Oral Oncology 137 (2023) 106307

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Magnetic resonance imaging based radiomics prediction of Human Papillomavirus infection status and overall survival in oropharyngeal squamous cell carcinoma

Paulien A. Boot [a], Steven W. Mes [b,c], Christiaan M. de Bloeme [a,b], Roland M. Martens [a,b], C. René Leemans [b,c], Ronald Boellaard [a,b], Mark A. van de Wiel [d,e], Pim de Graaf [a,b,*]

[a] Amsterdam UMC, Location Vrije Universiteit Amsterdam, Department of Radiology and Nuclear Medicine, De Boelelaan 1117, Amsterdam, the Netherlands
[b] Cancer Center Amsterdam, Imaging and Biomarkers, Amsterdam, the Netherlands
[c] Amsterdam UMC, Location Vrije Universiteit Amsterdam, Department of Otolaryngology – Head and Neck Surgery, De Boelelaan 1117, Amsterdam, the Netherlands
[d] Amsterdam UMC, Location Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, De Boelelaan 1117, Amsterdam, the Netherlands
[e] Amsterdam Public Health Research Institute, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Squamous cell carcinoma of head and neck
Oropharyngeal neoplasms
Magnetic resonance imaging
Human Papillomavirus
Biomarkers
Survival analysis
Factor analysis
Radiomics

ABSTRACT

Objectives: Human papillomavirus- (HPV) positive oropharyngeal squamous cell carcinoma (OPSCC) differs biologically and clinically from HPV-negative OPSCC and has a better prognosis. This study aims to analyze the value of magnetic resonance imaging (MRI)-based radiomics in predicting HPV status in OPSCC and aims to develop a prognostic model in OPSCC including HPV status and MRI-based radiomics.

Materials and methods: Manual delineation of 249 primary OPSCCs (91 HPV-positive and 159 HPV-negative) on pretreatment native T1-weighted MRIs was performed and used to extract 498 radiomic features per delineation. A logistic regression (LR) and random forest (RF) model were developed using univariate feature selection. Additionally, factor analysis was performed, and the derived factors were combined with clinical data in a predictive model to assess the performance on predicting HPV status. Additionally, factors were combined with clinical parameters in a multivariable survival regression analysis.

Results: Both feature-based LR and RF models performed with an AUC of 0.79 in prediction of HPV status. Fourteen of the twenty most significant features were similar in both models, mainly concerning tumor sphericity, intensity variation, compactness, and tumor diameter. The model combining clinical data and radiomic factors (AUC = 0.89) outperformed the radiomics-only model in predicting OPSCC HPV status. Overall survival prediction was most accurate using the combination of clinical parameters and radiomic factors (C-index = 0.72).

Conclusion: Predictive models based on MR-radiomic features were able to predict HPV status with sufficient performance, supporting the role of MRI-based radiomics as potential imaging biomarker. Survival prediction improved by combining clinical features with MRI-based radiomics.

Abbreviations: AABB, Axis aligned bounding box; AEE, Approximate Enclosing Ellipsoid; AUC, Area under the receiver-operator characteristics curve; C-index, Concordance index; DWI, Diffusion weighted imaging; FMradio, Factor modeling for radiomic data; GLDZM, Grey level distance zone matrix; HPV, Human papillomavirus; LR, Logistic regression; IBSI, Image biomarker standardization initiative; IHC, Immunohistochemistry; MRI, Magnetic resonance imaging; NOS, Not otherwise specified; OPSCC, Oropharyngeal squamous cell carcinoma; PCR, Polymerase chain reaction; RaCat, Radiomics Calculator; RF, Random Forest; RFE, Recursive feature elimination; ROI, Region of interest; SD, Standard deviation; SE, Standard error; SHAP, Shapley additive explanations; STIR, Short tau inversion recovery; T1W, T1-weighted; TORS, Transoral robotic surgery.

* Corresponding author at: Amsterdam UMC, Location Vrije Universiteit Amsterdam, Department of Radiology and Nuclear Medicine, De Boelelaan 1117, Amsterdam, the Netherlands.
E-mail address: p.degraaf@amsterdamumc.nl (P. de Graaf).

https://doi.org/10.1016/j.oraloncology.2023.106307
Received 22 October 2022; Received in revised form 28 December 2022; Accepted 8 January 2023
Available online 18 January 2023

Introduction

Head and neck cancer accounts for an estimated 4.9% of worldwide cancer incidence, making it the seventh most common cancer type [1]. Within this group, the incidence of oropharyngeal squamous cell carcinoma (OPSCC) has increased drastically by 57.3% in the United States from 1975 to 2014 [2], largely due to increased prevalence of human papillomavirus (HPV) related OPSCC. A distinction is made between HPV-positive and HPV-negative OPSCC because of their clinical and biological differences. HPV-positive OPSCC presents with smaller or even occult primary tumors, characterized by more well-defined borders, whereas HPV-negative tumors present with more advanced T-staging including increased invasion of adjacent structures [3]. Furthermore, HPV-positive OPSCC patients show better response to radio- and chemotherapy and have better overall survival, with 5-year survival rates of 80% versus 50% for HPV-negative patients [4–6]. Consequently, the American Joint Committee on Cancer classifies OPSCC into HPV-positive and HPV-negative tumors in their newest staging manual (eight edition) [7], hence the importance of accurate and reliable HPV status determination in OPSCC. The most widely applied HPV-detection method in OPSCC is p16 immunohistochemistry (IHC) [8]. However, approximately 17% of OPSCCs are p16-positive,

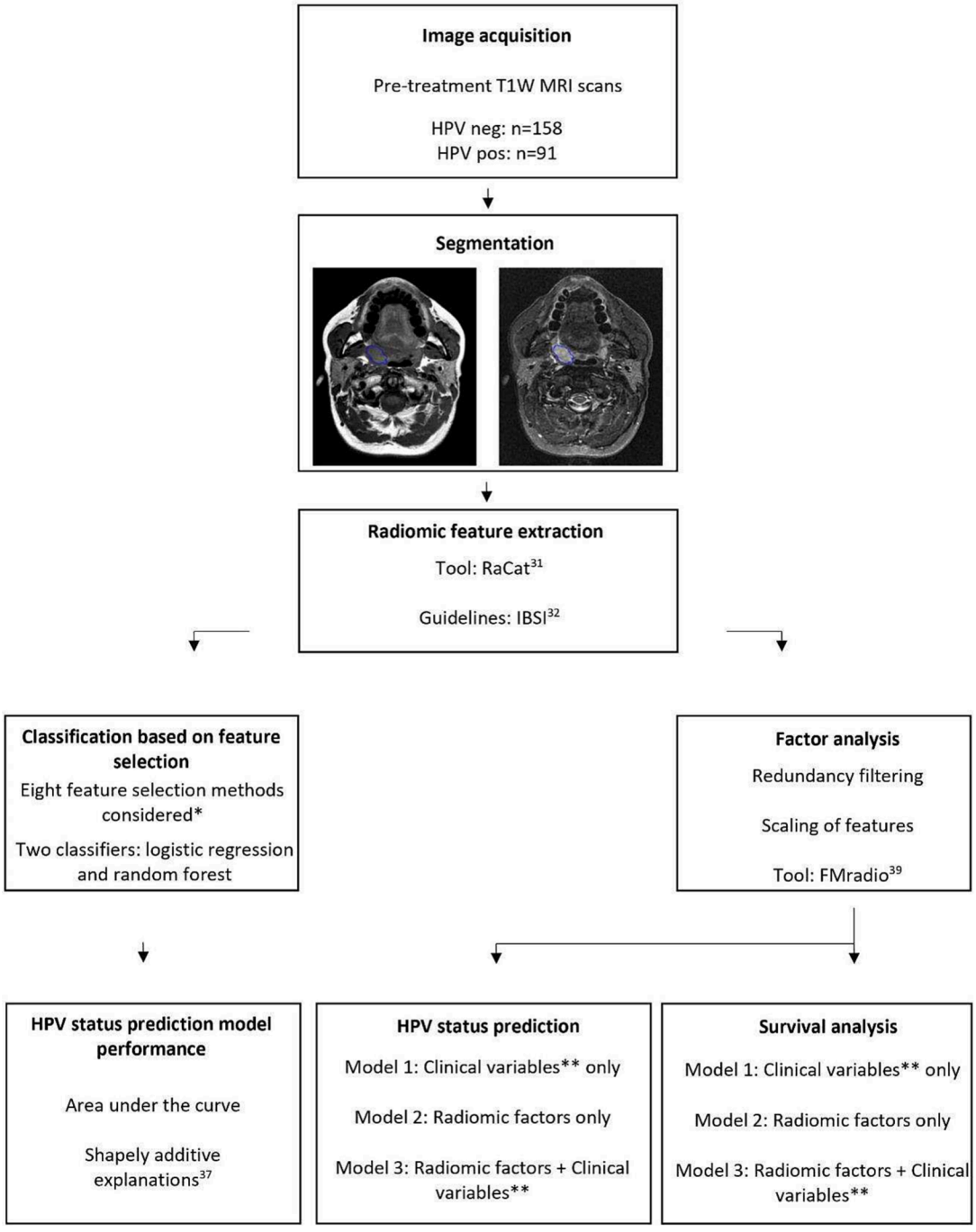

Fig. 1. Flowchart of study. *The methods considered were the firefly algorithm, lasso method, principal component analysis, recursive feature elimination (RFE)-logistic regression, RFE-random forest, RFE-support vector machine, univariate feature selection and no selection method **Clinical variables were tumor site, T-stadium, N-stadium, gender, age, and smoking status. Abbreviations: T1W: T1-weighted, MRI: Magnetic resonance imaging, HPV: Human Papillomavirus., RaCat: Radiomics Calculator, IBSI: Imaging biomarker standardization guidelin, FMradio: Factor modeling for radiomic data.

though HPV polymerase chain reaction (PCR) negative [9]. This group has similar inferior survival-rates as p16-negative OPSCC patients, hence the importance of identifying this specific group [10]. It is therefore advised to perform subsequent HPV PCR analysis when p16 IHC is positive [11].

A relatively new and promising tool in complementing HPV status prediction in OPSCC is radiomics-based phenotyping. Radiomics entails extracting an extensive number of quantitative features from medical imaging [12]. These features describe characteristics such as tumor signal intensity, shape and texture patterns [13]. Subsequently, the correlation of these quantitative features with clinical characteristics and their predictive and prognostic value can be assessed [14]. Exploring magnetic resonance imaging (MRI)-based radiomics of OPSCC might bring unidentified radiological, and thereby biological differences to light between HPV-positive and HPV-negative tumors. Moreover, radiomics might additionally reduce time-consuming and costly PCR analysis and provide for a non-invasive way of phenotyping OPSCC.

Most radiomics studies on OPSCC HPV status make use of computed tomography (CT) imaging [15–18], presumably due to the relative ease of data extraction and interpretation as well as standardization across scanners and vendors [19]. These CT-based radiomics studies on OPSCC show promising results regarding OPSCC prognosis and phenotyping. However, in many institutions MRI is the modality of choice for imaging OPSCC, due to detailed soft tissue contrast and ability to identify physical properties of tumors by application of separate sequence acquisition protocols [20]. Currently, only few studies have shown the use of MRI-based radiomics in predicting HPV status [21–23]. Moreover, numerous studies have shown MRI-based radiomic features can add information to clinical prognostic models of HPV-negative head and neck squamous cell carcinoma (HNSCC) or in heterogeneous cohorts [24–26]. However, the use of radiomics in prognostic models with HPV-positive tumors is less known. Most previous studies have additionally failed to include HPV PCR results in their research, ergo in this present study data including PCR status is available.

The primary aim of this study is therefore to develop an MRI-based radiomics signature of primary OPSCC to predict HPV status. The secondary aim is to develop an MRI-based radiomic prognostic model in OPSCC including HPV status and radiomics.

Materials and methods

Study population

Fig. 1 illustrates the study process. Three cohorts of OPSCC patients, treated in Amsterdam UMC, were acquired: 1) retrospective cohort of OPSCC patients treated from 2008 to 2012 [25], including HPV-positive OPSCC, initially excluded from the earlier study; 2) retrospective dataset of HNSCC patients treated from 2012 to 2016 [27]; 3) prospective dataset of OPSCC patients treated from 2014 to 2018 [28]. For all datasets IRB approval was obtained and the inclusion criteria were histologically proven OPSCC, available pre-treatment T1-weighted (T1W) MR imaging and treatment with curative intent. The exclusion criteria were inadequate MRI quality due to artifacts, previous locoregional treatment for head- and neck carcinoma and age under 18. For all patients, a validated algorithm for HPV detection was used: p16-IHC followed by a GP 5+/6+ PCR on the p16-IHC positive cases (sensitivity: 96%, specificity: 98%). [29]. Final TNM classification of the primary OPSCC was decided during the multidisciplinary team meetings at our institution, following the 7th edition of TNM classification [42].

MRI and segmentation

The datasets contained whole-lesion segmented - primary tumor only – OPSCCs, delineated on T1W MR imaging according to the previously published method [25]. All segmentations were completed manually by using the software program VelocityAI 3.1 (Varian Medical Systems, Inc. Palo Alto, USA). The newly included HPV-positive tumors, initially excluded from the retrospective dataset with data from 2008 to 2012 [25], were additionally delineated for this study by one observer SWM (five years of experience), supervised by PdG (fifteen years of experience). Fig. 2 shows a manual delineation example. Delineation was done on axial T1W imaging, with axial short tau inversion recovery (STIR), axial post-contrast T1W scans and diffusion weighted imaging (DWI) used for reference, when available. For all patients, endoscopy reports were available and used to facilitate manual delineation.

Radiomic feature extraction

The DICOM and RTstruct files exported from Velocity were converted into Nifti files using spyder 3.6.3 [30].

Previous to feature extraction, the MR images were resampled to 2x2x2 mm isotropic voxels and discretized using a fixed bin of 64 [31,32]. Feature extraction and subsequent feature calculations were performed using in-house build RaCat software [30], following the Image Biomarker Standardization Initiative (IBSI) [31,32] guidelines.

Statistical analysis

The difference of parameters between HPV subgroups was analyzed performing the independent samples T-test (i.e., age, voxel size); Chi-square test (i.e., gender) and Fischer's exact test (i.e., smoking, T-stage, N-stage, vendors, magnetic field strength, slice thickness). The difference of parameters between the three cohorts was analyzed using ANOVA testing (i.e., age, voxel size); Fischer's exact test (i.e., gender) and the Chi-square test (i.e., smoking, T-stage, N-stage, vendors, magnetic field strength and slice thickness). P values of <0.05 were considered statistically significant.

Classification based on feature selection

Hereafter, a machine learning framework in Python 3.6 was used to find the best fitting model [33]. Feature selection was applied to potentially improve generalizability and interpretability. We present results of eight feature selection methods (Table 2), as well as results of two classifiers: logistic regression (LR) and random forest (RF) classifier, the former being an standard benchmark, the latter a more flexible non-parametric algorithm able to outperform the former occasionally [34]. In absence of a separate, independent validation set, an approach with double cross validation is preferred over using a single random sample of the dataset for validation [35]. Therefore, this method was chosen to select the best-performing model in this study. For all 16 combinations, our model building framework is a standard, double cross-validation (CV) approach: the inner CV-loop is used to optimize the model in terms of selected features and hyper-parameter optimization, the outer CV-loop is used to test model performance. For both loops AUC (as evaluated on the left-out samples in the CV) is used as the evaluation criterion. We used a 5-fold CV for the inner CV-loop, and balanced, repeated 10-fold CV for the outer loop, using 5 repeats. Repeats were used to limit chance findings.

No oversampling was applied since the ratio of HPV-negative and HPV-positive patients was balanced. Features were scaled using z-score normalization, since the features follow a normal distribution, and standardization maintains useful information concerning possible outliers and makes the data less sensitive to them [36]. To evaluate model performance, a receiver operator characteristics curve was generated, and the AUC was calculated. The Brier score was used to assess model calibration and refinement (0.0 being optimal) [37]. Finally, twenty features with highest Shapley additive explanations [38] (SHAP) values were calculated. SHAP assigns each feature an importance value for a particular prediction, helping to interpret the outcome [38].

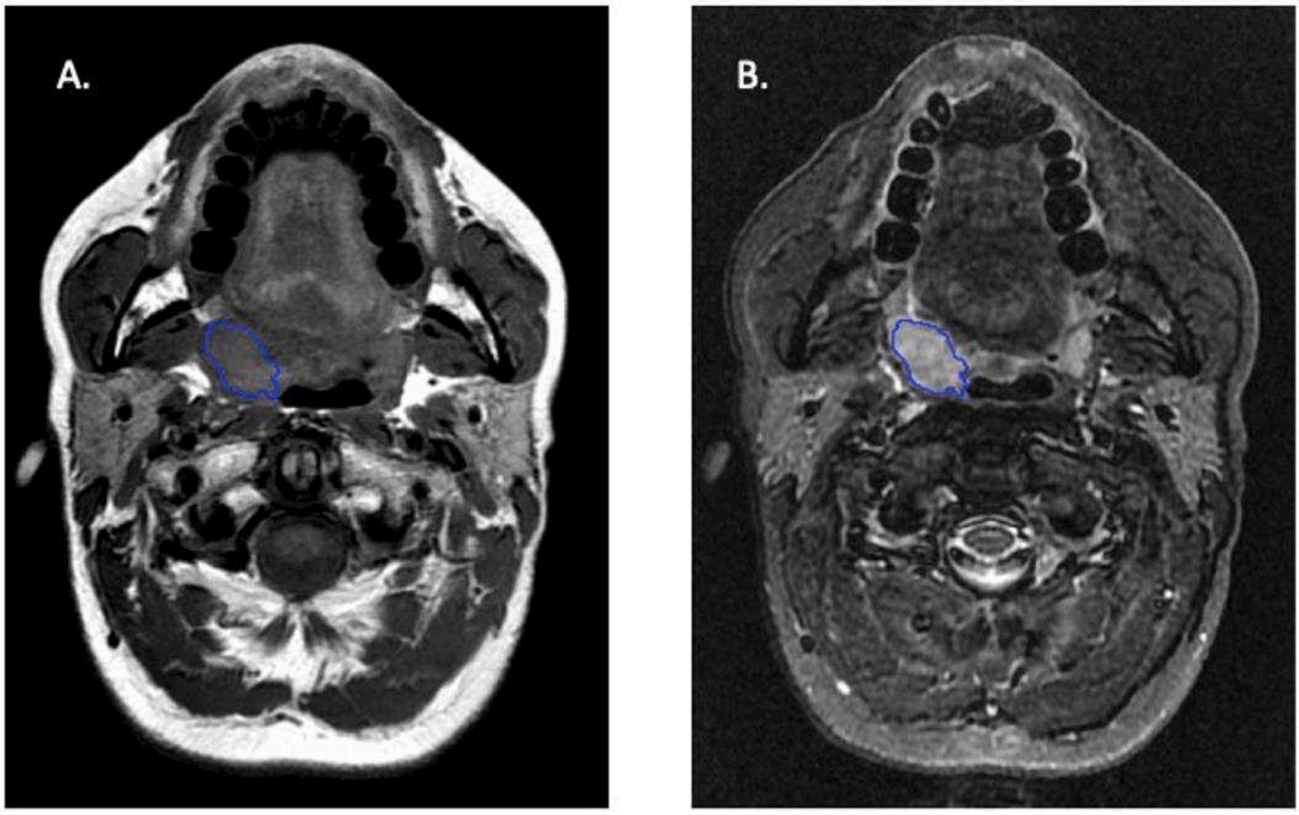

Fig. 2. Illustration of manual segmentation of a T2N0 right sided tonsillar tumor on T1W MRI (A) and STIR (B).

Factor analysis

Firstly, redundancy filtering was performed on the extracted features by removing the minimal number of features above a correlation threshold, which was set at 0.95. Subsequently, features were scaled (centered around 0 and variance 1) to prevent the possibility of features with the largest scale dominating the analysis. A regularized estimator of the correlation matrix between the scaled features was obtained, and a maximum likelihood factor analysis was performed on the matrix. Number of latent features was determined using the Kaiser-Guttman rule [39] on the regularized correlation matrix. Factor scores were obtained by regressing latent features on the observed data by way of the obtained factor solution. These steps were performed by using the R package "FMradio: Factor modeling for radiomic data", version 1 [40]. The proportion of variance explained by all factors must be appreciable, thus the threshold was set on 80%.

HPV status prediction

To assess the value of radiomics versus clinical variables in predicting HPV status, the factors were used in LR models. First, the constructed factors were used as predictors for a model with only radiomic factors as predictors and a model combining radiomic factors and clinical data. Additionally, a model was built with clinical data only. For the clinical model and the model combining clinical data and radiomic factors, only uniformly available clinical variables were used, for which some predictive power was expected. These variables were age at diagnosis, T-stage, N-stage, gender, and smoking status. To assess performance of the models, LR analyses were used to determine AUC's.

Survival analysis

Hereafter, the constructed factors were used to build a prognostic model for overall survival. Firstly, the prognostic performance of clinical parameters and radiomic factors was assessed separately by performing a cox regression analysis. Thereafter, clinical parameters and radiomic factors were combined in a multivariable regression analysis as well as an RF analysis. Predictive performance of the models was assessed by using the concordance index (C-Index). A risk stratification for overall survival was done using Kaplan Meier survival curves, which was divided into high (≥66%), medium (≥33–66%), and low risk (<33%). Analyses concerning prognostic modelling were performed with R.

Results

Patient characteristics

In total, 249 OPSCC patients were included, of which 91 HPV-positive- and 158 HPV-negative OPSCC. Table 1 shows results of p16 IHC and HPV PCR testing. Notably, eleven patients had positive p16 tests, but negative HPV PCR results, who were included in the HPV-negative group. Furthermore, patient- and scanner characteristics are reported per HPV status subgroup (Table 1) and per cohort (Supplemental Table 1). Age did not significantly differ between HPV subgroups, with a mean of 61 years ($p = 0.793$). In the HPV-negative group patients were more likely to be male ($p = 0.007$), and active smokers ($p < 0.0001$). For the HPV-positive group, tumors were almost exclusively located in tonsils and base of the tongue (48.4% and 50.5%, respectively). Most HPV-negative tumors were located in tonsils and base of the tongue as well (38.0% and 36.1%, respectively), yet 25.9% of tumors were located elsewhere in the oropharynx. For both tumor and nodal stage significant differences were found between HPV subgroups (both $p < 0.0001$). HPV-negative tumors presented with higher tumor stage: 60.1% of tumors T3 and T4, versus 34.1% for HPV-positive tumors. HPV-negative patients more frequently presented with N0 stage than HPV-positive patients (31.6% and 5.5%, respectively). For scanner

Table 1
Patient- and scanner characteristics, per HPV subgroup.

Patients	Total p = 249 (%)	HPV-positive p = 91 (%)	HPV-negative p = 158 (%)	P value [a]
P16 IHC/PCR results				
P16 IHC-	147 (59)			
P16 IHC+/PCR-	11 (4)			
P16 IHC+/PCR+	91 (37)			
Age [mean in years] (SD)	61 (8.2)	61 (8.4)	61 (8.1)	.793[c]
Gender				.007[d]
Female	78 (31.3)	19 (20.8)	59 (37.3)	
Male	171 (68.7)	72 (79.1)	99 (62.7)	
Smoking				.000[e]
Never smoked	48 (19)	33 (36)	15 (10)	
Stopped smoking	68 (27)	32 (35)	36 (23)	
Active smoker	116 (48)	26 (29)	107 (68)	
Tumor stage[b]				.000[e]
T1	32 (13)	20 (22)	12 (8)	
T2	91 (37)	40 (44)	51 (32)	
T3	47 (19)	14 (15)	33 (21)	
T4	79 (32)	17 (19)	62 (39)	
Nodal stage [b]				.000[e]
N0	55 (22)	5 (6)	50 (32)	
N1	45 (18)	14 (15)	31 (20)	
N2	129 (52)	61 (67)	68 (43)	
N3	18 (7)	10 (11)	8 (5)	
Tumor site				.000[e]
Tonsil	104 (42)	44 (48)	60 (38)	
Base of tongue	103 (41)	46 (51)	57 (36)	
Soft palate/Uvula	22 (9)	0 (0)	22 (14)	
Oropharynx NOS	20 (8)	1 (1)	19 (12)	
Vendors				.581[e]
GE	163 (66)	56 (62)	107 (68)	
Philips	55 (22)	21 (23)	34 (22)	
Siemens	31 (12)	14 (15)	17 (11)	
Magnetic field strength				.619[e]
1.0T	1 (0.4)	0 (0)	1 (1)	
1.5T	169 (68)	59 (65)	110 (70)	
3.0T	79 (32)	32 (35)	47 (30)	
Slice thickness				.001[e]
3 mm	53 (21)	15 (17)	38 (25)	
4 mm	171 (69)	58 (64)	113 (72)	
5 mm	25 (10)	18 (20)	7 (4)	
Voxel size X*Y*Z [mean] (SD)	1,051 (0.4)	1,126 (0.5)	1,007 (0.4)	.172[c]

[a] P value of differences between HPV-positive and HPV-negative group.
[b] TNM classification, 7th edition [43].
[c] Independent samples T-test.
[d] Chi square test.
[e] Fisher's exact test.
Abbreviations: IHC = Immunohistochemistry; PCR = Polymerase Chain Reaction; NOS = Not otherwise specified.

characteristics (i.e., vendors, magnetic field strength and voxel size) no significant differences were found (p = 0.581, p = 0.691, and p = 0.172, respectively).

HPV status prediction model building based on feature selection

Per tumor segmentation 498 features were extracted, describing intensity, morphology, and texture (Supplemental Table 2) [32]. Of the eight feature selection methods considered, univariate feature qualification showed best performance (Table 2). Univariate feature qualification rendered 49 features (Supplemental Table 3), which were subsequently used for model building. LR showed an AUC of 0.79 (standard deviation (SD) 0.10) and RF showed an equal AUC of 0.79 (SD 0.09) (Fig. 3). The Brier score of both models was 0.18 (SD 0.04). The twenty features with highest SHAP values for both analyses are shown in Fig. 4. Fourteen of the twenty most influential features were similar in both models, mainly concerning sphericity (e.g., spherical disproportion), intensity variation (e.g., coefficient of variation), compactness (e.g., compactness1) and diameter (e.g., minor- and major axis length) of the tumor. The SHAP plots show the values of features concerning compactness, volume density, flatness and sphericity to be greater for HPV-positive tumors. The values of features concerning intensity variation, non-uniformity, and diameter are greater for HPV-negative tumors.

Table 2
Performance of the eight considered feature selection methods and two classifiers.

Feature selection method	AUC Logistic Regression (standard deviation)	AUC Random Forrest (standard deviation)
Firefly Algorithm (FA)	0.62 (0.12)	0.65 (0.11)
Lasso	0.79 (0.10)	0.77 (0.10)
Principal Component Analysis (PCA)	0.78 (0.09)	0.74 (0.08)
RFE-Logistic Regression (RFE-LR)	0.78 (0.10)	0.79 (0.09)
RFE-Random Forrest (RFE-RF)	0.77 (0.10)	0.80 (0.09)
RFE-Support Vector Machine (RFE-SVM)	0.69 (0.10)	0.72 (0.10)
Univariate	0.79 (0.09)	0.79 (0.10)
None	0.77 (0.11)	0.77 (0.10)

Area Under the Curve scores for all eight considered feature selection methods, and for the two considered classifiers, namely logistic regression and random forest. Abbreviations: AUC = Area under the curve; RFE = Recursive Feature Elimination.

HPV status prediction model building based on factor analysis

For the factor analysis, redundancy filtering was applied on the 498 features to remove highly correlated features, which resulted in 100 features. With these features, ten factors were created, accounting for 82% of variation in the data. The exact content of each factor is presented in Supplemental Table 4. The ten factors were used to train a model with LR to predict OPSCC HPV status. Additionally, a model with only clinical variables was built. Both the models with radiomic factors only and clinical variables only performed with an AUC of 0.80. The combined model with clinical and radiomic data shows an AUC of 0.87. Significant covariates in the combined model were T-stage, N-stage, smoking status, gender, factors 1, 3, 5 and 6.

Overall survival

Finally, the ten radiomics factors were used to develop a model to predict overall survival of OPSCC patients. Since the number of events was small in the HPV-positive group, the decision was made to not split the group when fitting the survival models. HPV status was therefore employed as covariate in the model. For the multivariate Cox regression model, significant predictors for overall survival in the factors only model were factors 1, 3 and 6 and performed with a C-Index of 0.674 (SE = 0.03). When only clinical parameters were assessed, significant predictors were gender, smoking status, age, and HPV status, which performed with a C-Index of 0.679 (SE = 0.027). Significant predictors for overall survival were gender, HPV status and factors 1, 3, 6 and 7 when combining clinical parameters with radiomic factors, and this model was most predictive (C-Index = 0.72, SE = 0.026). Additionally, a random survival forest was explored, however this model performed less than the Cox analysis, with a C-Index of 0.65. The risk stratification into high, medium, and low risk for was constructed for overall survival (time to death): clinical parameters only (p < 0.0001), factors only (p = 0.0001) and combined model (p < 0.0001) (Fig. 5).

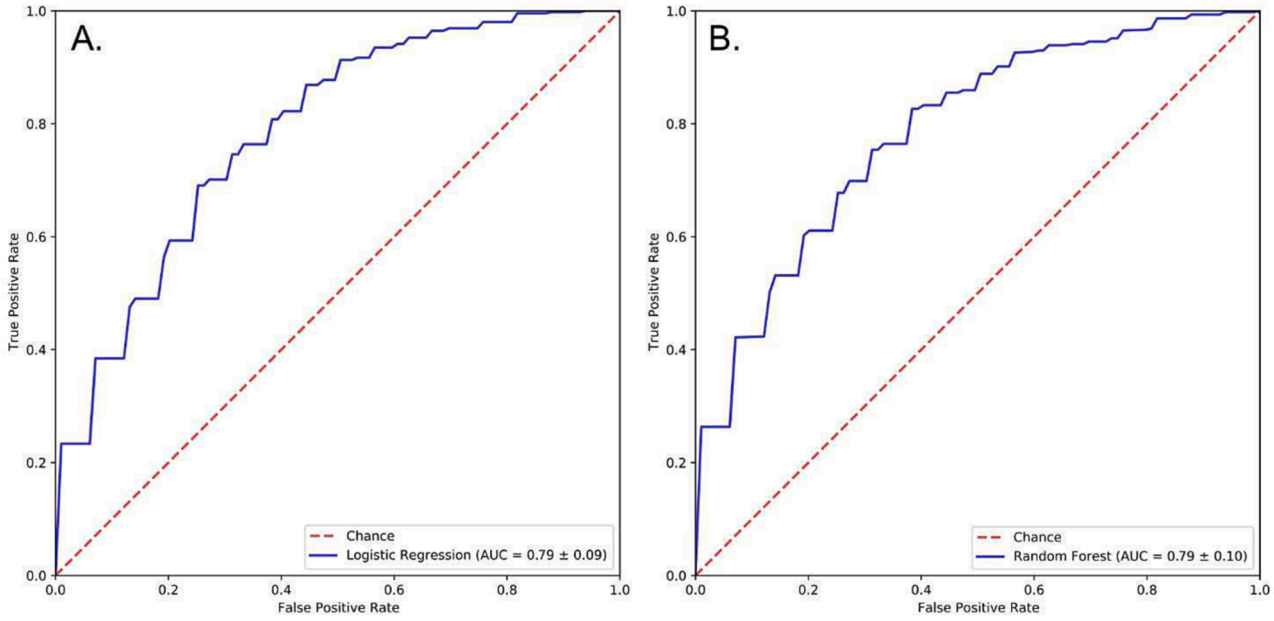

Fig. 3. ROC curves of HPV prediction by using radiomic features. After univariate feature qualification 49 features remained, which were used for model building through logistic regression and random forest analysis, the Fig. shows their performance. Logistic regression performed with an area under the curve of 0.79 ± 0.09 (A) and random forest performed with an area under the curve of 0.79 ± 0.10 (B).

Discussion

In the present study quantitative imaging features were extracted from T1W MR images of OPSCC patients and were firstly used to develop radiomics models to predict HPV infection status, with sufficient performance. Both the LR model and RF model based on univariate feature selection had similar performance (AUC = 0.79). Moreover, fourteen out of twenty most impactful features in both these models were similar, mostly concerning compactness, axis length, sphericity, and intensity variation. Additionally, a model was developed based on ten factors, with good predictive performance (AUC = 0.80), a model based on clinical data performed equally well (AUC = 0.80), and combining the factor model and clinical model showed even better performance (AUC = 0.89). Ideally, radiomics could be employed as a non-invasive test, however the increase of performance when combining the models was small and insufficiently accurate to replace p16/HPV PCR testing at this stage. Potentially, inclusion of additional MRI sequences could improve model performance.

The prognostic value of the developed radiomic factors was additionally assessed. Since the small number of events in the HPV-positive group, analyzing the prognostic value of the factors for HPV-positive patients separately was not worthwhile. However, the multivariable analysis did show HPV status analysis to be of importance as covariate in the analysis. For future research, a greater cohort of HPV-positive patients is necessary to determine if MRI-based radiomics can be of prognostic meaning in overall survival analysis of HPV-positive tumors.

The extracted MRI features reflect some biological differences between HPV-positive and HPV-negative OPSCC, demonstrating the feasibility of radiomics in HPV classification. The features compactness1, compactness2 and sphericity quantify the deviation of the region of interest (ROI) volume from a representative spheroid. This is congruent with HPV-positive OPSCC known to be sphere-like, meaning these carcinomas tend to have an abrupt transition between the tumor and the adjacent surface epithelium [3,41]. Furthermore, the features major- and minor axis length represent how far the volume extends along the largest and second largest axis. These axis features show to be greater for HPV-negative tumors, which is expected since HPV-negative OPSCC tend to present with higher T stages than HPV-positive OPSCC [42]. In the factor model, factor 5 was most influential and was partly based on the features volume density and axis length, which were also influential in both feature selection-based models.

Additionally, several less expected features a priori were of importance in the feature-based models. Firstly, the features 'coefficient of variation' and 'quartile coefficient' both are measures of dispersion of intensity distribution of the ROI. This dispersion is calculated by dividing the standard deviation by the mean of the intensity distribution. It seems HPV-negative OPSCC have greater dispersion of intensity distribution, which might translate to higher intensity heterogeneity. Secondly, in the LR model, the RF model and the factor-based model 'GLDZM zone distance non uniformity' features are of influence. The GLDZM counts the zones of linked voxels which share a specific grey level and possess the same distance to the ROI edge, capturing the relation between location and grey level. These features might likewise translate to greater heterogeneity in HPV-negative tumors. However, a relevant limitation becomes known since these extracted features are of mathematical nature. Consequently, clinical interpretability is limited and is subject to further research, e.g., to extensively explore correlation of extracted features and biological properties of OPSCC, for instance by correlation to genetics or histopathology. As follows, radiomic models might be of value in precision medicine since MRI-based radiomics seems to bring unseen variation of biological properties of carcinomas to light.

It must be acknowledged that some studies on MRI-based radiomics in OPSCC have been done previously. Sohn et al. show six radiomic features to have strong association with OPSCC HPV status, including the feature flatness, similar to this study, and their models showed similar performance with an AUC of 0.75. However, a limitation was their small and imbalanced sample size, with 10 HPV-negative and 42 HPV-positive cases [22]. Suh et al. extracted features from multiparametric MRI sequences in OPSCC and were successful in developing a radiomic signature of HPV status, with an AUC of 0.77 and 0.76 for their LR and RF model, respectively [21]. Nonetheless, their sample size was relatively small as well (n = 60). Additionally, Bos et al. proved their model based on contrast enhanced T1W MRI radiomic features can predict HPV status in OPSCC, with an AUC of 0.76 for their test set, with a more extensive database (n = 153) [23]. In their model the feature sphericity had significant impact, similar to this study. Besides the issue of relatively small datasets, previous studies fail to include HPV PCR

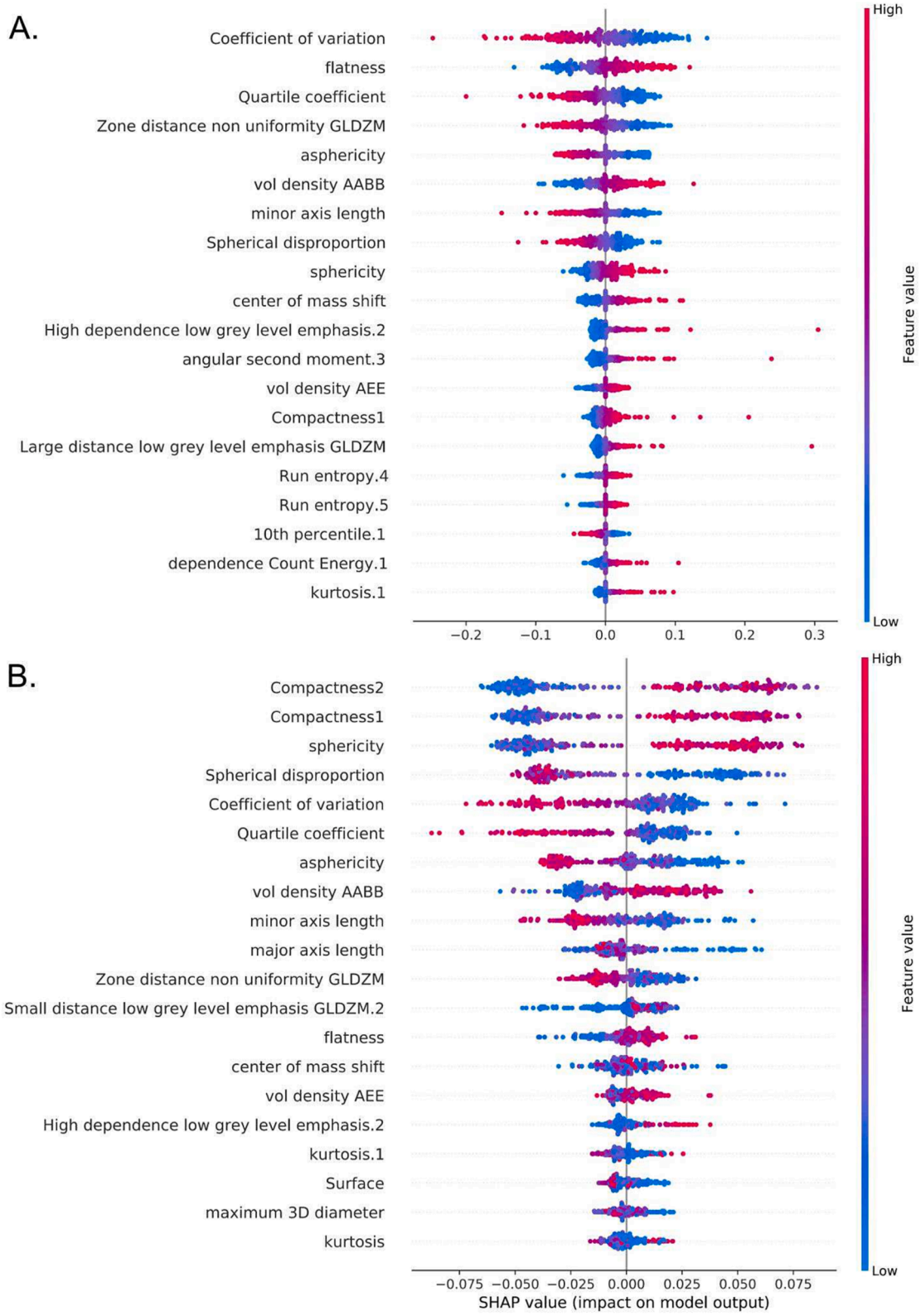

Fig. 4. Twenty features with highest SHAP value of logistic regression (A) and random forest (B). Each dot on the plot is a Shapley value for one ROI, per feature. The horizontal location (the Shapley value) expresses the effect that the observed level of a feature for a tumor has on the final predicted probability. Negative SHAP value (to the left) means negative impact, leading the model to predict 0 (i.e. HPV-negative). Positive SHAP value means positive impact on prediction, leading the model to predict 1 (i.e. HPV-positive). The color of the dots shows whether the value of the feature is high (in pink) or low (in blue) for that observation. The features are ranked in descending order showing the feature importance. Abbreviations: AABB: axis aligned bounding box; GLDZM: grey level distance zone matrix; AEE: approximate enclosing ellipsoid. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

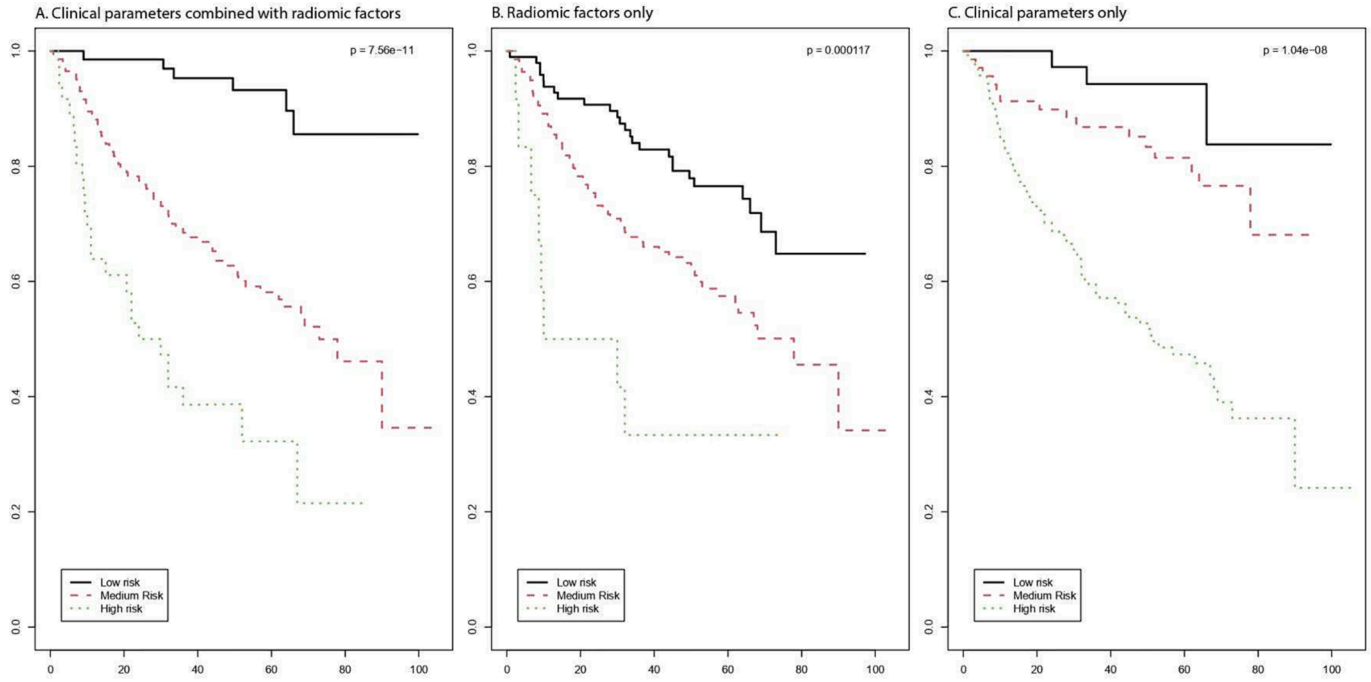

Fig. 5. Kaplan Meier overall survival curves for clinical parameters combined with radiomic factors ($p < 0.0001$) (A), for radiomic factors only ($p = 0.0001$) (B) and for clinical parameters only ($p < 0.0001$) (C). The significant predictors in the combined model were gender, HPV status and factors 1, 3, 6 and 7. The significant predictors in de radiomic factors only model were factors 1, 3 and 6. The significant predictors in the clinical model only were gender, smoking status, age, and HPV status.

results in their research. While in this study, not p16 IHC but the combination of p16 IHC and HPV PCR was leading in determining HPV status in OPSCC. Therefore, eleven patients with positive p16 results, but negative PCR results were now classified as OPSCC HPV-negative. An implication for further research is to investigate whether this specific group differs in terms of MRI-based radiomics.

Performance of previously conducted studies is similar to performance of the models in this study and some similar features are used. However, comparing different models must be done with caution since feature selection methods, imaging sequence used for feature extraction, choice of machine learning classifiers, and study population differs across studies. This highlights a critical issue in this field, namely the generalizability of the models. MRI-based radiomic studies in HNSCC is known to lack study design standardization, which can limit clinical relevance [19]. The IBSI aimed to address this issue by standardization of radiomics software and defining reporting guidelines, considered in the present study. Furthermore, although the models developed in this study are built on a relative extensive dataset (n = 249), with imaging obtained from 2008 to 2018 from different vendors, no validation in an external dataset was performed. A recommendation for further research is validating the models in an external dataset or developing a model on a multicenter dataset. Another possible limitation is variability in data due to inconsistencies in tumor delineation, but this has been assessed before. Mes et al. assessed a random subgroup of 30 OPSCCs when segmentation was done by two independent radiologists, and no significant differences in radiomic features were found [25]. Moreover, Martens et al. likewise found no significant different values and high interobserver correlation when two independent observers delineated OPSCCs [28]. Furthermore, many features in the models reflect the same properties of the ROIs, e.g., sphericity, asphericity, compactness1 and compactness2 all reflect how the ROI is sphere-like. Hence, for future analysis it might be interesting to perform multivariate feature selection analyses, to explore influence of separate features more extensively, or in multiple combinations. Finally, for further research it is advised to include data on pack years in prognostic models of OPSCC. Although we acknowledge the important impact of pack years on outcome in OPSCC, unfortunately, this data was only available for approximately half of the included patients in the current study.

Concluding, since the incidence of HPV-positive OPSCC is rising and clinical differences between HPV-positive and HPV-negative OPSCC are apparent, the importance of accurate discrimination between the two is evident. The radiomics models developed in the present study show sufficient performance for classification of HPV status in OPSCC, demonstrating the ability of radiomics in differentiating HPV status. Additionally, the specific extracted features of influence in the radiomics models provide use with more insight in radiological OPSCC properties. One might envision MRI-based radiomics to become a non-invasive, efficient, and possibly cost-reducing tool in phenotyping OPSCC, and tumors of other origin. Finally, the regression model concerning overall survival combining common clinical parameters and radiomics was more predictive than the models with clinical parameters or radiomics alone, hence implying additional value of incorporating radiomics in OPSCC prognosis.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, nor profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.oraloncology.2023.106307.

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 Nov 2018;68(6):394–424. https://doi.org/10.3322/caac.21492.

[2] Osazuwa-Peters N, Simpson MC, Massa ST, Adjei Boakye E, Antisdel JL, Varvares MA. 40-year incidence trends for oropharyngeal squamous cell carcinoma in the United States. Oral Oncol Nov 2017;74:90–7. https://doi.org/10.1016/j.oraloncology.2017.09.015.

[3] Cantrell SC, Peck BW, Li G, Wei Q, Sturgis EM, Ginsberg LE. Differences in imaging characteristics of HPV-positive and HPV-Negative oropharyngeal cancers: a blinded matched-pair analysis. AJNR Am J Neuroradiol Oct 2013;34(10):2005–9. https://doi.org/10.3174/ajnr.A3524.

[4] Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. J Natl Cancer Inst 2008;100(4):261–9. https://doi.org/10.1093/jnci/djn011.

[5] Lassen P, Eriksen JG, Hamilton-Dutoit S, Tramm T, Alsner J, Overgaard J. HPV-associated p16-expression and response to hypoxic modification of radiotherapy in head and neck cancer. Radiotherapy Oncol: J Eur Soc Therapeutic Radiol Oncol Jan 2010;94(1):30–5. https://doi.org/10.1016/j.radonc.2009.10.008.

[6] Ang KK, Harris J, Wheeler R, Weber R, Rosenthal DI, Nguyen-Tân PF, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010;363(1):24–35. https://doi.org/10.1056/NEJMoa0912217.

[7] Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA: Cancer J. Clin. Mar 2017;67(2):93–9. https://doi.org/10.3322/caac.21388.

[8] Jordan RC, Lingen MW, Perez-Ordonez B, He X, Pickard R, Koluder M, et al. Validation of methods for oropharyngeal cancer HPV status determination in US cooperative group trials. Am J Surg Pathol Jul 2012;36(7):945–54. https://doi.org/10.1097/PAS.0b013e318253a2d1.

[9] Prigge ES, Arbyn M, von Knebel DM, Reuschenbach M. Diagnostic accuracy of p16 (INK4a) immunohistochemistry in oropharyngeal squamous cell carcinomas: A systematic review and meta-analysis. Int J Cancer 2017;140(5):1186–98. https://doi.org/10.1002/ijc.30516.

[10] Rietbergen MM, Brakenhoff RH, Bloemena E, Witte BI, Snijders PJ, Heideman DA, et al. Human papillomavirus detection and comorbidity: critical issues in selection of patients with oropharyngeal cancer for treatment De-escalation trials. Ann Oncol Nov 2013;24(11):2740–5. https://doi.org/10.1093/annonc/mdt319.

[11] Pannone G., Rodolico V., Santoro A., Lo Muzio L., Franco R., Botti G., et al. Evaluation of a combined triple method to detect causative HPV in oral and oropharyngeal squamous cell carcinomas: p16 Immunohistochemistry, Consensus PCR HPV-DNA, and In Situ Hybridization. Infect Agent Cancer. Feb 29 2012;7:4. doi:10.1186/1750-9378-7-4.

[12] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol Dec 2017;14(12):749–62. https://doi.org/10.1038/nrclinonc.2017.141.

[13] Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. Magn Reson Imaging Nov 2012;30(9):1234–48. https://doi.org/10.1016/j.mri.2012.06.010.

[14] Giannitto C, Marvaso G, Botta F, Raimondi S, Alterio D, Ciardo D, et al. Association of quantitative MRI-based radiomic features with prognostic factors and recurrence rate in oropharyngeal squamous cell carcinoma. Neoplasma Nov 2020;67(6):1437–46. https://doi.org/10.4149/neo_2020_200310N249.

[15] Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G, et al. Computed Tomography Radiomics Predicts HPV Status and Local Tumor Control After Definitive Radiochemotherapy in Head and Neck Squamous Cell Carcinoma. Int J Radiat Oncol Biol Phys 2017;99(4):921–8. https://doi.org/10.1016/j.ijrobp.2017.06.002.

[16] Leijenaar RT, Bogowicz M, Jochems A, Hoebers FJ, Wesseling FW, Huang SH, et al. Development and validation of a radiomic signature to predict HPV (p16) status from standard CT imaging: a multicenter study. Br J Radiol Jun 2018;91(1086):20170498. https://doi.org/10.1259/bjr.20170498.

[17] Ou D, Blanchard P, Rosellini S, Levy A, Nguyen F, Leijenaar RTH, 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 Aug 2017;71:150–5. https://doi.org/10.1016/j.oraloncology.2017.06.015.

[18] Buch K, Fujita A, Li B, Kawashima Y, Qureshi MM, Sakai O. Using Texture Analysis to Determine Human Papillomavirus Status of Oropharyngeal Squamous Cell Carcinomas on CT. AJNR Am J Neuroradiol Jul 2015;36(7):1343–8. https://doi.org/10.3174/ajnr.A4285.

[19] Jethanandani A, Lin TA, Volpe S, Elhalawani H, Mohamed ASR, Yang P, 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.

[20] Leslie A, Fyfe E, Guest P, Goddard P, Kabala JE. Staging of squamous cell carcinoma of the oral cavity and oropharynx: a comparison of MRI and CT in T- and N-staging. J Comput Assist Tomogr 1999;23(1):43–9. https://doi.org/10.1097/00004728-199901000-00010.

[21] Suh CH, Lee KH, Choi YJ, Chung SR, Baek JH, Lee JH, et al. Oropharyngeal squamous cell carcinoma: radiomic machine-learning classifiers from multiparametric MR images for determination of HPV infection status. *Scientific reports*. Oct 16 2020;10(1):17525. doi:10.1038/s41598-020-74479-x.

[22] Sohn B, Choi YS, Ahn SS, Kim H, Han K, Lee SK, et al. Machine Learning Based Radiomic HPV Phenotyping of Oropharyngeal SCC: A Feasibility Study Using MRI. Laryngoscope Mar 2021;131(3):E851–6. https://doi.org/10.1002/lary.28889.

[23] Bos P, van den Brekel MWM, Gouw ZAR, Al-Mamgani A, Waktola S, Aerts H, et al. Clinical variables and magnetic resonance imaging-based radiomics predict human papillomavirus status of oropharyngeal cancer. Head Neck Feb 2021;43(2):485–95. https://doi.org/10.1002/hed.26505.

[24] Alfieri S, Romanò R, Bologna M, Calareso G, Corino V, Mirabile A, et al. Prognostic role of pre-treatment magnetic resonance imaging (MRI)-based radiomic analysis in effectively cured head and neck squamous cell carcinoma (HNSCC) patients. Acta Oncol Sep 2021;60(9):1192–200. https://doi.org/10.1080/0284186x.2021.1924401.

[25] Mes SW, van Velden FHP, Peltenburg B, Peeters CFW, Te Beest DE, van de Wiel MA, et al. Outcome prediction of head and neck squamous cell carcinoma by MRI radiomic signatures. Eur Radiol Nov 2020;30(11):6311–21. https://doi.org/10.1007/s00330-020-06962-y.

[26] 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 Aug 2019;117:193–8. https://doi.org/10.1016/j.ejrad.2019.06.019.

[27] Martens RM, Noij DP, Koopman T, Zwezerijnen B, Heymans M, de Jong MC, et al. Predictive value of quantitative diffusion-weighted imaging and 18-F-FDG-PET in head and neck squamous cell carcinoma treated by (chemo)radiotherapy. Eur J Radiol Apr 2019;113:39–50. https://doi.org/10.1016/j.ejrad.2019.01.031.

[28] Martens RM, Koopman T, Lavini C, Ali M, Peeters CFW, Noij DP, et al. Multiparametric functional MRI and (18)F-FDG-PET for survival prediction in patients with head and neck squamous cell carcinoma treated with (chemo) radiation. Eur Radiol Feb 2021;31(2):616–28. https://doi.org/10.1007/s00330-020-07163-3.

[29] Rietbergen MM, Leemans CR, Bloemena E, Heideman DA, Braakhuis BJ, Hesselink AT, et al. Increasing prevalence rates of HPV attributable oropharyngeal squamous cell carcinomas in the Netherlands as assessed by a validated test algorithm. Int J Cancer 2013;132(7):1565–71. https://doi.org/10.1002/ijc.27821.

[30] Phil T. Sikerdebaard/dcmrtstruct2nii: v1.0.19. Accessed 01012021, 2021. https://doi.org/10.5281/ZENODO.4037865.

[31] Pfaehler E, Zwanenburg A, de Jong JR, Boellaard R. RaCaT: An open source and easy to use radiomics calculator tool. PLoS One 2019;14(2):e0212223.

[32] Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology May 2020;295(2):328–38. https://doi.org/10.1148/radiol.2020191145.

[33] Cysouw MCF, Jansen BHE, van de Brug T, Oprea-Lager DE, Pfaehler E, de Vries BM, et al. Machine learning-based analysis of [(18)F]DCFPyL PET radiomics for risk stratification in primary prostate cancer. Eur J Nucl Med Mol Imaging Feb 2021;48(2):340–9. https://doi.org/10.1007/s00259-020-04971-z.

[34] Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinf 2018;19(1):270. https://doi.org/10.1186/s12859-018-2264-5.

[35] Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 2007;99(2):147–57. https://doi.org/10.1093/jnci/djk018.

[36] Raschka S. Python machine learning. 3 ed. Packt publishing ltd; 2015.

[37] Murphy AH. A new vector partition of the probability score. J Appl Meteorol Climatol 1973;12(4):595–600.

[38] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. 2017:4768-4777.

[39] Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas 1960;20(1):141–51.

[40] Peeters CF, Übelhör C, Mes SW, Martens R, Koopman T, de Graaf P, et al. Stable prediction with radiomics data. *arXiv preprint arXiv:190311696*. 2019.

[41] 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(1):48–54.

[42] Mellin H, Friesland S, Lewensohn R, Dalianis T, Munck-Wikland E. Human papillomavirus (HPV) DNA in tonsillar cancer: clinical correlates, risk of relapse, and survival. Int J Cancer 2000;89(3):300–4.

[43] Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Ann Surgical Oncol. Jun 2010;17(6):1471-4. doi:10.1245/s10434-010-0985-4.