

Radiomics in der Mamma MRT:  
Klassifikation kontrastmittelaufnehmender Läsionen nach ihrer Dignität mittels zweier  
klassischer Machine Learning Algorithmen

Von der Medizinischen Fakultät  
der Rheinisch-Westfälischen Technischen Hochschule Aachen  
zur Erlangung des akademischen Grades eines Doktors der Medizin  
genehmigte Dissertation

vorgelegt von

Hannah Sophia Schneider  
aus Singen am Hohentwiel

Berichter: Frau Privatdozentin Dr. med. Simone Schrading  
Frau Privatdozentin Dr. med. Laila Najjari

Tag der mündlichen Prüfung: 12.06.2020

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

## **Hinweis**

Diese Arbeit wurde als Teil der Veröffentlichung „Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-enhancing Lesions at Multiparametric Breast MRI. Radiology“, von Truhn, D, Schradang, S, Haarbüger, C, Schneider, H, Merhof, D, Kuhl, C, erschienen in Radiology. 2019 Feb;290(2):290-297, bereits vorab veröffentlicht.

# Inhaltsverzeichnis

<b>Hinweis</b> .....	<b>I</b>
<b>Inhaltsverzeichnis</b> .....	<b>II</b>
<b>Abkürzungsverzeichnis</b> .....	<b>IV</b>
<b>Einleitung</b> .....	<b>- 1 -</b>
<b>Brustkrebs und Mamma MRT</b> .....	<b>- 2 -</b>
Die Rolle der Mamma MRT in der Brustkrebsdiagnostik .....	- 2 -
Durchführung und Befundung der Mamma MRT .....	- 4 -
Diagnosesicherung mittels Histopathologie .....	- 7 -
<b>Radiomics</b> .....	<b>- 8 -</b>
Begriffsklärung .....	- 8 -
Methodik.....	- 11 -
<b>Zielsetzung</b> .....	<b>- 18 -</b>
<b>Material und Methoden</b> .....	<b>- 19 -</b>
<b>Studienpopulation</b> .....	<b>- 19 -</b>
<b>Durchführung und Befundung der Mamma MRT an der Uniklinik Aachen</b> .....	<b>- 20 -</b>
<b>Erhebung der Zieldaten</b> .....	<b>- 20 -</b>
<b>Radiomics</b> .....	<b>- 21 -</b>
Segmentierung mit MeVisLab 3.0 .....	- 21 -
Featureextraktion .....	- 23 -
Featureselektion .....	- 23 -
Validierung .....	- 24 -
Statistische Auswertung.....	- 25 -
<b>Ergebnisse</b> .....	<b>- 27 -</b>
<b>Studienkohorte</b> .....	<b>- 27 -</b>
<b>L1-Regularisierung und Hauptkomponentenanalyse</b> .....	<b>- 31 -</b>
Selektierte Features der L1-Regularisierung .....	- 31 -
Selektierte Features der Hauptkomponentenanalyse .....	- 33 -
<b>Statistische Auswertung</b> .....	<b>- 34 -</b>
<b>Diskussion</b> .....	<b>- 37 -</b>
<b>Selektierte Features</b> .....	<b>- 39 -</b>
<b>Limitationen</b> .....	<b>- 42 -</b>

<b>Ausblick .....</b>	<b>- 44 -</b>
<b>Zusammenfassung .....</b>	<b>- 47 -</b>
<b>Literatur .....</b>	<b>- 48 -</b>
<b>Anhang .....</b>	<b>- 54 -</b>
<b>Selektionsfrequenzen der L1-Regularisierung.....</b>	<b>- 54 -</b>
<b>Selektionsfrequenzen der Hauptkomponentenanalyse .....</b>	<b>- 62 -</b>

## Abkürzungsverzeichnis

ACR	American College of Radiology
ACS	American Cancer Society
AUC	Area under the Curve = Flächeninhalt unter der (ROC) Kurve
BI-RADS®	Beast Imaging Reporting and Data System
BRCA 1/2	Breast Cancer 1/2– zwei Tumorsuppressorgene
CADx	Computer Assisted Diagnostics
CNN	Convolutional Neural Network
CT	Computertomografie
DCIS	Duktales Carcinoma in Situ
ER	Östrogenrezeptor
ESMO	European Society for Medical Oncology
EUSOBI	European Society of Breast Imaging
GLCM	Gray Level Co-Occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
MIP	Maximumintensitätsprojektion
MRT	Magnetresonanztomografie
NCI	National Cancer Institute
PCA	Principal Component Analysis = Hauptkomponentenanalyse
PR	Progesteronrezeptor
ROC	Receiver Operating Characteristics
ROI	Region of Interest
VOI	Volume of Interest

## Einleitung

„Aus radiologischen Bildern quantitative Merkmale in großem Stil extrahieren, speichern und sie in klinischen Kontext setzen.“ – So kann das Gebiet der Radiomics beschrieben werden, welches sich in den letzten Jahren als inzwischen breites Forschungsfeld an der Schnittstelle zwischen digitaler Bildverarbeitung und der Radiologie entwickelt hat. Computerunterstützte Diagnosesysteme, welche mit konventionellen Methoden der Bilderkennung arbeiten, existieren zwar schon seit den 1960er Jahren, jedoch konnte erst in den letzten Jahrzehnten die Rechenleistung und Speicherkapazität von Computern so verbessert werden, dass ein breiter Einsatz bildverarbeitender Mittel in den radiologischen Disziplinen ermöglicht wurde (Giger 2018). Was genau versteht man nun unter Radiomics? Radiomics – ein Kofferwort bestehend aus „Radiologie“ und der Endsilbe „-omics“, wie sie in „genomics“ verwendet wird – bedeutet zunächst einmal die Konversion von medizinischen Bildinformationen zu speicher- und nutzbaren Daten. Das heißt, dass die visuellen Informationen aus CT, MRT oder weiteren Modalitäten, welche von einem menschlichen Betrachter qualitativ verarbeitet werden, zu quantifizierbaren (also in Zahlen fassbaren) und somit von Computern nutzbaren Informationen umgewandelt werden – und das in einer großen Menge. Ziel ist die Verknüpfung der Bild- mit klinischen, pathologischen oder genetischen Daten, um aus ihnen Aussagen über medizinisch relevante Parameter treffen zu können – ob redundant oder komplementär zu bereits etablierten Biomarkern. Dies geschieht über die Anwendung von Machine Learning Algorithmen, mithilfe derer Modelle entwickelt werden, welche die jeweiligen Aussagen möglichst zuverlässig treffen. Zukünftig erhofft man sich dadurch zum einen die Diagnoseunterstützung des Radiologen, zum anderen (und im Wesentlichen) wird die Einbettung von Radiomics in die Präzisionsmedizin, vor allem in die Onkologie, für die Entwicklung bildgebender in-vivo Biomarker von prognostischer und prädiktiver Bedeutung angestrebt (Aerts et al. 2014; Avanzo, Stancanello, and El Naqa 2017).

Bevor diese klinische Einbettung möglich werden kann, muss das Gebiet der Radiomics zunächst noch gründlich erforscht werden. Da für ein robustes und präzises Modell große Anzahlen an Untersuchungen notwendig sind, ist die Studienlage für die Mamma MRT, welche vergleichsweise selten durchgeführt wird, bisher eher dünn mit niedrigen

Fallzahlen. In dieser Arbeit wird eine Anwendung von Radiomics in der Mamma MRT untersucht mit dem Ziel, ein Modell zu entwickeln, welches intramammäre Tumoren in invasiven Brustkrebs und benigne Läsionen zu differenzieren vermag.

## Brustkrebs und Mamma MRT

### Die Rolle der Mamma MRT in der Brustkrebsdiagnostik

Mit etwa einem Viertel aller Krebserkrankungen ist Brustkrebs weltweit die häufigste Krebserkrankung in der weiblichen Bevölkerung. In Deutschland ist er unter Frauen die häufigste, in Nordamerika die zweithäufigste krebsbedingte Todesursache (Bray et al. 2018; Siegel, Miller, and Jemal 2019). Fast 70.000 Frauen jährlich erkranken deutschlandweit neu an Brustkrebs, was einer Inzidenz von 117 pro 100.000 (DeSantis et al. 2015; Robert-Koch-Institut 2017) entspricht. Im Schnitt wird bei jeder achten Frau in Deutschland in ihrem Leben ein Mammakarzinom diagnostiziert (kumulatives Lebenszeitrisiko 12.8%), in einem medianen Alter von 64 Jahren (Robert-Koch-Institut 2017). Seit Einführung des Mammografie-Screenings konnten die Überlebensraten zwar deutlich verbessert werden (derzeitige relative 5-Jahre-Überlebensrate in Deutschland von etwa 90%), allerdings profitieren nicht alle Frauen vom Screening-Programm: 20-40% der Mammakarzinome bei Frauen im Screening sind sogenannte Intervallkarzinome, das heißt sie werden im Zeitraum zwischen einer negativen Mammografie und der regelhaften Folgeuntersuchung klinisch manifest (Seely and Alhassan 2018). Die Mamma-MRT hingegen weist eine Sensitivität von zumeist über 90% (Range 75% bis 100%) auf, mit ihr können also fast alle Malignome der weiblichen Brust detektiert werden (Mann, Kuhl, and Moy 2019; Kuhl et al. 2017). Im Gegensatz zur rein anatomischen Darstellung der weiblichen Brust wie in der Mammografie können mit der MR-Mammografie zusätzliche, funktionale Informationen über das Brustgewebe gewonnen werden, welche auch über die Sonografie als dynamische Untersuchung mit gezielter Beurteilung der Gewebsdurchblutung hinausgehen. So kann in der MRT durch die Applikation von Kontrastmittel die Gewebepерfusion der gesamten Brust beurteilt werden und hierdurch Rückschlüsse auf eine mögliche Angiogenese, wie sie bei Malignomen typisch ist, gezogen werden (Bluff et al. 2009). Durch das schnelle Zellwachstum wird über Wachstumsfaktoren eine pathologische Angiogenese induziert, welche sich mit unorganisierten, atypisch verlaufenden, mäandernden und teilweise

dilatierten Tumorgefäßen präsentiert, welche darüber hinaus eine erhöhte Permeabilität und eine eingeschränkte Perfusion aufweisen (Viallard and Larrivee 2017). Diese Tumorangiogenese stellt sich in Form charakteristischer Kontrastmittelanreicherungen innerhalb der suspekten Läsion bezüglich des Musters und/oder des zeitlichen Verlaufs dar (Khalifa et al. 2014). Hierbei ist besonders für die Kontrastmittelkinetik bekannt, dass sie auf pathologische Veränderungen des Gewebes hinweist und somit einen hohen diagnostischen Wert besitzt (Kuhl 2007a; Baltzer, Dietzel, and Kaiser 2013; Tozaki, Igarashi, and Fukuda 2006; Mahoney et al. 2012). Trotz der eindeutigen Überlegenheit gegenüber den Standardverfahren in der Brustkrebsdiagnostik bleibt die Indikationsstellung für eine Mamma MRT streng (Sardanelli et al. 2011). Als Früherkennungsinstrument wird sie vom National Cancer Institute (NCI), der American Cancer Society (ACS) und der European Society for Medical Oncology (ESMO) lediglich bei Frauen mit einem erhöhten Risikoprofil (Lebenszeitrisko >20%), zum Beispiel bei BRCA-positiv getesteten Patientinnen und deren erstgradig Verwandten, empfohlen (Saslow et al. 2007; Senkus et al. 2015; NCI 2019). In den deutschen S3-Leitlinien wird für einen Einsatz der Mamma MRT zur Früherkennung bei Risikopatientinnen eine transparente Qualitätssicherung gefordert (Leitlinienprogramm Onkologie Deutsche Krebsgesellschaft 2017). Zur Primärdiagnostik kommt sie dann zum Einsatz, wenn bei fortbestehendem klinischen Verdacht konventionelle Bildgebungen inkonklusiv bleiben oder eine radiologisch-pathologische Uneinigkeit besteht. Etwas häufiger wird die Indikation bei bereits histologisch gesichertem Brustkrebs gestellt, da durch die MRT ein genaues Lokalstaging ermöglicht wird und somit die operative Therapie optimiert werden kann (Kuhl 2007b). Empfohlen wird sie jedoch nicht generell, sondern sollte bei bestimmten Tumorarten (z.B. lobuläre Histologie) und/oder dem Verdacht einer Multifokalität bzw. Multizentrität zur präoperativen Ausbreitungsdiagnostik erwogen werden (Senkus et al. 2015). Bei geplanter neoadjuvanter Chemotherapie ist eine MR-tomografische Verlaufskontrolle vor, während und nach dieser sinnvoll (Senkus et al. 2015; Kuhl 2007a). Bei Patientinnen mit einem erhöhten Rezidivrisiko, d.h. insbesondere bei jungen Patientinnen, Patientinnen mit hoher mammografischer Brustdichte oder familiärem Brustkrebs ist die Mamma MRT auch zur Nachsorge indiziert (van Bodegraven et al. 2017). Gründe für die seltene Durchführung sind vor allem die höheren Untersuchungskosten und die geringere Verfügbarkeit an radiologischen Diagnostikeinrichtungen, welche die Mamma MRT

anbieten. Die resultierende mangelnde Erfahrung in Befundung und Interpretation der MRT führt zu einer oft postulierten eher moderaten Spezifität dieser und zu einer hohen Falsch-Positiv-Rate. In früheren Studien wurde die Spezifität mit nur 37% bis zu 85% evaluiert, abhängig von der Studienpopulation und der MR-Technik – vor allem aber von der Expertise des Untersuchers (Niell et al. 2017). In der aktuellen Literatur können mit der Mamma MRT bei erfahrenem Untersucher hingegen Spezifitäten von 83% bis 98% erreicht werden und auch der positive prädiktive Wert der MRT ist mit 11-40% gleichwertig zur Mammografie (Mann, Kuhl, and Moy 2019; Kuhl et al. 2018). Das heißt, an spezialisierten Zentren kann durch Erfahrung in der Befundung vielen Frauen ein invasiver bioptischer Eingriff erspart werden, wenn in der MRT eine Malignität ausgeschlossen werden kann. Durch eine computerbasierte, präzise Diagnoseunterstützung ließe sich dies auf weniger erfahrene Zentren übertragen und somit die Spezifität der Mamma MRT im Gesamten erhöhen, wozu die Untersuchung quantitativer Bildmerkmale mittels Radiomics einen ersten Ansatz darstellt. Darüber hinaus könnte sich die Rolle der Mamma MRT in Zukunft durch die Nutzung bildgebender in-vivo Biomarker ändern, da eine potenzielle Nutzung zum Beispiel zur Prognoseabschätzung oder Therapiekontrolle durch Radiomics möglich werden kann.

## Durchführung und Befundung der Mamma MRT

Im Gegensatz zur standardisiert durchgeführten Mammografie existiert bis heute für die Mamma MRT kein Standard-Protokoll für ihre Durchführung. Durch die European Society of Breast Imaging (EUSOBI) werden Empfehlungen zur den Rahmenbedingungen und der Aufnahmetechnik gegeben, welche die folgenden Annahmen berücksichtigen und in ähnlicher Ausführung im „Breast Imaging Reporting and Data System“ (BI-RADS®) Katalog von 2013 aufgenommen wurden (Mann et al. 2008; Morris EA 2013).

- Um eine hohe räumliche Auflösung zu erreichen, sollte die Feldstärke mindestens 1,5 Tesla betragen und eine spezielle Brustspule genutzt werden; die maximale Schichtdicke beträgt 3 mm, die maximale Pixelgröße 1 mm. Die Aufnahmen erfolgen bilateral und zumeist in der axialen Ebene. Zusätzliche koronare Schnitte können zur besseren Beurteilung der axillären Lymphknotenstationen vonnöten sein.
- Zur Minimierung von Bewegungsartefakten wird die Brust in Aufnahmerichtung fixiert, das bedeutet für die axiale Schnittführung eine kraniokaudale Fixierung der Brust. Hierdurch wird auch eine Reduktion der aufzunehmenden Schnitte erreicht.

- Als Kontrastmittel wird in der Brustuntersuchung ein intravenös appliziertes, gadoliniumhaltiges, also T1-verkürzendes, Kontrastmittel eingesetzt, üblicherweise in einer Dosis von 1 mmol/kgKG und mit einer Rate von 3 ml/sec. Vor und nach Kontrastmittelapplikation werden mehrere T1-gewichtete, bevorzugt fettunterdrückte Puls-Sequenzen akquiriert, sodass man eine native und 3 bis 5 Post-Kontrastaufnahmen in 1- bis 2-minütigem Abstand (dynamische Serie) erhält. Außerdem werden Subtraktionsbilder angefertigt, um das Kontrastmittelverhalten zu untersuchen. Intensitätskurven der Kontrastmittelaufnahme für bestimmte Bildpunkte bzw. -regionen können berechnet werden. Übliche Untersuchungsabfolgen enthalten zusätzlich eine axiale T2-Sequenz.
- Ein verkürztes Untersuchungsprotokoll, welche die native T1-Sequenz und die erste Post-Kontrast-Aufnahme sowie deren Subtraktionsbild und Maximumintensitätsprojektion (MIP) beinhaltet, erbrachte in der Screening-Situation vergleichbare Sensitivitäten wie herkömmliche Protokolle (Kuhl et al. 2014).

Die Befundung der akquirierten Bilder durch den Radiologen erfolgt meist systematisch anhand des BI-RADS®-Katalogs, welcher Anfang der 1990er Jahre durch das American College of Radiology (ACR) zunächst für die Mammografie eingeführt und später für den Brustultraschall und die Mamma MRT erweitert wurde (Burnside et al. 2009). Nach mehreren Überarbeitungen wurde 2013 mit BI-RADS® 5 die aktuellste Ausgabe veröffentlicht (Morris EA 2013). Innerhalb des BI-RADS® findet sich neben der Einteilung von Brustdichte und Hintergrundanreicherung ein Lexikon verschiedener Deskriptoren, d.h. bildgebende Charakteristika einer intramammären Läsion, welche zwischen malignen und benignen Läsionen zu unterscheiden vermögen (Burnside et al. 2009). Hierbei wird nach verschiedenen Arten von Läsionen unterschieden (focus, mass lesion, non-mass enhancement, non-enhancing findings und weitere), welche zunächst identifiziert und anschließend anhand ihrer Form, Begrenzung und dem Anreicherungsverhalten des Kontrastmittels weiter beschrieben werden. Dabei sind besonders mass lesions (solide Tumoren) und non-mass enhancements (nicht-tumoröse Anreicherungen) von Interesse, da diese am ehesten malignitätsverdächtig sein können. Als mass lesion wird eine solide, raumeinnehmende, d.h. auch in den nativen Sequenzen abgrenzbare, kontrastmittelanreichernde Läsion bezeichnet, welche grundsätzlich benignen oder malignen Charakters haben kann. Findet sich in der Brustdrüse eine

solche tumoröse Läsion, wird diese anhand ihrer Form, Begrenzung, Binnenstruktur und dem intratumoralen Anreicherungsverhalten nach Kontrastmittelinjektion (subjektiv und/oder objektiv mittels Zeit-Intensitätskurve) beurteilt. Eine non-mass enhancement beschreibt eine Mehranreicherung des Brustdrüsengewebes ohne raumfordernde Komponente, d.h. ohne Korrelat in den nativen Sequenzen. Diese wird anhand ihres Verteilungsmusters (fokal, segmental, regional, diffus), des Anreicherungsmodus (homo- oder heterogen, "clumped" oder "clustered") und der Kontrastmittelkinetik beschrieben. Im Falle eines Malignoms unterscheiden sich tumoröse Läsionen und nicht-tumoröse Anreicherungen meist in ihrer Histopathologie: nicht-tumoröse Anreicherungen sind typisch für nicht-invasiven (duktales Carcinoma in situ = DCIS) oder lobulären Brustkrebs. Duktales Carcinoma vom "no-special type" stellt sich meist als solide tumoröse Läsion dar.

Typische Befunde, die für die Malignität einer mass lesion bzw. eines non-mass enhancements sprechen, sind (Lieberman et al. 2002; Kuhl 2007a; Gutierrez et al. 2009; Mahoney et al. 2012):

- Mass lesion:
  - spikuliert, asymmetrisch oder irreguläre Form mit unscharfer Berandung bzw. diffusem Wachstumsmuster
  - inhomogene Binnenstruktur und/oder Verteilung des Kontrastmittels mit Rim Enhancement
  - typische Kontrastmittelkinetik: schnelles Anfluten des Kontrastmittels mit nachfolgend schnellem Auswaschen oder einer Plateauphase.
- Non-mass enhancement:
  - segmental oder diffus verteiltes, rasch mehranreicherndes Areal
  - Clumped Enhancement

Bei sowohl tumorösen als auch nicht-tumorösen Läsionen ist besonders für die Kinetik der Kontrastmittelanreicherung, d.h. die Form der Zeit-Intensitätskurven, bekannt, dass sie gut zwischen malignen und benignen Läsionen zu differenzieren vermag (Kuhl 1999). Nach der Beschreibung der verschiedenen intramammären Läsionen gibt der befundende Radiologe in Zusammenschau aller Parameter eine Einschätzung, ob es sich um eine benigne oder maligne Läsion handelt. Diese wird in 6 BI-RADS®-Kategorien, mit jeweiligen Empfehlungen zum weiteren Vorgehen, unterteilt, wobei jeweils ein Zahlenwert pro Brustseite - nicht pro Läsion - vergeben wird:

Tabelle 1: MR-BI-RADS® Kategorien nach der 5. Ausgabe des ACR BI-RADS® Katalogs von 2013

BI-RADS® 1	negativ, d.h. keine intramammäre Läsion	Fortführung der regulären Früherkennung
BI-RADS® 2	sicher benigne Läsion	Fortführung der regulären Früherkennung
BI-RADS® 3	wahrscheinlich benigne Läsion	kurzfristige Kontrolle (nach 3-6 Monaten) empfohlen
BI-RADS® 4	suspekte Läsion	bioptische Sicherung sollte erwogen werden
BI-RADS® 5	sehr wahrscheinlich maligne Läsion	entsprechende Abklärung empfohlen
BI-RADS® 6	histologisch gesicherte maligne Läsion	entsprechende Therapieeinleitung empfohlen

### Diagnosesicherung mittels Histopathologie

Finden sich bei einer Patientin in der Bildgebung ein oder mehrere BI-RADS® 4 oder 5 Befunde, so wird in der Regel eine bildgesteuerte Biopsie durchgeführt um die Verdachtsdiagnose zu bestätigen oder auszuschließen. Nach einer ersten histopathologischen Untersuchung kann das entnommene Gewebe als benigne Läsion, Risikoläsion (d.h. Läsionen, welche ein gewisses Entartungsrisiko beherbergen), nicht-invasiven Brustkrebs (DCIS) oder invasiven Brustkrebs klassifiziert werden. Findet sich ein nicht-invasiver oder invasiver Brustkrebs, so folgt eine weitere Aufarbeitung des Gewebes, wobei standardmäßig histologischer Typ, Tumorgrad und Östrogenrezeptorstatus unterschieden werden. Beim invasiven Karzinom werden zusätzlich Progesteron- und Her-2-Rezeptorstatus sowie der Proliferationsmarker Ki-67 immunhistochemisch bestimmt (Senkus et al. 2015). Hierdurch lässt sich das Karzinom subtypisieren - denn Brustkrebs stellt nicht nur eine Krebsart, sondern eine heterogene Gruppe von Malignomen mit unterschiedlichen molekularen Merkmalen dar, welche sich hinsichtlich klinischer Präsentation, Prognose und des Ansprechens auf eine zytostatische Therapie unterscheiden (Sorlie et al. 2001; Voduc et al. 2010; Prat et al. 2015; Koh and Kim 2019). Üblicherweise werden 4 verschiedene molekulare Subtypen unterschieden: Luminal A, Luminal B, Her-2-überexprimierende Tumoren und basal-like

Tumoren (Reis-Filho and Puztai 2011; Goldhirsch et al. 2013; Anderson et al. 2014). Durch die routinemäßig durchgeführte Immunhistochemie kann eine annähernde Einteilung dieser Subtypen erfolgen. Diese entspricht nicht in Gänze der molekularen Einteilung, sondern ist eine Annäherung daran. Zum Beispiel entspricht er immunhistochemische Subtyp „triple-negativ“ nicht nur dem basal-like Subtyp, sondern enthält weitere spezielle molekulare Subtypen. Des Weiteren wird die Differenzierung zwischen Luminal A- und B-Tumoren, welche auf molekularer Ebene eindeutig ist, auf immunhistochemischer Ebene durch mehrere Eigenschaften charakterisiert. Hinzu kommt, dass die immunhistochemische Einteilung der Subtypen in der Literatur nicht einheitlich verwendet wird.

Eine Übersicht der aktuellen Definition der verschiedenen Subtypen nach Goldhirsch gibt die folgende Tabelle (Goldhirsch et al. 2011):

*Tabelle 2: Molekulare Subtypen des Mammakarzinoms nach ihren immunhistochemischen Markern nach Goldhirsch*

	<b>ER/PR</b>	<b>Her2</b>	<b>Ki67</b>
<b>Luminal A</b>	positiv	negativ	niedrig
<b>Luminal B</b>	positiv	negativ	hoch
	positiv	positiv	jedes
<b>Her2-positiv</b>	negativ	positiv	jedes
<b>Basal-like</b>	negativ	negativ	jedes

## Radiomics

### Begriffsklärung

*Radiomics* bezeichnet die Extraktion größerer Mengen quantitativer Bildfeatures aus Schnittbilduntersuchungen, das heißt die Konversion der Bildinformationen zu mathematisch definierten Bildmerkmalen für eine anschließende Weiterverarbeitung. Mit *quantitativen Bildfeatures* sind verschiedene mathematische Charakteristika einer Bildregion gemeint, zum Beispiel die Verteilung und Anordnung, Struktur oder Textur der Grauwerte innerhalb dieser. Für das Beispiel eines Tumors lassen sich über Bildfeatures dessen Phänotyp bezüglich Größe, Form, Berandung, Binnenstruktur, Textur und Kinetik der Kontrastmittelaufnahme und seine Umgebung quantifizieren –

also genau die Merkmale, die auch durch den Radiologen beurteilt werden. Darüber hinaus können weitere, für den Betrachter unsichtbare, komplexe Bildeigenschaften berechnet werden. Die Extraktion und Auswertung von Bildfeatures ist schon seit einigen Jahren in Form von computerunterstützten Diagnostiksystemen (CAD) insbesondere in der Mammografie sowohl zur Lokalisation auffälliger Areale (CAdE) als auch zur Klassifikation derer (CAdx) möglich (Giger 2010). Auch für die Mamma MRT wurden Diagnostiksysteme entwickelt, welche Läsionen automatisch detektieren und über extrahierte Features klassifizieren können. In der Regel werden dabei 20 bis 40 Features genutzt - vor allem jene, welche die Kinetik der Kontrastmittelanreicherung beschreiben, da für diese eine hohe Korrelation zur Malignitätswahrscheinlichkeit bekannt ist (Chen et al. 2010; Yang et al. 2015). Durch Radiomics werden diese Diagnosesysteme nun ausgebaut, indem eine deutlich größere Anzahl von Features extrahiert und Verknüpfungen zwischen diesen erstellt wird (Avanzo, Stancanello, and El Naqa 2017). Darüber hinaus wird angenommen, dass in den Bildern Informationen stecken, welche die pathophysiologischen Prozesse eines kranken Gewebes, zum Beispiel eines Tumors, widerspiegeln und diese Zusammenhänge über die Auswertung großer Mengen an Bilddaten gezeigt werden können (Gillies, Kinahan, and Hricak 2016). Dabei wird beispielsweise auf die Heterogenität der Tumoren auf biologischer (z.B. genetischer und histopathologischer) Ebene ein Augenmerk gelegt sowie auf deren Korrelation zu klinischen (z.B. prognostischen) Parametern. Die phänotypische Heterogenität kann mit bildgebenden Verfahren gut dargestellt und durch die Extraktion von Bildfeatures quantitativ erfasst werden, wobei sogenannte Texturfeatures eine zentrale Rolle spielen (Davnall et al. 2012; Gibbs and Turnbull 2003). Mit Radiomics kann es gelingen, einen Zusammenhang zwischen phänotypischer und biologischer Heterogenität sowie zwischen Phänotyp und Klinik herzustellen mit dem Ziel, Bildfeatures zu identifizieren, welche später in prognostischen und prädiktiven Modellen der Präzisionsmedizin genutzt werden können (Aerts et al. 2014; Giger 2018). Diese Zusammenhänge werden durch die Entwicklung mathematischer Modelle mittels Machine Learning Algorithmen untersucht, welche nach deren Training möglichst präzise Vorhersagen (Prädiktionen) über die Fragestellung liefern können. Um dieses Vorgehen genauer beleuchten zu können, müssen zunächst einige Begrifflichkeiten geklärt werden:

- *Machine Learning Algorithmus*: Meist parametrisches statistisches Modell, das Anhand einer Menge von Trainingsdaten parametrisiert wird und danach zur Prädiktion genutzt werden kann.
- *Statistisches Modell*: Mathematische Funktion, welche die Eingangsdaten miteinander verknüpft um eine Vorhersage über die Fragestellung zu treffen. Der Aufbau der Funktion - zum Beispiel linear oder höherdimensional, mit viel oder wenig Flexibilität - entscheidet über die Qualität der Vorhersagen, weshalb die Auswahl eines geeigneten Modells für die jeweilige Fragestellung essentiell ist.
- *Parameter*: Koeffizienten der mathematischen Funktion, über deren Anpassung das Modell trainiert wird.
- *Hyperparameter*: Übergeordnete definierte Parameter der mathematischen Funktion, über deren Anpassung verschiedene Trainingskonfigurationen entstehen. In diesem Beispiel zählt die Anzahl der Hauptkomponenten für die Hauptkomponentenanalyse oder die Höhe des Bestrafungsterms in der L1-Regularisierung zu den Hyperparametern.
- *Eingangsdaten*: Daten, anhand derer eine Vorhersage getroffen wird. Diese nennt man auch *Features* oder *Prädiktoren*. Im Bereich der Radiomics sind hiermit die extrahierten quantitativen Bildfeatures gemeint.
- *Vorhersagen oder Prädiktionen*: Daten, welche durch Anwendung des Modells vorhergesagt werden. Ob diese mit dem wahren Wert (*Ground Truth*), auch *Zieldaten* oder *Labels* genannt, übereinstimmen, beschreibt die Anpassungsgüte (das heißt die Genauigkeit) des Modells.
- *Trainingsdaten*: Datensatz mit bekannten Zieldaten, durch welche das Modell angepasst wird.
- *Validierungsdaten*: Datensatz mit bekannten Zieldaten, mit welchen während des Trainings verschiedene Hyperparameter hinsichtlich ihrer Prädiktionsleistung analysiert werden.
- *Analysedaten oder Testdaten*: Datensatz, auf welchem das trainierte Modell angewendet wird um dessen Prädiktionen mit den bekannten Zieldaten zu vergleichen.

## Methodik

Für die Verarbeitung der Bilddaten mittels Radiomics sind die folgenden Teilschritte notwendig, im Gesamten auch *Radiomics Pipeline* genannt:

1. Bild- und Datenakquise
2. Segmentierung
3. Featureextraktion
4. Featureselektion
5. Klassifikation/Regression

Im Detail:

### 1. Bild- und Datenakquise:

Die Bilddaten können aus jeder bildgebenden Modalität gewonnen werden, d.h. Radiomics ist sowohl für Röntgen-, Ultraschall, als auch für Schnittbildgebungen möglich. In der Regel werden zu Studienzwecken retrospektive Datensätze aus der klinischen Routine genutzt. Die rohen Bilddaten, welche für die Weiterverarbeitung genutzt werden, werden bei der Rekonstruktion des MRT Bildes mit 12 Bit quantisiert, das heißt für jeden Pixel kann einer von 4096 Grauwerten angenommen werden, welcher die Signalintensität kodiert. Durch Effekte wie Magnetfeldinhomogenitäten, Unterschiede in Sequenzen und herstellerbedingte Unterschiede in der Bildakquise können die Rohdaten derselben Gewebeeigenschaften verschiedene Werte annehmen, wodurch - anders als etwa bei der Computertomographie - keine quantitativen (im Sinne von vom Gerät unabhängigen) Werte vorliegen. In den meisten der bisherigen Studien wird davon ausgegangen, dass diese Fehler durch geeignete Vorverarbeitungsschritte auszugleichen sind, welche die Voxelwerte auch unterschiedlicher Akquisitionen vergleichbar machen. Die Akquise der klinischen, genomischen und histopathologischen Daten – je nach Fragestellung – muss vor allem korrekt sein und sollte so vollständig wie möglich erfolgen um eine spätere Integration der Daten zu erleichtern.

### 2. Segmentierung:

In einem zweiten Schritt wird auf den Bilddaten die Begrenzung des auszuwertenden Bildausschnitts (Region bzw. Volume of Interest) definiert. In den meisten Fällen bedeutet dies die Umrandung eines Tumors oder eines Gewebes mit der Möglichkeit, auch Sub-Regionen zu kennzeichnen. Gerade bei Malignomen gestaltet sich diese

Aufgabe als nicht trivial, da die Tumorränder oft unscharf/schlecht abgrenzbar sind. Die Segmentierung kann manuell oder teil- bis komplett automatisiert erfolgen. Die manuelle Segmentierung (d.h. ein Radiologe kennzeichnet auf einem Bild die genaue Begrenzung der Region) ist durch eine hohe Variabilität zwischen verschiedenen Untersuchern sowie einen hohen Zeitaufwand limitiert, wird aber bei den meisten Autoren als sogenannte „ground truth“, d.h. dem am ehesten der Wahrheit entsprechenden Wert, angesehen. Bei einer automatisierten Segmentierung besteht keine Variabilität, jedoch besteht nun die Gefahr, dass die durch den Computer errechnete Begrenzung inkorrekt ist. In der Literatur geht die Präferenz hin zu einem teilautomatisierten Verfahren (die Berandung wird durch einen Segmentierungsalgorithmus vorgeschlagen und durch den Untersucher optimiert) um Segmentierungen mit möglichst akkuraten Begrenzungen mit geringer Variabilität, hoher Reproduzierbarkeit und geringem Zeitaufwand zu ermöglichen (Gillies, Kinahan, and Hricak 2016).

### 3. Featureextraktion:

Durch die Anwendung verschiedener mathematischer Algorithmen auf die Grauwerte der einzelnen Pixel bzw. Voxel der segmentierten ROI/VOI lassen sich quantitative Bildfeatures extrahieren. Diese können sowohl den qualitativen Deskriptoren, mit welchen Radiologen arbeiten, nachempfunden als auch abstrakter oder verborgener Natur sein, das heißt für das menschliche Auge zunächst nicht erkennbar. Diese Features können in Formfeatures und statistische Features erster, zweiter und höherer Ordnung eingeteilt werden - je nach Literatur existieren verschiedene Bezeichnungen hierfür. Eine einfache Erklärung der Features werde ich im Folgenden geben, für mathematische Details sei auf weiterführende Literatur (Zwanenburg A) verwiesen. Die folgenden Beispiele beziehen sich auf die in Abbildung 1 dargestellten 4x4-Pixel-große Areale in 2D mit Grauwerten von 0 bis 3 und den zugehörigen Matrizen A und B.

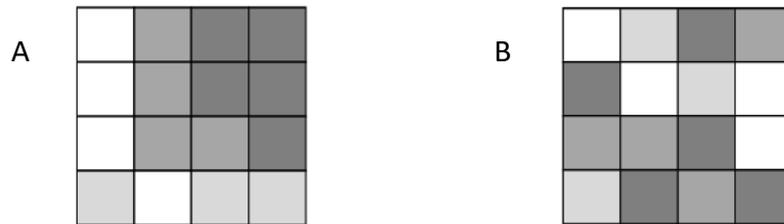


Abbildung 1: 4x4 Pixel Darstellung der Grauwerte 0 bis 3

$$A = \begin{pmatrix} 0 & 2 & 3 & 3 \\ 0 & 2 & 3 & 3 \\ 0 & 2 & 2 & 3 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & 1 & 3 & 2 \\ 3 & 0 & 1 & 0 \\ 2 & 2 & 3 & 0 \\ 1 & 3 & 2 & 3 \end{pmatrix}$$

1. *Formfeatures*: Über die Begrenzung der segmentierten Fläche bzw. des segmentierten Volumens kann dessen Form berechnet werden. Daraus lassen sich einfache Merkmale extrahieren, unter anderem Flächeninhalt bzw. Volumen, größter und kleinster Durchmesser, das Oberflächen-Volumen-Verhältnis oder die Ähnlichkeit mit bestimmten geometrischen Formen (zum Beispiel einer Kugel: "Sphärizität").
2. *Statistische Features erster Ordnung = Intensitätshistogramme*: Diese beschreiben die Verteilung der Grauwerte ohne deren räumliche Anordnung zu berücksichtigen. Dabei werden die Häufigkeiten der einzelnen Grauwerte als Histogramm dargestellt, aus welchem dann Mittelwert, Median, minimale und maximale Intensität sowie Gleichförmigkeit, Entropie, Schiefe und Kurtosis der Verteilung und weiteres berechnet werden können.

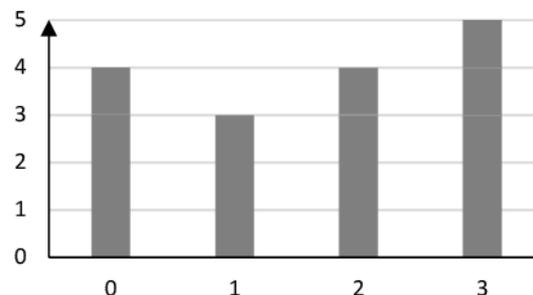


Abbildung 2 Intensitätshistogramm der Pixelanordnungen A und B

3. *Statistische Features zweiter Ordnung = Texturfeatures*: Über verschiedene Abhängigkeitsmatrizen und deren Transformationen können Beziehungen gleicher oder ungleicher Grauwerte zueinander innerhalb der ROI/VOI errechnet werden, indem diese Beziehungen in Matrizen aufgetragen und aus diesen weitere Features berechnet werden. So kann die räumliche Anordnung der

Grauwerte zueinander, das heißt die Textur oder Struktur der Region, beschrieben werden. Beispielsweise lässt sich dadurch ein homogener Tumor von einem heterogenen Tumor mit einer zentralen Nekrosezone oder eine grob- von einer feinkörnigen Struktur unterscheiden. Verwendete Matrizen sind (unter anderem):

- a. Gray Level Co-occurrence Matrix (GLCM): In den Matrizen wird aufgetragen, wie oft sich ein Pixel eines bestimmten Grauwertes in einer definierten Beziehung zu einem anderen Grauwert befindet. Dabei entspricht ein Eintrag der Anzahl, wie oft die Kombination zweier Grauwerte in einem definierten Abstand ( $\delta$ ) innerhalb der ROI/VOI vorkommt. Da dieser Abstand in allen Bildrichtungen (13 Richtungen in 3D, 4 Richtungen in 2D, angegeben durch Winkel  $\theta$ ) berechnet werden muss, entstehen mehrere Matrizen. Mit ihnen können große isointense Flächen von kontrastreichen Strukturen unterschieden werden. Die aus den GLCM extrahierten Features (Haralick Features) beschreiben beispielsweise die Fein- bzw. Grobheit, den Kontrast oder das Auftreten von Mustern in der Bildtextur. Als Beispiel die berechneten GLCM der Matrizen A und B:

$$C_A = \begin{pmatrix} 0 & 2 & 3 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 0 & 1 & 3 \\ 0 & 0 & 3 & 2 \end{pmatrix} \quad C_B = \begin{pmatrix} 0 & 3 & 0 & 2 \\ 3 & 0 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 2 & 2 & 4 & 0 \end{pmatrix}$$

für  $\delta = 1$  und  $\theta = 0^\circ$

- b. Gray Level Run Length Matrix (GLRLM): Hierbei wird in sogenannten "Runs" aufgetragen, wie oft ein bestimmter Grauwert in einer Richtung wiederholt wird, bevor ein anderer folgt, das heißt die Einträge entsprechen der Anzahl konsekutiver gleicher Grauwerte in einer Bildrichtung. Eine feinkörnige Bildtextur enthält also viele kurze Runs, eine grobe hingegen viele lange Runs. Die entsprechenden Features beschreiben zum Beispiel die Variabilität und Homogenität der Längen der Runs und der Grauwerte der Runs, wodurch Rückschlüsse auf die Beschaffenheit (z.B. Grobheit, Heterogenität oder Entropie der Bildtextur) gezogen werden können. Die berechneten GLRM für A und B sind:

$$R_A = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \end{pmatrix} R_B = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{pmatrix}$$

- c. Gray Level Size Zone Matrix (GLSZM): Diese Matrizen beschreiben „gray level zones“, d.h. Areale unmittelbar benachbarter Pixel desselben Grauwertes. Dabei wird aufgetragen, wie oft ein Areal eines bestimmten Grauwertes und einer bestimmten Größe in der ROI/VOI enthalten sind. Da die benachbarten Pixel in alle Richtungen gezählt werden, gibt es nur eine GLSZM für alle Bildrichtungen. Aus ihr werden über dieselben Formeln wie für die GLRLM Features berechnet, welche die Homogenität und Variabilität innerhalb der Grauwertareale widerspiegeln. Als Beispiel die berechneten GLSZM für A und B:

$$S_A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} S_B = \begin{pmatrix} 0 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 0 & 1 \end{pmatrix}$$

- d. Gray Level Dependence Matrix (GLDM): In diesen Matrizen werden Abhängigkeiten verschiedener Pixel dargestellt. Dabei wird aufgetragen, wie oft ein bestimmter Grauwert in einem definierten Abstand ( $\delta$ ) gemeinsam mit von ihm abhängigen Grauwerten in der ROI/VOI vorkommt, wobei mit Abhängigkeit ein bestimmter Grauwertunterschied ( $\alpha$ ) gemeint ist. Die berechneten Features stellen wiederum ein Maß für das Auftreten von Grauwertclustern und die Homogenität der Textur dar. Die GLDM für A und B lauten:

$$D_A = \begin{pmatrix} 0 & 2 & 2 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 3 & 1 \end{pmatrix} D_B = \begin{pmatrix} 0 & 4 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \\ 3 & 2 & 0 \end{pmatrix}$$

für  $\delta = 1$  und  $\alpha = 0$

4. *Statistische Features höherer Ordnung = Modell-/Transformationsbasierte Features:* Mit höhergradigen statistischen Features können über Modelle oder Transformationen bestimmte Anordnungsmuster der Grauwerte erkannt werden, zum Beispiel durch Vergleich der Pixelanordnungen mit Objekten. Hierzu werden vor der weiteren Verarbeitung bestimmte Filter (zum Beispiel Schwellenwerte) auf die Regionen gelegt. Beispiele beinhalten die

Fraktalanalyse, Minkowski Funktionen, Wavelets oder Laplace Transformationen.

Für die Featureextraktion existiert eine Reihe von Open Source Software, welche auf die Datensätze angewandt werden kann. In den jeweils beigefügten Materialien werden die Formeln zur Berechnung der Features angegeben. So bleibt die Featureextraktion transparent und reproduzierbar. Für den Fall, dass proprietäre Software genutzt wird, ist eine Offenlegung der Feature Deskriptoren und Formeln obligat um Reproduzierbarkeit zu gewährleisten.

#### 4. Featureselektion:

Ziel von Radiomics ist es, die große Anzahl der extrahierten Bildfeatures zu einem Wert bzw. einer Funktion zu fusionieren, welche Vorhersagen über eine bestimmte Fragestellung geben kann, zum Beispiel die Wahrscheinlichkeit für die Malignität eines Tumors (Giger 2018). Hierzu werden Modelle entwickelt, welche die für die jeweilige Fragestellung ideale Kombination und Gewichtung der Bildfeatures (auch „radiomic signature“ genannt) erlernen können (Erickson et al. 2017; Avanzo, Stancanello, and El Naqa 2017). Die Anzahl der extrahierten Bildfeatures bewegt sich heute im Bereich von mehreren Hundert bis einigen Tausend, die potentiell in das Modell eingespeist werden. Der Stichprobenumfang bewegt sich mit wenigen hundert Untersuchungen dagegen in deutlich geringeren Dimensionen: Meist werden nur wenige Hundert Untersuchungen durchgeführt. Das heißt die Anzahl der freien Parameter in den statistischen Modellen ist deutlich höher als die Anzahl der Datenpunkte, an denen die Modellparameter bestimmt werden – man sagt die Daten sind hochdimensional. Hieraus entstehen folgende Probleme: Erstens, nicht alle Features stehen in Korrelation zur Fragestellung. Und zweitens, einige der Features beinhalten die gleiche Information bezüglich der Fragestellung, sie sind also redundant (Multikollinearität). Nutzt man diese hochdimensionalen Daten für das statistische Modell, resultiert dies in einem „Overfitting“, das heißt das Modell passt sich zu genau auf die im Training genutzten Daten inklusive der irrelevanten und redundanten an – und kann im Test mit echten Daten nicht bestehen. Deshalb müssen die Dimensionen der Daten reduziert werden, indem irrelevante und redundante Features eliminiert werden und möglichst variante Features selektiert werden, sodass am Ende nur noch wenige (idealerweise deutlich weniger als Samples) Eingangsdaten genutzt werden. Diese Feature Selektion kann als

einzelner Schritt vor dem Training eines Modells erfolgen (zufällig oder gezielt), oder in das Lernmodell integriert sein.

#### 5. Klassifikation/Regression:

Für die Auswahl des geeigneten Lernmodells ist zunächst die Art der Fragestellung wichtig: Für quantitative Ausgabewerte werden Regressionsmodelle genutzt, für kategorische Ausgabewerte (Klassen) benötigt man Klassifikatoren, welche die Eingabedaten der Klasse zuordnen, zu der sie am wahrscheinlichsten gehören. Beispiele für Klassifikatoren sind Logistische Regression, k-nearest Neighbour oder Random Forests. Außerdem kann zwischen überwachten und unüberwachten Lernmodellen unterschieden werden. Bei überwachten Lernmodellen wird bei bekannten Ausgabedaten ein mathematisches Modell entwickelt, welches diese miteinander so korreliert, dass die Prädiktionen für ungesehene Eingabedaten möglichst korrekt getroffen werden können. Für überwachte Lernmodelle werden oft Regressionsmodelle oder Bayes Klassifikatoren genutzt. Im Gegensatz dazu sind die Zieldaten beim unüberwachten Lernen unbekannt mit dem Ziel, mithilfe des Modells Korrelationen innerhalb der Eingabedaten zu finden. So können Gruppierungen (Cluster) von Features identifiziert werden, welche „verwandte“ Features enthalten. Welcher Art diese „Verwandtschaft“ ist, hängt vom gewählten statistischen Modell ab. Im Anschluss lässt sich gegebenenfalls ein Zusammenhang zwischen den Clustern und möglichen Zieldaten untersuchen. Beispiele für die unüberwachte Form des maschinellen Lernens sind Nicht-negative Matrix Faktorisierung und Hauptkomponentenanalyse (Principal Component Analysis = PCA). Nach der Auswahl eines der verschiedenen Lernmodelle, wird dieses trainiert, das heißt seine Parameter auf die Trainingsdaten angepasst und danach mit Testdaten auf seine Anpassungsgüte geprüft. Hierzu können Resamplingmethoden, zum Beispiel das Bootstrapping, angewandt werden, mithilfe derer die Fehlerrate (für Klassifikatoren) bzw. der mittlere Fehler (für Regressionsmodelle) abgeschätzt wird. Sensitivität und Spezifität des Modells werden mittels Receiver Operating Characteristics (ROC) Kurve, deren Fläche unter der Kurve (Area under the curve = AUC) und die jeweiligen Konfidenzintervalle verglichen. Daraus entsteht ein mathematisches Modell, welches einige der Bildfeatures auswählt um über deren Kombination möglichst präzise Prädiktionen über die Zieldaten treffen zu können.

## Zielsetzung

Ziel dieser Arbeit ist die Anwendung von Radiomics Ansätzen in der Mamma-MRT in einem größeren Umfang als in bisher veröffentlichten Studien. Dabei werden mittels zweier verschiedener mathematischer Lernmodelle Klassifikatoren trainiert, welche für kontrastmittelanreichernde intramammäre Läsionen Prädiktionen über deren Malignitätswahrscheinlichkeit treffen können. Dazu werden zunächst Läsionen auf den Mamma-MRT-Untersuchungen identifiziert und segmentiert, um die extrahierten Bildfeatures im Anschluss mit den erhobenen klinischen Daten – womit insbesondere die Art der Läsion gemeint ist – zu korrelieren. Zur Klassifikation werden dabei zwei etablierte Machine Learning Algorithmen – die L1-Regularisierung sowie die Hauptkomponentenanalyse – genutzt. Die Genauigkeit der Prädiktionen wird letztendlich mittels doppelter Kreuzvalidierung und dem Bootstrapping als Resamplingmethode evaluiert. Um die Abhängigkeit der Prädiktionsleistung von der Läsions- und der Datensatzgröße zu untersuchen, wird dasselbe Verfahren auf Teilmengen der Gesamtdaten angewendet und die Ergebnisse verglichen. Letztendlich soll die Frage beantwortet werden, ob mittels klassischer Machine Learning Algorithmen eine „radiomic signature“ für intramammäre Läsionen in der Mamma-MRT gefunden werden, welche die Malignität dieser mit ausreichender Genauigkeit beschreibt.

## Material und Methoden

### Studienpopulation

Die Studienpopulation bestand aus MR-tomografisch anreichernden Läsionen, welche eindeutig als benigne oder maligne klassifiziert werden konnten. Hierzu wurden aus insgesamt 5687 diagnostischen Mamma MRTs, welche zwischen August 2011 und August 2015 am Universitätsklinikum Aachen durchgeführt und prospektiv befundet wurden, 1000 Untersuchungen zufällig ausgewählt und deren Bilder sowie Befunde überprüft. Das Fehlen einer anreichernden Läsion sowie zu große Bewegungsartefakte bzw. bewegungsabhängige Verschiebungen zwischen den einzelnen Sequenzen galten dabei als erste Ausschlusskriterien. Für die übrigen Untersuchungen wurden folgende anamnestische Daten erhoben: MR-Indikation, Alter und Menopausestatus, ggf. Zykluswoche, hormonelle oder antihormonelle Medikation, Risikoeinteilung anhand von Familien- und Eigenanamnese sowie relevante vorherige Therapien (beispielsweise Brusterhaltende Resektion und Bestrahlung nach Mammakarzinom). Außerdem wurde die MR-tomografische Hintergrundanreicherung sowie, falls bekannt, die mammografische Brustdichte erhoben.

Eingeschlossen wurden primär alle anreichernden Läsionen, die eindeutig als benigne oder maligne klassifiziert werden konnten: Die Validierung erfolgte anhand von histopathologischen Untersuchungen von Biopsie- (vor oder nach der MRT durchgeführt) und Operationspräparaten oder im Falle eines BI-RADS® 2 oder 3 Befundes anhand einer Follow-Up Untersuchung im MRT nach mindestens 24 Monaten. Alle Läsionen, für welche keine ausreichende Validierung per Histologie oder Follow-Up vorlag oder welche sich radiologisch nicht eindeutig dichotom kategorisieren ließen wurden aus der Studie ausgeschlossen. Histopathologisch gesicherte Risikoläsionen mit einem hohen Entartungsrisiko (= Borderline-Läsionen), zu welchen die high grade lobuläre intraepitheliale Neoplasie (LIN high grade oder CLIS), die flache epitheliale Atypie (FEA), die atypisch duktale Hyperplasie (ADH), radiäre Narben, benigne Phylloidtumoren sowie Papillome mit Atypien zählen, wurden ebenfalls ausgeschlossen. Zum einen handelt sich hierbei um eine heterogene Gruppe von Läsionen mit variablem Entartungsrisiko, zum anderen können bioptisch gesicherte Borderline-Läsionen fokale maligne Anteile enthalten (dies gilt insbesondere für die ADH und das CLIS), wodurch

keine eindeutige Klassifikation der Läsion in ihrer Validierung möglich ist (Rageth et al. 2016). Bei Patientinnen mit mehreren Läsionen desselben Typs (beispielsweise bei multifokalem Mammakarzinom oder multiplen Fibroadenomen), wurde jeweils eine dieser Läsionen zufällig selektiert und in die weitere Analyse eingeschlossen, um eine Verfälschung der Ergebnisse durch Messwiederholung zu vermeiden.

## Durchführung und Befundung der Mamma MRT an der Uniklinik Aachen

Die kontrastmittelunterstützten MR-Untersuchungen der Brust wurden in der oben genannten Zeitspanne anhand eines hausinternen Standardprotokolls durchgeführt, bestehend aus:

- axiale dynamische Gradientenecho-Serie ohne Fettunterdrückung (TR 250 ms, TE 4,6 ms, Flip Angle 90°, Akquisitionsmatrix 512x3x512, SENSE Faktor 2) mit Aufnahmen vor und 4-malig nach der Gabe von 0,1 mmol Gadobutrol pro Kilogramm Körpergewicht mit Berechnung der jeweiligen Subtraktionsbilder und Rekonstruktion von Maximumintensitätsprojektionen.
- axiale T2-gewichtete Turbospinecho-Sequenz

Dazu wurde ein 1.5-Tesla-System (Achieva; Philips Medical Systems, Best, Niederlande) mit einer bilateralen 4-Kanal-Brustspule (Invivo, Orlando, Florida, USA) genutzt, mit Fixierung der Brust in kraniokaudaler Richtung.

Die prospektive Befundung der Untersuchungen erfolgte durch verschiedene Fachärzte des Instituts für diagnostische und interventionelle Radiologie der Uniklinik Aachen mit Erfahrung in der Mamma-MR-Diagnostik von 7 bis 25 Jahren nach dem hausinternen Standard in Anlehnung an den MR-BI-RADS®-Katalog. Hierbei wurden neben den MR-BI-RADS®-Werten pro Brustseite zusätzliche Werte für jede auffällige Läsion vergeben, um eine multimodale sowie eine mehrzeitige Diagnostik zu vereinfachen. Für eine binäre Klassifikation durch den Radiologen wurden alle Läsionen der BI-RADS®-Kategorien 2 und 3 als benigne und alle Läsionen der BI-RADS®-Kategorien 4, 5 und 6 als maligne gewertet.

## Erhebung der Zieldaten

Je nach MR-BIRADS®-Kategorie wurden entweder Histopathologie oder Follow-Up als Zieldaten erhoben:

- BI-RADS® 2: Follow-Up durch mindestens eine befundunabhängige Mamma MRT später als 24 Monate nach der Indexuntersuchung.
- BI-RADS® 3: Follow-Up durch befundabhängige Kontrolluntersuchung; bei stationärem oder regredientem Befund weitere Mamma MRT später als 24 Monate nach der Indexuntersuchung, bei Befundprogredienz nachfolgende histopathologische Sicherung per Biopsie oder Probeentnahme
- BI-RADS® 4, 5: Histopathologische Diagnosesicherung nach der Indexuntersuchung per Biopsie, Probeentnahme oder operativer Entfernung
- BI-RADS® 6: Biopsische Diagnosesicherung bereits vor der Indexuntersuchung durchgeführt, Läsion noch eindeutig abgrenzbar

Histopathologische Beurteilung:

- B2-Läsion = benigne Läsion: Fibroadenom, Adenoseherd(e), Papillom, Lymphknoten, Mastopathie, fibrotisches oder entzündlich verändertes Brustparenchym
- B3-Läsion = Borderline-Läsion: CLIS, FEA, ADH, Papillom mit Atypien, benigner Phylloidtumor, radiäre Narbe
- B5a-Läsion = nichtinvasives Mammakarzinom: DCIS
  - Grading (low, intermediate, high grade)
  - Östrogenrezeptorexpression (Remmele-Score 0-12)
- B5b-Läsion = invasives Mammakarzinom:
  - Tumortyp (no-special type, lobulär, medullär, sonstige)
  - Tumorgrading (1-3)
  - Östrogen- und Progesteronrezeptorexpression (Remmele-Score 0-12)
  - Her-2-Rezeptorexpression (Score 0-3)
  - Proliferationsfraktion Ki-67 (0-100%)

## Radiomics

Segmentierung mit MeVisLab 3.0

Die ausgewählten Läsionen wurden stets durch denselben Untersucher per Hand segmentiert, wobei eine erste Segmentierung durch eine Doktorkandidatin (Hannah Schneider) erfolgte und diese im Anschluss von einer radiologischen Fachärztin mit mehr als 15 Jahren Mamma-MR-Erfahrung (PD Dr. med. Simone Schradung) korrigiert und verfeinert wurde. Hierzu wurde eine hauseigene Softwareerweiterung des



Abbildung 3 Screenshot der Segmentierung mit MeVisLab 3.0

Bildverarbeitungsprogramm MeVisLab 3.0 (MeVis Medical Solutions AG, Fraunhofer MEVIS, Bremen, Deutschland) genutzt, mit welchem das freihändige Einzeichnen und Verändern von Segmentierungsmasken ermöglicht wurde. So konnten die Läsionen zunächst grob umrandet und dieser Umriss sukzessive bis zum gewünschten Ergebnis verfeinert werden. Diese manuelle Segmentierung erfolgte schichtweise zumeist auf den Subtraktionsbildern, da hier eine gute Identifikation der Berandung einer Läsion möglich ist. Für den Fall, dass bewegungsabhängige Subtraktionsfehler vorlagen, wurde eine nicht-subtrahierte Post-Kontrastmittel-Aufnahme mit guter Darstellung des Läsionsumrisses gewählt. Aus den Segmentierungstapeln wurde im Anschluss das jeweilige 3-dimensionale Tumolvolumen erstellt und auf die übrigen Sequenzen ortskodiert übertragen. Vor der weiteren Verarbeitung der annotierten Bilder wurden die Signalintensitäten mittels Bias Field Correction N4ITK normiert und umskaliert.

## Featureextraktion

Es wurden mittels einer Software von Pyradiomics 16 Formfeatures des 3-dimensionalen Volumens sowie jeweils 19 Intensitäts- und 49 Texturfeatures (27 GLCM- und je 16 GLRM- und GLSZ-Features) aus der T2-gewichteten Sequenz, den Post-Kontrast-Aufnahmen und dem Subtraktionsbild extrahiert. Insgesamt ergaben sich somit 133 respektive 413 Intensitäts- und Texturfeatures. Die genauen Definitionen der Bildfeatures kann unter folgendem Link eingesehen werden: <http://pyradiomics.readthedocs.io/en/latest/features.html> (Stand 30.03.2019).

## Featureselektion

Zur Selektion von Features mit hoher Relevanz und geringer Redundanz wurden zwei statistische Modelle ausgewählt und ausgeführt:

### L1-Regularisierung

In diesem, auch *Lasso Regularisierung* genannten, überwachten Lernmodell zur Featureselektion wird die Regressionsanalyse (in der Regel per Methode der kleinsten Quadrate) um einen *Bestrafungsterm* erweitert, welcher die Regressionskoeffizienten regularisiert, indem deren Summe nach oben beschränkt wird. Die Stärke dieses Bestrafungsterms wird durch einen sogenannten Hyperparameter ( $\lambda$ ) gewichtet. Wird  $\lambda$  groß gewählt, führt dies zu hohen Bestrafungen falls viele Merkmale ausgewählt werden, wodurch möglichst wenige Merkmale in das Modell eingehen und dieses vereinfacht wird. Durch Kreuzvalidierung der L1-Regularisierung mit jeweils verschiedenen Schranken kann deren geeignete Höhe evaluiert werden. Gleichzeitig wird hierdurch erreicht, dass sowohl Features ausgeschlossen werden, für welche keine Korrelation mit den Zieldaten besteht als auch jene, welche sich kollinear zu anderen Features verhalten. Das heißt durch die L1-Regularisierung ergibt sich letztlich ein Regressionsmodell, welches idealerweise nur Features von Relevanz und ohne Redundanz erhält. Dieses lässt sich auch auf Klassifikationsprobleme übertragen.

### Hauptkomponentenanalyse

Die Hauptkomponentenanalyse ist ein unüberwachtes Lernmodell zur Dimensionsreduktion von hochdimensionalen Datensätzen durch die Zusammenfassung korrelierter Features zu sogenannten Hauptkomponenten. Dabei

wird die erste Hauptkomponente (eine Linearkombinationen von Features) so gewählt, dass sie die stärkste Varianz des Datensatzes enthält, das heißt innerhalb dieser Hauptkomponente befinden sich die Features mit der höchsten Streuung. Die zweite (und alle weiteren) Hauptkomponenten berechnen sich analog - allerdings werden nur Linearkombinationen betrachtet, welche unkorreliert zur ersten (bzw. zur vorherigen) Hauptkomponente sind; die Hauptvektoren stehen also orthogonal zueinander. Durch die weitere Iteration erhält man eine definierte Anzahl an unkorrelierten Hauptkomponenten (also Linearkombinationen aus Features), wobei diese Anzahl einen Hyperparameter des Modells darstellt.

Da beim unüberwachten Lernen die Features nicht in Korrelation mit den Zieldaten gesetzt werden, ergeben sich aus der Hauptkomponentenanalyse zwar unkorrelierte Hauptkomponenten, über deren Relevanz bezüglich einer möglichen Fragestellung allerdings keine Aussage getroffen werden kann. Sind die Zieldaten bekannt, kann durch ein überwachtes Modell eine Vorselektion relevanter Features erfolgen oder dieses auf die per Hauptkomponentenanalyse ermittelten Hauptkomponenten angewandt werden. In diesem Fall wurden mittels univariater p-Wert-Analyse zunächst 100 Features ausgewählt, das heißt für jedes Feature wurde berechnet, wie sehr dieses mit den Zieldaten korreliert (ohne Abhängigkeiten von anderen Features zu berücksichtigen). Die 100 Features mit den kleinsten P-Werten wurden dann der Hauptkomponentenanalyse zugeführt, in welcher die 10 ersten Hauptkomponenten berechnet wurden.

## Validierung

Zur Anpassung und Berechnung der Fehlerrate des jeweiligen Klassifikators wurde dieser doppelt 5- bzw. 10-fach kreuzvalidiert. Für eine einfach  $k$ -fache Kreuzvalidierung wird der gesamte Datensatz zufällig in  $k$  gleich große Teilmengen geteilt, wobei jeweils eine dieser Teilmengen als Validierungs- und die anderen  $k-1$  Teilmengen als Testdaten fungieren. Daraus ergeben sich  $k$  Kombinationen aus Trainings- und Testdaten, über welche das Modell  $k$ -fach angepasst und validiert wird - häufig sind 5- oder 10-fache Kreuzvalidierungen. Über die Mittelwerte der jeweiligen Fehlerquoten, Sensitivitäten und Spezifitäten kann die Anpassungsgüte des Modells abgeschätzt werden. Bei der doppelten Kreuzvalidierung werden die Daten in einer inneren und einer äußeren Schleife kreuzvalidiert, wobei die Gesamtdaten in Trainings-, Validierungs- und

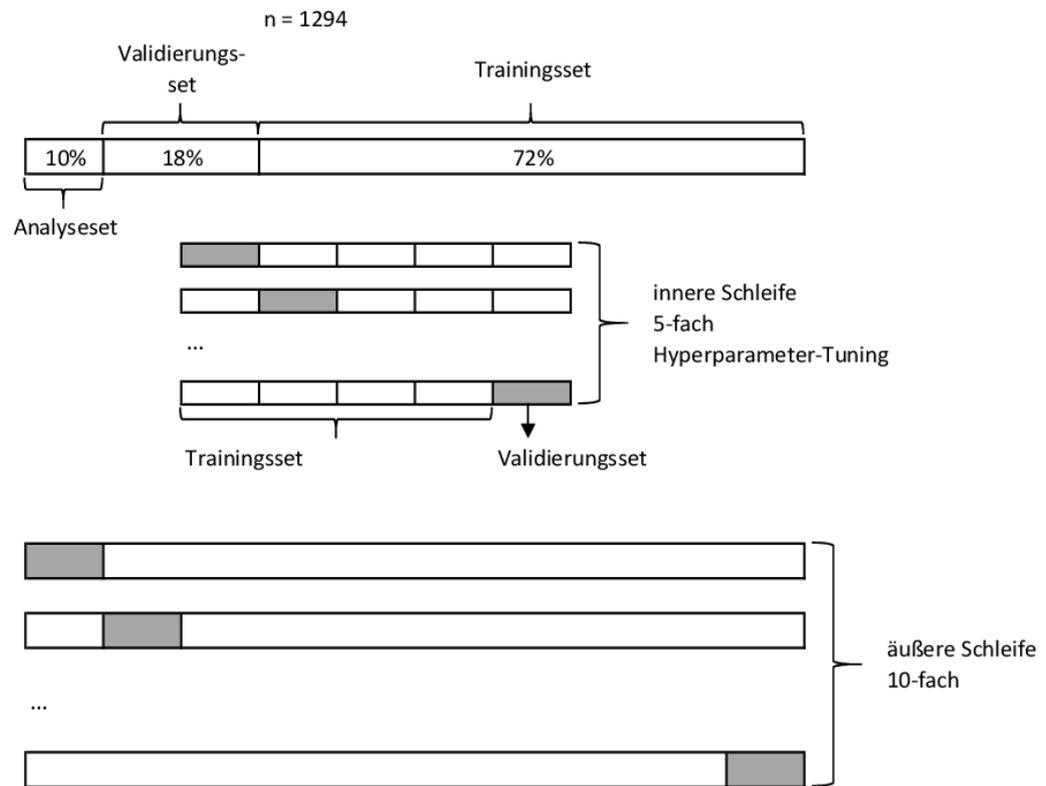


Abbildung 4: Doppelte Kreuzvalidierung mit 10-facher äußerer und 5-facher innerer Schleife

Testmengen aufgeteilt werden. In der inneren Schleife wird zunächst das Modell zunächst unter Ausschluss der Analysedaten mit verschiedenen Konfigurationen kreuzvalidiert. Dabei werden verschiedene Hyperparameter getestet, um deren beste Kombination auszuwählen und in die äußere Schleife einzupflegen. In dieser wird das trainierte, in der inneren Schleife am besten angepasste Modell dann auf den Analysedaten angewandt, sodass die Prädiktionsgenauigkeit des Modells errechnet werden können. In dieser Studie wurde die innere Schleife 5-fach und die äußere Schleife 10-fach kreuzvalidiert (Aufteilung 72%/18%/10% in Trainings-, Validierungs-, und Analysedatensätze), wodurch für jede Läsion eine Malignitätswahrscheinlichkeit berechnet und mit dem wahren Wert verglichen werden konnte.

### Statistische Auswertung

Zum Vergleich der Modelle in Radiomics wurden Receiver Operator Characteristic (ROC) Kurven erstellt, wobei die Datenpunkte hierbei verschiedene Cut-Off-Werte darstellen, ab welcher vom Algorithmus berechneten Malignitätswahrscheinlichkeit eine Läsion auch als maligne gewertet wurde. Als optimaler Cut-Off wurde dann der Wert definiert,

welcher sowohl eine hohe Sensitivität als auch eine hohe Spezifität besitzt, also den geringsten Abstand zur linken oberen Ecke der ROC-Kurve aufweist ( $\min m = (1 - \text{Sensitivität})^2 + (1 - \text{Spezifität})^2$ ). Um Standardabweichung und Konfidenzintervalle für die Fläche unter der Kurve (AUC), Sensitivität und Spezifität berechnen zu können, wurde mittels Bootstrapping-Verfahren ein 100.000-faches Resampling durchgeführt. Hierbei werden zufällige Stichproben aus der Gesamtmenge der Daten generiert, indem eine bestimmte Anzahl von Datenpunkten zufällig gezogen wird, der einzelne Datenpunkt nach seiner Ziehung allerdings weiterhin in der Gesamtmenge vorhanden ist (Ziehen mit Zurücklegen). So entstehen verschiedene zufällig verteilte Stichproben, welche den gleichen Datenpunkt auch mehrfach enthalten können. Für die jeweiligen Teilmengen wird nun die statistische Analyse durchgeführt, sodass sich eine Werteverteilung für AUC, Sensitivität und Spezifität ergibt, für welche sich Konfidenzintervalle berechnen lassen. Für den Vergleich der beiden Modelle in Radiomics untereinander (L1-Regularisierung vs. Hauptkomponentenanalyse) und zur Diagnostik durch den Radiologen (L1-Regularisierung vs. Radiologe und Hauptkomponentenanalyse vs. Radiologe) wurde das Signifikanzniveau im Sinne der Bonferroni-Korrektur für dreifache paarweise Vergleiche auf  $p = 0,05/3 = 0,017$  festgesetzt (Goeman and Solari 2014).

Um eine Abhängigkeit der Ergebnisse von der Stichprobengröße zu evaluieren, wurde die gesamte Radiomics Analyse zusätzlich auf einem zufällig gewählten Datensatz mit der halben Anzahl an Läsionen ( $n = 647$ ) durchgeführt. Eine weitere Analyse wurde für Läsionen mit einem maximalen Durchmesser  $< 2\text{cm}$  durchgeführt, zum einen, da für die Läsionsgröße ein positiver Vorhersagewert für die Malignität bekannt ist mit einem signifikanten Anstieg der Malignitätswahrscheinlichkeit für Läsionen  $> 12\text{mm}$  (Gutierrez et al. 2009; Kawai et al. 2018; Liberman et al. 2006), zum anderen um zu überprüfen, ob sich auch für kleine Läsionen Unterschiede in den Texturfeatures ergeben, welche von der Matrix- und somit von der Läsionsgröße abhängig sind.

## Ergebnisse

### Studienkohorte

Tabelle 3: Erhobene Daten für n=447 Patientinnen

Alter	64 [26-82]
Indikation	
- Früherkennung	152 (34%)
- Suspekter Befund	140 (31%)
- Staging	122 (27%)
- Nachsorge	33 (7%)
MR-ACR	
- 1	257 (58%)
- 2	141 (32%)
- 3	45 (10%)
- 4	3 (<1%)

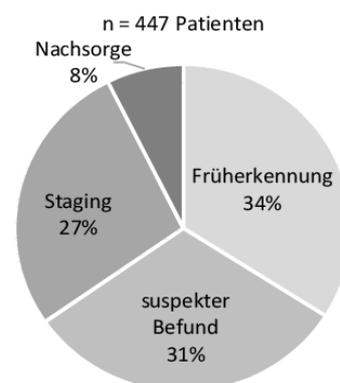


Abbildung 5: Indikationen für die Mamma MRT für n=447 Patientinnen

Insgesamt konnten 1294 anreichernde Läsionen aus Mamma MR Untersuchungen von 447 Patientinnen segmentiert und in die Untersuchung eingeschlossen werden. Die Patientinnen waren zum Zeitpunkt der Untersuchung im Durchschnitt 64 Jahre alt (Range 26-82 Jahre, Median 66 Jahre), 337 von ihnen (75,5%) befanden sich in der Postmenopause, 110 Patientinnen (24,5%) waren prämenopausal. Indikationen für die Durchführung der Mamma MRT waren: Früherkennung bei normalem oder erhöhtem familiären Risiko (n=152, 34%), Diagnostik bei suspektem Befund in der konventionellen Bildgebung (140 Patientinnen, 31%), Staging bei gesichertem Mammakarzinom (n=122, 27%), Nachsorge nach Mammakarzinom oder Borderline-Läsion (n=33, 7%). Die Evaluation der Hintergrundanreicherungen nach dem BI-RADS®-Katalog (MR-ACR) ergab eine fehlende Hintergrundanreicherung (MR-ACR 1) für 257 Patientinnen (58%), eine geringe Hintergrundanreicherung (MR-ACR 2) für 141 Patientinnen (32%), eine mittelstarke Hintergrundanreicherung (MR-ACR 3) für 45 Patientinnen (10%) und eine starke Hintergrundanreicherung (MR-ACR 4) für 3 Patientinnen (0,7%) (Vgl. Tabelle 3). Für 307 der 447 Patientinnen konnten vollständige anamnestische Angaben erhoben werden, welche Tabelle 4 entnommen werden können. Die Anzahl der pro Patientin

segmentierten Läsionen betragen: 221 Patientinnen mit einer Läsion, 66 Patientinnen mit 2 Läsionen, 53 Patientinnen mit 3 Läsionen, 39 Patientinnen mit 4 Läsionen und 68 Patientinnen mit 5 oder mehr Läsionen.

Table 4: Erhobene anamnestische Daten für n=307 Patientinnen

Anamnese vollständig	307 (100%)	Medikation	
		- Hormone	40 (13%)
		- Antihormone	9 (3%)
Familiäres Risiko		Eigenanamnese	
- keines	193 (63%)	- Brustgesund	269 (88%)
- gering erhöht	58 (19%)	- Z.n. B3 Läsion	10 (3%)
- mittelstark erhöht	41 (13%)	- Z.n. DCIS	11 (4%)
- stark erhöht	14 (5%)	- Z.n. Mamma-Ca	16 (5%)
- BRCA Mutation	1 (<1%)	- Z.n. Ovarial-Ca	1 (<1%)

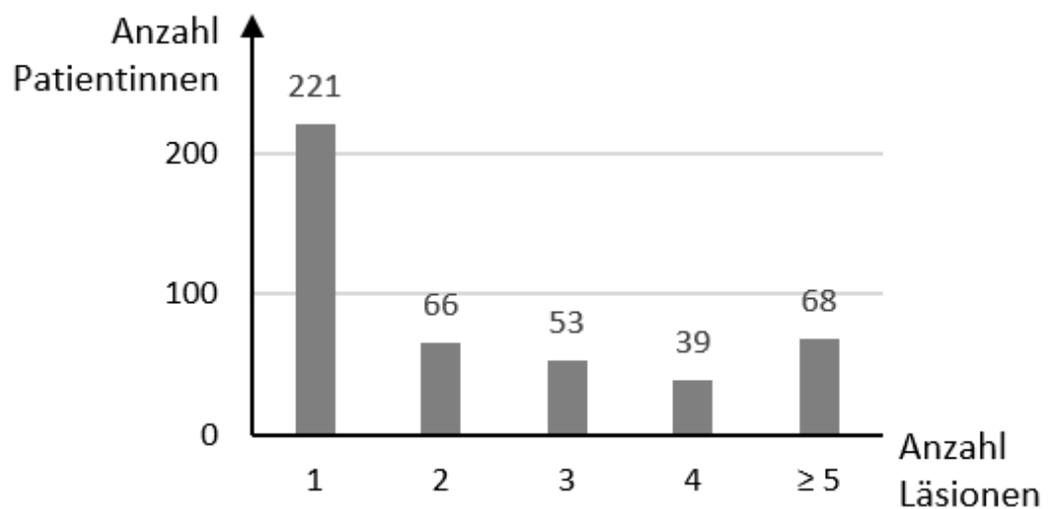


Abbildung 6: Verteilung der Läsionsanzahl pro Patientin

Tabelle 5: Validierung der n=1294 Läsionen

Benigne	507 (39%)	Maligne	787 (61%)
- Adenose	156 (31%)	- Karzinom	534 (68%)
- Fibroadenom	153 (30%)	- DCIS	253 (32%)
- Lymphknoten	63 (12%)		
- Papillom	20 (4%)		
- Weitere	115 (23%)		

507 der 1294 Läsionen (39,2%) wurden durch Histopathologie bzw. Follow-Up als benigne und 787 Läsionen (60,8%) als maligne identifiziert. Die benignen Läsionen ließen sich bildmorphologisch als Adenoseherde (n=156, 30,8%), Fibroadenome (n=153, 30,2%), intramammäre Lymphknoten (n=63, 12,4%), Papillome (n=20, 3,9%), sowie seltenere Entitäten wie Parenchyminseln, hormonell stimuliertes Parenchym, Mastopathie, randständig anreichernde Zysten, Narbengewebe, Serome sowie entzündlich bedingte Mehranreicherungen, welche zusammenfassend mit „weitere“ bezeichnet wurden (n=115, 22,7%), klassifizieren. Von den Malignomen waren 534 Läsionen (67,9%) invasive und 253 Läsionen (32,1%) nichtinvasive Karzinome (DCIS). Bei 704 Läsionen (55,4%) handelte es sich um solide Tumoren (mass lesions), wovon 67,9% Malignome waren (478 Läsionen). Von den 590 nicht-tumorösen Anreicherungen (non-mass enhancements, 45,6%) waren 61,7% maligne. Für 394 der 534 histologisch gesicherten invasiven Karzinome konnten vollständige histopathologische Daten erhoben werden, welche sich Tabelle 6 entnehmen lassen.

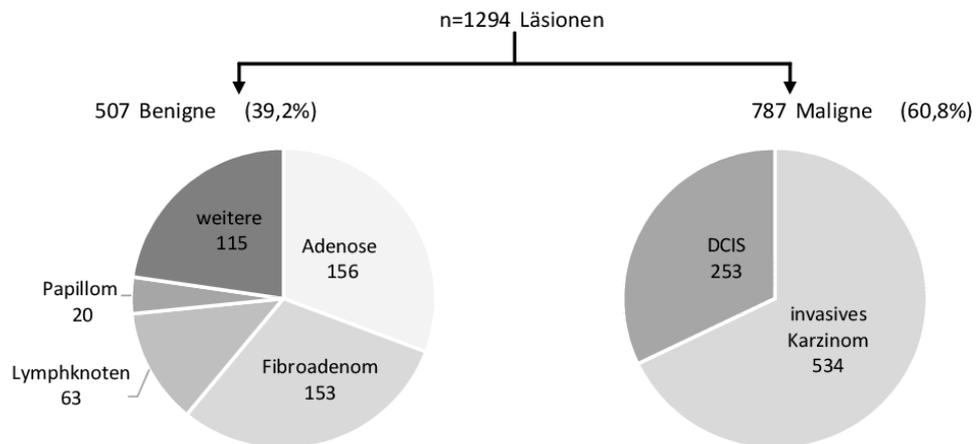


Abbildung 7: Läsionsarten der malignen und benignen Läsionen

Tabelle 6: Erhobene Histopathologien für n=394 Karzinome

Histologie vollständig		394 (100%)			
Differenzierung	- No Special Type	250 (63%)	Grading	- G1	68 (17%)
	- Lobulär	85 (22%)		- G2	263 (67%)
	- Gemischter Typ	26 (7%)		- G3	63 (16%)
	- Tubulär	13 (3%)	Rezeptorstatus	- Luminal	319 (81%)
	- Muzinös	8 (2%)		- Triple negativ	39 (10%)
	- Intrazystisch	3 (1%)		- Her2 positiv	36 (9%)
	- Sonstige	9 (2%)			

Im maximalen Durchmesser waren die Läsionen durchschnittlich 22,4 (+/- 10,1) mm groß, wobei sich Malignome mit 27,0 (+/- 11,3) mm größer präsentierten als benigne Läsionen mit 15,3 (+/- 6,6) mm.

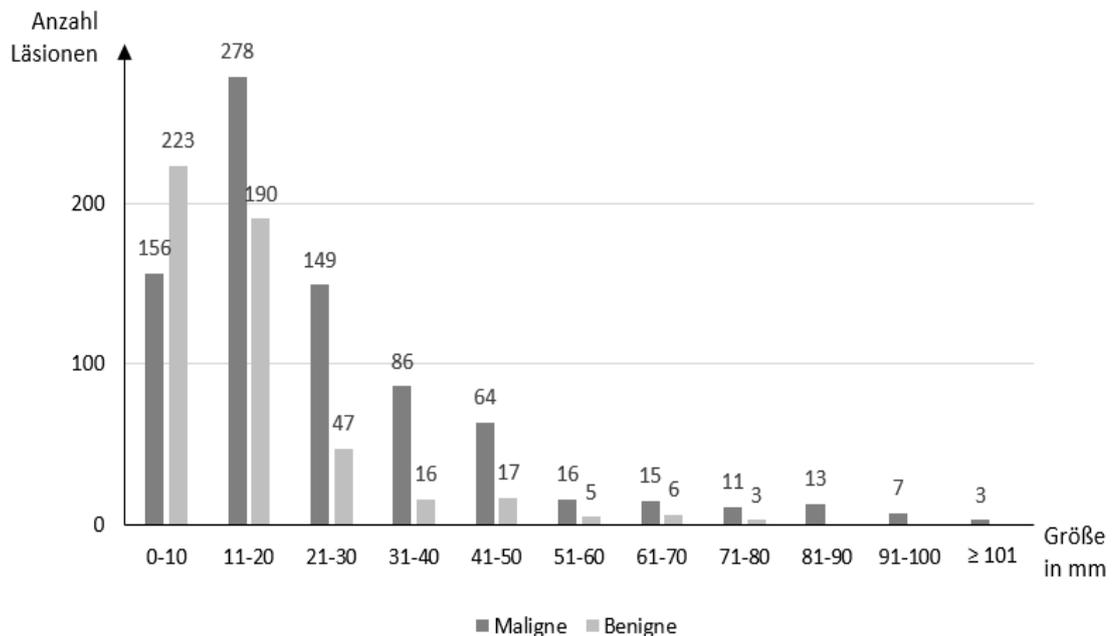


Abbildung 8: Absolute Größenverteilung nach Dignität

# L1-Regularisierung und Hauptkomponentenanalyse

## Selektierte Features der L1-Regularisierung

Durch Featureselektion mittels L1-Regularisierung konnten 47 relevante Bildfeatures identifiziert werden. Aus diesen wurden ein Formfeature, 6 Features des T2-gewichteten Bildes und 39 Features aus der dynamischen Serie selektiert. 23 dieser 47 Bildfeatures wurden durch die L1-Regularisierung innerhalb der 10-fachen Kreuzvalidierung jedes Mal ausgewählt und sind deshalb als besonders relevante Features zu werten. Diese sind:

- Formfeatures:
  - Sphärizität (Ähnlichkeit der Läsion mit einer Kugel)
- Statistische Features:
  - T2: 10. Perzentile (entspricht den Intensitäten der Areale der Läsion mit der deutlichsten Hypointensität)
  - T1: Interquartilsabstand (Intensitätsunterschied innerhalb der Läsion)

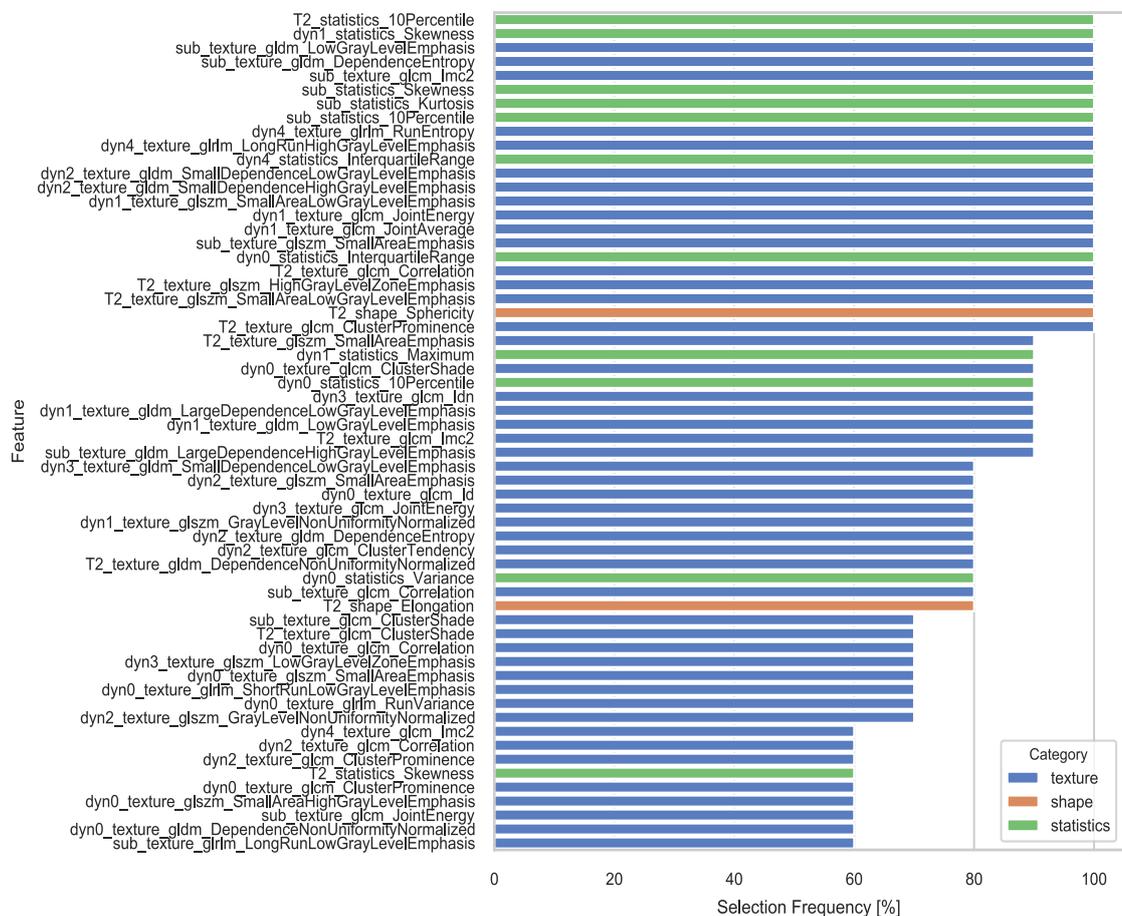


Abbildung 9: Selektierte Features der L1-Regularisierung nach ihrer Selektionshäufigkeit

- 1. Postkontrastaufnahme: Schiefe (starke Hyper- oder Hypointensität der Läsion)
- 4. Postkontrastaufnahme: Interquartilsabstand
- Subtraktionsbild: 10. Perzentile, Schiefe, Wölbung (Maß für die Streuung der Grauwerte)
- Texturfeatures
  - T2:
    - GLCM: Cluster Prominence (Maß für die Clusterbildung der Grauwerte), Correlation (Homogenität der Textur)
    - GLSZM: High Gray Level Zone Emphasis, Small Area Low Gray Level Emphasis (Maß für große hyperintense bzw. kleine hypointense Flächen)
  - 1. Postkontrastaufnahme:
    - GLCM: Joint Average (durchschnittliche Intensität), Joint Energy (Maß für die Homogenität der Textur)
    - GLSZM: Small Area Low Gray Level Emphasis
  - 2. Postkontrastaufnahme:
    - GLDM: Small Dependence High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis (Maße für den Kontrast benachbarter Grauwerte)
  - 4. Postkontrastaufnahme:
    - GLRLM: Long Run High Gray Level Emphasis (Flächenverteilung von hyperintensiven Grauwerten), Run Entropy (Maß für Heterogenität der Textur)
  - Subtraktionsbild:
    - GLCM: Informational Measure of Correlation 2 (Maß für die Komplexität der Textur)
    - GLDM: Dependence Entropy, Low Gray Level Emphasis (Maße für die Heterogenität bzw. Feinheit der Textur)
    - GLSZM: Small Area Emphasis (Maß für die Feinheit der Textur)

Weitere, häufig selektierte Features sind Texturfeatures aus sowohl den nativen Aufnahmen, den dynamischen Serien und den Subtraktionsbildern, sowie das Formfeature "Elongation", welches die Länglichkeit des VOI beschreibt und der

maximale Grauwert der ersten Postkontrastaufnahme. Eine detaillierte Auflistung der Features mit ihren Selektionsfrequenzen ist im Anhang aufgeführt.

## Selektierte Features der Hauptkomponentenanalyse

Die 17 durch die Hauptkomponentenanalyse stets selektierten Features sind:

- Formfeatures: Sphärizität
- Statistische Features: keine
- Texturfeatures:
  - T2:
    - GLSZM: Zone Entropy (Maß für die Heterogenität der Textur)
  - T1:
    - GLDM: Dependence Entropy (Maß für die Heterogenität der Textur)
  - 1. Postkontrastaufnahme:
    - GLCM: Joint Average (durchschnittliche Intensität)
    - GLDM: Dependence Entropy, Large Dependence High Gray Level Emphasis (Maß für hyperintense Areale)
    - GLSZM: Zone Entropy
  - 2. Postkontrastaufnahme:
    - GLCM: Informational Measure of Correlation 1 und 2 (Maße für die Komplexität der Textur)
    - GLSZM: Zone Entropy
  - 3. Und 4. Postkontrastaufnahme:
    - GLCM: Informational Measure of Correlation 1 und 2
  - Subtraktionsbild:
    - GLCM: Joint Entropy (Maß für die Heterogenität der Textur)
    - GLDM: Dependence Entropy
    - GLSZM: Zone Entropy



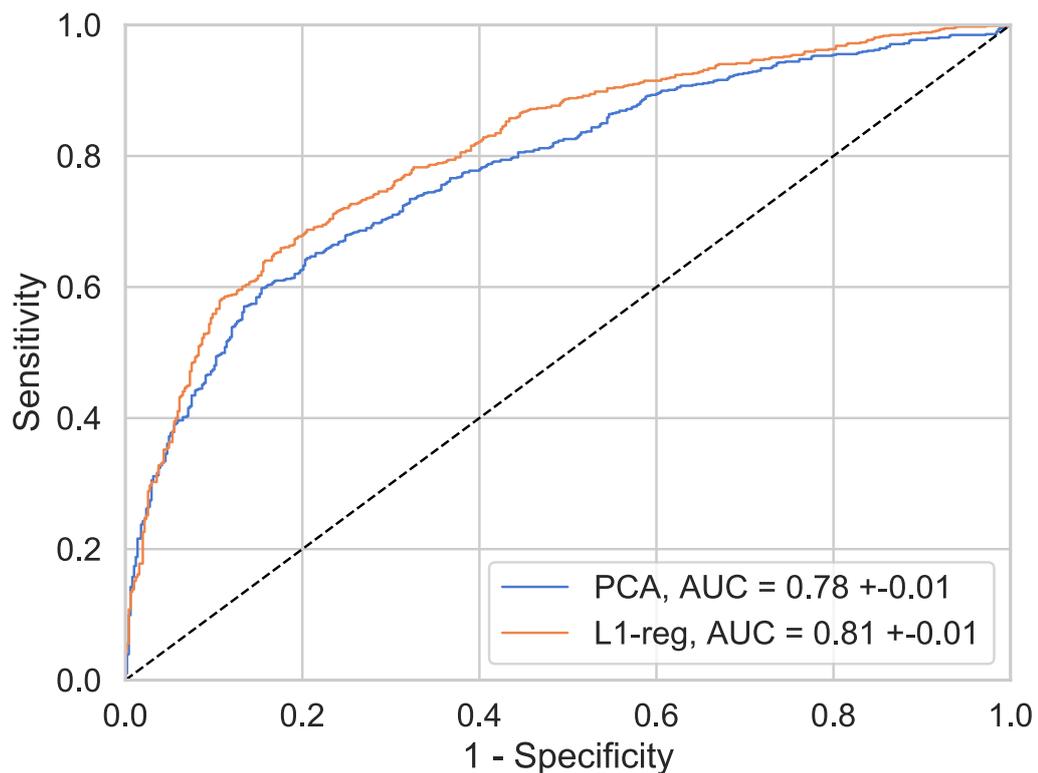


Abbildung 11: Receiver Operator Characteristics für L1-Regularisierung und Hauptkomponentenanalyse

Tabelle 3). Die Auswertung durch den Radiologen ergab hingegen eine AUC von 0,98 [0,97-0,99] mit einer Sensitivität von 99,7% [99,4-100%] und einer Spezifität von 86,4% [84,5-88,3%] und schnitt damit deutlich besser ab als beide Radiomics Ansätze ( $p < 0,001$ ). Die Auswertung der Modelle mit dem halben Datensatz von halber Größe ( $n=647$ ) ergab hinsichtlich der AUC keinen signifikanten Unterschied ( $p=0,05$  für L1-Regularisierung,  $p=0,06$  für Hauptkomponentenanalyse). Unter Ausschluss aller Läsionen mit einem Durchmesser von 2cm und größer (823 Läsionen) ergab die Analyse für die L1-Regularisierung eine AUC von 0,74 [0,71-0,77], für die Hauptkomponentenanalyse eine AUC von 0,70 [0,67-0,74] und für die BI-RADS®-Einteilung eine AUC von 0,98 [0,96-0,99] ergab.

Tabelle 7: Statistische Auswertung von L1-Regularisierung, Hauptkomponentenanalyse und Auswertung durch den radiologischen Experten, 95%-Konfidenzintervalle in eckigen Klammern

	<b>L1-Regularisierung</b>	<b>Hauptkomponentenanalyse</b>	<b>Radiologe</b>
<b>Sensitivität in %</b>	71,5 [69,0-74,0]	67,9 [65,4-70,4]	99,7 [99,4-100]
<b>Spezifität in %</b>	76,1 [73,8-78,4]	75,1 [72,7-77,5]	86,4 [84,5-88,3]
<b>AUC</b>	0,81 [0,79-0,84]	0,78 [0,75-0,80]	0,98 [0,97-0,99]
<b>AUC halber Datensatz</b>	0,80 [0,77-0,84]	0,78 [0,75-0,80]	0,96 [0,95-0,98]
<b>AUC Läsionen &lt;2cm</b>	0,74 [0,71-0,77]	0,70 [0,67-0,74]	0,98 [0,96-0,99]

## Diskussion

Mit dieser Arbeit konnte gezeigt werden, dass mittels Machine Learning Algorithmen eine Kombination quantitativer Bildmerkmale (“radiomic signature”) gefunden werden kann, welche in Korrelation zur Malignitätswahrscheinlichkeit einer Läsion steht. Dabei lag die Trefferwahrscheinlichkeit der Computeralgorithmen mit einer AUC von 0,81 bzw. 0,78, Sensitivitäten von 72% bzw. 68% und Spezifitäten von 76% bzw. 75% im klinisch akzeptablen Bereich. Zum Vergleich: Die Sensitivität und Spezifität der Mamma MRT bei routinemäßiger Befundung durch den Radiologen liegt je nach Studie bei 70 bis 100% respektive 60 bis 90% (Kriege et al. 2004; Leach et al. 2005; Lord et al. 2007; Weinstein et al. 2009). Verglichen mit der Befundung nach BI-RADS® durch die Radiologen in unserem Hause schnitten die Modelle jedoch deutlich schlechter ab. Hierfür birgt die überdurchschnittliche Expertise innerhalb der Abteilung, welche sich in den stets sehr hohen Sensitivitäten (93 bis 100 %) und Spezifitäten (94 bis 98%) der Mamma MRT in Publikationen aus der Uniklinik Aachen widerspiegelt, einen Erklärungsansatz (Kuhl et al. 2010; Kuhl et al. 2014). Das intuitive Argument, dass die Anzahl der Untersuchungen, mit welcher die Radiomics-Algorithmen trainiert wurde die Anzahl der durch den erfahrenen Radiologen zuvor betrachteten Mamma MRTs deutlich unterschreitet und somit vom Algorithmus gar keine vergleichbare „Erfahrung“ erwartet werden kann, lässt sich durch die zweite Auswertung mit halber Untersuchungszahl widerlegen: Da durch eine Halbierung der Fallzahl keine signifikante Verschlechterung der Klassifikationsleistung durch die Algorithmen erzielt wurde, ist auch keine Verbesserung dieser bei Erhöhung der Zahl an Trainingsdaten für den Algorithmus zu erwarten. Verglichen mit den bisherigen Publikationen, welche über Featureextraktion und –selektion eine Klassifikation hinsichtlich der Dignität von Läsionen in der Mamma-MR untersuchten, zeigte sich durch diese Studie keine Verbesserung der Klassifikationsleistung, obwohl innerhalb dieser Arbeit eine deutlich größere Anzahl sowohl an Untersuchungen als auch an extrahierten Features genutzt wurde. Beispielsweise konnten Bhooshan et al. durch Training eines Bayesian Neural Networks 48 Features aus T2-gewichteten und dynamischen Serien so kombinieren, dass eine AUC von 0,85 erreicht werden konnte, allerdings lag die Fallzahl mit insgesamt 196 Läsionen deutlich unter der Fallzahl dieser Arbeit (Bhooshan et al. 2011). Antropova et al. trainierten eine Support Vector Machine mit 38 extrahierten Features aus den

dynamischen Serien von 690 Läsionen und erreichten eine AUC von 0,86 (Antropova, Huynh, and Giger 2017). In beiden genannten Studien wurden kinetische Features einbezogen, welche die Kontrastmitteldynamik, das heißt den Intensitätsunterschied zwischen verschiedenen Post-Kontrast-Aufnahmen, beschreiben. Die extrahierten Texturfeatures beschränkten sich in beiden Studien auf maximal 14, aus den GLCM berechneten Features (Haralick Features). Dass in dieser Arbeit ohne die Integration kinetischer Features eine schlechtere Klassifikationsleistung erzielt wurde, könnte darauf schließen lassen, dass diese auch für Computeralgorithmen maßgeblich zur Differenzierung der Dignität einer Läsion beitragen, analog zur radiologischen Befundung, welche die Enhancement-Kurven der Läsionen einbezieht (Kuhl, Mielcareck, et al. 1999). Dass die Kontrastmittelkinetik für die Klassifikation durch den Radiologen letztlich jedoch nicht benötigt wird, wurde in den letzten Jahren durch die Etablierung eines verkürzten MRI-Protokolls gezeigt, bei welchem nur noch native T1-Sequenzen und eine Post-KM-Sequenz und die entsprechenden Rekonstruktionen (Subtraktion und MIP) zur Befundung genutzt werden ohne Einbußen bei der diagnostischen Genauigkeit gegenüber den Standard-Protokollen (Chhor and Mercado 2017). Es ließe sich auch argumentieren, dass eine Erweiterung der extrahierten Features aufgrund zu vieler Redundanzen die schlechtere Klassifikationsleistung bedingt. Allerdings muss berücksichtigt werden, dass sich die vorangegangenen Arbeiten auf solide Brustläsionen beschränkten, in dieser Arbeit jedoch sowohl tumoröse als auch nicht-tumoröse Anreicherungen eingeschlossen wurden – für letztere ist sowohl in der radiologischen Befundung als auch unter Einsatz von CAD-Systemen eine schlechtere Prädiktionsleistung der Dignität einer Läsion bekannt (Newell et al. 2010; Gutierrez et al. 2009). In einer Arbeit von Gallego-Ortiz und Martel konnte gezeigt werden, dass durch eine Unterscheidung tumoröser von nicht-tumorösen Läsionen eine bessere Klassifikationsleistung durch CAD-Systeme erbracht werden kann, als ohne eine vorherige Zuordnung der Läsionen zur einen oder anderen Läsionsart (Gallego-Ortiz and Martel 2016). Dies untermauert, dass im direkten Vergleich der Ergebnisse dieser Arbeit mit Arbeiten, welche nur solide tumoröse Läsionen einschließen, schlechtere Prädiktionsleistungen zu erwarten sind.

Da sich durch Reduzierung des Datensatzes um die Hälfte keine Verschlechterung der Ergebnisse zeigte, kann darauf geschlossen werden, dass die genutzten Modelle ab einer gewissen Anzahl von Daten gesättigt sind - sie haben sozusagen "ausgelernt" und

können sich durch weitere Erfahrung (das heißt größere Datensätze) nicht verbessern. Die möglichen Ursachen hierfür sind methodisch bedingt, wobei insbesondere die vordefinierten, handgemachten Features eine Limitation darstellen. Durch Anwendung komplexerer Machine Learning Methoden wie dem Deep Learning kann diese Sättigung der Prädiktionsleistung behoben werden, worauf später noch genauer eingegangen wird.

Betrachtet man die Größenverteilung der Läsionen fällt auf, dass mit dem Maximaldurchmesser einer Läsion auch die Wahrscheinlichkeit für eine Malignität steigt, analog zu den Ergebnissen von Kawai et al. (Kawai et al. 2018). In der Analyse kleiner Läsionen (Maximaldurchmesser <2cm) wurden mittels der Radiomics-Algorithmen schlechtere Klassifikationsleistungen erzielt als unter Einschluss aller Läsionen (AUC 0,81 bzw. 0,78 für alle Läsionen gegenüber AUC 0,74 bzw. 0,70 für Läsionen <2cm). Einen Erklärungsansatz hierfür bietet Abhängigkeit der meisten extrahierten Texturfeatures von der Matrixgröße und somit von der Läsionsgröße: je kleiner die Läsion, desto kleiner die Unterschiede in den berechneten Features verschiedener Texturen – und desto schlechter die Klassifikationsleistung. Des Weiteren lässt sich postulieren, dass sich größere und demnach fortgeschrittene Malignome deutlicher von benignen Läsionen unterscheiden als kleinere Karzinome, welche sich entweder in ihren anfänglichen Stadien befinden oder geringere Wachstumsraten aufweisen. Letztlich bleibt offen, ob sich für aggressivere Karzinome oder verschiedene Tumorentitäten Unterschiede in den radiomic signatures ergeben, da innerhalb dieser Arbeit keine Analyse mit verschiedenen Tumorgradings, Differenzierungen oder molekularen Subtypen durchgeführt wurde.

## Selektierte Features

Betrachtet man die von den Algorithmen selektierten Features genauer, fällt eine Überschneidung mit den im BI-RADS®-Katalog genannten und von Radiologen häufig genutzten qualitativen Bildmerkmalen zur Differenzierung maligner von benignen Läsionen auf. Insbesondere die mit beiden Modellen selektierten Features Sphärizität und GLCM Joint Average der ersten Postkontrastaufnahme spiegeln wichtige Kriterien der radiologischen Befundung wider: Das Formfeature Sphärizität drückt die Form und Begrenzung einer Läsion aus, wobei spikulierte, irreguläre Läsionen (niedrige Sphärizität) für Malignome sprechen und glatt und scharf begrenzte, ovaläre Läsionen

(hohe Sphärizität) am ehesten benigne sind. Mit dem Feature GLCM Joint Average der ersten Postkontrastaufnahme wird die durchschnittliche Intensität der Läsion beschrieben, wobei bei hohen Werten auf ein frühes "Aufleuchten" der Läsion, das heißt ein für Malignome typisches schnelles Anfluten des Kontrastmittels rückgeschlossen werden kann (Kuhl, Mielcareck, et al. 1999; Szabo et al. 2003).

Mittels der L1-Regularisierung wurden mit einer Anzahl von 47 relativ viele Features ausgewählt, welche auch trotz des genutzten Bestrafungsterms einige Redundanzen untereinander aufwiesen. Dabei wurden sowohl statistische als auch Texturfeatures selektiert. Die selektierten Features aus den T2-Sequenzen beschreiben dabei vor allem die Intensität (10. Perzentile) und ihre räumliche Verteilung innerhalb der Läsion (GLCM Cluster Prominence, GLSZM Features), was sich auch in der qualitativen Befundung durch höhere Malignitätswahrscheinlichkeiten bei T2-hypointensen, heterogenen Läsionen findet (Kuhl, Klaschik, et al. 1999; Malich et al. 2005). In der dynamischen Serie wurden vor allem solche Features ausgewählt, welche sowohl die starke frühe Kontrastmittelanreicherung (Schiefe, GLCM Joint Average der 1. Post-KM-Sequenz) als auch die läsionsinterne Heterogenität dieser beschreiben (GLCM Joint Energy der 1. Post-KM-Sequenz), analog zur Betrachtung durch den Radiologen, welcher in seiner Befundung insbesondere das Kontrastmittelanreicherungsverhalten des Tumors als Differenzierungsmerkmal zwischen malignen und benignen Läsionen nutzt. Aus dem statistischen Feature des Interquartilsabstands (aus der letzten Postkontrastaufnahme) kann ein Rim Enhancement mit einem Kontrast zwischen äußerer und innerer Zone der Läsion abgeleitet werden. Im Subtraktionsbild wurden neben Stärke und Verteilung der Intensitäten (Schiefe, 10. Perzentile, Kurtosis) als Maß für die Anreicherungsstärke ebenfalls Texturfeatures selektiert, welche die Heterogenität und Feinheit der Textur beschreiben (GLCM IMC2, GLDM Dependence Entropy, GLSZM Small Area Emphasis). Insgesamt kann gesagt werden, dass die selektierten Features der L1-Regularisierung die vom Radiologen betrachteten Tumormerkmale Form, frühes Anfluten des Kontrastmittels und inhomogene Kontrastmittelverteilung imitieren, was deren Klassifikationsvermögen unterstreicht. Die weitere Kontrastmitteldynamik im Sinne eines Auswaschens des Tumors, welches als typisch für Malignome gilt, wurde vom Modell hingegen nicht selektiert. Dies lässt sich zum einen damit erklären, dass die Kinetik mit den in dieser Arbeit extrahierten Features gar nicht erfasst wurde, da ein Auswaschen sich in den späten ohne Vergleich zu den frühen Post-KM-Sequenzen nicht

von einer von Beginn an mäßigen Kontrastmittelaufnahme unterscheidet. Zum anderen kann argumentiert werden, dass ein Auswaschen des Tumors in der Klassifikation der Läsionen von geringerer Bedeutung ist als das frühe, schnelle und inhomogene Enhancement – analog zu den Ergebnissen aus Studien mit verkürzten MR-Protokollen (Kuhl 2018; Chhor and Mercado 2017).

Im Gegensatz zur L1-Regularisierung wurden mit der Hauptkomponentenanalyse insgesamt zwar weniger Features selektiert, doch auch hier bestehen Redundanzen zwischen diesen. Beispielsweise wurden jeweils in der 2. bis 4. Postkontrastaufnahme die Features GLCM Informational Measure of Correlation 1 und 2 ausgewählt, welche beide die Komplexität der Textur beschreiben und untereinander hohe Korrelationen aufweisen. Interessanterweise wurde mit der Hauptkomponentenanalyse während der 10-fachen Kreuzvalidierung insgesamt nur 4 statistische Features – und diese maximal 5-mal innerhalb der 10-fachen Schleife – ausgewählt, und zwar zum einen die Hypointensität in der nativen T2-Sequenz (Minimum T2), zum anderen Features welche jeweils die Stärke und Intensitätsverteilung der frühen Kontrastmittelanreicherung widerspiegeln (Maximum und Range der 1. Post-KM-Sequenz bzw. Range der 2. Post-KM-Sequenz). Bei allen weiteren selektierten Features handelt es sich um Texturfeatures, welche in den nativen Aufnahmen und den Subtraktionsbildern die Heterogenität der Textur (GLSZM Zone Entropy der T2-Sequenz, GLDM Dependence Entropy der nativen T1-Sequenz, GLCM Joint Entropy, GLDM Dependence Entropy, GLSZM Zone Entropy der Subtraktionsbilder) und in der dynamischen Serie zusätzlich die frühe Kontrastmittelanreicherung sowie die gesamte Komplexität der Textur beschreiben (GLCM Joint Average, GLDM Dependence Entropy der 1. Post-KM-Sequenz; GLCM IMC 1 und 2 der 2., 3. und 4. Post-KM-Sequenz). Daraus lässt sich schließen, dass für die Hauptkomponentenanalyse neben der bereits diskutierten Form und frühen Kontrastmittelanreicherung vor allem das interne Signalverhalten der Läsion sowohl vor als auch nach Kontrastmittelapplikation relevant sind - analog zur Betrachtung der dynamischen Serie durch den Radiologen.

Als zukünftig mögliche Anwendung des Klassifikationsalgorithmus steht die Implementierung in ein computer-assistiertes Diagnostiksystem und die Anwendung im klinischen Alltag, wodurch eine Erhöhung der Sensitivität und vor allem der Spezifität der Mamma MRT auch für weniger erfahrene Befunder möglich wäre. Hierdurch ließe sich letztlich eine erhöhte Verfügbarkeit der Mamma MRT und deren breitflächigere

Anwendung, möglicherweise auch als Screening-Instrument für bestimmte Patientenkollektive, ermöglichen. Ebenfalls wäre durch den Einsatz des verkürzten MR-Protokolls, dessen diagnostische Wertigkeit mit dieser Arbeit unterstrichen wird, eine erhöhte Verfügbarkeit und eine deutliche Kostenreduktion der Mamma MRT möglich. Letztlich könnte eine erhöhte Detektionsrate vor allem mammografisch okkulten Mammakarzinome mit konsekutiver Senkung der brustkrebsspezifischen Mortalität möglich werden.

## Limitationen

Durch die Auswahl der Studienkohorte ergeben sich für die Fragestellung dieser Arbeit mehrere Limitationen: Erstens, bei der Auswahl der eingeschlossenen Läsionen wurde nicht zwischen tumorösen und nicht-tumorösen anreichernden Läsionen unterschieden. Grundlage hierfür ist die Annahme, dass es eine Signatur von Bildfeatures gibt, welche unabhängig von der Läsionsart ist. In der klinischen Praxis wird hingegen regelhaft zwischen den beiden Formen der Kontrastmittelanreicherung unterschieden (Morris EA 2013). In den bisherigen Studien zu Radiomics in der Mamma MRT wurde sich daher zumeist auf solide Tumoren beschränkt. Für computerassistierte Diagnostiksysteme konnte gezeigt werden, dass sie für solide Tumoren eine bessere Klassifikationsleistung aufweisen als für nicht-tumoröse Läsionen (Newell et al. 2010). Auch für den menschlichen Betrachter stellt die Einordnung nicht-tumoröser Anreicherungen die anspruchsvollere Aufgabe dar als die Detektion und Klassifizierung solider Tumoren, woraus eine niedrigere Sensitivität für die Detektion von duktalem Carcinoma in situ resultiert (Kuhl 2007a). Da in dieser Arbeit nicht zwischen den Läsionsarten unterschieden wurde, bleibt offen, ob durch eine Trennung eine bessere Leistung der Modelle möglich wäre oder ob die selektierten Features voneinander abweichen. Als zweiten Punkt ist der Einschluss schon biopsierter Läsionen zu nennen. Durch die Probenentnahme wird ein Gewebedefekt innerhalb der Läsion sowie eine Reaktion des umliegenden Gewebes verursacht, welches in den in der Regel kurz darauf durchgeführten MRT-Untersuchungen abgebildet wird ist. Da die MR Diagnostik in vielen Fällen bei schon bioptisch gesicherten malignen Tumoren durchgeführt wurde, sind in dieser Arbeit einige der Läsionen mit MR-tomografisch sichtbaren Stanzdefekten eingeschlossen wurden. Somit könnten Features existieren und selektiert worden sein,

welche diese Defekte widerspiegeln und als Zeichen für Malignität gewertet werden und somit die Ergebnisse verfälschen.

Die in dieser Arbeit verwendeten MR Aufnahmen der Brust wurden alle am gleichen Tomografen nach demselben Untersuchungsprotokoll akquiriert. Da durch verschiedene Geräte und verschiedene MR-Sequenzen Grauwertvariationen entstehen, werden auch die extrahierten Features, insbesondere die Texturfeatures, beeinflusst (Collewet, Strzelecki, and Mariette 2004). In einer Studie von Saha et al. konnte eine Abhängigkeit der berechneten Features durch verschiedene Aufnahmeparameter bzw. zwischen verschiedenen Gerätehersteller speziell für Mamma-MR-Tomografien gezeigt werden (Saha et al. 2017). Auch für CT-Untersuchungen konnte eine Instabilität der Features bei variablen Akquisitionsparametern gezeigt werden (Berenguer et al. 2018). Somit können die in dieser Arbeit trainierten Modelle zunächst nur auf hausinterne, nach dem gleichen Protokoll akquirierte Untersuchungen übertragen werden, da bei einer Anwendung auf externen Bildern ohne erneutes Training eine schlechtere Vorhersageleistung zu erwarten ist. Ein robustes, von der Aufnahmetechnik unabhängiges Modell, würde entweder eine universelle Normalisierung der Grauwerte (zum Beispiel anhand eines MR-Modells mit definierten Eigenschaften) voraussetzen oder könnte durch Training auf einem großen Datensatz aus verschiedenen Institutionen mit Geräten verschiedener Hersteller, Aufnahmeparametern und -protokollen erreicht werden (Collewet, Strzelecki, and Mariette 2004).

Durch die manuelle Segmentierung der Läsionen ergeben sich zweierlei Einschränkungen: zum einen ist der hohe zeitliche Aufwand mit einer durchschnittlichen Bearbeitungszeit von 20 Minuten pro Untersuchung - allein für die Segmentierung - im klinischen Alltag nicht angemessen, insbesondere im Hinblick auf zukünftige Studien größeren Datenumfangs. Zum anderen wird die Reproduzierbarkeit der Daten deutlich eingeschränkt. Auch wenn durch die manuelle Segmentierung durch einen Experten eine vermeintliche "ground truth" erreicht werden kann, wird in der Literatur trotzdem eine zumindest semiautomatische bis gänzlich automatisierte Segmentierung vorgezogen um die Studie vergleichbar zu machen, da für manuelle Segmentierungen eine hohe Variabilität zwischen unterschiedlichen Betrachtern bekannt ist (Menze et al. 2015). Bei Anwendung eines Modells aus handsegmentierten Daten auf einem neuen Datensatz, welcher von einem zweiten Untersucher segmentiert wurde, ist also eine Verfälschung der Ergebnisse bzw. eine schlechtere Vorhersageleistung des Modells zu

erwarten. Ein weiterer Punkt ist, dass eine Läsionsdetektion wie bei den computerassistierten Diagnostiksystemen, mit der manuellen Segmentierung nicht erfolgt. Der Algorithmus lernt zwar, definierte Regionen (ROIs) zu klassifizieren, nicht jedoch, diese innerhalb des Datensatzes zu erkennen. Somit wäre die Anwendung des Modells im klinischen Alltag zunächst von geringerem Nutzen, da die Läsionen zunächst vom Betrachter erkannt und markiert werden müssen. Eine zukünftige Erweiterung der Algorithmen auf die Detektion suspekter Läsionen aus den Rohdaten ist für eine Anwendung als routinemäßig eingesetzte Diagnostikunterstützung also obligat.

Durch die Auswertung vordefinierter, von Hand programmierter Bildfeatures wird dem selbstständigen Lernen des Computers Grenzen gesetzt, die durch das menschliche Verständnis der Bildverarbeitung entstehen. Wie in der Arbeit gezeigt, wurden durch diese Vorgehensweise hauptsächlich diejenigen Features selektiert, welche auch bei menschlicher Betrachtung ausschlaggebend für eine Klassifizierung der Läsionen nach ihrer Malignität sind. Möchte man die qualitative radiologische Befundung computerbasiert imitieren, so müssten in diesem Zusammenhang auch 4-dimensionale Features extrahiert werden, um die Anreicherungskinetik der Bildpunkte quantitativ zu erfassen. Hierfür können spatiotemporale Features, das heißt Features, welche die pixelweise Grauwertänderung und deren räumliche Anordnung berücksichtigen, berechnet werden. Für diese konnten in mehreren Studien zumindest univariate Korrelationen zur Malignitätswahrscheinlichkeit eines Brusttumors in MR-Tomografien und ein besseres Klassifikationsvermögen gegenüber Texturfeatures gezeigt werden (Milenkovic et al. 2013; Zheng et al. 2009; Agner et al. 2011). Durch den Einsatz vorprogrammierter, definierter Features wird jedoch ausgeschlossen, dass weitere, (noch) undefinierte Bildmerkmale existieren, welche von lernenden Algorithmen erkannt werden könnten. Um dies zu überwinden können Deep Learning Modelle genutzt werden, auf welche im nächsten Kapitel genauer eingegangen wird.

## Ausblick

Einige der oben genannten Limitationen können durch Anwendung von Deep bzw. Transfer Learning Methoden mithilfe von Convolutional Neural Networks (CNN) behoben werden. Hierbei handelt es sich um Softwarearchitektur, die den neuronalen Verknüpfungen im menschlichen Gehirn nachempfunden ist. Dabei sind verschiedene Datenpunkte in mehreren Ebenen (Layers) über Gewichtungen miteinander verbunden,

wobei durch das Training des Netzes die Gewichtungen angepasst werden. Da das Netzwerk selbstständig erkennt, welche Anteile und Charakteristika innerhalb einer Bildregion mit der Prädiktion korreliert sind, ist weder eine vorherige Definition von Bildfeatures noch eine Segmentation der Tumorgrenzen nötig - eine grobe Angabe über die Region of Interest zum Beispiel über eine Markierung des Tumormittelpunkts ist ausreichend. Damit das Netz ausreichend angepasst werden kann, sind jedoch wesentlich größere Datenmengen notwendig, als für herkömmliche Machine Learning Ansätze. Um die benötigte Datenmenge zu reduzieren, können auf Standardbildern (zum Beispiel von Landschaften) vortrainierte CNN genutzt werden, welche anschließend auf den eigentlichen Daten weiter trainiert werden ("Fine-Tuning"), wofür der Begriff Transfer Learning genutzt wird. Für eine genauere Erklärung der Funktionsweise neuronaler Netze sei auf weiterführende Literatur verwiesen (Topol 2019; LeCun, Bengio, and Hinton 2015; Litjens et al. 2017).

Dass mittels CNN durchaus bessere Klassifikationsergebnisse erzielt werden können als mit klassischen Machine Learning Ansätzen, konnte für verschiedene Modalitäten und Fragestellungen gezeigt werden. In der Arbeit von Antropova et al. konnten bessere Klassifikationsleistungen bezüglich der Malignität von Brustläsionen durch „deep learned“ - d.h. durch CNN extrahierte Features - sowohl in der Mamma-MRT (AUC 0,87) als auch in der Mammografie und der Mammasonografie gezeigt werden (Antropova, Huynh, and Giger 2017). Als Weiterführung dieser Doktorarbeit untersuchte die Arbeitsgruppe unter Truhn et al. die Prädiktionsleistung eines auf demselben Datensatz trainierten neuronalen Netzes mit signifikant besseren Ergebnissen als den beiden klassischen Radiomics Modellen (Truhn et al. 2018). Für die besten Klassifikationsergebnisse (AUC 0,88, Sensitivität 78%, Spezifität 85%) wurden durch das CNN dabei die Bilder der dynamischen Serie (native Aufnahme, erste und dritte Postkontrastaufnahme) genutzt. Außerdem konnte gezeigt werden, dass - im Unterschied zur L1-Regularisierung und Hauptkomponentenanalyse - durch einen größeren Datensatz die Leistung des Netzes verbessert werden kann. Zukünftig werden sich die CNN-basierten Klassifikatoren im klinischen Alltag unter verschiedenen Aufnahmebedingungen prospektiv erproben müssen, um ihre Leistung weiter zu verbessern und eine breite Anwendung als unterstützendes Diagnostikum ermöglichen zu können.

Darüber hinaus gilt als Zielsetzung der Radiomics die Korrelation mit weiteren prognostischen und tumorbiologischen Parametern. In der Mamma MRT wurden hinsichtlich des Brustkrebses bisher unter anderem Zusammenhänge zwischen Bildparametern und rezidivfreiem Überleben, Lymphknotenstatus oder molekularen Subtypen untersucht, mit teils widersprüchlichen Ergebnissen, wobei in den meisten der bisherigen Studien lediglich kleine Datensätze und wenige Bildfeatures genutzt wurden (Guo et al. 2015; Burnside et al. 2016). Beispielsweise konnten Fan et al. mittels 88 handgemachter Features eine Klassifikation anhand der vier verschiedenen molekularen Subtypen nachweisen, Grimm et al. nutzte 56 Features und kam zu dem Ergebnis, dass sich mit diesen lediglich Luminal A und Luminal B Subtypen voneinander unterscheiden lassen (Fan et al. 2017; Grimm, Zhang, and Mazurowski 2015). Eine von Dong et al. publizierte Vorhersage des Lymphknotenstatus durch Features aus diffusionsgewichteten Sequenzen oder die Vorhersage des Ki-67-Wertes anhand eines Radiomic-Scores der dynamischen Serie könnten bei Bestätigung der Ergebnisse und weiterführender Forschung in der Entwicklung eines in-vivo Biomarkers resultieren (Ma et al. 2018; Dong et al. 2017; Liang et al. 2018). Auch die bildmorphologische Korrelation mit molekulargenetischen Parametern, zum Beispiel dem etablierten OnkotypeDX Score, wird mittels Radiomics Analysen derzeit untersucht (Nam et al. 2019) Für onkologische Fragestellungen ist auch eine Darstellung der phänotypischen Tumorheterogenität (und deren Korrelation zu genotypischen bzw. molekularen Unterschieden) mittels Radiomics zukünftig interessant, da hierdurch präzise Biopsien und Therapieplanungen ermöglicht werden könnten (Gatenby, Grove, and Gillies 2013). Bis zur Anwendung im klinischen Alltag sind allerdings viele weitere, multizentrisch Untersuchungen mit größeren Datenmengen vonnöten.

## Zusammenfassung

Mit dieser Arbeit wurden zwei klassische Machine Learning Verfahren – die L1-Regularisierung und die Hauptkomponentenanalyse – auf ihre Fähigkeit überprüft, manuell segmentierte kontrastmittelanreichernde Läsionen aus Routine-MRT der Brust anhand ihrer quantitativen Bildmerkmale nach ihrer Dignität zu klassifizieren. Dazu wurden die Läsionen manuell segmentiert und mit dem per Histologie oder 2-jährigem Follow-Up validiertem Zielwert in Korrelation gesetzt. Zur Klassifikation wurden dabei die oben genannten Algorithmen zur Featureselektion und Klassifikation doppelt kreuzvalidiert und per Resampling auf ihre Genauigkeit hin überprüft. Es lässt sich schlussfolgern, dass mittels der ausgewählten Methoden zwar klinisch akzeptable Klassifikationsleistungen erzielt werden können, der radiologische Experte erzielte jedoch deutlich bessere Ergebnisse. Eine Auswertung auf dem halben Datensatz erbrachte keine schlechteren Ergebnisse, sodass geschlussfolgert werden kann, dass ein Training auf einer größeren Datenmenge keine Verbesserung im Sinne eines lernenden Systems ergibt. Eine Betrachtung der selektierten Features ergab die Erkenntnis, dass insbesondere Form und intratumorale Heterogenität der Kontrastmittelaufnahme, d.h. der Textur nach Kontrastmittelapplikation, wegweisend für die Unterscheidung benigner von malignen Läsionen sind. Eine klinische Anwendung der Radiomics Algorithmen ist zum einen die Einbettung in die Routine-Befundung als unterstützendes Diagnostikum. Zum anderen ist zukünftig eine Nutzung von bildgebenden Parametern als komplementärer Biomarker in der Präzisionsmedizin denkbar. Hierzu sind jedoch einige Weiterentwicklungen nötig, wobei insbesondere die Anwendung von Deep bzw. Transfer Learning mittels neuronaler Netze vielversprechend sind.

## Literatur

- Aerts, H. J., E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin. 2014. 'Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach', *Nat Commun*, 5: 4006.
- Agner, S. C., S. Soman, E. Libfeld, M. McDonald, K. Thomas, S. Englander, M. A. Rosen, D. Chin, J. Noshier, and A. Madabhushi. 2011. 'Textural kinetics: a novel dynamic contrast-enhanced (DCE)-MRI feature for breast lesion classification', *J Digit Imaging*, 24: 446-63.
- Anderson, W. F., P. S. Rosenberg, A. Prat, C. M. Perou, and M. E. Sherman. 2014. 'How many etiological subtypes of breast cancer: two, three, four, or more?', *J Natl Cancer Inst*, 106.
- Antropova, N., B. Q. Huynh, and M. L. Giger. 2017. 'A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets', *Med Phys*, 44: 5162-71.
- Avanzo, M., J. Stancanella, and I. El Naqa. 2017. 'Beyond imaging: The promise of radiomics', *Phys Med*, 38: 122-39.
- Baltzer, P. A., M. Dietzel, and W. A. Kaiser. 2013. 'A simple and robust classification tree for differentiation between benign and malignant lesions in MR-mammography', *Eur Radiol*, 23: 2051-60.
- Berenguer, R., M. D. R. Pastor-Juan, J. Canales-Vazquez, M. Castro-Garcia, M. V. Villas, F. Mansilla Legorburo, and S. Sabater. 2018. 'Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters', *Radiology*, 288: 407-15.
- Bhooshan, N., M. Giger, L. Lan, H. Li, A. Marquez, A. Shimauchi, and G. M. Newstead. 2011. 'Combined use of T2-weighted MRI and T1-weighted dynamic contrast-enhanced MRI in the automated analysis of breast lesions', *Magn Reson Med*, 66: 555-64.
- Bluff, J. E., S. R. Menakuru, S. S. Cross, S. E. Higham, S. P. Balasubramanian, N. J. Brown, M. W. Reed, and C. A. Staton. 2009. 'Angiogenesis is associated with the onset of hyperplasia in human ductal breast disease', *Br J Cancer*, 101: 666-72.
- Bray, F., J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal. 2018. 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA Cancer J Clin*, 68: 394-424.
- Burnside, E. S., K. Drukker, H. Li, E. Bonaccio, M. Zuley, M. Ganott, J. M. Net, E. J. Sutton, K. R. Brandt, G. J. Whitman, S. D. Conzen, L. Lan, Y. Ji, Y. Zhu, C. C. Jaffe, E. P. Huang, J. B. Freymann, J. S. Kirby, E. A. Morris, and M. L. Giger. 2016. 'Using computer-extracted image phenotypes from tumors on breast magnetic resonance imaging to predict breast cancer pathologic stage', *Cancer*, 122: 748-57.
- Burnside, E. S., E. A. Sickles, L. W. Bassett, D. L. Rubin, C. H. Lee, D. M. Ikeda, E. B. Mendelson, P. A. Wilcox, P. F. Butler, and C. J. D'Orsi. 2009. 'The ACR BI-RADS experience: learning from history', *J Am Coll Radiol*, 6: 851-60.
- Chen, W., M. L. Giger, G. M. Newstead, U. Bick, S. A. Jansen, H. Li, and L. Lan. 2010. 'Computerized assessment of breast lesion malignancy using DCE-MRI

- robustness study on two independent clinical datasets from two manufacturers', *Acad Radiol*, 17: 822-9.
- Chhor, C. M., and C. L. Mercado. 2017. 'Abbreviated MRI Protocols: Wave of the Future for Breast Cancer Screening', *AJR Am J Roentgenol*, 208: 284-89.
- Collewet, G., M. Strzelecki, and F. Mariette. 2004. 'Influence of MRI acquisition protocols and image intensity normalization methods on texture classification', *Magn Reson Imaging*, 22: 81-91.
- Davnall, F., C. S. Yip, G. Ljungqvist, M. Selmi, F. Ng, B. Sanghera, B. Ganeshan, K. A. Miles, G. J. Cook, and V. Goh. 2012. 'Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?', *Insights Imaging*, 3: 573-89.
- DeSantis, C. E., F. Bray, J. Ferlay, J. Lortet-Tieulent, B. O. Anderson, and A. Jemal. 2015. 'International Variation in Female Breast Cancer Incidence and Mortality Rates', *Cancer Epidemiol Biomarkers Prev*, 24: 1495-506.
- Dong, Y., Q. Feng, W. Yang, Z. Lu, C. Deng, L. Zhang, Z. Lian, J. Liu, X. Luo, S. Pei, X. Mo, W. Huang, C. Liang, B. Zhang, and S. Zhang. 2017. 'Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI', *Eur Radiol*.
- Erickson, B. J., P. Korfiatis, Z. Akkus, and T. L. Kline. 2017. 'Machine Learning for Medical Imaging', *Radiographics*, 37: 505-15.
- Fan, M., H. Li, S. Wang, B. Zheng, J. Zhang, and L. Li. 2017. 'Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer', *PLoS One*, 12: e0171683.
- Gallego-Ortiz, C., and A. L. Martel. 2016. 'Improving the Accuracy of Computer-aided Diagnosis for Breast MR Imaging by Differentiating between Mass and Nonmass Lesions', *Radiology*, 278: 679-88.
- Gatenby, R. A., O. Grove, and R. J. Gillies. 2013. 'Quantitative imaging in cancer evolution and ecology', *Radiology*, 269: 8-15.
- Gibbs, P., and L. W. Turnbull. 2003. 'Textural analysis of contrast-enhanced MR images of the breast', *Magn Reson Med*, 50: 92-8.
- Giger, M. L. 2010. 'Update on the potential of computer-aided diagnosis for breast cancer', *Future Oncol*, 6: 1-4.
- . 2018. 'Machine Learning in Medical Imaging', *J Am Coll Radiol*, 15: 512-20.
- Gillies, R. J., P. E. Kinahan, and H. Hricak. 2016. 'Radiomics: Images Are More than Pictures, They Are Data', *Radiology*, 278: 563-77.
- Goeman, J. J., and A. Solari. 2014. 'Multiple hypothesis testing in genomics', *Stat Med*, 33: 1946-78.
- Goldhirsch, A., E. P. Winer, A. S. Coates, R. D. Gelber, M. Piccart-Gebhart, B. Thurlimann, H. J. Senn, and members Panel. 2013. 'Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013', *Ann Oncol*, 24: 2206-23.
- Goldhirsch, A., W. C. Wood, A. S. Coates, R. D. Gelber, B. Thurlimann, H. J. Senn, and members Panel. 2011. 'Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011', *Ann Oncol*, 22: 1736-47.
- Grimm, L. J., J. Zhang, and M. A. Mazurowski. 2015. 'Computational approach to radiogenomics of breast cancer: Luminal A and luminal B molecular subtypes are associated with imaging features on routine breast MRI extracted using computer vision algorithms', *J Magn Reson Imaging*, 42: 902-7.

- Guo, W., H. Li, Y. Zhu, L. Lan, S. Yang, K. Drukker, E. Morris, E. Burnside, G. Whitman, M. L. Giger, Y. Ji, and Group Tcga Breast Phenotype Research. 2015. 'Prediction of clinical phenotypes in invasive breast carcinomas from the integration of radiomics and genomics data', *J Med Imaging (Bellingham)*, 2: 041007.
- Gutierrez, R. L., W. B. DeMartini, P. R. Eby, B. F. Kurland, S. Peacock, and C. D. Lehman. 2009. 'BI-RADS lesion characteristics predict likelihood of malignancy in breast MRI for masses but not for nonmasslike enhancement', *AJR Am J Roentgenol*, 193: 994-1000.
- Kawai, M., M. Kataoka, S. Kanao, M. Iima, N. Onishi, A. Ohashi, R. Sakaguchi, M. Toi, and K. Togashi. 2018. 'The Value of Lesion Size as an Adjunct to the BI-RADS-MRI 2013 Descriptors in the Diagnosis of Solitary Breast Masses', *Magn Reson Med Sci*, 17: 203-10.
- Khalifa, F., A. Soliman, A. El-Baz, M. Abou El-Ghar, T. El-Diasty, G. Gimel'farb, R. Ouseph, and A. C. Dwyer. 2014. 'Models and methods for analyzing DCE-MRI: a review', *Med Phys*, 41: 124301.
- Koh, J., and M. J. Kim. 2019. 'Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage', *Korean J Radiol*, 20: 69-82.
- Kriege, M., C. T. Brekelmans, C. Boetes, P. E. Besnard, H. M. Zonderland, I. M. Obdeijn, R. A. Manoliu, T. Kok, H. Peterse, M. M. Tilanus-Linthorst, S. H. Muller, S. Meijer, J. C. Oosterwijk, L. V. Beex, R. A. Tollenaar, H. J. de Koning, E. J. Rutgers, J. G. Klijn, and Group Magnetic Resonance Imaging Screening Study. 2004. 'Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition', *N Engl J Med*, 351: 427-37.
- Kuhl, C. 2007a. 'The current status of breast MR imaging. Part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice', *Radiology*, 244: 356-78.
- Kuhl, C. K. 2007b. 'Current status of breast MR imaging. Part 2. Clinical applications', *Radiology*, 244: 672-91.
- . 2018. 'Abbreviated breast MRI for screening women with dense breast: the EA1141 trial', *Br J Radiol*, 91: 20170441.
- Kuhl, C. K., A. Keulers, K. Strobel, H. Schneider, N. Gaisa, and S. Schrading. 2018. 'Not all false positive diagnoses are equal: On the prognostic implications of false-positive diagnoses made in breast MRI versus in mammography / digital tomosynthesis screening', *Breast Cancer Res*, 20: 13.
- Kuhl, C. K., S. Klaschik, P. Mielcarek, J. Gieseke, E. Wardelmann, and H. H. Schild. 1999. 'Do T2-weighted pulse sequences help with the differential diagnosis of enhancing lesions in dynamic breast MRI?', *J Magn Reson Imaging*, 9: 187-96.
- Kuhl, C. K., P. Mielcarek, S. Klaschik, C. Leutner, E. Wardelmann, J. Gieseke, and H. H. Schild. 1999. 'Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions?', *Radiology*, 211: 101-10.
- Kuhl, C. K., S. Schrading, K. Strobel, H. H. Schild, R. D. Hilgers, and H. B. Bieling. 2014. 'Abbreviated breast magnetic resonance imaging (MRI): first postcontrast subtracted images and maximum-intensity projection—a novel approach to breast cancer screening with MRI', *J Clin Oncol*, 32: 2304-10.
- Kuhl, C. K., K. Strobel, H. Bieling, C. Leutner, H. H. Schild, and S. Schrading. 2017. 'Supplemental Breast MR Imaging Screening of Women with Average Risk of Breast Cancer', *Radiology*, 283: 361-70.
- Kuhl, C., S. Weigel, S. Schrading, B. Arand, H. Bieling, R. Konig, B. Tombach, C. Leutner, A. Rieber-Brambs, D. Nordhoff, W. Heindel, M. Reiser, and H. H. Schild. 2010.

- 'Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of breast cancer: the EVA trial', *J Clin Oncol*, 28: 1450-7.
- Leach, M. O., C. R. Boggis, A. K. Dixon, D. F. Easton, R. A. Eeles, D. G. Evans, F. J. Gilbert, I. Griebisch, R. J. Hoff, P. Kessar, S. R. Lakhani, S. M. Moss, A. Nerurkar, A. R. Padhani, L. J. Pointon, D. Thompson, R. M. Warren, and Maribs study group. 2005. 'Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS)', *Lancet*, 365: 1769-78.
- LeCun, Y., Y. Bengio, and G. Hinton. 2015. 'Deep learning', *Nature*, 521: 436-44.
- Leitlinienprogramm Onkologie Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF. 2017. 'S3-Leitlinie Früherkennung, Diagnose, Therapie und Nachsorge des Mammakarzinoms, Version 4.0', <http://www.leitlinienprogramm-onkologie.de/leitlinien/mammakarzinom/>.
- Liang, C., Z. Cheng, Y. Huang, L. He, X. Chen, Z. Ma, X. Huang, C. Liang, and Z. Liu. 2018. 'An MRI-based Radiomics Classifier for Preoperative Prediction of Ki-67 Status in Breast Cancer', *Acad Radiol*, 25: 1111-17.
- Liberman, L., G. Mason, E. A. Morris, and D. D. Dershaw. 2006. 'Does size matter? Positive predictive value of MRI-detected breast lesions as a function of lesion size', *AJR Am J Roentgenol*, 186: 426-30.
- Liberman, L., E. A. Morris, M. J. Lee, J. B. Kaplan, L. R. LaTrenta, J. H. Menell, A. F. Abramson, S. M. Dashnaw, D. J. Ballon, and D. D. Dershaw. 2002. 'Breast lesions detected on MR imaging: features and positive predictive value', *AJR Am J Roentgenol*, 179: 171-8.
- Litjens, G., T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. W. van der Laak, B. van Ginneken, and C. I. Sanchez. 2017. 'A survey on deep learning in medical image analysis', *Med Image Anal*, 42: 60-88.
- Lord, S. J., W. Lei, P. Craft, J. N. Cawson, I. Morris, S. Walleiser, A. Griffiths, S. Parker, and N. Houssami. 2007. 'A systematic review of the effectiveness of magnetic resonance imaging (MRI) as an addition to mammography and ultrasound in screening young women at high risk of breast cancer', *Eur J Cancer*, 43: 1905-17.
- Ma, W., Y. Ji, L. Qi, X. Guo, X. Jian, and P. Liu. 2018. 'Breast cancer Ki67 expression prediction by DCE-MRI radiomics features', *Clin Radiol*, 73: 909 e1-09 e5.
- Mahoney, M. C., C. Gatsonis, L. Hanna, W. B. DeMartini, and C. Lehman. 2012. 'Positive predictive value of BI-RADS MR imaging', *Radiology*, 264: 51-8.
- Malich, A., D. R. Fischer, S. Wurdinger, J. Boettcher, C. Marx, M. Facius, and W. A. Kaiser. 2005. 'Potential MRI interpretation model: differentiation of benign from malignant breast masses', *AJR Am J Roentgenol*, 185: 964-70.
- Mann, R. M., C. K. Kuhl, K. Kinkel, and C. Boetes. 2008. 'Breast MRI: guidelines from the European Society of Breast Imaging', *Eur Radiol*, 18: 1307-18.
- Mann, R. M., C. K. Kuhl, and L. Moy. 2019. 'Contrast-enhanced MRI for breast cancer screening', *J Magn Reson Imaging*.
- Menze, B. H., A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. C.

- Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput. 2015. 'The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)', *IEEE Trans Med Imaging*, 34: 1993-2024.
- Milenkovic, J., K. Hertl, A. Kosir, J. Zibert, and J. F. Tasic. 2013. 'Characterization of spatiotemporal changes for the classification of dynamic contrast-enhanced magnetic-resonance breast lesions', *Artif Intell Med*, 58: 101-14.
- Morris EA, Comstock CE, Lee CH, et al. 2013. 'ACR BI-RADS® Magnetic Resonance Imaging', *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA, American College of Radiology.
- Nam, K. J., H. Park, E. S. Ko, Y. Lim, H. H. Cho, and J. E. Lee. 2019. 'Radiomics signature on 3T dynamic contrast-enhanced magnetic resonance imaging for estrogen receptor-positive invasive breast cancers: Preliminary results for correlation with Oncotype DX recurrence scores', *Medicine (Baltimore)*, 98: e15871.
- NCI. 2019. 'National Cancer Institute: Breast Cancer Screening', <https://www.cancer.gov/types/breast/hp/breast-screening-pdq>.
- Newell, D., K. Nie, J. H. Chen, C. C. Hsu, H. J. Yu, O. Nalcioglu, and M. Y. Su. 2010. 'Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement', *Eur Radiol*, 20: 771-81.
- Niell, B. L., P. E. Freer, R. J. Weinfurtner, E. K. Arleo, and J. S. Drukteinis. 2017. 'Screening for Breast Cancer', *Radiol Clin North Am*, 55: 1145-62.
- Prat, A., E. Pineda, B. Adamo, P. Galvan, A. Fernandez, L. Gaba, M. Diez, M. Viladot, A. Arance, and M. Munoz. 2015. 'Clinical implications of the intrinsic molecular subtypes of breast cancer', *Breast*, 24 Suppl 2: S26-35.
- Rageth, C. J., E. A. O'Flynn, C. Comstock, C. Kurtz, R. Kubik, H. Madjar, D. Lepori, G. Kampmann, A. Munding, A. Baege, T. Decker, S. Hosch, C. Tausch, J. F. Delaloye, E. Morris, and Z. Varga. 2016. 'First International Consensus Conference on lesions of uncertain malignant potential in the breast (B3 lesions)', *Breast Cancer Res Treat*, 159: 203-13.
- Reis-Filho, J. S., and L. Pusztai. 2011. 'Gene expression profiling in breast cancer: classification, prognostication, and prediction', *Lancet*, 378: 1812-23.
- Robert-Koch-Institut. 2017. 'Krebs in Deutschland für 2013/2014', [https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs\\_in\\_Deutschland/kid\\_2017/krebs\\_in\\_deutschland\\_2017.pdf?blob=publicationFile](https://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs_in_Deutschland/kid_2017/krebs_in_deutschland_2017.pdf?blob=publicationFile).
- Saha, A., X. Yu, D. Sahoo, and M. A. Mazurowski. 2017. 'Effects of MRI scanner parameters on breast cancer radiomics', *Expert Syst Appl*, 87: 384-91.
- Sardanelli, F., F. Podo, F. Santoro, S. Manoukian, S. Bergonzi, G. Trecate, D. Vergnaghi, M. Federico, L. Cortesi, S. Corcione, S. Morassut, C. Di Maggio, A. Cilotti, L. Martincich, M. Calabrese, C. Zuiani, L. Preda, B. Bonanni, L. A. Carbonaro, A. Contegiacomo, P. Panizza, E. Di Cesare, A. Savarese, M. Crecco, D. Turchetti, M. Tonutti, P. Belli, A. D. Maschio, and Study High Breast Cancer Risk Italian. 2011. 'Multicenter surveillance of women at high genetic breast cancer risk using mammography, ultrasonography, and contrast-enhanced magnetic resonance imaging (the high breast cancer risk italian 1 study): final results', *Invest Radiol*, 46: 94-105.

- Saslow, D., C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener, R. A. Smith, E. Warner, M. Yaffe, K. S. Andrews, C. A. Russell, and Group American Cancer Society Breast Cancer Advisory. 2007. 'American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography', *CA Cancer J Clin*, 57: 75-89.
- Seely, J. M., and T. Alhassan. 2018. 'Screening for breast cancer in 2018-what should we be doing today?', *Curr Oncol*, 25: S115-S24.
- Senkus, E., S. Kyriakides, S. Ohno, F. Penault-Llorca, P. Poortmans, E. Rutgers, S. Zackrisson, F. Cardoso, and Esmo Guidelines Committee. 2015. 'Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up', *Ann Oncol*, 26 Suppl 5: v8-30.
- Siegel, R. L., K. D. Miller, and A. Jemal. 2019. 'Cancer statistics, 2019', *CA Cancer J Clin*, 69: 7-34.
- Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning, and A. L. Borresen-Dale. 2001. 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', *Proc Natl Acad Sci U S A*, 98: 10869-74.
- Szabo, B. K., P. Aspelin, M. Kristoffersen Wiberg, T. Tot, and B. Bone. 2003. 'Invasive breast cancer: correlation of dynamic MR features with prognostic factors', *Eur Radiol*, 13: 2425-35.
- Topol, E. J. 2019. 'High-performance medicine: the convergence of human and artificial intelligence', *Nat Med*, 25: 44-56.
- Tozaki, M., T. Igarashi, and K. Fukuda. 2006. 'Positive and negative predictive values of BI-RADS-MRI descriptors for focal breast masses', *Magn Reson Med Sci*, 5: 7-15.
- Truhn, D., S. Schradang, C. Haarbuerger, H. Schneider, D. Merhof, and C. Kuhl. 2018. 'Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-enhancing Lesions at Multiparametric Breast MRI', *Radiology*: 181352.
- van Bodegraven, E. A., J. C. van Raaij, M. Van Goethem, and W. A. A. Tjalma. 2017. 'Guidelines and recommendations for MRI in breast cancer follow-up: A review', *Eur J Obstet Gynecol Reprod Biol*, 218: 5-11.
- Viallard, C., and B. Larrivee. 2017. 'Tumor angiogenesis and vascular normalization: alternative therapeutic targets', *Angiogenesis*, 20: 409-26.
- Voduc, K. D., M. C. Cheang, S. Tyldesley, K. Gelmon, T. O. Nielsen, and H. Kennecke. 2010. 'Breast cancer subtypes and the risk of local and regional relapse', *J Clin Oncol*, 28: 1684-91.
- Weinstein, S. P., A. R. Localio, E. F. Conant, M. Rosen, K. M. Thomas, and M. D. Schnall. 2009. 'Multimodality screening of high-risk women: a prospective cohort study', *J Clin Oncol*, 27: 6124-8.
- Yang, Q., L. Li, J. Zhang, G. Shao, and B. Zheng. 2015. 'A new quantitative image analysis method for improving breast cancer diagnosis using DCE-MRI examinations', *Med Phys*, 42: 103-9.
- Zheng, Y., S. Englander, S. Baloch, E. I. Zacharaki, Y. Fan, M. D. Schnall, and D. Shen. 2009. 'STEP: spatiotemporal enhancement pattern for MR-based breast tumor diagnosis', *Med Phys*, 36: 3192-204.
- Zwanenburg A, Leger S, Vallières M, Löck S. 'Image biomarker standardisation initiative', *arXiv preprint*: arXiv:1612.07003.

## Anhang

### Selektionsfrequenzen der L1-Regularisierung

Tabelle 8: Featureselektionsfrequenzen der L1-Regularisierung innerhalb der 10-fachen Kreuzvalidierung

Frequenz	Sequenz	Featureart	Feature
10/10	T2	Statistik	10Percentile
10/10	T2	Textur	glcm-ClusterProminence
10/10	T2	Textur	glcm-Correlation
10/10	T2	Textur	glszm-HighGrayLevelZoneEmphasis
10/10	T2	Textur	glszm-SmallAreaLowGrayLevelEmphasis
10/10	T2	Form	Sphericity
10/10	T1 nativ	Statistik	Interquartile Range
10/10	Post-KM 1	Statistik	Skewness
10/10	Post-KM 1	Textur	glcm-JointAverage
10/10	Post-KM 1	Textur	glcm-JointEnergy
10/10	Post-KM 1	Textur	glszm-SmallAreaLowGrayLevelEmphasis
10/10	Post-KM 2	Textur	gldm- SmallDependenceHighGrayLevelEmphasis
10/10	Post-KM 2	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
10/10	Post-KM 4	Statistik	InterquartileRange
10/10	Post-KM 4	Textur	glrlm-LongRunHighGrayLevelEmphasis
10/10	Post-KM 4	Textur	glrlm-RunEntropy
10/10	Sub	Statistik	10Percentile
10/10	Sub	Statistik	Kurtosis
10/10	Sub	Statistik	Skewness
10/10	Sub	Textur	glcm-lmc2
10/10	Sub	Textur	gldm-DependenceEntropy
10/10	Sub	Textur	gldm- LowGrayLevelEmphasis
10/10	Sub	Textur	glszm-SmallAreaEmphasis
9/10	T2	Textur	glcm-lmc
9/10	T2	Textur	glszm-SmallAreaEmphasis

9/10	T1 nativ	Statistik	10Percentile
9/10	T1 nativ	Textur	glcm-ClusterShade
9/10	Post-KM 1	Statistik	Maximum
9/10	Post-KM 1	Textur	gldm-LargeDependenceLowGrayLevelEmphasis
9/10	Post-KM 1	Textur	gldm-LowGrayLevelEmphasis
9/10	Post-KM 3	Textur	glcm-Idn
9/10	Sub	Textur	gldm-LargeDependenceHighGrayLevelEmphasis
8/10	T2	Textur	gldm-DependenceNonUniformityNormalized
8/10	T2	Form	Elongation
8/10	T1 nativ	Statistik	Variance
8/10	T1 nativ	Textur	glcm-Id
8/10	Post-KM 1	Textur	glszm-GrayLevelNonUniformityNormalized
8/10	Post-KM 2	Textur	glcm-ClusterTendency
8/10	Post-KM 2	Textur	gldm-DependenceEntropy
8/10	Post-KM 2	Textur	glszm-SmallAreaEmphasis
8/10	Post-KM 3	Textur	glcm-JointEnergy
8/10	Post-KM 3	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
8/10	Sub	Textur	glcm-Correlation
7/10	T2	Textur	glcm-ClusterShade
7/10	T1 nativ	Textur	glcm-Correlation
7/10	T1 nativ	Textur	gllm-RunVariance
7/10	T1 nativ	Textur	gllm-ShortRunLowGrayLevelEmphasis
7/10	T1 nativ	Textur	glszm-SmallAreaEmphasis
7/10	Post-KM 2	Textur	glszm-GrayLevelNonUniformityNormalized
7/10	Post-KM 3	Textur	glszm-LowGrayLevelZoneEmphasis
7/10	Sub	Textur	glcm-ClusterShade
6/10	T2	Statistik	Skewness
6/10	T1 nativ	Textur	glcm-ClusterProminence
6/10	T1 nativ	Textur	gldm-DependenceNonUniformityNormalized
6/10	T1 nativ	Textur	glszm-SmallAreaHighGrayLevelEmphasis

6/10	Post-KM 2	Textur	glcm-ClusterProminence
6/10	Post-KM 2	Textur	glcm-Correlation
6/10	Post-KM 4	Textur	glcm-Imc2
6/10	Sub	Textur	glcm-JointEnergy
6/10	Sub	Textur	glrlm-LongRunLowGrayLevelEmphasis
5/10	T2	Textur	glcm-Idn
5/10	T1 nativ	Textur	glcm-Imc1
5/10	Post-KM 1	Textur	glcm-ClusterProminence
5/10	Post-KM 1	Textur	glcm-MaximumProbability
5/10	Post-KM 1	Textur	gldm-DependenceEntropy
5/10	Post-KM 1	Textur	glszm-SizeZoneNonUniformityNormalized
5/10	Post-KM 3	Statistik	10Percentile
5/10	Post-KM 3	Statistik	Skewness
5/10	Post-KM 4	Statistik	10Percentile
5/10	Post-KM 4	Statistik	RobustMeanAbsoluteDeviation
5/10	Post-KM 4	Textur	glcm-Idmn
5/10	Sub	Statistik	InterquartileRange
5/10	Sub	Textur	glszm-SizeZoneNonUniformityNormalized
5/10	Sub	Textur	glszm-ZoneEntropy
4/10	T2	Statistik	glcm-Contrast
4/10	T2	Textur	glcm-JointEnergy
4/10	T2	Textur	gldm-LargeDependenceEmphasis
4/10	T2	Textur	glrlm-GrayLevelNonUniformityNormalized
4/10	T1 nativ	Statistik	Median
4/10	T1 nativ	Textur	glcm-DifferenceAverage
4/10	T1 nativ	Textur	glcm-MaximumProbability
4/10	T1 nativ	Textur	gldm-GrayLevelVariance
4/10	T1 nativ	Textur	gldm-LargeDependenceEmphasis
4/10	T1 nativ	Textur	gldm- SmallDependenceHighGrayLevelEmphasis
4/10	Post-KM 1	Statistik	Uniformity

4/10	Post-KM 3	Statistik	InterquartileRange
4/10	Post-KM 3	Textur	glcm-Id
4/10	Post-KM 4	Statistik	Minimum
4/10	Post-KM 4	Textur	glcm-DifferenceVariance
3/10	T2	Textur	gldm-LargeDependenceLowGrayLevelEmphasis
3/10	T2	Textur	glszm-GrayLevelVariance
3/10	T2	Textur	glszm-SizeZoneNonUniformity
3/10	Post-KM 1	Textur	glcm-ClusterShade
3/10	Post-KM 1	Textur	glcm-Imc1
3/10	Post-KM 1	Textur	glcm-InverseVariance
3/10	Post-KM 1	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
3/10	Post-KM 1	Textur	glszm-LargeAreaEmphasis
3/10	Post-KM 2	Statistik	RootMeanSquared
3/10	Post-KM 2	Textur	glcm-InverseVariance
3/10	Post-KM 3	Textur	glcm-Idm
3/10	Post-KM 4	Textur	gldm-DependenceNonUniformityNormalized
3/10	Post-KM 4	Textur	glszm-SmallAreaEmphasis
3/10	Sub	Textur	glcm-InverseVariance
3/10	Sub	Textur	gldm-DependenceNonUniformityNormalized
3/10	Sub	Textur	gldm-DependenceVariance
3/10	Sub	Textur	gldm-SmallDependenceHighGrayLevelEmphasis
2/10	T2	Statistik	Kurtosis
2/10	T2	Textur	glcm-Idmn
2/10	T2	Textur	glcm-MaximumProbability
2/10	T2	Textur	glrlm-LowGrayLevelRunEmphasis
2/10	T1 nativ	Statistik	Minimum
2/10	T1 nativ	Textur	glcm-Contrast
2/10	T1 nativ	Textur	glcm-Idmn
2/10	T1 nativ	Textur	gldm-DependenceVariance
2/10	T1 nativ	Textur	glszm-ZoneEntropy

2/10	Post-KM 1	Textur	gldm- LargeDependenceHighGrayLevelEmphasis
2/10	Post-KM 2	Statistik	10Percentile
2/10	Post-KM 2	Statistik	90Percentile
2/10	Post-KM 2	Textur	glcm-ClusterShade
2/10	Post-KM 2	Textur	glcm-JointEntropy
2/10	Post-KM 2	Textur	glrlm-RunLengthNonUniformityNormalized
2/10	Post-KM 2	Textur	glszm-LowGrayLevelZoneEmphasis
2/10	Post-KM 2	Textur	glszm-ZoneEntropy
2/10	Post-KM 3	Statistik	Kurtosis
2/10	Post-KM 3	Statistik	Uniformity
2/10	Post-KM 3	Textur	glcm-lmc2
2/10	Post-KM 3	Textur	glcm-InverseVariance
2/10	Post-KM 3	Textur	glcm-MaximumProbability
2/10	Post-KM 3	Textur	glrlm-ShortRunLowGrayLevelEmphasis
2/10	Post-KM 4	Statistik	MeanAbsoluteDeviation
2/10	Post-KM 4	Textur	glcm-Contrast
2/10	Post-KM 4	Textur	glszm-SmallAreaLowGrayLevelEmphasis
2/10	Sub	Textur	glcm-Autocorrelation
2/10	Sub	Textur	glcm-ClusterProminence
2/10	Sub	Textur	glcm-Contrast
2/10	Sub	Textur	glcm-ldmn
2/10	Sub	Textur	gldm-LargeDependenceLowGrayLevelEmphasis
2/10	Sub	Textur	glrlm-RunVariance
2/10	Sub	Textur	glszm-LargeAreaHighGrayLevelEmphasis
2/10	Sub	Textur	glszm-LargeAreaLowGrayLevelEmphasis
1/10	T2	Statistik	90Percentile
1/10	T2	Statistik	Maximum
1/10	T2	Statistik	MeanAbsoluteDeviation
1/10	T2	Statistik	Minimum
1/10	T2	Statistik	TotalEnergy

1/10	T2	Statistik	Uniformity
1/10	T2	Textur	glcm-JointAverage
1/10	T2	Textur	glcm-SumEntropy
1/10	T2	Textur	glcm-SumSquares
1/10	T2	Textur	gldm- SmallDependenceHighGrayLevelEmphasis
1/10	T2	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
1/10	T2	Textur	glrlm-LongRunEmphasis
1/10	T2	Textur	glrlm-LongRunHighGrayLevelEmphasis
1/10	T2	Textur	glrlm-LongRunLowGrayLevelEmphasis
1/10	T2	Textur	glrlm-ShortRunLowGrayLevelEmphasis
1/10	T2	Textur	glszm-GrayLevelNonUniformityNormalized
1/10	T2	Textur	glszm-LargeAreaHighGrayLevelEmphasis
1/10	T2	Textur	glszm-LargeAreaLowGrayLevelEmphasis
1/10	T2	Textur	glszm-LowGrayLevelZoneEmphasis
1/10	T2	Textur	glszm-SmallAreaHighGrayLevelEmphasis
1/10	T2	Textur	glszm-ZonePercentage
1/10	T2	Form	Maximum2DDiameterRow
1/10	T2	Form	MinorAxis
1/10	T2	Form	SurfaceArea
1/10	T2	Form	SurfaceVolumeRatio
1/10	T1 nativ	Statistik	Maximum
1/10	T1 nativ	Textur	glcm-ClusterTendency
1/10	T1 nativ	Textur	glcm-DifferenceVariance
1/10	T1 nativ	Textur	glcm-Idn
1/10	T1 nativ	Textur	glcm-InverseVariance
1/10	T1 nativ	Textur	gldm- LargeDependenceHighGrayLevelEmphasis
1/10	T1 nativ	Textur	gldm-LowGrayLevelEmphasis
1/10	T1 nativ	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
1/10	T1 nativ	Textur	glrlm-GrayLevelNonUniformityNormalized

1/10	T1 nativ	Textur	glrlm-LongRunLowGrayLevelEmphasis
1/10	T1 nativ	Textur	glrlm-RunEntropy
1/10	T1 nativ	Textur	glszm-GrayLevelNonUniformityNormalized
1/10	T1 nativ	Textur	glszm-GrayLevelVariance
1/10	T1 nativ	Textur	glszm-HighGrayLevelZoneEmphasis
1/10	T1 nativ	Textur	glszm-LargeAreaHighGrayLevelEmphasis
1/10	T1 nativ	Textur	glszm-LargeAreaLowGrayLevelEmphasis
1/10	T1 nativ	Textur	glszm-LowGrayLevelZoneEmphasis
1/10	T1 nativ	Textur	glszm-SmallAreaLowGrayLevelEmphasis
1/10	T1 nativ	Textur	glszm-ZoneVariance
1/10	Post-KM 1	Statistik	Kurtosis
1/10	Post-KM 1	Statistik	Median
1/10	Post-KM 1	Statistik	Range
1/10	Post-KM 1	Statistik	RootMeanSquared
1/10	Post-KM 1	Textur	glcm-DifferenceVariance
1/10	Post-KM 1	Textur	glcm-Idmn
1/10	Post-KM 1	Textur	glcm-JointEntropy
1/10	Post-KM 1	Textur	glcm-SumEntropy
1/10	Post-KM 1	Textur	gldm-DependenceNonUniformityNormalized
1/10	Post-KM 1	Textur	gldm-DependenceVariance
1/10	Post-KM 1	Textur	gldm-SmallDependenceHighGrayLevelEmphasis
1/10	Post-KM 1	Textur	glrlm-GrayLevelVariance
1/10	Post-KM 1	Textur	glrlm-LongRunLowGrayLevelEmphasis
1/10	Post-KM 1	Textur	glrlm-ShortRunLowGrayLevelEmphasis
1/10	Post-KM 1	Textur	glszm-GrayLevelVariance
1/10	Post-KM 1	Textur	glszm-HighGrayLevelZoneEmphasis
1/10	Post-KM 1	Textur	glszm-LowGrayLevelZoneEmphasis
1/10	Post-KM 1	Textur	glszm-SmallAreaEmphasis
1/10	Post-KM 1	Textur	glszm-ZonePercentage
1/10	Post-KM 1	Textur	glszm-ZoneVariance

1/10	Post-KM 2	Statistik	Kurtosis
1/10	Post-KM 2	Statistik	Maximum
1/10	Post-KM 2	Statistik	Range
1/10	Post-KM 2	Statistik	RobustMeanAbsoluteDeviation
1/10	Post-KM 2	Textur	glcm-Imc1
1/10	Post-KM 2	Textur	glcm-Imc2
1/10	Post-KM 2	Textur	glcm-JointAverage
1/10	Post-KM 2	Textur	glcm-MaximumProbability
1/10	Post-KM 2	Textur	gldm-DependenceVariance
1/10	Post-KM 2	Textur	gldm-LargeDependenceEmphasis
1/10	Post-KM 2	Textur	gldm-LargeDependenceLowGrayLevelEmphasis
1/10	Post-KM 2	Textur	gldm-SmallDependenceEmphasis
1/10	Post-KM 2	Textur	glszm-SmallAreaLowGrayLevelEmphasis
1/10	Post-KM 3	Statistik	90Percentile
1/10	Post-KM 3	Statistik	RootMeanSquared
1/10	Post-KM 3	Textur	glcm-ClusterProminence
1/10	Post-KM 3	Textur	glcm-ClusterShade
1/10	Post-KM 3	Textur	glcm-DifferenceEntropy
1/10	Post-KM 3	Textur	gldm-DependenceEntropy
1/10	Post-KM 3	Textur	gldm-DependenceNonUniformityNormalized
1/10	Post-KM 3	Textur	gldm-LargeDependenceLowGrayLevelEmphasis
1/10	Post-KM 3	Textur	glrlm-RunEntropy
1/10	Post-KM 3	Textur	glrlm-ShortRunEmphasis
1/10	Post-KM 3	Textur	glszm-SmallAreaHighGrayLevelEmphasis
1/10	Post-KM 3	Textur	glszm-ZoneEntropy
1/10	Post-KM 4	Statistik	90Percentile
1/10	Post-KM 4	Statistik	Kurtosis
1/10	Post-KM 4	Statistik	Maximum
1/10	Post-KM 4	Statistik	Skewness
1/10	Post-KM 4	Textur	glcm-ClusterProminence
1/10	Post-KM 4	Textur	glcm-ClusterShade

1/10	Post-KM 4	Textur	glcm-lmc1
1/10	Post-KM 4	Textur	glcm-InverseVariance
1/10	Post-KM 4	Textur	glcm-JointEntropy
1/10	Post-KM 4	Textur	gldm- LargeDependenceHighGrayLevelEmphasis
1/10	Post-KM 4	Textur	glrlm-ShortRunEmphasis
1/10	Post-KM 4	Textur	glrlm-ShortRunHighGrayLevelEmphasis
1/10	Post-KM 4	Textur	glszm-GrayLevelNonUniformityNormalized
1/10	Post-KM 4	Textur	glszm-ZonePercentage
1/10	Sub	Statistik	MeanAbsoluteDeviation
1/10	Sub	Statistik	RobustMeanAbsoluteDeviation
1/10	Sub	Statistik	Uniformity
1/10	Sub	Textur	glcm-DifferenceAverage
1/10	Sub	Textur	glcm-DifferenceEntropy
1/10	Sub	Textur	glcm-DifferenceVariance
1/10	Sub	Textur	glcm-JointEntropy
1/10	Sub	Textur	gldm-LargeDependenceEmphasis
1/10	Sub	Textur	gldm-SmallDependenceEmphasis
1/10	Sub	Textur	glrlm-LongRunHighGrayLevelEmphasis
1/10	Sub	Textur	glrlm-LowGrayLevelRunEmphasis
1/10	Sub	Textur	glszm-SmallAreaHighGrayLevelEmphasis
1/10	Sub	Textur	glszm-ZonePercentage

## Selektionsfrequenzen der Hauptkomponentenanalyse

*Tabelle 9: Featureselektionsfrequenzen der Hauptkomponentenanalyse innerhalb der 10-fachen Kreuzvalidierung*

<b>Frequenz</b>	<b>Sequenz</b>	<b>Featureart</b>	<b>Feature</b>
10/10	T2	Textur	glszm-ZoneEntropy
10/10	T2	Form	Sphericity
10/10	T1 nativ	Textur	gldm-DependenceEntropy

10/10	Post-KM 1	Textur	glcm-JointAverage
10/10	Post-KM 1	Textur	gldm-DependenceEntropy
10/10	Post-KM 1	Textur	gldm- LargeDependenceHighGrayLevelEmphasis
10/10	Post-KM 1	Textur	glszm-ZoneEntropy
10/10	Post-KM 2	Textur	glcm-lmc1
10/10	Post-KM 2	Textur	glcm-lmc2
10/10	Post-KM 2	Textur	glszm-ZoneEntropy
10/10	Post-KM 3	Textur	glcm-lmc1
10/10	Post-KM 3	Textur	glcm-lmc2
10/10	Post-KM 4	Textur	glcm-lmc1
10/10	Post-KM 4	Textur	glcm-lmc2
10/10	Sub	Textur	glcm-JointEntropy
10/10	Sub	Textur	gldm-DependenceEntropy
10/10	Sub	Textur	glszm-ZoneEntropy
9/10	T2	Textur	gldm-DependenceEntropy
8/10	T1 nativ	Textur	glszm-ZoneEntropy
7/10	Post-KM 2	Textur	glcm-JointAverage
7/10	Post-KM 2	Textur	gldm-DependenceEntropy
7/10	Post-KM 2	Textur	gldm- LargeDependenceHighGrayLevelEmphasis
6/10	Post-KM 1	Textur	glcm-Autocorrelation
6/10	Post-KM 1	Textur	gllm-LowGrayLevelRunEmphasis
5/10	T2	Statistik	Minimum
5/10	T2	Textur	glcm-lmc1
5/10	T2	Textur	glcm-lmc2
5/10	T2	Textur	gldm-SmallDependenceLowGrayLevelEmphasis

5/10	T2	Form	MajorAxis
5/10	T2	Form	Maximum2DDiameterColumn
5/10	T2	Form	Maximum2DDiameterRow
5/10	T2	Form	Maximum2DDiameterSlice
5/10	T2	Form	Maximum2DDiameter
5/10	T2	Form	MinorAxis
5/10	T2	Form	SurfaceVolumeRatio
5/10	T1 nativ	Textur	glcm-lmc1
5/10	T1 nativ	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
5/10	Post-KM 1	Statistik	Maximum
5/10	Post-KM 1	Statistik	Range
5/10	Post-KM 1	Textur	glcm-lmc1
5/10	Post-KM 1	Textur	glcm-lmc2
5/10	Post-KM 1	Textur	glcm-JointEntropy
5/10	Post-KM 1	Textur	glcm-MaximumProbability
5/10	Post-KM 1	Textur	gldm-HighGrayLevelEmphasis
5/10	Post-KM 1	Textur	gldm-LowGrayLevelEmphasis
5/10	Post-KM 1	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
5/10	Post-KM 1	Textur	glrlm-HighGrayLevelRunEmphasis
5/10	Post-KM 1	Textur	glrlm-LongRunHighGrayLevelEmphasis
5/10	Post-KM 1	Textur	glrlm-ShortRunHighGrayLevelEmphasis
5/10	Post-KM 1	Textur	glrlm-ShortRunLowGrayLevelEmphasis
5/10	Post-KM 1	Textur	glszm-LowGrayLevelZoneEmphasis
5/10	Post-KM 1	Textur	glszm-SmallAreaLowGrayLevelEmphasis
5/10	Post-KM 2	Textur	glcm-Autocorrelation
5/10	Post-KM 2	Textur	glcm-lmnn

5/10	Post-KM 2	Textur	gldm-HighGrayLevelEmphasis
5/10	Post-KM 2	Textur	gllm-HighGrayLevelRunEmphasis
5/10	Post-KM 2	Textur	gllm-LongRunHighGrayLevelEmphasis
5/10	Post-KM 2	Textur	gllm-ShortRunLowGrayLevelEmphasis
5/10	Post-KM 3	Textur	glcm-Autocorrelation
5/10	Post-KM 3	Textur	glcm-ldmn
5/10	Post-KM 3	Textur	glcm-ldn
5/10	Post-KM 3	Textur	glcm-JointAverage
5/10	Post-KM 3	Textur	gldm-DependenceEntropy
5/10	Post-KM 3	Textur	gldm-DependenceNonUniformityNormalized
5/10	Post-KM 3	Textur	gldm- LargeDependenceHighGrayLevelEmphasis
5/10	Post-KM 3	Textur	gllm-ShortRunLowGrayLevelEmphasis
5/10	Post-KM 3	Textur	glszm-ZoneEntropy
5/10	Post-KM 3	Textur	glszm-ZonePercentage
5/10	Post-KM 4	Textur	glcm-ldmn
5/10	Post-KM 4	Textur	glcm-ldn
5/10	Post-KM 4	Textur	gldm-DependenceEntropy
5/10	Post-KM 4	Textur	gldm-DependenceNonUniformityNormalized
5/10	Post-KM 4	Textur	gldm- LargeDependenceHighGrayLevelEmphasis
5/10	Post-KM 4	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
5/10	Post-KM 4	Textur	gllm-ShortRunLowGrayLevelEmphasis
5/10	Post-KM 4	Textur	glszm-ZoneEntropy
5/10	Post-KM 4	Textur	glszm-ZonePercentage
5/10	Sub	Textur	glcm-lmc1
5/10	Sub	Textur	glcm-MaximumProbability

5/10	Sub	Textur	gldm- LargeDependenceHighGrayLevelEmphasis
4/10	T1 nativ	Textur	glcm-lmc2
4/10	Post-KM 1	Textur	glcm-JointEnergy
4/10	Post-KM 2	Statistik	Range
4/10	Post-KM 2	Textur	gllm-ShortRunHighGrayLevelEmphasis
4/10	Sub	Textur	glszm-SmallAreaLowGrayLevelEmphasis
3/10	Post-KM 1	Textur	gllm-LongRunLowGrayLevelEmphasis
3/10	Post-KM 1	Textur	glszm-HighGrayLevelZoneEmphasis
3/10	Sub	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
2/10	T2	Textur	glszm-SmallAreaLowGrayLevelEmphasis
2/10	Post-KM 2	Textur	glcm-lcn
2/10	Post-KM 2	Textur	gldm-DependenceNonUniformityNormalized
2/10	Post-KM 3	Textur	gllm-LongRunHighGrayLevelEmphasis
2/10	Post-KM 4	Textur	glcm-JointAverage
2/10	Post-KM 4	Textur	gldm-SmallDependenceEmphasis
1/10	T2	Textur	glszm-HighGrayLevelZoneEmphasis
1/10	T1 nativ	Textur	gllm-ShortRunLowGrayLevelEmphasis
1/10	Post-KM 2	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
1/10	Post-KM 2	Textur	glszm-ZonePercentage
1/10	Post-KM 3	Textur	gldm-HighGrayLevelEmphasis
1/10	Post-KM 3	Textur	gldm-SmallDependenceEmphasis
1/10	Post-KM 3	Textur	gldm-SmallDependenceLowGrayLevelEmphasis
1/10	Sub	Textur	glcm-JointEnergy
1/10	Sub	Textur	glszm-LowGrayLevelZoneEmphasis

## **Auflistung eigener Publikationen**

Not all false positive diagnoses are equal: On the prognostic implications of false-positive diagnoses made in breast MRI versus in mammography / digital tomosynthesis screening. Kuhl, C, Keulers, A, Strobel, K, Schneider, H, Gaisa, N, Schrading, S. Breast Cancer Res. 2018 Feb 9;20(1):13.

## **Danksagung**

Mein besonderer Dank gilt meiner Doktormutter PD Dr. med. Simone Schrading, meinem Betreuer Dipl.-Phys. Dr. med. Daniel Truhn sowie M.Sc. Christoph Haarburger für die Unterstützung bei der Durchführung und Verfassung meiner Doktorarbeit.

Außerdem danke ich Univ.-Prof. Dr. med. Christiane Kuhl sowie Univ.-Prof. Dipl.-Ing. Dorit Merhof für die Ermöglichung der Mitarbeit an der zugehörigen Publikation.

Weiterer Dank für die Unterstützung bei der Verfassung der Doktorarbeit sowie die Begleitung durch mein Studium gilt Rachel Becker, Stefan Schmidt, Maximilian Fabry und meinen Eltern Marietta und Lothar Schneider.

## **Erklärung § 5 Abs. 1 zur Datenaufbewahrung**

Hiermit erkläre ich, dass die dieser Dissertation zu Grunde liegenden Originaldaten am Lehrstuhl für Bildverarbeitung der RWTH Aachen hinterlegt sind.

## Eidesstattliche Erklärung gemäß § 5 Abs. (1) und § 11 Abs. (3) 12. der Promotionsordnung

Hiermit erkläre ich, **Hannah Sophia Schneider** an Eides statt, dass ich folgende in der von mir selbstständig erstellten Dissertation „**Radiomics in der Mamma MRT: Klassifikation kontrastmittelaufnehmender Läsionen nach ihrer Dignität mittels zweier klassischer Machine Learning Algorithmen**“ dargestellten Ergebnisse erhoben habe:

Bei der Durchführung der Arbeit hatte ich folgende Hilfestellungen, die in der Danksagung angegeben sind:

	Doktorandin Hannah Sophia Schneider	Prakt. Betreuer Dipl. phys. Dr. med. Daniel Truhn	Kooperations- partner Christoph Haarburger, M.Sc.	Doktormutter PD Dr. med. Simone Schrading	Summe (%)
Studienüberwachung		50		50	100
Studiendesign/Konzeption	20	30	20	30	100
Anamneseerhebung	100				100
Auswertung Anamnese	100				100
Durchführung der Segmentierung	80			20	100
Durchführung Machine Learning	20	40	40		100
Interpretation der Datenauswertung	40	20	20	20	100

---

Unterschrift der Doktorandin/des Doktoranden

Als Betreuer der obigen Dissertation bestätige ich die Angaben von Hannah Sophia Schneider

---

Unterschrift der Doktormutter

## Lebenslauf

Name: Hannah Sophia Schneider  
Geboren: 28.06.1992 in Singen am Hohentwiel

### Schulbildung:

1998-2002 Grundschole Stockach  
2002-2011 Nellenburg-Gymnasium Stockach  
Abschluss: Allgemeine Hochschulreife, Note: 1,0  
2008-2009 Monaghan Collegiate School, Co. Monaghan, Irland  
Abschluss: Leaving Certificate, Note: 530/600 Punkten

### Universitäre Ausbildung:

2011-2018 RWTH Aachen, Studium der Medizin  
Abschluss: Staatsexamen 12/2018, Note: 2  
Ärztliche Approbation am 14.01.2019  
Praktisches Jahr: Innere Medizin Hämatologie/Onkologie und Gastroenterologie Uniklinik Aachen, Chirurgie Kantonsspital Schaffhausen, Wahlfach Radiologie Uniklinik Aachen  
2012-2015 RWTH Aachen und FU Hagen, Studium der Mathematik ohne Abschluss

### Berufslaufbahn:

2012-2015 Modellstudiengang Medizin Aachen  
Uniklinik Aachen  
Studentische Hilfskraft als Tutorin (Physik, Statistik, Physiologie)  
2014-2017 Klinik für Diagnostische und Interventionelle Radiologie,  
Uniklinik Aachen  
Studentische Hilfskraft in der AG Mammadiagnostik  
Seit 11/2016 Klinik für Diagnostische und Interventionelle Radiologie,  
Uniklinik Aachen  
Doktorandin in der AG Mammadiagnostik  
Seit 03/2019 Klinik für Diagnostische und Interventionelle Radiologie,  
Universitätsklinikum Freiburg  
Ärztin in Weiterbildung

### Sonstiges:

2016-2018 Stipendiatin im Programm „Die Hellsten Köpfe für die Radiologie“ der Deutschen Röntgengesellschaft