PATHOLOGICA 2024;116:379-389;
DOI: 10.32074/1591-951X-1027

Original article

A machine learning approach to predict HPV positivity of oropharyngeal squamous cell carcinoma

Silvia Varricchio[1*], Gennaro Ilardi[1*], Angela Crispino[1], Marco Pietro D'Angelo[2], Daniela Russo[1], Rosa Maria Di Crescenzo[1], Stefania Staibano[1**], Francesco Merolla[2**]

[1] Department of Advanced Biomedical Sciences, University of Naples "Federico II", Naples, Italy; [2] Department of Medicine and Health Sciences "V. Tiberio", University of Molise, Campobasso, Italy
**These authors equally contributed to this work*
***Co-senior authors*

Summary
HPV status is an important prognostic factor in oropharyngeal squamous cell carcinoma (OPSCC), with HPV-positive tumors associated with better overall survival. To determine HPV status, we rely on the immunohistochemical investigation for expression of the $P16^{INK4a}$ protein, which must be associated with molecular investigation for the presence of viral DNA. We aim to define a criterion based on image analysis and machine learning to predict HPV status from hematoxylin/eosin stain.
We extracted a pool of 41 morphometric and colorimetric features from each tumor cell identified from two different cohorts of tumor tissues obtained from the Cancer Genome Atlas and the archives of the Pathological Anatomy of Federico II of Naples. On this data, we built a random Forest classifier. Our model showed a 90% accuracy. We also studied the variable importance to define a criterion useful for the explainability of the model. Prediction of the molecular state of a neoplastic cell based on digitally extracted morphometric features is fascinating and promises to revolutionize histopathology. We have built a classifier capable of anticipating the result of p16-immunohistochemistry and molecular test to assess the HPV status of squamous carcinomas of the oropharynx by analyzing the hematoxylin/eosin staining.

Key words: HPV, OPSCC, Machine Learning, Computational Pathology, Digital Pathology, HistoQC, QuPath

Received: May 9, 2024
Accepted: December 9, 2024

Correspondence
Francesco Merolla
E-mail: francesco.merolla@unimol.it

How to cite this article: Varricchio S, Ilardi G, Crispino A, et al. A machine learning approach to predict HPV positivity of oropharyngeal squamous cell carcinoma. Pathologica 2024;116:379-389. https://doi.org/10.32074/1591-951X-1027

Introduction

The epidemiology of oropharyngeal squamous cell carcinoma (OPSCC) has changed profoundly in recent years. The incidence of OPSCC attributed to tobacco and alcohol exposure has gradually decreased, while its correlation with human papillomavirus (HPV) infection is becoming increasingly evident[1,2]. In Italy, it is estimated that 31% of OPSCC are attributed to HPV infection[2]. In contrast, lower fractions, < 10% and 2.4%, have been estimated for oral cavity and laryngeal carcinomas, respectively[1-3]. It is now clear that HPV-positive OPSCCs represent a biologically distinct entity.
OPSCC includes tumors of the tonsils, base of the tongue, soft palate, and throat. Clinically, it is evident that they have a better prognosis compared to HPV-negative carcinomas[4]. For this reason, the 8th edition of the UICC/AJCC staging system has defined HPV-positive and HPV-negative OPSCCs as separate entities with distinct molecular profiles, tu-

mor characteristics, and outcomes[5]. Furthermore, the increase in survival of these patients has encouraged clinical studies inboth North America (RTOG 1016) and Europe (De-ESCALaTE-HPV) to examine the possibility of reducing the intensity of curative therapies to mitigate both acute and late toxicity[6,7]. Therefore, it is of fundamental importance to identify the presence of HPV.

The AJCC 8th edition recommends using p16^{INK4a} immunohistochemical (IHC) analysis to evaluate HPV status[8,9]. However, it is crucial to stress that although p16^{INK4a} immunohistochemistry is a highly sensitive test, several studies have demonstrated that this technique is only moderately specific since p16^{INK4a} positivity may be associated with cell growth and not necessarily due to HPV infection[10]. A commonly adopted approach to accurately distinguish between HPV positivity and negativity involves a combination of IHC staining for p16^{INK4a} and in situ hybridization (ISH) genotyping for HPV. This dual analysis showed acceptable levels of sensitivity (97%) and specificity (94%)[11].

The future of HPV identification may go further by adopting approaches based on digital technologies and algorithms in pathology. Computational Pathology is the "third revolution in anatomical pathology"[12]. However, as of today, the approved digital models for clinical use are methods created from magnetic resonance (MR) or computed tomography (CT) data. Compared to these medical images, histological slide images contain more information: millions of different cells can be seen in a single histological slide, and their morphology and spatial arrangement provide more information than other medical images.

In addition, it has been successfully demonstrated that computational models can be trained to identify cancer subtypes and molecular features directly from histopathological images stained with hematoxylin and eosin (H&E), bypassing immunohistochemistry[13,14]. We previously described a Machine Learning approach based on cellular features to predict the Ki-67 positivity of oral squamous cell carcinoma cells[14]. Wang et al.[15] demonstrated the effectiveness of a deep learning approach in identifying HPV status in whole slide images of routine HNSCC sections stained with hematoxylin and eosin, achieving an AUROC of 0.9223 ± 0.0397.

Based on these premises, we studied a series of OPSCC classified based on HPV status. Following a computational approach, we analyzed a pool of characteristics extracted from each identified tumor cell. Subsequently, we built a random forest-based classifier to define the HPV status of OPSCCs on hematoxylin/eosin alone by anticipating the results of IHC and molecular analysis. A feature importance analysis approach also allowed us to extract the most relevant features for classification purposes to the advantage of the explainability of our model.

Materials and methods

Study population

In this study, data were obtained from two patient cohorts. The first was built from the publicly available dataset from the Cancer Genome Atlas (TCGA), from which 8 diagnostic H&E OPSCC whole-slide images (4 HPV-negative and 4 HPV-positive) were selected. The second dataset consisted of 27 tissue microarrays (TMA) (14 HPV-negative and 13 HPV-positive) and 57 whole-slide cases (39 HPV-negative and 18 HPV-positive OPSCCs stained with hematoxylin and eosin. These samples were retrieved from the archives of the Pathology Unit of the University "Federico II" of Naples. HPV positivity in these cases was determined through routine IHC p16^{INK4a} staining and INNO-LiPA® HPV Genotyping (Tab. I).

The full lists of cases analyzed in the present study are in Supplementary Tables I, II, and III.

p16^{INK4a} Immunohistochemistry

IHC evaluation of p16^{INK4a} was performed as previous-

Table I. Summary of Study Populations (nd: not detected).

	Year of diagnosis	Age	Sex		HistoQC	p16		INNOLIPA®		Tot.
			Male	Female		Negative	Positive	Negative	Positive	
TCGA cases	2009-2013	40-59	8	1	9/10	4	4	4	4	8
TMA cases	2009-2017	35-80	20	7	27/27	14	13	nd	nd	27
Whole-slide cases	2017-2022	23-80	41	16	57/58	39	18	7	6	57
Total	2009-2022	23-80	69	24	93/95	57	35	11	10	92

ly described[16-18]. Briefly, we used the Ventana Benchmark Ultra platform (Ventana Medical Systems Inc., Tucson, AZ) with the CINtec p16 kit (Roche Ltd AG, Heidelberg, Germany). Four µm tissue sections were deparaffinized and subjected to antigen retrieval using CC1 buffer (Cell Conditioning 1, Ventana Medical Systems) for 30 minutes. They were incubated with the prediluted CINtec p16^{INK4a} primary antibody (clone E6H4) for 20 minutes at room temperature and detected with Ultra View Universal Alkaline Phosphatase Red Detection Kit. Finally, after contrasting with hematoxylin II for 8 minutes and Bluing reagent for 4 minutes, sections were coverslipped through a synthetic medium (Entellan; Merck, Darmstadt, Germany). The positive control was a section of tonsillar squamous cell carcinoma with high p16^{INK4a} expression. The positivity index for p16^{INK4a} occurred through a binary evaluation such as "positive" or "negative". The test was scored positive if strong, homogeneous, and diffuse nuclear staining was present in more than 75% of the malignant cells. On the contrary, a negative evaluation was assigned when found to be non-continuous (especially not of basal and para-basal cells) or if exclusively cytoplasmic.

INNO-LiPA®

Of the 57 cases selected for the project, 14 had been subjected to the INNO-LiPA assay for genotyping. The test execution involved DNA extraction using the QIAamp DSP DNA FFPE Tissue kit, which uses silica membrane technology (QIAamp technology) to isolate and purify genomic DNA from formalin-fixed paraffin-embedded samples. We performed DNA extraction following the manufacturer's instructions.

Following DNA extraction and quantitation, we performed the INNO-LiPA HPV Genotyping test as previously described[16]. Briefly, a 65 bp fragment of the L1 region of the HPV genome was amplified using SPF/PCR. After 40 cycles of PCR, an amplified biotinylated sequence was obtained, which was analyzed using the Auto-LiPA machine. The biotinylated amplicons were hybridized with specific oligonucleotide probes immobilized on strips. Subsequently, alkaline phosphatase-conjugated with streptavidin was added, which bound all the biotinylated hybrids previously formed. The reaction was incubated with the chromogen BCIP/NBT (5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium), producing a purple precipitate, allowing visual interpretation of the result. This multi-parametric method consists of a single strip that simultaneously detects 32 HPV genotypes. Furthermore, the strip is equipped with reaction and hybridization control bands for both human and viral DNA, as well as 19 of the UNG (uracil N-glycosylase) system, which allows reduction of the number of false positives due to DNA contamination already amplified.

Workflow

The analysis process for training and classification is schematically represented in Figure 1. It includes several crucial phases, such as preprocessing, quality control, color normalization, object detec-

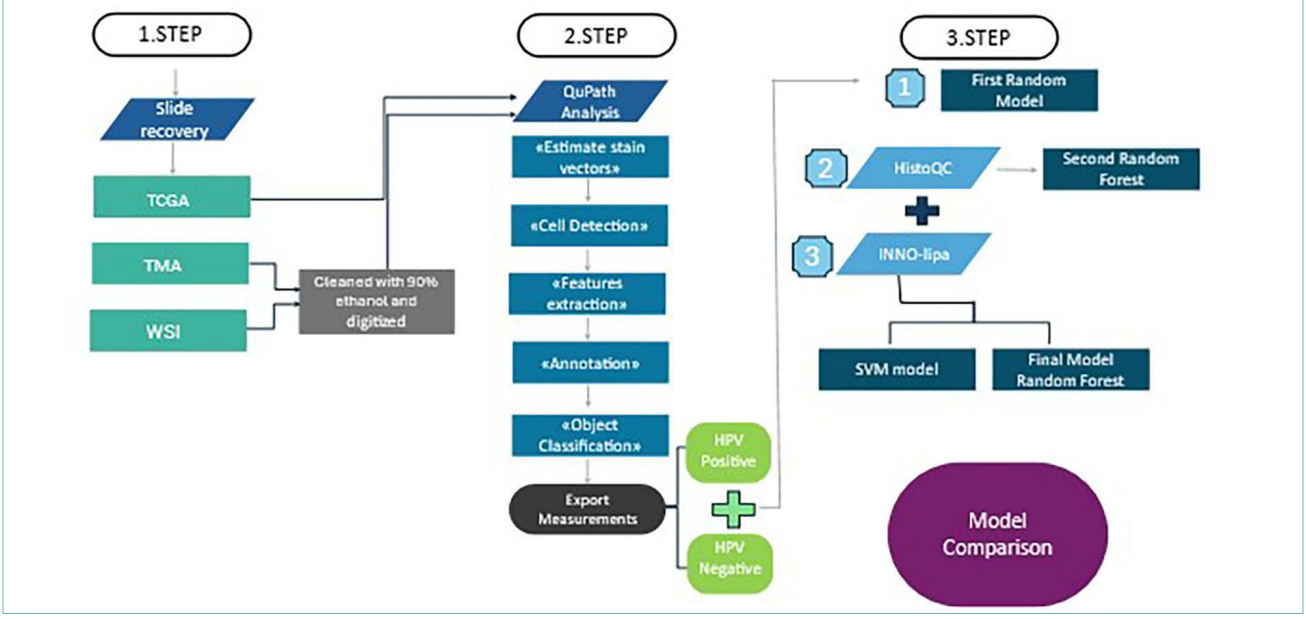

Figure 1. Workflow.

tion, and feature extraction, followed by the analysis of the extracted data. We conducted a thorough assessment for quality control to identify potential artifacts that could compromise or negatively impact the learning algorithms. We also performed color correction, which is essential to ensure that the method can generalize across inputs with different characteristics resulting from scanner protocols and staining variations. 41 morphometric and colorimetric features were extracted from each detected object (i.e., tumor cell) following manual segmentation of stromal and tumor areas (described in Tab. IV in Supplementary). The extracted values were analyzed using Random Forest and Support Vector Machines algorithms.

Quality Control (QC)

We conducted a rigorous quality control (QC) process on the Whole Slide Images (WSI) using the open-source software HistoQC[19]. This process was essential to identify possible artifacts or issues in the images, which could significantly impact classifier validation and machine learning-based analysis.

Through applying QC, we ensured that our model was applied only to regions of the images considered valid, ignoring those parts that, although they might have appeared intact at first glance, concealed imperfections that could compromise the analysis. This approach allowed us to identify and remove samples with artifacts while constructing subsequent classifiers.

The software-generated masks, where areas highlighted in pink were considered suitable for analysis, while those in green were excluded. It is important to emphasize that the images were analyzed without quality control to compare the classifier's behavior before and after applying QC (Fig. 2).

Normalization and feature extraction

The histological slides, stained with hematoxylin and eosin, were scanned using a Leica Aperio scanner (Leica Biosystems Nussloch GmbH) at 20x magnification. Before digitization, the glass slides were cleaned with 90% ethanol to prevent foreign elements from interfering with the analysis. The WSI obtained were analyzed using the open-source software QuPath[20], allowing us to normalize staining vectors, identify tumor cells, and extract features.

Firstly, we normalized the staining vectors by applying the "Estimate stain vectors" function, and then we ran a Cell Detection task on manually annotated tumor-containing ROIs (Fig. 3).

Groovy scripts available on the QuPath documentation webpage ["https://qupath.readthedocs.io/en/stable/", last accessed: 10/11/2023] were utilized and adapted for our purposes. We provided the complete script employed in Supplementary II.

Following a feature extraction step, we classified the detected objects with the "HPV positive" and "HPV

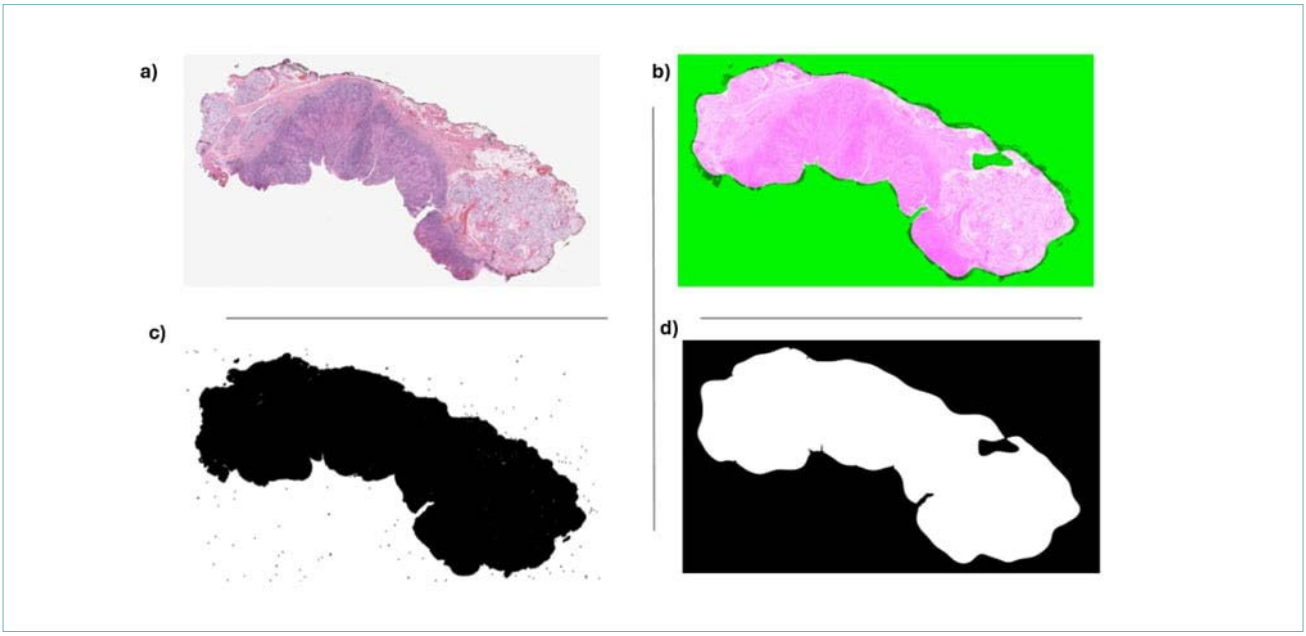

Figure 2. Qualitative control with HistoQC. (a) Original image; (b)(c)(d) HistoQC results mask; (b) The regions colored in pink are the portions of tissue found suitable for the analysis, and in green, those rejected. (c) The white mask segments the unusable area. (d) The white mask segments the usable part of the tissue.

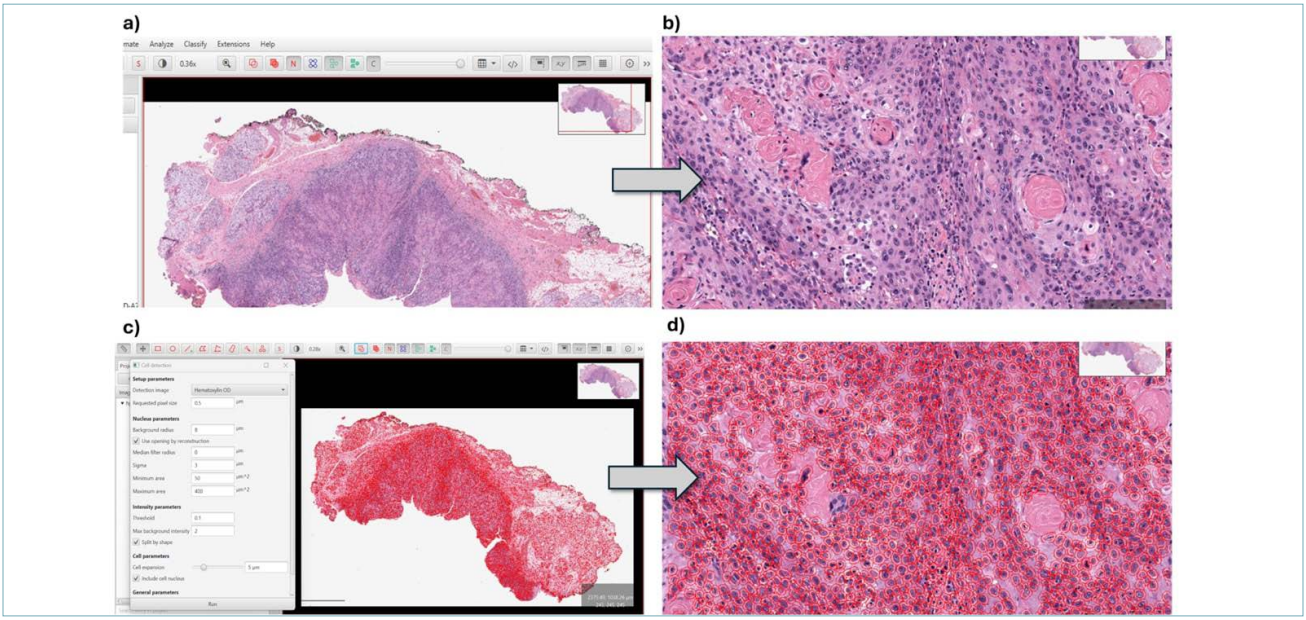

Figure 3. Cell detection with QuPath. (a) Example of the original image. (b) Following the run of the cell detection QuPath script, the software displays the detected cells with red outlines. (Higher magnification of the same region in (b) and (d)).

negative" labels, depending on $p16^{INK4a}$ IHC and genotyping analysis.

We examined the comprehensive feature datasets for the two classes to create a Random Forest classification model to predict HPV status.

Model Construction and Validation

Random Forest models

We utilized the features extracted from tissue samples to train various Random Forest models[21,22]. Each model consisted of a set of 500 classification trees, and to enhance the stability and accuracy of each model we employed bootstrap aggregation.

Bootstrap aggregation is a machine-learning strategy combining multiple versions of decision trees into a single random forest model. Each decision tree version was created from a random data sampling with replacement.

The dataset was randomly divided into a training set (80%) and a test set (20%). The Random Forest models were then trained on the training set. The performance of each model was assessed using the test set, which did not contain class labels.

We used metrics such as Accuracy, Precision, F1 score, and the percentage of correctly classified samples to evaluate and compare the models through a Confusion matrix.

The results of these assessments are reported in Table II, from which we selected the best-performing model.

We experimented with using and modifying the script provided in the sklearn library. [https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html, last accessed: 10/01/2024]. For additional details, including the adapted script and an expanded explanation of the methodology, please refer to Supplementary III.

Table II. Comparison of the results of Random Forest models.

Random Forest	Model-1	Model-2	Model-3
Sensitivity	0.83	0.83	0.893
Specificity	0.896	0.897	0.903
Accuracy	0.863	0.865	0.900
Precision:			
Negative	0.84	0.84	0.90
Positive	0.89	0.89	0.86

Support Vector Machine model

Furthermore, we used the features extracted from the tissue samples to create a support vector machine (SVM) model. The SVM algorithm works mainly on finding an optimal hyperplane that effectively separates the classes in the dataset, maximizing the distance between the closest samples (support vectors) to the plane. This hyperplane is known as the decision limit or optimal limit[23].

However, to handle the complexity of the data, we applied Principal Component Analysis (PCA) before using the SVM algorithm.
We set the PCA to retain 95% of the variance, which means that we are preserving 95% of the information contained in the original data. In the end, we obtained 22 components. As with the Random Forest, we started from pre-existing scripts provided in the Python sklearn library [https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html and last consulted: 10/01/2024] adequately customized to our specific needs (Supplementary V).
Next, we created the SVM model using the [scikit-learn library.https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html, last consulted: 10/01/2024]. We split the dataset randomly into a training set (85%) and a test set (15%). Employing the GridSearchCV function, we searched for the optimal combination of hyperparameters for the SVM model to maximize the model's accuracy. To ensure the reliability and robustness of our model evaluation, we employed 5-fold cross-validation. In this technique, the dataset is partitioned into 5 equal subsets, or folds, iteratively training the model on four folds while validating on the remaining fold. By combining the results from these iterations, we obtained a comprehensive assessment of the model's performance across different subsets of the data. The best model obtained from the optimization process uses the radial basis function (RFB) kernel.
Also, in this case, to evaluate the performance of the SVM model, we used metrics such as accuracy, precision, recall, F1 score, and the percentage of samples correctly classified through a Confusion matrix.
Further details on how we configured and used the SVM model with PCA can be found in Supplementary file V.

Results

Random Forest Models

Model-1: Visual QC

We developed Random Forest classification models using morphometric and colorimetric data extracted from detected tumor cells.
Model-1 was built by applying quality control based solely on visual analysis. Furthermore, the model was trained using HPV test-positive samples tested solely through IHC. Our classifier was created using 689,380 observations, of which 448,648 were positive for HPV and 240,732 were negative. Since the dataset exhibited class imbalance, we initially balanced the data. We subsequently randomly split the dataset into a training and a test set, with a 80/20 ratio (80% train, and 20% test).
This classifier achieved an accuracy of 86.3%, with an error rate of 0.138. However, when examining other validation parameters, as shown in Table III, this model's Precision, Recall, and F1-Score did not yield an optimal result. The confusion matrix in Figure 4 summarizes the result: 45,262 True Negatives, 41,449 True Positives, 5,203 False Positives, and 8,480 False Negatives. The number of False Negatives (FN) and False Positives (FP) is relatively high, indicating that the model does not correctly predict HPV status.

Table III. Results of Model-1 with quality control based solely on visual analysis.

First RF-Model	Precision	Recall	F1-score	Support
Negative	0.84	0.90	0.87	50470
Positive	0.89	0.83	0.86	49979
macro avg	0.87	0.86	0.86	100449
weighted avg	0.87	0.86	0.86	100449
Accuracy	**0.8637816205**			
OOB error	**0.138**			

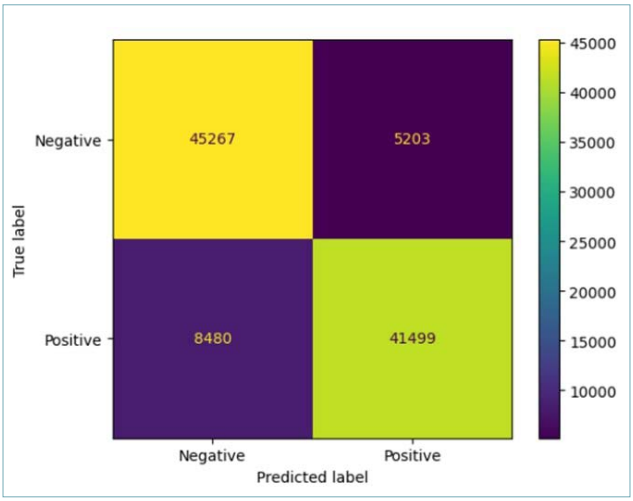

Figure 4. Confusion Matrix of Model-1.

Model-2: HistoQC as Quality Control Strategy

Therefore, we constructed a second model following a strict QC step with HistoQC. Comparing the two sets of results (Tabs. IV-V), Model-2 showed a slight improvement in accuracy compared to the first model, increasing from 86.3% to 86.5%. From the data obtained from the Confusion Matrix (Fig. 5), the second

model slightly decreases the number of False Positives compared to the first model (4997 compared to 5203), which is a positive outcome. Additionally, the number of false negatives is reduced compared to Model-1 (7995 compared to 8480 of the first model).

Our results confirmed the need for double testing in detecting HPV status, since only this way could we significantly separate the two classes based on HPV status.

These results highlight the critical importance of quality control and iteration in developing a machine-learning model, especially when working with data from biological samples such as tissue slides.

Model-3: Strict HPV Status Assessment with Combined p16^{INK4a} IHC and Genotyping Tests

We finally questioned whether the lack of consideration for the difference in performance between individual IHC tests and the combined use of p16^{INK4a} IHC and genotyping tests in detecting HPV positivity could have an impact. As a result, we built a third model, classifying HPV-positive cases based on the double test, IHC and molecular, from both our case sources (integrating cases from both the TCGA and the archive that underwent both tests). We examined 1,048,079 detected tumor cells, split into 583,390 negative and 464,689 positives.

We achieved an accuracy of 90%, significantly im-

Table IV. Evaluation of Model-2 with HistoQC in Quality Control process.

Second RF-Model	Precision	Recall	F1-score	Support
Negative	0.84	0.90	0.87	48366
Positive	0.89	0.83	0.86	47927
macro avg	0.87	0.86	0.86	96293
weighted avg	0.87	0.87	0.86	96293
Accuracy	**0.8650784584**			
OOB error	**0.135**			

Table V. Model-3. Optimizing HPV Detection with Combined IHC and Genotyping Tests.

Final RF-Model	Precision	Recall	F1-score	Support
Negative	0.89	0.91	0.90	95320
Positive	0.91	0.89	0.90	95242
macro avg	0.90	0.90	0.90	190562
weighted avg	0.90	0.90	0.90	190562
Accuracy	**0.9006202705**			
OOB error	**0.099**			

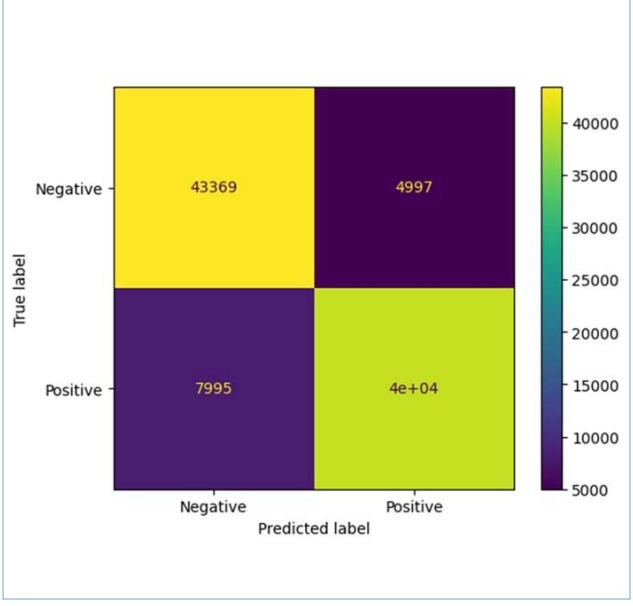

Figure 5. Confusion matrix of Model-2.

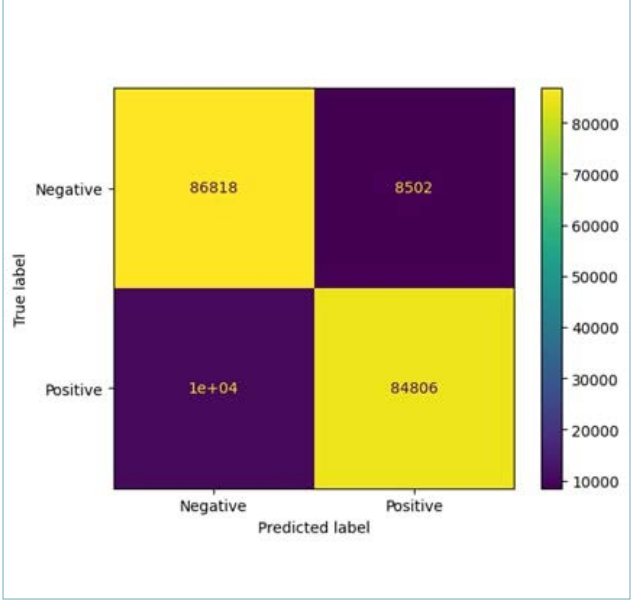

Figure 6. Confusion Matrix of Model-3.

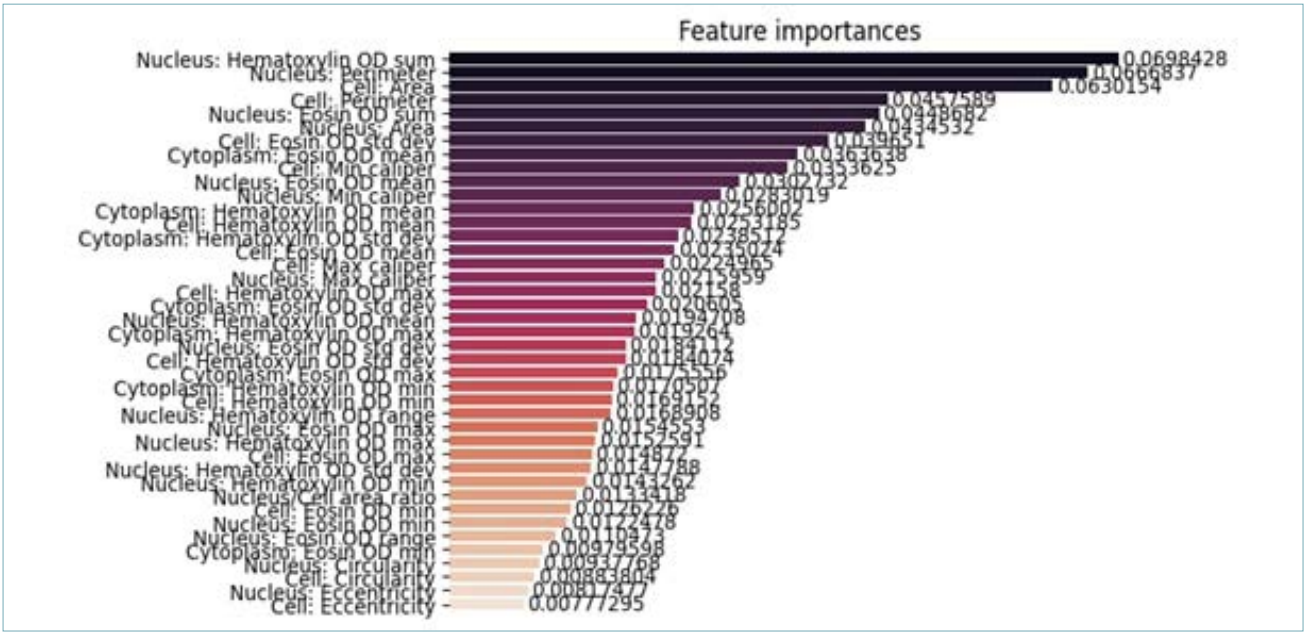

Figure 7. Variable importance determined by random forests classifier.

proving over the previous ones with the latter model. Furthermore, the error decreased to 9.9%, indicating that the model had become more robust. When examining the validation parameters, as shown in Table IV, the Precision, Recall, and F1-Score metrics are significantly better than in the previous models. The Confusion Matrix confirms this observation (Fig. 6): True Positives (84,806) and True Negatives (86,818) have increased, but we did not get a reduction in False Negatives (10,436) and False Positives (8502).

Furthermore, using Random Forest allowed us to generate a variable importance plot that analyzes the relative importance of different features in the model creation process. In particular, our observation that the average sum optical density of nuclear hematoxylin (with a weight of 6.9%) and the nuclear perimeter (with a weight of 6.6%) have a significant influence on distinguishing between the two classes suggests that these features are strongly correlated with HPV positivity.

Result SVM Model

The best SVM model was obtained using the GridSearchCV function to optimize the model parameters and achieved an overall accuracy of 88%.

Examining the validation parameters, shown in Table VI, we observed that for the "Negative" class, the model has an accuracy of 90%, a Recall of 85%, and an F1 score of 87%, while for the "Positive" class, the accuracy is 86%, Recall is 90%, and F1 score is 88%. Examining the validation parameters in Table V, we noticed that compared to the "Negative" class, the "Positive" class has slightly lower accuracy but higher Recall and F1 score. This suggests that the SVM model has a greater ability to correctly identify positives than negatives, although it sometimes makes errors.

In addition to the above observations, when analyzing the Confusion matrix (Fig. 8), we found that the model exhibited strong performance in correctly classifying a large number of true negatives (60,910) and true positives (64,544). However, it is important to note that there are still a considerable number of false positives (10,662) and false negatives (6808). In the SVM, we noticed a higher number of false positives, while we observed a higher number of false negatives in the Random Forest.

Table VI. Model-SVM.

SVM-Model	Precision	Recall	F1-score	Support
Negative	0.90	0.85	0.87	71572
Positive	0.86	0.90	0.88	71350
macro avg	0.88	0.88	0.88	142922
weighted avg	0.88	0.88	0.88	142922
Accuracy	**0.877779488**			
OOB error	**0.132**			

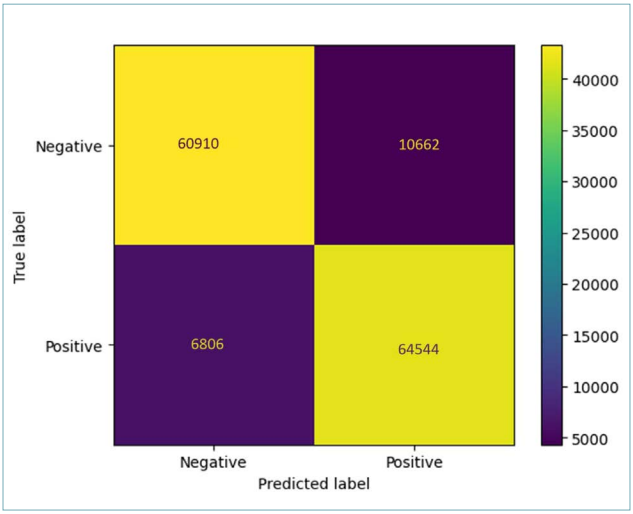

Figure 8. Confusion Matrix of Model-SVM.

Discussion

The prevalence of HPV-positive OPSCC varies across Europe, with the highest rates found in Nordic countries[24]. The distinction between HPV-positive and HPV-negative OPSCC is crucial for treatment planning, such as treatment de-escalation[25]. There is an urgent need for a validated and reproducible testing strategy[11,26,27].

Different testing methods have been proposed, with a range of sensitivity, specificity, and predictive values; the combination of $p16^{INK4a}$ IHC and DNA qPCR shows promising results[11,28]. These studies underscore the importance of accurate and reliable HPV status assessment in OPSCC.

The low cost and ease of use of IHC on formalin-fixed paraffin-embedded tissue have made $p16^{INK4a}$ IHC the most widely used test for the detection of HPV, compared to ISH and PCR-based tests[29].

As a molecular test, the INNO-LiPA® system, based on SPF10 PCR amplification and LiPA hybridization assays, identifies 32 HPV genotypes. In particular, the genotypes identified are 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68 (high risk); 26, 53, 66,70, 73, 82 (probable high risk); 6, 11, 40, 42, 43, 44, 54, 61, 62, 67, 81, 83, 89 (low risk).

Digital pathology, particularly whole slide scanning and image analysis, has revolutionized tissue biomarker research, accelerating biomarker discovery and the development of companion diagnostics[30]. This technology has been further enhanced using tissue microarray technology for high-throughput analysis of tissue-based biomarkers[31]. The shift toward quantitative assessment of tissue biomarkers has made applying quantitative image analysis a crucial tool in this field[32]. Integrating digital image analysis data with other data types, such as clinical and genomic, is also highlighted as a key aspect of this research[33].

Currently, most standard surgical pathology practices rely on histopathologically examining tiny samples. Smaller and smaller samples are being analyzed due to the pressure for earlier diagnosis and the need to lessen the invasiveness of diagnostic histopathology sampling procedures. Frequently, biospecimen consumption is decreased to save them for later special staining or molecular biology examination, which is required. With the help of QuPath, an open-source histopathology image analysis program, and hematoxylin and eosin-stained histopathology glass slides, our study shows that predicting HPV status in oropharyngeal squamous carcinoma is feasible. This allows to gather data regarding this important aspect of OPSCCs on standard basic staining (i.e., H&E).

Our machine-learning method concentrated on cellular characteristics that could differentiate HPV-positive OPSCC tumor cells from HPV-negative ones.

From our analysis, we applied two machine learning methods, building models using Random Forest and SVM, achieving acceptable margins of error.

Also, with Random Forest, we obtained information on the contribution of each variable in determining the probability of a tested sample falling into one class rather than the other. The variable importance is extremely interesting as it helps explain the model. Based on our analysis, it turned out that colorimetric features are crucial in classifying OPSCC HPV status. When using the model on other source datasets, our results make an accurate color normalization step mandatory.

Furthermore, the results confirmed the importance of distinguishing the HPV status using the double test (IHC and molecular), as only in this way could the two classes (HPV-positive and HPV-negative) be separated more significantly.

Although our results are promising, our work still has some limitations. A robust approach based on Deep Learning would have allowed us to build a more versatile classifier with probably wider implementation margins. Notwithstanding, with this work's "handcrafted" approach, we intended to thoroughly study the contribution of each single attribute in differential image analysis between the two classes of OPSCC. We enhanced the heterogeneity of our datasets, taking samples from different sources. We planned to increase variability further by involving other institutions in a multicentric work (in preparation).

Using QuPath as an analysis platform and, subse-

quently, model implementation allows us to bring the results of our work to a more accessible level close to the experience of the pathologist engaged in diagnostic routine.

Conclusions

Establishing HPV status of OPSCC is crucial to establishing correct therapy. HPV status cannot be defined based on the morphological observation of the histological preparations alone. Therefore, we require the IHC determination of the expression of the $p16^{INK4a}$ protein in addition to a molecular test to increase the specificity and sensitivity of the test. Taking advantage of digital pathology workflow, and applying machine learning techniques, we developed two types of classifier models: a Random Forest and an SVM. Both were trained to predict HPV status on H&E stained histopathology tissue slides.
Our models, based on Random Forest and SVM, show a margin of error of approximately 10% in predicting the HPV status of OPSCCs.

Conflicts of interest statement

The authors declare no conflicts of interest.

Funding

Rare cancers of the head and neck: a comprehensive approach combining genomic,immunophenotypic and computational aspects to improve patient prognosis and establish innovative preclinical models – RENASCENCE "(project codePNRR-TR1-2023-12377661)"

Authors' contributions

Conceptualization, F.M.; methodology, S.V., and G.I.; software, F.M.; validation, D.R., R.M.D.C.; formal analysis, F.M., and A.C.; investigation, S.V.; resources, G.I.; data curation, A.C.; writing---original draft preparation, F.M.; writing---review and editing, F.M. and A.C.; funding, S.S.; supervision, F.M. All authors reviewed the manuscript.

Ethical consideration

The study was performed according to the Declaration of Helsinki and in agreement with Italian law for studies based only on retrospective analyses on routine archival FFPE-tissue; a written informed consent from the living patient, following the indication of Italian DLgs No. 196/03 (Codex on Privacy), as modified by UE 2016/679 law of the European Parliament and Commission, was obtained at the time of surgery.

List of Abbreviations

CPATH: Computational Pathology
DL: Deep Learning
DP: Digital Pathology
H&E: Hematoxylin and Eosin
HPV: Human Papilloma Virus
IHC: Immunohistochemistry
ISH: In Situ Hybridization
ML: Machine Learning
OPSCC: Oropharyngeal Squamous Cell Carcinoma
PCA: Principal Component Analysis
QC: Quality Control
RF: Random Forest
RBF: Radial basis function
ROI: Region Of Interest
SVM: Support vector machine
TCGA: Tumor Cancer Genome Atlas
TMA: Tissue Micro Array
UICC/AJCC: Union for International Cancer Control/ American Joint Committee on Cancer
WSI: Whole Slide Image

References

1 De Flora S, La Maestra S. Epidemiology of cancers of infectious origin and prevention strategies. J Prev Med Hyg. 2015;56(1):E15-E20.

2 AIOM. LINEE GUIDA TUMORI DELLA TESTA E DEL COLLO. AIOM. Published December 31, 2021. Accessed January 22, 2024. https://www.aiom.it/linee-guida-aiom-2021-tumori-della-testa-e-del-collo/

3 Lechner M, Liu J, Masterson L, Fenton TR. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. Nat Rev Clin Oncol. 2022;19(5):306-327. https://doi.org/10.1038/s41571-022-00603-7

4 Ang KK, Harris J, Wheeler R, 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

5 Craig SG, Anderson LA, Schache AG, et al. Recommendations for determining HPV status in patients with oropharyngeal cancers under TNM8 guidelines: a two-tier approach. Br J Cancer. 2019;120(8):827-833. https://doi.org/10.1038/s41416-019-0414-9

6 Mehanna H, Rischin D, Wong SJ, et al. De-Escalation After DE-ESCALATE and RTOG 1016: A Head and Neck Cancer InterGroup Framework for Future De-Escalation Studies. J Clin Oncol. 2020;38(22):2552-2557. https://doi.org/10.1200/JCO.20.00056

7 Schache AG, Liloglou T, Risk JM, et al. Validation of a novel diagnostic standard in HPV-positive oropharyngeal squamous cell carcinoma. Br J Cancer. 2013;108(6):1332-1339. https://doi.org/10.1038/bjc.2013.63

8 Lewis JS, Beadle B, Bishop JA, et al. Human Papillomavirus Testing in Head and Neck Carcinomas: Guideline From the College of American Pathologists. Arch Pathol Lab Med. 2018;142(5):559-597. https://doi.org/10.5858/arpa.2017-0286-CP

9 Machczyński P, Majchrzak E, Niewinski P, et al. A review of the 8th edition of the AJCC staging system for oropharyngeal cancer according to HPV status. Eur Arch Otorhinolaryngol. 2020;277(9):2407-2412. https://doi.org/10.1007/s00405-020-05979-9

10 Prigge ES, Arbyn M, von Knebel Doeberitz M, et al. Diagnostic accuracy of p16INK4a immunohistochemistry in oropharyngeal

squamous cell carcinomas: A systematic review and meta-analysis. Int J Cancer. 2017;140(5):1186-1198. https://doi.org/10.1002/ijc.30516

[11] Schache AG, Liloglou T, Risk JM, et al. Evaluation of human papilloma virus diagnostic testing in oropharyngeal squamous cell carcinoma: sensitivity, specificity and prognostic discrimination. Clin Cancer Res Off J Am Assoc Cancer Res. 2011;17(19):6262-6271. https://doi.org/10.1158/1078-0432.CCR-11-0388

[12] Salto-Tellez M, Maxwell P, Hamilton P. Artificial intelligence—the third revolution in pathology. Histopathology. 2019;74(3):372-376. https://doi.org/10.1111/his.13760

[13] Echle A, Rindtorff NT, Brinker TJ, et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer. 2021;124(4):686-696. https://doi.org/10.1038/s41416-020-01122-x

[14] Martino F, Varricchio S, Russo D, et al. A Machine-learning Approach for the Assessment of the Proliferative Compartment of Solid Tumors on Hematoxylin-Eosin-Stained Sections. Cancers. 2020;12(5):1344. https://doi.org/10.3390/cancers12051344

[15] Wang R, Khurram SA, Walsh H, et al. A Novel Deep Learning Algorithm for Human Papillomavirus Infection Prediction in Head and Neck Cancers Using Routine Histology Images. Mod Pathol. 2023;36(12). https://doi.org/10.1016/j.modpat.2023.100320

[16] Ilardi G, Russo D, Varricchio S, et al. HPV Virus Transcriptional Status Assessment in a Case of Sinonasal Carcinoma. Int J Mol Sci. 2018;19(3):883. https://doi.org/10.3390/ijms19030883

[17] Russo D, Merolla F, Mascolo M, et al. FKBP51 immunohistochemical expression: A new prognostic biomarker for OSCC? Int J Mol Sci. 2017;18(2). https://doi.org/10.3390/ijms18020443

[18] Russo D, Di Crescenzo RM, Broggi G, et al. Expression of P16INK4a in Uveal Melanoma: New Perspectives. Front Oncol. 2020;10:562074. https://doi.org/10.3389/fonc.2020.562074

[19] Janowczyk A, Zuo R, Gilmore H, et al. HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. JCO Clin Cancer Inform. 2019;3:CCI.18.00157. https://doi.org/10.1200/CCI.18.00157

[20] Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. 2017;7:16878. https://doi.org/10.1038/s41598-017-17204-5

[21] AlSagri H, Ykhlef M. Quantifying Feature Importance for Detecting Depression using Random Forest. Int J Adv Comput Sci Appl. 2020;11(5). https://doi.org/10.14569/IJACSA.2020.0110577

[22] Hastie T, Friedman J, Tibshirani R. Boosting and Additive Trees. In: Hastie T, Friedman J, Tibshirani R, eds. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer; 2001:587-604. https://doi.org/10.1007/978-0-387-21606-5_10

[23] Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow - Aurélien Géron - Google Libri. SECOND EDITION. O'Reilly Media, Inc.; 2022. Accessed January 31, 2024. https://books.google.it/books?hl=it&lr=&id=X5ySEAAAQBAJ&oi=fnd&pg=PT10&dq=related:zLachf9ZOJ4J:scholar.google.com/&ots=yCZwvl40uM&sig=gZixMHaHCdkhIhh_O299rblOqgk&redir_esc=y#v=onepage&q&f=false

[24] Stjernstrøm KD, Jensen JS, Jakobsen KK, et al. Current status of human papillomavirus positivity in oropharyngeal squamous cell carcinoma in Europe: a systematic review. Acta Otolaryngol (Stockh). 2019;139(12):1112-1116. https://doi.org/10.1080/00016489.2019.1669820

[25] Boscolo-Rizzo P, Pawlita M, Holzinger D. From HPV-positive towards HPV-driven oropharyngeal squamous cell carcinomas. Cancer Treat Rev. 2016;42:24-29. https://doi.org/10.1016/j.ctrv.2015.10.009

[26] Moutasim KA, Robinson M, Thavaraj S. Human papillomavirus testing in diagnostic head and neck histopathology. Diagn Histopathol. 2015;21(2):77-84. https://doi.org/10.1016/j.mpdhp.2015.02.002

[27] Robinson M, Schache A, Sloan P, et al. HPV Specific Testing: A Requirement for Oropharyngeal Squamous Cell Carcinoma Patients. Head Neck Pathol. 2012;6(S1):83-90. https://doi.org/10.1007/s12105-012-0370-7

[28] Melchers LJ, Mastik MF, Samaniego Cameron B, et al. Detection of HPV-associated oropharyngeal tumours in a 16-year cohort: more than meets the eye. Br J Cancer. 2015;112(8):1349-1357. https://doi.org/10.1038/bjc.2015.99

[29] Fauzi FH, Hamzan NI, Rahman NA, et al. Detection of human papillomavirus in oropharyngeal squamous cell carcinoma. J Zhejiang Univ Sci B. 2020;21(12):961-976. https://doi.org/10.1631/jzus.B2000161

[30] Hamilton P, O'Reilly P, Bankhead P, et al. Digital and Computational Pathology for Biomarker Discovery. In: Badve S, Kumar GL, eds. Predictive Biomarkers in Oncology. Springer International Publishing; 2019:87-105. https://doi.org/10.1007/978-3-319-95228-4_7

[31] Van Eycke YR, Debeir O, Verset L, et al. High-throughput analysis of tissue-based biomarkers in digital pathology. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE; 2015:7732-7735. https://doi.org/10.1109/EMBC.2015.7320184

[32] Lara H, Li Z, Abels E, et al. Quantitative Image Analysis for Tissue Biomarker Use: A White Paper From the Digital Pathology Association. Appl Immunohistochem Mol Morphol. 2021;29(7):479-493. https://doi.org/10.1097/PAI.0000000000000930

[33] Hamilton PW, Bankhead P, Wang Y, et al. Digital pathology and image analysis in tissue biomarker research. Methods. 2014;70(1):59-73. https://doi.org/10.1016/j.ymeth.2014.06.015