Radiotherapy and Oncology 130 (2019) 10–17

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

CT imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer

Stefan Leger [a,b,c,*], Alex Zwanenburg [a,b,c], Karoline Pilz [a,b,c,h], Sebastian Zschaeck [a,b,d,e], Klaus Zöphel [f,g], Jörg Kotzerke [f,g], Andreas Schreiber [h], Daniel Zips [i], Mechthild Krause [a,b,c,d,j], Michael Baumann [a,b,c,d,j], Esther G.C. Troost [a,b,c,d,j,1], Christian Richter [a,b,d,j,1], Steffen Löck [a,b,d,1]

[a] OncoRay National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum Dresden – Rossendorf; [b] German Cancer Research Center (DKFZ), Heidelberg and German Cancer Consortium (DKTK) partner site Dresden, Dresden; [c] National Center for Tumor Diseases (NCT), Partner Site Dresden, Germany German Cancer Research Center (DKFZ), Heidelberg, Germany, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany, and Helmholtz Association / Helmholtz-Zentrum Dresden – Rossendorf (HZDR), Dresden, Germany; [d] Department of Radiotherapy and Radiation Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden; [e] Charité Universitätsmedizin Berlin, Department of Radiation Oncology; [f] Department of Nuclear Medicine, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden; [g] Helmholtz-Zentrum Dresden-Rossendorf, PET Center, Institute of Radiopharmaceutical Cancer Research; [h] Department of Radiotherapy, Hospital Dresden-Friedrichstadt; [i] Department of Radiation Oncology, Faculty of Medicine and University Hospital Tübingen, Eberhard Karls Universität Tübingen; and [j] Helmholtz-Zentrum Dresden – Rossendorf, Institute of Radiooncology OncoRay, Germany

ARTICLE INFO

Article history:
Received 16 November 2017
Received in revised form 27 June 2018
Accepted 24 July 2018
Available online 4 August 2018

Keywords:
Radiomic risk modelling
Computed tomography
Imaging during treatment
Patient stratification

ABSTRACT

Background and purpose: The development of radiomic risk models to predict clinical outcome is usually based on pre-treatment imaging, such as computed tomography (CT) scans used for radiation treatment planning. Imaging data acquired during the course of treatment may improve their prognostic performance. We compared the performance of radiomic risk models based on the pre-treatment CT and CT scans acquired in the second week of therapy.

Material and methods: Treatment planning and second week CT scans of 78 head and neck squamous cell carcinoma patients treated with primary radiochemotherapy were collected. 1538 image features were extracted from each image. Prognostic models for loco-regional tumour control (LRC) and overall survival (OS) were built using 6 feature selection methods and 6 machine learning algorithms. Prognostic performance was assessed using the concordance index (C-Index). Furthermore, patients were stratified into risk groups and differences in LRC and OS were evaluated by log-rank tests.

Results: The performance of radiomic risk model in predicting LRC was improved using the second week CT scans (C-Index: 0.79), in comparison to the pre-treatment CT scans (C-Index: 0.65). This was confirmed by Kaplan–Meier analyses, in which risk stratification based on the second week CT could be improved for LRC ($p = 0.002$) compared to pre-treatment CT ($p = 0.063$).

Conclusion: Incorporation of imaging during treatment may be a promising way to improve radiomic risk models for clinical treatment adaption, *i.e.*, to select patients that may benefit from dose modification.

Model-driven radiation oncology using high-throughput analyses of advanced imaging biomarkers (Radiomics) is an upcoming field in the era of patient-specific and individualised cancer therapy [1]. Radiomics has shown promising results in several studies using different image modalities, such as computed tomography (CT), positron emission tomography (PET) or magnetic resonance imaging (MRI) to predict survival outcome data [2–7]. Most radiomic models are based on pre-treatment imaging. Imaging during treatment may be of additional prognostic value, since it may reflect biological processes associated with therapy response, such as re-oxygenation and/or tumour shrinkage [8–10].

Several studies investigated the prognostic value of specific image biomarkers over time, *e.g.*, using PET imaging [8,9,11,12]. For patients with locally advanced head and neck squamous cell carcinoma (HNSCC), Hentschel et al. [13] showed that the decrease of the maximum standard uptake value (SUV) extracted from [^{18}F]-fluorodeoxyglucose (FDG)-PET imaging in treatment weeks 1 or 2 had a higher prognostic value than at baseline. Furthermore, Zips

* Corresponding author at: OncoRay – National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Fetscherstraße 74, PF 41, 01307 Dresden, Germany.
E-mail address: Stefan.Leger@oncoray.de (S. Leger).
[1] Shared senior authorship.

https://doi.org/10.1016/j.radonc.2018.07.020
0167-8140/

et al. [14] demonstrated the strong prognostic value of [^{18}F]-fluoromisonidazole (FMISO)-PET imaging parameters after weeks 1 and 2 of radiotherapy, which was recently validated [15].

In the field of radiomics, so far only a few studies have assessed the change or the prognostic value of advanced imaging biomarkers during the course of treatment. Cunliffe et al. [16] investigated the relationship between radiation dose characteristics and the change of CT-based radiomic features with the development of radiation pneumonitis using imaging before and after treatment. Recently, Fave et al. [17] showed that quantitative radiomic features derived from CT change significantly during treatment. However, they also found that these changes contain only limited prognostic value for patients with non-small cell lung cancer (delta radiomics). Van Timmeren et al. [18] described a feature selection methodology using cone beam CT (CBCT) to select reproducible delta radiomic features that are informative due to their change during treatment. However, the prognostic value of those features was not investigated. With those limited data available, additional studies are required to evaluate the possible improvement of prognostic radiomic models on CT imaging data acquired during treatment.

Therefore, the main objective of this study was to investigate the potential of radiomic risk models (for loco-regional tumour control; LRC, and overall survival; OS) trained on pre-treatment planning CT imaging in comparison to CT imaging in the second week of radiochemotherapy and to the combination of both data sets for patients with locally advanced HNSCC.

Material and methods

Patient data

Radiomic risk models were developed and validated on two different patient cohorts with 78 patients in total. All patients were diagnosed with histologically confirmed locally advanced HNSCC and received primary radiochemotherapy (RCT). The study design is presented in Fig. 1. The exploratory cohort consisted of 48 patients, treated within a prospective clinical trial (NCT00180180, [14,15]) at the University Hospital Dresden (UKD, Germany) between 2006 and 2012. The imaging data consisted of an FDG-PET/CT scan ($CT_{W0\text{-}FDG}$), which was used for treatment planning. Furthermore, FMISO-PET/CT scans were acquired 2–4 days after the planning CT scan, but prior to initiation of RCT ($CT_{W0\text{-}FMISO}$), and after a dose of 18–20 Gy (CT_{W2}, end of week 2 of RCT). The validation cohort consisted of 30 patients, who were treated at the UKD and the Hospital Dresden-Friedrichstadt between 2005 and 2009. Imaging in this cohort contained an FDG-PET/CT ($CT_{W0\text{-}FDG}$) scan for treatment planning and a subsequent CT scan after a dose of 18–20 Gy (CT_{W2}, end of week 2–3) during RCT. All imaging data were acquired with treatment masks in radiotherapy position.

Moreover, an additional cohort of 18 patients with HNSCC was included to assess the stability of radiomic features. These patients were treated within a prospective clinical trial at the UKD between 2014 and 2016 [DRKS00006007]. Imaging data and time points in this cohort were comparable to the exploratory cohort. This cohort was excluded from further analyses due to insufficient follow-up for the evaluation of LRC and OS. Ethical approval for the multicentre retrospective analyses of clinical and imaging data was obtained from the Ethics Committee at the Technische Universität Dresden, Germany (EK177042017). All analyses were carried out in accordance with the relevant guidelines and regulations. Informed consent was obtained from all patients.

Study design

In the present study, prognostic radiomic models were developed based on imaging data acquired (a) pre-treatment ($CT_{W0\text{-}FDG}$) and (b) after the second week of treatment (CT_{W2}). In addition, the combination of $CT_{W0\text{-}FDG}$ and CT_{W2} (c) was considered. Delta features were computed for (c) as $\Delta CT = CT_{W2}/CT_{W0\text{-}FDG}$, i.e. the ratio of feature values derived from CT_{W2} and $CT_{W0\text{-}FDG}$ for every feature. Three types of analyses were conducted based on the three feature sets: (I) Radiomic models were developed using different feature selection methods and machine learning algorithms to avoid incidental findings. The mean performance of the developed models was compared between feature sets (a)–(c). (II) The radiomic models with the highest performance within the exploratory cohort in pre-treatment were analysed in more detail for (a)–(c). Risk-based patient stratification was assessed and the correlations between features within the signatures and their expression values were analysed. (III) The

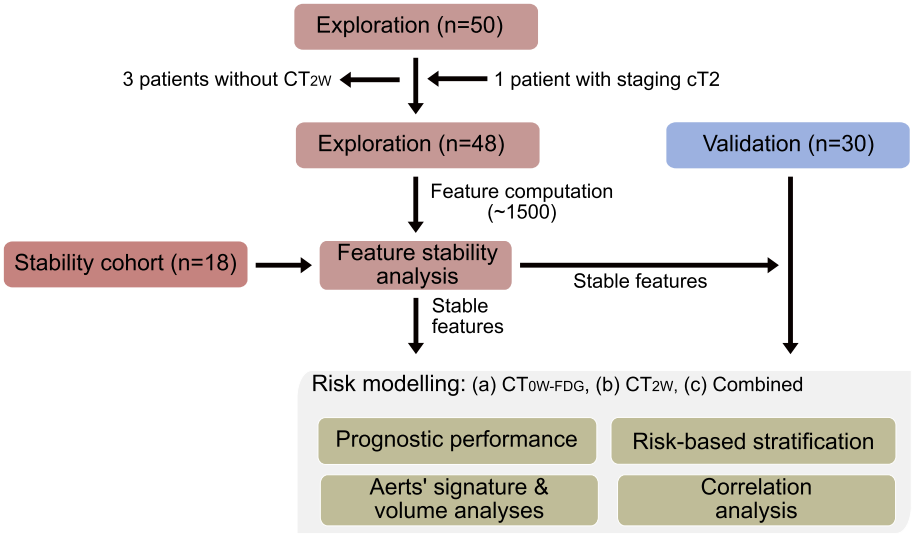

Fig. 1. Representation of the study design. Three cohorts were included. CT images from the exploratory cohort (48) and stability cohort (18) were used to identify a stable feature set. Subsequently, pre-treatment and in-treatment images from the exploratory cohort as well as their combination were used to train prognostic radiomic models. Prognostic models were also trained using the radiomic signature obtained by Aerts et al. [3] and the tumour volume. Prognostic model performance and patient risk group stratification were assessed on the validation cohort (30).

prognostic performance of radiomic models that were based on the signature obtained by Aerts et al. [3] and on the tumour volume was evaluated for both time points.

Image pre-processing and feature extraction

The gross tumour volume (GTV) of the primary tumour was manually delineated by a radiation oncologist in training (KP) and independently validated (EGCT) on each CT scan separately. The voxel spacing was resampled using trilinear image interpolation to an isotropic voxel size of $1.0 \times 1.0 \times 1.0\ mm^3$ to correct for differences in voxel spacing and slice thicknesses between cohorts [3,19]. The GTV mask was re-segmented to cover only soft tissue voxels between −150 and 180 Hounsfield units, thus removing voxels that contain air and bone tissue, which may otherwise affect feature expression. Additional images were created by applying spatial filtering to the base image to emphasise image characteristics such as edges and blobs. Eight additional images were created by applying a stationary coiflet-1 wavelet high-/low-pass filter along each of the three spatial dimensions [3,20]. One further image was created by applying a Laplacian of Gaussian (LoG) filter consisting of five different filter kernel widths (1.0 mm, 2.0 mm, 3.0 mm, 5.0 mm, 6.0 mm [21]). Finally, 18 statistical, 38 histogram-based and 95 texture features were extracted from the GTV within each image set (base image and 9 transformed images). 28 morphological features were computed within the base image only. Thus, 1538 features were computed in total. Image pre-processing and feature extraction were performed according to the guidelines of the image biomarker standardisation initiative [22]. A further description of the feature extraction can be found in Supplement S1.

Feature stability

Feature stability was assessed between the $CT_{W0\text{-}FDG}$ and $CT_{W0\text{-}FMISO}$ scans prior to model building and validation to reduce the influence of different CT acquisition parameters on the prognostic models (Table 1). The in-treatment scans (CT_{W2}) were acquired with a lower exposure than the pre-treatment images ($CT_{W0\text{-}FDG}$) in the exploratory and validation cohorts to limit patient radiation dose, while the acquisition parameters between the $CT_{W0\text{-}FMISO}$ pre-treatment and the CT_{W2} scans were similar. Therefore, feature stability was assessed using the $CT_{W0\text{-}FDG}$ and $CT_{W0\text{-}FMISO}$ images of the exploratory cohort and the additional cohort of 18 patients, leading to 66 patients in total. The $CT_{W0\text{-}FDG}$ and the $CT_{W0\text{-}FMISO}$ scans were rigidly registered with RayStation (version 6.0, RaySearch Laboratories AB, Stockholm, Sweden). Afterwards, the GTV was manually transferred from the $CT_{W0\text{-}FDG}$ to the $CT_{W0\text{-}FMISO}$ images. Imaging features with a Spearman rank correlation coefficient (SCC) ≥ 0.8 between $CT_{W0\text{-}FDG}$ and $CT_{W0\text{-}FMISO}$ were considered stable and used for feature selection and model building [23].

Feature selection methods and machine learning algorithms

In this study, 6 feature selection methods and 6 learning algorithms found as most reliable in a previous systematic evaluation [24] were used for prognostic modelling. The following feature selection methods were applied: Spearman correlation, mutual information maximisation (MIM), mutual information feature selection (MIFS), minimum redundancy maximum relevance (MRMR), random forest variable importance (RF-VI) and a forward feature selection based on Cox regression (multi-Cox). For model building we used Cox regression, boosting trees (BT) Cox models, boosting gradient linear models (BGLM) with Cox regression, random survival forests (RSF) and random forest using maximally selected rank statistics (MRF). Additionally, we investigated the full-parametric BT-Weibull model. A short description of these methods can be found in Supplement S2. All feature selection methods and machine learning algorithms are able to handle continuous time-to-event survival data.

Radiomics modelling framework

The radiomic risk models were generated within our radiomics modelling framework [24] (see Fig. 2) consisting of five major processing steps: feature pre-processing, feature selection, hyperparameter optimisation, model building and model validation. Feature pre-processing comprised z-normalisation of the imaging features and feature clustering. Features in the exploratory cohort were clustered using a hierarchical clustering algorithm to obtain an initial non-redundant set of biomarkers [4,25]. The distance between each pair of features was based on their correlation. All features that formed a cluster with intra-cluster correlation >0.90 were replaced by a new meta-feature with the mean value of all clustered features. For instance, clustering yielded 32 meta-features and 23 unclustered features for the pre-treatment images. After clustering, feature selection was conducted to identify the most relevant features. Feature selection was repeated on $n = 1000$ bootstrap samples (*i.e.*, .632 bootstrap method with replacement) of the exploratory cohort to reduce randomness in the selection of relevant features. Afterwards, the hyperparameters of the machine learning algorithms, such as signature size or algorithm-specific settings, were optimised using internal cross validation of the exploratory cohort. Model training was performed on $m = 1000$ bootstrap samples of the exploratory cohort

Table 1
Image acquisition parameters of the different cohorts.

Imaging parameter	Exploratory cohort			Stability cohort		Validation cohort	
	$CT_{W0\text{-}FDG}$	$CT_{W0\text{-}FMISO}$	CT_{W2}	$CT_{W0\text{-}FDG}$	$CT_{W0\text{-}FMISO}$	$CT_{W0\text{-}FDG}$	CT_{W2}
Voxel spacing *(x, y)*, in mm							
(0.85, 0.85)	0	1	0	0	0	0	0
(0.97, 0.97)	12	1	0	18	0	1	0
(1.36, 1.36)	36	46	48	0	18	29	30
z, in mm							
2.0/3.0/5.0	0/12/36	0/1/47	1/0/47	6/12/0	18/0/0	0/1/29	0/0/30
Reconstruction kernel							
B10s/B19f/B20f/B20s/B31f/NA	1/0/32/0/12/3	0/10/37/0/0/1	0/12/36/0/0/0	0/0/0/0/18/0	0/18/0/0/0/0	0/0/28/2/0/0	0/0/30/0/0/0
Mean exposure mAs	30.8	7.7	7.9	39.0	9.6	39.1	8.1
Manufacturer	Siemens						
Scanner model	Sensation 16/Biograph 16						
Mean exposure time, in ms	420–500						
Tube voltage in kV	120						

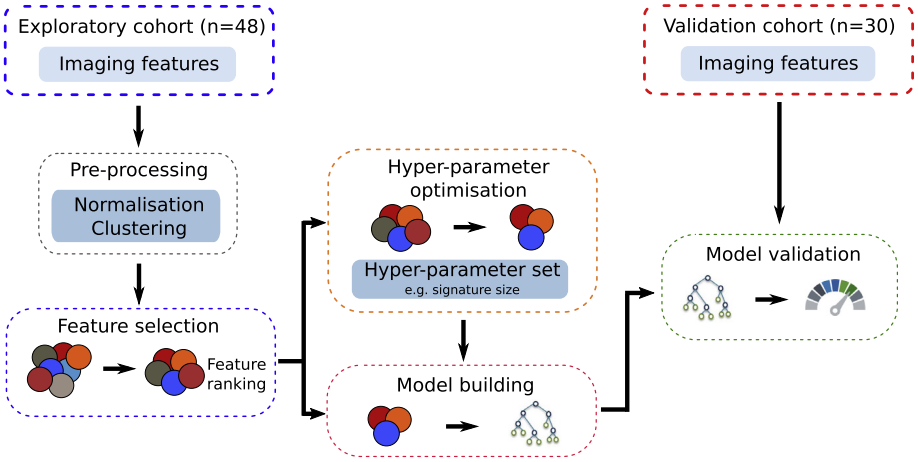

Fig. 2. Schematic overview of the radiomic modelling framework (RMF) consisting of five steps: feature pre-processing, feature selection to identify prognostic biomarkers, automatic hyper-parameter optimisation for each model using a 2-fold cross validation with 40 repetitions based on the exploratory cohort, model building and model validation.

for each combination of feature selection method and machine learning algorithm. Afterwards, an ensemble prediction was made by averaging the predicted risk scores for each model using data of the independent validation cohort [26]. Further details of the radiomics modelling framework are described in Supplement S3.

Statistical analyses

The clinical endpoints LRC (primary) and OS (secondary) were calculated from the first day of RCT to the date of event or censoring. Binary variables were compared between the patient cohorts using exact Fisher tests or X^2 tests, while differences in continuous variables were evaluated by Mann–Whitney-*U* tests. The performance of radiomic models was assessed using the concordance index (C-Index) [27,28]. The C-Index is a generalisation of the area under the curve for continuous time-to-event survival data, and ranges from 0.5 (random) to 1.0 (perfect prediction). The average model performance of the considered feature selection methods and machine learning algorithms was compared between the time points by a multi-level approach (MLA), which is described in Supplement S4. Patients were stratified into low and high risk groups based on the risk predicted by the radiomic models. The median risk on the exploratory cohort was determined and applied to stratify the validation cohort. Survival curves were estimated by the Kaplan–Meier method and compared by log-rank tests. All analyses were performed using SPSS 23 (IBM Corporation, Armonk, NY) and R language (R Foundation for Statistical Computing, Vienna, Austria). Two-sided tests were applied and p-values < 0.05 were considered statistically significant.

Results

The patient characteristics of the exploratory and validation cohort are summarised in Table 2. Median follow-up time was 28.8 months (range 1.3–70.3 months) for the exploratory cohort and 21.5 months (1.4–107.2 months) for the validation cohort. The 2-year LRC rate was 63.0% for the exploratory and 56.0% for the validation cohort ($p = 0.61$). Overall survival after 2 years was

Table 2
Patient characteristics of the exploratory and the validation cohort.

Variable	Exploratory cohort	Validation cohort	*p*-value
Gender (male/female)	41/7	27/3	0.73[1]
Age (median, range, in years)	53.5 (42.0–74.0)	54.5 (37.0–67.0)	0.79[3]
Clinical TNM Staging			
cT Stage 1/2/3/4/missing	0/1/17/30/0	3/8/9/9/1	<0.001[2]
cN stage 0/1/2/3/missing	4/6/37/1/0	4/2/21/2/1	0.43[2]
UICC stage 2010			
I/II/III/IV/missing	0/0/7/41/0	1/2/1/26/0	0.069[1]
Tumour volume (median, range, in cm^3)	40.67 (7.29–239.07)	23.49 (2.71–183.56)	0.94[3]
Dose (median, range, in Gy)	72.0 (69.0–72.0)	72.0 (70.6–76.8)	0.32[3]
HPV16 DNA negative/positive/missing	36/5/7	0/0/30	–
Number of events			
LRC	15	11	
OS	33	17	
Follow up time of patients alive (median, range, in months)	38.4 (23.8–70.3)	61.7 (7.8–107.2)	
Follow up time (median, range, in months)	28.8 (1.3–70.3)	21.5 (1.4–107.2)	
Fractionation scheme	PTV1: primary tumour and involved lymph nodes, 72 Gy in 6 weeks PTV2: elective nodes 59.4 Gy in 5 weeks PTV1 + 2: 30 Gy/2 Gy in 15 fx PTV1: 42 Gy/1.4 Gy (twice a day) in 30 fx PTV2: 29.4 Gy/1.4 Gy (twice a day) in 21 fx		

[1]Exact Fisher test, [2]X^2 test, [3]Wilcoxon–Mann–Whitney test.
Abbreviations: cT, clinical tumour stage. cN, clinical nodal stage. LRC, loco-regional tumour control. OS, overall survival.

50.0% for the exploratory and 53.0% for the validation cohort (p = 0.56). The corresponding Kaplan–Meier curves are shown in Supplementary Fig. S1. Patients in the validation cohort had a significantly lower clinical T stage ($p < 0.001$).

(I) Radiomic models were developed using imaging data acquired at (a) pre-treatment ($CT_{W0\text{-}FDG}$), (b) after the second week of treatment (CT_{W2}) and (c) their combination including delta features. 1538 imaging features were extracted per CT scan. Feature stability assessment reduced the feature set to 269 stable imaging features (SCC $\geq$0.8), consisting of 12 statistical, 18 morphological, 26 histogram-based and 213 texture features. Using these 269 stable features, radiomic models were developed for the endpoint LRC based on different feature selection methods and machine learning algorithms. Fig. 3 shows the C-Index for the exploratory and validation cohorts. The average performance of models developed using the in-treatment CT_{W2} images (C-Index: 0.73 ± 0.04, mean ± std) was significantly higher than the average performance of models based on the pre-treatment $CT_{W0\text{-}FDG}$ images (C-Index: 0.62 ± 0.04, MLA: p = 0.005). Using the combined feature set also led to improved results with a mean C-Index of 0.70 ± 0.05 compared to pre-treatment (MLA: p = 0.06). An additional internal cross-validation experiment on the entire dataset confirmed these results (C-Index $CT_{W0\text{-}FDG}$: 0.61 and CT_{W2}: 0.70, MLA: p = 0.16, Supplement S5).

(II) For the pre-treatment $CT_{W0\text{-}FDG}$ scans, Spearman feature selection combined with the BT-Cox model achieved a C-Index of 0.95 (95% confidence interval [0.92–1.00]) in exploration, and was thus further analysed. Its validation C-Index was 0.65 [0.51–0.79]. The model based on CT_{W2} scans showed a higher validation performance (0.79 [0.68–0.96]), while the model based on the combined feature set (ΔCT) performed similar to the baseline model (0.65 [0.49–0.88]). Hyper-parameters of these models are given in Supplement Table S4. Patients were stratified into low and high risk groups according to the median risk of the BT-Cox models (Fig. 4). The models trained on the pre-treatment $CT_{W0\text{-}FDG}$ scans and on the combined feature set were not able to stratify patients of the validation cohort with a significant difference in LRC (p = 0.063 and p = 0.19, respectively), whereas the model trained on the CT_{W2} scans led to a significant patient stratification (p = 0.002). Similar results were obtained for the other models (Supplementary Fig. S2). The signatures of above models generally consisted of texture features and morphological features, which were non-redundant (Supplement Table S2). Feature expressions are shown in Fig. 4 and details are given in Supplement Table S1.

(III) The in-treatment BT-Cox model based on Aerts' signature also led to improved prognostic performance on the validation cohort compared to the pre-treatment model (C-Index: CT_{W2}: 0.74 [0.61–0.91] and $CT_{W0\text{-}FDG}$: 0.66 [0.51–0.89], respectively). However, the median cut-offs resulting from both models were not able to stratify patients of the validation cohort into a low and high risk group with a significant difference in LRC

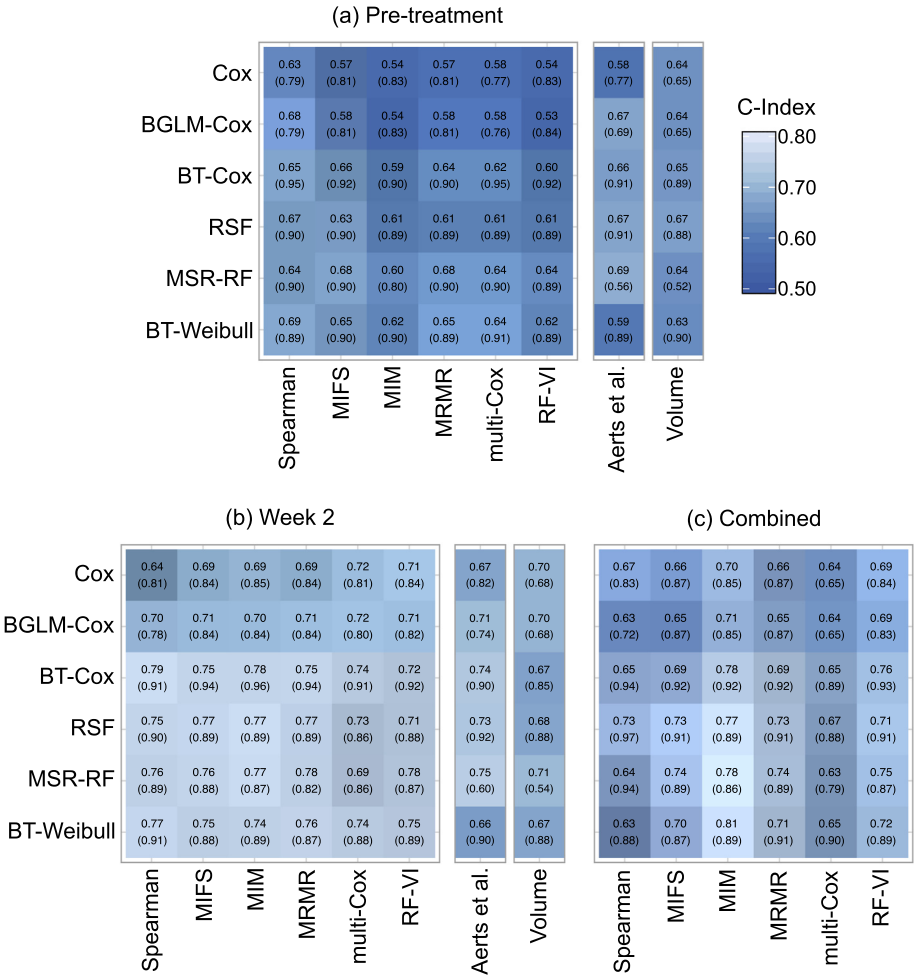

Fig. 3. Grid with concordance indices for predicting loco-regional tumour control for the validation cohort and the exploratory cohort (brackets). Radiomic models were developed using a feature selection method (rows) and a learning algorithm (columns), based on (a) pre-treatment $CT_{W0\text{-}FDG}$ imaging, (b) in-treatment CT_{W2} imaging and (c) the combined feature set. Furthermore, performances of the models based on Aerts et al. [3] signature and the tumour volume are shown. A detailed description of the algorithm abbreviations can be found in Supplement S2.

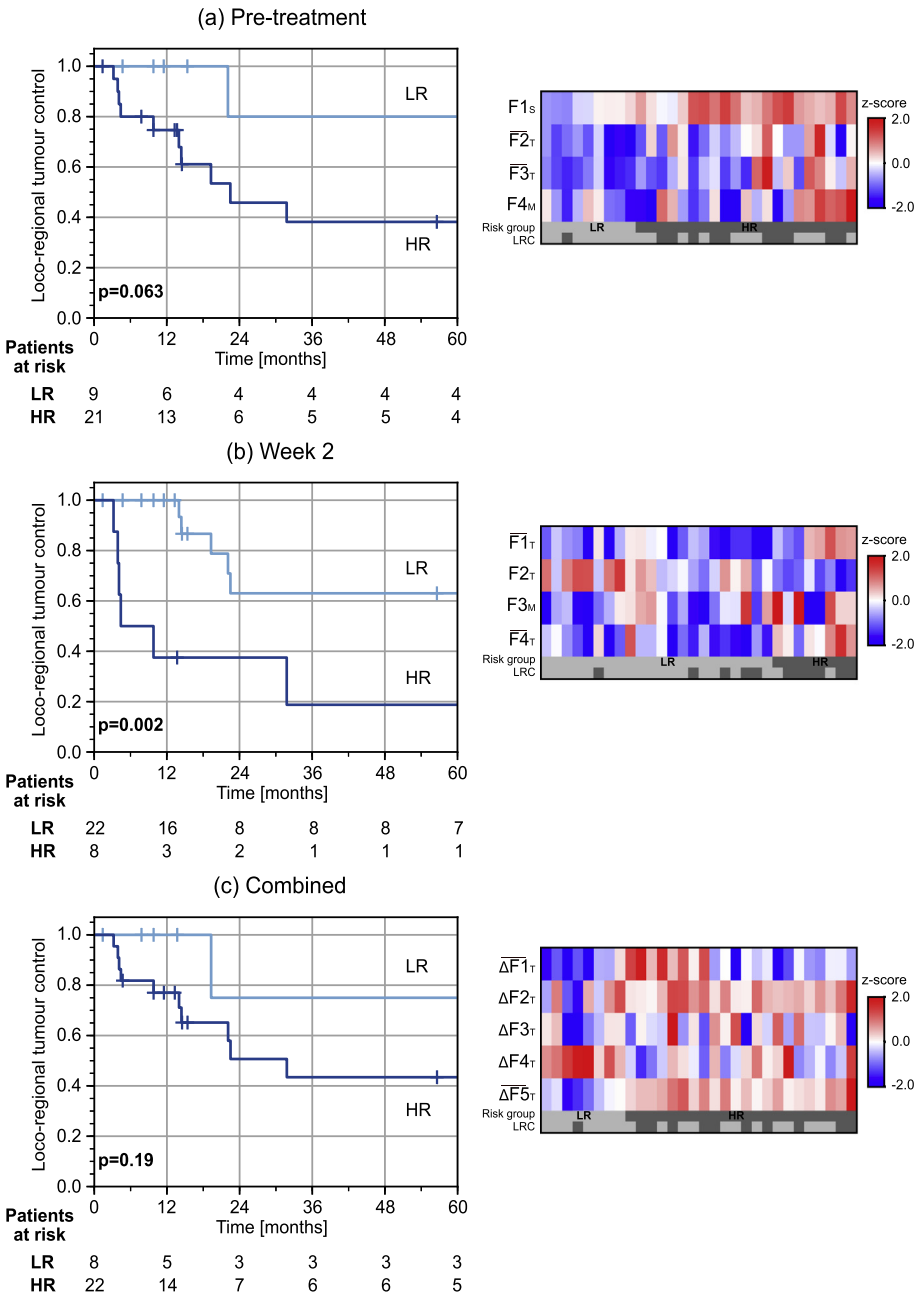

Fig. 4. Kaplan–Meier plots of loco-regional tumour control (left) and heatmaps (right) for patients of the validation cohort stratified into a low (LR) and a high (HR) risk group based on the median risk value determined on the exploratory cohort based on the BT-Cox model in combination with Spearman feature selection. The median cut-off values on the (a) pre-treatment $CT_{W0\text{-}FDG}$, (b) CT_{W2} and (c) combined feature set ($CT_{W0\text{-}FDG}$, CT_{W2}, ΔCT) were -0.41253, −0.03290 and −0.00081, respectively. LRC during follow-up (yes, light; no, dark) is shown. A detailed description of the feature abbreviations can be found in Supplementary Table S1.

($CT_{W0\text{-}FDG}$: $p = 0.53$ and CT_{W2}: $p = 0.30$). In addition, BT-Cox models were built using the tumour volume. On the validation cohort, these models achieved a performance of 0.65 [0.51–0.84] and 0.67 [0.53–0.88] for pre-treatment and in-treatment feature sets, respectively.

For the endpoint OS, (I) the model performances in the validation and exploratory cohorts are shown in Supplementary Fig. S3. (II) MIM feature selection combined with the BT-Cox model led to the best average models. Radiomic signatures from these models are described in Supplement Table S1. In validation, a similar performance was observed for the in-treatment and the pre-treatment models ($CT_{W0\text{-}FDG}$: 0.59 [0.45–0.77]) and CT_{W2}: 0.61 [0.46–0.77]). Kaplan–Meier analyses are summarised in Supplementary Figure S4. (III) The pre-treatment model based on Aerts' signature achieved a higher C-Index (0.66 [0.52–0.82]) than the in-treatment model (0.62 [0.48–0.80]). The in-treatment tumour volume (0.67, [0.54–0.83]) had a higher prognostic value than the baseline GTV (0.63, [0.50–0.80]).

Discussion

In this study, the prognostic value of radiomic risk models using CT images obtained during treatment was compared to models based on pre-treatment CT images. We showed that risk models trained on week 2 scans achieved a significantly higher prognostic performance for LRC and led to improved patient risk stratifications in comparison to pre-treatment CT scans. The improved performance was also observed for the signature of Aerts et al. [3].

Models based on in-treatment imaging showed a higher prognostic value than tumour volume, which performed similar to the pre-treatment models.

Changes in image feature expressions may be related to changes in tumour biology during the course of treatment. Such biological changes comprise RCT-induced re-oxygenation and shrinkage of the tumour, that have been associated with treatment response [12,29–31]. For example, Zips et al. [14] showed that the hypoxic volume and the tumour to background ratio obtained from FMISO-PET images have a strong association with loco-regional tumour control. Furthermore, they observed an improved prognostic performance after week 2 of RCT in comparison to the pre-treatment images, which was recently validated [15]. The CT-based radiomic signatures developed in this manuscript contained mostly texture and morphological features. We assessed the correlation of these features with tumour hypoxia, measured by the FMISO-PET imaging parameters hypoxic volume and the tumour to background ratio [14,15]. Most texture-based features from our signatures and from Aerts' signature were moderately correlated with one or more FMISO-PET hypoxia markers (Supplement Table S3). This correlation slightly increased in the second week of treatment, where FMISO-PET markers were previously shown to have improved prognostic value [14]. The correlation with hypoxia markers offers one explanation for the better performance of the in-treatment CT-based radiomic models and gives an indication that tumour hypoxia may be observable in macroscopic CT imaging. However, the links between imaging features, tumour hypoxia and other biological tumour mechanisms should be studied in closer detail, as these are generally poorly understood.

The combined feature set, consisting of pre-treatment, week 2 and delta CT imaging features, likewise led to improved model performance compared to models only based on pre-treatment CT-based features. However, the selected features were predominantly extracted from the second week CT scans, or were delta features, which underlines the importance of the in-treatment data. Moreover, signatures containing only features from the week 2 CT scans showed a higher performance (*e.g.*, MIM feature selection) for predicting LRC than signatures including delta features (*e.g.*, Spearman feature selection). This is in line with the study by Fave et al. [17], which showed that delta radiomic features are changing during treatment but provide limited additional prognostic information compared to baseline imaging.

For OS we observed similar results for in-treatment and pre-treatment imaging. Compared to LRC, the model performance was generally lower. This may be due to the fact that the cause of death was not necessarily cancer-related, which causes the OS endpoint to be comparatively noisy. Furthermore, the combined feature set led to a lower performance in the validation cohort than the in-treatment feature set. One explanation for the reduced performance could be the selected features. In the developed signature, the two delta features were discordantly expressed between the exploratory and the validation cohort (Supplementary Fig. S5). This may negatively effect the risk prediction. Interestingly, OS models based on Aerts' signature achieved the highest accuracy using pre-treatment CT scans. This result is reasonable, since this signature was developed for the prediction of OS using pre-treatment CT scans.

We performed a feature stability analysis prior to model building, as several acquisition parameters differed between the planning and week 2 CT scans, *e.g.*, the mean CT exposure settings. A lower CT exposure leads to increased image noise, which in turn affects imaging features. Therefore, feature stability was assessed prior to feature selection and model building to reduce the influence of such differences. Feature stability was measured by the Spearman rank correlation coefficient to consider non-linear correlation effects. Furthermore, the stability is also influenced by other factors, such as uncertainties in image registration or dissimilarities due to the GTV transfer. To further enhance the robustness, feature stability information, *e.g.*, from test re-test or multiple tumour delineations datasets, may be included in future [32,33]. Furthermore, initiatives such as the Quantitative Imaging Network (QIN) of the National Institute of Health may help to establish open and standardised protocols for image acquisition, reconstruction, and analysis [34–36].

A limitation of this study is the relatively low number of patients and the small number of events for both endpoints. At the participating institutions, CT scans during treatment are generally not acquired as part of the clinical routine, except for treatment re-planning. Therefore, only data from clinical imaging trials were available for analysis. We accounted for the limited data by comparing the average performance of several feature selection methods and machine learning algorithms between the time points. We also performed an additional cross-validation experiment on the combined cohort. Both analyses showed a prognostic advantage of in-treatment imaging. In addition, we repeated the analyses excluding 12 patients with differing CT-acquisition parameters in the exploratory cohort, leading to similar results (not shown). Further validation of our findings is planned through retrospective analysis of additional data sets from other centres and data recorded in an on-going prospective clinical study.

An alternative to in-treatment CT may be CBCT, which is routinely acquired in many centres during RCT for quality assurance, such as treatment position verification. However, the applicability of CBCT for radiomic risk modelling requires further investigation [37]. For instance, the image quality of CBCT is low in comparison to conventional CT imaging, and the limited field of view may not be large enough to cover large tumours. These limitations may negatively affect the accuracy of radiomic risk models [38], but their influence may be somewhat mitigated by improved image reconstruction algorithms.

Radiomic risk modelling may not be limited to prognostically predicting survival endpoints, but may be used to predict the risk for occurrence of late radiation-induced side effects as well. For instance, radiomic models were recently built to predict the occurrence of xerostomia and sticky saliva for HNSCC patients [39,40]. Incorporating imaging during treatment to predict late side effects may lead to a higher prognostic accuracy, *e.g.*, by capturing RCT-induced reactions of the normal tissue.

In the present study we showed that the incorporation of CT imaging acquired during treatment may be a promising way to improve radiomic risk models. Moreover, the investigated time point (second week of treatment) is suitable to make an early treatment adaptation in patients not responding to radio(chemo) therapy.

Conflict of interest

The authors state that the research presented in this manuscript is free of conflicts of interest.

Acknowledgments

The project is supported by the Federal Ministry of Education and Research (BMBF-0371N52).

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Competing financial interests

The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.radonc.2018.07.020.

References

[1] Baumann M, Krause M, Overgaard J, Debus J, Bentzen SM, Daartz J, et al. Radiation oncology in the era of precision medicine. Nat Rev Cancer 2016;16:234–49.
[2] El Naqa I, Grigsby PW, Apte A, Kidd E, Donnelly E, Khullar D, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit 2009;42:1162–71.
[3] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5.
[4] Parmar C, Leijenaar RTH, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. Sci Rep 2015;5:1–10.
[5] Kickingereder P, Götz M, Wick A, Neuberger U, Schlemmer H, Radbruch A, et al. OS4.6 Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. Neuro Oncol 2016;18.
[6] Song J, Liu Z, Zhong W, Huang Y, Ma Z, Dong D, et al. Non-small cell lung cancer: quantitative phenotypic analysis of CT images as a potential marker of prognosis. Sci Rep 2016;6:1–9.
[7] King AD, Chow K-K, Yu K-H, Mo FKF, Yeung DKW, Yuan J, et al. Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted MR imaging for the prediction of treatment response. Radiology 2013;266:531–8.
[8] Dietz A, Vanselow B, Rudat V, Conradt C, Weidauer H, Kallinowski F, et al. Prognostic impact of reoxygenation in advanced cancer of the head and neck during the initial course of chemoradiation or radiotherapy alone. Head Neck 2003;25:50–8.
[9] Ljungkvist ASE, Bussink J, Kaanders JHAM, van der Kogel AJ. Dynamics of tumor hypoxia measured with bioreductive hypoxic cell markers. Radiat Res 2007;167:127–45.
[10] Linge A, Lohaus F, Löck S, Nowak A, Gudziol V, Valentini C, et al. HPV status, cancer stem cell marker expression, hypoxia gene signatures and tumour volume identify good prognosis subgroups in patients with HNSCC after primary radiochemotherapy: A multicentre retrospective study of the German Cancer Consortium Radiation. Radiother Oncol 2016;121:364–73.
[11] van Putten LM. Tumour reoxygenation during fractionated radiotherapy; studies with a transplantable mouse osteosarcoma. Eur J Cancer 1968;4:172–82.
[12] Yaromina A, Kroeber T, Meinzer A, Boeke S, Thames H, Baumann M, et al. Exploratory study of the prognostic value of microenvironmental parameters during fractionated irradiation in human squamous cell carcinoma xenografts. Int J Radiat Oncol Biol Phys 2011;80:1205–13.
[13] Hentschel M, Appold S, Schreiber A, Abolmaali N, Abramyuk A, Dörr W, et al. Early FDG PET at 10 or 20 Gy under chemoradiotherapy is prognostic for locoregional control and overall survival in patients with head and neck cancer. Eur J Nucl Med Mol Imaging 2011;38:1203–11.
[14] Zips D, Zöphel K, Abolmaali N, Perrin R, Abramyuk A, Haase R, et al. Exploratory prospective trial of hypoxia-specific PET imaging during radiochemotherapy in patients with locally advanced head-and-neck cancer. Radiother Oncol 2012;105:21–8.
[15] Löck S, Perrin R, Seidlitz A, Bandurska-luque A. Residual tumour hypoxia in head-and-neck cancer patients undergoing primary radiochemotherapy, final results of a prospective trial on repeat FMISO-PET imaging. Radiother Oncol 2017;124:533–40.
[16] Cunliffe A, Armato SG, Castillo R, Pham N, Guerrero T, Al-Hallaq HA. Lung texture in serial thoracic computed tomography scans: Correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. Int J Radiat Oncol Biol Phys 2015;91:1048–56.
[17] Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, et al. Delta-radiomics features for the prediction of patient outcomes in non–small cell lung cancer. Sci Rep 2017;7:588.
[18] van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Lambin P. Feature selection methodology for longitudinal cone-beam CT radiomics. Acta Oncol (Madr) 2017;56:1537–43.
[19] Shafiq-ul-hassan M, Zhang GG, Latifi K, Ullah G, Hunt DC, Balagurunathan Y, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 2017;44:1050–62.
[20] Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities 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;60:5471–96.
[21] Coroller TP, Grossmann P, Hou Y, Rios-Velazquez E, Leijenaar RTH, Hermann G, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015;114:345–50.
[22] Zwanenburg A, Leger S, Vallières M, Löck S, Initiative for the IBS. Image biomarker standardisation initiative. arXiv:161207003 [csCV] 2016.
[23] Leijenaar RTH, Carvalho S, Velazquez ER, van Elmpt WJC, Parmar C, Hoekstra OS, et al. Stability of FDG-PET radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol (Madr) 2013;52:1391–7.
[24] Leger S, Zwanenburg A, Pilz K, Lohaus F, Linge A, Zöphel K, et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. Sci Rep 2017;7:13206.
[25] Walker M, Kublin JG, Zunt JR. Fast R functions for robust correlations and hierarchical clustering. J Stat Softw 2009;42:115–25.
[26] Dietterich TG. Ensemble methods in machine learning. Mult Classif Syst 2000;1857:1–15.
[27] Harrell FE, Lee KL, Mark DB. Prognostic/clinical prediction models: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Tutorials Biostat Stat Methods Clin Stud 2005;1:223–49.
[28] Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation. Stat Med 2004;23:2109–23.
[29] Wiedenmann NE, Bucher S, Hentschel M, Mix M, Vach W, Bittner MI, et al. Serial [18F]-fluoromisonidazole PET during radiochemotherapy for locally advanced head and neck cancer and its correlation with outcome. Radiother Oncol 2015;117:113–7.
[30] Stadler P, Feldmann HJ, Creighton C, Kau R, Molls M. Changes in tumor oxygenation during combined treatment with split-course radiotherapy and chemotherapy in patients with head and neck cancer. Radiother Oncol 1998;48:157–64.
[31] Linge A, Lock S, Gudziol V, Nowak A, Lohaus F, Von Neubeck C, et al. Low cancer stem cell marker expression and low hypoxia identify good prognosis subgroups in HPV(-) HNSCC after postoperative radiochemotherapy: A multicenter study of the DKTK-ROG. Clin Cancer Res 2016;22:2639–49.
[32] Zhao B, Tan Y, Tsai W-Y, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep 2016;6:1–7.
[33] Kim H, Park CM, Lee M, Park SJ, Song YS, Lee JH, et al. Impact of reconstruction algorithms on CT radiomic features of pulmonary tumors: analysis of intra- and inter-reader variability and inter-reconstruction algorithm variability. PLoS One 2016;11:1–11.
[34] Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC. Group F the. Quantitative imaging test approval and biomarker qualification: interrelated but distinct activities. Radiology 2011;259:875–84.
[35] Buckler AJ, Bresolin L, Dunnick MNR, Sullivan DC. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. Radiology 2011;258:906–14.
[36] Clarke LP, Nordstrom RJ, Zhang H, Tandon P, Zhang Y, Redmond G, et al. The quantitative imaging network: NCI's historical perspective and planned goals. Transl Oncol 2014;7:1–4.
[37] Van Timmeren JE, Leijenaar RTH, Van Elmpt W, Lambin P. Are planning CT radiomics and cone-beam CT radiomics interchangeable? Radiother Oncol 2016:446–7.
[38] van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Oberije C, Monshouwer R, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol 2017;123:363–9.
[39] van Dijk LV, Brouwer CL. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. Radiother Oncol 2016;122:185–91.
[40] van Dijk L, Noordzij W, Brouwer CL, Boellaard R, Burgerhof JG, Langendijk JA, et al. 18F-FDG PET image biomarkers improve prediction of late radiation-induced xerostomia. Radiother Oncol 2017.