

Association of quantitative MRI-based radiomic features with prognostic factors and recurrence rate in oropharyngeal squamous cell carcinoma

C. GIANNITTO^{1,*}, G. MARVASO^{2,3,*}, F. BOTTA⁴, S. RAIMONDI⁵, D. ALTERIO², D. CIARDO^{2,*}, S. VOLPE^{2,3}, F. DE PIANO⁶, E. ANCONA⁶, M. TAGLIABUE⁷, D. ORIGGI⁴, S. CHIOCCA⁸, F. A. MAFFINI⁹, M. ANSARIN⁷, V. BAGNARDI¹⁰, F. CATTANI⁴, F. NOLÈ¹¹, L. PREDA^{12,13}, R. ORECCHIA^{14,15}, E. CASSANO¹⁶, M. CREMONESI¹⁷, A. STARZYŃSKA¹⁸, M. BELLOMI^{1,3,5}, B. A. JERECZEK-FOSSA^{2,3,5}

¹Department of Radiology, IEO, European Institute of Oncology IRCCS, Milan, Italy; ²Division of Radiation Oncology, IEO, European Institute of Oncology IRCCS, Milan, Italy; ³Department of Oncology and Hemato-oncology, University of Milan, Milan, Italy; ⁴Unit of Medical Physics, IEO, European Institute of Oncology IRCCS, Milan, Italy; ⁵Molecular and Pharmaco-Epidemiology Unit, Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, Milan, Italy; ⁶Postgraduation School in Radiodiagnosics, University of Milan, Milan, Italy; ⁷Division of Otolaryngology and Head and Neck Surgery, IEO, European Institute of Oncology IRCCS, Milan, Italy; ⁸Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, Milan, Italy; ⁹Department of Pathology, IEO, European Institute of Oncology IRCCS, Milan, Italy; ¹⁰Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy; ¹¹Department of Medical Oncology, Division of Urogenital and Head and Neck Tumors, IEO, European Institute of Oncology IRCCS, Milan, Italy; ¹²Department of Clinical-Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Pavia, Italy; ¹³National Centre of Oncological Hadrontherapy (CNAO), Pavia, Italy; ¹⁴Scientific Directorate, IEO, European Institute of Oncology IRCCS, Milan, Italy; ¹⁵Department of Medical Imaging and Radiation Sciences, IEO, European Institute of Oncology IRCCS, Milan, Italy; ¹⁶Breast Imaging Unit, IEO, European Institute of Oncology IRCCS, Milan, Italy; ¹⁷Radiation Research Unit, IEO, European Institute of Oncology IRCCS, Milan, Italy; ¹⁸Department of Oral Surgery, Medical University of Gdansk, Gdansk, Poland

*Correspondence: delia.ciarDO@gmail.com

^{*}Contributed equally to this work. [§]Contributed equally to this work as senior authors.

Received March 10, 2020 / Accepted May 24, 2020

Radiomics focuses on extracting a large number of quantitative imaging features and testing both their correlation with clinical characteristics and their prognostic and predictive values. We propose a radiomic approach using magnetic resonance imaging (MRI) to decode the tumor phenotype and local recurrence in oropharyngeal squamous cell carcinoma (OPSCC). The contrast-enhanced T1-weighted sequences from baseline MRI examinations of OPSCC patients treated between 2008 and 2016 were retrospectively selected. Radiomic features were extracted using the IBEX software, and hierarchical clustering was applied to reduce features redundancy. The association of each radiomic feature with tumor grading and stage, HPV status, loco-regional recurrence within 2 years, considered as main endpoints, was assessed by univariate analysis and then corrected for multiple testing. Statistical analysis was performed with SAS/STAT® software. Thirty-two eligible cases were identified. For each patient, 1286 radiomic features were extracted, subsequently grouped into 16 clusters. Higher grading (G3 vs. G1/G2) was associated with lower values of GOH/65Percentile and GOH/85Percentile features ($p=0.04$ and 0.01 , respectively). Positive HPV status was associated with higher values of GOH/10Percentile ($p=0.03$) and lower values of GOH/90Percentile ($p=0.03$). Loco-regional recurrence within 2 years was associated with higher values of GLCM3/4-7Correlation ($p=0.04$) and lower values of GLCM3/2-1InformationMeasureCorr1 ($p=0.04$). Results lost the statistical significance after correction for multiple testing. T stage was significantly correlated with 9 features, 4 of which (GLCM25/180-4InformationMeasureCorr2, Shape/MeanBreadth, GLCM25/90-1InverseDiffMomentNorm, and GLCM3/6-1InformationMeasureCorr1) retained statistical significance after False Discovery Rate correction. MRI-based radiomics is a feasible and promising approach for the prediction of tumor phenotype and local recurrence in OPSCC. Some radiomic features seem to be correlated with tumor characteristics and oncologic outcome however, larger collaborative studies are warranted in order to increase the statistical power and to obtain robust and validated results.

Key words: radiomics, oropharyngeal squamous cell carcinoma, prognostic factors, Human Papilloma Virus (HPV)

Head and neck squamous cell carcinoma is a heterogeneous disease. Current classification based only on the anatomical site and tumor (T) stage fails to capture the biological heterogeneity of these malignancies, thus preventing an adequate patient stratification. The clinical scenario of oropharyngeal squamous cell carcinoma (OPSCC) has recently changed with the identification of Human Papilloma Virus (HPV) as a prognostic factor, with 5-year overall survival (OS) ranging from 50% in HPV-negative to 80% in HPV-positive OPSCC [1]. So far, there is no general consensus on the HPV identification, which is mainly based on p16 staining or polymerase chain reaction (PCR) amplification of HPV DNA. About 10–20% of OPSCC categorized as HPV-positive show poor treatment outcome, thus reducing the sensitivity of HPV prognostic value when used in clinical practice [2, 3]. Further potential prognostic factors relying on tumor molecular characterization could be derived from bioptic samples. The main limitations of this approach are i) the invasiveness, with potential side effects for the patient, ii) the non-repeatability of the analysis and iii) the potential misinterpretation of the results due to inherent heterogeneity of the tumor along with iv) the fairly smallness of the piece of tissue to study before the surgical treatment. Moreover, the psychological component of the patient should be considered: the risk of repeated invasive or unsafe tests, eventually with uncertain diagnostic results, could be burdensome [4]. Imaging analysis with radiomic characterization might potentially overcome these drawbacks, with a non-invasive, minimal, or non-detrimental approach (depending on the nature of radiations, ionizing or not), repeatability of the imaging analysis, and the ability to represent the whole tumor volume. Radiomics refers to the comprehensive evaluation of tumor phenotype by the means of a high throughput extraction of quantitative image features. Recent investigations into radiomics demonstrated the prognostic value of this analysis as a complementary tool for the optimization of treatment strategies [4]. The radiomic ability to identify intratumor heterogeneity may serve as a bridge from invasive to non-invasive diagnostic medicine useful for diagnosis, treatment response evaluation, and follow-up.

The majority of published studies performed the radiomic analysis on computer tomography (CT) imaging, due to easy data extraction and interpretation. Indeed, for CT images the imaging signal is directly related to tissue radiodensity, the variety of acquisition, and reconstruction parameters are more easily manageable, and standardization among different centers is more approachable as compared to magnetic resonance imaging (MRI-multiparametric imaging). Previously published CT-based works support the rationale for the application of radiomics in the setting of OPSCC. A recent study by Ranjbar et al. has compared the classification accuracy of machine learning and of a trained human eye in predicting HPV status using CT-based textural analysis and demonstrated that the accuracy of the best machine learning

model was significantly better than those of the two human controls ($p < 0.001$ and $p = 0.002$) [5]. Other authors managed to identify radiomic models for the prediction of HPV status and/or loco-regional control in either head-and-neck cancer as a whole [6, 7] or in homogenous subsets of OPSCC patients [8, 9].

Compared to CT studies, the appeal of developing MRI-based radiomic applications derives from i) the better soft-tissue contrast resolution, ii) the lack of dental filling artifacts, iii) the possibility of incorporating information from functional dedicated features and iv) the use of non-ionizing radiation. Despite these potential advantages, MRI-based radiomic studies suffer from technological issues that hinder procedure standardization [10]. In this pilot study, we retrospectively analyzed radiomic features extracted from T1-weighted MRI sequences of OPSCC patients. The aim is to decode tumor phenotype and treatment response using a radiomic approach with MRI by mean of quantitative features, particularly focusing on a possible correlation with HPV status, tumor grading, loco-regional control at 2 years.

Patients and methods

We retrospectively reviewed all consecutive patients treated in our Institute for OPSCC between 2008 and 2016. The study was part of a research project on the head and neck cancer radiotherapy notified to the Institutional Ethical Committee (notification number IEO N94/11). All patients signed written informed consent for the use of their anonymized data for clinical research and educational purpose.

Inclusion criteria. The patient inclusion criteria were as follows: diagnosis of OPSCC (all tumor stages); T \neq Tx; no previous treatment for head and neck cancer; patients treated between 2008 and 2016 with multiple modality (surgery with/without chemo-radiotherapy or concomitant chemo-radiotherapy); curative intention to treat; clinical follow-up at our Institution, records available; willingness and ability to sign a written informed consent for clinical research and educational purposes; availability of baseline MRI performed at our Institution; all cases were discussed at the institutional multidisciplinary tumor board. The image set inclusion criteria were as follows: availability of three-dimensional T1 weighted gradient-echo sequence acquired with isotropic voxel ($0.6 \times 0.6 \times 0.6$ mm³) after Gadolinium contrast injection (acquisition parameters: repetition time TR=7.4 ms, echo time TE=2.8 ms); tumor volume >1 cm³ on three-dimensional T1 weighted sequences; all images were acquired on an AvantoFit 1.5 T scanner (Siemens Healthineers, Erlangen, Germany) with a dedicated head-and-neck 20-channel phased array coil. The tumor region was identified and manually drawn on the three-dimensional images by two expert radiologists and extracted as RT Structure file on AW Server 3.2 workstation (GE Healthcare, Milwaukee, WI). The final patient cohort included in the analysis fulfilled both clinical and imaging criteria.

Clinical data. For all patients, a surgical specimen was obtained from a biopsy or surgical excision. Every specimen was fixed in buffered formalin, cut, and stained with a standard Hematoxylin & Eosin. Histological diagnosis was reported inclusive of histological type and grading following a Broder's method [11]. After that in cases almost all specimen was evaluated a p16^{ink4} (p16) staining that showed cytoplasmic and nuclear staining in more than 75% in almost all patients as reported in literature and suggestive of HPV infection. After that for a more accurate value for high-risk (HR-HPV) OPSCC, we had performed a polymerase chain reaction (PCR) for HPV searching an HPV DNA host integration. The HPV DNA positive cases confirm a diagnosis of SCC HPV related.

The cases positive for p16 staining were all positive for HPV DNA while the p16 negative OPSCCs were also negative for HPV DNA [12, 13].

The staging was assessed based on the clinical and radiological exams, following the 8th edition of the TNM classification of Malignant Tumors by the AJCC [14].

Loco-regional recurrence was assessed by clinical and/or radiological evaluation and was registered as the time from the end of treatment until the first occurrence of disease in oropharynx or neck. All patients were checked every 3 months after treatment with video-endoscopy, and every 6 months with MRI or PET based on clinical evaluation.

Extraction of MRI-based radiomic features. The MRI of the selected patients, along with the corresponding RT Structure file identifying the tumor, were uploaded in the Imaging Biomarker Explorer (IBEX) software [15].

The Norm_ThreeSigma pre-processing was applied before calculating the radiomic features, with default settings ($n=3$, BitDepth=8). In this way, for each patient, the range [average - 3σ , average + 3σ] of voxel intensities was linearly rescaled to the $[1-2^8]$ range. All the available features belonging to the following categories were calculated: Gray Level Cooccurrence Matrix (GLCM) 2.5 and 3 (where 2.5 indicates that the occurrence of the

intensity pairs is computed in 2D slice by slice and then summed from all 2D image slices, whereas 3 indicates that the occurrences are directly computed in 3D), Gray Level Run Length Matrix (GLRLM) 2.5, Gradient Orient Histogram (GOH), and Shape. In addition, 7 features from the Intensity Direct category (GlobalEntropy, GlobalUniformity, Kurtosis, Skewness, LocalEntropyMax, LocalEntropyMean, LocalEntropyMin) and the Busyness feature from the Neighbor Intensity Difference (NID) categories 2.5 and 3 were extracted. For GLCM, GLRLM and NID calculation, the [0-256] interval was re-sampled into 128 equally spaced bins (IBEX settings GrayLimits=256, NumLevels=128, or RangeMin=0, RangeMax=256, NBins=256). For the remaining IBEX settings, the default values proposed for each radiomic feature category were kept.

A dedicated routine written in MATLAB® environment (The MathWorks, Inc., Natick, MA) was used to eliminate identical features, which occurred twice or more in the dataset, due to equal definitions for features belonging to different categories.

Statistical analysis. Baseline characteristics of patients and tumors were expressed as frequency and percentages for categorical variables (gender, tumor stage, tumor site, HPV status, grading, type of treatment) and as median and range for continuous variables (age). The association of patient and tumor characteristics with grading (G1/G2 vs. G3), HPV (positive vs. negative), recurrence within 2 years (yes vs. no) and T stage (Tis/T1/T2 vs. T3/T4) was assessed with the Fisher Exact test for categorical variables and the non-parametric Wilcoxon two-independent samples test for continuous variables.

In order to reduce the dimensionality and to exclude redundant and highly correlated features, the radiomic features included in the analysis were grouped into clusters with a hierarchical clustering procedure, namely the VARCLUS procedure implemented in SAS/STAT® software version 9.4 (SAS Institute Inc., Cary, USA). The procedure stopped splitting when every cluster had a percentage of variation explained ≥ 0.80 .

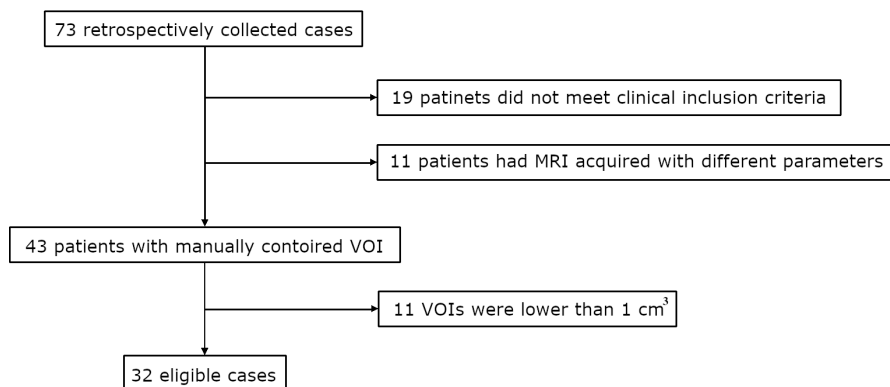


Figure 1. Diagram of the patient selection process.

Table 1. Baseline characteristics of the study population (n=32).

Characteristic	n (%)
Age at diagnosis [^] [years]	60 (42–89)
Sex	
Males	26 (81%)
Females	6 (19%)
Tumor stage (8 th ed.) [14]	
I	3 (9%)
II	3 (9%)
III	5 (16%)
IVA/IVB	21 (66%)
T stage*	
Tis	1 (3%)
T1	6 (19%)
T2	10 (31%)
T3	3 (9%)
T4	12 (38%)
N stage*	
Nx	2 (6%)
N0	5 (16%)
N1	5 (16%)
N2	19 (59%)
N3	1 (3%)
Tumor site	
Tonsil	17 (53%)
Base of tongue	11 (34%)
Vallecula	3 (9%)
Soft palate	1 (3%)
Grading	
1	3 (9%)
2	9 (28%)
3	19 (59%)
Missing	1 (3%)
HPV status	
Negative	9 (28%)
Positive	20 (63%)
Not classifiable	3 (9%)
Treatment	
Surgery	6 (19%)
Surgery+CHT+RT	5 (16%)
CHT+RT	18 (56%)
Surgery+RT	3 (9%)

Abbreviations: T = tumor; N = nodes; HPV = human papilloma virus; CHT = chemotherapy; RT = radiotherapy; [^]median (range); *either clinical or pathological

The univariate association of each radiomic feature with the main endpoints (namely grading, HPV status, recurrence within two years, and T) was assessed by the non-parametric Wilcoxon two-independent samples test. For each cluster, the feature with the lowest p-value was selected. False Discovery Rate (FDR) correction for multiple testing was applied and reported. The analyses were performed using SAS software (SAS Institute Inc., Cary, USA).

Results

Patient cohort. Of the 73 retrospectively collected patients, 32 were eligible for the analysis. The diagram in Figure 1 presents the entire selection process. Table 1 reports the main characteristics of the analyzed patients. Most patients (26/32, 81%) had a locally advanced stage, namely, stage III and IV tumors according to the American Joint Committee on Cancer/Union for International Cancer Control (AJCC/UICC) [14]. Seventeen patients (53%) had a diagnosis of tonsil cancer, 12 (37.5%) of the base of tongue cancer, and 1 (3%) of soft palate cancer. Twenty patients (62%) were HPV-positive, 9 patients were HPV-negative (28%), whereas HPV status was unknown in 3 patients (9%). In 25 out of 32 patients, the HPV DNA PCR confirmed the p16 HPV staining. In the remaining 7 samples, the HPV DNA was not performed. The treatment strategy, according to international guidelines and institutional multidisciplinary tumor board, was a surgery (mini-invasive robot-assisted surgery or open surgery eventually followed by adjuvant radiotherapy and/or chemo-radiotherapy based on pathological stage) or concurrent chemo-radiation in 14 and 18 cases, respectively. After a median 30-month follow-up (range 4–100 months), 29 patients were alive without disease, 2 patients died from tumor progression, and 1 died from other causes. Among the 20 patients with at least a 2-year follow-up, the loco-regional control rate was 80%. Patients with tumor grade G3 were older than patients with G1/G2 (median age 60 vs. 55, $p=0.04$, Supplementary Table S1); tonsil tumors were more frequently observed in HPV-positive than HPV-negative patients (70% vs 22%, $p=0.02$, Supplementary Table S2); smaller tumors (Tis/T1/2) were more frequently localized in tonsils (76%), while the more common localization of larger tumors (T3/T4) was the base of tongue (53%) ($p=0.04$, Supplementary Table S3).

Radiomic analysis. At first analysis, 1627 features for each patient were extracted. After eliminating identical features, 1291 features remained. In addition, 4 features from the Shape category were removed because meaningless for the study purpose (Number of Objects, Number of Voxels, Voxel Size, Mass), and the Compactness1 feature was also removed because of erroneous formula implementation in the software version in use, thus obtaining 1286 radiomic features for each patient. After hierarchical clustering analysis, 16 clusters were obtained (Figure 2).

Feature correlating with the endpoint. Figure 3 represents the box-plot for the radiomic features with the lowest p-value, for each clinical endpoint. As far as grading is concerned, patients with G3 had lower values of GOH/65Percentile and of GOH/85Percentile compared to patients with G1/G2 ($p=0.04$ and 0.01, respectively). However, after correction for multiple testing, the results were no more statistically significant (Table 2).

HPV-positive patients had higher values of GOH/10Percentile ($p=0.03$) and lower values of GOH/90Percentile

($p=0.03$). However, after correction for multiple testing, no radiomic feature retained statistical significance (Table 2).

Patients with higher T stages showed lower values of GLCM25/180-4InformationMeasureCorr2, GLCM3/11-7Dissimilarity and GOH/5Percentile ($p=0.0008$, $p=0.04$, and $p=0.02$, respectively) and higher values of Shape/MeanBreadth, GLCM25/90-1InverseDiffMomentNorm, GLCM3/6-7Correlation, GLCM3/0-7Correlation, GLCM3/6-1InformationMeasureCorr1, and GOH/90Percentile ($p=0.01$, $p=0.01$, $p=0.02$, $p=0.03$, $p=0.004$, $p=0.03$, respectively). After correction for multiple testing, GLCM25/180-4InformationMeasureCorr2, Shape/MeanBreadth, GLCM25/90-1InverseDiffMomentNorm, and GLCM3/6-1InformationMeasureCorr1 remain statistically significantly correlated with T stage ($q=0.01$, $q=0.03$, $q=0.03$, and $q=0.03$, respectively).

Finally, patients with loco-regional recurrence within 2 years had higher values of GLCM3/4-7Correlation ($p=0.04$) and lower values of GLCM3/2-1InformationMeasureCorr1 ($p=0.04$). Results were no more statistically significant after correction for multiple testing (Table 2).

Detailed data of statistical analysis and results are reported in the Supplementary materials.

Cluster analysis: HPV status and Grading. The two features associated with HPV status were found in clusters 8 and 15, whereas the two features associated with Grading were found in clusters 7 and 15. Clusters 7, 8, and 15 were entirely composed of features belonging to the GOH category. Moreover, the two features from cluster 15 (90Percentile, associated with HPV status, and 85Percentile, associated

with Grading) are linearly correlated, with $R^2=0.93$. Gradient Orient Histogram (GOH) is the histogram of the directions of voxel intensity variations. A very repetitive pattern is expected to provide a histogram with few directions, whereas a variegated (or noisy) pattern should yield a histogram with more different directions. In this sense, our results seem to suggest that both HPV-negative and low grading lesions may

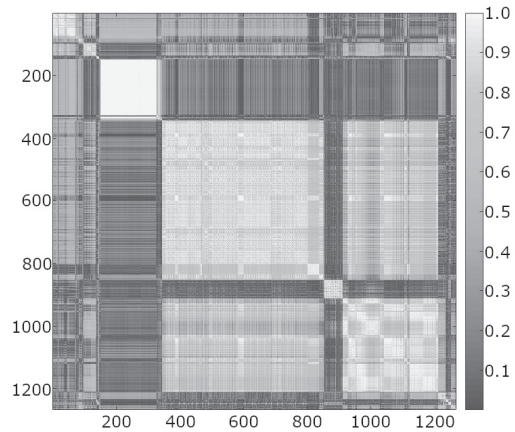


Figure 2. Heat map showing the clustering of the radiomic features. The 1286 radiomic features are ordered along the two-axis grouped into clusters. The color scale represents the absolute value of the Pearson correlation coefficient among each feature pair. The squares along the diagonal evidence the high intra-cluster correlation, whereas the blue areas evidence low correlation among different clusters (except for one couple, represented by clusters 3 and 5).

Table 2. Association between representative[^] radiomic features from 16 clusters and grading, HPV status, local recurrence within two years and T stage: univariate analysis.

Radiomic features (cluster)	Predictive factor classes (number of cases)		p-value	Q-value [°]
	G1/G2 (n=12)	G3 (n=19)		
Grading				
GOH/65Percentile (7)	118.45 (106.34–125.36)	115.64 (108.43–124.70)	0.04	0.35
GOH/85Percentile (15)	153.43 (148.39–157.37)	152.10 (145.30–153.43)	0.01	0.19
HPV status				
	Negative (n=9)	Positive (n=20)		
GOH/10Percentile (8)	14.74 (13.32–16.70)	16.54 (12.99–19.80)	0.03	0.22
GOH/90Percentile (15)	161.57 (158.20–164.48)	161.02 (156.37–162.18)	0.03	0.22
T stage				
	Tis/T1/T2 (n=17)	T3/T4 (n=15)		
GLCM25/180-4InformationMeasureCorr2 (1)	0.73 (0.47–0.89)	0.55 (0.39–0.7)	0.0008	0.01
Shape/MeanBreadth (2)	2.65 (1.58–10.43)	5.14 (2.84–14.66)	0.01	0.03
GLCM25/90-1InverseDiffMomentNorm (3) [*]	996.04 (992.53–997.74)	996.89 (995.71–998.46)	0.01	0.03
GLCM3/11-7Dissimilarity (5)	13.23 (10.09–20.2)	10.80 (7.48–15.44)	0.04	0.07
GOH/5Percentile (8)	7.35 (5.87–9.46)	6.34 (5.19–7.91)	0.02	0.06
GLCM3/6-7Correlation (9)	0.03 (–0.08–0.2)	0.10 (–0.02–0.33)	0.02	0.06
GLCM3/0-7Correlation (10)	0.08 (–0.12–0.32)	0.20 (–0.04–0.36)	0.03	0.06
GLCM3/6-1InformationMeasureCorr1 (13)	–0.12 (–0.21–0.07)	–0.10 (–0.13–0.07)	0.004	0.03
GOH/90Percentile (15)	160.35 (156.37–162.35)	161.57 (158.20–164.48)	0.03	0.06
Local recurrence				
	No (n=16)	Yes (n=4)		
GLCM3/4-7Correlation (12)	0.08 (–0.07–0.23)	0.15 (0.14–0.17)	0.04	0.31
GLCM3/2-1InformationMeasureCorr1 (13)	–0.21 (–0.31–0.16)	–0.23 (–0.25–0.23)	0.04	0.31

Abbreviations: GOH = Gradient Orient Histogram; GLCM = Gray Level Co-occurrence Matrix; [^]for each cluster, we reported the features with the lowest p-value. For equal p-values, we chose the first calculated feature; [°]False Discovery Rate adjusted p-values for the non-parametric Wilcoxon two-independent samples test; ^{*}values divided by 1000

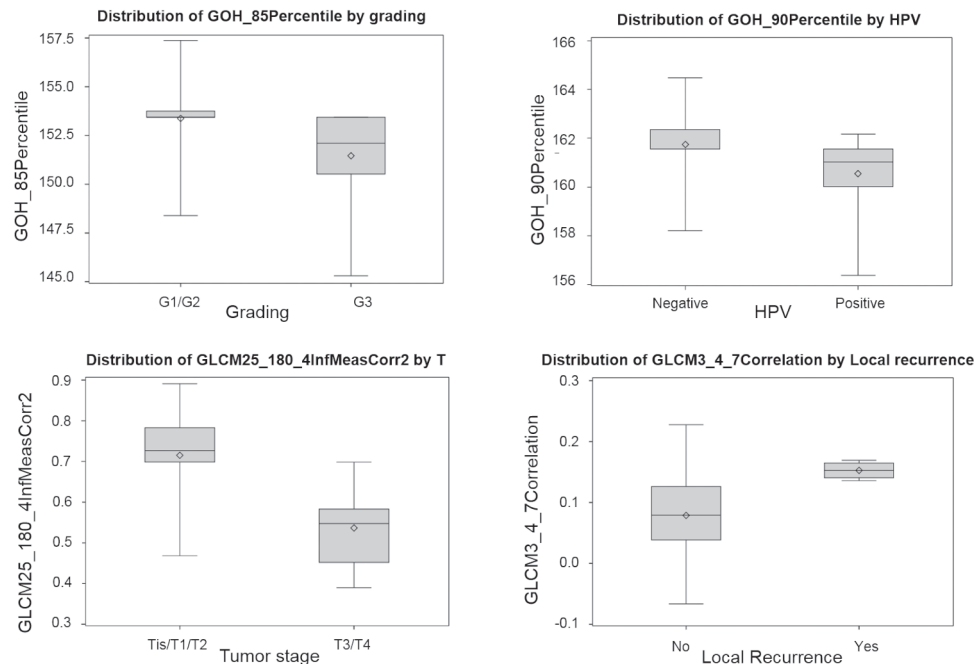


Figure 3. Box plot representing the distribution of the radiomic features A) GOH/85Percentile according to grading, B) GOH/90 Percentile according to HPV status, C) GLCM25/180-4InformationMeasureCorrelation2 according to T stage, D) GLCM3/4-7Correlation according to local recurrence. Minimum and maximum are depicted by whiskers, the box signifies the upper and lower quartiles, the median and the mean are represented respectively by a line and a small rhombus within the box.

be associated with a more variegated texture as compared to HPV-positive and High Grading lesions.

T-stage. The highest correlation with T-stage was obtained for the feature 180-4InformationMeasureCorr2 from the GLCM25 category, belonging to cluster 1 (p -value=0.0008, Q -value=0.01). Cluster 1 includes the features highly correlated with the lesion volume, and 180-4InformationMeasureCorr2 exhibits a moderate inverse correlation with volume as well ($R^2=0.73$), indirectly confirming the T-stage definition (T1/T2 stage tumors having a smaller volume than T3/T4 stage lesions). Similarly, the MeanBreath feature from the Shape category (cluster 2) measures the mean width of the lesion and is closely related to the size of the tumor. The two remaining features significantly correlating with T-stage are both texture features: 90-1InverseDiffMomentNorm from GLCM25 category in cluster 3 and 6-1InformationMeasureCorr1 from GLCM3 category in cluster 13. Within cluster 3, 90-1InverseDiffMomentNorm exhibits quite good linear correlation with the feature Homogeneity ($R^2=0.74$) and no correlation with volume, suggesting that T1/T2 stage lesions may have a less homogeneous texture as compared to T3/T4 lesions. Cluster 13 is entirely composed by InformationMeasureCorr features, calculated along different directions both in 2 and 3 dimensions; in particular, 6-1InformationMeasureCorr1 does not correlate with volume and may suggest that T1/T2 stage lesions have a different “randomness” degree as compared to T3/T4 lesions.

Local Recurrence within 2 years. The features associated with local recurrence within 2 years were found in cluster 12 (4-7 Correlation) and cluster 13 (2-1InformationMeasureCorr1), and both belong to the GLCM3 texture category. Cluster 12 is composed of only three Correlation features, calculated along different directions and with a different offset, highly correlated among them; our result suggests that lesions without local recurrence exhibit lower correlation among voxel intensities as compared to lesions with local recurrence within 2 years. Similarly, cluster 13 includes only InformationMeasureCorr type features; 2-1InformationMeasureCorr1 has a moderate linear correlation ($R^2=0.68$) with the 6-1InformationMeasureCorr1 features, belonging to the same cluster 13 and significantly associated to T-stage; smaller values of these features appear to be associated to both T1/T2 stage and to the presence of Local Recurrence.

Discussion

Our work provides a preliminary report on MRI-based radiomic analysis for a homogenous cohort of OPSCC patients and the results could, therefore, be applied to patients with similar characteristics and image settings. Although we could not identify any radiomic signature for the target variables, due to the small sample size, in the univariate analysis we still could demonstrate an association between features of the GLCM category with HPV status, grading, T

stage, and loco-regional control, which deserves to be further investigated on larger datasets. Our results evidenced that both HPV-negative and low grading lesions may be associated with a more variegated texture as compared to HPV-positive and High Grading lesions. The fact that both HPV status and grading were associated with the same category of radiomic features (GOH) indirectly indicated a relationship between the two clinical endpoints. This aspect is still debated, with controversial findings in the scientific literature [16, 17]. Our clinical data, despite not statistically significant (Supplementary Tables S1, S3), confirm the association between HPV-positive status and high grading already observed in previously published studies [18]. We also observed that HPV-positive subjects tend to have smaller mean Breadth values, which indicates that the tumor size is smaller in the HPV-positive subjects; the same result was obtained by Yu et al. on CT images [9]. Additionally, our data indicated that T1/T2 stage lesions may have a less homogeneous texture and a different “randomness” degree (represented by the InformationMeasureCorrelation feature) as compared to T3/T4 lesions and that smaller values of InformationMeasureCorrelation features appear to be associated to both T1/T2 stage and to the presence of Local Recurrence.

To the best of our knowledge, this is one of the first dedicated MRI-based radiomic studies for the prediction of stage and HPV status in OPSCC, and the first one including also grading and loco-regional control as clinical endpoints. Ren et al. [19] considered a different stage endpoint (TNM stage) and a more heterogeneous population as compared to the present work, including patients affected by squamous cell carcinoma in the whole head and neck district, not only OPSCC. Despite not directly comparable to ours, their encouraging although preliminary results, comprehensive of development and validation of a radiomic signature for the distinction of stage I–II from stage III–IV HNSCC, support the validity of the present analysis on T stage and the necessity of further investigation on the possible role of MRI radiomics in OPSCC staging. Other researchers investigated the correlation between HPV status and quantitative parameters extracted from both morphological MRI sequences and Apparent Diffusion Coefficient (ADC) maps [20–24]. Among them, Ravanelli et al. [20] performed a quantitative analysis of multiparametric MRI of OPSCC patients, including a contrast-enhanced T1 sequence similar to the one analyzed in this study, but they focused on a single slice of the tumor and calculated a limited number of features, not including the second-order statistics features investigated in this study. They concluded that only the mean ADC value inside the lesion was significantly correlated to HPV status after false discovery rate correction, and mean ADC was successfully included in a predictive model for HPV status in combination with smoking status. Integrating these findings with our results, which suggest a possible correlation between second-order statistics features (extracted from contrast-enhanced T1-weighted images) and HPV status, could lead to a new

model with improved predictive power. A systematic review recently published by Jethanandani et al. [10] highlighted that, as a matter of fact, head-and-neck MRI-based radiomics is currently a young and largely unexplored field, for different reasons. Not only the anatomical and biological complexity of the head-and-neck subsites is a source of variability, but more general concerns regarding discrepancies in study design, lack of a common ontology, insufficient reporting of data sourcing, modelling processing, and impact of the use of different MR scanners, acquisition parameters, and image analysis methodologies have been raised. The sample size is also an issue: radiomic studies need large datasets in order to increase the statistical power and obtain robust and reproducible results. As a result, head-and-neck MRI radiomic studies are hardly comparable and poorly generalizable. To move forward, collaboration among research groups and cooperative consortia are warranted in order to join data and better investigate methodological variances [25, 26].

In consideration of this, for the present study, we made a rigorous selection based on the clinical characteristics, and we adopted methodological choices aimed to maximize the robustness of the radiomic features extracted. Firstly, we strived to overcome the problem of imaging variability through a strict selection of imaging studies. Specifically, all MRI examinations considered for the analysis were acquired on the same scanner with fixed acquisition parameters (TE, TR, and voxel size). Secondly, we choose to apply a threshold on VOIs size based on a preliminary phantom analysis indicating that radiomic features lose repeatability at decreasing VOI size, and on the evidence that some radiomic features may lose their ability to quantify heterogeneity when calculated on a small amount of voxels, due to an under-sampling of the intensity distribution [27]. In this regard, including only patients with VOIs larger than 1 cm^3 was a conservative choice. Thirdly, we tested different pre-processing and IBEX settings for radiomic features calculation, aiming to investigate how they affect the radiomic analysis, and tentatively identify the procedure maximizing the ability to predict the clinical endpoints. Specifically, the whole analysis was repeated with or without applying the Norm_ThreeSigma pre-processing option available in IBEX, and discretizing at 128 or 256 equally spaced bins before calculating GLCM, GLRLM and NID matrices (data not shown). Interestingly, the different procedures did not lead to significant differences in our series. This may be interpreted as an index of the robustness of our results, but further data are needed to validate this assumption. Lastly, being aware that in the present study MRI did not undergo standardization of intensity scale (fundamental to make different patient examinations quantitatively comparable [28]), we choose to calculate only the radiomic features which are unaffected by the normalization procedure. A dedicated test confirmed that the features included in this study varied by less than 1% when calculated on identical images with different voxel intensities but fixed contrast (data not shown).

Despite these methodological precautions, we acknowledge that our retrospective study has some limitations. The sample size was not big enough to allow for internal validation of our findings and to obtain a radiomic signature for the investigated clinical parameters by applying multivariate analysis techniques. The combination of quantitative MRI-based radiomic features with other routine collected clinical variables may indeed increase the accuracy of the prediction of recurrence and tumor progression, however, the low number of local recurrence ($n=4$) did not allow us to create and validate a clinical-radiomic predictive model. Our exploratory univariate analysis of association could be nevertheless of great interest as a starting point for future, larger, and collaborative studies, possibly including both radiomic and clinical variables. The low number of patients, their heterogeneity (tumor stage, treatment approach, etc.), missing data typical for retrospective series (HPV status in some patients, tobacco-smoking habits, etc.), and the high number of radiomic features included in the analysis also rose the problem of multiple testing. We tried to overcome this issue in 2 steps: 1) with a multidimensionality reduction obtained with cluster analysis and aimed at excluding redundant features and 2) by applying and reporting FDR correction for multiple testing. The application of multiple testing correction, especially in explorative analyses, has been largely discussed also concerning the consequent statistical power reduction [29–32]. We believe that our approach of describing the performed tests of significance and reporting both the original and the FDR corrected p-values is appropriate as a primary analysis in the context of an explorative pilot study. Besides, despite our attempt to reduce variability, it was not possible to overcome the limitations inherent to all MRI-based radiomic studies: magnetic field inhomogeneity in the VOI, the variability of the patient position inside the coil, and variability of the tumor anatomical location inside the patient. Further studies on larger cohorts are warranted to quantify the impact of these parameters on textural analysis. Furthermore, a systematic analysis of the stability and robustness of the 1286 radiomic features was not performed. In principle, only radiomic features that remain stable under repeated measurements and are not significantly affected by acquisition settings should be selected for the analysis [33]. As a consequence, it cannot be excluded that some of the 1286 radiomic features considered for the analysis may not be robust or stable. This issue is currently under investigation in a dedicated phantom study, together with the implementation of methods to correct for magnetic field inhomogeneity [34] and the study of the dependence of feature robustness on VOI volume. The future outcomes of such analysis may actually lead to reconsider the criteria for patient inclusion in terms of VOI threshold and radiomic feature selection.

Future studies could also include the analysis of other MRI sequences in association with the contrast-enhanced T1-weighted sequence considered in this study. Indeed, the

study from Ren et al. [19] evidenced a higher performance of the TNM stage predictive model based on both contrast-enhanced T1- and T2-weighted images, as compared to the same model based on either of the two sequences, whereas Ravanelli et al. [20], among others, highlighted the significant role of ADC maps from DWI images in the HPV status prediction. The extension of the analysis to multiparametric MR would introduce, for each sequence, the challenges related to MR quantitative analysis, which may require different strategies since each sequence may exhibit different properties in terms of spatial resolution, signal to noise ratio, and possible anisotropy of signal sampling. This however will not represent a major impairment, considering the continuous improvement of radiomics methodology and the increasing efforts towards standardization initiatives. As a further challenge, the analysis of additional sequences would increase the number of model parameters, requiring a larger sample size to reach adequate statistical power for the study.

Overall, our experience confirms the current pitfalls of MRI-based radiomics and offers some possible methodological solutions, which may be discussed and reviewed in the setting of wider patient cohorts. We strongly believe that the application of MRI-based radiomics in clinical practice may provide additional insights into the biology of OPSCC and convey additional information to existing validated biomarkers, thus contributing to the creation of highly individualized algorithms for patient treatment and follow-up. As compared with CT, MRI offers a better anatomical definition of the tumor volume, which have been proven for both primary tumor and metastatic lymph nodes [35], and recognized as of prognostic value in the 8th edition of the TNM classification of Malignant Tumors by the AJCC [14]. Another promising aspect of MRI is the ability to convey cost-effective functional information for treatment planning and follow-up [36–38], which could be of immense clinical value if incorporated into validated radiomic models.

In conclusion, MRI-based radiomics is a feasible and promising approach for the prediction of local recurrence and HPV status in OPSCC. We provide a set of methodological proposals that might contribute to answering some of the still open issues in the field if discussed and applied in the setting of wider multicenter studies with a shared methodology. Arguably, the incorporation of radiomics into clinical practice may contribute to improving patient-tailored therapeutic pathways for OPSCC patients.

References

- [1] ANG KK, STURGIS EM. Human papillomavirus as a marker of the natural history and response to therapy of head and neck squamous cell carcinoma. *Semin Radiat Oncol* 2012; 22: 128–142. <https://doi.org/10.1016/j.semradonc.2011.12.004>

- [2] KECK MK, ZUO Z, KHATTRI A, STRICKER TP, BROWN CD et al. Integrative analysis of head and neck cancer identifies two biologically distinct HPV and three non-HPV subtypes. *Clin Cancer Res* 2015; 21: 870–881. <https://doi.org/10.1158/1078-0432.CCR-14-2481>
- [3] KELLY JR, HUSAIN ZA, BURTNES B. Treatment de-intensification strategies for head and neck cancer. *Eur J Cancer* 2016; 68: 125–133. <https://doi.org/10.1016/j.ejca.2016.09.006>
- [4] AERTS HJ. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol* 2016; 2: 1636–1642. <https://doi.org/10.1001/jamaoncol.2016.2631>
- [5] RANJBAR S, NING S, ZWART CM, WOOD CP, WEINDLING SM et al. Computed Tomography-Based Texture Analysis to Determine Human Papillomavirus Status of Oropharyngeal Squamous Cell Carcinoma. *J Comput Assist Tomogr* 2018; 42: 299–305. <https://doi.org/10.1097/RCT.0000000000000682>
- [6] BOGOWICZ M, RIESTERER O, BUNDSCHUH RA, VEIT-HAIBACH P, HÜLLNER M et al. Stability of radiomic features in CT perfusion maps. *Phys Med Biol* 2016; 61: 8736–8749. <https://doi.org/10.1088/1361-6560/61/24/8736>
- [7] OU D, BLANCHARD P, ROSELLINI S, LEVY A, NGUYEN F et al. Predictive and prognostic value of CT based radiomics signature in locally advanced head and neck cancers patients treated with concurrent chemoradiotherapy or bioradiotherapy and its added value to Human Papillomavirus status. *Oral Oncol* 2017; 71: 150–155. <https://doi.org/10.1016/j.oraloncology.2017.06.015>
- [8] LEIJENAAR RT, BOGOWICZ M, JOCHEMS A, HOEBERS FJ, WESSELING FW 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. <https://doi.org/10.1259/bjr.20170498>
- [9] YU K, ZHANG Y, YU Y, HUANG C, LIU R et al. Radiomic analysis in prediction of Human Papilloma Virus status. *Clin Transl Radiat Oncol* 2017; 7: 49–54. <https://doi.org/10.1016/j.ctro.2017.10.001>
- [10] JETHANANDANI A, LIN TA, VOLPE S, ELHALAWANI H, MOHAMED ASR et al. Exploring Applications of Radiomics in Magnetic Resonance Imaging of Head and Neck Cancer: A Systematic Review. *Front Oncol* 2018; 8: 131. <https://doi.org/10.3389/fonc.2018.00131>
- [11] BRODERS AC. Epithelioma of the cavities and internal organs of the head and neck. *Arch Surg* 1925; 11: 43–73. <https://doi.org/10.1001/archsurg.1925.01120130050004>
- [12] LEWIS JS JR, CHERNOCK RD, MA XJ, FLANAGAN JJ, LUO Y et al. Partial p16 staining in oropharyngeal squamous cell carcinoma: extent and pattern correlate with human papillomavirus RNA status. *Mod Pathol* 2012; 25: 1212–1220. <https://doi.org/10.1038/modpathol.2012.79>
- [13] BARNES L, EVESON JW, REICHART P, SIDRANSKY D. Pathology and Genetics of Head and Neck Tumours 2005. WHO Classification of Tumours, 3rd Edition; volume 9.
- [14] Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK et al, editors. *AJCC cancer staging manual*. 2017 8th ed. New York: Springer.
- [15] ZHANG L, FRIED DV, FAVE XJ, HUNTER LA, YANG J et al. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys* 2015; 42: 1341–1353. <https://doi.org/10.1118/1.4908210>
- [16] BEGUMS, CAO D, GILLISON M, ZAHURAK M, WESTRA WH. Tissue distribution of human papillomavirus 16 DNA integration in patients with tonsillar carcinoma. *Clin Cancer Res* 2005; 11: 5694–5699. <https://doi.org/10.1158/1078-0432.CCR-05-0587>
- [17] WESTRA WH. The changing face of head and neck cancer in the 21st century: the impact of HPV on the epidemiology and pathology of oral cancer. *Head Neck Pathol* 2009; 3: 78–81. <https://doi.org/10.1007/s12105-009-0100-y>
- [18] WESTRA WH. The Morphologic Profile of HPV-Related Head and Neck Squamous Carcinoma: Implications for Diagnosis, Prognosis, and Clinical Management. *Head Neck Pathol* 2012; 6: S48–54. <https://doi.org/10.1007/s12105-012-0371-6>
- [19] REN J, TIEN J, YUAN Y, DONG D, LI X et al. Magnetic resonance imaging based radiomics signature for the pre-operative discrimination of stage I-II and III-IV head and neck squamous cell carcinoma. *Eur J Radiol* 2018; 106: 1–6. <https://doi.org/10.1016/j.ejrad.2018.07.002>
- [20] RAVANELLI M, GRAMMATICA A, TONONCELLI E, MORELLO R, LEALI M et al. Correlation between human papillomavirus status and quantitative MR imaging parameters including Diffusion Weighted Imaging and texture features in oropharyngeal carcinoma. *Am J Neuroradiol* 2018; 39: 1878–1883. <https://doi.org/10.3174/ajnr.A5792>
- [21] Driessen JP, Van Bommel AJ, Van Kempen PM, Janssen LM, Terhaard CH et al. Correlation of human papillomavirus status with apparent diffusion coefficient of diffusion-weighted MRI in head and neck squamous cell carcinomas. *Head Neck* 2016; 38: E613.18. <https://doi.org/10.1002/hed.24051>
- [22] CHAN MW, HIGGINS K, ENEPEKIDES D, POON I, SYMONS SP et al. Radiologic differences between human papillomavirus-related and human papillomavirus-unrelated oropharyngeal carcinoma on diffusion-weighted imaging. *Otorhinolaryngol Relat Spec* 2016; 78: 344–352. <https://doi.org/10.1159/000458446>
- [23] NAKAHIRA M, SAITO N, YAMAGUCHI H, KUBA K, SUGASAWA M. Use of quantitative diffusion-weighted magnetic resonance imaging to predict human papillomavirus status in patients with oropharyngeal squamous cell carcinoma. *Eur Arch Otorhinolaryngol* 2014; 271: 1219–1225. <https://doi.org/10.1007/s00405-013-2641-7>
- [24] SCHOUTEN CS, DE GRAAF P, BLOEMENA E, WITTE BI, BRAAKHUIS BJ et al. Quantitative diffusion-weighted MRI parameters and human papillomavirus status in oropharyngeal squamous cell carcinoma. *Am J Neuroradiol* 2015; 36: 763–767. <https://doi.org/10.3174/ajnr.A4271>
- [25] OGUNDIMU EO, ALTMAN DG, COLLINS GS. Adequate sample size for developing prediction models was not simply related to events per variable. *J Clin Epidemiol* 2016; 76: 175–182. <https://doi.org/10.1016/j.jclinepi.2016.02.031>

- [26] BROOKS FJ, GRIGSBY PW. The effect of small tumor volumes on studies of intratumoral heterogeneity of tracer uptake. *J Nucl Med* 2014; 55: 37–42. <https://doi.org/10.2967/jnumed.112.116715>
- [27] ROTHMAN KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990; 1: 43–46.
- [28] NYUL LG, UDUPA JK. On standardizing the MR Image intensity scale. *Magn Reson Med* 1999; 42: 1072–1081. [https://doi.org/10.1002/\(sici\)1522-2594\(199912\)42:6<1072::aid-mrm11>3.0.co;2-m](https://doi.org/10.1002/(sici)1522-2594(199912)42:6<1072::aid-mrm11>3.0.co;2-m)
- [29] PERNEGER TV. What's wrong with Bonferroni adjustments. *BMJ* 1998; 316: 1236–1238. <https://doi.org/10.1136/bmj.316.7139.1236>
- [30] PERNEGER TV. Adjusting for multiple testing in studies is less important than other concerns. *BMJ* 1999; 318: 1288. <https://doi.org/10.1136/bmj.318.7193.1288a>
- [31] ARMSTRONG RA. When to use the Bonferroni correction. *Ophthalmic Physiol Opt* 2014; 34: 502–508. <https://doi.org/10.1111/opo.12131>
- [32] YIP SSF, AERTS HJWL. Applications and limitations of radiomics. *Phys Med Biol* 2016; 61: R150–R166. <https://doi.org/10.1088/0031-9155/61/13/R150>
- [33] SCALCO E, MARZI S, SANGUINETI G, VIDIRI A, RIZZO G. Characterization of cervical lymph-nodes using a multi-parametric and multi-modal approach for an early prediction of tumor response to chemo-radiotherapy. *Phys Med* 2016; 32: 1672–1680. <https://doi.org/10.1016/j.ejmp.2016.09.003>
- [34] TANG C, KOMAKULA S, CHAN C, MURPHY JD, JIANG W et al. Radiologic assessment of retropharyngeal node involvement in oropharyngeal carcinomas stratified by HPV status. *Radiother Oncol* 2013; 109: 293–296. <https://doi.org/10.1016/j.radonc.2013.09.001>
- [35] PARK JO, JUNG SL, JOO YH, JUNG CK, CHO KJ et al. Diagnostic accuracy of magnetic resonance imaging (MRI) in the assessment of tumor invasion depth in oral/oropharyngeal cancer. *Oral Oncol* 2011; 47: 381–386. <https://doi.org/10.1016/j.oraloncology.2011.03.012>
- [36] SWARTZ JE, DRIESSEN JP, VAN KEMPEN PMW, DE BREE R, JANSSEN LM et al. Influence of tumor and micro-environment characteristics on diffusion-weighted imaging in oropharyngeal carcinoma: A pilot study. *Oral Oncol* 2018; 77: 9–15. <https://doi.org/10.1016/j.oraloncology.2017.12.001>
- [37] CHAN SC, CHENG NM, HSIEH CH, NG SH, LIN CY et al. Multiparametric imaging using 18F-FDG PET/CT heterogeneity parameters and functional MRI techniques: prognostic significance in patients with primary advanced oropharyngeal or hypopharyngeal squamous cell carcinoma treated with chemoradiotherapy. *Oncotarget* 2017; 8: 62606–62621. <https://doi.org/10.18632/oncotarget.15904>
- [38] GREUTER MJ, SCHOUTEN CS, CASTELIJNS JA, DE GRAAF P, COMANS EF et al. Cost-effectiveness of response evaluation after chemoradiation in patients with advanced oropharyngeal cancer using 18F-FDG-PET-CT and/or diffusion-weighted MRI. *BMC Cancer* 2017; 17: 256. <https://doi.org/10.1186/s12885-017-3254-0>