Definition and validation of a radiomics signature for loco-regional tumour control in patients with locally advanced head and neck squamous cell carcinoma

Supplementary material:

Supplementary Table 1. Number of patients in each sub-cohort with the corresponding treatment period and allocation to the discovery or validation cohort. DKTK: German cancer research center, UKD: University Hospital Dresden.

Sub-cohort	Treatment period	Patient number	Allocation
Retrospective primary DKTK HNSCC cohort (6 partner sites) [1]	2005-2011	147	Discovery
Additional DKTK cohort from partner site Dresden (UKD) [2]	2002-2014	86	Discovery
FDG cohort from UKD and Radiotherapy centre Dresden Friedrichstadt	2005-2009	20	Validation
FMISO cohort (UKD) [3,4]	2006-2013	51	Validation
FMISO cohort from Tübingen [5]	2008-2013	14	Validation

Supplementary Table 2. CT acquisition and reconstruction settings for the discovery and validation cohort.

Image acquisition parameters		Discovery cohort (n=233)	Validation cohort (n=85)
Voxel spacing (x,y) in mm	(0.85,0.85)	1	0
	(0.87,0.87)	2	0
	(0.88,0.88)	1	0
	(0.90,0.90)	1	0
	(0.92,0.92)	1	0
	(0.93,0.93)	1	0
	(0.94,0.94)	3	0
	(0.96,0.96)	2	0
	(0.97,0.97)	3	0
	(0.98,0.98)	141	13
	(1.17,1.17)	21	0
	(1.27,1.27)	26	14
	(1.36,1.36)	29	58
z in mm	2	36	0
	2.5	22	0
	3	74	27
	3.75	1	0
	5	100	58
Reconstruction kernel	B10s	20	1
	B20f	3	51
	B20s	1	1
	B30f	2	0
	B30s	29	0
	B31f	19	12
	B31s	16	0
	B40f	1	0
	B40s	1	0
	B50s	9	0
	59.10.AB50	12	0
	Missing	120	20
Mean exposure mA		181.27 (Missing:59)	76.78 (Missing:14)
Manufacturer	Siemens	99	66
	MDS Nordion	79	5
	BrainLAB	2	0
	GE Medical Systems	31	14
	Picker International	2	0
	Philips	20	0
Scanner model	Biograph16	19	5
	Emotion	9	0
	Sensation16	4	53
	Helax TMS	79	0
	Somatom PLUS4	16	0
	Sensation Open	49	0
	PatXfer RT	2	0
	Lightspeed Ultra	23	0
	PQ5000	2	0
	Brilliance Big Bore	20	0
	Volume Zoom	2	0
	None	8	15
Mean exposure time in ms		733.75 (Missing:59)	508.75 (Missing:14)
Tube voltage in kV	120	86	71

|		
130	9	0
140	16	0
Missing	122	14

mA: milliamps, ms: milliseconds kV: kilovolts

Supplementary Table 3. Settings used for radiomics image processing of the CT scans and feature computation.

Image interpolation	
Interpolation method	Cubic spline
Voxel dimensions in mm^3	1 x 1 x 1
Anti-aliasing smoothing parameter β[6]	0.98
ROI interpolation	
Interpolation method	Cubic spline
Inclusion threshold	0.5
Discretisation	
Discretisation method	Fixed Bin Number (FBN) of 32 bins
Intensity Volume Histogram discretisation method	Fixed Bin Number (FBN) of 1000 bins
Image transformation	
Image filter	Mean-Intensity Laplacian of Gaussian (1,2,3,4,5 mm)
Texture matrices	
Grey-level Run Length Matrix (GLRLM)	Calculation method: 3D Merge method: volume merge (IBSI: IAZD)
Grey-level Size Zone Matrix (GLSZM)	Calculation method: 3D (IBSI: KOBO)
Neighbourhood Grey Tone Difference Matrix (NGTDM)	Calculation method: 3D (IBSI: KOBO)
Neighbourhood Grey Level Dependence Matrix (NGLDM)	Distance for neighborhood: 1.8 voxels Difference level: 0.0 Calculation method: 3D (IBSI: KOBO)
Grey Level Co-occurrence Matrix (GLCM)	Distance for neighborhood: 1.0 voxels Calculation method: 3D Merge method: volume merge (IBSI: IAZD)
Grey Level Distance Zone Matrix (GLDZM)	Calculation method: 3D (IBSI: KOBO)
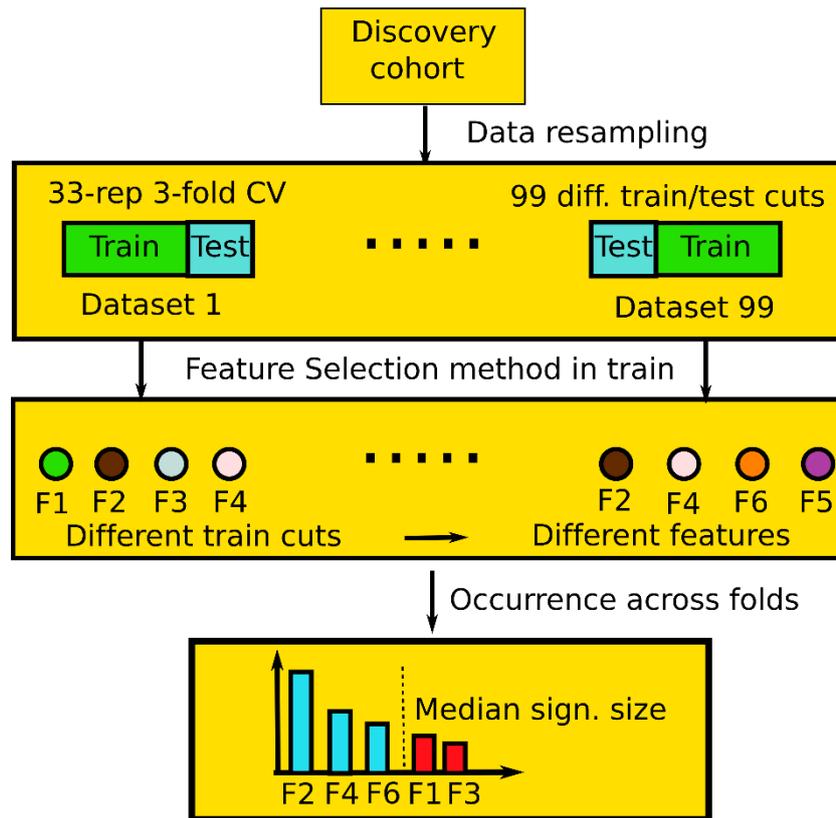
Supplementary Figure 1. Schematic overview of the feature selection performed within the machine learning framework using the discovery cohort. Data was split into cross-validation (CV) runs where feature selection (using Spearman, Minimum Redundancy Maximum Relevance (MRMR) and Lasso Cox methods) were performed for each set of CV training folds. Different features were chosen in each run. An aggregation was performed by ranking the features by occurrence from most frequent to least frequent across CV runs. Finally, only the features that had a ranking equal or higher to the median signature size across CV runs were chosen.
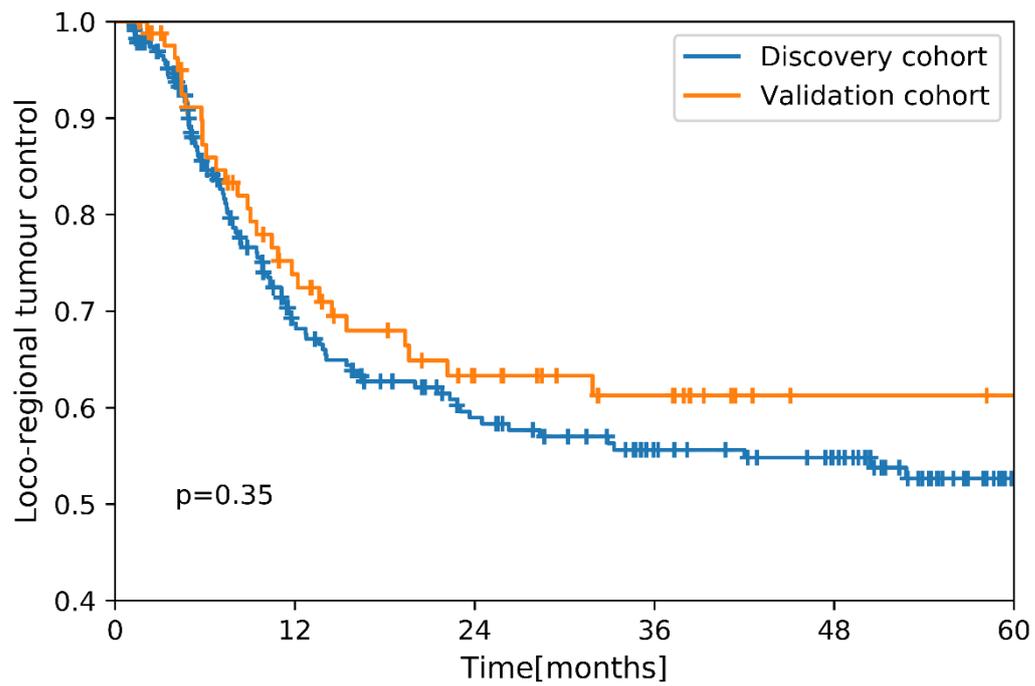
Supplementary Figure 2. Kaplan-Meier curves for discovery (blue) and validation (orange) cohort. There was no statistically significant difference between the cohorts regarding loco-regional tumour control (p-value calculated by log-rank test).

Supplementary Table 4. Association of clinical parameters with loco-regional control (LRC) via univariable Cox regression in the discovery cohort with concordance index (C-Index), hazard ratio (HR) with 95% confidence interval (CI) and p-value.

Clinical parameter	C-Index	HR (95% CI)	p-value
GTV (cm³)	0.59	1.336 [1.075-1.662]	**0.006**
Age (years)	0.53	0.990 [0.968-1.013]	0.42
Total dose (Gy)	0.55	0.764 [0.608-0.960]	**0.021**
Gender (male (ref) vs.female)	0.51	0.717 [0.298-1.726]	0.51
Tumour site (Oropharynx vs. others (ref))	0.50	0.999 [0.654-1.515]	1
UICC stage (2010) (<4 (ref) vs. 4)	0.52	1.871 [0.765-4.615]	0.15
Grading (<2 (ref) vs. >=2)	0.51	0.663 [0.243-1.895]	0.42
cT stage (<4 vs. 4 (ref))	0.52	0.820 [0.532-1.279]	0.38
cN stage (<2 vs. >=2 (ref))	0.53	0.708 [0.412-1.217]	0.21
p16 status (negative (ref) vs.positive)	0.52	0.537 [0.321-1.913]	0.11
HPV16 DNA (negative (ref) vs.positive)	0.55	0.434 [0.132-1.422]	0.32
Alcohol (non-pos (ref) vs. positive)	0.49	0.995 [0.658-1.505]	1
Smoking (non-pos (ref) vs. positive)	0.52	1.207 [0.693-2.102]	0.55

Supplementary Table 5. performances in cross-validation (CV) of the clinical features for every feature selection and model combination and occurrences of the 3 most-occurring features. Occurrences depend only on feature selection.

Model	Feature Selection	C-Index train (95% CI)	C-Index validation (95% CI)	Occurrences
Cox	Spearman	0.65 [0.59-0.70]	0.55 [0.50-0.61]	GTV:81.8%
				Dose:29.4%
				Alcohol:23.3%
Cox	MRMR	0.65 [0.59-0.70]	0.56 [0.49-0.62]	GTV:94.9%
				Dose:29.4%
				p16:21.3%
Cox	Lasso-Cox	0.66 [0.60-0.71]	0.55 [0.48-0.61]	GTV:98.7%
				Dose:27.4%
				N-stage:20.4%
BGLM Cox	Spearman	0.66 [0.60-0.71]	0.54 [0.47–0.61]	GTV:81.8%
				Dose:29.4%
				Alcohol:23.3%
BGLM Cox	MRMR	0.66 [0.60-0.71]	0.53 [0.46-0.59]	GTV:94.9%
				Dose:29.4%
				p16:21.3%
BGLM Cox	Lasso-Cox	0.66 [0.60-0.72]	0.53 [0.46-0.59]	GTV:98.7%
				Dose:27.4%
				N-stage:20.4%
RSF	Spearman	0.73 [0.69-0.79]	0.54 [0.46-0.60]	GTV:81.8%
				Dose:29.4%
				Alcohol:23.3%
RSF	MRMR	0.74 [0.69-0.78]	0.55 [0.49-0.61]	GTV:94.9%
				Dose:29.4%
				p16:21.3%
RSF	Lasso-Cox	0.74 [0.69-0.79]	0.54 [0.47-0.60]	GTV:98.7%
				Dose:27.4%
				N-stage:20.4%

Supplementary Table 6. Clinical model: table with model information for the Cox univariate regression on the discovery cohort with the chosen clinical feature. Information displayed is hazard ratio (HR) with 95% confidence interval (CI), p-value of model coefficient, z-shift (mean), z-scale (standard deviation) of the feature for z-transformation and λ parameter for Yeo-Johnson transform

Feature	HR [95% CI]	p-value	z-shift	z-scale	λ
GTV (cm³)	1.336 [1.075-1.662]	0.009	3.457	0.881	0

GTV: gross tumour volume

Supplementary Table 7. Cluster representative CT features along with their type and lower boundary of the 95% CI of the intraclass correlation coefficient (ICC). Definitions for the features can be found in [7].

Feature	Type	ICC low
loc_peak_loc	Local-intensity based	0.879
stat_mean	Statistical	0.979
stat_median	Statistical	0.997
stat_min	Statistical	0.771
stat_p10	Statistical	0.972
stat_p90	Statistical	0.992
stat_iqr	Statistical	0.994
stat_range	Statistical	0.882
stat_qcod	Statistical	0.960
stat_rms	Statistical	0.943
ivh_v10	Intensity Volume Hist.	0.986
ivh_diff_v25_v75	Intensity Volume Hist.	0.800
ih_skew_fbn_n32	Intensity Histogram	0.924
ih_rmad_fbn_n32	Intensity Histogram	0.880
ih_qcod_fbn_n32	Intensity Histogram	0.819
morph_pca_elongation	Morphological	0.994
morph_pca_flatness	Morphological	0.990
morph_vol_dens_ombb	Morphological	0.874
morph_vol_dens_aee	Morphological	0.952
morph_moran_i	Morphological	0.926
morph_geary_c	Morphological	0.942
cm_corr_d1_3d_avg_fbn_n32	Texture (GLCM)	0.881
cm_clust_shade_d1_3d_v_mrg_fbn_n32	Texture (GLCM)	0.939
cm_clust_prom_d1_3d_v_mrg_fbn_n32	Texture (GLCM)	0.913
cm_info_corr1_d1_3d_v_mrg_fbn_n32	Texture (GLCM)	0.909
cm_info_corr2_d1_3d_v_mrg_fbn_n32	Texture (GLCM)	0.821
cm_joint_entr_d1_3d_v_mrg_fbn_n32	Texture (GLCM)	0.808
rlm_glnu_norm_3d_avg_fbn_n32	Texture (GLRLM)	0.904
rlm_glnu_3d_mrg_fbn_n32	Texture (GLRLM)	0.897
dzm_ldhge_3d_fbn_n32	Texture (GLDZM)	0.927

dzm_zdnu_norm_3d_fbn_n32	Texture (GLDZM)	0.863
ngl_lgce_d1_a0.0_3d_fbn_n32	Texture (NGLDM)	0.908
ngl_hdlge_d1_a0.0_3d_fbn_n32	Texture (NGLDM)	0.838
ngl_dc_var_d1_a0.0_3d_fbn_n32	Texture (NGLDM)	0.781
log_loc_peak_glob	Local-intensity based.	0.856
log_stat_mean	Intensity-Volume Hist.	0.984
log_stat_skew	Statistical	0.923
log_stat_kurt	Statistical	0.936
log_stat_min	Statistical	0.942
log_stat_p10	Statistical	0.979
log_stat_p90	Statistical	0.966
log_stat_max	Statistical	0.839
log_stat_rms	Statistical	0.950
log_ivh_v50	Intensity Volume Hist.	0.791
log_ivh_i75	Intensity Volume Hist.	0.883
log_morph_integ_int	Morphological	0.924
log_morph_moran_i	Morphological	0.889
log_morph_geary_c	Morphological	0.950
log_cm_info_corr2_d1_3d_avg_fbn_n32	Texture (GLCM)	0.948
log_cm_corr_d1_3d_v_mrg_fbn_n32	Texture (GLCM)	0.824
log_cm_clust_prom_d1_3d_v_mrg_fbn_n32	Texture (GLCM)	0.878
log_cm_info_corr1_d1_3d_v_mrg_fbn_n32	Texture (GLCM)	0.924
log_rlm_glnu_norm_3d_avg_fbn_n32	Texture (GLRLM)	0.932
log_szm_hgze_3d_fbn_n32	Texture (GLSZM)	0.801
log_szm_glnu_3d_fbn_n32	Texture (GLSZM)	0.866
log_dzm_sdhge_3d_fbn_n32	Texture (GLDZM)	0.842
log_dzm_ldhge_3d_fbn_n32	Texture (GLDZM)	0.959
log_dzm_zdnu_3d_fbn_n32	Texture (GLDZM)	0.794
log_ngl_hdlge_d1_a0.0_3d_fbn_n32	Texture (NGLDM)	0.807
log_ngl_ldhge_d1_a0.0_3d_fbn_n32	Texture (NGLDM)	0.923
log_ngl_hdhge_d1_a0.0_3d_fbn_n32	Texture (NGLDM)	0.914
log_ngl_dc_var_d1_a0.0_3d_fbn_n32	Texture (NGLDM)	0.842

Hyperparameter optimisation:

Hyperparameter optimisation is conducted via the SMBO algorithm [8]. Beginning from a random configuration of hyperparameters, the objective is to optimise an objective function. A random forest is trained on the initial sets of hyperparameters to predict the objective score, and continuously updated as the hyperparameter space is sampled. The algorithm chooses new contender sets of hyperparameters based on the expected improvement of the objective and uses those hyperparameters in models trained using the same bootstraps of the training data as the current best hyperparameter set. If a contender set of hyperparameters is found that improves over the best-known set, this replaces the best-known set. This procedure is repeated until no further improvements are found, the parameter space is exhausted or a maximum number of iterations is reached.

The objective function tries to balance model performance (C-Index) in bootstrapped and OOB data. The objective function chosen for the study is given as:

$$F = s_{oob} - |s_{oob} - s_{ib}|$$

Where s_{ib} is the model performance score for the in-bag data and s_{oob} is the OOB performance. This function tries to balance performance between training and internal validation in order to avoid overfitting. This optimization is conducted for every CV fold, meaning that every fold has differently optimized parameters that are used for model building within that fold.

For the CT features, the hyperparameter optimisation was performed with the fixed parameter tumour volume (clinical signature), affecting which CT features were selected.

Supplementary Table 8. Hyperparameters selected for the CT feature set in the CV setting for every feature selection and model combination with the clinical signature fixed. Hyperparameter values are shown with median value and range.

| Model	Feature Selection	Hyperparameters (median [min,max])
Cox	Spearman	Signature size: 3 [1,40]
Cox	MRMR	Signature size: 3[1,7]
Cox	Lasso	Signature size: 3 [1,15]
BGLM	Spearman	Signature size: 2 [1,62] n_boost;1.276 [0.0075, 2.9998] learning_rate: -3.0120 [-4.9807, -0.0056]
BGLM	MRMR	Signature size: 1 [1,4] n_boost; 0.6196 [0.0032, 2.4989] learning_rate: -2.1218 [-4.9927, -0.0162]
BGLM	Lasso	Signature size: 2 [1,10] n_boost;2.0174 [0.0056, 2.9996] learning_rate: -1.1494 [-4.9990, -0.002]
RSF	Spearman	Signature_size: 13 [1, 60] n_tree: 9 [4, 10] sample_size: 0.423 [0.062, 0.997] m_try: 0.298 [0.002, 0.997] node_size: 19 [5, 50] n_split: 0 [0, 0] split_rule: logrank tree_depth: 6[1, 10]
RSF	MRMR	sign_size: 2 [1, 6] n_tree: 9 [5, 10] sample_size: 0.331 [0.075, 0.989] m_try: 0.2801 [0.0092, 0.9891] node_size:16 [5, 50] n_split: 0 [0, 0] split_rule: logrank tree_depth: 5 [1, 10]
RSF	Lasso	Signature size: 2 [1, 11] n_tree: 9 [4, 10] sample_size: 0.343 [0.032, 0.989] m_try: 0.2279 [0.0031, 0.9954] node_size: 20 [5, 50] n_split: 0 [0, 0] split_rule: logrank tree_depth: 4 [1, 10]

Permutation test and model information:

In order to assess feature importance in our final signature, permutation tests of 1000 bootstraps were conducted. Each feature was tested individually as follows. In each bootstrap, the feature values are randomly permuted. Risks were then predicted for the bootstrapped data with the permuted feature using the Cox model. A C-Index was then computed for the predicted risks. This leads to 1000 C-indices for each feature. The distribution of C-indices is then compared with the C-index of the unpermuted data. We derive a p-value for a one-sided test with the alternative hypothesis that permuting the feature decreases the C-index of the model predictions:

$$p_{per} = \frac{n_{CI_{permuted} \geq CI_{original}}}{N}$$

Feature importance is calculated for both the discovery and validation cohort in the manner described above.

Supplementary Table 9. Permutation p-values obtained for all three features in the final model in exploration and validation.

Feature	p-value discovery	p-value validation
GTV (cm³)	0.084	0.001
stat_p10	0.037	0.004
log_ngl_hdhge	0.068	0.25
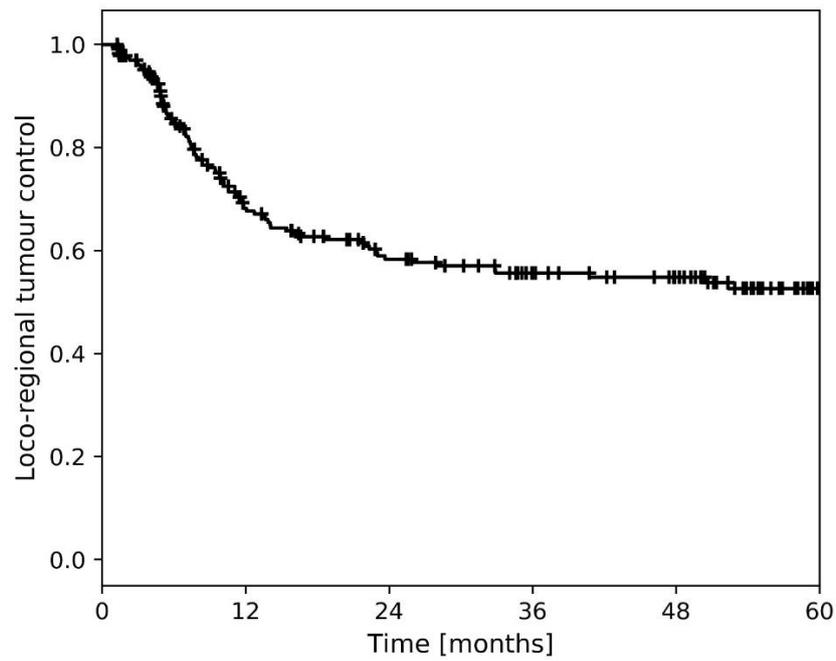
Supplementary Figure 3. Baseline loco-regional tumour control over a period of 60 months.

R packages:

The R packages employed in this study for feature selection and modelling are: rlang[9], stats [10], data.table [11], survival [12], cluster [13], randomForestSRC [14], mboost [15], ggplot2 [16] , gtable [17], glmnet [18], ranger [19], scales [20], stringl [21], xml2 [22].

References

1. 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 Oncology Group (DKTK-ROG). Radiother Oncol. 2016;121:364–73.

2. Linge A, Schmidt S, Lohaus F, Krenn C, Bandurska-Luque A, Platzek I, et al. Independent validation of tumour volume, cancer stem cell markers and hypoxia-associated gene expressions for HNSCC after primary radiochemotherapy. Clin Transl Radiat Oncol. 2019;16:40–7.

3. 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.

4. Löck S, Perrin R, Seidlitz A, Bandurska-Luque A, Zschaeck S, Zöphel K, et al. 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.

5. Zwirner K, Hilke FJ, Demidov G, Socarras Fernandez J, Ossowski S, Gani C, et al. Radiogenomics in head and neck cancer: correlation of radiomic heterogeneity and somatic mutations in TP53, FAT1 and KMT2D. Strahlentherapie und Onkol. 2019;195:771–9.

6. Zwanenburg A, Leger S, Agolli L, Pilz K, Troost EGC, Richter C, et al. Assessing robustness of radiomic features by image perturbation. Sci Rep. 2019;9:614.

7. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295:328–38.

8. Hutter F, Hoos HH, Leyton-Brown K. Sequential model-based optimization for general algorithm configuration. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2011.

9. Functions for Base Types and Core R and 'Tidyverse' Features [R package rlang version 0.4.8]. Comprehensive R Archive Network (CRAN); 2020; Available from: https://cran.r-project.org/package=rlang

10. R: The R Project for Statistical Computing . Available from: https://www.r-project.org/

11. Extension of 'data.frame' [R package data.table version 1.13.2]. Comprehensive R Archive Network (CRAN); 2020; Available from: https://cran.r-project.org/package=data.table

12. Therneau TM. Survival Analysis [R package survival version 3.2-7]. Comprehensive R Archive Network (CRAN); 2020; Available from: https://cran.r-project.org/package=survival

13. Maechler M, Struyf A, Hubert M, Hornik K, Studer M, Roudier P. cluster: Cluster Analysis Basics and Extensions. R Packag. version 2.1.0. 2019.

14. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat. 2008;2:841–60.

15. Model-Based Boosting [R package mboost version 2.9-3]. Comprehensive R Archive Network (CRAN); 2020; Available from: https://cran.r-project.org/package=mboost

16. Create Elegant Data Visualisations Using the Grammar of Graphics • ggplot2 .. Available from: https://ggplot2.tidyverse.org/

17. Arrange 'Grobs' in Tables [R package gtable version 0.3.0]. Comprehensive R Archive Network

(CRAN); 2019; Available from: https://cran.r-project.org/package=gtable

18. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw.; 2010;33:1–22.

19. Wright MN, Ziegler A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. J Stat Softw. 2017;77.

20. Wickham H, Seidel D. scales: Scale Functions for Visualization. R Packag version 110. 2019;

21. Software & Data | stringi Package for R — Marek Gagolewski. Available from: https://www.gagolewski.com/software/stringi/

22. Parse XML [R package xml2 version 1.3.2]. Comprehensive R Archive Network (CRAN); 2020; Available from: https://cran.r-project.org/package=xml2