Pre-treatment radiomic features predict individual lymph node failure for head and neck cancer patients

Zhai, Tian-Tian; Langendijk, Johannes A; van Dijk, Lisanne V; van der Schaaf, Arjen; Sommers, Linda; Vemer-van den Hoek, Johanna G M; Bijl, Henk P; Halmos, Gyorgy B; Witjes, Max J H; Oosting, Sjoukje F

Published in:
Radiotherapy and Oncology

DOI:
10.1016/j.radonc.2020.02.005

Publication date:
2020

Radiotherapy and Oncology 146 (2020) 58–65

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Pre-treatment radiomic features predict individual lymph node failure for head and neck cancer patients

Tian-Tian Zhai [a,b,*], Johannes A. Langendijk [a], Lisanne V. van Dijk [a], Arjen van der Schaaf [a], Linda Sommers [a], Johanna G.M. Vemer-van den Hoek [a], Henk P. Bijl [a], Gyorgy B. Halmos [c], Max J.H. Witjes [d], Sjoukje F. Oosting [e], Walter Noordzij [f], Nanna M. Sijtsema [a], Roel J.H.M. Steenbakkers [a]

[a] Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, The Netherlands; [b] Department of Radiation Oncology, Cancer Hospital of Shantou University Medical College, Shantou, China; [c] Department of Otorhinolaryngology, Head and Neck Surgery; [d] Department of Maxillofacial Surgery; [e] Department of Medical Oncology; and [f] Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 28 August 2019
Received in revised form 26 December 2019
Accepted 9 February 2020

Keywords:
Head and neck cancer
Radiomics
Prediction model
Individual nodal failure
Pre-treatment

ABSTRACT

Background and purpose: To develop and validate a pre-treatment radiomics-based prediction model to identify pathological lymph nodes (pLNs) at risk of failures after definitive radiotherapy in head and neck squamous cell carcinoma patients.

Materials and methods: Training and validation cohorts consisted of 165 patients with 558 pLNs and 112 patients with 467 pLNs, respectively. All patients were primarily treated with definitive radiotherapy, with or without systemic treatment. The endpoint was the cumulative incidence of nodal failure. For each pLN, 82 pre-treatment CT radiomic features and 7 clinical features were included in the Cox proportional-hazard analysis.

Results: There were 68 and 23 nodal failures in the training and validation cohorts, respectively. Multivariable analysis revealed three clinical features (T-stage, gender and WHO Performance-status) and two radiomic features (Least-axis-length representing nodal size and gray level co-occurrence matrix based - Correlation representing nodal heterogeneity) as independent prognostic factors. The model showed good discrimination with a c-index of 0.80 (0.69–0.91) in the validation cohort, significantly better than models based on clinical features ($p < 0.001$) or radiomics ($p = 0.003$) alone. High- and low-risk groups were defined by using thresholds of estimated nodal failure risks at 2-year of 60% and 10%, resulting in positive and negative predictive values of 94.4% and 98.7%, respectively.

Conclusion: A pre-treatment prediction model was developed and validated, integrating the quantitative radiomic features of individual lymph nodes with generally used clinical features. Using this prediction model, lymph nodes with a high failure risk can be identified prior to treatment, which might be used to select patients for intensified treatment strategies targeted on individual lymph nodes.

The optimal management of neck node metastases for head and neck squamous cell carcinoma (HNSCC) patients remains to be determined [1–4]. The clinical and radiographic complete nodal response rates after definitive radiotherapy with or without systemic treatment in node positive (N+) HNSCC patients are around 40–50% [1–3]. Neck dissection is generally recommended for patients without complete response (CR), reducing neck failure rates from 15–24% to 6–10% [1,4]. However, it is difficult to identify those patients without neck CR accurately. PET-CT guided surveillance is advised for treatment response assessment with negative predictive value of 95%, while the positive predictive value (PPV) of PET-CT is only around 50–80% [5–7]. The clinical consequence of this low PPV is that 20–50% patients with non-pathological lymph nodes will be diagnosed as pathological or equivocal by PET-CT. Other studies showed that approximately 30–40% of neck dissection specimens harbor viable tumor cells, meaning that most of these patients are over-treated with the risk of severe post-operative complications [8–11]. In addition, in N+ patients with radiographic CR in the neck, the risk of regional failure varies between 2% and 8% [1,12]. This low regional failure rate and the presumed possibility of salvage surgery suggests a wait-and-see policy for patients with CR in the neck. However, only around 20% of regional recurrences are surgically salvageable due to fibrosis in the neck after radiotherapy [13–15]. Therefore, treatment intensification in a selected high failure risk group combined

* Corresponding author at: Department of Radiation Oncology, University of Groningen, University Medical Center Groningen, P.O.Box 30001, 9300 RB Groningen, The Netherlands.
E-mail address: t.zhai@umcg.nl (T.-T. Zhai).

https://doi.org/10.1016/j.radonc.2020.02.005
0167-8140/

with wait-and-see for the low failure risk group might strike a balance between over- and under-treatment.

Around 91% of lymph node failures occur in the high-dose area, which corresponds to the initial nodal gross tumor volume area [8]. If the individual lymph nodes at risk of persistent or recurrent disease can be identified before treatment, selective intensified treatment regimens, like intensified radiation treatment or planned surgical dissection could be implemented for those lymph nodes with the highest risk of failure [3,9,16–18]. However, to be able to apply such strategies, it is essential to identify those pathological lymph nodes that have a high risk of persistence or recurrence [19].

Radiomics refers to the data values generated by the quantification of features describing intensity, shape and textural characteristics of a region of interest in medical images. Radiomics have shown the potential to predict survival, tumour response, side effects, virus status, and genomic information [20–23]. In our previous study, the quantitative computed tomography (CT) based radiomics of the gross tumor volume and pathological lymph nodes showed prognostic value for regional recurrence on the patient level [24]. To our knowledge, pre-treatment prediction of individual lymph node failures in HNSCC using radiomics has not been investigated so far. Since lymph node radiomic features provide information on the individual lymph node phenotypes, which might improve the performance of prediction models estimating the failure risk of each pathological lymph node. Therefore, the aim of this study was to test the hypothesis that the performance of a prediction model for individual nodal failure can be improved by adding radiomic features of individual lymph nodes to the prediction models, consisting of commonly used classical prognostic factors. Such a model could support decision-making not only for individual patients, but also for specific pathological lymph nodes.

Materials and methods

Patient selection and treatment

This was a retrospective analysis of prospectively acquired HNSCC patient data available at the University Medical Center Groningen (UMCG). This study was approved by the medical ethical committee of the UMCG. The study population consisted of 348 consecutive non-surgically treated clinically N+ HNSCC patients between July 2007 and June 2016. All patients were primarily treated with definitive radiotherapy to a total dose of 70 Gy with fractions of 2 Gy in 6–7 weeks, with or without chemotherapy or cetuximab. A more detailed description of the radiation protocol has been published previously [20,25]. The Appendix A presents the patient recruitment pathway as well as the exclusion criteria. In total, 165 patients treated before January 2013 was included in the training cohort and 112 patients treated thereafter was include in the validation cohort [26].

Clinical parameters

The clinical parameters considered as candidate predictors for nodal failure included: gender (male vs. female), T-stage (T3-T4 vs. T1-T2), N-stage (N2-N3 vs. N1), clinical stage (IV vs. III), treatment modality (radiotherapy only vs. radiotherapy with systemic treatment), WHO performance-status (WHO PS; 1–3 vs. 0) and age. All parameters were prospectively collected from our data registration program. T, N and clinical stage were defined according to the 7th edition of the American Joint Committee on Cancer (AJCC) Staging Manual [27]. Due to the important role of human papillomavirus (HPV) status in oropharyngeal cancer (OPC), HPV status (HPV− vs. HPV+) was included in a subgroup analysis of the OPC patients from the training and validation cohorts [28,29]. HPV-status was assessed by p16 immunohistochemistry and confirmed by DNA polymerase chain reaction in case of p16-positivity in OPC patients. In this study cohort, only one and two HPV+ OPC patients were treated with cetuximab in the training and validation cohort, respectively. Therefore, they were not discussed separately [30].

CT image acquisition, radiomic features extraction, and reproducibility evaluation

All patients underwent a standard contrast-enhanced planning CT-scan. The nodes were considered as pathological lymph nodes (pLNs) in cases of positive cytology, presence of necrosis, short-axis diameter ≥10 mm and/or FDG-PET positivity. All pLNs were delineated on the planning CT-scans by experienced head and neck radiation oncologists. Overall, 82 radiomic features were extracted from every pLNs using Matlab (R2014a; Mathworks, Natick, USA) with feature definition and calibration according to "Image biomarker standardisation initiative" and reported following REMARK guideline [31,32]. Scans from 18 patients were used for inter-observer and intra-observer radiomic reproducibility tests. The radiomic features with inter- and intra-class correlation coefficients (ICCs) > 0.75 were considered robust for delineation variation and were included in the further analysis. A more detailed description of the CT scan parameters, radiomic features and reproducibility evaluation is given in Appendix B.

Endpoints

The endpoint was the cumulative incidence of nodal failure, defined as residual or recurrent lymph node metastases within or overlapping with the primary pathological lymph node region before treatment. In contrast with regional or neck failure [33,34], in which the entire neck is considered, nodal failure refers to failure of each separate node. Residual disease was defined as a persistent node at a minimum of 12 weeks after treatment. A recurrent node was defined as a new pathological node after an initial complete response. Recurrent diseases outside the original pathological lymph node region were not considered as events in this analysis. Residual and recurrent diseases are managed similarly in clinic, and thus were analysed together [27]. All nodal failures were contoured on the follow-up CT or MRI scans, and the follow-up imaging was co-registered to the planning CT. Every nodal failure was linked to the original pLN. Nodal failure was confirmed by histological or cytological diagnosis, or obvious lymph nodes with ≥10 mm short-axis diameter, detected on at least two image modalities from CT, MRI, PET-CT and ultrasound. Time to event was defined as the date from the first day of radiotherapy to the date of nodal failure. The lymph nodes without failures that were removed by neck dissection were censored at the date of surgery and others were censored at the date of last follow-up. Patients received systematic follow-up, consisting of clinical head and neck examination and additional imaging in cases of suspicious findings, after treatment every 3 months in the first 2 years and every 6 months thereafter.

Data analysis

Model development and validation

Univariable cox-regression analysis was performed to assess clinical risk factors for nodal failure.

To reduce the probability of overfitting and multi-collinearity, pre-selection was performed for the radiomic features. If the Spearman rank-order correlation between pairs of radiomic features was >0.80, then the radiomic feature with the lower univariable association with the endpoint was excluded from further

analysis [20,35]. All clinical and pre-selected radiomic features were included in a multivariable Cox proportional hazard regression analysis (forward selection based on Likelihood ratio test, $p < 0.05$) to create multivariable clinical, radiomic and combined models.

The complete process of radiomic feature pre-selection and feature selection (multivariable model training) was repeated on 1000 bootstrap samples of the training set according to the TRIPOD guideline [26]. Only the most frequently selected variables were considered in the final clinical, radiomic and combined models. The concordance-index (c-index) was determined to assess the model's discriminative power.

The performances of the final clinical, radiomic and combined models were then tested with the validation cohort.

Nodal failure risk curves and nomogram

The baseline cumulative hazard function H0(t) of the combined model was described in the simplified look-up table in Appendix C, and a nomogram for nodal control probability estimation at 1 and 2 years after treatment was created. Nodal failure risk curves at 2 years were constructed based on the combined model. To determine the cut off points for the optimal positive and negative predictive values (PPV and NPV), all calculated nodal failure probabilities from training and validation cohorts at 2 years were compared with the actual nodal failures.

Subgroup analysis for oropharyngeal cancer

The same analysis procedure was repeated for the subgroup of OPC patients from the training cohort to create OPC-clinical, OPC-radiomic and OPC-combined models. The models trained with the HNSCC group and the OPC group patients were compared and externally validated with the OPC patients from the validation cohort.

Statistical analysis

Statistical analysis was conducted with the R software (version 3.2.1). Two tailed *p*-values <0.05 were considered statistically significant. The chi-square test was used to compare the categorical variables and an independent sample *t*-test was used to compare normally distributed variables between different groups. Model performance was calculated using the Harrell's c-index. The z-score test was used to test the difference between two c-indices. The Hosmer–Lemeshow (HL) test was used to test the calibration for the nodal failure risk at 2 years, *p*-values >0.05 represent good calibration.

Results

The training cohort consisted of 165 patients with 558 pLNs. There were 68 (12.2%) nodal failures in 37 (22.4%) patients during follow-up and the median follow-up was 36.1 (range: 2.9–130.2) months. The validation cohort consisted of 112 patients with 467 pLNs. There were 23 (4.9%) nodal failures in 19 (17.0%) patients during follow-up and the median follow-up was 30.8 (range: 3.7–65.8) months. The other failures including local failure, regional failure, distant metastasis and death were recorded on the patient level and summarized in Appendix D. Less events were seen in the validation cohort due to the shorter follow up and more HPV-positive OPC patients compared with the training cohort.

In total, 87 of 91 nodal failures occurred within 2 years after treatment. The clinical characteristics of the training and validation cohorts are summarized in Table 1. The patients in the validation cohort were older than those in the training cohort. There were significantly more patients with tumors originating from the hypopharynx and larynx and fewer HPV-positive OPC patients in the training cohort than that in the validation cohort. The two cohorts were well balanced regarding all other clinical characteristics.

The average ICCs of inter- and intra-observer agreement of all radiomic features were 0.94 and 0.93. There were 6 radiomic features with an ICC < 0.75 that were excluded from further analysis (Appendix E).

Gender, T-stage, N-stage and WHO PS showed significant associations with nodal failure in the univariable analysis (Appendix F). These parameters were included in the final clinical model as independent prognostic factors for nodal failure (Table 2).

Sixty of the 82 radiomic features were significantly associated with nodal failure in the univariable analysis. The pre-selection and feature selection (multivariable model training) were repeated on 1000 bootstrap samples (Appendix G). The radiomic features that were selected more than 750 times and that were significantly associated with nodal failure in the multivariable analysis were Least-axis-length of lymph node (LALLN, representing nodal size) and Correlation of gray level co-occurrence matrix (Corr-GLCM, representing nodal heterogeneity) (Table 2).

All clinical and radiomic features were included in the bootstrapped variable selection for the development of the combined model. All variables that were selected in the final clinical and radiomic models were also selected and remained significant in the final combined model except for N-stage (Table 2 and Appendix G). The performances of the final clinical, radiomic and combined models in the training and validation cohorts are summarized in Fig. 1. The c-index of the radiomic model was 0.84 (95% confidence interval (CI): 0.77–0.91), slightly but not significantly ($p = 0.093$) better than that of the clinical model 0.78 (95%CI: 0.71–0.85). When tested in the validation cohort, the c-index of the radiomic model was 0.79 (95%CI: 0.71–0.87) and was higher than that of the clinical model (0.69; 95%CI: 0.59–0.79). The combined model performed significantly better than the clinical model ($p < 0.001$) and radiomic model ($p = 0.003$), with a c-index of 0.90 (95% CI: 0.83–0.97) in the training cohort and 0.80 (95% CI: 0.69–0.91) in the validation cohort.

The relationship between the nodal failure risk and the radiomic features (LALLN and Corr-GLCM) is shown in Fig. 2 for female patients and Fig. 3 for male patients. A nomogram was developed as the graphic representation of the combined model and can be found in Appendix H. The HL-test for the probability of nodal failure at 2 years was not significant in the training cohort ($p = 0.51$) nor in the validation cohort ($p = 0.14$), indicating that there is a good agreement between the estimated nodal failure risk by using the model and actual nodal failure risk based on the datasets. By using cut-off values of the estimated risks of 60% and 10% for high- and low-risk groups, respectively, the PPV and NPV were 94.4% and 98.7% (Fig. 4). The lymph nodes were stratified into high-, intermediate- and low-risk groups according to the cut-off values.

The subgroup analysis included all OPC patients with known HPV-status. In the training cohort, 73 OPC patients with 268 LNs resulted in 32 (11.9%) nodal failures from 21 patients. In the validation cohort, 64 OPC patients with 274 LNs resulted in 15 (5.5%) nodal failures from 12 patients. Except for HPV status, no significant differences in patient characteristics were found between the two subgroups (Appendix I). Based on this analysis, we constructed OPC-clinical, OPC-radiomic and OPC-combined models as shown in Appendix J. HPV status was identified as a significant feature in the OPC-clinical model. However, in the combined model, the textural feature short run high grey level emphasis (SRHGE) of GLRLM was selected instead of HPV status because of its larger predictive performance. The OPC-clinical (c-index: 0.68; 95%CI: 0.54–0.82), OPC-radiomic c-index: 0.78; 95% CI:0.63–0.93) and OPC-combined (c-index: 0.78; 95%CI: 0.65–0.91) models performed similarly to the clinical (c-index: 0.68; 95%CI: 0.52–0.84),

Table 1
Characteristics of the head and neck squamous cell carcinoma patients in the training and validation cohorts.

Characteristic	Training cohort		Validation cohort		*p*-value
	n = 165	%	*n* = 112	%	
Age at diagnosis (median ± SD, years)	60 ± 9		64 ± 10		0.025[b]
Gender					0.302[c]
Female	45	27.3	37	33.0	
Male	120	72.7	75	67.0	
T-stage[a]					0.154[c]
T1	11	6.7	13	11.6	
T2	35	21.2	22	19.6	
T3	49	29.7	22	19.6	
T4	70	42.4	55	49.1	
N-stage[a]					0.226[c]
N1	33	20.0	19	17.0	
N2	123	74.5	91	81.3	
N3	9	5.5	2	1.8	
Clinical stage[a]					0.400[c]
III	25	15.2	13	11.6	
IV	140	84.8	99	88.4	
Treatment modality					0.369[c]
RT only	49	29.7	39	34.8	
RT with systemic treatment	116	70.3	73	65.2	
WHO PS					0.310[c]
0	105	63.6	65	58.0	
1	52	31.5	35	31.3	
2	7	4.2	10	8.9	
3	1	0.6	2	1.8	
Tumor site					0.006[c]
oral cavity	11	6.7	11	9.8	
oropharynx	75	45.5	70	62.5	
nasopharynx	4	2.4	5	4.5	
hypopharynx	33	20.0	11	9.8	
larynx	42	25.5	15	13.4	
HPV status					0.002[c]
OPC HPV-	48	29.1	30	26.8	
OPC HPV+	25	15.2	34	30.4	
OPC unknown	2	1.2	6	5.4	
Not OPC	90	54.5	42	37.5	

Abbreviations: T = tumor; N = lymph node; RT = radiotherapy; WHO PS = World Health Organization performance status; HPV = human papillomavirus; OPC = oropharyngeal cancer.
[a] According to the 7th edition of the AJCC/UICC staging system.
[b] *p*-Value was calculated using the independent sample *t*-test.
[c] *p*-Value was calculated using the chi-square test.

Table 2
Estimated coefficients (β) of clinical, radiomic and combined models.

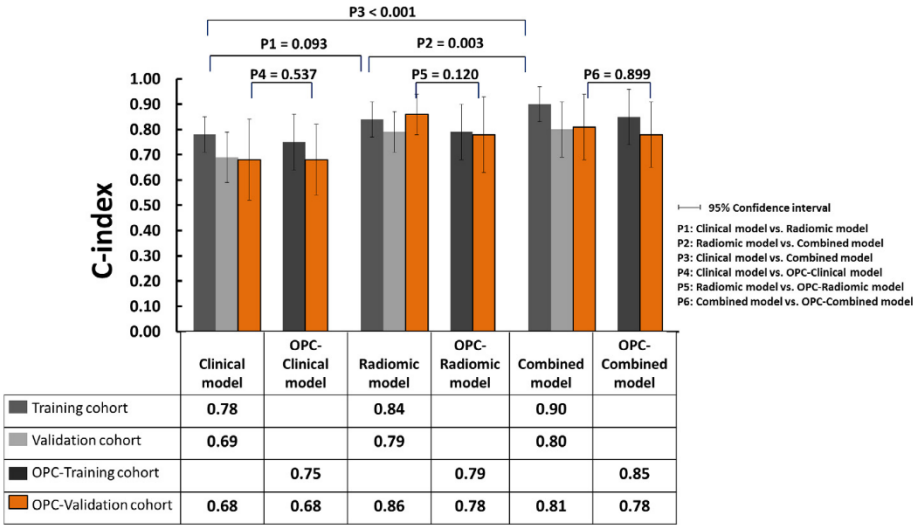
	Clinical model					Radiomic model					Combined model				
	β	Corrected β	HR	HR (95% CI)	*p*-Value	β	Corrected β	HR	HR (95% CI)	*p*-Value	β	Corrected β	HR	HR (95% CI)	*p*-Value
Gender (Male vs. Female)	2.39	2.08	10.97	2.67–45.09	<0.001						2.19	1.88	8.94	2.17–36.81	0.002
T-stage (T3-T4 vs. T1-T2)	1.24	1.08	3.46	1.38–8.73	0.008						1.08	0.93	2.96	1.16–7.50	0.023
N-stage (N2-N3 vs. N1)	0.98	0.85	2.66	1.53–4.63	<0.001						-	-	-	-	-
WHO PS (1–3 vs. 0)	0.90	0.78	2.46	1.46–4.16	0.001						1.06	0.91	2.90	1.71–4.91	<0.001
LALLN* (cm)						0.87	0.78	2.39	1.94–2.95	<0.001	0.83	0.71	2.31	1.89–2.82	<0.001
Corr-GLCM*						3.49	3.14	32.93	3.45–314.65	0.002	2.84	2.44	17.05	1.70–171.06	0.016

Abbreviations: T = tumor; N = lymph node; WHO PS = World Health Organization performance status; LALLN = Least axis length of lymph node; Corr-GLCM = Correlation of grey level co-occurrence matrix; HR = Hazard ratio; CI = confidence interval.
* Radiomic features, LALLN and Corr-GLCM are continuous variables.

radiomic (c-index: 0.86; 95%CI: 0.78–0.94) and combined models (c-index: 0.81; 95%CI: 0.68–0.94) with non-significant *p*-values of 0.537, 0.120, and 0.899) when they were tested in the subgroup of oropharyngeal cancer patients in the validation cohort, respectively (Fig. 1).

Discussion

To our knowledge, this is the first study to develop and validate a pre-treatment prediction model for individual nodal failures. By combining non-invasive quantitative radiomic features of



	Clinical model	OPC-Clinical model	Radiomic model	OPC-Radiomic model	Combined model	OPC-Combined model
Training cohort	0.78		0.84		0.90	
Validation cohort	0.69		0.79		0.80	
OPC-Training cohort		0.75		0.79		0.85
OPC-Validation cohort	0.68	0.68	0.86	0.78	0.81	0.78

Fig. 1. Prediction performances of models. P1, P2 and P3 showed the comparisons between the clinical, radiomic and combined models; P4, P5 and P6 showed the comparisons between the clinical and OPC-clinical models, radiomic and OPC-radiomic models, and combined and OPC-combined models on OPC-validation cohort (the orange bar). Abbreviation: OPC = oropharyngeal cancer.

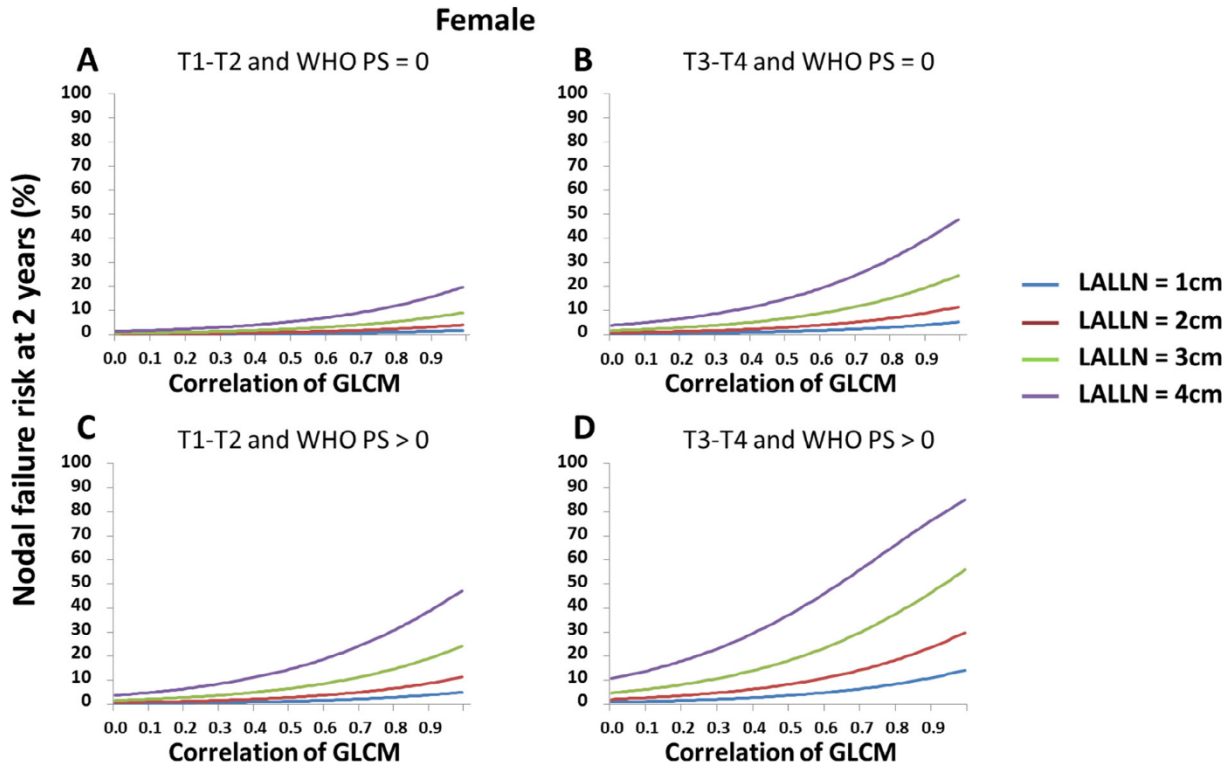


Fig. 2. The risk of nodal failure for female patients at 2 years. Abbreviations: WHO PS = WHO performance score; LALLN = Least-axis-length of lymph node; GLCM = Grey level co-occurrence matrix.

individual lymph nodes with clinical features, it is possible to classify pLNs as low- or high-risk. This might provide new options for defining more personalised treatment strategies.

The significant clinical features in the clinical model (T-stage, N-stage, gender and WHO PS) and OPC-clinical model (T-stage, gender and HPV status) are consistent with earlier reports [1,24,36]. Sixty radiomic features showed significant association with nodal failure in the univariable analysis. Out of the sixty radiomic features, the geometric feature (LALLN) and textural feature (Corr-GLCM) were the most frequently selected radiomic features in the 1000 bootstrap samples and identified as independent prognostic factors in the radiomic and combined models.

LALLN, representing the size of the lymph node, refers to the length of the shortest axis along which the lymph node is extended in three-dimensions (3D). This is consistent with the results reported by Vergeer et al., they found that the lymph node size was a prognostic factor for nodal control. In their study, the nodal volume was used to represent the nodal size [37]. Nodal volume

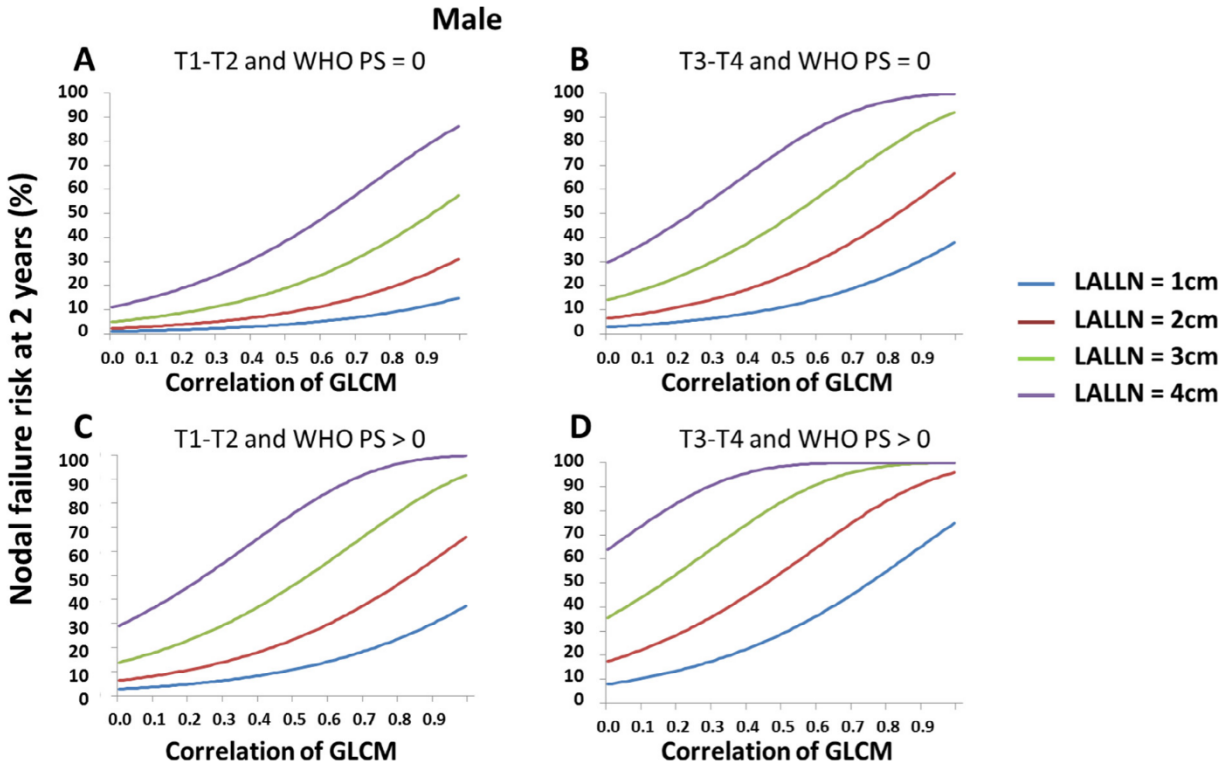


Fig. 3. The risk of nodal failure for male patients at 2 years. Abbreviations: WHO PS = WHO performance score; LALLN = Least axis length of lymph node; GLCM = Grey level co-occurrence matrix.

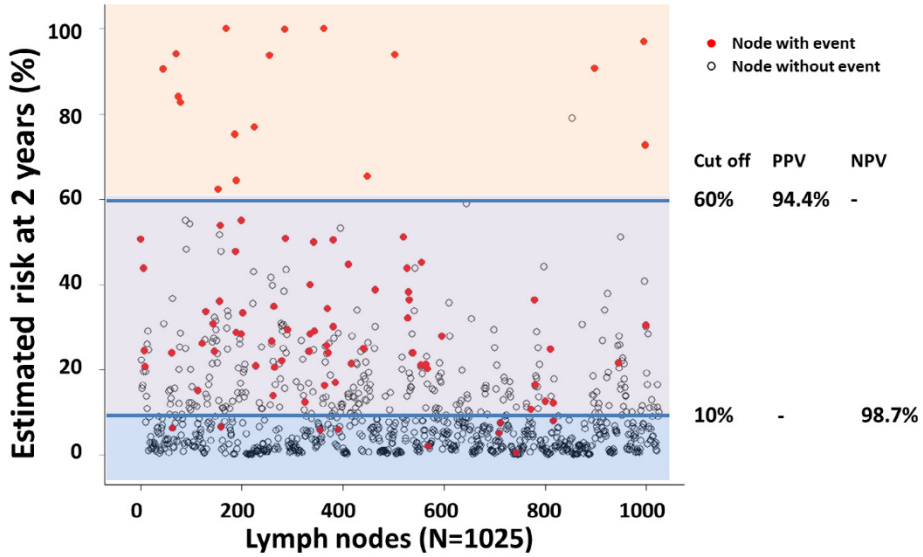


Fig. 4. The failure risk of all lymph nodes from training and validation cohorts at 2 years, calculated with the combined model. The pathological lymph nodes were stratified into three groups: high-risk (orange area), intermediate-risk (purple area) and low-risk (blue area) groups. Abbreviations: PPV = positive predictive value; NPV = negative predictive value.

was also associated with nodal control in the univariable analysis of this study and highly correlated with LALLN (0.84), but performed worse than LALLN in the prediction of nodal control, and nodal volume did not add prognostic information to the model with LALLN. Another prognostic feature was the short-axis diameter of the lymph node, which is also representative for the size of the lymph node. The short-axis diameter is frequently used in radiological reports and defined as the widest diameter perpendicular to the longest axis in the transverse plane. It was highly correlated (0.96) with LALLN and performed similarly to LALLN in this study. However, LALLN was the most selected radiomic feature (868 times of 1000 bootstrapped samples, short-axis diameter was selected 31 times only), indicating its robustness (Appendix G), and therefore was included in the model.

Patients with advanced N-stage tend to have larger LALLN and multiple lymph nodes. The patients with advanced N-stage (7th edition) showed worse nodal control in the clinical univariable and multivariable analysis of this cohort. We believe that the 8th

edition N-stage might have an even stronger association with the nodal failures since it includes extra-nodal extension (ENE) and HPV+ oropharyngeal cancer's new staging system [38]. This is unfortunately not possible to explore for this retrospective study due to the low accuracy of ENE evaluation using existing CT images [39]. However, for patients with more than one pathological lymph node, the use of LALLN for each lymph node is much more informative than the N-stage for individual nodal failure prediction. The combined model with LALLN performed significantly better than models with N-stage.

Corr-GLCM is a textural feature describing the heterogeneity of the lymph node. It describes the correlation of a reference voxel to its neighbours. Haralick et al. showed that Corr-GLCM could be used for quantifying heterogeneity and for distinguishing heterogeneous and homogeneous materials [40]. In this study, the lymph nodes with lower Corr-GLCM values had large areas of similar intensities, i.e. lower heterogeneity, and lower nodal failure risk. Intra-lymph node heterogeneity can be increased by the presence of necrosis, which can be recognized as an area with lower CT-intensities surrounded by an irregular rim of higher CT-intensities in contrast-enhanced CT images [41]. Therefore, Corr-GLCM may indicate necrosis status of lymph nodes.

The textural feature SRHGE, which replaced HPV-status, was selected for the OPC-combined model. This feature emphasises small areas with high CT-intensities (short run length with high grey levels). The association between radiomic features and HPV-status has been shown in a series of reports [42,43], and was also significant in this study with a p-value of 0.001. In the present study, a lower SRHGE was associated with a higher nodal failure risk. Lower SRHGE values are expected in volumes with lower contrast enhancement. Since contrast enhancement is related to local circulation, lower SRHGE values can be expected in hypo-vascular volumes with a higher risk of hypoxia [44,45]. Therefore, higher failure risk in lymph nodes with a lower SRHGE could be associated with hypoxia. Further research is necessary to explore the possible underlying biological mechanisms behind radiomic phenotypes [46,47], which might provide a non-invasive means of assessing these biological features.

In the management of neck node metastases in HNSCC patients, PET-CT surveillance has significantly improved the assessment of response status with a very high NPV of 95% [5–7]. In the randomized controlled PET-NECK trial reported by Mehanna et al., it was shown that PET-CT surveillance spared a neck dissection for approximately 80% of patients [7]. However, PET-CT at 8–12 weeks after radiotherapy has a low PPV of 50–80%, mainly due to radiation-induced inflammatory changes [5–7]. The clinical consequence of this low PPV is that 20–50% patients with non-pathological lymph nodes will be diagnosed as pathological or equivocal by PET-CT. In the PET-NECK trial, 66 out of 266 patients had an incomplete or equivocal imaging response, meaning that 20–50% (13–33 patients) patients have false positive PET-CT [5–7]. Therefore, performing surgical resection for all of those patients is not the optimal workflow. Our combined model based on clinical and radiomics features can identify the lymph nodes with a PPV for the high-risk group of 94.4% and a NPV for the low-risk group of 98.7% prior to treatment. If we could identify the high-risk lymph nodes before treatment, an intensified radiation schedule or lymph node targeted dissection before or after (chemo-)radiation could be arranged to avoid complex clinical decisions on re-irradiation or severe post-operative complications. For the low- and intermediate-risk lymph nodes, a wait-and-see policy could be applied when they have complete PET-CT response. Frequently imaging follow-up is recommended for those intermediate-risk lymph nodes that show incomplete PET-CT response. For the low-risk lymph nodes, a wait-and-see policy could be applied to avoid frequently imaging follow up. This hypothesis should be investigated in follow-up studies. Neck management could be modified by using the pre-treatment prediction model as a supplement to post-treatment PET-CT surveillance. Such a workflow might improve the nodal control rate in the high-risk patients and reduce the number of unnecessary lymph node dissections in the low-risk patients.

A limitation of the current study is that only contrast-enhanced CT images were used. The model is therefore not applicable to CT-scans without contrast-enhancement, since the textural features could differ from those in our study. However, when we used the model without the textural feature (corr-GLCM), model performance was good with a C-statistic of 0.90 in the training cohort and 0.79 in the validation cohort, indicating that the model without corr-GLCM could also be used. To further improve the current combined prediction model for lymph node failure, radiomic features from other image modalities such as MRI, PET-CT and ultrasound could be investigated, as well as the changes of the radiomic features between pre-, during- and post-treatment imaging [48]. Another limitation is the lack of histological confirmation of nodal failure in some of the cases, therefore at least two image modalities were used to confirm the diagnosis. Although we used training and validation datasets in this study, datasets from other institutions using different data acquisition protocols and CT-scanners are needed for further validation.

In conclusion, we developed a multivariable prediction model for nodal failures that can be applied to estimate the risk of failure for individual pathological lymph nodes, based on quantitative and non-invasive radiomic features describing the size and heterogeneity of the whole lymph node in combination with clinical features of the patient. This prediction model allows for an accurate prediction of failure for individual lymph nodes and could be used to guide decisions on treatment strategies customized for individual pathological lymph nodes.

Conflict of interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2020.02.005.

References

[1] Thariat J, Ang KK, Allen PK, Ahamad A, Williams MD, Myers JN, et al. Prediction of neck dissection requirement after definitive radiotherapy for head-and-neck squamous cell carcinoma. Int J Radiat Oncol Biol Phys 2012;82:e367–74.
[2] Brizel DM, Prosnitz RG, Hunter S, Fisher SR, Clough RL, Downey MA, et al. Necessity for adjuvant neck dissection in setting of concurrent chemoradiation for advanced head-and-neck cancer. Int J Radiat Oncol Biol Phys 2004;58:1418–23.
[3] Robbins KT, Shannon K, Vieira F. Superselective neck dissection after chemoradiation. Arch Otolaryngol Head Neck Surg 2007;133:486–9.
[4] Liauw SL, Mancuso AA, Amdur RJ, Morris CG, Villaret DB, Werning JW, et al. Postradiotherapy neck dissection for lymph node-positive head and neck cancer: the use of computed tomography to manage the neck. J Clin Oncol 2006;24:1421–7.
[5] Moeller BJ, Rana V, Cannon BA, Williams MD, Sturgis EM, Ginsberg LE, et al. Prospective risk-adjusted [18 F] fluorodeoxyglucose positron emission tomography and computed tomography assessment of radiation response in head and neck cancer. J Clin Oncol 2018;27:2509–15.
[6] Yao M, Smith RB, Graham MM, Hoffman HT, Tan H, Funk GF, et al. The role of FDG PET in management of neck metastasis from head and neck cancer after definitive radiation treatment. Int J Radiat Oncol Biol Phys 2005;63:991–9.
[7] Mehanna H, Wong WL, McConkey CC, Rahman JK, Robinson M, Hartley AGJ, et al. PET-CT surveillance versus neck dissection in advanced head and neck cancer. N Engl J Med 2016;374:1444–54.

[8] Studer G, Huber GF, Holz E, Glanzmann C. Less may be more: nodal treatment in neck positive head neck cancer patients. Eur Arch Oto-Rhino-Laryngology 2016;273:1549–56.
[9] Donatelli-Lassig AA, Duffy SA, Fowler KE, Ronis DL, Chepeha DB, Terrell JE. The effect of neck dissection on quality of life after chemoradiation. Otolaryngol – Head Neck Surg 2008;139:511–8.
[10] Machtay M, Moughan J, Trotti A, Garden AS, Weber RS, Cooper JS, et al. Factors associated with severe late toxicity after concurrent chemoradiation for locally advanced head and neck cancer: an RTOG analysis. J Clin Oncol 2008;26:3582–9.
[11] van den Bovenkamp K, Dorgelo B, Noordhuis MG, van der Laan BFAM, van der Vegt B, Bijl HP, et al. Viable tumor in salvage neck dissections in head and neck cancer: relation with initial treatment, change of lymph node size and human papillomavirus. Oral Oncol 2018;77:131–6.
[12] Goenka A, Morris LGT, Rao SS, Wolden SL, Wong RJ, Kraus DH, et al. Long-term regional control in the observed neck following definitive chemoradiation for node-positive oropharyngeal squamous cell cancer. Int J Cancer 2013;133:1214–21.
[13] Mabanta SR, Mendenhall WM, Stringer SP, Cassisi NJ. Salvage treatment for neck recurrence after irradiation alone for head and neck squamous cell carcinoma with clinically positive neck nodes. Head Neck 1999;21:591–4.
[14] McHam SA, Adelstein DJ, Rybicki LA, Lavertu P, Esclamado RM, Wood BG, et al. Who merits a neck dissection after definitive chemoradiotherapy for N2–N3 squamous cell head and neck cancer?. Head Neck 2003;25:791–8.
[15] Johansen S, Norman MH, Dale E, Amdal CD, Furre T, Malinen E, et al. Patterns of local-regional recurrence after conformal and intensity-modulated radiotherapy for head and neck cancer. Radiat Oncol 2017;12:87.
[16] Mukhija V, Gupta S, Jacobson AS, Eloy JA, Genden EM. Selective neck dissection following adjuvant therapy for advanced head and neck cancer. Head Neck 2008;31:183–8.
[17] Beitler JJ, Zhang Q, Fu KK, Trotti A, Spencer SA, Jones CU, et al. Final results of local-regional control and late toxicity of RTOG 90–03; a randomized trial of altered fractionation radiation for locally advanced head and neck cancer. Int J Radiat Oncol Biol Phys 2014;89:13–20.
[18] Lacas B, Bourhis J, Overgaard J, Zhang Q, Grégoire V, Nankivell M, et al. Role of radiotherapy fractionation in head and neck cancers (MARCH): an updated meta-analysis. Lancet Oncol 2017;18:1221–37.
[19] Caudell JJ, Torres-Roca JF, Gillies RJ, Enderling H, Kim S, Rishi A, et al. The future of personalised radiotherapy for head and neck cancer. Lancet Oncol 2017;18:e266–73.
[20] Zhai TT, van Dijk LV, Huang BT, Lin ZX, Ribeiro CO, Brouwer CL, et al. Improving the prediction of overall survival for head and neck cancer patients using image biomarkers in combination with clinical parameters. Radiother Oncol 2017;124:256–62.
[21] van Dijk LV, Brouwer CL, van der Schaaf A, Burgerhof JGM, Beukinga RJ, Langendijk JA, et al. CT image biomarkers to improve patient-specific prediction of radiation-induced xerostomia and sticky saliva. Radiother Oncol The Authors 2017;122:185–91.
[22] 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;5:4006.
[23] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749–62.
[24] Zhai TT, Langendijk JA, van Dijk LV, Halmos GB, Witjes MJH, Oosting SF, et al. The prognostic value of CT-based image-biomarkers for head and neck cancer patients treated with definitive (chemo-)radiation. Oral Oncol 2019;95:178–86.
[25] Van Der Laan HP, Van De Water TA, Van Herpt HE, Christianen MEMC, Bijl HP, Korevaar EW, et al. The potential of intensity-modulated proton radiotherapy to reduce swallowing dysfunction in the treatment of head and neck cancer: a planning comparative study. Acta Oncol (Madr) 2013;52:561–9.
[26] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. Eur Radiol 2015;67:1142–51.
[27] Pfister DG, Spencer S, Brizel DM, Burtness B, Busse PM, Caudell JJ, et al. Head and neck cancers, version 1.2015 featured updates to the NCCN guidelines. J Natl Compr Cancer Netw 2015;13:847–56.
[28] Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân F, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010;363:24–35.
[29] Dayyani F, Etzel CJ, Liu M, Ho CH, Lippman SM, Tsao AS. Meta-analysis of the impact of human papillomavirus (HPV) on cancer risk and overall survival in head and neck squamous cell carcinomas (HNSCC). Head Neck Oncol 2010;2:1–11.
[30] Mehanna H, Robinson M, Hartley A, Kong A, Foran B, Fulton-Lieuw T, et al. Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial. Lancet 2019;393:51–60.
[31] Zwanenburg A, Leger S, Vallières M, Löck S, Initiative for the IBS. Image biomarker standardisation initiative. 2016; Available from: http://arxiv.org/abs/1612.07003.
[32] Mcshane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. REporting recommendations for tumour MARKer prognostic studies (REMARK). Eur J Cancer 2005;41:1690–6.
[33] Van Der Schaaf A, Xu CJ, Van Luijk P, Van'T Veld AA, Langendijk JA, Schilstra C. Multivariate modeling of complications with data driven variable selection: guarding against overfitting and effects of data set size. Radiother Oncol 2012;105:115–21.
[34] Silva P, Homer JJ, Slevin NJ, Musgrove BT, Sloan P, Price P, et al. Clinical and biological factors affecting response to radiotherapy in patients with head and neck cancer: a review. Clin Otolaryngol 2007;32:337–45.
[35] Meinshausen Nicolai, Bühlmann Peter. Stability selection. J R Statist Soc B 2010;72:417–73.
[36] van den Brekel MWM, Bindels EMJ, Balm AJM. Prognostic factors in head and neck cancer. Eur J Cancer 2002;38(8):1041–3.
[37] Vergeer MR, Doornaert P, René Leemans C, Buter J, Slotman BJ, Langendijk JA. Control of nodal metastases in squamous cell head and neck cancer treated by radiation therapy or chemoradiation. Radiother Oncol 2006;79:39–44.
[38] Lydiatt W, O'Sullivan B, Patel S. Major changes in head and neck staging for 2018. Am Soc Clin Oncol Educ B 2018:505–14.
[39] Prabhu RS, Magliocca KR, Hanasoge S, Aiken AH, Hudgins PA, Hall WA, et al. Accuracy of computed tomography for predicting pathologic nodal extracapsular extension in patients with head-and-neck cancer undergoing initial surgical resection. Int J Radiat Oncol Biol Phys 2014;88:122–9.
[40] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:610–21.
[41] Miles KA. Tumour angiogenesis and its relation to contrast enhancement on computed tomography: a review. 1999;30:198–205.
[42] Leijenaar RTH, Bogowicz M, Jochems A, Hoebers F, Wesseling F, Huang SH, et al. Development and validation of a radiomic signature to predict HPV (p16) status from standard CT imaging: a multicenter study. Br J Radiol 2018;91:20170498.
[43] Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G, et al. Computed tomography radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma. Int J Radiat Oncol Biol Phys 2017;99:921–8.
[44] Muz B, Azab AK. The role of hypoxia in cancer progression, angiogenesis, metastasis, and resistance to therapy. 2015;3:83–92.
[45] Gottgens E, Ostheimer C, Span PN, Bussink J, Hammond EM. HPV, hypoxia and radiation response in head and neck cancer. Br J Radiol 2018;91:20180047.
[46] Aerts HJWL. The potential of radiomic-based phenotyping in precision medicine a review. JAMA Oncol 2016;2:1636–42.
[47] Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. Theranostics 2019;9:1303–22.
[48] Leger S, Zwanenburg A, Pilz K, Zschaeck S, Zöphel K, Kotzerke J, et al. CT imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer. Radiother Oncol 2019;130:10–7.