

Radiomics Analysis of Multiparametric PET/MRI for N- and M-Staging in Patients with Primary Cervical Cancer

Radiomics-Analyse anhand der multiparametrischen PET/MRT für das N- und M-Staging von Patientinnen mit primärem Zervixkarzinom

Authors

Lale Umutlu¹, Felix Nensa¹, Aydin Demircioglu¹, Gerald Antoch², Ken Herrmann³, Michael Forsting¹, Johannes Stefan Grueneisen¹

Affiliations

- 1 Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Essen, University of Duisburg-Essen, D-45147 Essen, Germany
- 2 Department of Diagnostic and Interventional Radiology, University Dusseldorf, Medical Faculty, D-40225 Dusseldorf, Germany
- 3 Department of Nuclear Medicine, University Hospital Essen, University of Duisburg-Essen, D-45147 Essen, Germany

Key words

machine learning, multiparametric PET/MRI, cervical cancer, radiomics analysis

received 28.08.2019

accepted 05.01.2020

Bibliography

DOI <https://doi.org/10.1055/a-1100-0127>

Published online: 30.4.2020

Fortschr Röntgenstr 2020; 192: 754–763

© Georg Thieme Verlag KG, Stuttgart · New York

ISSN 1438-9029

Correspondence

Dr. Lale Umutlu

Institut für Diagnostische und Interventionelle Radiologie und Neuroradiologie, Universitätsklinik Essen, Hufelandstr. 55, 45122 Essen, Germany

Tel.: ++49/2 01/72 38 45 27

Fax: ++49/2 01/7 23 15 63

lale.umutlu@uk-essen.de

ABSTRACT

Purpose The aim of this study was to investigate the potential of multiparametric ¹⁸F-FDG PET/MR imaging as a platform for radiomics analysis and machine learning algorithms based on primary cervical cancers to predict N- and M-stage in patients.

Materials and Methods A total of 30 patients with histopathological confirmation of primary and untreated cervical cancer were prospectively enrolled for a multiparametric ¹⁸F-FDG PET/MR examination, comprising a dedicated proto-

col for imaging of the female pelvis. The primary tumor in the uterine cervix was manually segmented on post-contrast T1-weighted images. Quantitative features were extracted from the segmented tumors using the Radiomic Image Processing Toolbox for the R software environment for statistical computing and graphics. 45 different image features were calculated from non-enhanced as well as post-contrast T1-weighted TSE images, T2-weighted TSE images, the ADC map, the parametric Ktrans, Kep, Ve and iAUC maps and PET images, respectively. Statistical analysis and modeling was performed using Python 3.5 and the scikit-learn software machine learning library for the Python programming language.

Results Prediction of M-stage was superior when compared to N-stage. Prediction of M-stage using SVM with SVM-RFE as feature selection obtained the highest performance providing sensitivity of 91 % and specificity of 92 %. Using receiver operating characteristic (ROC) analysis of the pooled predictions, the area under the curve (AUC) was 0.97. Prediction of N-stage using RBF-SVM with MIFS as feature selection reached sensitivity of 83 %, specificity of 67 % and an AUC of 0.82.

Conclusion M- and N-stage can be predicted based on isolated radiomics analyses of the primary tumor in cervical cancers, thus serving as a template for noninvasive tumor phenotyping and patient stratification using high-dimensional feature vectors extracted from multiparametric PET/MRI data.

Key points:

- Radiomics analysis based on multiparametric PET/MRI enables prediction of the metastatic status of cervical cancers
- Prediction of M-stage is superior to N-stage
- Multiparametric PET/MRI displays a valuable platform for radiomics analyses

ZUSAMMENFASSUNG

Zielsetzung Ziel dieser Studie war die Evaluierung des prädiktiven Potenzials der Radiomics-Analyse zur Bestimmung des N- und M-Stadiums des primären Zervixkarzinoms anhand multiparametrischer ¹⁸F-FDG-PET/MRT-Bildgebung.

Material und Methoden 30 Patientinnen mit einem histologisch gesicherten, primären und therapienaiven Zervixkarzinom unterzogen sich einer multiparametrischen ¹⁸F-FDG-

PET/MRT-Untersuchung unter Verwendung eines dedizierten Untersuchungsprotokolls des weiblichen Beckens. Nach Segmentierung der Primärtumoren wurden quantitative Bildparameter mittels der Radiomic-Image-Processing-Toolbox bestimmt. Insgesamt wurden 45 verschiedene quantitative Bildmerkmale jeweils anhand der T2-gewichteten TSE-Sequenzen, der nativen und kontrastmittelgestützten T1-gewichteten TSE-Sequenzen, der ADC-Map, verschiedenen Perfusionsparametern (Ktrans, Kep, Ve and iAUC) und den ¹⁸F-FDG-PET-Datensätzen für jeden Tumor extrahiert. Die statistische Analyse zur Bestimmung des N- und M-Stadiums erfolgte unter der Verwendung der Python 3.5 und Scikit-learn-Software-Bibliothek für maschinelles Lernen.

Ergebnisse Insgesamt zeigte sich eine höhere Genauigkeit zur Prädiktion des korrekten M-Stadiums im Vergleich zum N-Stadium. Zur Prädiktion des korrekten M-Stadiums zeigten sich unter der Verwendung von SVM und SVM-RFE zur Feature-Auswahl die besten Ergebnisse mit einer Sensitivität von 91 %, einer Spezifität von 92 % und einer Fläche unter der Kurve (AUC) von 0,97. Die höchste Genauigkeit für die Bestimmung des N-Stadiums erfolgte unter der Verwendung von RBF-SVM und MIFS zur Feature-Auswahl mit einer Sensi-

vität von 83 %, einer Spezifität von 67 % und einer Fläche unter der Kurve (AUC) von 0,82.

Schlussfolgerung Die Radiomics-Analyse von multiparametrischen PET/MR-Datensätzen ermöglicht eine präzise Prädiktion des M- und N-Stadiums von Patientinnen mit primärem Zervixkarzinom und könnte damit supportiv zur nichtinvasiven Tumor-Phänotypisierung und Patientenstratifizierung eingesetzt werden.

Kernaussagen:

- Die Radiomics-Analyse der multiparametrischen PET/MRT ermöglicht die Prädiktion des Metastasierungsstatus des Zervixkarzinoms.
- Die Prädiktion des M-Stadiums ist der Prädiktion des N-Stadiums überlegen.
- Die multiparametrische PET/MRT bietet eine valide Plattform für Radiomics-Analysen.

Citation Format

- Umutlu L, Nensa F, Demircioglu A et al. Radiomics Analysis of Multiparametric PET/MRI for N- and M-Staging in Patients with Primary Cervical Cancer. *Fortschr Röntgenstr* 2020; 192: 754–763

Introduction

Until recently, precision medicine by means of personalized cancer care was premised on the molecular characterization of tumor tissue derived from invasive procedures, such as biopsy sampling. Lately, a two-step combination comprising large-scale extraction of quantitative features from image data and machine learning has been proposed as a noninvasive method of phenotyping in solid tumors. This technique, now commonly referred to as “radiomics”, has the potential to complement current diagnostic workup with new imaging biomarkers for improved cancer detection, diagnosis, prediction of prognosis and treatment response [1, 2]. Bound to the nature of radiomics analysis in collecting an extensive volume of data, similar to genomic analyses, the application of machine learning algorithms is inevitable to process the large number of features [3]. These kinds of computational analysis methods are vital to assess the diagnostic and predictive potential of radiomics analysis and further establish it in clinical precision oncology.

Whereas previous investigations on radiomics analyses mostly put the focus on computer tomography (CT)-derived datasets [4–6], multiparametric positron emission tomography/magnetic resonance imaging (PET/MRI) facilitates the simultaneous acquisition of a significant number of complementary morphologic, metabolic and functional tissue parameters (► **Fig. 1**), providing a favorable acquisition platform for comprehensive radiomics signatures [7, 8]. Recent publications evaluated the use of radiomics analysis on sequentially obtained PET and MR data and reported promising results for tissue characterization and for predicting the prognosis of oncological patients [9–12]. Furthermore, recent studies on radiomics analysis on hybrid imaging revealed its potential as a promising imaging surrogate for N-sta-

ging, which may be of clinical relevance for patient stratification and in the case of unclear imaging findings to reduce unnecessary invasive sampling [13]. Hence, the aim of this study was to investigate the potential of PET/MR imaging as a platform for tumor phenotyping using radiomics and machine learning in patients with cervical cancer and to predict lymph node and distant metastases, in terms of N- and M-stage, based solely on the analysis of the primary cancer.

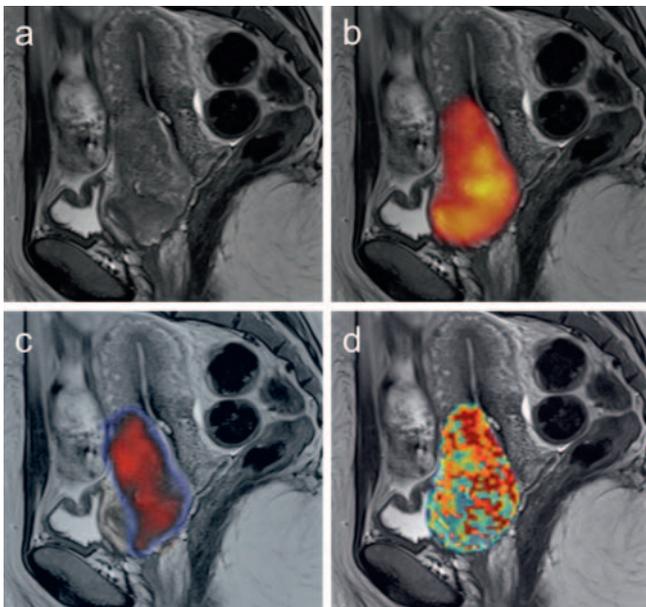
Materials and Methods

Patients

Ethical approval was granted by the local ethics committee as part of fundamental research on integrated PET/MRI (registry number: 11–4825-BO). A total of 30 consecutive patients (mean age: 48 ± 15 years) with therapy-naive, histopathologically confirmed primary cancer of the uterine cervix were prospectively enrolled between August 2014 and January 2017. All patients underwent simultaneous PET/MR imaging after written and informed consent was obtained and prior to the initiation of definitive treatment. Patient characteristics are shown in ► **Table 1**.

PET/MR imaging data acquisition and pre-processing

PET/MR examinations were performed in supine position on a 3 Tesla Biograph mMR scanner (Siemens Healthineers, Germany). Pelvic PET and MR data were acquired simultaneously in one bed position. All patients were instructed to fast for a period of at least 6 hours prior to the start of each examination to ensure adequate blood glucose levels (below 150 mg/dl). Data acquisition started with a mean



► **Fig. 1** PET/MR examination of a 47-year-old patient with a primary and untreated squamous cell cancer of the uterine cervix. Quantitative multiparametric analysis of simultaneously acquired morphological (**a**, T2w TSE) as well as metabolic (**b**, image fusion with ^{18}F -FDG PET data) and MR-derived functional parameters (image fusion with: **c**, inverted and color-coded ADC values; **d**, color-coded perfusion parameters (iAUC)).

► **Abb. 1** PET/MR-Untersuchung einer 47-jährigen Patientin mit einem primären unbehandelten Plattenepithelkarzinom des Gebärmutterhalses. Quantitative multiparametrische Analyse von simultan akquirierten morphologischen (**a**, T2w-TSE), metabolischen (**b**, Bildfusion mit ^{18}F -FDG-PET-Daten) und MR-basierten funktionellen Parametern (Bildfusion mit: **c**, invertierten und farbcodierten ADC-Werten; **d**, farbcodierten Perfusionsparametern (hier: iAUC)).

delay of 67 ± 16 min after a body weight-adapted dosage (2 MBq/kg bodyweight) of ^{18}F -fluorodeoxyglucose (^{18}F -FDG) was administered intravenously (mean activity of $136 \pm 36 \text{ MBq}$). For PET image reconstruction, the iterative ordered-subset expectation maximization algorithm was used, 3 iterations and 21 subsets, Gaussian filter with 4 mm full width at half maximum and a 344×344 image matrix. Attenuation correction of the PET datasets was performed automatically. Therefore, a four-compartment-model attenuation map (μ -map) was calculated based on fat-only and water-only datasets, as obtained by Dixon-based sequences. MR datasets of the female pelvis were acquired with a dedicated phased-array body surface coil. The dedicated MR imaging protocol as well as sequence parameters are displayed in ► **Table 2**. For dynamic contrast-enhanced (DCE) MR imaging, two sagittal pre-contrast T1w VIBE sequences with flip angles of 2 and 15 degrees were acquired. Subsequently, a body weight-adapted dosage of paramagnetic contrast agent (0.1 mmol/kg bodyweight Gadobutrol, Bayer Healthcare, Germany) was intravenously injected, followed by the acquisition of repetitive fat saturated T1w VIBE sequences (114 measurements).

A commercial software application (Tissue 4D; Siemens Healthineers, Germany) was used for pharmacokinetic modeling of dynamic contrast-enhanced (DCE), perfusion-weighted MRI data-

sets. In order to create DCE-MRI parameter maps with Tissue 4D, the following steps were performed: First, motion artifacts were reduced using elastic 3D motion correction. Then, T1 mapping was performed using two native VIBE acquisitions with varying flip angles (2° and 15°). Finally, pharmacokinetic models were calculated on a voxel-by-voxel basis using the one-compartment Tofts model with a simulated intermediate flow arterial input function. The following pharmacokinetic parameters were calculated and exported as DICOM parameter maps: Transfer constant of capillary permeability (K^{trans}), reflux constant (K_{ep}), extravascular extracellular volume fraction (V_e) and the initial area under the curve (iAUC) for the first 60 seconds.

Real-time calculation of trace-weighted images and apparent diffusion coefficient (ADC) maps from single-shot diffusion-weighted EPI sequences was automatically performed by the scanner system.

PET image units were converted to standardized uptake values (SUV) using the OsiriX MD medical image viewer. T1- and T2-weighted MR images were neither normalized nor rescaled.

Tumor segmentation

PET/MRI datasets were imported into the Fiji open source image processing package [14], using the ImageJ interface to the Orthanc DICOM server [15]. The primary gross tumor volume (GTV) was manually outlined by two experienced radiologists in consensus using Fiji on sagittal VIBE MR sequences 20 seconds after i.v. contrast agent injection. The delineated GTV was exported using voxel-wise binary image masks (0: off mask, 1: on mask). Original MR sequences, DCE-MRI parameter maps, ADC maps and reconstructed PET image data were resampled using the VIBE sequence as the reference in order to create one-to-one mapping of voxel coordinates between all image volumes and the binary GTV mask.

Image feature extraction

Quantitative imaging features were extracted using the Radiomic Image Processing Toolbox for R [16, 17]. A total of 45 different image features were calculated from non-enhanced as well as post-contrast T1-weighted TSE images, T1-weighted perfusion images, T2-weighted TSE images, ADC maps, parametric K^{trans} , K_{ep} , V_e and iAUC maps and PET images, respectively, totaling 450 features. These features included standard first order statistics (e.g. standard deviation, mean and energy), features derived from the gray level co-occurrence matrix (e.g. contrast, correlation dissimilarity and inverse difference moment) and features derived from the gray level run length matrix (e.g. gray level non-uniformity, long run emphasis and run percentage). The complete list of image features is documented in the Radiomic Image Processing Toolbox [16]. Images were discretized by binning the voxels into a fixed number of 32 bins. Furthermore, the overall longest diameter of the tumor – as performed in both the clinical routine and RECIST tumor assessment – was added to the image features.

Standard of reference

Histopathological verification of the primary cervical cancer was available in all 30 patients. As clinical evaluation is the primary and guideline-recommended staging approach for cervical cancer

► **Table 1** Characteristics of patients enrolled in this trial.

► **Tab. 1** Auflistung der Patientencharakteristiken.

patient	age	histology	tumor grade	T-stage	N-stage	M-stage	FIGO – stage	size (mm)
1	39	squamous cell carcinoma	G2	T1b	N0	M0	IB2	31
2	56	squamous cell carcinoma	G3	T3a	N1	M1	IIIC2	72
3	58	adenocarcinoma	G2	T2b	N0	M0	IIB	25
4	34	squamous cell carcinoma	G2	T1b	N0	M0	IB3	44
5	28	squamous cell carcinoma	G1	T1b	N0	M0	IB1	17
6	29	squamous cell carcinoma	G2	T1b	N0	M0	IB3	46
7	57	squamous cell carcinoma	G3	T4	N1	M1	IVB	56
8	71	squamous cell carcinoma	G2	T3b	N1	M0	IIIC1	48
9	40	squamous cell carcinoma	G2	T2a	N1	M0	IIIC1	52
10	60	adenocarcinoma	G3	T4	N1	M0	IVA	56
11	68	squamous cell carcinoma	G2	T1b	N0	M0	IB2	38
12	54	squamous cell carcinoma	G2	T2b	N1	M1	IIIC2	82
13	73	adenocarcinoma	G3	T1b	N0	M0	IB2	34
14	32	adenocarcinoma	G3	T1b	N0	M0	IB3	63
15	64	adenocarcinoma	G3	T3b	N1	M1	IVB	50
16	68	squamous cell carcinoma	G2	T2b	N0	M0	IIB	43
17	70	squamous cell carcinoma	G3	T2b	N1	M0	IIIC1	62
18	51	adenocarcinoma	G2	T1b	N1	M0	IIIC1	60
19	30	squamous cell carcinoma	G2	T1b	N0	M0	IB1	19
20	64	squamous cell carcinoma	G3	T1b	N0	M0	IB3	52
21	61	squamous cell carcinoma	G3	T2b	N1	M1	IIIC2	35
22	39	squamous cell carcinoma	G3	T1b	N1	M0	IIIC1	39
23	28	squamous cell carcinoma	G3	T2a	N1	M1	IIIC2	59
24	28	squamous cell carcinoma	G2	T2a	N1	M0	IIIC1	43
25	63	squamous cell carcinoma	G3	T3b	N1	M1	IVB	73
26	52	squamous cell carcinoma	G2	T1b	N0	M0	IB1	13
27	36	squamous cell carcinoma	G2	T1b	N0	M0	IB2	27
28	38	adenocarcinoma	G3	T2a	N1	M0	IIIC1	53
29	39	squamous cell carcinoma	G3	T1b	N0	M0	IB3	48
30	48	squamous cell carcinoma	G3	T1b	N0	M0	IB3	45

patients and histopathological verification of all suspect lymph node or distant metastases is not obligatory to initiate treatment [18, 19], a modified reference standard was chosen as defined in previous publications [20]. A consensus interpretation among a radiologist and nuclear medicine specialist was performed for each patient, taking into account the results of pelvic and/or para-aortic lymphadenectomy (available in 25/30 patients), results of pelvic and whole-body PET/MRI examinations, clinical staging results as well as findings in corresponding follow-up examinations. In this context, a suspicious lesion was considered as malignant when (i) a lesion disappeared or decreased in size and/or revealed a decreasing ¹⁸F-FDG accumulation under systemic therapy as

well as (ii) an increase/decrease in the number and size could be determined in subsequent examinations. Conversely, morphologically inconspicuous and PET-negative lesions in PET/MRI and imaging follow-up examinations were considered benign.

Statistical analysis

Statistical analysis and modeling was performed using radiomics analysis. The dimensionality of the feature space was reduced by five different methods [3]: Analysis of variance (Anova), Support vector machine recursive feature elimination (SVM-RFE), Mutual information (MIFS), t-score and Wilcoxon tests. The 20 highest

► **Table 2** MR sequence parameters (cor: coronal, sag: sagittal; ax: axial).

► **Tab. 2** MRT Sequenzparameter (cor: koronar, sag: sagittal, ax: axial).

	slice thickness (mm)	acquisition time	repetition time/echo time (msec)	field of view (mm)	phase FoV (%)	matrix size
T1w VIBE Dixon cor.	3.12	13 sec	3.6/1.23 (1 st) and 2.46 (2 nd)	500	65.6	192 × 79
T2w TSE ax.	4	2 min 33 sec	5820/114	400	71.9	512 × 192
T2w TSE sag.	4	4 min 08 sec	4930/114	300	71.9	512 × 240
T1w TSE ax.	4	1 min 22 sec	495/12	400	71.9	512 × 230
DW EPI (b-values: 0, 500, 1000 s/mm ²) ax.	5	2 min 06 sec	9900/82	420	75	160 × 90
T1w VIBE dynamic ax. with fat saturation	2.5	4 sec	4.421/1.29	350	68.8	512 × 246
T1w TSE ax. post-contrast with fat saturation	4	4 min 37 sec	588/12	400	75	512 × 230

ranking features were selected for further analysis. Each feature selection method was combined with several classifiers: Naïve Bayes, linear SVM, non-linear SVM (RBF-SVM), Random Forest and Gradient tree boosting (XGBoost). All classifiers were trained to predict N- or M-stage of the tumor and corresponding ROC curves were computed.

To accommodate for the small sample size, a stratified and nested 5-fold cross-validation procedure was repeated 40 times. In the inner loop, features were selected and the parameters of the classifier were optimized, while the resulting model was evaluated in the outer loop. The exact tuning procedure is summarized in ► **Table 3**. To estimate the final performance, all ROC curves were aggregated by pooling and the corresponding AUC was computed.

Statistical significance was tested by a permutation test with 1000 repeats. In addition, the ROC curves were tested against a constant model predicting the majority class by a bootstrap test. To gain insight into the specific choice of selecting 20 features, the number of features were varied in factors of 2 between 4 and 451.

Computations were performed using Python 3.5 and R 3.4.4.

Results

Patients and dataset

The multiparametric PET/MRI scans of all 30 patients were successfully completed, yielding 15 patients with N1 stage and 7 with M1 stage, each represented by 451 tumor image features.

N-stage results

The receiver operating characteristic (ROC) analysis of the pooled predictions revealed an area under the curve (AUC) ranging from 0.29 to 0.82 (► **Table 4**). No clear difference between the feature selection methods was found. The Mutual Information Feature Selection (MIFS) performed best when combined with RBF-SVM,

yielding the highest performance with a sensitivity of 83 %, a specificity of 67 % and an AUC of 0.82. The corresponding pooled ROC curve is displayed in ► **Fig. 2**. The lowest performance was seen in RBF-SVM when combined with SVM-RFE (AUC of 0.29). The permutation test indicated that the best model was highly significant ($p < 0.001$). The bootstrapping test indicated that the ROC curve was different from the ROC curve of the constant model ($p < 10^{-16}$).

M-stage results

Prediction of M-stage was superior when compared to N-stage. Receiver operating characteristic (ROC) analysis of the pooled predictions revealed an area under the curve (AUC) ranging from 0.66 to 0.97 (► **Table 5**). The features selected by SVM-RFE were among the best when combined with linear SVM or RBF-SVM. Naïve Bayes performed comparably well, indicating that the features are independent (after feature selection). Overall, there is no clear best feature selection method, although Mutual Information (MIFS) for feature selection apparently yielded inferior results.

Focusing on the best result, linear SVM with SVM-RFE obtained the highest performance. This model provided sensitivity of 91 % and specificity of 92 % with an AUC of 0.97. The corresponding pooled ROC curve is displayed in ► **Fig. 3**. The permutation test indicated that this model was highly significant ($p < 0.001$), while the bootstrapping test indicated that the ROC curve was different from the ROC curve of the constant model ($p < 10^{-16}$).

Image Features

In order to analyze the influence of the image features on the results, a correlation analysis on all 451 image features was performed. The analysis results of the full set of 451 image features revealed strong correlations between image features (► **Fig. 4**, left). Utilizing the feature selection step instead of removing correlated features aimed at compensation of potential noise by

► **Table 3** Parameters and tuning methods used for each classifier. As running times for tuning all parameters of the XGBoost classifier would be too high, stage-wise tuning was applied. The tuning procedure was independent of the stage to be predicted.

► **Tab. 3** Parameter- sowie Tuningmethoden für die einzelnen Klassifikatoren. Da die Laufzeiten des Tunings der Parameter des XGBoost-Klassifikators zu hoch waren, wurde ein schrittweises Tuning vorgenommen. Das Tuning war unabhängig vom vorherzusagenden Stadium.

classifier	parameters	tuning method
Naïve Bayes	none	none
linear SVM	$C \in [0..17]$	grid search with stratified 5-fold CV
RBF-SVM	$C \in [0..5]$ $\gamma \in [2^{-6}..5]$	grid search with stratified 5-fold CV
random forest	estimators {50, 250, 500} Min_samples_leaf {1..5} Max_depth {1, 2} Min_impurity_decrease $10^{[1,3,5,7,9]}$	grid search with stratified 5-fold CV
XGBoost	Max_depth {1..5} Min_child_weight {1, 2, 3} γ {0.0..0.9} subsample {0.15..0.95} colsample_bytree {0.1..0.95} learning_rate {0.01, 0.10, 0.2, 0.3, 0.5} N_estimators {100,250,500,1000,2500} Reg_alpha {0, 10^{-11} , 10^{-6} , 10^{-3} , 0.1, 1, 100}	tuning procedure following https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/

► **Table 4** Pooled AUC results (95 % confidence interval of the mean) for the 5-fold cross-validation with 40 repeats for N-stage.

► **Tab. 4** Gepoolte AUC-Werte (95 % Konfidenzintervall des Mittelwertes) für die 5-fache Kreuzvalidierung mit 40 Wiederholungen für das N-Stadium.

	naive bayes	random forest	RBF-SVM	linear SVM	XGBoost
anova	0.79 (95 % CI: 0.78, 0.80)	0.81 (95 % CI: 0.80, 0.82)	0.54 (95 % CI: 0.46, 0.57)	0.79 (95 % CI: 0.79, 0.81)	0.76 (95 % CI: 0.75, 0.78)
MIFS	0.79 (95 % CI: 0.78, 0.81)	0.75 (95 % CI: 0.74, 0.77)	0.82 (95 % CI: 0.79, 0.83)	0.81 (95 % CI: 0.80, 0.83)	0.68 (95 % CI: 0.67, 0.71)
SVM-RFE	0.82 (95 % CI: 0.81, 0.84)	0.81 (95 % CI: 0.79, 0.82)	0.29 (95 % CI: 0.24, 0.30)	0.79 (95 % CI: 0.77, 0.80)	0.76 (95 % CI: 0.75, 0.78)
t-score	0.79 (95 % CI: 0.78, 0.80)	0.81 (95 % CI: 0.80, 0.82)	0.54 (95 % CI: 0.46, 0.57)	0.79 (95 % CI: 0.79, 0.81)	0.76 (95 % CI: 0.75, 0.78)
wilcoxon	0.67 (95 % CI: 0.65, 0.68)	0.78 (95 % CI: 0.78, 0.79)	0.53 (95 % CI: 0.47, 0.55)	0.78 (95 % CI: 0.77, 0.79)	0.72 (95 % CI: 0.70, 0.74)

utilizing the information available in other correlated features, thus making classification more stable.

Analyses of the interactions of the feature selection method with the prediction task revealed strongly differing features for N- and M-stage. While feature selection methods in the case of M-stage exclusively preferred features that were derived from morphologic T2- and T1-weighted images, the preferred features in the case of N-stage were selected across all sequences including PET data, K^{trans} and ADC maps (► **Fig. 4**).

Varying the number of features revealed that slightly higher performing models could be found for N-stage (► **Fig. 5**, right), e. g. an AUC of 0.86 could be reached if no feature selection would

have been applied to the random forest model, and for M-stage (► **Fig. 5**, left), e. g. an AUC of 0.98 for linear SVM with SVM-RFE and 8 features.

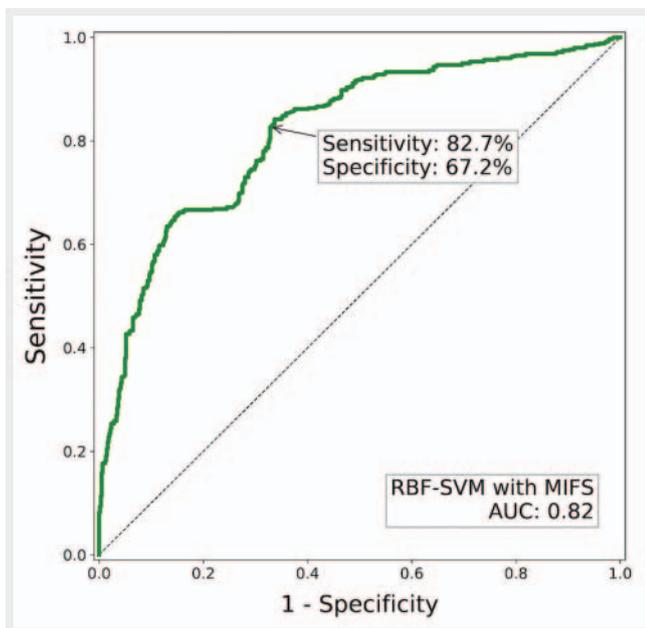
Discussion

The introduction of radiomics into scientific and clinical image analysis has caused an important shift in our understanding and utilization of medical images, transitioning from sole visual interpretation towards computer-based comprehensive quantification of tumor phenotypes using large numbers of quantitative image features

► **Table 5** Pooled AUC results (95 % confidence interval of the mean) for the 5-fold cross-validation with 40 repeats for M-stage.

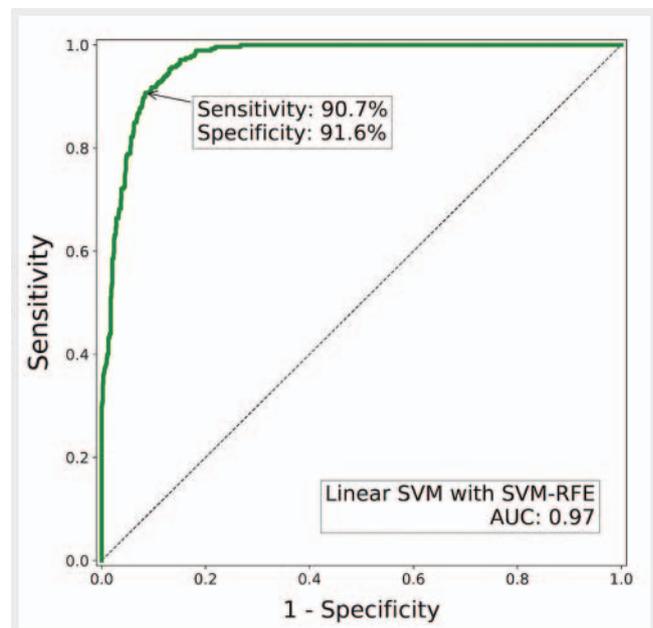
► **Tab. 5** Gepoolte AUC-Werte (95 % Konfidenzintervall des Mittelwertes) für die 5-fache Kreuzvalidierung mit 40 Wiederholungen für das M-Stadium.

	naive bayes	random forest	RBF-SVM	linear SVM	XGBoost
anova	0.93 (95 % CI: 0.92, 0.93)	0.89 (95 % CI: 0.88, 0.90)	0.92 (95 % CI: 0.91, 0.93)	0.93 (95 % CI: 0.92, 0.94)	0.93 (95 % CI: 0.92, 0.94)
MIFS	0.84 (95 % CI: 0.81, 0.86)	0.71 (95 % CI: 0.69, 0.73)	0.78 (95 % CI: 0.78, 0.81)	0.81 (95 % CI: 0.8, 0.84)	0.66 (95 % CI: 0.64, 0.68)
SVM-RFE	0.91 (95 % CI: 0.90, 0.92)	0.87 (95 % CI: 0.86, 0.89)	0.96 (95 % CI: 0.95, 0.97)	0.97 (95 % CI: 0.96, 0.97)	0.91 (95 % CI: 0.90, 0.92)
t-score	0.90 (95 % CI: 0.89, 0.91)	0.93 (95 % CI: 0.93, 0.94)	0.91 (95 % CI: 0.91, 0.92)	0.92 (95 % CI: 0.92, 0.93)	0.93 (95 % CI: 0.93, 0.94)
wilcoxon	0.93 (95 % CI: 0.93, 0.94)	0.92 (95 % CI: 0.92, 0.93)	0.93 (95 % CI: 0.93, 0.94)	0.94 (95 % CI: 0.94, 0.95)	0.91 (95 % CI: 0.90, 0.92)



► **Fig. 2** Pooled ROC curve for the RBF-SVM classifier with MIFS feature selection for the 5-fold cross-validation with 40 repeats for N-stage.

► **Abb. 2** Gepoolte ROC-Kurve des RBF-SVM-Klassifikators mit MIFS über die 5-fache Kreuzvalidierung mit 40 Wiederholungen zur Vorhersage des N-Stadiums.



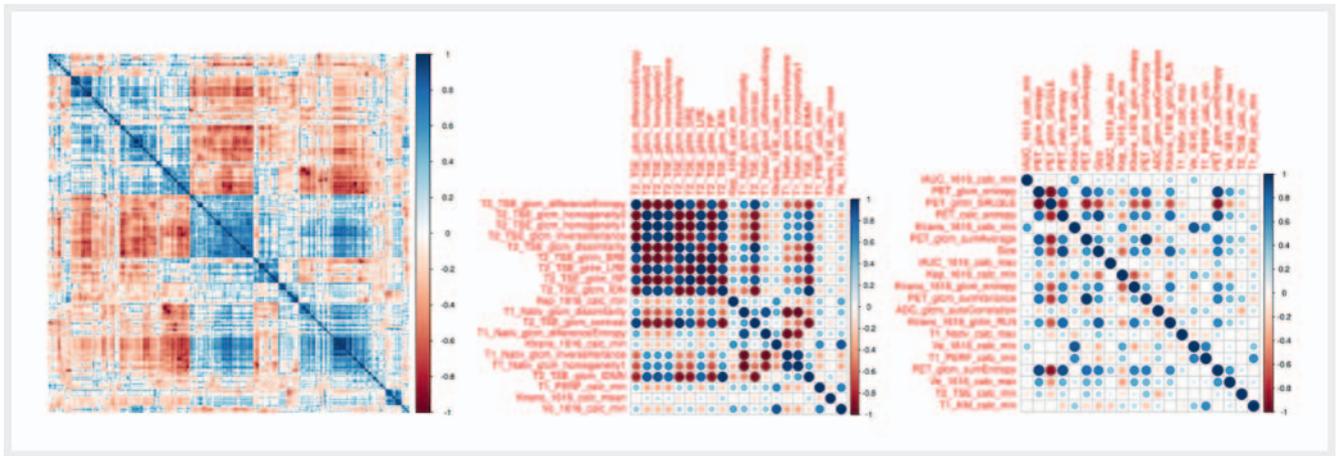
► **Fig. 3** Pooled ROC curve for the SVM classifier with SVM-RFE feature selection for the 5-fold cross-validation with 40 repeats for M-stage.

► **Abb. 3** Gepoolte ROC-Kurve des SVM-Klassifikators mit SVM-RFE für die 5-fache Kreuzvalidierung mit 40 Wiederholungen zur Vorhersage des M-Stadiums.

[1, 21]. Since then, a rapidly growing number of studies have found radiomic signatures to be predictive markers of underlying gene expression patterns, therapy response, relapse, patient survival and other clinical and histopathological outcomes, building a bridge between imaging and genomics, also known as radiogenomics [22–24]. Up to current status, this research is largely driven by novel hypotheses and machine learning procedures to be retrospectively tested mainly on computed tomography (CT) studies. While a large number of features can already be extracted from CT images, all

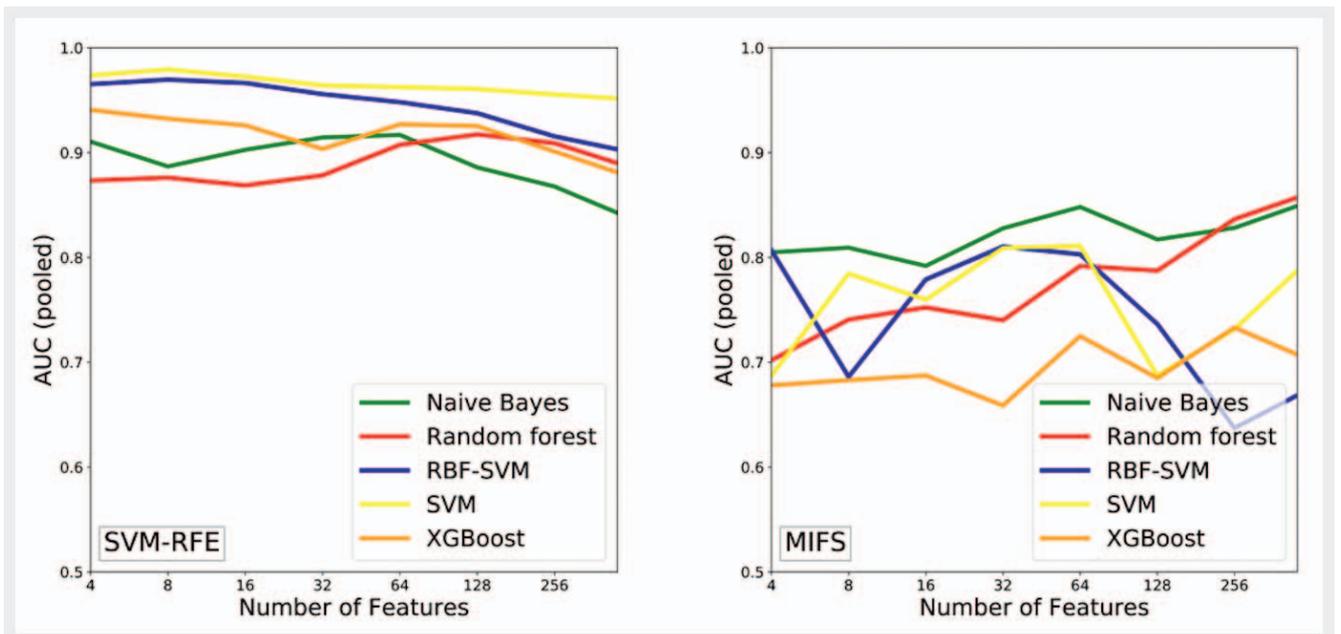
these features eventually originate from the same biophysical property of the respective tissue, X-ray attenuation and absorption, hence restricting the analysis to a constrained dimensionality. Recent studies underlined the potential of multi-dimensional imaging modalities such as MRI in enhancing the platform of obtained and quantifiable image features [25–27].

In the present study, we hypothesized that multiparametric PET/MR imaging would enable more comprehensive tumor phenotyping displayed in strong radiomic signatures as their respec-



► **Fig. 4** Graphical display of the clustered feature correlation matrix including all 451 image features (left). The matrix plot reveals strong correlations between a large number of features, as can be seen by the blocking structure. The most often selected features across all feature selection methods for the M-stage (middle) and the N-stage (right).

► **Abb. 4** Grafische Darstellung der Korrelationsmatrix aller 451 verwendeten quantitativen Bildmerkmale (links), von denen zahlreiche eine starke Korrelation aufweisen. Die am häufigsten ausgewählten Bildmerkmale zur Prädiktion des M- und N-Stadiums sind in der Mitte bzw. rechts dargestellt.



► **Fig. 5** Pooled AUC resulting from varying the number of features. Results from the best M-stage model using SVM-RFE as feature selection (left) as well as results from the best N-stage models using MIFS as feature selection (right).

► **Abb. 5** Gepoolte AUC-Werte, die sich aus der Variation der verschiedenen Bildmerkmale ergeben. Darstellung der besten Ergebnisse für das M-Stadium mit SVM-RFE als Feature-Auswahl (links) sowie Ergebnisse aus den besten N-Stadium-Modellen mit MIFS als Feature-Auswahl (rechts).

tive image features are based on a variety of morphological, functional and metabolic tissue properties derived from simultaneous multiparametric imaging. The results of our study underline the potential of multiparametric PET/MRI for the prediction of N- and M-stage based on a radiomics analysis, which was strictly restricted to the primary tumor in patients with cervical cancer. While utilization of SVM-RFE feature selection provided the high-

est performance for N- and M-staging, the sensitivity, specificity and AUC for M-stage were generally higher than for N-stage. This observation could potentially be explained by the fact that metastases in distant parts of the body usually represent more advanced disease often accompanied by more extensive morphological and genetic alterations of the primary tumor, which in turn could express more characteristic radiomic signatures.

Predicting N- and M-stage using radiomic signatures of the primary tumors was an adequate challenge to test our hypothesis. As outcome variables, these can be reliably assessed and are not affected by confounders like therapy regimens and hence, present valuable information with respect to the further therapeutic workup. A similar hypothesis on surrogate parameters for the prediction of mediastinal lymph node metastases based on ^{18}F -FDG PET/CT was recently investigated by Flechsig et al. [13]. Their results underline the predictive power of ^{18}F -FDG PET/CT with AUC values ranging from 0.89 for density and 0.82 for SUVmax as surrogate markers for N-staging in lung cancer patients. In another recent study by Kim et al., the authors reported MTV (metabolic tumor volume) and SUVmax to be independent factors for the prediction of lymph node metastases in rectal cancer patients [28]. In contrast to these previous publications, our results underline the potential of multiparametric PET/MRI as the imaging platform for soft-tissue-based tumors like cervical cancer, as the feature selection methods in case of N-stage preferred a combination of PET, perfusion and diffusion parameters and exclusively morphologic T1 and T2 parameters for M-stage. Our results indicate two important findings. First, the results of the radiomics analysis demonstrate the relatively minor importance of the tumor size for the prediction of N- and M-stage in cervical cancer, a characteristic that has been considered one of the most important and distinctive characterization features for tumor staging and therapy monitoring in conventional image analysis to date. Furthermore, comparable to previous studies on hybrid imaging, our results indicate and underline the predictive potential of hybrid imaging to provide surrogate markers not only for N- but also for M-staging [13, 28]. These kinds of imaging-based algorithmic analyses may help to identify targets for biopsies for N- or M-staging, optimize radiation planning regimens (in the case of extrapelvic lymph node metastases) and in the long term, maybe even reduce invasive staging techniques.

Nevertheless, while these initial results of radiomics-based prediction of N- and M-stage in cervical cancer underline the extensive potential of radiomics, limitations of our initial results include potential effects based on the heterogeneity of the investigated histopathological subtypes and tumor sizes as well as the small patient cohort. Hence, future studies should include larger patient cohorts as well as external validation and the inclusion of clinical and genetic parameters, to determine the true potential and clinical value of radiogenomics. Furthermore, although all of our results were internally validated by a repeated, nested cross-validation as well as a permutation test, it cannot be excluded that our models might be subjected to a certain amount of overfitting. This can only be verified by future studies, which should include larger patient cohorts as well as external validation. Inclusion of clinical and genetic parameters could also help to determine the true potential and clinical value of radiogenomics.

Conclusion

The results of this first pilot study on radiomics analysis of multiparametric PET/MR imaging data demonstrate the vast potential of this combination for image-based tumor phenotyping and its

potential to provide a promising imaging surrogate for preoperative N- and M-staging, treatment planning and risk stratification in patients with cervical cancer in a pretherapeutic setting.

CLINICAL RELEVANCE

- The metastatic status of primary cancers of the uterine cervix can be predicted based on multiparametric PET/MR-derived imaging data.
- The results of this pilot study demonstrate the potential for noninvasive image-based tumor phenotyping and patient stratification.
- The study results underline the predictive potential of radiomics analysis and its potential to be further established in clinical precision oncology.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- [1] 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
- [2] Lambin P, Rios-Velazquez E, Leijenaar R et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48: 441–446. doi:10.1016/j.ejca.2011.11.036
- [3] Parmar C, Grossmann P, Bussink J et al. Machine Learning methods for Quantitative Radiomic Biomarkers. *Sci Rep* 2015; 5: 13087. doi:10.1038/srep13087
- [4] Hawkins S, Wang H, Liu Y et al. Predicting Malignant Nodules from Screening CT Scans. *J Thorac Oncol* 2016; 11: 2120–2128. doi:10.1016/j.jtho.2016.07.002
- [5] Kalpathy-Cramer J, Mamomov A, Zhao B et al. Radiomics of Lung Nodules: A Multi-Institutional Study of Robustness and Agreement of Quantitative Imaging Features. *Tomography* 2016; 2: 430–437. doi:10.18383/j.tom.2016.00235
- [6] Grossmann P, Stringfield O, El-Hachem N et al. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife* 2017; 6: doi:10.7554/eLife.23421
- [7] Sarabhai T, Tschischka A, Stebner V et al. Simultaneous multiparametric PET/MRI for the assessment of therapeutic response to chemotherapy or concurrent chemoradiotherapy of cervical cancer patients: Preliminary results. *Clin Imaging* 2018; 49: 163–168. doi:10.1016/j.clinimag.2018.03.009
- [8] Lee DH, Kim SH, Im SA et al. Multiparametric fully-integrated ^{18}F -FDG PET/MRI of advanced gastric cancer for prediction of chemotherapy response: a preliminary study. *Eur Radiol* 2015; 26(26): 2771–2778. doi:10.1007/s00330-015-4105-5
- [9] Lohmann P, Kocher M, Ceccon G et al. Combined FET PET/MRI radiomics differentiates radiation injury from recurrent brain metastasis. *Neuroimage Clin* 2018; 20: 537–542. doi:10.1016/j.nicl.2018.08.024
- [10] Huang SY, Franc BL, Harnish RJ et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *NPJ Breast Cancer* 2018; 4: 24. doi:10.1038/s41523-018-0078-2

- [11] Lucia F, Visvikis D, Vallieres M et al. External validation of a combined PET and MRI radiomics model for prediction of recurrence in cervical cancer patients treated with chemoradiotherapy. *Eur J Nucl Med Mol Imaging* 2018; 46: 864–877. doi:10.1007/s00259-018-4231-9
- [12] Vallieres M, Freeman CR, Skamene SR et al. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015; 60: 5471–5496. doi:10.1088/0031-9155/60/14/5471
- [13] Flechsig P, Frank P, Kratochwil C et al. Radiomic Analysis using Density Threshold for FDG-PET/CT-Based N-Staging in Lung Cancer Patients. *Mol Imaging Biol* 2017; 19: 315–322. doi:10.1007/s11307-016-0996-z
- [14] Schindelin J, Arganda-Carreras I, Frise E et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; 9: 676–682. doi:10.1038/nmeth.2019
- [15] Jodogne SBC, Devillers M, Lenaerts E et al. Orthanc – A Lightweight, restful DICOM server for healthcare and medical research. IEEE 10th International Symposium on 2013. Available from: 2013 <https://ieeexplore.ieee.org/document/6556444/>
- [16] Carlson J. Radiomics: Radiomic Image Processing Toolbox. R package version 0.1.2. 2016. Available from: <https://cran.r-project.org/package=radiomics>
- [17] Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016. Available from: <https://www.r-project.org/>
- [18] Pecorelli S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynaecol Obstet* 2009; 105: 103–104
- [19] Pecorelli S, Zigliani L, Odicino F. Revised FIGO staging for carcinoma of the cervix. *Int J Gynaecol Obstet* 2009; 105: 107–108. doi:10.1016/j.ijgo.2009.02.009
- [20] Grueneisen J, Schaarschmidt BM, Beiderwellen K et al. Diagnostic value of diffusion-weighted imaging in simultaneous ¹⁸F-FDG PET/MR imaging for whole-body staging of women with pelvic malignancies. *J Nucl Med* 2014; 55: 1930–1935. doi:10.2967/jnumed.114.146886
- [21] Aerts HJ, Velazquez ER, Leijenaar RT et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014; 5: 4006. doi:10.1038/ncomms5006
- [22] Pinker K, Shitano F, Sala E et al. Background, current role, and potential applications of radiogenomics. *J Magn Reson Imaging* 2018; 47: 604–620. doi:10.1002/jmri.25870
- [23] Smith CP, Czarniecki M, Mehralivand S et al. Radiomics and radiogenomics of prostate cancer. *Abdom Radiol (NY)* 2018; 44: 2021–2029. doi:10.1007/s00261-018-1660-7
- [24] Grove O, Berglund AE, Schabath MB et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS One* 2015; 10: e0118261. doi:10.1371/journal.pone.0118261
- [25] Bonekamp D, Kohl S, Wiesenfarth M et al. Radiomic Machine Learning for Characterization of Prostate Lesions with MRI: Comparison to ADC Values. *Radiology* 2018. doi:10.1148/radiol.2018173064:173064
- [26] Ma W, Ji Y, Qi L et al. Breast cancer Ki67 expression prediction by DCE-MRI radiomics features. *Clin Radiol* 2018; 73: 909.e1–909.e5. doi:10.1016/j.crad.2018.05.027
- [27] Park H, Lim Y, Ko ES et al. Radiomics Signature on Magnetic Resonance Imaging: Association with Disease-Free Survival in Patients with Invasive Breast Cancer. *Clin Cancer Res* 2018; 24: 4705–4714. doi:10.1158/1078-0432.CCR-17-3783
- [28] Kim SH, Song BI, Kim BW et al. Predictive Value of [(18)F]FDG PET/CT for Lymph Node Metastasis in Rectal Cancer. *Sci Rep* 2019; 9: 4979. doi:10.1038/s41598-019-41422-8