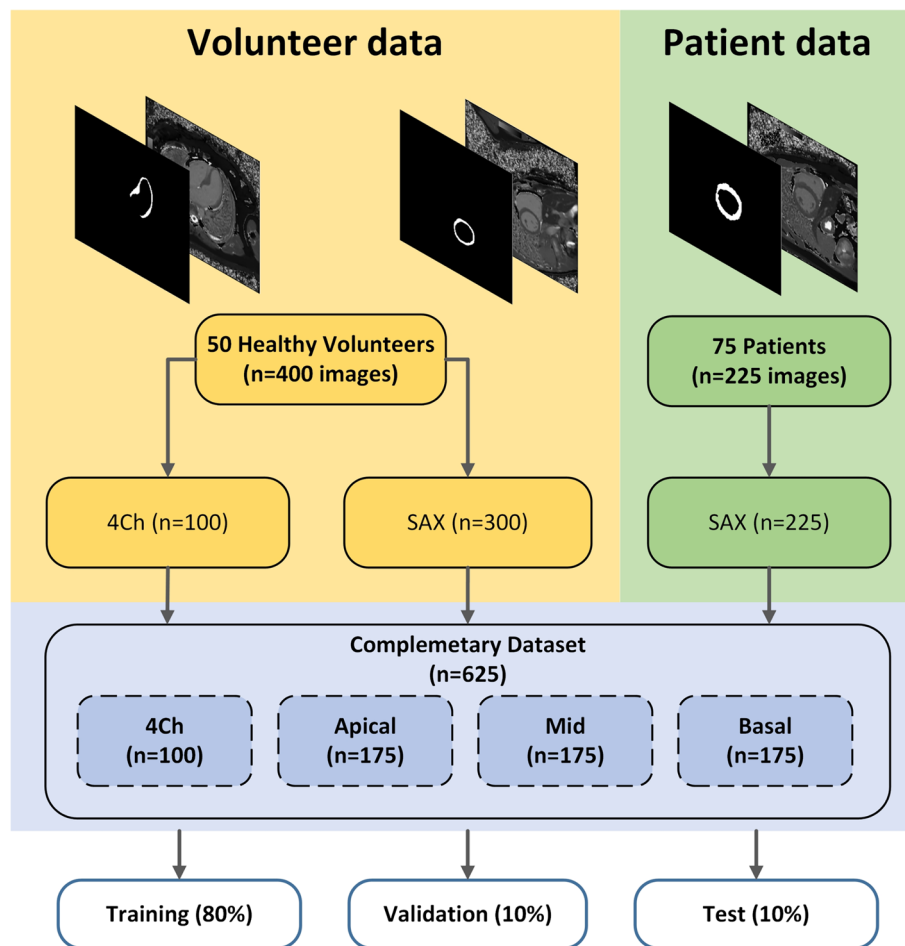


Figure 2. Description of complementary dataset: The visualization shows how the two independently assessed datasets of healthy subjects ($n = 50$) and patients ($n = 75$) from routine care were assembled, how many views exist and on which partitions the dataset is distributed.

Cohort/View	All	4Ch	API	MID	BAS
Training	500 (80%)	79 (12.64%)	138 (22.08%)	146 (23.36%)	137 (21.92%)
Validation	62 (10%)	8 (1.28%)	17 (2.72%)	14 (2.24%)	23 (3.68%)
Test	63 (10%)	13 (2.08%)	20 (3.20%)	15 (2.40%)	15 (2.40%)

Table 1. Data partitions overview. The complementary dataset collected and annotated by two raters is divided in a training, a validation and a test dataset where each subset contains four different views (four chamber view, apical, mid-ventricular, and basal).

Ethical approval and informed consent

This study was approved by the institutional review board (Ethics Committee, University Medical Center Rostock) and written informed consent was obtained from all 50 healthy volunteers free from cardiovascular disease prior to enrollment. For the additional 75 patients the requirement for informed consent was waived due to the retrospective nature of this part of the investigation. The study conformed to the ethical guidelines of the Helsinki Declaration (revised version from 2013).

Manual Image segmentation

Both data sets, volunteer data and patient data, were annotated independently by two experienced assessors. Manual segmentation was performed using an open source software solution (ITK-Snap, version 4.0.0).

Image preprocessing

We selected a U-Net architecture to automatically segment the image data. To achieve good results with smaller training data sets, U-Nets can be effectively trained in combination with data augmentation techniques such as mirroring, rotation, and scaling²¹. These properties are particularly helpful because in the first preprocessing step before training the resolution of the existing MRI images was scaled from $(256 \times 218, \text{pixel size } 1.41 \times 1.41 \text{ mm})$ to a uniform format $(512 \times 512, \text{pixel size } 0.70 \times 0.60 \text{ mm})$. For practical implementation, scaling makes sense in order to convert the data into a uniform input format and thus be independent of image sizes and pixel spacing. We initially considered performing image padding to avoid warping and maintain square pixels. However, this reduced segmentation accuracy and was therefore not implemented.

To ensure that the residual interpolation error for each individual segmentation mask remained as small as possible, we evaluated various scaling methods (bilinear, Lanczos, Bicubic, pixel-area relation, and nearest neighbor) and determined the influence on the error. For this purpose, mean T1 times were compared between original images (resolution 256×218 pixels) and scaled images (resolution 512×512). Nearest neighbor interpolation was found to introduce minimal error. No additional data augmentation techniques were used to increase the data volume. Grey-scale images were used without intensity normalization.

Network architecture

Convolutional layers were used in both the encoder and decoder parts, followed by Rectified Linear Unit (*ReLU*) or *LeakyReLU* activation functions. The encoder sublayers were followed by the *MaxPool* layer to extract global features from the input images, which led to loss of spatial information. In the decoder part, the convolutions were transposed to reconstruct the spatial resolution step by step and create detailed segmentation maps by transforming the activations in higher layers back to the original input size. Skip connections were inserted between the encoder and the decoder and were used to pass features from the encoder to the decoder. This is useful for combining both local and global information and obtaining fine details. The final layer had only one output channel to which the sigmoid activation function was applied to generate the pixel class probabilities, allowing the creation of binary segmentation maps.

The effects of two activation functions were compared during model performance evaluation, as proposed in a previous work²².

Therefore, the model was evaluated separately using each of these functions in image segmentation. The activation functions were implemented in the decoder block of the model. For the standard activation function *ReLU*, the output of the neuron corresponds to the level of activation if it is in the positive value range. The output of a neuron can therefore increase arbitrarily with increasing activation. If activations $x < 0$ the neuron remains deactivated. An updated version of *ReLU* was also used, the *LeakyReLU*. It is one of its upgrades and passes information with a scaling factor $\alpha = 0.1$.

Training

To train the deep learning model, cardiac T1 maps of the left ventricle in short-axis view (SAX) and four-chamber view (4Ch) with the corresponding segmentation masks were used as input. The complementary data set, consisting of 625 DICOM images, was randomly divided into three independent subsets: training, validation, and testing data with, 80% of the data set defined for training, 10% of the data set defined for interim validation (during training), and the remaining 10% of the data used for testing. Model training was performed on a Workstation (CPU i7-12,700 2.10 GHz / 32 GB RAM; NVIDIA RTX A2000 / 6 GB GRAM) with CUDA 11.6, Python Tensorflow 2.12.0, Numpy 1.23.5, PyDICOM 2.4.2 and Medpy 0.3.0.

During the training loop, the performance of each model was evaluated using binary accuracy (pixel-wise accuracy) and a metric (DSC and IOU). In particular, two metrics $DSC(S_1, S_2) = 2|S_1 \cap S_2| / (|S_1| + |S_2|)$ and $IOU(S_1, S_2) = |S_1 \cap S_2| / |S_1 \cup S_2|$ have been proposed in the relevant literature^{22–24} to measure the congruence of two pixel sets S_1 and S_2 ²⁵.

The training data set was divided into batches of 16. A training loop of 5 to 70 training epochs was implemented in steps of 5 epochs using the ADAM optimizer²⁶ in combination with variable activation layers and a decreasing learning rate.

Testing and performance evaluation

The mean T1 times of the ROIs for all trained models were calculated using the test data set, consisting of 63 images in different views (4Ch and SAX). This dataset can be viewed as an external validation dataset in previously unseen images. In addition, the absolute differences in T1 ($\Delta T1$) between the predicted segmentations and the ground truth segmentations as well as the metrics DSC and IOU were determined.

The best model was considered as the model minimizing the error in T1 time quantification ($\Delta T1$) and maximizing both metrics of segmentation accuracy ($\overline{DSC}$ and $\overline{IOU}$). For this purpose

a function $D(\Delta T1, DSC, IOU)$ is defined in Eq. (1), which was calculated for all 56 models and all 63 images from the test data set. The arithmetic averages of $\Delta T1$, $\overline{DSC}$ and $\overline{IOU}$ were calculated over all 63 absolute values for each model. Because $\Delta T1$ should be as small as possible and DSC and IOU as large as possible, the smallest $\Delta T1_j$ average and the largest of $\overline{DSC}_j$ and $\overline{IOU}_j$ were then sought and the minimum of the function D was determined using the least squares method.

We determined the best fitting model as the one minimizing the function D_j with the definition

$$D_j^2 := (\overline{\Delta T1}_j - \min(\overline{\Delta T1}_j))^2 + (\overline{DSC}_j - \max(\overline{DSC}_j))^2 + (\overline{IOU}_j - \max(\overline{IOU}_j))^2, \quad (1)$$

where $j = 1, \dots, 56$ is the model number.

In the second step, the identified best model was compared with the ground truth segmentations of the first rater (R1) and the second rater (R2) as well as the interrater variability, as shown in Fig. 3.

Statistical analysis

Geometrical segmentation accuracy was evaluated by DSC (Sørensen-Dice-Coefficient) and IOU (Intersection over union). Error in quantitative mapping results was quantified as the error in T1 relaxation times ($\Delta T1$). Bland–Altman analysis was performed to obtain limits of agreement and mean bias between model results. In order to compare model performance and inter-rater variability for the best model, differences in $\Delta T1$, DSC, and IOU were compared between model vs. human and human vs. human (interrater variability) using the Mann–Whitney U test. A value of $p < 0.05$ was considered statistically significant. Statistical analyses were performed using the Python (3.11.2) libraries Scipy 1.10.1 and Seaborn 0.12.2.

Results

Patient characteristics and CMR indications

Our cohort included 50 healthy volunteers (29 female, 21 male, mean age 39.4 ± 13.7 years) and 75 patients (29 female, 46 male, mean age 39.4 ± 13.7 years). The most common CMR indications in patients were known or suspected myocarditis, cardiomyopathy, and ischemic heart disease with 26 (34.7%), 15 (20.0%) and 12 (16.0%) patients, respectively.

Selection of the optimal model based on overall model performance

As a result of the image preprocessing, the maximum absolute interpolation error $\Delta T1$ for the scaling process from (256×218) to (512×512) resolution was $\Delta T1 = \pm 4$ ms, which was considered clinically irrelevant considering that T1 values typically range from 900 to 1200 ms and the limits of agreement for test–retest–reproducibility of T1 mapping are approximately -39 to $+34$ ms²⁰. As detailed in the methods section above, the selection of the optimal model was based on a comprehensive strategy minimizing error for the quantitative T1 results ($\Delta T1$) and optimize the segmentation accuracy (DSC, IOU). All three parameters were weighted equally. Figure 4 and Fig. 5 show the overall model performance for all four neural networks as a function of training epochs. A convergence of performance metrics over the number of training epochs is seen, particularly with regards to segmentation accuracy (Fig. 5) with differences becoming small between 25 and 55 training epochs.

Based on Eqs. (1)–, [EP35DSCORE] was determined to be the model with optimal overall performance based on $\Delta T1$, DSC, and IOU (Table 2).

The table shows the aggregated metrics and relaxation times for the ten best models for all of the 63 testing images in all views (4Ch and SAX). In this context, best means a minimal ($\Delta T1$) and maximum metrics.

The chosen model is marked in green in Fig. 4 and Fig. 5. For this model, mean error in T1 relaxation time was 10.6 ± 17.9 ms, mean DSC was 0.88 ± 0.07 , and mean IOU was 0.79 ± 0.09 . Specifically, this model shows very high agreement for T1 results with a narrow interquartile range (IQR) of $\Delta T1$ from -7 ms to 5 ms and an overall range of $\Delta T1$ from -25 ms to $+23$ ms (Fig. 4). Therefore, the [EP35DSCORE] model was selected for detailed performance evaluation.

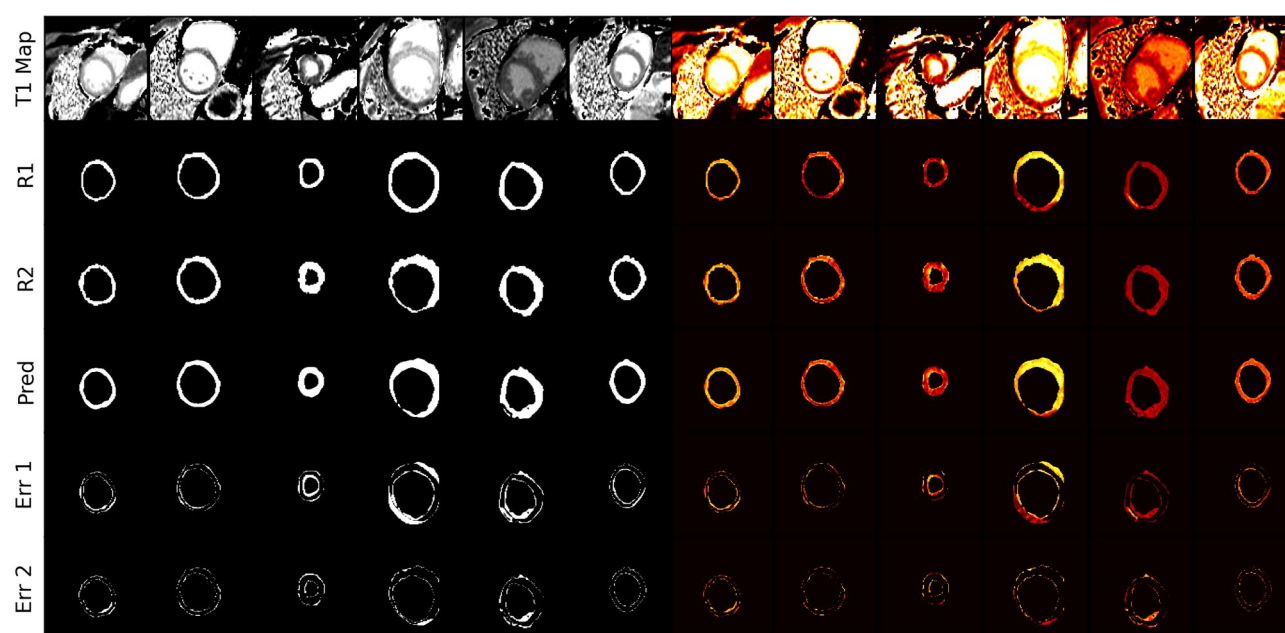


Figure 3. SAX ROI comparison between both Raters and Prediction: In the comparison between original T1 map both raters R1 and R2 and the predicted mask show some slight differences in lower rows Err1 (Prediction–R1 reference) and Err2 (Prediction–R2 reference), Binary masks (left) and intensity masks (right) with a zoom factor of 2.3.

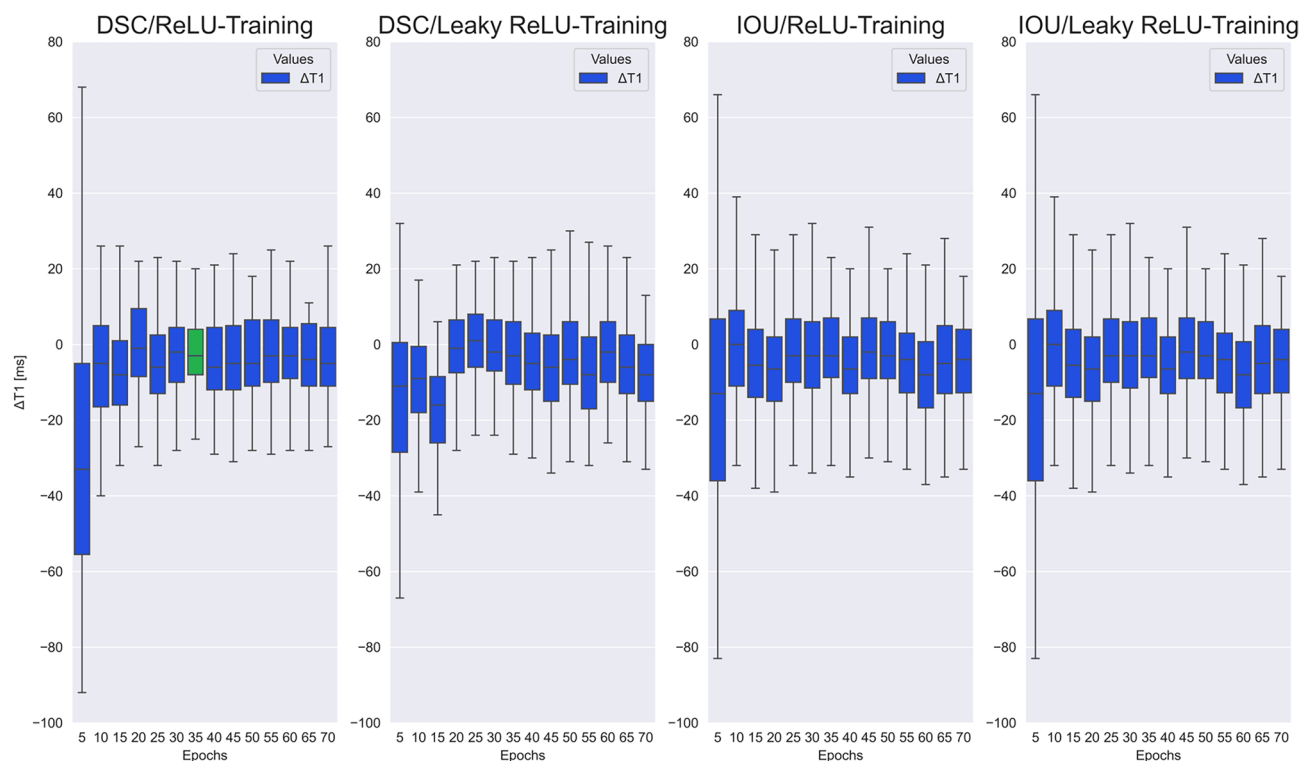


Figure 4. Test cohort model performance with $\Delta T1$: Performance measurement comparison for all 56 trained models with two activation functions (ReLU and Leaky ReLU) and two metrics (DSC and IOU) over all views (SAX and 4Ch) regarding the difference of the averaged ROI relaxation times $\Delta T1$ between prediction and ground truth test data set.

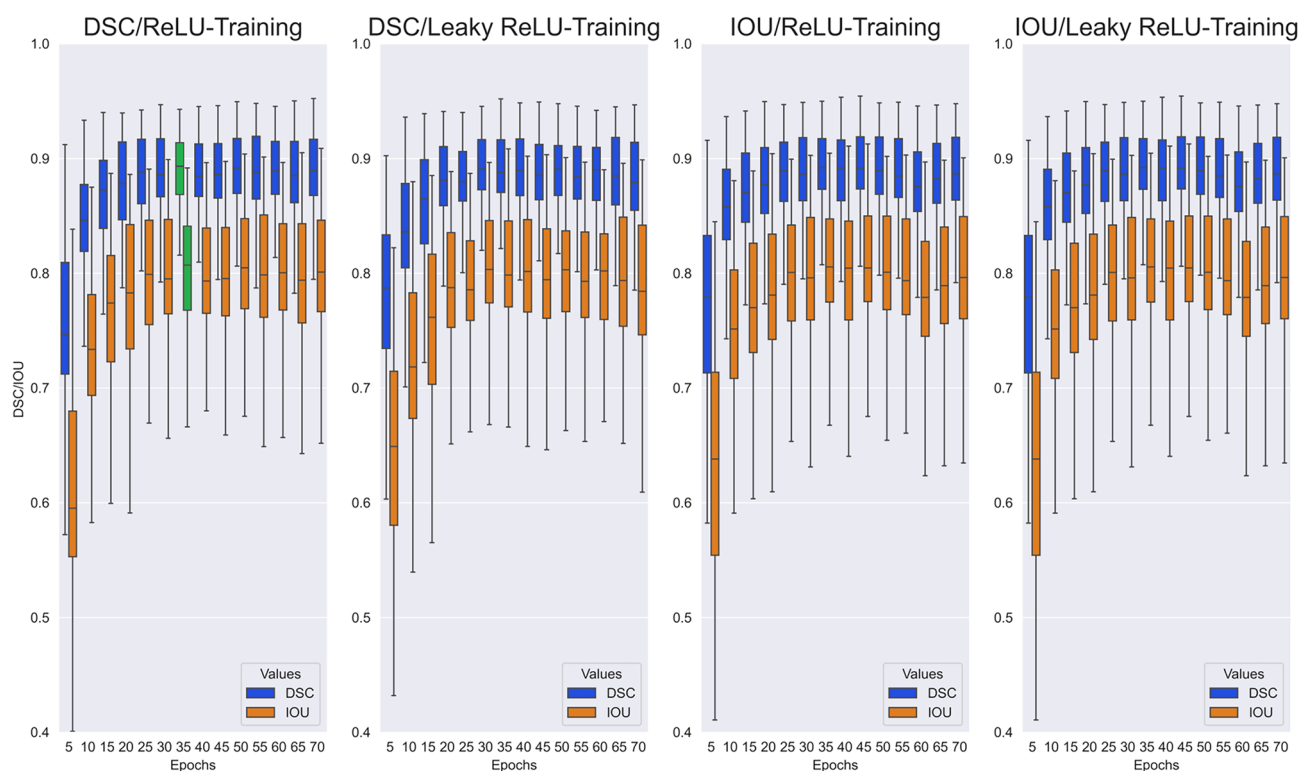


Figure 5. Test cohort model performance with Metrics (DSC/IOU): Performance measurement comparison for all 56 trained models with two activation functions (ReLU and Leaky ReLU) and two metrics (DSC and IOU) and over all views (SAX and 4Ch) regarding the segmentation overlapping between prediction and ground truth test data set also measured with DSC and IOU.

Model	$\Delta T1$	DSC	IOU	D
EP35DSCORE	10.571 (17.891)	0.879 (0.069)	0.790 (0.088)	0.013
EP50IOULR	10.714 (15.839)	0.885 (0.049)	0.797 (0.071)	0.143
EP55IOURE	10.746 (14.612)	0.877 (0.080)	0.787 (0.096)	0.175
EP45IOURE	11.063 (18.483)	0.887 (0.061)	0.801 (0.082)	0.492
EP25IOURE	11.095 (18.278)	0.881 (0.064)	0.791 (0.083)	0.524
EP50IOURE	11.222 (17.111)	0.881 (0.073)	0.794 (0.091)	0.651
EP45IOULR	11.302 (17.875)	0.884 (0.060)	0.796 (0.080)	0.730
EP70DSCORE	11.317 (15.861)	0.881 (0.073)	0.793 (0.091)	0.746
EP35IOULR	11.333 (19.683)	0.885 (0.058)	0.798 (0.080)	0.762
EP55DSCORE	11.413 (16.531)	0.879 (0.073)	0.789 (0.092)	0.841

Table 2. Model comparison for all views. Significant values are in bold.

Detailed performance evaluation for the optimal model

Figure 6 shows the performance of the model [EP35DSCORE] for all views. The IQR was $\Delta T1 \leq 15$ ms for all views with the deviations being predominantly negative in for the SAX views and predominantly positive in for the 4Ch views. Segmentation was equally accurate for long-axis views (mean error $\Delta T1$: 6.77 ± 8.3 ms, mean DSC: 0.89 ± 0.03 , mean IOU: 0.81 ± 0.04) as for short-axis images (mean error $\Delta T1$: 11.6 ± 19.7 ms, mean DSC: 0.88 ± 0.08 , mean IOU: 0.79 ± 0.1). On the short axis, performance for the apical slices was inferior (mean error $\Delta T1$: 15.9 ± 28.2 ms, mean DSC: 0.84 ± 0.11 , mean IOU: 0.73 ± 0.1) compared to basal (mean error $\Delta T1$: 7.9 ± 9.7 ms, mean DSC: 0.91 ± 0.02 , mean IOU: 0.84 ± 0.04) and mid-ventricular (mean error $\Delta T1$: 9.4 ± 12.3 ms, mean DSC: 0.89 ± 0.02 , mean IOU: 0.81 ± 0.04) slices.

Examples of predicted and ground truth masks are shown in Fig. 7. On Bland–Altman analysis, limits of agreement between model results and ground truth were from -35.5 to $+36.1$ ms (Fig. 8). In 92.1% of cases, the error $\Delta T1$ was within ± 20 ms.

Model performance in the context of inter-rater variability

We further aimed to quantitatively compare the variability between DL-based model vs. human evaluation to the inter-rater variability between two human assessors. The limits of agreement between model results and ground truth were from -35.5 to $+36.1$ ms (Fig. 8), this was superior to the agreement between two human assessors

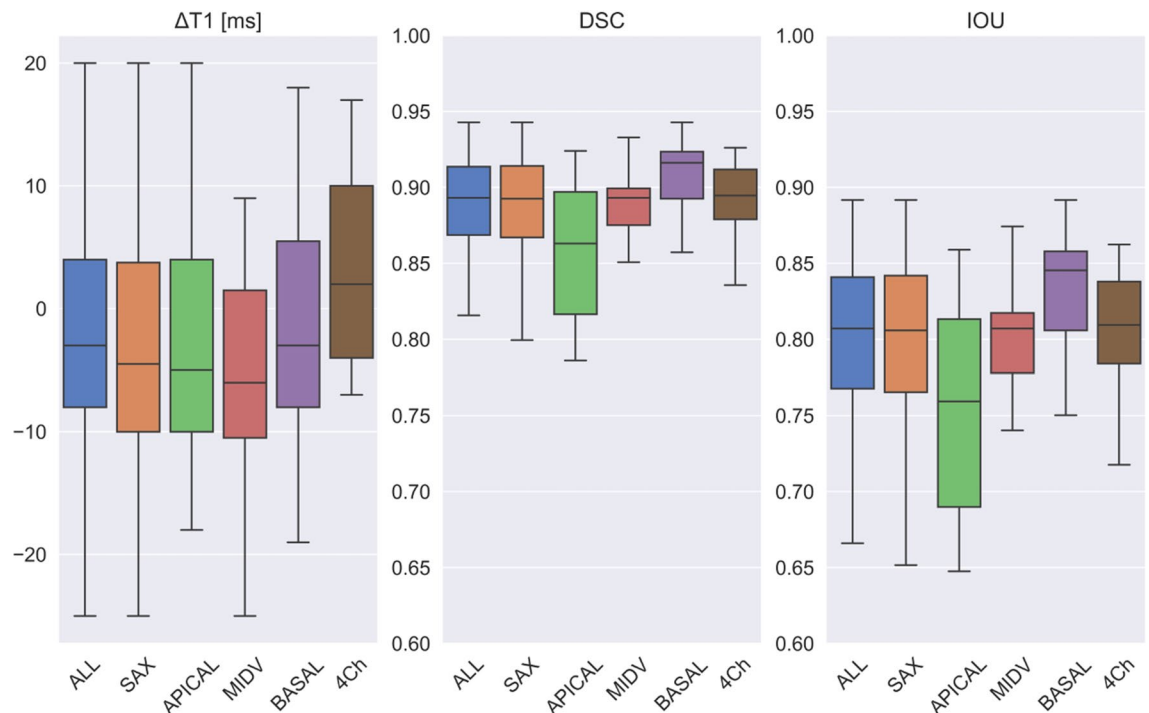


Figure 6. Model statistics [EP35DSCORE]: Boxplot results for all measured metrics ($\Delta T1$, DSC and IOU) in the test cohort between model predictions and ground truth mask, shown for all views, short-axis views only and separately for each view (apical, midventricular, basal, and 4Ch).

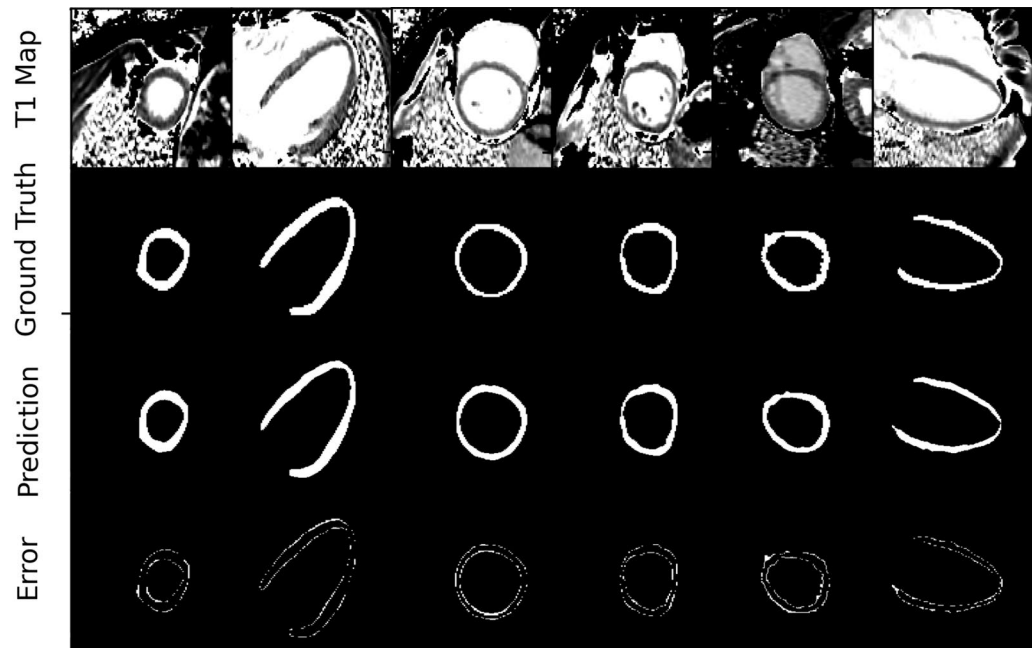


Figure 7. Best model [EP35DSCORE] predictions and ground truth masks: Example of comparison between predicted and ground truth binary masks and segmentation error with a zoom factor of 2.3.

(-34.7 to +59.1 ms). Furthermore, the model's mean prediction bias (0.29 ms) was significantly smaller than the mean interrater bias (-12.2 ms).

In a more detailed analysis of interrater variability, we further found that the DL-based model's segmentation accuracy correlated with inter-reader agreement: Both model accuracy and inter-reader agreement were inferior for apical slices and superior for basal and mid-ventricular slices (compare Fig. 6 and Fig. 9). For $\Delta T1$, we found that the error of model vs. human was significantly smaller than the difference human vs. human (interrater variability) specifically for apical slices ($p = 0.013$), while this was not significant for mid-ventricular and basal slices (Table 3). Geometrical segmentation accuracy of model vs. human was significantly higher than the agreement between two human segmentations for all short-axis slice positions.

Discussion

Related work

While there is abundant literature on fully automated segmentation in cardiac MRI, the majority of these studies have focused on cine sequences (aimed at visualizing and quantifying ventricular and atrial function). In contrast, mapping is performed to quantitatively evaluate biophysical properties of the myocardium, which can hint at underlying cardiac diseases. Only few studies have investigated automated segmentation of T1 maps. In one retrospective study a convolutional neural network was trained to segment myocardial T1 and T2 maps using an edge probability estimation approach¹⁸. This previous study was based on a different sequence type (mSASHA – with T1 and T2 maps acquired simultaneously), while our study is based on the more commonly used Modified Look-Locker inversion recovery (MOLLI) sequence. In contrast to our analysis, Howard and colleagues included only short-axis maps and chose a segment-analysis based on the AHA model of 16 myocardial segments. Like in our study, the agreement in segmentation between the neural network and human experts (DSC: 0.82–0.86) was comparable to the agreement between two human experts (DSC: 0.84).

Bhatt and colleagues employed synthetic contrast augmentation in the development and testing of a deep learning-based algorithm for segmentation of T1 maps¹⁴. Their method demonstrated high accuracy in myocardial segmentation (DSC: 0.81) and T1 calculation (correlation coefficient $R = 0.87$). Although they were able to show that this approach provides a small increase in accuracy, synthetic contrast augmentation substantially increases the complexity of the network architecture and computational demands. We therefore aimed to develop an algorithm which works on “raw” T1 maps without requiring artificial contrast augmentation.

Several other previous studies have developed algorithms for automated T1 map segmentation with notable difference in network architecture and the underlying imaging dataset^{15,16}. Two independent studies developed T1 map segmentation pipelines that combine both segmentation and quality-control steps^{13,17}. In this approach, not one but several models and combined models are employed to segment the myocardium on each dataset. From the resulting segmentations, a quality scoring model then selects the optimal segmentation based on predicted segmentation accuracy. In a similar approach, Puyol-Antón and colleagues developed a neural network with Bayesian inference for T1 map segmentation which additionally provides information on segmentation uncertainty¹⁹. This could be an interesting approach for clinical routine in which the human reader is alerted to review areas of uncertainty after a fully automated segmentation has been carried out.

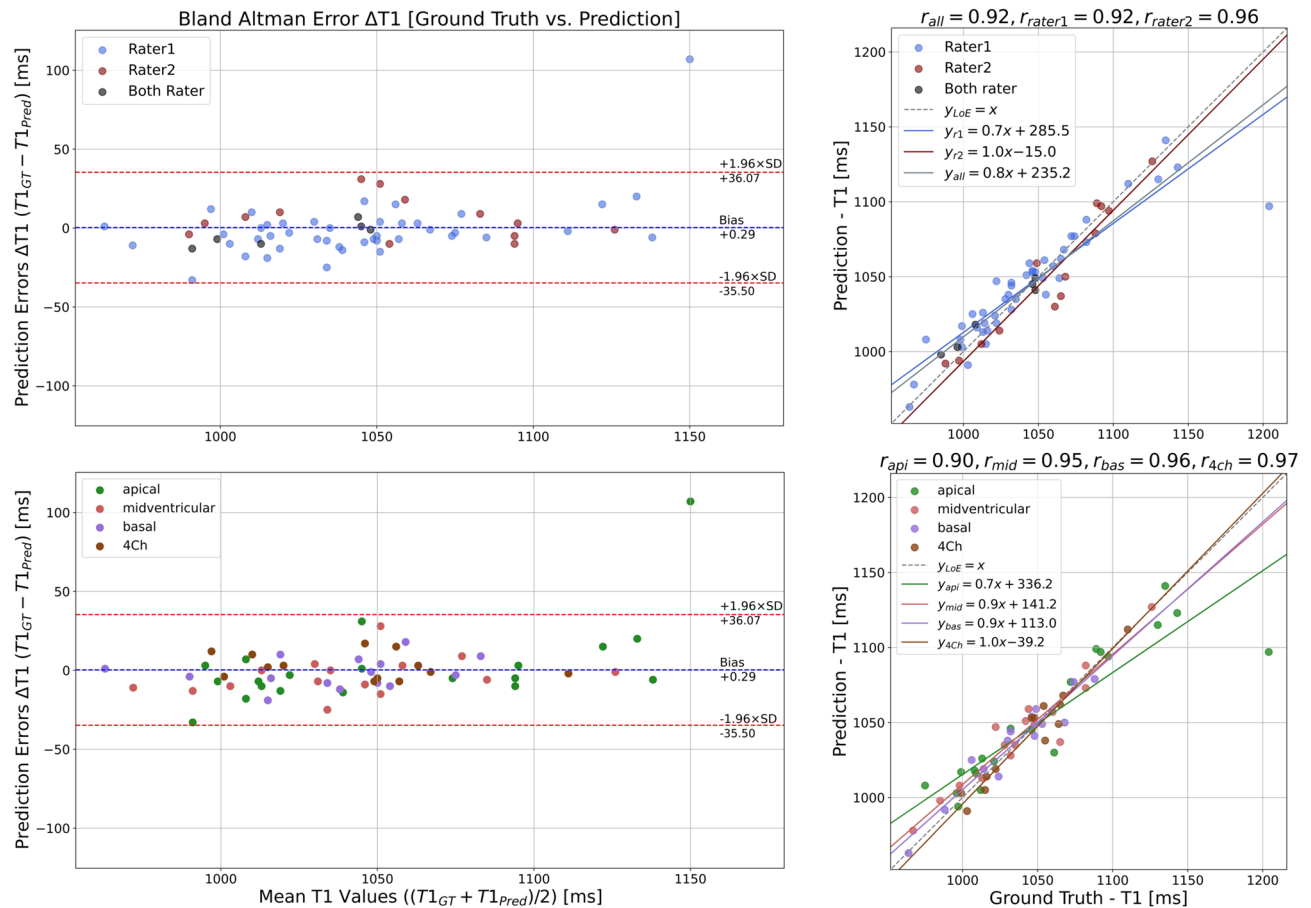


Figure 8. View- and Rater-clustered (top/bottom) model errors: The clustered Bland–Altman analyzes for $\Delta T1$ show all deviations for the test cohort divided for both raters (top) and divided by layer (apical, mid-ventricular, basal, and 4Ch) below. The linear dependencies between ground truth and predictions (right) show all linear fits in comparison to the line of equality ($y_{LoE} = x$), which represents an optimal reliability, for the test cohort, split for both raters (top) and split by layer (apical, midventricular, basal and 4Ch) below. This clarifies that the spread of measurement uncertainty becomes larger for smaller regions (apical and basal) and the model fits best for both raters.

Our study is in good agreement with the existing literature as we have shown that fully automated segmentation of the left ventricular myocardium is feasible using CNNs with a U-net architecture with an accuracy comparable to the agreement between two human experts. All previous studies on fully automated deep learning-based segmentation of T1 maps were limited to maps acquired in the short axis of the left ventricle, in which the left ventricular myocardium has the approximate shape of a closed circle. Our study is unique in including both short-axis and long-axis (4 chamber view) T1 maps. On long axis view, the left ventricular myocardium forms a U-shape. Thus, including both short- and long-axis views increases the complexity of the segmentation task. Interestingly, we found that segmentation performance was as good for long-axis views (mean DSC: 0.89 ± 0.03) as for short-axis images (0.88 ± 0.08). Accuracy in T1 quantification was slightly superior for long-axis views than for short-axis images (mean error $\Delta T1$: 6.77 ± 8.3 vs 11.6 ± 19.7 ms), largely driven by segmentation inaccuracies in the apical short-axis slices. In apical slices with very thin myocardium, even small deviations in segmentation can lead to relatively large differences in T1 times (Supplementary Material, Fig. S1).

Implications for practice

In contemporary clinical practice, T1 maps are often analyzed by manually drawing a region of interest in the interventricular septum. This introduces substantial inter-observer variability. Global analysis of myocardial T1 values across the myocardium has been shown to be non-inferior to placing a ROI in the septum, but requires contouring of the myocardium, which is time-consuming and again prone to inter-observer variability²⁰. There is considerable variability in the clinical use of T1 mapping in cardiac MRI. Most commonly, one or three short-axis slices are acquired although some centers acquire long-axis slices or both. Here, we propose a deep learning-based algorithm, which can segment the myocardium on T1 maps and provide mean quantitative results for myocardial T1 relaxation time without human interaction. The algorithm is intended for integration into the clinical workflow of cardiovascular imaging specialists. Integration of the algorithm either within the scanner software or into post-processing software with bidirectional access to the Picture Archiving and Communication

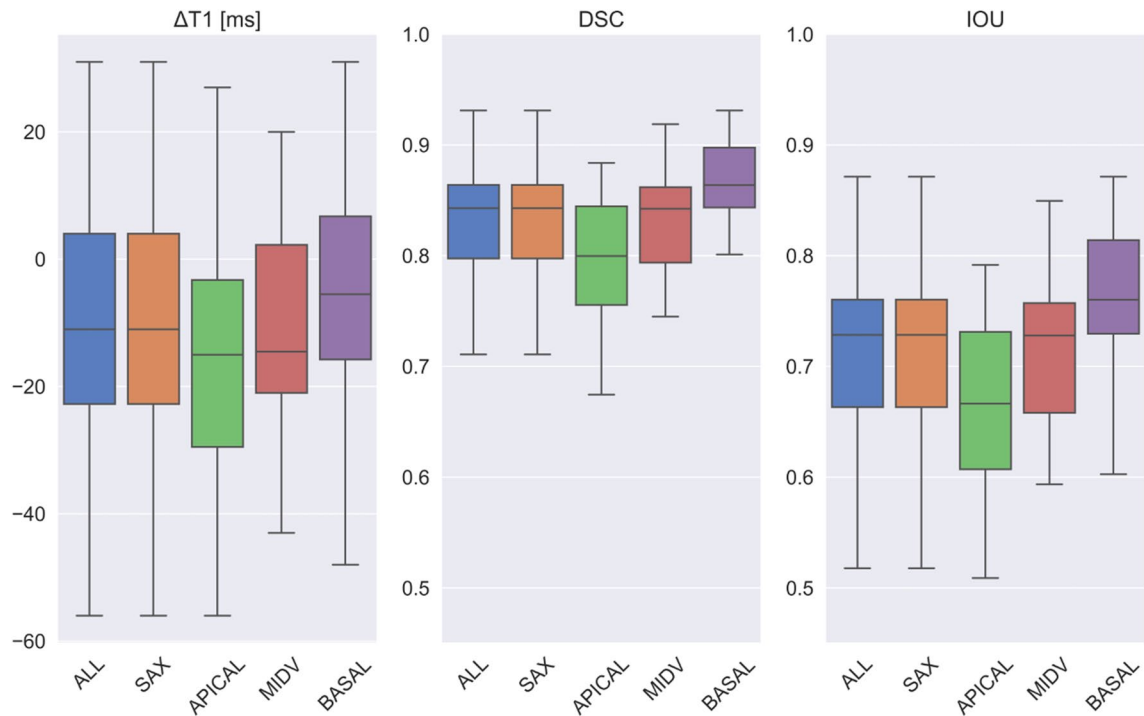


Figure 9. Interrater statistics: Boxplot results for 90 SAX slices for all measured metrics ($\Delta T1$, DSC and IOU) in the test cohort between both rater ground truth mask, shown for all short-axis views and separately for each view (apical, midventricular and basal).

		$\Delta T1$			DSC			IOU		
		Median	Range	p	Median	Range	p	Median	Range	p
SAX	Prediction	- 4.5	- 33—107	0.006	0.892	0.43—0.94	<0.001	0.806	0.27—0.89	<0.001
	Interrater	- 11.0	- 109—31		0.842	0.65—0.93		0.729	0.48—0.87	
API	Prediction	- 5.0	- 33—107	0.013	0.863	0.43—0.92	0.003	0.759	0.27—0.86	0.003
	Interrater	- 15.0	- 89—27		0.799	0.67—0.88		0.666	0.51—0.79	
MID	Prediction	- 6.0	- 25—28	0.11	0.893	0.85—0.93	<0.001	0.807	0.74—0.87	<0.001
	Interrater	- 14.5	- 109—20		0.842	0.65—0.92		0.728	0.48—0.85	
BAS	Prediction	- 3.0	- 19—18	0.56	0.916	0.86—0.94	<0.001	0.845	0.75—0.89	<0.001
	Interrater	- 5.5	- 48—31		0.864	0.75—0.93		0.760	0.60—0.87	

Table 3. Metrics of best model and interrater comparison. All metrics from best model were compared with interrater metrics for SAX slices. P-values are from Mann–Whitney U-Test. Significant values are in bold.

System (PACS) could ensure that results of T1 analysis can be directly sent to PACS and are ready before physicians even start reading the study.

Study limitations

Our study was limited to the global quantitative analysis of T1 maps, i.e. the average T1 values across segmented myocardium. This approach is suitable for the diagnosis of pathologies with diffuse changes of the myocardium, such as cardiac amyloidosis, Fabry’s disease and diffuse fibrosis. Other pathologies, most notably myocardial infarction and myocarditis, often manifest as focal changes in T1 values and may not lead to significant changes in mean T1. Therefore, even with automated analysis of mean T1 values, visual analysis of T1 maps for focal changes will still be required. All maps were derived from one 1.5 T MRI scanner of one vendor. Further work is required to confirm the performance of the algorithm for fully automated segmentation of T1 maps from other vendors or acquired at different field strength. Similarly, our study was limited to native T1 mapping. We chose to investigate the influence of activation function and metric on model performance. Other factors such as learning rate, the number of layers, the number of filters for each layer, and loss functions may also influence model performance and were not systematically assessed in our study. The performance was evaluated on a test cohort of previously unseen images from our center. External validation with data from other vendors and centers was beyond the scope of this study. Without further validation it is not clear if the algorithm will perform equally well for T2 maps, post-contrast T1 or extracellular volume (ECV) maps. Since this was a retrospective study,

we did not directly show that fully automated in-line analysis of the myocardial T1 maps is feasible in clinical routine. This will require the algorithm to be integrated into either the MRI console itself or into post-processing software connect to the hospital PACS.

Conclusion

The highest-performing algorithm developed in our study can segment the myocardium on T1 maps both in long-axis and short-axis orientation with very high accuracy. This allows for fully automated quantitative analysis of myocardial T1 maps.

Data availability

All data are available upon request from the corresponding author. The software source code is freely available on Github (<https://github.com/voxelacrobat/MRI-T1-Mapping-Segmentation>, Version 1.0).

Received: 17 April 2024; Accepted: 6 August 2024

Published online: 14 August 2024

References

- World Health Organisation. Cardiovascular diseases (CVDs) fact sheet. [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). Accessed 7 8 (2023).
- Messroghli, D. R. *et al.* Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). *J. Cardiovasc. Magn. Reson.* **19**, 75. <https://doi.org/10.1186/s12968-017-0389-8> (2017).
- Puntmann, V. O., Peker, E., Chandrasekhar, Y. & Nagel, E. T1 Mapping in characterizing myocardial disease: A comprehensive review. *Circ. Res.* **119**, 277–299. <https://doi.org/10.1161/CIRCRESAHA.116.307974> (2016).
- Bai, W. *et al.* Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J. Cardiovasc. Magn. Reson.* **20**, 65. <https://doi.org/10.1186/s12968-018-0471-x> (2018).
- Bernard, O. *et al.* Deep Learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?. *IEEE. Trans. Med. Imaging.* **37**, 2514–2525. <https://doi.org/10.1109/TMI.2018.2837502> (2018).
- Böttcher, B. *et al.* Fully automated quantification of left ventricular volumes and function in cardiac MRI: clinical evaluation of a deep learning-based algorithm. *Int. J. Cardiovasc. Imaging.* **36**, 2239–2247. <https://doi.org/10.1007/s10554-020-01935-0> (2020).
- Dangi, S., Linte, C. A. & Yaniv, Z. A distance map regularized CNN for cardiac cine MR image segmentation. *Med. Phys.* **46**, 5637–5651. <https://doi.org/10.1002/mp.13853> (2019).
- Ma, Z. *et al.* An iterative multi-path fully convolutional neural network for automatic cardiac segmentation in cine MR images. *Med. Phys.* **46**, 5652–5665. <https://doi.org/10.1002/mp.13859> (2019).
- Queirós, S. *et al.* Multi-centre validation of an automatic algorithm for fast 4D myocardial segmentation in cine CMR datasets. *Eur. Heart J. Cardiovasc. Imaging.* **17**, 1118–1127. <https://doi.org/10.1093/ehjci/jev247> (2016).
- Ruijsink, B. *et al.* Fully automated, quality-controlled cardiac analysis from CMR: Validation and large-scale application to characterize cardiac function. *JACC Cardiovasc. Imaging.* **13**, 684–695. <https://doi.org/10.1016/j.jcmg.2019.05.030> (2020).
- Suinesiaputra, A. *et al.* Fully-automated left ventricular mass and volume MRI analysis in the UK Biobank population cohort: evaluation of initial results. *Int. J. Cardiovasc. Imaging.* **34**, 281–291. <https://doi.org/10.1007/s10554-017-1225-9> (2018).
- Vigneault, D. M., Xie, W., Ho, C. Y., Bluemke, D. A. & Noble, J. A. Ω -Net (Omega-Net): Fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks. *Med. Image Anal.* **48**, 95–106. <https://doi.org/10.1016/j.media.2018.05.008> (2018).
- Arega, T. W. *et al.* Automatic uncertainty-based quality controlled T1 mapping and ECV analysis from native and post-contrast cardiac T1 mapping images using Bayesian vision transformer. *Med. Image Anal.* <https://doi.org/10.1016/j.media.2023.102773> (2023).
- Bhatt, N. *et al.* A deep learning segmentation pipeline for cardiac T1 mapping using MRI relaxation-based synthetic contrast augmentation. *Radiol. Artif. Intell.* <https://doi.org/10.1148/ryai.210294> (2022).
- Fahmy, A. S., El-Rewaidy, H., Nezafat, M., Nakamori, S. & Nezafat, R. Automated analysis of cardiovascular magnetic resonance myocardial native T1 mapping images using fully convolutional neural networks. *J. Cardiovasc. Magn. Reson.* **21**, 7. <https://doi.org/10.1186/s12968-018-0516-1> (2019).
- Farrag, N. A., Lochbihler, A., White, J. A. & Ukwatta, E. Evaluation of fully automated myocardial segmentation techniques in native and contrast-enhanced T1-mapping cardiovascular magnetic resonance images using fully convolutional neural networks. *Med. Phys.* **48**, 215–226. <https://doi.org/10.1002/mp.14574> (2021).
- Hann, E. *et al.* Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. *Med. Image Anal.* <https://doi.org/10.1016/j.media.2021.102029> (2021).
- Howard, J. P. *et al.* Automated inline myocardial segmentation of joint t1 and t2 mapping using deep learning. *Radiol. Artif. Intell.* <https://doi.org/10.1148/ryai.220050> (2023).
- Puyol-Antón, E. *et al.* Automated quantification of myocardial tissue characteristics from native T1 mapping using neural networks with uncertainty-based quality-control. *J. Cardiovasc. Magn. Reson.* **22**, 60. <https://doi.org/10.1186/s12968-020-00650-y> (2020).
- Böttcher, B. *et al.* Global and regional test-retest reproducibility of native T1 and T2 mapping in cardiac magnetic resonance imaging. *J. Magn. Reson. Imaging* **54**, 1763–1772. <https://doi.org/10.1002/jmri.27755> (2021).
- Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (eds Navab, N. *et al.*) (Springer International Publishing, 2015).
- Shen, T., Huang, F. & Zhang, X. CT medical image segmentation algorithm based on deep learning technology. *Math. Biosci. Eng.* **20**, 10954–10976. <https://doi.org/10.3934/mbe.2023485> (2023).
- Furtado, P. Testing segmentation popular loss and variations in three multiclass medical imaging problems. *J. Imaging* <https://doi.org/10.3390/jimaging7020016> (2021).
- Müller, D., Soto-Rey, I. & Kramer, F. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Res. Notes.* **15**, 210. <https://doi.org/10.1186/s13104-022-06096-y> (2022).
- Handels, H. *Medizinische Bildverarbeitung: Bildanalyse, Mustererkennung und Visualisierung für die computergestützte ärztliche Diagnostik und Therapie* | Studium 2nd edn. (Vieweg+Teubner, 2009).
- Kingma D.P. & Ba J. Adam: A Method for Stochastic Optimization. [Conference paper at ICLR 2015, accessed on 22 December 2014]. Available online: <https://arxiv.org/abs/1412.6980>.

Author contributions

M.M. and F.G.M. designed the study. S.I., B.B. and F.G.M. performed data curation and ground truth annotation. M.M. implemented the software frameworks for image postprocessing, model training and statistical analysis. All authors made substantial contributions to the analysis and interpretation of data. M.M., S.I. and F.G.M. wrote the original draft of the main manuscript text and M.M. and A.C.K. prepared Figs. 1–9. All authors substantively revised the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. The study was in part funded by the Federal Ministry of Education and Research (BMBF) through the Network University Medicine „NUM 2.0“ (grant number 01KX2121). The funders had no role in study design, data analysis, data interpretation or in drafting the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-69529-7>.

Correspondence and requests for materials should be addressed to F.G.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

CARDIAC

Open Access



Repeatability of radiomic features in myocardial T1 and T2 mapping

Mathias Manzke¹ , Fabian C. Laqua² , Benjamin Böttcher¹, Ann-Christin Klemenz¹ , Marc-André Weber¹ , Bettina Baeßler²  and Felix G. Meinel^{1*} 

Abstract

Purpose To investigate the test–retest repeatability of radiomic features in myocardial native T1 and T2 mapping.

Methods In this prospective study, 50 healthy volunteers (29 women and 21 men, mean age 39.4 ± 13.7 years) underwent two identical cardiac magnetic resonance imaging (MRI) examinations at 1.5 T. The protocol included native T1 and T2 mapping in both short-axis and long-axis orientation. For T1 mapping, we investigated standard (1.9×1.9 mm) and high (1.4×1.4 mm) spatial resolution. After manual segmentation of the left ventricular myocardium, 100 radiomic features from seven feature classes were extracted and analyzed. Test–retest repeatability of radiomic features was assessed using the intraclass correlation coefficient (ICC) and classified as poor ($\text{ICC} < 0.50$), moderate ($0.50\text{--}0.75$), good ($0.75\text{--}0.90$), and excellent (> 0.90).

Results For T1 maps acquired in short-axis orientation at standard resolution, repeatability was excellent for 6 features, good for 29 features, moderate for 19 features, and poor for 46 features. We identified 15 features from 6 classes which showed good to excellent reproducibility for T1 mapping in all resolutions and all orientations. For short-axis T2 maps, repeatability was excellent for 6 features, good for 25 features, moderate for 23 features, and poor for 46 features. 12 features from 5 classes were found to have good to excellent repeatability in T2 mapping independent of slice orientation.

Conclusion We have identified a subset of features with good to excellent repeatability independent of slice orientation and spatial resolution. We recommend using these features for further radiomics research in myocardial T1 and T2 mapping.

Key Points

Question *The study addresses the need for reliable radiomic features for quantitative analysis of the myocardium to ensure diagnostic consistency in cardiac MRI.*

Findings *We have identified a subset of radiomic features demonstrating good to excellent repeatability in native T1 and T2 mapping independent of slice orientation and resolution.*

Clinical relevance *Radiomic features have been proposed as diagnostic and prognostic biomarkers in various heart diseases. By identifying a subset of particularly reproducible radiomic features our study serves to inform the selection of radiomic features in future research and clinical applications.*

Keywords Radiomics, Cardiac magnetic resonance imaging, T1 mapping, T2 mapping, Reproducibility

*Correspondence:

Felix G. Meinel

felix.meinel@med.uni-rostock.de

¹Institute of Diagnostic and Interventional Radiology, Pediatric Radiology and Neuroradiology, University Medical Centre Rostock, Rostock, Germany

²Department of Diagnostic and Interventional Radiology, University Hospital Würzburg, Würzburg, Germany

Graphical Abstract

Repeatability of radiomic features in myocardial T1 and T2 mapping

How reproducible are radiomic features in myocardial native T1 and T2 mapping?

Data acquisition

- Two identical MRI examinations

Feature extraction

- 100 features from 7 classes

Repeatability analysis

- Feature robustness classification

50 healthy
volunteers

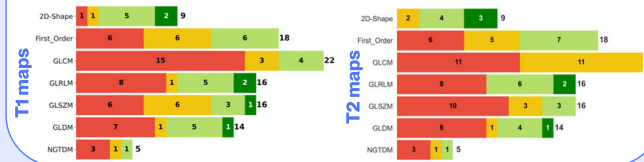


Heart



Single

Robustness of radiomic features on maps



We identified a subset of 15 radiomic features for T1 mapping and 12 features for T2 mapping with good-to-excellent repeatability independent of slice orientation and spatial resolution.

Eur Radiol (2024) Manzke M, Laqua FC, Böttcher B et al;
DOI: 10.1007/s00330-024-11337-8

EUROPEAN SOCIETY OF RADIOLOGY
European Radiology

Introduction

Cardiac magnetic resonance imaging (MRI) mapping sequences allow for quantitative tissue characterization of the myocardium [1]. Myocardial relaxation times reflects biophysical properties of the myocardial muscle cells and the surrounding extracellular space. Quantifying T1 or T2 relaxation times is therefore helpful to diagnose pathologies affecting the myocardium. Altered myocardial relaxation times can result from fibrosis, inflammation, edema or storage diseases [2].

In addition to evaluating mean global or regional T1 and T2 relaxation times, radiomics features have been used to extract detailed quantitative information from myocardial T1 and T2 maps. Radiomics features are essentially numerical data extracted from segmented regions of interest (ROI) within radiological images that capture a variety of aspects regarding image texture, shape, intensity, and patterns and may serve as imaging biomarkers of myocardial disease [3, 4]. Radiomics analysis of myocardial native T1 mapping has been used as diagnostic and prognostic biomarkers in a wide range of cardiac conditions. T1 mapping radiomics features have been investigated as diagnostic biomarkers to diagnose myocarditis [5], characterize the extracellular matrix in dilated cardiomyopathy [6], classify hypertrophic

cardiomyopathy genotypes [7] and phenotypes [8, 9], distinguish hypertrophic cardiomyopathy from hypertensive heart disease [10, 11] and differentiate cardiac tumors from thrombi [12]. As prognostic biomarkers, T1 radiomics features have been suggested to predict clinical course in dilated cardiomyopathy [13, 14], heart failure in systemic lupus erythematosus [15], functional recovery after myocardial infarction [16], and treatment response in pulmonary arterial hypertension [17]. The application of radiomics analysis of T2 maps has focused on the diagnosis of acute and chronic myocarditis [5].

An essential prerequisite for diagnostic or prognostic significance is the repeatability of the measurement results. Although the reproducibility of radiomics features has received significant attention [18–20] and has been studied for cardiac cine sequences [21], there is very limited data on the repeatability of radiomics features in myocardial T1 and T2 mapping. Only one previous study specifically investigated the reproducibility of myocardial T1 and T2 radiomics and found that only a small subset of myocardial radiomic features were reproducible [22]. This previous analysis focused on the overall performance of feature classes, not individual features. The influence of slice orientation (short-axis vs. long-axis view) and spatial resolution was also not investigated.

Therefore, the aim of this study was to systematically analyze the repeatability of individual radiomic features in myocardial native T1 and T2 mapping.

Material and methods

Patient selection and study design

We performed a post-hoc analysis of data from a prospective study of 50 healthy participants (29 female, 21 male, mean age 39.4 ± 13.7 years), without cardiovascular disease. Mean BMI was 23.9 ± 4.0 kg/m², and the mean heart rate was 72 ± 10 per minute, who underwent two identical cardiac MRI examinations. The distribution of demographic parameters is shown in Suppl. Fig. 1. All MRI examinations were performed on a 1.5-T MRI scanner (Avanto Fit Siemens Healthineers). Details of imaging protocols and the process of creating the T1 maps have been published previously [23].

Ethical approval and informed consent

This study was approved by the institutional review board (Ethics Committee, University Medical Center Rostock), and written informed consent was obtained from all volunteers prior to enrollment. The study conformed to the ethical guidelines of the Helsinki Declaration (revised version from 2013).

Image acquisition

Each volunteer underwent two MRI investigations on the same day with a break of at least 20 min between the examinations. All MRI examinations were performed on a 1.5-T MRI scanner (Avanto Fit, Siemens Healthineers). For T1 and T2 mapping, three short-axis (SAX) slices were acquired in a basal, midventricular, and apical position of the left ventricle. Additionally, long-axis T1 and T2 mapping was performed in the four-chamber view (4Ch). For T1 mapping, a standard MODified Look-Locker Inversion recovery (MOLLI) 5(3)3 sequence was used and acquired in two different spatial resolutions: standard resolution with 1.9×1.9 mm pixel size and high resolution with 1.4×1.4 mm pixel size. Further details of the imaging protocol have been published previously [23, 24].

Image processing

The left ventricular myocardium was manually segmented on all mapping data by one radiologist using open-source software (ITK-SNAP, version 3.8.0). Segmentation was performed on the original source images, not scanner-generated maps. We deliberately chose to exclude a thin subendocardial and subepicardial layer of the myocardium to avoid any risk of inadvertently including the left ventricular blood pool or epicardial fat in the segmentation.

Radiomics feature extraction

Radiomic features of the left ventricular myocardium were extracted using open-source Python packages (pyradiomics, version 3.0.1, SimpleITK 2.0.2, numpy 1.21.5) for all segmentations of the T1 and T2 mapping data. We analyzed all 100 radiomic features as provided by the software from the following seven feature classes: First Order Features, Shape Features (2D), Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Size Zone Matrix (GLSZM) Features, Gray Level Run Length Matrix (GLRLM) features, Neighboring Gray Tone Difference Matrix (NGTDM) features and Gray Level Dependence Matrix (GLDM) features (Table 1 and Supplementary Material).

Statistical analysis

Test–retest repeatability was assessed by using the intra-class correlation coefficient (ICC) as the primary parameter according to the definition of Shrout and Fleiss [25]. This criteria represents the feature robustness and, hence, the reproducibility [26]. For all radiomic features, ICCs were calculated with 95% confidence intervals from two-way random effects for the models of both measurements. Numerical ICC values were interpreted as follows: < 0.50: poor, 0.50–0.75: moderate, 0.75–0.90: good, > 0.90: excellent [27].

To further assess the reproducibility, we used the concordance correlation coefficient (CCC) [28, 29]. The Pearson correlation coefficient r was calculated to measure the linear relationship between both measurements.

Detailed results for all features were presented as tables for standard resolution and short-axis orientation as these can be considered the clinical standard. We further analyzed results also for long-axis T1 and T2 mapping (4Ch) and for high-resolution T1 mapping. For each of these settings and each of the seven feature classes, the number of features with excellent, good, moderate, and poor reproducibility was visualized as stacked bar charts.

We further aimed to identify a subset of radiomic features with good to excellent repeatability regardless of image orientation and spatial resolution. For T1 mapping, we identified the intersection of features that showed good to excellent repeatability ($ICC > 0.75$) in all four sequence variations: short-axis and long-axis views as well as standard and high spatial resolution. For T2 mapping, we determined the intersection of features with good to excellent repeatability ($ICC > 0.75$) in short-axis and long-axis views. The resulting selections of robust radiomic features were visualized and color-coded according to their ICC values over all inhabited feature sets. For these features, we further analyzed for differences in their numerical values between female and male participants as well as younger and older participants. For this purpose,

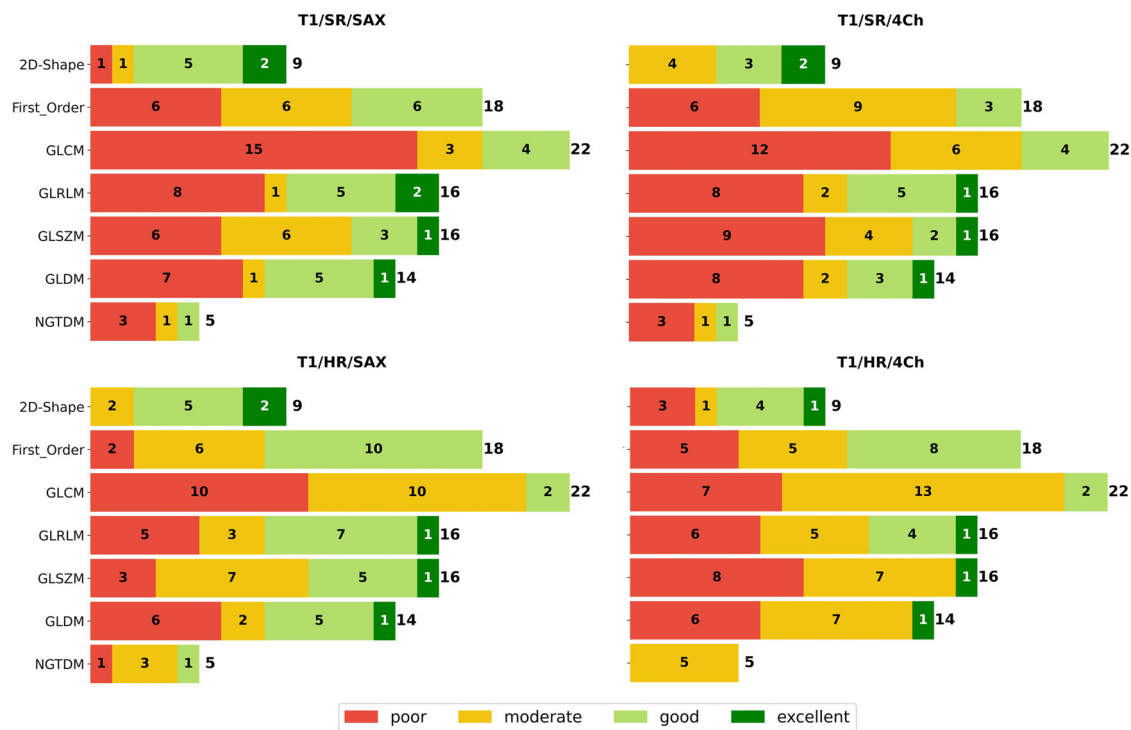


Fig. 1 Robustness of radiomic feature classes for T1 mapping: for all combinations of short-axis (SAX) or four-chamber view (4Ch) and high resolution (HR) or standard resolution (SR) T1 mapping, the figure shows the distribution of robustness for all seven radiomic feature classes considered: Shape Features (2D), First Order Features, Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Size Zone Matrix (GLSZM) Features, Gray Level Run Length Matrix (GLRLM) features, Neighboring Gray Tone Difference Matrix (NGTDM) features and Gray Level Dependence Matrix (GLDM). Green = excellent (ICC > 0.9), light green = good (ICC 0.75–0.90), yellow = moderate (ICC 0.5–0.75), red = poor (ICC < 0.5) robustness

the measured values were clustered according to gender and age group, and then the mean value was determined for each characteristic over both runs. The nonparametric Mann–Whitney *U*-test was used to test for gender- or age-specific differences.

Statistical analyses were performed using the Python (3.12.3) libraries Scipy 1.10.1 for Pearson Correlation Coefficient and Pingouin 0.5.4 for ICC as well as Seaborn 0.12.2. for heatmap plots (Figs. 2, 3, 5, and 6) and Matplotlib 3.9.2 for bar charts (Figs. 1 and 4). Within the Pingouin library, ICC type 2 (single random raters) was used.

Results

Repeatability of radiomic features in T1 mapping (standard setting)

T1 mapping acquired with standard (1.9 × 1.9 mm) in-plane spatial resolution in short-axis orientation was considered the clinical standard. For this setting, the repeatability of all radiomic features in each feature class is shown in Supplementary Tables 2–8 (Supplementary Material), sorted in descending order according to their robustness. We observed substantial differences in the

repeatability of radiomic features. Across all feature classes, repeatability was excellent for 6 features, good for 29 features, moderate for 19 features, and poor for 46 features. Within each feature class, the repeatability of individual features was highly variable, ranging from poor to good or excellent for all feature classes.

While we used the ICC as the primary parameter of repeatability and robustness, the results for CCC and Pearson correlation coefficient *r* show very consistent results for the robustness of individual radiomic features.

Influence of image orientation and spatial resolution on the repeatability of radiomic features in T1 mapping

The aggregated results of the entire T1 evaluation are summarized as stacked bar charts in Fig. 1. These charts show the percentage of features with excellent, good, moderate and poor robustness for each feature class and each of the investigated variations of T1 mapping techniques (short-axis and 4Ch in standard and high spatial resolution).

Independent of view or resolution, 2D Shape features were most reproducible, with at least 55.5% of features in this class showing excellent or good reproducibility. The

least robust feature classes were NGTDM and GLCM, with 0–20% and 9.1–18.2%, respectively, of features in this class showing excellent or good reproducibility.

The performances of all features are shown in Fig. 2 for all views and all resolutions (Fig. 3) gives a more detailed representation of those features that showed good to excellent reproducibility in at least one of the four sequence variations: short-axis and long-axis views, as well as standard and high spatial resolution. Overall, the reproducibility of radiomic features was superior in the short-axis view (SAX) than in the 4Ch: 32 features from 7 classes showed good to excellent reproducibility in short-axis T1 mapping, compared to only 16 features from 6 classes in the 4Ch (Fig. 3). The number of features with good to excellent reproducibility was 35 for SAX and standard resolution, 26 for 4Ch and standard resolution, 40 for SAX and high resolution and 22 for 4Ch and standard resolution, Fig. 3).

Selection of the most robust radiomic features in T1 mapping

We further determined the intersection of features that showed good to excellent repeatability ($ICC > 0.75$) in all four sequence variations: short-axis and long-axis views, as well as standard and high spatial resolution. This subset of particularly robust features is shown in bold in Fig. 3. In this analysis, we identified 15 features from 6 classes that perform excellent or good in both spatial resolutions and both short-axis and long-axis views.

For these features, we further analyzed for differences in their numerical values between female and male participants as well as younger and older participants, shown in Supplementary Tables 9 and 10 (Supplementary Material). All selected features except sphericity showed significant differences between men and women. For age, no significant differences were seen except for 2 shape features, which may have been confounded by a larger proportion of women in the older age group.

Repeatability of radiomic features in T2 mapping (standard setting)

T2 mapping acquired in short-axis orientation was considered the clinical standard. For this setting, the repeatability of all radiomic features in each feature class is shown in Supplementary Tables 11–17 (Supplementary Material), sorted in descending order according to their robustness. Substantial differences in the repeatability of radiomic features were seen also for T2 mapping. Across all feature classes, repeatability was excellent for 6 features, good for 25 features, moderate for 23 features, and poor for 46 features. The repeatability of individual features ranged from poor to good or excellent within most feature classes.

Influence of image orientation on the repeatability of radiomic features in T2 mapping

2D Shape features were the most reproducible, with 77.7% (SAX) and 33.3% (4Ch) of features in this class showing excellent or good reproducibility. All other feature classes showed very limited repeatability, with $\leq 50\%$ of features (SAX) and $\leq 25\%$ of features (4Ch) showing excellent or good repeatability (Fig. 4).

The performances of all features are shown in Fig. 5 for both views. Notably, none of the features in the GLCM feature class showed good or excellent reproducibility for T2 mapping. Figure 6. gives a more detailed representation of those features that showed good to excellent reproducibility in at least one of the views. Overall, the reproducibility of radiomic features was superior in the SAX compared to the 4Ch: 31 features from 6 classes showed good to excellent reproducibility in short-axis T2 mapping, compared to 14 features from 5 classes in the 4Ch (Fig. 6).

Selection of the most robust radiomic features in T2 mapping

We identified 12 features from 5 classes that showed excellent or good repeatability ($ICC > 0.75$) in both short-axis and long-axis T2 mapping. This selection of features is shown in bold in Fig. 6.

Overlap between the most robust radiomic features in T1 and T2 mapping

Seven of the 12 features identified as most reproducible for T2 mapping were also among the 15 most reproducible features for T1 mapping. These include 3 shape features (MinorAxisLength, Maximum2DDiameterSclie, Maximum2DDiameterColumn) and 3 GLRLM features (RunPercentage, RunLengthNonUniformityNormalized, GrayLevelNonUniformity) and 1 GLDM feature (GrayLevelNonUniformity).

Discussion

Although there is abundant literature on radiomics analysis in cardiac MRI, the majority of these studies have focused on the prognostic and diagnostic value of cardiovascular disease. Radiomics analysis of native T1 and T2 mapping has been applied to myocarditis [5], dilated cardiomyopathy [6, 13, 14], hypertrophic cardiomyopathy [7, 9–11], and myocardial infarction [16], among other pathologies. Only a few of the published papers investigated the reliability of radiomics features in cardiac MRI [19, 30].

In a previous study [30], the repeatability of radiomic features in quantitative T1 and T2 mapping was investigated in 6 healthy participants and 18 patients for the SAX and one resolution. Features with an ICC greater

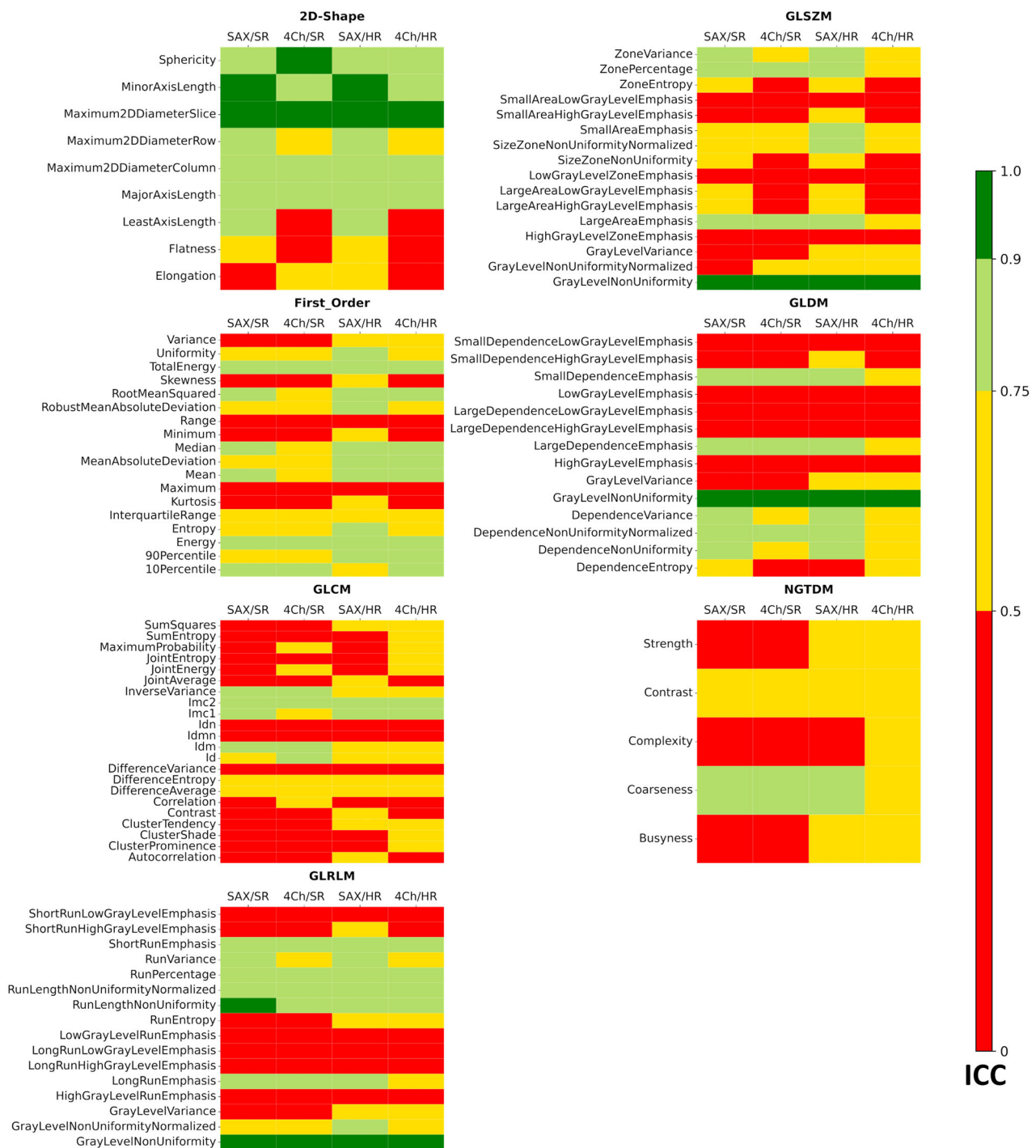


Fig. 2 Robustness of all radiomic features for T1 mapping: for all combinations of short-axis (SAX) or four-chamber view (4Ch) and high resolution (HR) or standard resolution (SR) T1 mapping, all features are color-coded depending on the ICC-value divided for Shape Features (2D), First Order Features, Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Size Zone Matrix (GLSZM) Features, Gray Level Run Length Matrix (GLRLM) features, Neighboring Gray Tone Difference Matrix (NGTDM) features and Gray Level Dependence Matrix (GLDM). Green = excellent (ICC > 0.9), light green = good (ICC 0.75–0.90), yellow = moderate (ICC 0.5–0.75), red = poor (ICC < 0.5) robustness

		SAX/SR	4Ch/SR	SAX/HR	4Ch/HR	
2D-Shape -	Sphericity	0.80	0.93	0.78	0.90	
2D-Shape -	MinorAxisLength	0.94	0.90	0.96	0.89	
2D-Shape -	Maximum2DDiameterSlice	0.96	0.96	0.95	0.96	
2D-Shape -	Maximum2DDiameterRow	0.78		0.81		
2D-Shape -	Maximum2DDiameterColumn	0.87	0.87	0.89	0.83	
2D-Shape -	MajorAxisLength	0.86	0.90	0.88	0.90	
2D-Shape -	LeastAxisLength	0.85		0.87		
NGTDM -	Coarseness	0.88	0.82	0.84		
GLSZM -	ZoneVariance	0.80		0.80		
GLSZM -	ZonePercentage	0.87	0.81	0.89		
GLSZM -	SmallAreaEmphasis			0.78		
GLSZM -	SizeZoneNonUniformityNormalized			0.80		
GLSZM -	LargeAreaEmphasis	0.82	0.78	0.84		
GLSZM -	GrayLevelNonUniformity	0.95	0.92	0.91	0.92	
GLRLM -	ShortRunEmphasis	0.88	0.79	0.87	0.77	
GLRLM -	RunVariance	0.87		0.86		
GLRLM -	RunPercentage	0.88	0.81	0.87	0.75	
GLRLM -	RunLengthNonUniformityNormalized	0.88	0.79	0.86	0.78	
GLRLM -	RunLengthNonUniformity	0.92	0.82	0.89	0.75	
GLRLM -	LongRunEmphasis	0.88	0.80	0.87		
GLRLM -	GrayLevelNonUniformityNormalized			0.79		
GLRLM -	GrayLevelNonUniformity	0.96	0.95	0.96	0.92	
GLDM -	SmallDependenceEmphasis	0.85	0.80	0.89		
GLDM -	LargeDependenceEmphasis	0.88	0.80	0.85		
GLDM -	GrayLevelNonUniformity	0.96	0.95	0.96	0.91	
GLDM -	DependenceVariance	0.82		0.76		
GLDM -	DependenceNonUniformityNormalized	0.86	0.80	0.85		
GLDM -	DependenceNonUniformity	0.77		0.77		
GLCM -	InverseVariance	0.76	0.76			
GLCM -	Imc2	0.87	0.76	0.84	0.78	
GLCM -	Imc1	0.85		0.82	0.78	
GLCM -	Idm	0.76	0.78			
GLCM -	Id		0.78			
First_Order -	Uniformity			0.80		
First_Order -	TotalEnergy	0.90	0.82	0.87	0.77	
First_Order -	RootMeanSquared	0.81		0.81	0.86	
First_Order -	RobustMeanAbsoluteDeviation			0.77		
First_Order -	Median	0.84		0.85	0.84	
First_Order -	MeanAbsoluteDeviation			0.76	0.75	
First_Order -	Mean	0.82		0.80	0.85	
First_Order -	Entropy			0.78		
First_Order -	Energy	0.89	0.82	0.87	0.77	
First_Order -	90Percentile			0.82	0.79	
First_Order -	10Percentile	0.90	0.76		0.82	

Fig. 3 Selected features for T1 mapping: only features with good to excellent reproducibility in at least one of the four sequence variations (short-axis and long-axis views as well as standard and high spatial resolution) are shown. ICC values are color-coded for both views (short-axis/four chambers) and both resolutions. ICC values ≤ 0.75 are not shown. The subset of 15 features that showed good to excellent repeatability (ICC > 0.75) in all four sequence variations is shown in bold

than 0.75 and limits of agreement $\pm 5\%$ were considered reproducible. Except for 90Percentile, all of the radiomic features found to be most reproducible for T1 mapping in their study also showed good to excellent reproducibility in short-axis standard-resolution T1 mapping in our cohort (Mean, Median, 10Percentile, RootMeanSquared, Imc2, RunLengthNonUniformityNormalized, RunPercentage and ShortRunEmphasis). For short-axis T2 mapping, we can also confirm that the four features identified as reproducible in this previous study (MaximumDiameter, RunLengthNonUniformityNormalized, RunPercentage ShortRunEmphasis), showed good to excellent reproducibility in our cohort in T2 mapping in

SAX. In contrast to this previous study, we were able to show that feature robustness varies by resolution (standard resolution or high resolution) and view, resulting in significantly fewer reproducible features for the 4Ch.

Another previous study assessed the repeatability of radiomic features in quantitative T1 and T2 mapping in phantom experiments, 10 healthy participants and 51 patients [31]. In this previous study, features with ICCs greater than or equal to 0.8 were considered reproducible. Consistent with our findings, they also observed that only a minority of radiomic features were reproducible. Several differences between this previous study and ours are noteworthy. The previous analysis [31] focused on the



Fig. 4 Robustness of radiomic feature classes for T2 mapping: for short-axis (SAX) and four-chamber view (4Ch) T2 mapping, the figure shows the distribution of robustness for all seven radiomics feature classes considered: Shape Features (2D), First Order Features, Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Size Zone Matrix (GLSZM) Features, Gray Level Run Length Matrix (GLRLM) features, Neighboring Gray Tone Difference Matrix (NGTDM) features and Gray Level Dependence Matrix (GLDM). Green = excellent (ICC > 0.9), light green = good (ICC 0.75–0.90), yellow = moderate (ICC 0.5–0.75), red = poor (ICC < 0.5) robustness

overall performance of feature classes, not individual features, and included an analysis of the influence of image filters. In contrast, our study investigated the repeatability of individual features and the influence of slice orientation (short-axis vs. long-axis view) and spatial resolution on the repeatability of radiomic features. We found that the number of reproducible features was higher for the short-axis than for the long-axis (4Ch) acquisition. This may be related to the orientation of the myocardial muscle fibers relative to the imaging plane or to the geometry of the circular region of interest (ROI) in short-axis mapping compared to the U-shaped ROI in the 4Ch. For spatial resolution, there was no clear trend favoring standard-resolution or higher-resolution maps. We conclude that the standard resolution used in clinical cardiac MR imaging is probably sufficient for radiomic analysis.

While we focused on test–retest repeatability, others have investigated the effect of image processing on radiomic features in T1 and T2 mapping. In patients with hypertrophic cardiomyopathy, Marfisi and colleagues found that image resampling, discretization, and filtering had a significant influence on radiomic features [32]. In their study, radiomic features from T2 maps were less sensitive to changes in image processing than those from T1 maps. A recent study based on a scan-rescan cohort of brain MRI data demonstrated that using a novel physics-informed discretization method substantially improved the number of reproducible radiomic features compared to a conventional fixed bin number discretization [33]. Applying physics-informed discretization methods to image processing in myocardial mapping may thus be a

promising avenue to improve the reproducibility of radiomic features.

In our study, we observed that only a limited number of radiomic features demonstrated sufficient reproducibility even though this test–retest study compared two identical MRI examinations performed on the same scanner with identical settings. In clinical practice and cardiovascular imaging studies, cardiac MR examinations are often performed on different scanners with differences in imaging parameters. It is plausible that reproducibility will be even more limited in such a less controlled setting. One previous study found that most radiomic features are indeed sensitive to changes in sequence parameters [22]. For myocardial T1 and T2 mapping, they found that radiomic features were more sensitive to changes in flip angle and slice thickness and less sensitive to changes in in-plane spatial resolution [22]. To address such differences in image acquisition parameters, normalization methods have been developed and shown to improve the robustness of radiomics analysis [34].

In our study, the feature classes 2D shape and GLRLM proved to be more reproducible than other feature classes in our study. Shape features are independent of the gray level intensity distribution in the ROI and are therefore only calculated for the underived image and mask. The segmentations of a randomly chosen subset of participants are shown as an example in Suppl. Fig. 2. Due to the high segmentation agreement by the same radiologist, the agreement between geometric properties of the regions of interest was very high, which is why the 2D shape features predominantly showed good results. The grayscale run length matrix (GLRLM) quantifies the grayscale runs,

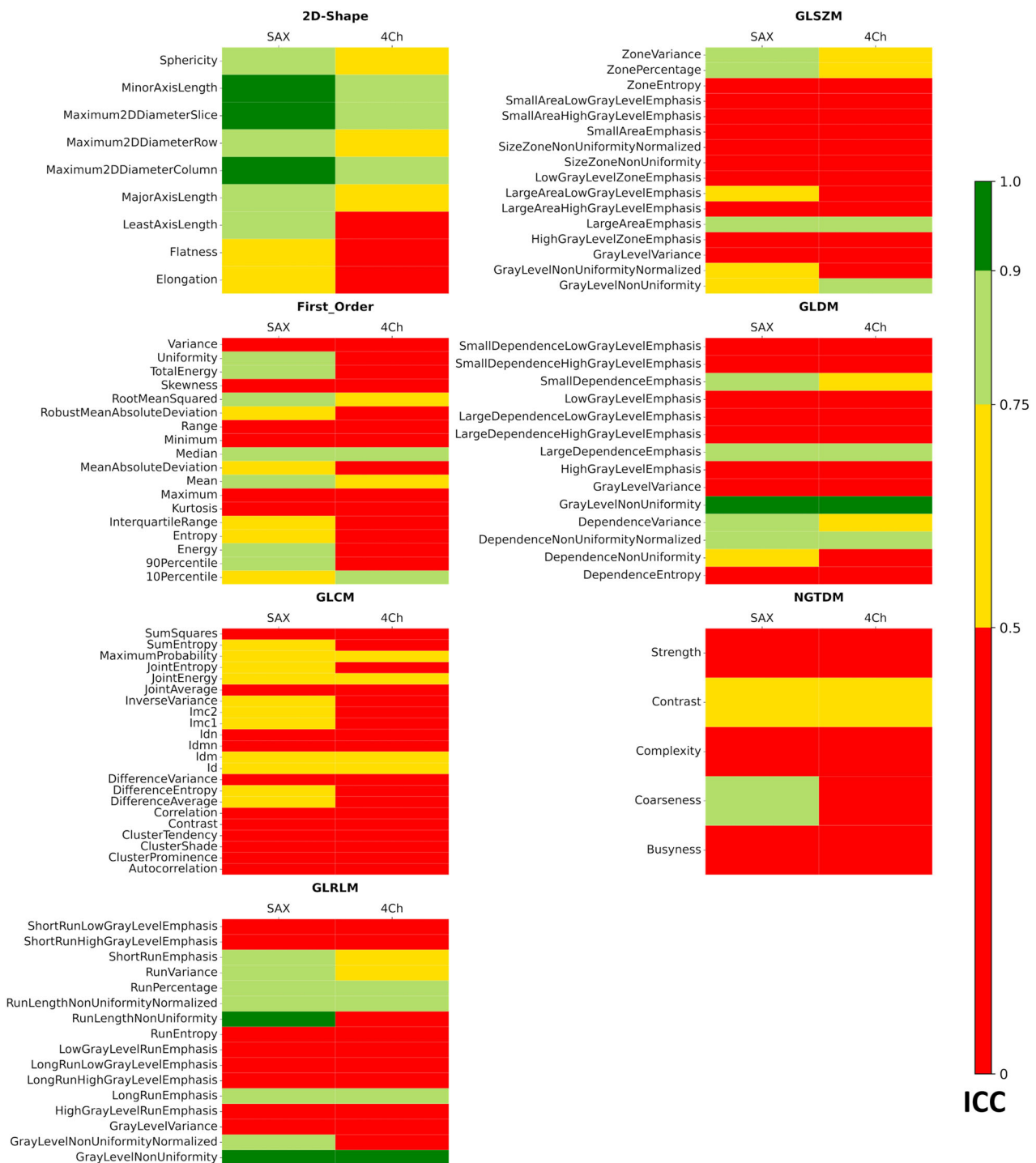


Fig. 5 Robustness of all radiomic features for T2 mapping: for short-axis (SAX) and four-chamber view (4Ch) T2 mapping, all features are color-coded depending on the ICC-value divided for Shape Features (2D), First Order Features, Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Size Zone Matrix (GLSZM) Features, Gray Level Run Length Matrix (GLRLM) features, Neighboring Gray Tone Difference Matrix (NGTDM) features and Gray Level Dependence Matrix (GLDM). Green = excellent (ICC > 0.9), light green = good (ICC 0.75–0.90), yellow = moderate (ICC 0.5–0.75), red = poor (ICC < 0.5) robustness

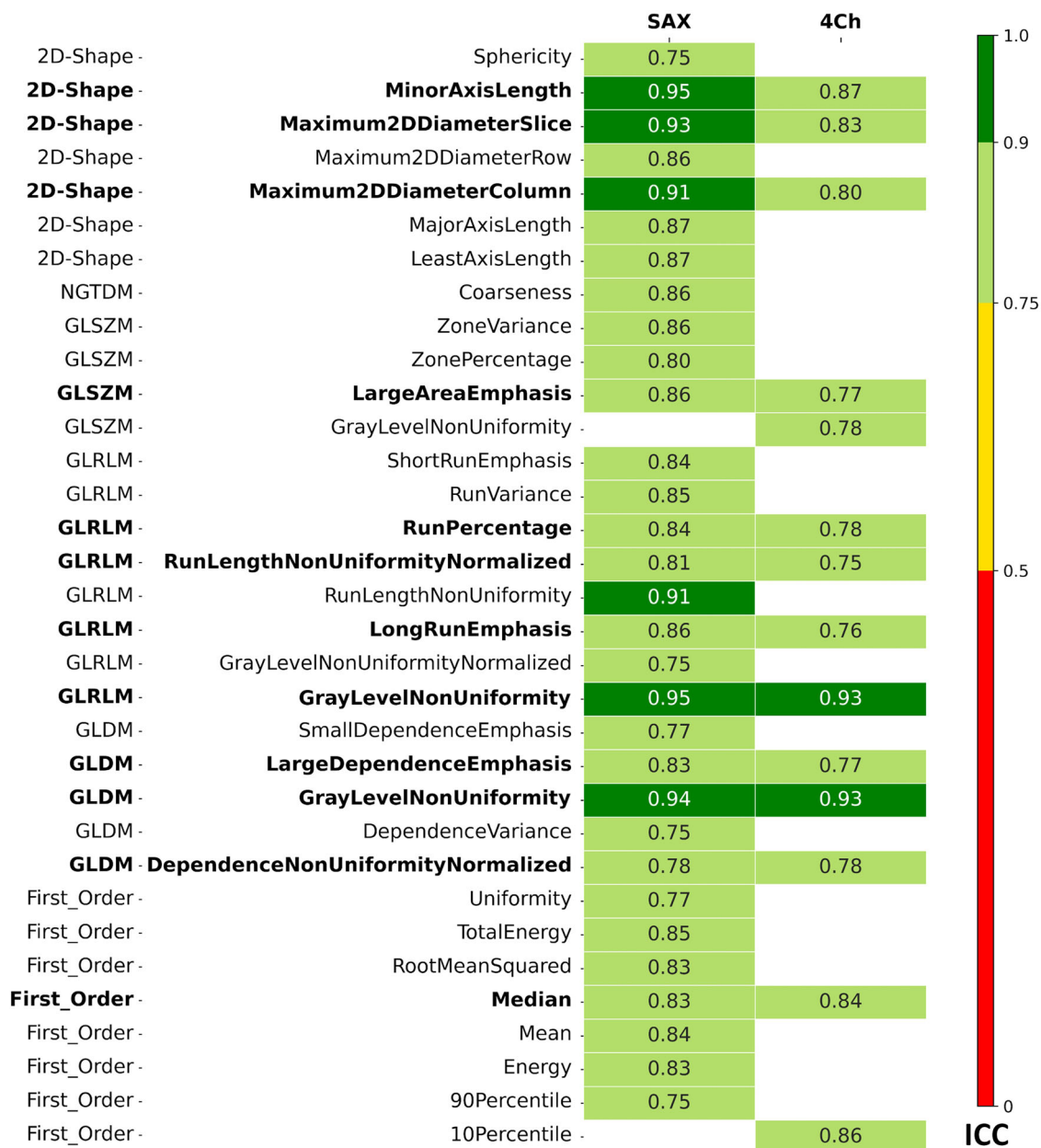


Fig. 6 Selected features for T2 mapping: only features with good to excellent reproducibility in at least one of the views (short-axis and 4-chamber view) are shown. ICC values are color-coded for both views (short-axis/4-chamber). ICC values ≤ 0.75 are not shown. The subset of 12 features that showed good to excellent repeatability (ICC > 0.75) in both views is shown in bold

which are defined as the length in pixels of consecutive pixels with the same grayscale value. The relatively good repeatability in this feature class may be related to the study cohort, which included only healthy subjects with relatively homogeneous T1 and T2 maps. The results may differ in patients with focal or diffuse myocardial disease.

Several of those features that we identified as reproducible have been investigated as diagnostic biomarkers.

Among six radiomic features found to discriminate between hypertensive heart disease and hypertrophic cardiomyopathy [11], two GLRLM features (RunLengthNonUniformity and ShortRunEmphasis) were confirmed as highly reproducible in our study. For the diagnosis of acute myocarditis, the combination of two texture features (kurtosis on T2 mapping and GrayLevelNonUniformity on T1 mapping) has been reported as

providing the highest diagnostic performance [5]. While GrayLevelNonUniformity on T1 mapping was among the subset of features with high repeatability in our study, we found poor repeatability of kurtosis on T2 mapping. This highlights the need for future radiomic research to focus on features that have been shown to be highly reproducible.

We chose the intraclass correlation coefficient (ICC) as the primary parameter of repeatability in our study. The CCC and Pearson correlation coefficient were used as ancillary parameters to confirm the results. The ICC proposed by Fisher [35], assumes that diagnostic test results follow a one-way ANOVA model with random subject effects and take repeated measurements into account. This model can be extended for replicated measurements and subject-rater interactions, with different versions discussed by Shrout and Fleiss [25]. The CCC proposed by Lin [28] measures agreement without relying on an ANOVA model. The CCC combines consistency (measured by Pearson correlation) and bias, assessing both the closeness of observations to the regression line and the alignment of the regression line to perfect fit [36]. Some authors [29, 36] discuss the limitations of the Pearson correlation coefficient r in assessing agreement between test scores, although it can measure the strength of a linear relationship.

In order to check the relationships between these three metrics of reproducibility, they were plotted graphically against each other (Suppl. Fig. 3). The ICC and CCC show a very strong linear relationship. The dependencies of ICC and CCC on the Pearson correlation coefficient r also indicate a linear relationship between the metrics with a small degree of scatter.

Implications for practice

In this work, we have identified 32 radiomics features for short-axis T1 mapping and 15 features for T1 mapping regardless of view and resolution, as well as 12 radiomic features for T2 mapping also independent of view, as particularly robust features with good to excellent repeatability. We recommend these parameters for further research on patients with heart disease. Radiomics capabilities could help distinguish between different heart diseases by revealing subtle differences in tissue texture, morphology, and function that are not visible to the human eye. Another use case could arise for monitoring the success of therapy by using the functions to quantify changes in heart tissue and function during and after treatment. This can help tailor therapy and maximize effectiveness. Nevertheless, further investigation of radiomics data is required to discover new biomarkers and develop new diagnostic and therapeutic approaches. By integrating radiomics capabilities into clinical trials,

they can help validate new therapies and improve patient care.

Study limitations

This prospective study was limited to healthy volunteers. Therefore, we cannot directly infer the reproducibility of radiomic features in patients with diffuse or focal myocardial disease. All maps were calculated using one manufacturer's 1.5-T MRI scanner. Further work is required to confirm the robustness of the recommended radiomic features for T1 and T2 mapping performed with other manufacturers or different field strengths. Likewise, our study was limited to native T1. Without further validation, it is not clear whether the high robustness features also perform equally well for post-contrast T1 maps or extracellular volume (ECV) maps.

Conclusion

In myocardial T1 and T2 mapping, only a minority of radiomic features have sufficient repeatability to qualify as a reliable imaging biomarker. In this study, we have identified a subset of radiomic features that show good to excellent repeatability independent of slice orientation and spatial resolution. We recommend considering these features for further radiomics research in myocardial T1 and T2 mapping.

Abbreviations

4Ch	Four-chamber view
CCC	Concordance correlation coefficient
GLCM	Gray Level Co-occurrence Matrix features
GLDM	Gray Level Dependence Matrix features
GLRLM	Gray Level Run Length Matrix features
GLSZM	Gray Level Size Zone Matrix features
ICC	Intraclass correlation coefficient
MRI	Magnetic resonance imaging
NGTDM	Neighboring Gray Tone Difference Matrix features
ROI	Region of interest
SAX	Short-axis view

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1007/s00330-024-11337-8>.

Funding

This study was funded by the Federal Ministry of Education and Research (BMBF) through the Network University Medicine "NUM 2.0" (grant number 01KX2121). The funders had no role in study design, data analysis, data interpretation, or manuscript drafting. Open Access funding enabled and organized by Projekt DEAL.

Data availability

All data are available upon request from the corresponding author.

Code availability

The software source code is available on GitHub (<https://github.com/voxelacrobot/Radiomics-on-cardiac-MRI-mapping/>).

Compliance with ethical standards

Guarantor

The scientific guarantor of this publication is Felix G. Meinel.

Conflict of interest

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry

One of the authors (Mathias Manzke) has significant statistical expertise.

Informed consent

Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval

Institutional Review Board approval was obtained (Ethics Committee, University Medical Center Rostock).

Study subjects or cohorts overlap

Previous publications based on the same study cohort have investigated the reproducibility of global and regional mean T1 and T2 values [23] and the influence of spatial resolution on mean T1 and T2 values and their reproducibility [24]. In this present work, we investigate the repeatability of radiomic features in T1 and T2 mapping. The radiomics data presented in this manuscript has not been previously reported.

Methodology

- Prospective
- Diagnostic or prognostic study
- Performed at one institution

Received: 22 July 2024 Revised: 19 November 2024 Accepted: 5 December 2024

Published online: 15 January 2025

References

- Messroghli DR, Moon JC, Ferreira VM et al (2017) Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: a consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). *J Cardiovasc Magn Reson* 19:75
- Puntmann VO, Peker E, Chandrasekhar Y, Nagel E (2016) T1 mapping in characterizing myocardial disease: a comprehensive review. *Circ Res* 119:277–299
- Hassani C, Saremi F, Varghese BA, Duddalwar V (2020) Myocardial radiomics in cardiac MRI. *AJR Am J Roentgenol* 214:536–545
- Ponsiglione A, Stanzione A, Cuocolo R et al (2022) Cardiac CT and MRI radiomics: systematic review of the literature and radiomics quality score assessment. *Eur Radiol* 32:2629–2638
- Baessler B, Luecke C, Lurz J et al (2019) Cardiac MRI and texture analysis of myocardial T1 and T2 maps in myocarditis with acute versus chronic symptoms of heart failure. *Radiology* 292:608–617
- Nakamori S, Amyar A, Fahmy AS et al (2024) Cardiovascular magnetic resonance radiomics to identify components of the extracellular matrix in dilated cardiomyopathy. *Circulation*. <https://doi.org/10.1161/CIRCULATIONAHA.123.067107>
- Wang J, Yang F, Liu W et al (2020) Radiomic analysis of native T1 mapping images discriminates between MYH7 and MYBPC3-related hypertrophic cardiomyopathy. *J Magn Reson Imaging* 52:1714–1721
- Antonopoulos AS, Boutsikou M, Simantiris S et al (2021) Machine learning of native T1 mapping radiomics for classification of hypertrophic cardiomyopathy phenotypes. *Sci Rep* 11:23596
- Neisius U, El-Rewaify H, Kucukseymen S et al (2020) Texture signatures of native myocardial T1 as novel imaging markers for identification of hypertrophic cardiomyopathy patients without scar. *J Magn Reson Imaging* 52:906–919
- Wang Z-C, Fan Z-Z, Liu X-Y et al (2024) Deep learning for discrimination of hypertrophic cardiomyopathy and hypertensive heart disease on MRI native T1 maps. *J Magn Reson Imaging* 59:837–848
- Neisius U, El-Rewaify H, Nakamori S, Rodriguez J, Manning WJ, Nezafat R (2019) Radiomic analysis of myocardial native T1 imaging discriminates between hypertensive heart disease and hypertrophic cardiomyopathy. *JACC Cardiovasc Imaging* 12:1946–1954
- Son J, Hong YJ, Kim S et al (2022) Radiomics feature analysis using native T1 mapping for discriminating between cardiac tumors and thrombi. *Acad Radiol* 29:S1–S8
- Chang S, Han K, Kwon Y et al (2023) T1 map-based radiomics for prediction of left ventricular reverse remodeling in patients with non-ischemic dilated cardiomyopathy. *Korean J Radiol* 24:395–405
- Zhang J, Xu Y, Li W et al (2023) The predictive value of myocardial native T1 mapping radiomics in dilated cardiomyopathy: a study in a Chinese population. *J Magn Reson Imaging* 58:772–779
- Wu L-M, Shi R-Y, Wu C-W et al (2022) A radiomic MRI based nomogram for prediction of heart failure with preserved ejection fraction in systemic lupus erythematosus patients: insights from a three-center prospective study. *J Magn Reson Imaging* 56:779–789
- Ma Q, Ma Y, Yu T, Sun Z, Hou Y (2021) Radiomics of non-contrast-enhanced T1 mapping: diagnostic and predictive performance for myocardial injury in acute ST-segment-elevation myocardial infarction. *Korean J Radiol* 22:535–546
- Wang Y, Lin L, Li X et al (2022) Native T1 mapping-based radiomics for noninvasive prediction of the therapeutic effect of pulmonary arterial hypertension. *Diagnostics (Basel)* 12:2492
- Akinci D'Antonoli T, Cuocolo R, Baessler B, Pinto Dos Santos D (2024) Towards reproducible radiomics research: introduction of a database for radiomics studies. *Eur Radiol* 34:436–443
- Baeßler B, Weiss K, Pinto Dos Santos D (2019) Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. *Invest Radiol* 54:221–228
- Woznicki P, Laqua F, Bley T, Baeßler B (2022) AutoRadiomics: a framework for reproducible radiomics research. *Front Radiol* 2:919133
- Raisi-Estabragh Z, Gkontra P, Jaggi A et al (2020) Repeatability of cardiac magnetic resonance radiomics: a multi-centre multi-vendor test-retest study. *Front Cardiovasc Med* 7:586236
- Jang J, El-Rewaify H, Ngo LH et al (2021) Sensitivity of myocardial radiomic features to imaging parameters in cardiac MR imaging. *J Magn Reson Imaging* 54:787–794
- Böttcher B, Lorbeer R, Stöcklein S et al (2021) Global and regional test-retest reproducibility of native T1 and T2 mapping in cardiac magnetic resonance imaging. *J Magn Reson Imaging* 54:1763–1772
- Dalmer A, Meinel FG, Böttcher B et al (2024) Native myocardial T1 mapping: influence of spatial resolution on quantitative results and reproducibility. *Quant Imaging Med Surg* 14:20–30
- Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 86:420–428
- Mayerhoefer ME, Materka A, Langs G et al (2020) Introduction to radiomics. *J Nucl Med* 61:488–495
- Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163
- Lin L-K (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45:255
- Chen C-C, Barnhart HX (2008) Comparison of ICC and CCC for assessing agreement for data without and with replications. *Comput Stat Data Anal* 53:554–564
- Marfisi D, Giannelli M, Marzi C et al (2024) Test-retest repeatability of myocardial radiomic features from quantitative cardiac magnetic resonance T1 and T2 mapping. *Magn Reson Imaging* 113:110217
- Jang J, Ngo LH, Mancio J et al (2020) Reproducibility of segmentation-based myocardial radiomic features with cardiac MRI. *Radiol Cardiothorac Imaging* 2:e190216
- Marfisi D, Tessa C, Marzi C et al (2022) Image resampling and discretization effect on the estimate of myocardial radiomic features

- from T1 and T2 mapping in hypertrophic cardiomyopathy. *Sci Rep* 12:10186
33. Zhao W, Hu Z, Kazerooni AF et al (2024) Physics-informed discretization for reproducible and robust radiomic feature extraction using quantitative MRI. *Invest Radiol* 59:359–371
34. Hu Z, Zhuang Q, Xiao Y et al (2021) MIL normalization—prerequisites for accurate MRI radiomics analysis. *Comput Biol Med* 133:104403
35. Fisher RA (1992) Statistical methods for research workers. In: Kotz S, Johnson NL (eds) *Breakthroughs in statistics*. Springer New York, New York, NY, pp. 66–70
36. Morgan CJ, Aban I (2016) Methods for evaluating the agreement between diagnostic tests. *J Nucl Cardiol* 23:511–513

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.