Research Paper

Radiomics-based prediction of survival in patients with head and neck squamous cell carcinoma based on pre- and post-treatment ^{18}F-PET/CT

Zheran Liu[1,*], Yuan Cao[2,*], Wei Diao[2], Yue Cheng[3], Zhiyun Jia[2], Xingchen Peng[1]

[1]Department of Biotherapy, Cancer Center, West China Hospital of Sichuan University, Chengdu 610041, China
[2]Department of Nuclear Medicine, West China Hospital of Sichuan University, Chengdu 610041, China
[3]Department of Radiology, West China Hospital of Sichuan University, Chengdu 610041, China
*Equal contribution

Correspondence to: Xingchen Peng, Zhiyun Jia; **email:** pxx2014@scu.edu.cn, zhiyunjia@hotmail.com
Keywords: ^{18}F-PET/CT, head and neck squamous cell carcinoma, radiomics, prognosis, machine learning
Received: March 11, 2020 **Accepted:** June 4, 2020 **Published:** July 16, 2020

ABSTRACT

Background: 18-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-PET/CT) has been widely applied for the imaging of head and neck squamous cell carcinoma (HNSCC). This study examined whether pre- and post-treatment ^{18}F-PET/CT features can help predict the survival of HNSCC patients.
Results: Three radiomics features were identified as prognostic factors. Radiomics score calculated from these features significantly predicted overall survival (OS) and disease-free disease (DFS). The clinicopathological characteristics combined with pre- or post-treatment nomograms showed better ROC curves and decision curves than the nomogram based only on clinicopathological characteristics.
Conclusions: Combining clinicopathological characteristics with radiomics features of pre-treatment PET/CT or post-treatment PET/CT assessment of primary tumor sites as positive or negative may substantially improve prediction of OS and DFS of HNSCC patients.
Methods: 171 patients who received pre-treatment ^{18}F-PET/CT scans and 154 patients who received post-treatment ^{18}F-PET/CT scans with HNSCC in the Cancer Imaging Achieve (TCIA) were included. Nomograms that combined clinicopathological features with either pre-treatment PET/CT radiomics features or post-treatment assessment of primary tumor sites were constructed using data from 154 HNSCC patients. Receiver operating characteristic (ROC) curves and decision curves were used to compare the predictions of these models with those of a model incorporating only clinicopathological features.

INTRODUCTION

Head and neck cancer, which manifests most often as head and neck squamous cell carcinoma (HNSCC), is the sixth most common malignancy, with an incidence of 650,000 cases and 330,000 deaths annually worldwide [1, 2]. HNSCC refers to a broad range of malignant tumors, including in the oral cavity, larynx, oropharynx, and hypopharynx [3]. The 5-year survival rate of patients with HNSCC is only about 60% and is lowest for those with tumors in the hypopharynx [4]. HNSCC is usually treated by surgical resection with or without adjuvant radiotherapy or by definitive radiotherapy with or without concurrent chemotherapy [5].

The stage of HNSCC, which is vital for guiding treatment decisions, is usually determined based on imaging of the head and neck with computed tomography (CT) or magnetic resonance imaging (MRI) [6]. Increasingly, 18-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography/CT, which provides both

anatomical and metabolic information, is used to distinguish benign from malignant disease, assess treatment response and detect recurrence [7, 8]. PET/CT can detect HNSCC with a sensitivity of 72-96% and specificity of 83-100% [8–10]. While this imaging modality continues to gain ground as a tool for diagnosing disease and assessing treatment response, whether it can predict patient prognosis is unclear.

It may be possible to predict prognosis based on quantifiable features in PET/CT scans taken before or after treatment [11–13]. For example, studies have linked the survival of patients with HNSCC, lymphoma, or non-small-cell lung cancer to mean and maximum standardized ^{18}F-FDG uptake values (SUV*mean*, SUV*max*), metabolic tumor value (MTV) and total lesion glycolysis (TLG). These PET/CT features reflect tumor metabolic activity and lesion size [14–18]. Using radiomics to predict the prognosis of cancer patients is in its infancy, so the consensus still lacks on what image features provide the most reliable predictions.

In the present study, we used a quantitative radiomics approach to extract imaging features from pre-treatment ^{18}F-FDG PET/CT scans of patients with HNSCC and a conventional approach to extract positive/negative findings from post-treatment scans. Then we combined each of these types of data with clinicopathological characteristics to generate models to predict survival. The predictive performance of these models was compared to that of a model-based only on clinico-pathological characteristics.

RESULTS

Patient characteristics and radiomic signatures

A total of 171 patients (training cohort = 115 and a validation cohort = 56) were analyzed for the construction of a Radiomics score (Rad-score) model based on pre-treatment PET/CT, and 154 patients were analyzed for the development of nomograms based on pre- or post-treatment PET/CT. The clinical characteristics of patients in the training and validation cohorts were summarized in Table 1. The correlations between extracted radiomics features were calculated and visualized by a correlation matrix (Figure 1). LASSO Cox regression was used to choose potential prognostic predictors from the 56 radiomics features in the training cohort (Figure 2). Three radiomics features were identified, and both the univariate and multivariate analyses of the selected features were performed to show the correlation of these features with patients' survival (Supplementary Table 1). Besides, the collinearity statistics demonstrated that the collinearity between selected features was acceptable (variance inflation factor: SHAPE_Sphericity: 1.102, NGLDM_Coarseness: 1.274, SMTV: 1.375). Then, these features were used to calculate Rad-score for each patient:

$$Rad-score = -0.3392 * SHAPE\_Sphericity + 0.3736 * NGLDM\_Coarseness + 1.5655 * SMTV$$

The optimal cut-off value of the Radscore was 0.01187901, and patients in the training and validation cohorts were accordingly classified as low- or high-risk. Supplementary Table 2 shows clinicopathological characteristics between patients with low and high risk.

In the pre-treatment Rad-score model, the Kaplan-Meier analysis showed that high risk was associated with significantly worse overall survival (OS) in the training cohort (HR 5.89, 95%CI 1.74-20.02, p = 0.004), validation cohort (HR 5.59, 95%CI 1.83-17.09, p = 0.003) and both cohorts together (HR 6.33, 95%CI 2.77-14.5, $p < 0.001$). Similar results were obtained for disease-free survival (DFS) in the training cohort (HR 7.04, 95%CI 1.93-25.68, p = 0.003), validation cohort (HR 5.10, 95%CI 1.61-16.17, p = 0.006) and both cohorts together (HR 6.844, 95%CI 2.90-16.13, $p <$ 0.001) (Figure 3). In the post-treatment negative/positive model, Kaplan-Meier analysis showed that a positive finding was significantly related to worse OS (HR 6.609, 95%CI 3.649-11.97, $p < 0.001$) and DFS (HR 8.169, 95%CI 4.453-14.99, $p < 0.001$) (Figure 4). Cox regression showed that both the pre-treatment Rad-score and post-treatment outcomes were significant independent predictors of both OS and DFS (Supplementary Table 3). Besides, we compared the concordance index (C-index, which is proportional to the survival-prediction ability of variables) between Rad-score and four conventional PET features (TLG, MTV, SUV*mean*, and SUV*max*). The results showed that the survival-prediction ability of the Rad-score was much better than not only each single conventional PET feature but also the combined of four (Supplementary Table 4).

Prediction of OS and DFS using models based on radiomic signatures

As a first step in constructing predictive models based on radiomic signatures, we created a conventional prediction model based only on clinical characteristics of 154 HNSCC patients according to inclusion and exclusion criteria. This conventional clinical model also served as a benchmark for assessing the prognostic performance of the radiomic models. The clinical model was constructed by initially including eight clinical characteristics (body mass index, age, T stage, N stage, AJCC stage, cancer site, histology grade, and smoking history), from which age and histology grade were subsequently excluded because they

Table 1. Demographics and clinicopathologic characteristics of patients with HSNCC.

Variables	Training cohort (N = 115)		Validation cohort (N = 56)	
	N	%	N	%
Gender				
Male	100	86.96	47	83.93
Female	15	13.04	9	16.07
Age (years)				
< 60	71	61.74	30	53.57
≥ 60	44	38.26	26	46.43
Tumor size				
≤ 4	94	81.74	27	48.21
> 4	57	49.57	29	51.79
Tumor Location				
Oropharynx	92	80.00	47	83.93
Larynx	13	11.30	6	10.71
Oral cavity	2	1.74	2	3.57
Hypopharynx	8	6.96	3	5.36
Differentiation status				
Well	13	11.30	4	7.14
Moderate	51	44.35	28	50.00
Poor and undifferentiat	51	44.35	24	42.86
T stage *				
T1	21	18.26	10	17.86
T2	37	32.17	17	30.36
T3	34	29.57	18	32.14
T4	23	20.00	11	19.64
N stage *				
N0	10	8.70	9	16.07
N1	11	9.57	8	14.29
N2a	5	4.35	4	7.14
N2b	56	48.70	19	33.93
N2c	28	24.35	9	16.07
N3	5	4.35	7	12.50
TNM stage *				
I	1	0.87	0	0.00
II	2	1.74	3	5.36
III	15	13.04	13	23.21
IVA	90	78.26	32	57.14
IVB	7	6.09	8	14.29

***** according to 7th AJCC stage system.
Abbreviations: HSNCC-head and neck squamous cell carcinoma; N-number.

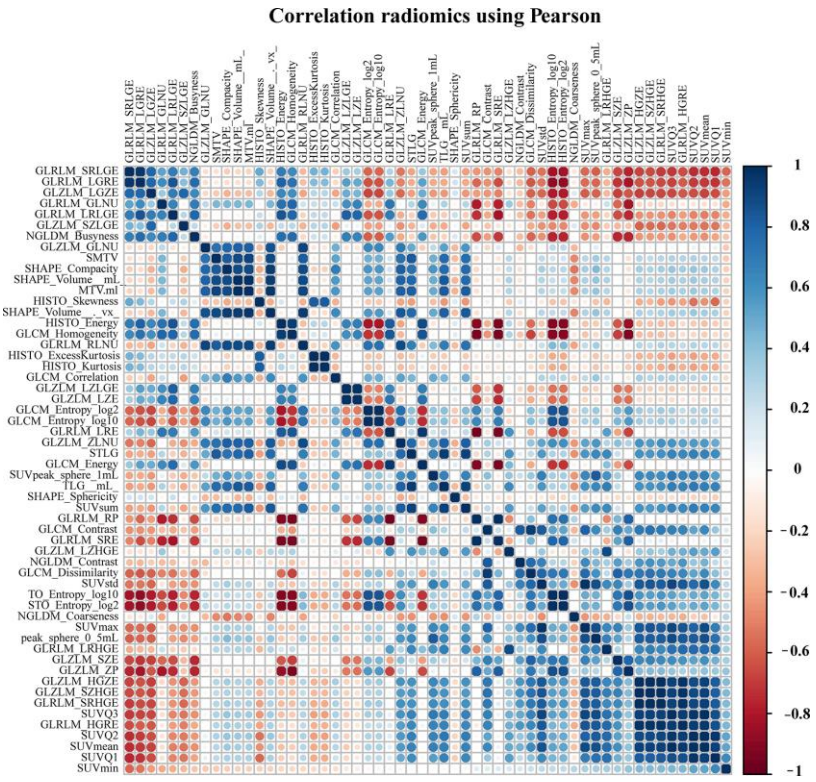


Figure 1. The correlation matrix between pre-processing radiomics features.

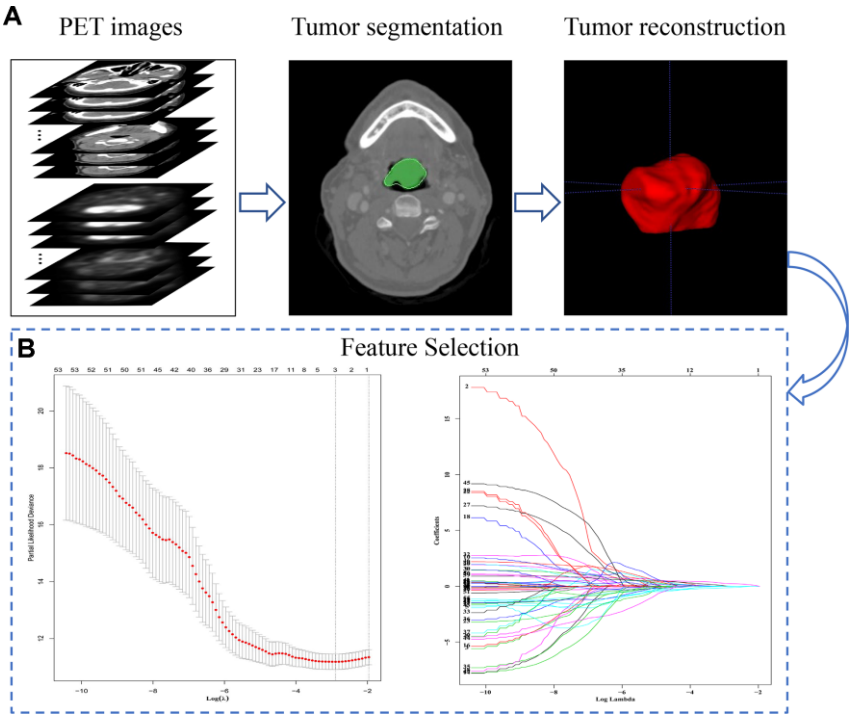


Figure 2. The extraction process and general characteristics of pre-treatment PET signatures with HNSCC patients. (**A**) The segmentation and reconstruction process of PET/CT images. (**B**) Demonstration of the varies of Lasso coefficient in different log (λ) sequence. A 15-fold cross validation were used to select the most optimal penalty parameter λ via minimum criteria. The minimum λ (λ = 0.05209914) were chose according to the criteria. Abbreviations: OS: overall survival. DFS: disease free survival.

did not satisfy the model's assumption of proportional hazards. The final clinical model contained body mass index, tumor location, and N stage, while other characteristics were excluded using a stepwise algorithm. This model was used to generate the corresponding reference OS and DFS nomograms (Supplementary Figure 1).

Radiomics signatures from the pre-treatment PET/CT scans were added to this conventional clinical model, and the corresponding model was used to generate OS and DFS nomograms (Table 2 and Figure 5). The C index indicated good discrimination of OS (C index 0.77, 95%CI 0.70-0.84) and DFS (C index 0.77, 95%CI 0.70-0.83). Calibration curves calculated for 3, 5, or 7 years showed good agreement with the OS and DFS nomograms.

Good results were also obtained when positive/negative findings based on post-treatment PET/CT were added to the conventional clinical model (Table 2 and Figure 6). The corresponding nomograms showed excellent accuracy and discrimination for OS (C index 0.822, 95%CI 0.767-0.877) and DFS (C index 0.832, 95%CI 0.781-0.883).

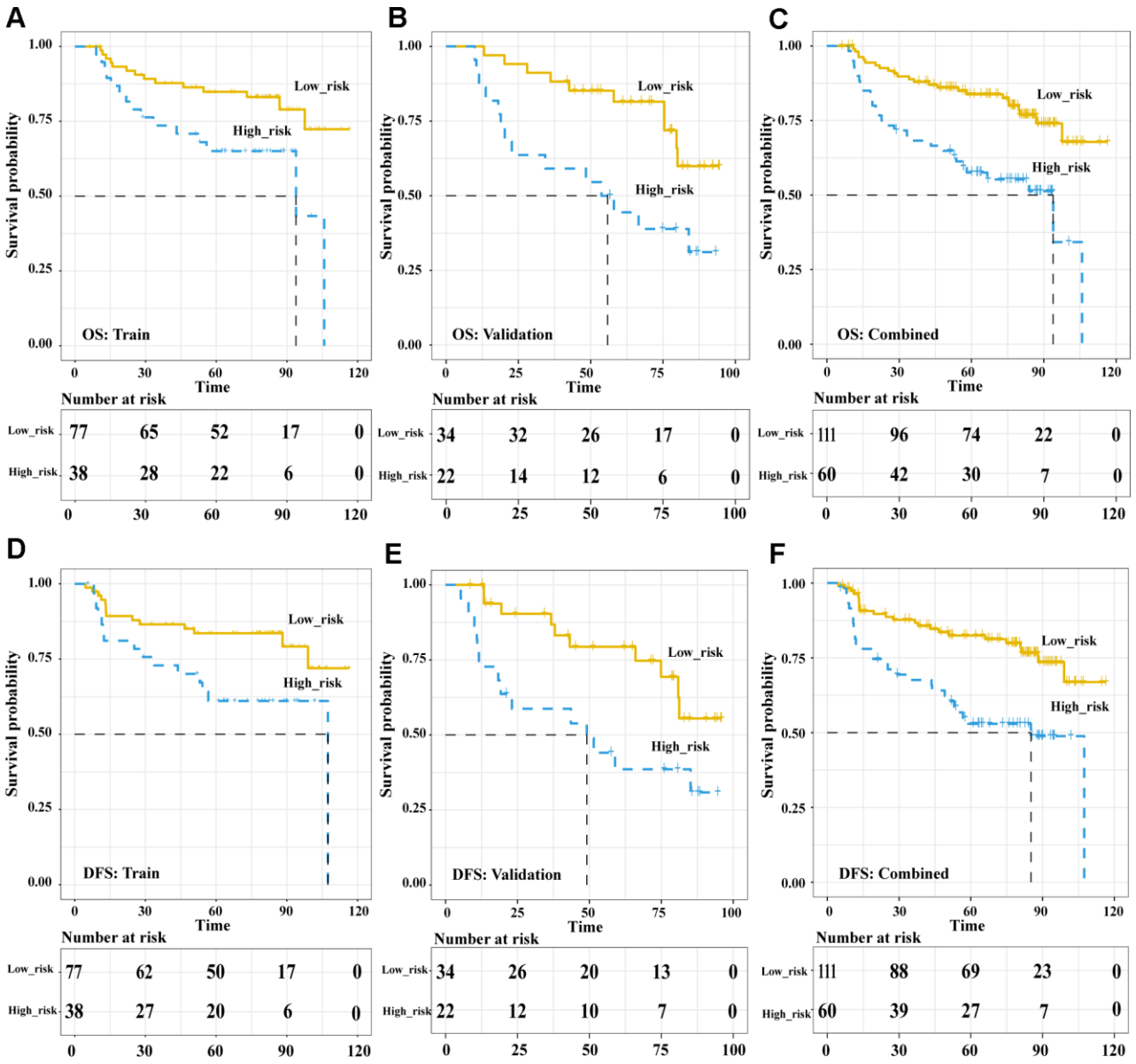


Figure 3. The Pre-treatment PET signatures could significantly stratify patients' OS and DFS. Kaplan-Meier survival analysis of pre-treatment Rad-score-defined risk levels in the training, validation cohorts and combined cohort. OS: the training cohort (**A**), validation cohort (**B**), and combined cohort (**C**). DFS: the training cohort (**D**), validation cohort (**E**), and combined cohort (**F**). Abbreviations: OS: overall survival. DFS: disease free survival.

Comparison of models

Comparison of ROC curves at 3, 5, and 7 years showed that the pre-treatment model predicted OS and DFS better than the conventional clinical model. In contrast, the post-treatment model performed significantly better than the pre-treatment model. Similarly, decision curves showed that the post-treatment model maximized clinical benefits for patients in the prediction of OS and DFS at 3, 5, and 7 years (Figures 7 and 8).

DISCUSSION

^{18}F-FDG-PET/CT radiomics signatures, which can capture spatial heterogeneity in tumors, have been applied as potential prognostic markers in many cancers, including gastric cancer [19], nasopharyngeal carcinoma [20], NSCLC [21], and HNSCC. HNSCC is a clinically heterogeneous disease, and few biomarkers are available for predicting tumor response to treatment or prognosis [22]. The present study used machine learning to identify 56 radiomics features in PET/CT scans of patients with HNSCC, and these features were significantly associated with OS and DFS. Combining some of these features with patients' clinicopathological characteristics allowed reliable and accurate predictions of OS and DFS, which were substantially better than those obtained based on clinicopathological characteristics alone. The models described here may help improve the design of treatment strategies in HNSCC and thereby lead to better patient prognosis.

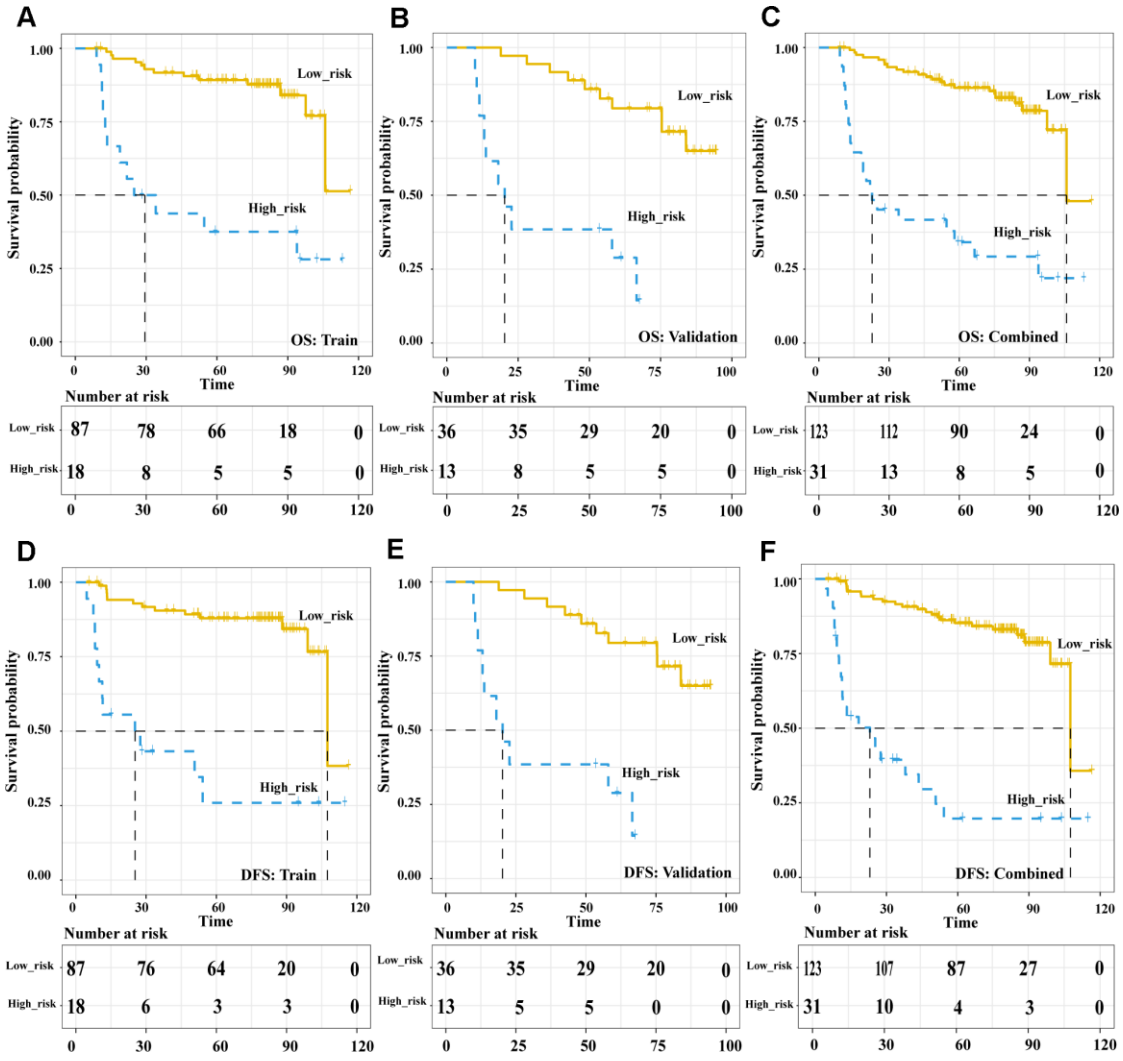


Figure 4. The Post-treatment PET outcomes is a powerful tool to stratify patients' OS and DFS. Kaplan-Meier survival analysis of post-treatment PET-outcome-defined risk levels in the training, validation cohorts and combined cohort. OS: the training cohort (**A**), validation cohort (**B**), and combined cohort (**C**). DFS: the training cohort (**D**), validation cohort (**E**), and combined cohort (**F**). Abbreviations: OS: overall survival. DFS: disease free survival.

Table 2. Multivariate Cox regression analyses for OS and DFS in the pre-treatment radiomics model and post-treatment PET model.

Variables	Overall survival		Disease-free survival	
	HR (95%CI)	*p*	HR (95%CI)	*p*
Pre-treatment Radiomics Model				
Rad score	3.28 (1.23-8.70)	0.017	3.43 (1.24-9.45)	0.017
N stage (vs. N0-2)	3.47 (1.40-8.61)	0.007	3.21 (1.29-7.95)	0.012
Cancer site (vs. Oropharynx)				
Hypopharynx	4.70 (1.96-11/28)	<0.001	3.90 (1.60-9.48)	0.002
Oral cavity	2.27 (0.30-17.00)	0.425	1.93 (0.26-14.41)	0.521
Larynx	1.25 (0.30-5.14)	0.756	1.19 (0.28-5.12)	0.812
Start-treatment BMI	0.93 (0.87-0.99)	0.025	0.924 (0.86-0.99)	0.019
Post-treatment PET Model				
PET outcome (vs. negative)	6.79 (3.69-12.47)	<0.001	8.26 (4.41-15.44)	<0.001
N stage (vs. N0-2)	5.87 (2.67-14.57)	<0.001	5.43 (2.20-13.37)	<0.001
Cancer site (vs. Oropharynx)				
Hypopharynx	6.40 (2.60-15.82)	<0.001	5.05 (2.02-12.64)	<0.001
Oral cavity	2.15 (0.28-16.30)	0.461	1.48 (0.20-11.21)	0.700
Larynx	2.17 (0.63-7.55)	0.221	2.33 (0.68-7.99)	0.179
Start-treatment BMI	0.91 (0.85-0.98)	<0.001	0.91 (0.85-0.98)	<0.001

Abbreviations: OS-overall survival; DFS-disease-free survival; HR-hazard ratio; BMI-body mass index; PET- positron emission tomography/computed tomography.

Accurately predicting prognosis is of great importance for optimizing treatment strategies in HNSCC, but it remains controversial. Several studies have attempted to assess the predictive value of radiomics information from CT and MRI images in HNSCC. Koun et al. [23] recruited 62 patients with HNSCC to evaluate the ability of pretreatment CT texture to predict treatment failure in patients with primary HNSCC treated with chemoradiotherapy. They found that three histogram features and four grey-level run length (GLRL) features predicted treatment failure in these patients. Yuan et al. [24] extracted 485 MRI-based radiomic features from 170 patients with HNSCC (85 in the training cohort, 85 in the validation cohort) and obtained higher C indices for the radiomics signature (0.73 for training and 0.71 for validation) and the nomogram (0.76 for training and 0.72 for validation) than the AJCC staging system (0.63 for training and 0.61 for validation). Their study established the feasibility of combining MRI-based radiomic signatures with clinical characteristics to predict prognosis in patients with HNSCC.

^{18}F-FDG-PET/CT has also been widely applied to predict survival in cancer patients because of its ability to provide information on tumor burden and aggressiveness. Bogowicz et al. [22] compared PET and CT radiomics for prediction of local tumor control in HNSCC, and they found PET to be more accurate than CT in predicting

tumor local control rate. Those authors highlighted the need to pay more attention to PET-based radiomic analysis for predicting prognosis. Kim et al. [25] examined the ability of PET/CT to predict treatment failure and guide clinical decision-making about salvage surgery. Despite their relatively small sample, they were able to predict OS and PFS reasonably well based on post-treatment PET findings. The optimal time to perform PET/CT on HNSCC patients and the optimal prognostic model for predicting survival remain unclear. Our study identified a pre-treatment Rad-score, comprising SHAPE_Sphericity, NGLDM_Coarseness, and standardized MTV (SMTV). This integrated PET/CT signature, when combined with clinicopathological

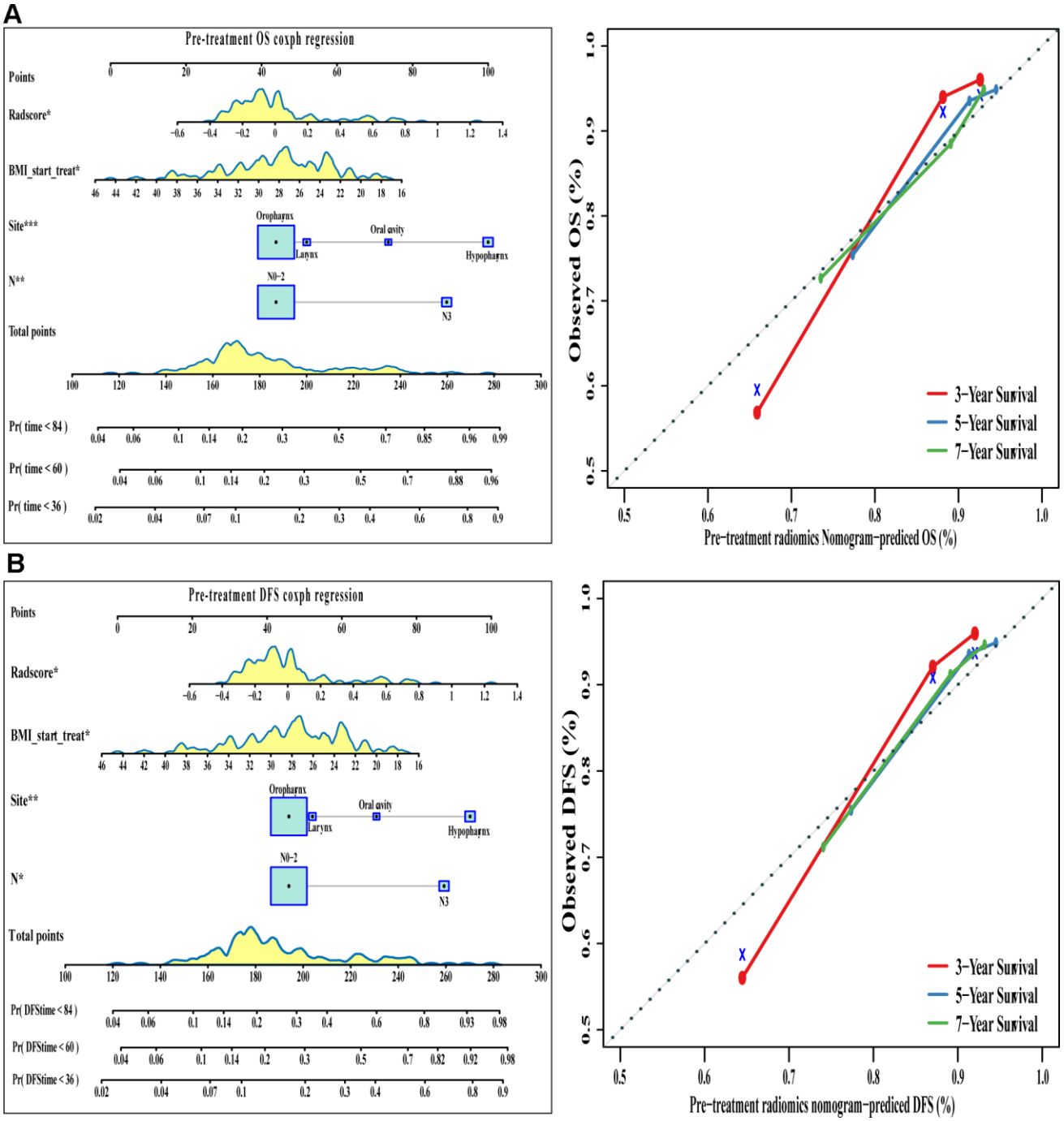


Figure 5. The visualization of OS and DFS survival models based on pre-treatment Rad-score combined with clinicopathologic characteristics. The constructed nomograms and their calibration plots to estimate the OS (**A**) and DFS (**B**) in 3, 5, and 7 years. Abbreviations: OS: overall survival. DFS: disease free survival.

characteristics, shows promise for predicting OS and DFS of HNSCC patients. Previous studies have demonstrated the prognostic significance of traditional PET quantitative parameters such as SUV [26], MTV, and TLG [27]. We found, however, that these parameters did not predict OS or DFS as well as the combination of our PET/CT radiomic signatures with a subset of clinico-pathological characteristics. These findings highlighted the potential role of PET/CT radiomic signatures that could play in the high throughput machine learning era. At the same time, our study suggested that post-treatment positive/negative findings may have even more prognostic potential than pre-treatment Rad-score when combined with clinicopathological characteristics.

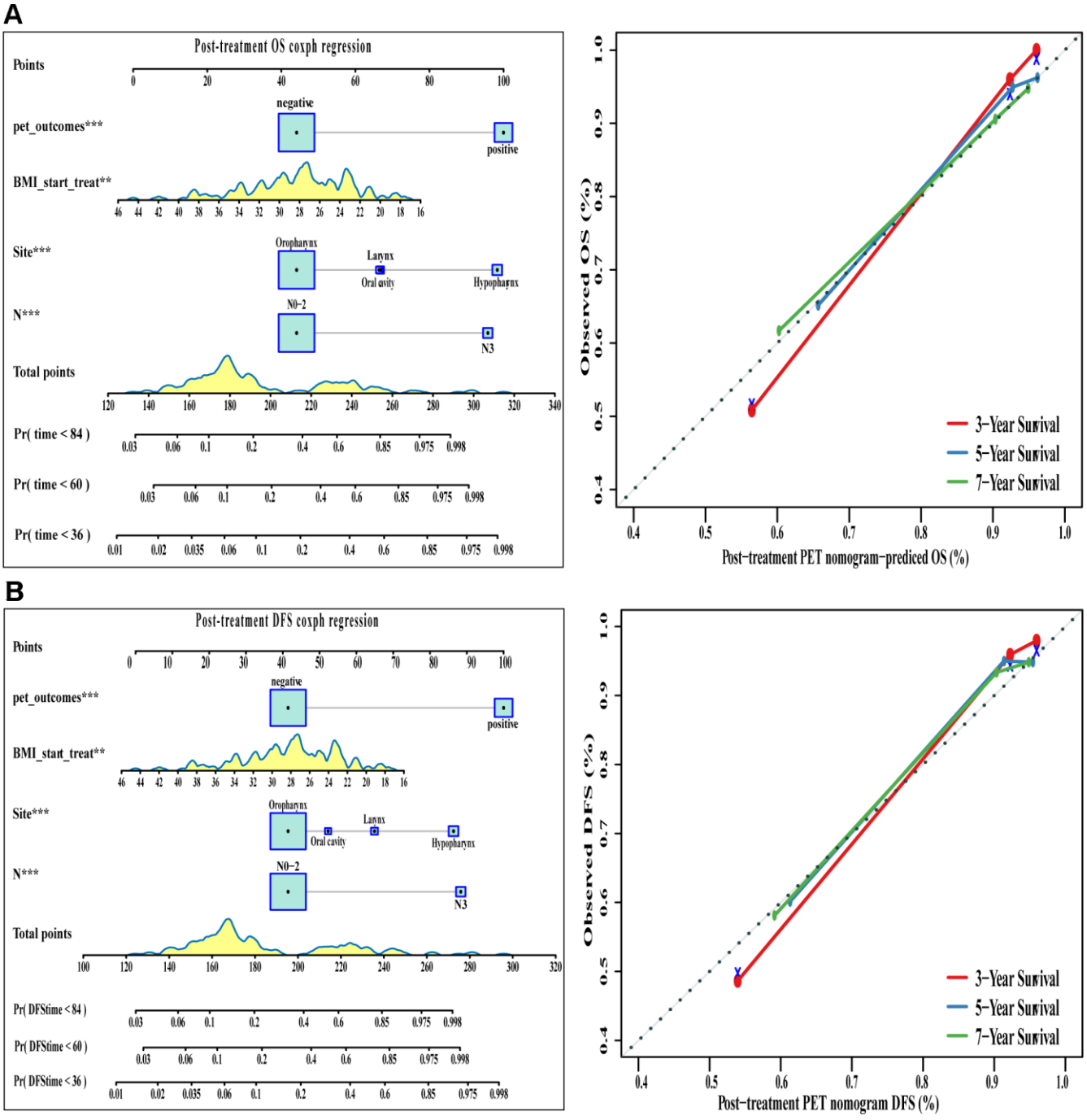


Figure 6. The visualization of OS and DFS survival models based on post-treatment PET signatures combined with clinicopathologic characteristics. The constructed nomograms and their calibration plots to estimate the OS (**A**) and DFS (**B**) in 3, 5, and 7 years. Abbreviations: OS: overall survival. DFS: disease free survival.

Our findings suggested the potential of PET/CT radiomic signatures to predict the prognosis of patients with HNSCC reliably. These promising results may partly reflect our efforts to control for heterogeneity in the patient population, which came from a single center with the same scanner. While this approach allows us to reduce potential confounding due to heterogeneity of patient characteristics and hospital practices, it also threatens the external validity of our results. Therefore, our findings should be verified and extended in larger, preferably multi-site patient populations.

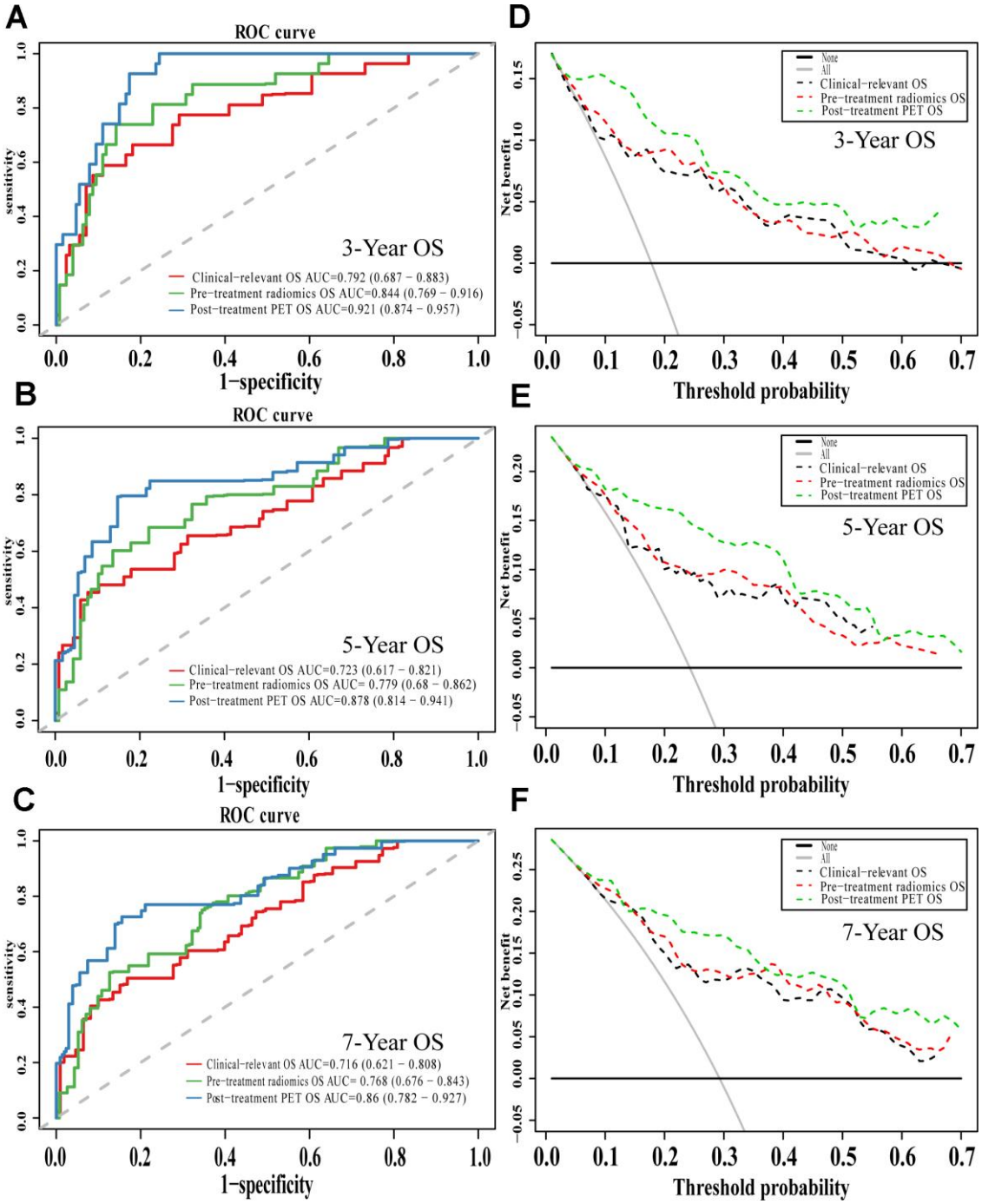


Figure 7. The evaluation of built OS nomograms. The ROC curves and DCA curves of the comparison between clinical-relevant, pre-treatment and post-treatment survival OS models in 3, 5, and 7 years. The ROC curves of 3-year survival (**A**), 5-year survival (**B**) and 7-year survival (**C**). The DCA curves of 3-year survival (**D**), 5-year survival (**E**) and 7-year survival (**F**). Abbreviations: OS: overall survival. ROC: receiver operator curve. DCA: decision curve analysis.

CONCLUSIONS

The present study using publicly available ^{18}F-PET/CT images suggests that combining clinicopathological characteristics with specific radiomic signatures from pre-treatment scans or with post-treatment assessment of primary tumor sites as positive or negative can predict OS and DFS of patients with HNSCC significantly better than clinicopathological characteristics alone.

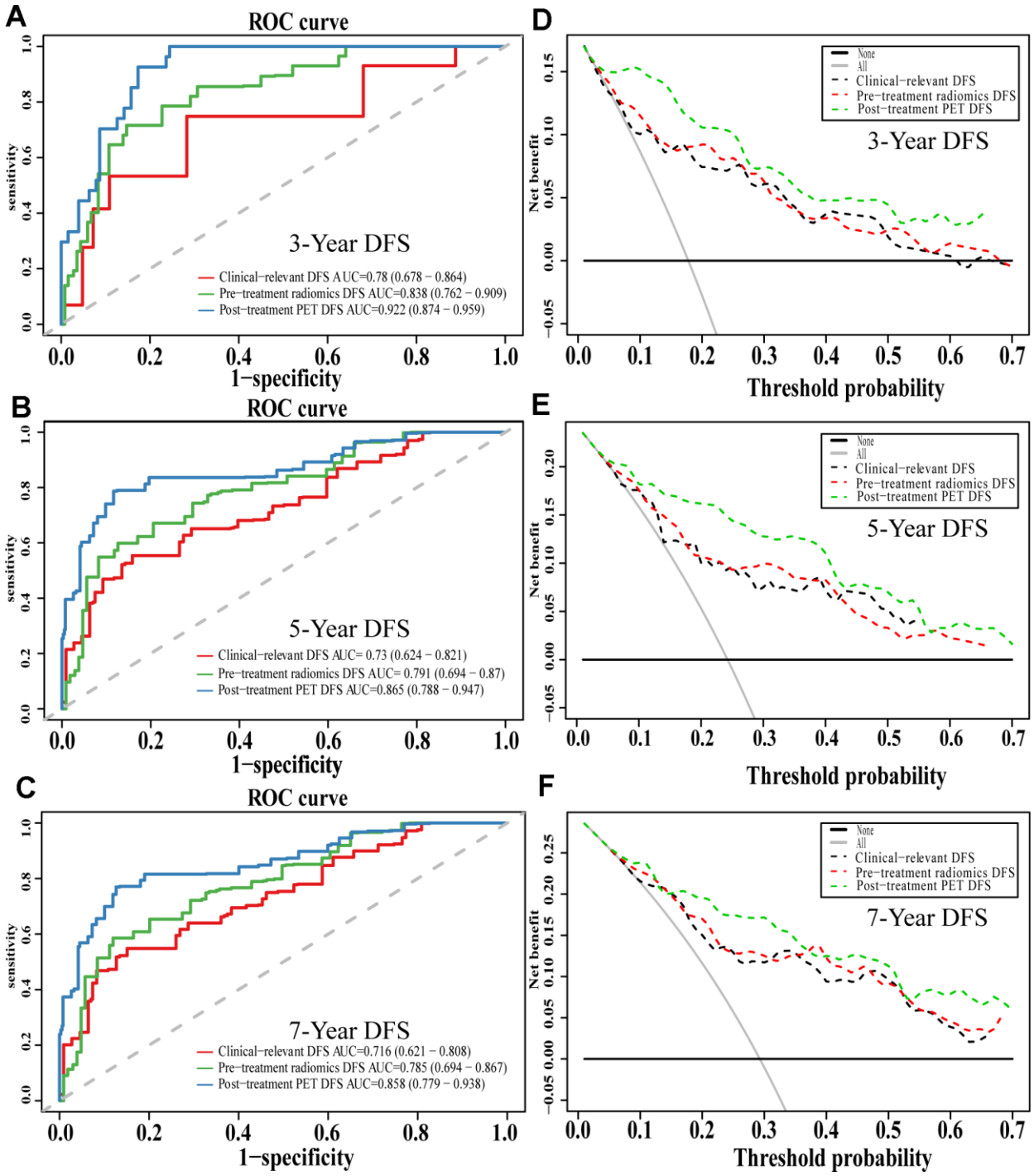


Figure 8. The evaluation of built DFS nomograms. The ROC curves and DCA curves of the comparison between clinical-relevant, pre-treatment and post-treatment survival DFS models in 3, 5, and 7 years. The ROC curves of 3-year survival (**A**), 5-year survival (**B**) and 7-year survival (**C**). The DCA curves of 3-year survival (**D**), 5-year survival (**E**) and 7-year survival (**F**). Abbreviations: DFS: disease free survival. ROC: receiver operator curve. DCA: decision curve analysis.

MATERIALS AND METHODS

Patient population

We extracted ^{18}F-FDG-PET/CT scans from the publicly available HNSCC dataset on The Cancer Imaging Achieve (TCIA) platform of the University of Texas MD Anderson Cancer Center [28] (http://www.cancerimagingarchive.net/). Of the total set of 2,840 consecutive patients with HNSCC treated with curative radiotherapy at the MD Anderson Cancer Center between 1 October 2003 and 31 August 2013 [29]. Two hundred fifteen patients overlapping in TCGA and TCIA databases were initially selected. Of these, 203 patients were included because they did not have a primary diagnosis of nasopharyngeal carcinoma, cancer of unknown primary site, or recurrent HNSCC. For the identification of pre-radiomics signatures, patients were excluded from the analysis if their pre-treatment PET/CT images were unavailable or the region of interest on their scans was too small to extract features. The rest of the patients were randomly divided into a training cohort and a validation cohort using the *caret* package in R 3.6.1 [30]. Finally, 171 patients with available pre-treatment PET scans and 154 patients with available post-treatment PET/CT scans were included in our study, according to the Data Descriptor [28]. For further identification of post-radiomics signatures and model construction, patients from the original cohort were included except for those who lacked the paired pre- and post-treatment PET/CT images (Supplementary Figure 2).

Pre-treatment PET/CT image analysis and feature extraction

The pre-treatment PET/CT images were segmented, and the features were extracted using LIFEx 4.0 (http://www.lifexsoft.org) [31]. The primary tumor without lymph nodes was segmented by two specialists in nuclear medicine (Y.C. and W.D.), who delineated a computer-generated volume of interest around voxels equal to or greater than 40% of SUV_{max} [32]. Noise in images was reduced by resampling FDG uptake values using 64 discrete values, boundary SUV values of 0 to 30, and a bin width of 0.47, based on typical SUVs for HNSCC tumors [33]. Data were extracted on 56 quantitative PET parameters, first-order intensity features, shape features, and texture indices (Supplementary Table 5 and Supplementary Material). Finally, texture features were investigated based on gray-level co-occurrence matrices, gray-level run-length matrices, neighborhood gray-tone difference matrix wavelet decompositions, and gray-level size zone matrices.

Post-treatment PET/CT image interpretation

The post-treatment PET/CT scans were reviewed independently by two specialists in nuclear medicine (Y.C. and W.D.), who determined whether the residual or recurrent disease was presented. Scans were judged negative if no focal increase in FDG uptake was evident, or if an increase in FDG uptake was apparent but could be attributed to physiological causes or the treatment [5]. Discrepancies between the independent assessments were resolved in consultation with a senior specialist in nuclear medicine (Z.Y.J.) and a radiation oncologist (X.C.P.). Pearson correlation analysis was performed to show the correlations between extracted radiomics features.

Feature selection and integration into a single Rad-score

Post-normalized Fifty-six radiomics features were entered into a "least absolute shrinkage and selection operator" (LASSO) algorithm [34] in a Cox regression model based on penalized maximum likelihood, to shrink the regression coefficients of most radiomics variables to zero. The λ is a penalty parameter that varies in each step of model fitting. Bootstrapping was used to cross-validate 1000 times to the built model and to select the variables most relevant to overall survival (OS) in the training cohort at an optimal λ. The minimum λ giving a minimum mean cross-validated error of the built model was determined, and the coefficients of the selected variables were identified at this λmin. Then a Rad-score for each patient was computed based on all LASSO-selected features using the following formula:

$$Rad-score = \sum_{i=1}^{n} Coefficient\ of\ \ feature\,(i) \\ *value\ of\ \ feature\,(i) \qquad (1)$$

where the coefficient of radiomics feature *(i)* was the coefficient determined in the regression model.

Data in the training set were used to generate a time-dependent receiver operating characteristic (ROC) curve by survivalROC Package in R to describe the ability of the Rad-score to predict OS, which was defined as the period from the first diagnosis to death. This curve was used to identify the optimal cut-off for the Rad-score, and patients whose Rad-scores were higher than this threshold were classified as "high risk," while those with Rad-scores equal to or lower than the threshold were classified as "low risk".

Model construction and evaluation

The following three models were used to predict OS and disease-free survival (DFS), defined as the period

from the first diagnosis to death due to HNSCC: a conventional clinical model, a pre-treatment PET/CT model, and a post-treatment PET/CT model. The *conventional clinical model* contained several pre-treatment clinical characteristics that have been linked to the survival of HNSCC patients [35]: body mass index, age, T stage, N stage, stage according to the 7th edition of the American Joint Committee on Cancer (AJCC) guidelines, tumor location, histology grade, and smoking history. The model was optimized in a stepwise manner based on the Akaike information criterion, after which time-dependent variables were excluded by applying an assumption of proportional hazards. The *pre-treatment model* was generated by adding the Rad-score to this conventional clinical model. The *post-treatment model* was created by adding positive/negative findings (based on post-treatment PET/CT scans) to the conventional clinical model. Three nomograms were constructed based on the three models.

The various models were assessed for their ability to predict OS or DFS at 3, 5- or 7-years using calibration curves and Harrell's concordance index (C index). The "bootstrap split" method [36] was applied with 1000 iterations. Models were also assessed and compared using ROC curves, and the overfitting risk was evaluated using the Akaike information criterion. A decision curve analysis (DCA) was conducted to help to determine which model is the best in clinical use by comparing benefits and the harms of false-positive and false-negative prediction on the same scale [37, 38].

Statistical analysis

Data were analyzed statistically using R 3.6.1 [30] and a significance threshold of $p = 0.05$. LASSO-based Cox regression was conducted using the *glmnet* package, while ROC curves and optimal cut-offs were generated using the *survivalROC* and *tdROC* packages [39, 40]. The Pearson's correlation analysis was conducted and visuliazed by *rattle* package. OS and DFS were calculated, and survival curves were plotted using Kaplan-Meier analysis; statistical inference about the survival difference between high- and low-risk patients was accomplished using the Cox regression statistic, and the analyses were performed using the *survival* package [41, 42]. Multivariate Cox models were constructed and evaluated using the *survival* and *pec* packages [43–45], while decision curves were analyzed using the *DCA* package. When appropriate, results were reported as hazard ratios (HRs) with associated 95% confidence intervals (CIs). Collinearity diagnostics were run using SPSS software, version 25.0 (IBM Corporation, Armonk, NY, USA) to ensure partial regression coefficients derived from regression analyses were estimated precisely and that the relative importance of each predictor for OS and DFS could be assessed reliably.

Abbreviations

AJCC: American Joint Committee on Cancer; BMI: body mass index; CT: computed tomography; C-index: the concordance index; DCA: Decision Curve Analysis; DFS: disease-free survival; 18F-FDG PET/CT: fluorine-18-fluorodeoxyglucose tomography/computed tomography; HNSCC: head and neck squamous cell carcinoma; HR: Hazard ratio; LASSO: The least absolute shrinkage and selection operator; MTV: metabolic tumor value; MRI: magnetic resonance imaging; OS: overall survival; ROC: receiver operating characteristic; SUV: standardized uptake values; TLG: total lesion glycolysis.

AUTHOR CONTRIBUTIONS

In the present study, X.P. and Z.J. were responsible for the study design and participated in evaluation of results. Y. Chen, D.W., Z.L., and Y.Cao participated in collection of study materials or patients. D.W., Z.L., and Y.Cao participated in collection and assembly of data. Z.L. and Y.Cao did the data analysis and interpretation. Z.L. and Y.Cao drafted the manuscript. X.P., Z.J, Z.L.and Y.Cao proofread the manuscript for important intellectual content. All authors contributed to manuscript preparation. All authors reviewed the report and approved the final version.

ACKNOWLEDGMENTS

The authors thank Prof. Peng Huang who works at Evidence Medical Center of Nanchang University for his help in statistics.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant Nos. 81971595 and 81771812 to Jia, 81672386 and 81402494 to Peng), Technology Innovation Project of Chengdu Science and Technology Bureau (No.2019-YF05-00459-SN) the Sichuan Science and Technology Program (2018SZ0391 to Jia) and the Innovation Spark Project of Sichuan University (No. 2019SCUH0003 to Jia).

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68:394–424. https://doi.org/10.3322/caac.21492 PMID:30207593

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA Cancer J Clin. 2017; 67:7–30. https://doi.org/10.3322/caac.21387 PMID:28055103

3. Nocon CC, Ajmani GS, Bhayani MK. A contemporary analysis of racial disparities in recommended and received treatment for head and neck cancer. Cancer. 2020; 126:381–89. https://doi.org/10.1002/cncr.32342 PMID:31580491

4. LAG Ries DM, M Krapcho, A Mariotto, BA Miller. SEER Cancer Statistics Review, 1975–2004. Bethesda, MD: National Cancer Institute. 2006.

5. Li Y, Awan MJ, Chang T, Lavertu P, Zender C, Rezaee R, Fowler N, Wasman J, Avril NE, Chen N, Machtay M, Yao M. Post-radiotherapy PET/CT for predicting treatment outcomes in head and neck cancer after postoperative radiotherapy. Eur J Nucl Med Mol Imaging. 2019; 46:794–800. https://doi.org/10.1007/s00259-019-4272-8 PMID:30680587

6. Jabour BA, Choi Y, Hoh CK, Rege SD, Soong JC, Lufkin RB, Hanafee WN, Maddahi J, Chaiken L, Bailet J. Extracranial head and neck: PET imaging with 2-[F-18]fluoro-2-deoxy-D-glucose and MR imaging correlation. Radiology. 1993; 186:27–35. https://doi.org/10.1148/radiology.186.1.8416578 PMID:8416578

7. Ng SH, Yen TC, Chang JT, Chan SC, Ko SF, Wang HM, Lee LY, Kang CJ, Wong AM, Liao CT. Prospective study of [18F]fluorodeoxyglucose positron emission tomography and computed tomography and magnetic resonance imaging in oral cavity squamous cell carcinoma with palpably negative neck. J Clin Oncol. 2006; 24:4371–76. https://doi.org/10.1200/JCO.2006.05.7349 PMID:16983105

8. Laubenbacher C, Saumweber D, Wagner-Manslau C, Kau RJ, Herz M, Avril N, Ziegler S, Kruschke C, Arnold W, Schwaiger M. Comparison of fluorine-18-fluorodeoxyglucose PET, MRI and endoscopy for staging head and neck squamous-cell carcinomas. J Nucl Med. 1995; 36:1747–57. PMID:7562038

9. Kau RJ, Alexiou C, Laubenbacher C, Werner M, Schwaiger M, Arnold W. Lymph node detection of head and neck squamous cell carcinomas by positron emission tomography with fluorodeoxyglucose F 18 in a routine clinical setting. Arch Otolaryngol Head Neck Surg. 1999; 125:1322–28. https://doi.org/10.1001/archotol.125.12.1322 PMID:10604409

10. Ng SH, Yen TC, Liao CT, Chang JT, Chan SC, Ko SF, Wang HM, Wong HF. 18F-FDG PET and CT/MRI in oral cavity squamous cell carcinoma: a prospective study of 124 patients with histologic correlation. J Nucl Med. 2005; 46:1136–43. PMID:16000282

11. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012; 48:441–46. https://doi.org/10.1016/j.ejca.2011.11.036 PMID:22257792

12. Wong AJ, Kanwar A, Mohamed AS, Fuller CD. Radiomics in head and neck cancer: from exploration to application. Transl Cancer Res. 2016; 5:371–82. https://doi.org/10.21037/tcr.2016.07.18 PMID:30627523

13. El Naqa I, Grigsby P, Apte A, Kidd E, Donnelly E, Khullar D, Chaudhari S, Yang D, Schmitt M, Laforest R, Thorstad W, Deasy JO. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit. 2009; 42:1162–71. https://doi.org/10.1016/j.patcog.2008.08.011 PMID:20161266

14. Kim TM, Paeng JC, Chun IK, Keam B, Jeon YK, Lee SH, Kim DW, Lee DS, Kim CW, Chung JK, Kim IH, Heo DS. Total lesion glycolysis in positron emission tomography is a better predictor of outcome than the international prognostic index for patients with diffuse large B cell lymphoma. Cancer. 2013; 119:1195–202. https://doi.org/10.1002/cncr.27855 PMID:23212736

15. Procházka V, Klugar M, Bachanova V, Klugarová J, Tučková D, Papajík T. Comparing the accuracy of quantitative versus qualitative analyses of interim PET to prognosticate hodgkin lymphoma: a systematic review protocol of diagnostic test accuracy. BMJ Open. 2016; 6:e011729. https://doi.org/10.1136/bmjopen-2016-011729 PMID:27496236

16. Albano D, Bosio G, Pagani C, Re A, Tucci A, Giubbini R, Bertagna F. Prognostic role of baseline 18F-FDG PET/CT metabolic parameters in burkitt lymphoma. Eur J Nucl Med Mol Imaging. 2019; 46:87–96.

https://doi.org/10.1007/s00259-018-4173-2
PMID:30276438

17. Hohenstein NA, Chan JW, Wu SY, Tahir P, Yom SS. Diagnosis, staging, radiation treatment response assessment, and outcome prognostication of head and neck cancers using PET imaging: a systematic review. PET Clin. 2020; 15:65–75. https://doi.org/10.1016/j.cpet.2019.08.010 PMID:31735303

18. Kostakoglu L, Chauvie S. Metabolic tumor volume metrics in lymphoma. Semin Nucl Med. 2018; 48:50–66. https://doi.org/10.1053/j.semnuclmed.2017.09.005 PMID:29195618

19. Jiang Y, Yuan Q, Lv W, Xi S, Huang W, Sun Z, Chen H, Zhao L, Liu W, Hu Y, Lu L, Ma J, Li T, et al. Radiomic signature of ^{18}F fluorodeoxyglucose PET/CT for prediction of gastric cancer survival and chemotherapeutic benefits. Theranostics. 2018; 8:5915–28. https://doi.org/10.7150/thno.28018 PMID:30613271

20. Peng H, Dong D, Fang MJ, Li L, Tang LL, Chen L, Li WF, Mao YP, Fan W, Liu LZ, Tian L, Lin AH, Sun Y, et al. Prognostic value of deep learning PET/CT-based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. Clin Cancer Res. 2019; 25:4271–79. https://doi.org/10.1158/1078-0432.CCR-18-3065 PMID:30975664

21. Li X, Yin G, Zhang Y, Dai D, Liu J, Chen P, Zhu L, Ma W, Xu W. Predictive power of a radiomic signature based on ^{18}F-FDG PET/CT images for EGFR mutational status in NSCLC. Front Oncol. 2019; 9:1062. https://doi.org/10.3389/fonc.2019.01062 PMID:31681597

22. Bogowicz M, Riesterer O, Stark LS, Studer G, Unkelbach J, Guckenberger M, Tanadini-Lang S. Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. Acta Oncol. 2017; 56:1531–36. https://doi.org/10.1080/0284186X.2017.1346382 PMID:28820287

23. Kuno XH, Qureshi XMM, Chapman XMN, Li XB and Andreu-Arasa XVC. (2016). CT Texture Analysis Potentially Predicts Local Failure in Head and Neck Squamous Cell Carcinoma Treated with Chemoradiotherapy. 54th Annual Meeting of American Society of Neuroradiology (ASNR).

24. Yuan Y, Ren J, Shi Y, Tao X. MRI-based radiomic signature as predictive marker for patients with head and neck squamous cell carcinoma. Eur J Radiol. 2019; 117:193–98. https://doi.org/10.1016/j.ejrad.2019.06.019 PMID:31307647

25. Kim R, Ock CY, Keam B, Kim TM, Kim JH, Paeng JC, Kwon SK, Hah JH, Kwon TK, Kim DW, Wu HG, Sung MW, Heo DS. Predictive and prognostic value of PET/CT imaging post-chemoradiotherapy and clinical decision-making consequences in locally advanced head & neck squamous cell carcinoma: a retrospective study. BMC Cancer. 2016; 16:116. https://doi.org/10.1186/s12885-016-2147-y PMID:26884055

26. Manca G, Vanzi E, Rubello D, Giammarile F, Grassetto G, Wong KK, Perkins AC, Colletti PM, Volterrani D. (18)F-FDG PET/CT quantification in head and neck squamous cell cancer: principles, technical issues and clinical applications. Eur J Nucl Med Mol Imaging. 2016; 43:1360–75. https://doi.org/10.1007/s00259-015-3294-0 PMID:26780912

27. Wong KH, Panek R, Welsh L, Mcquaid D, Dunlop A, Riddell A, Murray I, Du Y, Chua S, Koh DM, Bhide S, Nutting C, Oyen WJ, et al. The predictive value of early assessment after 1 cycle of induction chemotherapy with 18F-FDG PET/CT and diffusion-weighted MRI for response to radical chemoradiotherapy in head and neck squamous cell carcinoma. J Nucl Med. 2016; 57:1843–50. https://doi.org/10.2967/jnumed.116.174433 PMID:27417648

28. Grossberg A MA, Elhalawani H, Bennett W, Smith K, Nolan T, Chamchod S, Kantor M, Browne T, Hutcheson K, Gunn G, Garden A, Frank S, Rosenthal D, Freymann J, Fuller C. Data from Head and Neck Cancer CT Atlas. The Cancer Imaging Archive. 2017.

29. Grossberg AJ, Mohamed AS, Elhalawani H, Bennett WC, Smith KE, Nolan TS, Williams B, Chamchod S, Heukelom J, Kantor ME, Browne T, Hutcheson KA, Gunn GB, et al. Imaging and clinical data archive for head and neck squamous cell carcinoma patients treated with radiotherapy. Sci Data. 2018; 5:180173. https://doi.org/10.1038/sdata.2018.173 PMID:30179230

30. Team CR. R: A Language and Environment for Statistical Computing.

31. Orlhac F, Soussan M, Maisonobe JA, Garcia CA, Vanderlinden B, Buvat I. Tumor texture analysis in 18F-FDG PET: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. Journal of nuclear medicine : official publication, Society of Nuclear Medicine. 2014; 55:414–422. https://doi.org/10.2967/jnumed.113.129858 PMID:24549286

32. Dibble EH, Alvarez AC, Truong MT, Mercier G, Cook EF, Subramaniam RM. 18F-FDG metabolic tumor volume and total glycolytic activity of oral cavity and oropharyngeal squamous cell cancer: adding value to clinical staging. J Nucl Med. 2012; 53:709–15. https://doi.org/10.2967/jnumed.111.099531 PMID:22492732

33. Guezennec C, Robin P, Orlhac F, Bourhis D, Delcroix O, Gobel Y, Rousset J, Schick U, Salaün PY, Abgral R. Prognostic value of textural indices extracted from pretherapeutic 18-F FDG-PET/CT in head and neck squamous cell carcinoma. Head Neck. 2019; 41:495–502. https://doi.org/10.1002/hed.25433 PMID:30549149

34. SL Kukreja JL, MJ Brenne. A least absolute shrinkage and selection operator(LASSO) for nonlinear system identification. IFAC proceedings volumes. 2006; 39:814–819. https://doi.org/10.3182/20060329-3-AU-2901.00128

35. Denaro N, Russi EG, Merlano MC. Pros and cons of the new edition of TNM classification of head and neck squamous cell carcinoma. Oncology. 2018; 95:202–10. https://doi.org/10.1159/000490415 PMID:29975952

36. Xu Y, Goodacre R. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. J Anal Test. 2018; 2:249–62. https://doi.org/10.1007/s41664-018-0068-2 PMID:30842888

37. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006; 26:565–74. https://doi.org/10.1177/0272989X06295361 PMID:17099194

38. Zhang Z, Rousson V, Lee WC, Ferdynus C, Chen M, Qian X, Guo Y, and written on behalf of AME Big-Data Clinical Trial Collaborative Group. Decision curve analysis: a technical note. Ann Transl Med. 2018; 6:308. https://doi.org/10.21037/atm.2018.07.02 PMID:30211196

39. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000; 56:337–44. https://doi.org/10.1111/j.0006-341x.2000.00337.x PMID:10877287

40. Li L, Greene T, Hu B. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. Stat Methods Med Res. 2018; 27:2264–78. https://doi.org/10.1177/0962280216680239 PMID:27895266

41. Andersen P, Gill R. Cox's regression model for counting processes, a large sample study. Annals of Statistics. 1982; 10:1100–1120. https://doi.org/10.1214/aos/1176345976

42. Therneau T, Grambsch P. Modeling Survival Data: Extending the Cox Model. Springer-Verlag. 2000. https://doi.org/10.1007/978-1-4757-3294-8

43. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. Stat Med. 1999; 18:2529–45. https://doi.org/10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5 PMID:10474158

44. Gerds TA, Schumacher M. Efron-type measures of prediction error for survival analysis. Biometrics. 2007; 63:1283–87. https://doi.org/10.1111/j.1541-0420.2007.00832.x PMID:17651459

45. Schumacher M, Binder H, Gerds T. Assessment of survival prediction models based on microarray data. Bioinformatics. 2007; 23:1768–74. https://doi.org/10.1093/bioinformatics/btm232 PMID:17485430

SUPPLEMENTARY MATERIALS

GLCM

The grey level co-occurrence matrix (GLCM) [Haralick] takes into account the ar-rangements of pairs of voxels to calculate textural indices. The GLCM is calculated

from 13 different directions in 3D with a δ-voxel distance ($\|\underset{d}{\rightarrow}\|$) relationship between neighboured voxels. The index value is the average of the index over the 13 directions in space (X, Y, Z). Seven textural indices are computed from this matrix. An entry (i, j) of GLCM for one direction is equal to:

$$GLCM_{\Delta x,\Delta y}(i,j) = \frac{1}{Pairs\,ROI} \sum_{p=1}^{N-\Delta x}\sum_{q=1}^{M-\Delta y} \begin{cases} 1 if\ (I(p,q)=i,\ I(p+\Delta x, q+\Delta y)=j) \\ and\ I(p,q), I(p+\Delta x, q+\Delta y) \in ROI \\ 0 otherwise \end{cases}$$

where I (p, q) corresponds to voxel (p, q) in an image (I) of size N*M. The vector $\underset{d}{\rightarrow}$ = (Δx, Δy) covers the 4 directions (D1, D2, D3, D4, Figure 1.1) in 2D space or 13 directions (D1, D2, ..., D13, Figure 1.2) in 3D space and Pairs ROI corresponds to the number of all voxel pairs belonging to the region of interest (ROI).

GLCM_Homogeneity

is the homogeneity of grey-level voxel pairs.

GLCM_Homogeneity = Average over 13 (or 4) directions $\left(\sum_i\sum_j \frac{GLCM(i,j)}{1+|i-j|}\right)$

GLCM_Energy

also called Uniformity or Second Angular Moment, is the uniformity of grey-level voxel pairs.

GLCM_Energy = Average over 13 (or 4) directions $\left(\sum_i\sum_j GLCM(i,j)^2\right)$

GLCM_Contrast

also called Variance or Inertia, is the local variations in the GLCM.

GLCM_Contrast = Average over 13 (or 4) directions $\left(\sum_i\sum_j (i-j)^2 GLCM(i,j)\right)$

GLCM_Correlation

is the linear dependency of grey-levels in GLCM.

GLCM_Correlation = Average over 13 (or 4) directions $\left(\sum_i\sum_j \frac{(i-\mu i)\cdot(j-\mu j)\cdot GLCM(i,j)}{\sigma i\cdot\sigma j}\right)$

where μi or μj corresponds to the average on row i or column j and σi and σj correspond to the variance on row i or column j.

GLCM_Entropy_log10

is the randomness of grey-level voxel pairs.

GLCM_Entropylog10 = Average over 13 (or 4) directions $\left(-\sum_i\sum_j GLCM(i,j)\cdot\log 2(GLCM(i,j))+\in\right)$

where ε = 2e-16.

Be aware of the logarithm used in the formula.

GLCM_Entropy_log2

is the randomness of grey-level voxel pairs.

GLCM_Entropylog2 = Average over 13 (or 4) direction $\left(-\sum_i\sum_j GLCM(i,j)\cdot\log 10(GLCM(i,j))+\in\right)$

where ε = 2e-16.

GLCM_Dissimilarity

is the variation of grey-level voxel pairs.

GLCM_Dissimilarity = Average over 13 (or 4) directions $\left(\sum_i\sum_j |i-j|\cdot GLCM(i,j)\right)$

NGLDM

The neighborhood grey-level different matrix (NGLDM) [Amadasum1989] corresponds to the

difference of grey-levels between one voxel and its 26 neighbours in 3 dimensions (8 in 2D). Three texture indices can be computed from this matrix. An element (i, 1) of NGLDM corresponds to the probability of occurrence of level i and an element (i, 2) is equal to:

$$\text{NGLDM}(i,2) = \sum_p \sum_q \begin{cases} \left|\bar{M}(p,q) - i\right| \; if \; I(p,q) = i \\ 0 \; else \end{cases}$$

where $\bar{M}(p, q)$ is the average of intensities over the 26 neighbor voxels of voxel (p, q).

NGLDM_Coarseness

is the level of spatial rate of change in intensity.

$$\text{NGLDM\_Coarseness} = \frac{1}{\sum_i NGLDM(i,1) \cdot NGLDM(i,2)}$$

NGLDM_Contrast

is the intensity difference between neighbouring regions.

$$\text{NGLDM\_Contrast} = \left[\sum_i \sum_j NGLDM(i,1) \cdot NGLDM(j,1) \cdot (i,j)^2 \right] \cdot \frac{\sum_i NGLDM(i,2)}{E \cdot G \cdot (G-1)}$$

where E corresponds to the number of voxels in the Volume of Interest and G the number of grey-levels.

NGLDM_Busyness

is the spatial frequency of changes in intensity.

$$\text{NGLDM\_Busyness} = \frac{\sum_i NGLDM(i,1) \cdot NGLDM(i,2)}{\sum_i \sum_j \left| (i \cdot NGLDM(i,1) - j \cdot NGLDM(j,1) \cdot (i,j)^2) \right|}$$

with NGLDM (i, 1) ≠ 0, NGLDM(j, 1) ≠ 0

GLRLM

The grey-level run length matrix (GLRLM) [Xu] gives the size of homogeneous runs for each grey level. This matrix is computed for the 13 different directions in 3D (4 in 2D) and for each of the 11 texture indices derived from this matrix, the 3D value is the average over the 13 directions in 3D (4 in 2D). The element (i, j) of GLRLM corresponds to the number of homogeneous runs of j voxels with intensity i in an image and is called GLRLM(i, j) thereafter.

GLRLM_SRE, GLRLM_LRE

Short-Run Emphasis or Long-Run Emphasis is the distribution of the short or the long homogeneous runs in an image.

GLRLM_SRE = Average over 13 (or 4) directions

$$\left(\frac{1}{H} \sum_i \sum_j \frac{GLRLM(i,j)}{i^2} \right)$$

GLRLM_LRE = Average over 13 (or 4) directions

$$\left(\frac{1}{H} \sum_i \sum_j GLRLM(i,j) \cdot i^2 \right)$$

GLRLM_SRLGE, GLRLM_SRHGE

Short-Run Low Gray-level Emphasis or Short-Run High Gray-level Emphasis is the distribution of the short homogeneous runs with low or high grey-levels.

GLRLM_SRLGE = Average over 13 (or 4) directions

$$\left(\frac{1}{H} \sum_i \sum_j \frac{GLRLM(i,j)}{i^2 \cdot j^2} \right)$$

GLRLM_SRHGE = Average over 13 (or 4) directions

$$\left(\frac{1}{H} \sum_i \sum_j \frac{GLRLM(i,j) \cdot i^2}{j^2} \right)$$

GLRLM_LRLGE, GLRLM_LRHGE

Long-Run Low Gray-level Emphasis or Long-Run High Gray-level Emphasis is the distribution of the long homogeneous runs with low or high grey-levels.

GLRLM_LRLGE = Average over 13 (or 4) directions

$$\left(\frac{1}{H} \sum_i \sum_j \frac{GLRLM(i,j) \cdot j^2}{i^2} \right)$$

GLRLM_LRHGE = Average over 13 (or 4) directions

$$\left(\frac{1}{H} \sum_i \sum_j GLRLM(i,j) \cdot i^2 \cdot j^2 \right)$$

GLRLM_GLNUr, GLRLM_RLNU

Gray-Level Non-Uniformity for run or Run Length Non-Uniformity is the nonuniformity of the grey-levels or the length of the homogeneous runs.

GLRLM_GLNUr = Average over 13 (or 4) directions

$$\left(\frac{1}{H}\sum_i\left(\sum_j GLRLM(i,j)\right)^2\right)$$

GLRLM_RLNU = Average over 13 (or 4) directions

$$\left(\frac{1}{H}\sum_j\left(\sum_i GLRLM(i,j)\right)^2\right)$$

GLRLM_RP

Run Percentage measures the homogeneity of the homogeneous runs.

GLRLM_RP = Average over 13 (or 4) directions

$$\left(\frac{H}{\sum_i\sum_j\left(j\cdot GLRLM(i,j)\right)}\right)$$

GLZLM

The grey-level zone length matrix (GLZLM) Thibault] provides information on the size of homogeneous zones for each grey-level in 3 dimensions (or 2D). It is also named Grey Level Size Zone Matrix (GLSZM). From this matrix, 11 texture indices are computed. Element (i, j) of GLZLM corresponds to the number of homogeneous zones of j voxels with the intensity i in an image and is called GLZLM(i, j) thereafter.

GLZLM_SZE, GLZLM_LZE

Short-Zone Emphasis or Long-Zone Emphasis is the distribution of the short or the long homogeneous zones in an image.

$$\text{GLZLM\_SZE} = \frac{1}{H}\sum_i\sum_j\frac{GLZLM(i,j)}{j^2}$$

$$\text{GLZLM\_LZE} = \frac{1}{H}\sum_i\sum_j GLZLM(i,j)\cdot j^2$$

where H corresponds to the number of homogeneous zones in the Volume of Interest.

GLZLM_LGZE, GLZLM_HGZE

Low Gray-level Zone Emphasis or High Gray-level Zone Emphasis is the distribution of the low or high grey-level zones.

$$\text{GLZLM\_LGZE} = \frac{1}{H}\sum_i\sum_j\frac{GLZLM(i,j)}{i^2}$$

$$\text{GLZLM\_HGZE} = \frac{1}{H}\sum_i\sum_j GLZLM(i,j)\cdot i^2$$

GLZLM_SZLGE, GLZLM_SZHGE

Short-Zone Low Gray-level Emphasis or Short-Zone High Gray-level Emphasis is the distribution of the short homogeneous zones with low or high grey-levels.

$$\text{GLZLM\_SZLGE} = \frac{1}{H}\sum_i\sum_j\frac{GLZLM(i,j)}{i^2\cdot j^2}$$

$$\text{GLZLM\_SZHGE} = \frac{1}{H}\sum_i\sum_j\frac{GLZLM(i,j)\cdot i^2}{j^2}$$

GLZLM_LZLGE, GLZLM_LZHGE

Long-Zone Low Gray-level Emphasis or Long-Zone High Gray-level Emphasis is the distribution of the long homogeneous zones with low or high grey-levels.

$$\text{GLZLM\_LZLGE} = \frac{1}{H}\sum_i\sum_j\frac{GLZLM(i,j)\cdot j^2}{i^2}$$

$$\text{GLZLM\_LZHGE} = \frac{1}{H}\sum_i\sum_j GLZLM(i,j)\cdot i^2\cdot j^2$$

GLZLM_GLNUz, GLZLM_ZLNU

Gray-Level Non-Uniformity for zone or Zone Length Non-Uniformity is the nonuniformity of the grey-levels or the length of the homogeneous zones.

$$\text{GLZLM\_GLNUz} = \frac{1}{H}\sum_i\left(\sum_j GLZLM(i,j)\right)^2$$

$$\text{GLZLM\_ZLNU} = \frac{1}{H}\sum_j\left(\sum_i GLZLM(i,j)\right)^2$$

GLZLM_ZP

Zone Percentage measures the homogeneity of the homogeneous zones.

$$\text{GLZLM\_ZP} = \frac{H}{\sum_i\sum_j\left(j\cdot GLZLM(i,j)\right)}$$

SHAPE_SPHERICITY

is how spherical a Volume of Interest is. Sphericity is equal to 1 for a perfect sphere.

$$\text{SHAPE\_Sphericity} = \frac{\pi^{1/3}\cdot(6V)^{2/3}}{A}$$

where V and A correspond to the volume and the surface of the Volume of Interest based on the Delaunay triangulation.

SHAPE_Compacity

reflects how compact the Volume of Interest is.

$$\text{SHAPE\_Compacity} = \frac{A^{3/2}}{V}$$

where V and A correspond to the volume and the surface of the Volume of Interest based on the Delaunay triangulation.

SHAPE_Volume (mL and voxels)

is the Volume of Interest in mL and in voxels.

$$\text{SHAPE\_Volume} = \sum_i V_i$$

where Vi corresponds to the volume of voxel i of the Volume of Interest.

HISTOGRAM CALCULATION

To build a histogram HISTO, it is necessary to determine a bin width ("bin" parameter). The indices derived from the histogram will depend on this bin width parameter.

HISTO_Skewness

is the asymmetry of the grey-level distribution in the histogram.

$$\text{HISTO\_Skewness} = \frac{\frac{1}{E}\sum_i \left(HISTO(i) - \overline{HISTO} \right)^3}{\left(\sqrt{\frac{1}{E}\sum_i \left(HISTO(i) - \overline{HISTO} \right)^2} \right)^3}$$

where HISTO(i) corresponds to the number of voxels with intensity i, E the total number of voxels in the Volume of Interest and $\overline{HISTO}$ the average of grey-levels in the histogram.

HISTO_Kurtosis

reflects the shape of the grey-level distribution (peaked or flat) relative to a normal distribution.

$$\text{HISTO\_Kurtosis} = \frac{\frac{1}{E}\sum_i \left(HISTO(i) - \overline{HISTO} \right)^4}{\left(\frac{1}{E}\sum_i \left(HISTO(i) - \overline{HISTO} \right)^2 \right)^2}$$

where HISTO(i) corresponds to the number of voxels with intensity i, E the total number of voxels in the Volume of Interest and $\overline{HISTO}$ the average of grey-levels in the histogram

HISTO_Entropy_log10

reflects the randomness of the distribution.

$$\text{HISTO\_Entropylog10} = -\sum_i p(i) \cdot log_{10}\left(p(i) + \varepsilon \right)$$

where p(i) is the probability of occurrence of voxels with intensity i and ε = 2e-16.

HISTO_Entropy_log2

reflects the randomness of the distribution.

$$\text{HISTO\_Entropylog2} = -\sum_i p(i) \cdot log_2\left(p(i) + \varepsilon \right)$$

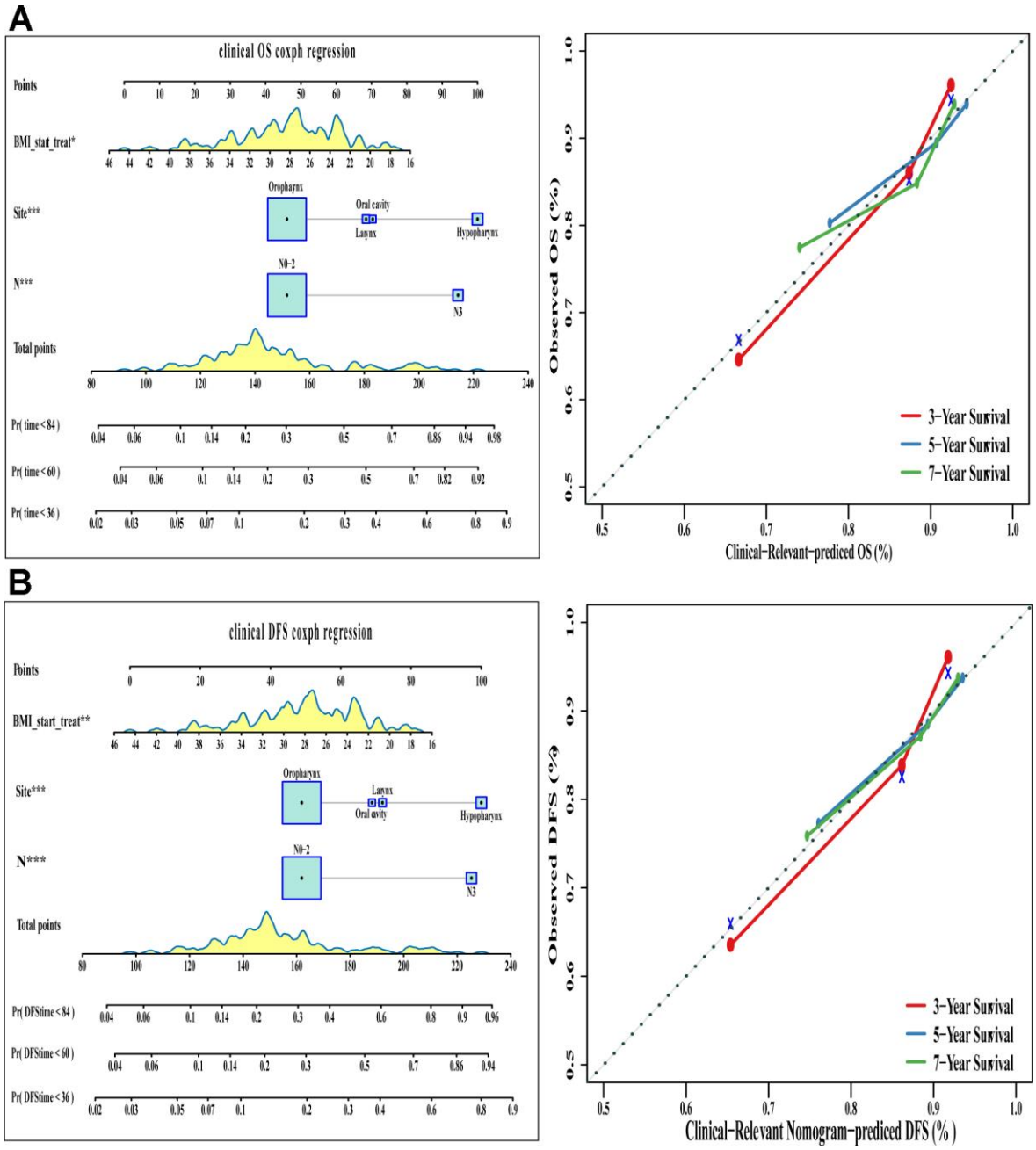
where p(i) is the probability of occurrence of voxels with intensity i and ε = 2e-16.

HISTO_Energy

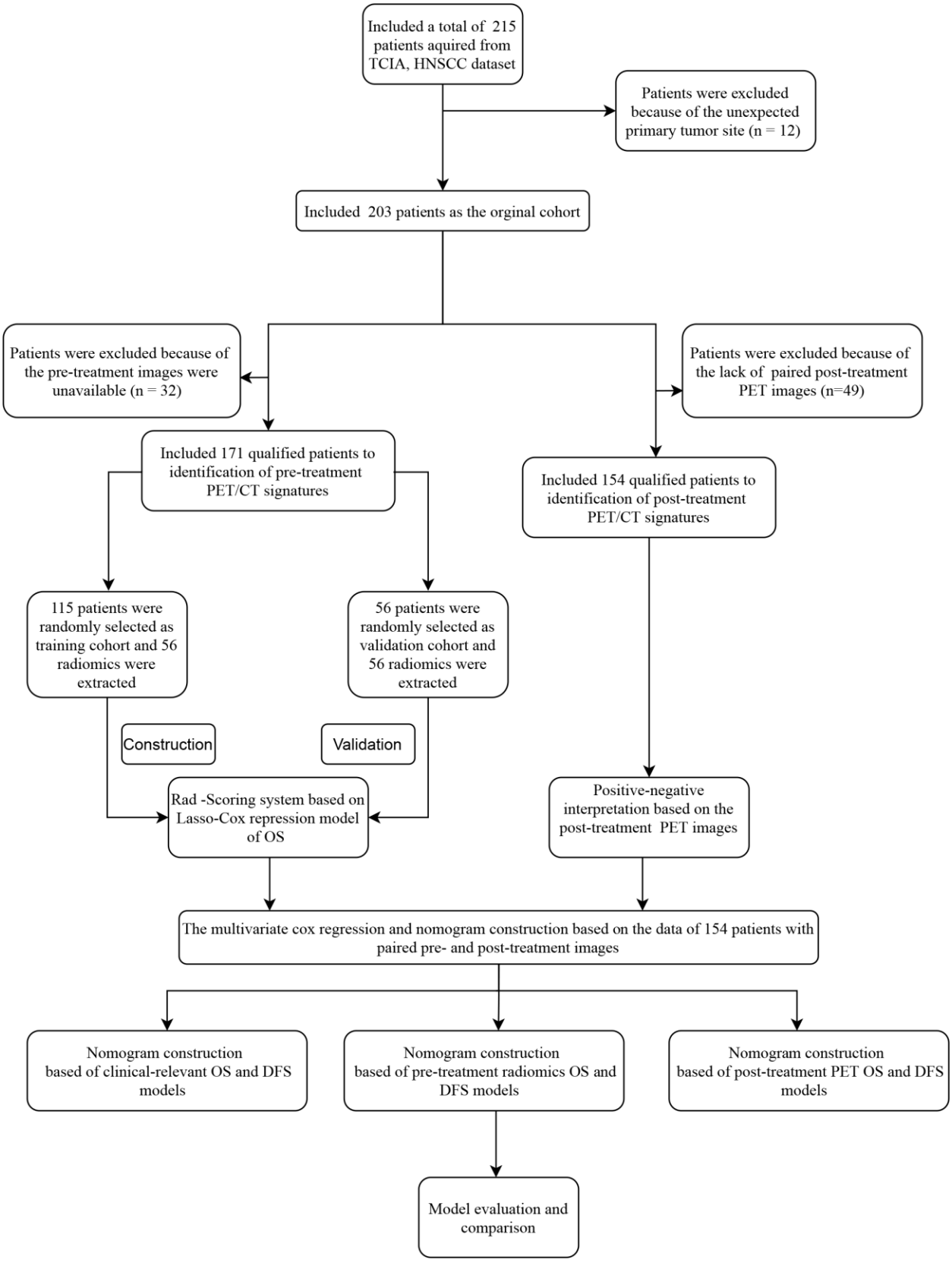
reflects the uniformity of the distribution.

$$\text{HISTO\_Energy} = \sum_i p(i)^2$$

Supplementary Figures



Supplementary Figure 1. The clinical-relevant survival nomograms and their calibration plots. (**A**) OS-based model. (**B**) DFS-based model.



Supplementary Figure 2. The flow chart of the study design.

Supplementary Tables

Supplementary Table 1. Univariate and multivariate Cox regression analyses for OS and DFS in the radiomics features.

Variables	Overall survival		Disease-free survival	
	HR (95%CI)	*p*	HR (95%CI)	*p*
Univariate analysis				
SHAPE_Sphericity	0.69 (0.48-1.01)	0.056	0.70 (0.49-1.02)	0.064
NGLDM_Coarseness	1.24 (0.96-1.61)	0.100	1.22 (0.95-1.58)	0.119
SMTV	1.42 (0.99-2.03)	0.060	1.48 (1.01 - 2.16)	0.046
Multivariate analysis				
SHAPE_Sphericity	0.71 (0.49-1.04)	0.081	0.72 (0.50-1.06)	0.096
NGLDM_Coarseness	1.45 (1.11-1.90)	0.003	1.43 (1.10-1.87)	0.008
SMTV	1.57 (1.05-2.33)	0.027	1.64 (1.08-0.49)	0.021

Abbreviations: OS-overall survival; DFS-disease-free survival; HR-hazard ratio; NGLDM-neighborhood grey-level different matrix; SMTV-standardized metabolic tumor value;

Supplementary Table 2. Clinical characteristics of patients according to the Rad-score in the combined training and validation cohorts.

Variables	Combined cohort (N = 171)			*p*-value
	N	Low Radscore	High Radscore	*p*
Gender				0.789
Male	147	96 (65.3%)	51 (34.7%)	
Female	24	15 (62.5%)	9 (37.5%)	
Age (years)				0.036
< 60	101	72 (71.3%)	29 (28.7%)	
≥ 60	70	39 (55.7%)	31 (44.3%)	
Tumor size				0.005
≤ 4	85	64 (75.3%)	21 (24.7%)	
> 4	86	47 (54.7)	39 (45.3%)	
Tumor Location				
Oropharynx	137	91 (66.4%)	46 (33.6%)	0.538
Larynx	19	12 (63.2%)	7 (36.8%)	
Oralcavity	4	3 (75.0%)	1 (25.0%)	
Hypopharynx	11	5 (45.5%)	6 (54.5%)	
Differentiation status				0.234
Well	17	4 (23.5%)	13 (76.5%)	
Moderate	79	54 (68.4%)	25 (31.6%)	
Poor and undifferentiated	75	44 (58.7%)	31 (41.3%)	
T Stage*				< 0.001
T1	31	20 (64.5%)	11 (35.5%)	
T2	54	44 (81.5%)	10 (18.5%)	
T3	52	36 (69.2%)	16 (30.8%)	
T4	34	11 (32.4%)	23 (67.6%)	
N Stage*				
N0	19	14 (73.7%)	5 (26.3%)	0.066
N1	19	14 (73.7%)	5 (26.3%)	
N2a	9	9 (100%)	0 (0%)	
N2b	75	46 (61.3%)	29 (38.7%)	
N2c	37	24 (64.9%)	13 (35.1%)	
N3	12	4 (33.3)	8 (66.7%)	
TNM stage*				0.044
I	1	1 (100%)	0 (0%)	
II	5	5 (100%)	0 (0%)	
III	28	22 (78.6%)	6 (21.4%)	
IVA	122	77 (63.1%)	45 (36.9%)	
IVB	15	6 (40%)	9 (60%)	

* according to 7th AJCC stage system.

Supplementary Table 3. Univariate Cox regression analyses for OS and DFS in the combined cohort.

Variables	Overall survival		Disease-free survival	
	HR (95%CI)	*p*	HR (95%CI)	*p*
Univariate Cox Regression Analysis				
PET outcome (vs. negative)	6.60 (3.65-11.97)	<0.001	8.17 (4.45-15.00)	<0.001
Rad-score	6.43 (2.68-15.4)	<0.001	6.81 (2.75-16.85)	<0.001
Start-treatment BMI	0.90 (0.85-0.97)	0.002	0.90 (0.85-0.96)	<0.001
Cancer site (vs. Oropharynx)				
Hypopharynx	5.85 (2.55-13.44)	<0.001	5.43 (2.37-12.39)	<0.001
Oral cavity	1.69 (0.23-12.54)	0.606	1.57 (0.21-11.6)	0.658
Larynx	2.68 (0.80-9.70)	0.114	2.58 (0.75-8.90)	0.135
N stage (vs. N0-2)	3.08 (1.36-6.97)	0.007	2.95 (1.29-6.73)	0.010
Age	1.05 (1.02-1.09)	0.001	1.05 (1.02-1.08)	0.002
Histologic grade (vs. well-moderate differentiated)	0.57 (0.31-1.06)	0.076	0.56 (0.30-1.05)	0.071
T stage (vs. T1-3)	1.76 (0.89-3.49)	0.105	1.66 (0.84-3.29)	0.146
7th AJCC Stage (vs. stage I-III)	0.68 (0.35-1.35)	0.273	0.66 (0.33-1.31)	0.232
Smoking history (vs. never)				
Less than 10 pack-years	1.43 (0.52-3.92)	0.485	1.51 (0.55-4.11)	0.422
greater or equal than 10 pack-years	1.13 (0.60-2.13)	0.705	1.13 (0.60-2.13)	0.699
Induction Chemotherapy (vs. no)	0.93 (0.44-1.55)	0.557	0.82 (0.44-1.53)	0.526
Concurrent Chemoradiotherapy (vs. no)	1.07 (0.59-1.95)	0.818	1.07 (0.59-1.94)	0.829

Abbreviations: OS-overall survival; DFS-disease free survival; HR-hazard ratio; BMI-body mass index; PET- positron emission tomography/computed tomography.

Supplementary Table 4. Univariate and multivariate Cox regression analyses for OS and DFS in the radiomics model and traditional PET model.

Variables	Overall survival			Disease-free survival		
	HR (95%CI)	*p*	C-index	HR (95%CI)	*p*	C-index
Rad-score	6.332 (2.765-14.5)	< 0.001	**0.641**	6.844 (2.904-16.13)	< 0.001	**0.643**
Conventional PET features						
Multivariate analysis			**0.588**			**0.576**
MTV	1.4069 (0.8850-2.236)	0.149		1.335 (0.826-2.159)	0.238	
TLG	0.9903 (0.558-1.758)	0.973		1.061 (0.592-1.902)	0.844	
SUV_{max}	0.7968 (0.442-1.437)	0.450		0.855 (0.477-1.534)	0.600	
SUV_{mean}	1.2836 (0.612-2.694)	0.509		1.141 (0.541-2.407)	0.729	
Univariate analysis						
MTV	1.408 (1.123-1.764)	0.003	0.567	1.400 (1.109-1.768)	0.004	0.562
TLG	1.286 (1.055-1.567)	0.013	0.579	1.281 (1.049-1.565)	0.015	0.580
SUV_{max}	1.045 (0.8141-1.341)	0.731	0.501	1.034 (0.805-1.329)	0.792	0.511
SUV_{mean}	1.105 (0.863-1.415)	0.430	0.530	1.082 (0.844-1.387)	0.535	0.536

Abbreviations: OS-overall survival; DFS-disease-free survival; HR-hazard ratio; PET- positron emission tomography/computed tomography; C-index-Concordance index; MTV-metabolic tumor value; TLG-total lesion glycolysis; SUV- standardized uptake values.

Supplementary Table 5. The image features extracted from PET/CT images of HNSCC patients.

PET paremeters	Intensity features	Shape	GLCM	GLRLM	NGTDM	GLZLM
SUV_{min}	HISTO_Skewness	SHAPE_Volume_ml	GLCM_Homogeneity	GLRLM_SRE	NGLDM_Coarseness	GLZLM_SZE
SUV_{mean}	HISTO_Kurtosis	SHAPE_Volume_vx	GLCM_Energy	GLRLM_LRE	NGLDM_Contrast	GLZLM_LZE
SUV_{std}	HISTO_ExcessKurtosis	SHAPE_Sphericity	GLCM_Contrast	GLRLM_LGRE	NGLDM_Busyness	GLZLM_LGZE
SUV_{max}	HISTO_Entropy_log10	SHAPE_Compacity	GLCM_Correlation	GLRLM_HGRE		GLZLM_HGZE
SUV_{sum}	HISTO_Entropy_log2		GLCM_Entropy_log10	GLRLM_SRLGE		GLZLM_SZLGE
SUVQ1	HIOTO_Energy		GLCM_Entropy_log2	GLRLM_SRHGE		GLZLM_SZHGE
SUVQ2			GLCM_Dissimilarity	GLRLM_LRLGE		GLZLM_LZLGE
SUVQ3				GLRLM_LRHGE		GLZLM_LZHGE
SUVpeak.sphere.0.5mL				GLRLM_GLNU		GLZLM_GLNU
SUVpeak.sphere.1mL				GLRLM_RLNU		GLZLM_ZLNU
MTV				GLRLM_RP		GLZLM_ZP
SMTV						
TLG						
STLG						