Original Article | Neuroimaging and Head & Neck

eISSN 2005-8330
https://doi.org/10.3348/kjr.2022.0397
Korean J Radiol 2023;24(1):51-61

Korean Journal of Radiology
KJR

Development and Testing of a Machine Learning Model Using ^{18}F-Fluorodeoxyglucose PET/CT-Derived Metabolic Parameters to Classify Human Papillomavirus Status in Oropharyngeal Squamous Carcinoma

Changsoo Woo[1], Kwan Hyeong Jo[2], Beomseok Sohn[1], Kisung Park[1, 3], Hojin Cho[4], Won Jun Kang[4], Jinna Kim[1], Seung-Koo Lee[1]

[1]Department of Radiology and Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea; [2]Department of Nuclear Medicine, Korea University Guro Hospital, Seoul, Korea; [3]Department of Mechanical Engineering, Pohang University of Science and Technology, Pohang, Korea; [4]Department of Nuclear Medicine, Severance Hospital, Yonsei University College of Medicine, Seoul, Korea

Objective: To develop and test a machine learning model for classifying human papillomavirus (HPV) status of patients with oropharyngeal squamous cell carcinoma (OPSCC) using ^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET-derived parameters in derived parameters and an appropriate combination of machine learning methods in patients with OPSCC.

Materials and Methods: This retrospective study enrolled 126 patients (118 male; mean age, 60 years) with newly diagnosed, pathologically confirmed OPSCC, that underwent ^{18}F-FDG PET-computed tomography (CT) between January 2012 and February 2020. Patients were randomly assigned to training and internal validation sets in a 7:3 ratio. An external test set of 19 patients (16 male; mean age, 65.3 years) was recruited sequentially from two other tertiary hospitals. Model 1 used only PET parameters, Model 2 used only clinical features, and Model 3 used both PET and clinical parameters. Multiple feature transforms, feature selection, oversampling, and training models are all investigated. The external test set was used to test the three models that performed best in the internal validation set. The values for area under the receiver operating characteristic curve (AUC) were compared between models.

Results: In the external test set, ExtraTrees-based Model 3, which uses two PET-derived parameters and three clinical features, with a combination of MinMaxScaler, mutual information selection, and adaptive synthetic sampling approach, showed the best performance (AUC = 0.78; 95% confidence interval, 0.46–1). Model 3 outperformed Model 1 using PET parameters alone (AUC = 0.48, p = 0.047) and Model 2 using clinical parameters alone (AUC = 0.52, p = 0.142) in predicting HPV status.

Conclusion: Using oversampling and mutual information selection, an ExtraTree-based HPV status classifier was developed by combining metabolic parameters derived from ^{18}F-FDG PET/CT and clinical parameters in OPSCC, which exhibited higher performance than the models using either PET or clinical parameters alone.

Keywords: *Human papillomavirus; Machine learning; Oropharynx; Positron emission tomography; Squamous cell carcinoma*

Received: November 21, 2021 **Revised:** September 28, 2022 **Accepted:** October 31, 2022

Corresponding author: Beomseok Sohn, MD, Department of Radiology and Research Institute of Radiological Science and Center for Clinical Imaging Data Science, Severance Hospital, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.
• E-mail: beomseoksohn@yuhs.ac; and
Kwan Hyeong Jo, MD, Department of Nuclear Medicine, Korea University Guro Hospital, 148 Gurodong-ro, Guro-gu, Seoul 08308, Korea.
• E-mail: phe_ea@naver.com

INTRODUCTION

The incidence of oropharyngeal squamous cell carcinoma (OPSCC), including OPSCC associated with human papillomavirus (HPV) infection, has increased in recent decades [1,2]. Patients with HPV have a more favorable prognosis and longer disease-free- and overall survival than OPSCC patients without HPV infection [3,4]. Accordingly, the American Joint Committee on Cancer has adjusted its staging manuals and classified HPV-associated OPSCC as a distinct entity [5]. HPV status in patients with oropharyngeal cancer is determined by histopathological, immunohistochemical, and molecular diagnoses [4,6]. However, these diagnoses are only feasible through biopsy, performed using laryngoscopes, or during surgical excision [7]. Predicting HPV status preoperatively based on imaging may aid in patient counseling and planning treatment strategies [8]. Additionally, imaging characteristics may provide further insights on the underlying pathophysiology of the distinct clinical course of OPSCC. Furthermore, if the number of specimens is not sufficient for diagnosis or the patient cannot undergo surgery, the HPV infection status of patients with OPSCC cannot be determined. Therefore, other non-invasive diagnostic methods may be helpful in treatment planning.

Many efforts have been made to predict HPV status in patients with OPSCC using machine-learning approaches based on imaging studies [9]. For example, radiomics-based investigations have used structural MRI, with and without diffusion-weighted imaging [10-12]. Few studies have used radiomics from ^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET to determine the HPV status [13]. However, the radiomics approach is vulnerable to issues related to reproducibility and is time-consuming because of the multiple preprocessing steps and feature extraction processes [14]. Recently, machine learning has become indispensable in the development of predictive models. The performance of machine-learning models varies depending on the combination of various feature transforms, feature selection, oversampling, and machine-learning algorithms. Thus, there is a need for machine-learning approaches with good performance that can be easily applied in real clinical situations.

This study aimed to develop and validate a diagnostic model for HPV status using ^{18}F-FDG PET/CT-derived parameters with an appropriate combination of feature selection, oversampling, and machine learning methods in OPSCC.

MATERIALS AND METHODS

Patient Population

This retrospective study was approved by our Institutional Review Board of Severance Hospital, which waived the requirement for informed consent (IRB No. 4-2021-1449) between January 2012 and February 2020, patients diagnosed with OPSCC who underwent ^{18}F-FDG PET/CT before treatment were identified using an electronic medical record (EMR) system. Patients with a history of treatment were excluded from the study. Patients without contrast-enhanced-CT or MRI data were excluded to ensure accurate tumor localization. Patients with suboptimal ^{18}F-FDG PET/CT findings, such as those with high blood glucose levels (> 150 mg/dL), were excluded [15]. Patients with damaged Digital Imaging and Communications in Medicine (DICOM) files were excluded from the study. The patient cohort was allocated to the internal training and validation sets. Patient data regarding age, sex, smoking history (pack-years [PY]), and tumor subsite were collected from the EMR system.

For the external test set, data were collected from patients diagnosed with OPSCC using ^{18}F-FDG PET/CT before treatment between February 2021 and December 2021 from two other tertiary general hospitals (Korea University GURO Hospital and Korea University ANAM Hospital). The same inclusion and exclusion criteria as for the internal set were used. Clinical data were collected from the EMR system of each hospital. Just like the interna set, the flowchart of the study is presented in Figure 1.

HPV Testing and Immunohistochemistry

Tumor specimens for immunohistochemical analysis were obtained from biopsies and surgery. HPV phenotyping was assessed in the internal training and validation sets based on the overexpression of p16 protein, which is strongly correlated with HPV positivity. Immunohistochemical analysis was performed using primary antibodies against p16 (E6H4; Roche MTM Laboratories AG). p16 positivity was defined as the presence of a diffuse staining pattern on the p16-stained slide specimens, whereas negativity was defined as either negative staining or staining of isolated cells or small cell clusters.

In the external test set, p16 immunohistochemistry analysis was performed at the Korea University GURO Hospital, and both p16 immunohistochemistry and polymerase chain reaction for HPV were used at the Korea

https://doi.org/10.3348/kjr.2022.0397

Korean Journal of Radiology
KJR

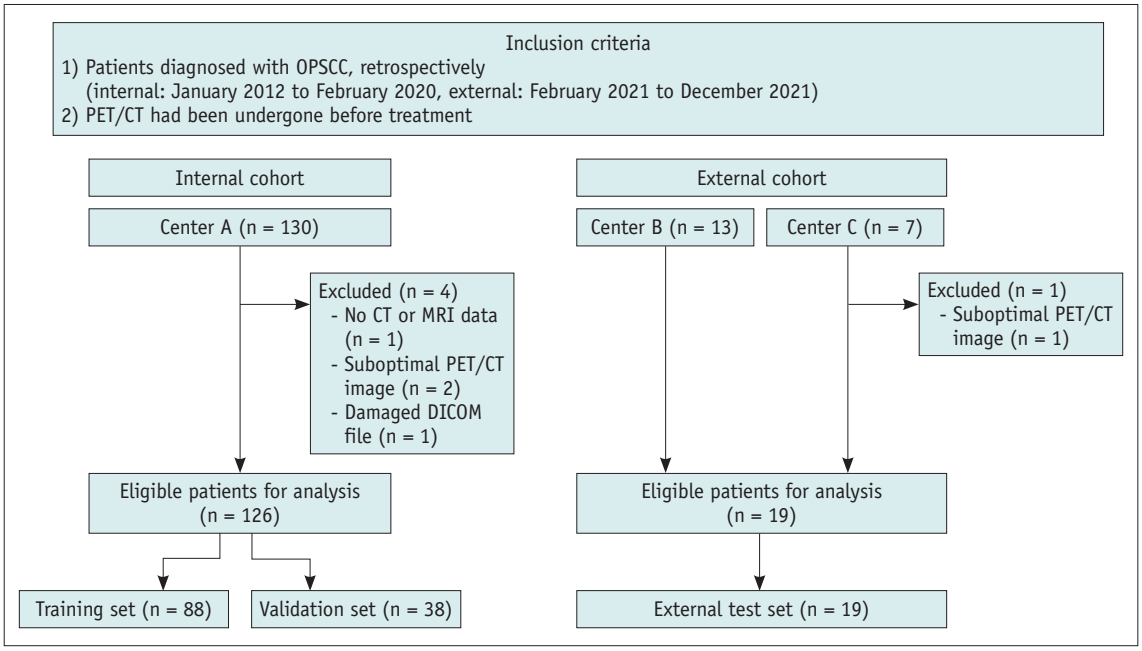

Fig. 1. Flowchart of patient enrollment. OPSCC = oropharyngeal squamous cell carcinoma

University ANAM Hospital [16].

^{18}F-FDG PET/CT Image Acquisition and Analysis

In an internal cohort, ^{18}F-FDG PET/CT was performed using Discovery PET/CT 600, 690, and 710 (General Electric Medical Systems). Before the ^{18}F-FDG injection, all patients fasted for at least 6 hours, and their blood glucose levels were confirmed to be ≤ 150 mg/dL. The ^{18}F-FDG was administered intravenously at a dose of 3.7–4.4 MBq/kg body weight. After the initial low-dose CT (Discovery 600, 690, 710:120 kVp, 30 mA), standard PET imaging was performed from the skull base to the mid-thighs, with a 1.5 minutes acquisition time per bed position.

For the external test set, PET/CT imaging of OPSCC patients from the Korea University GURO Hospital and Korea University ANAM Hospital was performed using Discovery MI (General Electric Medical Systems). Initial low-dose CT scans for attenuation correction (120 kVp, 30 mA) and PET scans of the same area were performed from the skull base to the proximal thighs with 1.5 minutes acquisition time per bed position. All other methods used for image acquisition and data analysis were adopted in the same manner as those used in the aforementioned institution.

The maximum standard uptake value (SUV_{max}) was the highest metabolic focus in all metabolically active primary tumors. Metabolic tumor volume (MTV) was automatically calculated by summing the total volume of voxels using a percentage threshold of 40% of the SUV_{max}. SUV_{mean} is the mean SUV of the volume of interest (VOI), and SUV_{peak} is defined as the average SUV within a small, fixed-size region of interest centered on a high-uptake part of the tumor. The total lesion glycolysis (TLG) of each primary tumor was calculated by multiplying the SUV_{mean} by the voxel number of the lesion. To obtain the tumor SUV_{max} to liver SUV_{mean} ratio (SUV_{max}-tumor to liver ratio [TLR]), the normal background liver SUV_{mean} was measured by drawing a 3-cm-sized spherical VOI in the right lobe of the liver. According to these definitions, metabolic and volume-based parameters such as SUVmax, SUV_{max}-TLR, SUVmean, SUV_{peak}, MTV, and TLG of the tumor VOI were derived. All semi-quantitative and volumetric measurements were conducted using the volume viewer software (MIM-7.0; MIM Software Inc.) by two experienced nuclear medicine specialists with > 10 years of experience (Fig. 2). Medical records, including the patient's HPV status, were not recorded during these measurements.

Machine Learning and Statistical Analysis

The internal patient cohort was randomly divided into the training and validation sets (7:3 ratio). After the training set was prepared, three groups of models were developed. First, we built Model 1, which used only the PET-derived parameters. Second, we built Model 2 with clinical features only, including age, sex, smoking history, and tumor subsite. Finally, Model 3 was developed using the PET-derived parameters and clinical features. A grid

 https://doi.org/10.3348/kjr.2022.0397

search was performed using a feature transform, feature selection, oversampling methods, and a training model to explore high-performance methods. Seven methods were used for the feature transform: 1) no transform, 2) standard (Z-score) transformation, 3) MinMax transformation, 4) power transformer, 5) quantile transformer, 6) normalized transformer (L1), and 7) normalized transformer (L2). Next, various features were selected using the SelectKBest method, based on 1) f-regression and 2) mutual information [17]. For oversampling, 1) random-oversampling examples, 2) synthetic minority oversampling technique (SMOTE), 3) borderline SMOTE, 4) adaptive synthetic sampling approach (ADASYN), 5) SMOTEEN, 6) SMOTETomek, and 7) no oversampling were investigated [18-21]. Finally,

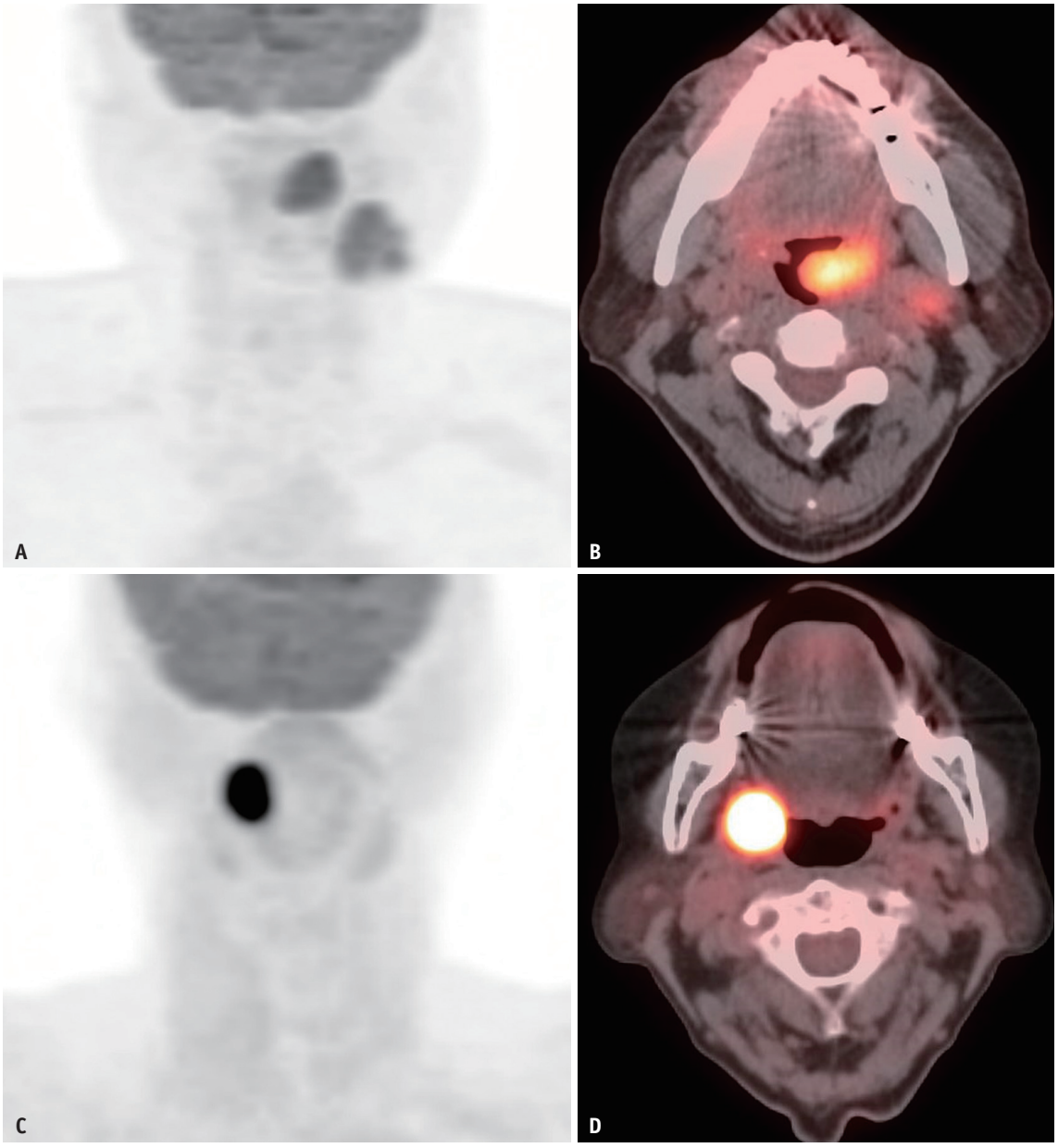

Fig. 2. Two representative cases show the different characteristics of HPV-associated OPSCC and HPV-negative OPSCC in ^{18}F-FDG PET/CT.
A, B. A 59-year-old male with HPV-associated OPSCC. Maximum intensity projection **(A)** and fused PET/CT **(B)** images. The primary tumor is seen in the left tonsil, and cystic nodal metastasis, which is more prevalent in HPV-associated OPSCC, is also present (tumor SUV_{max} 11.9, SUV_{max}-TLR 3.7). **C, D.** A 74-year-old male with HPV-negative OPSCC. Maximum intensity projection **(C)** and fused PET/CT **(D)** images showing the primary tumor in the right tonsil with intense ^{18}F-FDG uptake (tumor SUV_{max} 17.1, SUV_{max}-TLR 7.7). ^{18}F-FDG = ^{18}F-fluorodeoxyglucose, HPV = human papillomavirus, OPSCC = oropharyngeal squamous cell carcinoma, SUV_{max} = maximum standard uptake value, SUV_{max}-TLR = tumor SUV_{max} to liver SUV_{mean} ratio

https://doi.org/10.3348/kjr.2022.0397

Korean Journal of Radiology
KJR

six training models were explored: 1) random forest, 2) LightGBM, 3) ExtraTrees, 4) XGBoost, 5) AdaBoost, and 6) logistic regression [22-25]. All the possible combination models were developed using an internal training set. Ten-fold cross-validation was performed internally during the training process. After each model was developed, its performance was compared by using an internal validation set. The models with the three highest performances were selected for each group. The selected models were tested using an external test set, and the best model was chosen based on the area under the receiver operating characteristic (ROC) curve (AUC) in the external test set. The external test set AUC comparison for the three models was conducted using the Delong method. ROC curves for the three models were drawn based on the external test set.

To examine which parameter played an important role in the best model, we calculated the mean absolute Shapley value for each selected input feature using the Shapley additive explanations (SHAP) algorithm [26]. Additionally, we compared PET-derived parameters between the HPV-positive and HPV-negative groups. All processes up to this point were performed using Python 3 with the ScikitLearn library v0.21.2, and R software (version 3.5.1; R Foundation for Statistical Computing).

RESULTS

Among the 130 enrolled institutional patients, four were excluded (one without CT or MRI data, two with suboptimal PET images, and one with a damaged DICOM file). Finally, 126 patients were included in the analysis. In this internal cohort, 118 were men and eight were women (mean age, 60 years), and 103 (81.7%) and 23 (18.3%) were HPV-positive and HPV-negative, respectively. These patients were divided

Table 1. Patient Characteristics in the Training Set, Internal Validation Set, and External Test Set

	Internal Dataset			External Dataset	
	Training Set (n = 88)	Validation Set (n = 38)	*P**	Test Set (n = 19)	*P*†
Age, years	60.0 ± 8.5	60.1 ± 9.5	0.946‡	65.3 ± 11.8	0.075‡
Sex			1.000§		0.197§
Male	82 (93.2)	36 (94.7)		16 (84.2)	
Female	6 (6.8)	2 (5.3)		3 (15.8)	
HPV status			0.975‖		0.752§
Negative	16 (18.2)	7 (18.4)		4 (21.1)	
Positive	72 (81.8)	31 (81.6)		15 (78.9)	
Smoking history, pack-year	15.8 ± 20.6	20.3 ± 20.3	0.262‡	25.7 ± 16.5¶ (n = 12)	0.102‡
Tumor subsite			0.260‖		0.096‖
Tonsil	67 (76.1)	30 (78.9)		14 (73.7)	
BOT	21 (23.9)	7 (18.4)		4 (21.1)	
Soft palate	0 (0)	1 (2.6)		0 (0)	
PPW	0 (0)	0 (0)		1 (5.3)	
Staging					
T stage			0.120‖		0.965‖
T1	27 (30.7)	5 (13.1)		6 (31.6)	
T2	42 (47.7)	19 (50.0)		8 (42.1)	
T3	8 (9.1)	7 (8.0)		2 (10.5)	
T4 (T4a, T4b)	11 (12.5)	7 (8.0)		3 (15.8)	
N stage			0.058‖		0.001‖
N0	6 (6.8)	8 (21.1)		4 (21.1)	
N1	40 (45.5)	13 (34.2)		13 (68.4)	
N2 (N2a, N2b, N2c)	42 (47.7)	17 (44.7)		1 (5.3)	
N3	0 (0)	0 (0)		1 (5.3)	

Data are presented as means ± standard deviation or numbers of patients (%) unless specified otherwise. *Comparing training and internal validation set, †Comparing training and external set, ‡Calculated using *t* tests, §Calculated using Fisher's exact tests, ‖Calculated using chi-square tests, ¶Missing data of smoking history in 7 patients. BOT = base of tongue, HPV = human papillomavirus, PPW = posterior pharyngeal wall

https://doi.org/10.3348/kjr.2022.0397

into training (n = 88) and validation (n = 38) sets in a 7:3 ratio. The external cohort was enrolled from the following two institutions: 7 from the Korea University ANAM Hospital and 13 from the Korea University GURO Hospital. One patient with suboptimal PET images was excluded from the Korea University ANAM Hospital. Of the 19 included patients (6 from Korea University ANAM Hospital and 13 from Korea University GURO Hospital; 16 male and 3 female; mean age, 65.3 years), 15 (78.9%) and 4 (21.1%) were HPV-positive and HPV-negative, respectively. The 19 patients were allocated to the external test set (Fig. 1). The proportions of each HPV status did not differ significantly between the training and internal validation sets or between the training and external test sets. Age, sex, and smoking history did not significantly differ between the training and external test sets. Patient demographics in the training, internal validation, and external test sets are presented in Table 1. The smoking history of seven patients from the external test set was unavailable in the EMR system; therefore, they were considered missing values.

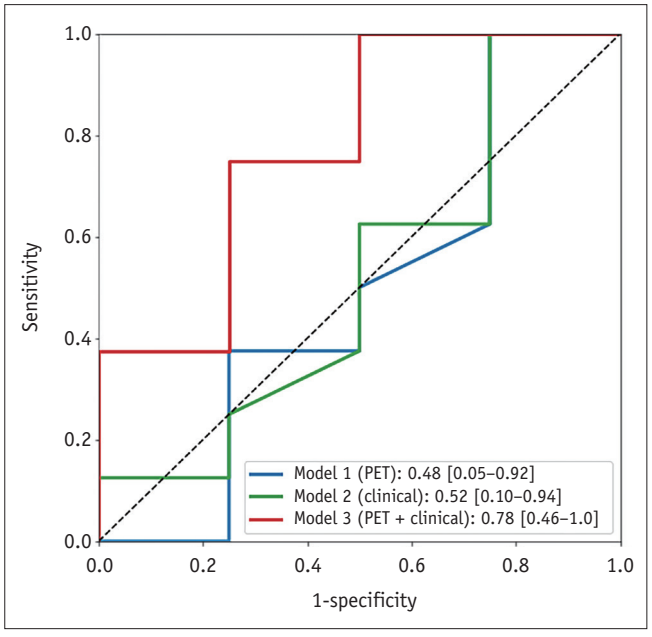

Fig. 3. ROC curves for the best Models 1, 2, and 3 in the external test set. The values in the box in the lower right corner are the area under the ROC curve [95% confidence interval]. ROC = receiver operating characteristic

A comparison of PET-derived parameters between HPV-positive and HPV-negative OPSCC patients showed a significant difference in the SUV_{max} in the internal dataset (all patients, n = 126). However, there was no statistical difference in the SUV_{max} in the external test set (Supplementary Table 1).

Various models were trained by combining various machine learning preprocessing methods and classifiers, and their performances in the validation set were calculated. Candidate models were drafted using the AUC in the internal validation set from each group using PET-derived parameters, clinical features, and both. ROC analysis was performed to identify the best models with the highest AUC in each group (Fig. 3). Regarding the AUC, the best model was Model 3, which used both PET-derived parameters and clinical features using the ExtraTreesClassifier, MinMaxScaler transformation, mutual information-based feature selection, and ADASYN oversampling. This model used five features: two PET-derived parameters (SUV_{max} and SUV_{max}-TLR) and three clinical features (age, smoking history, and tumor subsite). In the external test set, the area under the curve of the model was 0.78 in the external test set. The AUC of Model 3 was significantly higher than that of Model 1 (AUC = 0.48, p = 0.047). The AUC of Model 3 was also higher than that of Model 2; however, the difference was not statistically significant (AUC = 0.52, p = 0.142). The performance of the models for each group is summarized in Table 2. The sensitivity, specificity, accuracy, and F1 score of Model 3 were 100.0%, 50.0%, 83.3%, and 88.9%,

Table 2. Performance of Models 1, 2, and 3 in the Internal Validation Set and External Test Set

Model	ML Method	Internal Validation Set	External Test Set					
		AUC (95% CI)	AUC (95% CI)	*P*	Accuracy, %	Precision, %	Recall, %	F1 Score
Model 1 (PET)	RandomForest	0.71 (0.79–0.92)	0.48 (0.05–0.92)	0.047	50.0 (6/12)	62.5 (5/8)	62.5 (5/8)	0.63
Model 2 (clinical)	XGBoost	0.81 (0.67–0.95)	0.52 (0.10–0.94)	0.142	50.0 (6/12)	66.7 (4/6)	50.0 (4/8)	0.53
Model 3 (PET + clinical)	ExtraTrees	0.77 (0.59–0.94)	0.78 (0.46–1.00)	Reference	83.3 (10/12)	80.0 (8/10)	100.0 (8/8)	0.89

Model 1 used PET-derived parameters only, Model 2 used clinical features only, and Model 3 used both PET-derived parameters and clinical features. The cut off value of 0.5 was chosen to calculate accuracy, precision, recall, and F1 score. AUC = area under the receiver operating characteristic curve, ML = machine learning

https://doi.org/10.3348/kjr.2022.0397

Korean Journal of Radiology
KJR

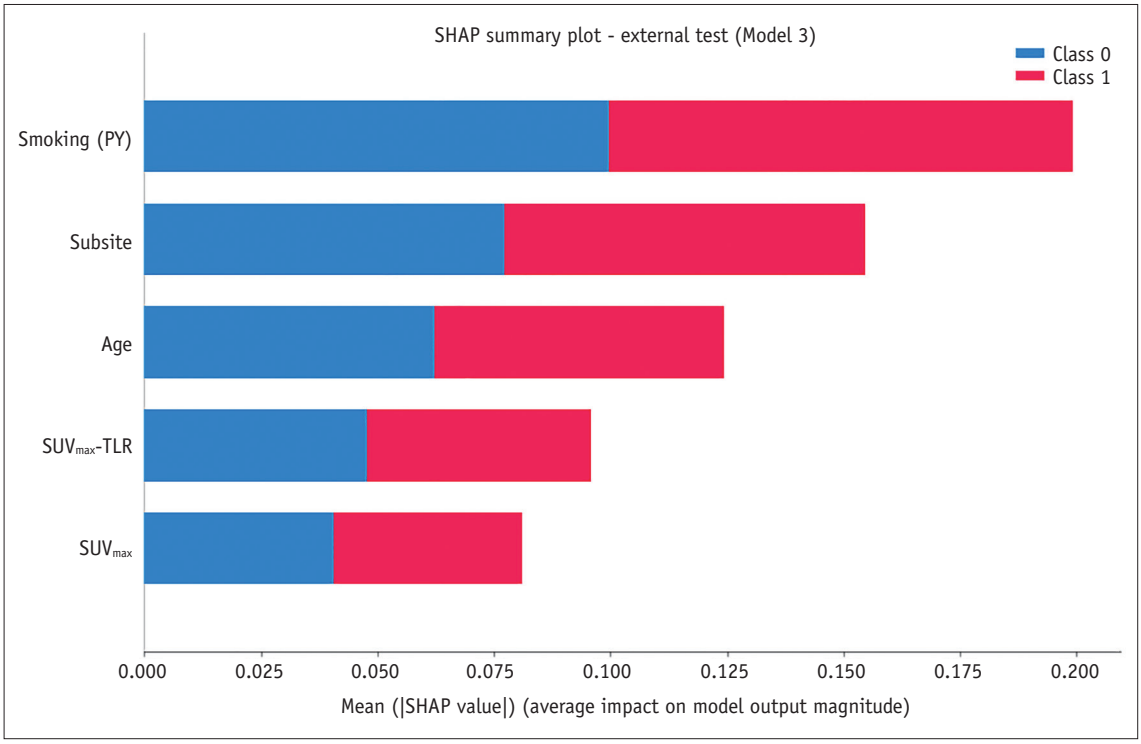

Fig. 4. Bar summary plot of mean absolute Shapley values for five parameters of the best Model 3. PY = pack-year, SHAP = Shapley additive explanations, SUV_{max} = maximum standard uptake value, SUV_{max}-TLR = tumor SUV_{max} to liver SUV_{mean} ratio

respectively.

Using the SHAP algorithm, the importance of the five parameters was visualized using a bar plot (Fig. 4). Smoking history had the most significant effect on prediction, followed by tumor subsite, age, SUV_{max}-TLR, and SUV_{max}. Analysis of the dot SHAP summary plots showed that a low PY for smoking history, tonsil location, low age, low SUV_{max}-TLR, and low SUV_{max} predicted HPV-positive OPSCC (Fig. 5).

DISCUSSION

In this study, an extratree-based model was developed to predict the HPV status of OPSCC using ^{18}F-FDG PET/CT-derived metabolic parameters by combining the MinMaxScaler feature transform and ADASYN oversampling. Our model predicted HPV status by incorporating clinical features and easily obtainable PET/CT parameters, and its performance in an external test set was comparable to that of prior research employing radiomic approaches [13].

Age, smoking history, and tumor site were included as clinical predictors of HPV status in our model. Smoking is a well-known risk factor for OPSCC. However, smoking is not a cofactor for HPV-mediated oropharyngeal carcinogenesis [27]. HPV-positive OPSCC is more common than HPV-negative OPSCC among non-smokers [28]. The prevalence of HPV-positive OPSCC decreases with age [29]. Most cases involve individuals aged < 65 years [30,31]. In a recent systematic review and meta-analysis, Haeggblom et al. [32] found considerable variation in HPV prevalence across the anatomic locations of the oropharynx. In tonsillar regions, the presence of a highly specialized crypt lymphoepithelium has been established as a permissive environment for HPV-driven carcinogenesis.

Various studies have been conducted using CT or MRI to determine HPV status. ^{18}F-FDG PET/CT is superior to MRI and CT for evaluating locoregional neck node metastasis and distant metastasis. The National Comprehensive Cancer Network guidelines also highlight the usefulness of ^{18}F-FDG PET/CT for head and neck cancer images [33-35]. ^{18}F-FDG PET/CT is distinct from CT and MRI in that it allows for direct imaging of metabolism. Therefore, it is worthwhile to investigate the broader applications of PET/CT, such as HPV prediction. ^{18}F-FDG PET/CT, in contrast, has some drawbacks, including non-uniformity in processing, difficulties in quality monitoring of equipment, and a higher cost than CT.

When comparing PET/CT-derived parameters, HPV-negative tumors showed a significantly higher SUV_{max} than HPV-positive tumors in our internal dataset, which is

 https://doi.org/10.3348/kjr.2022.0397

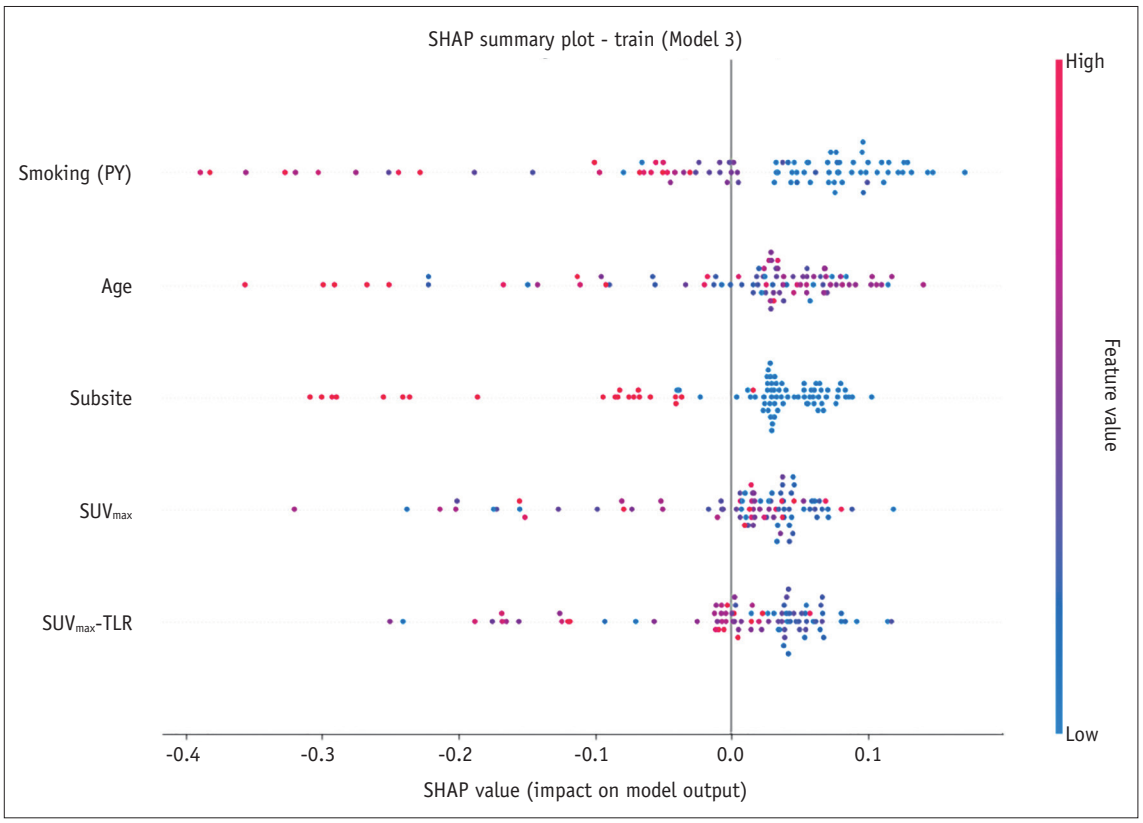

Fig. 5. Dot summary plots by SHAP in training set from the best Model 3. PY = pack-year, SHAP = Shapley additive explanations, SUV_{max} = maximum standard uptake value, SUV_{max}-TLR = tumor SUV_{max} to liver SUV_{mean} ratio

consistent with the results of previous literature [36]. The different histological characteristics of HPV-positive and HPV-negative OPSCC might account for the disparate SUV_{max} values, possibly because of the different distributions of hypoxic zones within the tumor [37,38]. Studies on the relationship between FDG parameters and glucose transporter (GLUT) in malignant tumors are ongoing [39-41]. Although it is expected that the HPV state, expression level of GLUT, and FDG avidity are related, many unknown factors remain, demonstrating the complexity of OPSCC pathophysiology [42,43]. SUV_{max} and SUV_{max}-TLR were the PET parameters used in the final model. SUV_{max}-TLR is a standardized method for semi-quantitative measurements to compensate for the weakness of the SUV values. This method may provide reliable and reproducible data as well as improve the PET characterization of tumors [44]. Few studies have attempted to classify HPV status using the radiomic approach of PET imaging [13]. Radiomics is time consuming because of the need for region-of-interest mapping, complex preprocessing, and feature extraction. This time-consuming and labor intensive process is a major disadvantage for clinical applications. Therefore, we only included parameters that were simple and generally easy-to-measure. In our model, which is faster and easier to apply, Only SUV_{max} and SUV_{max}-TLR were required.

A biopsy of the primary oropharyngeal lesion is required to diagnose OPSCC because many cases are treated surgically. Even after diagnosis, the process of obtaining tissue is inextricably included in the diagnosis and treatment. Therefore, the necessity of determining HPV status using ^{18}F-FDG PET/CT before biopsy and treatment and its clinical implications may be questioned. Nevertheless, as with many previous studies that used image-based gene prediction, this study began with exploratory objectives. Additionally, while an acceptable predictive performance to replace biopsy was not reached in this investigation, the AUC might be improved in our ongoing study, which combines MRI and PET. Moreover, the ability to double-check for possible biopsy sampling errors might be advantageous. Finally, if there is a discrepancy between the HPV status determined by actual biopsy and the predicted HPV status using images, we believe that future investigations should focus on determining the cause of the discrepancy.

Numerous image-based HPV classification studies have

https://doi.org/10.3348/kjr.2022.0397

Korean Journal of Radiology
KJR

recently been conducted using machine learning, which incorporates a diverse set of pre-processing methods and classification models. The performance of the model is determined by the classifier utilized, as well as the feature transformation, selection, and oversampling approaches. However, there is a risk of overfitting without sufficient external validation. Developing a model that learns and makes predictions using non-uniformly distributed data is a challenging task. The incidence of HPV-positive cancer has recently increased in OPSCC and now accounts for a significantly greater proportion than HPV-negative OPSCC [45,46]. When dealing with unbalanced data, proper oversampling should be used to avoid overfitting in directions with higher frequencies. In our investigation, we discovered that when the oversampling method was utilized, the performance was superior to when no oversampling was employed. We selected a tree-based model for the machine learning classifier. This finding is consistent with an earlier study utilizing PET/CT radiomics, which frequently uses XGBoost [13]. The tree-based classifier appeared to be appropriate for PET-based HPV predictions. This conclusion will be useful in future large-scale studies.

This study had some limitations. First, the model's performance for determining HPV status was insufficient to substitute tissue biopsy with a molecular diagnosis. Second, although our external test set was enrolled from two hospitals, the small number of patients, particularly in the external test set, is a major limitation of this investigation. In particular, the number of HPV-negative cases in the test set is small. This is well demonstrated by the absence of a statistical difference between Models 2 and 3, as well as an SUV_{max} difference between HPV-positive and HPV-negative cases in the external set. In future research, a larger test set size would be desirable to validate the robustness of the model. Images were obtained using the same reconstruction algorithm with the same settings whenever possible; however, using different scanners in different institutions was also a limitation. Finally, we used p16 as a surrogate marker for HPV positivity. However, p16 and HPV statuses can be mismatched [47,48].

In conclusion, using appropriate preprocessing and a classifier, we developed and tested an HPV status classifier model based on metabolic parameters derived from ^{18}F-FDG PET/CT and clinical parameters in OPSCC patients. It exhibited higher performance than models using either PET or clinical parameters alone.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2022.0397.

Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Conflicts of Interest

Jinna Kim who is on the editorial board of the *Korean Journal of Radiology* was not involved in the editorial evaluation or decision to publish this article. All remaining authors have declared no conflicts of interest.

Author Contributions

Conceptualization: Beomseok Sohn. Data curation: Beomseok Sohn, Kwan Hyeong Jo, Hojin Cho. Formal analysis: Beomseok Sohn, Kwan Hyeong Jo, Hojin Cho, Kisung Park. Funding acquisition: Beomseok Sohn. Investigation: Changsoo Woo, Beomseok Sohn, Kwan Hyeong Jo. Methodology: Beomseok Sohn, Kwan Hyeong Jo. Project administration: Beomseok Sohn, Kwan Hyeong Jo. Resources: Hojin Jo, Won Jun Kang, Jinna Kim, Seung-Koo Lee. Software: Beomseok Sohn, Kisung Park. Supervision: Seung-Koo Lee, Won Jun Kang. Validation: Beomseok Sohn, Kwan Hyeong Jo. Visualization: Kisung Park, Kwan Hyeong Jo, Beomseok Sohn. Writing—original draft: Changsoo Woo, Beomseok Sohn, Kwan Hyeong Jo. Writing—review & editing: all authors.

ORCID iDs

Changsoo Woo
https://orcid.org/0000-0002-2036-2135
Kwan Hyeong Jo
https://orcid.org/0000-0003-0363-9742
Beomseok Sohn
https://orcid.org/0000-0002-6765-8056
Kisung Park
https://orcid.org/0000-0002-7447-4247
Hojin Cho
https://orcid.org/0000-0002-8686-172X
Won Jun Kang
https://orcid.org/0000-0002-2107-8160
Jinna Kim
https://orcid.org/0000-0002-9978-4356

Seung-Koo Lee
https://orcid.org/0000-0001-5646-4072

Funding Statement
This study was supported by a faculty research grant of Yonsei University College of Medicine for (6-2021-0151).

REFERENCES

1. McDermott JD, Bowles DW. Epidemiology of head and neck squamous cell carcinomas: impact on staging and prevention strategies. *Curr Treat Options Oncol* 2019;20:43
2. Tumban E. A current update on human papillomavirus-associated head and neck cancers. *Viruses* 2019;11:922
3. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers* 2020;6:92
4. Taberna M, Mena M, Pavón MA, Alemany L, Gillison ML, Mesía R. Human papillomavirus-related oropharyngeal cancer. *Ann Oncol* 2017;28:2386-2398
5. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. *AJCC cancer staging manual*, 8th ed. New York: Springer, 2017
6. Solomon B, Young RJ, Rischin D. Head and neck squamous cell carcinoma: genomics and emerging biomarkers for immunomodulatory cancer treatments. *Semin Cancer Biol* 2018;52(Pt 2):228-240
7. You EL, Henry M, Zeitouni AG. Human papillomavirus-associated oropharyngeal cancer: review of current evidence and management. *Curr Oncol* 2019;26:119-123
8. Benson E, Li R, Eisele D, Fakhry C. The clinical impact of HPV tumor status upon head and neck squamous cell carcinomas. *Oral Oncol* 2014;50:565-574
9. Spadarella G, Ugga L, Calareso G, Villa R, D'Aniello S, Cuocolo R. The impact of radiomics for human papillomavirus status prediction in oropharyngeal cancer: systematic review and radiomics quality score assessment. *Neuroradiology* 2022;64:1639-1647
10. Ravanelli M, Grammatica A, Tononcelli E, Morello R, Leali M, Battocchio S, et al. Correlation between human papillomavirus status and quantitative MR imaging parameters including diffusion-weighted imaging and texture features in oropharyngeal carcinoma. *AJNR Am J Neuroradiol* 2018;39:1878-1883
11. 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* 2021;131:E851-E856
12. 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. *Sci Rep* 2020;10:17525
13. Haider SP, Mahajan A, Zeevi T, Baumeister P, Reichel C, Sharaf K, et al. PET/CT radiomics signature of human papilloma virus association in oropharyngeal squamous cell carcinoma. *Eur J Nucl Med Mol Imaging* 2020;47:2978-2991
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278:563-577
15. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* 2015;42:328-354
16. Begum S, 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
17. Pope P, Webster J. The use of an F-statistic in stepwise regression procedures. *Technometrics* 1972;14:327-340
18. Lunardon N, Menardi G, Torelli N. ROSE: a package for binary imbalanced learning. *R Journal* 2014;6:79-89
19. He H, Bai Y, Garcia EA, Li S. ADASYN: adaptive synthetic sampling approach for imbalanced learning. Proceedings of the International Joint Conference on Neural Networks, IJCNN 2008, part of the IEEE World Congress on Computational Intelligence, WCCI 2008; 2008 Jun 1-6; Hong Kong: IEEE; 2008:1322-1328
20. Han H, Wang WY, Mao BH. *Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning*. In: Huang DS, Zhang XP, Huang GB, eds. *Advances in intelligent computing. ICIC 2005. Lecture notes in computer science, vol 3644*. Berlin: Springer, 2005:878-887
21. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002;16:321-357
22. Simm J, De Abril IM, Sugiyama M. Tree-based ensemble multi-task learning method for classification and regression. *IEICE Trans Inf Syst* 2014;97:1677-1681
23. Schapire RE. *Explaining AdaBoost*. In: Schölkopf B, Luo Z, Vovk V, eds. *Empirical inference*. Berlin, Heidelberg: Springer, 2013
24. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4-9; Long Beach, CA, USA: NIPS; 2017:3149-3157
25. Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 Aug 13-17; San Francisco, CA, USA: KDD; 2016:785-794
26. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4-9; Long Beach, CA, USA: NIPS; 2017:4768-4777
27. D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, et al. Case-control study of human papillomavirus and oropharyngeal cancer. *N Engl J Med* 2007;356:1944-1956
28. Lechner M, Liu J, Masterson L, Fenton TR. HPV-associated

oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nat Rev Clin Oncol* 2022;19:306-327
29. Schache AG, Powell NG, Cuschieri KS, Robinson M, Leary S, Mehanna H, et al. HPV-related oropharynx cancer in the United Kingdom: an evolution in the understanding of disease etiology. *Cancer Res* 2016;76:6598-6606
30. Chaturvedi AK, Zumsteg ZS. A snapshot of the evolving epidemiology of oropharynx cancers. *Cancer* 2018;124:2893-2896
31. Zumsteg ZS, Cook-Wiens G, Yoshida E, Shiao SL, Lee NY, Mita A, et al. Incidence of oropharyngeal cancer among elderly patients in the United States. *JAMA Oncol* 2016;2:1617-1623
32. Haeggblom L, Ramqvist T, Tommasino M, Dalianis T, Näsman A. Time to change perspectives on HPV in oropharyngeal cancer. A systematic review of HPV prevalence per oropharyngeal subsite the last 3 years. *Papillomavirus Res* 2017;4:1-11
33. Pfister DG, Ang KK, Brizel DM, Burtness BA, Busse PM, Caudell JJ, et al. Head and neck cancers, version 2.2013. Featured updates to the NCCN guidelines. *J Natl Compr Canc Netw* 2013;11:917-923
34. Sohn B, Koh YW, Kang WJ, Lee JH, Shin NY, Kim J. Is there an additive value of 18 F-FDG PET-CT to CT/MRI for detecting nodal metastasis in oropharyngeal squamous cell carcinoma patients with palpably negative neck? *Acta Radiol* 2016;57:1352-1359
35. Kim SJ, Pak K, Kim K. Diagnostic accuracy of F-18 FDG PET or PET/CT for detection of lymph node metastasis in clinically node negative head and neck cancer patients; a systematic review and meta-analysis. *Am J Otolaryngol* 2019;40:297-305
36. Freihat O, Tóth Z, Pintér T, Kedves A, Sipos D, Cselik Z, et al. Pre-treatment PET/MRI based FDG and DWI imaging parameters for predicting HPV status and tumor response to chemoradiotherapy in primary oropharyngeal squamous cell carcinoma (OPSCC). *Oral Oncol* 2021;116:105239
37. Bose P, Brockton NT, Dort JC. Head and neck cancer: from anatomy to biology. *Int J Cancer* 2013;133:2013-2023
38. Wijsman R, Kaanders JH, Oyen WJ, Bussink J. Hypoxia and tumor metabolism in radiation oncology: targets visualized by positron emission tomography. *Q J Nucl Med Mol Imaging* 2013;57:244-256
39. Harshani JM, Yeluri S, Guttikonda VR. Glut-1 as a prognostic biomarker in oral squamous cell carcinoma. *J Oral Maxillofac Pathol* 2014;18:372-378
40. Botha H, Farah CS, Koo K, Cirillo N, McCullough M, Paolini R, et al. The role of glucose transporters in oral squamous cell carcinoma. *Biomolecules* 2021;11:1070
41. Baschnagel AM, Wobb JL, Dilworth JT, Williams L, Eskandari M, Wu D, et al. The association of (18)F-FDG PET and glucose metabolism biomarkers GLUT1 and HK2 in p16 positive and negative head and neck squamous cell carcinomas. *Radiother Oncol* 2015;117:118-124
42. Krupar R, Robold K, Gaag D, Spanier G, Kreutz M, Renner K, et al. Immunologic and metabolic characteristics of HPV-negative and HPV-positive head and neck squamous cell carcinomas are strikingly different. *Virchows Arch* 2014;465:299-312
43. Ribbat-Idel J, Perner S, Kuppler P, Klapper L, Krupar R, Watermann C, et al. Immunologic "cold" squamous cell carcinomas of the head and neck are associated with an unfavorable prognosis. *Front Med (Lausanne)* 2021;8:622330
44. Keramida G, Dizdarevic S, Bush J, Peters AM. Quantification of tumour (18) F-FDG uptake: normalise to blood glucose or scale to liver uptake? *Eur Radiol* 2015;25:2701-2708
45. Chaturvedi A, Engels E, Pfeiffer R, Hernandez B, Xiao W, Kim E, et al. Human papillomavirus (HPV) and rising oropharyngeal cancer incidence and survival in the United States. *J Clin Oncol* 2011;29:5529
46. Damgacioglu H, Sonawane K, Zhu Y, Li R, Balasubramanian BA, Lairson DR, et al. Oropharyngeal cancer incidence and mortality trends in all 50 states in the US, 2001-2017. *JAMA Otolaryngol Head Neck Surg* 2022;148:155-165
47. Shinn JR, Davis SJ, Lang-Kuhs KA, Rohde S, Wang X, Liu P, et al. Oropharyngeal squamous cell carcinoma with discordant p16 and HPV mRNA results: incidence and characterization in a large, contemporary United States cohort. *Am J Surg Pathol* 2021;45:951-961
48. Arsa L, Siripoon T, Trachu N, Foyhirun S, Pangpunyakulchai D, Sanpapant S, et al. Discrepancy in p16 expression in patients with HPV-associated head and neck squamous cell carcinoma in Thailand: clinical characteristics and survival outcomes. *BMC Cancer* 2021;21:504

 https://doi.org/10.3348/kjr.2022.0397