

# Wie geht Radiomics eigentlich? – Review

## How does Radiomics actually work? – Review

### Autoren

Ulrike Irmgard Attenberger<sup>1</sup>, Georg Langs<sup>2</sup>

### Institute

- 1 Department of Diagnostic and Interventional Radiology, Medical Faculty, University Hospital Bonn, Germany
- 2 Department of Biomedical Imaging and Image-guided Therapy, Computational Imaging Research Lab, Medical University of Vienna, Wien, Austria

### Key words

machine learning, radiomics, artificial intelligence, features

eingereicht 02.03.2020

akzeptiert 05.10.2020

online publiziert 02.12.2020

### Bibliografie

Fortschr Röntgenstr 2021; 193: 652–657

DOI 10.1055/a-1293-8953

ISSN 1438-9029

© 2020. Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14,  
70469 Stuttgart, Germany

### Korrespondenzadresse

Prof. Ulrike Irmgard Attenberger  
Department of Diagnostic and Interventional Radiology,  
University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn,  
Germany  
Tel.: +49/2 28/28 71 58 71  
ulrike.attenberger@ukbonn.de

### ZUSAMMENFASSUNG

Personalisierte Präzisionsmedizin setzt eine hochakurate Diagnostik voraus. Während die radiologische Forschung sich in den letzten Jahrzehnten mit Scanner- und Sequenztechnologien beschäftigt hat, rücken zunehmend Anwendungen der künstlichen Intelligenz in das wissenschaftliche Interesse, da sie die Möglichkeit der objektiven Quantifizierung und diagnostischen bzw. prognostischen Nutzung von Bildinformationen substanziiell erweitern könnten.

In diesem Zusammenhang beschreibt der Begriff „Radiomics“ die Extraktion quantitativer Merkmale aus Bilddaten wie zum Beispiel von Computertomografie- oder Magnetresonanztomografie-Untersuchungen. Diese Merkmale werden mithilfe von Modellen des maschinellen Lernens mit Vorhersagezielen wie Diagnose oder Prognose in Zusammenhang gebracht. Man geht davon aus, dass die integrative Beurteilung der so erhobenen Merkmalsmuster in Verbindung mit klinischen, molekularen und

genetischen Daten eine genauere Charakterisierung der Pathophysiologie von Erkrankungen sowie eine präzisere Vorhersage von Therapieansprechen und Outcome ermöglichen kann. In dieser Übersichtsarbeit werden der klassische Radiomics-Ansatz beschrieben und die bestehende sehr große Variabilität an Zugängen diskutiert. Abschließend werden Forschungsrichtungen skizziert, in die sich das von zunehmend enger Kollaboration zwischen Radiologie und Computerwissenschaften und der Notwendigkeit neuer Ausbildungskonzepte gekennzeichnete interdisziplinäre Feld bewegt. Ziel ist es, eine Grundlage für verantwortungsvollen, nachvollziehbaren Umgang mit eingebrachten Daten und angewandten Analysemethoden zu ermöglichen.

### Kernaussagen:

- Radiomics gewinnt einen immer größeren Stellenwert in der bildgebenden Forschung.
- Radiomics birgt ein großes Potenzial, den Erfordernissen der Präzisionsmedizin gerecht zu werden.
- Die Radiomicsanalyse unterliegt noch einer großen Variabilität.
- Wir benötigen eine qualitätsgesicherte Anwendung von Radiomics in der Medizin.

### Zitierweise

- Attenberger UI, Langs G, . How does Radiomics actually work? – Review. Fortschr Röntgenstr 2021; 193: 652–657

### ABSTRACT

Personalized precision medicine requires highly accurate diagnostics. While radiological research has focused on scanner and sequence technologies in recent decades, applications of artificial intelligence are increasingly attracting scientific interest as they could substantially expand the possibility of objective quantification and diagnostic or prognostic use of image information.

In this context, the term “radiomics” describes the extraction of quantitative features from imaging data such as those obtained from computed tomography or magnetic resonance imaging examinations. These features are associated with predictive goals such as diagnosis or prognosis using machine learning models. It is believed that the integrative assessment of the feature patterns thus obtained, in combination with clinical, molecular and genetic data, can enable a more accurate characterization of the pathophysiology of diseases and more precise prediction of therapy response and outcome.

This review describes the classical radiomics approach and discusses the existing very large variability of approaches. Finally, it outlines the research directions in which the interdisciplinary field of radiology and computer science is mov-

ing, characterized by increasingly close collaborations and the need for new educational concepts. The aim is to provide a basis for responsible and comprehensible handling of the data and analytical methods used.

## Einleitung

Die Anforderungen der personalisierten Präzisionsmedizin setzen eine hochakkurate Diagnostik voraus. Während die radiologische Forschung sich in den letzten Jahren und Jahrzehnten mit der Evaluation von Scanner- und Sequenztechnologien für eine akkuratere Erkrankungsdiagnostik befasste, stehen derzeit Implementierungen der künstlichen Intelligenz (KI) für eine optimierte Diagnostik im Zentrum des wissenschaftlichen Interesses. Die Implementierungsmöglichkeiten für KI-Techniken in der Radiologie sind vielseitig: automatisierte Läsionsdetektion und -charakterisierung, Erstellung von Biobanken, Dosisoptimierung, „structured reporting“ und Radiomics [2, 3]. Der Vollständigkeit halber nicht vergessen werden darf, dass KI-Techniken auch bei Scannern neuester Generation zur Optimierung der Datenakquisition selbst eingesetzt werden.

Der Begriff „Radiomics“ beschreibt die Extraktion von quantitativen Merkmalen aus Bilddaten, z. B. Computertomografie (CT), Magnetresonanztomografie (MRT), Positronen-Emissionstomografie (PET)-Untersuchungen und die Korrelation mit klinischen, genetischen oder molekularen Daten unter dem Einsatz von KI-Methoden wie maschinellem Lernen oder Deep Learning. Die Vorstellung erscheint vielversprechend: Über KI-Methoden können Informationen aus Bilddaten extrahiert werden, die weit über das Maß hinausgehen, das das menschliche Auge erfassen kann. Man geht davon aus, dass die integrative Beurteilung dieser aus den Bilddaten erhobenen Merkmale und Merkmalsmuster – oder „Features“ – mit klinischen, molekularen und genetischen Daten eine genauere Charakterisierung der Pathophysiologie von Erkrankungen sowie eine Aussage über Therapieansprechen und Outcome-Prädiktion ermöglichen kann. Die angewandten Techniken sind teilweise bereits seit Jahrzehnten bekannt, haben sich in den letzten Jahren aber substantiell weiterentwickelt und damit neue Zugänge zur maschinellen Verwertung von Bildinformationen eröffnet. So gehen Publikationen bis auf das Ende der 1940er-Jahre zurück, und Modelle wie neuronale Netze wurden auch in den 1980er-Jahren intensiv beforscht [4]. Optimierte Rechnerleistungen zusammen mit methodischen Fortschritten und zunehmender Verfügbarkeit großer Datenmengen, die das Trainieren von Modellen erleichtern, haben zu einem von beeindruckenden Ergebnissen gezeichneten Wiederaufnahmen dieser Arbeiten geführt [5]. Sie erlauben nun einen (zeit)effizienteren Einsatz dieser Techniken – eine Grundlage für spätere potenzielle klinische Implementierungen. Die Einsatzgebiete in der bildgebenden Diagnostik sind vielfältig und reichen von der onkologischen bis hin zur kardialen und muskuloskelettalen Diagnostik.

Radiomics gewinnt einen immer größeren Stellenwert in der bildgebenden Forschung, da es großes Potenzial birgt, den Erfordernissen der Präzisionsmedizin gerecht zu werden. In mehreren

Arbeiten wird eine Übersicht über die zugrunde liegenden Konzepte gegeben [6, 7]. Nicht vergessen werden darf jedoch, dass jeder einzelne Schritt der Radiomicsanalyse einer großen Variabilität unterliegt. Ein verantwortungsvoller, nachvollziehbarer Umgang mit den eingebrachten Daten und angewandten Analysemethoden ist daher unabdingbare Grundvoraussetzung. Aufgrund des ganz neuen Umgangs mit Bilddaten wird für die Zukunft eine noch engere Kollaboration mit Medical Imaging Computing Data Scientists erforderlich sowie eine Neustrukturierung der radiologischen Ausbildung.

Radiomics, das einen Teilbereich von KI-Implementierungsmöglichkeiten in der Radiologie beschreibt, folgt einem expliziten Schema, nach dem die Bilddaten prozessiert, segmentiert und analysiert werden. Diese Analyse wird im Rahmen dieses Übersichtsartikels dargestellt und erläutert.

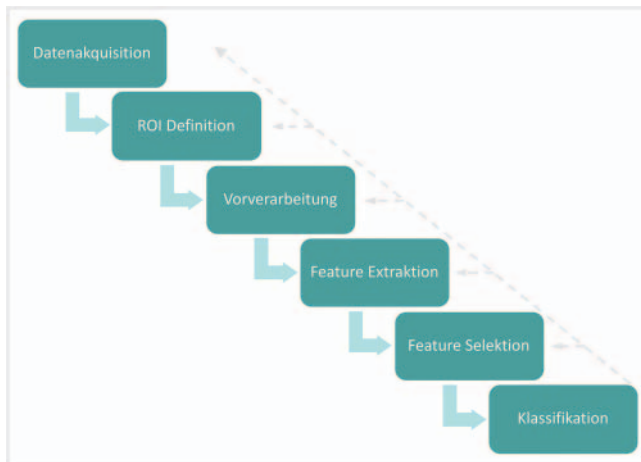
## Radiomics Hands-on

### Die 6 Phasen einer Radiomicsanalyse

Die Radiomicsanalyse lässt sich im Wesentlichen in 6 Schritte unterteilen: (i) die Datenakquisition, (ii) die Definition einer Region of Interest (ROI), (iii) das Daten(Pre-) Processing, (iv) die Feature-Extraktion, (v) die Selektion der für die Fragestellung relevanten Features und (vi) die Klassifikation (► **Abb. 1**) [8].

### Datenakquisition

Die Art und Weise, in der die Daten akquiriert werden, nimmt wesentlich Einfluss auf das Ergebnis der Radiomicsanalyse. Aus diesem Grund ist es erstrebenswert, Bildgebungsprotokolle zu verwenden, die standardisiert, reproduzierbar und vergleichbar sind [9]. So zeigte eine Studie von Waugh et al, dass eine höhere Time to Repetition (TR) eine bessere Diskrimination von Textur-Features in der Mamma-MRT ermöglicht [10]. Baessler et al. testeten in ihrer Publikation systematisch die Einflussfaktoren der Sequenzwahl in der MRT auf die Feature-Robustheit [11]. Die höchste Feature-Robustheit konnte mit einer hochaufgelösten FLAIR-Sequenz erzielt werden. Die im Vergleich akquirierte T2-gewichtete Sequenz mit geringerer Auflösung erzielte dagegen die schlechteste Feature-Stabilität. Auch unter den verschiedenen Feature-Gruppen (Matrices) gab es Unterschiede in der Robustheit. Die Shape- und GLZLM (Grey Level Zone Length Matrix)-Gruppen erzielten die höchste Robustheit, während die histogrammbasierten Features am wenigsten robust waren [11]. Aus diesem Grund fordern Lambin et al. ein stratifiziertes Vorgehen für die Datenselektion: detaillierte Offenlegung der Bildgebungssequenzen, robuste Segmentierung, z. B. durch multiple Auswerter, Phantomstudien und Bildgebung zu verschiedenen Zeitpunkten [9].



► **Abb. 1** Die 6 Phasen einer Radiomicsanalyse. Je nach Teil- oder Gesamtergebnis müssen Teilschritte oder auch die gesamte Analyse wiederholt werden.

## ROI-Definition

### Optimale ROI-Größe und Feature Maps

Nach der Datenakquisition erfolgt die Definition der Region of Interest (ROI), die den Bereich definiert, über den die weitere Analyse erfolgt. Die meisten Arbeiten mit Bezug zu Radiomics befassen sich mit Fragestellungen in der Onkologie, und die ROI wird typischerweise gesetzt, um die Lage einer Läsion zu identifizieren und die darauffolgende Analyse auf die Läsion anzuwenden. Auch dabei gibt es eine große Variabilität in der Methodik der ROI-Definition, die wiederum großen Einfluss auf das Ergebnis hat. Es können drei unterschiedliche ROI-Varianten gewählt werden: eine ROI, die der Kontur der Läsion folgt, eine, die die gesamte Läsion rechtwinklig umfasst (bounding box), und eine partielle ROI, die zentral in einem Ausschnitt der Läsion eingezeichnet wird [8]. Während bounding boxes leichter zu erstellen und oft auch ausreichend sind, erlauben präzise Segmentierungen von Läsionen eine Auswertung von Form und die genauere Analyse von Kontrasten an den Läsionsrändern. Neben der ROI-Form und -Platzierung hat auch die ROI-Größe einen wesentlichen Einfluss auf das Ergebnis. Sikiö et al. zeigten, dass ein Zusammenhang zwischen der Größe der ROI und der Feature-Stabilität besteht [12]. Bei einer räumlichen Auflösung von  $0,5 \times 0,7 \text{ mm}^2$  und einer Schichtdicke von 4 mm war die Feature-Stabilität bei einer  $80 \times 80$  Pixel großen ROI am geringsten; die stabilsten Features wurden mit einer ROI mit  $180 \times 180$  Pixeln erzielt [12].

### Methoden der Segmentierung

Die Segmentierung hat zwei Aufgaben: Erstens kann sie die Analyse spezifischer machen, indem explizit auf Merkmale inner- bzw. außerhalb einer Läsion zugegriffen werden kann; zweitens ist die Form der segmentierten Läsion selbst eine relevante Merkmalsquelle, die damit erschlossen wird. Die Segmentierung von Strukturen in medizinischen Bilddaten ist ein intensiv bearbeitetes Gebiet, und entsprechend gibt es unterschiedliche Möglichkeiten. Einer manuellen Segmentierung stehen automatisierte Segmen-

tierungsmethoden, die spezielle Segmentierungsalgorithmen wie z. B. Region Growing, Level-Sets für ebene Strukturen oder zuletzt auch erfolgreich Methoden des Deep Learning wie zum Beispiel sogenannte U-Nets nutzen [13, 14], gegenüber. Bis dato gilt die manuelle Segmentierung durch einen Expert-Reader als Goldstandard [15]. Problematisch dabei sind jedoch die Inter-Reader-Reliabilität, die Reproduzierbarkeit der durchgeführten Segmentierung und der Zeitaufwand, den das manuelle Segmentieren großer Datenmengen erfordert [16, 17]. Lambin et al. empfehlen zur Reduktion dieses Bias die Anfertigung multipler Segmentierungen, Multi-Reader-Analysen, den Ausschluss von Segmentierungen mit hohem Rauschen und die Verwendung von Daten unterschiedlicher Atemzyklen [9]. Grundsätzlich kann abhängig von den verfügbaren Daten Segmentierung sowohl in 2D- als auch in 3D-Bilddaten durchgeführt werden. Während eine 2D-Analyse weniger Differenzierung in der Form von Läsionen zulässt, ist sie unabhängiger von oft sehr variablen Aufnahmeparametern wie der Schichtdicke.

### Image Processing bzw. Pre-Processing

Vor der eigentlichen Feature-Extraktion steht das Image Pre-Processing. Dazu gehören je nach Datensatz die Interpolation, die Signalnormierung und die Grauwertreduktion.

Durch die Interpolation der Bilddaten kann die räumliche Auflösung in Daten vereinheitlicht werden. Studien zeigten, dass eine höhere Auflösung eine optimierte Feature-Extraktion erlaubt. In einer Arbeit von Mayerhoefer et al. wurden die besten Ergebnisse mit Interpolationsfaktoren von 2–4 erzielt [1]. Für die Signalnormierung ist die Vergleichbarkeit der im Rahmen der Analyse gewonnenen Merkmale relevant. In der Literatur werden im Wesentlichen drei Ansätze beschrieben: min/max, der Z-Score und  $\text{mean} \pm 3\sigma$  [18]. Die Methode „ $\text{mean} \pm 3\sigma$ “ bedeutet, dass die Intensitäten normalisiert werden innerhalb von  $\mu \pm 3\sigma$ , wobei  $\mu$  den Mittelwert der Grauwerte innerhalb der ROI beschreibt und  $\sigma$  die Standardabweichung. In Konsequenz werden Grauwerte, die sich außerhalb der Range  $[\mu - 3\sigma, \mu + 3\sigma]$  befinden, nicht für die Analyse berücksichtigt.

Die Reduktion der Grauwerte in der Form des sogenannten „Binnings“ bei der Merkmalsextraktion führt zu einer Verbesserung des Signal-zu-Rauschen(SNR)-Verhältnisses. Es bildet den im Bild vorkommenden Grauwertbereich in der Form von Häufigkeitsverteilungen ab. Grauwerte, die in der Literatur verwendet werden, sind 16, 32, 64, 128 und 256. Chen et al. empfehlen in ihrer Arbeit die Verwendung von 32 Grauwerten [19], Mahmoud-Ghonheim et al. empfehlen 128 [20].

Das Image Pre-Processing hat einen wesentlichen Einfluss auf die Feature-Robustheit. Wichtmann et al. untersuchten systematisch am Phantom den Einfluss der räumlichen Auflösung, der Grauwertreduktion und der Signalnormierung auf die Feature-Robustheit [21]. Sie konnten zeigen, dass nur 4 Features, Skewness (Histogramm), Volume [ml] (Shape), Volume [vox] (Shape) und run length nonuniformity [RLNU] (Grey Level Run Length Matrix, GLRLM), über die Variation aller Parameter robust blieben. Dies zeigt eindrücklich, dass dezidierte Empfehlungen für das Bildprocessing erforderlich sind.

► **Tab. 1** Überblick über die Features der einzelnen Matrizen [21].

First Order Statistics Features	Shape and size based Features	Textural Features	Wavelet Features
		<b>Grey-level Co-occurrence matrix based</b>	
Energy	Compactness	Autocorrelation	
Entropy	Maximum 3D Diameter	Cluster prominence	
Kurtosis	Spherical disproportion	Gray-level Run-length matrix based	
Maximum	Sphericity	Gray level non uniformity	
Mean	Surface area	Run Length Non uniformity	

### Feature-Extraktion

Features, die typischerweise für Radiomicsanalysen verwendet werden, können in 4 übergeordnete Gruppen eingeteilt werden: First-Order Statistics, Shape- und Texture-Features sowie Features, die mithilfe der Wavelet-Transformation relevanter Bildausschnitte gewonnen werden [16]. Zu der Gruppe der Texture Feature Matrices zählen die Matrices Grey Level Co-occurrence Matrix (GLCM), GLRLM, Grey Level Size Zone Matrix (GLSZM), Grey Level Dependence Matrix (GLDM) und Neighbouring Grey Tone Difference Matrix (NGTDM). Unter jeder dieser Matrices subsumieren sich wiederum multiple Features. Zu beachten ist, dass es eine große Variation in Nomenklatur, Methodik und Softwareimplementierung gibt [22]. Einen exemplarischen Überblick über Features der einzelnen Matrizes gibt ► **Tab. 1** [23]. Gleichzeitig finden Bemühungen zur Invarianz von Features gegenüber Protokollen und entsprechende Standardisierungsbestrebungen statt [24]. In die Auswahl von Merkmalsextraktoren fließt die Erwartung ein, welche Charakteristika für die Analyse relevant sind, und entsprechend werden oft Extraktoren gewählt oder konstruiert, die zum Beispiel gegenüber globaler Rotation oder sehr niederfrequenten Grauwertänderungen invariant sind.

Baessler et al. zeigten eindrucksvoll die diagnostische Aussagekraft von Texture-Features für die Diagnose von Myokarditiden mit dem MRT. In ihrer Arbeit konnten die Texture-Features in gleicher Weise wie herkömmliche MRT-Parameter Patienten mit einer bioptisch gesicherten Myokarditis von einer gesunden Kontrollgruppe differenzieren. Die herkömmlichen MRT-Parameter erlaubten allerdings im Gegensatz zu den Texture-Features keine Differenzierung zwischen gesunder Kontrollgruppe und Patienten mit negativer Biopsie, jedoch klinischem Verdacht einer Myokarditis. Da ergab sich lediglich für die Texture-Features, v. a. RLNU und Grey Level Non-Uniformity, ein statistisch signifikanter Unterschied [25]. Radiomics erlaubte im Vergleich zum derzeitigen

Standard eine genauere diagnostische Differenzierung zwischen Patienten mit einer Myokarditis und gesunder Kontrollgruppe.

### Feature-Selektion

Ein wesentliches Problem in der Radiomicsanalyse ist die Gefahr eines Overfittings der Daten. Diese tritt insbesondere auf, wenn die Zahl der Features die Zahl der Datensätze übersteigt. Dadurch wird die Aussagekraft der Analyse wesentlich eingeschränkt. Das Overfitting kann durch eine Reduktion der Dimensionalität vermieden werden, d. h. durch die Selektion von Features, die für die Analyse und Vorhersage verwendet werden. Das kann auf zwei Grundlagen erfolgen: Features, die reproduzierbar, robust, und nicht redundant sind, können ohne Wissen der Zielfragestellung ausgewählt werden und erlauben Feature-Reduktion ohne Bias [8, 16]. Feature-Auswahl basierend darauf, wie „informativ“, d. h. relevant ein Feature im Sinne der Fragestellung ist, stellt zwar eine effektive Strategie dar, birgt aber ebenfalls die Gefahr des Overfittings. Methoden aus dem maschinellen Lernen wie etwa Random Forests erlauben eine effektive Auswahl informativer Features bei gleichzeitiger Robustheit gegenüber großer Mengen nicht informativer Merkmale [26]. In diesem Fall ist aber, wie unten beschrieben, unbedingt eine Evaluation der schließlich resultierenden Vorhersagegenauigkeit auf einem unabhängigen Validierungsdatensatz notwendig, der weder zum Trainieren des Modells noch zur Auswahl der Features verwendet wurde [27].

Zur Testung der Stabilität von Features können Test-Re-Test-Datensätze verwendet werden. Nur die stabilen Features werden dann für die weitere Analyse eingesetzt. Zur Testung von Robustheit und Reproduzierbarkeit eignen sich der Concordance Correlation Coefficient (CCC), die Dynamic Range (DR) und der „Correlation Coefficient Across All Samples“. Arbeiten zeigten, dass die Zahl der Features damit wesentlich, z. B. von 397 auf 39, reduziert werden kann [16]. Darüber hinaus können Intra- und Inter-Observer-Variabilität durch den Intraclass Correlation Coefficient (ICC) und Bland-Altman-Plots getestet werden. Neben den hier angeführten statistischen Methoden können auch Verfahren des maschinellen Lernens wie etwa Random Forests zur Identifikation von relevanten Features für die Beantwortung der Fragestellung, z. B. die Differenzierung von gutartig/bösartig, eingesetzt werden.

### Klassifikation/Modelling

Am weitesten verbreitet sind derzeit Supervised-learning-Ansätze, d. h., eine Maschine wird anhand von Trainingsdatensätzen unter Kenntnis des Input-Vektors (Features) und des Output-Werts (Target) trainiert. Nach diesem Training wird der trainierte Algorithmus auf einen Testdatensatz angewendet. An dieser Stelle werden die extrahierten Merkmale zur Vorhersage verwendet, wobei eine Schlüsseleigenschaft der relevanten Methoden wie Support Vector Machines oder Random Forests ist, dass sie nicht nur den Zusammenhang zwischen isolierten Merkmalen und dem Vorhersageziel auswerten, sondern Merkmalgruppen als *multivariate Muster* ausnutzen können. An diesem Punkt findet derzeit auch ein sehr schneller Fortschritt statt, der, ermöglicht durch Techniken des Deep Learning, die Konstruktion von Features, ihre Auswahl und die Vorhersage zunehmend in gemeinsamen Modellen zusammenfasst.

## Validierung

Als letzter Schritt erfolgt die Validierung an einem Validierungsdatensatz. Mittels ROC/AUC(receiver operating characteristic/area under the curve)-Analyse wird die prädiktive Performance des Algorithmus getestet [28]. Wesentlich ist die Trennung zwischen Daten, die für das Training bzw. die Entwicklung der Vorhersagemodelle und die Auswahl von Features verwendet wurden, und solchen, die als Validierungsdaten dienen. Das ist notwendig, um eine überoptimistische Einschätzung der Vorhersagegenauigkeit sicherzustellen. Als Mittelweg kann eine sogenannte Kreuzvalidierung eingesetzt werden, bei der iterativ Training- und Testdatensatz separiert werden. Dabei muss in Betracht gezogen werden, dass die jeweiligen Testdaten Grundlage für Modellierungsentscheidungen sein konnten und damit keine gänzlich unabhängige Beurteilung zulassen – dazu ist ein getrennter Validierungsdatensatz notwendig.

Parmar et al. haben in ihrer Arbeit die Stabilität und die prädiktive Performance verschiedener Feature-Selektions- und Klassifikatormethoden getestet [28]. Ihre Ergebnisse zeigten, dass unter den unterschiedlichen Feature-Selektionsmethoden Wilcoxon test-based method (WLCX) und Mutual information maximization (MIM) die größte Stabilität erzielten. Unter den Klassifikatoren erzielte Bayesian die beste Performance mit einem AUC-Wert von 0,64 (SD ± 0,05).

Aufgrund der großen Variabilität der Radiomicsanalyse ist eine Standardisierung der Datenkollektion, der Evaluationskriterien und der Berichterstattung erforderlich. Dazu definieren Lambin et al. einen „Radiomics Quality Score (RQS)“ [9]. Der RQS beschreibt einen standardisierten Analyseablauf, beginnend mit der Datenselektion über die Bildgebung, die Feature Extraction, die Analyse und das Modelling sowie die Berichterstellung. Jeder dieser Schritte gliedert sich in mehrere Unterschritte, für die es jeweils Punkte gibt. Der Maximum Score (RQS Total), der erzielt werden kann, beträgt 36. Definition und Einführung eines RQS sind ein wesentlicher Schritt in die Richtung einer qualitätsgesicherten Anwendung von Radiomics in der Medizin, der der Variabilitätsproblematik der Analyse – die bereits bei der primären Bild-datenakquisition beginnt – durch ein dezidiertes Reporting der Einzelschritte begegnen will. Die Einführung eines RQS erscheint besonders relevant im Hinblick auf die erwartete künftige Konnektion von klinischen Decision-Support-Systemen mit Radiomicsdaten [9].

## Wohin geht die Reise?

Neben *klassischen* Radiomicsansätzen, die vordefinierte Features nutzen, nimmt die durch die rezente Entwicklung im Gebiet des Deep Learning entstandene Möglichkeit, Feature-Konstruktion und Vorhersagemodell-Training zu verknüpfen und mit effektiven Model-Architekturen umzusetzen, eine immer wichtigere Rolle in der Nutzung komplexer Bilddaten ein [7, 29]. Einerseits wird damit die Nutzung von Bildinformation möglich, die nicht durch die traditionellen Features abgedeckt wird, andererseits ergibt sich die Problematik der Interpretation von Deep-Learning-Modellen, deren Lösung zunehmend in den Fokus der Forschung gerät [30].

## Fazit

Radiomics nimmt einen immer größeren Stellenwert in der medizinischen Bildgebung ein, da sie großes Potenzial birgt, den Erfordernissen der Präzisionsmedizin gerecht zu werden. Nicht vergessen werden darf jedoch, dass jeder einzelne Schritt der Radiomicsanalyse einer großen Variabilität unterliegt. Ein verantwortungsvoller, nachvollziehbarer Umgang mit den eingebrachten Daten ist daher unabdingbare Grundvoraussetzung. Radiomics wird für die Zukunft eine noch engere Kollaboration mit Medical Imaging Computing Data Scientists erfordern sowie eine Neustrukturierung der radiologischen Ausbildung.

## Interessenkonflikt

Die Autorinnen/Autoren geben an, dass kein Interessenkonflikt besteht.

## Literatur

- [1] Mayerhoefer ME, Szomolanyi P, Jirak D et al. Effects of magnetic resonance image interpolation on the results of texture-based pattern classification: a phantom study. *Invest Radiol* 2009; 44: 405–411. doi:10.1097/RLI.0b013e3181a50a66
- [2] European Society of R. What the radiologist should know about artificial intelligence – an ESR white paper. *Insights Imaging* 2019; 10: 44 doi:10.1186/s13244-019-0738-2
- [3] Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2016; 278: 563–577. doi:10.1148/radiol.2015151169
- [4] LeCun Y, Boser B, Denker JS et al. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1989; 1: 541–551. doi:10.1162/neco.1989.1.4.541
- [5] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Lake Tahoe, Nevada: Curran Associates Inc. 2012
- [6] Mayerhoefer ME, Materka A, Langs G et al. Introduction to Radiomics. *J Nucl Med* 2020; 61: 488–495. doi:10.2967/jnumed.118.222893
- [7] Langs G, Rohrich S, Hofmanninger J et al. Machine learning: from radiomics to discovery and routine. *Radiologe* 2018; 58: 1–6. doi:10.1007/s00117-018-0407-3
- [8] Larroza A, Bodí V, Moratal D. Texture analysis in magnetic resonance imaging: review and considerations for future applications. *Assessment of cellular and organ function and dysfunction using direct and derived MRI methodologies*. 2016: 75–106
- [9] Lambin P, Leijenaar RTH, Deist TM et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology* 2017; 14: 749–762. doi:10.1038/nrclinonc.2017.141
- [10] Waugh SA, Lerski RA, Bidaut L et al. The influence of field strength and different clinical breast MRI protocols on the outcome of texture analysis using foam phantoms. *Med Phys* 2011; 38: 5058–5066. doi:10.1118/1.3622605
- [11] Baessler B, Weiss K, Pinto Dos Santos D. Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study. *Invest Radiol* 2019; 54: 221–228. doi:10.1097/RLI.0000000000000530
- [12] Sikiö M, Holli-Helenius KK, Ryymin P et al. The effect of region of interest size on textural parameters. In, 2015 9th International Symposium on Image and Signal Processing and Analysis (ISPA): IEEE 2015: 149–153
- [13] Brox T, Bruhn A, Weickert J. Variational motion segmentation with level sets. In, *European Conference on Computer Vision*: Springer 2006: 471–483

- [14] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In, International Conference on Medical image computing and computer-assisted intervention: Springer 2015: 234–241
- [15] Raykar VC, Yu S, Zhao LH et al. Supervised learning from multiple experts: whom to trust when everyone lies a bit. In, Proceedings of the 26<sup>th</sup> Annual international conference on machine learning 2009: 889–896
- [16] Kumar V, Gu Y, Basu S et al. Radiomics: the process and the challenges. *Magn Reson Imaging* 2012; 30: 1234–1248. doi:10.1016/j.mri.2012.06.010
- [17] Watadani T, Sakai F, Johkoh T et al. Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology* 2013; 266: 936–944. doi:10.1148/radiol.12112516
- [18] Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging* 2004; 22: 81–91. doi:10.1016/j.mri.2003.09.001
- [19] Chen W, Giger ML, Li H et al. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magn Reson Med* 2007; 58: 562–571. doi:10.1002/mrm.21347
- [20] Mahmoud-Ghoneim D, Alkaabi MK, de Certaines JD et al. The impact of image dynamic range on texture classification of brain white matter. *BMC Med Imaging* 2008; 8: 18 doi:10.1186/1471-2342-8-18
- [21] Wichtmann BD, Attenberger UI, Harder FN et al. Influence of image processing on the robustness of radiomic features derived from magnetic resonance imaging – a phantom study. 2019 ps://submissions.miramart.com/ISMRM2019/ViewSubmission.aspx?sbmlID=7334
- [22] Aerts HJ, Velazquez ER, Leijenaar RT et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications* 2014; 5: 1–9
- [23] Aerts HJ. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol* 2016; 2: 1636–1642. doi:10.1001/jamaoncol.2016.2631
- [24] Andrearczyk V, Depeursinge A, Müller H. Learning cross-protocol radiomics and deep feature standardization from CT images of texture phantoms. In, *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications: International Society for Optics and Photonics* 2019: 109540I
- [25] Baessler B, Luecke C, Lurz J et al. Cardiac MRI Texture Analysis of T1 and T2 Maps in Patients with Infarctlike Acute Myocarditis. *Radiology* 2018; 289: 357–365. doi:10.1148/radiol.2018180411
- [26] Breiman L. Random forests. *Machine Learning* 2001; 45: 5–32. doi:10.1023/A:1010933404324
- [27] Guyon I, Elisseeff A. An introduction to variable and feature selection. *Journal of machine learning research* 2003; 3: 1157–1182
- [28] Parmar C, Grossmann P, Bussink J et al. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep* 2015; 5: 13087 doi:10.1038/srep13087
- [29] Litjens G, Kooi T, Bejnordi BE et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017; 42: 60–88. doi:10.1016/j.media.2017.07.005
- [30] Holzinger A, Langs G, Denk H et al. Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery* 2019; 9: e1312 doi:10.1002/widm.1312